Integration of host, pathogen and microbiome -omics data for studying infectious diseases

Dissertation

in Partial Fulfilment of the Requirements for the Degree of

“Doctor of Philosophy” (PhD)

Submitted to the Council of the Faculty of Biological Sciences

of Friedrich Schiller University Jena

by Mohammadhassan Mirhakkak Esfahani

born on 10. April 1990 in Esfahan, Iran

Reviewers:

1. Prof. Dr. Gianni Panagiotou (Hans-Knoell-Institute Jena)
2. Prof. Dr. Michael Bauer (Jena University Hospital)
3. Prof. Dr. Adil Mardinoglu (King’s College London)

The defense of the doctoral thesis took place on November 25, 2022 in Jena.

Table of Contents

INTRODUCTION

1. **Infectious diseases**

Infection is the process of invasion of an organism with a disease-causing effect on the host. It is threatening human health for centuries and remains a serious global problem until this day. Early scripts describe the Black Death spreading in Europe in the 1300s. Not until many years after, the germ theory of disease (i.e., certain diseases are caused by the invasion of the body by microorganisms) was born in the 1500s. The scientific progress in the recognition of infectious diseases resulted in remarkable findings. In the nineteenth century, Sir John Scott Burden-Sanderson realized that bacteria cannot survive in a liquid growth culture covered by mold. In 1871, Joseph Lister discovered the negative influences of *Penicillium glaucum* on bacterial growth. In this regard, the word “antibiosis” was defined by Jean Paul Vuillemin in 1889. This referred to any biological relationship in which one living organism kills another to ensure its survival. For instance, the Penicillin discovery by Alexander Fleming in 1928 was one of the most remarkable findings in this area (Durand et al., 2019). Later in the middle of the 1900s, during the second world war, Hans Knoell achieved the penicillin production at a laboratory scale in Germany.

At the beginning of the twentieth century, infectious diseases caused a high mortality rate in the United States (Armstrong et al., 1999). However, it gradually decreased due to influencing factors such as better nutrition, better housing, improved hygiene and sanitation, immunizations (such as antibiotics administration), and safer food and water (Cohen, 2000). Although known infectious diseases could be reduced this way, new equally deadly infectious diseases emerged in recent decades. The emergence of newly recognized infections is mainly caused by factors influencing an increase in host susceptibility (Beldomenico & Begon, 2010), increase in disease transmission (Grassly & Fraser, 2008), and evolution of new organisms (Velavan & Meyer, 2020) or drug-resistant microorganisms (Srivastava et al., 2018), which may be not only bacterial and viral but also fungal in origin. The advent of new technologies and changes in patterns of social life inevitably resulted in factors that directly impact the transmission of infectious diseases and the susceptibility of humans to them. These factors include climate changes due to global warming, increased agricultural land usage, aging populations in developed and rapidly growing populations in 2nd and 3rd world countries, international travel and commerce, and breakdown of public health measures.

Given this global environment, the breakout of new infectious diseases appears unavoidable, as the recent COVID-19 pandemic has demonstrated with brutal force. Therefore, it is indispensable to recognize the serious challenge to overcome diseases

caused by pathogenic microorganisms on social and especially medical levels. Understanding virulence modes and developing efficient treatments against infectious disease is primarily hampered by constant microbial adaptation and evolution through the production of beneficial mutations and emersion of new microbial strains. Besides, microbial evolution potentially adds more complexity to recognizing microbial diversity in human microbiota, whose pivotal role in developing the infection has been appreciated for years.

1.1. Role of the microbiome in infectious diseases

The human microbiome substantially influences the onset and progression of infectious diseases as early as birth. Established interactions between the bacterial community and the host during the evolutionary stages play a critical role in maintaining the health status of individuals. Studies have shown that the microbiomes of healthy populations comprise diverse ranges of taxonomic compositions and functional potentials (Huttenhower et al., 2012; Qin et al., 2010). This complicates the identification of microbial compositions that deteriorate the health of an individual. In the context of infectious diseases, however, it is essential to understand microbial composition that potentially induces or prevents colonization of opportunistic pathogens. For instance, increasing gut bacterial diversity using oral delivery of fecal microbiota transplantation and preventing recurrent *Clostridium difficile* infection has been shown to improve health (Khanna et al., 2016). In general, host-microbe and microbe-microbe metabolic interactions are key factors in shaping a microbiome composition that can lead to the colonization of potential human pathogens. In other words, it causes an imbalance between the commensal and pathogenic microbiome which leads to dysbiosis. Inflammatory bowel disease, Rheumatoid arthritis, Systemic lupus erythematosus, and Ankylosing spondylitis are a few examples of inflammatory diseases that are caused by impaired interaction between intestinal microbiota and mucosal immune system (Shi et al., 2017). Furthermore, the effect of bacterial interactions on infectious diseases has been revealed in several studies. Bacterial interactions shape host fitness through life-history tradeoffs (Gould et al., 2017), potentially cause distinct microbiome of disease and healthy states (Van Rensburg et al., 2015), and change the severity of the infection diseases (Korgaonkar et al., 2013). In this regard, available nutrients from the consumed diet by individuals are key factors shaping the metabolic interaction among the microbial community

as, depending on the metabolic composition, they can be beneficial or detrimental for individual microorganisms.

The vast majority of the human microbiome is composed of bacterial species with very low diversity. However, the less abundant part of the microbiome (rare biosphere) (Dethlefsen et al., 2008) is more diverse and usually contains opportunistic pathogens causing health issues when the host is compromised. The rare biosphere especially includes the fungal microbiome (mycobiome) which is orders of magnitude smaller than the bacterial microbiome (Huffnagle & Noverr, 2013). The role of human mycobiome in human health is recently appreciated. Yet, there is little known about the dynamics of the mycobiome during health. Therefore, it appears vital, same as bacterial species, to understand the factors predisposing to the fungal infection, such as interaction with the microbiome through metabolism and diet effects.

1.2. Fungal infection

Among the 50,000 to 250,000 identified fungal species (Richardson & Warnock, 2012), around 600 are associated with human diseases (Brown, Denning, Gow, et al., 2012). Despite this relatively low diversity, over 150 million severe cases of fungal infection occur globally resulting in 1.7 million deaths per year (Bongomin et al., 2017). Besides, strain diversity of individual species is an underestimated factor influencing the spread of infection. The long-term therapeutic applications using antifungal drugs such as triazoles have elevated the number of drug-resistant fungal species and strains. Yet, the fungal infection threat has been underestimated by healthcare authorities (Fausto et al., 2019). Hence, fungal infection can be counted a global threat that needs more effective research to become neutralized.

Metabolism of fungal pathogens is a key factor in developing pathogenicity. They establish the infection by absorbing the local nutrients such as carbon and nitrogen sources. It has been shown that perturbation of carbon, nitrogen, or micronutrient assimilation attenuates fungal pathogenicity (Ene et al., 2014). Therefore, competition for assimilation of nutrition sources due to coexistence with the bacterial community and the host appears critical for the survival and pathogenicity of fungal pathogens. In this regard, identifying mechanisms by which nutrition is deprived of fungal pathogens can provide insights into how to counteract the pathogenicity of fungal pathogens. It has been shown that fungal infections caused by Candida, Cryptococcus, and Aspergillus species are the most prevalent causing

high mortality rate annually (Janbon et al., 2019). This dissertation aims to provide a systems biology perspective of how fungal pathogens (specifically *Candida albicans* and *Aspergillus fumigatus*) benefit or detriment from coexistence with the microbial community when metabolically interacting within this complex environment.

1.2.1. *Candida albicans*

Candida albicans is one of the ~600 fungal human pathogens existing on the earth (Brown, Denning, & Levitz, 2012). In most individuals, it remains as a harmless commensal in the human host body. However, it may cause infection under certain circumstances including an impaired immune system in immunocompromised individuals, dysbalanced microbiota, and an impaired intestinal barrier (Kumamoto et al., 2020). Diseases caused by opportunistic *C. albicans* range from superficial infections like candidiasis in oral cavity, skin, and vagina, to life-threatening systematic infections causing high morbidity and mortality rate. More than thousands of Candida species can cause candidiasis, however in almost of the patients, *C. albicans* is the most abundant isolate (Pfaller & Diekema, 2007). For instance, up to 95% of the oral candidiasis (OC) cases are caused by *C. albicans* (Vila et al., 2020). About 75% of women suffer from vulvovaginal candidiasis at least once in their life, which is mostly caused by this fungus (Denning et al., 2018). Gastrointestinal tract has been found to be one of the suitable niches for colonization of the Candida species increasing the risk of systematic candidiasis (Nash et al., 2017), especially with suppressed immune system. Furthermore, experimental models have indicated that *C. albicans* can enter the bloodstream by translocating through the intestinal barrier (Kadosh et al., 2016; Vautier et al., 2012). Therefore, the pathogenicity-leading factors of *C. albicans* in the intestinal tract can potentially cause systemic candidiasis.

The results obtained from mouse models and human patients have shown that the composition of microbiota is critical for establishment of *C. albicans* colonization (Panpetch et al., 2019; Urzua et al., 2008). Some studies have revealed that mice treatment with antibiotics, especially antibiotics targeting anaerobic bacteria eases the colonization of the fungus (Mishra & Koh, 2018; White et al., 2007). In fact, lack of the bacterial species antagonizing with *C. albicans* gives the opportunity for the fungus to overgrow and switch to the pathogenic state (Graf et al., 2019). Not only the drug treatment but also the diet influences the intestinal composition and *C. albicans* colonization by impacting metabolic regulations. *C. albicans* is able to simultaneously consume multiple nutrients. This nutritional

flexibility makes the fungus adapt to environments with nutritional limitations (Miramón et al., 2020). Nutrient limitations and competition between bacteria and *C. albicans* provide selective pressures that might result in the elimination of less-adapted bacteria and give the fungus opportunity for growth (Calderone & Fonzi, 2001). Therefore, the interplay between *C. albicans* and the microbiome appears to be a key factor in the commensal-to-pathogen shift of the fungus. A longitudinal study of a cohort of 178 preterm infants has revealed that *C. albicans* antagonizes *Klebsiella pneumoniae* whereas *Staphylococcus* spp. may antagonize the fungus (Rao et al., 2021). Another study of 942 human adults has illustrated a mutualistic relationship between fungal and bacterial kingdoms but failed to uncover correlations at the species level (Sun et al., 2021). Despite these insights, there is still a knowledge gap towards a holistic understanding of the functional interactions in the gastrointestinal tract leading to disturbance or maintenance of the predisposing factors of *C. albicans* pathogenicity. For instance, only a system-wide investigation of metabolic trade-offs and mechanisms involved in interactions can potentially provide new insights into the weakness and strength of this and likely other fungi. Altogether, there are different studies towards this aim, yet studies including a theoretical system-wide understanding paving the way for new approaches to combat the disease caused by fungal pathogens are still scarce. A major aim of this dissertation was therefore to predict disease modulating fungal interactions with the bacterial diversity in the gut at the genome-scale metabolic level, which is inaccessible by experiment alone. These insights have the potential to suggest potential key modulators of the fungus and thus open doors for designing new treatments against the infection.

1.2.2. *Aspergillus fumigatus*

Aspergillosis is the most frequent cause of nosocomial fungal infections (Marr et al., 2002; L. S. Wilson et al., 2002), which is caused by Aspergillus species. Out of over 600 species in this genus (Geiser, 2009), *A. fumigatus* is a primary cause of invasive infection (Brakhage, 2005) and occurs naturally in soils. The proliferation of the fungus in its natural habitat results in spreading numerous conidia into the air, which are inhaled daily by humans and animals. The healthy immune system typically responds adequately to entered conidia and eliminates them. However, the ability of the fungus to adapt in a new habitat, especially in immunocompromised individuals, is remarkable and has an adverse effect leading to severe infection (Latgé & Chamilos, 2020). The spectrum of Aspergillus diseases is considerably

broad-ranging from diseases happening in immunocompetent to immunocompromised patients. In terms of the underlying immune status of the host, Aspergillus diseases are classified into three groups of hypersensitivity allergy, structural disease with or without inflammation, and severe immunodeficiency (Latgé & Chamilos, 2020). Severe asthma with fungal sensitization (SAFS) is an Aspergillus-related type of asthma in immunocompetent patients which affects millions of patients with asthma (Brown, Denning, Gow, et al., 2012). Allergic bronchopulmonary aspergillosis (ABPA) is a hypersensitivity disease of the lung in patients with a high immune response that is mostly caused by *A. fumigatus*. Aspergillus colonization in asymptomatic patients with no evidence of invasive or allergic disease is quite common (Máiz et al., 2015; Mortensen et al., 2011). Chronic Pulmonary Aspergillosis (CPA) is a progressive type of aspergillosis that destroys lung tissue. It is provisionally estimated that there are three million patients with CPA worldwide (Brown, Denning, Gow, et al., 2012).

A. fumigatus needs to adapt to its environment to successfully colonize, e.g. new areas, such as human lungs. This adaptation results from the ability of the fungus to generate genetic variations through asexual, parasexual, and sexual cycles (Jianhua Zhang et al., 2021). Most importantly, the variations may be beneficial for the fungus to become drug resistant causing treatment failure. Hence, it is critical to characterize *A. fumigatus* population from different niches and the human host. To this end, the number of available sequencing data on the fungus has recently increased (Ahangarkani et al., 2020; Barber et al., 2021; Chen et al., 2020), leading to a substantial increase in knowledge of the genetic diversity of *A. fumigatus* at the strain resolution. However, it remains unclear how genetic diversity is linked to the phenotypic diversity of *A. fumigatus* strains and how this is leading to metabolic adaptation either in the human lung or environmental niches such as agricultural lands. Given that *A. fumigatus* faces an existing microbiome in the human lung, understanding the microbiome interaction with the fungus can provide insights into how the bacterial community influences the clearance or pathogenicity of the fungus. Metabolic crosstalk between the fungus and bacterial community is one important element of interaction. For example, *Pseudomonas aeruginosa* hinders the biofilm formation of *A. fumigatus* by releasing some small heat-stable molecules (Mowat et al., 2010). Another study showed the antagonistic effect of *Staphylococcus aureus* on *A. fumigatus* conidia in the co-culture of the fungus and the bacterium (Granillo et al., 2015). Yet, due to its remarkable genetic diversity, different metabolic potentials might be employed by the fungus to counteract any growth disturbing effects of the coexisting microbial community. These interactions are still poorly understood at a systems level and have not yet exploited systems

biology simulations. Hence, more effective research in this regard at the systematic level are needed. A major goal of this dissertation is therefore to provide insights into how *A. fumigatus* strains isolated from different niches differ in phenotypes and how the different strains establish a metabolic lung landscape including potentially affecting the lung microbiome in favor of fungal growth. Towards this aim computational system-wide *in silico* studies appeared suitable to study fungal metabolism at the genome scale.

2. **Data-driven computational systems biology approaches in infectious diseases**

Infectious disease is one of the most important and tractable targets in medical studies. The evolution of the human immune system and microorganisms contributing to infection and their interactions inevitably adds complexity to adjusting efficient treatment of infected patients. For instance, bacterial species and metabolites against a fungal pathogen have been identified by performing a systematic study of the human gut where fungal and bacterial species are interacting in patients with antibiotic treatment (Seelbinder et al., 2020). Due to the prevalence of such complexities, reductionist efforts are not ideal to gain a broad perspective of infectious diseases and for developing novel therapeutic approaches. One way to systematically understand these complexities, given the advent of big life science data, is system-wide computational approaches (systems biology) and particularly constraint-based or genome-scale metabolic modeling. These approaches are suited to integrate and combine multiple high-throughput omics- and pathogen-specific data to study respective infectious diseases in a systematic manner.

Systems biology can provide predictions that upon experimental confirmation or incompatibility lead to an iterative refinement of the model at hand and thus knowledge gain that is not feasible by experiment alone (Figure 1). It can be divided into two broad categories: ‘studying biology as a system’ and ‘modeling biology as a system’. The definition of a system in the former is the analysis of a large amount of data generated using current technologies (O’Malley & Dupré, 2005). In contrast, modeling aims to integrate and model biological data at a large, e.g. genome, scale to provide insights that pure data analysis alone cannot. Modeling can be approached both in a Bottom-up or Top-down manner. The Top-down approach aims to identify molecular interactions based on the molecular behavior observed in omics data. The Bottom-up approach infers the molecular mechanisms of the system based on interactions of known components in the system (Bruggeman & Westerhoff, 2007;

De Lorenzo & Galperin, 2009). The expected outcome of the study using either of the approaches is a model of the investigated system. The model derived from the Bottom-up approach gives a mechanistic understanding of how the system works (Arnold & Nikoloski, 2014). The Top-down approach, on contrary, represents a mathematical model that is capable to predict the behavior of a system given certain parameters (Molina-Mora et al., 2020). Both approaches have been frequently employed for metabolic network discovery (Singh & Lercher, 2020), which enables describing the interconversion of metabolites through encoded enzymes for achieving a particular objective in the cell. In this thesis, the bottom-up approach was used to build genome-scale metabolic models (GEMs) of fungal pathogens described in the next sections.

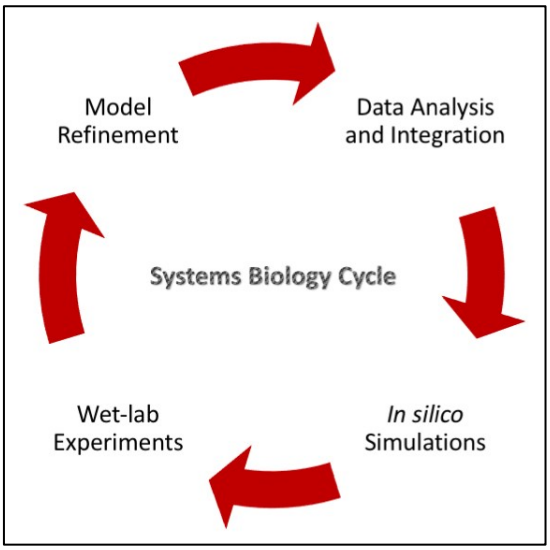


Figure 1. The cycle of Systems Biology. Complex biological phenomena can be identified by a combinations of modeling and wet-lab experiments. Regarding the research question and available data, a model is generated. Integrating the biological, the model generates hypotheses by implementing predictions. Accordingly, wet-lab experiments are designed and performed. If the model predictions contradict the wet-lab experiment results, the model is refined. This cycle iterates until a model with high predictive power is obtained.

In the context of infectious diseases, the cycle of systems biology makes hypotheses, tests them, and validates the results in order to gain comprehensive knowledge of host-pathogen responses, testable models of pathogenesis, diagnostic signatures, and potential drug targets (Aderem et al., 2011). Testable models include also metabolic models that are specifically useful for studying the metabolism of infection where the system's complexity is rather high due to the coexistence of the host with the microbial community.

Availability of high-throughput data and technologies is one of the crucial prerequisites for the implementation of a classical iterative systems biology cycle. This cycle includes integration of data, modeling the system at hand, and refining again upon optional further

data to reach a model of infection that allows for new hypothesis generation and testing. Technologies such as next-generation sequencing have given the opportunity to sequence transcriptomes, non-coding RNAs, and messenger RNAs. In addition, the advent of new technologies such as mass spectrometry has eased the generation of metabolomics, proteomics, lipidomics, and phenomics data. Besides analyzing the data to unravel the phenotypes associated with infection, this thesis mainly aims to integrate them into GEM reconstruction and analysis. Theoretical metabolic models can predict the phenotypic behavior of the target organism (described in detail in section 2.1). In this thesis, GEM reconstruction and analysis provide a testable platform for studying the role of metabolism in fungal infection. Among the aforementioned omics data, genomics, transcriptomics, metabolomics, and phenomics are the fundamental data sources used in this thesis and are briefly introduced next.

Genomics

Genomics enables the study of the genome and reveals its structure, functions, and evolution. It has been used to improve our understanding of human diseases in medical studies. To this end, a plethora of genome-wide association studies (GWAS) linked to human diseases have been conducted in recent years (Cano-Gamez & Trynka, 2020). These studies have explained how genetic variations are associated with disease-causing effects and identified mechanisms underlying complex diseases such as autoimmunities, neurodegenerations, and cardiovascular. Similarly, it has been used to study the genetic causes of infectious diseases. A GWAS prioritized human genes associated with an increased risk of developing systemic Candida infection (de Vries et al., 2020). Another study on *Helicobacter pylori* strains discovered genetic variants associated with gastric cancer risk (Berthenet et al., 2018).

Given the applicability of genomics analyses for describing the cell functions, functional genomics allows gaining knowledge about cell metabolism. This information is a primer to derive GEMs and even improve their qualities, ultimately improving our understanding of the cell phenotypes. Vital to the survival of human pathogens is interspecies phenotype variation occurring through evolutionary stages. Multi-strain GEM reconstruction and analysis (Monk et al., 2013; Seif et al., 2018) is a promising approach to explaining such population phenotypic variations, which can be derived from a large-scale functional genomics study at the strain level.

Transcriptomics

RNA sequencing (RNA-seq) is an evolving technology that makes utilization of next-generation sequencing (NGS) or deep sequencing to deal with transcriptome profiling (Haas & Zody, 2010). During the past decades, it has gradually replaced microarrays (Conway & Schoolnik, 2003) and tag-based transcriptome (Harbers & Carninci, 2005) analysis. Being both qualitative and quantitative is one of the biggest advantages of this technology. This ability fulfills measuring the expression levels of expressed genes with even low abundance levels. Better coverage and resolution compared to the previous methods is another advantage of this technique. Hence, it has become a better choice for researchers dealing with transcriptomics analysis.

RNA-seq has been widely employed for studying host-pathogen interactions. The RNA-seq datasets provide insight into mechanisms of host cell invasion and immune evasion. It, especially, assists in understanding infection dynamics. It concentrates on how pathogen regulates the genetic circuits in the host and how that is linked to pathogenesis. These studies may also open the door for the development of personalized therapies for patients suffering from different levels of infection. Many studies have shown the efficacy of the RNA-seq analysis to identify differentially expressed genes and metabolic and signaling pathways as biomarkers of infectious diseases by using functional enrichment methods such as gene ontology (GO) and gene set enrichment analysis (GSEA) (Alonso-Hearn et al., 2019; Y. Wang et al., 2014; J. A. C. Wilson et al., 2017). Yet, considering the transcriptomic data besides other approaches for investigating metabolism (e.g., metabolic modeling) as another layer can provide a better resolution for understanding the genotype-phenotype connection.

Metabolomics

Metabolomics is the study of chemical processes involving metabolites. Novel technologies including different types of mass spectrometry have enabled producing metabolic profiles giving metabolic cues from the studied organism. The human metabolome project (Wishart et al., 2007) is one of the remarkable studies in this regard. However, metabolomics rarely provides a clear explanation for a biological phenomenon on its own and out of context. Interpretation of metabolomics data can be enhanced by its integration into systems biology approaches. This includes inferring the regulatory mechanisms in a biological network (Bruggeman & Westerhoff, 2007), explaining the genetic variations (Suhre et al., 2011), and

especially integrating the data into GEMs to derive context-specific metabolic networks (Cho et al., 2019), identify different modes of metabolic regulations (Çakir et al., 2006), and simulate the dynamic intracellular metabolic changes (Bordbar et al., 2017). As metabolomics resembles the endpoint of biological processes, such as the end products of enzymatic reactions, it is especially suited to be included in data-driven systems biology studies of metabolism to improve GEMs' completeness and accuracy of its predictions.

Phenomics

Phenomics enables the systematic study of phenotypes of a given organism. It provides knowledge about phenomes emerging in response to different conditions such as environmental influences and genetic mutations. For instance, human phenotype ontology (Köhler et al., 2017) was created to provide knowledge about observed phenotype abnormalities linked to human diseases.

Phenotypic Microarray (PM) technology provides phenomics data enabling phenotypic data analysis on a large scale. PM is a high-throughput technology for testing a large number of cellular phenotypes at the same time. Each microarray is designed for testing a certain category of substrates. The ability of the PMs in detecting cell phenotypes has been validated for the first time by discovering undetected genetic changes in *Escherichia coli* (Bochner et al., 2001). The phenotypic characterization includes testing the ability of the target organism to catabolize basic nutrients such as carbon-, nitrogen-, sulfur- and phosphor-based substrates, as well as drug and chemical sensitivities. Color changes in wells indicate the intensity of growth due to redox reactions associated with cellular respiration, which occurs over hours (https://www.biolog.com/). These experiments dynamically record the growth signals represented as growth curves which need to be analyzed for correct interpretation. In general, this platform's applicability includes identifying growth/no-growth conditions and phenotypic differences when comparing different growth conditions or strains (Vaas et al., 2012; Vehkala et al., 2015).

Phenotypic microarray data has been widely used for uncovering the characteristics of mutations of different organisms for the genes whose functions are not well understood (Jiaqin Zhang & Biswas, 2009; Zhou et al., 2003). However, the usage advantage of this data is not only limited to that. It also provides valuable phenotypic information to tune and test GEMs' accuracy in predicting the catabolic capabilities of different substrates (i.e., carbon, nitrogen, sulfur, and phosphor sources) as it has been shown in many studies in the past

(Jensen et al., 2020; Liao et al., 2011; Norsigian et al., 2018; B. Wang et al., 2016). The quality of a metabolic model can then be improved by either filling the gaps (i.e., missing biochemical reactions and/or metabolites) or removing the false information from the model through the classical systems biology cycle.

2.1. Genome-scale metabolic modeling

With the emergence of genome sequencing technologies in the past years, the number of complete genome sequences has been increasing rapidly (Figure 2). Although annotation of genome sequences provides a huge amount of information about genome functions, the phenotypes of a given cell can only be understood to a limited extent at the molecular level based on genome annotation. In biotechnology and medicine applications, the relevant research questions that a researcher might require to address include the following: How does bacterial metabolism adapt to environmental changes? How does cell metabolism respond to genotype modifications? What are the side effects of a particular drug on human metabolism? Furthermore, confronting the information from the annotated genome and the observed phenotypes of a given cell might lead to contradictions. For example, it might occur that while an organism has been shown experimentally to grow on a certain medium, no sequence might be identified that associates with the uptake or production of a compound essential for the growth of the organism in the presence of the given medium. One approach to bridge those gaps between genome-derived biochemical information and metabolic phenotypes of the living cell is the reconstruction and analysis of GEMs (Palsson, 2015). These models are built based on the genetic information of the target organism and represent the association between a single or a group of gene/s and catalyzing a biochemical reaction in a concise manner.

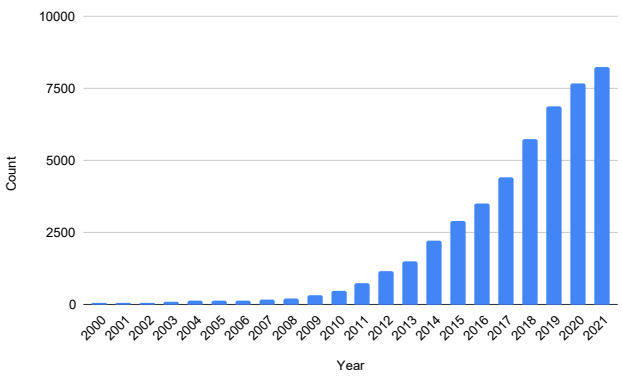


Figure 2. The number of peer-reviewed papers on "whole genome sequencing" published per year according to https://pubmed.ncbi.nlm.nih.gov.

Constraint-based GEMs are a group of models aiming at assessing the metabolic behavior of the target organism through analyzing the biochemical reaction fluxes under a certain environmental condition that ultimately shapes the physiological state of the underlying metabolic network. They represent a network of metabolites flowing through the edges of the network and converting to each other at nodes reflecting the role of reactions. Some of those metabolites can be exchanged with the environment. The minimum and maximum amount of flow of metabolites can be fixed to a certain extent while the network achieves balanced production and consumption of all intermediate metabolites. This condition is termed “steady-state,” meaning that it is assumed that there is no accumulation of molecules in the system. The system is set to achieve a certain objective such as maximization of the growth rate of the cell (Orth et al., 2010). Metabolic models provide a mathematical representation of such a system. The mathematical representation is based on the stoichiometric coefficients devoted to the molecules that are consumed or produced by a reaction. These coefficients impose constraints on the flow of metabolites through the metabolic network. All the coefficients for all the reactions in a network can be represented in a matrix. Considering that the system is running under the steady-state condition, the following represents the mathematical equation of the system:

$$\begin{aligned} &maximize\ (or\ minimize) \quad Z = f'v \\ &\qquad\qquad\qquad\qquad\quad such\ that\ S.v = 0 \\ &\qquad\qquad\qquad\qquad\quad v_j^L \leq v_j \leq v_j^U, \quad j = 1, 2, \dots, n \end{aligned} \tag{1}$$

where Z is the objective function, f is the vector of objective function coefficients, S is the stoichiometric matrix, v is the vector of reaction fluxes. S*v = 0 constrains the network for steady-state condition. L is the lower bound, and U is the upper bound for reaction j and v, v^L and v^U are flux vectors. The boundaries impose the thermodynamic constraints of all defined metabolic reactions. Ultimately, flux vector solutions can be obtained by using Linear programming with respect to the mathematically defined objective function (Z in equation 1) for which the objective flux value is optimized. However, alternative flux distributions towards the same objective function value exist, since metabolic models encompass more reactions than metabolites and thus potentially allow for equally possible routes through the given system (Mahadevan & Schilling, 2003) (Figure 3).

2.1.1. Genome-scale metabolic model generation

Generally, metabolic models substantially depend on the reconstruction efforts including manual curation to provide sufficient quality, genome coverage, and thus predictive power. Therefore, one has to ensure sufficient quality when either using published models (if possible) or when building a model from scratch. In principle, building a high-quality metabolic model can be broken down into four major steps (Thiele & Palsson, 2010): I) creating a draft reconstruction; II) manual reconstruction refinement; III) Conversion from reconstruction to mathematical model; IV) Network evaluation.

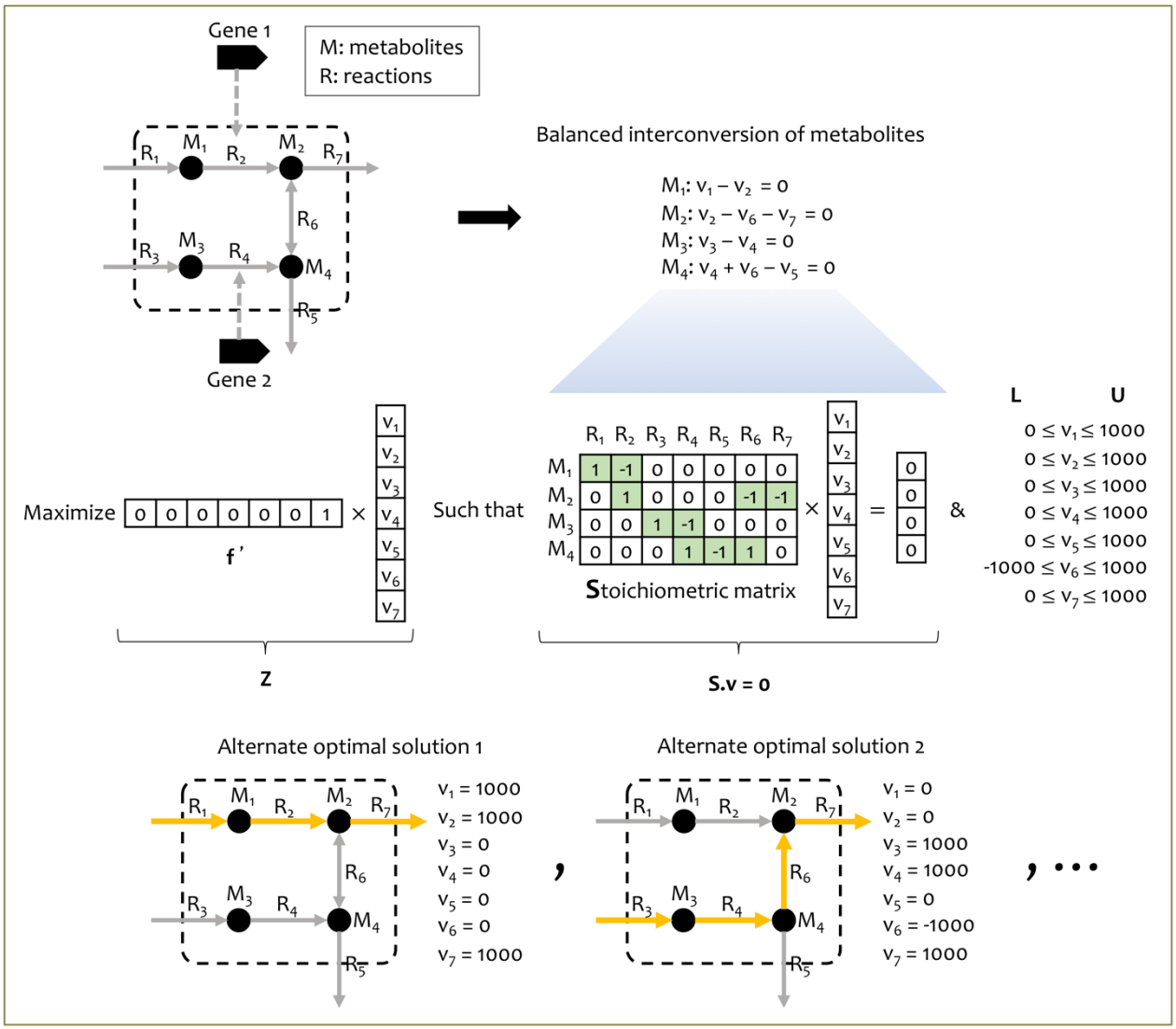


Figure 3. Toy metabolic network. The optimization problem can be solved to maximize the objective value of the system (V7 in the toy metabolic model) by using linear programming. Due to the larger number of reactions than metabolites (seven reactions and four metabolites in the toy model), the set of equations is underdetermined, and alternate optimal solutions exist.

In the first step, the draft reconstruction is created based on the genome annotation of the target organism. Therefore, the initial model contains the collection of genome-encoded metabolic functions. To this end, several tools and methods have been authored and developed. A group of tools is aiming at merely building the draft models from scratch without taking into account further refinements (Arakawa et al., 2006; Notebaart et al., 2006; Pinney et al., 2005). In contrast, there is another group of tools and methods to generate the draft models which consider some curation steps in order to represent partially refined models and consequently decrease the effort for further refinements (Agren et al., 2013; Devoid et al., 2013; Dias et al., 2015; Karp et al., 2016; Pitkänen et al., 2014; Swainston et al., 2011). Nevertheless, the automated draft reconstructions often lack the complete information relevant to metabolism and automated methods partially overlook standards for a high-quality generation of reconstructions (Faria et al., 2018; Simeonidis & Price, 2015). Hence, manual curation is essential in the second step of model creation.

In the second step, one must consider several refinements of reconstruction. Thermodynamically favorable directions of reactions should be identified. This avoids erroneous energy-generating cycles (EGCs) (Fritzemeier et al., 2017) in the network by which energy molecules are produced without taking any nutrients. In the case of studying the eukaryotic cells, genes and reactions should preferably be located in their subcellular compartments. Having constructed compartments with an optimal number of metabolite transporters avoids redundant exchanges of metabolites between the compartments and improves the accuracy of predictions. However, this refinement suffers from a lack of information on the location of enzymes activity in the cell and inner cell transporters. Prediction-based tools such as WoLF PSORT (Horton et al., 2007) and gap-filling tools (Bernstein et al., 2021) can accommodate it to some extent. Gene-protein-reaction (GPR) association is one of the fundamental features of genome-scale metabolic models by which the genotype-phenotype relationships can be revealed in simulations. Hence, these require careful verification in all curation steps to ensure predictive capabilities. The most conventional objective for microorganisms is growth since the typical goal of a cell is to maximize its reproduction (Feist & Palsson, 2010). Therefore, estimating the precise composition of biomass and incorporating the data into the model is vital to ensure the accurate performance of the model predictions. In general, such refinements, in particular manual curations, increase the quality of the reconstruction, especially for the details unseen by the draft reconstruction step.

After the manual curation, the reconstruction information is converted to a mathematical format for model evaluation. Assessing the physiological properties of the model is one of the major efforts in this step. *In-silico* simulations must be performed to inspect whether the model is able to capture the substrate usage capabilities, gene essentiality, and fermentation byproducts. This evaluation relies on the availability of biological data such as phenotypic data. Phenotypic microarray data shows whether the target organism can or cannot grow on individual substrates. Accordingly, the contradictions between the model predictions and the growth data can be fixed by inspecting the metabolic pathways in the model and correcting them by using the data obtained from the literature and databases.

2.2. Systems biology of microbiome

With the emergence of metagenome sequencing technologies, the number of advanced researches on the human microbiome has been rapidly increasing. Analyzing the metagenomics data, on its own, provides numerous insights regarding the association of microbial compositions with healthy and disease states such as infectious diseases. However, there is little knowledge about the interaction of individual microorganisms with other microbiome members and the host leading to disease and healthy states. Mathematical modeling, in general, is a promising approach for providing a mechanistic understanding of microbiome interactions and facilitates revealing the contribution of individual microbiome members to the disease and healthy states.

The microbiome can be mathematically modeled from different aspects. Some of the approaches focus on modeling based on sequence real abundance (Vandeputte et al., 2016) predicting e.g. co-occurrence of microorganisms or agent-based modeling to simulate the interactions between different communities (An et al., 2009). Investigating the metabolism dynamics using ordinary differential equations (ODEs) is also applied to model the cross-talks among microbiome members (Van Wey et al., 2016), but generally suffers from a lack of information on kinetic parameters, which are typically not available for all modeled species.

On the contrary, GEM analysis simplifies the simulations by considering the steady-state condition in the system. Despite this simplification assumption, it is a promising approach proven in different medical and industrial applications (Feist et al., 2007; Shlomi et al., 2011). For this reason, it has also rapidly become a popular approach in microbiome

studies (Chan et al., 2017; Shoaie et al., 2015). Recent developments in whole genome sequencing have resulted in the comprehensive generation of metabolic models for the microbiome bacterial species (Machado et al., 2018; Magnúsdóttir et al., 2017), facilitating the modeling of the interplay between microorganisms in the ecosystem. This thesis employs GEM analysis as a powerful mathematical modeling approach to study the role of the microbiome in developing fungal infectious diseases.

2.2.1. Metabolic modeling toward microbial community analysis

GEMs have been broadly used for simulating the microorganism-microorganism and microorganism-host interactions (Jamshidi & Raghunathan, 2015; Medlock et al., 2018). Simulating an interaction between two bacterial species can provide a mechanistic understanding of the interaction through metabolites exchanged between organisms. Simulation of a bacterium-bacterium interaction explained that butyrate production by *Faecalibacterium prausnitzii* is elevated with consuming acetic acid secreted from *Bifidobacterium adolescentis* (El-Semman et al., 2014). Variant methods have been developed for simulating the interactions among multiple microorganisms at the same time, which inevitably adds more complexity to modeling (Chan et al., 2017; Koch et al., 2019). This complexity increases when the target community is at a considerably large scale comprising hundreds of microorganisms.

In contrast, a pairwise simulation (Figure 4) provides mechanistic insights into the interaction between two organisms by applying a simpler mathematical formulation (Heinken & Thiele, 2015). This approach is especially useful for obtaining insights into the direct effect of one microorganism on another in the community. Each microbe is assumed to be a separate compartment communicating with the other one through a shared environment, which imitates a co-culture or complex ecosystem (e.g., human gut). The system's objective function is considered to be the maximization of the summation of biomass fluxes of both microorganisms as each one separately tries to maximize its reproduction. The predicted growth rate of each microbe can be compared to its predicted growth rate when growing alone. Hence, the simulation suggests either positive, negative, or neutral effects of one microbe on another in the interaction. As the interaction is encoded in exchanged metabolites between two microbes, it also provides an in-depth understanding of interaction at the

molecular level. In the context of infectious diseases, a pairwise interaction reveals how a pathogen can be benefited or harmed from interactions with individual microbes.

Metabolites play a critical role in interactions among microbiome members. For instance, long-term dietary influences the composition and activity of the microbes colonizing the gut (Walker et al., 2011). The metabolomics data can be integrated into GEM analysis of different species to infer the effect of the metabolites on corresponding cell phenotypes. For example, it may reveal how metabolites influence the reproduction and colonization of species through GEM analysis. Large cohort sizes in microbiome studies make comprehensive metabolomics profiling costly and logistically difficult to perform at a large scale. Hence, in case scenarios where the metabolomics data is lacking, a computational predictive tool is advantageous to this aim. Such methods predict the metabolomics profile reflecting the metabolic environment caused by microbial composition (Yin et al., 2020). Integration of the predicted metabolome profile into GEM analysis, therefore, can provide a better understanding of how the microbiome can metabolically influence the colonization of individual members, especially opportunistic pathogens existing in the biosphere.

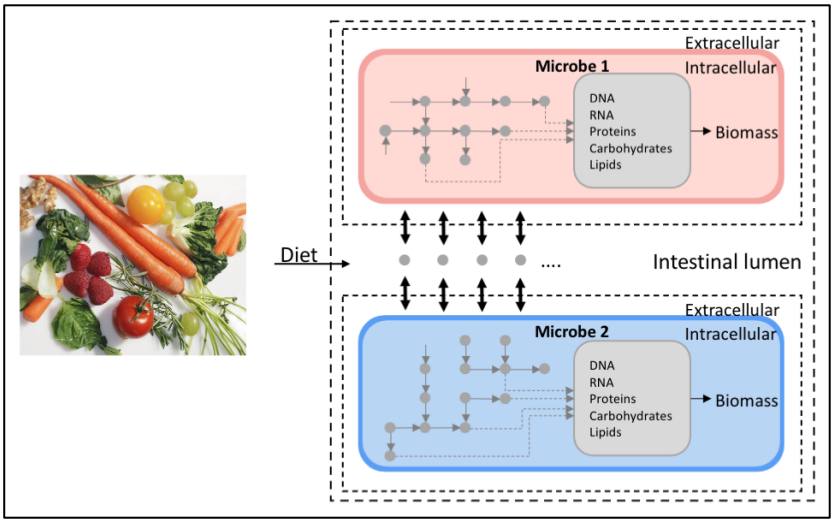


Figure 3. Pairwise interaction between two microbes. Each microbe reaches the maximum possible growth rate when living in interaction with another microbe. Each microbe can metabolically cross-talk with the paired one through the shared environment (e.g. intestinal lumen). The diet primarily specifies which metabolites are available to both GEMs.

3. **Objectives of this dissertation**

The aim of this dissertation is to apply computational tools and implement systems biology approaches with a focus on the metabolism of fungal pathogens (i.e., *Candida*

albicans and *Aspergillus fumigatus*) to uncover key factors in the colonization of the pathogens from a systems biology perspective and unseen by wet-lab experiments alone. Vital to data-driven modeling is the analysis of wet-lab experimental data to obtain better resolution for data interpretation. Due to the complexity of the topic, this thesis puts major effort into the systematic study of the role of metabolism in infection, especially by generating and analyzing GEMs of human fungal pathogens. Additionally, a part of this dissertation is dedicated to my contribution to adequate data analysis. As a proof of principle, the thesis starts with two studies showing the computational approaches' applicability to understand better or confirm findings obtained by analyzing omics data. Then, it continues with genome-scale metabolic model reconstruction and analysis of human fungal pathogens as the major part.

3.1. Research questions

Besides fungal infection, there are a plethora of common infectious diseases originating from viruses. Influenza and Rhinovirus are the predominant causes of the common cold. Though major differences in inflammatory or immune response have been detected between two common disease-causing viruses (To et al., 2018), a systematic transcriptomic analysis directly comparing influenza and Rhinovirus is lacking. Alongside performing the transcriptomics analysis, gene set enrichment analysis is carried out to address the following question: **What are the human phenotypes in response to infection with influenza and Rhinovirus?**

As human pathogens never live alone, this dissertation investigates the topic from a broader perspective by integrating the effect of the microbiome into the study. Microbiome composition has a high impact on predisposing infectious diseases. Consumption of antibiotics is very common for treating patients in the Intensive Care Unit (ICU). It has been shown that antibiotics have a significant impact on the microbial and metabolic composition of the human gut that might cause an infection-vulnerable microbiome (Seelbinder et al., 2020). Alongside a systematic investigation of microbiome samples of ICU patients to reveal the effect of antibiotic treatment on microbial and metabolic changes, genome-scale metabolic modeling was applied to address the following question: **How antibiotic treatment**

alters the contribution of individual bacterial species to the metabolic landscape in the human gut?

Next, my thesis continues with the major topic, i.e., fungal pathogens, to uncover the role of metabolism in colonization and overgrowth in connection with the human microbiome.

Recent studies have found associations between changes in the composition of the gut microbiota and the development of local *C. albicans* infection (da Silva Dantas et al., 2016; Förster et al., 2016). However, the potential of gut microbes and individual contributions of bacterial species to influence the development of systematic candidiasis by the fungus is not clear. Several studies have shown the impact of different bacterial species on *C. albicans* growth using wet-lab experiments (Shirtliff et al., 2009; Thein et al., 2006). For instance, the probiotic bacterium *Lactobacillus rhamnosus* is well-known to antagonize the fungus by inhibiting proliferation (Matsubara et al., 2016), the hypha formation (Allonsius et al., 2019), and biofilm formation (Rossoni et al., 2018). Yet the interaction for many gut bacterial species is unknown, as cultivation of many gut bacterial species is not doable. Moreover, metabolic trade-off and underlying mechanisms between the fungus and the bacteria causing growth inhibition and growth promotion are not well understood purely by experiments. *In silico* simulations can overcome such limitations and systematically provide insights into the fungus's metabolic interactions with the gut bacterial community.

Not only interaction with the microbiome but also particular functions of the fungus involved in cellular processes are crucial for its growth and colonization in the human gut. Regulating these processes is important for growth, proliferation, environmental adaptation, and stress responses (Ramírez-Zavala et al., 2017). Protein kinases play a fundamental role in regulating cellular processes. The *C. albicans* genome encodes two serine-arginine protein kinases, Sky1 and Sky2, and their functions are not yet understood. Investigating the phenotypes of the fungus in the presence and absence of encoding genes can reveal the function of these proteins.

In order to reveal important factors affecting the colonization of *C. albicans* in the human gut, a genome-scale metabolic model for the fungus was developed, validated, and used to analyze the metabolic interactions of the fungus with the human microbiome. In particular, the interaction of the fungus with *L. rhamnosus* was investigated in more detail by incorporating omics data into the study. Validation of the model was mainly accomplished by using phenotypic microarray data of *C. albicans* wild-type and some metabolic mutations. Likewise, phenotypic microarray data was used to reveal the functions of the aforementioned

protein kinases in the fungus. In this topic of the thesis, the following questions will be addressed:

- Which functional interactions in the human gut can promote or inhibit the growth of the fungus?

- Which metabolic pathways in the fungus respond to promoting and inhibiting influences?

- How can diet reflecting the metabolic environment influence *C. albicans*-microbiome interactions?

Given *Lactobacillus rhamnosus* as a particular antagonizing case, an in-depth investigation was carried out by *in silico* simulations in the presence of intestinal epithelial cells (IEC) and by integrating the omics data into the study. Regarding that, the following questions will be addressed:

- **What is the role of human epithelial cells in the antagonizing effect of *L. rhamnosus*?**

- **How metabolic pathways in the bacterium and IEC are coordinated to support the negative effect on the fungus and how the fungus metabolically responds to the negative effect?**

Lastly, the analysis of the phenotypic microarray for the protein kinases will address the following question:

- **What is the contribution of Sky1 and Sky2 proteins to the catabolic capability of *C. albicans* and what are the differences?**

Aspergillus fumigatus is a globally occurring environmental saprotrophic mold that poses a serious threat to hospitalized, particularly immunocompromised patients (Arastehfar et al., 2020). It remains largely unknown to which extent environmental or clinical *A. fumigatus* isolates possess specific characteristics due to the need of coping with environmental stresses or accessible nutrient profiles. A recent study from our group explored the genetic diversity of *A. fumigatus* to reveal a remarkably low core orthogroup (69%) shared by clinical

and environmental isolates (Barber et al., 2021). However, it is not yet clear whether the genetic diversity affects the phenotypic diversity of the isolates.

To this end, a platform of genome-scale metabolic models for 252 clinical and environmental *A. fumigatus* isolates and a pan-model was generated. This platform can be used to explain the phenotypic diversity of *A. fumigatus* isolates and assess the phenotypic behavior of the fungus under different environmental conditions circumventing the technical limitations and difficulties of the wet-lab experiments. In order to explore the applicability of the generated clinical genome-scale models, the interaction of the fungus with the lung microbiome of forty cystic fibrosis patients was analyzed. This part of the thesis will address the following questions:

- **What is the phenotypic diversity of the *A. fumigatus* isolates and what are the metabolic differences between clinical and environmental isolates?**

- **How *A. fumigatus* contributes to the shaping of the metabolic landscape of the lung microbiome in a favorable manner for fungal growth?**

OVERVIEW OF THE MANUSCRIPTS

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 July 2020
doi: 10.3389/fmicb.2020.01580

Comparative Transcriptomic Analysis of Rhinovirus and Influenza Virus Infection

Status: Published in Frontiers in Microbiology

Bibliography: Dissanayake, Thrimendra Kaushika, Sascha Schäuble, Mohammad Hassan Mirhakkak, Wai-Lan Wu, Anthony Chin-Ki Ng, Cyril CY Yip, Albert García López et al. "Comparative transcriptomic analysis of rhinovirus and influenza virus infection." *Frontiers in microbiology* (2020): 1580.

Summary: This study carried out a systematic transcriptomics analysis directly comparing Rhinovirus and Influenza. The systematic computational approaches identified differentially expressed genes and enriched pathways through gene set enrichment analysis (GSEA). It revealed the host phenotypes in response to the infection caused by Rhinovirus and Influenza.

GUT MICROBES
2021, VOL. 13, NO. 1, e1993598 (25 pages)
https://doi.org/10.1080/19490976.2021.1993598

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness

Status: Published in *Gut Microbes*

Bibliography: Marfil-Sánchez, Andrea, Lu Zhang, Pol Alonso-Pernas, Mohammad Mirhakkak, Melinda Mueller, Bastian Seelbinder, Yueqiong Ni et al. "An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness." *Gut Microbes* 13, no. 1 (2021): 1993598.

Summary: This study performed a comprehensive characterization of the microbiome, mycobiome, and functional potential of the gut community and individual species. It used a cohort of 75 individuals, including ICU patients under antibiotic administration. Alongside this systematic investigation, genome-scale metabolic modeling was applied to identify individual bacterial species' contribution to the human gut's metabolic landscape, altered by the antibiotic treatment.

The ISME Journal (2021) 15:1257–1270
https://doi.org/10.1038/s41396-020-00848-z

ISME

ARTICLE

Metabolic modeling predicts specific gut bacteria as key determinants for *Candida albicans* colonization levels

Status: Published in *the ISME journal*

Bibliography: Mirhakkak, Mohammad H., Sascha Schäuble, Tilman E. Klassert, Sascha Brunke, Philipp Brandt, Daniel Loos, Ruben V. Uribe et al. "Metabolic modeling predicts specific gut bacteria as key determinants for Candida albicans colonization levels." *The ISME journal* 15, no. 5 (2021): 1257-1270.

Summary: In this study, a genome-scale metabolic model for *Candida albicans* was generated. The model was used to identify bacterial species that potentially can inhibit or promote the growth of the fungus through pairwise simulations. Moreover, it suggested metabolites having positive or negative impact on *C. albicans* growth. *In vitro* experiment and metagenomics analysis of 26 fecal samples of the cancer patients supported the findings of *in silico* simulations.

Candida albicans SR-like protein kinases regulate different cellular processes: Sky1 is involved in control of ion homeostasis, while Sky2 is important for dipeptide utilization

Status: Accepted for publication in *Frontiers in Cellular and Infection Microbiology*

Authors: Philipp Brandt, Franziska Gerwien, Lysett Wagner, Thomas Krüger, Bernardo Ramírez-Zavala, Mohammad H. Mirhakkak, Sascha Schäuble, Olaf Kniemeyer, Gianni Panagiotou, Axel A. Brakhage, Joachim Morschhäuser and Slavena Vylkova

Summary: This study investigated the role of Serine-arginine (SR) protein kinases in the cellular functions of *Candida albicans* that have not been understood before. Different experimental growth assays and high-throughput phenotypic screening of SR gene knock-outs revealed the effect of Sky1 and Sky2 proteins on cellular processes influencing the fungal growth in different environmental and nutritional conditions.

Lactobacillus rhamnosus colonisation antagonizes _Candida albicans_ by forcing metabolic adaptations that compromise pathogenicity

Status: Revised version submitted to *Nature Communications*

Authors: Raquel Alonso-Roman, Antonia Last, Mohammad H. Mirhakkak, Jakob L. Sprague, Lars Möller, Peter Großmann, Katja Graf, Rena Gratz, Selene Mogavero, Slavena Vylkova, Gianni Panagiotou, Sascha Schäuble, Bernhard Hube, Mark S. Gresnigt

Summary: The colonizing and probiotic bacterium *Lactobacillus rhamnosus* can antagonize *Candida albicans*. The study investigated the interplay between *C. albicans*, *L. rhamnosus*, and intestinal epithelial cells (IEC) by integrating transcriptional and metabolic profiling, and reverse genetics. Untargeted metabolomics combined with constraint-based metabolic modeling identified antivirulence and antifungal metabolites and suggested that IEC fosters bacterial growth by releasing specific metabolites, which was confirmed experimentally.

Candida expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions

Status: In preparation

Authors: Bastian Seelbinder, Zoltan Lohinai, Ruben Vazquez-Uribe, Xiuqiang Chen, Silvia Lopez, Mohammad Mirhakkak, Anja Wellejus, Glen J. Weiss, Morten Sommer, Gianni Panagiotou

Summary: The study integrated mycobiome and shotgun metagenomics data from 75 lung cancer patients to demonstrate the critical role of gut bacteria in shaping mycobiome compositions. It identified an ecological signature associated with *Candida* expansion, addressing why many individuals with high levels of *Candida* do not develop systemic candidiasis. As a complementary layer, the study took advantage of constraint-based metabolic modeling to demonstrate the ability of bacterial species to produce short-chain fatty acids and lactic acid.

A pan-genome resembling genome-scale metabolic model platform of 252 *Aspergillus fumigatus* strains reveals growth dependencies from the lung microbiome

Status: In preparation

Authors: Mohammad H. Mirhakkak, Xiuqiang Chen, Tongta Sae-Ong, Lin Lin Xu, Thorsten Heinekamp, Oliver Kurzai, Amelia Barber, Axel Brakhage, Sebastien Boutin, Sascha Schäuble, Gianni Panagiotou

Summary: In this study, a platform of 252 *A. fumigatus* GEMs at the strain resolution was generated. It revealed the phenotypic diversity of *A. fumigatus* strains isolated from different hospitals and farms in Germany and explained the contribution of the fungus to the shaping of the metabolic landscape of the lung microbiome in a favorable manner for the fungal growth.

