



UNIONE EUROPEA
Fondo Sociale Europeo



UNIVERSITÀ
DEGLI STUDI
FIRENZE
Da un secolo, oltre.

DOTTORATO DI RICERCA IN BIOLOGIA EVOLUZIONISTICA ED ECOLOGIA

CICLO XXXVII - PON ex DM1061

Monitoring and modeling of microbial activity in wastewater treatment plants and identification of potential biotechnological targets

*Monitoraggio e modellizzazione dell'attività microbica in impianti per il trattamento delle acque reflue e
identificazione di potenziali target per l'ambito biotecnologico*

Settore Scientifico Disciplinare
BIOS/14A

Candidate:

Tania ALONSO-VÁSQUEZ

(firma)

Supervisor:

Marco FONDI, PhD

(firma)

Coordinatore:

Prof. Duccio CAVALIERI

(firma)

ANNI 2021/2024

Salvo eventuali più ampie autorizzazioni dell'autore, la tesi può essere liberamente consultata e può essere effettuato il salvataggio e la stampa di una copia per fini strettamente personali di studio, di ricerca e di insegnamento, con espresso divieto di qualunque utilizzo direttamente o indirettamente commerciale.

Ogni altro diritto sul materiale è riservato.

This thesis was made in collaboration with:

Gestione Impianti Depurazione Acque S.p.A.

Donatella FIBBI

Ester COPPINI



Vrije Universiteit Amsterdam

Prof. Matti GRALKA



Tania Alonso-Vásquez

*Monitoring and modeling of microbial activity
in wastewater treatment plants and identification
of potential biotechnological targets*

PhD dissertation, May 2025

Supervisor: Marco Fondi, PhD

University of Florence

Department of Biology

Via Madonna del Piano, 6

Sesto Fiorentino, 50019, Florence, Italy

Abstract

Microbial communities are complex ecosystems where organisms shape their environment through constant interaction. These interactions can range from friendly collaborations to fierce competition, and they form the basis of the community's structure, function and their surroundings.

Dynamic modeling is a powerful mathematical tool to describe these microbial interactions and to infer directionality and causality, as well as to predict the time-dependent properties of a system. In microbiology, it captures the dynamic behaviors of a microbiome in fluctuating environments and more than one model could be plausible for the system of interest, depending on the questions, the assumptions made and the available data.

There are several environments where microbial communities thrive, being wastewater treatment plants (WWTPs) among the most important. These are an essential part of modern society since they are designed to manage and treat wastewater coming from households, industries, and agriculture. Their main scope is to remove contaminants, including organic matter, harmful microorganisms, and pollutants such as metals or antimicrobials, so the water becomes safe for the environment or reuse.

In this thesis we aimed to describe and model the dynamics of the microbial communities living in a wastewater treatment plant. These could lead to the development of innovative and potentially economically advantageous biotechnological processes for wastewater treatment. To this aim, first, we conducted a bi-monthly sampling over two years at G.I.D.A. S.p.A. (Gestione Impianto Depu-

razione Acque), a WWTP located in Prato, Italy. We focused on the microbial compositions and function characterization, the different tanks' dynamic behaviors, ecological interactions, and potential biotechnological applications. And second, we conducted a comprehensive study of different mathematical models.

This thesis highlights two valuable points: *i)* the essential role of microbial communities in wastewater treatment and their broader environmental significance, and *ii)* the use of mathematical modeling to gain understanding of biological scenarios. By bridging experimental and computational approaches, it provides a robust framework for understanding and exploiting microbial dynamics. As global challenges related to water management, pollution, and antimicrobial resistance intensify, the insights gained from this work will play a crucial role in driving sustainable and innovative solutions in microbial ecology and biotechnology.

Contents

1	Introduction	11
1.1	Microbial communities	11
1.1.1	Microbial interactions	12
1.2	Dynamic community modeling	16
1.2.1	Generalized Lotka–Volterra	16
1.2.2	MacArthur consumer–resource model	20
1.2.3	Trait-based model	23
1.2.4	Individual-based model	26
1.2.5	Genome-scale Metabolic models	29
1.2.6	Deterministic or Stochastic models?	32
1.3	Wastewater treatment plants	34
2	Overview and aims of the work	39
2.1	The project	39
2.2	Aims	43
3	Results	45
3.1	NGS: from samples to microbiomes	45
3.1.1	GIDA: from 16S to microbial community functions	46
3.1.2	Metagenomes: How and who has an impact on marine protected areas?	68
3.1.3	Bacterial genomes: <i>Origanum vulgare</i> L.	104
3.2	Modeling microbial dynamics and metabolism	116

3.2.1	GSMMs and antimicrobial resistance	116
3.2.2	GSMM of cold adapted bacteria	137
3.2.3	How to study COVID impacts with GSMM	155
3.2.4	Dynamics of persister resuscitation	184
4	Conclusions and perspectives	195
4.1	Mathematical modeling	195
4.2	Understanding wastewater treatment plants	196
4.3	Future perspectives	198
	Bibliography	199
5	Overall activities	213
5.1	List of publications	213
5.2	Period of study and research abroad	214
5.3	Period of study and research in an enterprise	214
5.4	Congresses, seminars and workshops	214
5.5	Awards	216

Chapter 1

Introduction

1.1 Microbial communities

Microorganisms are an essential part of Earth's ecosystems. They can be found in every aquatic or terrestrial environment and fulfill important roles in biogeochemical cycles or in the complex process of intestinal carcinogenesis, for example (Falkowski et al. [2008](#); De Filippo et al. [2024](#)).

Microbial life shows an incredible ability to adapt, allowing it to thrive in a wide range of environments, even the most extreme ones. From the ocean depths to dry deserts, microorganisms find ways to survive and fill important ecological roles that help keeping ecosystems balanced and healthy. These characteristics also determine the composition of the bacterial community and the interactions they may present (Gupta et al. [2017](#)). This adaptability not only reflects how important they are to nature but also shows their potential use in fields like biotechnology and medicine. By studying how these organisms manage to live and grow under such different conditions, we can better understand the natural world and possibly discover new ways to solve environmental and industrial problems.

The environments where microorganisms thrive are as diverse as the adaptations they have developed. Such environments could range from extreme temperatures to animal or plant hosts. One example could be the psychrophilic

microorganisms, extremophiles capable of growing in low temperatures. These cold-adapted microorganisms are responsible for most of the biomass of extremely cold environments and play a key role in maintaining proper ecosystem functions (Fondi et al. 2015; Wilkins et al. 2013). *Pseudoalteromonas haloplanktis* TAC125, an Antarctic bacterial strain, is one of the most studied cold-adapted organisms. Investigating its metabolism can give fundamental insights into life in extreme environments and could play a crucial role for biological applications (Qi et al. 2019).

Clostridioides difficile, on the other hand, is a Gram-positive bacterium and it's typically found in normal human intestinal microbiota (Czepiel et al. 2019). This bacterium is nowadays the most important cause of healthcare-associated diarrhea for several reasons: i) it produces clostridial toxins, ii) it is resistant to a wide range of environments, including antibiotic exposure, and iii) the use of antibiotics favor its colonization and survival when there are low-level or non-resistant bacteria in the community (Czepiel et al. 2019; Markantonis et al. 2024). It is here when the interactions between the members of a microbial community take place and become crucial for treating these kind of infections. To address *C. difficile* infection, fecal microbiota transplantation (FMT) has proven to be an effective treatment (Staley et al. 2017), however to be sure about the characteristics the community should have, i.e. the taxa and the abundances that could optimize the eradication of the pathogen, the interactions between the main community members have to be elucidated (Qian et al. 2021).

1.1.1 Microbial interactions

Microbial communities are intricate ecosystems where organisms shape their environment through constant interactions. These can range from friendly collaborations to fierce competition, and they form the basis of the community's structure, and function, also shaping the surrounding environment.

Mutualism classically refers to any organism in a mutually beneficial relation-

ship with another (Drew et al. 2021). For instance, in the traditional biotechnological process of sourdough fermentation, lactic acid bacteria co-exist synergistically with yeasts. The yeast creates an anaerobic environment by producing carbon dioxide, which benefits the bacteria, while the bacteria produce metabolites that the yeast can utilize (Pérez-Alvarado et al. 2022). Similarly, bacteria known as Rhizobia form symbiotic relationships with legumes, fixing atmospheric nitrogen into a form the plants can use while receiving carbohydrates in return (Vaccaro et al. 2023).

Sometimes, only one partner benefits while the other remains unaffected and this relationship is known as commensalism (Ramanan et al. 2016). This can be seen in biofilms, where early microbial colonizers create a matrix that provides shelter and resources for other species without being directly harmed or helped themselves. Thanks to the physical proximity of the bacteria, they can communicate intimately through the exchange of metabolites and chemicals, developing a remarkable resilience through the shielding effect offered by the biofilm structure (Valiei et al. 2024).

However, interactions between organisms could also be negative: one organism could benefit at the expense of another, as in parasitism. For instance, pathogens like *Mycobacterium tuberculosis*, for its survival, modulate various host pathways, such as mitochondrial metabolism, apoptosis and necrosis, while causing harm to their hosts and ultimately triggering the disease (Daffé et al. 1999; Sengupta et al. 2023).

Competition is another critical dynamic in microbial communities, where organisms compete for scarce resources like nutrients or living space. This competition can be direct, as when bacteria produce antibiotics to kill competitors, like the streptomycin producing bacterium *Streptomyces griseus* (Vardanyan et al. 2006), or indirect, by simply outcompeting others for essential resources. In soil and the human gut, for example, where nutrients can be scarce, competition shapes the community, determining which microbes succeed (Goldford et al. 2018; Gibbs et

al. 2022).

Additionally, bacteria can be involved in social cheating, a form of intra-specific competition (Leinweber et al. 2017). These non-cooperating bacteria benefit from the resources produced by other bacteria, without share the costs of production. For instance, some bacteria produce antibiotics that tend to be effective against other species, and cheaters could benefit of it without being affected by the antibiotics themselves (Smith et al. 2019). The main problem is that cheaters have lower fitness costs, allowing them to outcompete other bacteria (Xenophontos et al. 2022).

Even if predation is something that we normally associate with macrofauna, in some cases, microbes engage in it. This is where one organism consumes another. For instance, the bacterium *Bdellovibrio bacteriovorus* utilizes whole cell invasion and digests other bacteria from within, acting as a microbial predator (Lovering et al. 2021).

Conversely, there are some organisms that harm another one without receiving any benefits, a phenomenon called amensalism. This can occur when microbes produce substances like lactic acid or antibiotics that inhibit the growth of some individuals in the population without directly benefiting the producer (Mougi 2016).

Microbes frequently engage in facilitation, also called cross-feeding or syntrophy, where one species benefits by consuming the metabolic byproducts of another. In anaerobic environments, fermentative bacteria, such as *Bacteroides thetaiotaomicron*, secrete acetate, formate, succinate, and propionate as major metabolic by-products, produce hydrogen, which secondary fermenters can make use of (Culp et al. 2023; Zachar et al. 2022).

These interactions result in microbial consortia performing metabolic functions collectively and are often defined by chemical or physical transformations: anabolic and catabolic processes are often distributed across different species which interact through the exchange of intermediate metabolites through the

environment (Graham et al. 2016; Qian et al. 2021; Daniels et al. 2023). These diverse interactions, that are also influenced by environmental factors such as temperature, pH, and nutrient availability, shape the environment the microbes reside in, and, as a consequence, individuals alter their metabolism to adjust to such changes. Therefore, interaction patterns, not only species, can change in time. Hence, the type of relationship one species or even one individual may have with another should be studied across spatio-temporal scales, as well as their evolutionary dynamics (Qian et al. 2021; van den Berg et al. 2022; Daniels et al. 2023). Understanding these complex relationships can shed light on how microbial systems regulate themselves, offering potential for applications in health, biotechnology, agriculture and environmental management.

The dynamics of a microbial community can present several complex cases, such as periodicities, chaos and stable states. In some cases, communities vary considerably in the initial stages of succession, but then reach an equilibrium in the abundances of the individuals present, i.e. a steady state (Faust et al. 2015; Gonze et al. 2018). However, microbial communities can exhibit multistability, that is the potential of microbial abundances to reach alternative stable states under the same environmental conditions. These can enable temporary transitions in community composition and function, and consequently impact the environment potentially leading to quasi-irreversible conditions (Lopes et al. 2024). While it is feasible to study microbial communities experimentally (Eng et al. 2019), it is not possible to recreate *in vitro* the actual complexity of real assemblages. Therefore, mathematical models are indispensable for linking community composition and connectivity to emergent functions and dynamics (van den Berg et al. 2022).

1.2 Dynamic community modeling

Dynamical systems are time-dependent processes and are typically described by ordinary differential equations (ODE) (Autorino et al. 2022). The theory that describes them originated in classical physics for modeling mechanical systems, however it has been used in many other fields, including macro and microecology, and an important property of this kind of systems is stability (Gerber 2014; Efendiev et al. 2006). In this way, dynamic modeling is a powerful mathematical tool to describe and to infer directionality and causality, as well as to predict the time-dependent properties of the system. In microbiology, it captures the dynamic behaviors of a microbiome in fluctuating environments (Gerber 2014; Qian et al. 2021). Multiple models can fit the system of interest, depending on the questions, the assumptions made and the available data (van den Berg et al. 2022).

1.2.1 Generalized Lotka–Volterra

The Lotka-Volterra model is a mathematical framework widely used to describe the dynamics of interacting populations within an ecosystem, and it is considered as a mechanistic dynamical model at the ecological scale. It was originally used to describe macrofaunal predator-prey dynamics in 1926 (Eq. 1.1). It describes how the fluctuation of the population of the prey x_i , e.g. snowshoe hare, depends on its birth rate b_i , on the population of the predator x_j , e.g. Canadian lynx, and on the rate at which the prey is consumed by the predator m_j (van den Berg et al. 2022; Gause 1935).

$$\frac{dx_i}{dx} = x_i b_i + x_i x_j m_j \quad (1.1)$$

As the equation 1.1 shows, the interactions described are limited to only two community-members. Hence, small modifications are needed in order to include

more than two community-members, as microbial communities require. These modifications define the generalized Lotka-Volterra (gLV) equations (Eq. 1.2), which are a set of non-linear, coupled, first-order differential equations. Here, the important parameters are the intrinsic growth rates μ_i instead of the birth rate, and the interaction coefficients α_{ij} between species i and j , typically represented as a vector and a matrix, respectively (Figure 1.1). The kind of interactions present in the community are described as positive or negative coefficients α (Table 1.1) and its magnitude indicates the strength. One of the primary applications of gLV in microbiology is to assess the stability and resilience of microbial communities under various environmental conditions.

$$\frac{dx_i}{dx} = x_i \mu_i + x_i \sum_{j=1}^N \alpha_{ij} x_j \quad (1.2)$$

Type of interaction	Effect of i over j	Effect of j over i
Mutualism	$\alpha_{ij} > 0$	$\alpha_{ji} > 0$
Commensalism	$\alpha_{ij} > 0$	$\alpha_{ji} = 0$
Parasitism	$\alpha_{ij} > 0$	$\alpha_{ji} < 0$
Competition	$\alpha_{ij} < 0$	$\alpha_{ji} \leq 0$
Predation	$\alpha_{ij} < 0$	$\alpha_{ji} > 0$
Amensalism	$\alpha_{ij} < 0$	$\alpha_{ji} = 0$
Cross-feeding	$\alpha_{ij} > 0$	$\alpha_{ji} \geq 0$

Table 1.1: According the type of interaction between community-members, the sign of the interaction coefficient could be positive, negative or equals to zero. See section 1.1 for more information.

The model has been widely used to predict how microbial communities respond to different environmental or anthropogenic changes, modifying or adding some elements to the equations. Here are reported three examples. Stein et al. (Stein et al. 2013) added an extra term to the gLV equations to describe the effect that some external perturbation may have on each of the community-members. Here, ε_{il} is the susceptibility i has to the time-dependent perturbation u_l , for instance, antibiotic exposure or changes in available nutrients (Eq. 1.3).

$$\frac{dx_i}{dx} = x_i\mu_i + x_i \sum_{j=1}^N \alpha_{ij}x_j + x_i \sum_{l=1}^P \varepsilon_{il}u_l \quad (1.3)$$

The subtraction of the second term of equation 1.2 takes into account the gain or loss of x_i exclusively as a result of the interaction it has with x_j , but not due to its intrinsic death rate. van den Berg et al. (van den Berg et al. 2022) added a third term that considers δ_i as the death rate of x_i (Eq. 1.4).

$$\frac{dx_i}{dx} = x_i\mu_i + x_i \sum_{j=1}^N \alpha_{ij}x_j - x_i\delta_i \quad (1.4)$$

Lastly, the nutrient availability could set a limit on the growth of the members of the community. This is called habitat carrying capacity, and it is defined as the maximum number of individuals that an environment can sustain (Wein et al. 2018). Pajon et al. (Pajon et al. 2023) modified the first term of equation 1.2 using K as a constant limiting the growth of x_i (Eq. 1.5).

$$\frac{dx_i}{dx} = x_i\mu_i \left(1 - \frac{x_i}{K}\right) + x_i \sum_{j=1}^N \alpha_{ij}x_j \quad (1.5)$$

The main advantage of the gLV model is its flexibility and ability to incorporate various types of interactions, as we showed. It can handle large communities with multiple interacting species, make it an ideal model to use for diverse microbial systems, such as gut or wastewater microbiomes. The parameters (growth rate and interaction coefficients) can be estimated using time-series data from *in vivo* or *in vitro* experiments through marker gene (e.g. 16S rRNA gene) amplicon sequencing or metagenomic sequencing. The interaction matrix can be easily interpreted (as shown in Table 1.1), offering valuable insights into community structure and dynamics (Qian et al. 2021; Gonze et al. 2018; Stein et al. 2013; Gerber 2014; Lima et al. 2020).

Despite the fact that gLV model is broadly used, it has some drawbacks. First of all, it assumes that community dynamics can be described only through pair-

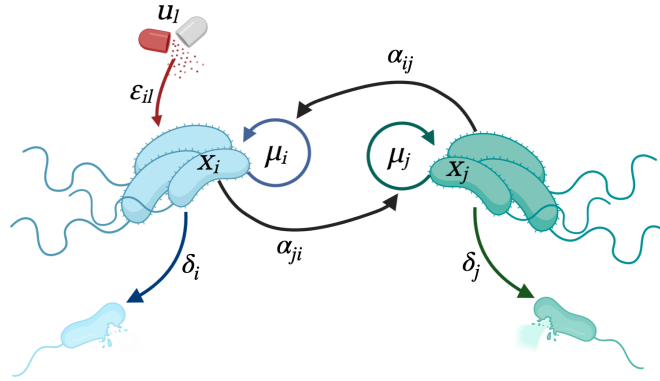


Figure 1.1: Generalized Lotka-Volterra model diagram of the possible modifications that can be made in order to take into account environmental or anthropogenic changes. For instance, the specific perturbation of antibiotic exposure (ϵ_{il}, u_i). Adapted from van den Berg et al. 2022.

wise interactions. Basically, its mathematical structure presumes that the growth of a community member depends only on its intrinsic growth rate and all pairwise interactions it has with other community members, potentially overlooking multi-species interactions that might be happening in the network (Qian et al. 2021). Second, all the interactions that gLV describes are static in time and space, and do not explicitly account for fluctuating environmental conditions, such as changes in pH, temperature, or nutrient levels, unless these are integrated as external variables. Third, if there is no carrying capacity taken into account, positive feedback loops could lead to an unconstrained growth, with the entire system becoming unfeasible (van den Berg et al. 2022). Finally, the number of interaction coefficients scales quadratically with the size of the community, requiring more computational and mathematical effort to fit the data.

To address these limitations, several extensions of the gLV model have been proposed. For example, stochastic gLV models (see section 1.2.6) incorporate random fluctuations to simulate environmental variability and stochastic events, providing a more realistic depiction of microbial dynamics, specially for small populations (Alonso-Vázquez, Giovannini, Garbini, et al. 2024). Another option could be the integration with metabolic models (see section 1.2.5). By combining gLV with genome-scale metabolic models, researchers can link microbial interac-

tions with metabolic functions, offering a deeper understanding of community functionality. Lastly, time-varying interaction coefficients allows the interaction coefficients α_{ij} to change over time, reflecting context-dependent interactions. The gLV model helps identify key members in the community and their roles. For instance, in wastewater treatment, certain microbial species are crucial for breaking down organic pollutants, and understanding their interactions could improve process efficiency. Case studies are shown in the Results chapter 3.

1.2.2 MacArthur consumer–resource model

Modeling the relationship between nutrients, toxins or signals and microbial communities could be used to elucidate how they share metabolites and divide the resources, and to what extent the interactions between members are determined by environmental changes (Bengtsson-Palme 2020). These dynamics tend to competition/mutualism, and they usually take place in non-equilibrium conditions, unlike the dynamics described by Lotka-Volterra. To this aim, Robert MacArthur developed the Consumer-Resource model (CRM) in 1969, where the main question is how species consume and deplete resources, excrete by-products and affect or are affected by signals (MacArthur 1970; van den Berg et al. 2022).

There are several studies that suggest that the availability of nutrients such as oxygen, nitrate, carbon dioxide or even light, shape the environment and constrain the composition of the microbial community. In this case, the composition of the microbial community is characterized in terms of metabolic functions rather than specific taxa (Gowda et al. 2022).

The model itself is based on two key elements: the consumers and the resources, hence it quantifies how resource use differs between consumers. In this way, it quantifies how resource utilization differs among consumers, focusing on resource overlap rather than direct interactions between community members. Essentially, if two members use a resource in a similar way, the probability of coexistence decreases. In equation 1.6, MacArthur describes the change in the

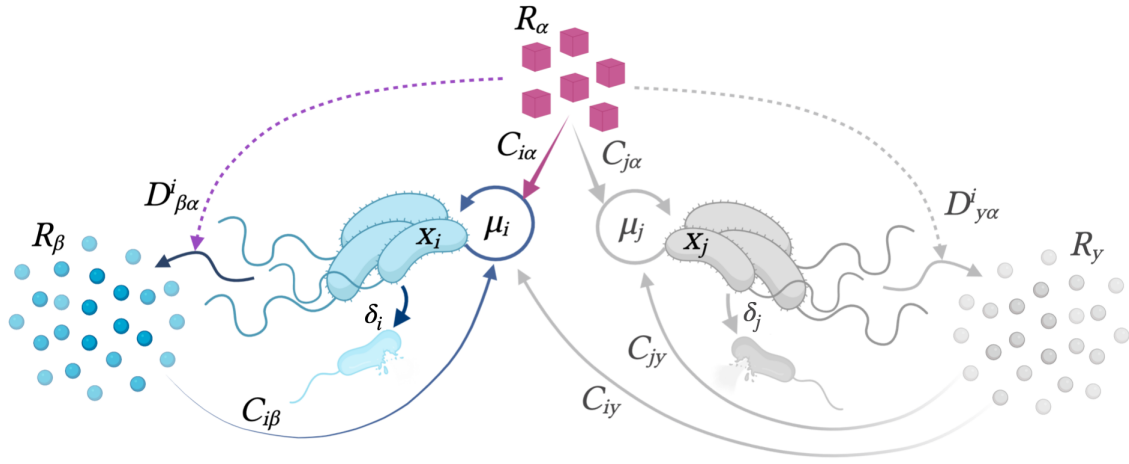


Figure 1.2: MacArthur Consumer Resource model diagram. This model describe the relationship between the consumer, i.e. species x_i in light blue, and its resources, i.e. R_α and R_β . Adapted from 1.4.

population of the consumer x_i as proportional to its growth rate μ_i and to the set of resources it could consume (Figure 1.2). This second term is composed by the concentration of the resource α R_α , the resource utilization matrix $C_{i\alpha}$, i.e. how species i uses resource α , and an energy factor $\Delta w_{i\alpha}$, i.e. a quality function that determines the amount of biomass the species i produces per unit of resource α consumed by i while maintaining energy balance. Additionally, the model incorporates a maintenance term that includes the death rate $\delta_i x_i$.

$$\frac{dx_i}{dt} = \mu_i x_i \left(\sum_{\alpha} R_{\alpha} C_{i\alpha} \Delta w_{i\alpha} \right) - \delta_i x_i \quad (1.6)$$

Let species i secrete metabolite β . Thus, the energy factor $\Delta w_{i\alpha}$ is defined as the energy density of metabolite α , w_{α} , minus the energy that species i uses to secrete metabolite β given the consumption of metabolite α . This energy is calculated with the stoichiometric matrix $D_{\beta\alpha}^i$ that specifies the units of metabolites β secreted by species i per unit of resource α consumed, multiplied by the energy density of metabolite β , w_{β} (Eq. 1.7).

$$\Delta w_{i\alpha} = w_{\alpha} - \sum_{\beta} D_{\beta\alpha}^i w_{\beta} \quad (1.7)$$

Just as the change in the population of the consumers is modeled according to the resource availability, the change in the concentration of each resource is modeled according to the population of its consumers. Equation 1.8 shows the dynamics of resource β as the balance between its secretion and consumption. The first term describes its secretion, it takes into account the stoichiometric matrix $D_{\beta\alpha}^i$, i.e. R_β secreted by i per unit of R_α , the resource utilization matrix of α , $C_{i\alpha}$, the concentration of metabolite α (needed for R_β secretion) and the abundance of species i . The second term instead describes its consumption taking into account the resource utilization matrix of R_β , $C_{i\beta}$, the concentration of metabolite β and the abundance of species i .

$$\frac{dR_\beta}{dt} = \sum_{\alpha, i} D_{\beta\alpha}^i C_{i\alpha} R_\alpha x_i - \sum_i C_{i\beta} R_\beta x_i \quad (1.8)$$

Additionally, the resources modeled using these equations may be renewable (e.g. light, nitrogen) or non-renewable (e.g. specific carbon sources), and can be used in different timescales (van den Berg et al. 2022). Modeling microbial communities often involves multiple species and resources, and even when parameters such as the uptake rates are difficult to estimate, they can be randomly sampled from statistical distributions. Cross-feeding dynamics can be included through stoichiometric coefficients to describe how the consumption of one metabolite is linked to the secretion of another, increasing the probability of coexistence rather than competition (Goldford et al. 2018; Marsland et al. 2019; Gibbs et al. 2022). Furthermore, the model can highlight how some species evolve to exploit specific resources efficiently, reducing competition and promoting biodiversity.

Despite the fact that MacArthur's model takes into account more parameters and environmental conditions than gLV model, it has its limitations. By setting the resource utilization matrices, it is automatically fixed to those specific conditions, however this could be overcome by including species' metabolic adapta-

tions, i.e. changing the preferences of nutrients. Another assumption is a steady supply, the resources are given at constant rates and unaffected by environmental or external factors. Plus, the species interact in an indirect mode, i.e. only through their shared resources, hence pairwise consumer-resource relationships. This model also assumes linear uptake rates, when uptakes could saturate at high resource availability, and this could be overcome by including saturation kinetics such as Michaelis-Menten one (Mancuso et al. 2021). However, the main challenge is to determine *a priori* chemical interactions, especially in complex or poorly characterized communities.

1.2.3 Trait-based model

Trait-based models describe the community through the functions and phenotypic traits present rather than species-wise. These approaches connect the individual traits of organisms to their impact on performance and overall community functions. Hence, it leads to a simplified community structure and often requires less parametrization effort than gLV or MacArthur CRM (Zakharova et al. 2019; Zakharova et al. 2020). The mechanistic understanding of the model is based on individuals' key traits, such as size, feeding, reproduction, etc. and associated trade-offs such as energetic needs and predation risks. The community dynamics are modeled according to how they are influenced by system-scale environmental changes (van den Berg et al. 2022).

These kind of models differ from traditional, ecological ones as they don't rely on *a priori* definitions of community structures or functional groups. Instead, they allow the emergence of new species or groups with novel trait combinations as environmental conditions change. One of the strong points of these models is the possibility of community members to both respond to and influence environmental gradients, offering a dynamic perspective on community-environment interactions (Lajoie et al. 2019). Unlike classic ecological models, the main focus of trait-based models is not on direct interactions between community mem-

bers or their implications for coexistence, like the generalized Lotka–Volterra or MacArthur consumer-resource frameworks, that determine the direct influence on fitness or the relationship with shared resources, respectively. Instead, they emphasize the relationship between community traits and environmental heterogeneity, and usually include predation as a secondary factor. A critical step in developing these models is selecting traits to include, which can be challenging when little is known about the community's environment. However, when small-scale environmental variations can be ignored, trait-based models are well suited at simulating adaptation, emergent phenotypic structures, and collective functions along environmental gradients by defining community members as their trait combinations (van den Berg et al. 2022).

Trait-based model equations are determined by two main characteristics, the dynamics of growth/death of an organism (Eq. 1.9), and the definition of its growth rate according to the environmental variability and the sensitivity the organism has for each trait (1.10).

$$\frac{dx_i}{dt} = (\mu_i - \delta_i) x_i \quad (1.9)$$

In the latter equation, the influence of an environmental factor, T , on the growth rate of species i is given by an Arrhenius-like equation. These type of equations are often used in physical chemistry to describe the energy of activation in certain circumstances, for instance, how reaction rates change according to temperature variations. Here, S_i is the sensitivity species i has to changes in T (gradient showed in Figure 1.3). Every set of trait-based equations are defined by trade-offs, like the intrinsic growth rate versus the sensitivity (Eq. 1.10). In other words, species i growth rate depends on the main resource R_α and its half-saturation constant $K_{i\alpha}$.

$$\mu_i = \frac{\mu_{i\alpha}^{max} R_\alpha}{R_\alpha + K_{i\alpha}} e^{S_i(T-T_{ref})} \quad (1.10)$$

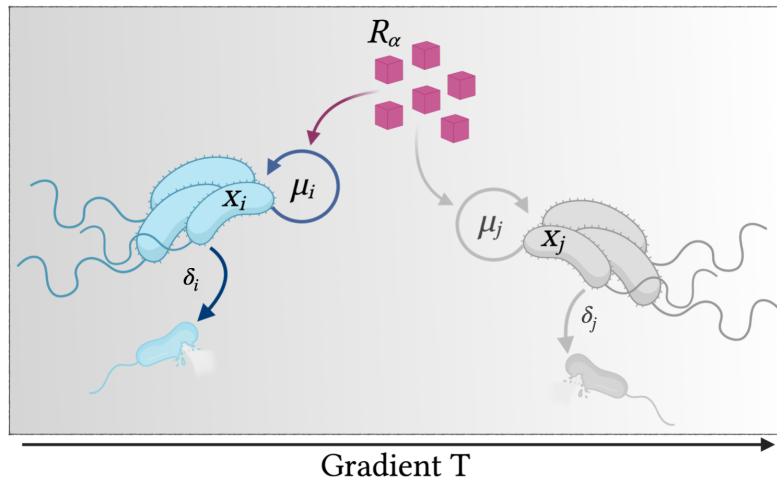


Figure 1.3: Trait-based model diagram. It is clearly shown the gradient of the environmental change of parameter T and the resources available. Adapted from van den Berg et al. 2022.

Even though trait-based models have been used, for example, to describe planktonic foraminifera using body size, calcification and feeding behavior as key traits (Grigoratou et al. 2019), they can come with notable limitations as well. One of the key challenges lies in their reliance on prior knowledge. To accurately represent a system, the model requires detailed information about the relevant environmental variables, how traits are assigned to species, and the trade-offs that exist between these traits. For instance, in equation 1.9, the resources are fixed and the trait/sensitivity must be defined a priori. This dependency can pose difficulties, especially in poorly understood ecosystems. Another limitation is their incomplete representation of evolutionary processes. These models often fail to capture the full dynamics of species adaptation and evolution over time, leaving a gap in understanding long-term community changes. Additionally, they lack explicit energetics, meaning they do not account for the energy balance or resource flow within the community, which can be crucial for accurately modeling ecological systems.

Cross-feeding interactions are not included in the model equations, a fundamental feature of many microbial communities where one species depends on the metabolic by-products of another. Furthermore, they assume that resources and

abiotic factors are independent, overlooking potential interactions or feedbacks between these elements. Species interactions, apart from predation, are similarly underrepresented, which limits the model's ability to fully capture the complexity of community dynamics. Finally, these models can be computationally demanding, requiring significant resources to simulate systems with large numbers of traits or environmental gradients. This may limit their accessibility for certain applications. Despite these challenges, trait-based models remain valuable tools, though they must be applied with an awareness of their inherent constraints.

1.2.4 Individual-based model

Individual-based models are defined from a bottom-up point of view. In other words, they describe the community by modeling a community member's interaction with its surroundings. This offers an impressive level of detail and flexibility, but comes with an important computational demand (Jørgensen et al. [2011](#)).

When it comes to microbial communities, these models operate at the resolution of individual cells, allowing for a fine-grained view of how they respond to their local environment. This higher level of detail allows the models to incorporate physical, chemical, and mechanical laws directly, ensuring they align with fundamental principles such as thermodynamics. Furthermore, by modeling all at a single cell level, they are naturally stochastic, capturing the inherent randomness in phenomena such as cell-cell interactions.

One of their strengths lies in their ability to include metabolites and other compounds as dynamic components of the system. These resources can interact, enabling the exploration of higher-order interactions within the community. The models also allow for cell-specific variations in growth and death rates, reflecting the heterogeneity often seen in real-world environments. By incorporating biochemical details, these models can predict concentration gradients of metabolites, toxins, or signaling molecules, even for compounds that are challenging to detect experimentally. Additionally, these characteristics make possible to cap-

ture micro-evolutionary dynamics, which tend to start from individual mutations (van den Berg et al. 2022).

Individual-based models are inherently modular, meaning additional biological or physical processes can be incorporated as needed. This adaptability makes them particularly useful for simulating systems with localized environmental differences or investigating the impact of specific factors at a microscopic scale. Their capacity to include such detailed layers of complexity provides researchers with a powerful framework to explore the dynamic and multifaceted nature of biological communities (Jayathilake et al. 2017).

The classic approach of individual-based models is to define a two-dimensional space where each cell is delimited inside a grid space. The cell i only gets to consume and grow on resources present within the strictly defined space, a grid cell (Figure 1.4). In equation 1.11 the growth of cell i is defined by the resources present in the same space and the energy they provide.

$$\frac{di}{dt} = \mu_i \left(\sum_{\alpha} \Delta w_{i\alpha} C_{i\alpha} R_{\alpha} \right) \quad (1.11)$$

However, as the cell grows, it can produce resource R_{β} while consuming resource R_{α} . Since they are being produced/consumed in a delimited space, there is the possibility of dispersion across the space via diffusion, which normally operates in a different timescale ($\partial t \ll dt$). This is described in equation 1.12, where the diffusion coefficient $D(R_{\alpha}, \mu)$ determines the rate at which the resource spreads and the Laplace operator $\nabla^2 R_{\alpha}(\mu, t)$ quantifies how the resource spreads out in space and in time.

$$\frac{\partial R_{\alpha}(\mu, t)}{\partial t} = \nabla [D(R_{\alpha}, \mu) \nabla R_{\alpha}(\mu, t)], \mu = (x, y) \quad (1.12)$$

However, this model has to take into account also the availability of resources and how they change in time. Equation 1.13 describes how resource R_{α} is diminished only by the consumption of cell i ($C_{i\alpha}$), and equation 1.14 describes how

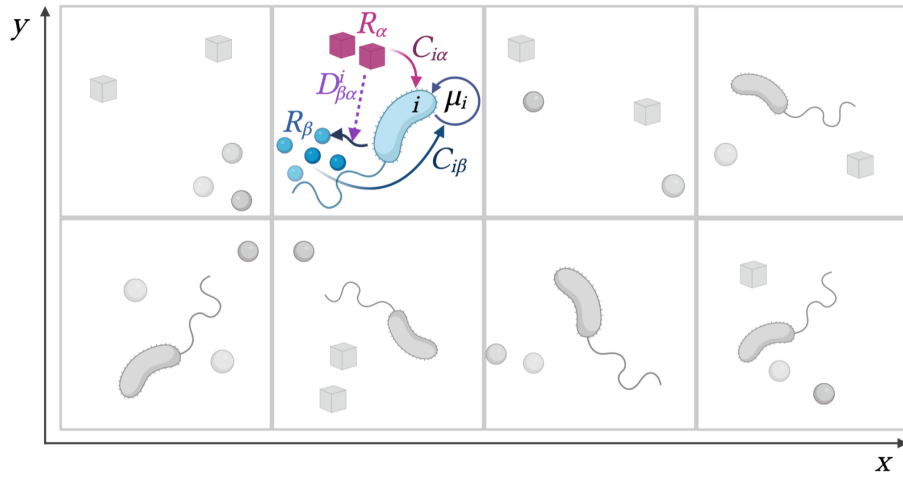


Figure 1.4: Individual-based model diagram. The area is divided in a grid, and the model describe the dynamics of each cell inside the cell-grid, including the resources present and their diffusion. Adapted from van den Berg et al. 2022.

resource R_β , the one produced by cell i , changes in time, taking into account also its production according to the resources consumed.

$$\frac{dR_\alpha}{dt} = - \sum_i C_{i\alpha} R_\alpha i \quad (1.13)$$

$$\frac{dR_\beta}{dt} = \sum_{\alpha, i} D_{\beta\alpha}^i C_{i\alpha} R_\alpha i - \sum_i C_{i\beta} R_\beta i \quad (1.14)$$

As expected, the computational demand individual-based models represent is their main drawback, since simulating large-scale, detailed systems with many interacting cells requires significant processing power and memory. Moreover, the robustness of the results generated by these type of models can be unclear. The outcomes often depend on the specific a priori assumptions and parameter choices made during the model's construction, which can lead to uncertainty about the reliability of the insights they provide. Additionally, these models require many parameters, making it difficult for researchers to interpret their biological or ecological meaning or to justify their values in a meaningful way.

The models also involve an expansive parameter space, increasing the risk of overfitting. When too many variables are adjusted to fit observed data, the model

may lose its predictive power and fail to generalize to new scenarios. Furthermore, these models cannot analytically resolve equilibrium conditions or their stability, making it challenging to understand the long-term behavior of the system or predict how it might respond to changes in initial conditions or external factors (Breckling [2002](#)).

Another concern is the need for a priori assumptions about the "rules" governing individual behavior. These assumptions are critical to the model but can introduce bias or limit its scope if they fail to capture the true complexity of the system. Finally, there is no universal framework for constructing or analyzing individual-based models, meaning that methodologies can vary significantly between studies. This lack of standardization prevents comparisons and reproducibility, and discourages their use (van den Berg et al. [2022](#)).

1.2.5 Genome-scale Metabolic models

Genome-scale metabolic models are a classical bottom-up view point of describing an organism, which relies on all its known metabolic reactions, genes and proteins of it. These models computationally describe, as the name states it, the metabolism of an organism through the gene-protein-reaction (GPR) associations.

Metabolic networks are initially reconstructed using information derived from the organism's genome and existing literature. This process includes several key steps, such as annotating the genome to identify its functions, assigning gene and reaction localizations, determining the organism's biomass composition, estimating its energy requirements, defining model constraints, identifying the relevant biochemical reactions, and determining their stoichiometry (Oberhardt et al. [2009](#); Baart et al. [2012](#); O'Brien et al. [2015](#)). The outcome of this process is a comprehensive database of chemical equations forming the reaction network in the target organism. These reactions are subsequently grouped into pathways, pathways are categorized into sectors (such as amino acid synthesis), and all components are ultimately assembled into genome-scale networks. In essence, network

reconstructions systematically compile and organize diverse information about a target organism into a coherent framework.

The next phase is to represent the reconstructed network in mathematical terms, forming the basis of a genome-scale model (GSMM). The mathematical representation depicts a stoichiometric matrix (S matrix), which includes all the coefficients of the metabolic reactions of the organism. In this matrix, the αr_1 element represents the stoichiometric coefficient of metabolite α (resource) in the reaction r_1 within the GSMM, in other words, columns correspond to metabolites and rows to reactions. A positive coefficient indicates that the metabolite is produced, while a negative coefficient signifies it is consumed. This mathematical framework enables precise modeling of the organism's metabolic processes and the most common simulation approach is flux balance analysis, which calculates the flow of metabolites through the network by constraining the space of solutions (Orth et al. 2010). This space is delimited by the resource availability, which is achieved by imposing upper and lower bounds for each reaction flux, and the optimization of an objective function in a steady state, which typically is the biomass production. Equation 1.15 shows the concentration of resource R_α according to its stoichiometric coefficients in the k th reaction of the entire set of reactions M , where each flux v_k is delimited by its lower and upper bounds, l_k and u_k , respectively.

$$\frac{dR_\alpha}{dt} = \sum_k^M S_{\alpha k} v_k = 0, \quad \text{where } l_k < v_k < u_k \quad (1.15)$$

When it comes to a microbial community, it is possible to build a community-level stoichiometric matrix which includes the information of the stoichiometric matrix of each species i , i.e. S_i (Figure 1.5). Flux balance analysis (FBA) is a powerful tool for predicting both the composition of microbial communities and the dynamics of metabolites within them. In the case of a microbial community model, this method works by solving a series of steady-state problems at discrete

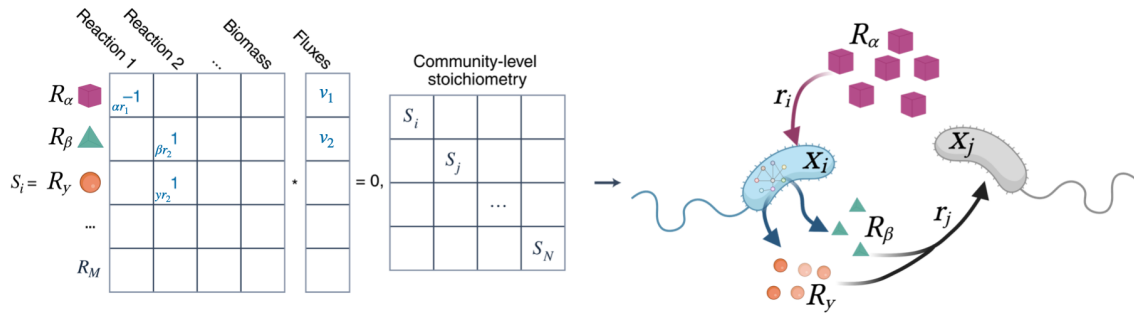


Figure 1.5: Genome-scale metabolic model. From left to right: a the stoichiometric matrix S_i , describing the GPR rules of species i and constrained to 0 (steady state); the community-level stoichiometry; the diagram of the resources interacting in the metabolism of species i . Adapted from 1.4.

time intervals, ensuring that metabolic reactions within the system remain balanced over time. The exchange of metabolites, such as the uptake and secretion by microbes, is captured using differential equations, adding a dynamic layer to the analysis.

Additionally, it is possible to model multiple species or genotypes simultaneously, offering a quantitative perspective on how microbial communities evolve and interact. Associating intracellular metabolic processes to community-level behaviors allows detailed predictions of how individual metabolic events shape broader ecological dynamics, describing also functional details of microbial communities. This can be done through the use of different -omic data, making it an essential tool for understanding the relationship between microbial metabolism and community composition over time (Kim et al. 2022; van den Berg et al. 2022; Muller et al. 2018).

As any other model, it has limitations. First of all, it is challenging to choose a biologically relevant objective function when simulating a community, although recently it has been described a methodology to build and evaluate GSMM of microbial communities (Tarzi et al. 2024). Second, there is no negative interactions apart from resource competition. Lastly, each GSMM requires extensive curation to achieve accuracy, especially when the target microbe is not related to a model organism since the available information is less (Coyte et al. 2015; Machado et al.

2021).

1.2.6 Deterministic or Stochastic models?

As we have seen so far, in microbial ecology, mathematical models are crucial tools to describe and predict the dynamics of microbial communities. These models can be broadly categorized into deterministic and stochastic approaches, each with distinct features and applications. Yet, it is possible to integrate stochastic processes into the deterministic models by using specific algorithms (Gonze et al. 2018).

Models such as the generalized Lotka-Volterra and MacArthur CRM are deterministic by nature, operate under the assumption that microbial community dynamics are predictable and are driven by fixed equations. These models rely on parameters like growth rates, interaction coefficients, and resource utilization rates, often assumed to be constant or influenced by external environmental parameters. For instance, gLV models use differential equations to represent the population dynamics of each species describing it through the use of pairwise interactions, and MacArthur models describe the dynamics by focusing on the resource overlap among species (see subsections 1.2.1 and 1.2.2).

Deterministic models are particularly useful for examining broad community behaviors and understanding the theoretical principles underlying microbial interactions. They excel at explaining how communities respond to perturbations or how environmental gradients shape the population structures. However, these models often simplify the intrinsic complexity of microbial ecosystems, as they do not account for random variability or higher-order interactions beyond pairwise relationships (Xu et al. 2021).

Stochastic models, on the other hand, incorporate randomness and uncertainty, reflecting the inherent variability of microbial communities. Their main characteristic consists in acknowledging that microbial interactions and environmental factors are not always predictable and that random events can influence

community dynamics. For instance, individual-based models are stochastic by nature. It takes a bottom-up approach by modeling every community member as an individual interacting agent/cell, their interactions with one another and their environment, by defining a set of ‘rules’ of any complexity (van den Berg et al. 2022).

Stochastic models can capture fine-scale heterogeneity, spatial variation, and local interactions, offering a more refined view of microbial ecosystems. In essence, these models set the rule that the transition of the system from state at time t to a new state at time $t + 1$ is random, hence it only can be specified the probability of being at a state n at time t , $P(n, t)$. To simulate this process we can use the statistically exact Gillespie algorithm (Cota et al. 2017; Black et al. 2012; Alonso-Vásquez, Giovannini, Garbini, et al. 2024).

Naturally occurring fluctuations are stochastic processes that these models can incorporate, such as random dispersal, mutations, and variations in environmental conditions, which deterministic models overlook. Thus, it is possible to study cross-feeding, resource competition, and emergent properties such as adaptation and coexistence (Black et al. 2012; Gonze et al. 2018). For instance, the dynamics of resource diffusion can be described using partial differential equations that include random variability, such as in the individual-based model equation 1.12.

As we said at the beginning of this section, ecological models can be described from a deterministic or a stochastic point of view, despite of their inherent nature. That is the case of generalized Lotka-Volterra. We used this framework and modeled the deterministic and stochastic scenarios to describe the dynamic of a three-member bacterial community with one persister-formation pathogenic bacteria (see section 3.2) (Alonso-Vásquez, Giovannini, Garbini, et al. 2024). Another example, is the MacArthur consumer-resource model, where stochastic colonization can be added (Marsland et al. 2019)

The trade-offs between these two approaches are significant. Deterministic models offer computational simplicity and analytic clarity but often lack the de-

tail needed to represent the full complexity of microbial ecosystems. Stochastic models, on the other hand, are more realistic and capable of capturing emergent properties, however they are computationally intensive and require careful parametrization to avoid overfitting. Additionally, stochastic models are less suited for making generalized predictions due to their dependence on probabilistic outcomes (Black et al. [2012](#)).

In practice, combining deterministic and stochastic elements can provide a more balanced modeling framework. For example, deterministic equations can describe broad community dynamics, while stochastic components introduce variability and account for fine-scale processes. This hybrid approach balances computational efficiency with ecological real scenarios, making it a powerful tool for describing the heterogeneity of microbial ecology (Xu et al. [2021](#); Song et al. [2021](#); He et al. [2021](#)).

1.3 Wastewater treatment plants

Later in this dissertation we will be focusing our attention on wastewater treatment plants (WWTPs), an essential part of modern society, designed to manage and treat wastewater coming from households, industries, and agriculture. Their main scope is to remove contaminants, including organic matter, harmful microorganisms, and pollutants such as metals or antimicrobials, so the water becomes safe for the environment or reuse. These plants play a crucial role in maintaining public health, preserving ecosystems, and providing sustainable water management in a world facing a climate emergency where water scarcity and pollution challenges are more frequent everyday (Moura et al. [2022](#); Bassin et al. [2021](#)). Moreover, modern WWTPs often aim to recover valuable resources. Nutrients like nitrogen and phosphorus can be extracted and reused as fertilizers, as well as the biogas that is generated during anaerobic digestion of sludge, since it could serve as a renewable energy source (Bhambhani et al. [2024](#)).

The design and construction of wastewater treatment plants is a multidisciplinary process that goes from engineering to biology. A typical WWTP consists of multiple tanks with a specific function in the water treatment workflow. The first step of the process is where wastewater enters the facility. A conventional treatment plant has to remove large debris, such as plastics, wood, and other solid wastes, which is normally done by using mechanical screens. This stage, known as preliminary treatment, prevents blockages and mechanical bulks in subsequent processing tanks (Kolpakova et al. 2024). In the next phase, primary treatment, sedimentation of heavier solids takes place. This heavier sediments settle to the bottom forming a sludge layer, while the lighter ones float to the surface and are removed. The clarified water moves to the secondary treatment stage, which is biologically driven. This stage often employs microorganisms to degrade organic pollutants in aerated tanks or through biofilm reactors (Elginöz et al. 2019). Some WWTPs have an extra step: tertiary treatment. This step has the aim to improve the characteristics of the water coming out of the secondary treatment and make it suitable for reuse (Zagklis et al. 2022). This process includes filtration, disinfection, and chemical treatments to remove residual nutrients, pathogens, and micropollutants such as pharmaceuticals or heavy metals. The treated water can then be safely released into rivers, lakes, or oceans, or even be reused for irrigation.

The construction of these plants must also take into account local environmental and population needs. Factors such as the volume of wastewater generated, the types of contaminants present (industry/household), and local regulatory standards dictate the design technicalities. In urban settings, large centralized plants are common, while rural areas may rely on decentralized or smaller-scale systems (Alasino et al. 2007).

The world's largest wastewater treatment facility is the Stickney Water Reclamation Plant, located near Chicago, Illinois, USA. Managed by the Metropolitan Water Reclamation District of Greater Chicago, it serves a population equivalent

of over 2.2 million people. The facility has a treatment capacity of approximately 5.3 million cubic meters of wastewater per day (Greater Chicago [n.d.](#)). Its capacity exemplifies the crucial role of WWTP in large, densely populated urban areas.

While WWTPs are essential for environmental protection, they also have significant impacts. Clearly, they reduce pollution loads in water bodies, helping biodiversity and ecosystem health, but the process can have downsides. The main problems are greenhouse gas emissions and methane release during anaerobic digestion, and that poorly managed sludge can contaminate soil and water sources (Hao et al. [2019](#)). Additionally, WWTPs also play a crucial role in addressing the removal of micropollutants and antibiotic-resistant bacteria, a latent risk for human and environmental health. Hence, it is important to develop technologies to mitigate these issues, not only from the wastewater treatment industry view point, but also from the biological and environmental sciences one (Su et al. [2021](#); Hao et al. [2019](#)).

There are various types of WWTPs, each suited to specific needs and contexts. The main two types of wastewater are *i*) the water that comes from domestic activities, also known as sewage water, and *ii*) the one that comes from manufacturing, industrial and commercial activities, also known as industrial wastewater. WWTPs may be designed to treat either a single type of wastewater or both. For instance, municipal WWTPs, handle domestic and industrial wastewater from urban areas, while industrial WWTPs treat the second type of wastewater. They often require specialized treatment for chemicals and heavy metals, depending on the type of industry. Decentralized systems work with small communities or remote areas wastewater, offering cost-effective solutions for localized wastewater management (Hospido et al. [2008](#)). Another type are membrane bioreactors (MBRs), which combine biological treatment with membrane filtration, removing organic and inorganic matters as a biological unit, obtaining high-quality treated water, often used in water reuse applications (Al-Asheh et al. [2021](#)).

One essential aspect of WWTPs are microorganisms, since they are the main agents driving the biological processes that remove organic matter and pathogens. The microbial community present is diverse, and includes bacteria, archaea, fungi, and protozoa, each playing specific roles in the degradation of contaminants. In the secondary treatment phase, microbes metabolize organic pollutants, converting them into biomass, carbon dioxide, and water. The species present in this process are determined by the characteristics of the pollutants, for example, aerobic bacteria dominate in systems like activated sludge, where oxygen is supplied to support their metabolism, while anaerobic bacteria and archaea break down organic matter to produce methane in anoxic environments (Meerbergen et al. 2017).

Microbial dynamics in WWTPs are complex and depend on the interactions between the populations present. Factors such as temperature, pH, oxygen availability, and the composition of the wastewater influence microbial community structure and function, along with stochastic processes present. Hence, developing mathematical models, as well as using -omic data can shed light to community structure and enhance microbial efficiency (Xia et al. 2018). These models can describe some of the WWTP features such as cross-feeding. For instance, in denitrification, one group of bacteria reduces nitrate to nitrite, while another completes the reduction to nitrogen gas (Lehtovirta-Morley 2018). This type of interactions typically lead to an overall process stability.

In conclusion, WWTPs are fundamental for modern society, since they can provide clean water, protect ecosystems, and recover valuable resources. It is crucial to have good practices when designing and building them, and to investigate the microbial communities' functionality, since big part of the process relies heavily on their dynamics. Given the ongoing importance of water management in our societies, it is necessary that WWTPs continue to evolve, integrating state-of-the-art technologies and microbial insights to address emerging challenges in an efficient way.

Chapter 2

Overview and aims of the work

2.1 The project

The rapid growth of the global population since the 20th century, along with accelerated urbanization, has inevitably led to an increase in the production of wastewater. In the last years, European directives are pushing for better management of wastewater treatment plants since they are critical for ensuring sustainable development in modern societies. The main component of wastewater treatment plants are microbial communities. These are essential for breaking down several substances in wastewater through their metabolism. Recent studies in several countries, including Italy, have shown that microbial composition is not influenced by location rather by stochastic processes, such as dispersion, and deterministic factors, such as temperature, pH, etc. However, the temporal changes in microbial composition is poorly understood, and knowing that microbial composition can change based on environmental factors, it was necessary a longitudinal study to gain a deeper understanding of the processes that affect WWTP efficiency throughout the year and to describe the dynamics of this specific communities.

This project was carried out in collaboration with G.I.D.A. S.p.A. (Gestione Impianti Depurazione Acque). G.I.D.A. is a wastewater treatment plant located

in Prato, Italy (Figure 2.1) and is conformed by several depuration systems across the province that are connected to the domestic and industrial users by a sewage network. However, this project involved the two biggest ones: Calice and Baciacavallo, where the resulting sludge from the purification process can amount up to 10,000 t/year and 30,000 t/year, respectively.

Calice is composed of two tanks dedicated to the treatment of urban wastewater (IDL, Impianto Depurazione Liquami), which receive domestic effluents from the western part of the city and the Municipality of Montemurlo, and tanks dedicated to handle non-hazardous liquid waste such as landfill leachate and sludge (ITL, Impianto Trattamento Liquami). The depuration process (IDL) consists of equalization, primary sedimentation, denitrification, biological oxidation, secondary sedimentation, clear final flocculation, sand filters and ozonation. The sludge line (ITL) consists of gravity thickening, mechanical dewatering with centrifuge and filter press. Baciacavallo is the centerpiece of the centralized depuration system being the oldest and biggest of all of them. The depuration process consists of equalization, primary sedimentation, biological oxidation, secondary sedimentation, clariflocculation and final refinement with ozone, for the reduction of color and residual surfactants (Figure 2.2a).

The project focused on researching and developing strategies and methods to identify the functional role of microbial strains naturally present in wastewater treatment plants, which could potentially be used as resources in the biotechnology sector. To achieve this, we conducted a bi-monthly sampling over a two-year period (February 2022 - January 2024; Figure 2.2b). Sampling was carried out at three different process stages at Calice depuration plant: ITL (sewage), denitrification, and biological oxidation. As well as for biological oxidation at the Baciacavallo plant (Shown in figure 2.2a). The number of tanks for each function are different. There are two tanks for ITL, two for denitrification and four for biological oxidation at Calice and Baciacavallo, each. We sampled each tank with six replicates at Calice, and with two replicates at Baciacavallo, making a total of

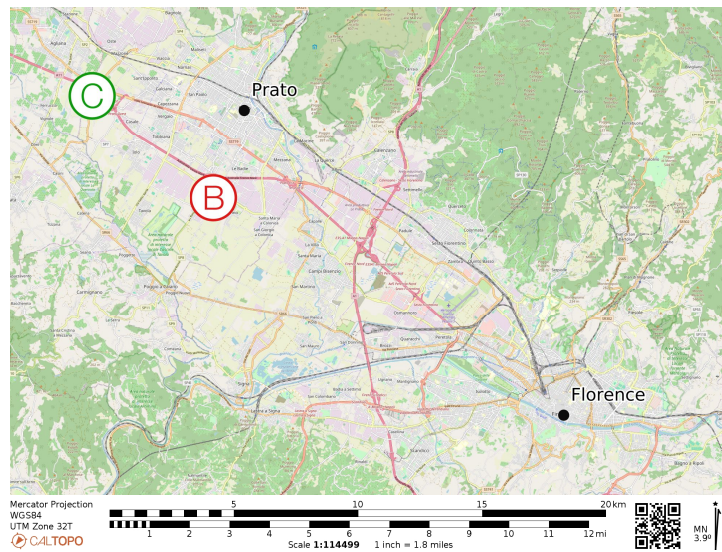


Figure 2.1: Map of the two depuration plants located in Prato, Italy. Calice in green (C) and Baciacavallo in red (B).

672 samples.

Molecular biology techniques were employed to characterize the microbiome found in activated sludge, assessing the presence and quantity of bacteria as well as their interaction dynamics. Hence, we performed DNA extraction from all the biological samples, followed by Next Generation Sequencing (NGS) of the 16S rRNA. We analyzed the sequencing data and obtained a precise identification of the bacterial component in wastewater. This allowed the detection of environmentally significant organisms and enabled the selection of representative organisms to monitor over time, observing changes in their composition and metabolic interactions, as referred in (Alonso-Vásquez, Giovannini, Fagorzi, et al. 2024).

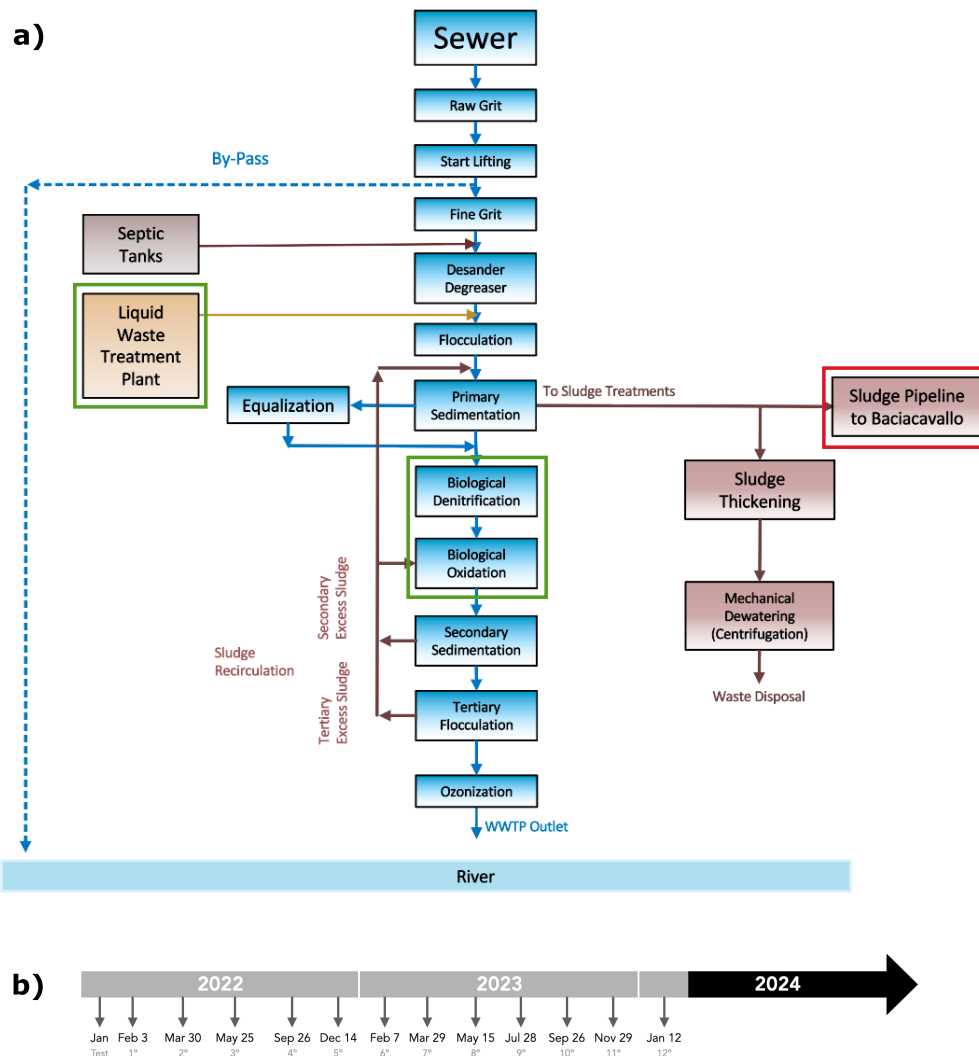


Figure 2.2: **a)** Depuration process of Calice and Baciacavallo. Blue-filled squares denote the treatment of urban wastewater (IDL), yellow-filled square denotes the treatment of non-hazardous liquid waste (ITL) and burgundy-filled squares to the right denote the sludge treatment at Baciacavallo. Sampled tanks are showed with green-outlined squares for Calice and a red-outlined square for Baciacavallo. **b)** Time line of sampling points.

2.2 Aims

This research project has as a long-term aim to develop innovative and potentially economically advantageous biotechnological processes for wastewater treatment. However, the short-term aims implied a thorough description of the communities and their environmental contexts over time, followed by the modeling of the dynamics present in the dominant community. In this way, the project focused on two main objectives:

- i)* Analysis and characterization of bacterial communities present in GIDA depuration plants.
 - Deepening the knowledge and expertise in NGS data analysis, including 16S rRNA and metagenomic samples.
 - Correlating environmental variables (e.g. temperature, pH, nutrient levels) with changes in microbial composition and function.
 - Identifying key environmental drivers that affect wastewater treatment efficiency.
- ii)* Modeling the dynamics of microbial communities.
 - Research on existing microbial community models.
 - Developing a theoretical model of a community and their dynamics before, during and after an environmental perturbation.

Chapter 3

Results

3.1 NGS: from samples to microbiomes

Understanding microbial communities is crucial for many areas of research, from environmental monitoring to medical microbiology. The diversity and complexity of bacterial populations in different habitats require specific methodologies to study them effectively. In this chapter, I will introduce key approaches used to analyze bacterial communities and their genomes, focusing on sequencing techniques and data analysis strategies. Subsections [3.1.1](#) and [3.1.2](#) show the analysis of environmental samples, one from wastewater and the other from marine sediments. While subsection [3.1.3](#) focuses on the analysis of bacterial strain genomes.

There are different approaches to study bacteria and their communities. In this section I will discuss three methods commonly used in this field. When sequencing environmental samples, we are almost certain that multiple bacterial species (or other taxa) will be present, and the analysis can be addressed by using targeted or shotgun sequencing. The first option requires a specific marker, depending on our research question. For instance, amplifying and sequencing the 16S rRNA gene is a common practice, as the workflow I used in (Alonso-Vásquez, Giovannini, Fagorzi, et al. [2024](#)) using wastewater samples. On the other hand, shotgun sequencing results in reads that may belong to any part of the genomes

present in the sample, requiring a different workflow. Assembling all the potential genomes within each sample is an option, alternatively the reads could also be mapped and classified using a reference database, for example following the workflow I used in (Alonso-Vázquez, Fagorzi, et al. 2024) with marine sediment samples. Finally, when we aim to characterize a specific organism, the data could be processed for *de novo* sequencing and assembly of the genome of a single species, as we did in Semenzato et. al. (Semenzato et al. 2022).

3.1.1 GIDA: from 16S to microbial community functions

The operation of wastewater treatment plants (WWTPs), as we saw in the introduction of this thesis (see section 1.3 of Chapter 1), relies on complex microbial communities that facilitate the degradation of organic matter, nutrient cycling and pollutant removal. In this study, we characterized the microbial dynamics in a WWTP located in Prato, Italy, managed by G.I.D.A. S.p.A. Sampling was conducted bi-monthly from January 2022 to November 2023, with additional sampling after a flood event happened in November 2023, resulting in 689 samples from 12 tanks across two main facilities: Calice and Baciacavallo, as described in Chapter 2. These facilities process up to 169,000 m³/day of wastewater, with Calice focusing on sewage and biological oxidation and Baciacavallo on biological oxidation.

We extracted the 16S rRNA gene from all the samples and Biomarker Technologies (BMKGENE) performed the metagenomic targeted sequencing, identifying 744 bacterial families, and we defined a core community formed by 51 families, representing 76% of the total relative abundance. These families, distributed across 13 phyla, exhibited functional roles vital to wastewater treatment, such as nitrification (e.g., *Nitrosomonadaceae*) and denitrification (e.g., *Nitrospiraceae*). Our analysis showed that environmental factors like nitrogen compounds, heavy metals, and pH shaped the community composition and its dynamics significantly. Seasonal variations were evident, with families such as *Rhodocyclaceae* and

Anaerolineaceae showing different abundance patterns linked to nitrogen cycling and organic matter degradation.

Additionally, we identified functional subcommunities within the core. A stable subcommunity exhibited minimal variance over time, suggesting resilience to environmental fluctuations. In contrast, a variable subcommunity demonstrated significant temporal shifts, reflecting sensitivity to changing conditions. These dynamics underscore the critical role of microbial diversity in maintaining treatment efficacy. We annotated the functional metabolic groups using the FAPROTAX database, which highlighted key metabolic processes such as aerobic chemoheterotrophy, denitrification, and nitrite respiration. These processes displayed seasonal fluctuations, particularly in oxidation and denitrification tanks.

Environmental measurements revealed significant differences among tanks. The sewage treatment tank showed pronounced variability due to high chemical oxygen demand. Oxidation and denitrification tanks exhibited greater stability, with microbial communities clustering by location and treatment stage. Seasonal trends influenced by temperature, nutrient availability, and heavy metals further differentiated microbial profiles, suggesting a possible relationship between environmental pressures and microbial adaptation.

Metagenomic analysis are able to link microbial composition to functional capabilities, which enables predictive modeling for system optimization. For instance, understanding the role of core taxa like *Comamonadaceae* and *Rhodocyclaceae* in phosphorus removal can guide targeted interventions to enhance nutrient cycling. The observed resilience of stable subcommunities suggests potential strategies for improving robustness under fluctuating conditions, such as those induced by urbanization or climate change.

Until now, this study highlights the importance of integrating microbial ecology into WWTP operations. It provides a foundation for sustainable wastewater management practices by understanding the relationships between microbial diversity, environmental factors and system performance. These insights are par-

ticularly relevant as global wastewater production rises, which lead to the need of efficient and adaptable treatment solutions. The comprehensive profiling of microbial communities, their functions and interactions at G.I.D.A. S.p.A. is an ongoing study, with much still to be explored and understood.

Dynamics and Core Microbial Community Functions in a WWTP

Alonso-Vázquez Tania¹, Giovannini Michele¹, Fagorzi Camilla¹, Perrin Elena¹, Mari Lorenzo², Berardi Chiara², Palli Laura², Macellaro La Franca Letizia², Coppini Ester², Fibbi Donatella², Fani Renato¹, Fondi Marco¹

¹*Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, 50019, Italy*

²*G.I.D.A. SpA, Via Baciavacallo 36, Prato, 59100, Italy*

Abstract

This study explores the microbial community dynamics in a wastewater treatment plant (WWTP) located in Prato, Italy. GIDA SpA operates this plant using activated sludge systems across several facilities, and this study focuses in the main two: Calice and Baciavacallo, treating up to 169,000 m³/day. Between January 2022 and November 2023, 689 samples were collected from three different treatment stages (sewage, oxidation and denitrification) to assess seasonal microbial diversity and environmental interactions. Metagenomic analyses identified 744 bacterial families, and we defined a core community with 51 families with an average abundance of 76%. Seasonal and functional variations within these communities were mapped using statistical models and redundancy analysis. Core families such as *Rhodocyclaceae* and *Anaerolineaceae* showed differential seasonal behaviour linked to environmental parameters like nitrogen levels, heavy metals, and pH. Functional profiling revealed roles in nitrification, denitrification, and organic matter degradation, critical for plant efficiency. The findings highlight significant temporal and spatial variations in microbial communities influenced by chemical and physical factors. This comprehensive profiling emphasizes the need for microbial ecology integration in WWTP optimization, particularly as global wastewater production increases. By linking microbial structure to functionality, this study provides actionable insights for enhancing sustainability and resilience in engineered ecosystems.

Keywords: Microbial ecology, Wastewater treatment plant, Sludge systems, Microbial Community Dynamics

1. Introduction

Microbial communities are fundamental components of various ecosystems, particularly in wastewater treatment plants (WWTPs) where they facilitate essential processes such as the degradation of organic matter, nutrient removal, and the detoxification of pollutants. These microbial populations, mainly formed of bacteria, archaea, fungi and viruses, collectively contribute to key biogeochemical cycles like those of carbon, nitrogen, sulfur, phosphorus, and certain metals. These cycles play a significant role in wastewater treatment processes such as nitrification, denitrification, oxidation and anaerobic digestion (Ju & Zhang, 2015; Wu et al., 2019; Zhang et al., 2012).

The structure of these communities, the diversity of microbes, and the interactions among dominant microorganisms play a crucial role in determining the performance and the functional stability of WWTPs (H. Qin et al., 2018; Zhang et al., 2012). Studying the dynamics of microbial communities in wastewater treatment facilities is crucial for improving how effectively these plants operate and stay stable. The way these microorganisms handle nutrients and pollutants plays a big role in treatment success, which becomes even more important as cities and industries expand (Rizzo et al., 2013). Generally, microbes in wastewater treatment plants are influenced by several environmental factors such as pH, temperature, and oxygen availability, as well as the presence of pollutants like pharmaceutical leftovers, personal care items, and antibiotic-resistant genes. These factors act as selective pressures, especially in highly contaminated environments like WWTPs impacting microbial communities both on the short and long term (Carles et al., 2022; Kumar et al., 2023) and posing challenges to the sustainability of existing treatment techniques.

Moreover, studying microbial dynamics in WWTPs provides insight into microbial ecology in engineered systems, helps predict responses to environmental changes, and supports the development of strategies to mitigate challenges like microbial resistance or the proliferation of harmful microbes (Wardi et al., 2023). Consequently, advancing our understanding of microbial communities in wastewater treatment is crucial for enhancing the performance and reliability of these systems in the face of increasing urbanization and industrialization.

In its most recent report, the United Nations noted that 187 billion cubic meters of wastewater were generated globally in 2022, while 220 billion cubic meters were treated — indicating that more wastewater was treated than produced. Despite this, only 58% of the

wastewater generated worldwide undergoes safe treatment, with Italy achieving a higher rate of 70% (2024).

In this work, we analysed the microbial communities present in a WWTP located in Prato, Italy. GIDA SpA (Gestione Impianti Depurazione Acque), is a domestic and industrial WWTP and has a biological core depuration process with activated sludge systems. These systems are widely used in WWTP and normally present an enormous microbial diversity that can vary worldwide (Liu et al., 2023; Wu et al., 2019). There are three main depuration systems by GIDA: *i)* Baciacavallo, the biggest treatment plant, treating up to 130,000 m³/day, reducing up to 100,000 kg/day of COD and 4,500 kg/day of surfactants; *ii)* Calice treats up to 39,000 m³/day, reducing up to 20,000 kg/day of COD and 1,100 kg/day of surfactants; and lastly *iii)* Val Bisenzio, separated in three lines, treating up to 25,000 m³/day, reducing up to 12,000 kg/day of COD and 350 kg/day of surfactants. We regularly sampled this WWTP from January 2022 to November 2023, for a total of about 700 samples that were representative of different tanks (i.e. different steps along the treatment pipeline) and used targeted metagenomics to explore both the diversity of the WWTP microbial community and, more interestingly, its dynamics and its association with seasonality and environmental parameters.

2. Materials and Methods

2.1. Sampling

Here, we have investigated the microbial communities of wastewater samples coming from GIDA SpA (Gestione Impianti Depurazione Acque), a WWTP located in Prato, Italy (Figure 1). The depuration system is composed by three main plants: Calice (P1), Baciacavallo (P2) and Val Bisenzio. The samples came from the first two plants, that cover up to 87% of the treated water per day by GIDA S.p.A. (169 m³/d out of 194 m³/d).

The depuration process is divided in two disposal lines that are parallelized, and it is composed by the entry of sewage and sewer wastes, separately. After their corresponding treatments, the process continues with the sludge treatment in Baciacavallo (biological oxidation treatment) and Calice, where it follows biological denitrification and then oxidation treatment.

We sampled 12 different depuration tanks: two tanks of sewage treatment, two of biological denitrification and four of biological oxidation in Calice, and four tanks biological oxidation in Baciacavallo (Table S1). Between January 2022 and November 2023, we collected every

two months six samples from each tank in Calice and two from each tank in Baciacavallo, having a total of 54 samples per time point. However, due to a flood that hit the area of Prato on November 4th 2023, we added three time points with 24 samples each, resulting in a total of 689 samples.