MANUSCRIPTS

No 1. Comparative Transcriptomic Analysis of Rhinovirus and Influenza Virus Infection

FORM 1

Manuscript No. 1

Manuscript title: Comparative transcriptomics analysis of Rhinovirus and Influenza virus infection

Authors: Thrimendra Kaushika Dissanayake, Sascha Schäuble, **Mohammad Hassan Mirhakkak**, Wai-Lan Wu1, Anthony Chin-Ki Ng, Cyril C. Y. Yip, Albert García López, Thomas Wolf, Man-Lung Yeung1,3,4,5, Kwok-Hung Chan, Kwok-Yung Yuen, Gianni Panagiotou and Kelvin Kai-Wang To

Bibliographic information: Dissanayake, Thrimendra Kaushika, et al. "Comparative transcriptomic analysis of rhinovirus and influenza virus infection." *Frontiers in microbiology* (2020): 1580.

The candidate is (Please tick the appropriate box.)

☐ First author, ☐ Co-first author, ☐ Corresponding author, ☒ Co-author.

Status: published

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Dissanayake, T.K.	15%		80%	25%	
Schäuble, S.	15%	40%		25%	
Mirhakkak, M.H.		25%		10%	
López, A.G.		15%		10%	
Wolf, T.		20%		15%	
Panagiotou, G.	35%			15%	40%
To, K.K.	35%				40%
Others			20%		20%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

FORM 2

Manuscript No. 1

Short reference Dissanayake *et al*., *Frontiers in microbiology* (2020)

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles): Gene set enrichment analysis

Figure(s) # 5	☒	**100%**
Figure(s) # 1-4	☐	**0%**

(Add more table boxes depending on the number of figures)

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 July 2020
doi: 10.3389/fmicb.2020.01580

Comparative Transcriptomic Analysis of Rhinovirus and Influenza Virus Infection

*Thrimendra Kaushika Dissanayake[1†], Sascha Schäuble[2†], Mohammad Hassan Mirhakkak[2], Wai-Lan Wu[1], Anthony Chin-Ki Ng[1], Cyril C. Y. Yip[1], Albert García López[2], Thomas Wolf[2], Man-Lung Yeung[1,3,4,5], Kwok-Hung Chan[1], Kwok-Yung Yuen[1,3,4,5], Gianni Panagiotou[2,6] and Kelvin Kai-Wang To[1,3,4,5]**

[1] Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China, [2] Systems Biology and Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany, [3] State Key Laboratory for Emerging Infectious Diseases, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China, [4] Department of Clinical Microbiology and Infection Control, The University of Hong Kong, Hong Kong, China, [5] Carol Yu Centre for Infection, The University of Hong Kong, Hong Kong, China, [6] Systems Biology and Bioinformatics Group, School of Biological Sciences, Faculty of Sciences, The University of Hong Kong, Hong Kong, China

OPEN ACCESS

Edited by:
Lars Kaderali, Universitätsmedizin Greifswald, Germany

Reviewed by:
Stefan Simm, Goethe Business School, Germany
Sabine Stegemann-Koniszewski, University Hospital Magdeburg, Germany

***Correspondence:**
Kelvin Kai-Wang To
kelvinto@hku.hk

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *28 February 2020*
Accepted: *17 June 2020*
Published: *21 July 2020*

Citation:
Dissanayake TK, Schäuble S, Mirhakkak MH, Wu W-L, Ng AC-K, Yip CCY, López AG, Wolf T, Yeung M-L, Chan K-H, Yuen K-Y, Panagiotou G and To KK-W (2020) Comparative Transcriptomic Analysis of Rhinovirus and Influenza Virus Infection. Front. Microbiol. 11:1580. doi: 10.3389/fmicb.2020.01580

Rhinovirus (RV) and influenza virus are the most frequently detected respiratory viruses among adult patients with community acquired pneumonia. Previous clinical studies have identified major differences in the clinical presentations and inflammatory or immune response during these infections. A systematic transcriptomic analysis directly comparing influenza and RV is lacking. Here, we sought to compare the transcriptomic response to these viral infections. Human airway epithelial Calu-3 cells were infected with contemporary clinical isolates of RV, influenza A virus (IAV), or influenza B virus (IBV). Host gene expression was determined using RNA-seq. Differentially expressed genes (DEGs) with respect to mock-infected cells were identified using the overlapping gene-set of four different statistical models. Transcriptomic analysis showed that RV-infected cells have a more blunted host response with fewer DEGs than IAV or IBV-infected cells. IFNL1 and CXCL10 were among the most upregulated DEGs during RV, IAV, and IBV infection. Other DEGs that were highly expressed for all 3 viruses were mainly genes related to type I or type III interferons (RSAD2, IDO1) and chemokines (CXCL11). Notably, ICAM5, a known receptor for enterovirus D68, was highly expressed during RV infection only. Gene Set Enrichment Analysis (GSEA) confirmed that pathways associated with interferon response, innate immunity, or regulation of inflammatory response, were most perturbed for all three viruses. Network analysis showed that steroid-related pathways were enriched. Taken together, our data using contemporary virus strains suggests that genes related to interferon and chemokine predominated the host response associated with RV, IAV, and IBV infection. Several highly expressed genes, especially ICAM5 which is preferentially-induced during RV infection, deserve further investigation.

Keywords: influenza, rhinovirus, transcriptomics analysis, ICAM5, interferons, cytokines

INTRODUCTION

Influenza virus is well-known to cause severe respiratory tract infection. The number of deaths associated with seasonal influenza virus infection has been estimated to be between 290,000 and 650,000 globally (Iuliano et al., 2018). Influenza pandemics and seasonal epidemics are associated with sudden surge in hospitalizations and deaths. Unlike influenza virus, rhinovirus (RV) has been relatively neglected because of the deep-rooted association with common cold. However, recent studies have shown that RV is a frequent cause of severe respiratory illnesses and is associated with a higher case-fatality rate than influenza virus infection (Jain et al., 2015; Hung et al., 2017; Ieven et al., 2018; Prill et al., 2018). RV is also the most commonly detected respiratory viruses among patients with lower respiratory tract infection. Outbreaks of severe RV infection have been reported (Marcone et al., 2017).

Clinically, both influenza virus and RV cause respiratory tract infection, and can be complicated by extrapulmonary disease (To et al., 2016a, 2019). However, there are important clinical differences between RV and influenza virus infection. We have previously found that hospitalized patients with RV infection are more likely to present with exacerbation of chronic lung diseases or complicated with extrapulmonary manifestations (To et al., 2018, 2019).

Host response to infection provides tremendous insights into the pathogenesis of an infection. Our previous study showed that a persistently dysregulated cytokine and chemokine response was associated with severe influenza A(H1N1)pdm09 infection (To et al., 2010). Our study in adult hospitalized patients showed that influenza virus was associated with a much more robust cytokine and chemokine response, especially CXCL10. In contrast, RV was associated with an exaggerated T_H2 response, characterized by an elevated eosinophil count and IL-5 (To et al., 2018). The avian influenza virus A(H7N9), which is associated with a much higher case-fatality rate than seasonal influenza virus, can directly infect human mononuclear cells and induce much more heightened cytokine response than seasonal influenza virus (Lee et al., 2017). Host gene expression has been utilized to assess host response. Peripheral blood gene expression studies have shown that influenza and RV patients exhibit different gene expression profile (Zaas et al., 2009; Zhai et al., 2015).

Since the primary site of damage occurs in the lung, the local host response also plays an important role in respiratory virus infection. Host gene response after RV infection has been previously compared to that of influenza virus infection in a human bronchial epithelial cell line BEAS-2B cells using microarray analysis (Kim et al., 2015). However, BEAS-2B has high basal production of interferon-stimulated genes which may affect the response of other host genes (Seng et al., 2014; Hillyer et al., 2018). In this study, we compared the host response between influenza virus and RV in a well-characterized lower airway cell line, Calu-3, using RNA-seq. Calu-3 cell line was chosen because it supports the growth of both influenza and RV (Rajan et al., 2013, 2014; To et al., 2016b). Furthermore, Calu-3 cell line has been used extensively for transcriptomic experiments previously because of high reproducibility (Aevermann et al., 2014; Menachery et al., 2014).

MATERIALS AND METHODS

Viruses

The viruses used in this study were isolated from patients in Hong Kong. Influenza A(H1N1) virus A/HK/415742/2009 and influenza B virus (IBV) B/HK/411989/2011 were used in our previous studies (Zheng et al., 2010; To et al., 2016a). RV 451892/2011 was isolated from a patient with pneumonia, and belongs to RV species A type 1A. Influenza A virus (IAV) and IBV were propagated in Madin Darby canine kidney (MDCK) cells at 37°C, while RV was propagated in RD cells at 33°C. Viruses were concentrated by ultracentrifugation and then resuspended in 1 ml of minimum essential medium (MEM) and Dulbecco's Modified Eagle Medium (DMEM) for influenza virus and RV, respectively. Plaque number was determined on MDCK cells for IAV and IBV, and on H1HeLa cells (ATCC CRL-1958) for RV.

Immunofluorescence Assay for Viral Protein Expression

Immunofluorescence assay for viral protein expression was performed as described previously with modifications (To et al., 2016). Briefly, IAV, IBV or RV was added to Calu-3 cells at 1 multiplicity of infection (MOI) and incubated at 37°C for 1 h. At 1 h post infection, cells were washed and minimum essential free medium was added. At 24 h post-infection, the seeded cells were fixed in chilled acetone at −20°C for 10 min and stained with fluorescein-tagged murine monoclonal antibodies against IAV, IBV (IAV and IBV DFA Reagent, D3® Ultra 8TM DFA Respiratory Virus Screening and Identification Kit, Diagnostic Hybrids, Inc., Quidel, United States) or pan-picornavirus proteins (LIGHT DIAGNOSTICS™ Pan-Enterovirus Reagent, Chemicon International, Temecula, CA, United States) at 37°C for 30 min and examined under fluorescence microscope.

Virus Replication in Calu-3 Cells

Virus infection in Calu-3 cells (ATCC® HTB-55™, passage number: 7) was performed as we described previously with modification (To et al., 2009). Calu-3 was infected with IAV, IBV, and RV at 1 MOI in DMEM-F12 medium. For the determination of MOI, the virus quantification was performed using plaque assay on MDCK cells for IAV and IBV, and H1Hela for RV. Culture supernatant was collected at pre-determined time points and real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR) for each virus was performed. RNA extraction and RT-qPCR were performed as we described previously with modifications (Zhao et al., 2018; Chan et al., 2019). Briefly, viral RNA was extracted using QIAamp® Viral RNA Mini Kit (Qiagen, Hilden, Germany). One step RT-qPCR was performed using AgPath-ID™ One-Step RT-PCR kit (Applied Biosystems, California, United States). The primers and probes for the detection of IAV, IBV, and RV were described previously with modifications (**Supplementary Table S1**; To

et al., 2017, 2018). Real-time RT-qPCR was performed using LightCycler® LC96 instrument (Roche, Mannheim, Germany). Duplicates were performed for each virus for each time point in two independent experiments. Statistical significance was calculated with two-way ANOVA.

Real-Time RT-PCR for Cytokines, Chemokines, and ICAM5

Total RNA was extracted from infected cells using Qiagen RNeasy Mini Kit (Qiagen, Hilden, Germany). Real time reverse transcription polymerase chain reaction (RT-PCR) was performed as described previously with modifications (Wei et al., 2016). Briefly, RNA was reverse-transcribed to cDNA using PrimeScript™ RT reagent kit (Takara Bio Inc., Shiga, Japan). Real time PCR assays for TNF-α, IL6, CXCL10, IFNβ, and ICAM5 were performed in LightCycler 96 system (Roche Applied Sciences, Indianapolis, United States) using primers and probes, and cycling condition as in **Supplementary Table S2**. The expression of house-keeping gene GAPDH was quantified in parallel for RNA normalization. The relative expression of the target genes was calculated by the ΔΔ Ct method. Statistical analysis was performed using PRISM® 6.0 for Windows. Duplicates were performed for each virus at each time point in two independent experiments for measuring cytokine, chemokine and ICAM5 expression. Statistical significance was calculated with two-way ANOVA. One-way ANOVA with Holm-Sidak's multiple comparisons was performed when comparing the fold change for ICAM5 expression in cells infected by different viruses or mock-infected cells test ($^{**}P < 0.001$, $^{***}P < 0.0001$).

Enzyme-Linked Immunosorbent Assay (ELISA) for Cytokines and Chemokines

Cell supernatant was collected from Calu-3 cell infected with IAV, IBV, and RV at 0, 6, 12, and 24 hpi in triplicates from 1 independent experiment. ELISA was done using Human DuoSet ELISA kits for IFN-λ1/λ2/λ3 (Catalog number DY7246, DY1587, DY5259) and CXCL10 (Catalog number DY266) (R&D Systems). Error bars represent SEM. Statistical significance was calculated with two-way ANOVA. Optical density for each well was measured at 450 nM (corrected for 570 nM during analysis) using Beckman Coulter DTX880 Multimode Detector.

RNA-Seq Library Preparation, Sequencing, and Analysis

Total RNA was extracted from two replicates for mock infection (control), RV, IAV, and IBV for time points 0, 6, 12, and 24 h post infection, respectively, using RNeasy (Qiagen Hilden, Germany). RNA quantity and quality were assessed using NanoDrop Spectrophotometer and Bioanalyzer. Library preparation and Illumina sequencing (paired-end sequencing of 101 bp) were performed at University of Hong Kong, Centre for Genomic Sciences (HKU, CGS). cDNA libraries were prepared by KAPA Stranded mRNA-Seq Kit. One microgram of total RNA was used as starting material. Manufacturer's protocol was followed. In brief, poly-A containing mRNA was collected by using poly-T oligo-attached magnetic beads. The purified mRNA was broken down into short fragments and was applied as template to synthesize the first-strand cDNA by using random hexamer-primer and reverse transcriptase. In the second strand cDNA synthesis, the mRNA template was removed and a replacement strand was generated to form the blunt-end double-stranded (ds) cDNA. The ds cDNA underwent 3′ adenylation and indexed adaptor ligation. The adaptor-ligated libraries were enriched by 10 cycles of polymerase chain reaction (PCR). The libraries were denatured and diluted to optimal concentration and applied in the cluster generation steps. HiSeq PE Cluster Kit v4 with cbot was used for cluster generation on the flow cell. Illumina HiSeq SBS Kit v4 was used for paired-end 101 bp sequencing. Whole dataset has been deposited in the NCBI Sequence Read Archive with accession number (PRJNA609228).

Bioinformatics Analysis

RNA-seq raw data were processed following the GEO2RNA-Seq pipeline (Seelbinder et al., 2019) a RNA-Seq pre-processing workflow and package for analyzing read files, trimming of raw reads, mapping on reference genomes, counting reads per gene and finding significant differentially expressed genes (DEGs). Quality of raw read data was checked using FastQC version 0.11.5. Reads were quality- and adapter-trimmed using Trimmomatic version 0.36. Reads were rRNA-filtered using SortMeRNA version 2.1 with a single rRNA database concatenated from all rRNA databases shipped with SortMeRNA. Reads were mapped against the human reference genome and corresponding annotation GRCh38 89 (2017-05-07, obtained from Ensembl) using TopHat2 version 2.1.0.

Reads per feature (gene) were counted using Rsubread's featureCounts version 1.20.6. The output off all pre-processing steps was reviewed using MultiQC version 1.1 (**Supplementary Data S1**). Additionally, human genome coverage and exome coverage per sample was calculated using R version 3.2.0 (**Supplementary Data S2**). Hierarchical clustering of MRN-normalized read counts using the unweighted pair group method with arithmetic mean (UPGM) metric was calculated with the "hclust" function from the R base package "stats" version 3.2.0. Principal component analysis (PCA) of MRN-normalized read counts was done with the "prcomp" function from the R base package "stats" version 3.2.0. DEGs were determined using four different tools, including DESeq (Anders and Huber, 2010), DESeq2 (Love et al., 2014), Limma (Ritchie et al., 2015), and EdgeR (Robinson et al., 2010). A gene was considered to be differentially expressed if the expression change was reported significantly different by all tools with a $p \leq 0.01$. Using the consensus identification of DEGs by the aforementioned four tools assures controlling the false positive rate and increases the specificity at the expense of reduced sensitivity. DEGs were summarized together with log2 of fold changes based on MRN, TPM or RPKM normalization (**Supplementary Data S3**). MRN normalization was used for further analysis throughout the manuscript, since MRN was reported to be superior over other normalization schemes (Maza et al., 2013). Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) was performed by using g:Profiler (Raudvere et al., 2019). Enrichment score

was calculated as the –log10 (*P*-value) as described previously (Morrison et al., 2014). Hierarchical clustering was performed for DEGs with an absolute log2 fold change greater than two using Ward's method as implemented by the R package "pheatmap," version 1.0.12.

RESULTS

Virus Infection in Calu-3 Cells

To confirm whether Calu-3 cells are susceptible to IAV, IBV, and RV infection, antigen expression and viral replication were determined. Immunofluorescence assay showed that IAV, IBV and RV could infect Calu-3 cells at 1 MOI (**Figure 1A**). Viral load assay showed that all three viruses could replicate in Calu-3 cells (**Figure 1B**). Next, cytokine and chemokine expression were measured to determine the host response of Calu-3 cells after IAV, IBV or RV infection. TNF-α, IL-6, CXCL10, and IFN-β were induced after infection with all three viruses (**Figure 1C**). ELISA for the IFN-λ1/λ2/λ3, and CXCL10 showed detection of proteins at 12 hpi for IBV and 24 hpi for IAV and RV (**Figure 1D**). Results shown in **Figures 1C,D** were consistent with RNA-seq data.

Longitudinal RNA-Seq During Infection

RNA-seq was performed to determine the gene expression profile during IAV, IBV, and RV infection. Using software from Illumina (bcl2fastq), sequencing reads were assigned into individual samples with each sample having an average throughput of 6.6 Gb (**Supplementary Table S3**) and a total throughput of 210 Gb. In terms of sequence quality, an average of 93% of the bases achieved a quality score of Q30 where Q30 denotes the accuracy of a base call to be 99.9%. MultiQC and mapping statistics table showed very good assignment rates of over 85% reads assigned to the human reference genome for all samples (**Supplementary Data S1, S2**).

A first global overview by hierarchical clustering of read counts showed a clear separation of the IAV, IBV, RV from-mock infection samples at 12 and 24 h post infection (hpi) (**Figures 2A,B**). While IBV showed already clear separation at 6 hpi, there was no clear separation between IAV, RV and mock-infected cells at 6 hpi, however (**Figure 2C**). Expression was notable for several mitochondrial genes and influenced explained variance in the PCAs (**Supplementary Figure S1**). In addition, several of these most influencing genes, e.g., *IFIT2*, are known to possess immune system functionality.

Differentially Expressed Genes (DEGs) Analysis

In order to elucidate the difference between RV and influenza virus infection, we compared the DEGs after RV, IAV, and IBV infection against mock and found a high agreement across all four tools used for DEG identification (**Supplementary Figure S2**). Of note, using the intersect of four different tools for identification of DEGs ensured high specificity, and thus reliability, at a tolerable sensitivity drop (**Supplementary Figure S2**). We observed a gradual increase in the number of significantly upregulated and downregulated DEGs for all three viral infections from 6 to 24 hpi (**Supplementary Figure S2**). When compared to mock-infected cells, the number of DEGs was lower for RV than those of IAV and IBV infection at each time point (**Supplementary Figure S3**). For IAV the number of DEGs increased from 147 to 2306 and for IBV from 1692 to 6788 over time. When compared to mock-infected cells, RV-infected cells gave rise to the lowest number of DEGs when compared to IAV or IBV-infected cells at all time points. In particular, there were only two DEGs for RV at 6 hpi, with very low log2 fold-change (<1.2) (**Supplementary Data S3**). These results suggest a delayed host response for RV when compared to influenza viruses.

Since only two DEGs were found for RV at 6 hpi, we focused on the comparison for 12 and 24 hpi (**Table 1** and **Supplementary Data S3**). At 12 and 24 hpi, interferon-related genes and CXCL10 were highly expressed for both RV and IAV/IBV infection. IFN-λ genes (IFNL1, IFNL2, IFNL3) were the most highly expressed among the interferon genes at both 12 and 24 h post-infection for all viruses. CXCL10 was the top DEGs for all 3 viruses at 24 hpi, but also within the top 6 DEGs at 12 hpi.

Next, we investigated the change over time for all 34 DEGs that originate from investigating the top 20 upregulated DEGs per virus infections vs. mock for 12 hpi. Thirty one out of 34 are also DEGs in the condition 12 hpi against 0 hpi in either mock, IAV, IBV, or RV. Eight of these 31 DEGs are shared by mock and all virus conditions, while four are shared by mock and IAV and IBV (**Supplementary Figure S4**). However, none of these DEGs were regulated in the same direction when comparing mock expression with virus infections for 12 hpi against 0 hpi. In fact, significant downregulation occurred for 12 of the 31 DEGs only in mock, whereas none were upregulated in mock at 12 hpi.

In addition to interferons and CXCL10, several other genes were also found to be highly expressed during RV and IAV/IBV virus infection. At 12 or 24 hpi, two genes with known immune-related functions (RSAD2 and IDO1) were found to be among the top 20 DEGs for all 3 viruses (**Table 1**). ICAM5 was more significantly induced in cells infected with RV than those infected with IAV or IBV at 12 hpi (11th top DEG for RV; log2 fold change 5.25) (**Table 1** and **Figure 3A**). We additionally performed ICAM5 monoplex real-time RT-PCR to confirm the expression of ICAM5 gene. ICAM-5 was expressed at a significantly higher level at 12 hpi for RV than that of IAV or IBV (**Figure 3B**).

Gene Set Enrichment Analysis (GSEA)

First, GSEA was performed for all three viruses at the 12 hpi time point. The majority of perturbed pathways were related to innate immune and virus response, interferons and cytokine signaling and revealed no major differentiation among different virus infections (**Table 2** and **Supplementary Data S4**). To dissect the enriched function in the transcriptomics response further, we focused on DEGs with a log2 fold change of at least two across of all time points. We identified three major sub-clusters of the remaining 1888 transcripts across all three viruses by hierarchical clustering. GSEA was performed for each individual sub-cluster and the major significantly enriched pathways were identified, respectively (**Figure 4** and **Supplementary Data S5**).

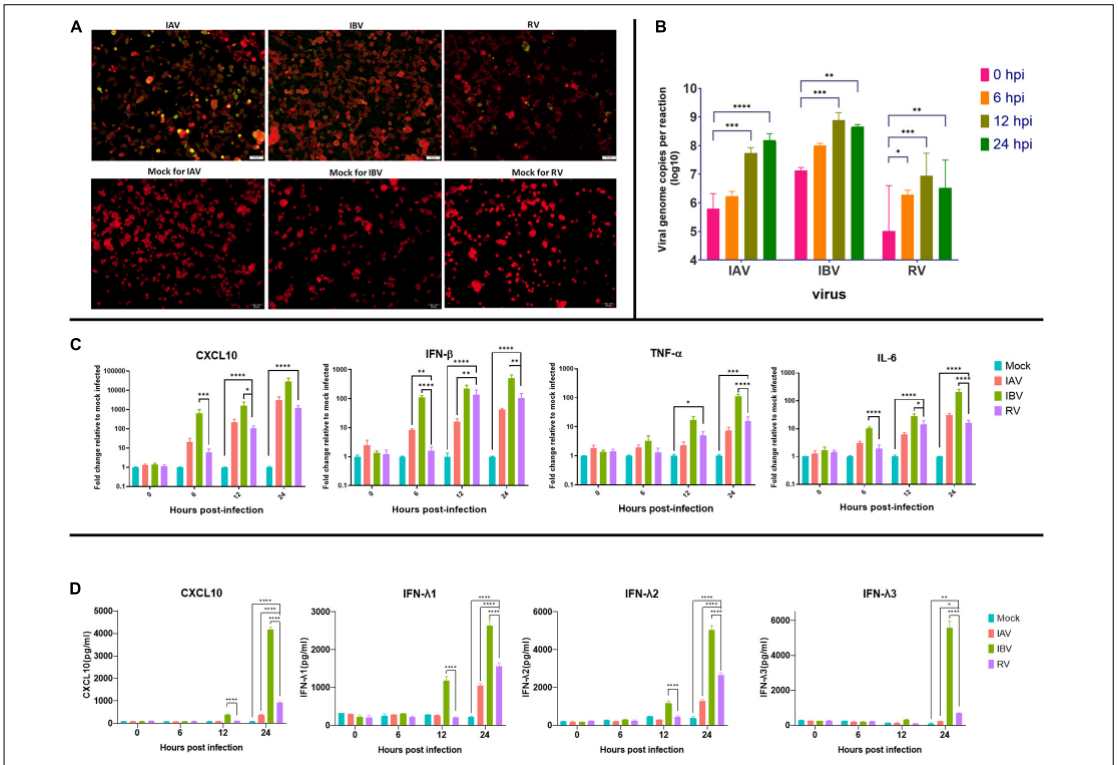


FIGURE 1 | Infection of influenza A virus (IAV), influenza B virus (IBV) or rhinovirus (RV) in Calu-3 cells. **(A)** Calu-3 cells were infected with IAV, IBV, or RV at 1 MOI. IAV, IBV, and RV antigen expression was determined at 24 h post infection. Antigen expression was determined using fluorescein-tagged murine monoclonal antibodies against IAV, IBV, or RV. Mock-infected cells stained with respective monoclonal antibodies against IAV, IBV, or RV are shown in the bottom row. White scale bar = 50 μm. **(B)** Multicycle growth assay. Calu-3 cells were infected with IAV, IBV, or RV at 1 MOI. Viral load was determined using real-time RT-qPCR. **(C)** Cytokine and chemokine expression of Calu-3 cells infected with IAV, IBV, or RV at 1 MOI. Cytokine expression was determined using real time RT-PCR. GAPDH was used for normalization of gene expression. **(D)** Cytokine and chemokine protein expression of Calu-3 cells infected with IAV, IBV, or RV at 1 MOI. Protein expression was determined using ELISA. Bars **(B,C)** represent means (error bars show standard error of mean) of duplicates in two independent experiments. Bars **(D)** represent means (error bars show standard error of mean) of triplicates in one independent experiment. Statistical significance (for **B–D**) was calculated with two-way ANOVA. (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$). hpi, hours post infection; MOI, multiplicity of infection.

This analysis showed that regulation of some major immune response biological processes like defense response to virus were induced by infection of all three virus types.

Since the number of DEGs specific to IBV infection was substantially higher (**Supplementary Figure S5**), we investigated enriched categories only for IBV regulated genes (**Supplementary Data S6**). Next to non-specific categories related to cell cycle organization, regulation and communication, we identified the enriched process "cellular response to stress," which included DEGs like *CCL2*, *IL1B,* or *CD34* that were not differentially regulated by IAV and RV (**Supplementary Data S7**).

We additionally interrogated the specific role of downregulated DEGs in our data set, but could only find non-specific cell cycle, cell organization and cell communication GO biological process categories, which showcase that the cells shift to response to the infecting viruses rather than continuing cell cycle processes (**Supplementary Figure S6** and **Supplementary Data S8**).

In addition, we interrogated if different viruses exclusively cause differentially regulated pathways enriched by DEGs. The analysis showed that steroid-related pathways which are specifically significantly regulated in RV-infected cells relative to mock-infected subjects at 12 hpi (**Supplementary Data S9**). These pathways include "cellular response to mineralocorticoid stimulus" (GO:0071389), "glucocorticoid mediated signaling pathway" (GO:0043402), "regulation of glucocorticoid mediated signaling pathway" (GO:1900169), and "response to dexamethasone" (GO:0071548) (**Supplementary Data S9**). We identified 13 DEGs, such as EGFR or FOXO3, to be common among steroid-relevant pathways (**Figure 5**). Interestingly, we find upregulated IFNB1 and ICAM1 among these with proposed steroid-pathway relevancy as well. Of

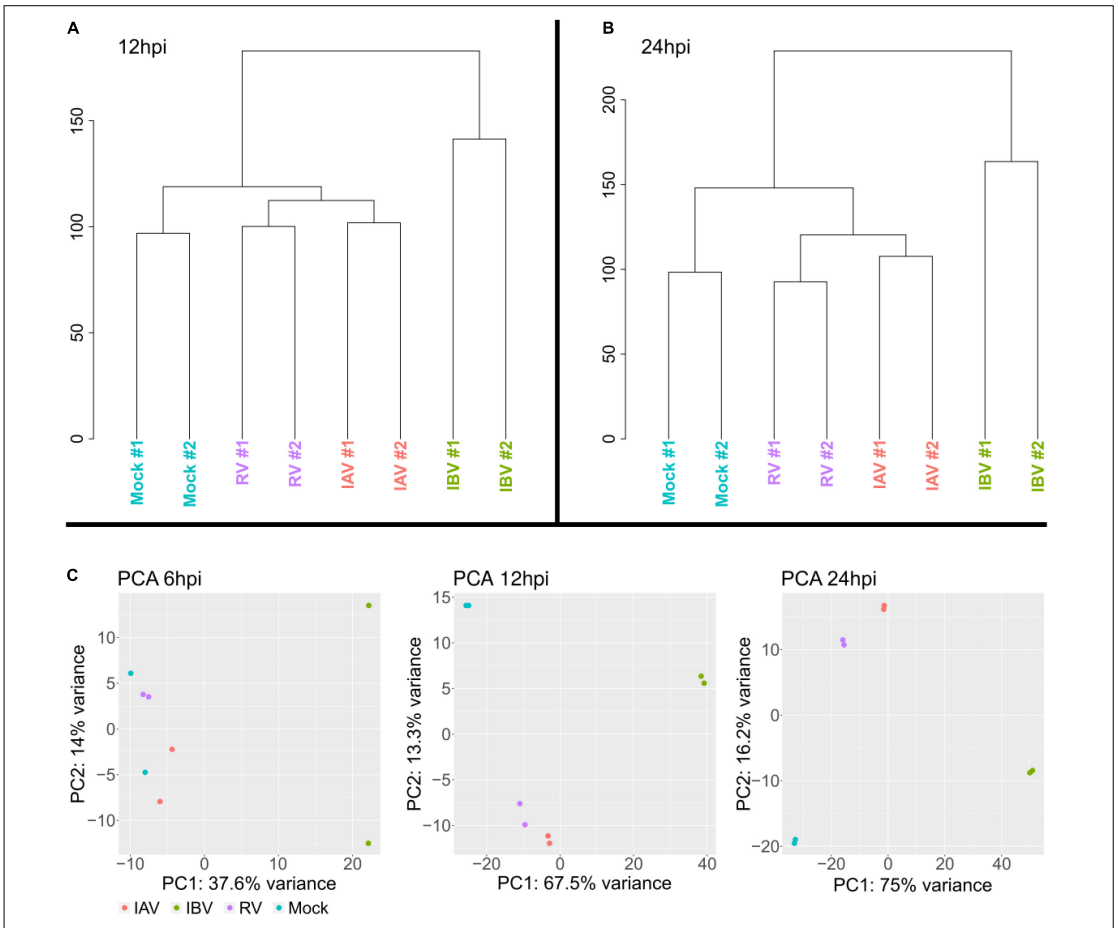


FIGURE 2 | Hierarchical clustering of gene expression for **(A)** 12 hpi time point and **(B)** 24 hpi time point showing a distinct gene expression for IAV, IBV, and RV. **(C)** Principal component analysis for 6, 12, and 24 hpi.

note, ICAM1 has been indicated to be increased by RV infection and identified to be a target for the development of therapeutic interventions for virus-induced asthma exacerbation (Papi and Johnston, 1999).

DISCUSSION

Both RV and influenza virus can cause severe respiratory tract infection (Hung et al., 2017). However, there are significant differences in their clinical manifestations and immune or inflammatory response (To et al., 2018). In this study, we used a transcriptomic approach to systemically compare the host response between RV and influenza virus infection in a human lower airway epithelial cell line. We have found that RV induced a more delayed and blunted host response than influenza virus infection. Interferon response and other innate immune response predominated for both RV and influenza virus infection. Interestingly ICAM5 was the only gene that was significantly upregulated for RV but not for influenza virus infection.

Our study is unique in several aspects. First, we used virus isolates that are isolated in recent years. Previous transcriptomic studies used classical strains that were isolated many years ago (Kim et al., 2015), and their results may not be representative of the contemporary viruses. Second, we infected Calu-3 cells using the same MOI to reduce the bias due to different virus titers.

The number of DEGs was much lower for RV than those of influenza virus infection. This is consistent with our previous study, which showed that the cytokine and chemokine responses were significantly lower for patients with RV infection than those with influenza virus infection (To et al., 2018). This is also consistent with the study by Zhai et al. (2015) that the

TABLE 1 | Top 20 upregulated DEGs when compared with mock-infected cells.

	IAV vs. mock		IBV vs. mock		RV vs. mock	
Hours post-infection	**Gene name**	**log2 (Fold Change)**	**Gene name**	**log2 (Fold Change)**	**Gene name**	**log2 (Fold Change)**
12	*IFNL2*	7.10	*IFNL1*	9.88	*IFNL2*	7.61
	IFNL1	7.09	*CXCL10*	9.63	*IFNL3*	7.38
	IFNL3	6.89	*IFNL2*	9.55	*IFNB1*	7.34
	RSAD2	6.76	*IFNL3*	9.36	*IFNL1*	7.14
	CXCL10	6.68	*IFNB1*	8.95	*AC011511.5*	6.13
	IFNB1	6.44	*CXCL11*	8.91	*FAP*	6.01
	IDO1	6.28	*IDO1*	8.59	*TULP2*	5.77
	BATF2	6.19	*CCL5*	8.40	*RPL7P19*	5.51
	MX2	5.92	*BATF2*	8.14	*ZBP1*	5.38
	IFIT2	5.86	*GBP5*	8.09	*AL133163.1*	5.36
	OAS2	5.76	*CXCL9*	8.01	*ICAM5*	5.25
	OR52K3P	5.58	*RSAD2*	7.89	*CXCL10*	5.12
	IFIT3	5.53	*GBP4*	7.70	*RSAD2*	5.02
	AC005515.1	5.53	*ZBP1*	7.42	*BATF2*	4.65
	IFIT1	5.52	*CX3CL1*	7.33	*IFIT2*	4.63
	GBP4	5.52	*AC005515.1*	7.24	*MX2*	4.61
	TRIM22	5.43	*LRP2*	7.18	*IFIT1*	4.48
	ZBP1	5.40	*MX2*	7.10	*OAS2*	4.46
	ETV7	5.40	*CD69*	6.99	*IFIT3*	4.35
	CMPK2	5.39	*CSF3*	6.94	*CMPK2*	4.31
24	*CXCL10*	10.32	*CXCL10*	12.08	*CXCL10*	8.90
	CXCL11	9.20	*CXCL11*	11.58	*IFNL3*	7.68
	ZBP1	9.01	*IFNL1*	10.44	*IFNL2*	7.62
	IDO1	8.56	*CCL5*	10.35	*CXCL11*	7.59
	IFNL2	8.31	*CSF3*	10.21	*ZBP1*	7.45
	KLHDC7B	8.26	*IFNL2*	10.07	*RSAD2*	7.21
	IFNL3	8.21	*IFNL3*	9.90	*MX2*	7.13
	IFNL1	8.11	*TNF*	8.97	*IFNL1*	6.92
	RSAD2	7.98	*RSAD2*	8.79	*IDO1*	6.90
	TRIM22	7.76	*IFNB1*	8.76	*OAS2*	6.89
	SLC15A3	7.69	*IL6*	8.73	*TRIM22*	6.79
	MX2	7.64	*ZBP1*	8.41	*CMPK2*	6.77
	CCL5	7.55	*IDO1*	8.33	*IFNB1*	6.71
	BST2	7.48	*AL021578.1*	8.28	*KLHDC7B*	6.53
	OAS2	7.27	*CX3CL1*	8.12	*SLC15A3*	6.47
	GBP4	7.21	*HSPA6*	8.04	*GBP4*	6.09
	AC005515.1	7.13	*CXorf49B*	7.84	*FAP*	6.01
	CX3CL1	7.01	*OASL*	7.83	*ETV7*	5.82
	CMPK2	6.94	*GBP4*	7.82	*BST2*	5.82
	ETV7	6.82	*CCL22*	7.71	*BATF2*	5.77

The log2 fold changes are based on MRN normalization (cf. ***Supplementary Data S3****).*

overall gene expression was much stronger for patients with influenza virus infection than those with RV infection. We do note that even though the same MOI were used for all viruses a higher viral load was observed for IBV (**Figure 1B**), which might be partially responsible for the elevated number of DEGs expressed in Calu-3 cells when challenged with IBV. This additional set of DEGs includes interesting genes, e.g., *CCL2*, *IL1B* or *CD34* (**Supplementary Data S7**), which should be further investigated.

IFNL1, which encodes IFN-λ1, was among the top DEG for all 3 viruses at 12 h post-infection. *IFNL2* and *IFNL3*, which encode IFN-λ2 and IFN-λ3, respectively, were also induced in RV or influenza virus infected cells, but were not detectable for mock-infected cells. IFN-λ is a type III interferon, which acts via the IFN-λ receptor (Syedbasha and Egli, 2017). Through ELISA, IFN-λ was shown to be significantly expressed mainly around 24 hpi across all viruses, IBV infected samples showed significant IFN-λ expression much earlier, at 12 hpi (**Figure 1D**).

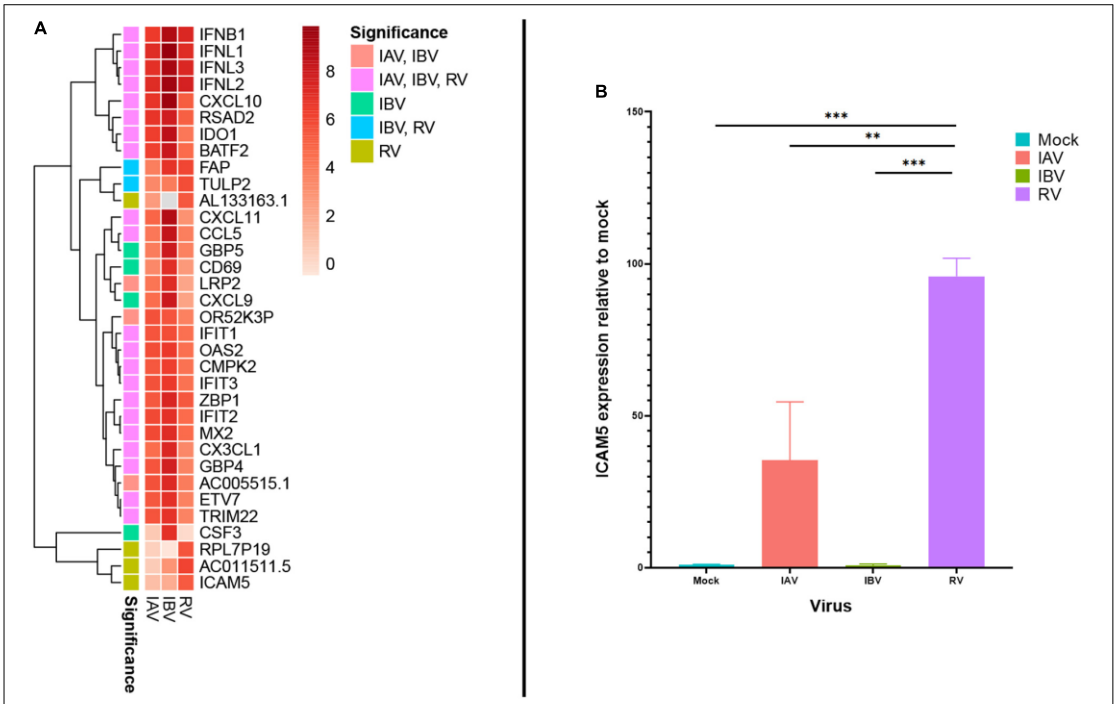


FIGURE 3 | (A) Heatmap for top-20 differentially expressed genes with an absolute log2 fold change ≥ 2 across all virus infections with significant infection annotated. **(B)** ICAM5 expression. Monoplex ICAM5 specific real time RT-PCR was performed for IAV, IBV, and RV-infected cells. $^{**}P < 0.01$; $^{***}P < 0.001$.

IFN-λ1 is an important antiviral cytokine. IFN-λ has been shown to be important for the immune defense against RV. Asthmatic patients are more prone to severe RV infection, which correlated with a poorer induction of IFN-λ in airway epithelial cells and alveolar macrophages isolated from asthmatic patients than those isolated from healthy individuals (Contoli et al., 2006). In a mouse model, IFN-λ has been shown to reduce influenza virus replication, modulate immune response and protect mice from IAV infection (Davidson et al., 2016). Inhibition of *IFNL3* has been shown to increase antibody response against IAV infection (Egli et al., 2014). In a rhesus macaque model infected with influenza virus A(H5N1), IFNλ genes were also found to be highly expressed in the lung tissues (Shinya et al., 2012).

CXCL-10 was highly expressed for all infections. This is compatible with our previous findings in patients, in which *CXCL10* expression was induced at high levels for both influenza virus and RV infection (To et al., 2018). This is also shown through the ELISA results, where we can see significant expression of CXCL10 at 24 hpi (**Figure 1D**). Our results also corroborate with the results from RNA-seq experiments on the nasopharyngeal swabs from patients with respiratory virus infection, in which *CXCL10* was one of the genes that could be used in identifying patients with respiratory virus infection (Landry and Foxman, 2018).

At 12 or 24 h, most of common genes among the top 20 DEGs were interferons or interferon-inducible genes *(IFNL1, IFNB1, RSAD2, IDO1)* or chemokines *(CXCL10, CXCL11)*. *RSAD2* and *IDO1* are mainly triggered by interferons as part of the concerted counteraction against viral infection (Duschene and Broderick, 2012; Gaelings et al., 2017). Interestingly, *IRF7*, though differentially regulated (**Supplementary Data S3**), was not among the top differentially regulated genes, although it possesses a pivotal role in virus triggered IFN type I induction. *GBP4* is among our top regulated genes across all viruses and was reported to interact with *IRF7* in a negative manner (Hu et al., 2011). *BATF2* is a transcription factor that controls the differentiation of dendritic cells. *BATF2* has been identified to be one of the biomarkers that can predict the progression of active tuberculosis for individuals who have close contact with tuberculosis patients (Roe et al., 2019). Batf2-/- mice had more severe *Trypansoma* infection (Kitada et al., 2017).

A previous study by Kim et al. (2015) compared the transcriptomic profiles between RV and influenza virus in a human bronchial cell line BEAS-2B. Similar to our study, they have demonstrated that IFNB1, CXCL10, CXCL11, CCL5 was upregulated in both IAV and RV infection. However, while our

TABLE 2 | Top 10 reactome pathways that enriched for IAV, IBV, or RV infection when compared with mock infection at 12 hpi using g:Profiler analysis.

	IAV			IBV			RV		
	Reactome pathway	**FDR-adjusted *p*-value**	**Enrichment score**	**Reactome pathway**	**FDR-adjusted p-value**	**Enrichment score**	**Reactome pathway**	**FDR-adjusted *p*-value**	**Enrichment score**
1	Interferon signaling	2.1E-27	26.69	Cell cycle	1.4E-20	19.86	Interferon signaling	4.0E-25	24.39
2	Cytokine signaling in immune system	4.1E-27	26.38	Influenza infection	1.7E-20	19.78	Interferon alpha/beta signaling	2.5E-22	21.60
3	Interferon alpha/beta signaling	1.3E-24	23.90	L13a-mediated translational silencing of ceruloplasmin expression	2.6E-20	19.58	Cytokine signaling in immune system	1.2E-17	16.92
4	Immune system	2.3E-21	20.64	Cytokine signaling in immune system	3.4E-20	19.47	Interferon gamma signaling	8.4E-13	12.08
5	Eukaryotic translation elongation	7.4E-20	19.13	Cap-dependent translation initiation	4.4E-20	19.35	Immune system	1.9E-08	7.72
6	Peptide chain elongation	7.4E-20	19.13	Eukaryotic translation initiation	4.4E-20	19.35	Interleukin-4 and Interleukin-13 signaling	6.1E-07	6.21
7	Viral mRNA translation	7.4E-20	19.13	Influenza life cycle	6.5E-20	19.18	Interleukin-10 signaling	8.5E-06	5.07
8	Nonsense mediated decay (NMD) enhanced by EJC	1.4E-19	18.85	Signaling by ROBO receptors	8.8E-20	19.06	Antiviral mechanism by IFN-stimulated genes	2.0E-05	4.70
9	Nonsense-mediated decay (NMD)	1.4E-19	18.85	Cell cycle, Mitotic	1.1E-19	18.95	FOXO-mediated transcription	9.4E-05	4.03
10	Influenza infection	1.6E-19	18.79	GTP hydrolysis and joining of the 60S ribosomal subunit	1.8E-19	18.74	Signaling by interleukins	1.5E-04	3.82

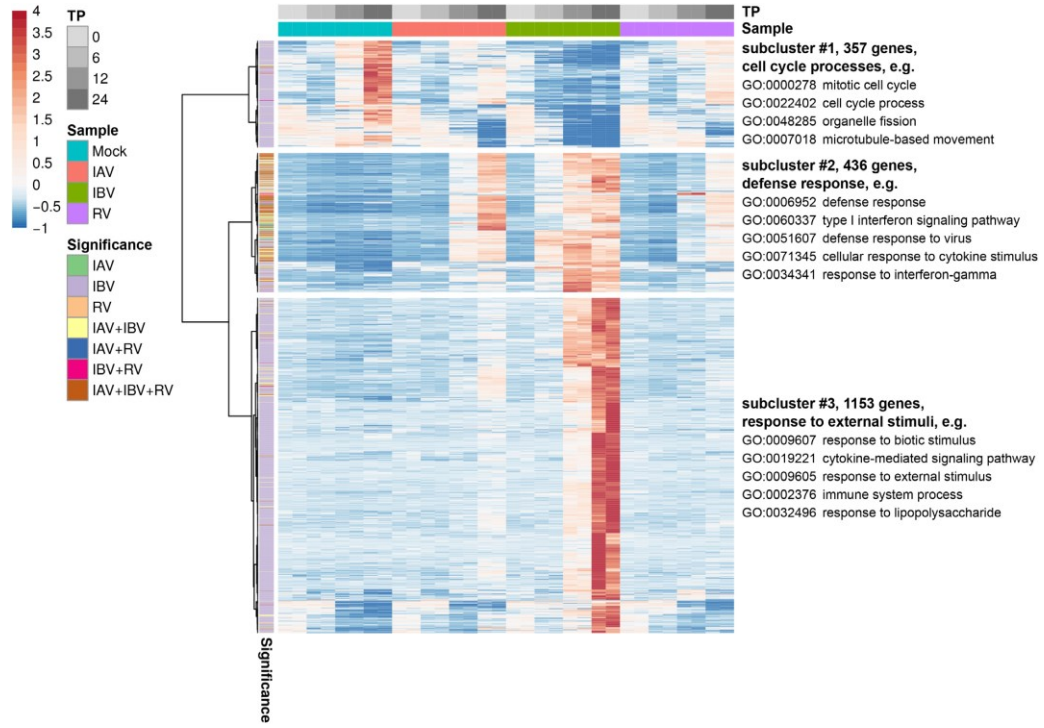


FIGURE 4 | Heatmap for differentially expressed genes with an absolute log2 fold change ≥ 2 across all virus infections. Three subclusters as determined by hierarchical clustering are shown and GSEA was performed for each individual sub-cluster and the major significantly enriched pathways are indicated.

current study in Calu-3 cells showed significant induction of IFNλ genes, the study by Kim et al. in BEAS-2B cell line did not show induction of these genes after infection. One major limitation associated with BEAS-2B cell line is that there is a high basal production of interferon-stimulated genes, such as IRF7, ISG15, MX1, STING, which may affect the response of other host genes and also lead to Influenza A virus resistance (Seng et al., 2014).

ICAM5 was strongly expressed during RV infection, which was verified by monoplex RT-qPCR. Although ICAM5 is unlikely to be a single factor to account for the difference between RV and influenza virus, the fact that ICAM5 is much more highly expressed in RV infected cells than both influenza A and influenza B virus strongly suggests that ICAM5 may play a major role in the pathogenesis of RV. ICAM5 is a known receptor for enterovirus D68, which also belongs to the *Picornaviridae* family (Wei et al., 2016). Nonetheless. the significance of ICAM5 on RV infection deserves further studies.

Our network pathway enrichment analysis showed that steroid-related pathways are enriched. RV is more likely to be associated with acute exacerbation of asthma than influenza virus infection (To et al., 2019). A previous study has shown that RV infection leads to steroid-resistance in airway epithelium (Papi et al., 2013). Therefore, our transcriptomic analysis reveals that the difference in steroid pathways may be associated with the clinical manifestations.

Though there are studies which have analyzed the transcriptome of influenza virus or RV infection, most did not compare these viruses together. In a study with experimental human infections, comparison of blood mRNA expression showed that *SOCS1* gene were uniquely expressed for RV infection when compared with IAV and RSV (Zaas et al., 2009). However, in our study, *SOCS1* was upregulated for all 3 viruses, with higher levels among influenza virus infected cells.

Our study has demonstrated that there are some important differences like *ICAM5* expression which may explain the clinical findings of these viral infections. To generalize on our findings future work should include additional subtypes from different viruses on top of the strains we used in the present study. Our investigation revealed a number of genes that are similarly expressed upon infection with any of the studied viruses. These genes may yield broad spectrum antivirals for the treatment of influenza virus and RV infection.

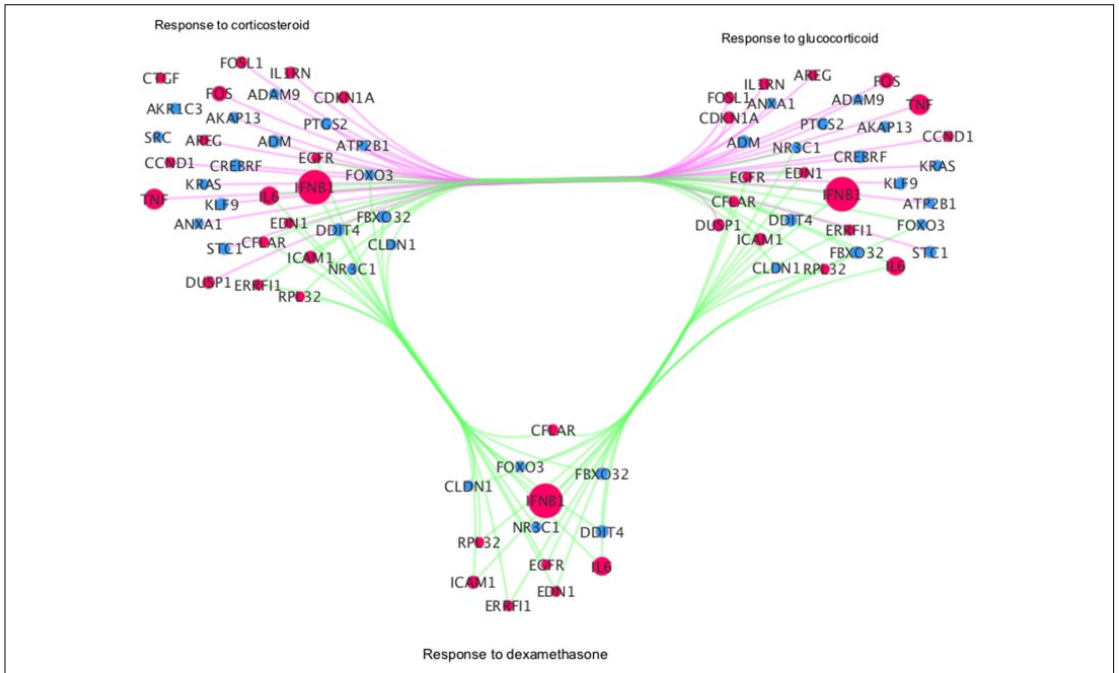


FIGURE 5 | Network visualization of GSEA derived categories related to steroid biosynthesis. Red: up-regulated, Blue: down-regulated; Size: log2 fold change. Green edges: gene is shared among all categories; magenta: genes are only shared among response to corticosteroid and response to glucocorticoid categories.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI Sequence Read Archive (accession number PRJNA609228).

AUTHOR CONTRIBUTIONS

TD and KT contributed to the conception and design of the work, data acquisition, data analysis, interpretation of the data, and drafting and revising the manuscript critically for intellectual content. M-LY and K-YY contributed to the conception and design of the work, data analysis, interpretation of the data, and revising the manuscript critically for intellectual content. GP contributed to the design of the work, data analysis, interpretation of the data, and revising the manuscript critically for intellectual content. SS, MM, W-LW, AN, CY, AL, TW, and K-HC contributed to the data analysis and interpretation of data, and revising the manuscript critically for intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Health and Medical Research Fund (HMRF), Food and Health Bureau, Hong Kong SAR Government (Ref. No. HKM-15-M03), and the donations of Richard Yu and Carol Yu, Michael Seak-Kan Tong, Respiratory Viral Research Foundation Limited, Hui Ming, Hui Hoy and Chow Sin Lan Charity Fund Limited, Chan Yin Chuen Memorial Charitable Foundation, and Marina Man-Wai Lee. GP, MM, and AL would like to thank the Deutsche Forschungsgemeinschaft (DFG) CRC/Transregio 124 "Pathogenic fungi and their human host: Networks of interaction," subproject B5 and INF.

ACKNOWLEDGMENTS

We acknowledge the Centre for Genomic Sciences of the University of Hong Kong for providing the sequencing services. We would like to thank Dr. Thomas Wolf for his contribution in the initial stages of the RNA-seq analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01580/full#supplementary-material

REFERENCES

Aevermann, B. D., Pickett, B. E., Kumar, S., Klem, E. B., Agnihothram, S., Askovich, P. S., et al. (2014). A comprehensive collection of systems biology data characterizing the host response to viral infection. *Sci Data* 1:140033.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Chan, W. M., Wong, L. H., So, C. F., Chen, L. L., Wu, W. L., Ip, J. D., et al. (2019). Development and evaluation of a conventional RT-PCR for differentiating emerging influenza B/Victoria lineage viruses with hemagglutinin amino acid deletion from B/Yamagata lineage viruses. *J. Med. Virol.* 92, 382–385. doi: 10.1002/jmv.25607

Contoli, M., Message, S. D., Laza-Stanca, V., Edwards, M. R., Wark, P. A., Bartlett, N. W., et al. (2006). Role of deficient type III interferon-lambda production in asthma exacerbations. *Nat. Med.* 12, 1023–1026. doi: 10.1038/nm1462

Davidson, S., McCabe, T. M., Crotta, S., Gad, H. H., Hessel, E. M., Beinke, S., et al. (2016). IFNlambda is a potent anti-influenza therapeutic without the inflammatory side effects of IFNalpha treatment. *EMBO Mol. Med.* 8, 1099–1112. doi: 10.15252/emmm.201606413

Duschene, K. S., and Broderick, J. B. (2012). Viperin: a radical response to viral infection. *Biomol. Concepts* 3, 255–266. doi: 10.1515/bmc-2011-0057

Egli, A., Santer, D. M., O'Shea, D., Barakat, K., Syedbasha, M., Vollmer, M., et al. (2014). IL-28B is a key regulator of B- and T-cell vaccine responses against influenza. *PLoS Pathog.* 10:e1004556. doi: 10.1371/journal.ppat.1004556

Gaelings, L., Soderholm, S., Bugai, A., Fu, Y., Nandania, J., Schepens, B., et al. (2017). Regulation of kynurenine biosynthesis during influenza virus infection. *FEBS J.* 284, 222–236. doi: 10.1111/febs.13966

Hillyer, P., Shepard, R., Uehling, M., Krenz, M., Sheikh, F., Thayer, K. R., et al. (2018). Differential responses by human respiratory epithelial cell lines to respiratory syncytial virus reflect distinct patterns of infection control. *J. Virol.* 92:e02202-17.

Hu, Y., Wang, J., Yang, B., Zheng, N., Qin, M., Ji, Y., et al. (2011). Guanylate binding protein 4 negatively regulates virus-induced type I IFN and antiviral response by targeting IFN regulatory factor 7. *J. Immunol.* 187, 6456–6462. doi: 10.4049/jimmunol.1003691

Hung, I. F., Zhang, A. J., To, K. K., Chan, J. F., Zhu, S. H., Zhang, R., et al. (2017). Unexpectedly higher morbidity and mortality of hospitalized elderly patients associated with rhinovirus compared with influenza virus respiratory tract infection. *Int. J. Mol. Sci.* 18:259. doi: 10.3390/ijms18020259

Ieven, M., Coenen, S., Loens, K., Lammens, C., Coenjaerts, F., Vanderstraeten, A., et al. (2018). Aetiology of lower respiratory tract infection in adults in primary care: a prospective study in 11 European countries. *Clin. Microbiol. Infect* 24, 1158–1163. doi: 10.1016/j.cmi.2018.02.004

Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., et al. (2018). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* 391, 1285–1300.

Jain, S., Self, W. H., Wunderink, R. G., Fakhran, S., Balk, R., Bramley, A. M., et al. (2015). Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N. Engl. J. Med.* 373, 415–427.

Kim, T. K., Bheda-Malge, A., Lin, Y., Sreekrishna, K., Adams, R., Robinson, M. K., et al. (2015). A systems approach to understanding human rhinovirus and influenza virus infection. *Virology* 486, 146–157. doi: 10.1016/j.virol.2015.08.014

Kitada, S., Kayama, H., Okuzaki, D., Koga, R., Kobayashi, M., Arima, Y., et al. (2017). BATF2 inhibits immunopathological Th17 responses by suppressing Il23a expression during Trypanosoma cruzi infection. *J. Exp. Med.* 214, 1313–1331. doi: 10.1084/jem.20161076

Landry, M. L., and Foxman, E. F. (2018). Antiviral response in the nasopharynx identifies patients with respiratory virus infection. *J. Infect. Dis.* 217, 897–905. doi: 10.1093/infdis/jix648

Lee, A. C. Y., To, K. K. W., Zhu, H., Chu, H., Li, C., Mak, W. W. N., et al. (2017). Avian influenza virus A H7N9 infects multiple mononuclear cell types in peripheral blood and induces dysregulated cytokine responses and apoptosis in infected monocytes. *J. Gen. Virol.* 98, 922–934. doi: 10.1099/jgv.0.000751

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Marcone, D. N., Carballal, G., Iraneta, M., Rubies, Y., Vidaurreta, S. M., and Echavarria, M. (2017). Nosocomial transmission and genetic diversity of rhinovirus in a neonatal intensive care unit. *J. Pediatr.* 193, 252.e1–255.e1. doi: 10.1016/j.jpeds.2017.09.013

Maza, E., Frasse, P., Senin, P., Bouzayen, M., and Zouine, M. (2013). Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments: a matter of relative size of studied transcriptomes. *Commun. Integr. Biol.* 6:e25849. doi: 10.4161/cib.25849

Menachery, V. D., Eisfeld, A. J., Schafer, A., Josset, L., Sims, A. C., Proll, S., et al. (2014). Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *mBio* 5:e1174-14.

Morrison, J., Josset, L., Tchitchek, N., Chang, J., Belser, J. A., Swayne, D. E., et al. (2014). H7N9 and other pathogenic avian influenza viruses elicit a three-pronged transcriptomic signature that is reminiscent of 1918 influenza virus and is associated with lethal outcome in mice. *J. Virol.* 88, 10556–10568. doi: 10.1128/jvi.00570-14

Papi, A., Contoli, M., Adcock, I. M., Bellettato, C., Padovani, A., Casolari, P., et al. (2013). Rhinovirus infection causes steroid resistance in airway epithelium through nuclear factor kappaB and c-Jun N-terminal kinase activation. *J. Allergy Clin.* 132, 1075.e6–1085.e6.

Papi, A., and Johnston, S. L. (1999). Rhinovirus infection induces expression of its own receptor intercellular adhesion molecule 1 (ICAM-1) via increased NF-kappaB-mediated transcription. *J. Biol. Chem.* 274, 9707–9720. doi: 10.1074/jbc.274.14.9707

Prill, M. M., Dahl, R. M., Midgley, C. M., Chern, S. W., Lu, X., Feikin, D. R., et al. (2018). Severe respiratory illness associated with rhinovirus during the enterovirus D68 outbreak in the United States, August 2014-November 2014. *Clin. Infect. Dis.* 66, 1528–1534. doi: 10.1093/cid/cix1034

Rajan, D., Gaston, K. A., McCracken, C. E., Erdman, D. D., and Anderson, L. J. (2013). Response to rhinovirus infection by human airway epithelial cells and peripheral blood mononuclear cells in an in vitro two-chamber tissue culture system. *PLoS One* 8:e66600. doi: 10.1371/journal.pone.0066600

Rajan, D., McCracken, C. E., Kopleman, H. B., Kyu, S. Y., Lee, F. E., Lu, X., et al. (2014). Human rhinovirus induced cytokine/chemokine responses in human airway epithelial and immune cells. *PLoS One* 9:e114322. doi: 10.1371/journal.pone.0114322

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Roe, J., Venturini, C., Gupta, R. K., Gurry, C., Chain, B. M., Sun, Y., et al. (2019). Blood transcriptomic stratification of short-term risk in contacts of tuberculosis. *Clin. Infect. Dis.* 70, 731–737. doi: 10.1093/cid/ciz252

Seelbinder, B., Wolf, T., Priebe, S., McNamara, S., Gerber, S., Guthke, R., et al. (2019). GEO2RNAseq: an easy-to-use R pipeline for complete pre-processing of RNA-seq data. *bioRxiv* [preprint]. doi: 10.1101/771063

Seng, L. G., Daly, J., Chang, K. C., and Kuchipudi, S. V. (2014). High basal expression of interferon-stimulated genes in human bronchial epithelial (BEAS-2B) cells contributes to influenza A virus resistance. *PLoS One* 9:e109023. doi: 10.1371/journal.pone.0109023

Shinya, K., Gao, Y., Cilloniz, C., Suzuki, Y., Fujie, M., Deng, G., et al. (2012). Integrated clinical, pathologic, virologic, and transcriptomic analysis of H5N1 influenza virus-induced viral pneumonia in the rhesus macaque. *J. Virol.* 86, 6055–6066. doi: 10.1128/jvi.00365-12

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Syedbasha, M., and Egli, A. (2017). Interferon lambda: modulating immunity in infectious diseases. *Front. Immunol.* 8:119. doi: 10.3389/fimmu.2017.00119

To, K. K., Hung, I. F., Li, I. W., Lee, K. L., Koo, C. K., Yan, W. W., et al. (2010). Delayed clearance of viral load and marked cytokine activation in severe cases

of pandemic H1N1 2009 influenza virus infection. *Clin. Infect. Dis.* 50, 850–859. doi: 10.1086/650581

To, K. K., Lau, C. C., Woo, P. C., Lau, S. K., Chan, J. F., Chan, K. H., et al. (2009). Human H7N9 virus induces a more pronounced pro-inflammatory cytokine but an attenuated interferon response in human bronchial epithelial cells when compared with an epidemiologically-linked chicken H7N9 virus. *Virol. J.* 13:42.

To, K. K., Lau, S. K., Chan, K. H., Mok, K. Y., Luk, H. K., Yip, C. C., et al. (2016a). Pulmonary and extrapulmonary complications of human rhinovirus infection in critically ill patients. *J. Clin. Virol.* 77, 85–91. doi: 10.1016/j.jcv.2016.02.014

To, K. K., Lau, C. C., Woo, P. C., Lau, S. K., Chan, J. F., Chan, K. H., et al. (2016b). Human H7N9 virus induces a more pronounced pro-inflammatory cytokine but an attenuated interferon response in human bronchial epithelial cells when compared with an epidemiologically-linked chicken H7N9 virus. *Virol. J.* 13:42.

To, K. K., Lu, L., Fong, C. H., Wu, A. K., Mok, K. Y., Yip, C. C., et al. (2018). Rhinovirus respiratory tract infection in hospitalized adult patients is associated with TH2 response irrespective of asthma. *J. Infect.* 76, 465–474. doi: 10.1016/j.jinf.2018.02.005

To, K. K., Lu, L., Yip, C. C., Poon, R. W., Fung, A. M., Cheng, A., et al. (2017). Additional molecular testing of saliva specimens improves the detection of respiratory viruses. *Emerg. Microbes Infect.* 6:e49.

To, K. K., Mok, K. Y., Chan, A. S., Cheung, N. N., Wang, P., Lui, Y. M., et al. (2016). Mycophenolic acid, an immunomodulator, has potent and broad-spectrum in vitro antiviral activity against pandemic, seasonal and avian influenza viruses affecting humans. *J. Gen. Virol.* 97, 1807–1817. doi: 10.1099/jgv.0.000512

To, K. K. W., Chan, K. H., Ho, J., Pang, P. K. P., Ho, D. T. Y., Chang, A. C. H., et al. (2019). Respiratory virus infection among hospitalized adult patients with or without clinically apparent respiratory infection: a prospective cohort study. *Clin. Microbiol. Infect.* 25, 1539–1545. doi: 10.1016/j.cmi.2019.04.012

Wei, W., Guo, H., Chang, J., Yu, Y., Liu, G., Zhang, N., et al. (2016). ICAM-5/telencephalin is a functional entry receptor for enterovirus D68. *Cell Host Microbe* 20, 631–641. doi: 10.1016/j.chom.2016.09.013

Zaas, A. K., Chen, M., Varkey, J., Veldman, T., Hero, A. O., Lucas, J., et al. (2009). Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* 6, 207–217. doi: 10.1016/j.chom.2009.07.006

Zhai, Y., Franco, L. M., Atmar, R. L., Quarles, J. M., Arden, N., Bucasas, K. L., et al. (2015). Host transcriptional response to influenza and other acute respiratory viral infections–a prospective cohort study. *PLoS Pathog.* 11:e1004869. doi: 10.1371/journal.ppat.1004869

Zhao, H., To, K. K. W., Chu, H., Ding, Q., Zhao, X., Li, C., et al. (2018). Dual-functional peptide with defective interfering genes effectively protects mice against avian and seasonal influenza. *Nat. Commun.* 9:2358.

Zheng, B., Chan, K. H., Zhang, A. J., Zhou, J., Chan, C. C., Poon, V. K., et al. (2010). D225G mutation in hemagglutinin of pandemic influenza H1N1 virus enhances virulence in mice. *Exp. Biol. Med.* 235, 981–988. doi: 10.1258/ebm.2010.010071

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

No 2. An Integrative Understanding of the Large Metabolic Shifts Induced by Antibiotics in Critical Illness

FORM 1

Manuscript No. 2

Manuscript title: An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness

Authors: Andrea Marfil-Sánchez, Lu Zhang, Pol Alonso-Pernas, **Mohammad H. Mirhakkak**, Melinda Mueller, Bastian Seelbinder, Yueqiong Ni, Rakesh Santhanam, Anne Busch, Christine Beemelmanns, Maria Ermolaeva, Michael Bauer, Gianni Panagiotou

Bibliographic information: Marfil-Sánchez, Andrea, et al. "An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness." *Gut Microbes* 13.1 (2021): 1993598.

The candidate is

☐ First author, ☐ Co-first author, ☐ Corresponding author, ☒ Co-author.

Status: Published

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Marfil-Sánchez	30%	40%		40%	
Zhang	30%	40%		40%	
Alonso-Pernas			20%		
Mirhakkak		10%		5%	
Mueller			20%		
Seelbinder		10%		5%	
Busch			20%		
Bauer	20%		20%		60%
Panagiotou	20%			10%	30%
Others			20%		10%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the faculty)

FORM 2

Manuscript No. 2

Short reference Marfil-Sánchez *et al.*, *Gut Microbes* (2021)

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures: My contribution in the publication is *in silico* constraint-based metabolic modeling.

Figure #3	☒	Approximate contribution: 5**%** (contribution to the part A)
Figure #5	☒	Approximate contribution: 25**%** (Figure part B for metabolic modeling)
Figures # 1, 2, 4, 6, and all supplementary figures	☐	**0%**

(Add more table boxes depending on the number of figures)

_____ Signature candidate

_____ Signature supervisor (member of the Faculty)

GUT MICROBES
2021, VOL. 13, NO. 1, e1993598 (25 pages)
https://doi.org/10.1080/19490976.2021.1993598

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness

Andrea Marfil-Sánchez[a,#], Lu Zhang[a,#], Pol Alonso-Pernas[b], Mohammad Mirhakkak[a], Melinda Mueller[b], Bastian Seelbinder[a], Yueqiong Ni[a], Rakesh Santhanam[a], Anne Busch[c], Christine Beemelmanns[d], Maria Ermolaeva[b], Michael Bauer[c,e], and Gianni Panagiotou[a,f,g]

[a]Systems Biology and Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany; [b]Leibniz Institute on Aging – Fritz Lipmann Institute, Jena, Germany; [c]Department of Anesthesiology and Intensive Care Medicine, Jena University Hospital, Jena, Germany; [d]Chemical Biology of Microbe-Host Interactions, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany; [e]Center for Sepsis Control and Care, Jena University Hospital, Jena, Germany; [f]Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, Hong Kong, China; [g]Lead Contact

ABSTRACT

Antibiotics are commonly used in the Intensive Care Unit (ICU); however, several studies showed that the impact of antibiotics to prevent infection, multi-organ failure, and death in the ICU is less clear than their benefit on course of infection in the absence of organ dysfunction. We characterized here the compositional and metabolic changes of the gut microbiome induced by critical illness and antibiotics in a cohort of 75 individuals in conjunction with 2,180 gut microbiome samples representing 16 different diseases. We revealed an "infection-vulnerable" gut microbiome environment present only in critically ill treated with antibiotics (ICU^+). Feeding of *Caenorhabditis elegans* with *Bifidobacterium animalis* and *Lactobacillus crispatus*, species that expanded in ICU^+ patients, revealed a significant negative impact of these microbes on host viability and developmental homeostasis. These results suggest that antibiotic administration can dramatically impact essential functional activities in the gut related to immune responses more than critical illness itself, which might explain in part untoward effects of antibiotics in the critically ill.

ARTICLE HISTORY
Received 25 February 2021
Revised 6 August 2021
Accepted 9 August 2021

KEYWORDS
Gut microbiota; antibiotics; critical illness; intensive care unit; metagenomics; metabolomics; ITS2

Introduction

Critical illness leads to the admission of more than 5 million patients per year to intensive care units (ICUs) in the United States alone. Intensive or invasive monitoring of ICU patients accounts for approximately 20% of the total US hospital cost, while the worldwide death rates for critically ill patients are increasing at a higher rate than any other common cause of death [1]. Almost half of ICU patients show symptoms related to an initial systemic inflammatory response syndrome (SIRS).[2] However, besides inflammation, signs of immune exhaustion or 'paralysis' might occur simultaneously.[3] A disbalance of pro- and anti-inflammatory responses can lead to an increased risk of infection [4] and related sepsis, which are responsible for nearly 60% of deaths in ICUs and account for approximately 40% of ICU costs.[5]

Vincent *et al.* [6] reported that while only 54% of ICU patients had suspected or proven infection, as many as 70% received at least one antibiotic, reflecting a rather "liberal use" within contemporary ICUs. This reflects an early antibiotic treatment – in order to avoid the deleterious impact of a missed or delayed antibiotic therapy if infection triggers organ dysfunction.[7] However, several studies showed that the benefits of antibiotics for prevention of infection, multi-organ failure, and death in the ICU are unclear.[8] Evidence suggests that approximately 37% of antibiotic treatments are unnecessary or not compliant with guidelines.[9] Infections and antibiotics can cause

CONTACT Gianni Panagiotou gianni.panagiotou@leibniz-hki.de Systems Biology and Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Beutenbergstraße 11A, Jena 07745, Germany; Michael Bauer michael.bauer@med.uni-jena.de Department of Anesthesiology and Intensive Care Medicine, Jena University Hospital, Jena, Germany; Maria Ermolaeva maria.ermolaeva@leibniz-fli.de Stress Tolerance and Homeostasis, Leibniz Institute on Aging - Fritz Lipmann Institute, Beutenbergstraße 11, Jena 07745, Germany
[#]These authors contributed equally to this work.
Supplemental data for this article can be accessed on the publisher's website.

a rapid loss of commensal gastrointestinal microbiota,[10] which can result in metabolic and immune disturbances in the critically ill. An important role of gut bacteria is the fermentation of dietary fiber into short-chain fatty acids (SCFAs), which play an important role in preserving gut integrity.[11] A decrease in SCFAs concentrations may result in colonization by pathogenic species. For example, several *Bacteroides* and *Bifidobacteria* species secrete SCFAs inhibiting pathogen growth.[12,13] Similarly, *Clostridium scindens* and *Ruminococcus obeum* produce secondary bile acids (BAs) that prevent the growth of *Clostridium difficile* [14] and *Vibrio cholerae*,[15] respectively. Gut bacteria also play an important role in the activation of host immunity against infections, both through innate and adaptive mechanisms.[16]

With advent of the concept of the human being as a "holobiont" and the perception of the gut microbiome as being highly relevant in the regulation of the immune system, attention to interventions affecting the microbiome is now given also in critical illness. Previous studies have focused predominantly on taxonomic information using 16S rRNA gene sequencing to identify differences between health and disease.[17,18]

Fungal constituents of the microbiome represent an overlooked but very important kingdom. Research is beginning to show that fungi are critical for maintaining systemic immunity and intestinal homeostasis.[19] The mycobiome of skin,[20] gut,[21] oral cavity,[22] and lungs,[23] among other anatomical sites, in healthy individuals has been characterized in previous studies. Based on these studies, it seems that between individuals and anatomic sites there is high variability in the human mycobiome diversity,[24] which is consistent with what we know from the Human Microbiome Project (HMP, 2012) for the bacterial microbiome. Most of the anatomic sites in humans are dominated by members of the *Basidiomycota* phylum, such as *Cryptococcus* spp., *Malassezia* spp. and *Filobasidium* spp., and the *Ascomycota* phylum including *Saccharomyces cerevisiae, Candida* spp., and *Cladosporium* spp.[25,26] Despite the recent findings showcasing the fungal role on host health, host–microbe and microbe–microbe interactions, only less than 0.5% of the published microbiome papers investigate or refer to the fungal population.[19] The mycobiota is increasingly recognized as a critical player in the development of human diseases, including inflammatory bowel disease, allergic airway diseases, skin disease, alcoholic liver disease, autoimmunity, neurological disorders, and metabolic syndrome.[27–30] In relation to microbe–microbe interactions, existing studies indicate that a competitive association between bacteria and fungi exists in the human gut. This was shown in antibiotic-treated subjects and germ-free mice, where an overgrowth of particular fungi in the gut and/or susceptibility to fungal infection was observed.[31] In addition, overgrowth of fungi due to antibiotics treatment has been associated with the development of allergic airway responses to *Aspergillus fumigatus* mold spores.[32] Part of the chemical warfare between bacteria and fungi is also the secretion of antifungal peptides from epithelial cells, which can be induced by commensal bacteria such as *Blautia producta* and *Bacteroides thetaiotaomicron*.[33] Previous research related to critically ill patients and mycobiome is very limited.[34–39] One recent study characterized the mycobiome of the lower respiratory tract of patients in ICU showing that *Candida* spp. dominated the fungal community in both with and without antibiotic therapy patient groups.[34] Another prospective pilot study showed an increase of *C. albicans* in the oral mycobiome after an admission to the ICU.[38] However, systematic investigation of the interactions between the gut mycobiome of critically ill patients and the bacteria functional activity is currently lacking.

Here, by a comprehensive characterization of the microbiome, mycobiome, and functional potential of the gut community and individual species, we demonstrate that even though antibiotics do not significantly disturb the bacterial and fungal composition of critically ill patients, as observed in healthy individuals,[40] they cause abundance changes in a handful of species that are highly connected with the production of SCFAs and BAs, allowing the expansion of pathogenic species.

Results

Highly distinct microbiome in ICU patients

To investigate the gut microbiome composition of critically ill patients, we initially included 70 critically ill patients. Of these, 54 were diagnosed with

probable or microbiologically confirmed infection as defined by Calandra *et al.* [41] (respiratory tract (n = 37), abdominal (n = 6), bones/soft tissue (n = 3), chest (n = 2), catheter associated (n = 1), urogenital (n = 1), and unknown (n = 4)). These 54 received antibiotic treatment, whereas 16 did not presented an infection and did not receive antibiotic treatment. Of those, we collected stool samples from 49 patients receiving antibiotic treatment (broad spectrum beta-lactam antibiotics; n = 19 meropenem and n = 30 piperacillin/tazobactam) and 14 patients without antibiotic treatment, as well as from 12 healthy human volunteers. Basic anthropometric and clinical characteristics of the participants are displayed in Supplementary File 1. There were no significant differences between the two ICU groups in gender, age, BMI, type of admission or surgery and length of ICU or hospital stay (continuous data were compared by the Student's t-test, dichotomous variables by the chi-squared test, a p-value < 0.05 was considered significant). Similarly, there were no significant differences between the healthy volunteers and ICU patients in the basic demographic characteristics.

We assessed the structure of the gut microbiome via shotgun metagenomics generating 39 million high-quality reads per sample on average (Supplementary Table 1). We used MetaPhlan2 [42] for taxonomic profiling, and we identified 428 species in total. Alpha-diversity measured as Shannon, Simpson, and Chao1 index dropped significantly in both antibiotics treated (ICU^+) (Wilcoxon rank-sum test, P = 8.6e-6, P = 5.3e-6, P = 3.8e-5, Shannon, Simpson and Chao1 indices, respectively) and untreated subjects (ICU^-) (Wilcoxon rank-sum test, P = .0037, P = .0025, P = .0054, Shannon, Simpson and Chao1 indices, respectively) compared to healthy individuals (Figure 1(a)). In contrast to what has been observed in healthy volunteers,[44,45] antibiotics administration had no significant impact on the alpha diversity when comparing the ICU^+ against the ICU^- patients (Wilcoxon rank-sum test, P = .79, P = .79, P = .64, Shannon, Simpson and Chao1 indices, respectively). The type of beta-lactam used had no significant impact either (Supplementary Figure 1).

We subsequently investigated the variation in the microbiome structure of the three groups by calculating the distance to centroid (Figure 1(b)). The gut microbiome of the ICU^+ group had significantly higher structural variation compared to both healthy individuals (Wilcoxon rank-sum test, P = 1.1e-12) and the ICU^- group even though it did not reach statistical significance (Wilcoxon rank-sum test, P = .056). Next, we calculated the beta-diversity (Bray–Curtis distance) of the three groups, which showed that although there was a high community dissimilarity between ICU^+ and Healthy (PERMANOVA, P = .001) and between ICU^- and Healthy (PERMANOVA, P = .001), the ICU^+ and ICU^- groups did not show significant differences (PERMANOVA, P = .104) (Figure 1(c), Supplementary Figure 9(a)). Similarly, to alpha diversity, when we examined the two beta-lactam groups (meropenem and piperacillin/tazobactam) separately, we did not observe any significant difference (Supplementary Figure 1, Supplementary Figure 9(b)). To further evaluate our findings, we repeated the taxonomic analysis using the Metagenomic Species method (MGS) [46] and we observed the same patterns as above (Supplementary Figure 2, Supplementary Figure 9(c)).

As evident from both, the alpha- and beta-diversity comparisons, the similarities in the structure of the gut microbiome of the ICU^+ compared to the ICU^- group triggered our interest to examine our cohort in relation to another, small publicly available dataset of critically ill patients.[10] In that study, the authors performed a longitudinal sampling of critically ill patients with and without antibiotics administration to study abundance changes in the resistome profile of known pathogens. For the comparative analysis, we applied a microbial source tracking algorithm, namely FEAST.[43] When using our cohort as a sink we noticed how progressively the public ICU^+ group becomes more taxonomically similar with our ICU^+ group, as well as the compositional similarities between the two ICU^- groups (Figure 1(d)). From the methodological point of view, this analysis confirms that FEAST [43] is a relatively sensitive method but also that despite the non-significant overall differences between ICU^+ and ICU^- groups there is still a characteristic microbiome signature due to the antibiotic administration. Considering the relatively small differences between the ICU^+ and

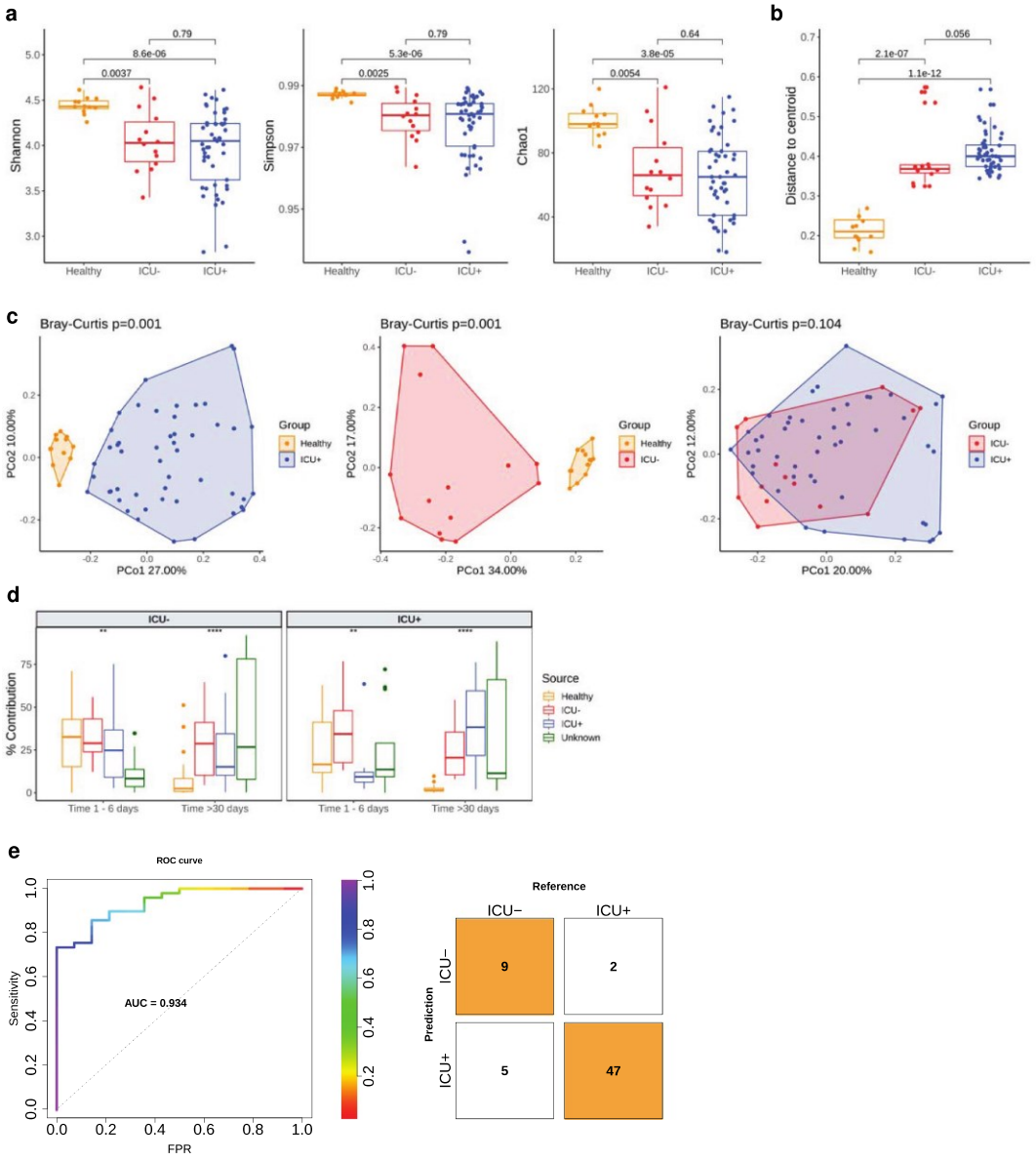


Figure 1. Distinct gut microbiota signatures in ICU patients. (a – b, d) Box plots showing the median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers) measurements. A comparison was considered significant if $P < .05$.(a) Alpha diversity of bacterial species using Shannon (left), Simpson (middle), and Chao1 (right) indices. Significant differences were determined using Wilcoxon rank-sum test.(b) Beta dispersion of bacterial species measured as the distance of the samples from one group to the group centroid in multivariate space. Significant differences were determined using Wilcoxon rank-sum test.(c) Principal component analysis (PCoA) of Bray-Curtis dissimilarity between bacterial species abundance profiles. Significant differences were determined using PERMANOVA and were considered significant if $P < .05$.(d) FEAST [43] estimation of microbial source contribution for each "sink". Here, sinks are species level relative abundances from samples from a publicly available ICU cohort.[10] For sources, we used species level relative abundances of our Healthy, ICU$^-$ and ICU$^+$ groups. Significant differences were determined using Kruskal-Wallis test (*P < .05; **P < .01, ***P < .001, ****P < .0001).(e) Receiver operating characteristic curve (ROC) of a cross-validated random forest classifier. The model was trained on taxonomic and functional profiles from 63 samples from the ICU$^-$ and ICU$^+$ groups. Model performance was summarized as area under ROC (AUC). The average AUC value and confusion matrix (threshold 0.5) are calculated based on a 5-fold cross validation results.

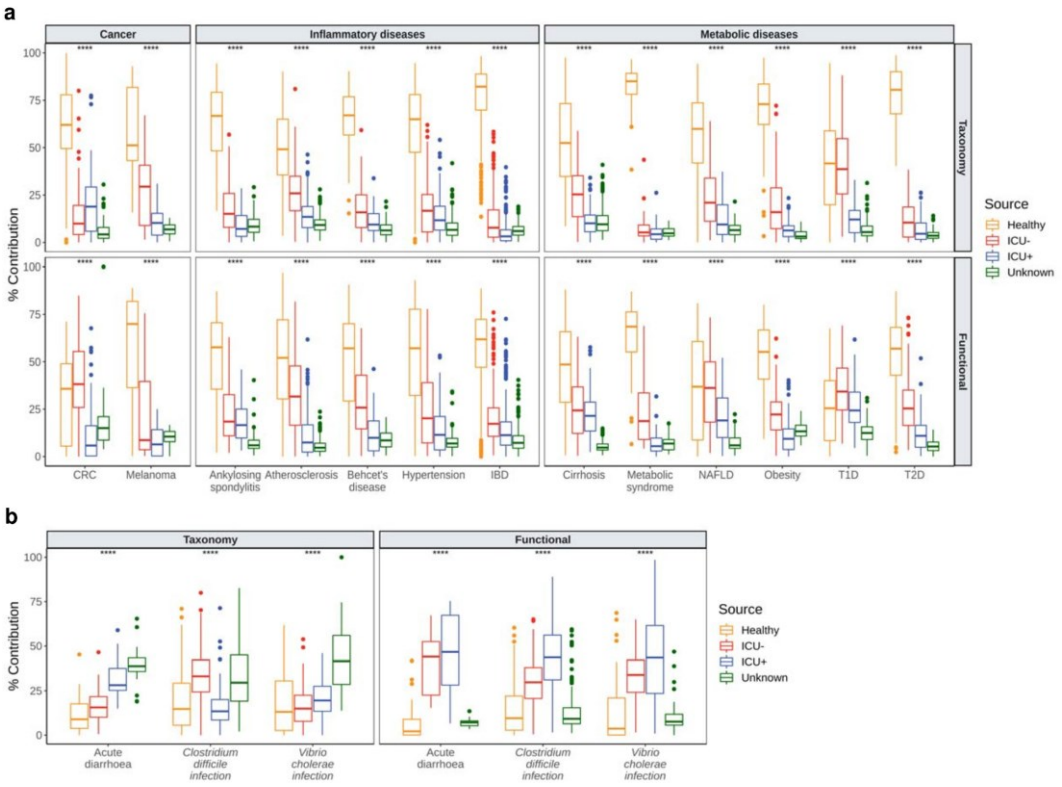


Figure 2. Comparative analysis of the microbiome of critically ill patients with other diseases. (a – b) FEAST [43] estimation of microbial source contribution for each "sink". Here, sinks are taxonomic and functional compositions from a panel of diseases. Source are taxonomic and functional compositions of samples from our Healthy, ICU^- and ICU^+ groups. Box plots show the median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), source contributions to sinks. (a) cancer, inflammatory and metabolic diseases and (b) infectious diseases. (a – b) Significant differences were determined using Kruskal-Wallis test (*P < .05; **P < .01, ***P < .001, ****P < .0001).

ICU^- groups, we next investigated whether these subtle differences in microbial signatures could be integrated into an algorithm to correctly classify the patients. To this end, a random forest classifier integrating species and pathways was developed and achieved an area under the receiver operating characteristic (ROC) curve (AUC) of 0.934 (Figure 1(e)) using 20 features (Supplementary Figure 3). *Enterococcus faecalis*, a known pathogen[48] whose abundance has been associated to an increase in susceptibility to *V. cholerae* infection,[15] and several amino acid pathways that can serve as precursors for the synthesis of short-chain fatty acids (SCFAs) were among the selected features.

Critically ill patients treated with antibiotics show an "infection-vulnerable" gut microbiome

We subsequently used HUMAnN2 [49] for functional profiling of the microbial communities and identified 483 pathways in total. To further evaluate the community taxonomic and functional characteristics of the critically ill patients, we compared our cohort with a panel of publicly available datasets. For that, we retrieved 2,180 taxonomically and functionally annotated samples from 16 human diseases using the curatedMetagenomicData [50] package in R and we applied FEAST,[43] treating our samples as sources and the disease panel as sinks. In these settings, FEAST [43] revealed that there is a significantly lower similarity

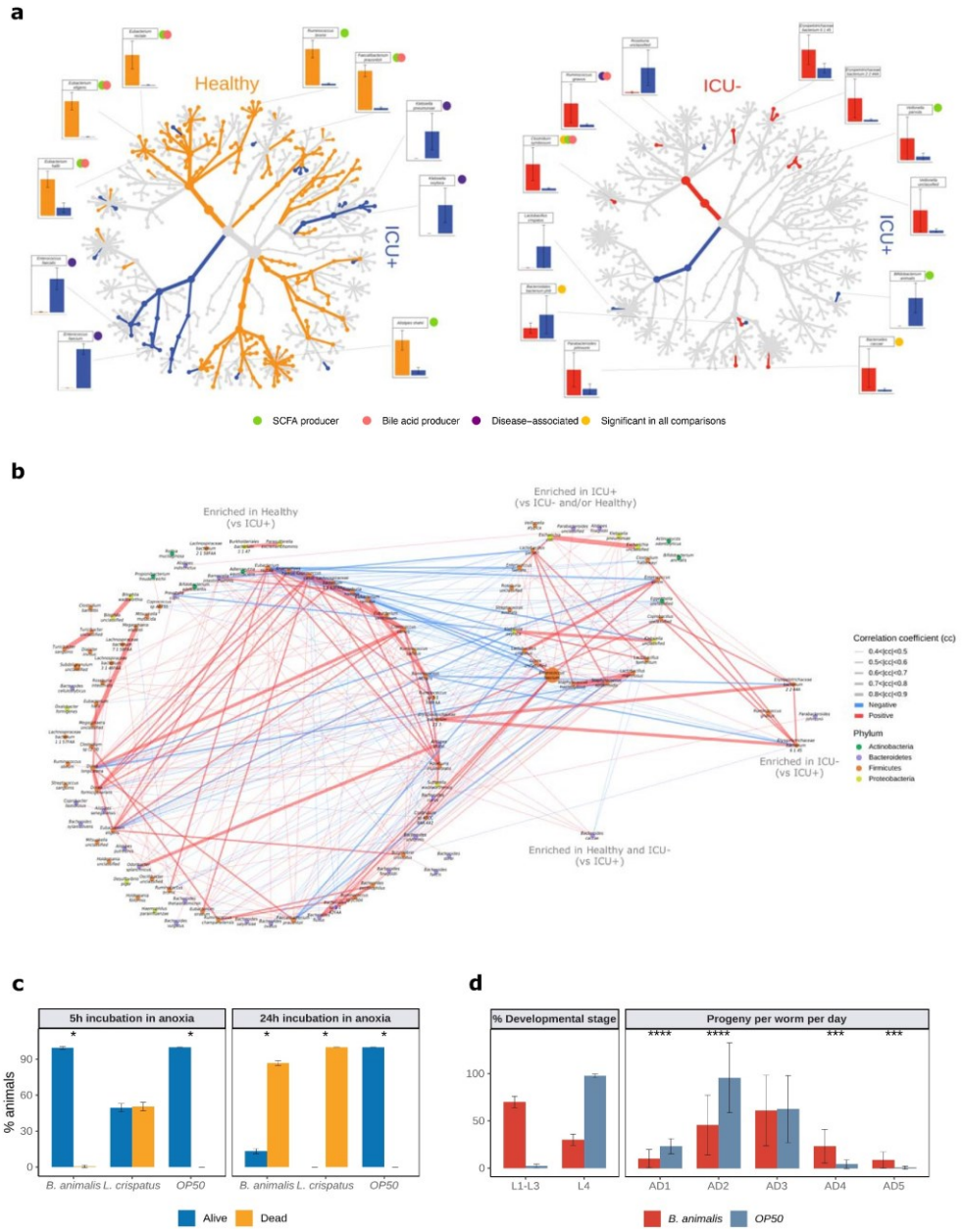


Figure 3. Differential abundance analysis reveals changes in bacteria with important functional properties in critically ill. (a) Taxonomic trees visualized using R package metacoder.[47] Only taxa differentially abundant between Healthy, ICU^+ or ICU^- ($P < .05$, Wilcoxon rank-sum test) are highlighted in the tree by color. Color of the taxa reflects the group with higher abundance. Bar plots show the relative abundances of significantly different (FDR<0.05, Wilcoxon rank-sum test) short-chain fatty acid producers (green circles), bile acid producers (red circles), disease-associated species (purple circles) or differentially abundant species identified in all pairwise comparisons (yellow circles).(b) Co-abundance networks of differentially abundant bacterial species ($P < .05$, Wilcoxon rank-sum test) between Healthy and ICU^+ and between ICU^+ and ICU^-. Only significant correlations ($P < .05$) with absolute correlation coefficient >0.4 were used for network construction. Nodes are colored based on their affiliated phyla. Node sizes reflect the mean abundance of the species. Edge colors reflect either negative correlation (blue) or positive correlation (red). Edge widths reflect the strength of the correlation. (c-d) Germ-free L1 larval stage *C. elegans* worms were populated with depicted bacterial strains in the anoxic chamber for indicated

times, followed by transfer to normoxia and UV-killed OP50 *E. coli* diet. In (C) host survival was measured after 24 h of normoxic culture and in (D – left panel) the timely development was assessed after 48 h of normoxic culture; (D – right panel) depicts reproductive aging of nematodes following anoxic reconstitution with *B. animalis*.

(Kruskal–Wallis, $P < .0001$) between the microbiome structure and function of individuals in the critically ill patients group with patients with cancer (colorectal cancer,[51–56] melanoma[57,58]), inflammatory (ankylosing spondylitis,[59] atherosclerosis,[60] Behcet's disease,[61] hypertension,[62] inflammatory bowel disease[46]), and metabolic diseases (cirrhosis,[63] metabolic syndrome,[64] nonalcoholic fatty-liver disease,[65] obesity,[66,67] Type-1-Diabetes,-[68,69] Type-2-Diabetes[70,71]) compared to Healthy individuals (Figure 2(a)). Due to this unexpected association, we expanded the range of diseases as sinks using infectious diseases and specifically 66 samples from patients with acute diarrhea,[72] *C. difficile* infection[73] and *V. cholerae* infection.[74] In contrast to what we observed with noninfectious diseases, the taxonomic composition of the critically ill patients is significantly more similar (Kruskal–Wallis, $P < .0001$) to infectious diseases than the gut composition of healthy individuals (Figure 2(b)). At the functional level, this difference is even more profound with the ICU$^+$ group having the highest similarity (Kruskal–Wallis, $P < .0001$) with all three infectious diseases, suggesting that critically ill patients may be at a risk of severe infections by gut pathogens and that antibiotic treatment increases significantly that risk.

Immunomodulatory metabolites and their microbial producers are depleted from critically ill patients exposed to antibiotics

We subsequently focused on the impact of antibiotic administration in the ICU in relation to abundance changes in microbes with a potential role in immune regulation and host-immune homeostasis. Significant differences were observed at the phylum level among the three groups. The Healthy group had, on average, higher abundance of *Bacteroidetes* (17.9%, Kruskal–Wallis, $P = .0063$), *Actinobacteria* (9.9%, Kruskal–Wallis, $P = .035$), and *Verrucomicrobia* (4.7%, Kruskal–Wallis, $P = .0098$) (Supplementary Figure 4(a)). At the genus level, a striking difference in the abundance of *Enterococcus* was found, with the ICU$^+$ group having the highest abundance compared to the other two groups (Kruskal–Wallis, $P = 3.6e\text{-}6$). In contrast, the relative abundance of *Blautia* was significantly higher in ICU$^-$ (Kruskal–Wallis, $P = .029$) (Supplementary Figure 4(a)).

At the species level, we found 106 and 80 species significantly differentially abundant (Wilcoxon rank-sum test, $P < .05$) between the Healthy and ICU$^+$ and Healthy and ICU$^-$ groups, respectively. The species enriched in the Healthy group include known SCFAs and BAs producers, such as *Ruminococcus bromii*,[76] *Faecalibacterium prausnitzii*,[77] *Eubacterium eligens*,[78] *Eubacterium hallii* [78] and *Eubacterium rectale*,[77,78] among others (Figure 3(a) and Supplementary Figure 4(b)). On the contrary, species enriched in ICU$^+$ and ICU$^-$ included known pathogens such as *Klebsiella pneumoniae*,[79] *Klebsiella oxytoca*,[79] *E. faecalis*,[48] and *Enterococcus faecium* [48] (Figure 3(a) and Supplementary Figure 4(b)). Comparing ICU$^+$ and ICU$^-$ groups, we found 12 species significantly differentially abundant (8 enriched in ICU$^-$ and 4 in the ICU$^+$). *R. gnavus, Clostridium symbiosum*, and *Veillonella parvula*, known SCFAs and BAs producers,[80–83] were enriched in ICU$^-$, whereas from the species enriched in ICU$^+$ (Figure 3(a)) *Bifidobacterium animalis* has been indicated as a SCFAs producer.[84] Since *B. animalis* was the only SCFA producer enriched in ICU$^+$, we retrieved the genome-scale metabolic model of *B. animalis* from the AGORA repository [85] and we simulated growth on an ICU media using flux balance analysis. We monitored the potential of *B. animalis* to produce SCFAs, including acetate, propionate, and butyrate; however, none of these compounds was predicted to be produced in our *in silico* simulations (data not shown).

We used the enriched species to reconstruct the species co-abundance network using all samples from our cohort. We observed a much more

intense within group communication between the Healthy-enriched species (vs. ICU$^+$) with 162 positive and 19 negative correlations (absolute correlation coefficient >0.4) compared to the species enriched in ICU$^+$ (vs Healthy and/or ICU$^-$) that showed 43 positive and no negative correlations (Figure 3(b) and Supplementary Table 2). *Alistipes putredinis, F. prausnitzii*, and *Bacteroides uniformis* were interconnected healthy-enriched species showing negative correlations with *Klebsiella* and *Staphylococcus* species, common sources of serious infections in ICU.[86] From the species enriched in ICU$^+$ compared to Healthy, *E. faecium, Staphylococcus epidermidis*, and *Lactobacillus rhamnosus* showed the highest number of correlations, suggesting that these may be important species to maintain the community structure. From the species enriched in ICU$^+$ compared to ICU$^-$, *Lactobacillus crispatus* showed positive correlations with known pathogens such as *Enterococcus* and *Klebsiella* species, as well as negative correlations with Healthy-enriched SCFAs and BAs producers including *F. prausnitzii, E. rectale*, and *A. shahii*.

To examine *in vivo* host effects of the species enriched in ICU$^+$ patients compared to ICU$^-$, we utilized nematode *C. elegans* as a model host. *C. elegans* recently emerged as a suitable simple model for discovering conserved host–microbiome interactions.[87] The experiments were conducted by reconstituting germ free L1 nematodes with the two, sufficiently annotated, bacterial strains enriched in ICU$^+$ (*B. animalis* and *L. crispatus*). We found that both strains had a strong negative impact on the host homeostasis: *L. crispatus* exerted direct toxicity in the model host (Figure 3(c)), and *B. animalis* instigated a delay in the *C. elegans* development indicative of the physiological stress (Figure 3(d)). By including heat-killed bacteria as an additional control and performing OD normalization across conditions, we found that host toxicity of *L. crispatus* requires live bacteria, while *B. animalis* rather acts as a passive stressor (Supplementary Figure 5). Interestingly, the nematodes, which overcame the developmental hindrance by *B. animalis*, displayed a delay in reproductive aging (Figure 3(d)) consistent with the putative probiotic effect of *B. animalis* in the animals that were able to conquer the initial stress caused by this bacterium. To probe the potential probiotic effect of *B. animalis* at the mechanistic level, we tested its ability to induce nuclear translocation of DAF-16/FOXO transcription factor – a conserved mediator of stress resistance and longevity extension.[90–92] We found that live *B. animalis* indeed had the strongest capacity to induce DAF-16 activation among all conditions tested (Supplementary Figure 6). Our results are thus in line with the previously reported ability of *B. animalis* to cause disease in immunocompromised human patients, while it acts as a probiotic in healthy humans,[93] demonstrating the physiological relevance of our nematode findings. Collectively, our *in vivo* tests indicate that the rarefication of the microbiome, which is exacerbated by antibiotics exposure in ICU$^+$ patients, facilitates the enrichment of microbes with potential of exerting direct detrimental effects on the host.

Functional shifts in the microbiome

Next, we compared MetaCyc pathway abundance to explore the gut microbiome functionality in the three groups. As shown in the ordination plot (Figure 4(a)), the differences between ICU$^+$ compared to the other 2 groups at the functional level are driven by SCFA biosynthesis, AA biosynthesis, and fermentation to SCFA, whereas the abundance of pathways related to nucleotide and nucleoside degradation and vitamin biosynthesis are the main drivers in the comparison of Healthy vs ICU$^-$. Therefore, we then focused on AAs, SCFAs, and BAs due to their important role in the regulation of the immune system and their influence in diseases associated with dysbiosis.[11] In total, we identified 117 significantly differentially abundant pathways related to the metabolism of AAs, SCFAs, and BAs (Wilcoxon rank-sum test, $P < .05$). Even though the majority of these pathways had a higher relative abundance in the Healthy compared with ICU$^-$ group, only a few pathways reached statistical significance. On the contrary, the differences in the relative abundances between the ICU$^+$ with the Healthy group were striking (Figure 4(b)). The relative abundance of 98 of the 117 unique pathways was significantly lower in ICU$^+$. Despite the small number of significantly differentially abundant species between ICU$^+$ and ICU$^-$, there were more differences in the metabolic pathways

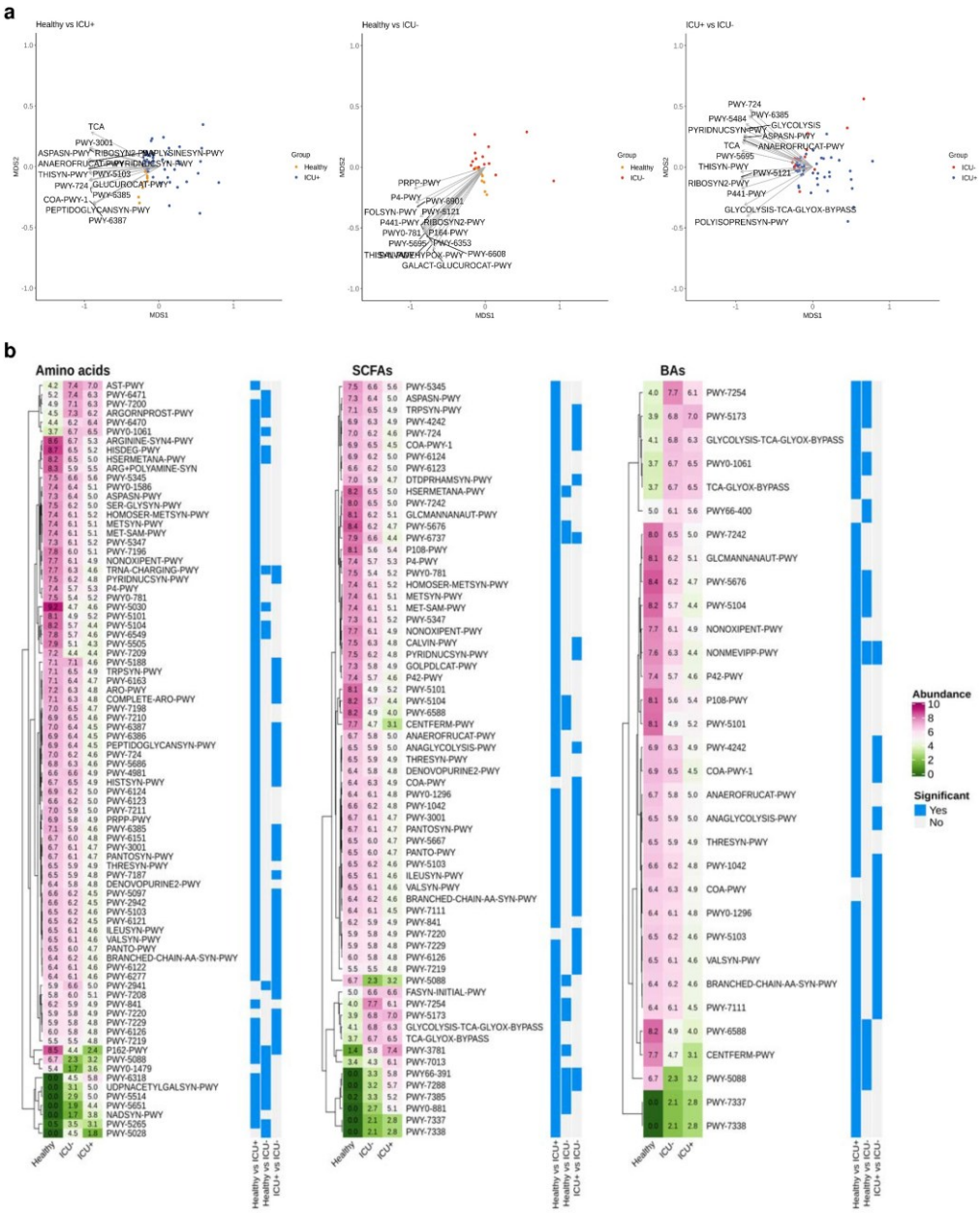


Figure 4. Functional shifts contributed by the microbiome. (a) Multidimensional scaling (MDS) plot of the samples based on the pathway abundances. The top 15 pathways with the strongest significant correlation with the overall ordination (FDR<0.05 with function envfit from the R package vegan [75]) are highlighted with arrows where the length of the arrows reflects the strength of the association.(b) Numbered heatmaps showing the relative abundances of differentially abundant bacterial MetaCyc pathways. Additional heatmaps indicate significance (blue tiles, $P < .05$, Wilcoxon rank-sum test). Only pathways related to short-chain fatty acid, bile acid and amino acid metabolism, as identified by manual curation, are shown.

between the two groups than between Healthy and ICU⁻, suggesting that antibiotic treatment has a stronger effect in the metabolism of AAs, SCFAs, and BAs than critical illness. Several of these pathways were selected in the random forest model to classify the two groups of ICU patients (Figure 1(e) and Supplementary Figure 3).

Metabolomics analysis

We then performed targeted metabolomic analysis and quantified the levels of 10 SCFAs and 27 BAs in 38 stool samples (see Methods). Consistently with the results from the MetaCyc pathway analysis, we observed a significant decrease in the abundance of SCFAs and BAs in the ICU⁺ group compared to the Healthy and ICU⁻ groups (Figure 5(a)). Of 10 identified SCFAs, 6 were significantly different between the groups (Figure 5(a)). Acetic acid, propionic acid, butyric acid, and valeric acid were significantly lower in the ICU⁺ group compared to both the Healthy and ICU⁻ groups (Wilcoxon rank-sum test, $P < .05$). The levels of these SCFAs were not found to differ significantly between the ICU⁻ and Healthy groups. A similar pattern was observed with the BAs; ketolithocholic acid, deoxycholic acid, glycolithocholic acid, hyodeoxycholic acid, isolithocholic acid, lithocholic acid, and ursodeoxycholic acid were all found significantly lower in abundance in ICU⁺ patients (Wilcoxon rank-sum test, $P < .05$) compared to both Healthy and ICU⁻ patients (Figure 5(a)). Among them, ursodeoxycholic acid is increasingly used in the clinical setting for a treatment of a variety of conditions.[94,95] Ketolithocholic acid, lithocholic acid, and ursodeoxycholic acid have been also found to provide resistance against *C. difficile* infections and to modulate the host inflammatory response during the infection.[96] The BA profiling of ICU⁺ patients may also explain the high similarity at the functional level revealed by FEAST with the *C. difficile* cohort (Figure 2(b)).