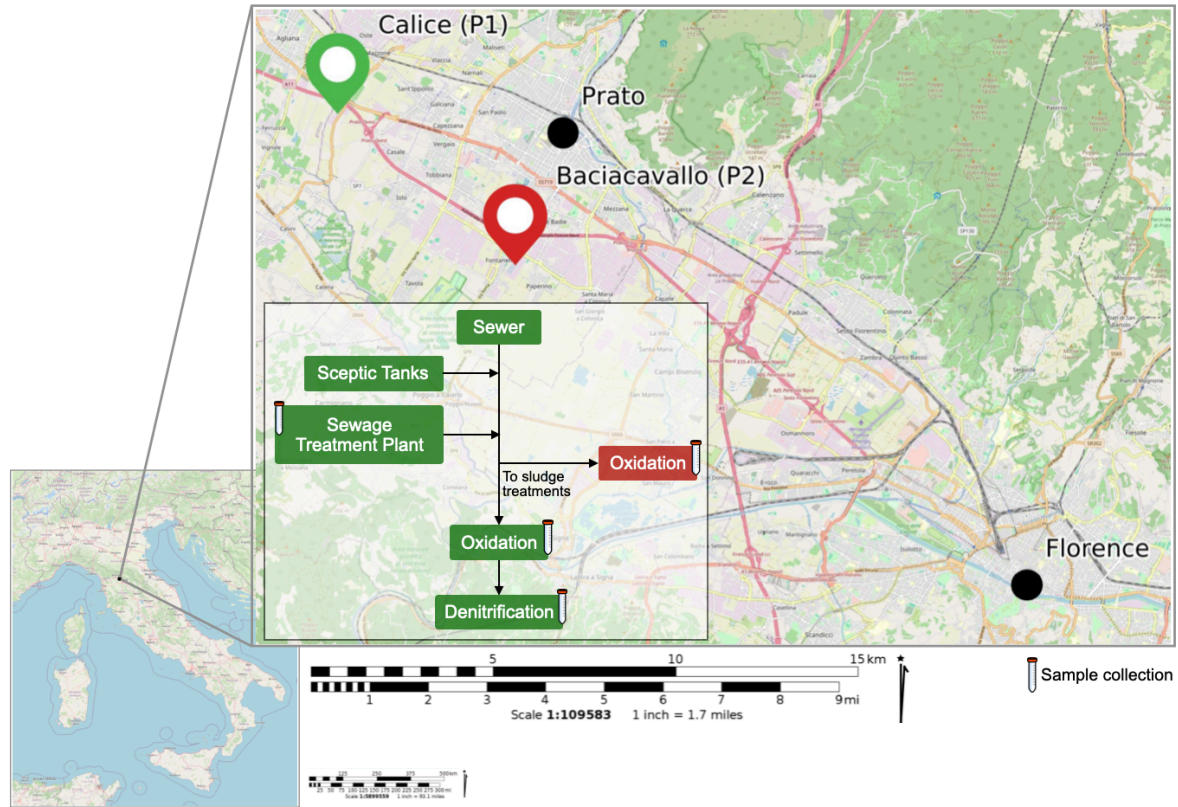


Figure 1. Map of the locations of the two depuration plants: Calice (P1, green) and Baciacavallo (P2, red). The depuration process is represented in a diagram, with the 4 points of sample collection represented: Sewage, Oxidation and Denitrification for Calice, and Oxidation for Baciacavallo.

2.2. Environmental parameters analysis

Chemical and physical factors were measured by GIDA S.p.A. using the Standard Methods for the Examination of Water and Wastewater (American Public Health Association, 1998) specifically for the chemical and biological oxygen demand (COD and BOD). Nitrogen was measured using different kits such as Hach LCK342 for nitrites, LCK339 for nitrates, LCK305 for ammonium and LCK138 for total nitrogen and organic nitrogen (NO_2 , NO_3 , NH_4 , N_{tot} , N_{org}). Phosphorus and chloride (P and Cl) were determined using a ThermoFisher Scientific Gallery – Automated Discrete Photometry Analyzer with the standard solutions of System Reagents for Environmental and Industrial Analysis. To measure heavy metals' presence, the

wastewater samples were subjected to digestion with HNO₃ 69% using the CEM MARS-6 microwave digestion system according to APAT IRSA-CNR 3010B. Determination was performed by atomic emission spectrometry inductively coupled to argon plasma with an optical detector, using the ICP - OES Thermo Fischer Scientific instrument, iCAP7200 Duo, according to APAT IRSA-CNR 3020 standard. Quantification is done with external CRM AccuStandard and LabMix24 Ultra Scientific calibration control and the specific wavelength of each element (Pb, Cu, Zn, Cr, Hg, As, Ni, Cd). Suspended solids were measured using a 10 µm filter. To measure the mixed liquor suspended solids (MLSS), a 50mL sample was filtered, dried and weighted. The sludge retention time (SRT) was calculated using the biomass in and out of the depuration system (Table S1).

Measured environmental data were analysed to test whether each environmental factor differed between the depuration tanks, and whether each environmental factor changed over time within individual tanks. We tested normality and homoscedasticity of the data using Shapiro-Wilkinson and Levene's tests, respectively. If the assumptions were met, we conducted a one-way analysis of variance (ANOVA) and a Tuckey's HSD post-hoc test. Otherwise, we used the non-parametric Kruskal-Wallis test and Dunn's post-hoc test with the Bonferroni correction. To assess if environmental factor changed over time within tanks, a separate analysis was conducted for each tank individually. All these analyses were conducted using R and the packages car and FSA.

We performed a redundancy analysis (RDA) to explore the relationship between the relative abundances of the main bacterial families and the environmental factors across time. We used this multivariate constrained ordination to identify which environmental variables best explained the variation in the microbial community composition for each tank. The 26 environmental factors measured were divided into three groups: Chemical factors (N_{tot}, NH₄, NO₂, NO₃, N_{org}, Cl, P, BOD, COD), heavy metals (Pb, Cu, Zn, Cr, Hg, As, Ni, Cd) and physical factors (Temperature, OD, pH, SRT, MLSS, TSS, Q_{tot}).

2.3. DNA extraction and sequencing

We collected 50 ml of wastewater and centrifugated it at 9000 rpm for 5 minutes. After removing the supernatant, the pellets were stored in Falcon tubes at +4°C. DNA extraction was performed using 250 mg (wet weight) of each sample, processed using DNeasy PowerLyzer PowerSoil (Qiagen) according to manufacturer's instructions, and stored at -20°C. The extracted DNA was visualized for quality and integrity on stained agarose gel and was

quantified using TECAN. PCR amplification and 16S rRNA sequencing of the V3 and V4 variable regions were performed by Biomarker Technologies (BMK) using Illumina NovaSeq and the prokaryotic primers 338F (5' – ACT CCT ACG GGA GGC AGCA – 3') and 806R (5' – GGA CTA CHV GGG TWT CTA AT – 3') (Zeng & An, 2021).

2.4. Data processing

Sequencing data were analysed using the QIIME2 software (version 2023.9.1), (Bolyen et al., 2019), the paired reads were trimmed and merged, quality filtered and denoised using the DADA2 plugin. To perform the classification, it was used the Silva database (release SSU 138) (Quast et al., 2012), and clustered into amplicon sequence variants (ASV). The classification confidence was ~93% identity from the database, obtaining a total of 199,720 ASVs. Alpha diversity indices such as Shannon-Wiener (species richness and evenness) and Chao1 (species richness) were calculated using the scikit-bio plugin from QIIME2.

We analysed the microbiome communities using the phyloseq package (version 1.48) for R 4.4.0 software (<https://www.r-project.org/>). We used both the taxonomy table and the ASV table, output from QIIME2, together with metadata table to form the phyloseq object, and used the microbiomeutilities package (version 1.00.17) for the analysis.

2.5. Identification and analysis of the core microbial community

The core community was selected at Family level based on average relative abundance and occurrence frequency of each taxa across all 689 samples. We chose the 200 most dominant families as the initial population. We used 25 different combinations of the occurrence (10%, 30%, 35%, 50% and 70%) and the average relative abundance (0.1%, 0.3%, 0.5%, 0.7% and 1%) across all samples for each family, as described in Cheng et al. (Cheng et al., 2024). Each combination resulted in a community with x number of families, and we chose the community that had all the families in at least 35% of the samples and an average relative abundance of 0.5%, designating it as the 'core community'. We then performed the microbiome analysis of the core community according to the abundances in different tanks per season. The diversity of the samples was assessed using principal coordinate analysis (PCoA) based on Bray-Curtis distance.

To assess the dynamics of the families within the core community, we measured the variance of the relative abundances over time (14 time points) for each family within each tank. First, we assessed whether the means of ASV relative abundances per family, tank and time point differed significantly using the Wilcoxon test or Student's t-test, depending on the

normality and homoscedasticity of the data. We then evaluated which families had more variance over time within each tank, for this we used a linear mixed-effects model (lmer package in R) defining time as a random effect. After a post-hoc test (emmeans package in R) we identified the families that showed the most significant temporal changes as the ‘variable subcommunity’ for each tank. We described its microbial composition at family and genus level. The ‘stable subcommunity’ was defined as the families with variance less than 10% within each tank.

Functional analysis was performed using the *Functional Annotation of Prokaryotic Taxa* (FAPROTAX) with its associated Python script ‘collapse_table.py’ (Louca et al., 2016) and the taxonomy and ASV detected by QIIME2 workflow. The latest version of the FAPROTAX database (1.2.10) was used for annotation. We integrated this data with the metadata to build a phyloseq object, which allow us to identify the functional profiles within the tanks across different seasons. In order to find out the corresponding families and genera that were found in each tank, we used the Faprotax flag --out_groups2records_table, to obtain the specific taxa assigned to each functional group. We then defined the 50 most dominant families and the 150 most dominant genera and matched them with their functional groups. Lastly, we analysed these taxa across different seasons within tanks.

3. Results

3.1. Overview of the depuration tanks diversity

The depuration tanks had different environmental characteristics according to their overall depuration function. In Table 1 we show all of the environmental factors that significantly differed between depuration tanks (first six rows) and that significantly changed over time within individual tanks (last four rows). Overall, P1–Sewage presented the highest number of significantly different environmental factors when compared to the rest of the depuration tanks (both oxidation and denitrification from the two depuration plants). This suggests that the sewage tank experiences more pronounced variations in environmental conditions, likely due to the nature of the waste it processes and the operational differences in its treatment approach. Although there are two oxidation tanks, they showed some differences that may be attributed to the depuration plants they are associated with.

Nearly all environmental factors presented a significant difference over time in all four tanks. This suggests nutrient cycling and changes in biomass, microbial activity and organic

matter processing. Nevertheless, a few environmental factors remained constant over time, such as Cd for P1–Denitrification; Pb, As, Ni and Cd for P1–Oxidation; NO₂ and Pb for P1–Sewage and finally, SO₄, S, Pb and Cd for P2–Oxidation.

Table 1. Number and name of environmental factors that presented a significative difference between tanks and over time ($p < 0.05$, see Methods for statistical analysis description). In grey, comparisons between P1–Sewage and the rest of the tanks.

Tanks	N.	Environmental factors with significative differences
P1-DEN vs. P1-OX	11	N _{tot} , NH ₄ , NO ₃ , N _{org} , SO ₄ , P, Pb, Cu, Zn, COD, TSS
P1-DEN vs. P1-SW	17	pH, N _{tot} , NH ₄ , NO ₂ , NO ₃ , N _{org} , P, Pb, Zn, Cr, As, Ni, Cd, Q _{tot} , COD, SRT, MLSS
P1-DEN vs. P2-OX	14	pH, N _{tot} , NH ₄ , NO ₃ , N _{org} , S, Pb, Cu, Zn, Hg, Q _{tot} , COD, SRT, TSS
P1-OX vs. P1-SW	18	pH, N _{tot} , NH ₄ , NO ₂ , NO ₃ , N _{org} , P, Cu, Zn, Cr, As, Ni, Cd, Q _{tot} , COD, SRT, MLSS, TSS
P1-OX vs. P2-OX	12	pH, N _{tot} , N _{org} , S, P, Cu, Zn, Hg, Q _{tot} , COD, SRT, TSS
P1-SW vs. P2-OX	19	N _{tot} , NH ₄ , NO ₂ , NO ₃ , N _{org} , S, P, Cu, Zn, Cr, Hg, As, Ni, Cd, Q _{tot} , COD, SRT, MLSS, TSS
P1-DEN in time	20	pH, N _{tot} , NH ₄ , NO ₂ , NO ₃ , N _{org} , S, P, Pb, Cu, Zn, Cr, Hg, As, Ni, Q _{tot} , COD, SRT, MLSS, TSS
P1-OX in time	17	pH, N _{tot} , NH ₄ , NO ₂ , NO ₃ , N _{org} , S, P, Cu, Zn, Cr, Hg, Q _{tot} , COD, SRT, MLSS, TSS
P1-SW in time	22	pH, N _{tot} , NH ₄ , NO ₃ , N _{org} , SO ₄ , S, Cl, P, Cu, Zn, Cr, Hg, As, Ni, Cd, Q _{tot} , BOD, COD, SRT, MLSS, TSS
P2-OX in time	19	pH, Temp., OD, N _{tot} , NH ₄ , NO ₂ , NO ₃ , N _{org} , P, Cu, Zn, Cr, Hg, As, Q _{tot} , COD, SRT, MLSS, TSS

Table 1. Environmental factors: Pb, lead; Cu, copper; Zn, zinc; Cr, chromium; Hg, mercury; As, arsenic; Ni, nickel; Cd, cadmium; N_{tot}, total nitrogen; NH₄, ammonium; NO₂, nitrite; NO₃, nitrate; N_{org}, organic nitrogen; Cl, chloride; P, phosphorus; BOD, biological oxygen demand; COD, chemical oxygen demand; Temp, temperature; OD, oxygen demand; pH; SRT, sludge retention time; MLSS, mixed liquor suspended solids; TSS, total suspended solids; Q_{tot}, treatment capacity or flow rate.

We next analysed the microbial diversity associated to each tank and how different they were in terms of hosted microbial community. Overall, the ASV present in the depuration tanks showed significant differences according to their respective process/function. The sewage treatment line clustered independently (in green, Figure 2), while the denitrification and oxidation tanks formed a separate cluster (fuchsia and dark violet, respectively, Figure 2). The second principal coordinate showed a location-dependant differentiation, with Calice positioned at the bottom and Baciacavallo at the top (triangle vs. circle, Figure 2).

Tanks' alpha diversity was measured using Shannon-Wiener and Chao1 indices. When evaluating the diversity taking into account the richness and evenness of the tanks (Shannon-Wiener) and the total number of ASV (Chao1) according to their function, the sewage treatment line had a significantly lower diversity (Table S2 a, Figure S1 a). Instead, the diversity across

sample points had no differences when it came to number of ASVs (Chao1), but according to the richness and evenness (Shannon-Wiener) there was a slight difference in the diversity of the input and middle points in Calice (Table S2 b, Figure S1 b).

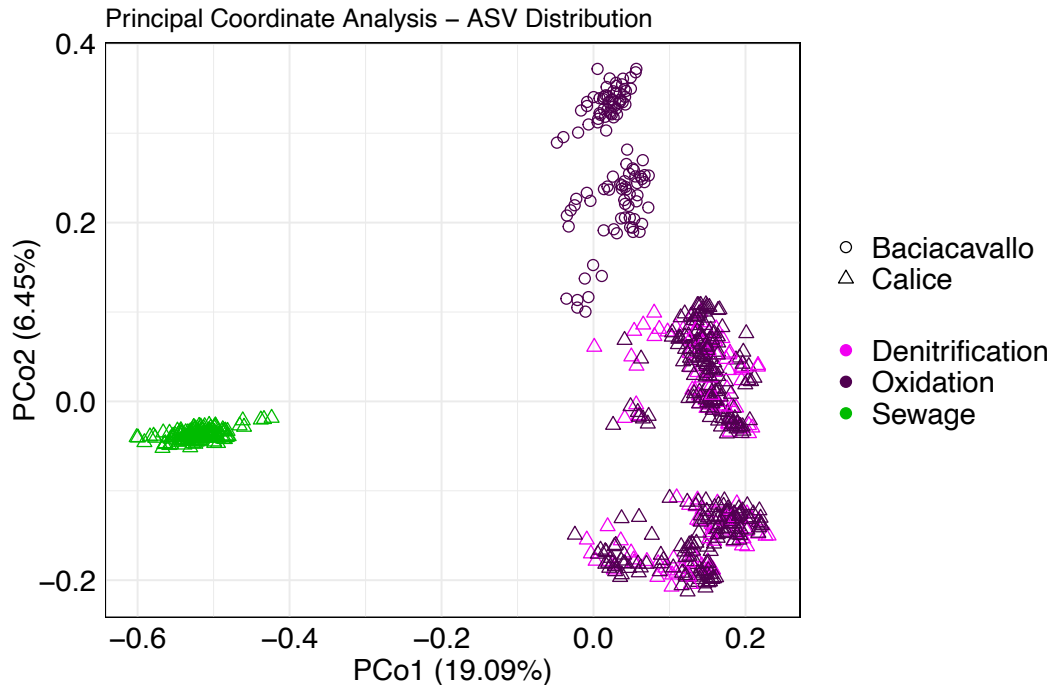


Figure 2. PCoA of the amplicon sequences variant distribution according to the type of depuration tank and the location.

3.2. Core community

QIIME2 pipeline identified a total of 744 different bacterial families, however we chose the 200 most dominant ones and looked for a core community. We used 25 different combinations of family occurrences (10%, 30%, 35%, 50% and 70%) and the average relative abundance (0.1%, 0.3%, 0.5%, 0.7% and 1%). This resulted in 25 different communities, each described by the number of families present (occurrences, Table S3 a) and the average relative abundance of all families that the community represents (Table S3 b). We chose the core community that had an average abundance higher than 0.5% and occurrences higher than 35%. This core community accounted for an average of 76% of the total relative abundance and included a total of 51 families, that belonged to 13 phyla (Figure 3). Our community included common families in depuration-sewage-treatment populations, such as *Comamonadaceae*, *Rhodocyclaceae*, *Saprospiraceae*, *Nitrospiraceae*, *Nitrosomonadaceae* (Begmatov et al., 2022; Chen et al., 2017; Cheng et al., 2024; Chun et al., 2021; Wardi et al., 2023).

In the sewage tank of Calice (P1 – Sewage, Figure 3), we found families with apparent seasonality behaviours. *Rhodocyclaceae*'s relative abundance was higher in winter and spring, and lower in summer and autumn. NS9_marine_group was less abundant in summer than in any other season, contrary to *Anaerolineaceae*, who was more abundant during summer (Figure 3).

The families that belong to the core community had a different relationship with each of the environmental factors. In the denitrification tank, seasonal variation was observed, with winter samples clustering near the RDA center, indicating stability, while spring and autumn samples were more dispersed, reflecting fluctuating conditions. Families like *Hyphomonadaceae*, *Gemmatimonadaceae*, PHOS-HE36, OM190, *Microscillaceae*, *Saprospiraceae*, Candidatus_*Kaiserbacteria* and NS9_marine_group showed positive correlations with NH₄, BOD, N_{tot}, P, MLSS, Zn and As, but negative correlations with Cu, Pb, NO₃, and Q_{tot}. In contrast, *Dojkabacteria*, *Anaerolineaceae* and *Moraxellaceae* correlated positively with Cu and Q_{tot} and negatively with MLSS. *Microtrichaceae* showed a clear positive influence by Pb, while *Chitinophagaceae* and SBR1031 positively correlated with Cr. Families like *Saccharimonadales*, *Steroidobacteraceae*, *Comamonadaceae* and *Caldilineaceae* formed a central cluster, suggesting moderate environmental influence (Figure S2).

Calice's oxidation tank displayed time-dependent patterns, with autumn, summer and winter samples clustering along RDA2, and spring 2022 samples forming two clusters along RDA1. Families like *Dojkabacteria*, *Anaerolineaceae*, and *Moraxellaceae* showed positive associations with Zn, NH₄, N_{tot}, TSS, MLSS, and Q_{tot}, but negative ones with Cu, NO₂ and pH. *Blastocatellaceae*, SBR1031, PHOS-HE36, OLB14 and *Gemmatimonadaceae* correlated positively with BOD, P, SRT, pH and slightly with NO₂, while showing negative associations with Cr, NO₃, MLSS and Q_{tot}. In contrast, Candidatus_*Kaiserbacteria*, AKYH767 and PeM15 were positively influenced by these factors. *Steroidobacteraceae*, *Chitinophagaceae*, *Comamonadaceae*, and *Rhodocyclaceae* clustered centrally, indicating mild environmental correlations (Figure S3).

In the sewage tank, spring samples varied along RDA1, autumn along RDA2, while winter and summer showed less variation. Most microbial families clustered in the center, suggesting low environmental correlation. However, *Chitinophagaceae*, *Microtrichaceae*, NS9_marine_group and *Hyphomonadaceae* were positively influenced by Zn, Cu, Ni, As, Cr and N_{tot}, while negatively influenced by Cl and NO₃. In contrast, *Rhodocyclaceae*, Candidatus_*Kaiserbacteria*, *Anaerolineaceae* and *Phycisphaeraceae* were positively

influenced by Cl and NO₃. *Moraxellaceae* correlated positively with NH₄ and pH but negatively with P, NO₃, Cl and most heavy metals. SRT positively influenced *Candidatus_Kaiserbacteria*, OLB14 and *Rhodocyclaceae*, while *Steroidobacteraceae* and NS9_marine_group were negatively influenced, showing positive correlation with TSS, MLSS and pH (Figure S4).

Finally, Baciacavallo's oxidation tank showed seasonal patterns similar to the sewage tank, with spring samples along RDA1, autumn along RDA2, and winter and summer clustering centrally. *Moraxellaceae* showed a positive correlation with Cu, Zn, COD and TSS, and a negative correlation with NO₂, NH₄, pH and temperature. *Candidatus_Kaiserbacteria* was positively influenced by NO₂, NH₄, pH, Q_{tot}, OD, SRT and temperature, but negatively by Cu, Zn, TSS, MLSS, COD, P and N_{tot}. *Microscillaceae*, *Microtrichaceae*, *Blastocatellaceae* and *Rhodocyclaceae* were positively influenced by As, Cr, NO₃, N_{tot}, OD, SRT, and Q_{tot}, and negatively by BOD, P and temperature. *Anaerolineaceae*, PHOS-HE36, NS9_marine_group and *Dojkabacteria* responded positively to BOD, P, MLSS and temperature, and negatively to NO₃, SRT, Q_{tot}, OD, As and Cr. Families such as OLB14 and *Sphingomonadaceae* showed no significant environmental correlations (Figure S5).

The communities present in the tanks showed location-dependent differentiation, specifically for the ones in the sewage depuration process. Figure 3 shows a decrease in some of the families' relative abundance of P1–Sewage. Specially in families like *Blastocatellaceae*, *Propionibacteriaceae*, *AKYH767*, *Bacteroidetes-vadinHA17*, *Ilumatobacteraceae*, *Intrasporangiaceae* and *Microtrichaceae*. This difference was evident at genus level as well, where genera corresponding to *Microtrichaceae* family had a higher abundance than in P1–Sewage or P2–Oxidation (Figure S6).

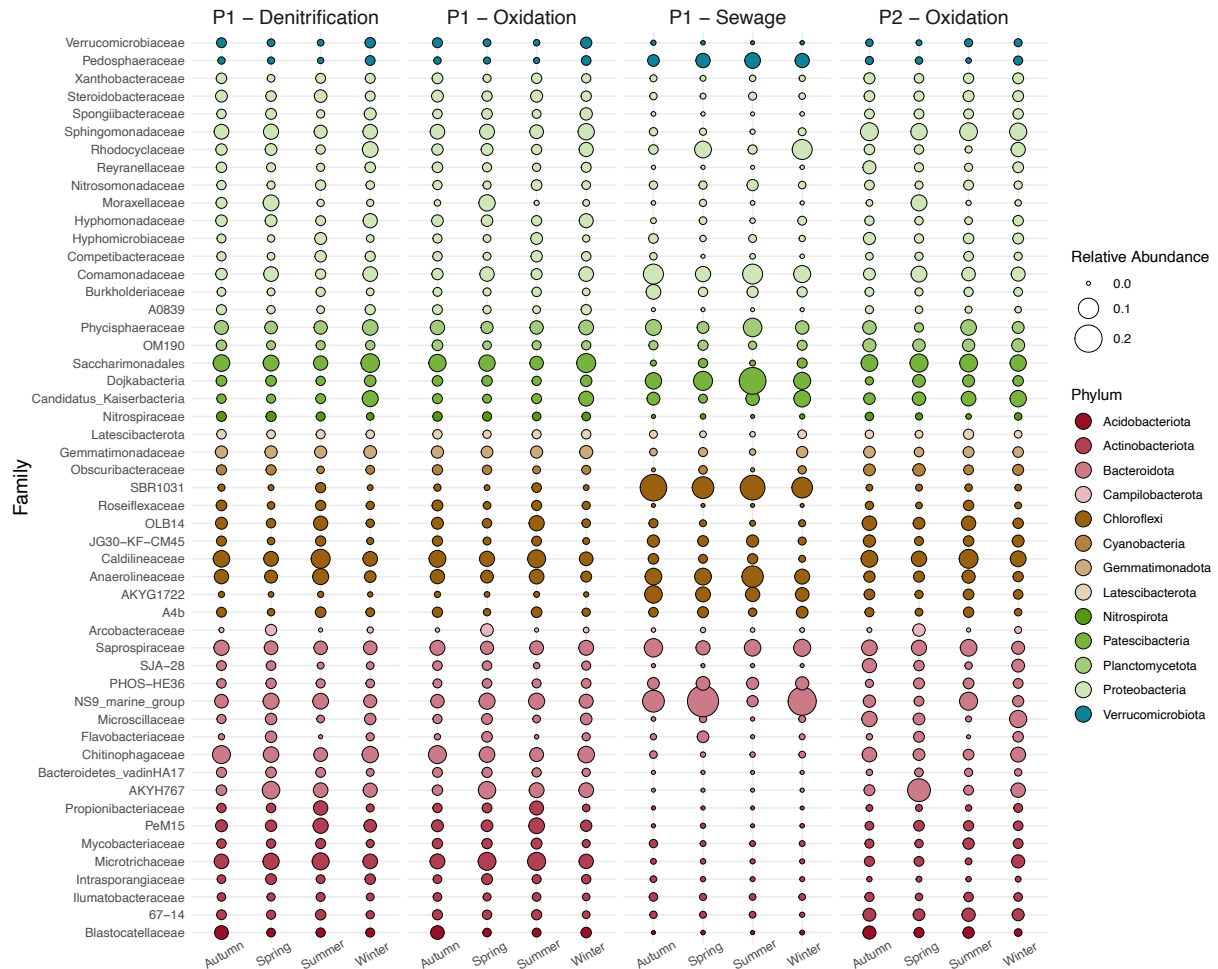


Figure 3. Relative abundance of the 51 families that form our core community. P1 corresponds to the depuration plant located in Calice, while P2, to the one located in Baciacavallo.

Since population dynamics showed a significant variability in time for each family in all tanks (Student's t-test or Wilcoxon Test, $p < 0.001$), we focused on the families that were significantly different from the rest of the population (Linear mixed-effects model, $p < 0.05$), such as *AKYG1722*, *AKYH767*, *Burkholderiaceae*, *Candidatus Kaiserbacteria*, *Dojkabacteria*, *Microscillaceae*, *Microtrichaceae*, NS9 marine group, *PHOS-HE36*, *Roseiflexaceae* and *SBR1031* (red in Figure 4a). All these families conformed the variable community, and Figure 4b shows how its relative abundances change in time. Additionally, there is a stable subcommunity in each of the tanks, this subcommunities were defined as the families whose variance was lower than 10% of the total variance of the community (blue in Figure 4a). Both the stable and the variable communities within tanks are listed in Table S4.

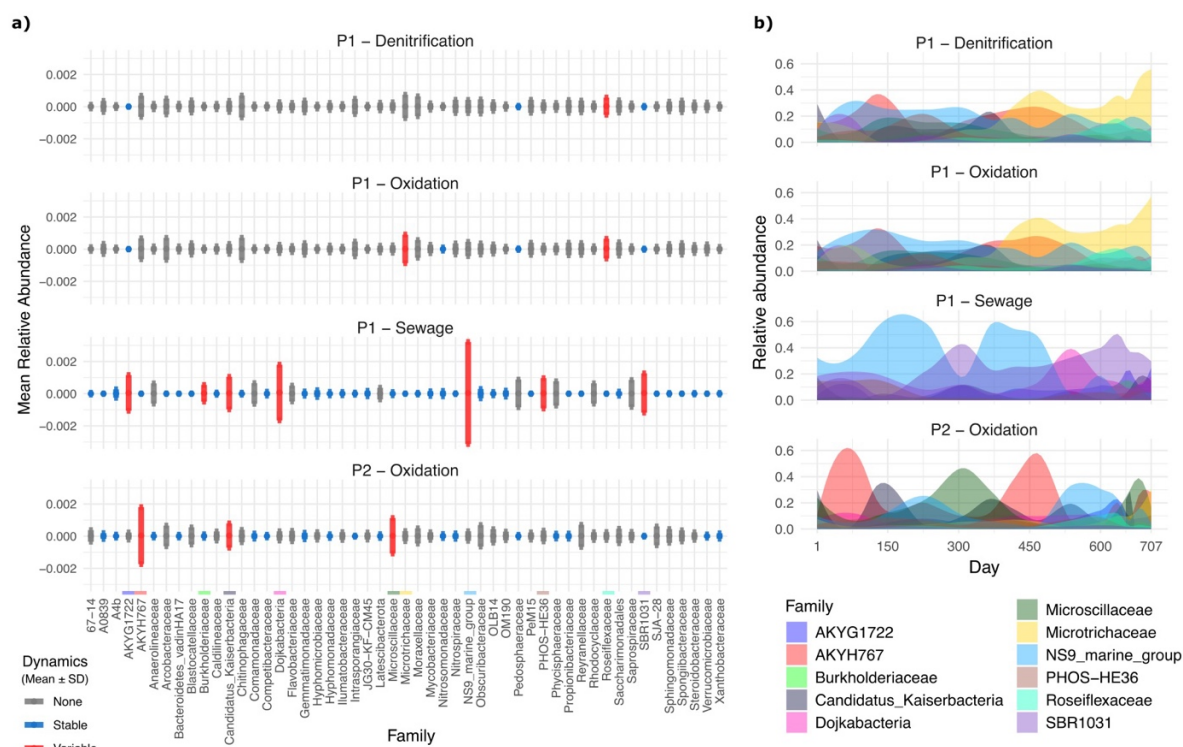


Figure 4. a) Variability of relative abundance of each family across 14 time points, represented by the mean and standard deviation. All the time distributions showed significant changes over time, with significant differences between families. In red, families that had a significant variability over time in contrast to the rest of the families. In blue, families that presented a variance lower than 10% compared to the rest of the community. b) Relative abundance of the variable subcommunity, changing across 14 time points (using the smooth function of R with a degree 7 polynomial).

The core community presented a different functional profile according to the studied tank. The functional profile of sewage tank was slightly different than the other tanks, including, for example, aerobic chemoheterotrophy that resulted more represented in oxidation and nitrification tanks. Together with nitrate reduction, and nitrogen, nitrate and nitrite respiration, aerobic chemoheterotrophy showed a clear seasonal pattern, increasing during winter and spring and decreasing in autumn and summer (Figure 5). This seasonal pattern appeared blurred in the other tanks. We found 51 families and 118 genera that had at least one of the main metabolic functions (Table S5.1 and S5.2, respectively). When analysing the family abundances according to their function across seasons within tanks we found relevant functions such as fermentation, chemoheterotrophy and nitrite respiration (Figure S7).

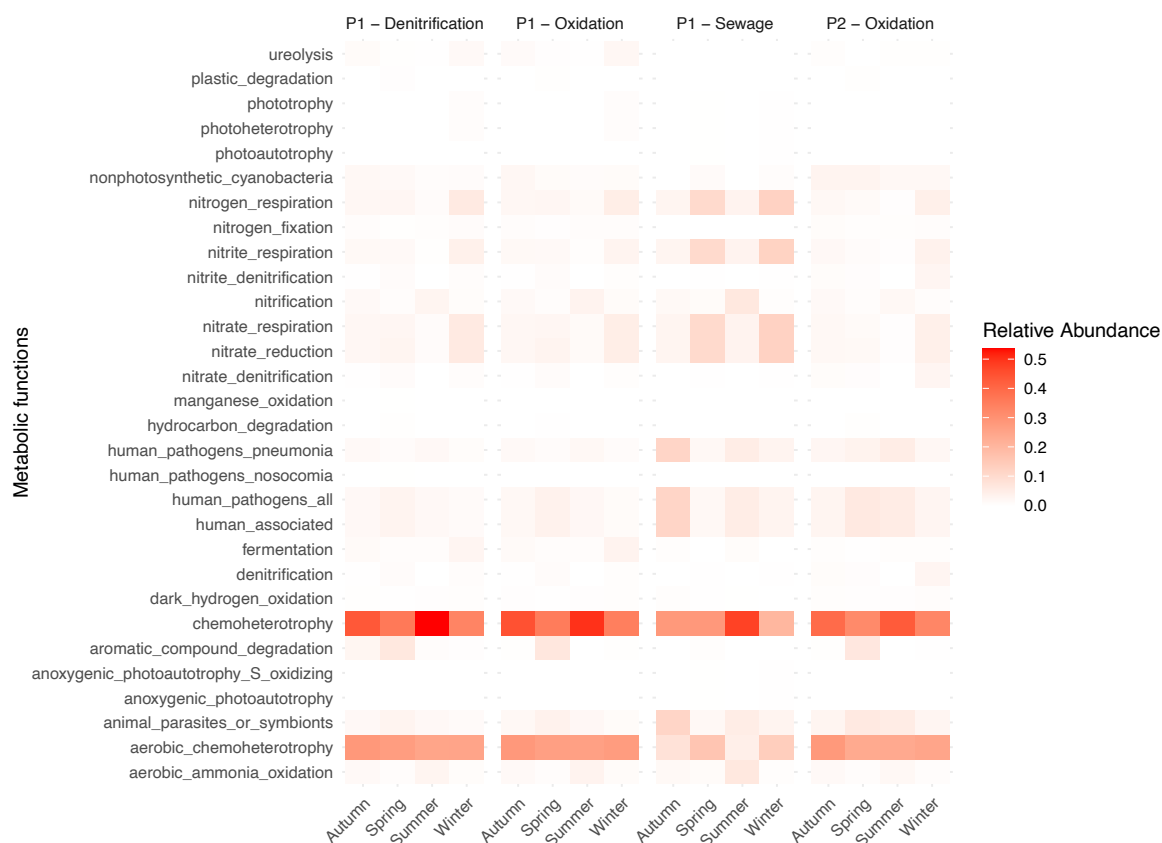


Figure 5. Functional profiles across seasons within tanks.

4. Discussion

The study of wastewater treatment plants is crucial for the upcoming years giving the actual climate crisis. In this study, we analyzed and characterized the bacterial community present in the depuration plant of Calice and Baciacavallo, Prato.

Overall, in a wastewater treatment plant, four groups of bacteria have crucial functions, i.e. as ammonia-oxidizing bacteria (AOB), nitrite-oxidizing bacteria (NOB), phosphate-accumulating bacteria (PAO) and glycogen accumulating organisms (GAO) (Begmatov et al., 2022). In our samples, we found taxa from each of the functional types. At family level, *Nitrosomonadaceae* is known to be AOB, and was indeed found in most of the tanks. We also found *Nitrosomonas* and *Nitrospira* genera to be more abundant in sewage tanks and indeed, it is known that, in this kind of tanks, ammonia-oxidizing bacteria are crucial for its main function, the nitrification process (Bellucci & Curtis, 2011; Kowalchuk & Stephen, 2001).

Taxa belonging to *Gammaproteobacteria* were very abundant in all steps of the process, including families like *Comamonadaceae*, *Steroidobacteraceae*, *Moraxellaceae*, *Sphingobacteriaceae* and *Rhodocyclaceae*. Importantly, most of the taxa from this class enable

the complete oxidation of ammonia into nitrate (W. Qin et al., 2024). The denitrification can take place in the following steps of the depuration, i.e. oxidation and denitrification tanks, and so we found in them nitrite-oxidizing bacteria *Nitrospiraceae*, with a higher abundance than in the sewage depurator.

Families such as *Rhodocyclaceae*, *Microtrichaceae*, *Gemmatimonadaceae*, *Obscuribacteraceae* are known to be phosphate-accumulators (Begmatov et al., 2022; Liu et al., 2023), and genera such as *Thauera*, *Dechloromonas*, *Zooglea*, *Candidatus Microthrix* and *Candidatus Obscuribacter* are mostly present in denitrification and oxidation processes. However, genus *Gemmatimonas* was only found in the sewage tank, suggesting a preference for the community and environmental conditions present in that tank.

When it comes to GAO, we found high abundances of *Competibacteraceae* in denitrification and oxidation tanks, while a very low abundance in the sewage tanks. This might be due to the considerable differences between the other three in terms of chemical factors and abundance variability. One possible reason could be related to the leachate that the sewage tank receives, which despite of being 2% of the treated volume, it represents the 50% of the COD in the plant.

The sampled tanks had different characteristics, such as chemical factors, heavy metal concentrations and bacterial groups. Groups like *Anaerolinaceae* and SBR1031 are anaerobic digesters and have potential as cellulose degrader or anaerobic syntrophs (Xia et al., 2016) and were present especially in denitrification and oxidation process, which suggest a preference for cross-feeding in that particular environment.

5. References

- American Public Health Association. (1998). *Standard Methods for the Examination of Water and Wastewater*. APHA; AWWA; WEF.
- Begmatov, S., Dorofeev, A. G., Kadnikov, V. V., Beletsky, A. V., Pimenov, N. V., Ravin, N. V., & Mardanov, A. V. (2022). The structure of microbial communities of activated sludge of large-scale wastewater treatment plants in the city of Moscow. *Scientific Reports*, 12(1), 3458. <https://doi.org/10.1038/s41598-022-07132-4>

380 Bellucci, M., & Curtis, T. P. (2011). Ammonia-Oxidizing Bacteria in Wastewater. In
 381 *Methods in Enzymology* (Vol. 496, pp. 269–286). Elsevier.
 382 <https://doi.org/10.1016/B978-0-12-386489-5.00011-7>
 383 Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A.,
 384 Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E.,
 385 Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-
 386 Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive,
 387 scalable and extensible microbiome data science using QIIME 2. *Nature*
 388 *Biotechnology*, 37(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
 389 Carles, L., Wullschleger, S., Joss, A., Eggen, R. I. L., Schirmer, K., Schuwirth, N., Stamm,
 390 C., & Tlili, A. (2022). Wastewater microorganisms impact microbial diversity and
 391 important ecological functions of stream periphyton. *Water Research*, 225, 119119.
 392 <https://doi.org/10.1016/j.watres.2022.119119>
 393 Chen, Y., Lan, S., Wang, L., Dong, S., Zhou, H., Tan, Z., & Li, X. (2017). A review: Driving
 394 factors and regulation strategies of microbial community structure and dynamics in
 395 wastewater treatment systems. *Chemosphere*, 174, 173–182.
 396 <https://doi.org/10.1016/j.chemosphere.2017.01.129>
 397 Cheng, Z., Xia, W., McKelvey, S., He, Q., Chen, Y., & Yuan, H. (2024). *Building Predictive*
 398 *Understanding of Microbial Ecology by Bridging Microbial Growth Kinetics and*
 399 *Microbial Population Dynamics*. <https://doi.org/10.1101/2024.07.25.605167>
 400 Chun, S.-J., Cui, Y., Baek, S. H., Ahn, C.-Y., & Oh, H.-M. (2021). Seasonal succession of
 401 microbes in different size-fractions and their modular structures determined by both
 402 macro- and micro-environmental filtering in dynamic coastal waters. *Science of The*
 403 *Total Environment*, 784, 147046. <https://doi.org/10.1016/j.scitotenv.2021.147046>

404 Ju, F., & Zhang, T. (2015). Bacterial assembly and temporal dynamics in activated sludge of
 405 a full-scale municipal wastewater treatment plant. *The ISME Journal*, 9(3), 683–695.
 406 <https://doi.org/10.1038/ismej.2014.162>

407 Kowalchuk, G. A., & Stephen, J. R. (2001). Ammonia-Oxidizing Bacteria: A Model for
 408 Molecular Microbial Ecology. *Annual Review of Microbiology*, 55(1), 485–529.
 409 <https://doi.org/10.1146/annurev.micro.55.1.485>

410 Kumar, V., Sehgal, R., & Gupta, R. (2023). Microbes and wastewater treatment. In
 411 *Development in Wastewater Treatment Research and Processes* (pp. 239–255).
 412 Elsevier. <https://doi.org/10.1016/B978-0-323-88505-8.00010-3>

413 Liu, X., Nie, Y., & Wu, X.-L. (2023). Predicting microbial community compositions in
 414 wastewater treatment plants using artificial neural networks. *Microbiome*, 11(1), 93.
 415 <https://doi.org/10.1186/s40168-023-01519-9>

416 Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the
 417 global ocean microbiome. *Science*, 353(6305), 1272–1277.
 418 <https://doi.org/10.1126/science.aaf4507>

419 Qin, H., Ji, B., Zhang, S., & Kong, Z. (2018). Study on the bacterial and archaeal community
 420 structure and diversity of activated sludge from three wastewater treatment plants.
 421 *Marine Pollution Bulletin*, 135, 801–807.
 422 <https://doi.org/10.1016/j.marpolbul.2018.08.010>

423 Qin, W., Wei, S. P., Zheng, Y., Choi, E., Li, X., Johnston, J., Wan, X., Abrahamson, B.,
 424 Flinkstrom, Z., Wang, B., Li, H., Hou, L., Tao, Q., Chlouber, W. W., Sun, X., Wells,
 425 M., Ngo, L., Hunt, K. A., Urakawa, H., ... Winkler, M.-K. H. (2024). Ammonia-
 426 oxidizing bacteria and archaea exhibit differential nitrogen source preferences. *Nature*
 427 *Microbiology*, 9(2), 524–536. <https://doi.org/10.1038/s41564-023-01593-7>

428 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner,
 429 F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data
 430 processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596.
 431 <https://doi.org/10.1093/nar/gks1219>

432 Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M. C., Michael, I., & Fatta-
 433 Kassinos, D. (2013). Urban wastewater treatment plants as hotspots for antibiotic
 434 resistant bacteria and genes spread into the environment: A review. *Science of The*
 435 *Total Environment*, *447*, 345–360. <https://doi.org/10.1016/j.scitotenv.2013.01.032>

436 Wardi, M., Slimani, N., Alla, A. A., & Belmouden, A. (2023). First study of the effect of
 437 wastewater treatment on microbial biodiversity at three wastewater treatment plants in
 438 Agadir, Morocco, using 16S rRNA sequencing. *Environmental Pollution*, *337*,
 439 122528. <https://doi.org/10.1016/j.envpol.2023.122528>

440 Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M. R., Li, Z.,
 441 Van Nostrand, J. D., Ling, F., Xiao, N., Zhang, Y., Vierheilig, J., Wells, G. F., Yang,
 442 Y., Deng, Y., Tu, Q., Wang, A., ... Zhou, J. (2019). Global diversity and
 443 biogeography of bacterial communities in wastewater treatment plants. *Nature*
 444 *Microbiology*, *4*(7), 1183–1195. <https://doi.org/10.1038/s41564-019-0426-5>

445 Xia, Y., Wang, Y., Wang, Y., Chin, F. Y. L., & Zhang, T. (2016). Cellular adhesiveness and
 446 cellulolytic capacity in Anaerolineae revealed by omics-based genome interpretation.
 447 *Biotechnology for Biofuels*, *9*(1), 111. <https://doi.org/10.1186/s13068-016-0524-z>

448 Zeng, Q., & An, S. (2021). Identifying the Biogeographic Patterns of Rare and Abundant
 449 Bacterial Communities Using Different Primer Sets on the Loess Plateau.
 450 *Microorganisms*, *9*(1), 139. <https://doi.org/10.3390/microorganisms9010139>

451 Zhang, T., Shao, M.-F., & Ye, L. (2012). 454 Pyrosequencing reveals bacterial diversity of
452 activated sludge from 14 sewage treatment plants. *The ISME Journal*, 6(6), 1137–
453 1147. <https://doi.org/10.1038/ismej.2011.188>
454

3.1.2 Metagenomes: How and who has an impact on marine protected areas?

In this study we show the methodology of both targeted and shotgun sequencing analysis (metagenomic survey) of several marine samples along the gradient from a port and a protected marine area. Describing the impact one could have on the other.

The impact that ports could have on their surroundings is evident due to high levels of pollution received from land-based sources, marine traffic and port infrastructure. This issue becomes of matter for marine ecosystems, especially for marine protected areas, such as Secche della Meloria Island, located approximately 3 miles offshore from Livorno harbour, Italy. Since the presence of pollutants as organic and inorganic chemicals in harbour areas have been well documented, as well as with anthropogenic activities like pesticides, personal care products and pharmaceuticals, to name a few, it is reasonable to be aware of the impact all these could have on the microbial communities living in the nearby areas.

In this study, we aimed to describe and analyze the bacterial groups present in the sea floor sediments, in particular the differential occurrences of bacteria along the sampling points and the presence of antibiotic resistance genes. These results shed light on the impact human activities may have on marine microbes with potential consequences for human and environment health.

We used two different pipelines for the analyses, one for 16S rRNA sequences (targeted) and one for metagenomic assembled genomes (shotgun). For the first one, we used DADA2 to filter low quality reads and denoised the sequences, removing the putative chimeras. The taxonomic classification was performed using the Decipher package and the Silva database. The diversity and microbiomes were analyzed using R packages iNEXT and ANACOM-BC2. For metagenomes, instead, we trimmed the reads with Sickle tool and then normalized the coverage to accelerate the metagenomes' assembly with SPAdes. We obtained the cover-

age of the contigs by mapping our reads into them using Bowtie2 and Samtools, and then proceeded to the construction of the metagenomic bins with MetaBAT2. In order to be sure of the quality of the metagenomic assembled genomes, we used CheckM to choose a threshold of completeness and contamination, and used those chosen bins to reduce gene redundancy with MMseqs2. Finally, we used the trimmed reads as input for Kraken2 to have the taxonomic profiles of each metagenomic sample.

Both the results from the analysis of metagenome assembled genomes and 16S rRNA amplicon sequences showed specific trends of bacterial groups distributed along the sampled gradient. Phyla such as NB1-j and Acidobacteriota showed an increment from the harbour to the marine protected area. Desulfobulbales, on the other hand, showed the opposite trend. As expected, the harbour area showed a massive presence of microbial taxa related to urban areas, such as the human pathogen Chlamydiae.

Additionally, we evaluate the presence of genes conferring resistance to antibiotics. A total of 111 genes were found in all the samples, however, the genes for the potential resistance to vancomycin were among the most abundant in all samples but the intermediate ones. Additionally, we did not find *Amycolaptosis orientalis* in the samples, a bacterium responsible for the production of the vancomycin. This suggested that the presence of vancomycin resistance genes in the marine protected area is not linked to the presence of vancomycin-producing bacteria.

This study highlighted the importance of microbiological surveillance of marine sediments to assess the impact of human activities, and how the analysis of both metagenomic assembled genomes and amplicon sequence variants may give a broad understanding of environmental conditions.

Metagenomic surveys shows a widespread diffusion of antibiotic resistance genes in a transect from urbanized to marine protected area

Authors: Tania Alonso-Vásquez¹, Camilla Fagorzi¹, Alessio Mengoni¹, Matteo Oliva²,
Duccio Cavalieri¹, Carlo Pretti^{2,3}, Lisa Cangioli¹, Giovanni Bacci^{1,*}, Alberto Ugolini¹

¹, Department of Biology, University of Florence, Sesto Fiorentino, 50019, Florence, Italy

², Interuniversity Center of Marine Biology and Applied Ecology (CIBM) "G. Bacci," Livorno, Italy

³, Department of Veterinary Science, University of Pisa, San Piero a Grado, Pisa, Italy

*, Author for correspondence: giovanni.bacci@unifi.it

Abstract

Ports are hot spots of pollution; they receive pollution from land-based sources, marine traffic and port infrastructures. Marine ecosystems of nearby areas can be strongly affected by pollution from port-related activities. Here, we investigated the microbiomes present in sea floor sediments along a transect from the harbour of Livorno (Central Italy) to a nearby marine protected area. Results of 16S rRNA amplicon sequencing and metagenome assembled genomes (MAGs) analyses indicated the presence of different trends of specific bacterial groups (e.g. phyla NB1-j, Acidobacteriota and Desulfobulbales)

along the transect, correlating with the measured pollution levels. Human pathogenic bacteria and antibiotic resistance genes (ARGs) were also found. These results demonstrate a pervasive impact of human port activities and highlight the importance of microbiological surveillance of marine sediments, which may constitute a reservoir of ARGs and pathogenic bacteria.

Keywords: Anthropic impact, marine microbiome, vancomycin resistance

1. Introduction

Pollution of coastal environments is posing serious concerns for its detrimental effects on the marine food web and human health¹⁻⁴. Harbour areas are typical hot spots in terms of pollution status^{5,6}. These areas receive pollution from land-based sources, marine traffic and port infrastructures. The presence of pollutants as organic (e.g., polycyclic aromatic hydrocarbons, PAHs) and inorganic (e.g., trace metals) chemicals in harbour areas has been well documented (for instance see⁷⁻¹⁰). Moreover, harbour areas normally insist over densely urbanized areas which can contribute additional pollution due to anthropogenic activities, as for instance pesticides, other persistent organic pollutants, emerging pollutants (e.g. personal care products, pharmaceuticals, plasticizers, microplastics, flame retardants, etc.) pathogenic microorganisms, antibiotic resistant bacteria, and antibiotic resistant genes (ARGs)¹⁰⁻¹². Additionally, polluted harbour areas may constitute a potential pollution source for the neighboring coastal areas, which can include areas deserving special attention in terms of biodiversity preservation, as those included in the list of Marine Protected Areas (MPAs)¹³.

43 Surveillance and impact assessment of anthropic activities of heavily urbanized coastal
44 areas, paying special attention to harbours, is then a primary activity in relation to either
45 human health or environmental protection. In the Mediterranean basin, densely
46 populated harbour areas may be very close to MPAs, and consequently their effects in
47 terms of spread of pollutants can influence biodiversity hotspots^{14,15}.

48 Bacteria influence the fate and distribution of contaminants in marine sediments by
49 providing metabolic activities able to degrade/alter pollutants (e.g. hydrocarbon
50 degradation, metal speciation). However, they are also affected by the presence of
51 contaminants which, in turn, may operate selective pressure toward specific microbial
52 taxa and metabolic functions. The massive use of antibiotics and the consequent spread
53 of antibiotic resistance genes in the marine (water and sediment) microbiomes are thus
54 creating concerns especially in relation to potential human pathogens¹⁶⁻¹⁸.

55 In the last ten years, the analysis of the impact of anthropic activities on marine sediments
56 has largely increased thanks to the application of metagenomic and metabarcoding
57 methods (for instance see¹⁹⁻²⁷). These investigations revealed correlations between the
58 microbiota of marine sediments and pollutants such as trace metals and persistent
59 organic compounds²⁸. These changes in microbiota likely impact sedimentary niche
60 changes, but also are associated with spread of pathogens and genes related to human
61 pathogens, as antibiotic resistance genes (ARGs)^{17,29}.

62 Here, we aimed at investigating the microbiomes present in sea floor sediments along a
63 transect from the harbour of Livorno (Central Italy) to the nearby MPA of Secche della
64 Meloria Island, paying attention to both the differential occurrence of bacteria taxa and
65 the presence of ARGs to assess the impact human activities may have on sea floor
66 sediment microbiomes with potential consequences for human health and environment.

2. Materials and Methods

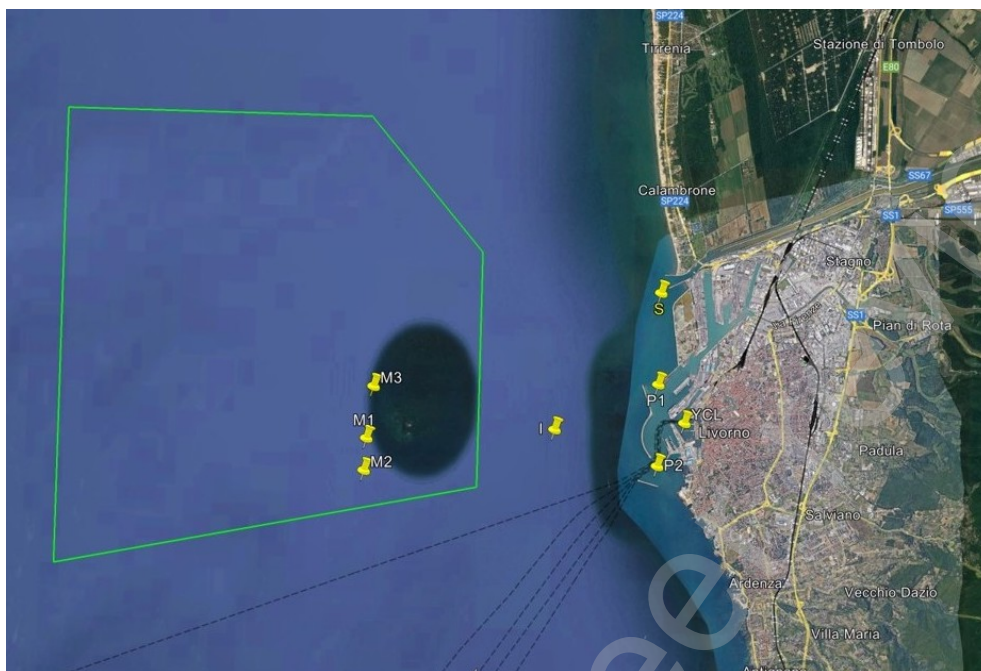
2.1. Sampling

We selected two study sites, Livorno harbour and the Marine Protected Area (MPA) named “Secche della Meloria”, a marine shoal located about 3 miles offshore from Livorno harbour. A total of 4 sampling sites were selected, comprehending a total of 8 sampling points.

During the summer of 2018, sediment samples were collected using a manually operated 5 L Van Veen grab. Sampling sites were selected as follows: 1) three sampling points inside the harbour mouth (harbour Yachting Club, YCL; harbour mouth north, P1; harbour mouth south, P2), 2) one sampling point between the harbour and MPA (Intermediate, I), 3) one sampling point in the proximity of the harbour (mouth of channel Scolmatore, S), and 4) three sampling points in the Marine Protected Area (M1, M2, M3) (Figure 1, Table S1 and S2.1). The grab was able to collect sediment from the top 20 cm layer. All samples were stored at -20°C.

Approximately 100 g aliquots from each sample were individually collected in HDPE jars and stored at -18 ± 1 °C for chemical analysis. Particle size distributions were determined by dry sieving the sediments at 60°C for 24 hours, using a series of sieves with varying mesh sizes based on the Udden-Wentworth classification³⁰. This classification categorizes the sediment into fractions as follows: <63 µm for pelitic fraction, 63–125 µm for very fine sand, 125–250 µm for fine sand, 250–500 µm for medium sand, 500–1000 µm for coarse sand, 1000–2000 µm for very coarse sand, and >2000 µm for gravel. All sediment parameters are shown as percentages of total dry sample weight. To simplify the data presentation, we grouped particle sizes into gravel (>2000 µm), sand (63–2000 µm), and pelitic fraction (<63 µm) (Table S2.1). We used MAR-3 CTD multiprobe

91 model to measure the main physico-chemical parameters of water (Idromarambiente,
92 Italy) (Table S2.2).



93
94 **Figure 1.** Sampling points. Geographical map of samplings, Harbour Yachting Club (YCL),
95 Harbour north and south mouth (P1 and P2), mouth of Scolmatore channel (S),
96 Intermediate (I) and MPA "Secche della Meloria" Island (M1, M2, M3). The green polygon
97 represents the limits of MPA "Secche della Meloria".

98 2.2. Chemical analysis

99 2.2.1. Inorganics (trace elements)

100 Prior to trace element analysis, the samples were dried at 40°C for 48 hours and sieved
101 through a 2 mm mesh sieve. The dried samples were then analyzed for Hg levels using a
102 DMA-80 Tri Cell (Milestone, FKV) following the EPA 7473 method³¹. For the analysis of
103 other elements, sediment samples were digested using the EPA 3051A³² method with an
104 Ethos 1 Microwave Digestion System (Milestone, FKV). Al, Cr, Cu, Fe, Ni, V, and Zn
105 analyses were conducted using an ICP Varian 720-ES (Agilent) in accordance with EPA

6010D method³³, while As, Cd and Pb were quantified using an AAS Varian SpectrAA240Z following EPA 7010 method³⁴. The limit of quantification (LOQ) for all elements, except Cr (0.01 mg/Kg DW) and Hg (0.005 mg/Kg DW), was 1 mg/Kg dry weight (DW). The average recovery rate was >93%. For values below the LOQ, data were adjusted by assigning a value equal to LOQ/2 as the method by Cohen and Ryan³⁵.

2.2.2. Organics (PAHs, PCBs, hydrocarbons C>12 and C<12)

Polychlorinated biphenyls (PCBs) standard stock solutions were prepared in acetone, polycyclic aromatic hydrocarbons (PAHs) in toluene and hydrocarbons in hexane, with concentrations ranging from 40–400 mg/L. All the stock solutions were stored at –18°C.

The extractions of the 16 PAHs and PCBs were carried out following the method by Ben Salem et al.³⁶, using the QuEChERS extraction method. An aliquot of the sample was mixed with ultrapure water and surrogate spiking standards (nitrobenzene-d5, 2-fluorobiphenyl, and p-terphenyl-d14) in a 50 mL polypropylene tube, then acetonitrile was added. The QuEChERS extraction salts kit was then added and energetically shaken. The centrifuged extract was treated with a QuEChERS purification salts kit, filtered, and concentrated into a 2 mL vial for GC analysis. The extracts were analyzed following the EPA 8270E method³⁷ for PAHs and PCBs, and ISO 16703 method³⁸ for C>12 hydrocarbons. C<12 hydrocarbons were analyzed using EPA 5021A 2014³⁹ and EPA 8015C 2007⁴⁰ methods.

The PAHs and PCBs analyses were conducted using a GC Trace 1300 (Thermo Scientific) coupled to a triple quadrupole mass spectrometer TSQ Duo with an electron impact ionization source (Thermo Scientific) and a TG-5MS column. C>12 and C<12 hydrocarbons analyses were performed with a GC Trace 1300, coupled to a TriPlus RSH

autosampler and flame ionization detector (Thermo Scientific) using a TG-5SilMS column. In all analyses, highly pure helium was used as carrier gas. The LOQ of PAHs and PCBs was 1 µg/Kg DW. For hydrocarbons C>12, the LOQ was 5 mg/Kg DW, and for hydrocarbons C<12, it was 1.5 µg/Kg fresh weight (FW). The average recovery rate was consistently above 92%. For values <LOQ, data were adjusted by assigning a value equal to LOQ/2, as recommended by Cohen and Ryan (1989)³⁵.

2.3. DNA extraction and sequencing

Environmental DNA was extracted from sediment samples using a DNeasy PowerSoil Pro (Qiagen). Nanodrop, fluorometry (Qubit, Life Technologies, USA) and spectrophotometric readings quantification were performed to determine quantities and quality of extracted DNA. Amplification of 16S rRNA gene was performed on V3-V4 region of bacterial 16S rRNA gene was amplified via PCR using specific primers (806R and 515F), as previously reported⁴¹. Amplicon sequencing was performed by IGA Technology service (Italy) on an Illumina MiSeq apparatus.

Metagenomic libraries preparation and DNA sequencing were performed following the standard pipelines for Illumina Novaseq6000, PE150 sequencing. Sequencing was performed by IGA Technology service (Italy). Raw sequence data reported in this study have been deposited in the National Center for Biotechnology Information (accession number: PRJNA1067305).

2.4. Amplicon sequence variant reconstruction

Microbial amplicon sequence variants (ASVs) were inferred by using the DADA2 pipeline version 1.22.0⁴² after PCR primer removal (cutadapt version 4.1⁴³). Briefly: low quality reads were filtered out by using the “filterAndTrim” function from DADA2 pipeline and

an error rate model was generated (“learnErrors” function). High-quality reads were denoised by using the “dada” function and putative chimeric sequences were identified and removed by using the “removeBimeraDenovo” function. The DECIPHER package was used to taxonomically classify every variant inferred against the Silva reference database (version 138.1)⁴⁴. Alpha diversity together with bacterial distribution across different sites were inspected by using the R packages iNEXT (version 3.0.0)⁴⁵ and ANCOM-BC2 (version 2.2.0)⁴⁶. Additional information about amplicon reconstruction pipeline and statistical analyses were reported in supplementary materials, section 2.

2.5. Metagenome analyses

Illumina reads were trimmed (Sickle, version 1.33⁴⁷) and their coverage was normalized using BBNorm to accelerate the metagenomes’ assembly (SPAdes, version 3.14.0⁴⁸). Metagenomes were mapped back to the assembled contigs by using Bowtie2 (version 2.4.5)⁴⁹ and binned into “metagenomic bins” with MetaBAT (version 2.2.15)⁵⁰. We assessed the completeness and contamination of each bin with CheckM (version 1.1.3)⁵¹ using the “lineage_wf” workflow. To reduce gene redundancy in metagenomes we used the easy-cluster pipeline implemented in the MMseqs2 suite (Many-against-Many sequence searching; version: 13.45111)⁵².

Using the trimmed reads, we performed the taxonomic characterization using Kraken2 (version 2.0.9-beta)⁵³ with Minikraken v2 as a reference database, and the microbiome analysis using the “Phyloseq” R package (version 4.0.3)⁵⁴. Alpha and beta diversity were calculated with the “iNEXT” package in R (see Figure S2 for the pipeline overview).

In order to evaluate the presence of ARGs in each sample, we used the genes previously obtained with MMSeq2 and The Comprehensive Antibiotic Resistance Database

(CARD)⁵⁵. We then performed the functional enrichment analysis by considering a cut-off of 10% of the abundance of the ARGs in each drug class and analyzed the genes present there according to the sampling site. Additional information about amplicon reconstruction pipeline and statistical analyses were reported in supplementary materials, section 3.

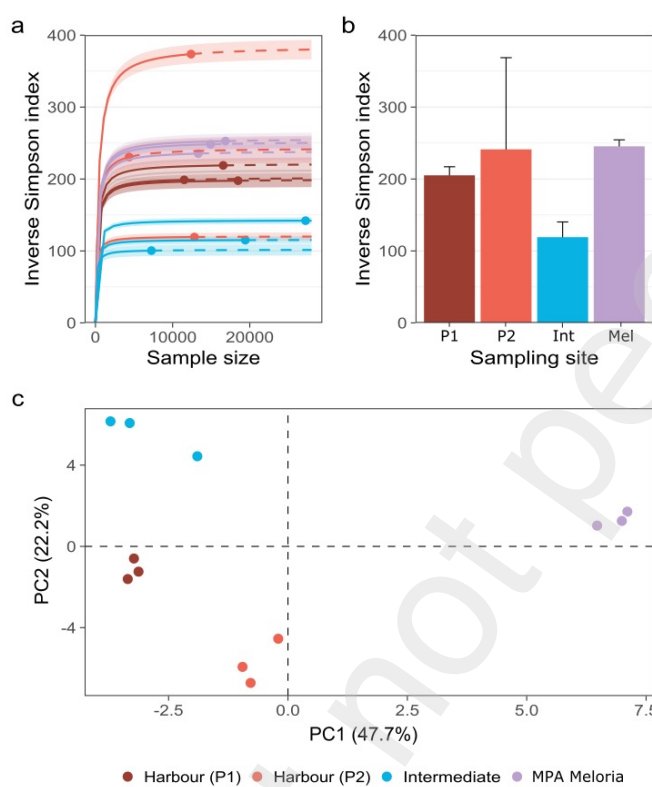
3. Results

3.1. Sea floor microbiota show a gradient related to anthropic pressure

From the 2472 ASVs profiled, 2435 (98.5%) were classified as bacteria whereas only 37 (1.5%) were classified as Archaea. When analyzing the effect of sampling sites (harbour, harbour P1 and P2, intermediate and Meloria) on the alpha diversity, no differences were found, suggesting no influence of the anthropic gradient (i.e. origin) (ANOVA, $F=2.43$, $df=3$, $p=0.14$; Table S3, Figure 2a and b, and S3). However, bacterial communities were differentially distributed along the anthropic gradient, showing the Meloria samples apart from the Intermediate and Harbour samples (ANOVA-like permutation test, $F=10.35$, $df=3$, $p<0.002$; Figure 2c, Table S4).

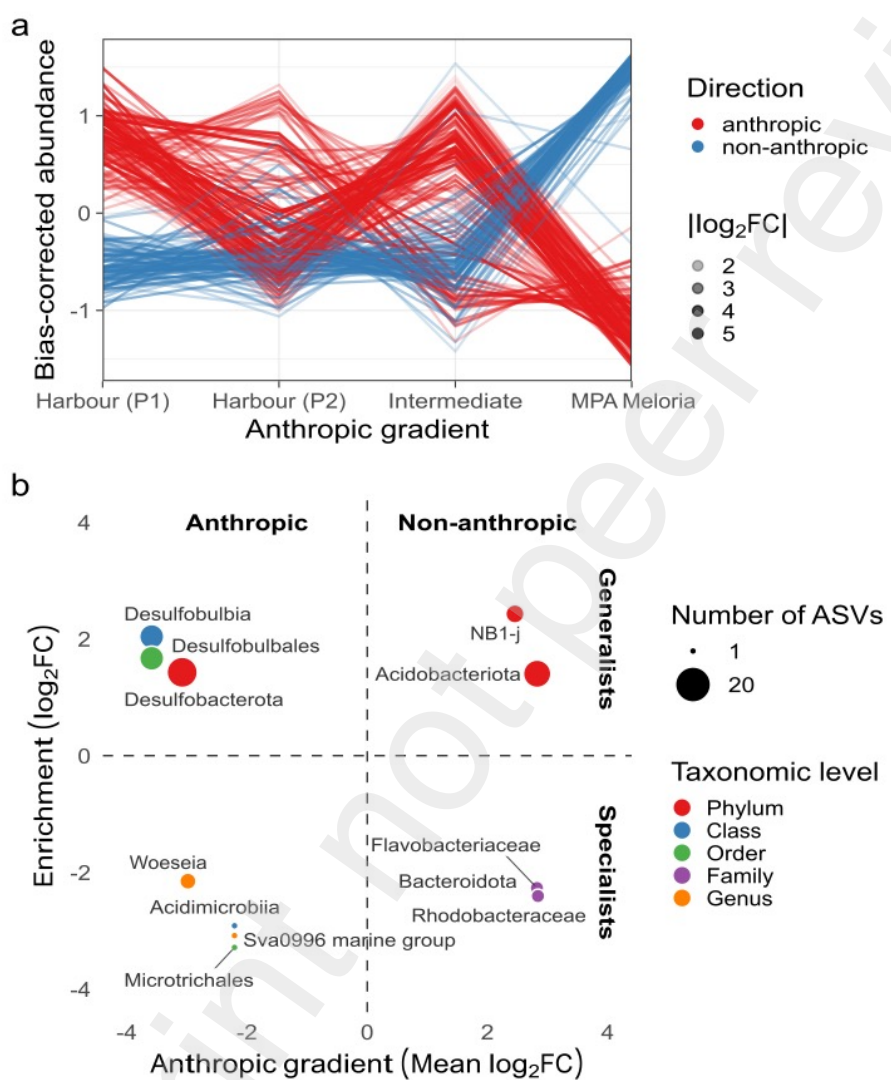
In order to analyze the impact of anthropic activities on microbial distributions, we assessed the amount of ASVs and their change according to the sampling site. We identified 197 (36.68%) ASVs present in the sediments that were influenced by the anthropic gradient: 121 increased along the gradient whereas 76 decreased (Figure 3a). We were able to detect a set of taxa which were enriched/depleted along the anthropic gradient (Figure 3b, Table S5). Phyla NB1-j and Acidobacteriota increased from the harbour to Meloria and were also enriched in number (non-anthropic generalists). On the other hand, the order Desulfobulbales (together with its higher taxonomic levels) increased from the Meloria to the harbour and was also enriched (anthropic generalists).

199 Other taxa increased along the two transects but were instead depleted. The phylum
200 Bacteroidota, together with the families Rhodobacteraceae and Flavobacteraceae
201 reported a higher abundance in the MPA Meloria area but were depleted in number (non-
202 anthropic specialists). Similarly, the marine group Sva-0996 and Woeseia genera,
203 together with order Microtrichales and its immediate higher taxonomic level, were
204 depleted in number but highly abundant in the harbour (anthropic specialists).



205 **Figure 2.** Microbial diversity estimation. **a)** The inverse Simpson index was reported with
206 increasing sampling effort for all samples. Interpolated, extrapolated, and observed
207 diversity were reported by using a solid line, a dotted line, and solid points. Diversity
208 across increasing sampling effort was reported together with the 98% confidence limits
209 of diversity (iNEXT package). **b)** Observed diversity with error bars representing the
210 standard deviation of the observed values of diversity in each group. Colors corresponds
211

212 to sample site. **c)** Redundancy analysis (RDA) of microbial assemblages representing all
 213 samples included in the study. All samples are reported by using a different color scale
 214 according to their sampling site. The variance explained by each axis is reported between
 215 parentheses next to the axis label.



216

217 **Figure 3.** Effect of anthropic gradient on microbial distribution. Microbial patterns along
 218 anthropic gradient were reported in panel **a** by using different colors: microbial ASVs
 219 showing an increasing pattern from MPA Meloria to the harbour were reported in red,
 220 whereas microbial ASVs reporting an increasing pattern from the harbour to the MPA

Meloria were reported in blue. Lines were colored according to the $\log_2$ fold-change values between the first sampling point in the harbour (P1) and the MPA Meloria. The same value was averaged and used in panel **b** as a measure of the effect of the anthropic gradient (x axis). Enriched taxa were reported by using the $\log_2$ fold-change values (y axis) and colored according to their taxonomic resolution. The size of each taxa corresponds to the number of ASVs significantly increasing/decreasing along the anthropic gradient whereas taxa names were reported next to each point.

3.2. Pollution affects taxonomies and functions of metagenomes

3.2.1. Biodiversity and taxonomy of sample microbiomes

To perform taxonomic identification and the following analysis of ARGs, we assembled metagenomic reads into contigs and classified contigs into potential taxonomies by binning to tetranucleotide frequencies (Supplemental Table S6). The number of bins from Meloria samples was lower with respect to those obtained from all other sites (Table S7.1), and from 43 bins in total, only 13 presented a completeness higher than 50% and a contamination lower than 10%, i.e. only bins that had at least half of the taxonomically assigned genome and a maximum of 10% of some other genome (Table S7.2).

Rarefaction curves reached a plateau for all samples, which indicates a satisfactory survey of the bacterial diversity in our metagenomic datasets (Figure S4). Hence, alpha diversity was estimated (Figure 4a and b, Table S8). Inverse Simpson index, used as a measure of alpha diversity, indicated a slightly higher diversity in samples from Scolmatore and the lowest values for Meloria samples. However, this difference was not statistically significant.

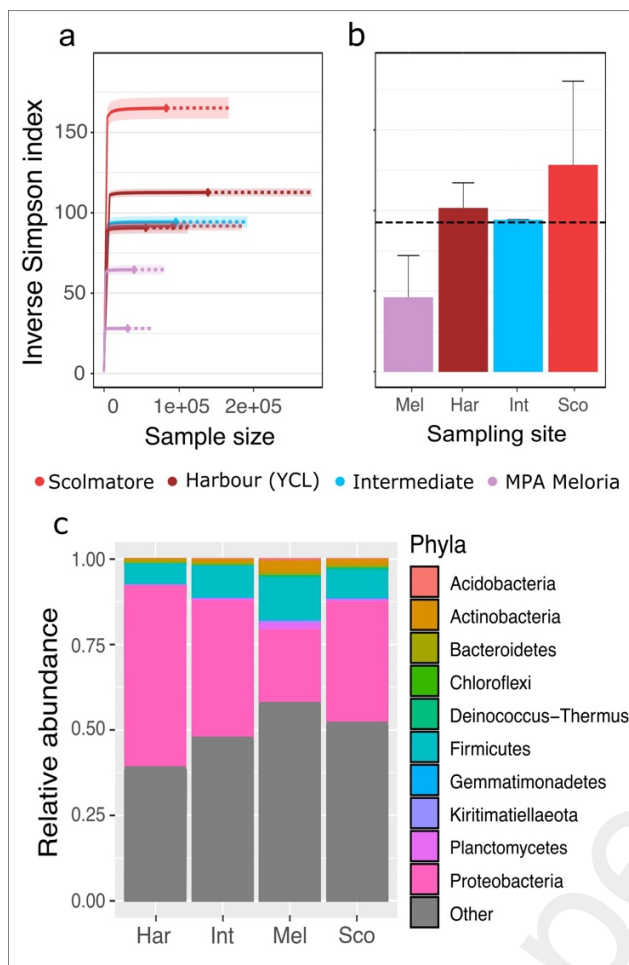


Figure 4. Microbial diversity estimation for the metagenome samples. **a)** Alpha diversity measures for all sampling was reported using the inverse Simpson index. Interpolated, extrapolated and observed diversity were reported by using a solid line, a dotted line and solid points. **b)** Observed diversity with error bars representing the standard deviation of the observed values of diversity in each group. Colors corresponds to sample origins. **c)** Relative abundances of dominant bacterial phyla. Phyla with a minor representation are shown in *Other*. Har, Harbour (YCL); Int, Intermediate; Mel, MPA Meloria; Sco, Scolmatore.

Concerning the taxonomic composition of the microbiomes, on the overall samples, 36 different phyla were identified (Figure 4c). Five phyla were dominant for all samples,

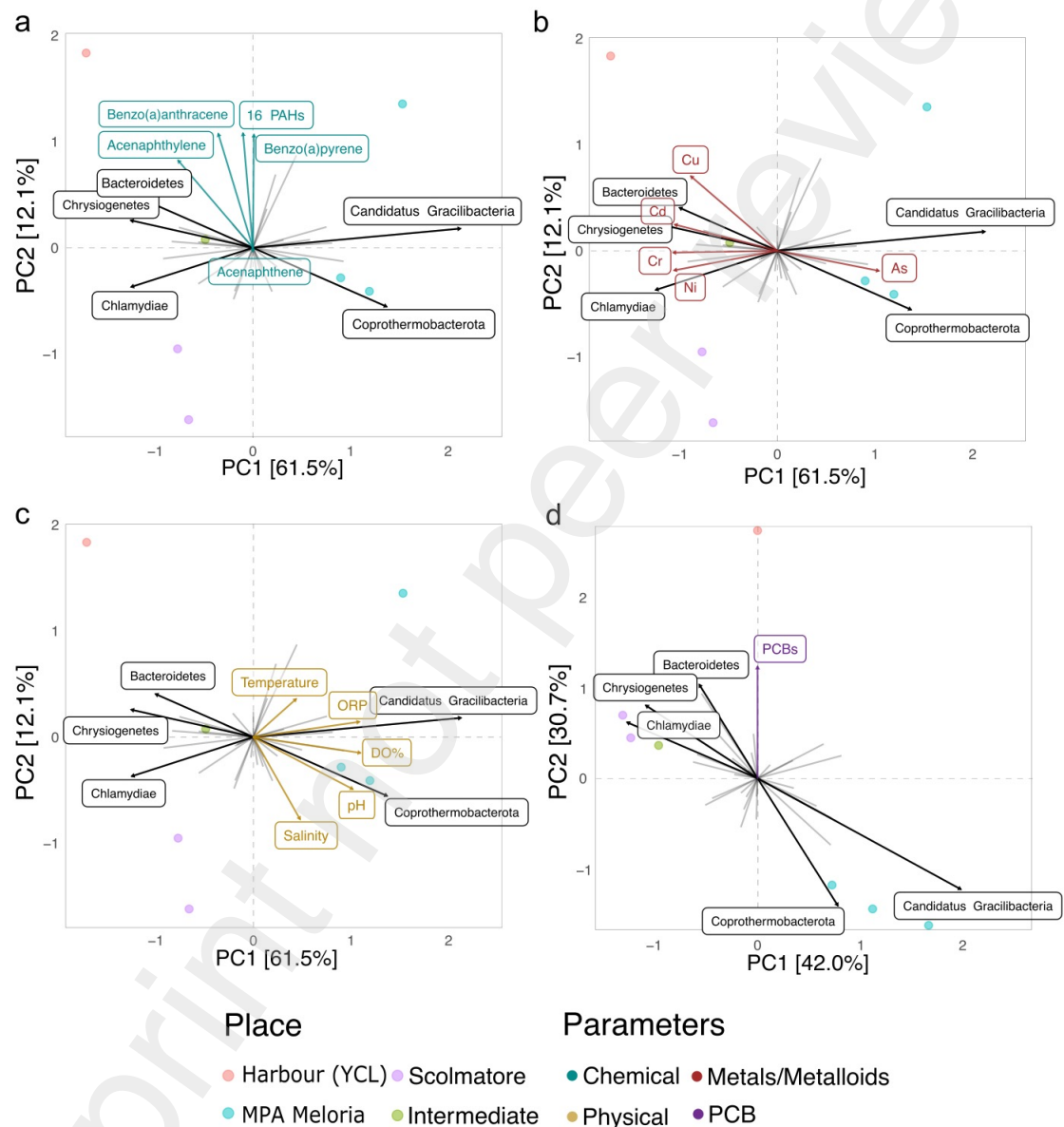
254 accounting for ~95% of dataset (Proteobacteria, Actinobacteria, Bacteroidetes,
255 Firmicutes and Planctomycetes).

256 The abundances of each phylum were slightly different depending on the sampling point
257 (Scolmatore, Harbour (YCL), Intermediate or MPA Meloria), the main differences being
258 found when comparing MPA Meloria to Harbour sediment samples (Proteobacteria:
259 44.3% vs. 76.4%; Actinobacteria: 27.7% vs. 5.1%; Firmicutes: 16.5% vs. 7.9%;
260 Bacteroidetes: 2.5% vs. 7.9%) (Figure S5).

261 **3.2.2. Effect of the environmental parameters and pollution on microbiome** 262 **diversity**

263 We analyzed the environmental parameters, including chemical, physical, trace metals
264 and metalloids and organic pollutants such as PAH, PCB and hydrocarbons (Table S2.2)
265 present in the points of sampling and the influence they can have on the microbiome
266 composition. We observed that the chemical, physical and trace metals and metalloids
267 parameters explained 61.5% of the variation of the samples, while the organic pollutants
268 explain the 42% of it (Figure 5). It appeared from this ordination analysis that different
269 bacterial groups explain the differences among microbiomes from the sampled
270 sediments. Bacteria belonging to Coprothermobacterota and Candidatus Gracilibacteria
271 are discriminating the microbiome of Meloria sample points. The chemicals that describe
272 the distribution of bacteria in each sample are benzo(a)pyrene, benzo(a)anthracene,
273 acenaphthene, acenaphthylene and 16 polycyclic aromatic hydrocarbons (PAHs). All of
274 them have a positive relationship with each other, except for acenaphthene, which has a
275 negative relationship with the other chemicals and weakly drives the variation of the
276 community (Figure 5a). The difference between the microbiome found in MPA Meloria
277 island and the rest of the samples can be explained by the set of trace metals and

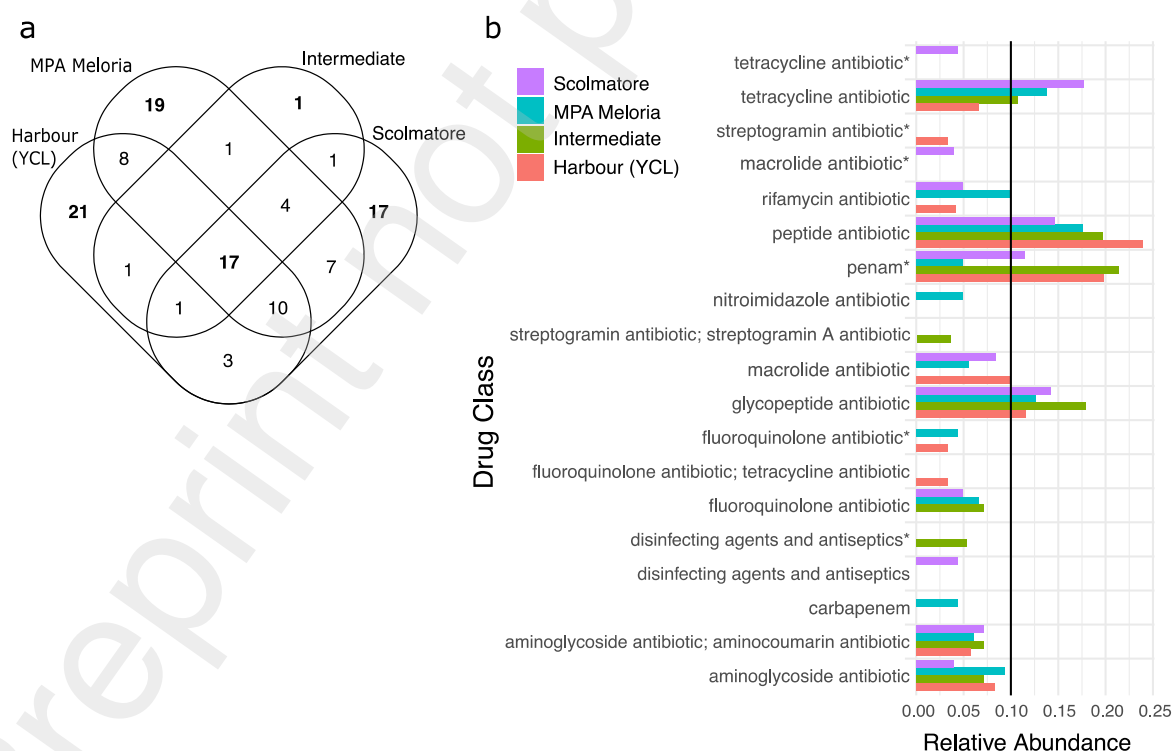
278 metalloids (Figure 5b), but also by factors such as temperature, oxidation/reduction
 279 potential, dissolved oxygen saturation, pH and salinity (Figure 5c). Finally, the sum of all
 280 polychlorinated biphenyls strongly describes the composition of Harbour samples and is
 281 negatively related with the Meloria samples (Figure 5b).



282
 283 **Figure 5.** Constrained principal component analysis for harbour, Scolmatore,
 284 intermediate and MPA Meloria samples. It is shown **a)** chemical, **b)** trace of metals and
 285 metalloids, **c)** physical and **d)** PCB parameters.

3.2.3. Identification of Antibiotic resistance genes (ARGs) in the microbiomes

We finally searched for the presence of ARGs for each sample. We found 61 ARGs in Harbour samples, 66 for Meloria samples, 26 for Intermediate samples and 60 ARGs for Scolmatore samples. There were found 17 ARGs present in all four sampling sites (Figure 6a) and quite a high number of ARGs were present exclusively in the Harbour samples (21 ARGs) and in the Meloria samples (19 ARGs; Table 1). The Intermediate and Scolmatore samples harbored 1 and 17 exclusive ARGs, respectively. Considering the type of ARGs identified and the possible resistance to the antimicrobial they may determine, we found several beta-lactamases encoding genes, which can determine resistance to beta-lactamic antibiotics and various other classes, including systems conferring resistance to antibiotics targeting RNA polymerase and ribosome functionality.



299 **Figure 6. a)** Venn diagram describing the antibiotic resistance genes identified
300 exclusively in each one of the sampling sites, along with the genes in common. **b)** Drug
301 classes accounting for, at least, 70% of all the identified ARGs per sample. The chosen cut-
302 off (0.1) was reported with a black line. Since a single class of antibiotics may contain
303 more than one type of drug (i.e. some genes have a broader antibiotic resistance profile),
304 the complete pattern of resistance was reported with an asterisk as defined below.

305 *Disinfecting agents and antiseptics*: fluoroquinolone antibiotic, disinfecting agents and antiseptics.
306 *Fluoroquinolone antibiotic*: fluoroquinolone antibiotic, cephalosporin, glycylicline, penam, tetracycline
307 antibiotic, rifamycin antibiotic, phenicol antibiotic, disinfecting agents and antiseptics. *Penam*: macrolide
308 antibiotic, fluoroquinolone antibiotic, penam, tetracycline antibiotic. *Streptogramin antibiotic*: macrolide
309 antibiotic, lincosamide antibiotic, streptogramin antibiotic, streptogramin A antibiotic, streptogramin B
310 antibiotic. *Tetracycline antibiotic*: macrolide antibiotic, tetracycline antibiotic, disinfecting agents and
311 antiseptics. *Macrolide antibiotic*: macrolide antibiotic, fluoroquinolone antibiotic, penam.

Table 1. Name of ARGs families present exclusively in MPA Meloria samples (19). It is shown the drug class they belong to.

Antibiotic Resistance Gene Family	Drug class
ole glycosyltransferase	macrolide antibiotic
PFM-like beta-lactamase	carbapenem
AIM beta-lactamase	cephalosporin; cephamycin; penam
msr-type ABC-F protein	macrolide antibiotic; streptogramin antibiotic
ampC-type beta-lactamase	cephalosporin; penam
ECM beta-lactamase	carbapenem
APH(6)	aminoglycoside antibiotic
16S rRNA methyltransferase (G1405)	aminoglycoside antibiotic
helicase-like RNA polymerase protection protein	rifamycin antibiotic
EC beta-lactamase	cephalosporin
Edeine acetyltransferase	peptide antibiotic; polyamine antibiotic
ANT(3'')	aminoglycoside antibiotic
PDC beta-lactamase	monobactam; carbapenem; cephalosporin
Bleomycin resistant protein	glycopeptide antibiotic
Cfr 23S ribosomal RNA methyltransferase	lincosamide antibiotic; streptogramin antibiotic; streptogramin A antibiotic; oxazolidinone antibiotic; phenicol antibiotic; pleuromutilin antibiotic
AAC(3); AAC(6')	aminoglycoside antibiotic
vanY; glycopeptide resistance gene cluster	glycopeptide antibiotic
PST beta-lactamase	carbapenem
vanJ membrane protein	glycopeptide antibiotic

312

313 From the drug classes that were analyzed, we considered *peptide antibiotic*, *penam**,
314 *glycopeptide antibiotic* and *tetracycline antibiotic* for sampling sites (using a cut-off = 0.1)
315 (Figure 6b). We analyzed the ARGs present in the resulted drug classes (Table S9). The
316 number of ARGs present in *Glycopeptide antibiotic* class was statistically different
317 between Meloria and Intermediate samples (Wilcoxon test, $p < 0.05$), but not for the other
318 sampling sites. In the case of *Peptide antibiotic* class, we only found that Intermediate and
319 Harbour samples were statistically different (Wilcoxon test, $p < 0.05$).

320 In the glycopeptide drug class we identified a high number of genes resistant to
321 vancomycin, specifically, from the 78 incidences all sampling sites, 22 were vancomycin
322 resistant genes in MPA Meloria, 14 in Harbour samples, 32 in Scolmatore samples and 10
323 in the intermediate point. Overall, we identified *vanR*, *vanT*, *vanH*, *vanJ*, *vanS*, *vanW* and
324 *vanY* genes, being the intermediate point the location where the lower amount and
325 variety of vancomycin resistance genes were found, with just *vanR*, *vanT* and *vanH*. This
326 suggests a preferential mobility of genes or of microbial taxa carrying vancomycin
327 resistance genes towards both extremes, the port and the MPA Meloria.

328 Among the ARGs present in the peptide drug class, we identified the presence of
329 daptomycin and defensin resistance genes, specifically *cls*, *liaS*, *pgsA* and *mprF*. However,
330 the amount of them was similar in all sampling points. However, ATP-binding cassette
331 (ABC) antibiotic efflux pump, General Bacterial Porin with reduced permeability to
332 peptide antibiotics and *pmr* phosphoethanolamine transferase genes were also present
333 in all the sampling points, showing a decay in the intermediate origin.

The tetracyclin and penam* drug classes presented a more stable distribution of ARGs for the four different origin samples.

4. Discussion

The One Health concept is promoting an integrated, multidisciplinary approach to balance and optimize the health of people and the environment⁵⁶. One Health is focusing on food and water safety, nutrition, the control of zoonoses, pollution management, and antimicrobial resistance⁵⁷. The evaluation of the degree and extent of the impact of human activities on the environment is then a central point in the policies for environmental protection and health. For instance, the massive spread of antibiotic resistance is becoming an increasing hot topic, expecting antibiotic resistance to be one the main cause of death worldwide within the next years⁵⁸⁻⁶⁰. Transmission paths of ARGs between human, environmental, agricultural and clinical reservoirs have been clearly demonstrated⁶¹⁻⁶⁴. Marine sediments and biofilms may constitute a reservoir of ARGs⁶⁵, however relatively few studies have been conducted for assessing presence of signature of impact of anthropic activities (including presence of ARGs) in marine sediments along well defined transects^{17,66,67}.

Here, we reported results from an analysis of microbial communities inhabiting marine sediments along a transect of human contamination due to port activities, in Livorno harbour, Italy. This location was chosen as a model of human pressure of human activities toward nearby Marine Protected Areas (MPA). The port of Livorno is a multi-purpose port of call of national importance. It is the main Italian port for forestry products, new cars and one of the most important for Ro-Ro traffic. The port lies at the southern end of an arched sandy beach about 60 km long. The most important rivers flowing within this physiographic unit are Magra, Serchio and Arno, the latter being the largest Tuscan river,

358 with an average annual flow of ca. 110 m³/s, and serving the densely urbanized areas of
359 Florence and Pisa⁶⁸. The mouth of the river Arno is located approximately 13 km to the
360 north of the port Livorno. Close to the port, the Scolmatore (spillaway) Channel,
361 constitutes an important body of contaminated freshwater. It is a channel of the Arno
362 river, 28.3 km long which ends approximately 3 km north to the Livorno port area. The
363 Meloria Island Marine Protected Area is located cs. 10 km in front of the port area. The
364 location of the MPA Meloria and the harbour of Livorno nearby create unique
365 opportunities to investigate along a transect from the port to the MPA the impact of
366 contamination due to human activities on the port/coastal area, with respect to the MPA.

367 Results obtained from both 16S rRNA amplicon sequencing and metagenomic analyses in
368 the microbiomes along the transect harbour of Livorno- MPA Meloria Island indicated the
369 presence of bacterial groups increasing along the gradient (e.g. phyla NB1-j and
370 Acidobacteriota), while other followed the opposite trend (the order Desulfobulbales).
371 These taxa may be related to the differential impact on harbour activities. Indeed,
372 Desulfobulbales were found more abundant in marine sediments rich in heavy-metals
373 and hydrocarbons⁶⁹, suggesting they could have functional relevance in marine
374 sediments that display high concentrations of such contaminants. NB1-j and
375 Acidobacteriota were already shown to be negatively correlated with marine sediment
376 contamination by heavy metals (Cd)^{70,71}. Similarly, *Woeseia* was found to be particularly
377 abundant in hydrocarbon polluted marine sediments⁷².

378 Although the small sampling size, these and more human pathogenic bacteria were
379 identified, having the Scolmatore area with the highest number of ARGs, followed by the
380 MPA Meloria, Harbour area and Intermediate sampling point with the lowest number.
381 Regarding the specific signature of human-associated bacteria, sediments of the

382 Scolmatore area showed a massive presence of microbial taxa which can be related to
383 urban areas. This is not surprising since the Scolmatore area is in direct contact with all
384 human-related pollutants present in the Arno river and collected in the Scolmatore
385 channel. In particular, Chlamydiae, a group of human pathogenic bacteria⁷³, were highly
386 present in the Scolmatore area, witnessing a strong human-related contamination in the
387 area.

388 The impact of urban area was also strongly supported by the metagenomic analysis. In
389 fact, the analysis of ARGs in MAGs derived from the metagenomic dataset identified a list
390 of 111 genes conferring resistance to 93 different antibiotics. The genes for the potential
391 resistance to vancomycin were among the most abundant. The proportion of such genes
392 over total ARGs decreased from the Scolmatore and Harbour to the intermediate point,
393 but were still very abundant in the sediment of the MPA of Meloria. Vancomycin is a
394 member of glycopeptide antibiotics and is considered the last resort for the treatment of
395 severe infections caused by bacterial pathogens such as *Enterococcus species*,
396 *Staphylococcus aureus*, and *Clostridium difficile*. Resistance to vancomycin arises from
397 various types of mechanisms, comprising multi-enzyme pathways or modification of cell
398 wall permeability^{74,75}. Genes *vanR*, *vanT*, *vanH*, *vanJ*, *vanS*, *vanW* and *vanY* were detected.
399 The gene *vanR*, along with *vanS*, activate *van* genes' expression in response to
400 extracellular glycopeptide antibiotics⁷⁶. The gene *vanT* encodes for a *Vibrio anguillarum*
401 LuxR-like transcriptional regulator, although is also found in *Streptococcus agalactiae*
402 and *Streptococcus anginosus* as part of *vanG* cluster^{77,78}. It was found that in *Enterococcus*
403 *gallinarum*, *vanT* encodes the VanT membrane bound serine racemase, which produces
404 D-Ser^{79,80}. The genes *vanH*, *vanA* and *vanX* promote the action of an alternate cell wall
405 biosynthesis pathway, leading to a mechanism of high-level resistance to vancomycin in

406 enterococci⁸¹. The gene *vanJ*, present in the actinobacterium *Streptomyces coelicolor*,
 407 confers resistance to the glycopeptide antibiotics teicoplanin and teicoplanin-like,
 408 compounds that are closely related to vancomycin. It encodes a membrane protein
 409 oriented with its C-terminal active site exposed to the extracytoplasmic space⁸². The gene
 410 *vanW* is present in the *vanB* gene cluster, also known as *vanB2*⁸³ and it is found in
 411 *Enterococcus faecium* isolates, harboring the transposone Tn5382⁸⁴. The gene *vanY*
 412 encodes a D,D-carboxypeptidase normally found in resistant enterococci⁸⁵. This gene is
 413 found within the *vanA* gene cluster for the lipoglycopeptide antibiotic resistance, present
 414 in *E. faecium* and *Streptococcus pneumoniae*⁸⁶, which is associated with the transposon
 415 Tn1546 and has been credited with the dissemination of *vanA* among *E. faecium* and other
 416 enterococci^{87,88}. We found in the metagenomes 10 species of *Enterococcus*, 40 of
 417 *Streptococcus* and 110 species of *Streptomyces* genus (Table S10). All the species
 418 mentioned above having one of the genes conferring resistance to vancomycin are
 419 present in the samples, except *E. gallinarum*. Consequently, from our findings we are not
 420 able to demonstrate a human-related (i.e. clinical) origin of vancomycin-resistance genes
 421 identified, considering that, at least for *vanJ*, this gene is present in nonpathogenic
 422 bacterial species and can be a normal part of the environmental microbiota. Additionally,
 423 we did not find *Amycolaptosis orientalis* in the samples. This bacterium is responsible for
 424 the production of the vancomycin⁸¹, suggesting that the presence of vancomycin
 425 resistance genes in Meloria samples is not linked to the presence of vancomycin-
 426 producing bacteria. Interestingly, we found four species of the genus *Desulfitobacterium*:
 427 *D. dehalogenans*, *D. dichloroeliminans*, *D. hafniense* and *D. metallireducens*. These are
 428 anaerobic bacteria often isolated in environments contaminated by halogenated organic
 429 compounds, which may bring attention towards the presence and possible
 430 decontamination activities of this group of bacteria in the sampled sediments.

5. Acknowledgements

This work was supported by Interreg-Maritime Italy-France IMPACT EU-funded project (PC Interreg V-A IFM 2014–2020, Prot. ISMAR n. 0002269).

6. Author Contribution Statement

Tania Alonso-Vásquez: Formal Analysis, Data Curation, Writing - Original Draft, Visualization. **Camilla Fagorzi:** Investigation, Resources. **Alessio Mengoni:** Writing - Original Draft, Conceptualization. **Matteo Oliva:** Conceptualization. **Duccio Cavalieri:** Supervision. **Carlo Pretti:** Conceptualization, Writing - Original Draft, Funding acquisition. **Lisa Cangioli:** Supervision. **Giovanni Bacci:** Conceptualization, Formal Analysis, Data Curation, Writing - Original Draft, Visualization. **Alberto Ugolini:** Writing - Original Draft, Supervision.

7. Glossary

ARGs = Antibiotic Resistance Genes; CARD = The Comprehensive Antibiotic Resistance Database; MAGs = Metagenome Assembled Genomes; PAHs = Polycyclic Aromatic Hydrocarbons; PCBs = Polychlorinated Biphenyls; MPAs = Marine Protected Areas; HDPE = High Density Polyethylene; DMA = Direct Mercury Analysis; EPA = Environmental Protection Agency; GC = Gas Chromatography; MS = Mass Spectrometer; LOQ = Limit of Quantification; DW = Dry Weight; FW = Fresh Weight; ASV = Amplicon Sequence Variants.

Sampling areas:

YCL = Harbour Yachting Club; P1 = Harbour Mouth North; P2 = Harbour Mouth South; I = Intermediate; S = Mouth of Channel Scolmatore; M1, M2, M3 = Marine Protected Area “Secche della Meloria” Island.

454 References

- 455 1. Halpern, B. S. *et al.* A Global Map of Human Impact on Marine Ecosystems. *Science* **319**,
 456 948–952 (2008).
- 457 2. Raffaelli, D., H., S. Intertidal ecology. *J. Mar. Biol. Ass.* **77**, 286–286 (1997).
- 458 3. Shahidul Islam, Md. & Tanaka, M. Impacts of pollution on coastal and marine
 459 ecosystems including coastal and marine fisheries and approach for management: a
 460 review and synthesis. *Marine Pollution Bulletin* **48**, 624–649 (2004).
- 461 4. McLachlan, A. & Brown A: *The Ecology of Sandy Shores*. (Elsevier, 2006).
- 462 5. Kanellopoulos, T. D. *et al.* Trace element contamination status of surface marine
 463 sediments of Greece: an assessment based on two decades (2001–2021) of data.
 464 *Environ Sci Pollut Res* **29**, 45171–45189 (2022).
- 465 6. Tamburini, E. *et al.* Impacts of Anthropogenic Pollutants on Benthic Prokaryotic
 466 Communities in Mediterranean Touristic Ports. *Front. Microbiol.* **11**, 1234 (2020).
- 467 7. Castrec, J. *et al.* Active and passive biomonitoring of trace elements, polycyclic aromatic
 468 hydrocarbons, and polychlorinated biphenyls in small Mediterranean harbours. *Marine*
 469 *Pollution Bulletin* **187**, 114578 (2023).
- 470 8. Dai, C. *et al.* Review on the contamination and remediation of polycyclic aromatic
 471 hydrocarbons (PAHs) in coastal soil and sediments. *Environmental Research* **205**, 112423
 472 (2022).
- 473 9. Lepland, A. *et al.* Sedimentation and chronology of heavy metal pollution in Oslo harbor,
 474 Norway. *Marine Pollution Bulletin* **60**, 1512–1522 (2010).
- 475 10. Lipizer, M. *et al.* Trace metals and polycyclic aromatic hydrocarbons in the Eastern
 476 Mediterranean sediments: Concentration ranges as a tool for quality control of large
 477 data collections. *Marine Pollution Bulletin* **185**, 114181 (2022).

- 478 11. Malins, D. C. *et al.* Toxic chemicals in sediments and biota from a creosote-polluted
479 harbor: relationships with hepatic neoplasms and other hepatic lesions in English sole (
480 *Parophrys vetulus*). *Carcinogenesis* **6**, 1463–1469 (1985).
- 481 12. Zhang, W., Ki, J.-S. & Qian, P.-Y. Microbial diversity in polluted harbor sediments I:
482 Bacterial community assessment based on four clone libraries of 16S rDNA. *Estuarine,*
483 *Coastal and Shelf Science* **76**, 668–681 (2008).
- 484 13. Horta E Costa, B. *et al.* A regulation-based classification system for Marine Protected
485 Areas (MPAs). *Marine Policy* **72**, 192–198 (2016).
- 486 14. De Oliveira Júnior, J. G. C., Campos-Silva, J. V., Santos, D. T. V., Ladle, R. J. & Da Silva
487 Batista, V. Quantifying anthropogenic threats affecting Marine Protected Areas in
488 developing countries. *Journal of Environmental Management* **279**, 111614 (2021).
- 489 15. Nunes, B. Z., Moreira, L. B., Xu, E. G. & Castro, Í. B. A global snapshot of microplastic
490 contamination in sediments and biota of marine protected areas. *Science of The Total*
491 *Environment* **865**, 161293 (2023).
- 492 16. Berglund, B. Environmental dissemination of antibiotic resistance genes and correlation
493 to anthropogenic contamination with antibiotics. *Infection Ecology & Epidemiology* **5**,
494 28564 (2015).
- 495 17. Chen, J., McIlroy, S. E., Archana, A., Baker, D. M. & Panagiotou, G. A pollution gradient
496 contributes to the taxonomic, functional, and resistome diversity of microbial
497 communities in marine sediments. *Microbiome* **7**, 104 (2019).
- 498 18. Czatzkowska, M., Wolak, I., Harnisz, M. & Korzeniewska, E. Impact of Anthropogenic
499 Activities on the Dissemination of ARGs in the Environment—A Review. *IJERPH* **19**,
500 12853 (2022).

- 501 19. Chen, B. *et al.* Metagenomic Profiles of Antibiotic Resistance Genes (ARGs) between
502 Human Impacted Estuary and Deep Ocean Sediments. *Environ. Sci. Technol.* **47**, 12753–
503 12760 (2013).
- 504 20. Imchen, M. *et al.* Comparative mangrove metagenome reveals global prevalence of
505 heavy metals and antibiotic resistome across different ecosystems. *Sci Rep* **8**, 11187
506 (2018).
- 507 21. Imchen, M. & Kumavath, R. Metagenomic insights into the antibiotic resistome of
508 mangrove sediments and their association to socioeconomic status. *Environmental*
509 *Pollution* **268**, 115795 (2021).
- 510 22. Mustafa, G. A., Abd-Elgawad, A., Ouf, A. & Siam, R. The Egyptian Red Sea coastal
511 microbiome: A study revealing differential microbial responses to diverse anthropogenic
512 pollutants. *Environmental Pollution* **214**, 892–902 (2016).
- 513 23. Nathani, N. M., Mootapally, C. & Dave, B. P. Antibiotic resistance genes allied to the
514 pelagic sediment microbiome in the Gulf of Khambhat and Arabian Sea. *Science of The*
515 *Total Environment* **653**, 446–454 (2019).
- 516 24. Nowrotek, M., Jałowiecki, Ł., Harnisz, M. & Płaza, G. A. Culturomics and metagenomics:
517 In understanding of environmental resistome. *Front. Environ. Sci. Eng.* **13**, 40 (2019).
- 518 25. Underwood, G. J. C., Dumbrell, A. J., McGenity, T. J., McKew, B. A. & Whitby, C. The
519 Microbiome of Coastal Sediments. in *The Marine Microbiome* (eds. Stal, L. J. & Cretoiu,
520 M. S.) vol. 3 479–534 (Springer International Publishing, Cham, 2022).
- 521 26. Wambua, S. *et al.* Cross-Sectional Variations in Structure and Function of Coral Reef
522 Microbiome With Local Anthropogenic Impacts on the Kenyan Coast of the Indian
523 Ocean. *Front. Microbiol.* **12**, 673128 (2021).

- 524 27. Yang, J. *et al.* Marine Sediment Bacteria Harbor Antibiotic Resistance Genes Highly
525 Similar to Those Found in Human Pathogens. *Microb Ecol* **65**, 975–981 (2013).
- 526 28. Xu, M. *et al.* Elevated nitrate enriches microbial functional genes for potential
527 bioremediation of complexly contaminated sediments. *The ISME Journal* **8**, 1932–1944
528 (2014).
- 529 29. Muyzer, G. & Stams, A. J. M. The ecology and biotechnology of sulphate-reducing
530 bacteria. *Nat Rev Microbiol* **6**, 441–454 (2008).
- 531 30. Wentworth, C. K. A Scale of Grade and Class Terms for Clastic Sediments. *The Journal of*
532 *Geology* **30**, 377–392 (1922).
- 533 31. EPA. Method 7473 (SW-846): Mercury in Solids and Solutions by Thermal
534 Decomposition, Amalgamation, and Atomic Absorption Spectrophotometry,. (2007).
- 535 32. EPA. Method 3051A (SW-846). Microwave Assisted Acid Digestion of Sediments,
536 Sludges, and Oils. (2007).
- 537 33. EPA. Method 6010D (SW-846). Inductively Coupled Plasma-Atomic Emission
538 Spectrometry. (2014).
- 539 34. EPA. Method 7010. Graphite furnace atomic absorption spectrophotometry. (2007).
- 540 35. Cohen, M. A. & Ryan, P. B. Observations Less than the Analytical Limit of Detection: A
541 New Approach. *JAPCA* **39**, 328–329 (1989).
- 542 36. Ben Salem, F., Ben Said, O., Duran, R. & Monperrus, M. Validation of an Adapted
543 QuEChERS Method for the Simultaneous Analysis of Polycyclic Aromatic Hydrocarbons,
544 Polychlorinated Biphenyls and Organochlorine Pesticides in Sediment by Gas
545 Chromatography–Mass Spectrometry. *Bull Environ Contam Toxicol* **96**, 678–684 (2016).
- 546 37. EPA. Method 8270E. Semivolatile organic compounds by gas chromatography/mass
547 spectrometry. (2017).

- 548 38. ISO. Method 16703. Soil quality – Determination of content of hydrocarbon in the range
549 C10 to C40 by gas chromatography. (2011).
- 550 39. EPA. Method 5021A (SW-846): Volatile Organic Compounds (VOCs) in Various Sample
551 Matrices Using Equilibrium Headspace Analysis. (2014).
- 552 40. EPA. Method 8015C (SW-846). Nonhalogenated Organics by Gas Chromatography.
553 (2007).
- 554 41. Abdelrhman, K. F. A. *et al.* A First Insight into the Gut Microbiota of the Sea Turtle
555 *Caretta caretta*. *Front. Microbiol.* **7**, (2016).
- 556 42. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon
557 data. *Nat Methods* **13**, 581–583 (2016).
- 558 43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
559 reads. *EMBnet j.* **17**, 10 (2011).
- 560 44. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data
561 processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012).
- 562 45. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation
563 of species diversity (Hill numbers). *Methods Ecol Evol* **7**, 1451–1456 (2016).
- 564 46. Lin, H. & Peddada, S. D. Analysis of compositions of microbiomes with bias correction.
565 *Nat Commun* **11**, 3514 (2020).
- 566 47. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for
567 FastQ files. (2011).
- 568 48. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De
569 Novo Assembler. *CP in Bioinformatics* **70**, e102 (2020).
- 570 49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
571 **9**, 357–359 (2012).

- 572 50. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient
573 genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
- 574 51. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
575 assessing the quality of microbial genomes recovered from isolates, single cells, and
576 metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
- 577 52. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for
578 the analysis of massive data sets. *Nat Biotechnol* **35**, 1026–1028 (2017).
- 579 53. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
580 *Genome Biol* **20**, 257 (2019).
- 581 54. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive
582 Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**, e61217 (2013).
- 583 55. Alcock, B. P. *et al.* CARD 2023: expanded curation, support for machine learning, and
584 resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic
585 Acids Research* **51**, D690–D699 (2023).
- 586 56. Banerjee, S. & Van Der Heijden, M. G. A. Soil microbiomes and one health. *Nat Rev
587 Microbiol* **21**, 6–20 (2023).
- 588 57. World Health Organization. Regional Office for Europe. *A Health Perspective on the Role
589 of the Environment in One Health*. World Health Organization. Regional Office for
590 Europe. <https://iris.who.int/handle/10665/354574>. (2022).
- 591 58. Laxminarayan, R. The overlooked pandemic of antimicrobial resistance. *The Lancet* **399**,
592 606–607 (2022).
- 593 59. Laxminarayan, R. *et al.* Antibiotic resistance—the need for global solutions. *The Lancet
594 Infectious Diseases* **13**, 1057–1098 (2013).

- 595 60. Murray, C. J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a
596 systematic analysis. *The Lancet* **399**, 629–655 (2022).
- 597 61. Fondi, M. & Fani, R. The horizontal flow of the plasmid resistome: clues from
598 inter-generic similarity networks. *Environmental Microbiology* **12**, 3228–3242 (2010).
- 599 62. Qamar, M. U. *et al.* Antibiotic-Resistant Bacteria, Antimicrobial Resistance Genes, and
600 Antibiotic Residue in Food from Animal Sources: One Health Food Safety Concern.
601 *Microorganisms* **11**, 161 (2023).
- 602 63. Walsh, T. R. A one-health approach to antimicrobial resistance. *Nat Microbiol* **3**, 854–
603 855 (2018).
- 604 64. Wang, X. & Li, T. Current problems with the antibiotic-resistant bacteria and
605 multiresistance bacteria. in *Degradation of Antibiotics and Antibiotic-Resistant Bacteria*
606 *from Various Sources* 89–115 (Elsevier, 2023). doi:10.1016/B978-0-323-99866-6.00016-
607 7.
- 608 65. Balcázar, J. L., Subirats, J. & Borrego, C. M. The role of biofilms as environmental
609 reservoirs of antibiotic resistance. *Front. Microbiol.* **6**, (2015).
- 610 66. Li, Y., Zheng, L., Zhang, Y., Liu, H. & Jing, H. Comparative metagenomics study reveals
611 pollution induced changes of microbial genes in mangrove sediments. *Sci Rep* **9**, 5739
612 (2019).
- 613 67. Lu, X.-M., Chen, C. & Zheng, T.-L. Metagenomic Insights into Effects of Chemical
614 Pollutants on Microbial Community Composition and Function in Estuarine Sediments
615 Receiving Polluted River Water. *Microb Ecol* **73**, 791–800 (2017).
- 616 68. de Girolamo, P. & Noli, A. The historical development of the Port of Livorno (Italy) and
617 its new Port Plan 2010 in advanced stage of elaboration. *Journal of Shipping and Ocean*
618 *Engineering* **1.2**, 77–100 (2011).

- 619 69. Dell'Anno, F. *et al.* Highly Contaminated Marine Sediments Can Host Rare Bacterial Taxa
620 Potentially Useful for Bioremediation. *Front. Microbiol.* **12**, 584850 (2021).
- 621 70. Guo, H., Nasir, M., Lv, J., Dai, Y. & Gao, J. Understanding the variation of microbial
622 community in heavy metals contaminated soil using high throughput sequencing.
623 *Ecotoxicology and Environmental Safety* **144**, 300–306 (2017).
- 624 71. Yu, C., Zhu, Z., Meng, K., Zhang, H. & Xu, M. Unveiling the impact and mechanisms of Cd-
625 driven ecological assembly and coexistence of bacterial communities in coastal
626 sediments of Yellow Sea. *Journal of Hazardous Materials* **460**, 132309 (2023).
- 627 72. Hamdan, H. Z., Ahmad, F. A., Zayyat, R. M. & Salam, D. A. Spatio-temporal variation of
628 the microbial community of the coast of Lebanon in response to petroleum hydrocarbon
629 pollution. *Marine Pollution Bulletin* **192**, 115037 (2023).
- 630 73. Collingro, A., Köstlbacher, S. & Horn, M. Chlamydiae in the Environment. *Trends in*
631 *Microbiology* **28**, 877–888 (2020).
- 632 74. Darby, E. M. *et al.* Molecular mechanisms of antibiotic resistance revisited. *Nat Rev*
633 *Microbiol* **21**, 280–295 (2023).
- 634 75. Stogios, P. J. & Savchenko, A. Molecular mechanisms of vancomycin resistance. *Protein*
635 *Science* **29**, 654–669 (2020).
- 636 76. Hong, H.-J., Hutchings, M. I. & Buttner, M. J. Vancomycin Resistance VanS/VanR Two-
637 Component Systems. in *Bacterial Signal Transduction: Networks and Drug Targets* (ed.
638 Utsumi, R.) vol. 631 200–213 (Springer New York, New York, NY, 2008).
- 639 77. Boyd, D. A. *et al.* VanG-Type Vancomycin-Resistant *Enterococcus faecalis* Strains Isolated
640 in Canada. *Antimicrob Agents Chemother* **50**, 2217–2221 (2006).

- 641 78. Croxatto, A. *et al.* VanT, a Homologue of *Vibrio harveyi* LuxR, Regulates Serine,
642 Metalloprotease, Pigment, and Biofilm Production in *Vibrio anguillarum*. *J Bacteriol* **184**,
643 1617–1629 (2002).
- 644 79. Arthur, M., Reynolds, P. & Courvalin, P. Glycopeptide resistance in enterococci. *Trends in*
645 *Microbiology* **4**, 401–407 (1996).
- 646 80. Courvalin, P. Vancomycin Resistance in Gram-Positive Cocci. *Clinical Infectious Diseases*
647 **42**, S25–S34 (2006).
- 648 81. Marshall, C. G., Lessard, I. A. D., Park, I.-S. & Wright, G. D. Glycopeptide Antibiotic
649 Resistance Genes in Glycopeptide-Producing Organisms. *Antimicrob Agents Chemother*
650 **42**, 2215–2220 (1998).
- 651 82. Novotna, G., Hill, C., Vincent, K., Liu, C. & Hong, H.-J. A Novel Membrane Protein, VanJ,
652 Conferring Resistance to Teicoplanin. *Antimicrob Agents Chemother* **56**, 1784–1796
653 (2012).
- 654 83. Lu, J.-J., Perng, C.-L., Ho, M.-F., Chiueh, T.-S. & Lee, W.-H. High Prevalence of VanB2
655 Vancomycin-Resistant *Enterococcus faecium* in Taiwan. *J Clin Microbiol* **39**, 2140–2145
656 (2001).
- 657 84. Lu, J.-J., Chang, T.-Y., Perng, C.-L. & Lee, S.-Y. The *vanB2* Gene Cluster of the Majority of
658 Vancomycin-Resistant *Enterococcus faecium* Isolates from Taiwan Is Associated with the
659 *pbp5* Gene and Is Carried by Tn 5382 Containing a Novel Insertion Sequence. *Antimicrob*
660 *Agents Chemother* **49**, 3937–3939 (2005).
- 661 85. Guffey, A. A. & Loll, P. J. Regulation of Resistance in Vancomycin-Resistant Enterococci:
662 The VanRS Two-Component System. *Microorganisms* **9**, 2026 (2021).

86. Vimberg, V., Zieglerová, L., Buriánková, K., Branny, P. & Balíková Novotná, G. VanZ
Reduces the Binding of Lipoglycopeptide Antibiotics to *Staphylococcus aureus* and
Streptococcus pneumoniae Cells. *Front. Microbiol.* **11**, 566 (2020).
87. Arthur, M., Molinas, C., Depardieu, F. & Courvalin, P. Characterization of Tn1546, a Tn3-
related transposon conferring glycopeptide resistance by synthesis of depsipeptide
peptidoglycan precursors in *Enterococcus faecium* BM4147. *J Bacteriol* **175**, 117–127
(1993).
88. Haas, W. *et al.* Genomic Analysis of Vancomycin-Resistant *Staphylococcus aureus*
Isolates from the 3rd Case Identified in the United States Reveals Chromosomal
Integration of the *vanA* Locus. *Microbiol Spectr* **11**, e04317-22 (2023).

3.1.3 Bacterial genomes: *Origanum vulgare* L.

This study shows the utility and the methodology of a *de novo* assembly of four endophytic bacterial strains' genomes, and the detection of different types of genes, such as antibiotic resistance or diesel fuel degradation (Semenzato et al. 2022). Nowadays, pathogens that have developed resistance to multiple drugs, elevating the risk to human health and the declining effectiveness of traditional antibiotics. This makes the search for alternatives a top priority. It is worth noticing the high diversity of natural bioactive compounds in medicinal plants, which may potentially open up new paths for therapy. Interestingly, the research outcomes have demonstrated that endophytic bacteria might actively participate in the development of these useful compounds, thus, the importance of the comprehension of the symbiotic relationship between them and the plant tissues.

The aim of our study was the isolation of four endophytic bacterial strains from *Origanum vulgare* L. tissues, a medicinal plant that has been recognized as a source of health benefits. The strains showed potential to curb the growth of opportunistic pathogens, which otherwise could lead to severe infections in patients with cystic fibrosis.

We conducted Average Nucleotide Identity (ANI) analysis with FastANI to compare newly assembled genomes, and checked the contaminants using ContEst16S. This tool generated a maximum likelihood phylogenetic tree based on 16S rRNA sequences, and we subsequently retrieved closely related genera from NCBI for further ANI comparisons. Species affiliation was validated through BLAST, analyzing 3000 bp random regions of each query genome. Thus, we identified secondary metabolite biosynthetic gene clusters using antiSMASH with stringent detection criteria and predicted resistomes using the Resistance Gene Identifier (RGI) from the CARD database. We found their genomes were replete with gene clusters responsible for the synthesis of several antibiotics like paeniodin, paenilarvins, polymyxin and paenicidin A, and therefore, they might have a role in their protection.

Lastly, we compared alkane hydroxylase genes (alkM, alkS, and alkT) with query genomes and verified the results using BLAST across genomes affiliated with specific bacterial strains, including *Metabacillus dongyingensis*, *Paenibacillus xylanexedens*, *Priestia megaterium*, and *Bacillus* sp. However, as much as the strains were capable of growing on the diesel fuel in a minimal media, no genes that were associated with the n-alkanes breakdown were detected. It points to the possibility that the bacteria might use the other pathways, potentially indirectly breaking down diesel using enzymes that cause the oxidation of aromatic compounds instead. The essence of this research was the understanding of the intricate relationships between plants and their microbial inhabitants.



Article

Genomic Analysis of Endophytic *Bacillus*-Related Strains Isolated from the Medicinal Plant *Origanum vulgare* L. Revealed the Presence of Metabolic Pathways Involved in the Biosynthesis of Bioactive Compounds

Giulia Semenzato ¹, Tania Alonso-Vázquez ¹, Sara Del Duca ¹, Alberto Vassallo ¹, Christopher Riccardi ¹, Marco Zaccaroni ¹, Nadia Mucci ², Anna Padula ², Giovanni Emiliani ³, Antonio Palumbo Piccionello ⁴, Anna Maria Puglia ⁴ and Renato Fani ^{1,*}

¹ Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, 50019 Florence, Italy; giulia.semenzato@unifi.it (G.S.); tania.alonsovasquez@unifi.it (T.A.-V.); sara.delduca@unifi.it (S.D.D.); alberto.vassallo@unicam.it (A.V.); christopher.riccardi@unifi.it (C.R.); marco.zaccaroni@unifi.it (M.Z.)

² Institute for Environmental Protection and Research, Via Ca' Fornacetta 9, Ozzano dell'Emilia, 40064 Bologna, Italy; nadia.mucci@isprambiente.it (N.M.); anna.padula@isprambiente.it (A.P.)

³ Institute for Sustainable Plant Protection (IPSP), National Research Council (CNR), Via Madonna del Piano 10, Sesto Fiorentino, 50019 Florence, Italy; giovanni.emiliani@ipsp.cnr.it

⁴ Department of Biological, Chemical and Pharmaceutical Sciences and Technologies-STEBICEF, University of Palermo, Viale delle Scienze Ed.17, 90128 Palermo, Italy; antonio.palumbopiccionello@unipa.it (A.P.P.); a.maria.puglia@unipa.it (A.M.P.)

* Correspondence: renato.fani@unifi.it; Tel.: +39-0554574742



Citation: Semenzato, G.; Alonso-Vázquez, T.; Del Duca, S.; Vassallo, A.; Riccardi, C.; Zaccaroni, M.; Mucci, N.; Padula, A.; Emiliani, G.; Palumbo Piccionello, A.; et al. Genomic Analysis of Endophytic *Bacillus*-Related Strains Isolated from the Medicinal Plant *Origanum vulgare* L. Revealed the Presence of Metabolic Pathways Involved in the Biosynthesis of Bioactive Compounds. *Microorganisms* **2022**, *10*, 919. <https://doi.org/10.3390/microorganisms10050919>

Academic Editors: Gustavo Santoyo, James White, Ajay Kumar and Virendra Mishra

Received: 31 March 2022

Accepted: 25 April 2022

Published: 27 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Multidrug-resistant pathogens represent a serious threat to human health. The inefficacy of traditional antibiotic drugs could be surmounted through the exploitation of natural bioactive compounds of which medicinal plants are a great reservoir. The finding that bacteria living inside plant tissues, (i.e., the endophytic bacterial microbiome) can influence the synthesis of the aforementioned compounds leads to the necessity of unraveling the mechanisms involved in the determination of this symbiotic relationship. Here, we report the genome sequence of four endophytic bacterial strains isolated from the medicinal plant *Origanum vulgare* L. and able to antagonize the growth of opportunistic pathogens of cystic fibrosis patients. The in silico analysis revealed the presence of gene clusters involved in the production of antimicrobial compounds, such as paeniodin, paenilavins, polymyxin, and paenicidin A. Endophytes' adaptation to the plant microenvironment was evaluated through the analysis of the presence of antibiotic resistance genes in the four genomes. The diesel fuel degrading potential was also tested. Strains grew in minimum media supplemented with diesel fuel, but no *n*-alkanes degradation genes were found in their genomes, suggesting that diesel fuel degradation might occur through other steps involving enzymes catalyzing the oxidation of aromatic compounds.

Keywords: bacterial endophytes; essential oil; antimicrobial resistance; microbiome; plant growth-promoting bacteria

1. Introduction

The rapid development of multidrug-resistant pathogens requires novel therapeutic chemical compounds, of which medicinal and aromatic plants represent a great reservoir. The discovery that microorganisms residing inside the plant tissues may produce similar, if not the same, bioactive compounds as their plant hosts could offer a new sustainable and promising source of bioactive molecules [1,2].

Endophytes are referred to as microorganisms that reside inside the inner tissues of plants without any visible signs of infection [3]. They can influence plant physiology by

facilitating nutrients acquisition, producing or modulating the levels of phytohormones, and inducing systemic resistance in their host [4]. The bacterial infection contemplates more phases regulated by the activation of gene expression in both plant and microorganisms: first, the production of root exudates attracts soil bacteria near the rhizosphere; once the bacteria have adhered to the surface of the roots, they can penetrate inside the internal parts of the plant and colonize different anatomical parts [5]. It has been demonstrated that different compartments of the same plant are inhabited by different endophytic communities: the cultivable bacteria obtained from the anatomical structures of the medicinal plants *Echinacea purpurea* L. and *Lavandula angustifolia* Mill. are characterized by a distinct structure and composition in terms of genera and endophytic strains [6,7]. The peculiar distribution of the endophytes within plant inner tissues suggests the existence of a selective pressure, but little is known about the forces involved in the shaping of the microbial communities.

The plant anatomical parts represent specific niches that require endophytes adaptation. Antimicrobial molecules produced by specialized compartments of medicinal plants could allow microorganisms with particular antimicrobial resistance characteristics to colonize that specific tissue; thus, essential oils produced by aromatic plants could act as a selective force driving microbial communities distribution inside the plant [8]. Moreover, bacterial endophytes themselves might influence the microenvironments of the plant [9]: the microorganisms can strongly influence the plant's bioactive molecules production, contribute to the secondary metabolism of their host, and directly synthesize antimicrobials [10].

In this context, metaomics and comparative genome analysis can be applied to the plant holobiont to unravel the mechanisms implied in the plant–microbiota interactions. Identification and characterization of genes involved in such a beneficial relationship are necessary to better exploit the mutualistic association between the two counterparts [11]. Functional aspects and molecular interactions can be delineated, enabling the prediction of endophytes' capacity to synthesize novel bioactive secondary metabolites and potential antibacterial drug candidates [4], able to overcome the global issue of antibiotic resistance.

In this work, we provided the genetic features of four endophytic strains belonging to a bacterial collection of isolates obtained from the medicinal plant *O. vulgare* L., widely known for its antimicrobial and antioxidant activity [12]. Based on the inhibiting activity exerted by some of the bacterial isolates tested against human pathogens [13], OVS6, OVL9, OVF10, and OVS21 were selected as good candidates for further characterization. Thus, their genomes sequence was obtained, in order to decipher the genetic features that underlie the plant–microbiota interactions and to better understand the potential pharmacological applications of endophytes.

2. Materials and Methods

2.1. Bacterial Strains and Growth Conditions

Plants of *Origanum vulgare* L., cultivated in the common garden “Giardino delle Erbe—Augusto Rinaldi Ceroni” (Casola Valsenio, Ravenna, Italy) were collected in July 2018. Endophytes and soil bacteria were isolated from the anatomical parts of the plant, (i.e., flower, leaf, and stem) and bulk soil, as described in Castronovo et al. [13]. One isolate from flower, one from leaf, and two from stem compartments were selected for the present work. Isolates are referred to as OV followed by F, L, or S for flower, leaf, and stem districts, respectively. Three of them were previously affiliated with the *Bacillus* genus (OVL9, OVS6, and OVS21), while OVF10 belongs to the genus *Paenibacillus* [13]. Each strain was cultured on tryptic soy agar (TSA) medium (BioLife Italiana, Milan, Italy) for 48 h at 30 °C.

2.2. Tests for the Diesel Degrading Potential of the Endophytes

The four endophytic strains were tested for their ability to grow in the presence of diesel fuel as the only carbon and energy source. A single colony of each endophyte was separately suspended in 100 µL of 0.9% *w/v* NaCl solution and then streaked on different minimum media agar plates supplemented with glucose or diesel. In particular, bacterial

strains were streaked on Minimum Davis (MMD, 1 g $(\text{NH}_4)_2\text{SO}_4$, 7 g K_2HPO_4 , 2 g KH_2PO_4 , 0.5 g $\text{Na}_3\text{-citrate}\cdot 2\text{H}_2\text{O}$, 0.1 g $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$, pH 7.2, per liter of deionized water) and M9 minimal medium (6 g Na_2HPO_4 , 3 g KH_2PO_4 , 0.5 g NaCl, 1 g NH_4Cl , 1 mL MgSO_4 1M, 0.5 mL CaCl_2 0.5M per liter of deionized water) containing 0.4% *v/v* diesel fuel or 1% *w/v* glucose as the sole carbon sources. Diesel fuel (Esso Italiana, Roma, Italy) was previously filtered through a 0.2 μm -pore-size filter (Sartorius) for sterilization and particle removal. TSA plates were used as growth control. Once streaked, the endophytes were incubated at 30 °C for 3 days. Positivity to this assay was assessed as the presence or the absence of visible growth, expressed in a range from 4 (complete growth) to 0 (absence of growth).

The test was then performed in liquid medium, using MMD and M9 supplemented with glucose (1% *w/v*) or diesel (0.4% *v/v*), and tryptic soy broth (TSB, BioLife) as growth control. Endophytes growth curves were performed in microplates. Cells were grown overnight at 30 °C under shaking (130 rpm) in TSB. The next morning, cell cultures were washed twice in saline solution (NaCl 0.9% *w/v*) and diluted to a starting O.D.₆₀₀ of 0.01 in each media. Cells were then incubated at 30 °C and the O.D.₆₀₀ measures were taken every hour for 72 h using the Infinite M Nano (Tecan, Männedorf, Switzerland) microplate reader. Each curve was performed in triplicate. The calculation of the areas under the growth curve (AUC) was obtained according to the chained trapezoidal rule [14]. Statistical significance of the differences between each condition was calculated through a non-parametric test and a post hoc analysis, Kruskal-Wallis and Dunn tests.

2.3. DNA Extraction

A single colony of each strain was inoculated in 10 mL of fresh TSB (Biolife) in a 50 mL tube and incubated at 30 °C overnight under shaking at 130 rpm. Bacterial cells were collected by centrifugation (15,500× *g* for 4 min) and the genomic DNA was extracted using PowerLyzer PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA), following the protocol provided by the manufacturer with some modifications to ensure optimal cell lysis. In particular, cells were resuspended in the PowerSoil Bead Solution in the presence of lysozyme (final concentration 1 mg/mL); after incubating for 1 h at 37 °C, PowerSoil Solution C1 and proteinase K (final concentration 0.5 mg/mL) were added and samples were furtherly incubated at 55 °C for 2 h before proceeding with the DNA purification steps.

2.4. Genome Sequencing

Nanopore sequencing was performed with a PCR-free approach following the native barcoding genomic DNA protocol provided by Oxford Nanopore Technologies (ONT) (version NBE_9065_v109_revY_14Aug2019). The gDNA of the strains were sequenced with other 8 non-related gDNA samples. Briefly, 1 μg of each input gDNA was repaired and end-prepped using the NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (E7180S, New England Biolabs). Upon purification with Agencourt AMPure XP beads (Beckman Coulter) on a magnetic separator, concentrations of DNA samples were determined using a Qubit 4 Fluorometer and Qubit dsDNA HS Assay Kit (ThermoFisher Scientific). Thus, 500 ng of each end-prepped DNA sample were barcoded using Native Barcoding Expansion 13–24 (EXP-NBD114, ONT) and NEB Blunt/TA Ligase Master Mix (M0367, New England Biolabs). After a purification step, equimolar amounts of barcoded DNA samples were pooled to have a total of 700 ng and were subjected to the adapter ligation. During the subsequent clean-up step, the DNA library was enriched with >3 kb-long fragments using the Long Fragment Buffer included in the Ligation Sequencing Kit (SQK-LSK109, ONT). DNA library was immediately sequenced; therefore, an R9.4.1 Flow Cell (FLO-MIN106D, ONT) was primed with a Flow Cell Priming Kit (EXP-FLP002, ONT). The library was loaded following the instruction provided by the protocol and sequencing was performed with a MinION MK1B (ONT) and the MinKNOW software v.21.10.4 for 72 h. Basecalling and demultiplexing were performed using Guppy v.4.3.4.

2.5. Genome Assembly and Bioinformatic Analysis

De novo assembly was accomplished using Canu assembler software v.2.1.1 [15] and the quality of contigs was evaluated by QUAST v.5.0.2 [16]. These procedures were performed in a Galaxy environment (<https://usegalaxy.eu>, accessed on 3 September 2021). The assembled genome sequences were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v.6.0 (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/, accessed on 22 March 2022). The Average Nucleotide Identity (ANI) analysis was performed using FastANI v.1.3, with default options [17]. The newly assembled genomes were uploaded to ContEst16S to assess the presence of contaminants. This service also generates a maximum likelihood phylogenetic tree using 16S rRNA sequences. Using this information, the genomic sequences of the genera that are more closely related to the bacterial strains under investigation were downloaded. These were retrieved from the NCBI and used as reference input for the ANI analysis. The affiliated species were confirmed with the Basic Local Alignment Search Tool (BLAST) by searching 3000 bp long random regions of each query genome.

2.6. Secondary Metabolites Gene Clusters Detection

The antiSMASH v.6.0.1 webserver was used for the identification of gene clusters involved in the biosynthesis of secondary metabolites in genomes of bacteria [18]. Each query genome (OVS6, OVL9, OVR10, and OVS21) was uploaded as a FASTA format, and the analysis was performed using a strict method of detection in order to identify only well-defined clusters containing genes with a significant alignment.

2.7. Antibiotic Resistance and Diesel Fuel Degradation Genes

The Resistance Gene Identifier from the Comprehensive Antibiotic Resistance Database (CARD: RGI) was used to predict resistome(s) from genome data based on homology and SNP models [19].

Alkane hydroxylases gene sequences, *alkM*, *alkS*, and *alkT*, were downloaded from the NCBI gene database, and then compared with OVS6, OVL9, OVF10, and OVS21 genomes. To verify the result, all three gene sequences were looked for in each one of the genome sequences affiliated to the query genomes (*Metabacillus dongyingensis*, *Paenibacillus xylanexedens*, and *Priestia megaterium*) and *Bacillus* sp. using the Basic Local Alignment Search Tool (BLAST).

3. Results

3.1. Genome Sequencing and Assembly

The results of genome assembly of the sequencing reads performed through Canu assembler plugin is shown in Table 1. The annotation of the genomic sequences was performed using the NCBI GenBank annotation pipeline. The complete genome sequences are available in GenBank under the accession numbers CP092335 (OVS6), JALHBO000000000 (OVL9), CP094668 (OVF10), JALHBP000000000 (OVS21).

Given the previous affiliation of OVS6, OVL9, and OVS21 to the genus *Bacillus* [13], their ANI analysis was performed using *Bacillus* sp. genomes downloaded from the NCBI assembly database (<http://www.ncbi.nlm.nih.gov/assembly/>, accessed on 7 March 2022) as references but showed low ANI values (<90%). In the light of this observation, the closest clades of the maximum likelihood phylogenetic trees obtained with ContEst16S were used as reference genomes. This resulted in higher ANI values and the affiliated species were confirmed with BLAST. Using 7 *Metabacillus* genomes for OVS6, 33 *Priestia* genomes for OVL9 and OVS21, and 116 *Paenibacillus* genomes for OVF10, the ANI analysis revealed that the endophytes are bound to *Metabacillus dongyingensis* with a 99.61 ANI value, *Priestia megaterium* with a 99.46 and 99.25 ANI values, and with *Paenibacillus xylanexedens* with a 95.25 ANI value, respectively (Table 2).

Table 1. General features of the genomes of the four endophytic strains.

	OVS6	OVL9	OVF10	OVS21
Assembly size	5,073,147	6,280,800	7,243,271	5,954,252
Number of contigs	1	12	1	11
GC%	40.09	37.54	45.89	37.68
Total genes	5510	7242	7168	6717
CDSs	5388	7053	7026	6496
rRNA	33	43	36	51
tRNA	84	138	102	162
ncRNA	5	8	4	8
Accession number	CP092335	JALHBO000000000	CP094668	JALHBP000000000

Table 2. Identity values for each endophytic strain.

	OVS6	OVL9	OVF10	OVS21
Species	<i>Metabacillus dongyingensis</i>	<i>Priestia megaterium</i>	<i>Paenibacillus xylanexedens</i>	<i>Priestia megaterium</i>
ANI value	99.61	99.46	95.25	99.25
BLAST ID%	100	99.89	92.81	99.57
BLAST query coverage%	100	100	99	100

3.2. Tests for The Diesel Fuel Degrading Potential of the Endophytes

Diesel fuel degrading potential of the endophytes was investigated, as described in Materials and Methods. As shown in Table 3, all the endophytes grew on MMD and M9 minimal agar media supplemented with glucose 1% *w/v* and, to a lesser extent, on the same media containing diesel fuel 0.4% *v/v*.

Table 3. Diesel fuel degrading potential of the endophytes on agar media. The values and the colors indicate different growth levels of the strains compared to the growth control, that is: complete (4), strong (3), intermediate (2), weak (1), absent (0).

	OVS6	OVL9	OVF10	OVS21
MMD + glucose	1	2	1	2
MMD + diesel fuel	1	1	1	1
M9 + glucose	2	3	1	3
M9 + diesel fuel	1	1	1	1
TSA	4	4	4	4

The test was also performed using liquid media, in triplicate. For each condition, endophytes growth curves were obtained, and the average AUC was calculated. Since data did not present a normal distribution and homoscedasticity, non-parametric tests (Kruskal–Wallis test and Dunn test) were used to reveal statistically significant differences between each growth condition.

As shown in Figure 1, the two different minimal media supplemented with glucose allowed the growth of all the endophytes, even though to a lesser extent than the TSB complete medium, except for OVS6 which had an impaired growth in MMD with glucose. Statistically significant differences between different growth conditions are highlighted in Figure 1. Endophytes growth in MMD supplemented with diesel fuel was not statistically different from the MMD with glucose condition, thus allowing us to hypothesize the ability of the strains to grow in the presence of diesel fuel as a carbon and energy source.

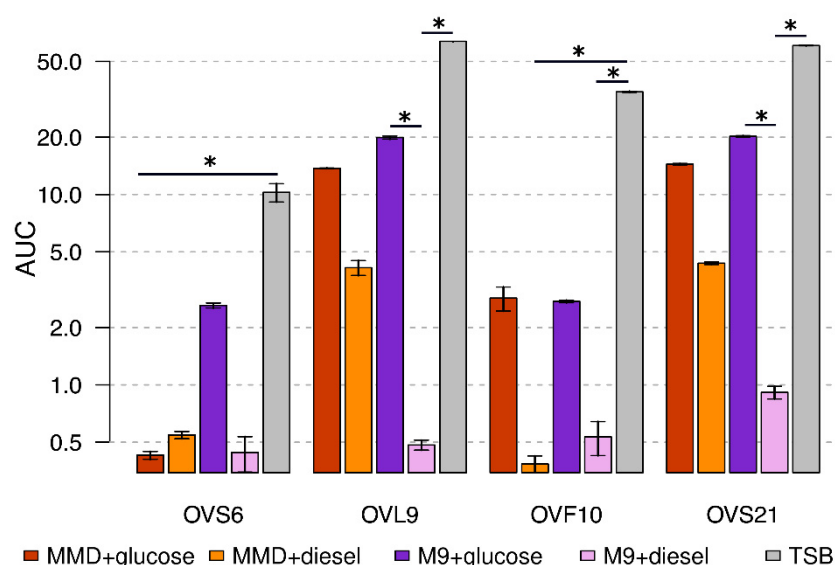


Figure 1. Diesel fuel degrading potential of the endophytes on liquid media. Values are expressed as the average values of area under the growth curve (AUC). Each color corresponds to a different growth medium, whereas bars represent standard errors. Asterisks show statistically significant differences between conditions ($p < 0.05$).

To validate the results obtained, alkane-degradation genes were searched in the genomes of the endophytes. There was no presence of *alkM*, *alkS*, and *alkT* genes in all four query genomes. Their absence was confirmed using the Basic Local Alignment Search Tool and, as reference genomes, those of *Metabacillus dongyingensis*, *Paenibacillus xylanexedens*, *Priestia megaterium*, and *Bacillus* sp.

3.3. Secondary Metabolites Gene Clusters

The analysis of secondary metabolites biosynthetic gene clusters (BGCs) predicted several types of BGCs for each genome; however, only the ones with a percentage equal to or higher than 50% of similarity to clusters available in the database are reported (Table 4). For the OVS6 genome, the analysis revealed a lasso peptide BCG, accountable for the production of paeninodin; a 100% similarity was found to the paeninodin BCG from *Paenibacillus dendritiformis* C454. For OVL9 and OVS21 genomes, a terpene BCG was found, with a 50% similarity to the carotenoid BCG from *Halobacillus halophilus* DSM 2266. Finally, the analysis of the OVF10 genome showed three BCGs. The first two were attributed to the non-ribosomal peptide synthetase cluster, finding a 100% similarity to the paenilarvins BCG from *Paenibacillus larvae* subsp. *larvae* DSM 25430, and to the polymyxin BCG from *Paenibacillus polymyxa*. The third cluster was a lanthipeptide-class-I BCG, with an 85% similarity with the paenicidin A BCG from *Paenibacillus polymyxa*.

Table 4. Biosynthetic gene clusters identified using AntiSMASH and having the higher percentage of similarity. Nucleotide positions within the genome are reported.

Genome	Type	From (nt)	To (nt)	Most Similar Known Cluster (% Similarity)
OVS6	Lasso peptide	2840151	2864059	Paeninodin (100%)
OVL9	Terpene (contig 11)	4862360	4883208	Carotenoid (50%)
OVF10	Non-ribosomal peptide	4914549	4966941	Paenilarvins (100%)
	synthetase cluster	6908170	6988910	Polymyxin (100%)
OVS21	Class I Lanthipeptide	270456	296759	Paenicidin A (85%)
	Terpene (contig 1)	4569666	4590514	Carotenoid (50%)

3.4. Antibiotic Resistance Genes

Using the Resistance Gene Identifier from the Comprehensive Antibiotic Resistance Database (CARD: RGI), for the OVF10 genome, one antibiotic resistance gene was found with a strict criterion algorithm. It obtained an identity of the matching region of 84.32% and 100% of the length of the reference sequence. This gene belongs to the “Llm 23S ribosomal RNA methyltransferase” antibiotic resistance orthology (ARO) and antimicrobial resistance (AMR) gene family and it is linked to the lincosamide antibiotic class. The reference genome to which the query gene matched has—as a resistance mechanism—the antibiotic target alteration.

Using the loose criterion for the other three query genomes, several genes were identified; however, only the ones with a high percentage of identity of matching region and length of reference sequence are reported here. For the OVS6 genome, the RGI found two matches. The first one obtained an identity of the matching region of 80.74% and 100.34% of the length of the reference sequence. It belongs to the “*vgbB*” ARO and “streptogramin *vgb* lyase” AMR gene family and it is linked to the streptogramin antibiotic class. It uses antibiotic inactivation as a resistance mechanism. The second one obtained an identity of the matching region of 80.6% and 99.57% of the length of the reference sequence. It belongs to the “*vanRM*” ARO and “glycopeptide resistance gene cluster, *vanR*” AMR gene family and it is related to the glycopeptide antibiotic class. It contemplates target alteration as a resistance mechanism. Lastly, the same hit was found for OVL9 and OVS21 genomes, which had an identity of the matching region of 80.45% and 80.53%, respectively, and 100.59% of the length of the reference sequence for both genomes. The hit gene belongs to the “*Staphylococcus aureus rpoB* mutants conferring resistance to rifampicin” ARO and to the “rifamycin-resistant beta-subunit of RNA polymerase (*rpoB*)” AMR gene family. It uses antibiotic target alteration and antibiotic target replacement as resistance mechanisms, and it belongs to the rifamycin antibiotic class. Since the loose algorithm works outside the detection model cut-offs, the found genes may require further study.

4. Discussion

Endophytes, microorganisms that reside in the internal tissues of living plants, have been recognized as potential untapped sources of novel natural products that could find applications in medicine, agriculture, and industry [20]. Endophytes are known to be able to synthesize some similar, if not the same, chemical compounds like the ones found in medicinal plant extracts, probably as an adaptation to the plant microenvironment [21]. Medicinal and aromatic plants have always been exploited for their natural biologically active compounds, so they can represent the starting point to investigate endophytes' secondary metabolites biosynthesis potential [2].

The aim of this work was to determine and analyze the complete genome sequence of four endophytic bacterial strains isolated from stem or leaves of *O. vulgare* plants in order to identify genes involved in the biosynthesis of antimicrobial compounds and/or responsible for resistance to antibiotics.

The genome analysis of the four *O. vulgare* endophytes revealed some relevant genetic features providing insights into their chemical potential as antibiotics producers and helping to identify genetic traits that could be associated with the endophytic lifestyle.

The ANI analysis allowed to shed light on the taxonomical position of the isolates, previously identified as *Bacillus* spp. (OVS6, OVL9, and OVS21) and *Paenibacillus* sp. (OVF10). The four *O. vulgare* endophytes belong to species already isolated from the aerial parts of other plants or their rhizosphere. Based on the ANI analysis results, OVS6 was identified as *Metabacillus dongyingensis*. This microorganism has been identified as a novel plant growth-promoting bacterium isolated from the saline-alkaline rhizosphere of *Ulmus pumila* and it is able to promote the growth of *Zea mais* L. under saline conditions [22]. Genome sequences from OVL9 and OVS21 showed a high degree of similarity to the species *Priestia megaterium*, formerly known as *Bacillus megaterium* [23]. *B. megaterium* isolated from the rhizosphere of maize plants demonstrated a growth-promoting effect on *Arabidopsis thaliana* that involved

auxin and ethylene production [24]. Finally, OVF10 was identified as *Paenibacillus xylanexedens*. Species of the genus *Paenibacillus* are well-known plant endophytes able to exert a growth-promoting activity on their host and protect it from phytopathogens through the biosynthesis of antibacterial compounds [25]. Thus, it can be suggested that these *O. vulgare* endophytes could be plant growth-promoting bacteria, exerting a beneficial effect on their host.

Diesel fuel degrading potential tests revealed the capability of the four endophytic strains to grow in the presence of diesel fuel as the sole carbon and energy source. The essential oil composition of *O. vulgare* plant from which the endophytes were extracted revealed the presence of 73.5% sesquiterpene hydrocarbons, 17.6% monoterpene hydrocarbons, 4.8% oxygenated sesquiterpenes, and 3.7% oxygenated monoterpenes [13]. The capability of the endophytes to grow on media containing diesel fuel could suggest their adaptation to the hydrocarbon components of *O. vulgare* essential oil, leading to the hypothesis that its composition might represent one of the factors involved in the plant-endophytes symbiosis [8]. For this reason, alkane hydroxylases genes (*alk* genes) have been searched in the endophytes' genomes. Even though these genes have not been identified, it cannot be excluded that the strains' capability to grow on diesel fuel might depend on other gene clusters involved in different hydrocarbons metabolism.

Resistance Gene Identifier provided information about the antibiotic resistance profiles of the four isolates. Data obtained can be related to the adaptation mechanisms that endophytes require in order to colonize plant inner tissues, which are microenvironments rich in antimicrobial molecules.

The evaluation of secondary metabolites biosynthetic gene clusters revealed a wide spectrum of compounds. Concerning the OVS6 genome, the analysis revealed a lasso peptide biosynthetic gene cluster (BGC) accountable for the production of paenidin; a 100% similarity was found with the paenidin BGC from *Paenibacillus dendritiformis* C454. Lasso peptides are a class of ribosomally-synthesized and post-translationally modified peptides (RiPPs), known for their antimicrobial and antiviral activities [26]. OVF10 genome contains two non-ribosomal peptide synthase clusters, finding a 100% similarity with the paenilavins BGC from *Paenibacillus larvae* subsp. *larvae* DSM 25430, and with the polymyxin BGC from *Paenibacillus polymyxa*. It also harbors a lanthipeptide-class-I BCG, with an 85% similarity with the paenidin A BGC from *Paenibacillus polymyxa*. Paenilavins, members of the iturins family, are characterized by a strong antifungal activity, able to contrast the growth of human and plant and pathogenic fungi [27]. Polymyxins are well-known antibacterial molecules with a narrow antibacterial spectrum, mainly against common Gram-negative bacteria. They are active against most members of the Enterobacteriaceae family, including *Escherichia coli*, *Enterobacter* spp., *Klebsiella* spp., *Citrobacter* spp., *Salmonella* spp., and *Shigella* spp. [28]. On the contrary, paenidin belongs to the antibiotic class of lantibiotics, which are active against Gram-positive bacteria [29]. The identified BCGs might be involved in the growth inhibitory activity that the four endophytes exert against ten selected strains belonging to the *Burkholderia cepacia* complex, opportunistic pathogens able to infect immunocompromised cystic fibrosis patients, and other Gram-positive and Gram-negative human pathogens [13].

These data represent a further step in the comprehension of the role of the bacterial endophytes in medicinal plant secondary metabolism and might shed light on their biotechnological application in pharmaceutical and agricultural fields. To test whether bacterial endophytes are involved in the promotion of *O. vulgare* plant growth and/or influence the biosynthesis of secondary metabolites that can then flow in the plant essential oil/extract, an in vitro model for *O. vulgare* plants can be set up. In this way axenic *O. vulgare* plants might be inoculated with bacterial endophytic strains isolated from the same plants, thus allowing the comparison of the behavior of inoculated plants vs. control, (i.e., non-inoculated ones) from different viewpoints: growth, production of secondary metabolite, volatile organic compounds. The setup of this in vitro model is in progress.