To identify which species were mainly responsible for the differences in the levels of the measured SCFAs and BAs in ICU⁺ compared to the other two groups, we performed growth rate analysis [89] and Spearman's correlation between the 13 metabolites and 38 species. The species selected were either having (i) significantly higher abundance in the Healthy group compared to ICU⁺ but not compared to the ICU⁻ group or (ii) significantly higher abundance in the ICU⁻ compared to the ICU⁺ group (Figure 5(b)). As shown in Figure 5b, *A. putredinis, Lachnospiraceae bacterium 2 1 58FAA* and *Lachnospiraceae bacterium 1 1 57FAA* showed positive correlations with SCFAs and/or BAs and were predicted through Flux Variability Analysis (FVA) [88] to secrete SCFAs. Several *Bacteroides* species, including *B. uniformis*, were actively growing (GRiD [89] > 1), they showed positive correlations with the measured SCFAs and they were predicted through FVA [88] to secrete SCFAs supporting the identified correlation (Figure 5(b)). Since BA metabolism is not included in these metabolic models, we analyzed the BA biosynthesis potential of the species showing high correlation with BAs using differential analysis of enzymes involved in the BA biosynthesis pathway. The enzymes *cbh* and *baiN* were found in the genomes of *Bacteroides xylanisolvens, L. bacterium 2 1 58FAA, R. obeum*, and *R. gnavus*. The abundance levels of *B. xylanisolvens* and *R. obeum cbh* and the abundance levels of *R. obeum baiN* were found significantly higher in the Healthy group compared to the ICU⁺ (Wilcoxon rank-sum test, $P < .05$, data not shown). *R. obeum* has been previously shown to have an inhibitory effect on *V. cholerae* due to its capacity to degrade virulence-activating signals in the gut through the synthesis of bile salt hydrolases.[15]

Candida species and resistant genes flourish in critically ill patients

Since the microbiome composition and functional profile of critically ill patients appeared to be dysbiotic, we investigated next if this leads to systematic changes in the structure of the fungal community, with emphasis on *Candida* species and the antibiotic resistance gene (ARG) levels in the three groups. We built high-quality libraries for ITS2 sequencing of 74 available stool samples from the Healthy, ICU⁺ and ICU⁻ groups. We estimated the fungal relative abundance using the DADA2 pipeline.[97] The taxonomic profiling revealed that *Ascomycota* (85.63%) was the most abundant fungal phyla, followed by *Basidiomycota* (9.84%) and *Muromycota* (3.44%) (Figure 6(a)). There was no significant difference in the alpha diversity (Shannon, Simpson, Chao1) between any of the two groups and the type of beta-lactam

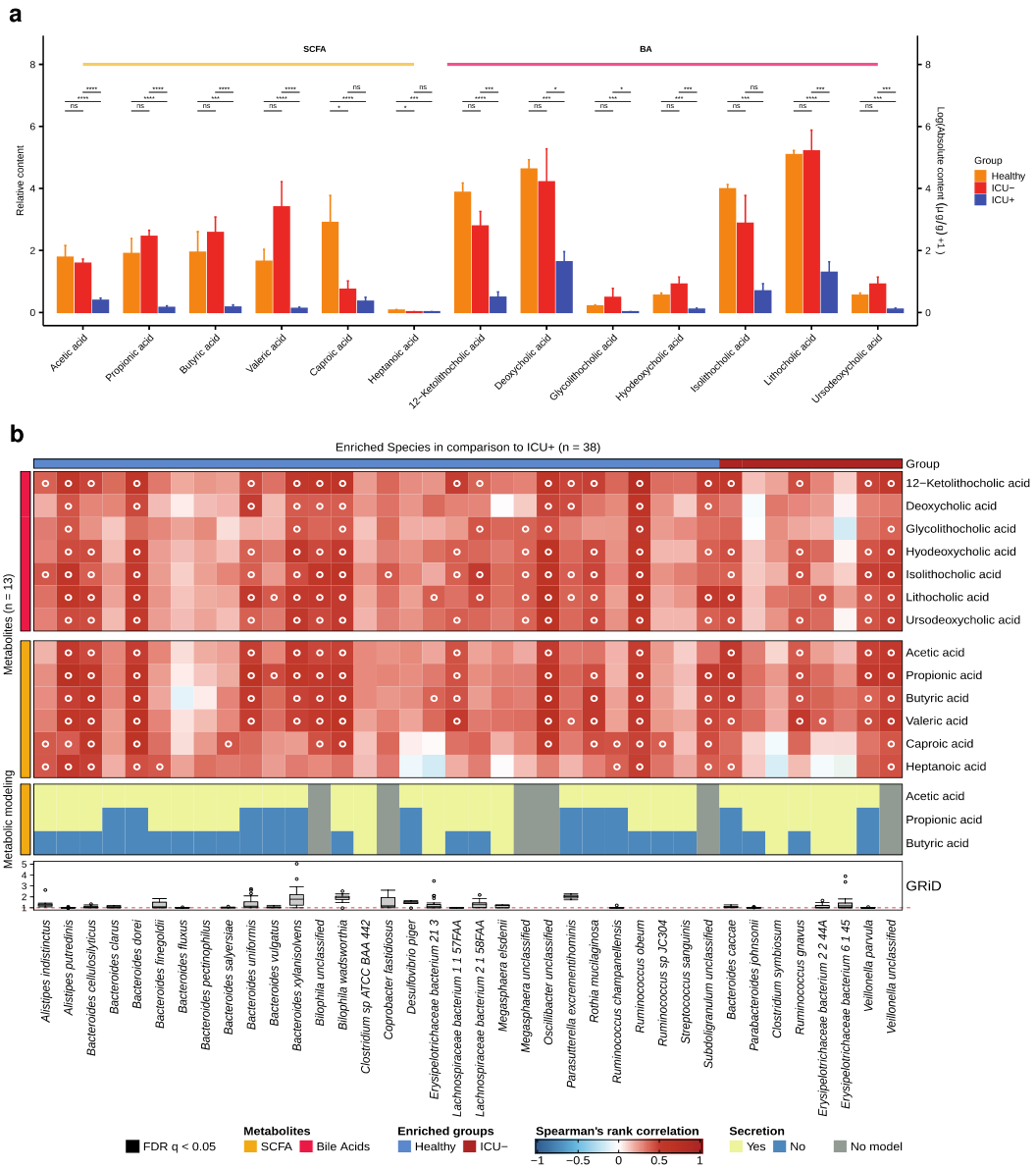


Figure 5. Associations between gut microbiome and short-chain fatty acids and bile acids. (a) Content of SCFAs and BAs ($\log_e$) in three groups are plotted as bar plots. Wilcoxon rank-sum test, *P < .05, **P < .01, ***P < .001, ****P < .0001, ns: not significant.(b) Heatmap showing Spearman's correlations between significantly different metabolites and previously identified significant species. The species selected were either having (i) significantly higher abundance in the Healthy group compared to ICU^+ or (ii) significantly higher abundance in the ICU^- compared to the ICU^+ group. Metabolites category is shown with row annotation and the origin of enriched species is shown with column annotation. Correlations with FDR<0.05 are marked with a circle. Heatmap in the lower panel shows the secretion of short-chain fatty acids obtained by flux variability analysis (FVA) [88] of available bacterial metabolic models. Unavailable models of corresponding species are shown in gray. Bottom boxplots show the results from the Growth Rate Analysis (GRiD).[89] The red dashed line indicates GRiD [89] value 1.

administrated in the ICU$^+$ group did not appear to influence this pattern (Wilcoxon rank-sum test, $P > .05$) (Supplementary Figure 7). However, when we calculated the Bray–Curtis distance of the three groups we observed a high community dissimilarity between ICU$^+$ and Healthy (PERMANOVA, $P = .001$) and between ICU$^-$ and Healthy (PERMANOVA, $P = .001$), whereas the ICU$^+$ and ICU$^-$ groups did not show significant differences (PERMANOVA, P = .703) (Figure 6(a), Supplementary Figure 10(a)). We subsequently investigated differences in fungal species relative abundance between the three groups. In total 19, 11, and 3 fungal species were found significantly different (Wilcoxon rank-sum test, $P < .05$) in the comparisons between Healthy vs ICU$^+$, Healthy vs ICU$^-$, ICU$^+$ vs ICU$^-$, respectively (Figure 6(b)). The *Candida* genus, which includes species that are opportunistic pathogens, was found to have the greatest number of differentially abundant species between the three groups. *Candida albicans, Candida glabrata, Candida pseudolambica*, and *Candida tropicalis* were all found to have the highest abundance in ICU$^+$ patients (Figure 6(b)).

To determine whether antibiotic treatment exerts selective pressure on the resistome as a whole, we analyzed the change in Pfams related to the resistome and mobilome, as well as the abundance differences of ARGs between the three groups. In the abundance comparison between the Healthy and the ICU$^+$ groups, there were 71 statistically significant Pfams (Wilcoxon rank-sum test, $P < .05$) related to the resistome and/or mobilome, with 48 of them being more abundant in the ICU$^+$ and 23 in the Healthy group (Supplementary Figure 8(a)). Interestingly, the differences between ICU$^-$ and Healthy groups were also large with 30 and 16 Pfams being more abundant in the ICU$^-$ and the Healthy group, respectively (Supplementary Figure 8(a)). In the comparison between ICU$^+$ and ICU$^-$ there were 19 significant Pfams with 14 being more abundant in the ICU$^+$ and 5 in the ICU$^-$ group (Supplementary Figure 8(a)).

Subsequently, we annotated the ARGs in the three groups using deepARG.[98] The overall ARG abundance profile of the ICU$^+$ and ICU$^-$ groups using the Bray–Curtis distances indicates that there is no substantial perturbation during antibiotic treatment (PERMANOVA, $P = .261$), but there are significant differences between ICU$^+$ and Healthy (PERMANOVA, $P = .001$) and between ICU$^-$ and Healthy (PERMANOVA, $P = .001$) (Supplementary Figure 8(b), Supplementary Figure 10(b)). Comparing the total accumulative relative ARGs abundance led to the same conclusion; critically ill patients have already a unique resistome profile compared to healthy individuals, which is only marginally disturbed by antibiotics administration (Wilcoxon rank-sum test, $P = 5.97e\text{-}8$, Supplementary Figure 8(c)). Considering the polypharmacology approaches often applied in ICU, our observation is in agreement to recent studies suggesting that human targeted drugs can significantly impact on the gut resistome profile.[99] Nevertheless, an abundance comparison of the individual ARGs revealed 24 ARGs, including 3 beta-lactams (SHV, PENA, AMPC), significantly higher in the ICU$^+$ group and only 1 in the ICU$^-$ group (dabestr, Confidence Interval (CI) = 95%, Supplementary Table 3). Among the ARGs that exhibit the highest abundance differences (dabestr, 95%) between the ICU$^+$ and ICU$^-$ groups were AAC(6′)-I (aminoglycoside), ADEC (multidrug), ERMB (macrolide, lincosamide and streptogramin), VANS (glycopeptide), and MSRC (macrolide, lincosamide and streptogramin) (Supplementary Figure 8(d)).

Discussion

Sepsis, acute respiratory distress syndrome (ARDS), and multi-organ failure represent common conditions most frequently driven by an inappropriate host response to pathogens of the critically ill. These conditions are responsible for immense global mortality accompanied by a tremendous economic burden [1]. While breakthroughs of molecular medicine have revolutionized treatment in oncology and rheumatology, in critical illness research endeavors of decades have not resulted in any targeted therapies. In practice, intensive care can be considered mostly as supportive and antibiotics are a cornerstone of care for patients with sepsis, i.e. infection-driven organ dysfunction.[7]

The off-target effects of antibiotics on the microbiome are, however, also particularly obvious in the ICU population. Antibiotics often fail to resolve

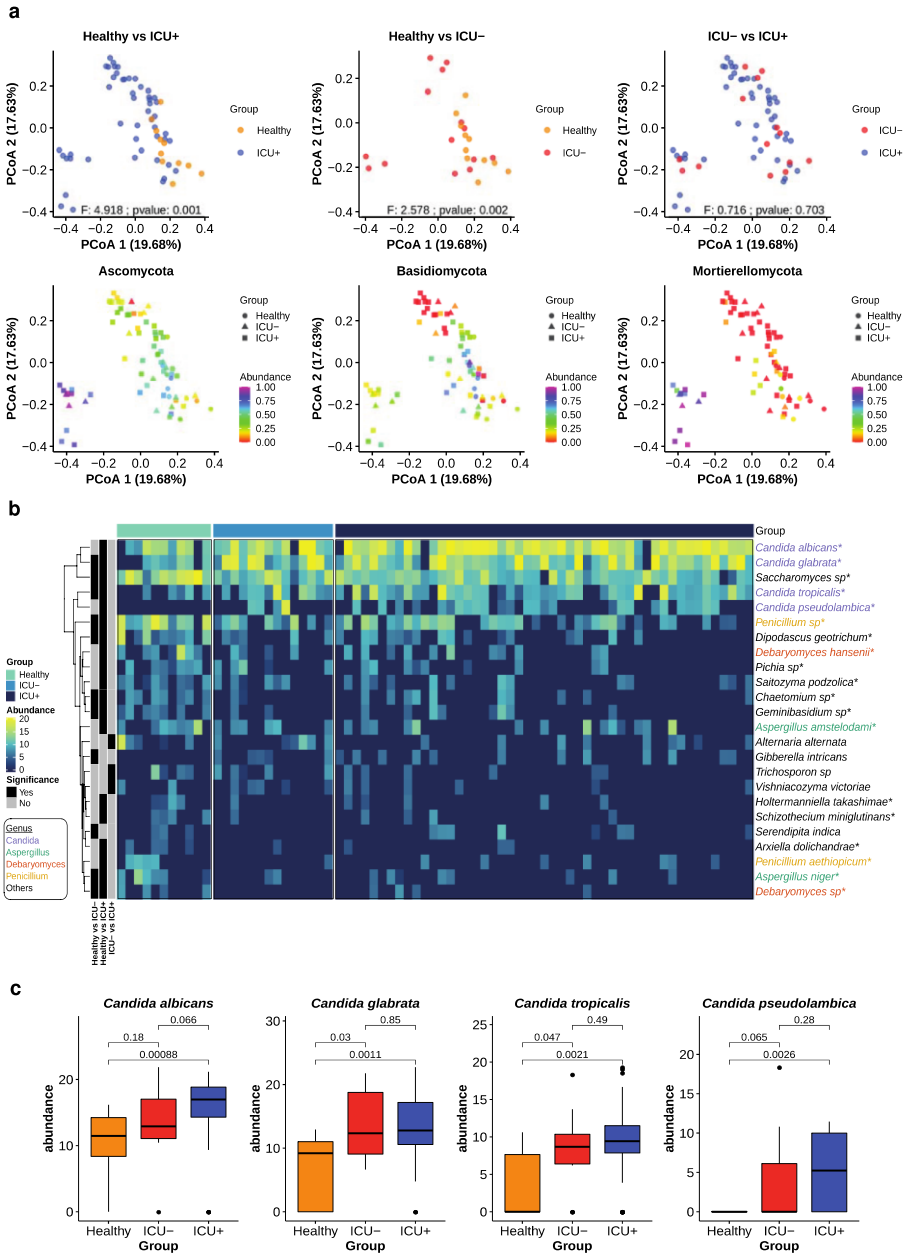


Figure 6. ICU patients have distinct mycobiome profiles. (a) Upper – PCoA plots of the gut mycobiome based on species-level Bray-Curtis distances. Color indicates the groups. Significant differences were determined using PERMANOVA ($P < .05$). Lower – the same PCoA plots as upper panel highlighting the top 3 most abundant fungal phyla (*Ascomycota, Basidiomycota* and *Mortierellomycotina*). Color scale indicates the relative abundance of each phylum. Groups are defined by shapes.(b) Heatmap showing the normalized abundances of significantly differentially abundant species in Healthy vs ICU^+, Healthy vs ICU^-, ICU^+ vs ICU^- comparisons ($P < .05$, Wilcoxon rank-sum test). Cumulative Sum Scaling (CSS) was used for species abundances normalization. The annotation bars on the left indicate the comparisons in which the species was found to be significant. The names of species are colored according to their affiliated genera. Differentially abundant species in ICU^+ vs Healthy comparison after FDR correction (FDR<0.25) are marked with *.(c) Boxplots comparing the normalized abundances of specific *Candida* species (Wilcoxon rank-sum test).

organ failure despite evidence of infection and, even more concerning, are frequently administered not to miss an occult infection. While the resulting increase of multi-resistant bacteria is an obvious problem, the negative impact on the "holobiont" in the ICU have largely been ignored. More to the point, considerations in intensive care are more dwelling around the early days of introduction of antimicrobial therapy when Paul Ehrlich propagated the concept of "therapia sterilisans magna" where only "parasitotropic" effects in the absence of "organotropic" effects of drugs were envisioned.[100] Thus, molecular therapies for these common and lethal diseases are desperately needed and depend on a better understanding of systems biology of the host metabolome–microbiome interplay.[101] Up to now, the downstream consequences, such as host inflammation and cellular damage, and not the upstream sources, in particular the complex microbial ecosystems that reside in and on the human body, have been the priority of research. Nevertheless, two recent studies in mice [102,103] indicate that Fecal Material Transplantation (FMT) and specific species in the gut microbiome could prevent sepsis opening up new clinical research avenues.

Here, we present evidence from a human study that the selective pressures to which critically ill patients are exposed (parenteral nutrition, polypharmacy, including e.g., proton pump inhibitors, shock states requiring invasive life support measures, such as catecholamine treatment) shape the microbiome of these patients in a unique way with highly distinct characteristics compared to healthy or other disease states, including metabolic, inflammatory, or malignant diseases. As we have shown, the microbiome structure and function of critically ill patients resembles signatures mainly observed in severe infections such as *C. difficile* and *V. cholerae*. This already "infection-vulnerable" microbiome structure in critical illness becomes severely dysbiotic after antibiotic treatment with an observed depletion of SCFAs, including propionate, butyrate, and acetate, and BAs. Similarly, in an elegant mice study *Kim et al.*, [102] demonstrated using FMT from healthy littermates that high levels of butyrate (and potentially also propionate), provided mainly by *Bacteroidetes*, can rescue from lethal sepsis caused by a pathogenic mixture of *K. oxytoca, E. faecalis, Serratia marcescens* and *C. albicans* isolated from a septic patient. Three of these pathogenic species, *K. oxytoca, E. faecalis* and *C. albicans*, were found in significantly higher abundance in our critically ill patients compared to healthy individuals. However, the poor similarity in gut microbial taxonomic abundances between human and mice [104] highlights the importance of investigating the relationship between critical illness and gut microbiota in human clinical samples. For example, while Kim *et al.*, [102] hypothesized that critical illness itself may result in the depletion of gut butyrate. In our human study, this was not the case and only after antibiotics administration we observed a dramatic decrease in the SCFA levels. Nevertheless, we observed a significant change in taxonomy also in ICU^{-} compared to Healthy related to inflammation, such as a significant increase in the abundance of *Staphylococcus* and *Enterococcus*, which have been reported as key factors for the development of signs of systemic inflammation, nosocomial infection and complications in the ICU course.[48,86]

Our study has several limitations. Critically ill patients represent a heterogeneous patient group characterized by comorbidities, past infections and age, leakage of alveolar, and intestinal barriers as well as impaired defense and repair mechanisms.[105,106] Low flow states up to overt shock, autonomic dysfunction, and lack of sufficient options for enteral feeding, e.g., due to dysphagia or impaired consciousness affect gut function, transit time, and defecation.[107] Many drugs that are applied routinely in addition to antibiotics, such as proton pump inhibitors or catecholamines can affect gut function substantially.[99] Nevertheless, while rodent studies allow to control many of the aforementioned confounders, only clinical studies allow to address the impact of antibiotics on the gut microbiome as it relates to clinically meaningful outcomes. Thus, we aimed at control of confounders through inclusion of patient cohorts requiring critical care but not receiving antibiotics.

Patients in ICUs represent a relatively small subgroup of hospitalized patients, but they reflect a specific at-risk population that accounts for approximately 25% of all hospital-acquired infections.[108] Infection as a typical complication of critical illness increases length of ICU stay,

morbidity, mortality, and costs.[109] Moreover, the rise in multidrug-resistant strains prompts attention on nonantibiotic strategies in the prevention and treatment of nosocomial infections, such as pro- or synbiotics. However, mechanisms of action in the vulnerable population of critically ill patients requires further investigation into the mechanisms that shape the gut microbiome. From our study we suggest that the loss of a handful of species, that are highly connected with the production of SCFAs and BAs, during antibiotic administration in the ICU allows the expansion of pathogenic species, which exhibit potential to cause direct hindrance of host homeostasis. Despite the availability of more advanced antibiotics in ICUs death rates from sepsis following nosocomial infections keep increasing, indicating that these antibiotics do not increase survival but instead they produce a highly dysbiotic gut ecosystem that allows more aggressively resistant and lethal pathogens to thrive.[1] These changes are likely associated with or even to promote a state of "protracted critical illness", a frequent observation in patients discharged after prolonged intensive care and characterized by persistent systemic infection. It seems warranted to design studies that aim to restore the gut microbiome or replace key metabolites, such as SCFAs or BAs, in this vulnerable patient population to restore homeostasis of the "metaorganism" after discharge from intensive care.

Materials and methods

Study design

A prospective observational study was undertaken from May 2018 until January 2019 at the Jena University Hospital. Adult critically ill patients either treated with systemic antimicrobial therapy (piperacillin/tazobactam or meropenem) for at least 2 days or without any systemic antimicrobial therapy within the last 7 days were eligible for this study. Patients with inflammatory bowel disease, major bowel resection, selective decontamination of the oral and digestive tract, oral vancomycin therapy, immunocompromised patients, history of chemotherapy during the last 6 months, or known travel history to areas of high antimicrobial resistance within the last 4 weeks were excluded from this study. The need for informed consent was waived by the IRB, since this investigation did not involve any intervention at the patient. The collected basic data were used only for calculating mean values per group and were not linked to the individual metagenomic profiles. Stool samples from healthy volunteers (>50 years old) with no antimicrobial therapy within the last 6 months served as a control group. Healthy volunteers interested in participating were invited for a consultation with medical doctors in the Jena University Hospital. Their health status and prior antibiotic use was self-reported. The basic demographic data for the healthy cohort is provided in Supplementary File 1. We collected up to 100 ml feces which were sampled immediately after natural defecation. Fecal specimens were transferred into two sterile containers, one that was mixed with liquid thioglycolate medium supplemented with catalase and 10% glycerol and one without buffer solution. Both containers were stored at −80°C. Each patient contributed with only one stool sample.

DNA extraction from stool samples

All stool samples were processed by Novogene (UK). DNA was extracted using the following protocol: Stool samples were thoroughly mixed with 900 µL of CTAB lysis buffer. All samples were incubated at 65°C for 60 min before being centrifuged at 12000 × g for 5 min at 4°C. Supernatants were transferred to fresh 2-mL microcentrifuge tubes and 900 µL of phenol: chloroform:isoamyl alcohol (25:24:1, pH = 6.7; Sigma-Aldrich) was added for each extraction. Samples were mixed thoroughly prior to being incubated at room temperature for 10 min. Phase separation occurred by centrifugation at 12,000 × g for 15 min at 4°C, and the upper aqueous phase was re-extracted with a further 900 µL of phenol:chloroform:isoamyl alcohol. Next, samples were centrifuged at 12,000 × g for 10 min at 4°C, and the upper aqueous phases were transferred to fresh 2-mL microcentrifuge tubes. The final extraction was performed with 900 µL of chloroform:isoamyl alcohol (24:1), and layer separation occurred by centrifugation at 12,000 × g for 15 min at 4°C. Precipitation of DNA was achieved by adding the upper phase

from the last extraction step to 450 µL of isopropanol (Sigma-Aldrich) containing 50 µL of 7.5 M ammonium acetate (Fisher). Samples were incubated at – 20°C overnight, although shorter incubations (1 h) produced lower DNA yields. Samples were centrifuged at 7500 × g for 10 min at 4°C, and supernatants were discarded. Finally, DNA pellets were washed three times in 1 mL of 70% (v/v) ethanol (Fisher). The final pellet was air-dried and re-suspended in 200 µL of 75 mM TE buffer (pH = 8.0; Sigma-Aldrich).

Library preparation and sequencing for metagenomics

Sequencing library was generated based on Illumina technologies and following manufactures' recommendations. Index codes were added to each sample. Briefly, the genomic DNA was randomly fragmented to a size of 350 bp, then DNA fragments were narrowly size selected with sample purification beads. The selected fragments were then end polished, A-tailed, and ligated with adapter. These fragments were filtered with beads again and amplified by PCR reaction. At last, the library was analyzed for size distribution and quantified using real-time PCR. The library was then to be sequenced on an Illumina platform Novaseq 6000 (Novogene) with paired-end reads of 150 bp.

Internal transcribed spacer sequencing

The concentration of genomic DNA was determined by Qubit, and the DNA quality was checked on the gel. 200 ng of DNA was used as input for PCR reaction with corresponding primer set specifically binding to different hypervariable regions. Each primer set had a unique barcode. PCR product was then run on the gel and DNA fragment with the proper amplification size was cut and purified. The purified PCR product was then used as template for library preparation. The PCR products were pooled together with equal amount and then end polished, A-tailed, and ligated with the adapter. These fragments were filtered with beads again. After PCR reaction (to make library fully double strand), the library was analyzed for size distribution and quantified using real-time PCR. The library was then to be sequenced on Hiseq2500.

Metabolomics analysis

We performed targeted metabolomics analysis for 38 of the 75 available samples. The remaining samples were destroyed during a prolonged stay in the customs during the COVID-19 pandemic.

Quantification of SCFAs: SCFAs were extracted by addition of 2 mg ultra-pure water pr. mg of sample. The samples were vortex mixed for 1–2 min until suspension is reached, and centrifuged at max speed for 10 min at 4°C. The supernatant was transferred to a spinX centrifuge filter and centrifuged for additional 5 min at 4°C. The filtrate was collected and stored at –20°C until analysis. Sample analysis was carried out by MS-Omics as follows. Samples were acidified using hydrochloric acid, and deuterium labeled internal standards where added. All samples were analyzed in a randomized order. Analysis was performed using a high polarity column (ZebronTM ZB-FFAP, GC Cap. Column 30 m x 0.25 mm x 0.25 µm) installed in a GC (7890B, Agilent) coupled with a quadrupole detector (5977B, Agilent). The system was controlled by ChemStation (Agilent). Raw data was converted to netCDF format using Chemstation (Agilent), before the data was imported and processed in Matlab R2014b (MathWorks, Inc.) using the PARADISe software described by Johnsen et al.[110]

Quantification of BAs: Bile acids were extracted by addition of 4 mg methanol pr. mg of sample. The samples were vortex mixed for 1–2 min until suspension is reached, and centrifuged at max speed for 10 min at 4°C. The supernatant was transferred to a spinX centrifuge filter and centrifuged for additional 5 min at 4°C. In a HPLC vial, 285 µL filtrate is mixed with 15 µL internal standard. The final volume of filtrate of CS10768 and CS10798 where below 285 µl. Therefore, 200 µl were combined with 85 µl Methanol and 15 µl internal standard. Sample analysis was carried out by MS-Omics as follows. The analysis was carried out using a Thermo Scientific Vanquish LC coupled to Thermo Q Exactive HF MS. An electrospray ionization interface was used as ionization source. Analysis was performed in negative ionization mode. The chromatographic separation of bile acids was carried out on a Waters Acquity HSS T3 1.8 µm 2.1 × 150 mm (Waters). The column was thermostated at 30°C. The mobile phases consisted of (A) ammonium

acetate 10 mmol/l, and (B) methanol:acetonitrile (1:1, v/v). Bile acids were eluted by increasing B in A from 45% to 100% for 16 min. Flow rate was 0.3 min. Peak areas were extracted using Tracefinder 4.1 (Thermo Scientific). Identification of compounds were based on accurate mass and retention time of authentic standards.

Data processing

Quality control of sequence data

Quality control to remove low-quality reads was performed as described previously.[111] Briefly, all Illumina primer/adapter/linker sequences were removed. Subsequently, low-quality regions (consecutive regions with Phred quality <20) were trimmed. Finally, all reads were mapped to the human genome with BWA version 0.7.4 [112] and reads with >95% identity and 90% coverage were removed as human DNA contamination.

Taxonomic profiling

Taxonomic annotation of the high-quality reads was performed using MetaPhlAn2[42] version 2.7.7 with default settings, generating taxonomic relative abundances. Bacterial community profiles were constructed at phylum, genus and species level for further analyses.

Taxonomic annotation of fungal ITS was performed using the DADA2 pipeline[97] version 1.14 with default parameters including adapter removal, quality filtering and trimming, dereplication of identical reads, read-pair merging, ITS2 extraction and chimera removal. Remaining reads were binned as operational taxonomic units and aligned to the UNITE fungi database using RDP classifier.[113] All samples were then normalized by cumulative sum scaling using R package metagenomeSeq.[114]

Functional annotation

The HUMAnN2 pipeline [49] version 0.11.2 was used for functional annotation of the high-quality reads after the quality control. The quantified pathway and gene family abundances in the units of RPKs (read per kilobase) were then normalized to copies per million (CPM) units by the provided HUMAnN2 script, resulting in transcript-per-million-like (TPM) normalization. Gene families were then regrouped to Pfam domains for further analyses.

Abundance comparisons

Species, pathways, and Pfams were filtered by 10% prevalence across all samples and their relative abundances were used for statistical comparisons between the three groups. Differentially abundant features were identified by the Wilcoxon rank-sum test and were considered significantly differentially abundant if the *P*-value was less than 0.05.

Differentially abundant phyla and genera were identified by the Kruskal-Wallis test and were considered significantly differentially abundant if the *P*-value was less than 0.05.

Metagenomics sequences from HUMAnN2 profiles

Gene family abundances were clustered using mgs-canopy[46] version 1.0 software with standard parameters. Gene family clusters were considered metagenomic sequences (MGS) if they had at least 700 genes. Taxonomic annotation of MGS was done using species annotation information available for each gene family.

We calculated contributions of each species to an MGS. An MGS was annotated to the species with the largest contribution if: the gene contribution of that species was more than 50% and the second largest species was “unclassified” or contributed less than 10%.

Diversity analysis

Alpha diversity indices Shannon, Simpson, and Chao1 were calculated using the R packages vegan[75] and fossil[115] based on relative species abundance. Wilcoxon rank-sum test was used to test for significant differences in alpha diversity. For estimating community dissimilarities, Bray–Curtis distances were calculated using the R package vegan[75] based on the relative species abundance. To test for significant differences in the microbial composition, permutational multivariate analysis of variance (PERMANOVA), as implemented in the function adonis from R package vegan,[75] was used to analyze beta-diversity.

Co-abundance networks

The relative abundance table for significantly different species was processed using SparCC [116] for co-abundance network inference. Species–species correlation coefficients were estimated as the

average of 20 inference iterations and 100 permutations were used for the pseudo *P*-value calculation. For the visualization of the co-abundance network, only interactions with an absolute correlation coefficient >0.4 were used.

Metabolic modeling

To estimate the availability and composition of metabolites in ICU patients, the nutrition fed in ICU (https://www.fresenius-kabi.com/de/ernaehrung/fresubin-original) was considered. Metabolic composition of complex products such as fish oil was described by vmh diet designer (https://www.vmh.life/#nutrition/dietdesigner). The human genome-scale metabolic model Recon3D 3.01 [117] simulated based on the ICU specific diet (Supplementary File 2) was used to predict metabolites that can potentially be secreted by the host. Flux Variability Analysis (FVA) [88] was used to determine feasible exchange reaction flux bounds that support metabolite secretion alongside optimal growth rate. Identified metabolites were assumed to be available for the bacterial species and strains to be consumed. Genome-scale metabolic models of the studied species and strains were collected from two different gut model repositories, AGORA 1.03 (https://www.vmh.life) [85] and CarveMe.[118] Taking into account the availability of ICU diet compounds and potential host-secreted metabolites in the gut, the maximum amount of SCFAs (acetic acid, propionic acid, and butyric acid) production by bacterial species and strains were predicted by applying FVA [88] again alongside achieving maximum ATP yield as objective for the available bacterial metabolic models.

Abundance of ARGs

The metagenomic reads were analyzed using the deepARG pipeline,[98] which uses deep learning to identify and quantify ARGs. Reads were compared to the provided DeepARG-DB database using a prediction model to evaluate sequence similarities and predict antibiotic resistance. The pipeline was run in short sequence mode with a minimum probability cutoff of 0.8, an identity cutoff of 80%, an E-value cutoff of 1e-10 and a minimum coverage of 50%.

Testing for significant differences in ARG abundance was performed using R package dabestr [119] with a confidence interval of 95%.

Bacterial growth rate estimation

Bacterial growth rate was calculated using the growth rate index (GRiD) [89] version 1.2. The algorithm first calculates the coverage of all contigs of a reference genome in the sample, sorts them from high to low, and reorders them to two groups, placing an *ori*-containing contig at start and a *ter*-containing contig at the mid-region of the genome. Next, it calculates coverage drops across a sliding 10 Kb window, with values representing the coverage ratio of the peak and trough of the curve. High values represent faster growth rates.

Random forest model

A Random Forest classifier was built to classify ICU patients into ICU^- and ICU^+ based on bacterial taxonomic profiles and pathways. The model was implemented using R package caret [120] with all bacterial species and pathway abundances as input features. The model was trained after centering and scaling the data and removing variables with near zero variance, using a tune length of 10 and fivefold cross-validation as resampling method, the rest of the parameters were left as default. Feature importance were calculated using function varImp from R package caret.[120] A random forest model was then built using only the 20 most important features. R package PRROC [121] was used for ROC calculation and plot.

Bacterial exposure and survival assay in C. elegans

Bacterial strains used were *Escherichia coli* OP50, *Lactobacillus crispatus* (DSM 20356), and *Bifidobacterium animalis* (DSM 20104). All strains except *E. coli* were acquired from the German Collection of Microorganisms and Cell Cultures and cultured following the supplier guidelines. Bacterial stocks were kept at −80°C until use. *C. elegans* strain used was N2 wild-type isolate obtained from the Caenorhabditis Genetics Center.

Bacterial stocks were inoculated in anoxic broth (MRS for *L. crispatus* and BSM for *B. animalis*) and incubated at 37°C for 72 h. 150 µL of each bacterial culture was spread onto medium sized NGM agar plates (5,5 cm diameter) and incubated for 24 h at 37°C in an anaerobic container (BD GasPak™ EZ

container systems) prior to worm addition. *E. coli* was grown on NGM in normoxia. UV-killed OP50 was produced by exposing NGM plates to UV light (320 nm) in a Chemi-Doc XRS+ transilluminator (BioRad) for 10 min.

Age-synchronized germ-free worms were obtained by collecting eggs from gravid adults upon treatment with alkaline hypochlorite solution (composition per liter: 200 mL 1 M KOH, 250 mL bleach, and 500 mL ddH2O) and letting the eggs hatch overnight at 20°C in M9 buffer. Synchronized L1 larvae were seeded onto NGM plates containing respective bacterial strains (4 plates per strain). Approximately 150 worms were seeded per plate. Plates were incubated at 20°C in an anaerobic container for 5 h or 24 h. After incubation worms were washed with M9 buffer and transferred to UV-killed OP50 plates. UV-killed OP50 plates were incubated at 20°C in normoxic conditions for 24 h prior to survival assessment. Heat-killed bacteria was obtained by submerging bacterial cultures at 80°C for 60 min and OD600 was normalized to 0,2 before placing the cultures onto NGM plates. Live bacteria was also subjected to OD600 normalization in all tests, which involved heat-killed control conditions.

Survival rate was assessed by screening all the worms present in each plate. A worm was regarded as dead if it did not respond to gentle touch with a platinum wire. Survival was expressed as percentage of the total number worms. Each experiment was performed 3 times.

Developmental fitness and reproductive aging assays in C. elegans

Bifidobacterium animalis was grown for 72 h on anoxic BSM broth at 37°C. Afterward, 150 uL of bacterial culture was spread onto medium-sized NGM plates (4 plates) and incubated for 24 h at 37°C in an anaerobic container (BD GasPak™ EZ container systems) prior to worm addition. *E. coli* OP50 was grown on NGM at normoxic conditions. *C. elegans* N2 population was synchronized as described above and approximately 150 worms were seeded on either *B. animalis* or OP50 plates. Plates were incubated for 24 h at 20°C in the anaerobic container. Developmental assay was carried out as previously described.[122] In brief, 30 worms per bacterium were put individually onto small UV-killed OP50 plates and incubated for 48 h at 20°C before developmental stage of each worm was visually assessed. Reproductive aging assay was carried out as described previously.[123] In brief, after incubation with *B. animalis* or OP50 the worms were washed with M9 and let to develop until L4 stage on UV-killed OP50 plates at 20°C, normoxia. At this moment 25 randomly picked worms (per condition) were transferred individually onto small-sized UV-killed OP50 plates. Every day the brood size of each worm was determined (sum of eggs and L1s) and parent worm was transferred to new plate until egg laying ceased. These experiments were performed 3 times.

DAF-16 nuclear translocation assay

Nematodes expressing DAF-16::GFP fusion protein were obtained from the *Caenorhabditis Genetics Center* (strain TJ356). Bacterial strains used were the same as described above. Anoxic broth was inoculated with an aliquot of actively growing bacterial culture and incubated at 37°C for 48 h. OD600 of all living cultures was normalized to 0,1. 150 µL of each bacterial culture was spread onto medium sized NGM agar plates (5,5 cm diameter) and incubated for 24 h at 37°C in an anaerobic container (BD GasPak™ EZ container systems) prior to worm addition. *E. coli* was grown on NGM in normoxia. Heat-killed bacteria was generated as described above and OD600 was normalized to 0,2 prior to seeding onto NGM plates.

Age-synchronized germ-free worms were obtained as described above and grown until L4 stage on UV-killed OP50. L4 worms were washed with M9 buffer and transferred to NGM plates containing bacteria. Approximately 150 worms were seeded per plate. Plates were incubated at 20°C in an anaerobic container (BD GasPak™ EZ container systems) on either live or heat-killed bacteria for 5 h. Control plates were incubated in normoxia on live bacteria only. After incubation, 30 worms were picked from each condition and transferred to empty NGM plates for imaging. Imaging was carried out using a ZEISS Axio Zoom.V16 microscope equipped with fluorescence light. Imaged worms were sorted into three categories (nuclear, intermediate, cytosolic) depending on the localization of the GFP tagged DAF-16 transcription factor. This experiment was performed three times.

Statistical analysis

To determine differential abundance of taxonomic, functional, and metabolic features between groups Wilcoxon two-tailed rank-sum test was applied when analyzing the differences between two groups, whereas Kruskal–Wallis test was used when more than two groups were compared. Correlation between microbial taxa and metabolites was assessed by Spearman's correlation. The R package dabestr[119] was used to test differential abundance of ARGs. Significant differences in source contributions to sinks using FEAST[43] were assessed using Wilcoxon two-tailed rank sum test. To assess differences in alpha diversity and beta dispersion, Wilcoxon two-tailed rank sum test was used, whereas PERMANOVA was used for beta diversity.

Disclosure statement

The authors declare that they have no competing interests.

Funding

GP, MB, AMS, AB, ME, and LZ would like to thank Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) CRC/Transregio 124 'Pathogenic fungi and their human host: Networks of interaction', subproject B5 and INF and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Germany's Excellence Strategy - EXC 2051 – Project-ID 390713860 for financial support. LZ and CB have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Project: 802736 MORPHEUS).

Availability of data and material

Raw sequence files were deposited into the NCBI Sequence Read Archive (SRA) under accession number PRJNA670323.

Ethics approval and consent to participate

The study was approved by the local ethics committee in Jena, Germany (5289-10/17). Informed consent was waived by the ethics committee because feces were only collected after natural defecation and no patient data were obtained.

Author contributions

GP and MB designed the study. MB was involved in sample collection and processing. GP designed and supervised the bioinformatics analysis and edited the manuscript. AMS and LZ carried out the data analysis and data integration. MM performed the metabolic modelling. BS and YN helped with the interpretation of results and revised the manuscript. RS carried out the quality control and pre-processing of the metagenomic data, and together with AB and CB helped with the interpretation of results. PAP, MM and ME designed and performed the animal studies. All authors read and approved the final manuscript.

References

1. Milbrandt E, Kersten A, Rahim M, Dremsizov T, Clermont G, Cooper L, Angus D. 2008. Growth of intensive care unit resource use and its estimated cost in medicare. Crit Care Med. 36(9):2504–2510. doi:10.1097/CCM.0b013e318183ef84.
2. Brun-Buisson C. 2000. The epidemiology of the systemic inflammatory response. Intensive Care Med [Internet]. 26(S1):S064–74. doi:10.1007/s001340051121.
3. Adib-Conquy M, Cavaillon JM. 2009. Compensatory anti-inflammatory response syndrome. Thromb Haemost. 101(1):36–47. doi:10.1160/TH08-07-0421.
4. Boomer JS, To K, Chang KC, Takasu O, Osborne DF, Walton AH, Bricker TL, Jarman SD, Kreisel D, Krupnick AS, et al. 2011.Immunosuppression in patients who die of sepsis and multiple organ failure. JAMA - J Am Med Assoc. 306(23):2594–2605. doi:10.1001/jama.2011.1829.
5. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. 2001. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med. 29 (7):1303–1310. doi:10.1097/00003246-200107000-00002.
6. Vincent J-L, Sakr Y, Singer M, Martin-Loeches I, Machado FR, Marshall JC, Finfer S, Pelosi P, Brazzi L, Aditianingsih D, et al. 2020. Prevalence and outcomes of infection among patients in intensive care units in 2017. JAMA [Internet]. 323(15):1478. doi:10.1001/jama.2020.2717.
7. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, Kumar A, Sevransky JE, Sprung CL, Nunnally ME, et al. 2017. Surviving sepsis campaign. Crit Care Med. 45:486–552.
8. Silvestri L, De La Cal M, Van Saene H. Selective decontamination of the digestive tract: the mechanism of action is control of gut overgrowth. Intensive Care Med. 2012;38.

9. Fridkin S, Baggs J, Fagan R, Magill S, Pollack LA, Malpiedi P, Slayton R, Khader K, Rubin MA, Jones M, et al. 2014. Vital signs: improving antibiotic use among hospitalized patients. MMWR Morb Mortal Wkly Rep [Internet]. 63:194–200. Available from. http://www.ncbi.nlm.nih.gov/pubmed/24598596%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4584728
10. Ravi A, Halstead FD, Bamford A, Casey A, Thomson NM, van Schaik W, Snelson C, Goulden R, Foster-Nyarko E, Gm S, et al. 2019. Loss of microbial diversity and pathogen domination of the gut microbiota in critically ill patients. Microb Genomics [Internet]. 5(9).
11. McKenney PT, Pamer EG. 2015. From hype to hope: the gut microbiota in enteric infectious disease. Cell [Internet] ; 163(6):1326–1332. doi:10.1016/j.cell.2015.11.032.
12. Jacobson A, Lam L, Rajendram M, Tamburini F, Honeycutt J, Pham T, Van Treuren W, Pruss K, Stabler SR, Lugo K, et al. 2018. A gut commensal-produced metabolite mediates colonization resistance to salmonella infection. Cell Host Microbe [Internet] ; 24(2):296–307.e7. doi:10.1016/j.chom.2018.07.002.
13. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T, et al. 2011.Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature. 469(7331):543–547. doi:10.1038/nature09646.
14. Kang JD, Myers CJ, Harris SC, Kakiyama G, Lee I-K, Yun B-S, Matsuzaki K, Furukawa M, Min H-K, Bajaj JS, et al. 2019. Bile acid 7α-Dehydroxylating gut bacteria secrete antibiotics that inhibit clostridium difficile: role of secondary bile acids. Cell Chem Biol Internet]. 26 (1):27–34.e4. doi:10.1016/j.chembiol.2018.10.003.
15. Alavi S, Mitchell JD, Cho JY, Liu R, Macbeth JC, Hsiao A. 2020. Interpersonal gut microbiome variation drives susceptibility and resistance to cholera infection. Cell [Internet]. 181(7):1533–1546.e13. doi:10.1016/j.cell.2020.05.036.
16. Hand TW. 2016. The role of the microbiota in shaping infectious immunity. Trends Immunol Internet. 37 (10):647–658. doi:10.1016/j.it.2016.08.007.
17. Armour CR, Nayfach S, Pollard KS, Sharpton TJ. 2019. A metagenomic meta-analysis reveals functional signatures of health and disease in the human gut microbiome. mSystems. 4(4):1–15. doi:10.1128/mSystems.00332-18.
18. McDonald D, Ackermann G, Khailova L, Baird C, Heyland D, Kozar R, Lemieux M, Derenski K, King J, Vis-Kampen C, et al. 2016. Extreme dysbiosis of the microbiome in critical illness. mSphere [Internet]. 1(4).
19. Hernández-Santos N, Klein BS. 2017. Through the scope darkly: the Gut mycobiome comes into focus. Cell Host Microbe. 22(6):728–729. doi:10.1016/j.chom.2017.11.013.
20. Zhang E, Tanaka T, Tajima M, Tsuboi R, Nishikawa A, Sugita T. 2011. Characterization of the skin fungal microbiota in patients with atopic dermatitis and in healthy subjects. Microbiol Immunol. 55(9):625–632. doi:10.1111/j.1348-0421.2011.00364.x.
21. Huseyin CE, O'Toole PW, Cotter PD, Scanlan PD. 2017. Forgotten fungi-the gut mycobiome in human health and disease. FEMS Microbiol Rev.
22. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. PLoS Pathog. 6(1):e1000713. doi:10.1371/journal.ppat.1000713.
23. Nguyen LDN, Viscogliosi E, Delhaes L. 2015. The lung mycobiome: an emerging field of the human respiratory microbiome. Front Microbiol. 6. doi:10.3389/fmicb.2015.00089.
24. Witherden EA, Moyes DL. Mycrobiome and gut inflammation: implications in gut disease. In: Immunity and inflammation in health and disease. Elsevier; 2018. p. 271–280.
25. van Woerden HC, Gregory C, Brown R, Marchesi JR, Hoogendoorn B, Matthews IP. 2013. Differences in fungi present in induced sputum samples from asthma patients and non-atopic controls: a community based case control study. BMC Infect Dis. 13(1). doi:10.1186/1471-2334-13-69.
26. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD. 2013. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One.
27. Sokol H, Leducq V, Aschard H, Pham HP, Jegou S, Landman C, Cohen D, Liguori G, Bourrier A, Nion-Larmurier I, et al. 2017. Fungal microbiota dysbiosis in IBD. Gut. 66(6):1039–1048. doi:10.1136/gutjnl-2015-310746.
28. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosh D, Panzer AR, Lamere B, Rackaityte E, Lukacs NW, et al. 2016. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat Med. 22(10):1187–1191. doi:10.1038/nm.4176.
29. Velegraki A, Cafarchia C, Gaitanis G, Iatta R, Boekhout T. 2015. Malassezia infections in humans and animals: pathophysiology, detection, and treatment. PLoS Pathog. 11(1):e1004523. doi:10.1371/journal.ppat.1004523.
30. Yang AM, Inamine T, Hochrath K, Chen P, Wang L, Llorente C, Bluemel S, Hartmann P, Xu J, Koyama Y, et al. 2017. Intestinal fungi contribute to development of alcoholic liver disease. J Clin Invest. 127(7):2829–2841. doi:10.1172/JCI90562.
31. Dollive S, Chen YY, Grunberg S, Bittinger K, Hoffmann C, Vandivier L, Cuff C, Lewis JD, Wu GD, Bushman FD. 2013. Fungi of the murine

gut: episodic variation and proliferation during antibiotic treatment. PLoS One. 8(8):e71806. doi:10.1371/journal.pone.0071806.

32. Noverr MC, Noggle RM, Toews GB, Huffnagle GB. 2004. Role of antibiotics and fungal microbiota in driving pulmonary allergic responses. Infect Immun. 72 (9):4996–5003. doi:10.1128/IAI.72.9.4996-5003.2004.
33. Fan D, Coughlin LA, Neubauer MM, Kim J, Kim MS, Zhan X, Simms-Waldrip TR, Xie Y, Hooper LV, Koh AY. 2015. Activation of HIF-1α and LL-37 by commensal bacteria inhibits Candida albicans colonization. Nat Med.
34. Krause R, Halwachs B, Thallinger GG, Klymiuk I, Gorkiewicz G, Hoenigl M, Prattes J, Valentin T, Heidrich K, Buzina W, et al. 2016.Characterisation of Candida within the mycrobiome/microbiome of the lower respiratory tract of ICU patients. PLoS One. 11(5):1–29. doi:10.1371/journal.pone.0155033.
35. Krause R, Moissl-Eichinger C, Halwachs B, Gorkiewicz G, Berg G, Valentin T, Prattes J, Högenauer C, Zollner-Schwetz I. 2017. Mycrobiome in the lower respiratory tract – a clinical perspective. Front Microbiol. 07. doi:10.3389/fmicb.2016.02169.
36. Arastehfar A, Carvalho A, Nguyen MH, Hedayati MT, Netea MG, Perlin DS, Hoenigl M. 2020. COVID-19-Associated Candidiasis (CAC): an underestimated complication in the absence of immunological predispositions? J Fungi. 6(4):211. doi:10.3390/jof6040211.
37. Lagunes L, Rello J. 2016. Invasive candidiasis: from mycobiome to infection, therapy, and prevention. Eur J Clin Microbiol Infect Dis. 35(8):1221–1226. doi:10.1007/s10096-016-2658-0.
38. Watkins RR, Mukherjee PK, Chandra J, Retuerto MA, Guidry C, Haller N, Paranjape C, Ghannoum MA. 2017. Admission to the intensive care unit is associated with changes in the oral mycobiome. J Intensive Care Med. 32(4):278–282. doi:10.1177/0885066615627757.
39. Heisel T, Nyaribo L, Sadowsky MJ, Gale CA. 2019. Breastmilk and NICU surfaces are potential sources of fungi for infant mycrobiomes. Fungal Genet Biol. 128:29–35. doi:10.1016/j.fgb.2019.03.008.
40. Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer AC, De Oliveira Lino FS, Chan KF, Loos D, Imamovic L, et al. 2020. Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. Microbiome. 8(1). doi:10.1186/s40168-020-00899-6.
41. Calandra T, Cohen J. 2005. The international sepsis forum consensus conference on definitions of infection in the intensive care unit. Crit Care Med. 33(7):1538–1548. doi:10.1097/01.CCM.0000168253.91200.83.
42. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods [Internet]. 12(10):902–903. doi:10.1038/nmeth.3589.
43. Shenhav L, Thompson M, Joseph TA, Briscoe L, Furman O, Bogumil D, Mizrahi I, Pe'er I, Halperin E. 2019. FEAST: fast expectation-maximization for microbial source tracking. Nat Methods. 16(7):627–632. [Internet]. doi:10.1038/s41592-019-0431-x.
44. Abeles SR, Jones MB, Santiago-Rodriguez TM, Ly M, Klitgord N, Yooseph S, Nelson KE, Pride DT. 2016. Microbial diversity in individuals and their household contacts following typical antibiotic courses. Microbiome [Internet]. 4(1):39. doi:10.1186/s40168-016-0187-9.
45. Burdet C, Nguyen TT, Duval X, Ferreira S, Andremont A, Guedj J, Mentré F, Cherng B, Su Gin DC, Rama ND, et al. 2019. Impact of antibiotic gut exposure on the temporal changes in microbiome diversity. Antimicrob Agents Chemother [Internet]. 63. https://aac.asm.org/content/63/10/e00820-19 .
46. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, et al. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol. 32(8):822–828. doi:10.1038/nbt.2939.
47. Foster ZSL, Sharpton TJ, Grünwald NJ. 2017. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. PLOS Comput Biol [Internet]. 13(2):e1005404. doi:10.1371/journal.pcbi.1005404.
48. Fiore E, Van Tyne D, Gilmore MS. 2019. Pathogenicity of Enterococci. Microbiol Spectr [Internet]. 7(4).
49. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, et al. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods [Internet]. 15(11):962–968. doi:10.1038/s41592-018-0176-y.
50. Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, Beghini F, Malik F, Ramos M, Dowd JB, et al. 2017. Accessible, curated metagenomic data through experimentHub. Nat Methods [Internet]. 14 (11):1023–1024. doi:10.1038/nmeth.4468.
51. Hannigan GD, Duhaime MB, Ruffin MT, Koumpouras CC, Schloss PD. 2018. Diagnostic potential and interactive dynamics of the colorectal cancer virome. MBio. 9(6). doi:10.1128/mBio.02248-18.
52. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, et al. 2015. Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat Commun. 6 (1):6528. doi:10.1038/ncomms7528.

53. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, Beghini F, Manara S, Karcher N, Pozzi C, et al. 2019. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med. 25:667–678.
54. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork P, Sinha R. 2016. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. PLoS One. 11(5):e0155362. doi:10.1371/journal.pone.0155362.
55. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, et al. 2017. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 66(1):70–78. doi:10.1136/gutjnl-2015-309800.
56. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, et al. 2014. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 10(11):766. doi:10.15252/msb.20145645.
57. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto PA, Vicente D, Hoffman K, Wei SC, et al. 2018. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 359(6371):97–103. doi:10.1126/science.aan4236.
58. Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre M-L, Luke JJ, Gajewski TF. 2018. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science. 359(6371):104–108. doi:10.1126/science.aao3290.
59. Wen C, Zheng Z, Shao T, Liu L, Xie Z, Le Chatelier E, He Z, Zhong W, Fan Y, Zhang L, et al. 2017. Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. Genome Biol. 18(1):142. doi:10.1186/s13059-017-1271-6.
60. Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, et al. 2017.The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun. 8(1):845. doi:10.1038/s41467-017-00900-1.
61. Ye Z, Zhang N, Wu C, Zhang X, Wang Q, Huang X, Du L, Cao Q, Tang J, Zhou C, et al. 2018.A metagenomic study of the gut microbiome in Behcet's disease. Microbiome. 6(1):135. doi:10.1186/s40168-018-0520-6.
62. Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, Wu S, Liu W, Cui Q, Geng B, et al. 2017.Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome. 5(1):1–19. doi:10.1186/s40168-016-0222-x.
63. Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, et al. 2014. Alterations of the human gut microbiome in liver cirrhosis. Nature [Internet]. 513(7516):59–64. 10.1038/nature13568.
64. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, Huerta-Cepas J, Nieuwdorp M, Salojärvi J, Voigt AY, et al. 2016.Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science. 352(6285):586–589. doi:10.1126/science.aad8852.
65. Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, Dulai PS, Caussy C, Bettencourt R, Highlander SK, et al. 2017.Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. Cell Metab. 25(5):1054–1062.e5. doi:10.1016/j.cmet.2017.04.001.
66. Louis S, Tappu R-M, Damms-Machado A, Huson DH, Bischoff SC. 2016. Characterization of the gut microbial community of obese patients following a weight-loss intervention using whole metagenome shotgun sequencing. PLoS One. 11(2):e0149564. doi:10.1371/journal.pone.0149564.
67. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, et al. 2013.Richness of human gut microbiome correlates with metabolic markers. Nature. 500(7464):541–546. doi:10.1038/nature12506.
68. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen A-M, Peet A, Tillmann V, Pöhö P, Mattila I, et al. 2015.The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe. 17(2):260–273. doi:10.1016/j.chom.2015.01.001.
69. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, Wampach L, Schneider JG, Hogan A, De Beaufort C, et al. 2016. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. Nat Microbiol. 2:16180. doi:10.1038/nmicrobiol.2016.180.
70. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, Nielsen J, Bäckhed F. 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 498(7452):99–103. doi:10.1038/nature12198.
71. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al. 2012.A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 490(7418):55–60. doi:10.1038/nature11450.
72. Kieser S, Sarker SA, Sakwinska O, Foata F, Sultana S, Khan Z, Islam S, Porta N, Combremont S, Betrisey B, et al. 2018. Bangladeshi children with acute diarrhoea show faecal microbiomes with increased Streptococcus abundance, irrespective of diarrhoea aetiology. Environ Microbiol [Internet]. 20(6):2256–2269. doi:10.1111/1462-2920.14274.
73. Vincent C, Miller MA, Edens TJ, Mehrotra S, Dewar K, Manges AR. 2016. Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and Clostridium difficile colonization or infection. Microbiome [Internet]. 4(1):12. doi:10.1186/s40168-016-0156-3.