Author Contributions: Conceptualization, A.P.P., G.E., A.M.P. and R.F.; Data curation, A.V.; Formal analysis, T.A.-V. and C.R.; Investigation, G.S., S.D.D., A.V., N.M. and A.P.; Project administration, R.F.;

Resources, S.D.D., M.Z. and R.F.; Supervision, G.E., A.P.P., A.M.P. and R.F.; Visualization, G.S., C.R. and T.A.-V.; Writing—original draft, G.S., T.A.-V., S.D.D., A.V. and C.R.; Writing—review and editing, M.Z., N.M., A.P., G.E., A.P.P., A.M.P. and R.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Alvin, A.; Miller, K.I.; Neilan, B.A. Exploring the potential of endophytes from medicinal plants as sources of antimycobacterial compounds. *Microbiol. Res.* **2014**, *169*, 483–495. [\[CrossRef\]](#) [\[PubMed\]](#)
- Castronovo, L.M.; Vassallo, A.; Mengoni, A.; Miceli, E.; Bogani, P.; Firenzuoli, F.; Fani, R.; Maggini, V. Medicinal plants and their bacterial microbiota: A review on antimicrobial compounds production for plant and human health. *Pathogens* **2021**, *10*, 106. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hardoim, P.R.; Van Overbeek, L.S.; Berg, G.; Pirttilä, A.M.; Compant, S.; Campisano, A.; Döring, M.; Sessitsch, A. The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiol. Mol. Biol. Rev.* **2015**, *79*, 293–320. [\[CrossRef\]](#) [\[PubMed\]](#)
- Burrage, S.G.; Jeon, J. Applications of endophytic microbes in agriculture, biotechnology, medicine, and beyond. *Microbiol. Res.* **2021**, *245*, 126691. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pinski, A.; Betekhtin, A.; Hupert-Kocurek, K.; Mur, L.A.J.; Hasterok, R. Defining the genetic basis of plant–endophytic bacteria interactions. *Int. J. Mol. Sci.* **2019**, *20*, 1947. [\[CrossRef\]](#)
- Chiellini, C.; Maida, I.; Emiliani, G.; Mengoni, A.; Mocali, S.; Fabiani, A.; Biffi, S.; Maggini, V.; Gori, L.; Vannacci, A.; et al. Endophytic and rhizospheric bacterial communities isolated from the medicinal plants *Echinacea purpurea* and *Echinacea angustifolia*. *Int. Microbiol.* **2014**, *17*, 165–174. [\[CrossRef\]](#)
- Emiliani, G.; Mengoni, A.; Maida, I.; Perrin, E.; Chiellini, C.; Fondi, M.; Gallo, E.; Gori, L.; Maggini, V.; Vannacci, A.; et al. Linking bacterial endophytic communities to essential oils: Clues from *Lavandula angustifolia* mill. *Evid. Based Complement. Altern. Med.* **2014**, *2014*, 1–16. [\[CrossRef\]](#)
- Checucci, A.; Maida, I.; Bacci, G.; Ninno, C.; Bilia, A.R.; Biffi, S.; Firenzuoli, F.; Flamini, G.; Fani, R.; Mengoni, A. Is the plant-associated microbiota of *Thymus* spp. adapted to plant essential oil? *Res. Microbiol.* **2017**, *168*, 276–282. [\[CrossRef\]](#)
- Maggini, V.; Miceli, E.; Fagorzi, C.; Maida, I.; Fondi, M.; Perrin, E.; Mengoni, A.; Bogani, P.; Chiellini, C.; Mocali, S.; et al. Antagonism and antibiotic resistance drive a species-specific plant microbiota differentiation in *Echinacea* spp. *FEMS Microbiol. Ecol.* **2018**, *94*, fty118. [\[CrossRef\]](#)
- Brader, G.; Compant, S.; Mitter, B.; Trognitz, F.; Sessitsch, A. Metabolic potential of endophytic bacteria. *Curr. Opin. Biotechnol.* **2013**, *27*, 30–37. [\[CrossRef\]](#)
- Kaul, S.; Sharma, T.; Dhar, M.K. ‘Omics’ tools for better understanding the plant–endophyte interactions. *Front. Plant Sci.* **2016**, *7*, 955. [\[CrossRef\]](#) [\[PubMed\]](#)
- Coccimiglio, J.; Alipour, M.; Jiang, Z.-H.; Gottardo, C.; Suntres, Z. Antioxidant, antibacterial, and cytotoxic activities of the ethanolic *Origanum vulgare* extract and its major constituents. *Oxidative Med. Cell. Longev.* **2016**, *2016*, 1–8. [\[CrossRef\]](#) [\[PubMed\]](#)
- Castronovo, L.M.; Calónico, C.; Ascrizzi, R.; Del Duca, S.; Delfino, V.; Chioccioli, S.; Vassallo, A.; Strozza, I.; De Leo, M.; Biffi, S.; et al. The Cultivable Bacterial Microbiota Associated to the Medicinal Plant *Origanum vulgare* L.: From Antibiotic Resistance to Growth-Inhibitory Properties. *Front. Microbiol.* **2020**, *11*, 862. [\[CrossRef\]](#)
- Del Duca, S.; Puglia, A.M.; Calderone, V.; Bazzicalupo, M.; Fani, R. Effect of Non-Lethal Selection on Spontaneous Revertants of Frameshift Mutations: The *Escherichia coli* *hisF* Case. *Microorganisms* **2022**, *10*, 692. [\[CrossRef\]](#)
- Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [\[CrossRef\]](#)
- Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 5114. [\[CrossRef\]](#)
- Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [\[CrossRef\]](#)
- Alcock, B.P.; Raphenya, A.R.; Lau, T.T.Y.; Tsang, K.K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.-L.V.; Cheng, A.A.; Liu, S.; et al. CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2020**, *48*, D517–D525. [\[CrossRef\]](#)

20. Strobel, G.; Daisy, B.; Castillo, U.; Harper, J. Natural Products from Endophytic Microorganisms. *J. Nat. Prod.* **2004**, *67*, 257–268. [CrossRef]
21. Martínez-Klimova, E.; Rodríguez-Peña, K.; Sánchez, S. Endophytes as sources of antibiotics. *Biochem. Pharmacol.* **2016**, *134*, 1–17. [CrossRef] [PubMed]
22. Yin, Z.; Wang, X.; Hu, Y.; Zhang, J.; Li, H.; Cui, Y.; Zhao, D.; Dong, X.; Zhang, X.; Liu, K.; et al. *Metabacillus dongyingensis* sp. nov. Is Represented by the Plant Alkaline Soil and Enhances the Growth of *Zea mays* L. under Salt. *Msystems* **2022**, *8*, e01426-21.
23. Gupta, R.S.; Patel, S.; Saini, N.; Chen, S. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel Bacillaceae genera, by phylogenomics and comparative genomic analyses: Description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5753–5798. [CrossRef]
24. López-Bucio, J.; Campos-Cuevas, J.C.; Hernández-Calderón, E.; Velásquez-Becerra, C.; Fariás-Rodríguez, R.; Macías-Rodríguez, L.I.; Valencia-Cantero, E. *Bacillus megaterium* rhizobacteria promote growth and alter root-system architecture through an auxin- and ethylene-independent signaling mechanism in *Arabidopsis thaliana*. *Mol. Plant-Microbe Interact.* **2007**, *20*, 207–217. [CrossRef]
25. Rybakova, D.; Cernava, T.; Köberl, M.; Liebminger, S.; Etemadi, M.; Berg, G. Endophytes-assisted biocontrol: Novel insights in ecology and the mode of action of *Paenibacillus*. *Plant Soil* **2015**, *405*, 125–140. [CrossRef]
26. Hegemann, J.; Zimmermann, M.; Xie, X.; Marahiel, M.A. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Accounts Chem. Res.* **2015**, *48*, 1909–1919. [CrossRef] [PubMed]
27. Sood, S.; Steinmetz, H.; Beims, H.; Mohr, K.I.; Stadler, M.; Djukic, M.; von der Ohe, W.; Steinert, M.; Daniel, R.; Müller, R. Paenilarvins: Iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. *ChemBioChem* **2014**, *15*, 1947–1955. [CrossRef]
28. Poirel, L.; Jayol, A.; Nordmann, P. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. *Clin. Microbiol. Rev.* **2017**, *30*, 557–596. Available online: <https://www.ncbi.nlm.nih.gov/pubmed/28275006> (accessed on 30 March 2022). [CrossRef]
29. Lohans, C.T.; Huang, Z.; van Belkum, M.J.; Giroud, M.; Sit, C.S.; Steels, E.M.; Zheng, J.; Whittall, R.M.; McMullen, L.M.; Vederas, J.C. Structural characterization of the highly cyclized lantibiotic paenicidin a via a partial desulfurization/reduction strategy. *J. Am. Chem. Soc.* **2012**, *134*, 19540–19543. [CrossRef]

3.2 Modeling microbial dynamics and metabolism

Genome-scale metabolic models (GSMMs) are powerful tools for studying metabolism at a systems level, providing insights into microbial and human physiology. By integrating multi-omics data and computational modeling, I explored metabolic adaptations in different contexts, including environmental microorganisms and human health. This section presents applications of GSMMs in different contexts, such as antimicrobial resistance in subsection 3.2.1. Additionally, in subsections 3.2.2 and 3.2.3, I highlight how computational modeling can be used to explore complex biological processes or even describe microbial community dynamics (subsection 3.2.4).

During my doctoral research, I had the opportunity to study and model microbial and human metabolism using GSMMs. First of all, I performed an extensive research on how GSMMs are used to understand antimicrobial resistance and the possible paths of action we may have ahead (Alonso-Vásquez, Fondi, et al. 2023). I explored the metabolic characteristics and potential of a cold-adapted marine bacterium by integrating transcriptomic and metabolomic data into its GSMM (Riccardi et al. 2023). In addition, following the same workflow, I integrated metabolomic data into a human GSMM in order to study how COVID-19 impacts human metabolism (Ghini et al. 2023), this work gave me the opportunity to contrast metabolic models from different organisms. Lastly, I developed and analyzed a theoretical framework to study the dynamics of a microbial community, especially the impact of population size fluctuations on bacterial persister cells (Alonso-Vásquez, Giovannini, Garbini, et al. 2024). On the following pages I will explain a little further what was done in each project.

3.2.1 GSMMs and antimicrobial resistance

The discovery of antimicrobials was a breakthrough in medicine and in everyday life, however, the growing number of pathogens resistant to these molecules has

become a global urgency in the last decades. Among all the strategies being explored to fight this challenge, a particularly promising one lies in investigating how cells' metabolism responds to antimicrobial exposure and how this response influences the efficacy of these drugs. Current antimicrobials inhibit several cellular processes, such as DNA replication, transcription, protein synthesis or cell wall biosynthesis, and so microbial metabolism undergoes significant changes in response. These metabolic changes are key predictors of treatment outcomes and a potential reservoir of potential drug targets or adjuvants.

Despite its potential, understanding the metabolic response of microbial cells remains challenging due to the complexity of metabolic networks, especially if there is a metabolic flux reorganization. To overcome this, computational modeling approaches have been developed and are becoming increasingly popular. Genome-scale metabolic models are the reconstruction of the metabolic network of an organism, which takes into account the genome annotation, the information available in several databases (for genes, metabolites, reactions and pathways) and further manual curation (see section [1.2.5](#)).

In this review we focused on the application of GSMMs in exploring the intertwined relationship between microbial metabolism and antimicrobial agents. We gathered around twenty different studies, the most recent case studies at the time. These works highlighted the use of GSMMs to investigate the metabolic adaptations of microbes during antimicrobial exposure. This approach shed light on the potential of these models to uncover new therapeutic targets and improve treatment strategies.

Review

Understanding Antimicrobial Resistance Using Genome-Scale Metabolic Modeling

Tania Alonso-Vásquez , Marco Fondi  and Elena Perrin 

Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto F.no, 50019 Florence, Italy; tania.alonsovasquez@unifi.it (T.A.-V.); marco.fondi@unifi.it (M.F.)

* Correspondence: elena.perrin@unifi.it; Tel.: +39-055-4574760

Abstract: The urgent necessity to fight antimicrobial resistance is universally recognized. In the search of new targets and strategies to face this global challenge, a promising approach resides in the study of the cellular response to antimicrobial exposure and on the impact of global cellular reprogramming on antimicrobial drugs' efficacy. The metabolic state of microbial cells has been shown to undergo several antimicrobial-induced modifications and, at the same time, to be a good predictor of the outcome of an antimicrobial treatment. Metabolism is a promising reservoir of potential drug targets/adjuvants that has not been fully exploited to date. One of the main problems in unraveling the metabolic response of cells to the environment resides in the complexity of such metabolic networks. To solve this problem, modeling approaches have been developed, and they are progressively gaining in popularity due to the huge availability of genomic information and the ease at which a genome sequence can be converted into models to run basic phenotype predictions. Here, we review the use of computational modeling to study the relationship between microbial metabolism and antimicrobials and the recent advances in the application of genome-scale metabolic modeling to the study of microbial responses to antimicrobial exposure.

Keywords: metabolic modeling; antimicrobial resistance; bacterial metabolism



Citation: Alonso-Vásquez, T.; Fondi, M.; Perrin, E. Understanding Antimicrobial Resistance Using Genome-Scale Metabolic Modeling. *Antibiotics* **2023**, *12*, 896. <https://doi.org/10.3390/antibiotics12050896>

Academic Editor: Marc Maresca

Received: 2 March 2023

Revised: 28 April 2023

Accepted: 6 May 2023

Published: 11 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Antimicrobial resistance (AMR), the ability of a microorganism to resist the action of one or more antimicrobial agents, is one of the major public health problems of this century [1]. Antimicrobial-resistant microbes are currently estimated to claim ~700,000 deaths per year, and this mortality rate is predicted to increase to 10 million per year by 2050 [1]. Consequently, antimicrobial resistance has been identified as one of the most important challenges to human health by several national and international bodies.

For bacteria, the current antimicrobials inhibit a narrow spectrum of cellular processes (e.g., DNA replication, transcription, protein synthesis, and cell wall biosynthesis) [2]. Their action is traditionally seen as a linear process in which the antimicrobial enters the cell, reaches and interacts with its target, and stops the growth or kills the bacteria [3]. Accordingly, it is generally assumed that antimicrobial resistance relies on a few specific microbial genes [4]. However, this description represents only the immediate effects of the antimicrobial. In many cases, what is not known and is still the subject of debate, is how antimicrobials actually kill bacterial cells [5]. Microbial metabolism is impacted by antimicrobials [4], and in recent years, the close link between bacterial metabolism, antimicrobial action, and antimicrobial resistance has increasingly emerged. Indeed, on the one hand, antimicrobials alter the metabolic phenotype (metabotype [6]) of bacterial cells, giving rise to an altered metabotype. On the other hand, the metabotype of bacteria (normal or altered) influences their susceptibility to antimicrobials (Figure 1) [2]. Consequently, the antimicrobial efficacy can be enhanced by altering the metabolic state of bacteria. Approaches that target the bacterial central metabolism for drug development or that use

the control of nutrients' availability as a mechanism for the selection of antimicrobial-tolerant strains (for a definition of the differences between tolerance and resistance, see below) are gaining more and more attention [2].

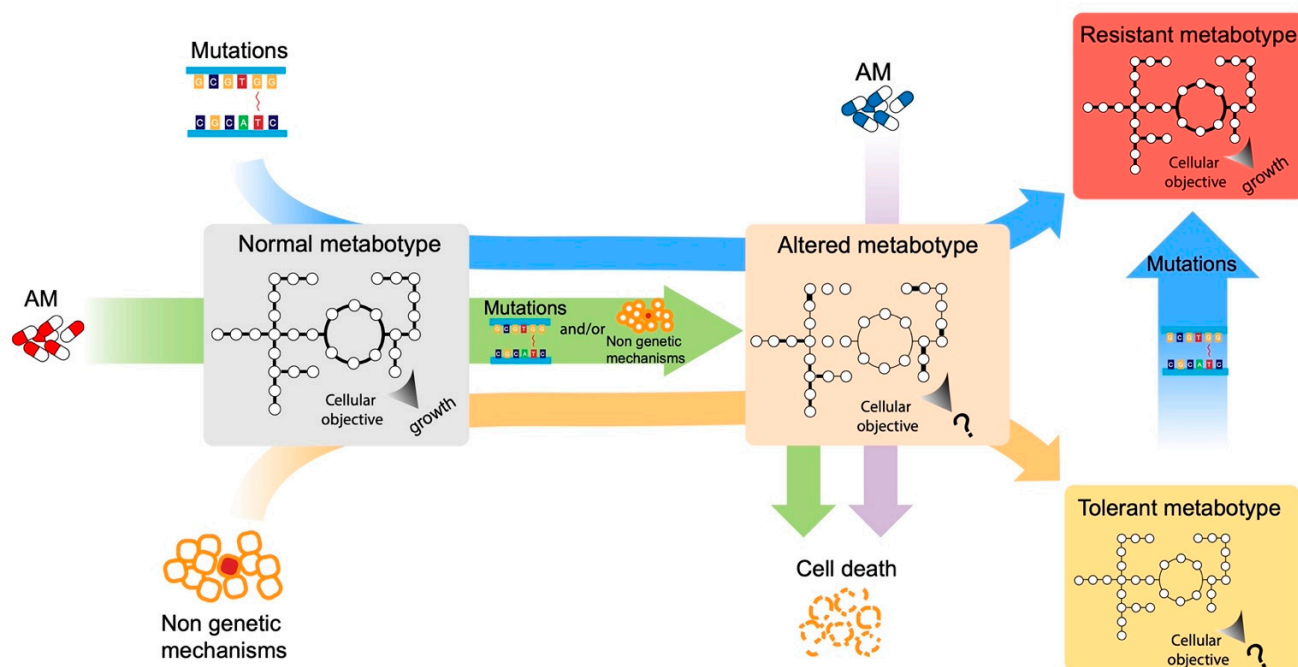


Figure 1. Possible paths to resistant and tolerant metabolotypes. Here, we show the hypothetical effects of mutations, antimicrobials' exposure, and non-genetic mechanisms on microbial metabolotypes, including the effects on the single reactions (thicker links among the nodes) and the most likely objective function ("growth" or "?").

Overall, antimicrobials can affect bacterial metabolism in three ways. First, they have a direct effect on metabolism that affects their efficacy, which seems to be different depending on whether the antimicrobial is bactericidal or bacteriostatic (even though it appears that all antimicrobials are ultimately bactericidal, and the difference between the two types of antimicrobials is only the rates at which they kill bacteria) [2,5].

The action of bactericidal antimicrobials on their primary targets causes damage to other essential macromolecules (nucleic acids, proteins, and membrane lipids) within the cell, resulting in the induction of stress response pathways. Stress responses increase metabolic activity to meet the corresponding energy demands. This results in the production of toxic metabolic byproducts such as reactive species, which damage macromolecules, and leads to the induction of additional stress response pathways. Once again, the overall cellular metabolic load results increased. The alteration of the metabolotype, therefore, creates a cyclical process that ends with cell death [7–15]. Wong et al. [16] recently suggested that the interactions of the toxic metabolic bioproducts with the membrane induce a loss of membrane integrity, which results in cytoplasmic condensation through the leakage of cytoplasmic contents, and then in cell death. Alternative cell death pathways may involve cellular damage to nucleic acids and proteins resulting from both the primary drug–target interaction and from the subsequent generation of reactive metabolic byproducts [16]. A similar process, involving reactive metabolic byproducts, seems to contribute to antimicrobial lethality, also under anaerobic conditions [17].

Bacteriostatic antimicrobials, instead, inhibit protein biosynthesis or transcription in certain contexts, leading to a decrease in metabolic activity and subsequent cell stasis, thus, again resulting in an altered metabolotype [14,18]. As mentioned before, however, bacteriostatic antimicrobials can also kill bacteria, probably depending, among other things,

on the number of ribosomes. Indeed, when the number of ribosomes is reduced, ribosome-targeting antimicrobials seem to become increasingly bactericidal, suggesting a kind of lethal protein synthesis threshold. However, this could not be the ultimate mechanism by which these ribosome-targeting ‘bacteriostatic’ antimicrobials kill bacteria [5]. Therefore, the altered metabolotypes resulting from antimicrobial treatment contribute to its final outcomes. The second and the third ways in which antimicrobials can influence bacterial metabolism are indirect. First, antimicrobial treatments generally involve the acquisition of resistance, through mutations or horizontal gene transfer, this often come with a fitness cost for the bacterial cell (Figure 1) [19,20]. Consequently, the acquisition of compensatory mutations that counterbalance the decreased fitness is a crucial step for the success of resistant strains [21,22]. These mutations generally restore normal growth, preserving resistance, and their number and type varies with the organism and the particular environmental conditions under which compensation occurs, indicating that fitness costs are dependent on the habitat and on the metabolic adaptation required for colonizing such a habitat [23]. Regarding the third mode, it has recently been demonstrated that, in addition to the acquisition of the classical mechanisms of resistance (target modification, drug inactivation, and drug transport), antimicrobials can also induce mutations in metabolic genes [24]. These metabolic mutations can confer resistance and are prevalent in clinical pathogens, suggesting that metabolic adaptation may represent a mechanism of resistance that confers tolerance, but may also mitigate the downstream toxic aspects of antimicrobials (Figure 1) [24].

On the other hand, the metabolic state of the bacterial cells can affect the antimicrobials’ efficacy in various ways. It must be stated that, in addition to the classical molecular mechanisms of antimicrobial resistance, bacterial cells can counteract antimicrobials in several ways, each of which relies on a general metabolic downregulation. Indeed, the term resistance is usually used to “describe the inherited ability of microorganisms to grow at high concentrations of an antimicrobial, irrespective of the duration of the treatment, and is quantified by the minimum inhibitory concentration (MIC) of the particular antimicrobial” [25]. If this resistant phenotype is observed only in a subpopulation of cells, it is known as hetero-resistance [26]. The term tolerance, instead, is used to describe “the ability, whether inherited or not, of microorganisms to survive transient exposure to high concentrations of an antimicrobial without a change in the MIC, which is often achieved by slowing down an essential bacterial process” [25]. It has been demonstrated that tolerance often evolves during frequent and intermittent antimicrobial treatments, and that its emergence often promotes the development of resistance [27]. Finally, the term persistence is used when only a subpopulation of a clonal bacterial population is able to survive exposure to high concentrations of an antimicrobial, without any genetic mutations [25]. Indeed, if persisters are isolated and regrown in the presence of antimicrobials, they display the same pattern of susceptibility as the original population [28].

The metabolic state of bacterial cells (normal or altered metabolotypes, Figure 1) is mainly involved in mechanisms of tolerance to the antimicrobials’ action. Indeed, most of the so-called “phenotypic resistance”, in which metabolism has a fundamental role, is mechanisms of tolerance rather than resistance. Indeed, this term refers to all the transient situations in which a bacterial population, susceptible to an antimicrobial, becomes resistant without any genetic change taking place; therefore, resulting as not inheritable [29]. For example, the growth rate is an important parameter that affects the susceptibility to antimicrobials of bacterial populations. Resting cells are less susceptible to antimicrobials than metabolically active cells, especially to the action of bactericidal antimicrobials [4]. However, a recent study showed that the metabolotype of the cell, instead of growth, better correlates with antimicrobial lethality, suggesting that antimicrobials should also be able to kill non-growing bacteria if metabolism is active, and that the metabolic response, following the initial interaction of an antimicrobial with its target, influences the bacterial response to the antimicrobial action [30].

Consequently, in all the conditions of an altered metabotype, in which the metabolism is poor or not active, bacterial cells are less susceptible to antimicrobials' action (Figure 1). For example, during a stringent response, the accumulation of (p)ppGpp determines a global switch on bacterial metabolism that also modulates their response to antimicrobials [31–34], while bacterial persisters (whose formation also involved (p)ppGpp), as mentioned above, are subpopulations of metabolically repressed cells, which survive antimicrobial treatment, although lacking genetically encoded antimicrobial resistance determinants [35–38]. Additionally, during swarming motility, bacterial cells undergo a metabolic shift which makes them less susceptible to antimicrobials, through a mechanism that is poorly understood but that might be related to changes in the cell envelope [39–43].

Another example of phenotypic resistance is growth in biofilms, microbial communities embedded in an extracellular polymeric matrix [44–46]. Bacterial cells grown in biofilm are less susceptible to antimicrobials through a combination of resistance and tolerance mechanisms. Indeed, bacterial cells in biofilm are more resistant to antimicrobials than planktonic cells due to higher levels of both spontaneous and stress-induced mutagenesis, but also of horizontal gene transfer [45]. In addition, biofilm cells are also more tolerant to antimicrobials than their planktonic counterparts through a combination of different known mechanisms, in which the metabotypes of the cells play a key role. Indeed, biofilms are characterized by a gradient of nutrients and oxygen which decrease passing from the outermost to the innermost layers [46]. Consequently, they are composed of subpopulations with different metabotypes: metabolically active populations are located on the oxygenated and richer in nutrients surface of the biofilm, while non-growing subpopulations reside in the central anoxic and low in nutrient zones [45,46]. This stratified bacterial physiology corresponds to stratified layers of susceptibility to antimicrobials, with internal cells generally more tolerant than external ones [45]. In addition, the metabolic active cells at the surface exhibit increased expression of antimicrobial resistance genes, while the metabolically inactive subpopulation exhibits reduced or negligible expression of the antimicrobial targets and a reduced antimicrobial uptake [45]. Moreover, the gradient of nutrients and oxygen also induces a progressive activation of stringent and SOS stress responses that impairs the efficacy of antimicrobials, contributing to the antimicrobial tolerance of biofilms [45]. Finally, the presence of the biofilm matrix slows antimicrobials' penetration and also favors their degradation due to the presence in the matrix of antimicrobial-modifying enzymes [45].

However, besides tolerance, bacterial metabolism is also involved in antimicrobial resistance. For example, classical elements involved in intrinsic resistance to antimicrobials, such as chromosomally encoded β -lactamases or multi-drug efflux pumps, play an important role in bacterial physiology, and they are not just an adaptive response to the presence of antimicrobials [4]. Furthermore, mutations in genes involved in cellular metabolism can contribute to intrinsic resistance [47]. Since, as stated before, bacterial metabolism is part of the antimicrobial-induced cell death process, mutations in the genes involved in these pathways (both their impairment and their overproduction) can influence the bacterial susceptibility to antimicrobials [4]. Moreover, all the regulatory genes involved in the control of metabolism, global regulators [48], signal transduction pathways controlled by two-component systems [49], and regulatory RNA [50] can influence bacterial susceptibility to antimicrobials. Finally, the metabolic adaptation that bacteria undergo during the colonization of a new environment (for example, during the early stages of host infection) might select antimicrobial-resistant bacteria (even in the absence of selective pressure with antimicrobials), thus further highlighting the existence of a tight link between bacterial metabolism and their susceptibility to antimicrobials [4].

In contrast to, for example, single mutations on specific genes that lead to antimicrobial resistance phenotypes, a causal relationship and/or mechanistic understanding of metabolism-dependent resistance/tolerance is much harder to achieve. Essentially, this is due to the inherent complexity of metabolic networks in which thousands of elements (reactions and metabolites) give rise to an intricate set of interconnections.

One of the most powerful tools to study cellular metabolism at the system level is represented by genome-scale metabolic models (GSMMs) [51]. The starting point to work with GSMMs is the reconstruction of the metabolic network of the organism of interest, starting from its genome annotation. The methods that can be implemented to reconstruct metabolic networks include KBase server, Model SEED, or CarveMe [52–54]. The resulting models are then validated using several databases (such as KEGG, PubMed, and BiGG, the more the better [55–57]), genomes from related microorganisms, and further manual curation using experimental data as potential benchmarks. Ultimately, GSMMs are formal representations of cellular metabolism that include genes, enzymes, reactions, and metabolites, describing the associated gene–protein reaction (GPR) rules [58]. This facilitates computation and prediction of phenotypes through techniques such as Flux Balance Analysis (FBA) [59]. Using the optimization of an objective function of interest (usually cellular growth), these models are capable of predicting the activity (rate) of each reaction in the model. Despite being far from flawless, these simulations permit speculation on the role of specific reactions/pathways in the simulated conditions. Indeed, one of the most promising applications of GSMMs is the possibility, by key interventions on the model itself, to include environmental conditions in the modeling framework, i.e., researchers can opportunely tune some model parameters to represent the actual (experimental) intracellular and extracellular contexts in their GSMMs. The two most valuable examples in this context are represented by: (i) the possibility to define the nutritional landscape of the model by setting the boundaries of uptake reactions to mimic experimental nutrients' availability, and (ii) the use of transcriptomic (as well as other -omics) data to define the set of enzymes that the cell is actually expressing *in vivo*. Both these methodologies constrain the model and create context-specific representations of the bacterial metabolism, according to the conditions used to obtain such extra information (nutrients' availability and gene/protein/metabolite abundance). Since the first model ever developed, for *Haemophilus influenzae* [60], several groups have continued with this task, constructing models for other bacteria. Originally, GSMM have been exploited in metabolic engineering of biotechnologically relevant strains, as the optimization of natural products encoded by biosynthetic gene clusters (BGCs) [61,62].

GSMMs, however, have been progressively applied to many other areas of research. Recently, several works have illustrated the tools for the development of metabolic networks of Gram-negative pathogens and the metabolism of priority pathogens reported by the World Health Organization, the possible drug targets for them (antimicrobial pharmacology), and the awareness of the spread of antimicrobial resistance pathogens [63–65]. Here, we rather focus on the study of the intertwined relationship between microbial metabolism and antimicrobial resistance (including the development of novel drug targets) [66], and through a survey of currently available case studies, we show how GSMMs can improve its understanding.

2. Case Studies

We have performed an operational classification of available case studies on the use of GSMMs to study antimicrobial (AM) exposure/resistance into antimicrobials: non-informed vs. informed (AM-non-informed vs. AM-informed) works. In the first group, we include all those works that did not use experimental data obtained from cells exposed to antimicrobial compounds to constrain and refine the model predictions. Conversely, all the studies that used experimental data (mostly -omics data) obtained from cells exposed to the antimicrobial molecule(s) to constrain the GSMM of the corresponding microorganisms (Figure 2) will fall in the second group. This latter group of studies will include, for example, those that used transcriptomic data obtained from microbial cells exposed to the antimicrobial compound to “inform” the model on which reactions can be used to represent the antimicrobial-exposed metabolic network. In other words, this approach permits to “turn on/off” those reactions whose genes are/are not expressed or active during the exposure to the antimicrobials.

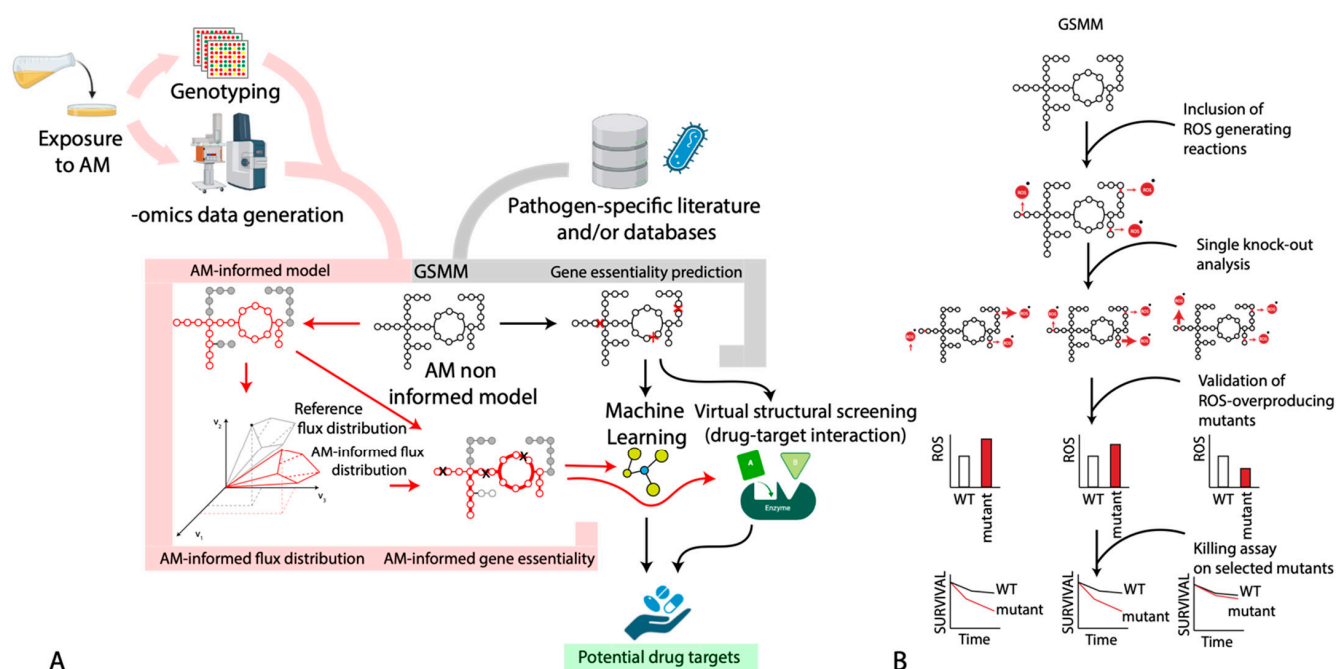


Figure 2. Schematic representation of AM-informed and AM-non-informed GSMM. **(A)** The combination of automatic metabolic reconstruction tools, literature, and strain-specific experiments leads to the creation of general, strain-specific GSMM models that can be further used for downstream simulations, leading, for example, to the prediction of possible drug targets. The integration with -omics data obtained from ad hoc designed experiments permits the generation of context-specific GSMM, capable of mimicking in a more reliable way the metabolic state of cells in the tested experimental conditions. Such AM-informed models can then be used to study the differences in flux distributions under such stressful conditions and/or to obtain to a more focused identification of condition-specific gene essentiality patterns. **(B)** The approach proposed by Brynildsen et al. [67], to demonstrate the link between ROS production and antibiotic efficacy. Including ROS-producing reactions in the original *Escherichia coli* GSMM and using simulations of gene knockouts, they identified targets (genes) whose removal would lead to an increased intracellular production of ROS. These mutant strains were then (successfully) tested for increased ROS production and increased sensitivity to antibiotics' killing.

All the metabolic reconstructions cited in this work and that have been used in anti-microbial resistance studies are reported in Supplementary Material Table S1.

2.1. AM-Non-Informed Models

Studying the metabolism of bacteria with the aim to prevent or fight AMR can be carried out by performing analyses such as gene/reaction or metabolite essentiality (GE and ME, respectively) in order to identify possible points of metabolic weakness of the cell. This approach removes a gene/reaction (or a metabolite in case of ME analysis) from the metabolic network and then simulates growth with FBA. In other words, every time a gene, reaction, or metabolite is deleted, an FBA is performed using the biomass production as the objective function. In 2009, in one of the first works of this kind, the *Acinetobacter baumannii*'s AYE GSMM was reconstructed and used in combination with a filtering framework of essential metabolites (EMFilter) to predict the most effective drug targets [68]. The first part of the analysis predicted 211 essential metabolites and, on this pool, the EMFilter was applied. The EMFilter is an algorithm that follows four steps. First, it removes currency metabolites that participate in reactions shared by several organisms, such as ATP and NADH. In the *A. baumannii* AYE case study, 179 metabolites remained after this step. Second, it selects essential metabolites that are surrounded by at least

two metabolite-consuming reactions, in order to enhance the drug impact (97 metabolites remained). Third, it prunes essential metabolites that participate in human metabolism as well (22 metabolites remained). Fourth, it removes essential metabolites that are being consumed by human homologous enzymes, in order to avoid the possible interaction between the drug and the human (rather than the microbial) drug target. After all the steps of the filtering pipeline, the enzymes involved in the consumption of nine essential metabolites were predicted as final drug-target candidates. As a confirmation, some of the enzymes identified by this approach have been previously reported as drug targets.

Analogously to the *A. baumannii* AYE case study, Jenior's group [69] decided to use this methodology to explore the possible metabolic pathway targets of *Clostridioides difficile*. They reconstructed the GSMs of two strains (630 and R20291), and after the GE analysis, they compared the prediction of essential genes for growth in silico and in vitro, obtaining 89.1% and 88.9% accuracies, respectively. This group identified the minimum subset of metabolites necessary for growth, and subsequently, the carbon sources needed. Since *C. difficile* utilizes phase variation to generate phenotypic (rough- or smooth-edge colony morphology) and metabolic heterogeneity to maximize its fitness and virulence expression, Jenior's group aimed to characterize the differences between in vitro and in vivo conditions using context-specific metabolic reconstructions derived from expression data. They found differences in alanine transport and utilizations, as well as in glycolysis' transporters and metabolic reactions during periods of increased virulence expression. With the gene essentiality analysis, they found numerous reactions that were essential only for the smooth phenotype (especially in the pentose phosphate pathway, PPP), and surprisingly, no unique essential reactions for the rough phenotype. Finally, by comparing the predicted core metabolic activity of high- and low-sporulation conditions, Jenior's group described how both N-acetylneuraminate and cytidine are being consumed during infection and decrease sporulation. These findings could help to find solutions to AM use in order to mitigate *C. difficile* infections.

López-López's group [70] also performed gene essentiality analysis in order to study what could be the best way to develop drugs for ampicillin-resistant *H. influenzae*. They refined its previous GSM, and after the GE analysis, they found 11 genes of the fatty acid biosynthesis pathway (FASII pathway), 6 genes for the phospholipid biosynthesis pathway, and 12 genes encoding enzymes involved in lipid A biosynthesis. Since the FASII pathway generates acyl carrier proteins (key components of the bacterial membrane), these authors proposed that genes related to this pathway could work as a drug target, specifically, beta-ketoacyl-ACP, which initiates fatty acid elongation cycles.

To link the (complex) carbon sources that one microbe might encounter in vivo and the overall metabolic reprogramming during the infection process, Payne et al. [71] studied the metabolic responses of *Pseudomonas aeruginosa* to different mucins. Mucins are polymeric, highly glycosylated proteins produced by airway epithelial cells and submucosal glands that play a key role in the microbe clearance and infection prevention. Even though these proteins are not considered as antimicrobial agents, it is clear that the mucin-drug interaction may have a remarkable impact on drug absorption since mucus is the first barrier that drugs must overcome to be adsorbed [72]. This impact could be at the metabolic level, i.e., by changing the metabolic state of bacteria. Payne's group developed a high-quality GSM for *P. aeruginosa* (iPau21) that has 40 genes, 24 metabolites, and 76 reactions more than the previous model, iPau1129. This new high-quality model was constrained using transcriptomic data from the literature of four different conditions: casamino acids (no mucin exposure), MUC5AC, MUC5B, and mucin-glycans, all of which are present in different body parts prone to *P. aeruginosa* infection. Since the experiments were carried out with strain PAO1, its genes in the transcriptomic dataset were mapped to PA14 orthologs, and then the data were integrated with the iPau21 model. The integration of the data was performed using RIPTiDe [73], a tool developed to integrate transcriptomic abundances using parsimony of overall fluxes to identify the most cost-effective usage of metabolism that best reflects the maximization of the biomass production, which was the objective

function. These four conditions led to four context-specific metabolic models that had a similar growth rate (less than five percent change). Payne's group compared the metabolic mechanisms of all four conditions using a non-metric multidimensional scaling (NMDS). It showed that, even when there was no significant difference in the growth rate, exposure to mucin MUC5B caused the largest shift in the metabolic pathways utilized for growth, with MUC5AC in second place. Specifically, the fumarate and propionate metabolism differed between the control and the MUC5B models. This suggested that such a metabolic shift might be involved in the resistance mechanism of *P. aeruginosa*. Thus, they proposed AMs that target proteins for fumarate and propionate metabolism in order to combat mucin-resistant bacteria. Dahal et al. [74] recently refined the *P. aeruginosa* model (iSD1509) and added a whole new pathway for ubiquinone-9 biosynthesis, required for anaerobic growth. They validated the model in its totality using experimental data from different works, including Dunphy's [75]. In order to understand how the metabolism could change under AM stress, Dahal et al. [74] performed flux sampling analysis using the biomass flux as an objective function, but constrained to 90% of the biomass flux value obtained by the FBA simulation, to mirror the presence of AM stress. In other words, the new biomass flux was constrained to $0.9 \times$ the biomass flux value from FBA. This method aims to obtain several flux sets from the space of solutions in order to obtain a more robust result for each flux sustained by each reaction, and they indeed confirmed previous results in that the TCA cycle was found to be significantly upregulated in the fumarate-supplemented medium compared to the glyoxylate-supplemented one. Hence, the model correctly differentiated between metabolites that increased drug lethality versus those that did not and offered mechanistic explanations for these responses. In conclusion, confirming Payne's group's results [71], supplementing the medium with fumarate caused a higher oxygen uptake rate due to a higher TCA cycle activity, leading to a higher drug efficiency.

Just as Payne et al. addressed the environmental–drug interaction scenario, evaluating how mucins change the metabolic state of bacteria, Chung et al. [76] proposed a different, metabolism-aware way to treat the multidrug-resistant (MDR) *Klebsiella pneumoniae* using polymyxin B in low doses, plus exogenous metabolites as adjuvants, which have a lower risk of resistance development. Chung's group reconstructed the GSMM of four different *K. pneumoniae* strains, one polymyxin-resistant, one polymyxin-susceptible, and two polymyxin-susceptible but MDR. They studied seven different metabolites that were previously identified as significant metabolites perturbed by the combination with polymyxin B, at the level of both gene expression and metabolism of the same pathway. In silico predictions using flux sampling (i.e., without imposing an arbitrary objective function) revealed that the presence of at least three of these exogenous metabolites (3-phosphoglycerate (3PG), ribose 5-phosphate (R5P), and uridine 5'-diphospho-N-acetylglucosamine (UACGAM)) enhanced the central metabolism (higher fluxes of glycolysis, PPP, and TCA cycle), increasing bacterial growth. Since the growth rate is one of the factors that determines the phenotype of susceptibility to antimicrobials, with a slow growth rate being associated with low antimicrobial activity, 3PG, R5P, and UACGAM were good candidates for a non-antimicrobial plus polymyxin treatment. Chung's group confirmed these results with time-kill studies, where bacteria displayed a decrease in the growth rate when using polymyxin B plus 3PG, R5P, or UACGAM, but not for the rest of the metabolites. Ultimately, this study provided evidence of the possibility that an altered metabolism could change the antimicrobial efficacy, also promoting the use of state-of-the-art modeling approaches to fully comprehend molecular adaptations to antimicrobial exposure.

An additional methodology to look for potential drug targets in different organisms is the combination of flux simulation-based techniques (e.g., FBA) with protein structure-based ones, such as structure-based virtual screening (SBVS). The latter is one of the most promising in silico techniques for drug design due to its robustness, and it is capable of predicting the most likely interaction between two molecules to form a stable complex [77], e.g., the drug and its target. Cesur and colleagues [78] used this protocol to find potential drug targets in *K. pneumoniae* through the constraint-based analysis of its GSMM. They

simulated three host microenvironments by setting specific bounds to each reaction flux and set the biomass production as the objective function. Combining GE and ME analyses on this GSMM, Cesur's group performed a preliminary screening that led to the identification of possible enzymatic targets, for which they identified the cellular compartments/function and performed a "druggability" assessment, i.e., to estimate the affinity of the protein to bind drug-like chemical compounds. Out of this selection procedure, one enzyme (encoded by the *kdsA* gene) stood out. For this target, using the computer-aided drug design approach (SBVS), Cesur's group obtained two possible compounds, one of which showed to be quite promising as it is a derivative of coumarin with low acute toxicity and already proven high antimicrobial activity. Nazarshodeh and colleagues [79] also used the SBVS method to identify drug effectiveness, but for *E. coli*. To this aim, they used a GSMM-PRO, i.e., a genome-scale metabolic model integrated with protein structures. Combining GE analysis (biomass production as the objective function) with a search for informative protein–ligand complexes, they obtained 70 essential genes, accounting for 92 protein–ligand complexes, which were proposed as drug targets. To perform SBV screening, the authors used the 3D structures of FDA-approved drugs and ranked them against each essential target (ideally more than one for potential polypharmacology cases). Finally, their pipeline suggested eight known antimicrobials as possible drugs to target the essential genes found, proposing a new therapeutic indication for these reported drugs (drug repurposing).

A possible advancement for this work is represented by the simulation of drug epistasis, as drug combinations can illuminate new possible therapeutic strategies that might be inaccessible to single-target drugs, possibly eliminating functional redundancies exhibited by metabolic networks. In 2016, Krueger's group [80] studied the possibility of extending FBA modeling in order to simulate drug effects over multiple doses. The methodologies implemented by this group (FBA-res and FBA-div) optimize the biomass production and account for the possibility that when a drug inhibits a reaction, (i) accumulated mass will slow down the upstream reaction thermodynamically (FBA-res), or that (ii) upstream reactions cannot sense or adjust to an inhibited downstream reaction on the timescale of the drug inhibition, and the substrate is wasted, creating less biomass over time (FBA-div). While these two theoretical approaches led to similar results to knockout simulations for single-agent effects, they had very different predictions for combinations with the approach simulating a flux-diversion to a waste reservoir (FBA-div), predicting potent antimicrobial synergies targeting metabolism. Indeed, FBA-div was able to model the strong antimicrobial synergies that target serial enzymes within a pathway, being able to predict synergies between three or more drugs.

2.2. AM-Informed Models

One of the most worrisome threats in clinical microbiology is the increasing frequency of colistin resistance in *A. baumannii* [81]. Besides the important quest of new strategies to tackle the issue, we also chose this case study because, in our opinion, the work performed in the last years to tackle the issue of AMR in *A. baumannii* nicely underlines two main features of genome-scale metabolic modeling: (i) the scalability and versatility of GSMMs to be used with different experimental (-omics) datasets (even from previous studies, an important feature in the context of data reuse), and (ii) the possibility to constantly update and (possibly) improve pre-existing models once new data are available. In 2017, to address new possible solutions for the treatment of infected patients and to overcome the appearance of resistant phenotypes, Presta and colleagues [82] reconstructed the model of AMR *A. baumannii* ATCC 19606 to perform a systems-level study of the antimicrobial response in this bacterium. Specifically, to propose new targets for AM drugs, they performed GE analysis. Since the exposure to the antimicrobial might alter the metabolotype (Figure 1), the authors used MADE (Metabolic Adjustment by Differential Expression) [83] to integrate the transcriptomic data obtained during colistin exposure (after 15 and 60 min of exposure [84]) with *A. baumannii*'s GSMM, leading [84] to four context-specific metabolic models (treated and untreated conditions at 15 and 60 min). After an FBA optimization using biomass

production as the objective function, they compared the flux distributions for each treated part and its untreated counterpart using the flux ratio and looked for the pathways these reactions belonged to. The changes were mainly observed at the level of sugar and nucleotide metabolism and fatty acid biosynthesis. Besides the analysis of general trends, the combination of transcriptomic data and GSMM might in principle lead to the identification of those genes that become essential under specific growth conditions (context-specific essential genes), in this case following the exposure to colistin. In other words, the proteins encoded by these genes might represent promising drug targets when coupled with colistin treatment. Analyzing the essential genes for each condition, they found 9 and 5 genes that are essential only after 15 and 60 min of colistin exposure, respectively, and removing the genes with human orthologues (in order to exclude them as drug targets), they finally identified five essential genes, whose proteins could represent specific drug targets.

The model reconstructed in this work (iLP844) led to further studies of *A. baumannii*'s metabolic network. Indeed, it was taken as a reference for the reconstruction of an improved model of the same strain (ATCC19606) [85] and for an *A. baumannii* AYE strain model (iCN718) [86]. Both groups used iLP844 to compare and validate their own model, by analyzing the shared genes and its growth on different carbon sources. Additionally, Norsigian's group [86] used it to check and confirm reaction reversibility, and to better describe the transport reactions present in both models. This refinement allowed to interrogate the model on the interplay of multiple metabolic pathways under colistin treatment in *A. baumannii* and to shed light on some key points to optimize polymyxin combination therapy.

Another example is that of Banerjee and Raghunathan [87], who studied the case of metabolic reprogramming when *Chromobacterium violaceum*, an opportunistic human pathogen, is exposed to antimicrobials. They reconstructed *C. violaceum*'s model (iDB858). After experimentally measuring several cell rates, they constrained different uptake values of the model (glucose, violacein secretion, molar growth rate, ATPM, and oxygen) to represent three different conditions: wildtype and exposure to chloramphenicol and to streptomycin, both together and independently. Using the biomass production as the objective function, performing FBA, FVA, and GE, they described a rewired central and redox metabolism in the presence of each antimicrobial; specifically, high levels of NAD recycling provided by pyruvate, suggesting a reprogramming of metabolism to compensate for the stress consequences.

Researchers can take advantage of the already developed GSMMs for the organism under investigation. In this case, the integration of -omics data for the reconstruction of context-specific models (representing different metabolotypes) can be performed in a more or less straightforward fashion. This is the case of Dunphy's group [75], who used iPau1129 *P. aeruginosa*'s GSMM [88] and lab-evolved AMR lineages of carbon source uptake data from a previous work [89] to contextualize the potential impact of gene deletion in antimicrobial-resistant mutations (by setting the upper and lower bounds of the model according to the experimental data). Dunphy and colleagues profiled the metabolotypes of four AMR lineages of *P. aeruginosa*: a lab-evolved phenotype, and piperacillin- (PIP), tobramycin- (TOB), and ciprofloxacin (CIP)-resistant strains. Each strain had been grown on 190 unique carbon sources, but there were only 14 growth-supporting carbon sources common for all four lineages, revealing a strong metabolic reprogramming following the exposure to each of the tested compounds. Since in Yen's work they found that the PIP-evolved lineage contained a large deletion of 343 genes, much higher than for the other lineages, Dunphy's group decided to analyze the role of each gene deletion and its impact on the metabolic functions and cellular growth. They did so by performing a GE analysis in silico using the biomass production as the objective function and repeated it on M9 minimal media and 42 different carbon sources, independently. The next step was to look for the intersection between predicted essential genes and the mutated genes in the PIP-evolved strain in order to predict resistance-specific essential genes and to identify incorrect model predictions. In conclusion, although the iPau1129 GSMM did not include several AMR

genes, it was possible for Dunphy's group to identify potential mutations impacting growth phenotypes, such as the loss of L-leucine catabolism in PIP-evolved metabolotypes.

Not surprisingly, most of the research concerning the effect of the cellular metabolic state on antimicrobial resistance emergence and/or antimicrobial efficacy has been carried out on *Escherichia coli*. In 2017, Zampieri et al. [23] used the most recent (to date) *E. coli* metabolic reconstruction (iJO1366) [90] in order to understand the impact of metabolic changes in the development of antimicrobial resistance. They selected three antimicrobials with different modes of action: ampicillin, chloramphenicol, and norfloxacin, and then let four independent lineages of the wildtype evolve resistance in minimal medium, either with glucose or with acetate as the sole carbon source. Then, they performed metabolome-based predictions using *E. coli*'s GSMM, and systematically maximized or minimized fluxes for each reaction of the model, i.e., they iteratively changed the objective function to calculate the shadow prices (the estimation of its sensitivity to changes in the availability of all individual metabolites). Negative shadow prices mean limiting metabolites for the reaction. In this way, reactions with overrepresentation of altered limiting metabolites could be those with an active role in the evolution of resistance. They observed that most of the evolved metabolic characteristics were present in pathways not directly affected by the antimicrobial. For the chloramphenicol-glucose-evolved populations, the evolved metabolic characteristics involved sugar transport, oxygen uptake, and CoA formation, while for the ampicillin-glucose-resistant populations, they involved nucleotide metabolism, serine biosynthesis, and cell wall recycling. The authors found an important difference between the ampicillin-glucose-evolved and ampicillin-acetate-evolved populations, with the first one being eight times more sensitive to fosfomycin, and the latter more tolerant to it. Thus, apparently, metabolic adaptation to antimicrobial exposure strongly depends on the nutritional context.

Recently, several groups have proposed the combination of GSMM and machine learning (ML) in order to broaden the space of predictions. ML is a method that handles the automatic learning of machines without explicit programming (black box) and has been widely used in the field of bioinformatics, and recently, in systems biology as well [91]. For example, in 2020, Kavvas et al. [92] proposed the integration of these two methods to obtain results to a higher resolution, i.e., not just to analyze gene presence-absence, but also to analyze the links of allele variations to observed phenotypes, such as antimicrobial resistance. Kavvas' group worked with *Mycobacterium tuberculosis* (TB) and built a genetic variant matrix (G) containing the information of all drug-tested TB strains and all their allelic variants. The abundance of AMR genes in the latest TB model (iEK1011) [93] justified the use of this model for their analysis. The Metabolic Allele Classifier (MAC), as they named it, is an allele-parameterized form of FBA, which takes the following as constraints: (i) an antimicrobial-specific objective function, (ii) a steady-state space, and (iii) upper and lower bounds for each reaction flux, in function to the information in G. The last two points describe strain-specific separate polytopes for resistance and susceptible strains within the overall flux space, and the first point is defined by the objective function that best separates both polytopes, where the optimal solution is the orthogonal vector to the plane that divides them, which is calculated with ML instead of the linear program. The MAC yields a drug susceptibility binary status as a final output: "susceptible" or "resistance" phenotype to a particular drug. Kavvas et al. tested three different AM compounds: pyrazinamide, para-aminosalicylic acid, and isoniazid, and performed pathway enrichment analysis out of the 197 alleles obtained, resulting in the identification of processes that were known AM mechanisms. Overall, Kavvas' group's work shed light on the possibility of studying allele variations-phenotype relations in a determined strain using just one GSMM of the same bacterium. Another example of the integration of the two methods (GSMM + ML) was reported by Percy et al. [94]. These authors aimed to reveal and describe the systemic relationship between the genetic determinants of AMR and metabolic evolutionary adaptations in *E. coli*. Using a set of unique *E. coli* genomes to analyze the AMR phenotypic variability, they developed a

ML-integrated k-mer and SNP approach: they scanned entire genomes against selected phenotypes, allowing for the identification of arbitrary numbers of genomic features ranked on the strength of the correlation with the AMR and the susceptible phenotype (named the Gradient Boosting Classifier). These genetic determinants, such as AMR genes and specific mutations and alleles to specific phenotypes, were assessed using the GSMM: they used FBA to constrain the model and predict the effects (different metabolotypes, see Figure 1) of the genetic determinants previously identified by performing gene deletion analysis in the metabolic network. Pearcy's group identified 20 genes essential for growth under rich environment conditions, and the ones that had the highest importance depicted by the ML method may be promising dual-mechanism antimicrobial targets: they would inhibit essential metabolic pathways while reducing the bacterium's ability to adapt.

A problem that is sometimes difficult to address is the impact of the combinations of different drugs and/or how the drug efficacy (or the efficacy of drug combinations) changes depending on the growth conditions. To address this issue, Chung et al. [95] proposed an approach named Condition-specific Antibiotic Regimen Assessment using Mechanistic Learning (CARAMeL). This method simulates context-specific metabolic fluxes using GSMMs, constrained with environmental conditions, -omics data, and metabolite composition from control and mono- or multi-treatments. Obtained (simulated) fluxes are then used to train a ML model to predict interaction outcomes for novel drug combinations (using the Random Forests algorithm). Using *E. coli*'s and *M. tuberculosis*' GSMMs, Chung's group assessed their methodology using several scenarios with multiple drugs and clinical trials, inferring a set of differentially regulated genes by calculating the differential fitness (from chemogenomic data) and expression (from transcriptomic data). Then, they assigned four scores for each condition and trained ML models to associate the described profiles to drug combination outcomes, as interaction scores. After comparing the predictions with experimental data (-omics and literature) and assessing the performance for synergy and antagonism scenarios, the authors' approach revealed to be accurate in predicting combination therapy outcomes. Afterwards, they assessed how a pathogen adapts to a first drug and how this, in turn, influences its sensitivity to a subsequent drug. Using data from evolved *E. coli* in single-drug treatments over different timespans and subsequent treatments, CARAMeL once again yielded robust predictions, also when addressing the model sensitivity to a broad range of drug combinations and growing media.

Even when there have been other methodologies that do not use GSMMs [96], the joint effort of each of the approaches described in this review gives a broad perspective of the relationship between the metabolic state of bacteria and the effect of an antimicrobial. As described in the previous section, the nutrient contents of the bacterial medium can alter the metabolic state of bacteria, inducing different responses under AM stress, just as the results of [81,93] showed. Additionally, integrating different types of -omics data, GSMM, and ML, defining a context-specific objective function [92,97], and/or performing GE and ME analysis not only helps to understand the different AM metabolotypes in bacteria, but also helps to shed light on the drug development and drug repurposing areas [71,74,78,79], in order to find antimicrobials able to kill bacteria with active metabolism, even if there is no growth. However, there are still some scenarios to address, such as the study of non-protein-coding regions, since it has been found that they can confer resistance as well [94,98].

Lastly, we would like to refer to a case study that used GSMMs predictions to enhance the killing activity of antibiotics. Brynildsen et al. [67,99] exploited the link existing between ROS production and antibiotic killing activity and used genome-scale metabolic modeling to identify metabolic targets that could lead to the overproduction of such ROS species and that could potentiate the action of several different antibiotics. Specifically, the metabolic network of *E. coli* was systematically perturbed (through the simulation of single-gene knockouts), and its flux distribution was analyzed to identify targets predicted to increase ROS production, i.e., genes whose removal from the WT would increase the intracellular ROS concentration. Remarkably, such computationally predicted targets were

experimentally validated and demonstrated to confer increased susceptibility to oxidants and to killing by antibiotics. A schematic representation of the overall pipeline used in this work is reported in Figure 2B.

3. Future Directions

The use of genome-scale metabolic modeling has improved our understanding of the intertwined relationship between metabolism and antimicrobials. We believe that the most useful application of GSMMs resides in using them as scaffolds for the interpretation of -omics datasets, as this helps in reducing the space of possible solutions, for example, after an FBA simulation, and to identify the most likely set of active reactions in a specific condition. Indeed, the speed and ease-of-use of constraint-based simulations comes at the cost of many (sometimes unrealistic) constraints that must be imposed on the model to make it mathematically solvable. Among them, the one that most impacts the use of genome-scale metabolic modeling for studying the response of microbial cells to cellular perturbation is probably the identification of a proper objective function, i.e., the reaction in the metabolic network that describes the primary objective of the cell in a specific growth condition [100]. While it seems reasonable to assume that, in laboratory conditions (e.g., growth in a bioreactor), the generation of new biomass would be the main cellular objective of the population, we think that this prerequisite is hardly met in a stressful condition such as the exposure to antimicrobials. For this reason, FBA simulations that encompass the use of a biomass-oriented objective function to mimic an antimicrobial-exposed/tolerant metabolite might not be capable of adequately representing the pool of active reactions in the cell. While the majority of the studies (also reviewed in this work) have not directly tackled this issue up to now, others have observed that the exposure to antimicrobials and the evolution of resistance may introduce novel and still unexplored metabolic constraints and cellular objectives. This, in turn, hampers the identification of the actual objective function(s) that drive metabolic adaptation to antimicrobial exposure [23]. Computational approaches can be used to mitigate this effect, for example, by assuming that altered metabolite concentrations in antimicrobial-resistant strains may reflect an attempt to redirect intracellular fluxes toward specific but unknown metabolic objectives to drive and compensate for resistance [23,101]. In [92], the problem of correctly capturing the cellular objective of the antimicrobial-resistant phenotypes was circumvented by integrating GWAS information with the *E. coli* metabolic model and deriving an “antimicrobial-specific objective” function that provides novel insights regarding the metabolic basis of the mutations involved in antimicrobial resistance. This was achieved by determining the set of potential constraints imposed by an allele through discretization of the flux solution space; thus, deriving the metabolotypes that more closely resemble the stressful conditions faced by cells. To tackle the same issue, i.e., the definition of a proper objective function under stressful conditions, Montezano et al. [97] studied the response of the *M. tuberculosis* bacterium when exposed to mefloquine, and integrated condition-specific proteomic data with its GSMM model to derive an antimicrobial-informed biological objective function, i.e., the actual cellular overall objective when facing antimicrobial-induced stress.

When experimental information on the cellular state in the presence of antimicrobials are not available, the exploration of the possible solution space through sampling methods (e.g., random flux sampling [101–104]) allows to define the possible set of active reactions without the necessity of defying a biomass-oriented objective function. In particular, random flux sampling can be used to characterize the solution space within a GSMM, allowing the identification of a statistically significant number of solutions that have been uniformly distributed throughout the entire solution space. In this way, an estimated probability distribution for each reaction’s flux in the network can be obtained without necessarily assuming growth as the primary objective function. Accordingly, random flux sampling might be a more reliable approach to simulate the metabolism of bacteria that are exposed to challenging conditions, such as the presence of antimicrobial molecules, in

which the survival and the maintenance of basic metabolic functions with minimal cost may be more realistic objectives.

Population-level metabolic heterogeneity (i.e., cell-to-cell variation in metabolism within an isogenic population) is known to play a key role in the response of microbes to antimicrobial stress [105], and several pathogens have been shown to take on a distinct metabolic state in vivo to tolerate antimicrobials. Canonical genome-scale metabolic modeling approaches typically represent the averaged metabolic state of entire populations and, as such, are not suited to capture single-cell metabolic differences. Additionally, this within-population variability seems to have a stochastic basis [106], another feature whose modeling with a constraint-based approach has been seldomly attempted in the past. Recently, however, computational methods to tackle both the metabolic heterogeneity within a population and the stochastic nature of certain cellular processes and their impact on metabolism have been developed [107,108]. Besides revealing important basic insights into the possible consequences of the stochasticity that is implicit in some biological circuits, these methods pave the way for their exploitation in the study of the metabolic state of bacterial cells in a population in the presence of antimicrobials. Since, for example, persisters are known for their lowered metabolic state, such an approach could indicate which pathways should be revived to reduce the fraction of dormant persister cells, and thus to develop metabolic-oriented and more effective therapies. Indeed, there could be different and independent ways through which metabolic dormancy is reached (i.e., downregulation of different metabolic pathways) within a population, and genome-scale metabolic modeling might help in identifying which of them is/are actually playing a role in persisters' survival. In this context, methods such as BacDrop [109] that permit to elucidate the heterogeneous responses of microbial cells to antimicrobial stress using single-cell RNA-sequencing on millions of single bacterial cells could permit adding a further layer of information (constraints) to the model, thus refining the predictions. Overall, there seems to be a methodological gap in the repertoire of computational approaches able to simulate the (sub-optimal) metabolic landscape of persister cells. Additional effort is thus required to develop tools and theoretical frameworks to generate testable hypotheses on the mechanisms through which tolerance/persistence is achieved at the whole population level.

As we saw in the case studies, the modeling of antimicrobial-induced metabolic changes has recently been boosted by the combination of ML methods with GSMMs. There are (at least) three different ways for integrating ML with constraint-based metabolic modeling (see Sahu et al. [110] for a thorough review), and ML can be used to identify those cellular components (e.g., genes, transcripts, metabolites) that should actually be included in the corresponding metabolic model [111,112]. Other ML methods can be directly integrated with the FBA algorithm [113]. Finally, the outcomes (flux distributions) of an FBA simulation can be analyzed and further refined by using ML. In the context of antimicrobial stress, a flux-based ML model was recently used to simulate the impact of metabolic heterogeneity on drug interactions and to predict the outcomes of novel and unseen combinations of drugs in *E. coli* and *M. tuberculosis*. It is important to stress that the relevance of such hybrid models (and of GSMMs in general) resides in their capability of providing a mechanistic understanding of how biological circuits (metabolism in this case) respond to environmental stresses/fluctuations, and which strategies are the most promising in preventing antimicrobial resistance/tolerance/persistence [114].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antibiotics12050896/s1>, Table S1: Metabolic reconstructions cited in this work and that have been used in antimicrobial resistance studies. References [23,68–71,74–76,78–80,82,87,92,94,95] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, T.A.-V., M.F., and E.P.; writing—original draft preparation, T.A.-V., M.F., and E.P.; writing—review and editing, T.A.-V., M.F., and E.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a PRIN-MUR (RESEARCH PROJECTS OF RELEVANT NATIONAL INTEREST—2020 Call), grant number: 20208LLXEJ.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Murray, C.J.; Ikuta, K.S.; Sharara, F.; Swetschinski, L.; Aguilar, G.R.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; et al. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *Lancet* **2022**, *399*, 629–655. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stokes, J.M.; Lopatkin, A.J.; Lobritz, M.A.; Collins, J.J. Bacterial Metabolism and Antibiotic Efficacy. *Cell Metab.* **2019**, *30*, 251–259. [\[CrossRef\]](#) [\[PubMed\]](#)
- Roemhild, R.; Bollenbach, T.; Andersson, D.I. The Physiology and Genetics of Bacterial Responses to Antibiotic Combinations. *Nat. Rev. Microbiol.* **2022**, *20*, 478–490. [\[CrossRef\]](#) [\[PubMed\]](#)
- Martínez, J.L.; Rojo, F. Metabolic Regulation of Antibiotic Resistance. *FEMS Microbiol. Rev.* **2011**, *35*, 768–789. [\[CrossRef\]](#) [\[PubMed\]](#)
- Baquero, F.; Levin, B.R. Proximate and Ultimate Causes of the Bactericidal Action of Antibiotics. *Nat. Rev. Microbiol.* **2021**, *19*, 123–132. [\[CrossRef\]](#)
- Gavaghan, C.L.; Holmes, E.; Lenz, E.; Wilson, I.D.; Nicholson, J.K. An NMR-Based Metabonomic Approach to Investigate the Biochemical Consequences of Genetic Strain Differences: Application to the C57BL10J and Alpk:ApfCD Mouse. *FEBS Lett.* **2000**, *484*, 169–174. [\[CrossRef\]](#)
- Adolfson, K.J.; Brynildsen, M.P. Futile Cycling Increases Sensitivity toward Oxidative Stress in *Escherichia coli*. *Metab. Eng.* **2015**, *29*, 26–35. [\[CrossRef\]](#)
- Belenky, P.; Ye, J.D.; Porter, C.B.M.; Cohen, N.R.; Lobritz, M.A.; Ferrante, T.; Jain, S.; Koray, B.J.; Schwarz, E.G.; Walker, G.C.; et al. Bactericidal Antibiotics Induce Toxic Metabolic Perturbations That Lead to Cellular Damage. *Cell Rep.* **2015**, *13*, 968–980. [\[CrossRef\]](#)
- Cho, H.; Uehara, T.; Bernhardt, T.G. Beta-Lactam Antibiotics Induce a Lethal Malfunctioning of the Bacterial Cell Wall Synthesis Machinery. *Cell* **2014**, *159*, 1300–1311. [\[CrossRef\]](#)
- Dwyer, D.J.; Belenky, P.A.; Yang, J.H.; MacDonald, I.C.; Martell, J.D.; Takahashi, N.; Chan, C.T.Y.; Lobritz, M.A.; Braff, D.; Schwarz, E.G.; et al. Antibiotics Induce Redox-Related Physiological Alterations as Part of Their Lethality. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E2100–E2109. [\[CrossRef\]](#)
- Foti, J.J.; Devadoss, B.; Winkler, J.A.; Collins, J.J.; Walker, G.C. Oxidation of the Guanine Nucleotide Pool Underlies Cell Death by Bactericidal Antibiotics. *Science* **2012**, *336*, 315–319. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hong, Y.; Zeng, J.; Wang, X.; Drlica, K.; Zhao, X. Post-Stress Bacterial Cell Death Mediated by Reactive Oxygen Species. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 10064–10071. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kohanski, M.A.; Dwyer, D.J.; Hayete, B.; Lawrence, C.A.; Collins, J.J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **2007**, *130*, 797–810. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lobritz, M.A.; Belenky, P.; Porter, C.B.M.; Gutierrez, A.; Yang, J.H.; Schwarz, E.G.; Dwyer, D.J.; Khalil, A.S.; Collins, J.J. Antibiotic Efficacy Is Linked to Bacterial Cellular Respiration. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8173–8180. [\[CrossRef\]](#)
- Vatansever, F.; de Melo, W.C.M.A.; Avci, P.; Vecchio, D.; Sadasivam, M.; Gupta, A.; Chandran, R.; Karimi, M.; Parizotto, N.A.; Yin, R.; et al. Antimicrobial Strategies Centered around Reactive Oxygen Species—Bactericidal Antibiotics, Photodynamic Therapy, and Beyond. *FEMS Microbiol. Rev.* **2013**, *37*, 955–989. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wong, F.; Stokes, J.M.; Cervantes, B.; Penkov, S.; Friedrichs, J.; Renner, L.D.; Collins, J.J. Cytoplasmic Condensation Induced by Membrane Damage Is Associated with Antibiotic Lethality. *Nat. Commun.* **2021**, *12*, 2321. [\[CrossRef\]](#)
- Wong, F.; Stokes, J.M.; Bening, S.C.; Vidoudez, C.; Trauger, S.A.; Collins, J.J. Reactive Metabolic Byproducts Contribute to Antibiotic Lethality under Anaerobic Conditions. *Mol. Cell* **2022**, *82*, 3499–3512.e10. [\[CrossRef\]](#)
- Lin, X.; Kang, L.; Li, H.; Peng, X. Fluctuation of Multiple Metabolic Pathways Is Required for *Escherichia coli* in Response to Chlorotetracycline Stress. *Mol. Biosyst.* **2014**, *10*, 901–908. [\[CrossRef\]](#)
- Dahlberg, C.; Chao, L. Amelioration of the Cost of Conjugative Plasmid Carriage in *Escherichia coli* K12. *Genetics* **2003**, *165*, 1641–1649. [\[CrossRef\]](#)
- Melnyk, A.H.; Wong, A.; Kassen, R. The Fitness Costs of Antibiotic Resistance Mutations. *Evol. Appl.* **2015**, *8*, 273–283. [\[CrossRef\]](#)
- Levin, B.R.; Lipsitch, M.; Perrot, V.; Schrag, S.; Antia, R.; Simonsen, L.; Walker, N.M.; Stewart, F.M. The Population Genetics of Antibiotic Resistance. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **1997**, *24* (Suppl. S1), S9–S16. [\[CrossRef\]](#)
- Marciano, D.C.; Karkouti, O.Y.; Palzkill, T. A Fitness Cost Associated with the Antibiotic Resistance Enzyme SME-1 Beta-Lactamase. *Genetics* **2007**, *176*, 2381–2392. [\[CrossRef\]](#)
- Zampieri, M.; Enke, T.; Chubukov, V.; Ricci, V.; Piddock, L.; Sauer, U. Metabolic Constraints on the Evolution of Antibiotic Resistance. *Mol. Syst. Biol.* **2017**, *13*, 917. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Lopatkin, A.J.; Bening, S.C.; Manson, A.L.; Stokes, J.M.; Kohanski, M.A.; Badran, A.H.; Earl, A.M.; Cheney, N.J.; Yang, J.H.; Collins, J.J. Clinically Relevant Mutations in Core Metabolic Genes Confer Antibiotic Resistance. *Science* **2021**, *371*, eaba0862. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Brauner, A.; Fridman, O.; Gefen, O.; Balaban, N.Q. Distinguishing between Resistance, Tolerance and Persistence to Antibiotic Treatment. *Nat. Rev. Microbiol.* **2016**, *14*, 320–330. [\[CrossRef\]](#)
26. Andersson, D.I.; Nicoloff, H.; Hjort, K. Mechanisms and Clinical Relevance of Bacterial Heteroresistance. *Nat. Rev. Microbiol.* **2019**, *17*, 479–496. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Sulaiman, J.E.; Lam, H. Evolution of Bacterial Tolerance under Antibiotic Treatment and Its Implications on the Development of Resistance. *Front. Microbiol.* **2021**, *12*, 617412. [\[CrossRef\]](#)
28. Balaban, N.Q.; Helaine, S.; Lewis, K.; Ackermann, M.; Aldridge, B.; Andersson, D.I.; Brynildsen, M.P.; Bumann, D.; Camilli, A.; Collins, J.J.; et al. Definitions and Guidelines for Research on Antibiotic Persistence. *Nat. Rev. Microbiol.* **2019**, *17*, 441–448. [\[CrossRef\]](#)
29. Levin, B.R.; Rozen, D.E. Non-Inherited Antibiotic Resistance. *Nat. Rev. Microbiol.* **2006**, *4*, 556–562. [\[CrossRef\]](#)
30. Lopatkin, A.J.; Stokes, J.M.; Zheng, E.J.; Yang, J.H.; Takahashi, M.K.; You, L.; Collins, J.J. Bacterial Metabolic State More Accurately Predicts Antibiotic Lethality than Growth Rate. *Nat. Microbiol.* **2019**, *4*, 2109–2117. [\[CrossRef\]](#)
31. Srivatsan, A.; Wang, J.D. Control of Bacterial Transcription, Translation and Replication by (p)PpGpp. *Curr. Opin. Microbiol.* **2008**, *11*, 100–105. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Pesavento, C.; Hengge, R. Bacterial Nucleotide-Based Second Messengers. *Curr. Opin. Microbiol.* **2009**, *12*, 170–176. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Wu, J.; Long, Q.; Xie, J. (P)PpGpp and Drug Resistance. *J. Cell. Physiol.* **2010**, *224*, 300–304. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Jain, V.; Kumar, M.; Chatterji, D. PpGpp: Stringent Response and Survival. *J. Microbiol. Seoul Korea* **2006**, *44*, 1–10.
35. Lewis, K. Persister Cells. *Annu. Rev. Microbiol.* **2010**, *64*, 357–372. [\[CrossRef\]](#)
36. Balaban, N.Q.; Merrin, J.; Chait, R.; Kowalik, L.; Leibler, S. Bacterial Persistence as a Phenotypic Switch. *Science* **2004**, *305*, 1622–1625. [\[CrossRef\]](#)
37. Shah, D.; Zhang, Z.; Khodursky, A.; Kaldalu, N.; Kurg, K.; Lewis, K. Persisters: A Distinct Physiological State of *E. Coli*. *BMC Microbiol.* **2006**, *6*, 53. [\[CrossRef\]](#)
38. Prax, M.; Bertram, R. Metabolic Aspects of Bacterial Persisters. *Front. Cell. Infect. Microbiol.* **2014**, *4*, 148. [\[CrossRef\]](#)
39. Kim, W.; Killam, T.; Sood, V.; Surette, M.G. Swarm-Cell Differentiation in *Salmonella enterica* Serovar Typhimurium Results in Elevated Resistance to Multiple Antibiotics. *J. Bacteriol.* **2003**, *185*, 3111–3117. [\[CrossRef\]](#)
40. Overhage, J.; Bains, M.; Brazas, M.D.; Hancock, R.E.W. Swarming of *Pseudomonas aeruginosa* Is a Complex Adaptation Leading to Increased Production of Virulence Factors and Antibiotic Resistance. *J. Bacteriol.* **2008**, *190*, 2671–2679. [\[CrossRef\]](#)
41. Yeung, A.T.; Torfs, E.C.; Jamshidi, F.; Bains, M.; Wiegand, I.; Hancock, R.E.; Overhage, J. Swarming of *Pseudomonas aeruginosa* Is Controlled by a Broad Spectrum of Transcriptional Regulators, Including MetR. *J. Bacteriol.* **2009**, *191*, 5592–5602. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Kim, W.; Surette, M.G. Swarming Populations of *Salmonella* Represent a Unique Physiological State Coupled to Multiple Mechanisms of Antibiotic Resistance. *Biol. Proced. Online* **2003**, *5*, 189–196. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Irazoki, O.; Campoy, S.; Barbé, J. The Transient Multidrug Resistance Phenotype of *Salmonella enterica* Swarming Cells Is Abolished by Sub-Inhibitory Concentrations of Antimicrobial Compounds. *Front. Microbiol.* **2017**, *8*, 1360. [\[CrossRef\]](#)
44. Sauer, K.; Stoodley, P.; Goeres, D.M.; Hall-Stoodley, L.; Burmølle, M.; Stewart, P.S.; Bjarnsholt, T. The Biofilm Life Cycle: Expanding the Conceptual Model of Biofilm Formation. *Nat. Rev. Microbiol.* **2022**, *20*, 608–620. [\[CrossRef\]](#)
45. Ciofu, O.; Moser, C.; Jensen, P.Ø.; Høiby, N. Tolerance and Resistance of Microbial Biofilms. *Nat. Rev. Microbiol.* **2022**, *20*, 621–635. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Jo, J.; Price-Whelan, A.; Dietrich, L.E.P. Gradients and Consequences of Heterogeneity in Biofilms. *Nat. Rev. Microbiol.* **2022**, *20*, 593–607. [\[CrossRef\]](#)
47. Fajardo, A.; Linares, J.F.; Martínez, J.L. Towards an Ecological Approach to Antibiotics and Antibiotic Resistance Genes. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **2009**, *15* (Suppl. S1), 14–16. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Corona, F.; Martinez, J.L. Phenotypic Resistance to Antibiotics. *Antibiotics* **2013**, *2*, 237–255. [\[CrossRef\]](#)
49. Bhagirath, A.Y.; Li, Y.; Patidar, R.; Yerex, K.; Ma, X.; Kumar, A.; Duan, K. Two Component Regulatory Systems and Antibiotic Resistance in Gram-Negative Pathogens. *Int. J. Mol. Sci.* **2019**, *20*, 1781. [\[CrossRef\]](#)
50. Dersch, P.; Khan, M.A.; Mühlen, S.; Görke, B. Roles of Regulatory RNAs for Antibiotic Resistance in Bacteria and Their Potential Value as Novel Drug Targets. *Front. Microbiol.* **2017**, *8*, 803. [\[CrossRef\]](#)
51. Kim, W.J.; Kim, H.U.; Lee, S.Y. Current State and Applications of Microbial Genome-Scale Metabolic Models. *Curr. Opin. Syst. Biol.* **2017**, *2*, 10–18. [\[CrossRef\]](#)
52. 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.; et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Henry, C.S.; DeJongh, M.; Best, A.A.; Frybarger, P.M.; Lindsay, B.; Stevens, R.L. High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models. *Nat. Biotechnol.* **2010**, *28*, 977–982. [\[CrossRef\]](#) [\[PubMed\]](#)