74. David LA, Weil A, Ryan ET, Calderwood SB, Harris JB, Chowdhury F, Begum Y, Qadri F, LaRocque RC, Turnbaugh PJ. 2015. Gut microbial succession follows acute secretory diarrhea in humans. MBio [Internet]. 6 (3):e00381–15. doi:10.1128/mBio.00381-15.
75. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci [Internet]. 14 (6):927–930. doi:10.1111/j.1654-1103.2003.tb02228.x.
76. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. 2019. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. MBio [Internet]. 10(1):e02566–18. doi:10.1128/mBio.02566-18.
77. Heinken A, Ravcheev DA, Baldini F, Heirendt L, Fleming RMT, Thiele I. 2019. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome. 7(1):1–18. doi:10.1186/s40168-019-0689-3.
78. Mukherjee A, Lordan C, Ross RP, Cotter PD. 2020. Gut microbes from the phylogenetically diverse genus Eubacterium and their various contributions to gut health. Gut Microbes [Internet]. 12(1):1802866. doi:10.1080/19490976.2020.1802866.
79. Podschun R, Ullmann U. 1998. Klebsiella spp. as Nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin Microbiol Rev Internet]. 11(4):589–603. doi:10.1128/CMR.11.4.589.
80. Devlin AS, Fischbach MA. 2015. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat Chem Biol [Internet]. 11(9):685–690: doi:10.1038/nchembio.1864.
81. Liu H, Tian R, Wang H, Feng S, Li H, Xiao Y, Luan X, Zhang Z, Shi N, Niu H, et al. 2020. Gut microbiota from coronary artery disease patients contributes to vascular dysfunction in mice by regulating bile acid metabolism and immune activation. J Transl Med [Internet]. 18 (1):382. doi:10.1186/s12967-020-02539-x.
82. Morrison DJ, Preston T. 2016. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes [Internet]. 7 (3):189–200. doi:10.1080/19490976.2015.1134082.
83. Mirković B, Murray MA, Lavelle GM, Molloy K, Azim AA, Gunaratnam C, Healy F, Slattery D, McNally P, Hatch J, et al. 2015. The role of short-chain fatty acids, produced by anaerobic bacteria, in the cystic fibrosis airway. Am J Respir Crit Care Med [Internet]. 192(11):1314–1324. doi:10.1164/rccm.201505-0943OC.
84. Horiuchi H, Kamikado K, Aoki R, Suganuma N, Nishijima T, Nakatani A, Kimura I. 2020. Bifidobacterium animalis subsp. lactis GCL2505 modulates host energy metabolism via the short-chain fatty acid receptor GPR43. Sci Rep [Internet]. 10(1):4158. doi:10.1038/s41598-020-60984-6.
85. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, et al. 2017. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol [Internet]. 35(1):81–89. doi:10.1038/nbt.3703.
86. Pons MJ, Ruiz J. 2019. Current trends in epidemiology and antimicrobial resistance in intensive care units. J Emerg Crit Care Med [Internet]. 3:5. doi:10.21037/jeccm.2019.01.05.
87. Han B, Sivaramakrishnan P, Lin -C-CJ, Neve IAA, He J, Tay LWR, Sowa JN, Sizovs A, Du G, Wang J, et al. 2017. Microbial genetic composition tunes host longevity. Cell [Internet]. 169:1249–1262.e13. http://www.ncbi.nlm.nih.gov/pubmed/28622510
88. Mahadevan R, Schilling CH. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng [Internet]. 5(4):264–276. doi:10.1016/j.ymben.2003.09.002.
89. Emiola A, Oh J. 2018. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. Nat Commun [Internet]. 9 (1):4956. doi:10.1038/s41467-018-07240-8.
90. Libina N, Berman JR, Kenyon C. 2003. Tissue-specific activities of C. elegans DAF-16 in the regulation of lifespan. Cell. 115(4):489–502. doi:10.1016/S0092-8674(03)00889-4.
91. Murphy CT. 2006. The search for DAF-16/FOXO transcriptional targets: approaches and discoveries. Exp Gerontol. 41(10):910–921. doi:10.1016/j.exger.2006.06.040.
92. Kenyon CJ. 2010. The genetics of ageing. Nature.
93. Jungersen M, Wind A, Johansen E, Christensen JE, Stuer-Lauridsen B, Eskesen D. 2014. The science behind the probiotic strain bifidobacterium animalis subsp. lactis BB-12(®). Microorganisms [Internet]. 2:92–110. (2): doi:10.3390/microorganisms2020092.
94. Paumgartner G. 2002. Ursodeoxycholic acid in cholestatic liver disease: mechanisms of action and therapeutic use revisited. Hepatology [Internet]. 36:525–531. (3): doi:10.1053/jhep.2002.36088.
95. Keely SJ, Steer CJ, Lajczak-McGinley NK. 2019. Ursodeoxycholic acid: a promising therapeutic target for inflammatory bowel diseases? Am J Physiol Liver Physiol [Internet]. 317:G872–81.
96. Theriot CM, Bowman AA, Young VB. 2016. Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for clostridium difficile spore germination and outgrowth in the large intestine. mSphere. 1(1). doi:10.1128/mSphere.00045-15.
97. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods [Internet]. 13(7):581–583. doi:10.1038/nmeth.3869.

98. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 6(1):1–15. doi:10.1186/s40168-018-0401-z.
99. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, et al. 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature [Internet]. 555 (7698):623–628. doi:10.1038/nature25979.
100. Sörgel F. 2004. The return of Ehrlich's 'Therapia magna sterilisans' and other Ehrlich concepts? Chemotherapy. 50(1):6–10. doi:10.1159/000077277.
101. Khaliq W, Großmann P, Neugebauer S, Kleyman A, Domizi R, Calcinaro S, Brealey D, Gräler M, Kiehntopf M, Schäuble S, et al. 2020. Lipid metabolic signatures deviate in sepsis survivors compared to non-survivors. Comput Struct Biotechnol J. 18:3678–3691. doi:10.1016/j.csbj.2020.11.009.
102. Kim SM, DeFazio JR, Hyoju SK, Sangani K, Keskey R, Krezalek MA, Khodarev NN, Sangwan N, Christley S, Harris KG, et al. 2020. Fecal microbiota transplant rescues mice from human pathogen mediated sepsis by restoring systemic immunity. Nat Commun [[Internet]. 11:1–11. doi:10.1038/s41467-020-15545-w.
103. Singer JR, Blosser EG, Zindl CL, Silberger DJ, Conlan S, Laufer VA, DiToro D, Deming C, Kumar R, Morrow CD, et al. 2019. Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis. Nat Med [Internet]. 25 (11):1772–1782. doi:10.1038/s41591-019-0640-y.
104. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. 2015. How informative is the mouse for human gut microbiota research? Dis Model Mech [Internet]. 8:1–16. doi:10.1242/dmm.017400.
105. Esper AM, Martin GS. 2011. The impact of comorbid conditions on critical illness. Crit Care Med [Internet]. 39 (12):2728–2735. doi:10.1097/CCM.0b013e318236f27e.
106. Huber-Lang M, Lambris JD, Ward PA. 2018. Innate immune responses to trauma. Nat Immunol [Internet]. 19(4):327–341. doi:10.1038/s41590-018-0064-8.
107. Ladopoulos T. 2018. Gastrointestinal dysmotility in critically ill patients. Ann Gastroenterol [[Internet]. Available from]. http://www.annalsgastro.gr/files/journals/1/earlyview/2018/ev-03-2018-10-AG3394-0250.pdf
108. Alonso-Echanove J, Gaynes RP. Scope and magnitude of nosocomial ICU infections. [Internet]. 2001.
109. Barrasa-Villar JI, Aibar-Remón C, Prieto-Andrés P, Mareca-Doñate R, Moliner-Lahoz J. 2017. Impact on morbidity, mortality, and length of stay of hospital-acquired infections by resistant microorganisms. Clin Infect Dis [Internet]. 65 (4):644–652. :doi:10.1093/cid/cix411.
110. Johnsen LG, Skou PB, Khakimov B, Bro R. 2017. Gas chromatography – mass spectrometry data processing made easy. J Chromatogr A [Internet]. 1503:57–64. Available from. https://linkinghub.elsevier.com/retrieve/pii/S0021967317306489 .
111. Li J, Sung CYJ, Lee N, Ni Y, Pihlajamäki J, Panagiotou G, El-Nezami H. 2016. Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice. Proc Natl Acad Sci [Internet]. 113(9):E1306–15. doi:10.1073/pnas.1518189113.
112. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Oxford Univ Press.
113. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol [Internet] ; 73(16):5261–5267. doi:10.1128/AEM.00062-07.
114. Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial marker-gene surveys. Nat Methods [Internet]. 10(12):1200–1202. doi:10.1038/nmeth.2658.
115. Vavrek MJ. 2011. fossil: palaeoecological and palaeogeographical analysis tools. Palaeontol Electron. 14.
116. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 8(9):1–11. doi:10.1371/journal.pcbi.1002687.
117. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK, et al. 2018. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol [Internet]. 36(3):272–281. doi:10.1038/nbt.4072.
118. Machado D, Andrejev S, Tramontano M, Patil KR. 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res [Internet]. 46(15):7542–7553. doi:10.1093/nar/gky537.
119. Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. 2019. Moving beyond P values: data analysis with estimation graphics. Nat Methods [Internet]. 16 (7):565–566. doi:10.1038/s41592-019-0470-3.
120. Kuhn M. 2008. Caret package. J Stat Softw. 28.
121. Grau J, Grosse I, Keilwagen J. 2015. PRROC: computing and visualizing Precision-recall and receiver operating characteristic curves in R. Bioinformatics. 31 (15):2595–2597. doi:10.1093/bioinformatics/btv153.
122. Xiong H, Pears C, Woollard A. 2017. An enhanced C. elegans based platform for toxicity assessment. Sci Rep [Internet]. 7(1):9839. doi:10.1038/s41598-017-10454-3.
123. Chawla DG, Shah RV, Barth ZK, Lee JD, Badecker KE, Naik A, Brewster MM, Salmon TP, Peel N. 2016. Caenorhabditis elegans glutamylating enzymes function redundantly in male mating. Biol Open [Internet]. 5:1290–1298. (9): doi:10.1242/bio.017442.

No 3. Metabolic Modeling Predicts Specific Gut Bacteria as Key Determinants for *Candida albicans* Colonization Levels

FORM 1

Manuscript No. 3

Manuscript title: Metabolic modeling predicts specific gut bacteria as key determinants for *Candida albicans* colonization levels

Authors: Mohammad H. Mirhakkak, Sascha Schäuble, Tilman E. Klassert, Sascha Brunke, Philipp Brandt, Daniel Loos, Ruben V. Uribe, Felipe Senne de Oliveira Lino, Yueqiong Ni, Slavena Vylkova, Hortense Slevogt, Bernhard Hube, Glen J. Weiss, Morten O. A. Sommer, Gianni Panagiotou

Bibliographic information: Mirhakkak, Mohammad H., et al. "Metabolic modeling predicts specific gut bacteria as key determinants for Candida albicans colonization levels." *The ISME journal* 15.5 (2021): 1257-1270.

The candidate is (Please tick the appropriate box.)

☐ First author, ☒ Co-first author, ☐ Corresponding author, ☐ Co-author.

Status: published

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Mirhakkak, M.H.	30%	40%		30%	
Schäuble, S.	30%	40%		30%	20%
Brandt, P.			35%		
Loos, D.		20%		10%	
Panagiotou, G.	40%			20%	40%
Others			65%	10%	40%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

FORM 2

Manuscript No. 3

Short reference Mirhakkak *et al*., *The ISME Journal* (2021)

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles): Reconstruction and analysis of *C. albicans* GSMM including single and pairwise simulations. Generation of ordinal regression model.

Figures # 1 and 2	☒	**100%**
Figure # 3	☒	Approximate contribution: 5**0%** (Figure parts C and D)

(Add more table boxes depending on the number of figures)

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

The ISME Journal (2021) 15:1257–1270
https://doi.org/10.1038/s41396-020-00848-z

ISME

ARTICLE

Metabolic modeling predicts specific gut bacteria as key determinants for *Candida albicans* colonization levels

Mohammad H. Mirhakkak[1] · Sascha Schäuble[1] · Tilman E. Klassert[2] · Sascha Brunke[3] · Philipp Brandt[4] · Daniel Loos[1] · Ruben V. Uribe[5] · Felipe Senne de Oliveira Lino[5] · Yueqiong Ni[1] · Slavena Vylkova[4] · Hortense Slevogt[2] · Bernhard Hube[3,6] · Glen J. Weiss[7] · Morten O. A. Sommer[5] · Gianni Panagiotou[1]

Received: 15 April 2020 / Revised: 6 November 2020 / Accepted: 18 November 2020 / Published online: 15 December 2020

Abstract

Candida albicans is a leading cause of life-threatening hospital-acquired infections and can lead to Candidemia with sepsis-like symptoms and high mortality rates. We reconstructed a genome-scale *C. albicans* metabolic model to investigate bacterial-fungal metabolic interactions in the gut as determinants of fungal abundance. We optimized the predictive capacity of our model using wild type and mutant *C. albicans* growth data and used it for in silico metabolic interaction predictions. Our analysis of more than 900 paired fungal–bacterial metabolic models predicted key gut bacterial species modulating *C. albicans* colonization levels. Among the studied microbes, *Alistipes putredinis* was predicted to negatively affect *C. albicans* levels. We confirmed these findings by metagenomic sequencing of stool samples from 24 human subjects and by fungal growth experiments in bacterial spent media. Furthermore, our pairwise simulations guided us to specific metabolites with promoting or inhibitory effect to the fungus when exposed in defined media under carbon and nitrogen limitation. Our study demonstrates that in silico metabolic prediction can lead to the identification of gut microbiome features that can significantly affect potentially harmful levels of *C. albicans*.

Highlights

- Genome-scale model reconstruction of *C. albicans* with 89% growth accuracy.
- Mutualism and parasitism are the most common predicted *C. albicans*-gut bacteria interactions.
- Metagenomic sequencing and in vitro assays reveal modulators of fungal growth.
- *Alistipes putredinis* potentially prevents elevated *C. albicans* levels.

Introduction

The fungus *Candida albicans* is found on the mucosal surfaces of at least 50–70% of healthy adults [1] and is a classic opportunistic pathogen. It resides as a harmless commensal but can become pathogenic in immunocompromised patients or under microbial dysbiosis [2, 3]. *C.*

These authors contributed equally: Mohammad H. Mirhakkak, Sascha Schäuble

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-020-00848-z) contains supplementary material, which is available to authorized users.

✉ Gianni Panagiotou
Gianni.Panagiotou@leibniz-hki.de

[1] Systems Biology & Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute, 07745 Jena, Germany

[2] ZIK Septomics, Jena University Hospital, 07745 Jena, Germany

[3] Department of Microbial Pathogenicity Mechanisms, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute, 07745 Jena, Germany

[4] Septomics Research Center, Friedrich Schiller University, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute, 07745 Jena, Germany

[5] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[6] Institute for Microbiology, Friedrich Schiller University, 07743 Jena, Germany

[7] MiRanostics Consulting, 85755 Oro Valley, AZ, USA

albicans causes up to 300,000 deaths per annum worldwide with an increasing number of individuals at risk [4]. Therefore, efforts to understand drivers of commensal or pathogenic outcomes of this fungus have intensified.

Recent studies found links between alterations in the composition and functionality of the gut microbiota and development of local or systemic *C. albicans* infections [5, 6]. A microbial tryptophan metabolic pathway appears to preserve immune physiology at mucosal surfaces by promoting indole-3-aldehyde production that contributes to IL-22 transcription [7]. Other gut microbial products such as bacteriocin are directly active against *C. albicans* [8]. A study of rectal samples from a cohort of 150 children linked gut microbiota to *Candida* prevalence, with a relative reduction in *Candida* species in children who received probiotics along with broad spectrum antibiotics [9]. Neither the dynamics of *Candida* species in the human gut nor the specific microbial contributors to the observed reduction have been studied, however. Based on these diverse findings, the commensal status of *C. albicans* appears to be related to the global taxonomy and functionality of the host microbiome.

A promising approach to analyzing interactions between *C. albicans* and gut microbial species uses genome-scale metabolic models (GSMMs) GSMMs have improved the biotechnological productivity of bacteria [10–12], revealed plant metabolic processes [13], and elucidated the Crabtree effect in yeast [14–16] and the Warburg effect in cancer cells [17]. Recent pioneering studies have developed high-quality GSMMs for gut bacteria that enable in silico analysis of gut metabolic functions and interactions [18, 19]. These resources have advanced the study of gut microbes and their respective pairwise interactions but have not yet been used to study interactions with opportunistic fungal pathogens such as *C. albicans*. The potential of gut microbes to influence the overall fitness of the fungus must be elucidated to support development of prophylactic or therapeutic strategies to control *C. albicans*.

We constructed a GSMM of *C. albicans*, starting with an automatically generated template model [20, 21]. We substantially improved its performance with manual curation and adaptation to phenotype microarray experiments. We used both, publicly available data [22] and new phenotypic microarray data for both wild type and mutant *C. albicans* strains. Our model predictions surpassed those of other GSMMs for species closely related to *C. albicans* that could serve as proxies for this fungus. We used the GSSM to simulate in silico pairwise metabolic interactions between *C. albicans* and each of 910 gut bacteria models. We challenged our predictions in vitro by growing *C. albicans* in carbon or nitrogen limited defined media in the presence of predicted fungal growth affecting metabolites. We further validated our results with stool samples from 24 human subjects, using metagenomics and internal transcribed spacer (ITS) sequencing to identify bacterial species associated with significant effects on *C. albicans* metabolism and growth. Finally, we assessed fungal growth in bacterial spent media experiments.

Materials and methods

Model reconstruction

To generate the *C. albicans* GSMM, we used the *C. albicans* metabolic model automatically reconstructed by the CoReCo pipeline as a template. In brief, CoReCo combines information from multiple data sources into a unified database and evaluates the probability of any reaction occurring in the target organism by computing a score for each enzyme based on sequence homology [20]. We refined the model in four consecutive steps (Fig. 1A) that included the identification and removal of duplicate metabolites, determination and resolve of erroneous energy-generating cycles [23], adaptation to phenotypic microarray data and exchange reaction modification based on flux variability analysis (FVA, Supplementary Material, Supplementary Data S1) [24].

Pairwise simulations

Pairwise simulations adapted from Heinken and Thiele [25] were performed using 818 AGORA 1.03 GSMMs [18] downloaded from (https://www.vmh.life) and 92 CarveMe gut bacterial GSMMs [19]. In brief, the *C. albicans* GSMM was paired to individual bacterial GSMMs and subsequently optimized by simultaneously maximizing *C. albicans* and bacteria biomass reactions. Interaction type was determined by taking the optimized growth rate in the pair compared to the growth rate of the individual GSMM into account (Supplementary Material).

To evaluate dissimilarities between promoters and inhibitors of *C. albicans* from different phyla, Bray-Curtis distances were calculated based on flux distributions of individual bacteria from pFBA simulations on different media [18] (https://www.vmh.life, Supplementary Table S1). Fisher's exact tests were calculated to determine if a specific phylum was enriched with species that either inhibited or promoted *C. albicans* growth. The python code and metabolic models for simulating pairwise GSMMs are available at https://github.com/mohammadmirhakkak/Candida-albicans-microbiome-interaction.

Phenotypic microarray experiments

C. albicans wild type and multiple mutant strains (Table 1) were pre-grown on YPD (1% yeast extract, 2% peptone,

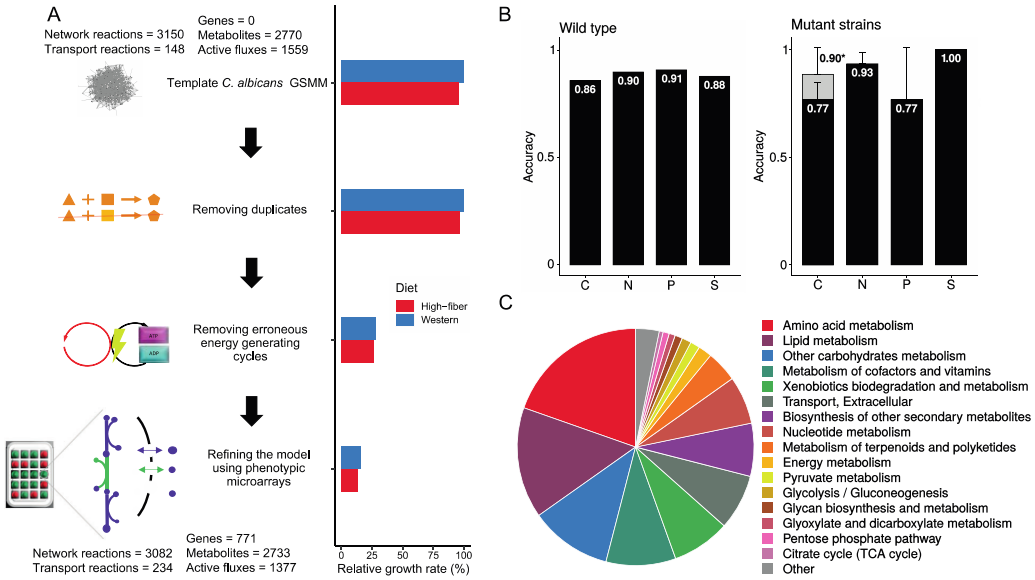


Fig. 1 *Candida albicans* GSMM reconstruction. A Based on a template [20] a manually curated metabolic model was achieved in several steps. Adjustment of model features such as modifying metabolic reactions or resolving energy cycles and model impact are indicated. Relative growth rates show relative differences to the template GSMM growth rate. **B** Benchmark results for model optimization using phenotypic microarray data for *C. albicans* growth. Light gray bar indicates accuracy on carbon media without arginine mutants that show growth on phenotypic microarray data without additional arginine (see main text for details). Accuracy was calculated as number of growth experiments that agree with model predictions across all growth experiments that were simulated. C carbon, N nitrogen, P phosphor, S sulfur. **C** Assigned pathway distribution of the final model.

Table 1 Selected *Candida albicans* strains for phenotypic microarrays experiments.

Strain	Genotype	Reference
SC5314	Prototroph	[65]
CEC2908	*ura3Δ::λimm434/ura3Δ::λimm434 his1Δ::hisG/HIS1 arg4Δ::hisG/ARG4 ADH1/adh1::pTDH3-carTA::SAT1*	[66]
SN87	*leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm434 IRO1/iro1Δ::imm434*	[67]
SN152	*leu2Δ/leu2Δ +LEU2, his1Δ/his1Δ +HIS1, arg4Δ/arg4Δ, URA3/ura3Δ::imm434, IRO1/iro1Δ::imm434*	[68]
JRC12	*arg1Δ::FRT/arg1Δ::FRT*	[69]
JRC38	*arg3Δ::FRT/arg3Δ::FRT*	[69]
CFG318	*NEUT5L/neut5l::FRT, put2Δ/put2Δ*	[47]

2% glucose, 2% agar) plates. Phenotype microarrays were performed by using microarray plates, reagents and devices according to the manufacturer Biolog, Inc. (Hayward, CA, USA) instructions. Briefly, *C. albicans* cells were taken directly from YPD agar plates and diluted in sterile dH2O to 62% transmittance as measured by a turbidimeter (Biolog, Inc.). Next, cells were combined with inoculating fluid IFY-0 base (1.2x), redox dye mix D (75x) (Biolog Inc.), and further supplemented with either Glucose, L-glutamic acid, potassium phosphate or sodium sulfate (Sigma-Aldrich) were required. 100 µl of the respective mixture (83.33% IFY-0 base, 1.33% redox dye D, 8.33% supplements, 2.08% cells, and if required 3.12% Glucose and dH2O) was added to each well of a Biolog Phenotype Microarray 96-well plate for fungi to test for metabolic activity in the presence of carbon sources (PM1 and PM2A), nitrogen sources (PM3B), and phosphorus and sulfur sources (PM4A). The phenotype microarrays were incubated at 30 °C in an OmniLog multiple plate reader in order to prevent hyphae formation that otherwise perturb growth measurements. Reduction of the redox dye, an indicator for metabolic activity, was measured kinetically every 15 min at an OD of 750 nm for 24–48 h. Each experiment was performed twice. Data analysis was done

using the R opm package for OmniLog phenotype microarray datasets [26].

In vitro experiments

Growth of C. albicans in presence of metabolites

To determine the effect of the metabolites on *C. albicans* growth, the clinical isolate SC5314 was grown overnight at 30 °C in YPD complex medium (1% yeast extract, 2% peptone, 2% glucose). 30 °C were chosen to prevent hyphae formation provoked at higher temperatures, which otherwise perturb growth curve measurements. Yeast cells were washed three times with sterile H_2O by centrifugation for 5 min at 4,200 × *g*. Test medium was composed of 1× yeast nitrogen base (YNB, Formedium) with either (standard) 0.25% NH_4SO_4/2% glucose, (C limited) 0.25% NH_4SO_4/0.25% glucose, (N limited) 0.008% NH_4SO_4/2% glucose, or (C/N low) 0.016% NH_4SO_4/0.5% glucose. Test substances were obtained from Sigma-Aldrich and were dissolved in H_2O at (nitrite) 156 mM, (desoxyadenosine) 100 mM, (sodium formate) 1 M, (putrescine) 500 mM, (L-asparagine) 100 mM, or (L-proline) 15.6 mM. Assays were performed as 1:2 dilution series in 96 well plates (TPP, flat bottom) and were composed of 180 µl test medium, 10 µl test substance, 10 µl yeast solution (1:10 dilution in H_2O, final OD_{600} of 0.1). Initial pH was verified to be at the expected ≈5.8. Growth was followed over at least 24 h using a Tecan infinite 200 multiwell plate reader set to 30 °C, with measurements at 600 nm every 15 min following 10 s of orbital shaking. All measurements were performed in triplicate from independent overnight cultures at different days. Growth was evaluated as area under the curve (AUC, trapezoidal method using GraphPad prism 8.1.2; baseline at mean of first three measurement) over 24 h and expressed as percent change compared to control setups (H_2O instead of test substance) in the same medium. AUCs were determined for three replicates, and mean change compared to controls is shown with standard deviations (SD) as error bars.

Strains and culture conditions

A. putredinis (DSM17216), *B. ovatus* (ATCC 8483), *B. vulgatus* (ATCC8482), *E. lenta* (DSM2243), *P. copri* (DSM18205), *R. torques* (ATCC27756), *E. coli* (MG1655 and BAA-1161), *P. corporis* (DSM18810), *C. albicans* (SC5314/ATCC MYA-2876), *C. albicans* (ATCC 10231) and *C. albicans* (ATCC 18804) were grown at 37 °C under anaerobic conditions (gas mixture, 95% N_2 and 5% H_2) in prereduced modified GAM (mGAM, Nissui Pharmaceutical Co. Ltd.) broth for liquid cultures or broth supplemented with agar for growth on plates. 37 °C and anaerobic conditions were chosen to resemble best the natural environment of these gut bacteria.

Sterile bacterial spent media

Bacterial strains were grown for 48 h in GAM broth, then subcultured 1:50 in fresh GAM broth and grown for 48 h in anaerobic conditions at 37 °C to resemble the gut environment. Cultures were centrifuged at 11,000 × *g* for 5 min and spent media removed without disturbing the pellet. Spent media were passed through a 0.2 µm syringe filter to remove remaining bacteria. After filtration, the pH of the spent media was analyzed using an electronic pH-meter (Supplementary Table S2). 1% (v/v) of Phosphate-buffered saline (PBS) was added in each experiment to maintain a constant pH.

Growth assays

An overnight culture of *C. albicans* was grown aerobically in mGAM media at 37 °C. Aerobic conditions were chosen to enable sufficient growth of *C. albicans*. Cells were then subcultured at a 1:1000 dilution into 150 µl of sterile spent bacterial media in different proportions: 75 and 100%. The spent media were diluted in fresh 25% mGAM broth for 75% spent media proportion and PBS (1% v/v), accordingly. The fermentation was performed in flat-bottom, 96-well plates. The plates were incubated for 24 h at 37 °C, with continuous orbital shaking at 900 rpm. 37 °C did not induce notable hyphae formation in growth assays and therefore did not perturb growth measurements on spent media. Cell densities were measured every 10 min at optical density 600 nm (OD600) using a microtiter reader (BioTek ELx800). Growth rates were calculated by plotting log of OD measurements in log phase and calculating the slope of the time points in log phase where r^2 was closest to 1, using at least 12 time points over 2 h. Growth inhibition was determined as the growth rate of *C. albicans* in spent bacterial media normalized to the growth rate in the corresponding mGAM fresh media dilution.

Microbiome profiling

Bacterial and fungal species profiles were generated for stool samples from a human cohort of 24 individuals at Western Regional Medical Center, Goodyear, Arizona, USA. Samples were collected after signed informed consent under a protocol approved by the Western Institutional Review Board (WIRB protocol number 20140271, Pallyup, Washington, USA). All subjects had been diagnosed with different types of cancer, with heterogeneous stage, treatment, and histological findings. Metagenomic sequencing

was performed at BGI, Hong Kong S.A.R., China, as described in Qin et al. [27], and ITS amplicon sequencing of the mycobiome was performed at ZIK Septomics, University Hospital Jena, Thuringia, Germany (extended details in Supplementary Material). Bacterial reads were obtained using 150-bp Illumina PE whole metagenome sequencing. Species profiling was by MetaPhlAn 2.7.6 [28]. after applying an in-house pipeline for quality control [29] and removal of human reads. Fungal reads were obtained using 250-bp Illumina PE ITS1 amplicon sequencing (extended details in Supplementary Material). The DADA2 ITS Pipeline Workflow 1.8 was followed for amplicon sequence variants [30]. Mothur classifier [31] called from QIIME 1.9.1 [32] and the UNITE database 7.2 [33] were used for fungal taxonomic assignments. Bacterial species were considered for further analysis, if a GSMM model was available, the species was part of the Open Tree of Life 10.4 [34] and it was prevalent in at least 4 of 26 samples. Correlations of bacterial abundance and growth rates to *C. albicans* abundance were tested using two-sided Spearman correlation ($p < 0.05$). The bacterial and fungal profiles are available under the ENA Study Nr. PRJEB33756.

Partial spearman correlation was computed within R using the package PResiduals (v0.2–6).

Ordinal regression model

Relative abundance values for *C. albicans* from human samples were grouped into two sets determined by a 20% relative abundance threshold. To predict *C. albicans* abundance levels, three binomial ordinal regression models were generated using as independent variables the bacterial abundance values, predicted interaction coefficients, and products of multiplying abundance values by predicted interaction coefficients. For model building, we either selected five bacterial species that significantly correlated with *C. albicans* by relative abundance and two that significantly correlated by GRiD value, or selected bacteria based on whether they increase performance as described in the main text. The predictive power of each model was assessed by determining the true/false positive/negative and accuracy values and by analyzing receiving operating characteristic curves using the R package ROCR (ver. 1.0–7).

Results and discussion

Reconstruction of a *Candida albicans* genome-scale metabolic model

To develop a *C. albicans* GSMM, we started with a model automatically generated by the Comparative ReConstruction (CoReCo) pipeline [20]. The initial *C. albicans* CoReCo model (BioModels ID 1604280052) comprised 2770 metabolites and 3298 reactions, of which 3150 were network and 148 transport reactions (Fig. 1A). In addition to a unique reaction set, the initial CoReCo GSMM for *C. albicans* contained multiple nontrivial duplicate reactions and metabolites. Typically, these involved marginally differing metabolite names such as *L-Glutamate* and *Glutamate* that had not been automatically detected by the CoReCo platform (Fig. 1A, Supplementary Data S1, Tables S1, S3, and S4).

We also curated energy-generating cycles (EGCs) that created energy compounds such as ATP without requiring nutrient uptake [23]. We resolved these infeasible EGCs by identifying and correcting implausible reaction directionalities using metabolic pathway databases such as BioCyc (Supplementary Table S4). For example, in our initial *C. albicans* model, we found an ATP-producing EGC that involved phosphate rather than pyrophosphate as indicated by BRENDA, Biocyc and KEGG databases. Correcting the involved reaction acetoacetate:CoA ligase resolved this particular EGC and additional EGCs while maintaining a viable biomass flux. We corrected six reactions by either changing metabolite usage or reaction directionality based on KEGG or BioCyc information (Supplementary Table S4). Resolving EGCs also reduced the flux through the biomass reaction towards 1.4, which is closer to the biomass reaction flux for other fungal models such as the yeast consensus model [35, 36] (Fig. 1A).

Next, we adapted our model to multiple phenotypic microarray growth experiments of *C. albicans*. These comprise up to 1440 different defined media experiments with diverse carbon, nitrogen, phosphorus or sulfate sources. We used a publicly available dataset with different *C. albicans* phenotypes [22] and created additional phenotypic microarray growth experiments including several mutant strains for e.g., different arginine biosynthesis steps (see Methods for details). Of the metabolites in the dataset, 455 mapped to metabolites in our *C. albicans* GSMM. First, we adapted our model to growth experiments that were in agreement between data from Ene et al. [22] and our own prototrophic *C. albicans* wild type strain SC5314 dataset for *C. albicans*. Of note, using only compatible growth measurements ensured robust growth information across different temperatures applied in our and in the published data by Ene et al. [22] (cf. Materials and Methods). Second, we adapted our model to all phenotypic growth experimental data of *C. albicans* mutant strains for those growth conditions that yielded the same results for both our wild type and the Ene et al. dataset in the prior step (Supplementary Table S4). These refinement steps included, for example, enabling ammonia production from urea or via lysine degradation, which was not initially present in our model

(Supplementary Tables S4 and S5). Based on the growth data, we added 107 potentially feasible exchange reactions involving metabolites associated with *C. albicans* growth. By applying flux variability analysis [24] we also identified five exchange reactions that shuttle 4-aminobenzoate, folinic acid, 7,8-diaminononanoate, hexadecanal, and hydrogensulfite in and out of the network, but do not support growth and removed these from the GSMM accordingly. Overall, we achieved a high compatibility between our model predictions and the phenotypic microarray growth experiments with nitrogen and sulfur predictions reaching above 90% accuracy over all mutant experiments (Fig. 1B, Supplementary Table S4). Of note, model predictions for carbon source experiments of three different *C. albicans* mutants for arginine biosynthesis (*arg1*Δ, *arg3*Δ and *arg4*Δ) predicted no growth rate for all carbon sources, since arginine is essential for the biomass objective function. Surprisingly, up to 33% of the associated phenotypic microarray data showed fungal growth, despite the inability to synthesize arginine due to the knock out. This might be due to recycling of available proteins in e.g., fungal vacuoles, as the defined growth media itself does not contain arginine, unless specifically tested. Considering only growth experiments that show no growth for arginine biosynthesis mutant strains our model accuracy for carbon sources reaches 90% across all tested *C. albicans* mutant (Fig. 1B).

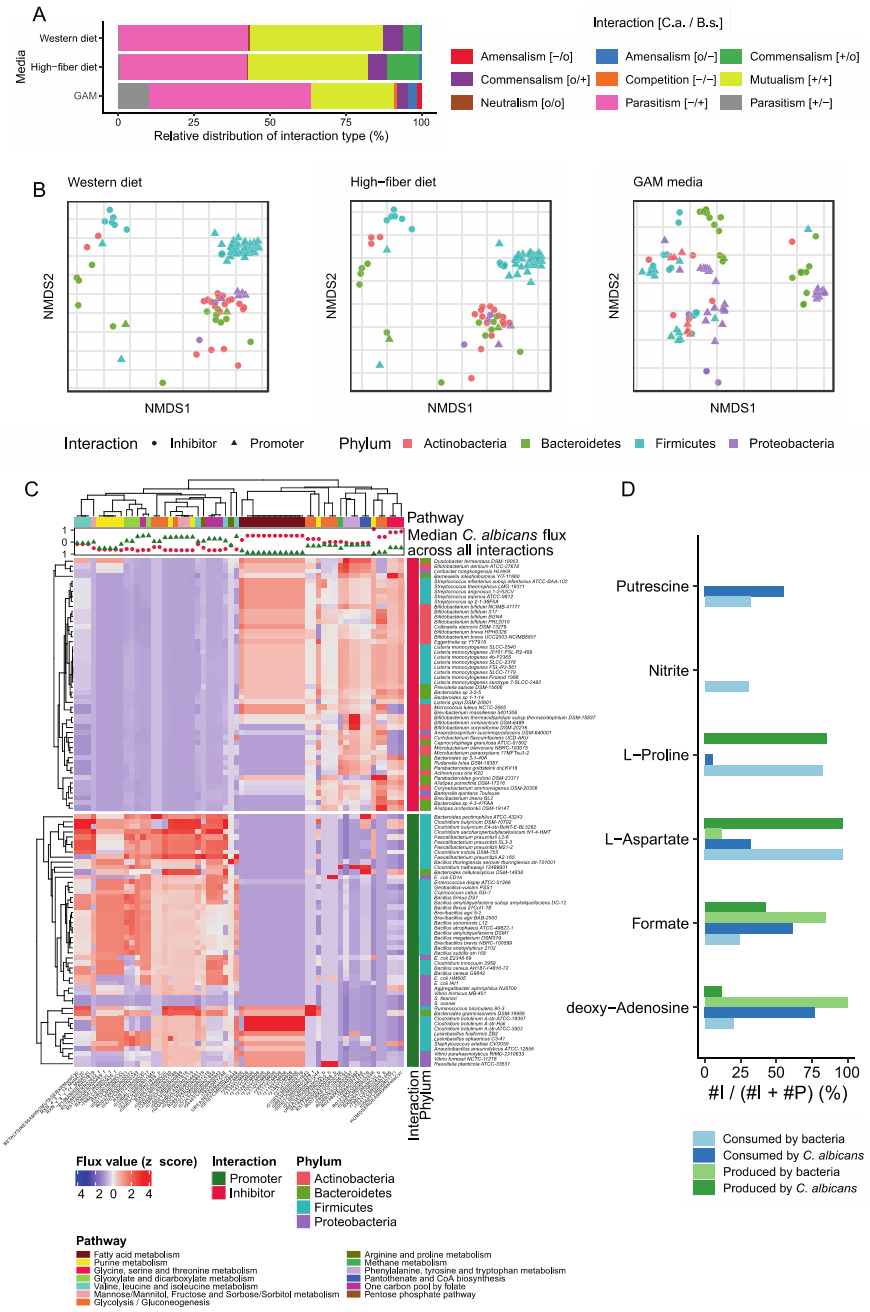
Finally, we added gene annotation for ~1500 reactions and associated individual gene to reaction rules using the KEGG database. We also added pathway association for all reactions if available and unified pathway associations across resources. This step resulted in 83 pathway definitions, including all essential pathways such as in central carbon, amino acid, and lipid metabolism (Fig. 1C). Overall, our model refinements resulted in an addition of 771 genes, led to a final model comprising 3082 metabolic reactions (−68 compared to draft model) and 2733 metabolites (−37 metabolites compared to draft model) and a reduced active flux flexibility towards biomass by 11.7% (Fig. 1A, Supplementary Data S1, Tables S1, S3, and S4).

Pairwise growth simulations predict gut bacteria modulating essential *C. albicans* metabolic activity

Next, we generated in silico metabolic interaction predictions about *C. albicans* coupled to gut bacteria GSMM models. Using 910 publicly available GSMMs for gut microbial species [18, 19], we performed pairwise metabolic analysis by linking our *C. albicans* model with each gut microbe GSMM [25] (Supplementary Table S1). The majority of GSMMs from different sources [18, 19] gave compatible growth rate predictions. To streamline further analysis, we continued with the assembly of gut organisms through reconstruction and analysis (AGORA) GSMMs (https://www.vmh.life) [18] unless models were available from both sources. In the latter case, we continued with the CarveMe model versions, since these are refined to bacterial growth data across 19 different media including extended pathway gap correction [37] and are based on a manually curated template model [19].

We simulated growth on two different media compositions that resemble typical Western and high-fiber diets (https://www.vmh.life) [18]. We identified the interaction type of each *C. albicans*-gut microbe pair by analyzing differences in predicted growth rates compared to growth rates derived from individual simulations using a flux balance analysis approach [38] (Methods, Pairwise simulations; Supplementary Material, Table S1). Predictions based on Western and high-fiber diets gave similar interaction-type distributions: mutualism (positive growth effect for both, *C. albicans* and paired bacteria) and parasitism (here, negative effect on *C. albicans* growth, positive growth effect on bacteria) were the most abundant (>81% of observed interaction types, Fig. 2A). Other interaction predictions included commensalism, in which both *C. albicans* and bacterial growth was promoted without negative effects on the respective paired microorganism (12% and 17% for Western and high-fiber diets, respectively). Only a few examples of parasitism in which *C. albicans* exerted negative effects on gut bacteria or amensalism (no growth effect on *C. albicans*, negative effect on bacteria) or neutralism (no growth effect on both, *C. albicans* and gut bacteria) were observed. We further examined prediction accuracy with additional simulations on standard Gifu anaerobic medium (GAM), which was used for in vitro validation experiments (Supplementary Table S6). Growth simulations on GAM were feasible for only 200 GSMM models, as all other models yielded no biomass formation. GAM simulations identified parasitism with negative effects on bacteria for up to 16% of viable bacteria models, while parasitism with negative effects on *C. albicans* (53%) and mutualism (28%) were again the most commonly occurring interaction types (Fig. 2A). Our interaction-type analysis based on predicting growth rate differences for microorganisms in individual and paired model setups hints that the majority of gut bacteria have either a mutualistic or a parasitic relationship with *C. albicans* irrespective of medium.

To analyze if predicted interactions that promoted or inhibited *C. albicans* growth were phylum- or diet-specific effects, we performed non-metric multidimensional scaling analysis (NMDS) based on parsimonious flux balance analysis (pFBA)-derived values for the top 50 promoting and inhibiting bacterial species [39] (Supplementary Table S1). The predicted top 50 inhibiting bacterial species comprised mainly species from the genus Bifidobacterium



and Listeria, but also included several bacteria from the Bacteriodetes phylum including e.g., *Alistipes putredinis*. In contrast, the predicted top 50 *C. albicans* promoting bacteria included many bacteria with the Bacillus, Bacteroides, Clostridium or Vibrio genus (Supplementary Table S1). After removal of a few outliers (up to seven GSMMs comprising several *Escherichia coli* strains, among others) we observed distinct grouping of the Firmicutes phylum on

◀ **Fig. 2 Pairwise in silico interaction experiments. A** Distribution of interaction type for *C. albicans* (C.a.) and bacterial species (B.s.). Interactions have positive (+), negative (−) or no (o) effect on growth rates of fungus or bacteria as indicated for interaction types. **B** Non-metric multidimensional scaling (NMDS) plots of bacterial reaction flux rates for top 50 *C. albicans*-inhibiting and -promoting bacteria simulated for three different media (Western and high-fiber diet, Gifu anaerobic media (GAM)). **C** Metabolic reactions of *C. albicans* with the most substantially differing flux rates of *C. albicans* when paired with top 50 inhibiting or promoting bacteria. Top: median *C. albicans* flux rate differences across all bacterial species paired with *C. albicans*. **D** Analysis for selected metabolites based on exchange reaction fluxes of simulated fungal–bacterial pairs for top 50 promoting or inhibiting bacterial species (cf. Supplementary Table S5). *x*-axis indicates the percentage of exchange reaction fluxes with bacteria that inhibit *C. albicans* growth.

both Western and high-fiber diets in the NDMS plots, with a largely predicted positive effect on *C. albicans* growth (Fig. 2B). The majority of Proteobacteria (for 71% of the respective GSMMs on Western, and 73% on high-fiber diet) were predicted to show a positive effect on *C. albicans* growth as well and show similar flux distributions to some Bacteroidetes and most Actinobacteria. In contrast, the predicted impact of Bacteroidetes was primarily negative (for 87% GSSMs on Western, and 86% on high-fiber diet) on *C. albicans* growth with differing flux distributions for both, Western and high-fiber diet (Fig. 2B). All Actinobacteria were predicted to inhibit *C. albicans* growth as well and showed similar flux distributions to some Bacteroidetes and most Proteobacteria for Western and high-fiber diets. Of note, in silico simulations on GAM revealed notable differences of flux distributions compared to Western or high-fiber diet (e.g., 53% of the paired Actinobacteria showed promoting effects on *C. albicans*) showing that GAM based simulations differ to some degree from these diets (Fig. 2B). The predicted interaction type was significantly dependent on phyla (chi-squared test, p value = 2.5e−9 on Western diet) with Actinobacteria (false discovery rate-corrected Fisher's exact test p value = 7.0e−7) and Firmicutes (false discovery rate-corrected Fisher's exact test p value = 1.5e−5) primarily responsible for the interaction type. Altogether, our analyses suggested that under Western and high-fiber diets, species from specific phyla had positive (Firmicutes, Proteobacteria), or negative (Actinobacteria, Bacteriodetes) effects on *C. albicans* growth with distinct flux distributions particularly for predicted growth-promoting Firmicutes species.

Next, we investigated *C. albicans*-specific reaction fluxes derived from paired fungal–bacterial model simulations. Paired bacteria were again selected based on their predicted impact on *C. albicans* growth rates such that its difference between paired fungal–bacterial and individual model simulations is most pronounced and comprised bacteria from four different phyla (Fig. 2C, Supplementary Fig. S1). From the obtained flux distributions *C. albicans* reactions were selected based on the most pronounced flux differences between *C. albicans* paired with predicted growth inhibiting and growth-promoting bacterial GSMMs (Methods, Pairwise simulations). Of note, predicted median flux differed notably for many reactions comparing the top inhibiting and promoting bacteria to all paired gut microbes (Fig. 2C, upper panel). These changes were particularly present in sugar, fatty acid, folate, and small amino acid associated pathways, but also in glycolysis and hint towards a shifted flux in the selected bacteria that dominantly affect *C. albicans* growth. Altered reaction fluxes were identified across major pathways including carbon, amino acid, purine and fatty acid metabolism. Central metabolites such as alpha-ketoglutarate, pyruvate and glutamate were balanced towards net production or consumption, depending on the paired gut microbe in the in silico simulation. *C. albicans* growth is affected by modulating carbohydrate metabolism [40] or the availability of amino acids such as leucine or valine [41] and might be used by gut microbiota to prevent or promote *C. albicans* growth. Specifically, we predicted elevated reaction fluxes that consumed L-glutamate in *C. albicans* when paired with fungal growth-promoting bacteria via the aminotransferases Glycine:2-oxoglutarate aminotransferase and branched-chain amino acid aminotransferase. Aminotransferases were studied before in the context of Candida infection and were attributed to the nutritional versatility of Candida species [42, 43]. Moreover, glutathione synthesized from L-glutamate is important in counteracting oxidative stress [44]. Predicted reaction fluxes including amino acids were notably different among top promotors and inhibitors and may serve as potential targets for identifying antifungal agents [45]. These findings suggest that gut bacteria that potentially perturb *C. albicans* growth may cause metabolic shifts in *C. albicans* towards L-glutamate-promoting reactions, which may allow the fungus to evade harmful oxidative stress levels.

Next, we focused on investigating fluxes of metabolite exchange reactions. Exchange reactions allow to shuttle metabolites in and out of a joint compartment in our paired in silico models. These joint compartments serve as a connection between the respective bacterial and the fungus model and allow to predict potentially *C. albicans* growth rate influencing metabolites (Fig. 2D, Supplementary Table S7). To identify such metabolites we specifically filtered for exchange reaction fluxes of metabolites that are primarily present for either *C. albicans* growth inhibiting or promoting bacteria as derived from our paired fungal–bacterial model simulations. Again, many amino acids like L-proline, or L-aspartate, but also other factors such as nitrite or putrescine are predicted to be notably differentially consumed by fungus or bacteria.

In vitro experiments and metagenomics analyses support metabolic dependencies of *C. albicans*

To test the quality of our in silico analysis, we investigated *C. albicans* growth in the presence of metabolites, performed metagenomic sequencing for 24 individuals and assessed *C. albicans* growth in bacterial spent media.

First, we grew *C. albicans* in the presence of metabolites and investigated the growth-promoting or -inhibiting effect of these metabolites under different carbon and/or nitrogen availabilities (Fig. 3A). We selected six metabolites that were either primarily consumed or produced by either fungus or bacteria in our paired metabolic in silico simulations (Fig. 2D, Supplementary Table S7). We hypothesized that metabolites that e.g., are predicted to be consumed by fungal growth inhibiting bacteria are either withheld from the fungus or cannot be metabolized by the fungus and are thus beneficial for the bacteria to outgrow *C. albicans*. Likewise, metabolites that are predicted to be produced by bacteria or consumed by the fungus when *C. albicans* is paired with growth inhibiting bacteria might provide clues of metabolites with a potential negative effect on *C. albicans* growth. Nitrite showed severe growth inhibiting effects on *C. albicans*, irrespective of available carbon and nitrogen source concentrations. The same inhibiting influence was observed to a lesser extent for putrescine and for the formic acid salt, sodium formate. Both putrescine and formic acid were also predicted to be primarily consumed by *C. albicans* when paired with fungal growth inhibiting bacteria. In contrast, L-Aspartate and

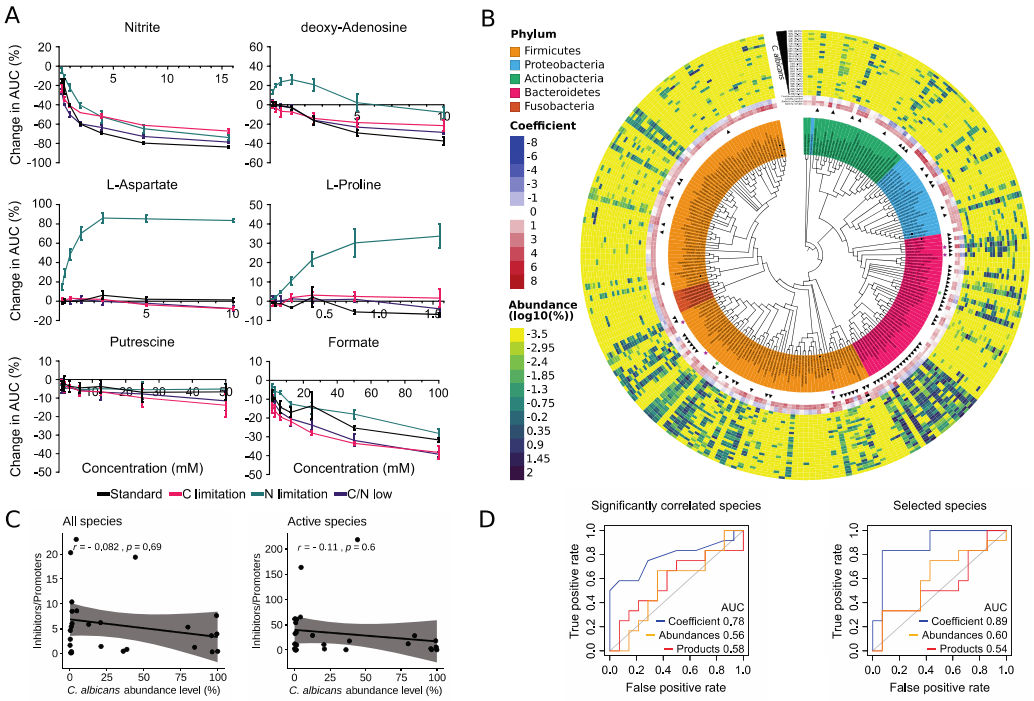


Fig. 3 Experimental and clinical data supporting in silico predictions. A Area under the curve (AUC) measurements for fungal growth in presence of selected metabolites in a series of concentration dilution experiments. AUCs were determined for three replicates. Mean change compared to medium-only controls is shown with standard deviations (SD) as error bars. **B** Bacterial abundance and growth rates were obtained using MetaPhlAn2 and GRiD 1.2, respectively. Modeled species were arranged according to the Open Tree of Life 10.4 [34]. Annotation rings from inner to outer: Significant correlations between *C. albicans* abundance and bacterial abundance (magenta stars) or bacterial GRiD score (green stars, Spearman's coefficient, $p < 0.05$); species with GRiD score greater than 1 in at least one sample (black triangles); in silico interaction coefficients from GSMM analysis (blue to red); sample bacterial abundance ($N = 26$, yellow to purple) sorted by *C. albicans* abundance (highest abundance is outermost ring). **C** Regression performance for all (left panel) or active (GRiD value > 1 in at least one patient, right panel) species. Each dot is a ratio of inhibitors to promoters for a patient. Values for inhibitors and promoters were calculated by summing products of bacterial abundance x in silico coefficient for *C. albicans*. **D** Area under the curve (AUC) of ordinal regression analysis for seven species with significant correlations with *C. albicans* abundance or GRiD values as shown in **B** (left panel). AUC of ordinal regression analysis for 29 selected species (see main text). The regression model performance was achieved by using GSMM based metabolic coefficients (Coefficients), bacterial relative abundance (Abundances) or the product of both (Products).