54. Machado, D.; Andrejev, S.; Tramontano, M.; Patil, K.R. Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities. *Nucleic Acids Res.* **2018**, *46*, 7542–7553. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2022**, *50*, D20–D26. [\[CrossRef\]](#)
57. King, Z.A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J.A.; Ebrahim, A.; Palsson, B.O.; Lewis, N.E. BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models. *Nucleic Acids Res.* **2016**, *44*, D515–D522. [\[CrossRef\]](#)
58. Passi, A.; Tibocha-Bonilla, J.D.; Kumar, M.; Tec-Campos, D.; Zengler, K.; Zuniga, C. Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data. *Metabolites* **2022**, *12*, 14. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Orth, J.D.; Thiele, I.; Palsson, B.O. What Is Flux Balance Analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248. [\[CrossRef\]](#)
60. Edwards, J.S.; Palsson, B.O. Systems Properties of the Haemophilus Influenzae Rd Metabolic Genotype. *J. Biol. Chem.* **1999**, *274*, 17410–17416. [\[CrossRef\]](#)
61. Saini, D.K.; Rai, A.; Devi, A.; Pabbi, S.; Chhabra, D.; Chang, J.-S.; Shukla, P. A Multi-Objective Hybrid Machine Learning Approach-Based Optimization for Enhanced Biomass and Bioactive Phycobiliproteins Production in *Nostoc* sp. CCC-403. *Bioresour. Technol.* **2021**, *329*, 124908. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Swayambhu, G.; Moscatello, N.; Atilla-Gokcumen, G.E.; Pfeifer, B.A. Flux Balance Analysis for Media Optimization and Genetic Targets to Improve Heterologous Siderophore Production. *iScience* **2020**, *23*, 101016. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Chung, W.Y.; Zhu, Y.; Mahamad Maifiah, M.H.; Shivashekaregowda, N.K.H.; Wong, E.H.; Abdul Rahim, N. Novel Antimicrobial Development Using Genome-Scale Metabolic Model of Gram-Negative Pathogens: A Review. *J. Antibiot.* **2021**, *74*, 95–104. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Zhu, Y.; Zhao, J.; Li, J. Genome-Scale Metabolic Modeling in Antimicrobial Pharmacology. *Eng. Microbiol.* **2022**, *2*, 100021. [\[CrossRef\]](#)
65. Sertbas, M.; Ulgen, K.O. Genome-Scale Metabolic Modeling for Unraveling Molecular Mechanisms of High Threat Pathogens. *Front. Cell Dev. Biol.* **2020**, *8*, 566702. [\[CrossRef\]](#)
66. Fang, X.; Lloyd, C.J.; Palsson, B.O. Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications. *Nat. Rev. Microbiol.* **2020**, *18*, 731–743. [\[CrossRef\]](#)
67. Brynildsen, M.P.; Winkler, J.A.; Spina, C.S.; MacDonald, I.C.; Collins, J.J. Potentiating Antibacterial Activity by Predictably Enhancing Endogenous Microbial ROS Production. *Nat. Biotechnol.* **2013**, *31*, 160–165. [\[CrossRef\]](#)
68. Kim, H.U.; Kim, T.Y.; Lee, S.Y. Genome-Scale Metabolic Network Analysis and Drug Targeting of Multi-Drug Resistant Pathogen *Acinetobacter baumannii* AYE. *Mol. Biosyst.* **2010**, *6*, 339–348. [\[CrossRef\]](#)
69. Jenior, M.L.; Leslie, J.L.; Powers, D.A.; Garrett, E.M.; Walker, K.A.; Dickenson, M.E.; Petri, W.A.; Tamayo, R.; Papin, J.A. Novel Drivers of Virulence in Clostridioides Difficile Identified via Context-Specific Metabolic Network Analysis. *Msystems* **2021**, *6*, e00919–21. [\[CrossRef\]](#)
70. López-López, N.; León, D.S.; de Castro, S.; Díez-Martínez, R.; Iglesias-Bexiga, M.; Camarasa, M.J.; Menéndez, M.; Nogales, J.; Garmendia, J. Interrogation of Essentiality in the Reconstructed Haemophilus Influenzae Metabolic Network Identifies Lipid Metabolism Antimicrobial Targets: Preclinical Evaluation of a FabH β -Ketoacyl-ACP Synthase Inhibitor. *Msystems* **2022**, *7*, e01459–21. [\[CrossRef\]](#)
71. Payne, D.D.; Renz, A.; Dunphy, L.J.; Lewis, T.; Dräger, A.; Papin, J.A. An Updated Genome-Scale Metabolic Network Reconstruction of *Pseudomonas aeruginosa* PA14 to Characterize Mucin-Driven Shifts in Bacterial Metabolism. *NPJ Syst. Biol. Appl.* **2021**, *7*, 37. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Butnaru, C.; Barbero, N.; Pacheco, D.; Petrini, P.; Visentin, S. Mucin Binding to Therapeutic Molecules: The Case of Antimicrobial Agents Used in Cystic Fibrosis. *Int. J. Pharm.* **2019**, *564*, 136–144. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Jenior, M.L.; Moutinho, T.J., Jr.; Dougherty, B.V.; Papin, J.A. Transcriptome-Guided Parsimonious Flux Analysis Improves Predictions with Metabolic Networks in Complex Environments. *PLoS Comput. Biol.* **2020**, *16*, e1007099. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Dahal, S.; Renz, A.; Dräger, A.; Yang, L. Genome-Scale Model of *Pseudomonas aeruginosa* Metabolism Unveils Virulence and Drug Potentiation. *Commun. Biol.* **2023**, *6*, 165. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Dunphy, L.J.; Yen, P.; Papin, J.A. Integrated Experimental and Computational Analyses Reveal Differential Metabolic Functionality in Antibiotic-Resistant *Pseudomonas aeruginosa*. *Cell Syst.* **2019**, *8*, 3–14.e3. [\[CrossRef\]](#)
76. Chung, W.Y.; Abdul Rahim, N.; Mahamad Maifiah, M.H.; Hawala Shivashekaregowda, N.K.; Zhu, Y.; Wong, E.H. In Silico Genome-Scale Metabolic Modeling and in Vitro Static Time-Kill Studies of Exogenous Metabolites Alone and with Polymyxin B against *Klebsiella pneumoniae*. *Front. Pharmacol.* **2022**, *13*, 880352. [\[CrossRef\]](#)
77. Maia, E.H.B.; Assis, L.C.; de Oliveira, T.A.; da Silva, A.M.; Taranto, A.G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* **2020**, *8*, 343. [\[CrossRef\]](#)
78. Cesur, M.F.; Siraj, B.; Uddin, R.; Durmuş, S.; Çakir, T. Network-Based Metabolism-Centered Screening of Potential Drug Targets in *Klebsiella pneumoniae* at Genome Scale. *Front. Cell. Infect. Microbiol.* **2020**, *9*, 447. [\[CrossRef\]](#)
79. Nazarshodeh, E.; Marashi, S.-A.; Gharaghani, S. Structural Systems Pharmacology: A Framework for Integrating Metabolic Network and Structure-Based Virtual Screening for Drug Discovery against Bacteria. *PLoS ONE* **2021**, *16*, e0261267. [\[CrossRef\]](#)
80. Krueger, A.S.; Munck, C.; Dantas, G.; Church, G.M.; Galagan, J.; Lehár, J.; Sommer, M.O.A. Simulating Serial-Target Antibacterial Drug Synergies Using Flux Balance Analysis. *PLoS ONE* **2016**, *11*, e0147651. [\[CrossRef\]](#)

81. Gonzalez-Villoria, A.M.; Valverde-Garduno, V. Antibiotic-Resistant *Acinetobacter baumannii* Increasing Success Remains a Challenge as a Nosocomial Pathogen. *J. Pathog.* **2016**, *2016*, 7318075. [\[CrossRef\]](#)
82. Presta, L.; Bosi, E.; Mansouri, L.; Dijkshoorn, L.; Fani, R.; Fondi, M. Constraint-Based Modeling Identifies New Putative Targets to Fight Colistin-Resistant *A. Baumannii* Infections. *Sci. Rep.* **2017**, *7*, 3706. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Jensen, P.A.; Papin, J.A. Functional Integration of a Metabolic Network Model and Expression Data without Arbitrary Thresholding. *Bioinformatics* **2011**, *27*, 541–547. [\[CrossRef\]](#)
84. Henry, R.; Crane, B.; Powell, D.; Deveson Lucas, D.; Li, Z.; Aranda, J.; Harrison, P.; Nation, R.L.; Adler, B.; Harper, M.; et al. The Transcriptomic Response of *Acinetobacter baumannii* to Colistin and Doripenem Alone and in Combination in an in Vitro Pharmacokinetics/Pharmacodynamics Model. *J. Antimicrob. Chemother.* **2015**, *70*, 1303–1313. [\[CrossRef\]](#)
85. Zhu, Y.; Zhao, J.; Maifiah, M.H.M.; Velkov, T.; Schreiber, F.; Li, J. Metabolic Responses to Polymyxin Treatment in *Acinetobacter baumannii* ATCC 19606: Integrating Transcriptomics and Metabolomics with Genome-Scale Metabolic Modeling. *Msystems* **2019**, *4*, e00157–18. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Norsigian, C.J.; Kavvas, E.; Seif, Y.; Palsson, B.O.; Monk, J.M. ICN718, an Updated and Improved Genome-Scale Metabolic Network Reconstruction of *Acinetobacter baumannii* AYE. *Front. Genet.* **2018**, *9*, 121. [\[CrossRef\]](#)
87. Banerjee, D.; Raghunathan, A. Constraints-Based Analysis Identifies NAD⁺ Recycling through Metabolic Reprogramming in Antibiotic Resistant *Chromobacterium violaceum*. *PLoS ONE* **2019**, *14*, e0210008. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Bartell, J.A.; Blazier, A.S.; Yen, P.; Thøgersen, J.C.; Jelsbak, L.; Goldberg, J.B.; Papin, J.A. Reconstruction of the Metabolic Network of *Pseudomonas aeruginosa* to Interrogate Virulence Factor Synthesis. *Nat. Commun.* **2017**, *8*, 14631. [\[CrossRef\]](#)
89. Yen, P.; Papin, J.A. History of Antibiotic Adaptation Influences Microbial Evolutionary Dynamics during Subsequent Treatment. *PLoS Biol.* **2017**, *15*, e2001586. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Orth, J.D.; Conrad, T.M.; Na, J.; Lerman, J.A.; Nam, H.; Feist, A.M.; Palsson, B.Ø. A Comprehensive Genome-Scale Reconstruction of *Escherichia coli* Metabolism—2011. *Mol. Syst. Biol.* **2011**, *7*, 535. [\[CrossRef\]](#)
91. Yang, A.; Zhang, W.; Wang, J.; Yang, K.; Han, Y.; Zhang, L. Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1032. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Kavvas, E.S.; Yang, L.; Monk, J.M.; Heckmann, D.; Palsson, B.O. A Biochemically-Interpretable Machine Learning Classifier for Microbial GWAS. *Nat. Commun.* **2020**, *11*, 2580. [\[CrossRef\]](#)
93. Kavvas, E.S.; Seif, Y.; Yurkovich, J.T.; Norsigian, C.; Poudel, S.; Greenwald, W.W.; Ghatak, S.; Palsson, B.O.; Monk, J.M. Updated and Standardized Genome-Scale Reconstruction of *Mycobacterium tuberculosis* H37Rv, IEK1011, Simulates Flux States Indicative of Physiological Conditions. *BMC Syst. Biol.* **2018**, *12*, 25. [\[CrossRef\]](#)
94. Percy, N.; Hu, Y.; Baker, M.; Maciel-Guerra, A.; Xue, N.; Wang, W.; Kaler, J.; Peng, Z.; Li, F.; Dottorini, T. Genome-Scale Metabolic Models and Machine Learning Reveal Genetic Determinants of Antibiotic Resistance in *Escherichia coli* and Unravel the Underlying Metabolic Adaptation Mechanisms. *Msystems* **2021**, *6*, e00913–20. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Chung, C.H.; Chandrasekaran, S. A Flux-Based Machine Learning Model to Simulate the Impact of Pathogen Metabolic Heterogeneity on Drug Interactions. *PNAS Nexus* **2022**, *1*, pgac132. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Stokes, J.M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N.M.; MacNair, C.R.; French, S.; Carfrae, L.A.; Bloom-Ackerman, Z.; et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180*, 688. [\[CrossRef\]](#)
97. Montezano, D.; Meek, L.; Gupta, R.; Bermudez, L.E.; Bermudez, J.C.M. Flux Balance Analysis with Objective Function Defined by Proteomics Data—Metabolism of *Mycobacterium tuberculosis* Exposed to Mefloquine. *PLoS ONE* **2015**, *10*, e0134014. [\[CrossRef\]](#)
98. Kambli, P.; Ajbani, K.; Nikam, C.; Sadani, M.; Shetty, A.; Udawadia, Z.; Georghiou, S.B.; Rodwell, T.C.; Catanzaro, A.; Rodrigues, C. Correlating Rrs and Eis Promoter Mutations in Clinical Isolates of *Mycobacterium tuberculosis* with Phenotypic Susceptibility Levels to the Second-Line Injectables. *Int. J. Mycobacteriol.* **2016**, *5*, 1–6. [\[CrossRef\]](#)
99. Van Acker, H.; Coenye, T. The Role of Reactive Oxygen Species in Antibiotic-Mediated Killing of Bacteria. *Trends Microbiol.* **2017**, *25*, 456–466. [\[CrossRef\]](#)
100. Dikicioglu, D.; Kırdar, B.; Oliver, S.G. Biomass Composition: The “Elephant in the Room” of Metabolic Modelling. *Metab. Off. J. Metab. Soc.* **2015**, *11*, 1690–1701. [\[CrossRef\]](#)
101. Megchelenbrink, W.; Huynen, M.; Marchiori, E. OptGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks. *PLoS ONE* **2014**, *9*, e86587. [\[CrossRef\]](#)
102. Herrmann, H.A.; Dyson, B.C.; Vass, L.; Johnson, G.N.; Schwartz, J.-M. Flux Sampling Is a Powerful Tool to Study Metabolism under Changing Environmental Conditions. *NPJ Syst. Biol. Appl.* **2019**, *5*, 32. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Wiback, S.J.; Famili, I.; Greenberg, H.J.; Palsson, B.Ø. Monte Carlo Sampling Can Be Used to Determine the Size and Shape of the Steady-State Flux Space. *J. Theor. Biol.* **2004**, *228*, 437–447. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Bordel, S.; Agren, R.; Nielsen, J. Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes. *PLoS Comput. Biol.* **2010**, *6*, e1000859. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Balaban, N.Q.; Gerdes, K.; Lewis, K.; McKinney, J.D. A Problem of Persistence: Still More Questions than Answers? *Nat. Rev. Microbiol.* **2013**, *11*, 587–591. [\[CrossRef\]](#)
106. Manuse, S.; Shan, Y.; Canas-Duarte, S.J.; Bakshi, S.; Sun, W.-S.; Mori, H.; Paulsson, J.; Lewis, K. Bacterial Persisters Are a Stochastically Formed Subpopulation of Low-Energy Cells. *PLoS Biol.* **2021**, *19*, e3001194. [\[CrossRef\]](#)
107. Tourigny, D.S.; Goldberg, A.P.; Karr, J.R. Simulating Single-Cell Metabolism Using a Stochastic Flux-Balance Analysis Algorithm. *Biophys. J.* **2021**, *120*, 5231–5242. [\[CrossRef\]](#)

108. Damiani, C.; Di Filippo, M.; Pescini, D.; Maspero, D.; Colombo, R.; Mauri, G. PopFBA: Tackling Intratumour Heterogeneity with Flux Balance Analysis. *Bioinformatics* **2017**, *33*, i311–i318. [[CrossRef](#)]
109. Ma, P.; Amemiya, H.M.; He, L.L.; Gandhi, S.J.; Nicol, R.; Bhattacharyya, R.P.; Smillie, C.S.; Hung, D.T. Bacterial Droplet-Based Single-Cell RNA-Seq Reveals Antibiotic-Associated Heterogeneous Cellular States. *Cell* **2023**, *186*, 877–891.e14. [[CrossRef](#)]
110. Sahu, A.; Blätke, M.-A.; Szymański, J.J.; Töpfer, N. Advances in Flux Balance Analysis by Integrating Machine Learning and Mechanism-Based Models. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4626–4640. [[CrossRef](#)]
111. Rana, P.; Berry, C.; Ghosh, P.; Fong, S.S. Recent Advances on Constraint-Based Models by Integrating Machine Learning. *Curr. Opin. Biotechnol.* **2020**, *64*, 85–91. [[CrossRef](#)]
112. Zampieri, G.; Vijayakumar, S.; Yaneske, E.; Angione, C. Machine and Deep Learning Meet Genome-Scale Metabolic Modeling. *PLoS Comput. Biol.* **2019**, *15*, e1007084. [[CrossRef](#)] [[PubMed](#)]
113. Bhadra, S.; Blomberg, P.; Castillo, S.; Rousu, J. Principal Metabolic Flux Mode Analysis. *Bioinforma. Oxf. Engl.* **2018**, *34*, 2409–2417. [[CrossRef](#)] [[PubMed](#)]
114. Yang, J.H.; Wright, S.N.; Hamblin, M.; McCloskey, D.; Alcantar, M.A.; Schrübbers, L.; Lopatkin, A.J.; Satish, S.; Nili, A.; Palsson, B.O.; et al. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **2019**, *177*, 1649–1661. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

3.2.2 GSMM of cold adapted bacteria

Genome-scale metabolic models can be constrained using -omic data, that is information that describe how the genes are influenced by certain conditions. In other words, the models are being contextualized in order to have a better representation of the bacterial metabolism in specific cases.

Since microbial communities are constantly subjected to environmental changes, with temperature variations being among the most significant factors influencing their behavior, studying their metabolism in specific scenarios is crucial. This is particularly relevant in the context of global warming, but it is also evident in cyclic scenarios, such as seasonal changes in sea-surface temperatures. Here is where we can use -omic data to gain insights into how microorganisms respond at the cellular level to these changes, so we can deepen our understanding of their strategies for adapting to fluctuating environments.

In this study, we explored the mechanisms by which a cold-adapted marine bacterium maintains metabolic homeostasis when grown at two distinctly different temperatures, 15°C and 0°C. To achieve this, we quantified both its intracellular and extracellular central metabolomes and analyzed transcriptomic changes under these conditions, the two growth temperatures.

Using MATLAB 2019a, COBRA toolbox and REMI method we integrated the transcriptomic data as TPM gene expression and the metabolomic data as relative concentration values for metabolites into the genome-scale metabolic reconstruction of *Pseudoalteromonas haloplanktis* TAC125. This is an Antarctic bacterium that has received much attention in the last decade due to the interest in characterizing its cold-adaptation and nutritional adaptation strategies. In this way, we estimated the differences between conditions by predicting the flux of each reaction present in the pathways TCA cycle, glycolysis, PPP and fatty acids metabolism.

The resulted fluxes of our constrained models showed a remarkable robustness in the bacterium's central metabolism, with the concentrations of key metabolites remaining relatively stable despite the substantial temperature difference. In

other words, there were more than 600 differentially expressed genes, belonging to different functional categories, however the intracellular concentrations of circa 90% of the analyzed metabolites (34 in total) were statistically similar in both temperature conditions (Spearman correlation, P value > 0.05). These striking results suggests that *P. haloplanktis* TAC125 has a remarkable capability to buffer the temperature shifts, producing a stability supported by extensive transcriptomic reprogramming, involving significant changes in the expression of hundreds of metabolic genes.

Additionally, our findings pointed out that the metabolic adaptation is mainly happening due to shifts in a small number of critical intermediates, such as phosphoenolpyruvate (PEP), and to the interactions between TCA cycle and fatty acids metabolism pathways. These results highlighted the complex interplay between gene expression and metabolic processes, which contributes to the resilience of core metabolism. Our study also emphasizes the value of combining advanced multidisciplinary approaches to shed light on the molecular mechanisms underlying microbial responses to environmental changes, especially when facing a global warming scenario.



Metabolic Robustness to Growth Temperature of a Cold-Adapted Marine Bacterium

 Christopher Riccardi,^a  Marzia Calvanese,^b  Veronica Ghini,^{c,d}  Tania Alonso-Vásquez,^a  Elena Perrin,^a  Paola Turano,^c  Giorgio Giurato,^{e,f}  Alessandro Weisz,^{e,f}  Ermenegilda Parrilli,^b  Maria Luisa Tutino,^b  Marco Fondi^{a,g}

^aDepartment of Biology, University of Florence, Florence, Italy

^bDipartimento di Scienze Chimiche, University of Naples, Napoli, Italy

^cCenter of Magnetic Resonance (CERM), University of Florence, Florence, Italy

^dConsorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMMP), Florence, Italy

^eDepartment of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Baronissi, Italy

^fGenome Research Center for Health - GRGS, Baronissi, Italy

^gCentro interdipartimentale per lo Studio delle Dinamiche Complesse (CSDC), University of Florence, Italy

Christopher Riccardi, Marzia Calvanese, and Veronica Ghini contributed equally to this study. Author order was determined alphabetically based on their first name.

ABSTRACT Microbial communities experience continuous environmental changes, with temperature fluctuations being the most impacting. This is particularly important considering the ongoing global warming but also in the “simpler” context of seasonal variability of sea-surface temperature. Understanding how microorganisms react at the cellular level can improve our understanding of their possible adaptations to a changing environment. In this work, we investigated the mechanisms through which metabolic homeostasis is maintained in a cold-adapted marine bacterium during growth at temperatures that differ widely (15 and 0°C). We have quantified its intracellular and extracellular central metabolomes together with changes occurring at the transcriptomic level in the same growth conditions. This information was then used to contextualize a genome-scale metabolic reconstruction, and to provide a systemic understanding of cellular adaptation to growth at 2 different temperatures. Our findings indicate a strong metabolic robustness at the level of the main central metabolites, counteracted by a relatively deep transcriptomic reprogramming that includes changes in gene expression of hundreds of metabolic genes. We interpret this as a transcriptomic buffering of cellular metabolism, able to produce overlapping metabolic phenotypes, despite the wide temperature gap. Moreover, we show that metabolic adaptation seems to be mostly played at the level of few key intermediates (e.g., phosphoenolpyruvate) and in the cross talk between the main central metabolic pathways. Overall, our findings reveal a complex interplay at gene expression level that contributes to the robustness/resilience of core metabolism, also promoting the leveraging of state-of-the-art multi-disciplinary approaches to fully comprehend molecular adaptations to environmental fluctuations.

IMPORTANCE This manuscript addresses a central and broad interest topic in environmental microbiology, i.e. the effect of growth temperature on microbial cell physiology. We investigated if and how metabolic homeostasis is maintained in a cold-adapted bacterium during growth at temperatures that differ widely and that match measured changes on the field. Our integrative approach revealed an extraordinary robustness of the central metabolome to growth temperature. However, this was counteracted by deep changes at the transcriptional level, and especially in the metabolic part of the transcriptome. This conflictual scenario was interpreted as a transcriptomic buffering of cellular metabolism, and was investigated using genome-scale metabolic modeling. Overall, our findings reveal a complex interplay at gene expression level that contributes to the

Editor Rachel Mackelprang, California State University, Northridge

Copyright © 2023 Riccardi 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 Marco Fondi, marco.fondi@unifi.it.

The authors declare no conflict of interest.

Received 12 November 2022

Accepted 26 January 2023

Published 27 February 2023

robustness/resilience of core metabolism, also promoting the use of state-of-the-art multi-disciplinary approaches to fully comprehend molecular adaptations to environmental fluctuations.

KEYWORDS cold-adaptation, genome-scale modeling, metabolomics, transcriptomics

Microorganisms are able to colonize virtually every environmental niche on Earth (1). They have adapted for millions of years prospering under conditions such extreme as water boiling or freezing points, high radiation, acidic or alkaline pH values, heavy metal pollution, and high salinity (2). Growth temperature, in particular, is one of the environmental parameters that mostly impact the physiology of microorganisms and that is thought to have played a key role in their adaptation, selection, and diversification (3). Given the geological history of our planet, it is reasonable to think that adaptation to changing temperatures has independently occurred many times in evolution; consequently, there exists a vast array of molecular strategies for this purpose, disseminated in the microbial kingdom (4–6). Their characterization is key in this phase of Earth's life, as global change is imposing rapid/dramatic modifications in basic environmental parameters (including temperature) that, in turn, will solicitate the activation of such temperature-adaptation related pathways in microbial communities. Understanding which genes get activated or which compounds get secreted in the environment following an increase of water temperature will help us model and predict the scenarios of microbial communities in a changing environment.

Without necessarily invoking global change, the need to rewire cellular networks in response to temperature shifts is likely a common feature in natural microbial communities. The temperature in the Southern Ocean, for example, is anywhere from -2 to 10°C, since Antarctic water temperature fluctuation responds to the seasonal advance and retreat of sea ice (7). Therefore, marine microorganisms are exposed to seasonal oscillations in temperature. Upper-ocean microbes can experience a higher variability in sea-surface temperature and their working temperatures exceed the *in situ* Eulerian temperature range by up to 10°C (8). Recent findings demonstrate how even upper-ocean microbes experience along-trajectory temperature variability up to 10°C greater than seasonal fluctuations as a result of large-scale climate variability, indicating a remarkable thermal tolerance by the drifting microbial populations in fluctuating marine environments (8, 9). From a cellular viewpoint, previous studies of cold- and heat-adapted microbes have revealed a variety of molecular adaptations that allow their activity and survival under extreme conditions. These initially involve the change in the expression of specific gene sets but, ultimately, the regulatory changes imposed by temperature increase/decrease are mostly implemented at the metabolic level, since this represents, in the words of Prof. Oliver Fiehn, “the ultimate response of biological systems to genetic or environmental changes” (10). These variegated adaptation strategies observed in temperature-stressed microorganisms indicate that: (i) each microorganism may follow a peculiar route to maintain cellular homeostasis when facing temperature fluctuations (11–13), and (ii) there exists intense cross talk between regulatory and metabolic networks in the response thereof (14).

Here, we specifically investigate the molecular mechanisms through which metabolic homeostasis is maintained in a marine bacterium that (in its natural settings) is known to experience such broad seasonal temperature fluctuations. Indeed, rather than studying the consequences of a cold shock event (a circumstance that rarely occurs in nature), we here focus on the comparison of the main cellular networks in cells growing at 2 different temperatures. This was performed through the integration of metabolomic and transcriptomic data unitedly with genome-scale metabolic modeling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 (PhTAC125). This bacterium (15) has been isolated from an Antarctic coastal seawater sample collected in the vicinity of the French Antarctic station Dumont d'Urville, Terre Adélie (66° 40' S; 140° 01' E) and has received much attention in the last decade due to the interest

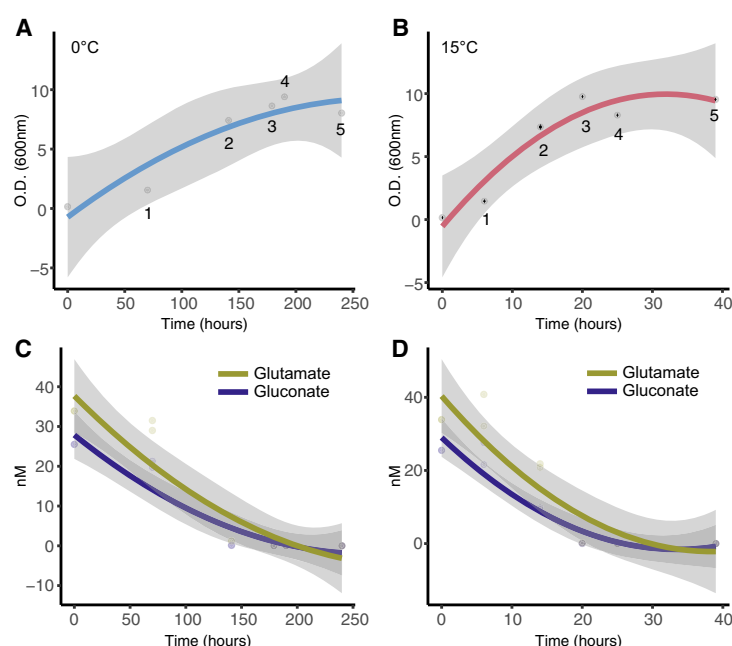


FIG 1 (A) Growth curve of *PhTAC125* at 0°C. Numbers indicate the sampling points for metabolomic and transcriptomic experiments. (B) Growth curve of *PhTAC125* at 15°C. Numbers indicate the sampling points for the metabolomic experiments. The transcriptome of *PhTAC125* growing cells was sampled at time point 1. (C) Glutamate and gluconate uptake at 0°C. (D) Glutamate and gluconate uptake at 15°C.

in characterizing its cold-adaptation and nutritional adaptation strategies, as well as its biotechnological potential (16–19).

We show that different growth temperatures induce broad transcriptional changes that involve genes of many key metabolic pathways. This transcriptional rewiring, however, is scarcely reflected at the level of the core metabolism, as most key central metabolites show overlapping trends at the 2 tested temperatures. The obtained -omics data were used to compute the flux distributions sustaining growth at low and high temperature, and this provided a mechanistic understanding of the possible adaptation strategies to temperature fluctuations.

RESULTS

We cultivated *PhTAC125* cells in a bioreactor and sampled the 0 and 15°C growth curves at 5 different time points (Fig. 1A and B) that overall resembled the same physiological conditions for the 2 experiments. For the 0°C growth, we sampled the following time points: 70, 141, 179, 190, and 240 h. The 15°C curve was sampled at 6, 14, 20, 25, and 39 h. The average growth rate of the 2 cultures varied, being 0.11 (standard deviation (s.d.) 0.011) h^{-1} and 0.016 (s.d. 6.4×10^{-5}) h^{-1} at 15° and 0°C, respectively. The same calculation was repeated and limited to the exponential phases of the two cultures, yielding 0.27 (s.d. 0.006) h^{-1} and 0.027 (s.d. 0.0009) h^{-1} at 15° and 0°C, respectively. The samples obtained were used for intracellular and extracellular metabolites quantification through nuclear magnetic resonance (NMR) (see Materials and Methods). For 2 of these time points (Fig. 1A and B, labeled as “1”), we also performed quantified gene expression levels using RNA-Seq.

PhTAC125 metabolome is qualitatively and quantitatively robust to temperature shift. Overall, we assigned and analyzed the concentration of 34 intracellular metabolites in the 2 growth curves. These metabolites represented key intermediates of central metabolic pathways, such as the TCA cycle, amino acids biosynthesis, glycolysis, Pentose Phosphate Pathway (PPP), and nucleic acids biosynthesis. Our NMR metabolomic approach was not able to discriminate between the oxidized and reduced forms of NAD and NADP, so we generally refer to NADX and NADPX in the rest of the manuscript. First, we monitored the overall trends of their intracellular concentrations with

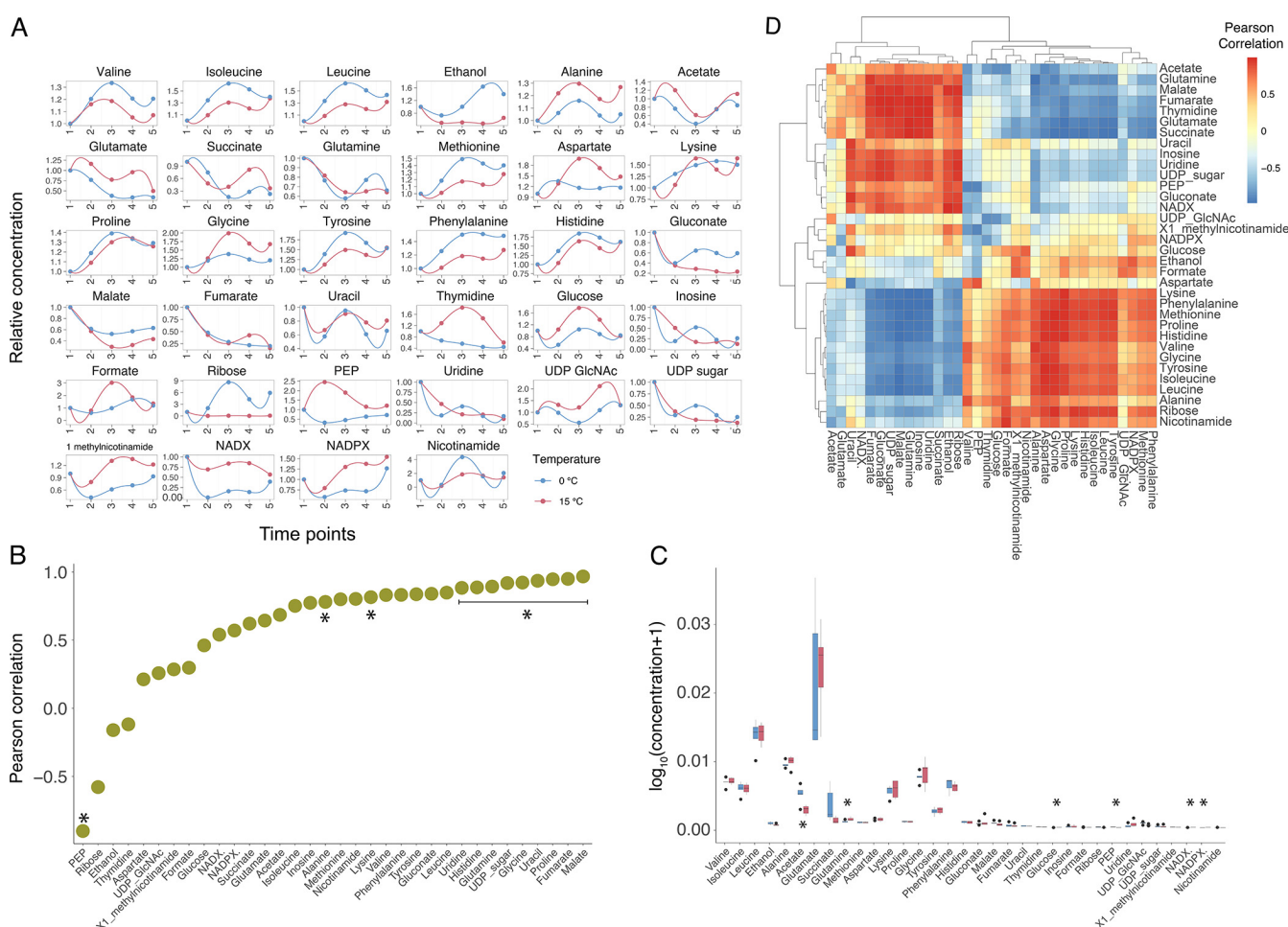


FIG 2 (A) Normalized intracellular metabolites concentration across the 5 time points. (B) Correlation of each intracellular metabolite at 0 and 15°C (asterisks indicate Spearman correlation P value < 0.05). (C) Comparison of the concentration of each intracellular metabolite at 0 and 15°C (asterisks indicate statistically significant correlations, i.e., P value < 0.05). (D) All-against-all correlations between intracellular metabolites at 0 and 15°C.

respect to the beginning of the growth experiment (Fig. 2A) by computing the Pearson product moment for each metabolite across the five time points and the relative statistical support (Spearman correlation, P value < 0.05 (Fig. 2B)). Strikingly, for about 90% of the analyzed metabolites, we found a positive correlation between the variation of their concentrations during the 2 separate growth experiments (at 0 and 15°C). These positive values ranged from 0.21 in the case of aspartate to 0.96 in the case of malate. Ethanol and thymidine displayed a Pearson product moment correlation (PPM) quite close to zero (-0.16 and -0.1, respectively), thus showing no correlation in the two experiments. Finally, the concentrations of PEP and ribose were negatively correlated in the 2 growth experiments, with PPM of -0.9 and -0.57, respectively. The negative correlation of PEP was supported statistically (Spearman correlation, P value = 0.03692).

To qualitatively assess the change in concentration of each metabolite against each other, we computed the all-against-all correlation among the metabolites identified in our study. The results of this analysis are reported in Fig. 2D. We observed 2 main blocks of metabolites: 1 embedding amino acids (with the exception of glutamate and glutamine), and 1 mostly including intermediates of the main central metabolic pathways (TCA, PPP, nucleic acids biosynthesis). Each of these clusters was characterized by a strong (and statistically supported) correlation among the representatives of the same cluster, and an equally relevant anti-correlation with the members of the other cluster. Overall, the intracellular concentration of amino acids was shown to increase over time, whereas the level of, for example, TCA and PPP intermediates was shown to

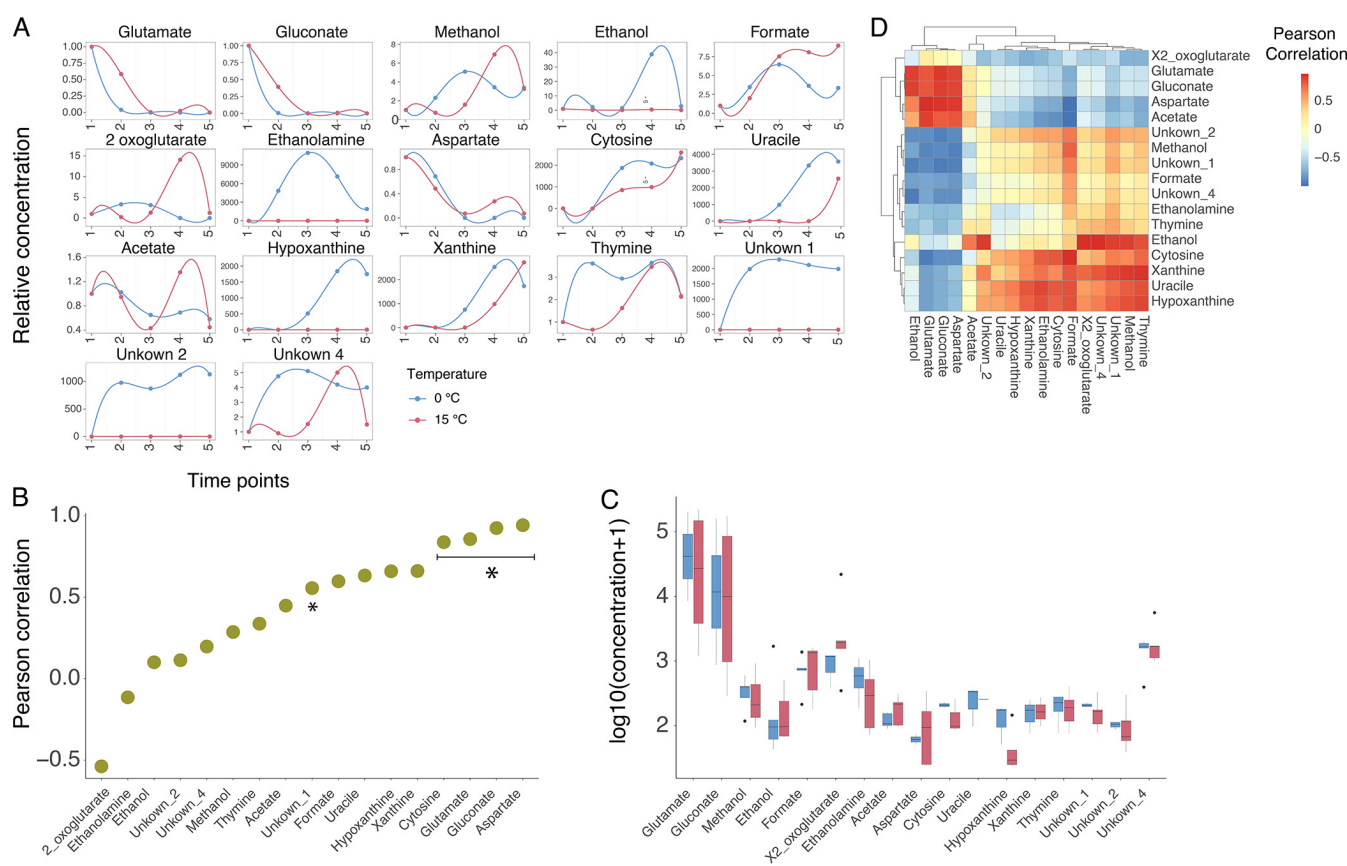
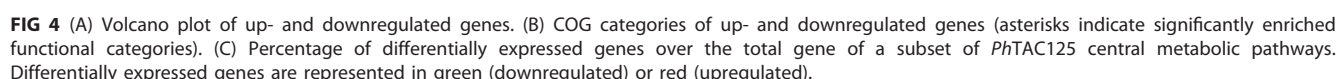


FIG 3 (A) Normalized extracellular metabolites concentration across the 5 time points. (B) Correlation of each intracellular metabolite at 0 and 15°C (asterisks indicate Spearman correlation P value < 0.05) (C) Comparison of the concentration (in a.u.) of each extracellular metabolite at 0 and 15°C (asterisks indicate statistically significant correlations, i.e., P value < 0.05). (D) All-against-all correlations between extracellular metabolites at 0 and 15°C.

decrease throughout the growth experiment. To assess whether the growth temperatures had a role in determining the observed trends in metabolite concentrations, we computed an all-against-all correlation for each metabolite at 0° and 15°C separately, and then evaluated whether these 2 correlation matrices differed significantly. The t -tests performed supported no significant differences between the means of these 2 comparisons (P values = 0.74 and 0.89, respectively) suggesting that, despite the large difference in growth temperatures, the overall dynamics of intracellular core metabolites were maintained, pointing to an apparent structural metabolic robustness to growth temperature. This robustness was also conserved quantitatively, as the average concentration of each metabolite was maintained similar across the 2 different growth experiments (Fig. 2C). Indeed, except for 6 metabolites out of 34 (acetate, glutamine, glucose, PEP, NADX, and NADPX, t test, P value < 0.05 , but P value > 0.05 after correction for multiple testing with the Bonferroni method), no other average metabolite concentration showed a significant difference at 0 and 15°C (Fig. 2C and Fig. S1).

A similar scenario was observed for the pool of (17) quantified extracellular metabolites (Fig. 3A). Consistently with growth conditions, glutamate and gluconate concentrations decreased over time until their exhaustion from the growth medium (Fig. 3A), a trend that (i) indicated the almost simultaneous consumption of these 2 nutrients and (ii) was observed at both 0 and 15°C. Out of the 17 extracellular metabolites that were quantified in our experiments, 16 showed a positive correlation between their trends at 0 and 15°C (Fig. 3B). In particular, 10 of them displayed a PPM above 0.5. The only metabolite whose trend differed in the 2 experiments was 2-oxoglutarate (PPM = -0.4). This situation was mirrored at the quantitative level (Fig. 3C). Indeed, none of the extracellular metabolites showed a statistically significant difference between 0°



The 0 and 15°C transcriptomes of PhTAC125. We next characterized the transcriptomes of *PhTAC125* at 0 and 15°C. We sampled both curves during exponential growth (Fig. 1A and B, sample points “1”), and sequenced the transcriptome using RNA-Seq technology (see Materials and Methods section). The main features of transcriptomic data are shown in Table S1 and Fig. S2 and S3.

Differentially expressed genes (DEGs) were identified using a log₂ fold change of 1 (or -1) and an adjusted *P* value of 0.05 as thresholds. Overall, we identified 607 differentially expressed genes in the comparison between growth at 15° and 0° C, with an almost identical amount of up- and downregulated genes at 15°C (304 and 303 genes, respectively). The outliers of up- and downregulated genes are shown in Fig. 4A and described in Table 1. Remarkably, out of 607 DEGs, 359 were metabolic protein-coding genes (roughly 59%).

Here we first describe specific features of outlier DEGs, and then focus our attention on broader functional categories that embedded the highest fraction of differentially expressed genes. Among the top 5 downregulated genes, 2 lacked a clear functional annotation (PSHA_RS12380 and PSHA_RS08940), and further experiments are required to understand their role in the adaptation to growth at warmer temperatures. Two of them (PSHA_RS07950 and PSHA_RS06965) were annotated as TonB-dependent receptors, suggesting their involvement in the uptake and transport of large substrates, possibly siderophores complexes and/or vitamins. The remaining gene (PSHA_RS11830) is annotated as NAD-dependent succinate-semialdehyde dehydrogenase, responsible for the conversion of succinate-semialdehyde to succinate. The downregulation of this gene is in line with the lower intracellular concentration of succinate at 15°C, as discussed later in the text. Strikingly, among the over-expressed genes during growth at 15°C, we found a prevalence of cold shock related proteins (4 out of 5) and a RNase R encoding gene (PSHA_RS14645). Although no functional studies have been done on cold shock responses and cold shock proteins (CSPs) from psychrophilic bacteria,

TABLE 1 The list of the top-five down- and upregulated genes and their functions

Locus tag	log ₂ FC	P _{adj} value	Annotation
PSHA_RS12380	−6.58	2.43e-100	Hypothetical protein
PSHA_RS07950	−5.99	1.52e-65	TonB-dependent receptor
PSHA_RS06965	−5.53	6.72e-07	TonB-dependent receptor
PSHA_RS11830	−4.44	2.64e-27	NAD-dependent succinate-semialdehyde dehydrogenase
PSHA_RS08940	−4.33	1.85e-53	PA2169 family four-helix-bundle
PSHA_RS14635	7.93	1.99e-63	Cold-shock protein
PSHA_RS14640	6.73	4.02e-13	Cold-shock protein
PSHA_RS14630	5.94	2.03e-38	Cold-shock protein
PSHA_RS14645	5.48	9.73e-31	Ribonuclease R
PSHA_RS16715	3.94	1.25e-21	Cold-shock protein

similarities with the CSPs that are produced in mesophiles have been observed. In particular, the over-expressed CSPs in *PhTAC125* share an amino acid sequence identity ranging between 63.8% and 65.6% with the mesophilic *cspE* of *E. coli*. Furthermore, these proteins contain highly conserved RNA-binding motifs, RNP1 (K-G-F-G-F-I) and RNP2 (V-F-V-H-F) (20, 21), indicated by a black box in Fig. S4.

PSHA_RS14630, PSHA_RS14635, PSHA_RS14640 and PSHA_RS16715 contain 1 highly conserved nucleic acid-binding domain, called cold shock domain (CSD, <http://pfam.xfam.org/family/PF00313>), which is annotated as RNA chaperone/anti-terminator. Moreover, the NCBI identifies both genes PSHA_RS14635 and PSHA_RS14640 as encoding for the same protein (WP_011329604.1), while the others encode for proteins with a different identifier but virtually same sequence. Cold shock proteins would normally counteract the deleterious effects of temperature drop, enabling the cells to grow at low temperatures (22). Jiang et al. proposed the role of *cspA* as an RNA chaperone capable of melting the RNA secondary structure (23), thereby enhancing translation of mRNAs at low temperatures. However, counterintuitively with their given name, not all members of the CSP family are cold-inducible, and their expression is activated upon different stresses (24). For this reason, CSPs might be required for bacterial adaptation to environmental changes. The fifth upregulated protein, RNase R, belongs to the RNR family. In *Escherichia coli*, the RNase R consists of a central nuclease domain, 2 cold shock (CSD) domains near the N-terminal region of the protein, an S1 domain and a highly basic C-terminal region (25). Cairrão et al. showed that the *mnr* gene is co-transcribed with flanking genes as an operon induced under cold shock in *E. coli* (locus tag b4179) (26). There is an important analogy here since, although *PhTAC125*'s operon map isn't available, its top 4-upregulated genes are mapped consecutively on the genome, with coding sense on the same strand (+), spaced by an average distance of 216 nucleotides. The homology to a variety of domains involved in stress response, together with our observation of a strong upregulation during growth at 15°C, may indicate that *PhTAC125* strives to mediate the elimination of detrimental secondary structures and a temperature rise promotes the expression of enzymes that are required for the correct processing of rRNA precursors.

Extending the analysis to a broader functional level, we identified significant differences in 12 COG categories. Specifically, intracellular trafficking, secretion and vesicular transport, signal transduction mechanisms, cell motility, cell wall/membrane/envelope biogenesis, replication, recombination and repair and cell cycle control, cell division, and chromosome partitioning functional categories were found to be over-represented among upregulated genes during growth at 15°C. This matches the results from recent works on the characterization of the evolutionary pathways responsible for thermal adaptation, and the overall notion that one of the main effects of temperature increases on cell physiology consists in the disruption of membrane integrity caused by increased fluidity (27, 28). Overexpression of cell wall and membrane biogenesis-related genes might help overcome this feature associated with growth at higher temperatures.

Conversely, genes involved in post-translational modification, protein turnover,

chaperones, translation, ribosomal structure and biogenesis, lipids, nucleic acids and amino acids transport and metabolism, as well as energy production functional categories, were over-represented among downregulated genes during growth at 15°C (Fig. 4B). We believe that this analysis depicts the system-level adaptation of bacterial life to diverse growth temperatures, and is in line both with the physiological features observed in this work and with previously obtained data on cold/warm adaptation. Indeed, the significant over-expression of genes broadly related to cell replication (COG categories L, M, and D) is in line with the higher growth rate observed at 15°C in respect to 0°C, over the entire growth period (Fig. 1 and B), and in the specific time interval where RNA was sampled. Further, the over-expression of cell motility at warmer temperatures is in perfect agreement with previous assays on *PhTAC125* motility that had shown a reduced swimming capability of this strain at 0°C (18).

Surprisingly, the upregulation of cell growth related processes is counteracted by a general downregulation of cellular metabolism. Key processes of *PhTAC125* core metabolism such as amino acid, lipid, and nucleotide metabolism together with energy production/conversion, were found to be downregulated (Fig. 4C). As mentioned previously, among 607 DEGs, 359 (59.14%) were metabolic genes. Zooming in at the level of the single pathways (Fig. 4C) revealed that the downregulation affected most of the key central pathways of *PhTAC125* metabolic network (including, for example, glycolysis, TCA cycle, PPP, and amino acids biosynthesis genes). In these pathways, the number of downregulated genes strongly outpaced that of upregulated genes (Fig. 4C). Thus, despite consistent intracellular and extracellular metabolic profiles between the 2 tested growth temperatures, the underlying expression of metabolic genes showed an opposite trend, with a remarkable number of DEGs in the 2 growth conditions. For example, the intracellular pools of key TCA intermediates (fumarate and malate) (Fig. 2C) were shown to be similar between 0 and 15°C, despite more than 50% of the TCA genes being differentially expressed in the 2 conditions (Fig. 4C). Similarly, the intracellular concentration of amino acids was shown to be comparable between the 2 conditions (Fig. 2C), whereas the level of expression of their corresponding biosynthetic genes was shown to be significantly different (Fig. 4C).

Accordingly, we hypothesize that the transcriptional network of *PhTAC125* provides a buffering mechanism through which metabolic homeostasis is maintained, at least at the level of the central metabolism. In the next section we will combine metabolomic and transcriptomic data with a genome-scale metabolic reconstruction of *PhTAC125* to unravel the intimate mechanisms through which this balance is achieved.

Genome-scale modeling of growth temperature adaptation. A genome-scale metabolic reconstruction exists for the strain *PhTAC125* (16). We, thus, exploited this resource to get a mechanistic interpretation of metabolic homeostasis. First, we checked whether the model was able to represent the experimentally determined phenotypes. We thus constrained the *PhTAC125* genome-scale reconstruction with uptake rates of glutamate and gluconate measured at 0 and 15° degrees, and ran an FBA simulation to predict the cellular growth rates. Average uptake rates of glutamate and gluconate were respectively computed between T1 and T3 for both growth curves (that is after 14 and 141 h for 15 and 0°C, see Materials and Methods), and resulted in 0.10 and 0.08 mmol/gCDW·h⁻¹ at 0°C and 0.47 and 0.62 mmol/gCDW·h⁻¹ at 15°C.

As shown in Fig. 5A, at 0°C the results of this simulation (0.021h⁻¹) were in line with the measured growth rates (0.027h⁻¹). Conversely, at 15°C a discrepancy was observed between the experimental growth rate and the one predicted *in silico*, 0.27 versus 0.13 h⁻¹, respectively. Although there could be many possible explanations for this, we hypothesize that constraining the model only based on the uptake rates of the provided nutrients may not be enough to correctly represent the metabolic phenotype of the cells exposed to different temperatures. For this reason, and to provide a more realistic picture of how metabolic homeostasis is maintained at the level of the central metabolism, we combined metabolomic and transcriptomic data with the genome-scale modeling of *PhTAC125* metabolism. We decided to focus on T1 (beginning of the

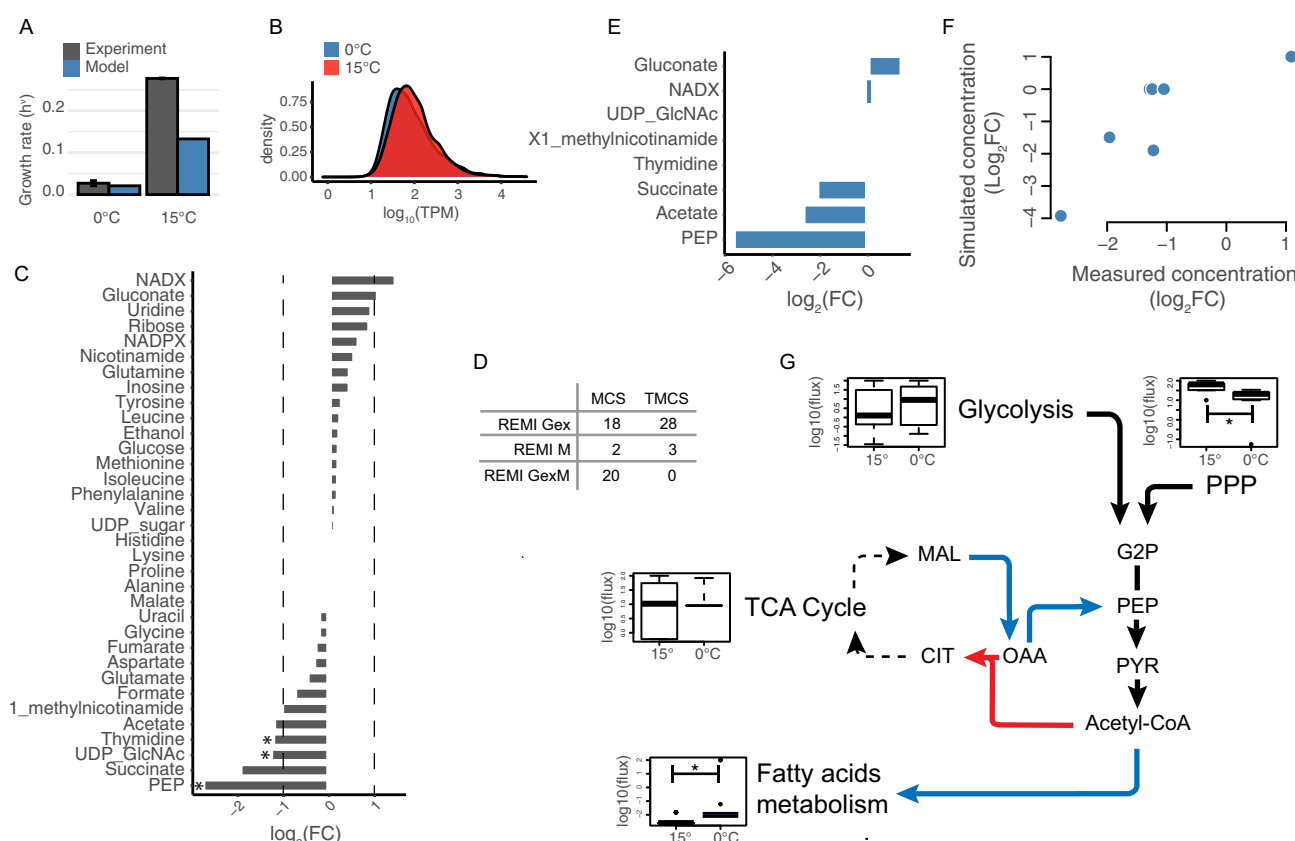


FIG 5 (A) Comparison between experimental and simulated growth rates. (B) Distribution of TPM values of gene expression at T1. (C) $\log_2\text{FC}$ of metabolites intracellular concentration at T1 during growth at 15°C versus growth at 0°C. (D) REMI output in terms of Maximal and Theoretical Consistency Scores following transcriptomic and metabolomic data integration. (E) Genome-scale modeling prediction of metabolites $\log_2\text{FC}$ concentration showing a $|\log_2\text{FC}| > 1$ during the metabolomic experiment. (F) Correlation between $\log_2\text{FC}$ of simulated (y axis) versus measured (x axis) internal metabolites concentration. (G) Working model of key metabolic adjustments at the 2 different temperatures. Blue and red arrows indicate fluxes predicted to increase during growth at 0° and 15°C, respectively. Boxplots represent the average flux values for each pathway. Asterisks indicate a significant (i.e., P value < 0.05) Kolmogorov-Smirnov statistical test.

exponential growth phase) as this time point should better resemble the cellular physiological state in which FBA assumptions hold the most (i.e., metabolic steady state). Then, we computed the $\log_2$ fold change ($\log_2\text{FC}$) of central intracellular metabolites in the 2 conditions (Fig. 5C). Most (76%) of the metabolites showed a $|\log_2\text{FC}|$ lower than 1, confirming an overall robustness of the central metabolism to growth temperature. Only 8 metabolites displayed a $|\log_2\text{FC}|$ lower/greater than 1 in the contrast between 15° and 0°C metabolomic data, and for only 3 of them we obtained a statistical support (P value < 0.05) (Fig. 5C). More specifically, NADX and gluconate were the 2 metabolites showing a $\log_2\text{FC} > 1$ (thus being more abundant at 15°) whereas PEP, succinate, UDP-GlcNAc, thymidine, acetate and X1- methyl-nicotinamide displayed a $\log_2\text{FC} < -1$ (thus being more abundant at 0°).

We then constrained the model using all available experimental data obtained in this work using REMI (29) (see Materials and Methods). Briefly, this approach allows translating gene expression and metabolite abundance data resulting from a “perturbation” experiment into differential flux distributions among the 2 resulting conditions. In our case, we used data (metabolomes and transcriptomes sampled at the same time point) from 15° and 0°C growths to analyze the systemic metabolic alteration(s) occurring in this pair of conditions. FC and TPM values (Fig. 5B and C) were used to constrain the model with metabolic and expression data, respectively. Further, the model was also constrained by setting the boundaries of uptake reactions to represent the actual medium used for the experiment (Schatz medium with glutamate and gluconate as the sole carbon sources, see Materials and Methods). The outcome of data integration

into the metabolic reconstruction is summarized by REMI through the computation of the theoretical maximum consistency score (TMCS) and the maximum consistency score (MCS). The first indicates the number of available omics data (for metabolites and reactions), whereas the latter represents the number of those constraints that are consistent with fluxes, and could be integrated into REMI models. As a result, the MCS is always equal to or smaller than the TMCS. In other words, MCS is the largest fraction of available data (metabolomics and transcriptomics) that could be incorporated into an FBA model from a given set of constraints (the abundance of metabolites and transcripts), while ensuring that the model still achieves the required metabolic functionalities and remains feasible. TMCS indicates the number of genes and metabolites with available relative abundance values that can potentially (either because above the specified threshold or their actual inclusion in the metabolic reconstruction) be integrated into the model. The proportion between TMCS and MCS obtained in this work is comparable to that from other studies where REMI was used for the same purposes (29) (Fig. 5D). Out of data integration and FBA simulations, we obtained 2 distinct flux distributions, i.e., the flux distribution resembling growth at 0°C and the 1 theoretically accounting for the growth at 15°C. The differences between the 2 will represent the most likely metabolic alterations in response to growth temperature and, consequently, will highlight those pathways/reactions that contribute the most to maintaining the observed metabolic homeostasis. First, we checked whether our simulations were accounting for the actual differences between 0° and 15°C intracellular metabolomes. We, thus, computed a matching coefficient (Simple Matching Coefficient (SMC)) between the predicted and measured $\log_2\text{FC}$ of each of the 34 internal metabolites, accounting for how many times the model correctly predicted the increase (or decrease) of its internal concentration. Overall, we found an SMC of 71% between simulated and measured metabolic data, revealing that the model is capable of accounting for most of the central metabolome rewiring in respect to growth temperature. Then, we focused on those metabolites that showed a marked change between growth at 15 and 0°C (i.e., for which $|\log_2\text{FC}| > 1$) (Fig. 5C). Figure 5F shows that, except for 3 metabolites (UDP-GlcNAc, Thymidine, and 1- methylnicotinamide) for which the model does not predict any difference between the 2 growth conditions, for the remaining 5 metabolites the model correctly predicts a higher (NAD⁺ and gluconate) or lower (acetate, PEP and succinate) internal production at 15°. Overall, we found a significant, positive correlation of 0.72 between the measured and the predicted FC of internal concentrations of the 8 metabolites for which $|\log_2\text{FC}| > 1$ (Fig. 5F) (Spearman correlation, P value = 0.04).

Next, we examined to what extent the central, interconnected pathways showed significantly altered flux distributions. While no clear signal could be identified for glycolysis and TCA cycle (mirroring what we observed with DEGs analysis), both PPP and fatty acids metabolism showed significant differences in their fluxes between 0° and 15° growth simulations. As for fatty acids metabolism, fluxes representative of growth at 0°C were significantly higher (Kolmogorov-Smirnov test, P value = 8.96×10^{-10}) than those resembling metabolism at 15°C (Fig. 5F). The opposite was observed for PPP simulated fluxes which displayed, on average, significantly lower values at 0° vs 15°C (Kolmogorov-Smirnov test, P value = 0.0030). A sustained activity of PPP at 15° is in line with the increased internal concentration of gluconate and ribose (2 key PPP intermediates) at this temperature (Fig. 5C). Similarly, an increase in fatty acids biosynthesis (and fatty acids metabolism in general) is in line with previous simulations and experiments concerning the involvement of this process in the adaptation to growth at (relatively) low temperatures (16).

We then focused on the analysis of simulated metabolic fluxes around the metabolite that showed the highest degree of variation, PEP. We asked which metabolic rewiring could lead to an increased production of this metabolite at 0°C. The analysis of fluxes revealed an increased activity at 0°C with respect to 15°C of both the reaction converting L-malate to 2-oxaloacetate ([S]-malate:NAD⁺ oxidoreductase) and the reaction converting 2-oxaloacetate to PEP (phosphate:oxaloacetate carboxy-lyase). Also,

ATP:pyruvate, H₂O phosphotransferase reaction displayed an increase in flux at 0°C. Overall, this would allow the partial redirection of key TCA cycle intermediates to the production of PEP. At the same time, we also recorded an increased flux at 0°C in the reaction redirecting acetyl-CoA to fatty acids biosynthesis. Conversely, reactions leading to the production of citrate from OAA and Acetyl-CoA resulted to be less active at 0°C.

Overall, our experimentally constrained simulations seem to suggest a working model for the metabolic adaptation to growth at different temperatures. Considering growth at 0°C, an increased level of intracellular PEP observed experimentally might be functional to its conversion to acetyl-CoA and its consequent tunneling into fatty acids metabolism (Fig. 5G), and might be the outcome of fluxes redirection from the TCA cycle to PEP production, rather than from glycolysis or PPP. Consistent with this idea is the overall increase of key TCA intermediates (Fig. 5A) measured during growth at 0°C with respect to growth at 15°C that suggests an overall increased activity of this central pathway in the 0°C growth condition.

DISCUSSION

In this work, we have studied how a cold-adapted bacterium rewires its central metabolism when growing at 2 distinct temperatures that overall resemble a seasonal shift. This was done by characterizing a pool of 34 intracellular and 17 extracellular central metabolites during 5 different time points of its growth curves (i.e., at 0°C and 15°C) and by evaluating gene expression during the initial stages of its exponential phases. To our surprise, the top 5 upregulated genes at 15°C contain a cold shock domain, which is usually recruited to counteract the deleterious effects of temperature drop. A possible explanation for the apparent paradox where CspA is activated, not only following cold stress, but also under non-stress and other stress conditions which entail a downregulation of bulk gene expression and protein synthesis is presented in (30). CSPs have been found in almost all types of bacteria and are mainly induced after a rapid temperature downshift to regulate the adaptation to cold stress but are also present under normal conditions to regulate other biological functions (31). For example, in *E. coli*, only 4 (*cspA*, *cspB*, *cspG*, and *cspI*) of the 9 CSP genes are cold-induced (32). Two of them, *cspE* and *cspC*, are constitutively expressed at physiological temperatures, and act as 'housekeeping RNA chaperones' to modulate the global gene expression (33, 34). Furthermore, these CSPs are also involved in the transcription anti-termination mechanism, which is based upon preventing the formation of secondary structures on the nascent mRNA (35–37). In *PhTAC125*, four cold shock-like proteins *cspE* are upregulated during growth at 15°C. This indicates that these CSP proteins may not play a role in cold-adaptation in *PhTAC125* but, analogously to *E. coli*, they may act as chaperones by destabilizing secondary structures in target RNA at high temperature so that the single-stranded state of target RNA is maintained. This may then enable efficient transcription and translation.

From a broader perspective, the comparison between the 0°C and 15°C transcriptomes produced more than 600 differentially expressed genes, assigned to a dozen of different functional categories. On the other hand, the intracellular concentration of nearly 90% of the analyzed metabolites correlates positively in the 2 conditions. We interpret this as the capability of *PhTAC125* regulatory network to buffer the temperature shift to produce strikingly similar metabolic phenotypes, despite the temperature gap. This is also reflected in the conservation of the overall metabolic network structure (i.e., the presence of correlated clusters of metabolites) (Fig. 2C and Fig. 3C). Remarkably, this conservation involved metabolites from many different pathways, such as amino acids metabolism, TCA cycle, PPP, and glycolysis, suggesting that such robustness is propagated at the level of the entire metabolism and is not restricted to specific pathways. Studies on the metabolic response to growth at different temperatures in other microorganisms have generally shown larger variability among intracellular pools of metabolites (38), and/or the tendency to activate genetic mechanisms that shut down metabolism under longer-term high temperature stress (39). This is not the case for

PhTAC125 that, through a global transcriptional buffering, maintains very consistent trends of its intracellular and extracellular metabolomes. This mirrors, for example, the response of *E. coli* to genetic and environmental perturbations and its capability to maintain metabolite levels stable, reflecting the rerouting of fluxes in the metabolic network (40). A few metabolites stood out in this conservation of metabolite pools, including PEP and ribose (showing opposite trends at 0 and 15°C), and ethanol and thymidine (showing no correlation between the 2 experiments). Also, NAD⁺ and NADP⁺ levels differed in the 2 conditions, as shown in Fig. 2A.

To unravel this complex interplay and produce a mechanistic interpretation of this adaptation, -omics data (together with data on carbon sources uptake rates) were used to constrain a genome-scale metabolic reconstruction and derive the most-likely metabolic phenotypes at these 2 temperatures. This approach provided evidence that most of the metabolic adaptation is probably played at the level of the phosphoenolpyruvate (PEP)–pyruvate–oxaloacetate node and specifically involves the increase of TCA fluxes and their redirection to PEP production. Consistent with this idea, is the overall reduction of key TCA intermediates at 15°C (Fig. 5A) and the observation that, upon growth on non-glycolytic compounds (e.g., glutamate and gluconate) the cycle intermediates malate or oxaloacetate must be converted to pyruvate and PEP for the synthesis of glycolytic intermediates. In our case, however, the availability of an increased pool of PEP would rather allow its conversion to acetyl-CoA, and its consequent tunneling into fatty acids metabolism (Fig. 5G). Also the overexpression of genes involved in valine, leucine, and isoleucine degradation (Fig. 4C) at 0°C can be associated with the higher activity of fatty acids metabolism, as the degradation of these 2 amino acids leads to the production of acetyl-CoA (41). Consistently with this observation, our model predicts an increased activity of fatty acids metabolism at 0° (Fig. 5G). The importance of fatty acids metabolism in growth at low temperatures is largely known. For example, the fluidity of the rigidified membrane (42) imposed by low temperatures can be restored through the modulation of the (i) saturated and unsaturated fatty acids, (ii) fatty acid chain length, and (iii) the proportion of *cis* to *trans* fatty acids thus, ultimately, through the modulation of fatty acids metabolism (as observed in this work). This also validates and extends previous findings on general cold-adaptation strategies (16, 43, 44), pinpointing the role of specific metabolic reactions in this process.

We also found an increased PPP activity at 15° with respect to 0°C (Fig. 5G). Increased PPP activity signifies increased NADPH production, and is linked with many biosynthetic processes. Interestingly, high activity of the PPP pathway has been described as an adaptive mechanism (in soil bacteria) to temperature stress (45), in particular by linking the requirement of sugar units for biofilm formation to the overall activity in this metabolic pathway. We know for a fact from the work of (18) that *PhTAC125* displays strong propensity to form more biofilm at 15°C versus 0°C when grown in the same minimal medium used in this work. Also, we found a remarkable fraction of overexpressed biofilm-related genes (Fig. 4C), reinforcing the idea of a temperature-mediated switch to this kind of lifestyle. Thus, we infer that a similar mechanism is at play in this marine bacterium where the redirection of fluxes in PPP at 15°C would be the metabolic basis for the formation of cellular aggregates at higher temperatures. Moreover, the increased PPP activity can respond to other cell requirements linked to the DNA replication recombination and repair, since the ribose produced by PPP is necessary for these cellular activities that result more active at 15°C (Table S2), as suggested by transcriptomic analysis (Fig. 5B).

Very recently, Macarena Toll-Riera et al. published (28) a work in which *PhTAC125* was evolved at increasingly higher temperatures to study the evolutionary potential of upper thermal tolerance, and characterize the genomic basis of temperature adaptation. Interestingly, the authors found almost no metabolic genes among those that apparently provided an increased fitness at higher temperatures. Indeed, clones that were selected for their improved growth at higher temperatures (up to 30°C) had mainly mutations in proteases-encoding genes or in genes involved in chromosome

copy number reduction, energy production and conversion and cell wall biosynthesis. Thus, apparently, the metabolism of *PhTAC125* was not the primary target of mutations that led to an increased fitness in the tested conditions. This, in turn, may indicate the presence of an already optimized and plastic metabolism that allows growth in a broad range of temperatures, in line with the metabolic robustness that we have characterized in this work.

Conclusions. The aim of our experiments was to evaluate the effect of growth temperatures on cellular homeostasis, particularly at the metabolic level. We have shown that a cold-adapted marine bacterium expresses 2 very similar metabolic phenotypes in response to 2 widely different temperatures, by adjusting the expression of key enzymes and fine-tuning the intracellular concentration of key intermediates. This illuminated on the possible molecular mechanisms that marine microbes may use to adapt to broad (seasonal) temperature changes. In particular, we showed that specific changes at the level of regulatory circuits can buffer such variations, and maintain the underlying metabolic network robust to temperature fluctuations. In the future, it will be interesting to investigate whether similar mechanisms are at play in natural microbial assemblages.

MATERIALS AND METHODS

Strains and growth conditions. *PhTAC125* (46) cells were grown in a 1.5 L GG medium (47) in a Stirred Tank Reactor 3 L fermenter (Applikon) connected to an eZ2 Bio Controller (Applikon) at 2 different temperatures (0°C and 15°C). This medium contains glutamate and gluconate as the only carbon sources. The bioreactor was equipped with the standard pH-, pO₂-, level- and temperature sensors for the bioprocess monitoring. For the growths of the *PhTAC125* bacterium, the pre-culture was centrifuged (6000 × g, 20 min, 4°C); the cells were washed twice with fresh medium, and then used to inoculate the bioreactor with a starting OD₆₀₀ of 0.2, in aerobic conditions (50% and 30% dissolved oxygen at 15°C and 0°C, respectively), in stirring (500 rpm at 15°C and 250 rpm at 0°C). The bacterial culture was carried out at 15°C for 40 h or at 0°C for 240 h. Each culture condition was repeated three times. Cell growth was monitored, measuring the OD₆₀₀ about every 2 h in the experiments at 15°C, and every 8 h at 0°C. Three different measurements were performed at each time point for each biological replicate. For intra- and extracellular metabolites analysis, the samples were taken in triplicate at 5 different time points during the growth in GG medium at 2 different temperatures (70 h, 141 h, 179 h, 190 h, and 240 h at 0°C, and 6.5 h, 14 h, 20 h, 25 h, and 39 h at 15°C). For the analysis of extracellular metabolites, aliquots (1 mL) of cell cultures were harvested during the growth and centrifuged for 15 min at 13,000g at 4°C; the supernatant was recovered, filtered (Filtropur 0.2 µm, SARSTED AG & Co. KG) and stored at −80°C. The analysis of intracellular metabolites was performed on 60 OD₆₀₀ pellets recovered during the growth by centrifuging for 20 min at 6,000 rpm at 4°C.

Transcriptomics. For the RNA-Seq experiment, 1 OD₆₀₀ pellets were recovered during the exponential growth phase (~1/1.5 OD/mL) at 15°C and 0°C by centrifugation (10 min, 13,000g, 4°C). Cell pellets were washed in RNase-free PBS three times and stored at −80°C. Total RNA was isolated from the cells using the Direct-zol RNA Kit (Zymo Research) following the manufacturer's instructions. Contaminating genomic DNA was then removed through treatment with RNase-free DNase I (Roche).

RNA sequencing. RNA concentration was measured using Quant-IT RNA assay kit-high sensitivity and a Qubit Fluorometer (Life Technologies), and its quality and integrity assessed with the Agilent 4200 TapeStation System (Agilent Technologies). Indexed libraries were prepared starting from 400 ng of total RNA according to Universal Prokaryotic RNA-Seq Library Prep kit (Tecan). Final libraries were sequenced at a concentration of 1.7 pM/lane on the NextSeq 500 platform (Illumina Inc) in paired mode 2 × 75bp. For each experimental condition, 3 biological replicates were prepared.

RNA-Seq data analysis. The 12 samples were assessed for base call quality and adapter content using fastp (48), allowing down to a mean quality threshold of 20 (i.e., probability of incorrect base call of 1 in 100) and minimum read length of 40 nucleotides. A median of 98.9% of the reads passed quality check, indicating that sequencing was carried out pristinely, and our data were biologically reliable. Salmon (49) index was built on *PhTAC125*'s transcriptome free of tRNA and rRNA sequences, providing the entire genomic sequence as background decoy to account for possible underlying DNA contamination. CDS and genomic FASTA were retrieved on the NCBI, using assembly accession [GCF_000026085.1](https://www.ncbi.nlm.nih.gov/assembly/GCF_000026085.1). Moreover, CDS and genomic files were concatenated with data from pMEGA plasmid (NZ_MN400773.1) and pMtBL plasmid (NZ_AJ224742.1). The transcriptome quantification step was performed using the `validateMappings` flag, to ensure a sensitive selective alignment of the sequencing reads. The resulting mapping rates ranged between 56.8% and 66.21% (average of 62.47%), whereas roughly half of the reads represented ribosomal DNA sequences (discarded for downstream analyses). The integration of transcript-level abundance estimates from Salmon with the data analysis pipeline was performed using R package *tximport* (R Core Team, 2022, <https://www.R-project.org/>, R version 4.0.3). The testing of changes in the overall transcriptional output was handled through the statistical engine in R package DESeq2, with an expected proportion of false positives set at 5%. The volcano scatterplot showing statistical significance versus magnitude of change for 607 differentially expressed genes was produced using the library EnhancedVolcano from the R Bioconductor 3.14 suite.