L-Proline showed concentration-dependent fungal growth-promoting effects specifically under nitrogen limitation. For both metabolites, we predicted a net consumption by *C. albicans* growth inhibiting bacteria. It is noteworthy that, while we did not test for morphological changes, proline and putrescine are known inducers of hyphae formation in *C. albicans*, which may therefore be influenced by bacterial production and consumption of these metabolites [46–48]. Interestingly, desoxy-adenosine showed a growth-promoting effect at low concentrations, most pronounced under nitrogen limitation, whereas higher metabolite concentrations had negative effects on *C. albicans* growth. In our in silico model we predicted its production by *C. albicans* growth inhibting bacteria. A bacterial production of desoxy-adenosine may therefore lead to sufficiently high concentration levels of this metabolite that might restrict growth of the fungus. In summary, our in silico analysis of metabolite exchanges between fungus and bacteria and the shown experimental data provide a concept for predicting and testing potentially fungal growth modulating metabolites.

Next, we investigated whether metabolic interactions are the main driver of the observed abundance-based associations of gut bacteria and *C. albicans* by analyzing stool samples from a cohort of 24 cancer patients (Supplementary Table S8). We assessed the structure of the gut microbiome in samples via shotgun metagenomic sequencing, generating 118.1 Gbp of sequencing data with an average of 2 × 26.2 million paired-end reads per sample. Taxonomic profiling revealed that Bacteroidetes (54.3%) and Firmicutes (36.3%) were the most abundant phyla, followed by Proteobacteria (5.83%), Verrucomicrobia (1.46%) and Actinobacteria (1.34%). From the 400 bacterial species identified in our samples, we retrieved GSMMs for 247 (Fig. 3B, Supplementary Fig. S2). The mean relative abundance of *C. albicans* within the fungal community as revealed by ITS1 amplicon sequencing of all cancer patient samples was 39.7% over all patient samples, while individual samples covered a broad relative abundance range from 0.1% to 99.8% relative *C. albicans* abundance. Only five bacterial species had a relative abundance with a significant correlation by Spearman's coefficient to *C. albicans* relative abundance across all patients (*Clostridium scindens* $\rho = 0.45$, *Hungatella hathewayi* $\rho = 0.55$, *Flavonifacfractor plautii* $\rho = 0.46$, *Barnesiella intestinihominis* $\rho = -0.47$ and *Alistipes putredinis* $\rho = -0.42$, Fig. 3B). Although the influence of external factors such as gender, age or therapy type cannot be completely ruled out, this suggests that, at least in terms of relative abundance compared to other fungal species, a limited set of gut microbes might influence *C. albicans* prevalence levels more than the remaining gut microbiome constituting bacteria. Of note, the directionality of the observed significant correlation was in accordance with the predicted paired-growth relationship by the GSMMs for all five bacteria (Supplementary Table S1). We also calculated Growth Rate InDex (GRiD) [49] to estimate in situ growth rates for bacteria species and analyze correlation to *C. albicans* relative abundance. We observed a finite GRiD value for 98 of the species in Fig. 3B (cf. Supplementary Table S9). GRiD values for two bacteria significantly correlated with *C. albicans* abundance (positive for *Roseburia inulinivorans* and negative for *Parabacteroides distasonis*, Fig. 3B) in the same direction as predicted by our GSMMs.

We next investigated to what extent fungal growth is influenced by spent media of selected bacterial strains for which we established cultivation protocols in our lab (Methods). Among the species selected for in vitro spent media experiments, *A. putredinis* had the strongest inhibitory effect in Western-diet conditions in in silico predictions (Supplementary Table S1). This species was also found significantly negatively correlated with *C. albicans* abundance in the human gut (Fig. 3B). Of note, adjusting for the external factors gender, age, ethnicity or immunotherapy application in our metagenomic dataset, affected this significant negative correlation only slightly (Supplementary Table S8). The addition of spent media from *A. putredinis* to *C. albicans* cultures (cf. Methods for details) resulted in up to 23% growth inhibition of the fungus (Supplementary Table S2). Interestingly, both butanoic and propanoic acid, two short chain fatty acids (SCFAs) with fungistatic properties produced by the gut microbiome [50, 51] were found with slight elevated concentrations in the spent media of *A. putredinis* (202 µM and 116 µM compared to 148 µM and 88 µM in modified GAM for propanoic and butanoic acid, respectively, Supplementary Table S2). *A. putredinis* is a reported producer of propionate presumably next to succinate [51]. Of note, for pairwise in silico simulations over all possible combinations of *A. putredinis* and any bacteria that are present in our human samples, we predicted that in 63% of the paired models *A. putredinis* can secrete propanoic acid. These data indicate that to a certain extend *A. putredinis* might contribute to global propanoic but also butanoic acid levels, two health promoting SCFAs [52]. In addition, *A. putredinis* is dominant in the fecal microbiota of healthy humans [53] and important for the maintenance of a healthy intestinal barrier [54]. Generally, the genus Alistipes shows disease protective effects against a number of diseases including fibrosis, colitis, cancer or cardiovascular disease [55]. Our results hint that *A. putredinis* is also potentially able to prevent elevated *C. albicans* levels, and we suggest more in-depth studies of *A. putredinis* in conjunction with *C. albicans*. To further examine whether the observed inhibition was a methodological artifact we performed additional experiments with species showing either positive or negative interaction with *C. albicans*. For example, our

in silico predictions show that *Ruminococcus torques* contributes positively to *C. albicans* growth, while also a non-significant positive correlation between this species and *C. albicans* levels was observed from the metagenomics data (Fig. 3B, Supplementary Table S8). Indeed, in our spent media experiment we also observed a positive effect of *Ruminococcus torques* on *C. albicans* growth (15% for 100% spent media, 12% for 75% spent media, Supplementary Table S2). Other strains showed varied levels of *C. albicans* growth inhibition which is in agreement with our metagenomic and spent media data except for *Escherichia coli*. For *E. coli* we observe a disagreement between a predicted positive growth effect on *C. albicans*, a low positive correlation in the metagenomics data, and a negative effect in the spent media experiments. Interestingly, *E. coli* was found to produce a soluble fungicidal factor, which cannot be captured by our metabolic model and might explain the discordance to our spent media results [56].

To further support the idea that metabolic interactions with a few bacterial species might be sufficient to determine *C. albicans* colonization levels, we investigated, by ordinal regression analysis, if the ratio of abundance x growth coefficients for all bacterial inhibitors vs. promoters correlated with *C. albicans* levels (Methods, Ordinal regression model). No significant correlation was seen when considering all species (Fig. 3C, left), indicating that the use of all species as predictors does not allow for a good model to predict *C. albicans* levels. Of note, the ratio of inhibitors to promoters calculated using only species that were active in at least one patient sample according to GRiD was not significantly correlated to *C. albicans* abundance levels as well (Fig. 3C, right). The critical role of metabolic interactions between the limited set of 7 significantly correlating bacterial species identified above to *C. albicans* relative abundances was evident in the ordinal regression model we developed using the interaction growth coefficients and relative abundances of the bacteria. We investigated if GSMM computed growth coefficients, relative bacterial abundance, or both were good predictors for *C. albicans* levels (Method, Ordinal regression models). We obtained the highest performance of 0.78 Area Under the Receiver Operating Characteristics (AUROC) using growth prediction coefficients from our GSMM computation (Fig. 3D, left). Using only relative bacteria abundance resulted in 0.56 AUROC, whereas a model using both interaction coefficients and growth data resulted in 0.58 AUROC. In addition to using bacteria that show significant correlations to *C. albicans* relative abundances as model features, we filtered candidate bacteria by discarding first the bacteria prevalent in at least 10%, but not more than 90% of the samples. This reduces the number of feature candidates to 121 bacteria. We selected next 10 subgroups of our samples by ignoring ~10% samples in each subset, such that each sample was once not part of the subset. For each subset we started with one bacterium and assessed model performance consecutively by adding further bacteria until all bacteria were included as features. Bacteria that caused a drop in model accuracy or regression slope were discarded, followed by another iteration of the model performance evaluation. These steps were repeated until no performance-impairing bacteria were left in the feature list. The union of the gained feature candidates across all 10 subsets resulted in a feature list of 57 gut bacteria, which included, among others, *A. putridinis* (Supplementary Table S10). We evaluated the regression model with these bacteria and obtained 0.89 AUROC using only growth prediction coefficients from our GSMM computation (Fig. 3D, right). Using only relative bacteria abundance or both, interaction coefficients and growth data, resulted again in a performance drop down to 0.60 and 0.54 AUROC, respectively. These results demonstrate that GSMM analysis based interaction coefficients could be used as predictors of ordinally scaled *C. albicans* levels. Our results indicate that the intrinsic metabolic interactivity between a fungus and bacteria contains valuable information about the performance of classification models. This information should be more extensively applied in future classification studies.

In conclusion we expanded the concept of using in silico metabolic interaction calculations to accurately predict pairwise beneficial or detrimental effects on co-existing organisms [18, 25, 57] to bacterial-fungal interaction predictions based on studies suggesting that key gut species might determine beneficial outcomes in patients with a range of diseases [5, 6, 9]. Selected metabolite experiments and shotgun metagenomics sequencing back our in silico modeling concept based on pairwise metabolic interaction simulations. Further studies of our predictions for major metabolic pathways, for example for carbon compounds and amino acids, may elucidate the specific mechanisms of these influences. Selected metabolite measurements in defined media could be used further for accurate predictions of potential metabolite candidates that are preferentially used by e.g., gut microbes over *C. albicans* and can hint towards bacterial species that specifically secrete *C. albicans*-inhibiting metabolites. Taken together, our findings support that specific gut bacteria influence gut colonization by *C. albicans*. Moreover, our analysis indicates that it may be possible to design synthetic communities with only a few bacterial species that could then influence essential metabolic activities of *C. albicans* and prevent fungal overgrowth. Further refinements of our model including compartmentalization complemented by comprehensive single-knockout studies of bacteria or the fungus may further improve the predictive capacity. Also additional diets and growth media compositions beyond the three used in this study may be tested, since e.g., GAM compositions

may vary, therefore influence in silico predictions and in general, might not reflect in vivo conditions as well as high fiber, western or other common human diets. Despite pairwise interactions were shown to be key drivers of the dynamics of microbial communities [58], additional in silico simulations of multiple interactions with e.g., the recently published MAMBO algorithm [59] need to be addressed, to potentially extend our understanding of the intricate relationship between *C. albicans* and the gut microbiota and its effect on *C. albicans* levels. Finally, tools that incorporate spatial information [60, 61] could determine the impact of niche colonization by gut fungal and bacterial species. In the present study we specifically focused on the intricate relationship between (gut) bacteria and *C. albicans* to elucidate their relationship independent of host factors in order to keep free parameters in a feasible range. Host factors are key modulators of fungal–bacterial or fungal–bacterial–host interactions as could be shown in other studies [3, 62–64]. Though adding considerable complexity to the setup, an extension of our conceptual approach to study specifically metabolic modes of fungal–bacterial interaction with host factors might play an important complementary role as long as the predictive capacity can be supported by sufficient data around the triangle host, bacteria and fungus. In summary, our strategy of studying fungal–bacterial relationships in the gut using an in silico, metabolism-driven approach already yielded promising results. Our approach adds another useful layer of in silico predictions that can contribute to stratify identification of potentially clinically relevant gut bacteria in face of the steadily growing amount of high throughput data. Ultimately, including metabolic in silico analysis could promote additional systems-biology and systems-medicine studies that focus on fungal infections and their often lethal implications to humans.

Acknowledgements GP, MM, SV and PB would like to thank Deutsche Forschungsgemeinschaft (DFG) CRC/Transregio 124 "Pathogenic fungi and their human host: Networks of interaction", subproject B5 and C2 (DFG project number 210879364). RVU, FSdOL acknowledge funding from The Novo Nordisk Foundation under NFF grant number: NNF10CC1016517 and The Novo Nordisk Foundation, Challenge programme, CaMiT under grant agreement: NNF17CO0028232. We also acknowledge funding by the German Ministry for Education and Science in the program Unternehmen Region (BMBF 03Z22JN11). We thank the patients for providing their stool samples and Chris Tachibana for language revision of our manuscript.

Compliance with ethical standards

Conflict of interest GJW is an employee of SOTIO and former employee of Unum Therapeutics, reports personal fees from MiRanostics Consulting, Paradigm, Genomic Health, Angiex, Imaging Endpoints II, IBEX Medical Analytics, Spring Bank Pharmaceuticals, Pfizer, IDEA Pharma, GLG Council, Guidepoint Global, Ignyta, and Circulogene; has received travel reimbursement from Cambridge HealthTech Institute, GlaxoSmith Kline, and Tesaro; has ownership interest in MiRanostics Consulting, Exact Sciences, and Circulogene; and has a patent for methods and kits to predict prognostic and therapeutic outcome in small cell lung cancer issued. All other authors declare no competing interests.

References

1. Kullberg BJ, Arendrup MC. Invasive Candidiasis. N Engl J Med. 2015;373:1445–56.
2. Perlroth J, Choi B, Spellberg B. Nosocomial fungal infections: epidemiology, diagnosis, and treatment. Med Mycol. 2007;45:321–46.
3. Fan D, Coughlin LA, Neubauer MM, Kim J, Kim MS, Zhan X, et al. Activation of HIF-1α and LL-37 by commensal bacteria inhibits Candida albicans colonization. Nat Med. 2015;21:808–14.
4. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. Sci Transl Med. 2012;4:165rv13.
5. Förster TM, Mogavero S, Dräger A, Graf K, Polke M, Jacobsen ID, et al. Enemies and brothers in arms: Candida albicans and gram-positive bacteria. Cell Microbiol. 2016;18:1709–15.
6. da Silva Dantas A, Lee KK, Raziunaite I, Schaefer K, Wagener J, Yadav B, et al. Cell biology of Candida albicans–host interactions. Curr Opin Microbiol. 2016;34:111–8.
7. Zelante T, Iannitti RG, Cunha C, De Luca A, Giovannini G, Pieraccini G, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity. 2013;39:372–85.
8. Kaewsrichan J, Peeyananjarassri K, Kongprasertkit J. Selection and identification of anaerobic lactobacilli producing inhibitory compounds against vaginal pathogens. FEMS Immunol Med Microbiol. 2006;48:75–83.
9. Kumar S, Bansal A, Chakrabarti A, Singhi S. Evaluation of efficacy of probiotics in prevention of candida colonization in a PICU-a randomized controlled trial. Crit Care Med. 2013;41:565–72.
10. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. 2007;3:121.
11. Feist AM, Palsson BØ. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat Biotechnol. 2008;26:659–67.
12. Trinh CT, Srienc F. Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol. Appl Environ Microbiol. 2009;75:6696–705.

13. Gomes de Oliveira Dal'Molin C, Quek L-E, Saa PA, Nielsen LK. A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. Front Plant Sci. 2015;6:4.
14. van Hoek MJA, Merks RMH. Redox balance is key to explaining full vs. partial switching to low-yield metabolism. BMC Syst Biol. 2012;6:22.
15. Pfeiffer T, Morley A. An evolutionary perspective on the Crabtree effect. Front Mol Biosci. 2014;1:17.
16. Nilsson A, Nielsen J. Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. Sci Rep. 2016;6:22264.
17. Shlomi T, Benyamini T, Gottlieb E, Sharan R, Ruppin E. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. PLoS Comput Biol. 2011;7:e1002018.
18. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol. 2016;35:81–9.
19. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 2018;46:7542–53.
20. Castillo S, Barth D, Arvas M, Pakula TM, Pitkänen E, Blomberg P, et al. Whole-genome metabolic model of Trichoderma reesei built by comparative reconstruction. Biotechnol Biofuels. 2016; 9:252.
21. Pitkänen E, Jouhten P, Hou J, Syed MF, Blomberg P, Kludas J, et al. Comparative genome-scale reconstruction of gapless metabolic networks for present and ancestral species. PLoS Comput Biol. 2014;10:e1003465.
22. Ene IV, Lohse MB, Vladu AV, Morschhäuser J, Johnson AD, Bennett RJ. Phenotypic profiling reveals that *Candida albicans* opaque cells represent a metabolically specialized cell state compared to default white cells. mBio. 2016;7:e01269–16.
23. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLOS Comput Biol. 2017;13:e1005494.
24. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003;5:264–76.
25. Heinken A, Thiele I. Systems biology of host-microbe metabolomics. Wiley Interdiscip Rev. 2015;7:195–219.
26. Vaas LAI, Sikorski J, Hofner B, Fiebig A, Buddruhs N, Klenk HP, et al. Opm: an R package for analysing OmniLog® phenotype microarray data. Bioinformatics. 2013;29:1823–4.
27. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490:55–60.
28. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015;12:902–3.
29. Li J, Sung CYJ, Lee N, Ni Y, Pihlajamäki J, Panagiotou G, et al. Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice. Proc Natl Acad Sci USA. 2016;113: E1306–15.
30. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
31. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41.
32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6.
33. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, et al. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol. 2013;22:5271–7.
34. Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, et al. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc Natl Acad Sci USA. 2015;112: 12764–9.
35. Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol. 2008;26:1155–60.
36. Aung HW, Henry SA, Walker LP. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. Ind Biotechnol. 2013; 9:215–28.
37. Tramontano M, Andrejev S, Pruteanu M, Klünemann M, Kuhn M, Galardini M, et al. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nat Microbiol. 2018;3:514–22.
38. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28:245–8.
39. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol Syst Biol. 2010;6:390.
40. Askew C, Sellam A, Epp E, Hogues H, Mullick A, Nantel A, et al. Transcriptional Regulation of Carbohydrate Metabolism in the Human Pathogen Candida albicans. PLoS Pathog. 2009;5: e1000612.
41. Kraidlova L, Schrevens S, Tournu H, Van Zeebroeck G, Sychrova H, Van Dijck P, et al. Characterization of the Candida albicans amino acid permease family: Gap2 is the only general amino acid permease and Gap4 Is an S-Adenosylmethionine (SAM) transporter required for SAM-induced morphogenesis. mSphere. 2016;1:e00284–16.
42. Brunke S, Seider K, Richter ME, Bremer-Streck S, Ramachandra S, Kiehntopf M, et al. Histidine degradation via an aminotransferase increases the nutritional flexibility of Candida glabrata. Eukaryot Cell. 2014;13:758–65.
43. Rząd K, Milewski S, Gabriel I. Versatility of putative aromatic aminotransferases from Candida albicans. Fungal Genet Biol. 2018;110:26–37.
44. Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, et al. The gut microbiota modulates host amino acid and glutathione metabolism in mice. Mol Syst Biol. 2015;11:834.
45. Jastrzębowska K, Gabriel I. Inhibitors of amino acids biosynthesis as antifungal agents. Amino Acids. 2015;47:227–49.
46. Dabrowa N, Taxer SSS, Howard DH. Germination of Candida albicans induced by proline. Infect Immun. 1976;13:830–5.
47. Silao FGS, Ward M, Ryman K, Wallström A, Brindefalk B, Udekwu K, et al. Mitochondrial proline catabolism activates Ras1/cAMP/PKA-induced filamentation in Candida albicans. PLoS Genet. 2019;15:e1007976.
48. Herrero AB, López MC, García S, Schmidt A, Spaltmann F, Ruiz-Herrera J, et al. Control of filament formation in Candida albicans by polyamine levels. Infect Immun. 1999;67:4870–8.
49. Emiola A, Oh J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. Nat Commun. 2018;9:4956.
50. Cottier F, Tan ASM, Xu X, Wang Y, Pavelka N. MIG1 Regulates Resistance of Candida albicans against the Fungistatic Effect of Weak Organic Acids. Eukaryot Cell. 2015;14:1054–61.
51. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol. 2017;19:29–41.
52. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes. 2016;7:189–200.

53. Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet J-P, et al. Towards the human intestinal microbiota phylogenetic core. Environ Microbiol. 2009;11:2574–84.
54. Kuhn KA, Schulz HM, Regner EH, Severs EL, Hendrickson JD, Mehta G, et al. Bacteroidales recruit IL-6-producing intraepithelial lymphocytes in the colon to promote barrier integrity. Mucosal Immunol. 2018;11:357–68.
55. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The genus alistipes: gut bacteria with emerging implications to inflammation, cancer, and mental health. Front Immunol 2020;11:906.
56. Cabral DJ, Penumutchu S, Norris C, Morones-Ramirez JR, Belenky P. Microbial competition between escherichia coli and candida albicans reveals a soluble fungicidal factor. Microb Cell. 2018;5:249–55.
57. Heinken A, Sahoo S, Fleming RMT, Thiele I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. Gut Microbes. 2013;4:28–40.
58. Venturelli OS, Carr AV, Fisher G, Hsu RH, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol Syst Biol. 2018;14.
59. Garza DR, Verk MC, Huynen MA, Dutilh BE. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. Nat Microbiol. 2018;3:456–60.
60. Biggs MB, Papin JA. Novel multiscale modeling tool applied to pseudomonas aeruginosa biofilm formation. PLoS One. 2013;8:e78011.
61. Bauer E, Zimmermann J, Baldini F, Thiele I, Kaleta C. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. PLOS Comput Biol. 2017;13:e1005544.
62. Liang SH, Anderson MZ, Hirakawa MP, Wang JM, Frazer C, Alaalm LM, et al. Hemizygosity enables a mutational transition governing fungal virulence and commensalism. Cell Host Microbe. 2019;25:418–31.e6.
63. Allert S, Förster TM, Svensson CM, Richardson JP, Pawlik T, Hebecker B, et al. Candida albicans-induced epithelial damage mediates translocation through intestinal barriers. mBio. 2018;9:e00915–18.
64. Leonardi I, Li X, Semon A, Li D, Doron I, Putzel G, et al. CX3CR1+, mononuclear phagocytes control immunity to intestinal fungi. Science. 2018;359:232–6.
65. Gillum AM, Tsay EY, Kirsch DR. Isolation of the Candida albicans gene for orotidine-5′-phosphate decarboxylase by complementation of S. cerevisiae ura3 and E. coli pyrF mutations. Mol Gen Genet. 1984;198:179–82.
66. Chauvel M, Nesseir A, Cabral V, Znaidi S, Goyard S, Bachellier-Bassi S, et al. A Versatile overexpression strategy in the pathogenic Yeast Candida albicans: Identification of regulators of morphogenesis and fitness. PLoS ONE 2012;7:e45912.
67. Noble SM, Johnson AD. Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen candida albicans. Eukaryotic Cell 2005;4:298–309.
68. Homann OR, Dea J, Noble SM, Johnson AD. A Phenotypic profile of the candida albicans regulatory network. PLoS Genet 2009;5:e1000783.
69. Jiménez-López C, Collette JR, Brothers KM, Shepardson KM, Cramer RA, Wheeler RT, et al. Candida albicans induces arginine biosynthetic genes in response to host-derived reactive oxygen species. Eukaryotic Cell 2012;12:91–100.

No 4. *Candida albicans* SR-like protein kinases regulate different cellular processes: Sky1 is involved in control of ion homeostasis, while Sky2 is important for dipeptide utilization

FORM 1

Manuscript No. 4

Manuscript title: *Candida albicans* SR-like protein kinases regulate different cellular processes: Sky1 is involved in control of ion homeostasis, while Sky2 is important for dipeptide utilization

Authors: Philipp Brandt, Franziska Gerwien, Lysett Wagner, Thomas Krüger, Bernardo Ramírez-Zavala, **Mohammad H. Mirhakkak**, Sascha Schäuble, Olaf Kniemeyer, Gianni Panagiotou, Axel A. Brakhage, Joachim Morschhäuser, Slavena Vylkova

Bibliographic information: Brandt, Philipp, et al. "*Candida albicans* SR-like protein kinases regulate different cellular processes: Sky1 is involved in control of ion homeostasis, while Sky2 is important for dipeptide utilization." *Frontiers in Cellular and Infection Microbiology* (2022)

The candidate is (Please tick the appropriate box.)

☐ First author, ☐ Co-first author, ☐ Corresponding author, ☐ Co-author.

Status: Accepted for publication

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Brandt, P.	30 %	35 %	60 %	50 %	
Gerwien, F.		5 %	5 %		
Wagner, L.		10 %	5 %	10 %	
Krüger, T.		25 %	15 %	5 %	
Ramírez-Zavala, B.			15 %		
Mirhakkak, M.H.		20 %		5 %	
Schäuble, S.		5%		5 %	
Kniemeyer, O.					10 %
Panagiotou, G.					10 %
Brakhage, A.A.					10 %
Morschhäuser, J.	20 %				20 %
Vylkova, S.	50 %			25 %	50 %
Others					
Total:	100 %	100 %	100 %	100 %	100 %

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

FORM 2

Manuscript No. 4

Short reference Brandt *et al*., *Frontiers in Cellular and Infection Microbiology* (Accepted for publication)

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles): My contribution to the publication is Phenotypic microarray data analysis.

Figure #3	☒	**100%**
Figures #1-2, 4-8, S1-S2	☒	**0%**

(Add more table boxes depending on the number of figures)

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

Candida albicans SR-like protein kinases regulate different cellular processes: Sky1 is involved in control of ion homeostasis, while Sky2 is important for dipeptide utilization

Philipp Brandt [1], Franziska Gerwien [1], Lysett Wagner [1], Thomas Krüger [2], Bernardo Ramírez-Zavala [3], Mohammad H. Mirhakkak [4], Sascha Schäuble [4], Olaf Kniemeyer [2], Gianni Panagiotou [4,5], Axel A. Brakhage [2], Joachim Morschhäuser [3] and Slavena Vylkova [1] *

[1]Septomics Research Center, Friedrich Schiller University and Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany
[2]Molecular and Applied Microbiology, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany
[3]Institute for Molecular Infection Biology, University of Würzburg, Germany
[4]Systems Biology and Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany
[5]Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, China

*** Correspondence:**
Slavena Vylkova
Slavena.Vylkova@leibniz-hki.de

Keywords: *Candida albicans*; protein kinases; Sky1, Sky2, dipeptide transport, ion homeostasis

Abstract

Protein kinases play a crucial role in regulating cellular processes such as growth, proliferation, environmental adaptation and stress responses. Serine-arginine (SR) protein kinases are highly conserved in eukaryotes and regulate fundamental processes such as constitutive and alternative splicing, mRNA processing and ion homeostasis. The *Candida albicans* genome encodes two (Sky1, Sky2) and the *Candida glabrata* genome one homolog (Sky1) of the human SR protein kinase 1, but their functions have not yet been investigated. We used deletion strains of the corresponding genes in both fungi to study their cellular functions. *C. glabrata* and *C. albicans* strains lacking *SKY1* exhibited higher resistance to osmotic stress and toxic polyamine concentrations, similar to their ortholog Sky1 in *Saccharomyces cerevisiae*. Deletion of *SKY2* in *C. albicans* resulted in impaired utilization of various dipeptides as the sole nitrogen source. Subsequent phosphoproteomic analysis identified the di- and tripeptide transporter Ptr22 as a potential Sky2 substrate. The transcript levels of this transporter remained constant regardless of the genomic background or the stimulus, suggesting that it is regulated on a post-transcriptional level. Sky2 seems to be involved in Ptr22 regulation since overexpression of *PTR22* in the *sky2*Δ mutant restored the ability to grow on dipeptides and made the cells more susceptible to the dipeptide antibiotics Polyoxin D and Nikkomycin Z. Altogether, our results demonstrate that *C. albicans* and *C. glabrata* Sky1 protein kinases are functionally similar to Sky1 in *S. cerevisiae*, whereas *C. albicans* Sky2 is important for regulation of dipeptide uptake via Ptr22.

Introduction

The human fungal pathogen *Candida albicans* colonizes various host niches, such as the oral cavity, the gastrointestinal and urogenital tracts, and the skin of healthy individuals. In immunocompromised patients and other susceptible individuals it can cause mucosal or systemic infections (Odds, 1988). Both its commensal and pathogenic lifestyles are regulated via complex networks in which protein kinases play an essential role. The *C. albicans* genome comprises 108 predicted protein kinases, many of which were shown to regulate cellular growth and proliferation, resistance to environmental stresses, and the expression of virulence attributes (Monge et al., 2006; Hogan and Sundstrom, 2009; Ramírez-Zavala et al., 2017). Despite their importance for commensalism and pathogenicity, the exact role of more than half of the protein kinases in *C. albicans* remains unknown.

The serine-arginine (SR) protein kinase (SRPK) subfamily is highly conserved from yeast to humans. They commonly catalyze the phosphorylation of mRNA regulatory proteins enriched in serine/arginine recognition motifs and therefore termed SR proteins (Zahler et al., 1992). Human SRPK1-3 are critical for the regulation of both constitutive and alternative splicing, mRNA nuclear export and stability, as well as translational control via shuttling of SR proteins to the cytoplasm (Zhou and Fu, 2013). In addition, human SRPK1 is also exploited during viral infections to facilitate the viral cell cycle and can also act as a tumor suppressor by modulating the state of the Akt kinase, a hallmark of several cancers like prostate, breast, and lung cancer (Hayes et al., 2007; Karakama et al., 2010; Odunsi et al., 2012; Wang et al., 2014).

Fungal SRPK-like kinases can be found across the whole fungal kingdom, including the phyla Ascomycota, Basidiomycota, Kickxellomycotina, Mortierellomycotina, and Mucoromycotina. However, their functions were studied only in the plant pathogens *Fusarium graminearum*, *Physarum polycephalum*, and *Puccinia striiformis f. sp. tritici* (Liu et al., 2009; Cheng et al., 2015; Wang et al., 2018), and in the model yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Both model yeasts have a single SRPK member that is homologous to the human SRPK1: *S. pombe* Dsk1 regulates mitosis and pre-mRNA splicing (Takeuchi and Yanagida, 1993; Tang et al., 1998; Tang et al., 2007), whereas *S. cerevisiae* Sky1 (ScSky1) broadly regulates cellular functions. For example, ScSky1 phosphorylates Npl3, an SR-like protein involved in histone H2B ubiquitination, pre-mRNA splicing, and export of mRNA from the nucleus (Moehle et al., 2012). In addition, ScSky1 regulates polyamine transport and is involved in ion homeostasis and salt tolerance (Erez and Kahana, 2001; Forment et al., 2002). Importantly, its reported targets lack the typical SR protein structure found in mammalian SRPK substrates which are characterized by one or two N-terminal RNA recognition motifs and a C-terminus enriched in SR/RS sites, typically as a stretch of repeats (Ghosh and Adams, 2011). Instead, ScSky1 targets contain up to three randomly positioned RNA recognition motifs and the SR/RS sites are dispersed throughout the protein sequence, making the identification of SRPK substrates difficult. Due to these differences, ScSky1 and its targets are termed “SR-like” (Gilbert et al., 2001).

According to the *Candida* Genome Database (CGD), the *C. albicans* genome encodes two ScSky1 orthologs - C1_06090C (orf19.2436) and C2_06600W (orf19.35), termed CaSky1 and CaSky2 in this article. In contrast, the *C. glabrata* genome encodes one ortholog, CAGL0F03905g, termed CgSky1 in this article. The functions of these predicted protein kinases have not yet been investigated. Hence, in this work we focus on functional characterization of the SRPK1 homologs in *C. albicans* and *C. glabrata*.

In this study, we confirm that CaSky1 and CaSky2 are homologs of the human SRPK1 protein. Likewise, other pathogenic *Candida* spp. including *Candida auris*, *Candida parapsilosis* and *Candida tropicalis* possess two SRPK1 homologs, except for *C. glabrata,* which has only a single member – CgSky1. We investigated their cellular roles by utilizing strains lacking these proteins. Cg*sky1*Δ and Ca*sky1*Δ mutants were resistant to osmotic stress and toxic polyamine concentrations compared to the wild type strains, which indicates a conserved functional similarity to ScSky1. In contrast, Ca*sky2*Δ mutants were impaired in utilization of dipeptides, a feature important for metabolic adaptation to human host niches (Dunkel et al., 2013). Phosphoproteomic analysis of *C. albicans sky1*Δ and *sky2*Δ

mutants revealed little overlap in the potential phosphorylation target proteins of these two kinases. Among the candidate target proteins of CaSky2 was Ptr22, a di- and tripeptide transporter in *C. albicans*. This transporter appeared to be regulated post-transcriptionally, since the transcript levels remained constant regardless of the genomic background. Furthermore, overexpression of *PTR22* in the Ca*sky2*Δ mutant restored the ability to grow on dipeptides as the sole nitrogen source and rendered the cells more susceptible to the dipeptide antibiotics Polyoxin D and Nikkomycin Z. These results demonstrate the distinct functions of the two SR-like protein kinases in *C. albicans*, with CaSky1 regulating ion homeostasis and CaSky2 being involved in regulation of dipeptide uptake via the Ptr22 transporter.

Materials and Methods

Strains and growth conditions

The *C. albicans*, *C. glabrata*, and *S. cerevisiae* strains used in this study are listed in Supplementary Table 1. All strains were stored as frozen stocks containing 20% glycerol at -80°C and sub-cultured on YPD agar plates (1% yeast extract, 2% peptone, 2% glucose) at 30°C for 2 days. Strains were routinely grown in YPD liquid medium at 30°C overnight with shaking at 180 rpm.

Candida albicans strain construction

C. albicans deletion mutants and complemented stains were constructed as described previously (Ramírez-Zavala et al., 2017). *C. albicans PTR22* overexpression strains were generated using the ApaI-SacII fragment from plasmid pPTR22E1 to integrate *PTR22* under the control of the *ADH1* promoter in the wild-type strain SC5314 and the *sky2*Δ mutants as described previously (Reuss et al., 2004; Dunkel et al., 2013). The correct genomic integration of all constructs was confirmed by Southern hybridization with the upstream and downstream flanking sequences. All primers are listed in Supplementary Table 2.

Candida glabrata strain construction

The generation of Cg*sky1*Δ was conducted with a PCR-based Gibson Assembly cloning approach (NEB) according to the manufacturers' protocol. Purified PCR fragments for the puC19 vector backbone, the Cg*SKY1* 5' flank (~1000 bp), a barcoded nourseothricin resistance cassette (*NAT1*) with constant flanking regions (U1 and D1) derived from the mutant 7G6 from Schwarzmüller *et al*. (Schwarzmüller et al., 2014), and the Cg*SKY1* 3' flank (~1000 bp) were fused into one vector. The deletion construct was verified by sequencing, then PCR-amplified and used to transform the *C. glabrata* wild-type strain ATCC2001 by a modified heat shock method (with 45°C heat shock for 15 min) (Sanglard et al., 1996). The transformants were plated onto YPD agar plates containing 200 µg/ml nourseothricin and positive knockout strains were verified by control PCRs. All primers are listed in Supplementary Table 2.

Saccharomyces cerevisiae strain construction

The *S. cerevisiae* wild-type strain BY4741 (Y00000) and the *sky1*Δ mutant (YMR216C) were obtained from Euroscarf (www.euroscarf.de). Both strains were transformed with the plasmid pHLUM (Addgene, Massachusetts, USA) by a LiAc/SS carrier DNA/PEG method to restore the non-auxotrophic strain background as described previously (Gietz and Schiestl, 2007).

Multiple alignment analysis

The alignment was generated using 'MUSCLE Alignment' implemented in Geneious Prime (v2020.1.1.) using default settings. Sequence data originates from GenBank (NP_003128, NP_013943) and the *Candida* Genome Database (C1_06090C, C2_06600W, CAGL0F03905g). The

data presented in Figure 1A was exported from Geneious Prime. For Figure 1B the alignment was restricted to the kinase domain only and illustrated by applying the R-package 'ggmsa' (by Guangchuang Yu, https://CRAN.R-project.org/package=ggmsa) using the color scheme "Clustal".

High-throughput phenotypic screen

C. albicans strains (Supplementary Table 1) were pre-grown on YPD plates. The high-throughput phenotypic screen was performed using Phenotype MicroArrays for microbial cells (PM) plates, reagents, and devices according to the manufacturers' instructions (Biolog, Inc., USA). Briefly, *C. albicans* cells were scraped from YPD agar plates and the cell number was adjusted in sterile dH_2O to 62% transmittance as measured by a turbidimeter (Biolog, Inc., USA). The cells were added to inoculating fluid IFY-0 base (1.2x), redox dye mix D (75x) (Biolog Inc., USA), and further supplemented with potassium phosphate, sodium sulfate and either glucose or glutamic acid (Sigma-Aldrich). 100 µl of the respective mixture was added to each well of PM plates for fungi (PM1-10 and PM21-25) to test for metabolic activity in the presence of different carbon sources, nitrogen sources, supplements, and chemicals. Plates were incubated at 37°C in an OmniLog multiple plate reader (Biolog, Inc., USA). Metabolic activity was determined by reduction of the redox dye and kinetically measured every 15 min at an OD of 750 nm for 24 – 48 h. Each experiment was performed in duplicates.
The respective negative control was substracted from the growth signals in each Phenotype MicroArray. Negative values were replaced by zero, if present. Next, each array's growth signals were categorized into two groups: active growth and non-active growth, as proposed by Vehkala *et al*. (Vehkala et al., 2015). In brief, a data fitted logistic curve represented active growth, while a line without showing exponential growth phase characteristics was interpreted as non-active growth in the investigated time frame. The method was repeated for each replicate separately. Hence, a given substrate's growth signal was detected as non-active or active growth if the analysis identified the same behavior for both replicates. Potential growth was identified if not all replicates for a given tested metabolite showed active growth. Next, a one-way ANOVA and a post-hoc Tukey's test were applied to identify the significant growth differences per substrate across the different *C. albicans* strains. For statistical analysis, the last time point per substrate per phenotypic microarray assay was chosen. Grouping the growth signals was implemented in R version 3.6.0 with the pipeline proposed by Vehkala *et al*., (Vehkala et al., 2015) built upon the opm package version 1.3.77 (Vaas et al., 2013).

Growth assays

YPD overnight cultures of the wild-type and mutant strains were centrifuged (4,000 × g, 5 min) and washed twice with dH_2O. Susceptibility to osmotic stress and stress caused by polyamines was tested by spot dilution assays on YPD agar plates as described in the respective figure legends. Briefly, strains were adjusted to an optical density (OD_{600}) of 1.0 and 5 µl of each 10-fold serial dilutions was spotted onto the respective plates and incubated for 2 days at either 30°C or 37°C.
Growth curves in liquid media were performed in YCB medium (11.7 g/L yeast carbon base, pH 5.0) containing 10 mM of the respective dipeptides. YPD overnight cultures of the respective *C. albicans* strains were centrifuged (4,000 × g, 5 min) and washed twice with dH_2O. Strains were adjusted to an OD_{600} of 0.01 in 12 ml YCB medium containing 10 mM of the respective dipeptides and incubated for 48 h at 37°C shaking at 180 rpm. The OD was measured after 24 h and 48 h. In addition, growth curves were performed in SD medium (0.17% (w/v) yeast nitrogen base without ammonium sulfate and amino acids, 0.5% (w/v) ammonium sulfate, 2% (w/v) glucose, pH 5.0) supplemented with 140 µg/ml Nikkomycin Z (Sigma-Aldrich) or 130 µg/ml Polyoxin D (Sigma-Aldrich) to test for sensitivity to these antibiotics. Strains were adjusted to an OD_{600} of 0.02 in 100 µl of the respective medium in 96-well plates. Plates were incubated at 37°C in an Infinite 200 Pro plate reader (Tecan, Germany) and

absorbance was measured every 20 min at an optical density of 600 nm for 24 h or 48 h. Each strain was measured in triplicates and the means and standard derivations were calculated.

qRT-PCR

YPD overnight cultures of the wild-type and mutant strains were centrifuged (4000 × g, 5 min) and washed twice with dH_2O. Strains were adjusted to an optical density (OD_{600}) of 0.1 in 30 ml YCB medium containing 10 mM of the respective dipeptides and grown for 4 h at 37°C. Shock frozen cell pellets were used for RNA isolation with a phenol-chloroform based method previously described by Martin *et al*. 2011 (Martin et al., 2011). RNA quality and quantity were determined using the Bioanalyzer device (Agilent) and the Nanodrop (Thermo Fisher Scientific), respectively. The following digest of residual gDNA was performed for 8 µg high-quality RNA with the Turbo DNase Kit (Ambion). After purification with the RNA Clean & Concentrator Kit (Zymo Research), cDNA synthesis was conducted with the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific) for 2 µg RNA. qRT-PCR was conducted in technical and biological triplicates using the Brilliant II SYBR Green system (Agilent) with diluted cDNA samples (1:10) and genomic DNA from the SC5314 WT strain (1:10 dilutions ranging from 100 µg/ml to 0.1 µg/ml) to determine primer efficiency. The qRT-PCR was run at the Agilent Stratagene Mx3005 device (Software MxPro) and further analyzed using the Δct method. The indicated expression is shown relative to the housekeeping gene *MED15*. All primers are listed in Supplementary Table 2.

Proteomics sample preparation

YPD overnight cultures of the wild type and mutant strains were adjusted to an optical density (OD_{600}) of 0.1 in 900 ml YPD medium and grown for 4 h at 37°C. Cells were disrupted by using mortar and pestle with liquid nitrogen. Cell debris were homogenized in lysis buffer (1% (w/v) SDS, 150 mM NaCl, 100 mM TEAB (triethyl ammonium bicarbonate), one tablet each of cOmplete Ultra Protease Inhibitor Cocktail and PhosSTOP). After addition of 0.5 µl Benzonase nuclease (250 U/µl) the samples were incubated at 37°C in a water bath sonicator for 30 min. Proteins were separated from insolubilized debris by centrifugation (15 min, 18000 × *g*). Each 6 mg of total protein per sample was diluted with 100 mM TEAB to gain a final volume of 4 ml. Subsequently, cysteine thiols were reduced and carbamidomethylated in one step for 30 min at 70°C by addition of 120 µl of 500 mM TCEP (tris(2-carboxyethyl)phosphine) and 120 µl of 625 mM 2-chloroacetamide (CAA). The samples were further cleaned up by methanol-chloroform-water precipitation using the protocol of Wessel and Flügge (Wessel and Flügge, 1984). Protein precipitates were resolubilized in 5% trifluoroethanol of aqueous 100 mM TEAB and digested overnight (18 h) with a Trypsin+LysC mixture (Promega) at a protein to protease ratio of 33:1. Each sample was divided in 6 × 1 mg used for the phosphopeptide enrichment and 200 µg initial protein used as reference for proteome analysis. Samples were evaporated in a SpeedVac. The reference proteome sample was resolubilized in 50 µl of 0.05% TFA in H_2O/ACN 98/2 (v/v) filtered through Ultrafree-MC 0.2 µm PTFE membrane spin filters (Merck-Millipore). The filtrate was transferred to HPLC vials and injected into the LC-MS/MS instrument for further analysis.

Phosphopeptide enrichment

Phosphopeptides were enriched by using TiO_2+ZrO_2 TopTips (Glygen Corp., Columbia, MD, USA). TopTips were loaded with 1 mg protein isolate using 6 TopTips per biological replicate after equilibration with 200 µl Load and Wash Solution 1, LWS1 (1% trifluoroacetic acid (TFA), 20% lactic acid, 25% acetonitrile (ACN), 54% H_2O). TopTips were centrifuged at 1500 rpm (~200 × *g*) for 5 min at room temperature. After washing with 200 µl LWS1, the TiO_2/ZrO_2 resin was washed with 25% ACN and subsequently the phosphopeptides were eluted with 200 µl $NH_3 \cdot H_2O$ (NH_4OH), pH 12. The alkaline solution was immediately evaporated using a vacuum concentrator (Eppendorf). The phosphoproteome samples were resolubilized in 50 µl of 0.05% TFA in H_2O/ACN 98/2 (v/v) filtered

through Ultrafree-MC 0.2 µm PTFE membrane spin filters (Merck-Millipore). The filtrate was then transferred to HPLC vials and injected into the LC-MS/MS instrument for further analysis.

LC-MS/MS analysis

Each sample was measured in triplicates (3 analytical replicates of 3 biological replicates of a reference proteome fraction and a phosphoproteome fraction). LC-MS/MS analysis was performed on an Ultimate 3000 nano RSLC system connected to a QExactive HF mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Peptide trapping for 5 min on an Acclaim Pep Map 100 column (2 cm x 75 µm, 3 µm) at 5 µl/min was followed by separation on an analytical Acclaim Pep Map RSLC nano column (50 cm × 75 µm, 2µm). Mobile phase gradient elution of eluent A (0.1% (v/v) formic acid in water) mixed with eluent B (0.1% (v/v) formic acid in 90/10 acetonitrile/water) was performed using the following gradient for the more hydrophilic phosphoproteome samples: 0-5 min at 4% B, 15 min at 7% B, 50 min at 10% B, 100 min at 14% B, 150 min at 25% B, 190 min at 60% B, 205-215 min at 96% B, 215-240 min at 4% B. The reference proteome gradient was as follows: 0-4 min at 4% B, 10 min at 7% B, 50 min at 12% B, 100 min at 16% B, 150 min at 25% B, 175 min at 35% B, 200 min at 60 %B, 210-215 min at 96% B, 215.1-240 min at 4% B.
Positively charged ions were generated at a spray voltage of 2.2 kV using a stainless-steel emitter attached to the Nanospray Flex Ion Source (Thermo Fisher Scientific). The quadrupole/orbitrap instrument was operated in Full MS / data dependent MS2 Top15 mode. Precursor ions were monitored at m/z 300-1500 at a resolution of 120,000 FWHM (full width at half maximum) using a maximum injection time (ITmax) of 120 ms and an AGC (automatic gain control) target of 3×10^6. Precursor ions with a charge state of z=2-5 were filtered at an isolation width of *m/z* 1.6 amu for further HCD fragmentation at 30% normalized collision energy (NCE). MS2 ions were scanned at 15,000 FWHM (ITmax=100 ms, AGC= 2×10^5) using a fixed first mass of *m/z* 120 amu. Dynamic exclusion of precursor ions was set to 30 s and the minimum AGC target for Precursor ions selected for HCD fragmentation was set to 1e3. The LC-MS/MS instrument was run by Chromeleon 7.2, QExactive HF Tune 2.8 and Xcalibur 4.0 software.

Protein database search

Tandem mass spectra were searched against the UniProt database (2021/07/19) (YYYY/MM/DD); https://www.uniprot.org/proteomes/UP000000559) of *Candida albicans* SC5314 using Proteome Discoverer (PD) 2.4 (Thermo) and the algorithms of Mascot 2.4.1 (Matrix Science, UK), Sequest HT (version of PD2.4), MS Amanda 2.0, and MS Fragger 3.2. Two missed cleavages were allowed for the tryptic digestion. The precursor mass tolerance was set to 10 ppm and the fragment mass tolerance was set to 0.02 Da. Modifications were defined as dynamic Met oxidation, phosphorylation of Ser, Thr, and Tyr, protein N-term acetylation as well as static Cys carbamidomethylation. A strict false discovery rate (FDR) < 1% (peptide and protein level) and a search engine score of >30 (Mascot), > 4 (Sequest HT), >300 (MS Amanda) or >8 (MS Fragger) was required for positive protein hits. The Percolator node of PD2.4 and a reverse decoy database was used for q value validation of spectral matches. Only rank 1 proteins and peptides of the top scored proteins were counted. Label-free protein quantification was based on the Minora algorithm of PD2.4 using the precursor abundance based on intensity and a signal-to-noise ratio >5. Normalization was performed by using the total peptide amount method. Imputation of missing quan values was applied by using abundance values of 75% of the lowest abundance identified per sample. For the reference proteome analysis used for master protein abundance correction of the phosphoproteome data, phosphopeptides were excluded from quantification. Differential protein abundance was defined as a fold change of >4, ratio-adjusted pvalue <0.05 (pvalue/log4ratio) and at least identified in 3 of 3 replicates. Differential phosphopeptide abundance was defined as a fold change of >4, ratio-adjusted pvalue <0.05 (pvalue/log4ratio) and at least identified in 2 of 3 replicates. The mass spectrometry proteomics data have been deposited to the

ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD027612.

Results

Sky1 and Sky2 are the sole predicted SR-like protein kinases in *C. albicans*

According to the *Candida* Genome Database, the *C. albicans* genome encodes two predicted SR protein kinases, based on sequence similarity to the sole SR-like protein kinase in *S. cerevisiae*, ScSky1. These orthologs correspond to the proteins encoded by C1_06090C (48.5% identity to ScSky1) and C2_06600W (46.5% identity to ScSky1), named in this study CaSky1 and CaSky2, respectively. To ensure that these are the only SRPKs in the pathogenic fungus and to identify homologs in other pathogenic *Candida* spp., we performed a BLASTp search analysis against the human SR protein kinase 1 protein (SRPK1) (NCBI Reference Sequence: NP_003128.3). Interestingly, except for *C. glabrata*, the most common pathogenic *Candida* spp. possess two SRPK homologs (Supplementary Figure 1). Not surprisingly, the CAGL0F03905g-derived protein in *C. glabrata* CgSky1 has 60.35% identity to ScSky1 (NCBI protein BLAST (Basic Local Alignment Search Tool)). This illustrates the phylogenetic relationships between the examined species: *C. glabrata* is more closely related to *S. cerevisiae* compared to the other *Candida* species that belong to the CTG clade (Muñoz et al., 2018). Therefore, we decided to include CgSky1 in our further analysis. The human SRPKs contain a well-conserved kinase domain at the C-terminus that is divided by a large, non-conserved insert domain (about 250 amino acids) (Ghosh and Adams, 2011). Multiple sequence alignment showed that the domains that contain the ATP binding sites, the activation loop and the kinase-docking sites are particularly highly conserved between *H. sapiens*, *S. cerevisiae*, *C. glabrata*, and *C. albicans* (Figure 1). Interestingly, the CaSky2 protein has a considerable number of additional amino acid residues on the N-terminal side of the kinase domain (alignment position 29-57, 94-108 and 153-320) that are not present in the other proteins (Figure 1B).

SKY1 deletion in *C. glabrata* and *C. albicans* results in higher resistance to osmotic and polyamine stress

S. cerevisiae Sky1 has been characterized as a regulator of polyamine transport and ion homeostasis, as mutants lacking *SKY1* are more resistant to toxic cations and polyamine concentrations (Erez and Kahana, 2001). Ion homeostasis is also a crucial process for *C. albicans* and *C. glabrata* growth and proliferation, survival in the host, virulence mechanisms and resistance to antifungal drugs (Llopis-Torregrosa et al., 2016; Li et al., 2018). To test whether CgSky1, CaSky1, and CaSky2 are involved in similar cellular functions we generated *C. glabrata* and *C. albicans* mutant strains lacking *SKY1* (Cg*sky1*Δ, Ca*sky1*Δ) and *C. albicans* strains lacking both *SKY1* and *SKY2* (Ca*sky1*Δ/*sky2*Δ). In addition, we used available *C. albicans sky2*Δ (Ca*sky2*Δ) mutants (Ramírez-Zavala et al., 2017). As a control we used the *S. cerevisiae* BY4741 wild-type strain and a *sky1*Δ mutant (Sc*sky1*Δ) from Euroscarf (www.euroscarf.de). In accordance with previous observations, the Sc*sky1*Δ mutant was more resistant to osmotic stress caused by NaCl or LiCl and to high concentrations of the polyamine spermine (Erez and Kahana, 2001; Forment et al., 2002). Similar phenotypes were observed for *C. glabrata* mutants lacking *SKY1* (Figure 2). Deletion of *SKY1* in *C. albicans* resulted in high resistance to LiCl and spermine, whereas lack of *SKY2* rendered the cells partially resistant to these stressors. None of the tested strains had altered resistance to NaCl (Figure 2). Altogether, these results demonstrate the functional conservation of Sky1 among the examined species.

C. albicans SKY2 is required for growth on various dipeptides as the sole nitrogen source

Since we noted phenotypic differences between the two *C. albicans* SRPKs CaSky1 and CaSky2 we decided to apply a high-throughput phenotypic screen to further characterize their cellular roles. Given that the two independently generated Ca*sky1*Δ and Ca*sky2*Δ mutant strains showed the same

phenotypes, we utilized only strain A of the respective mutant for this analysis. The Cas*ky1*Δ and Cas*ky2*Δ mutant strains were screened for metabolic activity on 904 different nutrients and supplements and compared to the metabolic activity of the wild-type strain SC5314. This approach enabled us to identify additional phenotypic differences of the Cas*ky1*Δ and Cas*ky2*Δ mutants compared to the control strain and between each other. Overall, there were fewer and less pronounced phenotypic differences between the Cas*ky1*Δ mutant and the wild type as opposed to the Cas*ky2*Δ mutant vs. wild-type strain. For instance, the Cas*ky2*Δ mutant exhibited either no active growth or a significantly reduced metabolic activity on multiple dipeptide combinations as a nitrogen source (Figure 3). The Cas*ky2*Δ mutant was able to utilize only 104 out of the tested 268 dipeptides (38.8%), whereas the wild-type strain and the Cas*ky1*Δ mutant were able to utilize 146 (54.5%) and 125 (46.6%) dipeptides, respectively. Interestingly, the Cas*ky2*Δ mutant was as capable as the wild type in utilization of tripeptides (Figure 3), single amino acids and more complex nitrogen sources (data not shown), showing that the dipeptide growth defect is quite specific.

To validate the observations from the phenotypic screen, we grew *C. albicans* SC5314 wild type, Cas*ky2*Δ mutants, and the *SKY2* complemented strains on several peptide combinations as the sole nitrogen source where we observed a growth defect for the Cas*ky2*Δ mutant in the phenotypic screen (Figure 3): alanine-phenylalanine (Ala-Phe), phenylalanine-serine (Phe-Ser), and methionine-glycine (Met-Gly). Indeed, the Cas*ky2*Δ mutants had significantly reduced growth compared to the wild-type strain on these particular dipeptides (Figure 4). The *SKY2* complemented strains either partially restored the Cas*ky2*Δ mutant phenotype (Ala-Phe) or exhibited growth comparable to the Cas*ky2*Δ mutant (Phe-Ser, Met-Gly). This could be due to either the observed differences between the two *SKY2* alleles (data not shown) or to a gene dosage effect. Taken together, these results confirm the findings from the phenotypic screen that *C. albicans* Sky2 is required for assimilation of dipeptides as a nitrogen source.