Metabolomics. ^1H NMR-based metabolomic analyses were performed on cell lysates and growth media to monitor the intracellular metabolites, and the uptake and release of the extracellular metabolites, respectively, by measuring their concentration levels in samples collected at different time points during the cell growth.

Medium samples were prepared in 5.00 mm NMR tubes by mixing 60 μL of a potassium phosphate buffer (1.5 M K_2HPO_4 , 100% (vol/vol) $^2\text{H}_2\text{O}$, 10 mM sodium trimethylsilyl [2,2,3,3- $^2\text{H}_4$]propionate (TMSP), pH 7.4), and 540 μL of each growth medium.

Cell lysate samples were prepared in 5.00 mm NMR tubes by mixing 60 μL of $^2\text{H}_2\text{O}$ and 540 μL of samples. All the NMR spectra were recorded using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency, and equipped with a 5 mm PATXI ^1H - ^{13}C - ^{15}N and ^2H -decoupling probe including a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). A BTO 2000 thermocouple served for temperature stabilization at the level of approximately 0.1 K at the sample. Before measurement, samples were kept for 5 min inside the NMR probe head, for temperature equilibration at 300 K.

^1H NMR spectra were acquired with water peak suppression and a standard NOESY pulse sequence using 128 scans, 65536 data points, a spectral width of 12019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mixing time of 0.1 s.

The raw data were multiplied by a 0.3 Hz exponential line broadening before applying Fourier transformation. Transformed spectra were automatically corrected for phase and baseline distortions. All the spectra were then calibrated to the reference signal of TMSP at δ 0.00 ppm using TopSpin 3.5 (Bruker BioSpin srl).

The metabolites, whose peaks in the spectra were well resolved, were assigned, and their levels analyzed using a dedicated R script developed in-house. In total, 34 and 17 metabolites were identified and quantified in the cell lysate and in the growth medium spectra, respectively. The assignment was performed using an internal ^1H NMR spectral library of pure organic compounds (BBIOREFCODE, Bruker BioSpin), stored reference NMR spectra of metabolites, and spiking experiments. Matching between new NMR data and databases was performed using the Assure NMR software (Bruker BioSpin). The relative concentrations of the various metabolites were calculated by integrating the corresponding signals in defined spectral ranges, using in-house developed R 3.0.2 scripts. Similarly, downstream data analysis was performed using R. Raw data, post-Raw data, and post-processing codes are made available at <https://github.com/combogenomics/MetRob015>.

Uptake rates of gluconate and glutamate at the 2 different temperatures were computed as the ratio between the average growth rate between T1 and T3 (μ) and the biomass yield (λ). The former was computed as:

$$\mu = \frac{\log OD_{T3} - \log OD_{T1}}{T3 - T1}$$

The latter was computed according to the following relationship:

$$\lambda = \frac{\text{Biomass } (\frac{\text{g}}{\text{L}})}{\text{Consumed C source (mol)}}$$

Biomass was obtained from OD values as described in (50), using 0.74 as a scaling factor for the growth at 15° and 0.66 for the growth at 0°C.

Genome-scale metabolic modeling. The genome-scale metabolic reconstruction used in this work is the one recently used in (51). All the simulations were run in MATLAB 2019a, using the COBRA toolbox (52) version 2.7.4. Metabolomic and transcriptomic data were integrated using REMI method (29), providing TPM gene expression and relative concentration values for metabolites at time point T1. Both for metabolites and gene expression, we selected the top 3% as upregulated and the bottom 3% as downregulated (REMI default is 5%). Consistently with the original REMI publication, we used the 2-fold change as the cutoff threshold to identify the significant gene expression and metabolite changes. All the other parameters were left as default. The predicted internal concentration of each metabolite was computed using the *computeFluxSplits* function implemented in the COBRA toolbox. This function computes the relative contributions of fluxes to the net production and consumption of a specific set of metabolites included in the model. Statistical tests on flux distributions were calculated using R.

To estimate the difference in the activity of the main central metabolic pathways at 0° and 15°C, we computed the predicted flux of each reaction in these pathways (TCA cycle, glycolysis, PPP, and fatty acids metabolism) at the 2 different temperatures, and averaged this number of reactions by the total number of reactions included in that pathway. From this latter analysis, we excluded those reactions that (i) had a flux equal to zero (were inactive) in both conditions and (ii) had a different sign (i.e., changed direction) at the 2 temperatures.

Data availability. The authors confirm that the data supporting the findings of this study are available within the article and/or its supplemental materials. Collection of sequence data produced in this study is available under the BioProject [PRJNA886636](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA886636).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.

FIG S2, PDF file, 0.1 MB.

FIG S3, PDF file, 0.04 MB.

FIG S4, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.01 MB.

TABLE S2, PDF file, 0.02 MB.

ACKNOWLEDGMENTS

We thank Dr. Assunta Sellitto for technical assistance for RNA-Seq libraries preparation.

P.T. and V.G. acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and, specifically, the CERM/CIRMMP Italy Centre.

This study was supported by Regione Campania, Progetto GENOMAeSALUTE (POR CAMPANIA FESR 2014/2020, azione 1.5; CUP: B41C17000080007), by PNRA (Programma Nazionale di Ricerche in Antartide) grant PNRA18_00075 and PNRA18_00335, and by a PRIN-MUR (RESEARCH PROJECTS OF RELEVANT NATIONAL INTEREST– 2020 Call) Project: 20208LLXEJ.

REFERENCES

- Rothschild LJ, Mancinelli RL. 2001. Life in extreme environments. *Nature* 409:1092–1101. <https://doi.org/10.1038/35059215>.
- Shu W-S, Huang L-N. 2022. Microbial diversity in extreme environments. *Nat Rev Microbiol* 20:219–235. <https://doi.org/10.1038/s41579-021-00648-y>.
- Hegedus S, Csonka J. 2010. Astrobiology: physical origin, biological evolution, and spatial distribution, p 27–57. Nova Science Publishers, New York, NY.
- Bennett AF, Lenski RE. 1997. Phenotypic and evolutionary adaptation of a model bacterial system to stressful thermal environments, p 135–154. *In* Bijlsma R, Loeschke V (ed), *Environmental Stress, Adaptation and Evolution*. Birkhäuser Basel, Basel, Switzerland.
- Saarienen K, Laakso J, Lindström L, Ketola T. 2018. Adaptation to fluctuations in temperature by nine species of bacteria. *Ecol Evol* 8:2901–2910. <https://doi.org/10.1002/ece3.3823>.
- Walworth NG, Zakem EJ, Dunne JP, Collins S, Levine NM. 2020. Microbial evolutionary strategies in a dynamic ocean. *Proc Natl Acad Sci U S A* 117:5943–5948. <https://doi.org/10.1073/pnas.1919321117>.
- Auger M, Morrow R, Kestenare E, Sallée J-B, Cowley R. 2021. Southern Ocean *in-situ* temperature trends over 25 years emerge from interannual variability. *Nat Commun* 12:514. <https://doi.org/10.1038/s41467-020-20781-1>.
- Doblin MA, van Sebille E. 2016. Drift in ocean currents impacts intergenerational microbial exposure to temperature. *Proc Natl Acad Sci U S A* 113:5700–5705. <https://doi.org/10.1073/pnas.1521093113>.
- Kim H, Ducklow HW. 2016. A decadal (2002–2014) analysis for dynamics of heterotrophic bacteria in an Antarctic coastal ecosystem: variability and physical and biogeochemical forcings. *Front Mar Sci* 3. <https://doi.org/10.3389/fmars.2016.00214>.
- Fiehn O. 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171. <https://doi.org/10.1023/A:1013713905833>.
- Lu H, Ulanov AV, Nobu M, Liu W-T. 2016. Global metabolomic responses of *Nitrosomonas europaea* 19718 to cold stress and altered ammonia feeding patterns. *Appl Microbiol Biotechnol* 100:1843–1852. <https://doi.org/10.1007/s00253-015-7095-y>.
- Czajka JJ, Abernathy MH, Benites VT, Baidoo EEK, Deming JW, Tang YJ. 2018. Model metabolic strategy for heterotrophic bacteria in the cold ocean based on *Colwellia psychrerythraea* 34H. *Proc Natl Acad Sci U S A* 115:12507–12512. <https://doi.org/10.1073/pnas.1807804115>.
- Kurdril P, Phuengcharoen P, Senachak J, Saree S, Hongsthong A. 2020. Revealing the key point of the temperature stress response of *Arthrospira platensis* C1 at the interconnection of C- and N- metabolism by proteome analyses and PPI networking. *BMC Mol and Cell Biol* 21:43. <https://doi.org/10.1186/s12860-020-00285-y>.
- Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L. 2010. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol* 6:364. <https://doi.org/10.1038/msb.2010.18>.
- Birolo L, Tutino ML, Fontanella B, Gerday C, Mainolfi K, Pascarella S, Sannia G, Vinci F, Marino G. 2000. Aspartate aminotransferase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125. *Eur J Biochem* 267:2790–2802. <https://doi.org/10.1046/j.1432-1327.2000.01299.x>.
- Fondi M, Maida I, Perrin E, Mellera A, Mocali S, Parrilli E, Tutino ML, Liò P, Fani R. 2015. Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125: Modelling of *P. haloplanktis* TAC125 metabolism. *Environ Microbiol* 17:751–766. <https://doi.org/10.1111/1462-2920.12513>.
- Mocali S, Chiellini C, Fabiani A, Decuzzi S, de Pascale D, Parrilli E, Tutino ML, Perrin E, Bosi E, Fondi M, Lo Giudice A, Fani R. 2017. Ecology of cold environments: new insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. *Sci Rep* 7:839. <https://doi.org/10.1038/s41598-017-00876-4>.
- Ricciardelli A, Casillo A, Vergara A, Balasco N, Corsaro MM, Tutino ML, Parrilli E. 2019. Environmental conditions shape the biofilm of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Microbiol Res* 218:66–75. <https://doi.org/10.1016/j.micres.2018.09.010>.
- Parrilli E, Tedesco P, Fondi M, Tutino ML, Lo Giudice A, de Pascale D, Fani R. 2021. The art of adapting to extreme environments: the model system *Pseudoalteromonas*. *Phys Life Rev* 36:137–161. <https://doi.org/10.1016/j.plrev.2019.04.003>.
- Landsman D. 1992. RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res* 20:2861–2864. <https://doi.org/10.1093/nar/20.11.2861>.
- Yamanaka K, Fang L, Inouye M. 1998. The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Mol Microbiol* 27:247–255. <https://doi.org/10.1046/j.1365-2958.1998.00683.x>.
- Keto-Timonen R, Hietala N, Palonen E, Hakakorpi A, Lindström M, Korkeala H. 2016. Cold shock proteins: a minireview with special emphasis on Csp-family of enteropathogenic *Yersinia*. *Front Microbiol* 7:1151. <https://doi.org/10.3389/fmicb.2016.01151>.
- Jiang W, Hou Y, Inouye M. 1997. CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J Biol Chem* 272:196–202. <https://doi.org/10.1074/jbc.272.1.196>.
- Caballero CJ, Menendez-Gil P, Catalan-Moreno A, Vergara-Irigaray M, García B, Segura V, Irurzun N, Villanueva M, Ruiz de Los Mozos I, Solano C, Lasa I, Toledo-Arana A. 2018. The regulon of the RNA chaperone CspA and its auto-regulation in *Staphylococcus aureus*. *Nucleic Acids Res* 46:1345–1361. <https://doi.org/10.1093/nar/gkx1284>.
- Phadtare S, Severinov K. 2010. RNA remodeling and gene regulation by cold shock proteins. *RNA Biol* 7:788–795. <https://doi.org/10.4161/rna.7.6.13482>.
- Cairão F, Cruz A, Mori H, Arraiano CM. 2003. Cold shock induction of RNase R and its role in the maturation of the quality control mediator SsrA/tmRNA. *Mol Microbiol* 50:1349–1360. <https://doi.org/10.1046/j.1365-2958.2003.03766.x>.
- Angilletta MJ, Jr. 2009. Thermal adaptation. Oxford University Press, England.
- Toll-Riera M, Olombrada M, Castro-Giner F, Wagner A. 2022. A limit on the evolutionary rescue of an Antarctic bacterium from rising temperatures. *Sci Adv* 8:eabk3511. <https://doi.org/10.1126/sciadv.abk3511>.
- Pandey V, Hadadi N, Hatzimanikatis V. 2019. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into

- thermodynamically consistent metabolic models. *PLoS Comput Biol* 15: e1007036. <https://doi.org/10.1371/journal.pcbi.1007036>.
30. Brandi A, Pon CL. 2012. Expression of *Escherichia coli* cspA during early exponential growth at 37°C. *Gene* 492:382–388. <https://doi.org/10.1016/j.gene.2011.10.047>.
 31. Horn G, Hofweber R, Kremer W, Kalbitzer HR. 2007. Structure and function of bacterial cold shock proteins. *Cell Mol Life Sci* 64:1457–1470. <https://doi.org/10.1007/s00018-007-6388-4>.
 32. Xia B, Ke H, Inouye M. 2001. Acquisition of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in *Escherichia coli*: S1 domain complements cold-shock protein deficiency. *Mol Microbiol* 40:179–188. <https://doi.org/10.1046/j.1365-2958.2001.02372.x>.
 33. Phadtare S, Inouye M. 2001. Role of CspC and CspE in regulation of expression of RpoS and UspA, the stress response proteins in *Escherichia coli*. *J Bacteriol* 183:1205–1214. <https://doi.org/10.1128/JB.183.4.1205-1214.2001>.
 34. Phadtare S, Tadigotla V, Shin W-H, Sengupta A, Severinov K. 2006. Analysis of *Escherichia coli* global gene expression profiles in response to overexpression and deletion of CspC and CspE. *J Bacteriol* 188:2521–2527. <https://doi.org/10.1128/JB.188.7.2521-2527.2006>.
 35. Bae W, Xia B, Inouye M, Severinov K. 2000. *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. *Proc Natl Acad Sci U S A* 97:7784–7789. <https://doi.org/10.1073/pnas.97.14.7784>.
 36. Phadtare S, Inouye M, Severinov K. 2002. The nucleic acid melting activity of *Escherichia coli* CspE is critical for transcription antitermination and cold acclimation of cells. *J Biol Chem* 277:7239–7245. <https://doi.org/10.1074/jbc.M111496200>.
 37. Phadtare S, Severinov K. 2005. Nucleic acid melting by *Escherichia coli* CspE. *Nucleic Acids Res* 33:5583–5590. <https://doi.org/10.1093/nar/gki859>.
 38. Guan Y, Yin D, Du X, Ye X. 2018. Metabolomics approach used for understanding temperature-related pectinase activity in *Bacillus licheniformis* DY2. *FEMS Microbiology Lett* 365. <https://doi.org/10.1093/femsle/fny255>.
 39. Kim S, Kim Y, Suh DH, Lee CH, Yoo SM, Lee SY, Yoon SH. 2020. Heat-responsive and time-resolved transcriptome and metabolome analyses of *Escherichia coli* uncover thermo-tolerant mechanisms. *Sci Rep* 10:17715. <https://doi.org/10.1038/s41598-020-74606-8>.
 40. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M. 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316:593–597. <https://doi.org/10.1126/science.1132067>.
 41. Massey LK, Sokatch JR, Conrad RS. 1976. Branched-chain amino acid catabolism in bacteria. *Bacteriol Rev* 40:42–54. <https://doi.org/10.1128/br.40.1.42-54.1976>.
 42. Los DA, Murata N. 2004. Membrane fluidity and its roles in the perception of environmental signals. *Biochim Biophys Acta* 1666:142–157. <https://doi.org/10.1016/j.bbame.2004.08.002>.
 43. Graumann PL, Marahiel MA. 1999. Cold shock response in *Bacillus subtilis*. *J Mol Microbiol Biotechnol* 1:203–209.
 44. Ting L, Williams TJ, Cowley MJ, Lauro FM, Guilhaus M, Raftery MJ, Cavicchioli R. 2010. Cold adaptation in the marine bacterium, *Sphingopyxis alaskensis*, assessed using quantitative proteomics. *Environ Microbiol* 12:2658–2676. <https://doi.org/10.1111/j.1462-2920.2010.02235.x>.
 45. Bore EK, Apostel C, Halicki S, Kuzyakov Y, Dippold MA. 2017. Microbial metabolism in soil at subzero temperatures: adaptation mechanisms revealed by position-specific ¹³C labeling. *Front Microbiol* 8:946. <https://doi.org/10.3389/fmicb.2017.00946>.
 46. Médigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN, Cheung F, Cruveiller S, D'Amico S, Duilio A, Fang G, Feller G, Ho C, Mangenot S, Marino G, Nilsson J, Parrilli E, Rocha EPC, Rouy Z, Sekowska A, Tutino ML, Vallenet D, von Heijne G, Danchin A. 2005. Coping with cold: the genome of the versatile marine Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* 15:1325–1335. <https://doi.org/10.1101/gr.4126905>.
 47. Sannino F, Giuliani M, Salvatore U, Apuzzo GA, de Pascale D, Fani R, Fondi M, Marino G, Tutino ML, Parrilli E. 2017. A novel synthetic medium and expression system for subzero growth and recombinant protein production in *Pseudoalteromonas haloplanktis* TAC125. *Appl Microbiol Biotechnol* 101:725–734. <https://doi.org/10.1007/s00253-016-7942-5>.
 48. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
 49. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419. <https://doi.org/10.1038/nmeth.4197>.
 50. Giuliani M, Parrilli E, Ferrer P, Baumann K, Marino G, Tutino ML. 2011. Process optimization for recombinant protein production in the psychrophilic bacterium *Pseudoalteromonas haloplanktis*. *Process Biochemistry* 46:953–959. <https://doi.org/10.1016/j.procbio.2011.01.011>.
 51. Fondi M, Gonzi S, Dziurzynski M, Turano P, Ghini V, Calvanese M, Colarusso A, Lauro C, Parrilli E, Tutino ML. 2021. Modelling hCDKL5 heterologous expression in bacteria. *Metabolites* 11:491. <https://doi.org/10.3390/metabo11080491>.
 52. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, et al. 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 14:639–702. <https://doi.org/10.1038/s41596-018-0098-2>.

3.2.3 How to study COVID impacts with GSMM

In the last twenty years there have been developed several metabolic models and not only of bacteria, but of a wide range of organisms, and their complexity varies according to the complexity of the organism's metabolism. In the present study, the aim was to report a detailed and comprehensive characterization of the metabolomic and lipoproteomic fingerprint of hospitalized COVID-19 patients, hence, to use the GSMM approach, it was necessary to use the human metabolic reconstruction. The latest human GSMM is based on the integrated metabolism of 26 organs of the human body, plus six blood cells types, and there are two reconstructions, since it provides sex-specific whole-body metabolism: *Harvey* is the male GSMM while *Harvetta* the female one.

A total of 605 EDTA-plasma samples were collected during various waves of the pandemic, each associated with the emergence of different SARS-CoV-2 variants, various disease severity and reflecting diverse vaccination statuses among the patients. Most of the samples came from patients during the acute phase of the infection, and a third of them (154 out of 510) were sampled after at least two months after the first negative swab. The rest of the samples (95) came from previously hospitalized recovered subjects. As a result of the different samplings, the data were classified as mild, severe or fatal infection.

Using the metabolomic information as input data, we contextualize both *Harvey* and *Harvetta* models to provide a systems-level understanding of the impact of COVID-19 on human metabolism. The metabolomic data was divided between male and female patients to have sex-specific datasets and then divided based on disease condition. We defined a *healthy* condition by not constraining the GSMM with metabolomic data. We calculated the fold change of each metabolite for the following conditions: recovered vs. mild, recovered vs. severe, recovered vs. fatal. We then assessed the significance of the changes by using the Wilcoxon test, and those metabolites that resulted statistically different in the examined conditions were used to constrain the models. More in detail, we modified the upper

and lower bounds (UB and LB, respectively) of each i_{th} exchange metabolite using the default boundary value (BV) and the fold change, as follows:

$$\begin{aligned}
 UB > 0 \wedge LB < 0 & \begin{cases} FC > 0; & x_{ub} = BV * FC \text{ and } x_{lb} = \frac{BV}{FC} \\ FC < 0; & x_{ub} = \frac{BV}{|FC|} \text{ and } x_{lb} = BV * |FC| \end{cases} \\
 UB \wedge LB < 0 & \begin{cases} FC > 0; & x_{ub} \wedge x_{lb} = \frac{BV}{FC} \\ FC < 0; & x_{ub} \wedge x_{lb} = BV * |FC| \end{cases} \\
 UB \wedge LB > 0 & \begin{cases} FC > 0; & x_{ub} \wedge x_{lb} = BV * FC \\ FC < 0; & x_{ub} \wedge x_{lb} = \frac{BV}{|FC|} \end{cases}
 \end{aligned}$$

Where x_{ub} and x_{lb} represent actual upper and lower bounds of the constrained reaction in the model after data integration.

This model revealed several set of pathways with different characteristics, *i*) those with an increased activity concordantly with disease severity both in male and female, *ii*) those with a higher metabolic activity in the comparison between healthy and fatal states, *iii*) those that displayed sex-specific increase in their activity and *iv*) those showing opposite trends in the sex-specific simulations. However, all these results suggested that the disease induces a metabolic signature throughout the human body, with the severity of the condition determining the extent of the metabolic disruptions. In other words, those showing the most pronounced dysregulation exhibit concentration trends that strongly correlate with disease severity. Another notable finding is the marked metabolic reprogramming observed across multiple organs, which points out an increased overall energetic demand of the organism following infection. Furthermore, the metabolites correlating the fatal outcome of the disease can be proposed as predictive and prognostic biomarkers.

This integrative study not only provided insights into the biochemical foun-

dations of COVID-19 but also highlighted, once again, the utility of multidisciplinary methods, such as metabolomics and computational modeling, in unraveling the systemic effects of infectious diseases.

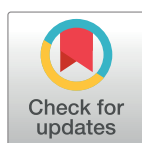
RESEARCH ARTICLE

COVID-19: A complex disease with a unique metabolic signature

Veronica Ghini^{1,2}, Walter Vieri^{1,3}, Tommaso Celli⁴, Valentina Pecchioli^{1,2}, Nunzia Boccia⁴, Tania Alonso-Vázquez³, Lorenzo Pelagatti⁴, Marco Fondi³, Claudio Luchinat^{1,2,5}, Laura Bertini⁴, Vieri Vannucchi⁴, Giancarlo Landini^{4*}, Paola Turano^{1,2,5*}

1 Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino Florence, Italy, **2** Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Florence, Italy, **3** Department of Biology, University of Florence, Sesto Fiorentino, Florence, Italy, **4** Internal Medicine, Santa Maria Nuova Hospital, Florence, Florence, Italy, **5** Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Sesto Fiorentino Florence, Italy

* giancarlo.landini@uslcentro.toscana.it (GL); turano@cerm.unifi.it (PT)



OPEN ACCESS

Citation: Ghini V, Vieri W, Celli T, Pecchioli V, Boccia N, Alonso-Vázquez T, et al. (2023) COVID-19: A complex disease with a unique metabolic signature. *PLoS Pathog* 19(11): e1011787. <https://doi.org/10.1371/journal.ppat.1011787>

Editor: Hung Nguyen, University of Central Florida, UNITED STATES

Received: April 24, 2023

Accepted: October 30, 2023

Published: November 9, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1011787>

Copyright: © 2023 Ghini et al. 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 use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: This study is available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench, <https://www.metabolomicsworkbench.org> where it has been

Abstract

Plasma of COVID-19 patients contains a strong metabolomic/lipoproteomic signature, revealed by the NMR analysis of a cohort of >500 patients sampled during various waves of COVID-19 infection, corresponding to the spread of different variants, and having different vaccination status. This composite signature highlights common traits of the SARS-CoV-2 infection. The most dysregulated molecules display concentration trends that scale with disease severity and might serve as prognostic markers for fatal events. Metabolomics evidence is then used as input data for a sex-specific multi-organ metabolic model. This reconstruction provides a comprehensive view of the impact of COVID-19 on the entire human metabolism. The human (male and female) metabolic network is strongly impacted by the disease to an extent dictated by its severity. A marked metabolic reprogramming at the level of many organs indicates an increase in the generic energetic demand of the organism following infection. Sex-specific modulation of immune response is also suggested.

Author summary

Metabolites and lipoproteins are the main components of human plasma and their concentration can be determined by nuclear magnetic resonance. In COVID-19 patients there are significant alterations in the concentration of several of these molecules. Using the plasma of more than 600 subjects (510 patients sampled in the acute phase of the infection plus 95 independent recovered subjects), we demonstrate that the dysregulation of these molecules is a function of the disease severity but it is not affected by either the SARS-CoV-2 variants or the vaccination status. The disease signature is particularly evident in those cases that subsequently evolve towards a fatal outcome, and could have prognostic value. Building on this large amount of data, we propose a metabolic reconstruction of the disease using a sex-specific multi-organ metabolic model.

assigned Study ID ST002404. The data can be accessed directly via its Project DOI: <http://dx.doi.org/10.21228/M89T2Q>.

Funding: This work was funded by Regione Toscana, project COMETA, Bando COVID-19 (<https://www.regione.toscana.it/-/bando-ricerca-covid-19-toscana>) to GL and PT. The project also covers the salaries of VG, WV, VP, TC and NB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

From the year 2020, the resilience of worldwide national health systems was profoundly challenged by the Coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The viral infection is characterized by a broad spectrum of clinical manifestations from an asymptomatic or pauci-symptomatic disease (in more than 80% of subjects), to interstitial pneumonia and acute respiratory distress syndrome requiring hospitalization and even ventilation of the patients [1–3]. Moreover, it has been demonstrated that a significant subset of patients developed concomitant multi-organ dysfunctions, including acute kidney and liver injuries, thromboembolism and sepsis that contribute to a fatal outcome [4,5]. Additionally, some comorbidities have been proposed as risk factors of disease severity and fatal outcome [6,7]. Although vaccines are now available and have demonstrated high efficacy in decreasing the severity of SARS-CoV-2, vaccination does not prevent SARS-CoV-2 transmission. The diffusion of the different viral variants as well as the phenomenon of the so called “Long-COVID”, i.e. long-term effects associated to COVID-19 infection [8], further complicate this scenario.

Massive worldwide efforts by research groups using omics sciences have been made to unravel the disease mechanisms and to identify biomarkers of the disease severity [9–13]. In this framework, ¹H NMR [14–17] plays a role for its ability to reveal a complex blood plasma signature exhibiting the presence of a strong fingerprint of the COVID-19 disease. Highly reproducible alterations of a large number of blood metabolites and lipoprotein parameters were identified as markers of the disease, suggesting the reprogramming of important metabolic pathways aimed at the energy supply for viral replication and for host immunological response [18–29]. However, how the administration of vaccines and the development of the different viral variants impact on the disease fingerprint, remain poorly characterized aspects.

Here, we report a detailed and comprehensive characterization of the metabolomic and lipoproteomic fingerprint of plasma samples of > 500 hospitalized COVID-19 patients, with different disease severities, infected with different viral variants and with different vaccination status. Sex-specific differences as well as the contribution of several comorbidities are also analysed. Our data deeply extend a first metabolomic/lipoproteomic characterization of the disease published at the beginning of 2022 [18], performed on a smaller number of subjects of the same cohort, infected before a significant spread of the δ variant and before widespread COVID-19 vaccination. While confirming previously identified severity markers [18–28], we establish for the first time a correlation between the levels of a few metabolites and lipoproteins and the fatal outcome of the disease. These molecules can therefore be proposed as predictive and prognostic biomarkers. The observed changes are interpreted through simulations of the overall metabolic state of the human body with a recently developed sex-specific multi-organ metabolic model [30], which until now has been used to predict known biomarkers of inherited metabolic diseases in different biofluids. Based on this reconstruction and using as input the changes in the metabolome observed in our cohort, we obtained a comprehensive view of the impact of COVID-19 on the entire human metabolism, which represents a step forward with respect to previous models based on not cell type-specific and not sex-specific flux balance analyses [31].

Results

In this study, EDTA-plasma samples from 510 COVID-19 positive patients (COVID-19 group) were collected during the period 20/06/2020–17/06/2022. The demographic and clinical data of the cohort are reported in [S1 Table](#). This population well represents the incidence of the disease in Tuscan hospitals along the course of the pandemic, in terms of sex, age, severity, and main risk factors/comorbidities. One hundred and fifty-four of these individuals were also

evaluated at a follow-up visit; this group is hereafter named as the “follow-up group”, [S2 Table](#). The reference group COVID-19-R ([S1 Table](#)) is constituted by 95 recovered subjects who had contracted the infection during the first wave and did not show any symptom of long COVID at the follow-up visit, when they were sampled. Consistently their metabolite and lipoprotein levels fall within the normality ranges [[19](#)].

Fingerprint of the main COVID-19 variants

To evaluate whether the metabolic and lipoprotein fingerprint of the disease was significantly changed as a function of the different variants of the virus, we selected 3 groups of samples. Based on the data of the statistical incidence of the various SARS-Cov-2 variants in the Tuscany Region (TR) (TR does not systematically determine the variants of each patient through genomic sequencing), only patients infected in periods characterized by defined dominant variants were included. In particular, we considered: i) 87 samples collected when the Wild-Type (WT), α , and β variants were prevalent (01/2020-06/2021), the “wt- α - β group”—these samples are new with respect to those included in our previous study [[18](#)]; ii) 91 samples collected when the δ variant was dominant, i.e. the “ δ group” (07/2021-12/2021); iii) 44 samples collected when the o variant was dominant (01/2022-02/2022), the “o group”. Together, the three groups account for a total of 222 samples. The detailed demographic and clinical characteristics of these subjects are reported in [Fig 1A](#).

From the multivariate statistics (see [Methods](#)) no clear differences emerge among the metabolomic fingerprints of the three groups of variants; instead, all of them can be almost perfectly discriminated from COVID-19 recovered subjects ([S1 Fig](#)). The three groups show very similar trends for both metabolite and lipoprotein levels, [Figs 1B–1C and 2](#). Conversely, eleven out of 25 metabolites and 16 out of 30 lipoprotein main parameters and main fractions result to be significantly different from the recovered subjects, regardless of the variant, [Figs 1B–1C and 2](#).

The three ketone bodies (3-hydroxybutyrate, acetoacetic acid and acetone) along with the amino acids Phe, Met and Ile, the sugars glucose and mannose, and the glycoproteins GlycA and GlycB, are significantly increased in all COVID-19 groups. Citric acid and acetic acid, instead, show decreased levels in all COVID-19 groups. Histidine is the only molecule that shows a different direction of changes among the three groups with respect to the COVID-19-R group; it decreased in the wt- α - β group while it increased in the δ and o groups.

For a few other metabolites (Val, Tyr, lactic acid, creatine, Leu, Gln, formic acid, Ala) we observe the same trends of changes among the three groups, although their variations are not consistently significant. The lack of significance could result from different number of samples in the various groups combined with relatively small fold changes ($|\text{Log}_2(\text{FC})| < 1$).

As for lipoproteins, all the parameters, except free cholesterol (Chol) and apolipoprotein B100 (Apo B100) associated to IDL, show the same trends along all the variants, [Fig 2](#). Significantly decreased level of LDL- and HDL-Chol, total Apo A1 and Apo A2, Apo A1-HDL, Apo A2-HDL, triglycerides (TG)-HDL and phospholipids (PhL)-LDL, -IDL and -HDL were observed in all the COVID-19 groups with respect to the recovered subjects. Increased values of the ratio Apo B100/Apo A1 and of TG-LDL are also coherently observed. Moreover, significantly increased levels of free Chol-LDL are observed only in wt- α - β and δ groups, whereas increased levels of the Apo B100-VLDL are observed only in the wt- α - β and o variants. A significant decrement of ApoB-LDL is instead only reported for the δ group. Lipoprotein sub-fractions and particle numbers are reported in [S2 Fig](#).

Fingerprint of the disease as a function of the vaccination status

Among the 510 COVID-19 positive hospitalized subjects, 71 were vaccinated (with two or more doses of DNA or RNA vaccine, administered at least 15 days before the first positive

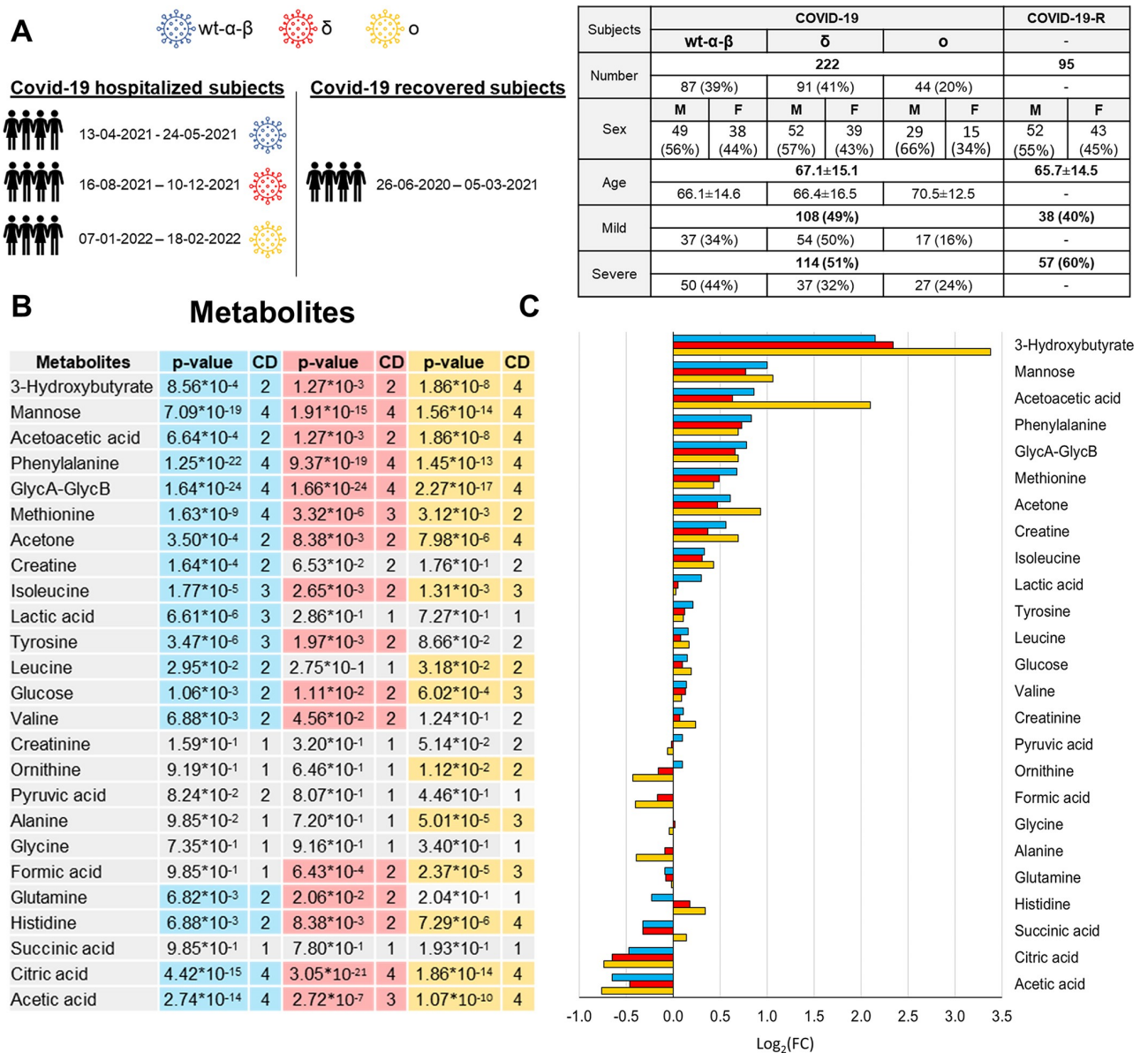


Fig 1. Metabolomic profiling of the main COVID-19 variants. (A) Main demographic and clinical characteristics of the enrolled subjects (right panel); for each group of subjects, the sample collection period was reported (left panel). (B) List of the metabolites quantified in plasma samples. The p-values and Cliff's Delta effects size are reported for the comparison between each of the three groups of variants with respect to the COVID-19-R group; p-values <0.05 are highlighted. (C) Values of Log₂ fold change (FC) of quantified metabolites. Positive/negative values have higher/lower concentration in plasma samples from each group of variants with respect to the COVID-19-R group. Colour coding: wt- α - β group (cyan); δ group (red); o group (orange).

<https://doi.org/10.1371/journal.ppat.1011787.g001>

swab). With the aim of characterizing if the vaccine introduces some important alterations in the metabolomic/lipoproteomic fingerprint of the disease, we compared these samples (hereafter named "VAX group") with the samples of 80 COVID-19 positive patients unvaccinated collected in the same period ("NO-VAX group"). The demographic and clinical data of these groups are reported in Fig 3A.

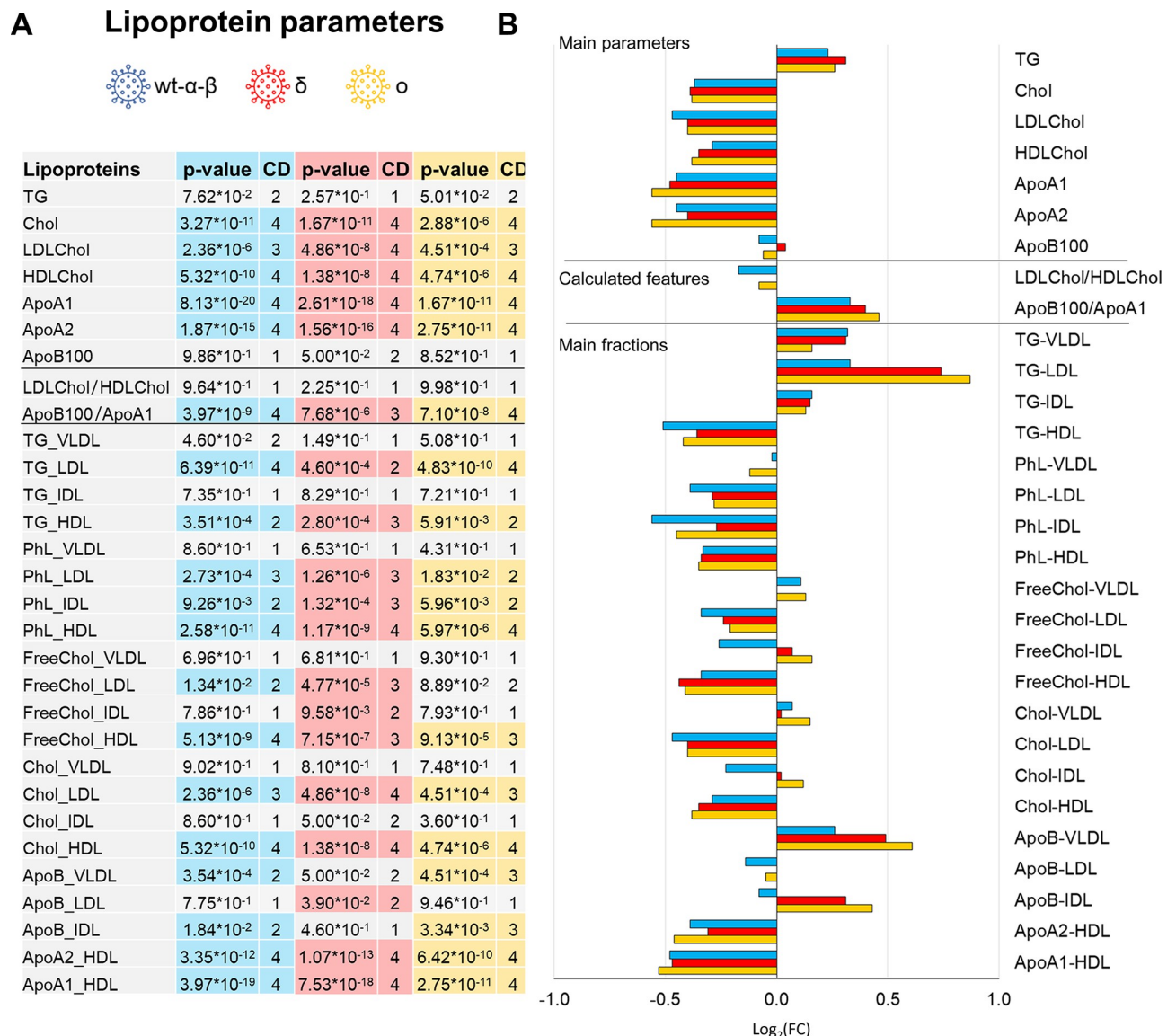


Fig 2. Lipoproteomic profiling of the main COVID-19 variants. A) List of lipoprotein parameters (main parameters, calculated features and main fractions) quantified in plasma samples. The p-values and Cliff's Delta effects size are reported for the comparison between each of the three groups of variants with respect to the COVID-19-R group; p-values < 0.05 are highlighted. B) Values of Log₂ fold change (FC) of quantified lipoprotein parameters (main parameters, calculated features and main fractions). Positive/negative values have higher/lower concentration in plasma samples from each group of variants with respect to COVID-19-R group. Colour coding: wt-α-β group (cyan); δ group (red); o group (orange).

<https://doi.org/10.1371/journal.ppat.1011787.g002>

Neither a significant clustering between VAX and NO-VAX groups is observed by multi-variate statistics (PCA and RF, [S3 Fig](#)) nor significantly different levels of metabolites or lipoproteins are found between the two groups. The two groups show the same metabolic and lipoproteomic alterations when compared to the COVID-19-R group ([Fig 3B–3C](#)). Only succinate shows an opposite trend, being significantly increased when considering the VAX group vs. the recovered subjects and decreased (but not significantly, p-value > 0.05) for the comparison between the NO-VAX group vs. the recovered subjects. Lipoprotein sub-fractions and particle numbers are reported in [S4 Fig](#).

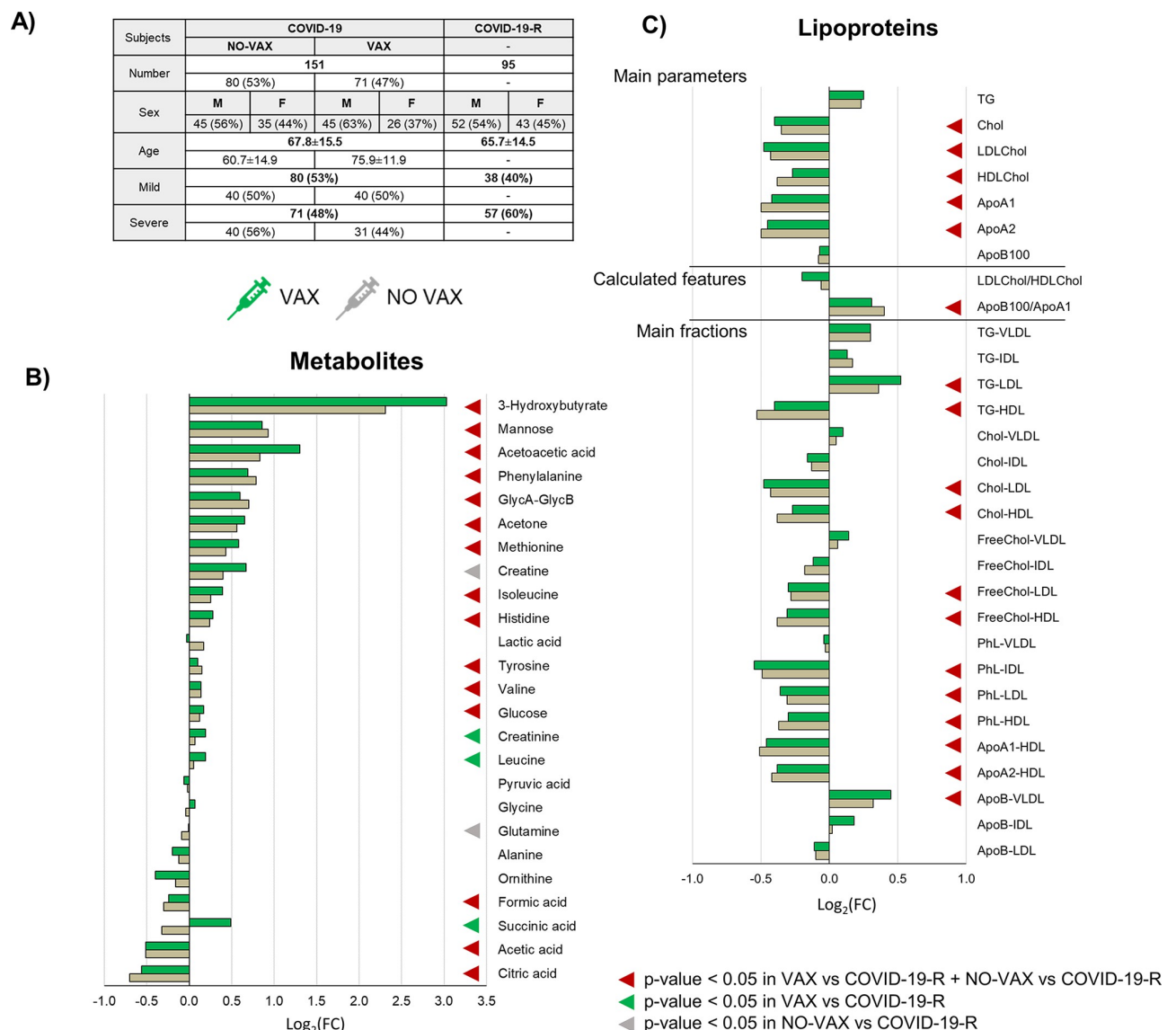


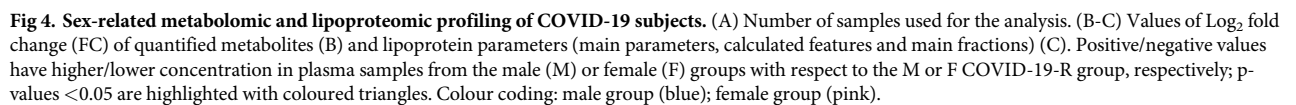
Fig 3. Metabolomic and lipoproteomic profiling of the VAX and NO-VAX groups. (A) Main demographic and clinical characteristics of the subjects. (B-C) Values of Log₂ fold change (FC) of quantified metabolites (B) and lipoprotein parameters (main parameters, calculated features and main fractions) (C). Positive/negative values have higher/lower concentration in plasma samples from the VAX or NO-VAX groups with respect to COVID-19-R group; p-values <0.05 are highlighted with coloured triangles. Colour coding: VAX group (green); NO-VAX (grey).

<https://doi.org/10.1371/journal.ppat.1011787.g003>

Sex-specific differences in COVID-19 disease

Sex differences in the disease fingerprint are hereafter evaluated. Since our data did not show significant differences in the metabolomic and lipoproteomic fingerprints associated with the main variants of the virus or with the vaccination status, all the 510 samples (274 from the previous study plus 236 newly enrolled) collected from COVID-19 patients are analyzed together, [S1 Table](#).

The presence of sex-specific differences in the plasma metabolomic and lipoproteomic profiles of healthy subjects has been characterized in detail [32]. Thus, to extract only COVID-19



related differences, 287 male and 223 female COVID-19 patients are compared to 52 male and 43 female recovered subjects, respectively, [Fig 4A](#).

Again, the strong and solid metabolic fingerprint induced by the infection can be detected in both male and female COVID-19 subjects, [Fig 4B–4C](#). No significant differences in the dysregulation trends of either metabolites or lipoprotein parameters emerge between the two groups, with the only exceptions of formic acid and Gln that are slightly but significantly decreased only for male patients. Lipoprotein sub-fractions and particle numbers are reported in [S5 Fig](#).

Global metabolomic/lipoproteomic fingerprint of COVID-19 disease and prognostic biomarkers

Since the previously tested potential confounding factors (variants, vaccination status, sex) did not show any relevant influences on the metabolomic/lipoproteomic fingerprint of the disease, we were prompted to perform a global and comprehensive characterization of the metabolic dysregulation induced by the viral infection. Thus, all the 510 samples collected from the positive subjects are compared to the 95 COVID-19- recovered subjects ([S1 Table](#)). In parallel, the impact of the disease severity was also evaluated. To this purpose all patients were classified as mild or severe, according to the respiratory symptoms manifested in the acute phase of the infection (see [Methods](#)).

Importantly, the RF model built on all samples shows a significant differential clustering between the whole COVID-19 group and the COVID-19-R group with 93.4% accuracy, 94.3% sensitivity, and 88.4% specificity ([Fig 5A](#)). This comparison highlights 17 significantly dysregulated metabolites ([Fig 5B–5C](#)) and 77 lipoprotein parameters ([Fig 6](#)). The levels of 3-hydroxybutyrate, mannose, acetoacetic acid, creatine, Phe, acetone, GlycA-GlycB, Met, Ile, glucose, lactic acid, Val, Tyr, Leu are increased in COVID-19 subjects with respect to the COVID-19-R group; the levels of Gln, citric acid and acetic acid are, instead, decreased. Among them, the concentrations of 8 metabolites result significantly increased/decreased in the comparison between mild and severe COVID-19 subjects ([Fig 5B–5C](#)). This trend is particularly evident for mannose, phenylalanine, GlycA-GlycB and citric acid ([S6A Fig](#)). Regarding the lipoprotein panel ([Figs 6 and S6B](#)), for the main parameters the COVID-19 group is characterized by a significant increment of total TG and a decrement of total Chol, HDL-Chol, LDL-Chol, as well as of ApoA1 and ApoA2. Looking at the level of the main- and sub-fraction composition, the infection leads to: *i*) a decrement of all the components associated with HDL (TG, Ph, Chol, Free Chol, ApoA1 and ApoA2), with HDL3 and, above all, HDL4, as the most affected subfractions; *ii*) an increment of TG-LDL (all subfractions) and a decrement of Chol-, free Chol- and Ph- in all LDL subfractions but LDL2 and LDL6 (LDL4 is the most affected sub-fraction). *iii*) an increment of TG-VLDL (all subfractions) and of ApoB-VLDL. As in the case of metabolites, several parameters show clear trends according to the disease severity; this is shown in [S6 Fig](#), which reports the metabolites and lipoprotein parameters (main parameters, calculated figures and main fractions) that have a p -value < 0.05 and a large Cliff's delta effect size in the comparison between the COVID-19 and COVID-19-R groups, and whose levels are also differentially altered between mild and severe patients.

The presence of clear trends in the concentration levels of some metabolites and lipoproteins according to the disease severity led us to consider the 40 severe subjects (28 wt- α - β , 6 δ , 5 o and 1 not attributable to a specific variant) with a fatal disease as a separate group, hereafter named “fatal group” ([S3 Table](#)). Interestingly, the levels of glucose, mannose, phenylalanine and of the three ketone bodies (3-hydroxybutyrate, acetone and acetoacetate) significantly increase going from the mild to the fatal group ([Figs 7 and S7](#)). Mild-to-fatal trends are also observed for the concentration changes of several parameters associated to the lipoprotein sub-fractions HDL4, LDL4, LDL5, LDL6 and VLDL5 ([Figs 7 and S7](#)). Among all of them, the fatal group shows significantly higher levels of all parameters associated to LDL5 and LDL6 but TG, with LDL5 parameters having the highest Cliff's delta values in the comparison between severe and fatal groups.

Comorbidities-dependent variations

Finally, the contribution of the patient's comorbidities on the dysregulated levels of metabolites and lipoproteins was also evaluated. [S1 Table](#) lists the number of subjects affected by the

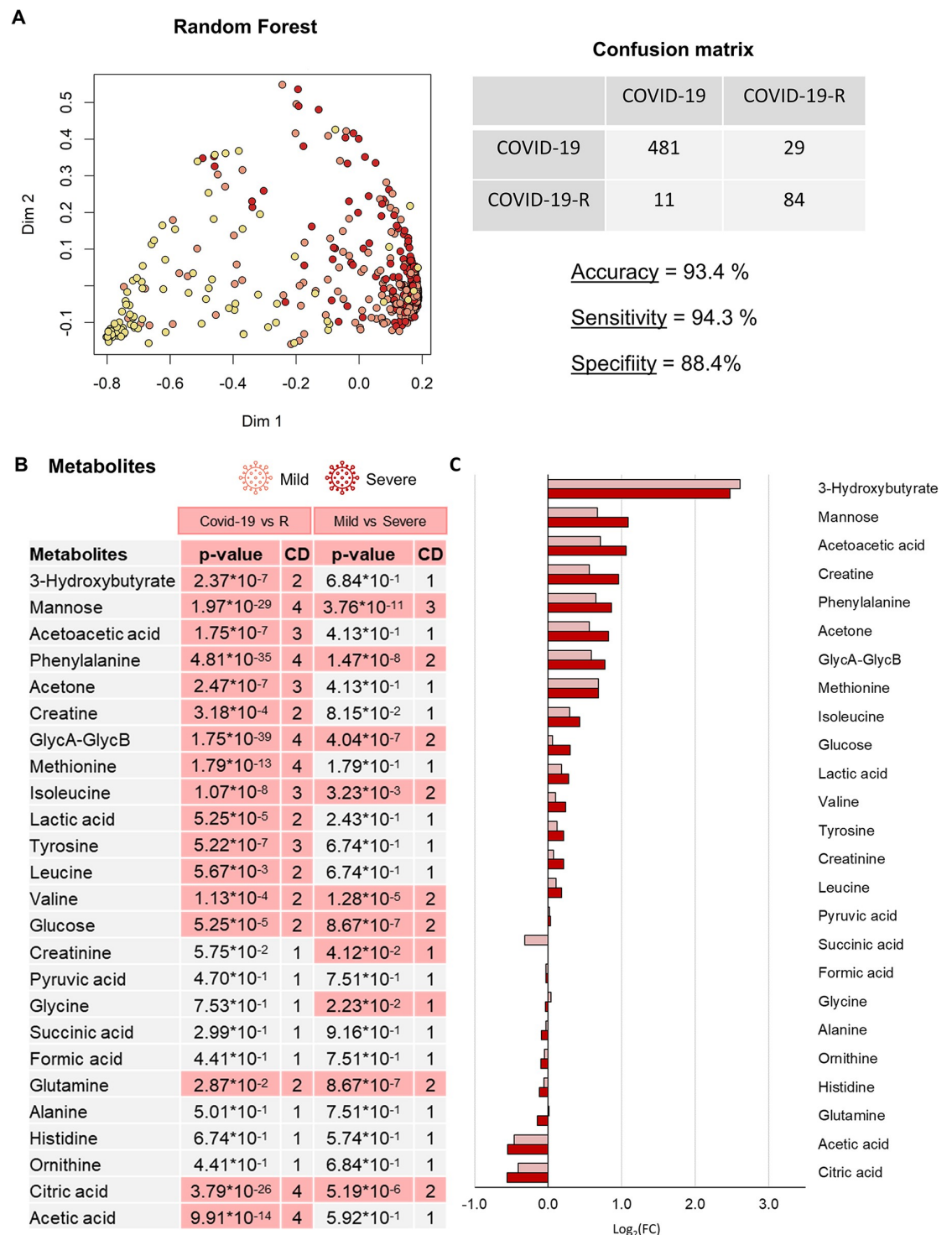


Fig 5. Metabolomic alterations in COVID-19 patients associated with clinical severity. (A) Proximity plots of the RF model discriminating COVID-19 patients (mild light red dots, severe red dots), and COVID-19-R subjects (yellow dots) using bucketed NOESY spectra. The confusion matrix with accuracy, specificity and sensitivity values. Notably, the misclassified patients all belong to the mild group. (B) List of the metabolites quantified in plasma samples. The p-values and Cliff's Delta effect-size are reported for the comparison between the COVID-19 group and the COVID-19-R group and for the comparison between mild and severe COVID-19 subjects; p-

values <0.05 are highlighted. (C) Values of Log_2 fold change (FC) of quantified metabolites. Positive/negative values have higher/lower concentration in plasma samples from mild or severe subjects with respect to the COVID-19-R group. Colour coding: mild (light red); severe (red).

<https://doi.org/10.1371/journal.ppat.1011787.g005>

comorbidities that can be considered as the main risk-factors of severe prognosis [6,7], i.e. asthma or chronic obstructive pulmonary disease (COPD), cardiovascular diseases (such as coronary artery disease (CAD), congestive heart failure (CHF), hypertension, type 2 diabetes (T2DM), dyslipidaemia, chronic kidney disease (CKD) and immune deficiency). Not unexpectedly, subjects affected by T2DM are characterized by higher glucose and mannose concentrations and those affected by CKD have high levels of creatinine (S8 Fig). Whereas creatinine was not found significantly different between the COVID-19 and COVID-19-R groups, high levels of mannose and glucose are very important markers of the COVID-19 signature and even for the discrimination between severe and fatal patients. Importantly, the observed differences in the levels of these two sugars are preserved when diabetic subjects are discarded from the analysis (S9 Fig). It is also important that the trend of increasing concentration for mannose and even more for glucose when going from mild to fatal is more evident for the T2DM subjects than in the no-T2DM patients (S9 Fig).

The follow-up cohort

For 154 out of 510 COVID-19 patients, a plasma sample was also collected during the follow-up visit, 2–6 months after the first negative swab (follow-up group S2 Table). Thus, for these patients two plasma samples were available, one at the onset of the acute infection (T1) and the second months after the negative swab (T2). They include 111 patients from the wt- α - β group (including 72 subjects with T1 analyzed already in the previous study), 33 from the δ group, 7 from the o group, and 3 not attributable to a specific variant. None of these subjects was diagnosed with long-COVID. The availability of two samples from the same individual allowed us to use a paired test. The same results emerge as from the above-described comparisons between the two independent groups of COVID-19 positive subjects and the COVID-19-R subjects. Consistently, no significant differences in metabolite and lipoprotein concentrations are detected when comparing the T2 samples with those from the COVID-19-R subjects. Indeed, the dysregulated molecules during the infection (i.e., at T1), are essentially reverted to within the normality range in samples collected at T2, S10 Fig. This behaviour is independent of the SARS-CoV-2 variant.

Genome-scale metabolic modelling

To investigate the impact of COVID-19 on the potential system-level functional metabolic shifts, we contextualized the most recent genome-scale metabolic model (GEM) of the human metabolism with the data obtained in this work. This approach represents a step forward with respect to the available literature data, because it is based on a multi-organ, sex-specific reconstruction. We initially focus on the male model analysis due to the stronger support provided by the higher number of fatal cases with respect to the female ones (27 vs. 13). Following the procedure illustrated in Methods, we generated healthy- (recovered), mild-, severe- and fatal-GEMs and used the underlying (predicted) flux distributions to understand the functional metabolism and metabolic pathways activity that define the severity of the disease. Using the metabolomic data and the most likely contribution of each organ (i.e. inputs and by-products) to the overall blood circulation, we contextualized the human metabolic network.

In analysing the enrichment/depletion of a given metabolite in the blood with respect to a specific organ, a key question arises: is the higher concentration of that specific metabolite in

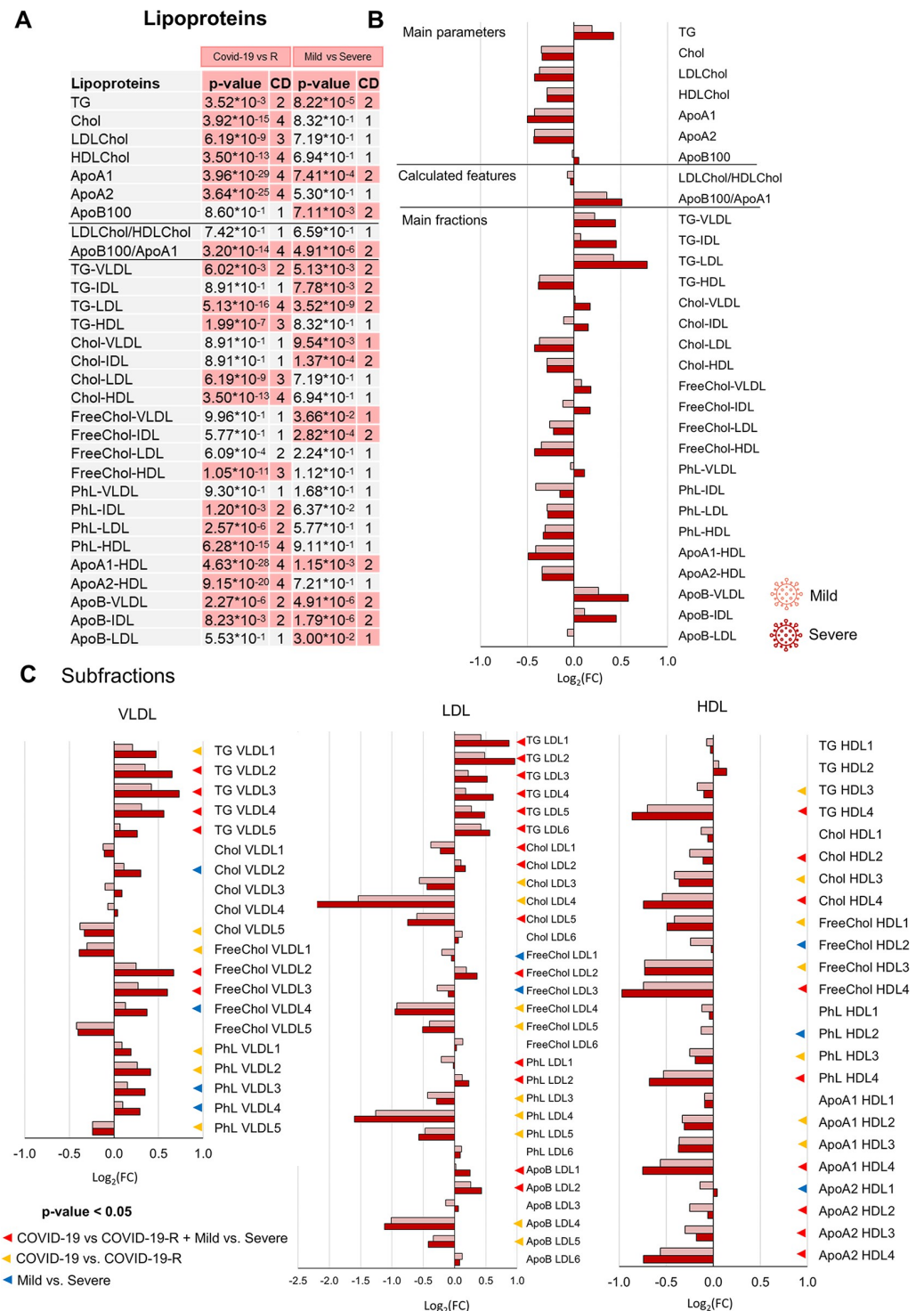


Fig 6. Lipoproteomic alterations in COVID-19 patients associated with clinical severity. (A) List of lipoprotein parameters (main parameters, calculated figures and main fractions) quantified in plasma samples. The p-values and Cliff's Delta effect size are reported for the comparison between the COVID-19 group with respect to the COVID-19-R group and for the comparison between mild and severe COVID-19 subjects; p-values < 0.05 are highlighted. (B-C) Values of Log₂ fold change (FC) of quantified lipoprotein main parameters, calculated features and main fractions (B) and lipoprotein subfractions (C). Positive/

negative values have higher/lower concentration in plasma samples from the COVID-19 group with respect to the COVID-19-R group. Colour coding: mild (light red); severe (red).

<https://doi.org/10.1371/journal.ppat.1011787.g006>

the plasma due to the reduced uptake by the corresponding organ or, conversely, it is due to an increased secretion by the same organ? A similar reasoning could be formulated in the case of lower metabolite concentration in the blood that could be the result of an increased uptake or a reduced secretion. We adopted the approach proposed by Dillard et al. (2022)[31] in

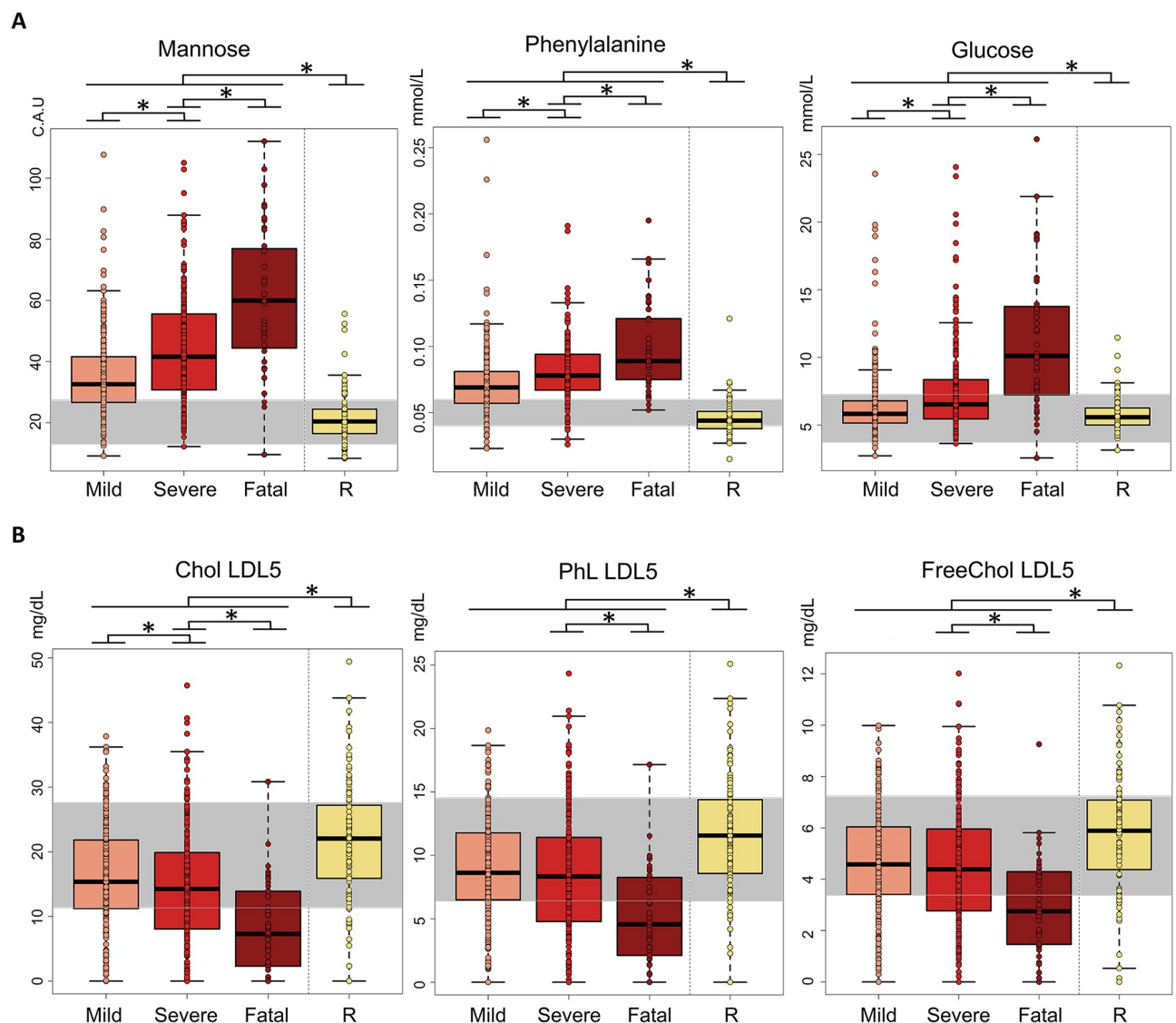


Fig 7. Markers of fatal events. Box plots of the concentration levels for (A) metabolites and (B) lipoprotein parameters (main parameters, calculated figures and main fractions) that have a p-value < 0.05 and a large Cliff's delta effect size in the comparison between mild, severe and fatal COVID-19 groups. The concentration levels in COVID-19-R subjects are also reported as control values. In each plot, the grey stripe embraces the concentration range in the reference "healthy" population. Colour coding: mild (light red); severe (red); fatal (dark red) COVID-19-R (yellow). * indicates p-value < 0.05: the upper line indicates the statistical significance between all the COVID-19 subjects and the COVID-19-R group; the lower lines indicate statistical significance between pairs of severity groups.

<https://doi.org/10.1371/journal.ppat.1011787.g007>

modelling COVID-19 with not cell-type specific Recon3D and according to which, in an active disease state, increased metabolite production is expected to be more plausible than reduced uptake. Thus, for example, they arbitrarily set to positive exchange bounds (thus simulating secretion) all the model metabolites identified as significantly higher in the severe COVID-19 disease state in order to simulate forced production.

In the first place, we asked whether i) healthy-, mild-, severe- and fatal-GEMs differed among each other and ii) metabolic dysregulation somehow followed the severity of the disease. **Fig 8A** shows that, at least in terms of active/inactive reactions (i.e. reactions carrying/not carrying flux) broad differences exist among the simulated metabolic states. The clustering, however, revealed a pattern that is compatible with a progressive dysregulation that matches the underlying severity of COVID-19 disease (healthy<mild<severe<fatal).

We then focused our attention on those pathways that showed an increase in their activity following the severity of the disease (**Figs 8B and S11**). More in detail, we selected those pathways whose reactions showed, on average, an increase in their flux values from the mild to the fatal states (in **Fig 8B** the healthy state is also included as an additional reference). Among those pathways exhibiting relatively high flux values across all the conditions (lower cluster of

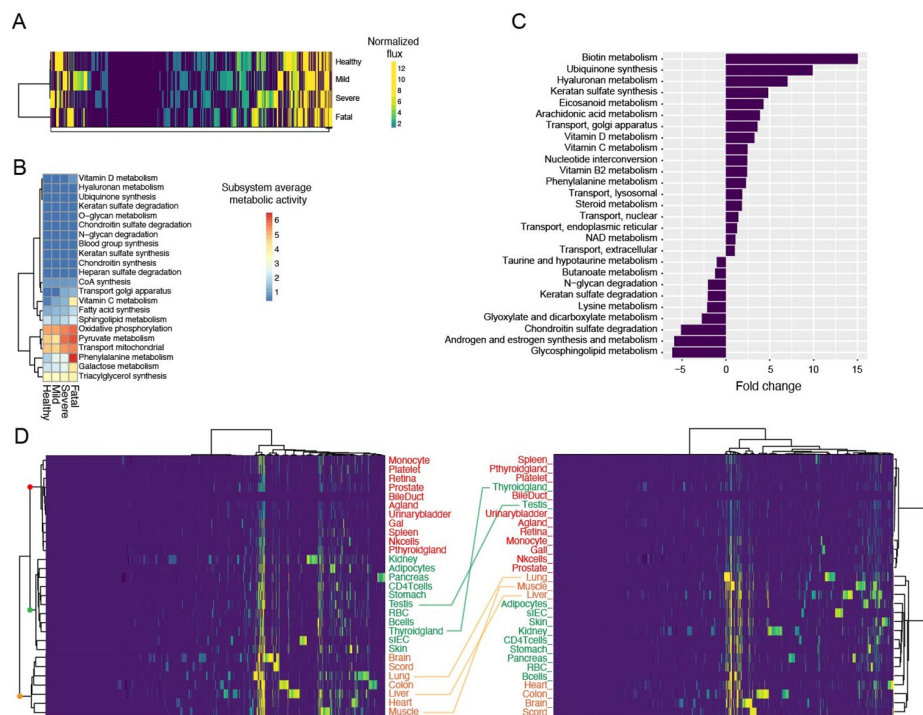


Fig 8. Genome-scale metabolic modelling of male subjects. A) Clustering of context-specific models. Each column represents a reaction in the human metabolic network and its color accounts for the (normalized) activity of that specific reaction in the corresponding context-specific model. Models are clustered according to their profile of active/inactive reactions. B) Heatmap that includes those metabolic processes (subsystems) showing an overall flux increase that paralleled that of disease severity (from mild to severe). Their activity in the healthy-model is also included in the figure for clarity. C) Fold change of average metabolic activity for each subsystem in the healthy vs. fatal condition. D) Heatmaps showing the clustering of the different organs according to the activity of each reaction (left healthy state, right fatal state). As in the panel A, each column represents a reaction in the human metabolic network and its color accounts for the (normalized) activity of that specific reaction in the corresponding organ. In the heatmap on the left, three main clusters are labelled with a red, green and orange dot and the organs belonging to each cluster are labelled accordingly and this color code is maintained in the heatmap on the right. Organs changing clusters in the two heatmaps are connected with a line.

<https://doi.org/10.1371/journal.ppat.1011787.g008>

the dendrogram of [Fig 8B](#)), a few pathways stood out from this analysis, i.e. oxidative phosphorylation and the metabolic pathways of phenylalanine, pyruvate, galactose and triacylglycerol.

The fact that fatal- and healthy-GEMs showed profoundly different metabolic states ([Fig 8A](#)) and that the activity of some metabolic pathways seems to correlate with COVID-19 severity ([Fig 8B](#)), prompted us to compare these two flux distributions and find those pathways (possessing at least five metabolic reactions) that differ in the two most extreme conditions (healthy vs. fatal [Fig 8C](#)). This analysis revealed the presence of 18 and 9 pathways that increase and decrease their activity in the fatal state with respect to the healthy one, respectively. The top three metabolisms that showed an increase in the fatal conditions with respect to the healthy one included biotin metabolism, ubiquinone synthesis and hyaluronan metabolism. On the contrary, the pathways that showed a marked decrease in their activity with respect to the healthy condition were glycosphingolipid metabolism, androgen and estrogen metabolism and chondroitin sulphate degradation.

The organ-specific structure of the human metabolic model adopted here offered the possibility to investigate the impact of COVID-19 on the metabolism of each organ represented therein. We thus clustered the different organs based on the activity of their metabolic reactions in the healthy and fatal states ([Fig 8D](#)). This analysis revealed a clear cut between three main clusters whose structure is overall conserved in the two extreme states. The first (red cluster in [Fig 8D](#)) embeds somehow peripheral organs, while the other two clusters (labelled with green and orange colours) include, among the others, all the vital organs (that are mostly included in the orange cluster, [Fig 8D](#)). Importantly, considering the healthy state (left heatmap of [Fig 8D](#)), the fluxes in the first clusters appear, on average, lower than in the second one, suggesting a lower metabolic activity in these organs with respect to the ones included in the second cluster. In the fatal metabolic state (right heatmap of [Fig 8D](#)) this situation appears to be more pronounced, with key organs showing an overall higher activity with respect to those embedded in the red cluster, compatibly with the disease state that characterized these patients. More in general, besides the conservation of the main clustering topology that subdivides the organs in three main clusters, we noticed a marked reshuffling of the relative order of each organ-specific metabolic network within these clusters, especially for some organs (thyroid gland, testis, lung, liver and muscle). From an analogous comparison between severe and fatal states ([S11 Fig](#)), we can establish that the main contribution to the metabolic dysregulation is associated to the progression from severe illness to the fatal state.

Also, we compared the sex-specific models obtained by integrating male and female models on the corresponding metabolic reconstructions. The overall structure of the models was maintained between sexes, both when considering the overall flux distribution ([Figs 8A and S12](#)) and the organ-level metabolic activity ([Figs 8D and S12](#)). Pathways increasing their activity concordantly with disease severity in both male and female models included N-glycan degradation, keratan sulphate synthesis, oxidative phosphorylation, and phenylalanine metabolism. We also found a pool of subsystems that showed a positive fold-change (i.e. higher metabolic activity) in the comparison between healthy and fatal states, both in male and female models. These were ubiquinone synthesis, vitamin D, C and B2 metabolism, and eicosanoid metabolism. Interestingly, we also found pathways that displayed sex-specific increase in their activity (e.g. arachidonic acid metabolism for male models and fatty acid oxidation, squalene and cholesterol synthesis in female models) or that showed opposite trends in the sex-specific simulations (e.g. androgen and estrogen synthesis and metabolism, decreasing in males and increasing in females).

Discussion

In this study we used an untargeted ^1H NMR approach to assess the metabolomic/lipoproteomic changes in the EDTA-plasma samples of a large Italian cohort of subjects (the COMETA cohort) with mild to fatal COVID-19. The advantages of using NMR with respect to the more sensitive mass spectrometry approach consist in being untargeted, highly reproducible, intrinsically quantitative and, in the case of blood derivative analyses, in allowing the automated profiling of a large ensemble of lipoprotein subfractions.

In a previous study from this consortium, 274 samples of the COMETA cohort were collected from COVID-19 patients hospitalized during the first wave of the pandemic (06/2020–04/2021), i.e. before a significant spread of the δ variant and before a massive vaccination campaign [18]. Here, samples from a much larger cohort (236 additional patients in the acute phase of the infection) and collected in a more expanded time frame (06/2020–06/2022) became available; they include 87 new patients from the wt- α - β group; 91 from the δ group; 44 from the o group, and 14 not attributable to a specific variant. Together, they form a cohort of 510 patients. The new analyses contribute to the advancement of the state of the art in three main aspects:

1. We demonstrate that the metabolomic/lipoproteomic signature of the disease is robust and does not change in a large and heterogeneous cohort of patients, sampled during the different waves of COVID-19 infection corresponding to the spread of different variants and having different vaccination status. Interestingly, most metabolites and lipoprotein parameters show the same trends of changes along all the variants. The larger increase in ketone bodies observed for the o variant can be attributed to the higher percentage of fatal events in this subgroup. Histidine, together with IDL-associated Chol and Apo B100 are the only molecules that increase in the δ and o groups but decrease in the wt- α - β group. No significant correlations (Pearson test) were found among these molecules. The decrement of histidine was already reported also in the other metabolomic studies based on samples collected before the spread of the δ variant [19,20,27]; to our knowledge no detailed information is yet available for the other variants. The signature of the disease is completely independent from the vaccination status of the subjects; indeed, infected vaccinated subjects show the same metabolic and lipoproteomic changes observed for non-vaccinated subjects (with the only notable exception of succinate, which needs to be further analyzed in future studies). Independence of the metabolomic profile from the vaccination status has recently been mentioned by other authors [29]. Notably, all the VAX subjects here considered had been vaccinated since more than 2 weeks, thus ensuring that the specific changes of the lipoproteome observed following vaccination had become negligible [33].
The metabolomic/lipidomic signature of the disease does not show any relevant difference between males and females, indicating that the strong disease signature overcomes any sex-specific signature. The only relevant variation between the two sexes concerns a decrease in formate and Gln in male patients; these two molecules are known to be important for the modulation of the immune response [34,35].
Lastly, we could also confirm in a set of 154 subjects sampled at 2 different time points that, after 2–6 months from negativization, in the absence of any clinical long-term symptoms, the dysregulated features all revert toward their normal value, indicating the substantial metabolic healing of COVID-19 clinically-recovered subjects.
2. The metabolomic and lipoproteomic changes induced by the SARS-CoV-2 infection that are described here are largely coherent with those obtained in our previous studies and also observed by other research groups worldwide [18–29]. The high reproducibility among all

the published data demonstrates that they accurately reflect pathophysiological changes associated to COVID-19. Additionally, here we can discriminate subjects who will progress towards a fatal outcome based on the levels of some molecules, that can be considered as prognostic markers. These subjects are characterized by the highest plasma levels of mannose and phenylalanine but also by very high levels of glucose and of the three ketone bodies, along with the largest dysregulation of several lipoprotein parameters. Notably, the highest fold changes for both glucose and mannose are observed for patients with T2DM, a pre-existing comorbidity identified as an important risk factor for COVID-19 associated mortality [7,36].

3. In the literature, the overall metabolic changes observed in COVID-19 have been tentatively explained in terms of a complex host-virus interaction aimed at the energy supply for both viral replication and host immunological response, leading to a dysregulation of carbohydrate and lipid metabolism and to mitochondrial dysfunction [18–28]. Also, metabolic modelling has shown to be a useful tool to systemically explore the metabolic impact of COVID-19 and to prioritize specific metabolic reactions in the search of novel treatment strategies [31,37–41]. The release in 2021 of the first multi-organ and sex-specific genome-scale human metabolic reconstruction offered us the possibility to i) more accurately constraint the model in a knowledge-based way; explore ii) the metabolic impact of COVID-19 on the entire metabolism and iii) the sex-dependent metabolic response to the disease. Based on NMR metabolomic data, our reconstruction shows that, globally, the human (male and female) metabolic network is strongly impacted by the disease, and that the extent of changes in metabolic configuration follows the disease severity (Fig 8A). Also, we found a marked metabolic reprogramming at the level of many organs embedded in our reconstruction. Indeed, despite the three main clusters that were conserved in the two cases (red, green, and orange clusters in Fig 8D) contained the same organs, i) the relative order of this clustering is different in the two conditions (red and green clusters clustering together in the healthy condition but not in the fatal one) and ii) the relative position of the organs inside the same cluster differ significantly in the comparison between healthy and fatal heatmap. Functionally speaking, a few pathways occurred multiple times, both among those showing an overall activity trend that mirrored disease severity, and among those that showed the highest fold change in the healthy vs. fatal comparison. These included oxidative phosphorylation and the metabolic pathways of phenylalanine, pyruvate, galactose, and triacylglycerol. Remarkably, i) these pathways include the possible disease markers experimentally identified by our NMR approach (e.g. phenylalanine and mannose) and ii) a large body of literature identified perturbations of the metabolites belonging to these pathways as possible metabolic responses to the COVID-19 disease state [42–46]. Our data also suggest the presence of system-level differences in the metabolic network of males and females in response to COVID-19 disease. A few pathways, whose changes were consistent between sexes (ubiquinone synthesis, vitamin D, C and B2 metabolism and eicosanoid metabolism), indicate an increase in the generic energetic demand of the organism following the severity of the disease (for both viral replication and host immunological response). Many other pathways showed marked sex-related differences. This is the case for example of the metabolic activities of reproductive steroid hormones that, in our sex-specific simulations, showed opposite trends (decreasing in males and increasing in females, Figs 8C and S12, respectively). Importantly, the immunomodulatory effects of estrogens in different viral infections are largely known [47], as well as their involvement in the response to COVID-19 [48,49]. Similarly, phenylalanine metabolism (despite following the same trend of disease severity in both males and females) was shown to be more active in men with respect to

women in healthy vs. fatal contrast (**Figs 8C and S12C**). This confirms previous findings on the sex dependency of this metabolic pathway during COVID-19 disease and on its possible connection with increased inflammatory status [50,51]. On the one hand, these findings highlight the importance of these specific metabolic pathways for monitoring disease progression, and, on the other hand, they also promote the use of systems biology tools for identifying non-trivial consequences of metabolic changes at the whole-body level.

Materials and methods

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki; it was approved by Comitato Etico Regionale per la Sperimentazione Clinica della Toscana—sezione Area Vasta Centro, code “18436_bio”. Written informed consent for inclusion was obtained from each subject before enrolment in the study.

Patients’ recruitment and sample collection

All the subjects included in this study were enrolled in the framework of the COMETA project, funded by the Tuscany Region, Italy. A total of 605 subjects (**S1 Table**) were recruited at the Santa Maria Nuova hospital of the Azienda USL Toscana Centro, in Florence (Italy) in the period between 20/06/2020 and 17/06/2022, i.e. during the first three waves of the COVID-19 pandemic in Italy. They include two different groups of subjects, namely:

- i. 510 COVID-19 hospitalized patients, that resulted positive for SARS-CoV-2 infection (with molecular nasopharyngeal swab), infected with different variants of the virus and with various disease severity; for all them a blood sample was collected during the acute phase of the infection. Additionally, for 154 out of 510 COVID-19 patients, we collected a second plasma sample during a follow-up visit (2–6 months after the first negative swab). This group is named as the “follow-up group”, **S2 Table**.
- ii. 95 COVID-19 recovered subjects (COVID-19-R group), previously hospitalized for SARS-CoV-2 infection (all during the first wave of pandemic); the blood of these subjects, who do not show any symptoms of persistent illness and could be considered as fully recovered, was collected 2–6 months after test negativization, during the follow-up visit (for this group we do not have the samples at the moment of the infection).

COVID-19 patients were classified as mild or severe according to the respiratory symptoms in the acute phase of the infection. All the subjects not requiring treatment with oxygen (or not requiring supplemental oxygen with respect to the treatment in progress before infection) or requiring oxygen treatment mask (Ventimask, VM) or nasal prongs with $\text{FiO}_2 \leq 40\%$ were classified as mild; patients requiring non-invasive ventilation (NIV) or MV with high $\text{FiO}_2 > 40\%$ or requiring orotracheal intubation (OTI) were classified as severe. Blood withdrawal occurred before NIV / VM ($\text{FiO}_2 > 40\%$) or OTI. Forty of these patients had a fatal outcome (**S3 Table**).

For COVID-19 recovered subjects, blood samples have been collected at the moment of admission to the hospital, when patients were swabbed regardless of the cause of hospitalization, and before the start of any therapy or other medical interventions (oxygen mask, ventilation, parenteral nutrition).

All plasma samples were collected, processed and stored according to ISO standards (ISO 23118: 2021), designed for high quality biological samples for metabolomic analysis [52–54].

The metabolomic analyses were performed on EDTA-plasma samples. Before analysis, the EDTA-plasma samples were stored at -80°C in the repository of the da Vinci European Biobank, which offered a conservation service (daVEB, DOI: [10.5334/ojb.af](https://doi.org/10.5334/ojb.af), <https://www.unifi.it/vp-11370-da-vinci-european-biobank.html>, Italy).

NMR analysis and spectral processing

NMR samples were prepared and recorded according to standard procedures for serum/plasma samples for metabolomics analysis [14,15].

NMR spectra for all the samples were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz of Larmor proton frequency and equipped with a PATXI $^1\text{H} - ^{13}\text{C} - ^{15}\text{N}$ and ^2H decoupling probe including a z-axis gradient coil, automatic tuning-matching (ATM) and an automatic, refrigerated sample changer (SampleJet, Bruker BioSpin). A BTO 2000 thermocouple served to stabilize the temperature to a level of approximately 0.1 K in the sample. Before measurement, the samples were kept for 5 minutes inside the NMR probe head, for temperature equilibration at 310 K. For each sample, three one-dimensional ^1H NMR spectra were acquired with water peak suppression and different pulse sequences [14]: i) standard NOESY 1Dpresat ii) standard 1D CPMG iii) standard 1D diffusion-edited. The parameters of each experiment are reported in [S4 Table](#). Free induction decays were multiplied by an exponential function equivalent to a 0.3 Hz line-broadening factor before applying Fourier transform. Transformed spectra were automatically corrected for phase and baseline distortions and calibrated at the glucose doublet at δ 5.24 ppm using TopSpin 4.1 (Bruker BioSpin).

Statistical analysis

All the statistical analyses were performed using the R software (R. 3.0.2).

The multivariate analyses were applied on NOESY binned spectra. To this aim, each spectrum was segmented into 0.02 ppm chemical shift bins, from 10.00 to 0.2 ppm, with the exclusion of EDTA resonances (regions: 2.53–2.60, 2.68–2.73, 3.07–3.24, 3.58–3.64 ppm) and water signal (region: 4.40–5.00 ppm); the corresponding spectral areas were integrated (Assure-NMR software, Bruker BioSpin). Unsupervised Principal Component Analysis (PCA) was used as an exploratory analysis to obtain a preliminary outlook of the data (visualization in a reduced space, presence of clusters or outliers).

Different types of multivariate statistics (i.e. Random Forest (RF) and Orthogonal Projections to Latent Structures-Discriminant Analysis (OPLS-DA)), were tested for supervised classification. Despite, the obtained models were essentially independent of the multivariate method used, here we reported the results obtained by RF algorithm because it allows stratified samplings, to ensure equal representation in the case of unbalanced groups [14].

The accuracy, sensitivity, and specificity of all calculated models were assessed according to the standard definitions.

Twenty-five metabolites, were assigned in all the spectra and their concentrations analysed. The assignment procedure was performed using an ^1H NMR spectra library of pure organic compounds (BBIORFCODE, Bruker BioSpin), public databases, e.g. the Human Metabolome Database (<https://hmdb.ca/>), and stored reference ^1H NMR spectra of metabolites. Metabolites were analysed using the In Vitro Diagnostics research (IVDr) B.I.-Quant PS tool (Bruker, BioSpin). For metabolites that are not present in the IVDr list, the respective areas were integrated using a R script developed in-house. The IVDr Lipoprotein Subclass Analysis B.I.-LISA tool (Bruker, BioSpin) was used to extract one hundred fourteen parameters associated to lipoproteins (main parameters, calculated features, main fractions, sub-fractions and particle numbers)

[55]. As done in a previous work from the COMETA consortium [18], a reference population of EDTA-plasma samples from 177 (86 males and 91 females) healthy subjects was used to calculate the “healthy” deviation range (mean $\pm$ SD) for each metabolite and lipoprotein.

The non-parametric Wilcoxon-Mann-Whitney was used to determine the significantly different parameters between the different groups of subjects. The obtained p-values were adjusted for multiple tests using False Discovery Rate Correction (FDR) according to the Benjamini-Hochberg method; an adjusted p-value <0.05 was considered statistically significant. Effect size (Ef) was also calculated; the magnitude is assessed using the thresholds provided in Romano et al. [56], that is $|Ef| < 0.147$ “negligible—1”, $|Ef| < 0.33$ “small—2”, $|Ef| < 0.474$ “medium—3”, otherwise “large—4”.

Genome-scale metabolic modelling

Metabolomics data were integrated with the most recent human GEM reconstruction, which is based on the integrated metabolism of 26 main organs of the human body and 6 blood cells types and provides sex-specific whole-body metabolism; in practice, two reconstructions are currently available “Harvey” and “Harvetta” accounting for the male and female metabolic models, respectively [30]. Here we apply this reconstruction to interpret the metabolic changes observed in plasma as a function of COVID-19 severity. We took advantage of the fact that in the updated human reconstruction the values of the exchange reactions (i.e. those reactions that account for the uptake/release of the nutrients by each organ) are the result of extensive literature search and accurately define the metabolic exchanges of each organ with the overall blood circulation. Thus, the direction of the exchange reactions (uptake or release) were set according to Thiele et al. (2020) [30]. The extent of the exchanges (i.e. the rates) were instead defined according to the metabolomic data obtained in this work. More in detail, we post-processed the metabolomic dataset as follows. The metabolomic data were first divided between male and female patients in order to be able to integrate each of the two datasets in the corresponding sex-specific metabolic reconstruction and then divided on the basis of the severity condition (recovered, mild, severe and fatal); as for the rest of this study and in accordance to previous findings [18], the recovered subjects are taken as healthy controls. Then, for each metabolite/lipoprotein, the mean among each sample and the fold changes (FC) among each of the following contrasts was computed: recovered vs. mild, recovered vs. severe, recovered vs. fatal. This latter step was performed using the *foldchange* function of the R package *gtool*. The Wilcoxon test was then used to assess the significance of the changes among the different conditions and the R function *p.adjust* was used to compute the False Discovery Rate (FDR) values. The FC of the plasma metabolites/lipoproteins that resulted to be statistically different in the examined conditions were used to constraint the model as described below.

Using COBRApy [57] (version 0.26) and selecting the Gurobi solver (version 9), we used Flux Balance Analysis (FBA) to create a standard “healthy” flux distribution by leaving the model unconstrained, simulating a Mediterranean diet and optimizing for the biomass objective function of the model. The other flux distributions (mild, severe and fatal) were computed modifying the upper and lower boundaries (UB and LB, respectively) using plasma metabolites FCs. More in details, for each i_{th} exchange metabolites, we set the UB and LB as follows:

$$UB > 0 \wedge LB < 0 \left\{ \begin{array}{ll} FC > 0; & x_{ub} = BV \cdot FC \\ & x_{lb} = BV \cdot 1/FC \\ FC < 0; & x_{ub} = BV \cdot 1/|FC| \\ & x_{lb} = BV \cdot |FC| \end{array} \right.$$

$$UB \wedge LB < 0 \begin{cases} FC > 0; x_{ub} \wedge x_{lb} = BV \cdot 1/FC \\ FC < 0; x_{ub} \wedge x_{lb} = BV \cdot |FC| \end{cases}$$

$$UB \wedge LB > 0 \begin{cases} FC > 0; x_{ub} \wedge x_{lb} = BV \cdot FC \\ FC < 0; x_{ub} \wedge x_{lb} = BV \cdot 1/|FC| \end{cases}$$

Where UB_i and LB_i represents the upper and lower boundaries for the i_{th} metabolite of all the exchange metabolites in the reconstruction, FC represents the fold change of the i_{th} metabolite in a specific contrast, x_{ub} and x_{lb} represent the actual upper and lower values of the constrained reaction in the model after metabolomic data integration, BV represents the default boundary value in the original model [30], respectively. From the four hypothetical flux distributions that we obtained (healthy, mild, severe and fatal) we pruned those reactions that display no flux in any of the four conditions. We used the remaining reactions to investigate the metabolic changes at the whole-body level that better account for the measured metabolic changes. To efficiently summarize our data, we took advantage of the model-embedded association between each reaction and its corresponding subsystem (i.e. cellular process) and computed a subsystem average metabolic activity (S.A.M.A.) as follows:

$$S.A.M.A. = \frac{\sum_{i=1}^N |f_i|}{N}$$

Where f_i represents the function of the i_{th} of the N reactions included in a given subsystem.

Similarly to what described above for each metabolite in the metabolomic dataset, FC and p -values were computed for each subsystem in the contrasts recovered-mild, recovered-severe and recovered-fatal. Heatmaps were reconstructed using *pheatmap*, and selecting the Ward's (*ward.D*) clustering method.

Supporting information

S1 Table. Demographic and clinical characteristics of the COMETA cohort. List of abbreviations: M: Males; F: Females; COPD: Chronic Obstructive Pulmonary Disease; CAD: Coronary Artery Disease; CHF: Congestive Heart Failure; T2DM: Type 2 Diabetes; CKD: Chronic Kidney Disease.
(XLSX)

S2 Table. Demographic and clinical characteristics of the follow-up group. List of abbreviations: M: Males; F: Females; COPD: Chronic Obstructive Pulmonary Disease; CAD: Coronary Artery Disease; CHF: Congestive Heart Failure; T2DM: Type 2 Diabetes; CKD: Chronic Kidney Disease.
(XLSX)

S3 Table. Demographic and clinical characteristics of the COVID-19 patients with a fatal outcome. List of abbreviations: M: Males; F: Females; COPD: Chronic Obstructive Pulmonary Disease; CAD: Coronary Artery Disease; CHF: Congestive Heart Failure; T2DM: Type 2 Diabetes; CKD: Chronic Kidney Disease.
(XLSX)

S4 Table. Parameters of the NMR experiments acquired for each plasma-EDTA samples at 600 MHz spectrometers.

(XLSX)

S1 Fig. Multivariate analysis of the main COVID-19 variants. A) PCA Score Plot based on bucketed NOESY spectra of the main three COVID-19 variant groups and COVID-19-R group. B) Proximity plots of the RF model discriminating the COVID-19 variant groups and COVID-19-R subjects using bucketed NOESY spectra. The confusion matrix and the accuracy value are reported. Colour coding: wt- α - β group (cyan); δ group (red); o group (orange); COVID-19-R (yellow).

(TIF)

S2 Fig. Lipoproteomic alterations (subfractions and particle numbers) in the main COVID-19 variants. Values of Log₂ fold change (FC) of quantified lipoprotein parameters (particle numbers and HDL, LDL, and VLDL subfractions). Positive/negative values have higher/lower concentration in plasma samples from each of the three variant COVID-19 groups with respect to COVID-19-R group. p-values <0.05 are highlighted with colored squares. Colour coding: wt- α - β group (cyan); δ group (red); o group (yellow).

(TIF)

S3 Fig. Multivariate analysis of the VAX and NO-VAX groups. A) PCA Score Plot based on bucketed NOESY spectra. B) Proximity plots of the RF model discriminating the COVID-19 VAX group and the COVID-19 NO-VAX group using bucketed NOESY spectra. The confusion matrix and the accuracy value are reported. Colour coding: VAX group (green); NO-VAX group (grey).

(TIF)

S4 Fig. Lipoproteomic alterations (subfractions and particle numbers) in the VAX and NO-VAX groups. Values of Log₂ fold change (FC) of quantified lipoprotein parameters (particle numbers and HDL, LDL, and VLDL subfractions). Positive/negative values have higher/lower concentration in plasma samples from the VAX or NO-VAX groups with respect to the COVID-19-R group; p-values <0.05 are highlighted with coloured triangles. Colour coding: VAX group (green); NO-VAX (grey).

(TIF)

S5 Fig. Sex-related lipoproteomic alterations (subfractions and particle numbers) of COVID-19 subjects. Values of Log₂ fold change (FC) of quantified lipoprotein parameters (particle numbers and HDL, LDL, and VLDL subfractions). Positive/negative values have higher/lower concentration in plasma samples from male (M) or female (F) groups with respect to M or F COVID-19-R subjects, respectively; p-values <0.05 are highlighted with coloured triangles. Colour coding: male group (blue); female group (pink).

(TIF)

S6 Fig. Markers of clinical severity. Box plots of the concentration levels for (A) metabolites and (B) lipoprotein parameters (main parameters, calculated figures and main fractions) that have a p-value < 0.05 and a large Cliff's delta effect size in the comparison between COVID-19 and COVID-19-R groups and whose levels are also significantly altered between mild and severe patients. In each plot, the grey stripe embraces the concentration range in the reference "healthy" population. Colour coding: mild (light red); severe (red); COVID-19-R (yellow).

*indicates p-value < 0.05: the upper line indicates the statistical significance between all the COVID-19 subjects and the COVID-19-R group; the lower line indicates statistical

significance between mild and severe subjects.
(TIF)

S7 Fig. Markers of fatal events. Box plots of (A) metabolites and (B) lipoproteins concentration levels in COVID-19 positive subjects grouped according to disease severity. The concentration levels in COVID-19-R subjects are also reported. Colour coding: mild (light red); severe (red); fatal (dark red); recovered subjects (yellow). * indicates p-value < 0.05: the upper line indicates the statistical significance between all the COVID-19 subjects and the COVID-19-R group; the lower lines indicate statistical significance between pairs of severity groups.
(TIF)

S8 Fig. Comorbidities-dependent variations. Box plots of Mannose, Glucose and Creatinine concentration levels in the 510 COVID-19 positive subjects grouped as a function of the main comorbidities. The concentration levels in COVID-19-R subjects are also reported as control values (yellow bar). In each plot, the grey stripe covers the concentration range in a "healthy" population. * indicates p-value < 0.05: the upper line indicates the statistical significance between all the COVID-19 subjects and the COVID-19-R group; the lower line indicates that at least one of the comorbidity groups (circled in red) is significantly different from all the others. List of abbreviations: COPD: Chronic Obstructive Pulmonary Disease; CAD: Coronary Artery Disease; CHF: Congestive Heart Failure; T2DM: Type 2 Diabetes; CKD: Chronic Kidney Disease.
(TIF)

S9 Fig. Comorbidities-dependent variations of severity markers. Box plots of (A) mannose and (B) glucose concentration levels in COVID-19 positive subjects grouped according to the grade of the disease severity. The concentration levels in COVID-19-R subjects are also reported. Left panels: all the subjects; middle panels: the T2DM subjects were excluded from the analysis; right panels: only T2DM subjects. In each plot the grey stripe covers the concentration range in a "healthy" population. Colour coding: mild (light red); severe (red); fatal (dark red); recovered subjects (yellow). * indicates p-value < 0.05: the upper line indicates the statistical significance between all the COVID-19 subjects and the COVID-19-R group; the lower lines indicate statistical significance between pairs of severity groups.
(TIF)

S10 Fig. Time-dependence of the severity markers in the follow-up group (S2 Table). Box plots of the concentration levels for (A) metabolites and (B) lipoproteins that have a p-value < 0.05 and a large Cliff's delta effect size in the comparison between the plasma samples collected at the moment of the acute infection (T1) and the samples collected at the follow-up visit (T2). The concentration levels in COVID-19-R subjects are also reported as control values. In each plot, the grey stripe embraces the concentration range in the reference "healthy" population. Colour coding: T1 (red); T2 (blue); COVID-19-R (yellow). * indicates p-value < 0.05: the upper line indicates the statistical significance between all the COVID-19 subjects and the COVID-19-R group; the lower line indicates statistical significance between T1 and T2.
(TIF)

S11 Fig. Comparison between severe and fatal states through genome-scale metabolic modelling. Heatmaps showing the clustering of the different organs according to the activity of each reaction (left severe, right fatal state). Each column represents a reaction in the human metabolic network and its color accounts for the (normalized) activity of that specific reaction in the corresponding organ. In the heatmap on the left, three main clusters are labelled with a red, green and orange dot and the organs belonging to each cluster are labelled accordingly and this color code is maintained in the heatmap on the right. Organs changing clusters in the

two heatmaps are connected with a line.
(TIF)

S12 Fig. Genome-scale metabolic modelling of female subjects. A) Clustering of context-specific models. Each column represents a reaction in the human metabolic network and its color accounts for the (normalized) activity of that specific reaction in the corresponding context-specific model. Models are clustered according to their profile of active/inactive reactions. B) Heatmap that includes those metabolic processes (subsystems) showing an overall flux increase that paralleled that of disease severity (from mild to severe). Their activity in the healthy-model is also included in the figure for clarity. C) Fold change of average metabolic activity for each subsystem in the healthy vs. fatal condition. D-E) Heatmaps showing the clustering of the different organs according to the activity of each reaction (D- healthy state, E- fatal state). As in in the panel A, each column represents a reaction in the human metabolic network and its color accounts for the (normalized) activity of that specific reaction in the corresponding organ.
(TIF)

Acknowledgments

P.T., V.G., V.P. and C.L. acknowledge their participation to the NMR International COVID-19 Research Network and the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

Author Contributions

Conceptualization: Paola Turano.

Formal analysis: Veronica Ghini, Walter Vieri, Valentina Pecchioli, Tania Alonso-Vásquez.

Funding acquisition: Giancarlo Landini, Paola Turano.

Investigation: Veronica Ghini, Tommaso Celli, Valentina Pecchioli.

Methodology: Veronica Ghini, Marco Fondi, Claudio Luchinat, Paola Turano.

Resources: Tommaso Celli, Nunzia Boccia, Lorenzo Pelagatti, Laura Bertini, Vieri Vannucchi.

Software: Marco Fondi.

Supervision: Marco Fondi, Claudio Luchinat, Laura Bertini, Vieri Vannucchi, Giancarlo Landini, Paola Turano.

Writing – original draft: Veronica Ghini, Marco Fondi, Paola Turano.

Writing – review & editing: Walter Vieri, Claudio Luchinat.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The Lancet*. 2020; 395: 470–473. [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9) PMID: 31986257
2. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021; 19: 141–154. <https://doi.org/10.1038/s41579-020-00459-7> PMID: 33024307
3. Ayres JS. A metabolic handbook for the COVID-19 pandemic. *Nat Metab*. 2020; 2: 572–585. <https://doi.org/10.1038/s42255-020-0237-2> PMID: 32694793
4. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun*. 2020; 109: 102433. <https://doi.org/10.1016/j.jaut.2020.102433> PMID: 32113704

5. Zaim S, Chong JH, Sankaranarayanan V, Harky A. COVID-19 and Multiorgan Response. *Curr Probl Cardiol.* 2020; 45: 100618. <https://doi.org/10.1016/j.cpcardiol.2020.100618> PMID: 32439197
6. Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, et al. Comorbidity and its Impact on Patients with COVID-19. *SN Compr Clin Med.* 2020; 2: 1069–1076. <https://doi.org/10.1007/s42399-020-00363-4> PMID: 32838147
7. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med.* 2020; 8: e21. [https://doi.org/10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8) PMID: 32171062
8. Proal AD, VanElzakker MB. Long COVID or Post-acute Sequelae of COVID-19 (PASC): An Overview of Biological Factors That May Contribute to Persistent Symptoms. *Frontiers in Microbiology.* 2021; 12: 1494. <https://doi.org/10.3389/fmicb.2021.698169> PMID: 34248921
9. Montaldo C, Messina F, Abbate I, Antonioli M, Bordoni V, Aiello A, et al. Multi-omics approach to COVID-19: a domain-based literature review. *Journal of Translational Medicine.* 2021; 19: 501. <https://doi.org/10.1186/s12967-021-03168-8> PMID: 34876157
10. Jia H, Liu C, Li D, Huang Q, Liu D, Zhang Y, et al. Metabolomic analyses reveal new stage-specific features of COVID-19. *European Respiratory Journal.* 2022;59. <https://doi.org/10.1183/13993003.00284-2021> PMID: 34289974
11. Danlos F-X, Grajeda-Iglesias C, Durand S, Sauvat A, Roumier M, Cantin D, et al. Metabolomic analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. *Cell Death Dis.* 2021; 12: 1–11. <https://doi.org/10.1038/s41419-021-03540-y> PMID: 33707411
12. Demichev V, Tober-Lau P, Lemke O, Nazarenko T, Thibeault C, Whitwell H, et al. A time-resolved proteomic and prognostic map of COVID-19. *Cell Systems.* 2021; 12: 780–794.e7. <https://doi.org/10.1016/j.cels.2021.05.005> PMID: 34139154
13. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell.* 2020; 182: 59–72.e15. <https://doi.org/10.1016/j.cell.2020.05.032> PMID: 32492406
14. Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, et al. High-Throughput Metabolomics by 1D NMR. *Angew Chem Int Ed.* 2019; 58: 968–994. <https://doi.org/10.1002/anie.201804736> PMID: 29999221
15. Takis PG, Ghini V, Tenori L, Turano P, Luchinat C. Uniqueness of the NMR approach to metabolomics. *TrAC Trends in Analytical Chemistry.* 2019; 120: 115300. <https://doi.org/10.1016/j.trac.2018.10.036>
16. Nagana Gowda GA, Raftery D. Recent Advances in NMR-Based Metabolomics. *Anal Chem.* 2017; 89: 490–510. <https://doi.org/10.1021/acs.analchem.6b04420> PMID: 28105846
17. Larive CK, Barding GA, Dinges MM. NMR Spectroscopy for Metabolomics and Metabolic Profiling. *Anal Chem.* 2015; 87: 133–146. <https://doi.org/10.1021/ac504075g> PMID: 25375201
18. Ghini V, Meoni G, Pelagatti L, Celli T, Veneziani F, Petrucci F, et al. Profiling metabolites and lipoproteins in COMETA, an Italian cohort of COVID-19 patients. *PLOS Pathogens.* 2022; 18: e1010443. <https://doi.org/10.1371/journal.ppat.1010443> PMID: 35446921
19. Meoni G, Ghini V, Maggi L, Vignoli A, Mazzoni A, Salvati L, et al. Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. *PLOS Pathogens.* 2021; 17: e1009243. <https://doi.org/10.1371/journal.ppat.1009243> PMID: 33524041
20. Bruzzzone C, Bizkarguenaga M, Gil-Redondo R, Diercks T, Arana E, García de Vicuña A, et al. SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience.* 2020; 23: 101645. <https://doi.org/10.1016/j.isci.2020.101645> PMID: 33043283
21. Kimhofer T, Lodge S, Whitley L, Gray N, Loo RL, Lawler NG, et al. Integrative Modeling of Quantitative Plasma Lipoprotein, Metabolic, and Amino Acid Data Reveals a Multiorgan Pathological Signature of SARS-CoV-2 Infection. *J Proteome Res.* 2020; 19: 4442–4454. <https://doi.org/10.1021/acs.jproteome.0c00519> PMID: 32806897
22. Lodge S, Nitschke P, Kimhofer T, Coudert JD, Begum S, Bong S-H, et al. NMR Spectroscopic Windows on the Systemic Effects of SARS-CoV-2 Infection on Plasma Lipoproteins and Metabolites in Relation to Circulating Cytokines. *J Proteome Res.* 2021; 20: 1382–1396. <https://doi.org/10.1021/acs.jproteome.0c00876> PMID: 33426894
23. Ballout RA, Kong H, Sampson M, Otvos JD, Cox AL, Agbor-Enoh S, et al. The NIH Lipo-COVID Study: A Pilot NMR Investigation of Lipoprotein Subfractions and Other Metabolites in Patients with Severe COVID-19. *Biomedicines.* 2021; 9: 1090. <https://doi.org/10.3390/biomedicines9091090> PMID: 34572275
24. Masuda R, Lodge S, Nitschke P, Spraul M, Schaefer H, Bong S-H, et al. Integrative Modeling of Plasma Metabolic and Lipoprotein Biomarkers of SARS-CoV-2 Infection in Spanish and Australian COVID-19

- Patient Cohorts. *J Proteome Res.* 2021; 20: 4139–4152. <https://doi.org/10.1021/acs.jproteome.1c00458> PMID: 34251833
25. Julkunen H, Cichońska A, Slagboom PE, Würtz P. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *eLife.* 2021; 10: e63033. <https://doi.org/10.7554/eLife.63033> PMID: 33942721
 26. Baranovicova E, Bobcakova A, Vysehradsky R, Dankova Z, Halasova E, Nosal V, et al. The Ability to Normalise Energy Metabolism in Advanced COVID-19 Disease Seems to Be One of the Key Factors Determining the Disease Progression—A Metabolomic NMR Study on Blood Plasma. *Applied Sciences.* 2021; 11: 4231. <https://doi.org/10.3390/app11094231>
 27. Bizkarguenaga M, Bruzzone C, Gil-Redondo R, SanJuan I, Martin-Ruiz I, Barriales D, et al. Uneven metabolic and lipidomic profiles in recovered COVID-19 patients as investigated by plasma NMR metabolomics. *NMR Biomed.* 2021; e4637. <https://doi.org/10.1002/nbm.4637> PMID: 34708437
 28. Correia BSB, Ferreira VG, Piagge PMFD, Almeida MB, Assunção NA, Raimundo JRS, et al. 1H qNMR-Based Metabolomics Discrimination of Covid-19 Severity. *J Proteome Res.* 2022; 21: 1640–1653. <https://doi.org/10.1021/acs.jproteome.1c00977> PMID: 35674498
 29. Gil-Redondo R, Conde R, Bizkarguenaga M, Bruzzone C, Lain A, González-Valle B, et al. An NMR-Based Model to Investigate the Metabolic Phenoreversion of COVID-19 Patients throughout a Longitudinal Study. *Metabolites.* 2022; 12: 1206. <https://doi.org/10.3390/metabo12121206> PMID: 36557244
 30. Thiele I, Sahoo S, Heinken A, Hertel J, Heirendt L, Aurich MK, et al. Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Molecular Systems Biology.* 2020; 16: e8982. <https://doi.org/10.15252/msb.20198982> PMID: 32463598
 31. Dillard LR, Wase N, Ramakrishnan G, Park JJ, Sherman NE, Carpenter R, et al. Leveraging metabolic modeling to identify functional metabolic alterations associated with COVID-19 disease severity. *Metabolomics.* 2022; 18: 51. <https://doi.org/10.1007/s11306-022-01904-9> PMID: 35819731
 32. Bell JA, Santos Ferreira DL, Fraser A, Soares ALG, Howe LD, Lawlor DA, et al. Sex differences in systemic metabolites at four life stages: cohort study with repeated metabolomics. *BMC Medicine.* 2021; 19: 58. <https://doi.org/10.1186/s12916-021-01929-2> PMID: 33622307
 33. Ghini V, Maggi L, Mazzoni A, Spinicci M, Zammarchi L, Bartoloni A, et al. Serum NMR Profiling Reveals Differential Alterations in the Lipoproteome Induced by Pfizer-BioNTech Vaccine in COVID-19 Recovered Subjects and Naïve Subjects. *Frontiers in Molecular Biosciences.* 2022; 9. <https://doi.org/10.3389/fmolb.2022.839809>
 34. Calder PC, Yaqoob P. Glutamine and the immune system. *Amino Acids.* 1999; 17: 227–241. <https://doi.org/10.1007/BF01366922> PMID: 10582122
 35. Pietzke M, Meiser J, Vazquez A. Formate metabolism in health and disease. *Molecular Metabolism.* 2020; 33: 23–37. <https://doi.org/10.1016/j.molmet.2019.05.012> PMID: 31402327
 36. Flemming A. Diabetes enhances viral loads in COVID-19. *Nat Rev Immunol.* 2022; 22: 410–410. <https://doi.org/10.1038/s41577-022-00748-7> PMID: 35688879
 37. Bannerman BP, Júlvez J, Oarga A, Blundell TL, Moreno P, Floto RA. Integrated human/SARS-CoV-2 metabolic models present novel treatment strategies against COVID-19. *Life Science Alliance.* 2021;4. <https://doi.org/10.26508/lsa.202000954> PMID: 34353886
 38. Yaneske E, Zampieri G, Bertoldi L, Benvenuto G, Angione C. Genome-scale metabolic modelling of SARS-CoV-2 in cancer cells reveals an increased shift to glycolytic energy production. *FEBS Letters.* 2021; 595: 2350–2365. <https://doi.org/10.1002/1873-3468.14180> PMID: 34409594
 39. Kishk A, Pacheco MP, Sauter T. DCcov: Repositioning of drugs and drug combinations for SARS-CoV-2 infected lung through constraint-based modeling. *iScience.* 2021; 24: 103331. <https://doi.org/10.1016/j.isci.2021.103331> PMID: 34723158
 40. Renz A, Widerspich L, Dräger A. Genome-Scale Metabolic Model of Infection with SARS-CoV-2 Mutants Confirms Guanylate Kinase as Robust Potential Antiviral Target. *Genes.* 2021; 12: 796. <https://doi.org/10.3390/genes12060796> PMID: 34073716
 41. Cheng K, Martin-Sancho L, Pal LR, Pu Y, Riva L, Yin X, et al. Genome-scale metabolic modeling reveals SARS-CoV-2-induced metabolic changes and antiviral targets. *Molecular Systems Biology.* 2021; 17: e10260. <https://doi.org/10.15252/msb.202110260> PMID: 34709707
 42. Krishnan S, Nordqvist H, Ambikan AT, Gupta S, Sperk M, Svensson-Akusjärvi S, et al. Metabolic Perturbation Associated With COVID-19 Disease Severity and SARS-CoV-2 Replication. *Molecular & Cellular Proteomics.* 2021; 20: 100159. <https://doi.org/10.1016/j.mcpro.2021.100159> PMID: 34619366
 43. Páez-Franco JC, Maravillas-Montero JL, Mejía-Domínguez NR, Torres-Ruiz J, Tamez-Torres KM, Pérez-Fragoso A, et al. Metabolomics analysis identifies glutamic acid and cystine imbalances in COVID-19 patients without comorbid conditions. Implications on redox homeostasis and COVID-19

- pathophysiology. *PLOS ONE*. 2022; 17: e0274910. <https://doi.org/10.1371/journal.pone.0274910> PMID: [36126080](#)
44. Filippas-Ntekouan S, Liberopoulos E, Elisaf M. Lipid testing in infectious diseases: possible role in diagnosis and prognosis. *Infection*. 2017; 45: 575–588. <https://doi.org/10.1007/s15010-017-1022-3> PMID: [28484991](#)
 45. Wei X, Zeng W, Su J, Wan H, Yu X, Cao X, et al. Hypolipidemia is associated with the severity of COVID-19. *J Clin Lipidol*. 2020; 14: 297–304. <https://doi.org/10.1016/j.jacl.2020.04.008> PMID: [32430154](#)
 46. Shao M-M, Shi M, Du J, Pei X-B, Gu B-B, Yi F-S. Metabolic Landscape of Bronchoalveolar Lavage Fluid in Coronavirus Disease 2019 at Single Cell Resolution. *Frontiers in Immunology*. 2022; 13. Available: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.829760> PMID: [35350779](#)
 47. Abramenko N, Vellieux F, Tesařová P, Kejlik Z, Kaplānek R, Lacina L, et al. Estrogen Receptor Modulators in Viral Infections Such as SARS-CoV-2: Therapeutic Consequences. *International Journal of Molecular Sciences*. 2021; 22: 6551. <https://doi.org/10.3390/ijms22126551> PMID: [34207220](#)
 48. Pinna G. Sex and COVID-19: A Protective Role for Reproductive Steroids. *Trends Endocrinol Metab*. 2021; 32: 3–6. <https://doi.org/10.1016/j.tem.2020.11.004> PMID: [33229187](#)
 49. Groban L, Wang H, Sun X, Ahmad S, Ferrario CM. Is Sex a Determinant of COVID-19 Infection? Truth or Myth? *Curr Hypertens Rep*. 2020; 22: 62. <https://doi.org/10.1007/s11906-020-01073-x> PMID: [32852624](#)
 50. Escarcega RD, Honarpisheh P, Colpo GD, Ahnstedt HW, Couture L, Juneja S, et al. Sex differences in global metabolomic profiles of COVID-19 patients. *Cell Death Dis*. 2022; 13: 1–12. <https://doi.org/10.1038/s41419-022-04861-2> PMID: [35568706](#)
 51. Luporini RL, Pott-Junior H, Di Medeiros Leal MCB, Castro A, Ferreira AG, Cominetti MR, et al. Phenylalanine and COVID-19: Tracking disease severity markers. *Int Immunopharmacol*. 2021; 101: 108313. <https://doi.org/10.1016/j.intimp.2021.108313> PMID: [34741868](#)
 52. Ghini V, Abuja PM, Polasek O, Kozera L, Laiho P, Anton G, et al. Metabolomic Fingerprints in Large Population Cohorts: Impact of Preanalytical Heterogeneity. *Clinical Chemistry*. 2021; 67: 1153–1155. <https://doi.org/10.1093/clinchem/hvab092> PMID: [34223627](#)
 53. Ghini V, Abuja PM, Polasek O, Kozera L, Laiho P, Anton G, et al. Impact of the pre-examination phase on multicenter metabolomic studies. *New Biotechnology*. 2022; 68: 37–47. <https://doi.org/10.1016/j.nbt.2022.01.006> PMID: [35066155](#)
 54. Ghini V, Quaglio D, Luchinat C, Turano P. NMR for sample quality assessment in metabolomics. *New Biotechnology*. 2019; 52: 25–34. <https://doi.org/10.1016/j.nbt.2019.04.004> PMID: [31022482](#)
 55. Jiménez B, Holmes E, Heude C, Tolson RF, Harvey N, Lodge SL, et al. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by ¹H NMR Spectroscopy in a Multilaboratory Trial. *Anal Chem*. 2018; 90: 11962–11971. <https://doi.org/10.1021/acs.analchem.8b02412> PMID: [30211542](#)
 56. Rosenthal R. Parametric measures of effect size. *The handbook of research synthesis*. New York, NY, US: Russell Sage Foundation; 1994. pp. 231–244.
 57. Ebrahim A, Lerman JA, Pålsson BO, Hyduke DR. COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*. 2013; 7: 74. <https://doi.org/10.1186/1752-0509-7-74> PMID: [23927696](#)

3.2.4 Dynamics of persister resuscitation

The study of resistance, tolerance, and persistence to antibiotics has traditionally focused on single microbial isolates. In recent years, there is growing recognition that interactions among microbes could significantly influence the outcomes of antibiotic treatments. In fact, there is evidence that suggests that these interactions are particularly important in shaping the emergence of persistence and tolerance within microbial populations. It has been described that persister cells, after reached a dormancy state and survived the antibiotic treatment, they may start a resuscitation process. In other words, after the antibiotic treatment, persister cells exit from the persistence state and start re-population (at a lower rate than before antibiotic treatment) (Mohiuddin et al. 2020). The processes of resuscitation and pathogen clearance are intricately tied to the dynamics of small populations size of the pathogen (persister) species. Several studies have suggested that even when the pathogen population decreases considerably, this might not be enough to eradicate these bacteria (Jones et al. 2006; Hara-Kudo et al. 2011; Gullberg et al. 2011). While the behavior of large populations is largely deterministic, small populations are subject to stochastic fluctuations, introducing unpredictability into their dynamics.

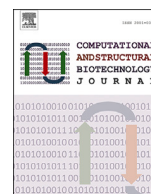
In this work, we explore how interactions within microbial communities affect the timing, mechanisms, and success of persister resuscitation, as well as how noise influences these processes. To investigate these dynamics, we developed a theoretical model of a three-member microbial community, focusing on how cell-to-cell interactions influence pathogen clearance. We use the generalized Lotka-Volterra framework and modeled it from a deterministic point of view, however, to examine the interplay between ecological relationships, population size, and the stochastic processes that dominate small populations, we describe the gLV model as a set of stochastic rules.

Briefly, the stochastic framework describes that in the time interval $[t, t + dt)$, where t is time and dt an infinitesimally small time step, the transition rate of

the system going from a state x to a new state y will change by no more than one cell. Since we have several possible transitions (depending on how many species we have in our community), the probability of each to happen depends on the interaction coefficient and a random number in the interval $(0, 1)$. This probability is a Markovian stochastic process, i.e. at a given time t the conditional probability of a future state y depends only on the immediate previous state x (Erban et al. 2020).

Using this mathematical framework, we modeled the fluctuations happening in the persisters' population. Our findings emphasize the critical role of microbial interactions in determining the resilience of populations following antibiotic treatment. While loose interactions generally resulted in the presence of the persisters-generating species in the final communities, stronger interactions were able to completely eradicate such species from the community. On the other hand, we saw that population fluctuations, driven by stochasticity, could shape the capacity of microbes to persist, recover, and ultimately resist clearance, i.e. eventually dominate the community after the antibiotic treatment. This scenario was never observed using the determinist framework.

These insights underscore the importance of considering microbial communities as dynamic and interconnected systems when studying antibiotic efficacy and designing therapeutic strategies. By integrating ecological and stochastic perspectives, this study provides a deeper understanding of the factors driving microbial survival and recovery in the face of antibiotic challenges. Furthermore, this study highlights the crucial role of mathematical models to study and understand the dynamics of microbial communities.



Research Article

An ecological and stochastic perspective on persisters resuscitation

Tania Alonso-Vázquez^a, Michele Giovannini^a, Gian Luigi Garbini^a, Mikolaj Dziurzynski^a, Giovanni Bacci^a, Ester Coppini^b, Donatella Fibbi^b, Marco Fondi^{a,*}

^a Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, 50019, Italy

^b G.I.D.A. SpA, Via Baciavalle 36, Prato, 59100, Italy



ARTICLE INFO

MSC:
46N60
37N25

Keywords:
Persisters
Microbial interactions
Microbial communities

ABSTRACT

Resistance, tolerance, and persistence to antibiotics have mainly been studied at the level of a single microbial isolate. However, in recent years it has become evident that microbial interactions play a role in determining the success of antibiotic treatments, in particular by influencing the occurrence of persistence and tolerance within a population. Additionally, the challenge of resuscitation (the capability of a population to revive after antibiotic exposure) and pathogen clearance are strongly linked to the small size of the surviving population and to the presence of fluctuations in cell counts. Indeed, while large population dynamics can be considered deterministic, small populations are influenced by stochastic processes, making their behaviour less predictable. Our study argues that microbe-microbe interactions within a community affect the mode, tempo, and success of persister resuscitation and that these are further influenced by noise. To this aim, we developed a theoretical model of a three-member microbial community and analysed the role of cell-to-cell interactions on pathogen clearance, using both deterministic and stochastic simulations. Our findings highlight the importance of ecological interactions and population size fluctuations (and hence the underlying cellular mechanisms) in determining the resilience of microbial populations following antibiotic treatment.

1. Introduction

For a long time, antibiotic resistance, tolerance and persistence have been studied from the point of view of single organisms and how each drug could affect their genome, metabolism and growth rate. However, in recent years, the study of microbial communities and how their ecological interactions could alter the efficacy of antibiotics has gained ground. In fact, mounting evidence suggests that such interactions play a crucial role in the emergence and maintenance of antibiotic persistence/tolerance and in the actual efficiency of antibacterial treatments [1–6]. The work by De Vos et al. (2017), for example, characterised a large set of cooperative and competitive interactions in polymicrobial urinary tract infections, showing that *Proteus mirabilis* predominantly protect some isolates from antibiotics while harmed others in the absence of them. They highlighted the importance of interactions in i) maintaining the stability of the underlying population and ii) affecting the survival of community members during antibiotic exposure [1]. Mizrahi et al. (2023) showed that the biggest benefit usually goes to the more sensitive strains, allowing them to grow in concentrations higher than the minimum for inhibition. Hence, the evolution of antibiotic-

resistant and tolerant bacteria can be modified depending on the type and strength of interspecies interactions. Sensitive strains that manage to grow under antibiotic stress tend to have a decrease in the death rate, coupled with a decrease in the growth rate, leading to tolerance; i.e. lower growth rate, more tolerance. However, we need to consider that there must be a slow antibiotic degradation in the environment for the tolerance to evolve [5]. While resistance is inherited and may be acquired by horizontal gene transfer of resistance-encoding genes or mutations, tolerance is the general ability of a population to survive longer to antibiotic treatments by having a lower death and survival rate, without any genetic modifications. There are cells that not only exhibit tolerance to antibiotics but also persist in their presence without undergoing genetic changes. These persister cells are a subpopulation of a growing bacterial population that survives antibiotic treatment by entering a dormant state, which is a reversible, non-inherited phenomenon [7–9]. Recent evidence suggests that low ATP levels induce protein aggregation and thus non or slow-growing bacterial cells, stochastically favouring persisters formation [10]. Although this low-energy mechanism may be present in several, if not all, bacterial species, this is not the only mechanism that produces persister cells. Toxin-antitoxin

* Corresponding author.

E-mail address: marco.fondi@unifi.it (M. Fondi).

<https://doi.org/10.1016/j.csbj.2024.12.002>

Received 12 July 2024; Received in revised form 2 December 2024; Accepted 3 December 2024

(TA) loci, such as the HipA system in *Escherichia coli*, have been implicated in the formation of persister cells [11–14]. Although its role in this process has been controversial, with conflicting evidence regarding the exact mechanisms involved, HipA might indeed be correlated with persistence [15,16]. The activation of the bacterial SOS response is yet another mechanism that has been associated with bacterial persistence [17]. Overall, it has been estimated that the fraction of persisters within a microbial population may be typically less than 1–2%, depending on, among other factors, the antibiotics used, the growth phase and the taxonomy of the microbes under study [18,19,9,20,21,16]. Once cells have reached a dormancy state of persistence and survived the antibiotic treatment, they may start a resuscitation process: their exit from the persistent state to repopulate, upon cessation of antibiotic exposure [22]. The phenotypic features that allow persister resuscitation are starting to gain attention, since this process may represent an attractive target for antibiotic treatments [23]. Although resuscitation dynamics of a group of persisters has long been considered to be constant over time [24], a recent work suggested that cells resuscitate exponentially rather than with a fixed rate [25]. In 2023, Fang and colleagues modelled the dynamics of resuscitation by single-cell tracking and showed that, in most cases, this does not depend on the treatment duration but only on the antibiotic concentration (tested with ampicillin, imipenem and ciprofloxacin). One essential characteristic that may help the cell is its ability to detoxify the intracellular environment by coupling the inner-membrane ABC transporters with the outer-membrane protein TolC. In other words, changes in the efflux strongly delay the resuscitation process. Interestingly, they observed that persister partitioning is antibiotic-specific. They described two processes that changes the resuscitation of persisters cells: depletion of drug-target binding proteins (cell-wall targeting antibiotics) and the activation of SOS-response (quinolone antibiotics), resulting in triangular structural defects and filamentation/polar partitioning, respectively [26,27,25]. This leads to a population formed by healthy, damaged and failed cells (their proportion depending on the antibiotic treatment) that, as a whole, display a reduced growth rate in respect to the pre-antibiotic exposure phase [25].

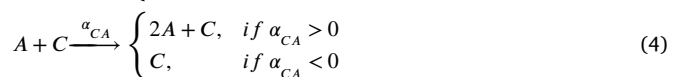
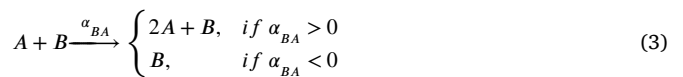
The problem of resuscitation and, more in general, pathogens clearance, is strongly connected to the small population size of the pathogen species. There are evidences that suggest that, even if the pathogen population has been drastically reduced by the antibiotic treatment, this may not be enough as failure to completely eradicate a population of bacteria may lead to undesired consequences. The reduced population can lead both to i) infection relapse [28,29] and ii) antibiotic resistance development [30]. Clearly, both scenarios complicate the subsequent treatments. While the dynamics of large populations can be thought of as a deterministic process, the dynamics of small populations may be less predictable and affected by factors that are negligible in the first scenario. The underlying probability of these stochastic processes has been described in different ways [31–33].

Here, we argue that microbe-microbe interactions in a microbial community represent an additional factor affecting the mode, tempo and success of persisters resuscitation. Since antibiotic exposure drastically reduces the size of sensitive species (virtually to the size of persisters pool), even mildly negative interactions may impede the resuscitation of persisters once their population size has dropped under a certain threshold. At the same time, noise may contribute to this complex scenario, as stochastic fluctuations in cell numbers could interfere with the dynamics of the ecological network and influence the resuscitation of the persister pool. To investigate these issues, we developed a theoretical model of a simple microbial community that includes one pathogenic, antibiotic-sensitive and persisters-generating species together with two other interacting community members and analysed how the interplay among the three players (as well as other model parameters) affects the resuscitation of the persisters, also when considering the role of noise.

2. Theory

In this work, we model a theoretical community made of three interacting species, namely species *A*, *B* and *C* (Fig. 1A). Each member of the community grows according to a specific growth rate (μ) and dies according to a specific death rate (δ). Species *C* represents a hypothetical pathogen, sensitive to a given antibiotic X (ABX) while *A* and *B* represent two other species that are not sensitive to ABX. Species *A*, *B* and *C* interact with each other and the strength of these interactions is given by coefficients of the type $\alpha_{i,j}$, accounting for the effect of species *i* on species *j* (Fig. 1B). During growth, species *C* generates persisters *P* with a rate γ . Persisters *P* do not grow ($\mu_P = 0$), are not sensitive to antibiotic and do not interact with species *A* and *B*. We then consider the scenario in which the community is exposed to antibiotic for a given amount of time (from t_{start} to t_{end}). During this time period, the growth rate of species *C* is set to 0 ($\mu_C = 0$), an antibiotic killing rate is introduced (ξ), all the interactions affecting species *C* are set to 0 ($\alpha_{A,C} = \alpha_{C,A} = \alpha_{B,C} = \alpha_{C,B} = 0$) and persisters production rate is set to 0 ($\gamma = 0$). After the exposure to the antibiotic treatment (i.e. when simulation time $t > t_{end}$) we simulate the resuscitation of persisters *P* by converting *P* into *C* with a rate equal to $\alpha_{res} e^{\beta t}$ as reported in [25] for ampicillin treatment in *E. coli* and *Salmonella enterica* and setting back to their original values all the interaction parameters ($\alpha_{i,j}$) and the persisters generation rate (γ). As for *C* growth rate after antibiotic exposure, we considered the fact that cells resuscitated from the pool of persisters sometimes manifest a damaged phenotype [25] and set the post-antibiotic growth rate μ_C to the 75% of its original value μ_C . Simulations then are continued until all the species have reached their steady state (if any). We used this simple and general modelling framework to evaluate the role of specific parameters (e.g. the duration of the antibiotic treatment, the persisters production rate and, more importantly, the effect of species interactions) on the capability of pathogen species *C* to restore normal growth after the antibiotic-induced perturbation.

The model can be formalised by the following set of equations. We report here the case for species *A*, which is analogous for species *B* and *C*:



Clearly if any of the above mentioned α is equal to 0, no changes in *A* will occur. In particular, for species *C* we also have persisters formation with rate γ :



Furthermore, during antibiotic exposure, we include the following equations to the system to account for the antibiotic killing rate ξ :



Finally, after the exposure to the antibiotic, we modelled persisters resuscitation as follows:



This set of equations leads to the following set of four ordinary differential equations (ODEs):

Table 1

Values of the model parameter together with their source.

Parameter	Description	Value	Unit	Reference
μ	Growth rate	0.7	h^{-1}	[35]
t_{start}, t_{end}	Time of ABX exposure	12, 18	h	-
δ	Death rate	0.1	h^{-1}	[36]
γ	Persisters generation rate	0 - 0.1	h^{-1}	[21]
ξ	ABX killing rate	0 - 2.0	h^{-1}	[37]
α_{res}	Resuscitation rate	0.2	-	[25]
ρ	Growth defect of resuscitation rate	0.75	-	[25]

$$[\dot{A}] = \mu[A] + \alpha_{BA}[AB] + \alpha_{CA}[AC] + \alpha_{AA}[A]^2 - \delta[A] \quad (10)$$

$$[\dot{B}] = \mu[B] + \alpha_{AB}[BA] + \alpha_{CB}[BC] + \alpha_{BB}[B]^2 - \delta[B] \quad (11)$$

$$[\dot{C}] = \mu[C] + \alpha_{AC}[CA] + \alpha_{BC}[CB] + \alpha_{CC}[C]^2 - \delta[C] - \gamma[C] \quad (12)$$

$$[\dot{P}] = \gamma[C] - \delta[P] \quad (13)$$

When $t_{start} < t < t_{end}$, i.e. during antibiotic exposure, equation (12) changes to:

$$[\dot{C}] = \mu[C] + \alpha_{AC}[CA] + \alpha_{BC}[CB] + \alpha_{CC}[C]^2 - \delta[C] - \gamma[C] - \xi[C] \quad (14)$$

And when $t > t_{end}$, i.e. after antibiotic exposure, equations (12) and (13) change to:

$$[\dot{C}] = \alpha_{res}e^{\beta t}[P] + \rho\mu[C] + \alpha_{AC}[CA] + \alpha_{BC}[CB] + \alpha_{CC}[C]^2 - \delta[C] - \gamma[C] \quad (15)$$

$$[\dot{P}] = \gamma[C] - \delta[P] - \alpha_{res}e^{\beta t}[P] \quad (16)$$

where ρ is a scaling factor to account for the growth defect of resuscitated C cells. The values of the parameters are reported in Table 1. We also formulated the deterministic model of the community represented by the Eqs. (1) - (9) as an ensemble of stochastic rules in a hypothetical volume (V) of 1000 μ l, a value that is in line with previous experiments/models on microbial community dynamics [34]. Mathematically, at each time t the system is fully described by $\mathbf{x} = (n_A, n_B, n_C, n_P)$, a vector whose components are discrete quantities that identify the total number of elements of species i (for $i = A, B, C, P$). To proceed in the analysis, it is necessary to introduce the transition rates $T(\mathbf{y}|\mathbf{x})$ from the state $\mathbf{x}$ to the new state $\mathbf{y}$, where $\mathbf{y}$ is a state compatible with the equations listed above. These transitions originate from Eqs. (1) - (9) and their explicit forms are as follows.

$$T_1(n_A + 1|n_A) = \mu \cdot n_A \quad (17)$$

$$T_2(n_B + 1|n_B) = \mu \cdot n_B \quad (18)$$

$$T_3(n_C + 1|n_C) = \mu \cdot n_C \quad (19)$$

$$T_4(n_A - 1|n_A, n_B) = \frac{\alpha_{BA}}{V} \cdot n_A \cdot n_B, \quad \text{if } \alpha_{BA} < 0 \quad (20)$$

$$T_5(n_A + 1|n_A, n_B) = \frac{\alpha_{BA}}{V} \cdot n_A \cdot n_B, \quad \text{if } \alpha_{BA} > 0 \quad (21)$$

$$T_6(n_A - 1|n_A, n_C) = \frac{\alpha_{CA}}{V} \cdot n_A \cdot n_C, \quad \text{if } \alpha_{CA} < 0 \quad (22)$$

$$T_7(n_A + 1|n_A, n_C) = \frac{\alpha_{CA}}{V} \cdot n_A \cdot n_C, \quad \text{if } \alpha_{CA} > 0 \quad (23)$$

$$T_8(n_B - 1|n_B, n_A) = \frac{\alpha_{AB}}{V} \cdot n_B \cdot n_A, \quad \text{if } \alpha_{AB} < 0 \quad (24)$$

$$T_9(n_B + 1|n_B, n_A) = \frac{\alpha_{AB}}{V} \cdot n_B \cdot n_A, \quad \text{if } \alpha_{AB} > 0 \quad (25)$$

$$T_{10}(n_B - 1|n_B, n_C) = \frac{\alpha_{CB}}{V} \cdot n_B \cdot n_C, \quad \text{if } \alpha_{CB} < 0 \quad (26)$$

$$T_{11}(n_B + 1|n_B, n_C) = \frac{\alpha_{CB}}{V} \cdot n_B \cdot n_C, \quad \text{if } \alpha_{CB} > 0 \quad (27)$$

$$T_{12}(n_C - 1|n_C, n_A) = \frac{\alpha_{AC}}{V} \cdot n_C \cdot n_A, \quad \text{if } \alpha_{AC} < 0 \quad (28)$$

$$T_{13}(n_C + 1|n_C, n_A) = \frac{\alpha_{AC}}{V} \cdot n_C \cdot n_A, \quad \text{if } \alpha_{AC} > 0 \quad (29)$$

$$T_{14}(n_C - 1|n_C, n_B) = \frac{\alpha_{BC}}{V} \cdot n_C \cdot n_B, \quad \text{if } \alpha_{BC} < 0 \quad (30)$$

$$T_{15}(n_C + 1|n_C, n_B) = \frac{\alpha_{BC}}{V} \cdot n_C \cdot n_B, \quad \text{if } \alpha_{BC} > 0 \quad (31)$$

$$T_{16}(n_C - 1, n_P + 1|n_C, n_P) = \gamma \cdot n_C \quad (32)$$

$$T_{17}(n_A - 1|n_A) = \delta \cdot n_A \quad (33)$$

$$T_{18}(n_B - 1|n_B) = \delta \cdot n_B \quad (34)$$

$$T_{19}(n_C - 1|n_C) = \delta \cdot n_C \quad (35)$$

$$T_{20}(n_C - 1|n_C) = \xi \cdot n_C \quad (36)$$

$$T_{21}(n_P - 1|n_P) = \delta \cdot n_P \quad (37)$$

$$T_{22}(n_C + 1, n_P - 1|n_C, n_P) = \alpha_{res}e^{\beta t} \cdot n_P \quad (38)$$

$$T_{23}(n_A - 1|n_A) = \frac{\alpha_{AA}}{V} \cdot n_A \cdot (n_A - 1), \quad \text{if } \alpha_{AA} < 0 \quad (39)$$

$$T_{24}(n_A + 1|n_A) = \frac{\alpha_{AA}}{V} \cdot n_A \cdot (n_A - 1), \quad \text{if } \alpha_{AA} > 0 \quad (40)$$

$$T_{25}(n_B - 1|n_B) = \frac{\alpha_{BB}}{V} \cdot n_B \cdot (n_B - 1), \quad \text{if } \alpha_{BB} < 0 \quad (41)$$

$$T_{26}(n_B + 1|n_B) = \frac{\alpha_{BB}}{V} \cdot n_B \cdot (n_B - 1), \quad \text{if } \alpha_{BB} > 0 \quad (42)$$

$$T_{27}(n_C - 1|n_C) = \frac{\alpha_{CC}}{V} \cdot n_C \cdot (n_C - 1), \quad \text{if } \alpha_{CC} < 0 \quad (43)$$

$$T_{28}(n_C + 1|n_C) = \frac{\alpha_{CC}}{V} \cdot n_C \cdot (n_C - 1), \quad \text{if } \alpha_{CC} > 0 \quad (44)$$

where, in the definition of each $T(n_i \pm 1|n_i)$, we included only those species actually taking part in the corresponding equation. Clearly, as in the deterministic formulation of the model, if any of the above mentioned α is equal to 0, no changes in the corresponding species will occur. Once the transition rates have been defined, we can define the master equation accounting for the probability of observing the system in its state $\mathbf{x}$ at time t , given the transition rates that describe the probability of passing from state $\mathbf{x}$ to state $\mathbf{y}$. In its general form, the chemical master equation can be written as follows:

$$\frac{d}{dt} P(\mathbf{x}, t) = \sum_{\mathbf{y} \neq \mathbf{x}} T(\mathbf{x}|\mathbf{y})P(\mathbf{y}, t) - T(\mathbf{y}|\mathbf{x})P(\mathbf{x}, t) \quad (45)$$

We will now introduce one step operator ϵ (or van Kampen operator) for each species in the model in order to be able to compact the formulation of the chemical master equation shown in Eq. (45).

$$\epsilon_A^\pm = f(n_A \pm 1, n_B, n_C, n_P) \quad (46)$$

$$\epsilon_B^\pm = f(n_A, n_B \pm 1, n_C, n_P) \quad (47)$$

$$\epsilon_C^\pm = f(n_A, n_B, n_C \pm 1, n_P) \quad (48)$$

$$\epsilon_P^\pm = f(n_A, n_B, n_C, n_P \pm 1) \quad (49)$$

where f is an arbitrary function that depends on the state $\mathbf{x}$. The subscript of each step operator ϵ embeds in the pedix the information on which model species increases or decreases following the corresponding chemical reaction (e.g. ϵ_A for species A).

Accordingly, the explicit form of the chemical master equation governing the evolution of the discrete stochastic model assumes the following form:

$$\begin{aligned} \frac{d}{dt} P(\mathbf{x}, t) = & \left[(\epsilon_A^- - 1)T_1(n_A + 1|n_A) + (\epsilon_B^- - 1)T_2(n_B + 1|n_B) \right. \\ & + (\epsilon_C^- - 1)T_3(n_C + 1|n_C) + (\epsilon_A^+ - 1)T_4(n_A - 1|n_A, n_B) \\ & + (\epsilon_A^- - 1)T_5(n_A + 1|n_A, n_B) + (\epsilon_A^+ - 1)T_6(n_A - 1|n_A, n_C) \\ & + (\epsilon_A^- - 1)T_7(n_A + 1|n_A, n_C) + (\epsilon_B^+ - 1)T_8(n_B - 1|n_B, n_A) \\ & \left. + (\epsilon_B^- - 1)T_9(n_B + 1|n_B, n_A) + (\epsilon_B^+ - 1)T_{10}(n_B - 1|n_B, n_C) \right] \end{aligned}$$

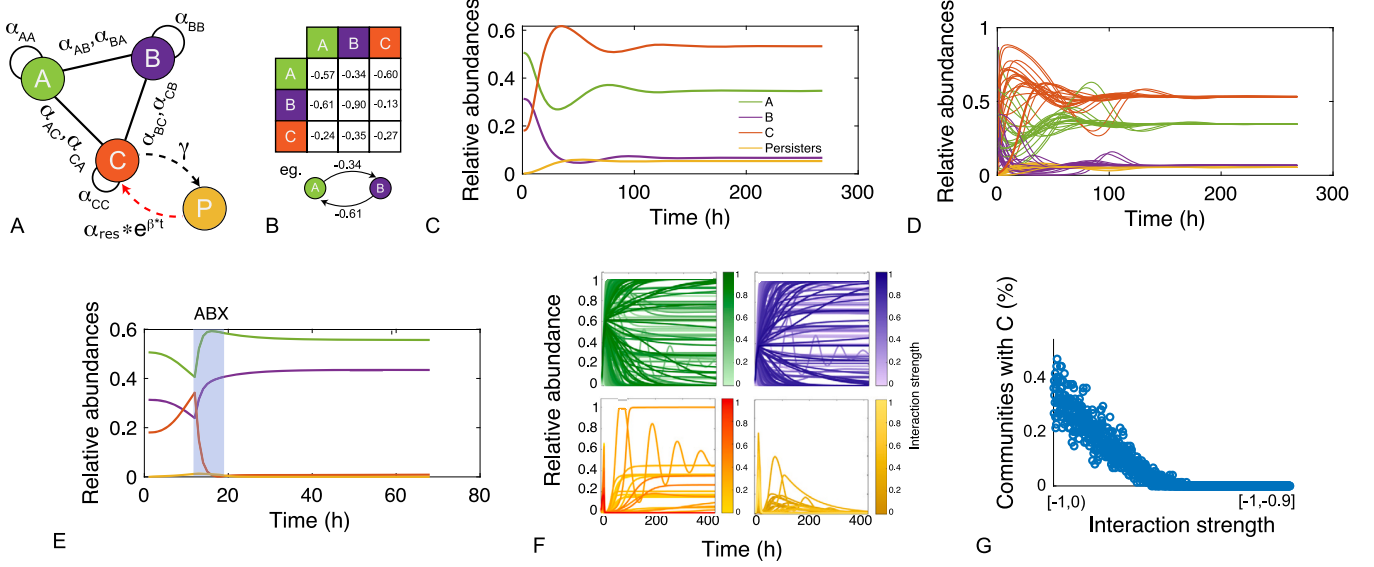


Fig. 1. Schematic representation of the three-member community modelled herein. (A) Network describing the dynamics of the species *A* (in green), *B* (in purple) and *C* (in orange), and the possibility of the persister *P* formation (in yellow). (B) Random interaction matrix that produces a stable non-zero steady state. (C) Sample simulation with a final co-existence of all the three species. (D) Dynamics of the community with different initial abundances for each species lead to the same steady state. (E) Dynamics of the community under antibiotic (ABX) stress in the interval 12–18 h. (F) Changes in the final composition of abundances resulting from varying the strength of the interactions among the three community members (i.e. changing the interaction matrix; line opacity indicates interaction strength). (G) Percentage of communities in which species *C* disappears when the interaction strength increases.

$$\begin{aligned}
 & +(\epsilon_B^- - 1)T_{11}(n_B + 1|n_B, n_C) + (\epsilon_C^+ - 1)T_{12}(n_C - 1|n_C, n_A) \\
 & +(\epsilon_C^- - 1)T_{13}(n_C + 1|n_C, n_A) + (\epsilon_C^+ - 1)T_{14}(n_C - 1|n_C, n_B) \\
 & +(\epsilon_C^- - 1)T_{15}(n_C + 1|n_C, n_B) + (\epsilon_C^+ \epsilon_P^- - 1)T_{16}(n_C - 1, n_P + 1|n_C, n_P) \\
 & +(\epsilon_A^+ - 1)T_{17}(n_A - 1|n_A) + (\epsilon_B^+ - 1)T_{18}(n_B - 1|n_B) \\
 & +(\epsilon_C^+ - 1)T_{19}(n_C - 1|n_C) + (\epsilon_C^+ - 1)T_{20}(n_C - 1|n_C) \\
 & +(\epsilon_P^+ - 1)T_{21}(n_P - 1|n_P) + (\epsilon_C^- \epsilon_P^+ - 1)T_{22}(n_C + 1, n_P - 1|n_C, n_P) \\
 & +(\epsilon_A^+ - 1)T_{23}(n_A - 1|n_A) + (\epsilon_A^- - 1)T_{24}(n_A + 1|n_A) \\
 & +(\epsilon_B^+ - 1)T_{25}(n_B - 1|n_B) + (\epsilon_B^- - 1)T_{26}(n_B + 1|n_B) \\
 & +(\epsilon_C^+ - 1)T_{27}(n_C - 1|n_C) + (\epsilon_C^- - 1)T_{28}(n_C + 1|n_C) \Big] P(\mathbf{x}, t)
 \end{aligned}$$

3. Results and discussion

3.1. The role of microbe-microbe interactions in persisters resuscitation

Microbial interactions can profoundly affect the dynamics of a microbial community, leading to very different end-points of each species [38]. We started by testing the effect of different interaction matrices on the final steady state of the community. We thus randomly sampled 100K interaction matrices in the interval $[-1, 1]$ and, for each of them, computed the end-point of the community. We found that, as expected [38], different interactions matrices led to different community end-points and to very different dynamics of each species in the model described by Eqs (10)–(13) (Fig. 1F).

Among all the (randomly assembled) matrices, we identified and preliminary selected a matrix that could produce *i*) a stable steady state, *ii*) a community in which each species reached a non-zero final state and *iii*) where the pathogenic species showed the highest abundance at the end of simulations (Fig. 1B and C). Not surprisingly, this matrix was composed by negative interactions, as positive interactions are generally thought to be rare and to destabilise microbial communities [39,40] (although recent evidence suggests that positive interactions may be less rare than we think [41]). Also, the values of these interaction matrices fall well in the range of other interaction matrices derived from other

works [42,1]. Finally, consistent with the underlying biology, the diagonal of this matrix is composed of negative entries as each of these species would eventually reach carrying capacity even in the absence of other species. Importantly, the final composition of the community did not change regardless of the initial abundance of each species (Fig. 1D). We then simulated the exposure to the antibiotic in the interval $t_{start} = 12$ h and $t_{end} = 18$ h. According to the model implementation (Fig. 1E), species *C* is affected by such perturbation and reaches a very low level of (relative) abundance. The other species react to such change by increasing (*A* and *B*) or decreasing (*P*) their abundance, as determined by the model structure. After antibiotic exposure, species *C* starts increasing again and the community gets to its new steady state. Using the end-points of such perturbed community as the initial points for a new simulation (without exposing it to the antibiotic) led to a final steady state identical to that of the initial simulation shown in Fig. 1C (See Figure S1). We next asked whether the structure of the community, i.e. the strength of the interactions among the three members, influenced the new steady state reached by species *C* following antibiotic exposure. We thus tested a wide range of possible interaction matrices in the interval $[-1, 0]$. As shown in Fig. 1F, the final composition of species *C* is profoundly affected by variations in the strength of the interactions among the members of the community. We analysed this aspect more thoroughly by sampling random interaction matrices in an increasingly narrower range and computing the final community composition for each of them (as the percentage of steady state communities including species *C*). We found that species *C* gets progressively excluded from the community with the increasing of the overall strength of microbe-microbe interactions (Fig. 1G). So, according to our model, the strength of the interactions of two other community members over a pathogenic, persisters-generating species may be enough to prevent the resuscitation of persisters following antibiotic treatment. This suggests that ecological interactions may play a role in determining the success of persisters resuscitation.