Phosphoproteome analysis reveals a distinct set of potential protein targets for *C. albicans* Sky1 and Sky2 protein kinases

We next performed a phosphoproteome analysis to identify the potential substrates responsible for the functional differences between CaSky1 and CaSky2. In brief, we grew the strains in YPD medium, extracted the proteins, and performed an LC-MS/MS analysis of tryptic peptides. TiO_2/ZrO_2-mediated phosphopeptide enrichment facilitated the identification and quantification of phosphopeptides with serine, threonine, and/or tyrosine phosphorylation. Phosphopeptide abundances were corrected against the corresponding master protein abundances (obtained from the non-enriched fraction) to enable a site-specific quantification. With this approach, we were able to identify 3946 proteins, 1663 phosphoproteins, 7243 phosphopeptides and 7727 phosphosites in total (Figure 5A). Of those, 19 proteins were less abundant and lacked phosphorylation at specific sites in both Cas*ky1*Δ and Cas*ky2*Δ strains compared to the wild-type strain (Supplementary Table 3,4). For example, phosphorylation of the nucleolar GTP binding protein 1 (Nog1) at S471 and S473 and the GTPase Bms1 at S767 was detected neither in the Cas*ky1*Δ nor in the Cas*ky2*Δ mutant. Furthermore, both mutants exhibited altered phosphorylation of the S81 site of the putative protein kinase Hrk1, which has a predicted role in cellular ion homeostasis. We generated *hrk1*Δ mutants and tested their resistance to osmotic stress and toxic polyamine concentrations. Interestingly, *hrk1*Δ mutants were as resistant as the Cas*ky1*Δ mutants to high LiCl and spermine concentrations (Supplementary Figure 2), suggesting that *C. albicans* Hrk1 could act downstream of Sky1 in a pathway that coordinates cellular responses to osmotic stress and toxic polyamine levels.

Aside from the small number of shared potential protein targets of CaSky1 and CaSky2, we identified many phosphopeptides that were differentially abundant in only one of the examined protein kinase mutants. For the Cas*ky1*Δ mutant we identified 268 phosphopeptides assigned to 224 proteins with significantly altered abundance (175 more abundant, 49 less abundant) compared to the wild type

(Figure 5B). For instance, phosphorylation of the putative membrane protein Lem3 at S33, S34, S35, and S41 and phosphorylation of the putative mRNA export protein Elf1 at S1126 was not detected in the Ca*sky1*Δ mutant, but found in the wild-type strain. For the Ca*sky2*Δ mutant, we identified 237 phosphopeptides assigned to 210 proteins significantly altered in abundance (23 more abundant, 187 less abundant) compared to the wild type (Figure 5B). Among the most liekly direct or indirect targets were the transcription factors Fcr1 and Sfl1. We could not detect phosphorylation of Fcr1 at S113 or S114, S198, S329, S369, S370, S373 and T376 or phosphorylation of Sfl1 at S641 in the Ca*sky2*Δ mutant, despite detectable phosphorylation in the wild type. Another potential target was the dipeptide/ tripeptide transporter Ptr22, for which phosphorylation was not detected at S2, T3, and S39 in the Ca*sky2*Δ mutant compared to the wild type. Furthermore, phosphorylation of certain phosphorylation sites of several protein kinases (Cst20, Gin4, Hsl1, Kic1, Ptk2, Sak1, Sat4, Swe1 and orf19.846) could not be detected in the Ca*sky2*Δ mutant but was detectable in the wild type. This suggests that CaSky2 is either part of a regulatory protein kinase network or that deletion of *SKY2* affects several pathways in the cell. Furthermore, the experimental data demonstrate a role of CaSky2 in regulation of dipeptide transport and support the role of CaSky1 in regulating ion homeostasis and RNA metabolism, a function shared with ScSky1 and CgSky1.

PTR22 overexpression overcomes the growth defect of the Ca*sky2*Δ mutant on dipeptides as the sole nitrogen source

Phosphoproteome analysis identified the dipeptide/tripeptide transporter Ptr22 as a potential direct or indirect target of CaSky2, which could explain the prominent growth defect of Ca*sky2*Δ mutants on dipeptides as the sole nitrogen source (Figure 3). To test whether Sky2 regulates dipeptide utilization via the di- and tripeptide transporter Ptr22, we generated strains overexpressing *PTR22* in the wild type and in the Ca*sky2*Δ mutant background and monitored their growth on selected dipeptides as the sole nitrogen source. As a control, we used the *ptr22*Δ mutant, which has a known growth defect on dipeptides (Dunkel et al., 2013). As expected, the strains lacking *PTR22* were unable to grow on the tested dipeptides (Figure 6). Strains overexpressing *PTR22* under the control of the constitutive *ADH1* promoter grew very well on the tested media yielding 4 times more cells (as measured by OD_{600}) compared to wild-type at 24 h with only slight further increase at 48 h. In comparison, the wild type strain growth on dipeptides at 48 h was similar to that of the *PTR22* overexpressing strains at 24 h. Importantly, overexpression of *PTR22* under control of the constitutive *ADH1* promoter rescued the growth defect of the Ca*sky2*Δ mutants on all dipeptides tested (Figure 6).
The nucleoside antibiotics Nikkomycin Z and Polyoxin D are antifungal agents that act as competitive inhibitors of chitin biosynthesis. It has been shown that *C. albicans* deletion mutant strains with defective uptake of di- and tripeptides are resistant to nikkomycins, polyoxins, and bacilysin (Mehta et al., 1984; Payne and Shallow, 1985). Therefore, we examined whether CaSky2 regulates the uptake of Nikkomycin Z or Polyoxin D via Ptr22. All strains grew well on SD medium (Figure 7). *C. albicans* Ptr22 was required for the growth inhibitory effects of the two antibiotics tested, as Nikkomycin Z and Polyoxin D did not affect the growth of the *ptr22*Δ mutant but did affect that of the wild-type strain (Figure7). Strains lacking *SKY2* also exhibited resistance to both antibiotics, but at a lower level compared to the *ptr22*Δ mutant. Overexpression of *PTR22* in the Ca*sky2*Δ mutant abolished its resistance to both antibiotics (Figure 7).

C. albicans PTR22 is transcriptionally not altered in the Ca*sky2*Δ mutant unlike other peptide transporters

The *C. albicans* genome encodes two dipeptide/tripeptide transporters (*PTR2* and *PTR22*) and eight oligopeptide transporters (*OPT1*-*OPT8*), of which OPT1-5 have been identified as the major oligopeptide transporters (Reuss and Morschhäuser, 2006). We performed qRT-PCR analysis to investigate whether these genes are differentially regulated in the *C. albicans* wild type and the Ca*sky2*Δ mutants grown in YCB medium with Ala-Phe, Phe-Ser or Met-Gly as the sole nitrogen

source. *PTR22* was expressed equally in all strains and media conditions (Figure 8). In all tested dipeptide media, *PTR2* and *OPT5* were expressed at lower levels compared to the other peptide transporters. Differences in the expression of *PTR2* and *OPT5* between the wild type and the Ca*sky2*Δ mutants were observed in Ala-Phe (showing lower expression in the Ca*sky2*Δ mutants) and Met-Gly (with higher expression levels in the Ca*sky2*Δ mutant strain B), while *PTR22* and *OPT1-4* were not differentially expressed in all strains. In contrast, in medium with Phe-Ser as the sole nitrogen source, the transcript levels of the peptide transporter genes *OPT1-4* were increased in the Ca*sky2*Δ mutants compared to the wild type, but these differences were statistically significant only for *OPT2* and *OPT3* (Figure 8). These results suggest that Ptr22 is likely regulated at the post-transcriptional level, e.g., by Sky2-dependent phosphorylation, and that impaired regulation of this transporter triggers the induction of other, probably non-specific, permeases to compensate for the altered dipeptide uptake.

Discussion

The subfamily of SR protein kinases, defined by preferential phosphorylation of mRNA-binding proteins that contain SR/RS-enriched sequences, is highly conserved from yeasts to humans (Giannakouros et al., 2011). While the function of SRPK has been described in detail in the model yeasts *S. cerevisiae* and *S. pombe,* where they control fundamental cellular processes such as mRNA processing and mitosis, their roles in human pathogenic fungi have not been previously investigated. This is the first work that examines their function in pathogenic *Candida* spp., specifically the SR protein kinases Sky1 and Sky2 in *C. albicans* and Sky1 in *C. glabrata*. While both CgSky1 and CaSky1 appear to have similar functions to the sole SRPK in *S. cerevisiae* ScSky1, CaSky2 has both expanded and diverged repertoire of target proteins, such as proteins involved in nutrient utilization.

While both model yeasts *S. cerevisiae* and *S. pombe* have a single SR protein kinase 1 homolog, gene expansions of SRPK kinases have been reported for filamentous fungi like *Aspergillus nidulans* (seven paralogous genes), *Neurospora crassa* (five), and several dermatophytes (18 to 34) (Martinez et al., 2012; De Souza et al., 2013). It is speculated that the response to a broad range of environmental conditions and the higher number of introns per gene in filamentous fungi has led to an increased importance of RNA processing resulted in such expansion (De Souza et al., 2013). Pathogenic *Candida* spp. are also confronted with rapidly changing and diverse host environments. However, alternative splicing in *Candida* spp. likely plays only a minor role in host adaptation, well reflected by the low frequency of introns in *Candida* spp. genomes compared to other pathogenic fungi (Sieber et al., 2018). Furthermore, the duplication of SRPK genes seems to have occurred independently at different time points in the evolution of different taxa, since SRPKs lack a prominent one-to-one correspondence between the sequences, as proposed by Giannakouros *et al*. (Giannakouros et al., 2011). The same authors reported the observation concerning the ‘spacer region’ that is characteristic for the SRPK subfamily. This ‘spacer region’ is highly diverse in sequence length, and possibly in function, as the spacer seems to be unique for each SRPK family member. As illustrated in Figure 1, there is certain sequence diversity in the middle part of the kinase domain between the different homologs. However, CaSky2 shows a unique extension of the N-terminus of the kinase domain, indicative of possible functional differentiation. Indeed, *C. albicans* and *C. glabrata SKY1* deletion strains shared the resistance to high salt and toxic polyamine concentrations to the *S. cerevisiae sky1*Δ mutant. A subsequent high-throughput phenotypic screen confirmed the substantial phenotypic differences between CaSky1 and CaSky2, supporting the notion of their divergent cellular functions. Whether these differences are due to the expanded ‘spacer region’ of CaSky2 would require further investigations.

The main feature of human SRPKs is that they catalyze the phosphorylation of proteins enriched in serine/arginine recognition motifs, which earned them the name SR proteins (Zahler et al., 1992). In *S. cerevisiae* the confirmed protein targets of the SRPK1 homolog Sky1 also have multiple SR/RS sites, but those typically are randomly distributed throughout the protein and therefore referred to as SR-like. For example, ScSky1 phosphorylates the SR-like RNA-binding protein Npl3, which

possesses eight SR/RS sites dispersed within the C-terminus of the protein (Siebel et al., 1999; Gilbert et al., 2001). Similarly, many of the direct or indirect target proteins of CaSky1 and CaSky2 identified in this study contain multiple randomly dispersed SR/RS sites. For example, the two proteins orf19.2459 or orf19.5051, which are among the 19 common potential substrates of CaSky1 and CaSky2, contain 20 and 16 SR/RS sites, respectively. Interestingly, the *C. albicans* ortholog of the *S. cerevisiae* Sky1 direct target ScNpl3, CaNpl3, was not among the potential phosphorylation targets of either CaSky1 or CaSky2. This is likely due to the substantially shorter protein length and the presence of only three SR/RS sites of CaNpl3 compared to the longer ScNpl3 protein that contains eight SR/RS sites. Thus, although direct evidence that CaSky1 and CaSky2 can phosphorylate proteins is still lacking, our data suggests that both are functional SR-like protein kinases.

Of the potential phosphorylation targets for both CaSky1 and CaSky2, we examined Hrk1, a protein kinase with a predicted role in cellular ion homeostasis. Deletion of *HRK1* resulted in the same resistance to LiCl and spermine as the deletion of Ca*SKY1*. The functional relationship between Hrk1 and CaSky1 requires further investigation. In addition to its role in regulation of ion homeostasis, CaSky1, similar to ScSky1, is potentially involved in RNA metabolism, as the putative mRNA export protein Elf1 was one of the most promising phosphorylation targets of CaSky1. Although we have not examined this in great details, *C. glabrata* Sky1 also seems to exert similar cellular functions as ScSky1 and CaSky1. Thus, at least in these three species, Sky1 appears to regulate ion and possibly mRNA homeostasis.

Our analysis revealed substantial functional differences between CaSky1 and CaSky2, with CaSky2 having multiple distinct potential targets and phenotypes. Among the most potential targets for CaSky2 were the SR/RS-enriched transcription factors Fcr1, a zinc cluster transcription factor and negative regulator of fluconazole, ketoconazole and brefeldin A resistance, and Sfl1, which is involved in the negative regulation of morphogenesis, flocculation, and virulence (Bauer and Wendland, 2007; Li et al., 2007; Shen et al., 2007). In addition, phosphorylation at specific phosphorylation sites of multiple protein kinases could not be identified in the Ca*sky2*Δ mutant compared to the SC5314 control strain, including Cst20, Gin4, Hsl1, Kic1, Ptk2, Sak1, Sat4, Swe1 and orf19.846. The Snf1-activating protein kinase Sak1 is a key regulator of metabolic adaptation and *in vivo* fitness in *C. albicans* (Ramírez-Zavala et al., 2017). Furthermore, we identified other potential direct or indirect CaSky2 target proteins that are involved in metabolic processes, such as nutrient uptake, e.g. the dipeptide/tripeptide transporter Ptr22 with no detected phosphorylation at S2, T3 and S39 in the Ca*sky2*Δ mutant compared to the wild type. Based on this, we speculate that CaSky2 plays an important role in signalling pathways that coordinate metabolism, morphogenesis and virulence. In this context, we were able to verify the functional relationship between Sky2 and Ptr22.

In *C. albicans* utilization of di- and tripeptides is mediated by Ptr2 and Ptr22, and partially by the oligopeptide transporter Opt1, while uptake of oligopeptides is mediated by the oligopeptide transporters Opt1-5 (Reuss and Morschhäuser, 2006; Dunkel et al., 2013). During growth on dipeptides as the sole nitrogen source, the transcript levels of *PTR22* remained unaltered, whereas *PTR2* and *OPT1-5* genes were differentially expressed in a substrate-specific manner in the absence of CaSky2. Thus, compensatory mechanisms might be in effect, which could explain the observed intermediate growth of the Ca*sky2*Δ mutant on dipeptides compared to the inhibited growth of the *ptr22*Δ mutant strain. Furthermore, our phosphoproteomics data suggest that Ptr22 is regulated on post-transcriptional level (e.g. by phosphorylation). Since overexpression of *PTR22* in the *C. albicans sky2*Δ mutant completely reversed the growth defect on dipeptides as the sole nitrogen source, we hypothesize that Ptr22 phosphorylation is required either for protein processing and/or its function.

Deletion of *SKY2* in *C. albicans* rendered the cells more resistant to the nucleoside antibiotics Nikkomycin Z and Polyoxin D, both of which are most likely taken up via peptide transporter(s) and which inhibit chitin biosynthesis in fungi (Yadan et al., 1984; Payne and Shallow, 1985). Consistently, strains lacking the dipeptide transporter Ptr22 were not susceptible to these antibiotics, while wild type growth was completely inhibited. The Ca*sky2*Δ mutants exhibited resistance to Nikkomycin Z and

Polyoxin D, but not as much as the *ptr22*Δ mutants. Although both antibiotics are relatively similar in structure, they have different levels of effectiveness, e.g. in *S. cerevisiae* Polyoxin D inhibits the chitin synthetases 1 and 2 more effectively than Nikkomycin Z (Cabib, 1991). While both antibiotics have been investigated for their potential use in treatment of mycoses in plants, only Nikkoymcin Z was tested on animal mycoses. For example, in a murine model the promising effect of Nikkomycin Z to treat coccidioidomycosis, histoplasmosis, and blastomycosis has been shown (Hector et al., 1990). Currently, Nikkomycin Z is in phase 2 clinical trial for treatment of coccidioidomycosis in humans (Nix et al., 2009). As the Ptr22 dipeptide transporter is required for Nikkomycin Z uptake in *C. albicans*, stimulation of dipeptide uptake should increase the efficacy of this antibiotic.
In conclusion, our results clearly demonstrate that the two *C. albicans* SR-like kinases Sky1 and Sky2 regulate distinct cellular processes. Sky1 regulates ion homeostasis as its *S. cerevisiae* Sky1 ortholog, whereas CaSky2 is important for utilization of dipeptides.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

PB, FG, TK, BR and OK designed and performed the wet lab experiments. PB, LW, MM and SS performed in silico analyses. PB, FG, LW, TK, MM and SS analyzed the data. PB, FG, LW and MM designed the figures. GP, AB, JM and SV discussed the results and supervised the project. PB, LW and SV wrote the manuscript in consultation with TK, BR, SS and JM.

Funding

This study was funded by the German Ministry for Education and Science in the program Unternehmen Region (BMBF 03Z22JN11) (to SV) and the German Research Foundation (DFG) through the TRR 124 FungiNet, "Pathogenic fungi and their human host: Networks of Interaction," DFG project number 210879364, project C2 (to SV and JM), B5 (to GP), project INF (to GP), and project Z2 (to OK).

Acknowledgments

We gratefully thank Matthew Blango for the critical reading of the manuscript and all members of the research group Host Fungal Interfaces at ZIK Septomics for the helpful discussions.

Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD027612.

References

Bauer, J., and Wendland, J. (2007). *Candida albicans* Sfl1 suppresses flocculation and filamentation. *Eukaryot Cell* 6(10)**,** 1736-1744. doi: 10.1128/ec.00236-07.

Cabib, E. (1991). Differential inhibition of chitin synthetases 1 and 2 from *Saccharomyces cerevisiae* by polyoxin D and nikkomycins. *Antimicrob Agents Chemother* 35(1)**,** 170-173. doi: 10.1128/aac.35.1.170.

Cheng, Y., Wang, X., Yao, J., Voegele, R.T., Zhang, Y., Wang, W., et al. (2015). Characterization of protein kinase PsSRPKL, a novel pathogenicity factor in the wheat stripe rust fungus. *Environ Microbiol* 17(8)**,** 2601-2617. doi: 10.1111/1462-2920.12719.

De Souza, C.P., Hashmi, S.B., Osmani, A.H., Andrews, P., Ringelberg, C.S., Dunlap, J.C., et al. (2013). Functional Analysis of the *Aspergillus nidulans* Kinome. *PLOS ONE* 8(3)**,** e58008. doi: 10.1371/journal.pone.0058008.

Dunkel, N., Hertlein, T., Franz, R., Reuß, O., Sasse, C., Schäfer, T., et al. (2013). Roles of different peptide transporters in nutrient acquisition in *Candida albicans*. *Eukaryot Cell* 12(4)**,** 520-528. doi: 10.1128/ec.00008-13.

Erez, O., and Kahana, C. (2001). Screening for modulators of spermine tolerance identifies Sky1, the SR protein kinase of *Saccharomyces cerevisiae*, as a regulator of polyamine transport and ion homeostasis. *Mol Cell Biol* 21(1)**,** 175-184. doi: 10.1128/mcb.21.1.175-184.2001.

Forment, J., Mulet, J.M., Vicente, O., and Serrano, R. (2002). The yeast SR protein kinase Sky1p modulates salt tolerance, membrane potential and the Trk1,2 potassium transporter. *Biochim Biophys Acta* 1565(1)**,** 36-40. doi: 10.1016/s0005-2736(02)00503-5.

Ghosh, G., and Adams, J.A. (2011). Phosphorylation mechanism and structure of serine-arginine protein kinases. *Febs j* 278(4)**,** 587-597. doi: 10.1111/j.1742-4658.2010.07992.x.

Giannakouros, T., Nikolakaki, E., Mylonis, I., and Georgatsou, E. (2011). Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. *Febs j* 278(4)**,** 570-586. doi: 10.1111/j.1742-4658.2010.07987.x.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2(1)**,** 31-34. doi: 10.1038/nprot.2007.13.

Gilbert, W., Siebel, C.W., and Guthrie, C. (2001). Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *Rna* 7(2)**,** 302-313. doi: 10.1017/s1355838201002369.

Hayes, G.M., Carrigan, P.E., and Miller, L.J. (2007). Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. *Cancer Res* 67(5)**,** 2072-2080. doi: 10.1158/0008-5472.Can-06-2969.

Hector, R.F., Zimmer, B.L., and Pappagianis, D. (1990). Evaluation of nikkomycins X and Z in murine models of coccidioidomycosis, histoplasmosis, and blastomycosis. *Antimicrob Agents Chemother* 34(4)**,** 587-593. doi: 10.1128/aac.34.4.587.

Hogan, D.A., and Sundstrom, P. (2009). The Ras/cAMP/PKA signaling pathway and virulence in *Candida albicans*. *Future Microbiol* 4(10)**,** 1263-1270. doi: 10.2217/fmb.09.106.

Karakama, Y., Sakamoto, N., Itsui, Y., Nakagawa, M., Tasaka-Fujita, M., Nishimura-Sakurai, Y., et al. (2010). Inhibition of hepatitis C virus replication by a specific inhibitor of serine-arginine-rich protein kinase. *Antimicrobial agents and chemotherapy* 54(8)**,** 3179-3186. doi: 10.1128/aac.00113-10.

Li, Y., Su, C., Mao, X., Cao, F., and Chen, J. (2007). Roles of *Candida albicans* Sfl1 in hyphal development. *Eukaryot Cell* 6(11)**,** 2112-2121. doi: 10.1128/ec.00199-07.

Li, Y., Sun, L., Lu, C., Gong, Y., Li, M., and Sun, S. (2018). Promising Antifungal Targets Against *Candida albicans* Based on Ion Homeostasis. *Front Cell Infect Microbiol* 8**,** 286. doi: 10.3389/fcimb.2018.00286.

Liu, S., Kang, K., Zhang, J., Ouyang, Q., Zhou, Z., Tian, S., et al. (2009). A novel *Physarum polycephalum* SR protein kinase specifically phosphorylates the RS domain of the human SR

protein, ASF/SF2. *Acta Biochim Biophys Sin (Shanghai)* 41(8)**,** 657-667. doi: 10.1093/abbs/gmp054.

Llopis-Torregrosa, V., Hušeková, B., and Sychrová, H. (2016). Potassium Uptake Mediated by Trk1 Is Crucial for *Candida glabrata* Growth and Fitness. *PLoS One* 11(4)**,** e0153374. doi: 10.1371/journal.pone.0153374.

Martin, R., Moran, G.P., Jacobsen, I.D., Heyken, A., Domey, J., Sullivan, D.J., et al. (2011). The *Candida albicans*-specific gene *EED1* encodes a key regulator of hyphal extension. *PLoS One* 6(4)**,** e18394. doi: 10.1371/journal.pone.0018394.

Martinez, D.A., Oliver, B.G., Gräser, Y., Goldberg, J.M., Li, W., Martinez-Rossi, N.M., et al. (2012). Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. *mBio* 3(5)**,** e00259-00212. doi: 10.1128/mBio.00259-12.

Mehta, R.J., Kingsbury, W.D., Valenta, J., and Actor, P. (1984). Anti-*Candida* activity of polyoxin: example of peptide transport in yeasts. *Antimicrobial Agents and Chemotherapy* 25(3)**,** 373-374. doi: doi:10.1128/AAC.25.3.373.

Moehle, E.A., Ryan, C.J., Krogan, N.J., Kress, T.L., and Guthrie, C. (2012). The yeast SR-like protein Npl3 links chromatin modification to mRNA processing. *PLoS Genet* 8(11)**,** e1003101. doi: 10.1371/journal.pgen.1003101.

Monge, R.A., Román, E., Nombela, C., and Pla, J. (2006). The MAP kinase signal transduction network in *Candida albicans*. *Microbiology (Reading)* 152(Pt 4)**,** 905-912. doi: 10.1099/mic.0.28616-0.

Muñoz, J.F., Gade, L., Chow, N.A., Loparev, V.N., Juieng, P., Berkow, E.L., et al. (2018). Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species. *Nature Communications* 9(1)**,** 5346. doi: 10.1038/s41467-018-07779-6.

Nix, D.E., Swezey, R.R., Hector, R., and Galgiani, J.N. (2009). Pharmacokinetics of nikkomycin Z after single rising oral doses. *Antimicrob Agents Chemother* 53(6)**,** 2517-2521. doi: 10.1128/aac.01609-08.

Odds, F.C. (1988). *Candida and Candidosis.* Elsevier Science Health Science Division.

Odunsi, K., Mhawech-Fauceglia, P., Andrews, C., Beck, A., Amuwo, O., Lele, S., et al. (2012). Elevated expression of the serine-arginine protein kinase 1 gene in ovarian cancer and its role in Cisplatin cytotoxicity *in vitro*. *PLoS One* 7(12)**,** e51030. doi: 10.1371/journal.pone.0051030.

Payne, J.W., and Shallow, D.A. (1985). Studies on drug targeting in the pathogenic fungus *Candida albicans*: peptide transport mutants resistant to polyoxins, nikkomycins and bacilysin. *FEMS Microbiology Letters* 28(1)**,** 55-60. doi: 10.1111/j.1574-6968.1985.tb00763.x.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47(D1)**,** D442-d450. doi: 10.1093/nar/gky1106.

Ramírez-Zavala, B., Mottola, A., Haubenreißer, J., Schneider, S., Allert, S., Brunke, S., et al. (2017). The Snf1-activating kinase Sak1 is a key regulator of metabolic adaptation and *in vivo* fitness of *Candida albicans*. *Mol Microbiol* 104(6)**,** 989-1007. doi: 10.1111/mmi.13674.

Reuss, O., and Morschhäuser, J. (2006). A family of oligopeptide transporters is required for growth of *Candida albicans* on proteins. *Mol Microbiol* 60(3)**,** 795-812. doi: 10.1111/j.1365-2958.2006.05136.x.

Reuss, O., Vik, A., Kolter, R., and Morschhäuser, J. (2004). The *SAT1* flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene* 341**,** 119-127. doi: 10.1016/j.gene.2004.06.021.

Sanglard, D., Ischer, F., Monod, M., and Bille, J. (1996). Susceptibilities of *Candida albicans* multidrug transporter mutants to various antifungal agents and other metabolic inhibitors. *Antimicrob Agents Chemother* 40(10)**,** 2300-2305. doi: 10.1128/aac.40.10.2300.

Schwarzmüller, T., Ma, B., Hiller, E., Istel, F., Tscherner, M., Brunke, S., et al. (2014). Systematic phenotyping of a large-scale *Candida glabrata* deletion collection reveals novel antifungal tolerance genes. *PLoS Pathog* 10(6)**,** e1004211. doi: 10.1371/journal.ppat.1004211.

Shen, H., An, M.M., Wang de, J., Xu, Z., Zhang, J.D., Gao, P.H., et al. (2007). Fcr1p inhibits development of fluconazole resistance in *Candida albicans* by abolishing *CDR1* induction. *Biol Pharm Bull* 30(1)**,** 68-73. doi: 10.1248/bpb.30.68.

Siebel, C.W., Feng, L., Guthrie, C., and Fu, X.D. (1999). Conservation in budding yeast of a kinase specific for SR splicing factors. *Proc Natl Acad Sci U S A* 96(10)**,** 5440-5445. doi: 10.1073/pnas.96.10.5440.

Sieber, P., Voigt, K., Kämmer, P., Brunke, S., Schuster, S., and Linde, J. (2018). Comparative Study on Alternative Splicing in Human Fungal Pathogens Suggests Its Involvement During Host Invasion. *Front Microbiol* 9**,** 2313. doi: 10.3389/fmicb.2018.02313.

Takeuchi, M., and Yanagida, M. (1993). A mitotic role for a novel fission yeast protein kinase dsk1 with cell cycle stage dependent phosphorylation and localization. *Mol Biol Cell* 4(3)**,** 247-260. doi: 10.1091/mbc.4.3.247.

Tang, Z., Tsurumi, A., Alaei, S., Wilson, C., Chiu, C., Oya, J., et al. (2007). Dsk1p kinase phosphorylates SR proteins and regulates their cellular localization in fission yeast. *Biochem J* 405(1)**,** 21-30. doi: 10.1042/bj20061523.

Tang, Z., Yanagida, M., and Lin, R.J. (1998). Fission yeast mitotic regulator Dsk1 is an SR protein-specific kinase. *J Biol Chem* 273(10)**,** 5963-5969. doi: 10.1074/jbc.273.10.5963.

Vaas, L.A.I., Sikorski, J., Hofner, B., Fiebig, A., Buddruhs, N., Klenk, H.-P., et al. (2013). opm: an R package for analysing OmniLog® phenotype microarray data. *Bioinformatics* 29(14)**,** 1823-1824. doi: 10.1093/bioinformatics/btt291.

Vehkala, M., Shubin, M., Connor, T.R., Thomson, N.R., and Corander, J. (2015). Novel R Pipeline for Analyzing Biolog Phenotypic Microarray Data. *PLOS ONE* 10(3)**,** e0118392. doi: 10.1371/journal.pone.0118392.

Wang, G., Sun, P., Gong, Z., Gu, L., Lou, Y., Fang, W., et al. (2018). Srk1 kinase, a SR protein-specific kinase, is important for sexual reproduction, plant infection and pre-mRNA processing in *Fusarium graminearum*. *Environ Microbiol* 20(9)**,** 3261-3277. doi: 10.1111/1462-2920.14299.

Wang, P., Zhou, Z., Hu, A., Ponte de Albuquerque, C., Zhou, Y., Hong, L., et al. (2014). Both decreased and increased SRPK1 levels promote cancer by interfering with PHLPP-mediated dephosphorylation of Akt. *Mol Cell* 54(3)**,** 378-391. doi: 10.1016/j.molcel.2014.03.007.

Wessel, D., and Flügge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 138(1)**,** 141-143. doi: 10.1016/0003-2697(84)90782-6.

Yadan, J.C., Gonneau, M., Sarthou, P., and Le Goffic, F. (1984). Sensitivity to nikkomycin Z in *Candida albicans*: role of peptide permeases. *J Bacteriol* 160(3)**,** 884-888. doi: 10.1128/jb.160.3.884-888.1984.

Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* 6(5)**,** 837-847. doi: 10.1101/gad.6.5.837.

Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122(3)**,** 191-207. doi: 10.1007/s00412-013-0407-z.

Figure legends

Figure 1. Protein alignment of SRPK homologs: *H. sapiens* SR protein kinase 1, *S. cerevisiae*, *C. glabrata* and *C. albicans*. The annotation of the protein features (the bars above the corresponding alignment section) is based on the annotation of the human SRPK 1 (NP_003128.3). **(A)** Overview alignment of the entire protein, where the kinase domains are indicated by the grey bars above. **(B)** Detailed alignment of the kinase domain (grey bar above) intercepted by the 'spacer' region. The protein kinase domain contains ATP binding sites (turquoise), activation loop (magenta) and kinase-docking sites (yellow). All regions required for SRPK1 function are highly conserved among the fungal species. The conserved alignment sites are highlighted within the alignment applying Clustal X default coloring.

Figure 2. Growth of *S. cerevisiae* and *C. glabrata sky1*Δ mutants, and of *C. albicans sky1*Δ and *sky2*Δ mutants on high salt and toxic spermine levels. YPD overnight cultures of the strains were adjusted to an optical density (OD_{600}) of 1.0. For *C. albicans* and *C. glabrata* two independently generated mutants were tested. Serial 10-fold dilutions were spotted on YPD agar plates containing the indicated stressor and incubated for 2-3 days at 30°C (*S. cerevisiae*) or 37°C (*C. albicans* and *C. glabrata*). *S. cerevisiae*/*C. glabrata*/*C. albicans* plates contained NaCl (1 M/1.7 M/1.8 M), LiCl (0.3 M/0.18 M/0.5 M) or spermine (2 mM/8 mM/12.8 mM).

Figure 3. *C. albicans* Sky2 is required for growth on various dipeptides as the sole nitrogen source. The metabolic activities of the *C. albicans* SC5314 wild type, Ca*sky1*Δ and Ca*sky2*Δ mutant strains were measured kinetically every 15 minutes for 24 h at 30°C by utilizing Biolog™ phenotypic microarray plates for fungi. Upper panel: Each growth signal was categorized as active (purple) or non-active (yellow) growth Lower panel: ANOVA and Tukey's test were applied to detect significantly differences in growth (log2 fold change) between the strains.

Figure 4. CaSky2 is important for utilization of selected dipeptides as the sole nitrogen source. YPD overnight cultures were adjusted to an optical density (OD_{600}) of 0.01 in YCB medium containing 10 mM of the indicated dipeptide as a nitrogen source and incubated at 37°C. The OD_{600} was measured after 24 h and 48 h. Both the A strain and the B strain of the mutant strains are displayed. The values shown are the calculated mean and standard deviation of 3 biological replicates. For statistical analysis, a two-way ANOVA was performed (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; compared to SC5314).

Figure 5. Phosphoproteome analysis revealed a distinct set of putative substrates for *C. albicans* Sky1 and Sky2. YPD overnight cultures were adjusted to an optical density (OD_{600}) of 0.1 in YPD medium and incubated at 37°C for 4 h. Cells were harvested, lysed and proteins were tryptically digested, followed by phosphopeptide enrichment based on TiO2/ZrO2 resin (a non-enriched fraction served as reference). Three biological replicates were used per strain. **(A)** Overview of the identified proteins, phosphoproteins and phosphosites. **(B)** Phosphoprotein abundance of *sky1*Δ vs. WT and *sky2*Δ vs WT. The values represent all identified phosphoproteins with an abundance ratio of >log4 and p<0.05.

Figure 6. *PTR22* overexpression rescued the *sky2*Δ mutant phenotype. YPD overnight cultures were adjusted to an optical density (OD_{600}) of 0.01 in YCB medium containing 10 mM of the indicated dipeptide as a nitrogen source and incubated at 37°C. The OD_{600} was measured after 24 h and 48 h. The values shown are the calculated mean and standard deviation of 3 biological replicates. For statistical analysis, a two-way ANOVA was performed (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; compared to SC5314). **(A)** The growth of the wild type and the A mutant strains is shown. **(B)** The growth of the wild type and the B mutant strains is shown.

Figure 7. Strains lacking *SKY2* showed resistance to the antibiotics Nikkomycin Z and Polyoxin D. Growth of the wild type strain SC5314 and mutant strains in SD medium and SD medium supplement with either 140 µg/ml Nikkomycin Z or 130 µg/ml Polyoxin D. YPD overnight cultures were adjusted to an optical density (OD_{600}) of 0.02 in the respective media and incubated at 37°C for 24 h. The growth (OD_{600}) was determined every 20 min at an optical density of 600 nm. Shown is the mean of each strain calculated from 3 technical replicates. For a better overview, the standard deviation is not shown. **(A)** The growth of the wild type and the A mutant strains is shown. **(B)** The growth of the wild type and the B mutant strains is shown.

Figure 8. Transcript levels of the dipeptide/tripeptide and oligopeptide transporter genes following growth on dipeptides as the sole nitrogen source. YPD overnight cultures of the wild type strain and the s*ky2*Δ mutant strains were adjusted to an optical density (OD_{600}) of 0.1 in YCB medium containing 10 mM of **(A)** Ala-Phe, **(B)** Phe-Ser or **(C)** Met-Gly and incubated at 37°C for 4 h. Cells were harvest and used for preparation of total RNA. The transcript levels of the dipeptide/tripeptide transporter genes *PTR2* and *PTR22,* and the oligopeptide transporter genes *OPT1*, *OPT2*, *OPT3*, *OPT4* and *OPT5* were assessed in biological and technical triplicates by qRT-PCR. Each bar represents the mRNA level of the indicated gene expressed as means ± SEM relative to the *MED15* expression. A two-way ANOVA was used to calculate the statistical relevance: (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).

Supplementary Figure 1. BLASTp of the human SRPK1 protein (NCBI Reference Sequence: NP_003128.3). Except for *C. glabrata* the most common pathogenic *Candida* spp. possess two SRPK homologs, all having a high level of homology to *S. cerevisiae* Sky1 protein kinase. [species |strain |GenBank number |model name on CGD]

Supplementary Figure 2: Deletion of *HRK1* confers resistance to high salt and toxic spermine levels. YPD overnight cultures of the strains were adjusted to an optical density (OD_{600}) of 1.0. For *hrk1*Δ two independently generated mutants were tested. Serial 10-fold dilutions were spotted on YPD agar plates containing the indicated stressor and incubated for 2 days at 37°C. Plates contained 1.8M NaCl, 0.3 M LiCl or 12.8 mM spermine.

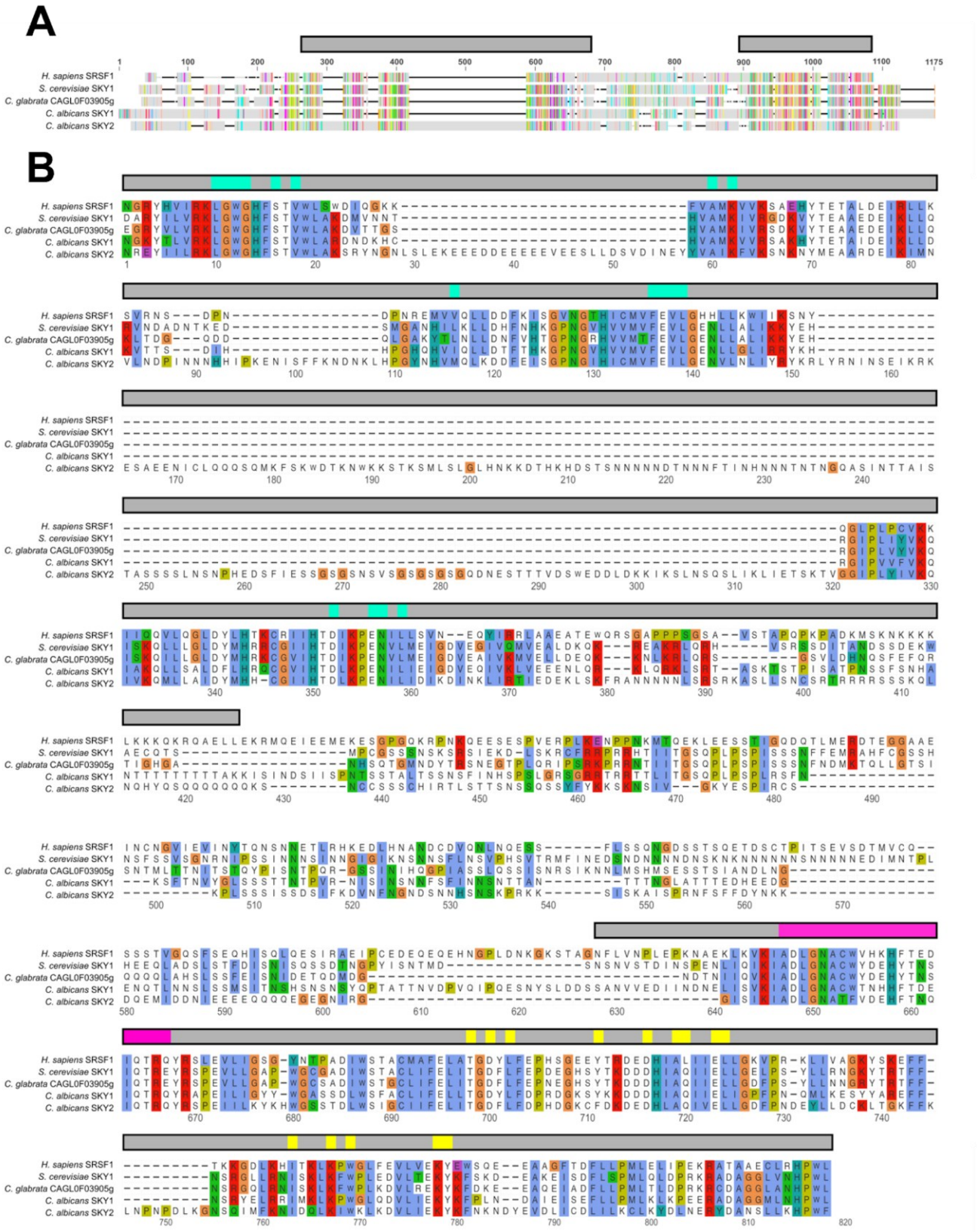

A
B
H. sapiens SRSF1
S. cerevisiae SKY1
C. glabrata CAGL0F03905g
C. albicans SKY1
C. albicans SKY2

Figure 1

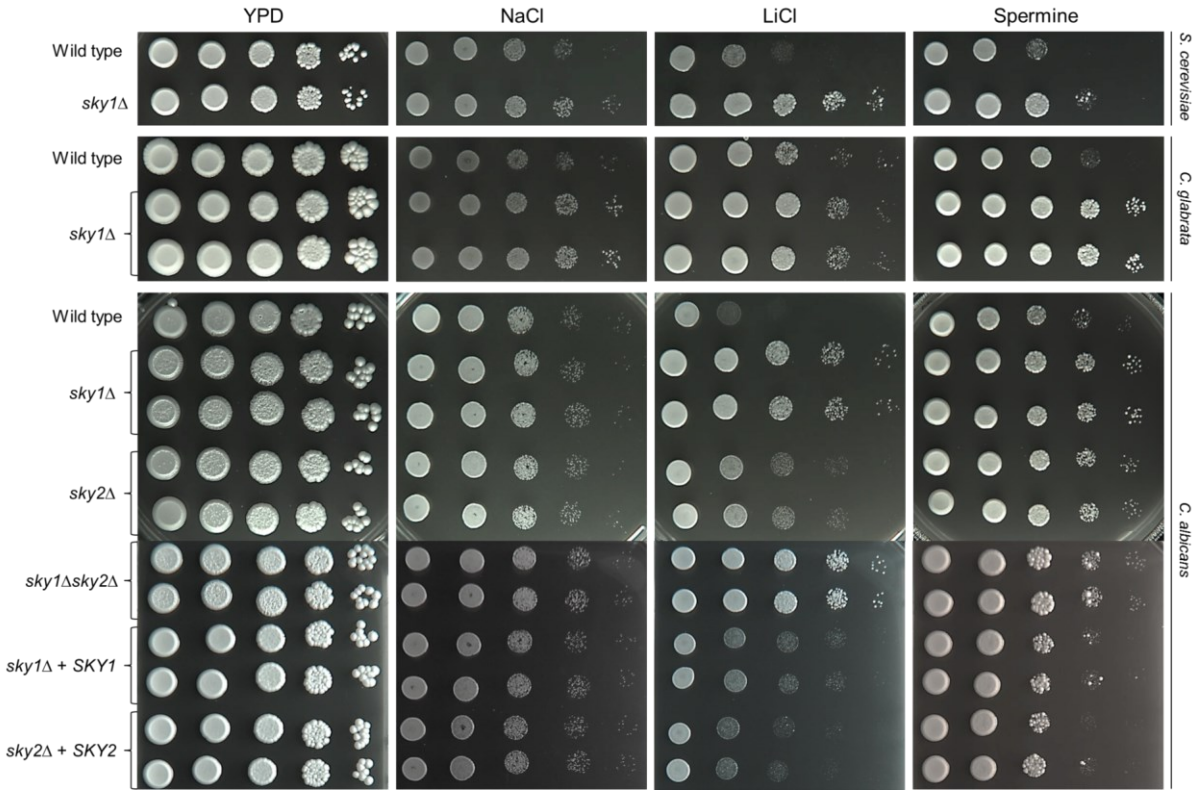


Figure 2

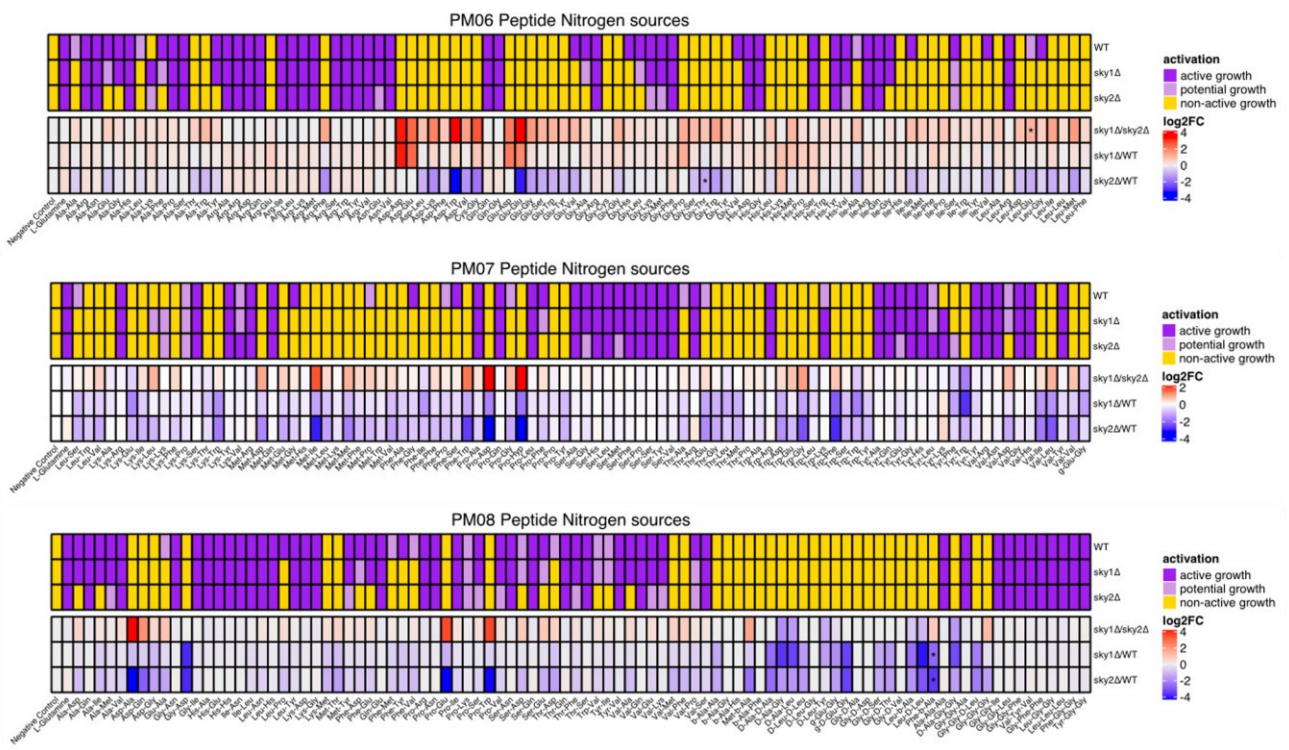


Figure 3

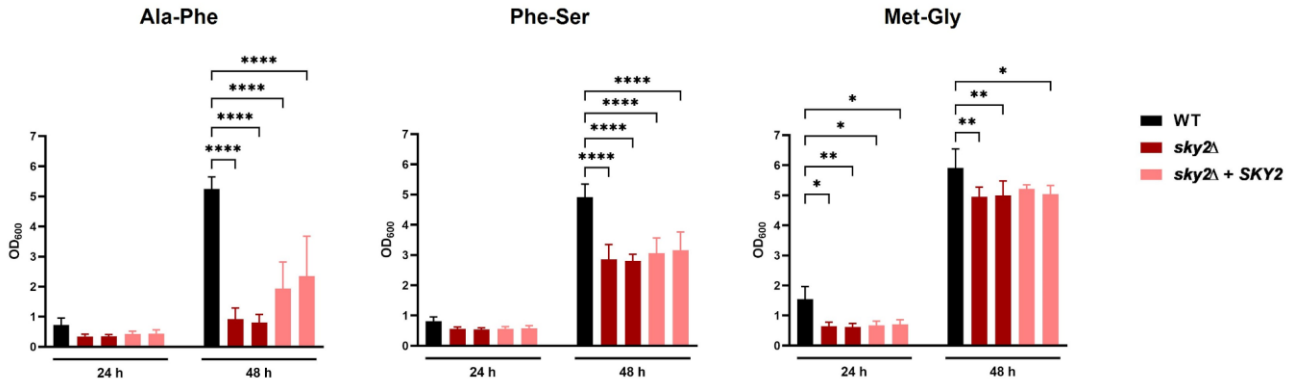


Figure 4

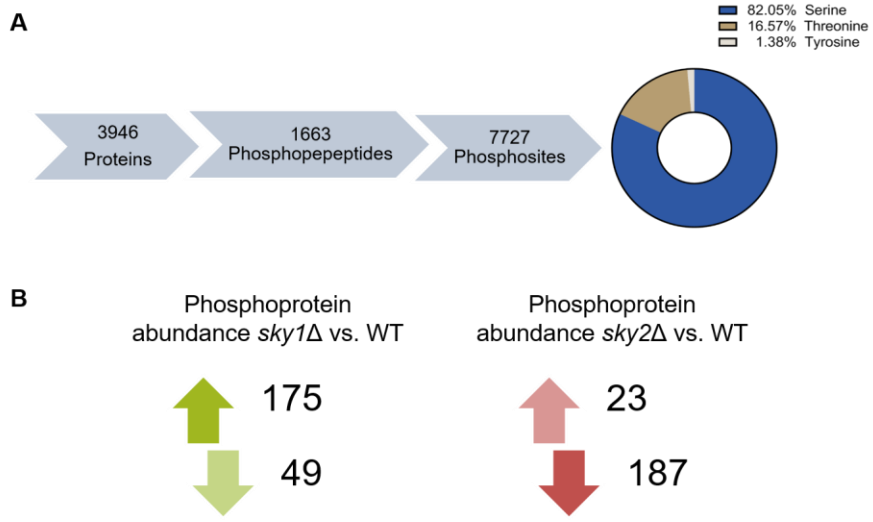


Figure 5

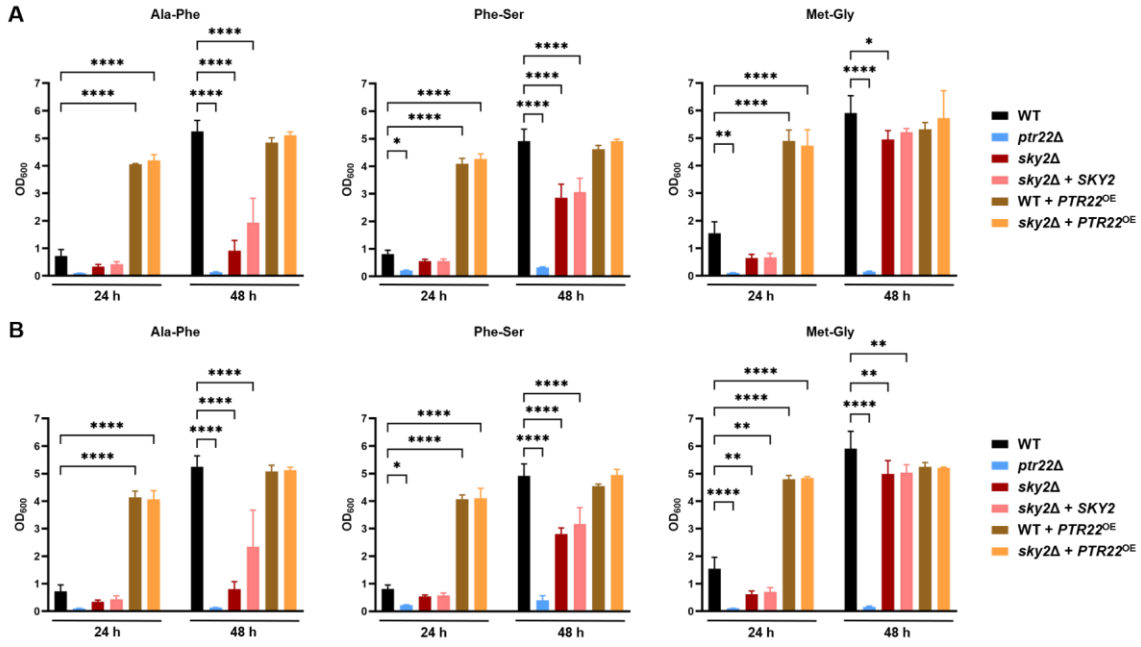


Figure 6

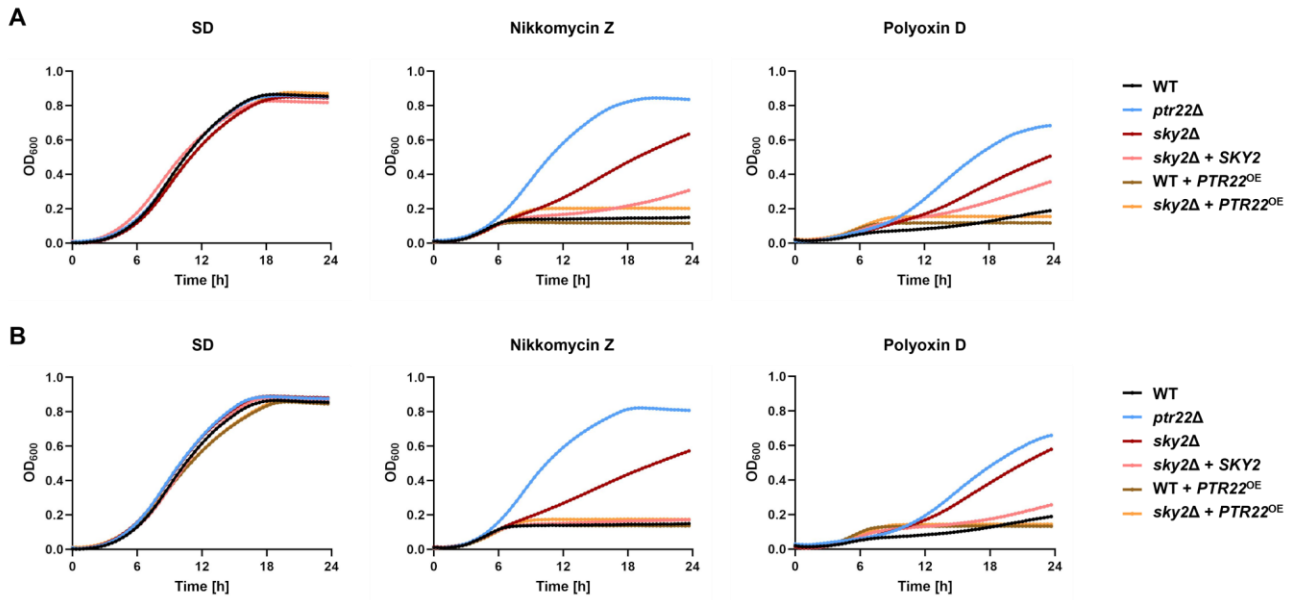

A
SD
Nikkomycin Z
Polyoxin D
OD600
Time [h]
WT
ptr22Δ
sky2Δ
sky2Δ + SKY2
WT + PTR22OE
sky2Δ + PTR22OE
B
SD
Nikkomycin Z
Polyoxin D
OD600
Time [h]
WT
ptr22Δ
sky2Δ
sky2Δ + SKY2
WT + PTR22OE
sky2Δ + PTR22OE