We then tested whether changes in the main parameters of the model (namely antibiotic exposure time t_{end} , antibiotic killing rate ξ and persister generation rate γ) could have any impact on the end-point values of species *C*. Furthermore, as we have previously shown that the interaction matrices can have a role in determining the final state of the

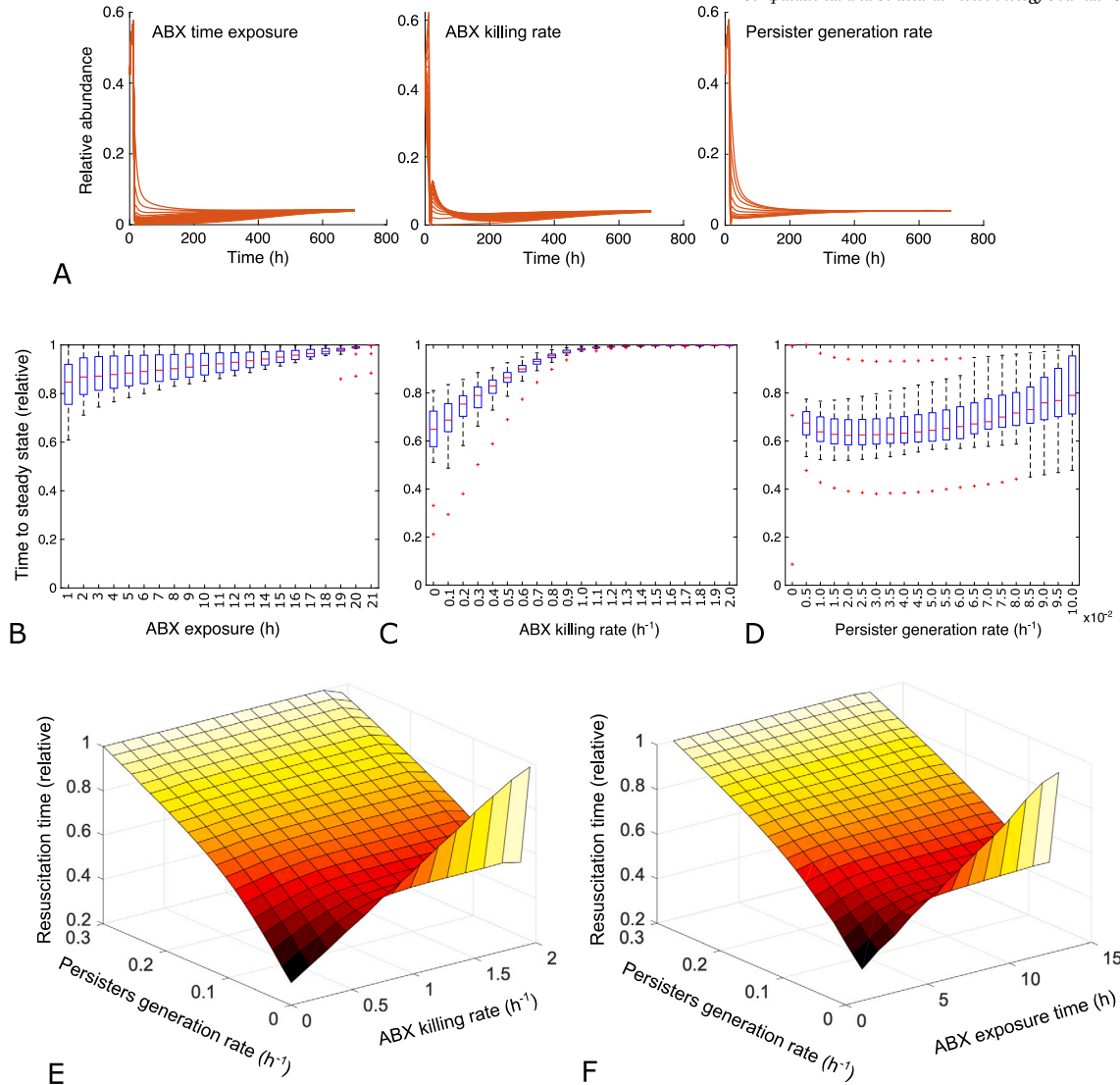


Fig. 2. Changes in the main parameters of the model. A) Changes in antibiotic exposure time t_{end} , antibiotic killing rate ξ and persister generation rate γ have no impact on the end-point values of species C. B-D) Impact on the post-antibiotic dynamic of species C of antibiotic exposure time (B), its strength (C) and persister generation rate (γ) (D). E) The effect of different antibiotic killing rate on the optimal resuscitation time. F) The effect of antibiotic exposure time on the optimal resuscitation time.

corresponding community, we tested the impact of each parameter's value on a random sample of 1000 interaction matrices. As shown in Fig. 2A, changes in antibiotic time exposure, antibiotic killing rate and rate of persister formation did not significantly alter the final end-point of species C. However, what seemed to be affected was the so-called resuscitation time of persisters [22,25], i.e. the time it takes for persister cells to start growing again after the antibiotic effect has vanished. Indeed, some differences among the effects of the tested parameters could be observed when considering the post-antibiotic dynamic of species C (Fig. 2A). To investigate this aspect more thoroughly, we measured the effect of the same three parameters on the time it took for the species C to get back to its steady state after exposing the community to the antibiotic. As shown in Fig. 2B and C, varying two of the three parameters led to expected consequences. Indeed, increasing the exposure time of the antibiotic (Fig. 2B) and its strength (Fig. 2C) is reflected in a progressively longer resuscitation time. When considering persisters generation rate, instead, we noticed a different trend: resuscitation time was maximal with γ equal to zero (no generation of persisters) and reached similarly high levels with a γ value of $0.1 h^{-1}$. In between, we identified the most efficient response (i.e. the fastest of resuscita-

tion time) with γ values ranging between 0.02 and $0.03 h^{-1}$ (Fig. 2D). This suggests the presence of an optimum in persister generation rate, i.e. in the fraction of the population that is stochastically converted to persisting (dormant) cells. Higher values are not optimal likely because they reduce the population to extremely low values so that only persister cells are responsible for the resuscitation process and, as previously mentioned, their growth restart is likely to be impaired in respect to healthy cells [25]. Conversely, γ values that are too low to prevent the generation of enough persister cells to ensure the resuscitation after the exposure to the antibiotic treatment. Interestingly, among all the persisters generation rates tested, those that correspond to the apparently optimal resuscitation strategies fall in the range of the common persisters formation rate found in a wide range of microbial species [22].

Next, we asked whether the presence of an optimal resuscitation time was somehow linked to the values we chose for the other two parameters (ξ and t_{end}). Fig. 2E shows the effect of different antibiotic killing rate on the optimal resuscitation time. While we found that the presence of such an optimum was maintained regardless of the antibiotic killing rate, we also found that the optimal range narrows as the antibiotic killing rate increases. In other words, the stronger the killing power of the antibiotic,

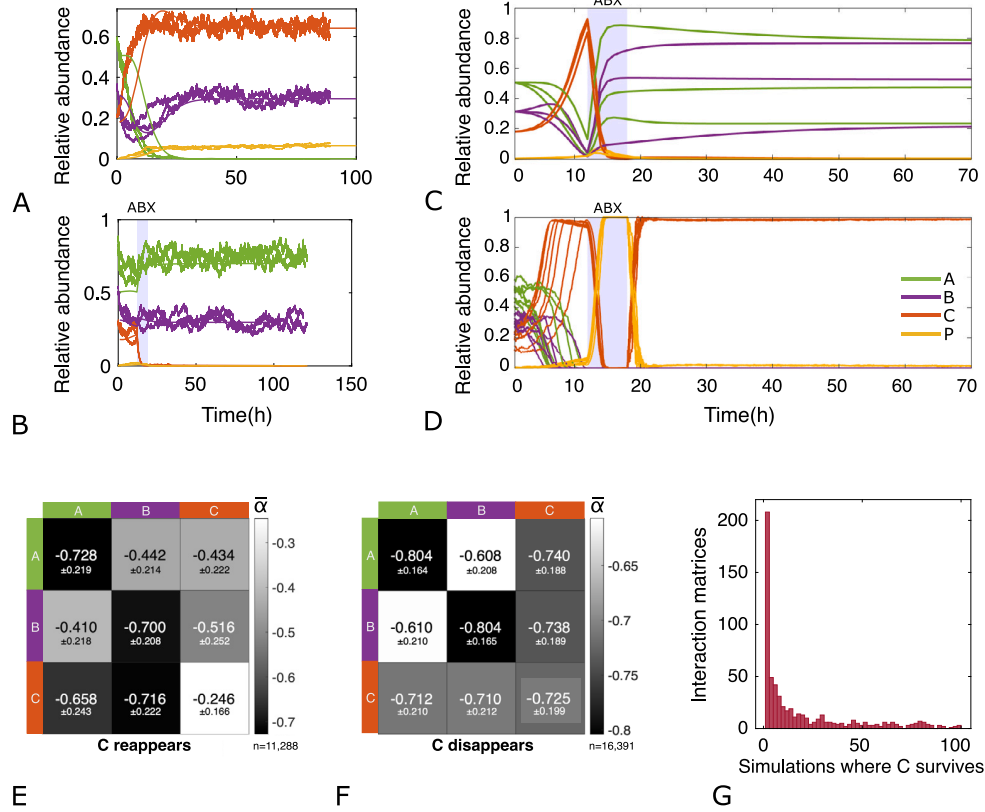


Fig. 3. The effect of fluctuations in a three-member community. A-B) A sample of deterministic (straight line) and stochastic (step line) simulations. It is shown the overlap between the two models, both in a non-antibiotic (A) and antibiotic (B) scenario. However, stochastic effects may become relevant when the size of species C population reaches very low values. C) Deterministic simulation of nine communities using different interaction matrices that resulted in the extinction of species C. D) Using the same interaction matrices, species C reappears when modelling the community dynamics stochastically while A and B disappear from the community. Not all interaction matrices led to persist resuscitation, and thus to species C reappearance. Here we show in E) the average α values that lead to it ($n=11,288$ for each α), and in F) the average α values that lead species C to extinction even under the fluctuation system ($n=16,391$ for each α). In both matrices the standard deviation is reported. Furthermore, on G) we show the number of interaction matrices that led to the survival of species C when modelling the dynamics stochastically, and the number of simulations that used that specific matrix. In other words, there is a set of interaction coefficients that are more common than others (Species A: green, B: purple, C: red, P: yellow).

the more important it is to maintain the fraction of persister cells in the optimal range in order to provide a faster resuscitation time. While moving from this optimal range could have no effect (or even positive effects) in the presence of a weak antibiotic, doing so in the presence of a strong one may lead to significantly longer resuscitation times (Fig. 2E). We noticed the same trend also when examining the role of different γ values with varying antibiotic exposure time (Fig. 2F).

3.2. Accounting for noise in population dynamics affects persisters resuscitation

Up to now, we have modelled our antibiotic-perturbed community using a deterministic approach. Next, we simulated the same communities using a stochastic framework in which random fluctuations of species abundances are taken into account. A sample of such simulations is reported in Fig. 3, where we show that the stochastic simulations overlap with the deterministic ones both in a non-antibiotic and antibiotic scenario (Fig. 3A and B, respectively). While this agreement between stochastic and deterministic models was expected for large bacterial populations, when the size of species C population reaches very low values, stochastic effects may become relevant in determining C resuscitation. Thus, we reasoned that, when species C reaches very low population values following antibiotic exposure, stochasticity may begin to play an important role in determining the final end-point of the community. More specifically, we asked whether noise could impact the clearance of species C from the community when the deterministic

model predicted so (i.e. a “no resuscitation” scenario). In other words, can random fluctuations in bacterial density counteract the effect of social interactions? In this regard, we specifically focused our attention on those simulation runs in which, following antibiotic exposure, the pathogen species C got extinct or approached very low abundance value (relative abundance lower than 1%). We explored 10^5 different interaction matrices randomly sampled in the interval $[-0.01, -1]$ and kept only those matrices that, following a 6 hours exposure to the antibiotic ($t_{start} = 12$ and $t_{end} = 18$), deterministically predicted a species C steady state with a relative abundance lower than 1%. These represent very interesting cases in which the structure of the community (i.e. the interactions among its members) are key to prevent the resuscitation of persisters. However, when dealing with low population sizes, stochastic fluctuations may lead to unpredictable steady states. In the present context, this may translate into a non zero probability of pathogen clearance and, in the worst case scenario, in the revival of the susceptible population. Let $\mathcal{M}$ be the set of such matrices (i.e. those that lead to the extinction of species C using the deterministic approach) and n_C the number of cells of species C:

$$\mathcal{M} = \{M = [\alpha_{ij}] \text{ where } i, j = A, B, C \mid n_C^{det} < 1\% \} \quad (50)$$

We identified a total of 26,116 of such matrices and a snapshot of the deterministic dynamics of some of the underlying communities is reported in Fig. 3C. Then, to account for fluctuations in the number of cells for each species, we performed 100 simulations for each of the

26,116 interaction matrices using the stochastic rules described above. Each simulation had one interaction matrix and specific transition probabilities, i.e. the probability of each reaction to happen. In other words, when considering an infinitesimal time interval, we can assure that only one of the Eq. (17) - (44) will take place, accounting for the increase or decrease in A , B , C or P population by one cell. Once we had a simulation using one of the interaction matrices in $\mathcal{M}$, we checked whether the stochastic end-points for the species C matched those predicted by the deterministic simulations. Interestingly, while in the vast majority of the simulations this turned out to be the case, we found instances in which the stochastic end-point of C was higher than that predicted by the deterministic model (Fig. 3D and C, respectively). Approximately, 1% of these simulations resulted in persisters' resuscitation and the survival of the antibiotic-sensitive species C . Let S be the subset of $\mathcal{M}$ that lead species C to survive using the stochastic framework:

$$S = \{S \in \mathcal{M} \mid n_C^{sto} > 0\} \quad (51)$$

The average α_{ij} values in S showed that the self-interactions of A and B tend to be stronger than C self-interaction, i.e. $\bar{\alpha}_{A,A}, \bar{\alpha}_{B,B} < \bar{\alpha}_{C,C}$, since they are negative interactions (Fig. 3E). On the other hand, average α_{ij} values in S^c , i.e. the interaction matrices that led species C to the extinction and to the overlap with the deterministic model, showed the opposite trend. Self-interaction of species C was stronger than the others, as well as the interactions between species A and B . In other words, there was an increase of $\bar{\alpha}_{C,C}$, $\bar{\alpha}_{A,B}$ and $\bar{\alpha}_{B,A}$ (Fig. 3F). These results highlight the importance of self-interactions in the fate of the perturbed species. From a biological stand point, since self-interactions account for the capability of a species to regulate its own abundance and/or a larger or smaller carrying capacity, this might underscore the importance of population-level features such as quorum sensing and/or nutrients exploitation in determining the relapse of species C in our community. However, S is not conformed by unique S matrices since there are several repeated elements, due to the different transition probabilities. In Fig. 3G we show the number of matrices that described the interactions of more than one simulation and let species C to survive. There were several interaction matrices S that described a small number of simulations where species C survived, and a small number of S that described a large number of them. The distribution of these matrices described how the interactions between the three members of the community have to be in order for species C to reappear, regardless the transition probabilities of each reaction. In Figure S2 we can see the S matrices that described the interactions of more than 94% of the communities simulated for each matrix, showing the α_{ij} values. In general, these matrices are characterised by a loose self interaction of species C , especially the ones that led C to survive 99% or 100% of the times, with $\alpha_{C,C} \approx 0$.

We next tested whether the state of the system, i.e. whether it has reached an equilibrium or not, may influence the dynamics of the entire community. Indeed, providing the antibiotic in the initial stages of species growth may result in the simultaneous application of two transient effects (antibiotic exposure and exponential growth of the species). We thus performed an additional set of simulations in which we allowed the system to get to the equilibrium (around $t = 70$ h) before adding the antibiotic (maintaining the exposure length as in previous simulations, i.e. 6 hours). Results of these simulations are shown in Figure S3 and were in good agreement with those shown in Fig. 3. However, one interesting features emerged. Postponing the antibiotic treatment to a time point where the species have reached an equilibrium point apparently reduces the probability of pathogen relapse in the community (see Figure S3A). Indeed, we found more instances in which the final community still embeds the pathogenic species in respect to simulations where the antibiotics was introduced from time $t = 12$ to $t = 18$ hours (see Figure S3E). Recent research has shown that early, aggressive, and sustained antibiotic regimens are often more effective at reducing relapse rates because they minimize the window of opportunity for persisters to repopulate. Infections involving pathogens like *Pseudomonas*

aeruginosa demonstrate that improper timing of treatment can lead to the evolution of persister-enriched populations, complicating eradication efforts [43,44].

4. Conclusions

In this work, we have assembled a simple theoretical community and studied the role of *i*) ecological interactions, *ii*) stress-related parameters and *iii*) stochastic fluctuations on the fate of persisters following a simulated antibiotic treatment. Our simulations indicated a key role of ecological interactions on the outcome of persisters once the stress (the antibiotics) vanished. While loose interactions generally resulted in the presence of the persisters-generating species in the final communities, tighter interactions concurred at the complete eradication of such species from the community. Interestingly, while antibiotic strength (efficacy) and treatment exposure correlated positively with persisters resuscitation, the rate at which persisters are formed displayed an optimum around the value commonly found in nature for this phenomenon. Finally, we specifically analysed those communities in which the ecological interactions concurred to the eradication of the pathogen using a stochastic framework that accounted for fluctuations at the level of cell counts. In this case, we observed that, in some cases, the pathogen species could resuscitate and eventually dominate the community following antibiotic treatment, a scenario that was never observed using the deterministic model. The relapse of bacterial infections due to persisters resuscitation is known to occur during real microbial infections eradication [25,45,16] and our simulations indicate that population size fluctuations might have a role therein and, more in general, in the resilience of microbial populations following a stressful perturbation. This work represents a first attempt to quantify the impact of the bacterial "social network" [46] on the survival and resuscitation of persisters in natural communities. In strictly clinical settings, the patterns we observed match our current knowledge derived from available experimental data. Overall, studies on antibiotic susceptibility that take microbial interaction networks and polymicrobial ecosystems into consideration are illuminating new bacterial survival strategies and therapeutic targets [47], including, for example, the observation that the persistence of one specific strain is negatively correlated with a close relative that shares the same metabolic necessities, i.e. both strains negatively interact one with the other [48,49]. Other relevant cases in which the importance of microbial interactions on infection relapse has been shown experimentally include the interaction between the gut microbiota and *Clostridioides difficile* [50–52], *Pseudomonas aeruginosa* and *Stenotrophomonas maltophilia* in Cystic Fibrosis (CF) [53], *Helicobacter pylori* and *Lactobacillus* spp. in the stomach [54,55] and polymicrobial infections of urinary tract [56,57]. All these examples underscore the importance of providing a theoretical background to the dynamics of poly-microbial infections and on the consequences of perturbations therein. While we have obtained preliminary results that confirm the key role of microbe-microbe interactions in community dynamics, further work will be necessary to upscale our model to resemble real-world communities and/or to set up a mock community that resembles the one modelled here. Indeed, for example, we here assumed that the non-pathogenic members of the community (A and B) were not affected by the ABX treatment and this clearly biased the outcomes of our simulations. What if some of the interacting members of the community are also affected by the antibiotics (as it may happen with broad-range antibiotic treatments)? Further, it will be interesting to address the role of ecological interactions in persisters resuscitation when the size of the community (including the number of pathogenic, persister-generating species) increases. Indeed, species richness and the strength of species interactions have been shown to regulate community dynamics [58]. Moreover, varying the antibiotic treatment strategy in a more drastic way could produce different outcomes from those that we obtained here. Antibiotics with a broad spectrum of action (or the use of multiple antibiotics) and/or the manipulation of antibiotic killing time have been

shown to induce different responses in the target microbial community [59,60]. Similarly, stochastic modelling has shown that the scenario of density regulation and the mode of action of the drug influence the survival of a resistant subpopulation [33]. Would these features hold true also when ecological interactions and stochastic fluctuations are considered at the same time? Finally, the modelling framework developed here is general enough to be used, in the future, for the study of microbial communities dynamics in other contexts in which the anthropogenic impact, among the others, is known to cause significant alterations to the community profile. These could include, for example, waste water treatment plants [61], the gut microbiota [62,63], agroecosystems [64] and natural systems in general [64].

CRedit authorship contribution statement

Tania Alonso-Vázquez: Writing – review & editing, Validation, Conceptualization. **Michele Giovannini:** Writing – review & editing, Conceptualization. **Gian Luigi Garbini:** Writing – review & editing, Conceptualization. **Mikolaj Dziurzynski:** Writing – review & editing, Conceptualization. **Giovanni Bacci:** Writing – review & editing, Conceptualization. **Ester Coppini:** Supervision, Funding acquisition. **Donatella Fibbi:** Supervision, Funding acquisition. **Marco Fondi:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

Material and methods

All the simulations were carried out using MATLAB v. 2022b. All the codes used in this work are available at Zenodo (<https://doi.org/10.5281/zenodo.12598803>).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Marco Fondi reports financial support was provided by Italian Ministry of University and Research (Grant code: 2022Z88RK4). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

National Recovery and Resilience Plan, Mission 4 Component 2 – Investment 1.4 – National Center for HPC, Big Data and Quantum Computing – funded by the European Union – NextGenerationEU – CUP (B83C22002830001) is acknowledged. M.F. acknowledges the funding from Project “Microbiome Assembly: A multilevel approach to predict establishment and variation of microbiomes” funded the Italian ‘Bando Prin 2022 - D.D. n. 104 del 02-02-2022’ by Italian Ministry of University and Research (Grant code: 2022Z88RK4). T.A.V. is supported by a PON PhD fellowship by the Italian Ministry of University and Research, “National Operational Programme on Research and Innovation 2014-2020” (CCI 2014IT16M2OP005), funds FSE REACT-EU, “Action IV.4 – PhD programmes and research contracts on innovation topics, and Action IV.5 – PhD programmes on green topics”. M.D. was funded by European Molecular Biology Organization Postdoctoral Fellowship (grant no. ALTF 565-2021).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.csbj.2024.12.002>.

References

- [1] de Vos MGJ, Zagorski M, McNally A, Bollenbach T. Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. *Proc Natl Acad Sci* 2017;114(40):10666–71.
- [2] Adamowicz EM, Muza M, Chacón JM, Harcombe WR. Cross-feeding modulates the rate and mechanism of antibiotic resistance evolution in a model microbial community of *Escherichia coli* and *Salmonella enterica*. *PLoS Pathog* 2020;16(7):e1008700.
- [3] Pauli B, Oña L, Hermann M, Kost C. Obligate mutualistic cooperation limits evolvability. *Nat Commun* 2022;13(1):337.
- [4] Bottery MJ, Matthews JL, Wood AJ, Johansen HK, Pitchford JW, Friman V-P. Inter-species interactions alter antibiotic efficacy in bacterial communities. *ISME J* 2021;16(3):812–21.
- [5] Mizrahi SP, Goyal A, Gore J. Community interactions drive the evolution of antibiotic tolerance in bacteria. *Proc Natl Acad Sci* 2023;120(3).
- [6] Denk-Lobnig M, Wood KB. Antibiotic resistance in bacterial communities. *Curr Opin Microbiol* 2023;74:102306.
- [7] Wood TK, Knabel SJ, Kwan BW. Bacterial persister cell formation and dormancy. *Appl Environ Microbiol* 2013;79(23):7116–21.
- [8] Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. *Science* 2004;305(5690):1622–5.
- [9] Balaban NQ, Helaine S, Lewis K, Ackermann M, Aldridge B, Andersson DI, et al. Definitions and guidelines for research on antibiotic persistence. *Nat Rev Microbiol* 2019;17:441–8.
- [10] Manuse S, Shan Y, Canas-Duarte SJ, Bakshi S, Sun W-S, Mori H, et al. Bacterial persisters are a stochastically formed subpopulation of low-energy cells. *PLoS Biol* 2021;19(4):e3001194.
- [11] Christensen-Dalsgaard M, Jørgensen MG, Gerdes K. Three new *relE*-homologous *mRNA* interferences of *Escherichia coli* differentially induced by environmental stresses. *Mol Microbiol* 2010;75(2):333–48.
- [12] Vázquez-Laslop N, Lee H, Neyfakh AA. Increased persistence in *Escherichia coli* caused by controlled expression of toxins or other unrelated proteins. *J Bacteriol* 2006;188(10):3494–7.
- [13] Dörr T, Vulić M, Lewis K. Ciprofloxacin causes persister formation by inducing the *tisB* toxin in *Escherichia coli*. *PLoS Biol* 2010;8(2):1–8.
- [14] Fisher RA, Gollan B, Helaine S. Persistent bacterial infections and persister cells. *Nat Rev Microbiol* 2017;15:453–64.
- [15] Huang CY, Gonzalez-Lopez C, Henry C, Mijakovic I, Ryan KR. *hipA* toxin-antitoxin systems mediate persistence in *Caulobacter crescentus*. *Sci Rep* 2020;10(2865).
- [16] Niu H, Gu J, Zhang Y. Bacterial persisters: molecular mechanisms and therapeutic development. *Signal Transduct Targeted Ther* 2024;9(174).
- [17] Dörr T, Lewis K, Vulić M. *Sos* response induces persistence to fluoroquinolones in *Escherichia coli*. *PLoS Genet* 2009;5(12):1–9.
- [18] Levin BR, Concepción-Acevedo J, Udekwi KI. Persistence: a copacetic and parsimonious hypothesis for the existence of non-inherited resistance to antibiotics. *Curr Opin Microbiol* 2014;21:18–21.
- [19] Brauner A, Friedman O, Gefen O, Balaban NQ. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat Rev Microbiol* 2016;14:320–30.
- [20] Podlesek Z, Žgur Bertok D. The *dna* damage inducible *sos* response is a key player in the generation of bacterial persister cells and population wide tolerance. *Front Microbiol* 2020;11(1785).
- [21] Salcedo-Sora JE, Kell DB. A quantitative survey of bacterial persistence in the presence of antibiotics: towards antipersister antimicrobial discovery. *Antibiotics* 2020;9(8):508.
- [22] Mohiuddin SG, Kavousi P, Orman MA. Flow-cytometry analysis reveals persister resuscitation characteristics. *BMC Microbiol* 2020;20:1–13.
- [23] Marques CN, Morozov A, Planzos P, Zelaya HM. The fatty acid signaling molecule *cis*-2-decenoic acid increases metabolic activity and reverts persister cells to an antimicrobial-susceptible state. *Appl Environ Microbiol* 2014;80(22):6976–91.
- [24] Harms A, Maisonneuve E, Gerdes K. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* 2016;354(6318):aaf4268.
- [25] Fang X, Allison KR. Resuscitation dynamics reveal persister partitioning after antibiotic treatment. *Mol Syst Biol* 2023;19(4):e11320.
- [26] Potluri L-P, de Pedro MA, Young KD. *Escherichia coli* low-molecular-weight penicillin-binding proteins help orient septal *ftsZ*, and their absence leads to asymmetric cell division and branching. *Mol Microbiol* 2012;84(2):203–24.
- [27] Raghunathan S, Chimthanawala A, Krishna S, Vecchiarelli AG, Badrinarayanan A. Asymmetric chromosome segregation and cell division in *dna* damage-induced bacterial filaments. *Mol Biol Cell* 2020;31(26):2920–31.
- [28] Jones RM, Nicas M, Hubbard AE, Reingold AL. The infectious dose of *Coxiella burnetii* (Q fever). *Appl Biosafety* 2006;11(1):32–41.
- [29] Hara-Kudo Y, Takatori K. Contamination level and ingestion dose of foodborne pathogens associated with infections. *Epidemiol Infect* 2011;139(10):1505–10.
- [30] Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, et al. Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathog* 2011;7(7):e1002158.
- [31] Alonso AA, Molina I, Theodoropoulos C. Modeling bacterial population growth from stochastic single-cell dynamics. *Appl Environ Microbiol* 2014;80(17):5241–53.
- [32] Teimouri H, Kolomeisky AB. Theoretical investigation of stochastic clearance of bacteria: first-passage analysis. *J R Soc Interface* 2019;16(152):20180765.

- [33] Czuppon P, Day T, Débarre F, Blanquart F. A stochastic analysis of the interplay between antibiotic dose, mode of action, and bacterial competition in the evolution of antibiotic resistance. *PLoS Comput Biol* 2023;19:1–20.
- [34] Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. *Science* 2018;361(6401):469–74.
- [35] Bren A, Glass DS, Kohanim YK, Mayo A, Alon U. Tradeoffs in bacterial physiology determine the efficiency of antibiotic killing. *Proc Natl Acad Sci* 2023;120(51):e2312651120.
- [36] Vega NM, Gore J. Stochastic assembly produces heterogeneous communities in the *Caenorhabditis elegans* intestine. *PLoS Biol* 2017;15(3):e2000633.
- [37] Handwerker S, Raucher B, Altarac D, Monka J, Marchione S, Singh KV, et al. Nosocomial outbreak due to enterococcus faecium highly resistant to vancomycin, penicillin, and gentamicin. *Clin Infect Dis* 1993;16(6):750–5.
- [38] Hu J, Amor DR, Barbier M, Bunin G, Gore J. Emergent phases of ecological diversity and dynamics mapped in microcosms. *Science* 2022;378(6615):85–9.
- [39] Foster KR, Bell T. Competition, not cooperation, dominates interactions among culturable microbial species. *Curr Biol* 2012;22(19):1845–50.
- [40] Ghoul M, Mitri S. The ecology and evolution of microbial competition. *Trends Microbiol* 2016;24(10):833–45.
- [41] Kehe J, Ortiz A, Kulesa A, Gore J, Blainey PC, Friedman J. Positive interactions are common among culturable bacteria. *Sci Adv* 2021;7(45):eabi7159.
- [42] Stein RR, Bucci V, Toussaint NC, Buffie CG, Räscher G, Pamer EG, et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol* 2013;9(12):e1003388.
- [43] Gjini E, Paupério FFS, Ganusov VV. Treatment timing shifts the benefits of short and long antibiotic treatment over infection. *Evol Med Public Health* 2020;2020(1):249–63.
- [44] La Rosa R, Johansen HK, Molin S. Persistent bacterial infections, antibiotic treatment failure, and microbial adaptive evolution. *Antibiotics* 2022;11(3).
- [45] Wainwright J, Hobbs G, Nakout I. Persister cells: formation, resuscitation and combative therapies. *Arch Microbiol* 2021;203.
- [46] Fernandez M, Riveros JD, Campos M, Mathee K, Narasimhan G. Microbial social networks. *BMC Genomics* 2015;16:1–13.
- [47] Rocha-Granados MC, Zenick B, Englander HE, Mok WW. The social network: impact of host and microbial interactions on bacterial antibiotic tolerance and persistence. *Cell Signal* 2020;75:109750.
- [48] Horrocks V, King OG, Yip AYG, Marques IM, McDonald JAK. Role of the gut microbiota in nutrient competition and protection against intestinal pathogen colonization. *Microbiology August* 2023;169(8).
- [49] Maldonado-Gómez MX, Martínez I, Bottacini F, O’Callaghan A, Ventura M, van Sinderen D, et al. Stable engraftment of *Bifidobacterium longum* ah1206 in the human gut depends on individualized features of the resident microbiome. *Cell Host Microbe* 2016;20(4):515–26.
- [50] Kachrimanidou M, Tsintarakis E. Insights into the role of human gut microbiota in *Clostridioides difficile* infection. *Microorganisms* 2020;8(2).
- [51] Lee A, Yoo JS, Yoon E-J. Gut microbiota and new microbiome-targeted drugs for *Clostridioides difficile* infections. *Antibiotics* 2024;13(10). <https://www.mdpi.com/2079-6382/13/10/995>.
- [52] Hromada S, Venturelli OS. Gut microbiota interspecies interactions shape the response of *Clostridioides difficile* to clinically relevant antibiotics. *PLoS Biol* 2023;21(5):1–39.
- [53] Filkins LM, O’Toole GA. Cystic fibrosis lung infections: polymicrobial, complex, and hard to treat. *PLoS Pathog* 2016;11(12):1–8.
- [54] Sitkin S, Lazebnik L, Avalueva E, Kononova S, Vakhtov T. Gastrointestinal microbiome and *Helicobacter pylori*: eradicate, leave it as it is, or take a personalized benefit-risk approach? *World J Gastroenterol* 2022;28(7).
- [55] Fiorani M, Tohumcu E, Del Vecchio LE, Porcari S, Cammarota G, Gasbarrini A, et al. The influence of *Helicobacter pylori* on human gastric and gut microbiota. *Antibiotics* 2023;12(4).
- [56] Gaston JR, Johnson AO, Bair KL, White AN, Armbruster CE. Polymicrobial interactions in the urinary tract: is the enemy of my enemy my friend? *Infect Immun* 2021.
- [57] Keogh D, Tay WH, Ho YY, Dale JL, Chen S, Umashankar S, et al. Enterococcal metabolite cues facilitate interspecies niche modulation and polymicrobial infection. *Cell Host Microbe* 2016;20(4):493–503.
- [58] Amor DR, Ratzke C, Gore J. Transient invaders can induce shifts between alternative stable states of microbial communities. *Sci Adv* 2020;6(8):eaay8676.
- [59] Nguyen J, Lara-Gutiérrez J, Stocker R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol Rev* 2021;45(4):fuaa068.
- [60] Cogan N, Rath H, Kommerein N, Stump S, Stiesch M. Theoretical and experimental evidence for eliminating persister bacteria by manipulating killing timing. *Microbiol Lett* 2016;363(23):fww264.
- [61] McCance W, Surapaneni A, Hampton J, Wang S, Currell M. Decoupling anthropogenic vs. natural impacts at a wastewater treatment plant situated on acid sulfate soils. *Sci Total Environ* 2022;821:153357.
- [62] Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Mannerås-Holm L, et al. Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat Med* 2017;23(7):850–8.
- [63] Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH, et al. Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J* 2014;8(11):2218–30.
- [64] Yang T, Lupwayi N, Marc S-A, Siddique KH, Bainard LD. Anthropogenic drivers of soil microbial communities and impacts on soil biological functions in agroecosystems. *Glob Ecol Conserv* 2021;27:e01521.

Chapter 4

Conclusions and perspectives

We can tell a story about microbial communities by integrating genetic information from bacteria with contextual data, such as where the DNA comes from, when it was retrieved, and their ecological roles, according to literature. All this information is generally derived from the analysis of Next-Generation Sequencing (NGS) data, complemented by metadata and previous studies. While we are confident in our ability to extract meaningful knowledge from these analyses, it has become evident that there is still more to understand, and that further studies require more advanced tools.

4.1 Mathematical modeling

Throughout my doctoral project, I worked extensively with NGS data and noticed there are many unanswered questions that could be addressed through the use of mathematical modeling. These models allow us to make predictive analyses, describe and forecast microbial behaviors and interactions, and thus expand our understanding of microbial ecosystems. In other words, microbial modeling provides a framework to interpret and predict simple and complex interactions within these systems.

A significant aspect of this thesis was the integration of modeling approaches to enhance our understanding of microbial community dynamics. By applying mathematical models, from gLV or Consumer-Resource, to Individual-based or genome-scale metabolic models (GSMM) (see section 3.2), and analyzing interactions at community-level, we can link microbial metabolism with ecological behaviors. These models can provide quantitative or qualitative insights into how microbial communities respond to perturbations, such as shifts in nutrient loads or exposure to antimicrobial agents. Furthermore, dynamic community modeling frameworks, together with deterministic or stochastic setups, facilitate predictions of community stability and resource utilization.

One example is the Generalized Lotka-Volterra (gLV) model. It has shown to be an adaptable and valuable tool in this field (Dedrick et al. 2023). By analyzing the interaction coefficients α_{ij} between the species present, we can determine whether a microbial community is likely to return to its original state after perturbations, such as changes in nutrient availability, temperature shifts, or exposure to antibiotics. As methodologies and computational techniques evolve, the gLV model will remain useful for our understanding of microbial ecosystems, due to its flexibility to include more species and its capability of implementing a stochastic framework (see section 3.2.4), driving innovations in biotechnology and sustainability.

As gLV models, mathematical modeling in general has shown to have the ability to capture complex interactions. This has made it essential across various areas, such as lake ecosystems, human health, food-associated microbiota and, as this thesis shows, wastewater treatment microbial communities.

4.2 Understanding wastewater treatment plants

I had the opportunity to work together with a wastewater treatment plant in Prato, Italy. In the context of the two major plants of G.I.D.A. S.p.A. (Gestione

Impianto Depurazione Acque), we conducted a bi-monthly sampling over two years. We focused on microbial compositions and functions characterization, the tanks' dynamic behaviors, ecological interactions, and potential biotechnological applications.

NGS of the 16S rRNA gene, lead us to the identification of a core microbial community comprising 51 families responsible for key metabolic processes such as nitrification, denitrification, and organic matter degradation. Interestingly, these processes corresponded to the functionality of the specific tanks. The findings revealed that these communities are shaped by both deterministic factors, such as nutrient availability and pH, and stochastic processes, including environmental fluctuations (seasonality).

Our results suggest a complex set of interactions between microbial diversity and WWTP functionality. Stable subcommunities demonstrated resilience to environmental perturbations, highlighting the robustness of wastewater treatment systems. Variable subcommunities, in contrast, exhibited sensitivity to environmental changes, reflecting their role in maintaining adaptability. This duality indicate the importance of microbial diversity in ensuring consistent and efficient wastewater processing, even under fluctuating operational conditions.

This work is a good starting point to explore the broader implications of microbial research in WWTPs since it has the potential to address global challenges such as antimicrobial resistance and resource recovery. For instance, identifying microbial strains capable of phosphorus removal and methane production suggests novel approaches for improving resource recovery processes. Additionally, we could also focus on antimicrobial resistance genes to explore the role of WWTPs as reservoirs and transmission pathways for resistant bacteria, proposing new techniques for monitoring and mitigation strategies.

4.3 Future perspectives

Despite its contributions, this thesis encountered several challenges. The reliance on 16S rRNA sequencing, while powerful, provides limited functional resolution. Future studies could be improved from integrating shotgun metagenomics or transcriptomics for a more comprehensive understanding of microbial activities. Furthermore, to accurately picture the complexity of microbial interactions, we could enhance spatial and temporal resolution in sampling since this could ensure a more representative dataset of community dynamics.

This research opens several promising directions. One of them could be the development of synthetic microbial consortia, informed by the core and variable community structures identified in this study. This could provide targeted solutions for optimizing WWTP performance. Additionally, the application of different modeling frameworks to analyze temporal microbial data may uncover hidden patterns, forecasting ecological behaviors and driving innovations in maintenance and operational efficiency.

In conclusion, this thesis highlights two valuable points: *i)* microbial communities as an essential part of wastewater treatment and their broader environmental significance, and *ii)* mathematical models as a useful tool to gain understanding of different biological scenarios. By bridging experimental and computational approaches, it provides a robust framework for understanding and exploiting microbial dynamics. As global challenges related to water management, pollution, and antimicrobial resistance intensify, the insights gained from this work will play a crucial role in driving sustainable and innovative solutions in microbial ecology and biotechnology.

Bibliography

Falkowski, Paul, Tom Fenchel, and Edward F Delong (2008). "The microbial engines that drive Earth's biogeochemical cycles". In: *Science* 320 (5879), pp. 1034–1039. DOI: [10.1126/science.1153213](https://doi.org/10.1126/science.1153213).

De Filippo, Carlotta et al. (2024). "Gut microbiota drives colon cancer risk associated with diet: a comparative analysis of meat-based and pesco-vegetarian diets". In: *Microbiome* 12.180. DOI: [10.1186/s40168-024-01900-2](https://doi.org/10.1186/s40168-024-01900-2).

Gupta, Ankit, Rasna Gupta, and Ram Lakhan Singh (2017). "Microbes and Environment". In: *Principles and Applications of Environmental Biotechnology for a Sustainable Future*. Ed. by Ram Lakhan Singh. Springer Singapore, pp. 43–84. ISBN: 978-981-10-1866-4. DOI: [10.1007/978-981-10-1866-4_3](https://doi.org/10.1007/978-981-10-1866-4_3).

Fondi, Marco et al. (2015). "Genome-scale metabolic reconstruction and constraint-based modelling of the Antarctic bacterium *seudoalteromonas haloplanktis* TAC125". In: *Environmental Microbiology* 17.3, pp. 751–766. DOI: [10.1111/1462-2920.12513](https://doi.org/10.1111/1462-2920.12513).

Wilkins, David et al. (2013). "Key microbial drivers in Antarctic aquatic environments". In: *FEMS Microbiology Reviews* 37.3, pp. 303–335. DOI: [10.1111/1574-6976.12007](https://doi.org/10.1111/1574-6976.12007).

Qi, Weihong et al. (2019). "New insights on *Pseudoalteromonas haloplanktis* TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies". In: *Scientific Reports* 9.16444. DOI: [10.1038/s41598-019-52832-z](https://doi.org/10.1038/s41598-019-52832-z).

- Czepiel, Jacek et al. (2019). "Clostridium difficile infection: review". In: *European Journal of Clinical Microbiology & Infectious Diseases* 38.7, pp. 1211–1221. DOI: [10.1007/s10096-019-03539-6](https://doi.org/10.1007/s10096-019-03539-6).
- Markantonis, John E., John T. Fallon, Rajat Madan, and Md Zahidul Alam (2024). "Clostridioides difficile Infection: Diagnosis and Treatment Challenges". In: *Pathogens* 13.2. ISSN: 2076-0817. DOI: [10.3390/pathogens13020118](https://doi.org/10.3390/pathogens13020118).
- Staley, Christopher et al. (2017). "Community dynamics drive punctuated engraftment of the fecal microbiome following transplantation using freeze-dried, encapsulated fecal microbiota". In: *Gut Microbes* 8.3, pp. 276–288. DOI: [10.1080/19490976.2017.1299310](https://doi.org/10.1080/19490976.2017.1299310).
- Qian, Yili, Freeman Lan, and Ophelia S Venturelli (2021). "Towards a deeper understanding of microbial communities: integrating experimental data with dynamic models". In: *Current Opinion in Microbiology* 62, pp. 84–92. ISSN: 1369-5274. DOI: [10.1016/j.mib.2021.05.003](https://doi.org/10.1016/j.mib.2021.05.003).
- Drew, Georgia C., Emily J. Stevens, and Keyla C. King (2021). "Microbial evolution and transitions along the parasite–mutualist continuum". In: *Nature Reviews Microbiology* 19, pp. 623–638. DOI: [10.1038/s41579-021-00550-7](https://doi.org/10.1038/s41579-021-00550-7).
- Pérez-Alvarado, Omar et al. (2022). "Role of lactic acid bacteria and yeasts in sourdough fermentation during breadmaking: Evaluation of postbiotic-like components and health benefits". In: *Frontiers in Microbiology* 13. DOI: [10.3389/fmicb.2022.969460](https://doi.org/10.3389/fmicb.2022.969460).
- Vaccaro, Francesca and Alessio Mengoni (2023). "Legume-Rhizobium Symbiosis and Beyond: Producing Synthetic Communities for Increasing Crop Production Under Climate Change Challenges". In: *Bioinoculants: Biological Option for Mitigating global Climate Change*. Ed. by Surender Singh, Radha Prasanna, and Kumar Pranaw. Singapore: Springer Nature Singapore, pp. 147–162. ISBN: 978-981-99-2973-3. DOI: [10.1007/978-981-99-2973-3_6](https://doi.org/10.1007/978-981-99-2973-3_6).
- Ramanan, Rishiram, Byung-Hyuk Kim, Dae-Hyun Cho, Hee-Mock Oh, and Hee-Sik Kim (2016). "Algae–bacteria interactions: Evolution, ecology and emerging

- applications". In: *Biotechnology Advances* 34.1, pp. 14–29. ISSN: 0734-9750. DOI: <https://doi.org/10.1016/j.biotechadv.2015.12.003>.
- Valiei, Amin, Andrew Dickson, Javad Aminian-Dehkordi, and Mohammad R. K. Mofrad (2024). "Metabolic interactions shape emergent biofilm structures in a conceptual model of gut mucosal bacterial communities". In: *npj Biofilms Microbiomes* 10.99. DOI: [10.1038/s41522-024-00572-y](https://doi.org/10.1038/s41522-024-00572-y).
- Daffé, M. and G. Etienne (1999). "The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity". In: *Tubercle and Lung Disease* 79.3, pp. 153–169. ISSN: 0962-8479. DOI: [10.1054/tuld.1998.0200](https://doi.org/10.1054/tuld.1998.0200).
- Sengupta, Shatabdi et al. (2023). "Chapter 2 - Modulation of host pathways by *Mycobacterium tuberculosis* for survival". In: *Bacterial Survival in the Hostile Environment*. Ed. by Ashutosh Kumar and Shivendra Tenguria. Developments in Applied Microbiology and Biotechnology. Academic Press, pp. 15–33. ISBN: 978-0-323-91806-0. DOI: [10.1016/B978-0-323-91806-0.00003-5](https://doi.org/10.1016/B978-0-323-91806-0.00003-5).
- Vardanyan, R.S. and V.J. Hruby (2006). "32 - Antibiotics". In: *Synthesis of Essential Drugs*. Ed. by R.S. Vardanyan and V.J. Hruby. Amsterdam: Elsevier, pp. 425–498. ISBN: 978-0-444-52166-8. DOI: <https://doi.org/10.1016/B978-044452166-8/50032-7>.
- Goldford, Joshua E. et al. (2018). "Emergent simplicity in microbial community assembly". In: *Science* 361, pp. 469–474. DOI: [10.1126/science.aat1168](https://doi.org/10.1126/science.aat1168).
- Gibbs, Theo, Yifan Zhang, Zachary R. Miller, and James P. O'Dwyer (Sept. 2022). "Stability criteria for the consumption and exchange of essential resources". In: *PLOS Computational Biology* 18.9, pp. 1–20. DOI: [10.1371/journal.pcbi.1010521](https://doi.org/10.1371/journal.pcbi.1010521).
- Leinweber, Anne, R Fredrik Inglis, and Rolf Kümmerli (2017). "Cheating fosters species co-existence in well-mixed bacterial communities". In: *The ISME Journal* 11. DOI: doi.org/10.1038/ismej.2016.195.

- Smith, Parker and Martin Schuster (2019). "Public goods and cheating in microbes". In: *Current Biology* 29.11, R442–R447. ISSN: 0960-9822. DOI: [10.1016/j.cub.2019.03.001](https://doi.org/10.1016/j.cub.2019.03.001).
- Xenophontos, Constantinos, W. Stanley Harpole, Kirsten Küsel, and Adam Thomas Clark (2022). "Cheating Promotes Coexistence in a Two-Species One-Substrate Culture Model". In: *Frontiers in Ecology and Evolution* 9. ISSN: 2296-701X. DOI: [10.3389/fevo.2021.786006](https://doi.org/10.3389/fevo.2021.786006).
- Lovering, Andrew L. and R. Elizabeth Sockett (2021). "Microbe Profile: Bdellovibrio bacteriovorus: a specialized bacterial predator of bacteria". In: *Microbiology* 167.4. ISSN: 1465-2080. DOI: [10.1099/mic.0.001043](https://doi.org/10.1099/mic.0.001043).
- Mougi, Akihiko (2016). "The roles of amensalistic and commensalistic interactions in large ecological network stability". In: *Scientific Reports volume* 6.29929. DOI: [10.1038/srep29929](https://doi.org/10.1038/srep29929).
- Culp, Elizabeth J. and Andrew L. Goodman (2023). "Cross-feeding in the gut microbiome: Ecology and mechanisms". In: *Cell Host & Microbe* 31.4, pp. 485–499. ISSN: 1931-3128. DOI: [10.1016/j.chom.2023.03.016](https://doi.org/10.1016/j.chom.2023.03.016).
- Zachar, István and Gergely Boza (2022). "The Evolution of Microbial Facilitation: Sociogenesis, Symbiogenesis, and Transition in Individuality". In: *Frontiers in Ecology and Evolution* 10. ISSN: 2296-701X. DOI: [10.3389/fevo.2022.798045](https://doi.org/10.3389/fevo.2022.798045).
- Graham, Emily B. et al. (2016). "Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes?" In: *Frontiers in Microbiology* 7. DOI: [10.3389/fmicb.2016.00214](https://doi.org/10.3389/fmicb.2016.00214).
- Daniels, Michael, Simon van Vliet, and Martin Ackermann (2023). "Changes in interactions over ecological time scales influence single-cell growth dynamics in a metabolically coupled marine microbial community". In: *ISME* 17, pp. 406–416. DOI: [10.1038/s41396-022-01312-w](https://doi.org/10.1038/s41396-022-01312-w).

- van den Berg, Naomi I. et al. (2022). "Ecological modelling approaches for predicting emergent properties in microbial communities". In: *Nature Ecology & Evolution* 6, pp. 855–865. DOI: [10.1038/s41559-022-01746-7](https://doi.org/10.1038/s41559-022-01746-7).
- Faust, Karoline, Leo Lahti, Didier Gonze, Willem M de Vos, and Jeroen Raes (2015). "Metagenomics meets time series analysis: unraveling microbial community dynamics". In: *Current Opinion in Microbiology* 25, pp. 56–66. ISSN: 1369-5274. DOI: [10.1016/j.mib.2015.04.004](https://doi.org/10.1016/j.mib.2015.04.004).
- Gonze, Didier, Katharine Z Coyte, Leo Lahti, and Karoline Faust (2018). "Microbial communities as dynamical systems". In: *Current Opinion in Microbiology* 44, pp. 41–49. ISSN: 1369-5274. DOI: [10.1016/j.mib.2018.07.004](https://doi.org/10.1016/j.mib.2018.07.004).
- Lopes, William, Daniel R. Amor, and Jeff Gore (2024). "Cooperative growth in microbial communities is a driver of multistability". In: *Nature Communications* 15, p. 4709. DOI: [10.1038/s41467-024-48521-9](https://doi.org/10.1038/s41467-024-48521-9).
- Eng, Alexander and Elhanan Borenstein (2019). "Microbial community design: methods, applications, and opportunities". In: *Current Opinion in Biotechnology* 58, pp. 117–128. DOI: <https://doi.org/10.1016/j.copbio.2019.03.002>.
- Autorino, Camilla and Nicoletta I. Petridou (2022). "Critical phenomena in embryonic organization". In: *Current Opinion in Systems Biology* 31, p. 100433. ISSN: 2452-3100. DOI: [10.1016/j.coisb.2022.100433](https://doi.org/10.1016/j.coisb.2022.100433).
- Gerber, Georg K. (2014). "The dynamic microbiome". In: *FEBS Letters* 588.22. The gut microbiome, pp. 4131–4139. ISSN: 0014-5793. DOI: [10.1016/j.febslet.2014.02.037](https://doi.org/10.1016/j.febslet.2014.02.037).
- Efendiev, M., S. Zelik, and A. Miranville (2006). "Dissipative Dynamical Systems of Infinite Dimension". In: *Encyclopedia of Mathematical Physics*. Ed. by Jean-Pierre Francoise, Gregory L. Naber, and Tsou Sheung Tsun. Oxford: Academic Press, pp. 101–110. ISBN: 978-0-12-512666-3. DOI: [10.1016/B0-12-512666-2/00095-X](https://doi.org/10.1016/B0-12-512666-2/00095-X).

- Gause, G. F. (1935). "Experimental Demonstration of Volterra's Periodic Oscillations in the Numbers of Animals". In: *Journal of Experimental Biology* 12. DOI: [10.1242/jeb.12.1.44](https://doi.org/10.1242/jeb.12.1.44).
- Stein, Richard R. et al. (Dec. 2013). "Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota". In: *PLOS Computational Biology* 9. DOI: [10.1371/journal.pcbi.1003388](https://doi.org/10.1371/journal.pcbi.1003388).
- Wein, Tanita et al. (2018). "Carrying Capacity and Colonization Dynamics of *Curvibacter* in the Hydra Host Habitat". In: *Frontiers in Microbiology* 9. DOI: [10.3389/fmicb.2018.00443](https://doi.org/10.3389/fmicb.2018.00443).
- Pajon, Camryn et al. (2023). "Interactions between metabolism and growth can determine the co-existence of *Staphylococcus aureus* and *Pseudomonas aeruginosa*". In: *eLife* 12. Ed. by Vaughn S Cooper and Aleksandra M Walczak. ISSN: 2050-084X. DOI: [10.7554/eLife.83664](https://doi.org/10.7554/eLife.83664).
- Lima, Laís O. et al. (Mar. 2020). "Modeling of the Coral Microbiome: the Influence of Temperature and Microbial Network". In: *mBio* 11.2. DOI: [10.1128/mbio.02691-19](https://doi.org/10.1128/mbio.02691-19).
- Alonso-Vásquez, Tania, Michele Giovannini, Gian Luigi Garbini, et al. (2024). "An ecological and stochastic perspective on persisters resuscitation". In: *Computational and Structural Biotechnology Journal* 27, pp. 1–9. ISSN: 2001-0370. DOI: [10.1016/j.csbj.2024.12.002](https://doi.org/10.1016/j.csbj.2024.12.002).
- Bengtsson-Palme, Johan (2020). "Microbial model communities: To understand complexity, harness the power of simplicity". In: *Computational and Structural Biotechnology Journal* 18, pp. 3987–4001. ISSN: 2001-0370. DOI: [10.1016/j.csbj.2020.11.043](https://doi.org/10.1016/j.csbj.2020.11.043).
- MacArthur, Robert (1970). "Species packing and competitive equilibrium for many species". In: *Theoretical Population Biology* 1.1, pp. 1–11. ISSN: 0040-5809. DOI: [10.1016/0040-5809\(70\)90039-0](https://doi.org/10.1016/0040-5809(70)90039-0).

- Gowda, Karna, Derek Ping, Madhav Mani, and Seppe Kuehn (2022). "Genomic structure predicts metabolite dynamics in microbial communities". In: *Cell* 185.3, 530–546.e25. ISSN: 0092-8674. DOI: [10.1016/j.cell.2021.12.036](https://doi.org/10.1016/j.cell.2021.12.036).
- Marsland III, Robert et al. (Feb. 2019). "Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities". In: *PLOS Computational Biology* 15.2, pp. 1–18. DOI: [10.1371/journal.pcbi.1006793](https://doi.org/10.1371/journal.pcbi.1006793).
- Mancuso, Christopher P, Hyunseok Lee, Clare I Abreu, Jeff Gore, and Ahmad S Khalil (Sept. 2021). "Environmental fluctuations reshape an unexpected diversity-disturbance relationship in a microbial community". In: *eLife* 10, e67175. DOI: [10.7554/eLife.67175](https://doi.org/10.7554/eLife.67175).
- Zakharova, L., K.M. Meyer, and M. Seifan (2019). "Trait-based modelling in ecology: A review of two decades of research". In: *Ecological Modelling* 407, p. 108703. ISSN: 0304-3800. DOI: [10.1016/j.ecolmodel.2019.05.008](https://doi.org/10.1016/j.ecolmodel.2019.05.008).
- (2020). "Combining trait- and individual-based modelling to understand desert plant community dynamics". In: *Ecological Modelling* 434, p. 109260. ISSN: 0304-3800. DOI: [10.1016/j.ecolmodel.2020.109260](https://doi.org/10.1016/j.ecolmodel.2020.109260).
- Lajoie, Geneviève and Steven W. Kembel (2019). "Making the Most of Trait-Based Approaches for Microbial Ecology". In: *Trends in Microbiology* 27.10, pp. 814–823. ISSN: 0966-842X. DOI: [10.1016/j.tim.2019.06.003](https://doi.org/10.1016/j.tim.2019.06.003).
- Grigoratou, M. et al. (2019). "A trait-based modelling approach to planktonic foraminifera ecology". In: *Biogeosciences* 16.7, pp. 1469–1492. DOI: [10.5194/bg-16-1469-2019](https://doi.org/10.5194/bg-16-1469-2019).
- Jørgensen, Sven Erik and Brian D. Fath (2011). "9 - Individual-Based Models". In: *Fundamentals of Ecological Modelling*. Ed. by Sven Erik Jørgensen and Brian D. Fath. Vol. 23. Developments in Environmental Modelling. Elsevier, pp. 291–308. DOI: [10.1016/B978-0-444-53567-2.00009-0](https://doi.org/10.1016/B978-0-444-53567-2.00009-0).

- Jayathilake, Pahala Gedara et al. (Aug. 2017). "A mechanistic Individual-based Model of microbial communities". In: *PLOS ONE* 12.8, pp. 1–26. DOI: [10.1371/journal.pone.0181965](https://doi.org/10.1371/journal.pone.0181965).
- Breckling, Broder (2002). "Individual-Based Modelling Potentials and Limitations". In: *The Scientific World Journal* 2.1, p. 684985. DOI: <https://doi.org/10.1100/tsw.2002.179>.
- Oberhardt, Matthew A, Bernhard Ø Palsson, and Jason A Papin (2009). "Applications of genome-scale metabolic reconstructions". In: *Molecular Systems Biology* 5.1, p. 320. DOI: [10.1038/msb.2009.77](https://doi.org/10.1038/msb.2009.77).
- Baart, Gino J. E. and Dirk E. Martens (2012). "Genome-Scale Metabolic Models: Reconstruction and Analysis". In: *Neisseria meningitidis: Advanced Methods and Protocols*. Ed. by Myron Christodoulides. Totowa, NJ: Humana Press, pp. 107–126. DOI: [10.1007/978-1-61779-346-2_7](https://doi.org/10.1007/978-1-61779-346-2_7).
- O'Brien, Edward J., Jonathan M. Monk, and Bernhard O. Palsson (2015). "Using Genome-scale Models to Predict Biological Capabilities". In: *Cell* 161.5, pp. 971–987. ISSN: 0092-8674. DOI: [10.1016/j.cell.2015.05.019](https://doi.org/10.1016/j.cell.2015.05.019).
- Orth, Jeffrey D, Ines Thiele, and Bernhard Palsson (2010). "What is flux balance analysis?" In: *Nature Biotechnology* 28.3, pp. 245–248. DOI: [10.1038/nbt.1614](https://doi.org/10.1038/nbt.1614).
- Kim, Minsuk, Jaeyun Sung, and Nicholas Chia (2022). "Resource-allocation constraint governs structure and function of microbial communities in metabolic modeling". In: *Metabolic Engineering* 70, pp. 12–22. ISSN: 1096-7176. DOI: [10.1016/j.ymben.2021.12.011](https://doi.org/10.1016/j.ymben.2021.12.011).
- Muller, Emilie E.L. et al. (2018). "Using metabolic networks to resolve ecological properties of microbiomes". In: *Current Opinion in Systems Biology* 8, pp. 73–80. ISSN: 2452-3100. DOI: [10.1016/j.coisb.2017.12.004](https://doi.org/10.1016/j.coisb.2017.12.004).
- Tarzi, Chaimaa, Guido Zampieri, Neil Sullivan, and Claudio Angione (2024). "Emerging methods for genome-scale metabolic modeling of microbial communities". In: *Trends in Endocrinology & Metabolism* 35.

- Coyte, Katharine Z., Jonas Schluter, and Kevin R. Foster (2015). "The ecology of the microbiome: Networks, competition, and stability." In: *Science* 350.6261, pp. 663–6. DOI: [10.1126/science.aad2602](https://doi.org/10.1126/science.aad2602).
- Machado, Daniel et al. (2021). "Polarization of microbial communities between competitive and cooperative metabolism." In: *Nature ecology & evolution* 5.2. DOI: [10.1038/s41559-020-01353-4](https://doi.org/10.1038/s41559-020-01353-4).
- Xu, Li, Denis Patterson, Ann Carla Staver, Simon Asher Levin, and Jin Wang (2021). "Unifying deterministic and stochastic ecological dynamics via a landscape-flux approach." In: *PNAS Proceedings of the National Academy of Sciences of the United States of America* 118.24. DOI: [10.1073/pnas.2103779118](https://doi.org/10.1073/pnas.2103779118).
- Cota, Wesley and Silvio C. Ferreira (2017). "Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks". In: *Computer Physics Communications* 219, pp. 303–312. ISSN: 0010-4655. DOI: [10.1016/j.cpc.2017.06.007](https://doi.org/10.1016/j.cpc.2017.06.007).
- Black, Andrew J. and Alan J. McKane (2012). "Stochastic formulation of ecological models and their applications". In: *Trends in Ecology & Evolution* 27.6, pp. 337–345. ISSN: 0169-5347. DOI: [10.1016/j.tree.2012.01.014](https://doi.org/10.1016/j.tree.2012.01.014).
- Song, Yuqian et al. (2021). "Different deterministic versus stochastic drivers for the composition and structure of a temperate grassland community". In: *Global Ecology and Conservation* 31, e01866. ISSN: 2351-9894. DOI: [10.1016/j.gecco.2021.e01866](https://doi.org/10.1016/j.gecco.2021.e01866).
- He, Xuelian, Calum Brown, and Luxiang Lin (2021). "Relative importance of deterministic and stochastic processes for beta diversity of bird assemblages in Yunnan, China". In: *Ecosphere* 12.6. DOI: [10.1002/ecs2.3545](https://doi.org/10.1002/ecs2.3545).
- Moura, Leonardo and Renata C. Picão (2022). "Chapter 23 - Removal of antimicrobial resistance determinants from wastewater: a risk perspective on conventional and emerging technologies". In: *Emerging Contaminants in the Environment*. Ed. by Hemen Sarma, Delfina C. Dominguez, and Wen-Yee Lee. El-

- sevier, pp. 603–642. ISBN: 978-0-323-85160-2. DOI: [10.1016/B978-0-323-85160-2.00023-8](https://doi.org/10.1016/B978-0-323-85160-2.00023-8).
- Bassin, João Paulo et al. (2021). “Chapter 16 - The impact of wastewater treatment plants on global climate change”. In: *Water Conservation in the Era of Global Climate Change*. Ed. by Binota Thokchom, Pengpeng Qiu, Pardeep Singh, and Parameswar K. Iyer. Elsevier, pp. 367–410. ISBN: 978-0-12-820200-5. DOI: [10.1016/B978-0-12-820200-5.00001-4](https://doi.org/10.1016/B978-0-12-820200-5.00001-4).
- Bhambhani, Anurag, Oriana Jovanovic, Arjen van Nieuwenhuijzen, Jan Peter van der Hoek, and Zoran Kapelan (2024). “Introducing a new method to assess the benefits of resources recovered from wastewater to the natural environment”. In: *Sustainable Production and Consumption* 46, pp. 559–570. ISSN: 2352-5509. DOI: [10.1016/j.spc.2024.03.016](https://doi.org/10.1016/j.spc.2024.03.016).
- Kolpakova, Valentina et al. (2024). “Design and construction of wastewater treatment facilities for small sewerage facilities”. In: *Case Studies in Chemical and Environmental Engineering* 9, p. 100774. ISSN: 2666-0164. DOI: [10.1016/j.cscee.2024.100774](https://doi.org/10.1016/j.cscee.2024.100774).
- Elginöz, Nilay, Muhammed Alzaboot, Fatos Germirli Babuna, and Gulen Iskender (2019). “Construction of a large water treatment plant: appraisal of environmental hotspots”. In: *Desalination and Water Treatment* 172, pp. 309–315. ISSN: 1944-3986. DOI: [10.5004/dwt.2019.25107](https://doi.org/10.5004/dwt.2019.25107).
- Zagklis, Dimitris P. and Georgios Bamos (2022). “Tertiary Wastewater Treatment Technologies: A Review of Technical, Economic, and Life Cycle Aspects”. In: *Processes* 10.11. DOI: [10.3390/pr10112304](https://doi.org/10.3390/pr10112304).
- Alasino, Noelia, Miguel C. Mussati, and Nicolás Scenna (2007). “Wastewater Treatment Plant Synthesis and Design”. In: *Industrial & Engineering Chemistry Research* 46. DOI: [10.1021/ie0704905](https://doi.org/10.1021/ie0704905).
- Greater Chicago, Metropolitan Water Reclamation District of (n.d.). *Stickney Water Reclamation Plant*. <https://mwrdd.org/locations/stickney-water-reclamation-plant>. Accessed.

- Hao, Xiaodi et al. (2019). "Environmental impacts of resource recovery from wastewater treatment plants". In: *Water Research* 160, pp. 268–277. ISSN: 0043-1354. DOI: [10.1016/j.watres.2019.05.068](https://doi.org/10.1016/j.watres.2019.05.068).
- Su, Du, Weiwei Ben, Bjarne W. Strobel, and Zhimin Qiang (2021). "Impacts of wastewater treatment plant upgrades on the distribution and risks of pharmaceuticals in receiving rivers". In: *Journal of Hazardous Materials* 406, p. 124331. ISSN: 0304-3894. DOI: [10.1016/j.jhazmat.2020.124331](https://doi.org/10.1016/j.jhazmat.2020.124331).
- Hospido, Almudena, Ma Teresa Moreira, and Gumersindo Feijoo (2008). "A comparison of municipal wastewater treatment plants for big centres of population in Galicia (Spain)". In: *The International Journal of Life Cycle Assessment* 13. DOI: [10.1065/lca2007.03.314](https://doi.org/10.1065/lca2007.03.314).
- Al-Asheh, Sameer, Marzieh Bagheri, and Ahmed Aidan (2021). "Membrane bioreactor for wastewater treatment: A review". In: *Case Studies in Chemical and Environmental Engineering* 4, p. 100109. ISSN: 2666-0164. DOI: [10.1016/j.cscee.2021.100109](https://doi.org/10.1016/j.cscee.2021.100109).
- Meerbergen, Ken et al. (2017). "Assessing the composition of microbial communities in textile wastewater treatment plants in comparison with municipal wastewater treatment plants". In: *MicrobiologyOpen* 6.1. DOI: [10.1002/mbio3.413](https://doi.org/10.1002/mbio3.413).
- Xia, Yu, Xianghua Wen, Bing Zhang, and Yunfeng Yang (2018). "Diversity and assembly patterns of activated sludge microbial communities: A review". In: *Biotechnology Advances* 36.4, pp. 1038–1047. ISSN: 0734-9750. DOI: [10.1016/j.biotechadv.2018.03.005](https://doi.org/10.1016/j.biotechadv.2018.03.005).
- Lehtovirta-Morley, Laura E (2018). "Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together". In: *FEMS Microbiology Letters* 365 (9). DOI: [10.1093/femsle/fny058](https://doi.org/10.1093/femsle/fny058).
- Alonso-Vásquez, Tania, Michele Giovannini, Camilla Fagorzi, et al. (2024). "Dynamics and Core Microbial Community Functions in a WWTP". In: *Unpublished manuscript*.

- Alonso-Vásquez, Tania, Camilla Fagorzi, et al. (2024). "Metagenomic Surveys Shows a Widespread Diffusion of Antibiotic Resistance Genes in a Transect from Urbanized to Marine Protected Area". In: *SSRN*. DOI: [10.2139/ssrn.4841579](https://doi.org/10.2139/ssrn.4841579).
- Semenzato, Giulia et al. (2022). "Genomic Analysis of Endophytic Bacillus-Related Strains Isolated from the Medicinal Plant *Origanum vulgare* L. Revealed the Presence of Metabolic Pathways Involved in the Biosynthesis of Bioactive Compounds". In: *Microorganisms* 10.5. DOI: [10.3390/microorganisms10050919](https://doi.org/10.3390/microorganisms10050919).
- Alonso-Vásquez, Tania, Marco Fondi, and Elena Perrin (2023). "Understanding Antimicrobial Resistance Using Genome-Scale Metabolic Modeling". In: *Antibiotics* 12.5. ISSN: 2079-6382. DOI: [10.3390/antibiotics12050896](https://doi.org/10.3390/antibiotics12050896).
- Riccardi, Christopher et al. (2023). "Metabolic Robustness to Growth Temperature of a Cold- Adapted Marine Bacterium." In: *mSystems* 8.2. DOI: [10.1128/msystems.01124-22](https://doi.org/10.1128/msystems.01124-22).
- Ghini, Veronica et al. (Nov. 2023). "COVID-19: A complex disease with a unique metabolic signature". In: *PLOS Pathogens* 19.11, pp. 1–26. DOI: [10.1371/journal.ppat.1011787](https://doi.org/10.1371/journal.ppat.1011787).
- Mohiuddin, Sayed Golam, Pouria Kavousi, and Mehmet A. Orman (2020). "Flow-cytometry analysis reveals persister resuscitation characteristics". In: *BMC Microbiology* 20.202. DOI: [10.1186/s12866-020-01888-3](https://doi.org/10.1186/s12866-020-01888-3).
- Jones, Rachael M., Mark Nicas, Alan E. Hubbard, and Arthur L. Reingold (2006). "The Infectious Dose of *Coxiella Burnetii* (Q Fever)". In: *Applied Biosafety* 11.1, pp. 32–41. DOI: [10.1177/153567600601100106](https://doi.org/10.1177/153567600601100106).
- Hara-Kudo, Y. and K. Takatori (2011). "Contamination level and ingestion dose of foodborne pathogens associated with infections". In: *Epidemiology and Infection* 139.10, pp. 1505–1510. DOI: [10.1017/S095026881000292X](https://doi.org/10.1017/S095026881000292X).
- Gullberg, Erik et al. (July 2011). "Selection of Resistant Bacteria at Very Low Antibiotic Concentrations". In: *PLOS Pathogens* 7.7, pp. 1–9. DOI: [10.1371/journal.ppat.1002158](https://doi.org/10.1371/journal.ppat.1002158).

- Erban, Radek and S. Jonathan Chapman (2020). *Stochastic Modelling of Reaction–Diffusion Processes*. Cambridge Texts in Applied Mathematics. Cambridge University Press.
- Dedrick, Sandra, Vaishnavi Warriar, Katherine P. Lemon, and Babak Momeni (2023). “When does a Lotka-Volterra model represent microbial interactions? Insights from in vitro nasal bacterial communities”. In: *mSystems* 8.3. DOI: [10.1128/msystems.00757-22](https://doi.org/10.1128/msystems.00757-22).

Chapter 5

Overall activities

5.1 List of publications

- Alonso-Vásquez, T. et al. Dynamics and Core Microbial Community Functions in a WWTP. Unpublished manuscript (2024). (Page [48](#))
- Alonso-Vásquez, T. et al. Metagenomic Surveys Shows a Widespread Diffusion of Antibiotic Resistance Genes in a Transect from Urbanized to Marine Protected Area. Submitted paper (2024). (Page [69](#))
- Semenzato, G., Alonso-Vásquez, T. et al. Genomic Analysis of Endophytic Bacillus-Related Strains Isolated from the Medicinal Plant *Origanum vulgare* L. Revealed the Presence of Metabolic Pathways Involved in the Biosynthesis of Bioactive Compounds. *Microorganisms* 10 (2022). (Page [105](#))
- Alonso-Vásquez, T., Fondi, M. & Perrin, E. Understanding Antimicrobial Resistance Using Genome-Scale Metabolic Modeling. *Antibiotics* 12 (2023) (Page [117](#))
- Riccardi, C., Calvanese, M., Ghini, V., Alonso-Vásquez, T. et al. Metabolic Robustness to Growth Temperature of a Cold-Adapted Marine Bacterium. *mSystems* 8 (2023). (Page [138](#))

- Ghini, V., Vieri, W., Celli, T., Pecchioli, V., Boccia, N., Alonso-Vásquez, T. et al. COVID-19: A complex disease with a unique metabolic signature. *PLOS Pathogens* 19, 1–26 (2023) (Page [157](#))
- Alonso-Vásquez, T. et al. An ecological and stochastic perspective on persister resuscitation. *Computational and Structural Biotechnology Journal* (2024). (Page [185](#))

5.2 Period of study and research abroad

- Internship at Systems Biology Lab. Vrije Universiteit Amsterdam.
Amsterdam, Netherlands
Prof. Matti Gralka
From 01/04/2023 to 02/10/2023

5.3 Period of study and research in an enterprise

- Internship at GIDA S.p.A. *Gestione Impianti Depurazione Acque S.p.A.*
Prato, Italy.
During the months of 10-12/2023, and 4-6,9-10/2024

5.4 Congresses, seminars and workshops

- Eleventh Edition of the Tuscany Day of Bioinformatics and Systems Biology BIOINFORMATIHA 11 Florence, Italy (14/10/2024).
Workshop: "Computational models of biological systems (dynamic and constraint based models)"
- Annual Meeting of the Society for Molecular Biology and Evolution SMBE 2024 Puerto Vallarta, Mexico (07-11/07/2024).
Poster presentation.

- Cortona Procarioti 2024 Cortona, Italy (27-29/06/2024).
Talk: "A perspective on how to model microbial communities dynamics"
- 20th Annual Meeting of the Bioinformatics Italian Society BITS. Trento, Italy (14-14/06/2024).
Talk: "An ecological perspective on persisters resuscitation"
- Microbiology 2023 - XXXIV SIMGBM Congress Cagliari, Italy (21-24/09/2024).
Poster presentation.
- Annual Meeting of the Society for Molecular Biology and Evolution SMBE 2023 Ferrara, Italy (23-27/07/2023).
Poster presentation.
- Seminar at University of Gdańsk, Poland Gdańsk, Poland (16/11/2022).
Seminar: "Metabolic Networks: Constrained based models construction"
- Ninth Edition of the Tuscany Day of Bioinformatics and Systems Biology BIOINFORMATIHA 9 Pisa, Italy (16/09/2022).
Talk: "Modeling the metabolic consequences of antimicrobial exposure in *Escherichia coli*"
- International Union of Microbiological Societies Congress (IUMS 2022 - On-line Edition) Rotterdam, Netherlands (20-22/07/2022).
Poster presentation.
- FEMS Conference on Microbiology 2022 Belgrade, Serbia (30/06/2024 - 02/07/2024).
Poster presentation.
- Cortona Procarioti 2022 Cortona, Italy (23-25/06/2022).
Talk: "Modelling the metabolic consequences of antimicrobial exposure"
- ELIXIR Workshop on Metabolomics and Integrative Omics Bari, Italy (04/2022).

5.5 Awards

- Best Young Speaker Award Awarding institution: Bioinformatics Italian Society. "An ecological perspective on persisters resuscitation" at the 20th Annual Meeting of the Bioinformatics Italian Society (BITS).
Trento, Italy. 14/06/2024.



UNIONE EUROPEA
Fondo Sociale Europeo



*Ministero dell'Università
e della Ricerca*



PON
RICERCA
E INNOVAZIONE
2014 - 2020

REACT EU

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020, risorse FSE REACT-EU
Azione IV.4 "Dottorati e contratti di ricerca su tematiche dell'innovazione"
e Azione IV.5 "Dottorati su tematiche Green"