Figure 7

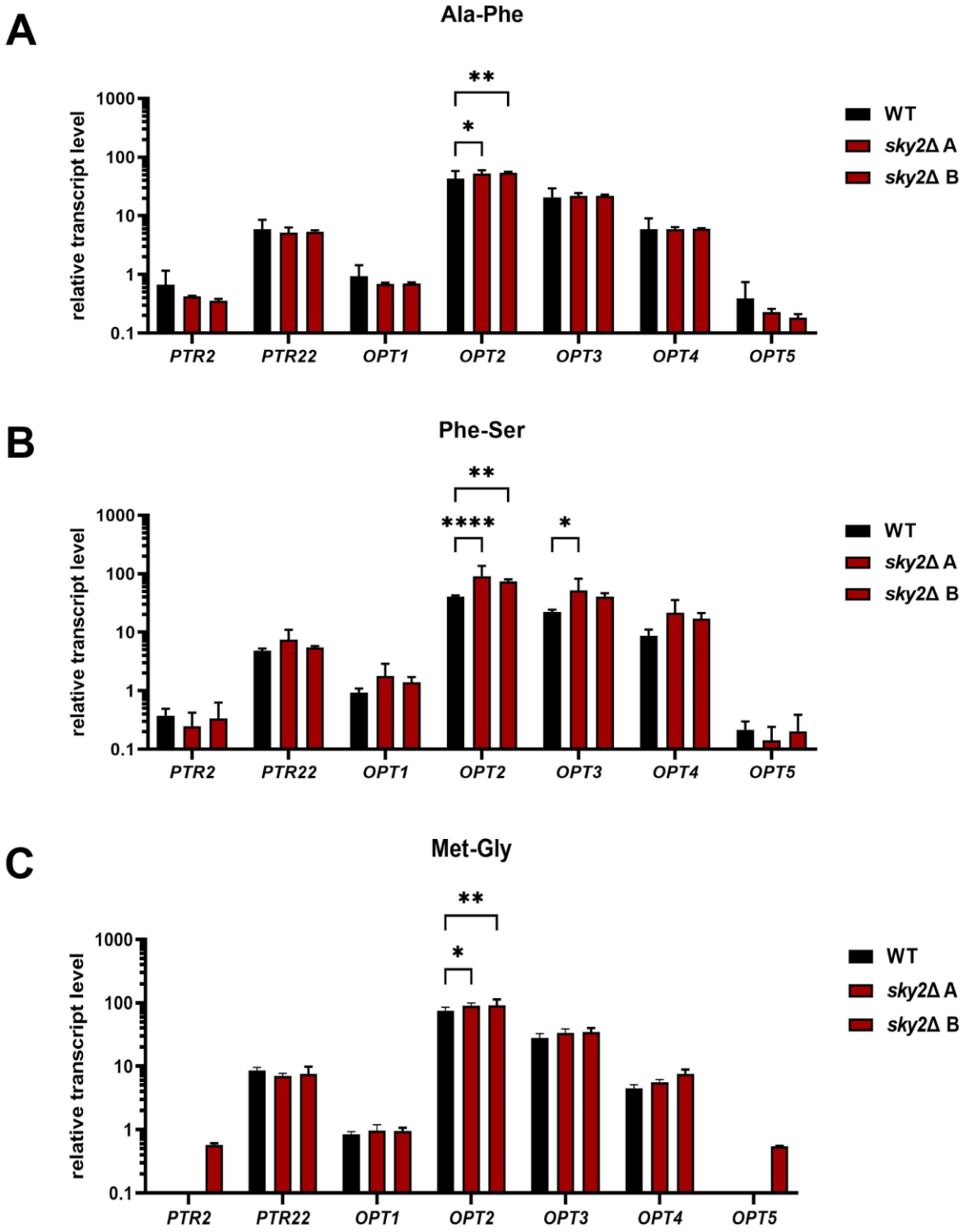
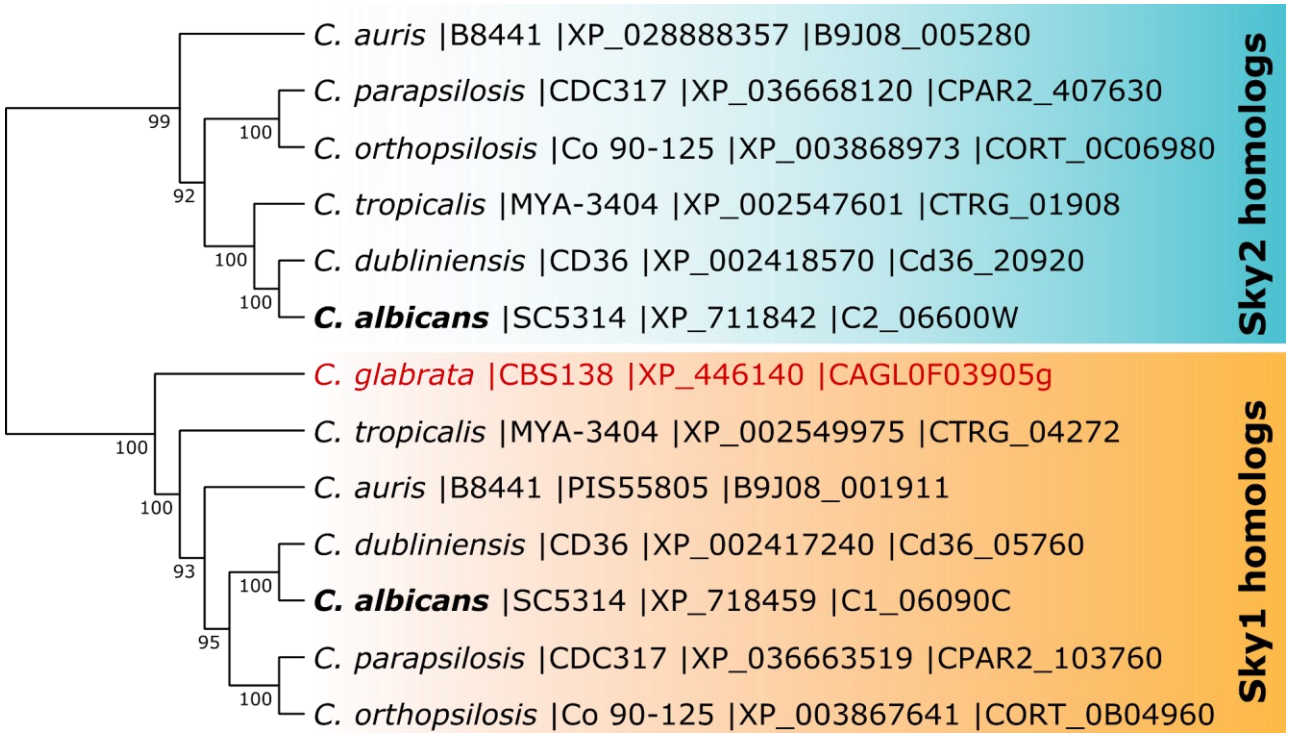
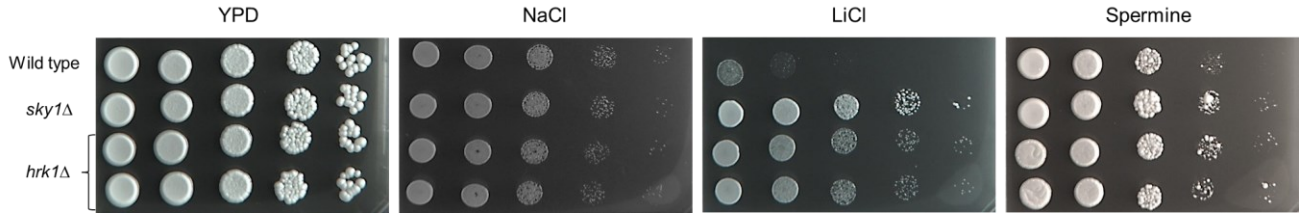

A
Ala-Phe
relative transcript level
1000
100
10
1
0.1
**
*
PTR2
PTR22
OPT1
OPT2
OPT3
OPT4
OPT5
WT
sky2Δ A
sky2Δ B
B
Phe-Ser
relative transcript level
1000
100
10
1
0.1
**
****
*
PTR2
PTR22
OPT1
OPT2
OPT3
OPT4
OPT5
WT
sky2Δ A
sky2Δ B
C
Met-Gly
relative transcript level
1000
100
10
1
0.1
**
*
PTR2
PTR22
OPT1
OPT2
OPT3
OPT4
OPT5
WT
sky2Δ A
sky2Δ B

Figure 8



Supplementary figure 1



Supplementary figure 2

No 5. *Lactobacillus rhamnosus* colonization antagonizes *Candida albicans* by forcing metabolic adaptations that compromise pathogenicity

FORM 1

Manuscript No. 5

Manuscript title: Lactobacillus rhamnosus colonisation antagonizes Candida albicans by forcing metabolic adaptations that compromise pathogenicity

Authors: Raquel Alonso-Roman, Antonia Last, **Mohammad H. Mirhakkak**, Jakob L. Sprague, Lars Möller, Peter Großmann, Katja Graf, Rena Gratz, Selene Mogavero, Slavena Vylkova, Gianni Panagiotou, Sascha Schäuble, Bernhard Hube, Mark S. Gresnigt

Bibliographic information (if published or accepted for publication: Citation):

The candidate is (Please tick the appropriate box.)

☐ First author, ☐ Co-first author, ☐ Corresponding author, ☒ Co-author.

Status: revised version submitted to Nature communications

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Alonso-Roman, R.	10%	25%	60%	30%	
Last, A.		5%	30%	20%	
Mirhakkak, M.H.		40%		10%	
Schäuble, S.	20%	10%		10%	
Panagiotou, G.					5%
Hube, B.	20%				70%
Gresnigt, M.S.	50%	20%		30%	5%
Others			10%		20%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

FORM 2

Manuscript No. 5

Short reference Alonso-Roman *et al*., (revision version submitted to Nature communications)

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures: My contribution in the publication is *in silico* constraint-based metabolic modeling for the triangle of *Candida albicans*, *Lactobacillus rhamnosus*, and Intestinal epithelial cells.

Figures # S3, S5	☒	**100%**
Figure # 2	☒	Approximate contribution: **30%** (Figure parts D and G)
Figure # 3	☒	Approximate contribution: **20%** (Figure part E)
Figure # 6	☒	Approximate contribution: **50%** (Flux variability analysis)
Figures # 1, 4, 5, 7, S1, S2, S4, S6-S8	☐	**0%**

(Add more table boxes depending on the number of figures)

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

Lactobacillus rhamnosus colonisation antagonizes *Candida albicans* by forcing metabolic adaptations that compromise pathogenicity

Raquel Alonso-Roman[1], Antonia Last[1], Mohammad H. Mirhakkak[2], Jakob L. Sprague[1], Lars Möller[1], Peter Großmann[2], Katja Graf[1,3], Rena Gratz[1], Selene Mogavero[1], Slavena Vylkova[4], Gianni Panagiotou[2,5], Sascha Schäuble[2], Bernhard Hube[1,6], Mark S. Gresnigt[7]
Corresponding author: bernhard.hube@hki-jena.de

Affiliations:
1: Department of Microbial Pathogenicity Mechanisms, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Jena, Germany
2: Systems Biology and Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Jena, Germany
3: Dynamic42 GmbH, Jena, Germany
4: Septomics Research Centre, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knoell-Institute, Jena, Germany
5: Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, China
6: Institute of Microbiology, Friedrich Schiller University, Jena, Germany
7: Junior Research Group Adaptive Pathogenicity Strategies, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Jena, Germany

Key words: *Candida albicans*, *Lactobacillus rhamnosus*, fungal-host interactions, transcriptome, metabolome, *in silico* genome scale metabolic modelling

4. ABSTRACT

Intestinal microbiota dysbiosis can initiate overgrowth of commensal *Candida* species – a major predisposing factor for disseminated candidiasis. Normally, commensal bacteria such as *Lactobacillus rhamnosus* can antagonize *C. albicans* pathogenicity. To uncover the molecular mechanisms of bacterial antagonism, we investigated the interplay between *C. albicans*, *L. rhamnosus,* and intestinal epithelial cells (IECs) by integrating transcriptional and metabolic profiling, and reverse genetics.
Untargeted metabolomics and *in silico* modelling suggested that IECs foster bacterial growth metabolically, leading to production of antivirulence compounds. The bacterial growth on IECs also modified the metabolic environment, which included the removal of *C. albicans'* favoured nutrient sources. Consequently, *C. albicans* was forced to transcriptionally rewire its metabolism and cell biology, this was associated with the reprogramming of defined virulence-related genes.
This suggests that intestinal colonization with bacteria can antagonize *C. albicans* by reshaping the metabolic environment forcing metabolic adaptations that reduce fungal pathogenicity.

5. INTRODUCTION

The yeast *Candida albicans* is a common commensal of the intestinal mycobiota[1,2]. However, certain predisposing conditions, such as a dysbalanced microbiota and a compromised immune system can favour its shift from a commensal to a pathogenic stage[3]. Consequently, *C. albicans* can translocate through the intestinal epithelium into the bloodstream[4,5], resulting in disseminated candidiasis associated with high mortality rates[6]. Specifically, an intact epithelial barrier, the immune system, and the commensal intestinal microbiota play crucial roles in maintaining *C. albicans* commensalism by suppressing overgrowth and pathogenicity[3,7,8]. Bacteria belonging to different genera, such as *Bifidobacterium*, *Streptococcus* and, more extensively, *Lactobacillus,* have been studied with regard to their antagonistic effects against *C. albicans*[9,10].

Lactobacillus species have been discussed as oral probiotics to prevent candidemia in premature newborns[11,12] and are also being explored as potential therapies against vulvovaginal candidiasis[13–15]. *In vitro* studies have demonstrated that lactobacilli can antagonize proliferation[9], hypha formation[16–21], biofilm formation[18,20–23], or even kill *C. albicans*[24,25]. However, these studies often do not recapitulate the role of the host or its epithelial barriers as active partners in the interaction. Previously, we demonstrated that the antagonistic effects of *L. rhamnosus* on intestinal epithelial cells (IECs) are associated with reduced *C. albicans* invasion, and damage[26]. *L. rhamnosus* colonization of IECs prevented fungal outgrowth and reduced the number of host cell-associated *C. albicans* cells through shedding[26]. Similarly, *L. rhamnosus* colonization in an intestine-on-chip-model reduced the fungal burden, epithelial damage, and fungal translocation into a surrogate bloodstream compartment[27].

However, the molecular mechanisms underlying the antagonistic effects of *L. rhamnosus* colonization against *C. albicans* pathogenicity remain, to a large extent, unresolved. Several antagonistic mechanisms of lactobacilli have been described using models excluding host cells, such as peroxide production[28] or lactic and acetic acid production[25,29]. However, these antagonistic mechanisms did not play a role in the protection of host cells[26]. More complex *in vitro* models are required to dissect the interactions between the host, lactobacilli, and *C. albicans* in a physiologically relevant manner to understand the principles of the observed protective activities.

We took a systems-biology approach involving multi-omics (transcriptome/metabolome) profiling, *in silico* metabolic modelling, and *in vitro* infection biology to uncover how *L. rhamnosus* colonization of IECs mediates protection against *C. albicans* infection. This included dissection of the metabolic crosstalk between epithelial cells, *L. rhamnosus*, and *C. albicans.* Using this approach, we elucidated how multifactorial changes in the metabolic environment force *C. albicans* transcriptional adaptation, resulting in a reduced capacity to damage epithelial cells.

6. RESULTS

L. rhamnosus suppresses C. albicans pathogenicity

We recently showed that *L. rhamnosus* reduced the number of *C. albicans* cells in contact with IECs[26], which may reduce invasion and damage of the tissue. Infections with a lower number of *C. albicans* cells also resulted in a reduced fungal burden in the tissue after infection **(Fig. 1A)**. Interestingly, this reduced fungal burden in the absence of *L. rhamnosus* still induced more damage than when lactobacilli were present **(Fig. 1B)**, demonstrating that reduced numbers of host-cell associated *C. albicans* alone are insufficient to mediate protection.
We previously showed that killed *L. rhamnosus* failed to protect IECs against *C. albicans* induced damage[26]. To confirm that colonization is only protective when lactobacilli are metabolically active and proliferating, the bacteria were killed using antibiotics four hours post infection (hpi). Killing *L. rhamnosus* after colonization still abolished their protective effect **(Fig. 1C)**. However, physical separation of *L. rhamnosus* from IECs and *C. albicans* did not impair the protection **(Fig. 1D)**, suggesting that the environment plays a key role.

Host-bacterial-fungal metabolic crosstalk

A detailed insight into the metabolic interactions between *C. albicans*, *L. rhamnosus*, and IECs was obtained using untargeted metabolomics. In the supernatants, 235 metabolites were identified, and their relative abundances were quantified. From these metabolites, 89 were present in the culture media, suggesting that the remaining metabolites are derived from metabolic activity of IECs, *C. albicans*, and *L. rhamnosus.*
The metabolic environment slightly differed between uninfected IECs and *C. albicans*-infected IECs **(Fig. 2A I)**. However, at 6 and 12 hpi *L. rhamnosus* colonization of IECs removed 53 and 46 metabolites and resulted in the appearance of 73 and 74 metabolites, compared to uninfected IECs, respectively **(Fig. 2A II, Fig. 2B)**. Interestingly, during *C. albicans* infection of *L. rhamnosus*-colonized IECs only 12 metabolites were altered. **(Fig. 2A IIII, Fig. 2B).** Unsupervised hierarchical clustering of the metabolome revealed distinct clusters of metabolites in the model at 6 hpi **(Fig. 2C, Tab. S1)** and 12 hpi **(Fig. S1, Tab. S2)**, which are associated with specific dynamics. At 6 hpi, cluster 2 and 3 **(Fig. 2C)** represent metabolites produced by IECs. The metabolites in cluster 3, however, are no longer present following *L. rhamnosus* colonization, suggesting that they are consumed by *L. rhamnosus.* Metabolites in cluster 1 were only present when *L. rhamnosus* colonized IECs, and thus may be involved in their protective effect. Similar clusters were identified at 12 hpi **(Fig. S1)**. Enrichment analysis showed that at both 6 and 12 hpi, cluster 1 was enriched in metabolites involved in pyrimidine metabolites. At 12 hpi, cluster 4 (cluster 3 at 6 hpi) was enriched in metabolites for valine, leucine and isoleucine biosynthesis **(Fig. S2)**.

Intestinal epithelial cells foster L. rhamnosus growth

L. rhamnosus cannot grow in KBM unless IECs are present, yet active growth is essential for their antagonistic potential **(Fig. 1C)**[26]**.** Physical separation of *L. rhamnosus* from the epithelial cells did not compromise protection **(Fig. 1D)** nor bacterial growth **(Fig. 2D).** Likewise, IEC-spent medium supported *L. rhamnosus* growth **(Fig. 2D)**, validating that epithelial-secreted metabolites foster bacterial growth. By simulating *L. rhamnosus* biomass formation *in silico* with flux balance analysis (FBA)[30], IEC supernatants were predicted to sustain an increased bacterial biomass, while the culture medium (KBM) did not **(Fig. 2E, supplementary data 2).** Thus, the IEC-derived metabolites, which disappear upon colonization **(Fig. 2C cluster 3, Tab.S1)**, are likely those that foster *L. rhamnosus* growth. From the different metabolite types in this cluster at least one representative was tested *in vitro* for its capacity to foster *L. rhamnosus* growth. Supplementation with fructose, xanthine, nicotinamide, nicotinic acid, 3-methyl-2-oxobutyrate, 3-methyl-2-oxovalerate, N-acetylgalactosamine, carnosine, and amino acids did not promote growth individually **(Fig. 2F)**.

Carnitine and citric acid supplementation slightly supported *L. rhamnosus* growth **(Fig. 2F)**, whereas combinations of citric acid with gamma-glutamylalanine or carnitine, and especially all three in combination, supported *L. rhamnosus* growth to rates similar as on IECs **(Fig. 2F)**.
Additionally, reconstructed genome-scale metabolic models (GEMs) were simulated for IECs, *C. albicans,* or *L. rhamnosus*[31,32], monitoring both biomass production and feasible reaction fluxes by again applying FBA and also flux variability[33] analysis (FVA, **supplementary data 2**). Since IECs are persistently present in the infection model, GEMs were parametrized with supernatant-specific metabolites. We investigated the GEMs used for IECs, *C. albicans*, and *L. rhamnosus* to uncover reaction flux shifts in individual metabolic pathways across different nutritional conditions including blank media and supernatants of IECs, *L. rhamnosus* or *C. albicans* **(Fig. 2G; Fig. S3)**. The *in silico* simulation using FVA revealed markedly altered *L. rhamnosus* metabolic pathway activities on IEC supernatants *vs.* culture medium, including nucleotide, Nicotinamide adenine dinucleotide (NAD), lipid-related metabolism, and Coenzyme A (CoA) synthesis. In addition, diverse amino acid metabolic pathways showed differing feasible flux ranges in response to IEC-derived metabolite profiles **(Fig. 2G)**. In comparison, changes in metabolic pathway activity were less prevalent in IECs **(Fig. S3A).** Here, only a few amino acids (including tyrosine and phenylalanine), ubiquinone and taurine pathways next to the generic protein assembly/degradation metabolic subsystem showed shifts in pathway activity of at least 40% upon *L. rhamnosus* affected supernatants. These simulation results suggest that *L. rhamnosus* utilizes IEC-secreted metabolites without inducing drastic metabolic changes in the host. These results did not change upon varying thresholds for required flux activity changes in the compared conditions **(supplementary data 3).**

L. rhamnosus-derived metabolites antagonize C. albicans

L. rhamnosus-mediated protection against *C. albicans* cytotoxicity is contact-independent **(Fig. 1D)**, suggesting that soluble molecules mediate the antagonistic effects. To test this hypothesis, *L. rhamnosus* was grown independent of host cells in the supplemented KBM and *C. albicans* was grown in the conditioned supernatants. The *L. rhamnosus* conditioned supernatants, reduced *C. albicans* filamentation and even induced a transition from hyphae to yeast **(Fig. S4A)**. Metabolites specifically secreted when *L. rhamnosus* colonized IECs (cluster 1, **Fig. 2C,** cluster 1 **Fig. S1B, Tab. S2**) were investigated for their effects against *C. albicans*. Literature searches revealed that several of these metabolites, including phenyllactic acid[34], mevalonolactone[35], 2-hydroxyisocaproic acid (HICA)[36,37], and 3-hydroxyoctanoate[38], have previously been reported to have antifungal potential. We validated the effects of phenyllactic acid and 2-hydroxyisocaproic acid, and observed that these metabolites changed the pH of the medium. Nevertheless, compared to pH-adjusted medium, an impact on *C. albicans* filamentation was still observed. Further screening of the *L. rhamnosus* colonization-derived metabolites revealed several additional metabolites that differentially influenced *C. albicans* filamentous growth **(Fig. S4B, Fig. 3A)**. Notably, cytosine induced a hypha-to-yeast transition, characterized by a complete halt of hyphal growth at approximately 8 h, after which cells continued to proliferate in the yeast morphology **(Fig. 3A, supplementary videos 1 and 2)**. The metabolites identified also reduced filamentation and induced a subtle shift towards yeast growth when administered at lower concentrations and in combination with each other **(Fig. 3B)**. After 4 h, hyphae were significantly shorter when the combination of metabolites was present **(Fig. 3C)**. In line with the impaired filamentation, *C. albicans*-induced damage of IECs at 24 hpi was reduced **(Fig. 3D)**. Causing the most striking phenotype, cytosine on its own was evaluated for its potential to inhibit *C. albicans*-induced damage and translocation. Both host cell damage as well as *C. albicans* translocation were reduced **(Fig. 3E, F)**.

L. rhamnosus induces a hostile environment for C. albicans

The reduced *C. albicans* growth observed when infecting *L. rhamnosus*-colonized IECs[26] was also reflected by *in silico* FBA[30] of the genome-scale metabolic model of *C. albicans*. By using our

established metabolome data-driven genome-scale metabolic models of IECs and *L. rhamnosus*[31,32] we predicted potential metabolite secretion or uptake **(supplementary data 2 and 3)**. These predictions were cross-checked with phenotypic growth data of *C. albicans*[39] and metabolomics fold changes for different conditions **(Fig. S5)**.

Glucose, fructose, mannose, and N-acetylglucosamine/N-acetylgalactosamine support efficient *C. albicans* growth[39]. However, these metabolites were depleted upon *L. rhamnosus* colonization **(Fig. 3G, Fig. S5)**. Instead, alternative carbon sources, like 3-4-hydroxyphenyllactate, lactate, N-acetylglutamate, and malate showed approximately 5-fold higher concentrations, but these supported *C. albicans* growth to a lesser degree[39].

Amino acids are favourable nitrogen sources to sustain *C. albicans* growth[39] as well as carbon sources when glucose is consumed, yet *L. rhamnosus* colonization reduced the availability of most favoured amino acids **(Fig. 3G)** and increased N-acetylated amino acids. *In silico* GEM based analyses predicted that amino acids are taken up by IECs and *L. rhamnosus* in the majority of cases **(Fig. S5)**. The relative decrease of amino acids upon *C. albicans* infection suggests that *C. albicans* uses them as an alternative, but less favoured carbon source. This was accompanied by a depletion of the favoured phosphorus source choline phosphate during *L. rhamnosus* colonization **(Fig. 3G)**. Collectively, *in silico* and metabolomics analyses revealed that *L. rhamnosus* colonization depleted favoured carbon, nitrogen, and phosphorus sources of *C. albicans*, which were replaced by alternative carbon or nitrogen sources forcing *C. albicans* into a suboptimal growth milieu. This could explain the 3-fold growth reduction when *C. albicans* infects *L. rhamnosus*-colonized epithelium **(Fig. 1A)**.

Further *in silico* GEM simulations provided clues on activity of individual pathways in both *C. albicans* and IECs **(Fig. S3)**. Reaction flux ranges of *C. albicans* glycolysis, TCA cycle, or pyruvate metabolic pathways changed by 20-40% when simulated on colonized *vs.* uncolonized IECs supernatant **(Fig. S3B)**. This suggests shifts in *C. albicans* metabolism due to the altered metabolic environment. *C. albicans* oxidative phosphorylation was predicted to substantially alter in IECs supernatants, and even further change in the presence of *L. rhamnosus,* adding evidence that its energy maintenance might be affected **(Fig. S3B)**. Additional changes were predicted in lipid, sulfur, and nucleotide related metabolic pathways, including biotin, butanoate, purine, and pyrimidine metabolism, suggesting a comprehensive metabolic shift in *C. albicans*, which begins to utilize alternative carbon sources.

L. rhamnosus forces fungal transcriptional metabolic adaptation

In silico simulations suggested that *C. albicans* adapts to cope with the changed metabolic environment. We hypothesized that these metabolic adaptations require transcriptional reprogramming. Differential gene expression of *C. albicans* during infection of *L. rhamnosus*-colonized IECs was investigated by transcriptional profiling. Unsupervised hierarchical clustering **(Fig. S6A)** and principal component analysis (PCA) **(Fig. 4A)** revealed a distinct gene expression pattern upon *L. rhamnosus* colonization, but only 1.3% of the up-regulated and 1.6% of the down-regulated genes overlapped across time points **(Fig. 4B, C)**.

L. rhamnosus killed with antibiotics at 4 hpi did not induce noticeable transcriptional reprogramming of *C. albicans*, as exemplified by missing shifts in the first two PCA-defining principal components and gene expression profiles, **(Fig. 4D, S6B)**, underlining that only live *L. rhamnosus* enforces transcriptional adaptation.

Next, we explored the transcriptional changes for hints pointing towards fungal metabolic adaptation. Gene Ontology (GO) enrichment analysis revealed that a variety of metabolic processes were down-regulated at 6 hpi, while the opposite was observed at 24 hpi **(Fig. 5A)**. As the GO terms did not yield very specific insights into the transcriptional metabolic adaptation, we analysed the expression of transcription factors regulating metabolic adaptation as well as key metabolic processes.

Consistent with the limited favoured carbohydrate availability **(Fig. 3G, Fig. S5)**, the transcription factor genes *MIG1* and *TYE7* were significantly up-regulated at 6 hpi, possibly to compete for any remaining favoured carbohydrates. Nevertheless, at 24 hpi *MIG1* and *TYE7* as well as *GAL4, SUC1, HMO1*, and *STD1* were significantly down-regulated. Despite increased expression of *MIG1* and *TYE7* early during infection, genes encoding major glycolysis enzymes such as *PFK1* were down-regulated during the entire infection **(Fig. 5B)**. Instead, the transcription factor gene *STP2,* regulator of amino-acid metabolism, showed increased expression early during infection. Consequently, expression of the amino-acid permease gene *GAP2* was increased throughout infection. In line with this, specific TCA cycle genes (*CIT1, ACO1, and MDH1*), genes in the glyoxylate shunt (*ICL1, MLS1*), and genes in the gluconeogenesis pathway (*PCK1, FBP1*) were up-regulated at 6 hpi **(Fig. 5B)**.

To contextualize *C. albicans* metabolic reprogramming across multiple analysis levels, we overlaid our metabolome-driven *in silico* modelling predictions **(Fig. S3B, supplementary data 2)** with our transcriptome and metabolome data **(Fig. 6, Tab. S3)**. *In silico* analysis based on the metabolite availability at 12 hpi predicted reduced *C. albicans* glycolytic activity upon *L. rhamnosus* colonization, consistent with the down-regulation of transcription factors regulating carbon metabolism at 24 hpi, and key glycolysis genes at 6 and 24 hpi **(Fig. 5B)**. *In silico* prediction also reflected most parts of the available transcriptional and metabolic data concerning the TCA cycle **(Fig. 6)**. Both our metabolome and transcriptome data suggested relevance for the glyoxylate shunt, which *C. albicans* likely uses early during infection to bypass a full TCA cycle to use available amino acids as carbon sources to compensate for the absence of favoured nutrients. At 24 hpi *C. albicans* transcriptomic data suggested that the bias towards the glyoxylate shunt was lost and *C. albicans* tried to compensate by re-establishing a full TCA cycle and thus energy metabolism. *In silico* metabolic flux predictions at 12 hpi supported this notion and suggested decreased flux rates in citrate synthase and in succinate dehydrogenase opposed to increased flux in succinyl-CoA synthetase. Finally, several parts of the oxidative phosphorylation pathway were predicted to be less active in the presence of *L. rhamnosus*. XTT assays, which assess mitochondrial dehydrogenase activity, suggested a reduced mitochondrial metabolic activity when *C. albicans* is cultured in supernatants of *L. rhamnosus*-colonized IECs **(Fig. S7A)** or in *L. rhamnosus*-conditioned medium independent of host cells **(Fig. S7B)**. In addition, two of the metabolites observed after *L. rhamnosus* colonization reduced mitochondrial dehydrogenase activity **(Fig. S7C)**.

This was supported by the transcriptional data suggesting a trend towards an overall down-regulation of genes in complex 1 and 4 of the oxidative phosphorylation pathway **(Fig. S7D, Fig. 6)**.

Our metabolome and transcriptome data combined with *in silico* metabolic flux predictions **(Fig. 3G, Fig. S5)** suggest that *C. albicans* undergoes drastic metabolic adaptations in response to the suboptimal nutritional environment induced by *L. rhamnosus* colonization.

Dysregulation of C. albicans virulence-relevant genes

Changes in fungal energy metabolism could be linked to reduced pathogenicity. Decreased *C. albicans* damage potential in the presence of *L. rhamnosus* could be mediated by the differential regulation of genes required for metabolic adaptation when these are linked to virulence. Therefore, 70 *C. albicans* mutants from a gene deletion mutant library[40] **(Tab. S4)**, corresponding to a range of differentially regulated genes, were tested for their damage potential. The mutants *kre5*Δ/Δ, *ptp3*Δ/Δ, *orf19.4292*Δ/Δ, *ahr1*Δ/Δ, and *ace2*Δ/Δ were attenuated in IEC damage potential **(Fig. 7A)**. The *ptp3*Δ/Δ mutant additionally showed reduced growth **(Fig. 7B)** and impaired filamentation **(Fig. 7C)**. In addition, *ypt7*Δ/Δ, *orf19.4459*Δ/Δ, *hyr1*Δ/Δ, *opt7*Δ/Δ, *rbe1*Δ/Δ, *pdk2*Δ/Δ, *orf19.7328*Δ/Δ, *zcf27*Δ/Δ, and *rgs2*Δ/Δ showed a hyper-damaging phenotype **(Fig. 7A)**. Comparison of expression and mutant damage potential revealed a correlation between down-regulation and a reduced damage capacity of mutants of the corresponding genes with the exception of *kre5*Δ/Δ. Except *ypt7*Δ/Δ and *orf19.7328*Δ/Δ, genes corresponding to the hyper-damaging mutants were significantly up-regulated

by *L. rhamnosus* at 6 or 12 hpi, suggesting that these represent potential antivirulence genes[41]. Collectively, this supports the hypothesis that the transcriptionally-regulated metabolic adaptions to *L. rhamnosus*-colonized IECs compromises the expression of virulence and antivirulence genes, thereby reducing the pathogenic potential of *C. albicans*.

7. DISCUSSION

Here, we investigated metabolic and molecular aspects of *L. rhamnosus*-mediated protection against *C. albicans* pathogenicity. While colonization with *L. rhamnosus* reduced the number of *C. albicans* in contact with the epithelium[26], reduced inocula still cause significant damage, hinting at additional mechanisms mediating the protection. We discovered that this contact-independent protection was associated with *Lactobacillus*-induced changes in the metabolic environment. These metabolic changes forced *C. albicans* to transcriptionally reprogram its metabolism, and these transcriptional changes were intertwined with genes required for pathogenicity.

Lactobacillus species shape the intestinal environment by consuming and releasing metabolites[42–45]. Antibiotic treatment reduced short chain fatty acid and secondary bile salt levels, while increasing carbohydrates and sugar alcohols, which enhanced *C. albicans* filamentation and colonization rates in mice[46–49]. We demonstrate that *L. rhamnosus* colonization not only creates an antagonistic environment, but also maintains it over time. Consistent with the rapid changes in the environment upon disturbances of the microbiota *in vivo*[46], a loss of the protective effect was observed upon killing the bacteria with antibiotics even after infection.

Metabolically active, proliferating *L. rhamnosus* cells are required for the antagonistic effect towards *C. albicans*[26], but the cell culture medium alone does not support *L. rhamnosus* growth. Combined *in silico* genome-scale metabolic simulations and *in vitro* experiments have been used before to shed light on the role of gut microbiota towards host metabolic disease[39,50]. Here, analysing both in conjunction revealed that epithelium-derived metabolites foster *L. rhamnosus* growth by providing nutritional support in specific pathways. This underscores that the host metabolic activity is required to support *L. rhamnosus* growth and its antagonistic effects in our model. The interplay of consumption of IEC-derived metabolites and subsequent metabolic activity of *L. rhamnosus*, highlights that *L. rhamnosus* may be metabolically specialized to survive in the presence of epithelial cells. Such interactions have been described to play a key role in the human gut, where the intestinal epithelium provides metabolites that selectively sustain beneficial members of the microbiota[51].

L. rhamnosus colonization of epithelial cells drastically changed the metabolic environment, an observation that can also be made in *L. rhamnosus*-colonized gnotobiotic mice[45]. We observed that specific antifungal and antivirulence compounds were secreted. *Lactobacillus* species are well known for their production of antimicrobial metabolites, which have been extensively studied outside the context of the host. Several metabolites that we detected upon *L. rhamnosus* colonization were previously characterized for their antifungal and antivirulence potential including: phenyllactic acid, mevalonolactone, 2-hydroxyisocaproic acid and 3-hydroxyoctanoate[35–38]. However, the recently characterized filamentation inhibiting metabolite β-carboline[21], was not detected in our metabolome, yet this may be attributable to the culture conditions or its degradation by host cells.

We further identified several metabolites that affect *C. albicans* filamentous growth alone and in combination. The metabolite cytosine, which was observed to induce a hyphae-to-yeast transition, was validated for its antivirulence properties, in terms of reducing epithelial damage and translocation. Although the underlying mechanisms of the effects caused by the individual metabolites are not yet understood, it underscores that metabolic interplay between *C. albicans* and antagonistic bacteria contributes to promoting *C. albicans* commensalism.

In addition, *L. rhamnosus* colonization changed nutrient availability. Comparative analysis of metabolite availability from this study with previously published metabolic phenotyping of *C. albicans*[39] revealed a depletion of preferred carbon and nitrogen sources, such as amino acids or glucose, and an enrichment of less-favoured carbon sources, such as lactate or malate. The utilization of different carbon sources can drastically influence *C. albicans* fitness and pathogenicity[52].

Even though single metabolites can potently inhibit *C. albicans* pathogenicity mechanisms such as filamentation, we believe that the promotion of commensalism by the bacterial microbiota is multifactorial. Both the production of antivirulence metabolites as well as alterations in the metabolic

environment may equally promote commensalism and even depend on each other. Moreover, *L. rhamnosus* has been described to produce chitin degrading proteins[16] and exopolysaccharides[17] that can inhibit hyphal morphogenesis. Likely a combination of metabolic antagonism and specific effector functions underly the potent pathogenicity inhibiting effects of *L. rhamnosus.*

While our study only includes a single bacterial member of the microbiota, we believe that similar metabolic interactions and competition by additional antagonistic bacteria underly the strong association between a healthy microbiota and commensalism. Moreover, multiple antagonistic bacterial species may act synergistically in promoting *C. albicans* commensalism.

Although we followed a one-model-at-a-time simulation approach, our *in silico* analysis revealed changes in key metabolic pathways and regulator genes of *C. albicans*, which we also found in our *C. albicans* transcriptome data. Further, more sophisticated community modelling simulating all three GEMs simultaneously were largely in agreement with our single GEM simulations, but may be investigated in more depth in future work.

Several studies suggest that *C. albicans* metabolism and virulence are interconnected[53–56], similarly to other fungal pathogens[57,58]. A variety of proteins regulate metabolism as well as virulence, including Yck2[56,59], Tye7[60,61], Gal4[60], Mig1[62], Mig2[62], and Ace2[63,64].

Supporting our hypothesis that alternative carbon sources reduce *C. albicans* pathogenicity, several down-regulated genes, especially in the late phase of infection, were associated with carbohydrate metabolism[62]. Mig1, an essential regulator in the glucose repression pathway[62], was up-regulated early during infection of *L. rhamnosus*-colonized IECs, but down-regulated later, when only alternative carbon sources were available. However, existing studies demonstrated that a *mig1*Δ/Δ deletion mutant only showed attenuated virulence when *MIG2* was simultaneously deleted[62]. Similarly, the deletion of the glycolysis-regulating genes *TYE7* and *GAL4*, only showed attenuated virulence in a *Galleria melonella* infection model when a corresponding double deletion mutant (*gal4*Δ/Δ/*tye7*Δ/Δ) was investigated[60]. Both genes were down-regulated at later time points in our model.

Based on this, we believe that individual deletion or down-regulation of these metabolic transcription factors can be compensated by redundancy, which secures a high level of metabolic flexibility for *C. albicans*. However, when more than one of these genes are not expressed *C. albicans* loses its metabolic flexibility, which is associated with reduced pathogenicity. In line with this, a closer look at the carbohydrate catabolism revealed the down-regulation of glycolysis-relevant genes and an up-regulation of several TCA cycle and glyoxylate shunt genes starting at 6 h. Interestingly, this matches the metabolic phenotype of *ace2*Δ/Δ[64]. However, the *ace2*Δ/Δ mutant has a large cell morphology defect[64], which is associated with attenuated virulence[63]. Nevertheless, the phenotype of this mutant suggests that transcription factors like Ace2 affect both virulence as well as metabolism.

Several genes of the oxidative phosphorylation pathway were supressed upon *L. rhamnosus* colonization, such as *ADH1* or *COX2*. Impaired oxidative phosphorylation could limit ATP production, which may reduce fungal growth, filamentation, and virulence. A link between respiration and fungal pathogenesis has been described[65] and specific correlations between *C. albicans* oxidative phosphorylation and pathogenicity mediated by *ADH1* have been observed[66].

The impact of transcriptional changes on *C. albicans* virulence was validated using deletion mutants. Deletion mutants of the genes *PTP3, AHR1, ACE2,* and *orf19.4292,* which are down-regulated during infection of *L. rhamnosus*-colonized IECs*,* exhibited an attenuated damage potential. Moreover, *ptp3*Δ/Δ was compromised in growth and filamentation. *PTP3* encodes a protein tyrosine phosphatase required for hyphal maintenance[67]. The transcription factor Ahr1 regulates the virulence genes *ALS3* and *ECE1*[68], but can also repress the white-to-opaque switch[69]. *Orf19.4292* encodes the prevacuolar trafficking protein Pep12, which is essential for virulence in mice[70]. The transcriptional regulator Ace2 regulates glycostress metabolism[64] and its role in virulence is discussed above. Our data underscores that down-regulation of each of these genes could be sufficient to reduce

pathogenicity. This supports the notion that *C. albicans* metabolic and transcriptional adaptations upon infection of *L. rhamnosus*-colonized IECs are intertwined with pathogenic potential.
Other studies also observed changes in the *C. albicans* gene expression induced by lactobacilli. During oral epithelium infection, *C. albicans* induced expression of genes associated with diverse metabolic pathways[71]. *L. plantarum*, *L. helveticus,* and *L. crispatus* also down-regulated hypha-associated genes (HAGs)[72,73]. As the majority of HAGs were not affected in our dataset, we hypothesize that niche-specific modes of action exist for the diverse *Lactobacillus* species.
Collectively, our results demonstrate that protection by *L. rhamnosus* colonization is a multifactorial process that synergistically affects *C. albicans* growth and pathogenicity. Different aspects of this complex interaction have been individually assessed in the past. Here we provide a multilevel comprehensive picture of the interplay between *C. albicans* and its antagonist *L. rhamnosus* in a human, gut-like experimental setup. The metabolic and transcriptional insights into the antagonistic potential of a single member of the microbiota underline the importance and complexity of a balanced intestinal microbiota that keeps *C. albicans* in its commensal state.
Our combined metabolome data, *in silico* metabolic modelling, transcriptome and mutant screening, as well as *in vitro* validations, provide fundamental, contextualized insights into how *C. albicans* pathogenicity can be controlled or prevented.

8. ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2051 – Project-ID 390713860. MSG was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft - DFG) Emmy Noether Program (project no. 434385622 / GR 5617/1-1). AL, KG, and BH were supported by the Infect ERA-NET Program FunComPath (BMBF 031L0001A) and the Centre for Sepsis Control and Care (CSCC, BMBF 01EO1002). JLS, BH, MHM, PG, GP, SS, SV, and MSG are further supported by the DFG within the Collaborative Research Centre (CRC)/Transregio (TRR) 124 “FungiNet” projects C1, C2, B5, and INF (DFG project number 210879364). SM and BH are further supported by a grant from the Wellcome Trust (215599_Z_19_Z). Authors would like to thank Dr. Sascha Brunke for his assistance in data management and providing Duplos.

9. MATERIALS AND METHODS

In vitro model and infection

An *in vitro* intestinal *C. albicans*-infection model was used to perform the experiments as previously described[26]. C2BBe1 (ATCC®CRL2102) and HT29-MTX (ATCC HTB-38; CLS, Lot No. 13B021) cells were seeded in collagen I coated (10 µg/ml, 2 h at room temperature [RT]; Thermo Scientific) 6-well, 24-well or 96-well plates at a ratio of 70:30 with a total cell density of 4×10^5 cells/well (6-well), 1×10^5 cells/well (24-well) and 2×10^4 cells/well (96-well, transwell). Cells were used for experiments after 14 days of differentiation in Dulbecco's Modified Eagle's Medium (DMEM; Gibco, Thermo Scientific) supplemented with 10% foetal calf serum (FCS; Bio & Sell), 10 µg/ml Holotransferrin (Calbiochem, Merck), and 1% non-essential amino acids (Gibco, Thermo Scientific) with medium exchange three times per week. Cell lines have been authenticated *via* commercial STR profiling (Eurofins Genomic) and checked for mycoplasma contaminations using a PCR mycoplasma test kit (PromoKine) according to the manufacturer's instructions.
For colonization, DMEM was exchanged for 1.3 ml (6-well) or 50 µl (96-well) serum-free Keratinocyte Basal Medium (KBM) (Lonza, Basel, Switzerland) and monolayers were colonized with 1.3 ml (6-well), 250 µl (24-well) or 50 µl (96-well) *Lactobacillus rhamnosus* ATCC 7469 (OD 0.2 in KBM) for 18 h prior to infection. For contact-independent colonization, DMEM was exchanged for 600 µl in the bottom of a 24-well plate and a transwell insert with 125 µl *L. rhamnosus* (OD 0.4 in KBM) was placed on top.
Fungal infection was performed with 1.3 ml (6-well), 250 µl (24-well) or 50 µl (96-well) *C. albicans* WT SC5314 (4×10^5 cells/ml in KBM). For the samples with antibiotic treatment, 500 µg/ml Gentamicin (Merck) and 1× PenStrep (Gibco, Thermo Scientific) were added 4 hpi. For infection in presence of metabolites, a solution of each metabolite in KBM was prepared fresh and sterile filtered. Then, concentration was adjusted and 50 µl (96-well, transwells) of this solution with *C. albicans* (MOI 1) were used to infect the cells for 24 h. Wells with only medium, *L. rhamnosus*, or *C. albicans* in the presence or absence of the host cells served as controls. Infected cells and controls were incubated at 37°C with 5% CO_2. Data from *in vitro* damage, fungal burden, and translocation assays were analysed using GraphPad prism version 8. Data from at least 3 biological replicates were analysed for statistical significance using a t-test or a one-way ANOVA with multiple comparisons. Statistical significance is depicted in the figures: * = $p \leq 0.05$, ** = $p \leq 0.01$, or *** = $p \leq 0.001$.

C. albicans CFU quantification

To determine how many *C. albicans* cells were present 24 hpi, CFU quantification was performed in 96-well plates. Supernatants were collected and IECs were treated for 5 min with 0.2% Triton-X-100 (Sigma-Aldrich) to lyse the host cells and release adherent fungal cells. After detaching adherent host cells *via* scraping with a pipette tip, the lysate was added to the respective supernatant. Wells were washed twice with PBS. The final samples were diluted appropriately with PBS and plated on YPD agar with 1× PenStrep (Gibco, Thermo Fisher Scientific) and incubated at 30°C until adequate growth for CFU counting was reached (24 h).

Quantification of cytotoxicity (LDH)

The host cell damage was determined by measuring the activity of cytoplasmic LDH[74] **(Fig. S8)**. LDH activity was quantified in the supernatant of infected IEC monolayers in 96-well plates 24 hpi using the Cytotoxicity Detection Kit (Roche) according to the manufacturer's instructions. LDH from rabbit muscle (5 mg/ml, Roche) was used to generate a standard curve for the determination of LDH concentrations. The background LDH activity control level of uninfected IECs was subtracted from the test conditions.

Metabolome analysis

Supernatants for untargeted metabolomics were collected 6 and 12 h after *C. albicans* infection **(Fig. S8)**. 500 µl of the supernatant was collected from 6-well plates, centrifuged, snap-frozen in liquid nitrogen, and stored at -80°C until analysis. Samples were analysed and interpreted by Metabolon (Morrisville, US). Experiments included three technical replicates and five independent experiments were performed in total.
Raw metabolome data was rescaled to set the median equal to 1, and the missing values imputed with the minimum. Data were loaded in R version 1.2.5019[75], rows were normalized and Euclidian distances were calculated. The heatmap was generated with the "pHeatmap" package v1.0.12[76] and used to manually obtain the optimal number of clusters. The distance matrix was subjected to hierarchical cluster analysis using the complete linkage agglomeration method and metabolites were classified according to their cluster. Colour bars indicating cluster or condition were added to the dendrograms using the R package "dendextend"[77]. Metabolite enrichment analysis were conducted in MetaboAnalyst 5.0[78] across KEGG pathways using overrepresentation analysis (ORA) analysis. All results across all mentioned clusters are presented and significant hits are indicated. FDR correction was done per cluster and time point, and FDR ≤ 0.1 was considered. Proportional Euler diagrams were done using the R package "eulerr"[79].

L. rhamnosus growth

L. rhamnosus was cultured in MRS Broth for 48 h at 37°C with 5% CO_2 and 1% O_2. Afterwards, cells were washed, and 3000 cells/ml were inoculated in the different media and IECs were colonized. At 0, 24, and 48 h independent wells were resuspended, and the content was appropriately diluted and plated on MRS agar. Plates were incubated for 48 h at 37°C with 5% CO_2 and 1% O_2 until CFU quantification. Tested metabolites are shown in Tab.1.

Live-cell imaging

C. albicans (1×10^4 cell/ml) in KBM with or without supplementation of individual metabolites or in combination was incubated for 24 h at 37°C with 5% CO_2 inside the Cell Discoverer 7 microscope (Zeiss), in which a bright field picture was taken every hour.

Translocation assay

To determine translocation through the epithelial barrier, infections were performed in transwell inserts (Sarstedt) with a pore size of 5 µm. 24 hpi, zymolyase (260 U/ml) was added to the bottom compartment and incubated for 2 h at 37°C with 5% CO_2. Afterwards, samples were diluted in PBS, plated on YPD agar, and incubated at 30°C for 24 h.

Transcriptional profiling

After 6 and 24 h of *C. albicans* infection in 6-well plates, RNA isolation of *C. albicans* was performed **(Fig. S8)**. At the appropriate time points, 650 µl RLT buffer was added to the wells and the plates were frozen in liquid nitrogen immediately. After thawing, fungal and host cells were collected *via* scraping. The collected material was centrifuged and fungal RNA isolation was performed on the pellet according to a previously described protocol[80]. RNA quantities were determined with a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific) and RNA quality was verified with an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA was subsequently converted into Cy5-labeled cRNA (Cy5 CTP; GE Healthcare, United Kingdom) using a QuickAmp labelling kit (Agilent). Samples were co-hybridized with a common Cy3-labeled reference (RNA from mid-log-phase-grown *C. albicans* SC5314[81]) on Agilent arrays (*C.a.*: AMADID 026869), scanned in a GenePix 4200AL with GenePix Pro 6.1 (Auto PMT; pixel size, 5 µm). Differentially expressed genes (DEGs) (*p*-value: 0.05; Log_2 fold change) were analysed with GeneSpring 14.9 (Agilent) and the *Candida* Genome Database (CGD; http://www.candidagenome.org). Out of the 6130 *C. albicans* genes, 5125 genes were used for the analysis after filtering on the minimal fluorescent signal and subtraction of the background signal. Genes were considered differentially

regulated when they had a Bonferroni-corrected *p*-value of < 0.05 and a Log_2 fold change of more than 1 or less than -1.
Gene expression data were exported from GeneSpring and loaded in R version 3.6.2[75], rows were normalized and Euclidian distances were calculated. The heatmap was generated with the "pHeatmap" package v1.0.12[76] and the distance matrix was subjected to hierarchical cluster analysis using the complete linkage agglomeration method. Colour bars indicating condition were added to the dendrograms using the R package "dendextend"[77]. The PCA was calculated using the R function "prcomp". Graphs were generated using the "ggbiplot" package v 0.55[82]. A proportional Venn diagram of the DEGs was drawn using the "eulerAPE" application v2.0.3[83].
GO term enrichment of differentially expressed genes was analysed using the GO-Term Finder on the *Candida* genome database[84], which uses a hypergeometric distribution with Multiple Hypothesis Correction (Bonferroni Correction) to calculate *p*-values. Subsequently, the significantly enriched GO-terms were processed using REVIGO[85] (similarity: Tiny (0.4); database: whole Uniprot; semantic similarity measure: SimRel) to remove overlapping and redundant GO-terms.

Bioinformatics

GEMs for *C. albicans*, human, and *L. rhamnosus* metabolism were used to simulate and analyse different growth scenarios *in silico*. Specifically, the recently published model for *C. albicans*[39] was downloaded from the supplementary material of the publication. The GEMs for *Lactobacillus rhamnosus* LMS2-1 for *L. rhamnosus*[32], and Recon3D 3.01[31], a comprehensive generic GEM of human metabolism used to simulate human intestinal epithelial cells, were downloaded from www.vmh.life. Metabolomics data were used to modulate feasible nutrition uptake for each model via respective exchange reactions as defined by each GEM. Feasible uptake rates for available metabolites were adapted from the metabolome measurements across all investigated conditions. Feasible uptake flux ranges for each metabolite in our GEMs were kept in the range [0, 1000] mmol/g(DW)h. The metabolite concentrations for each sample were transformed into this range based on the metabolite glutamine showing the highest concentration in the 12 hpi *L. rhamnosus* supernatant compared to all measured metabolites and all samples. The uptake rate of glutamine was set to 1000 mmol/g(DW)h accordingly, whereas all others were set to the respective proportion to the maximum glutamine value. The biomass function of each GEM was used as objective function for all metabolic modelling simulations. To obtain objective function values mimicking an anaerobic environment (oxygen influx prohibited) as well as feasible reaction flux ranges supporting at least 90% of the objective function flux, we applied flux balance analysis (FBA) and flux variability analysis (FVA) across all tested conditions for all tested GEMs[30,33]. All GEM analyses were done in COBRApy[86] using Python 3.6.4 and the IBM ILOG CPLEX Optimizer (version 12.8).

Mitochondrial activity assessment with XTT

To obtain the *L. rhamnosus*-conditioned supernatants, a culture of *L. rhamnosus* was washed and adjusted to an OD_{600} of 0.2 in KBM, and 1ml was used to colonize a confluent layer of differentiated intestinal epithelial cells (C2BBe1:HT29-MTX, 70:30) for 24 h (37°C, 5% CO_2). *L. rhamnosus* was also adjusted to an OD_{600} of 0.2 in KBM+ (KBM supplemented with 5 mM citric acid, 5 mM carnitine and 5 mM gamma-glutamyl-alanine) and incubated in a 24-well plate independently of host cells for 24 h (37°C, 5% CO_2). After 24 h, conditioned media was filtered. *C. albicans* was washed and adjusted to a concentration of 1×10^5 or 1×10^6 cell/ml in the different media and incubated in a 96-well plate for 1h (37°C, 5% CO_2). After that time, the plate was centrifuged and supernatants were exchanged with XTT reagent (PBS with 0.2 mg/ml XTT, VWR Life Sciences and 1.1 µg/ml menadione sodium bisulfite, Roth). The plate was then incubated for 2h at 37°C. Afterwards, 100 µl of the supernatants were transferred to a new plate and absorbance at 492 nm measured.
To compare XTT absorbance to biomass, a crystal violet assay was performed afterwards. The supernatant from each well was removed and the plate was left to air dry. Afterwards, 150 µl crystal

violet 1% (Sigma) was added. After 45 minutes incubation at room temperature, crystal violet was removed, and the wells were washed three times with distilled water. Crystal violet was then solubilized with 200 µl ethanol 99% for 45 min. Afterwards, 100 µl of the supernatants were transferred to a new plate and absorbance was measured at 550 nm.
The same procedure was used to assess the effect of the metabolites in the mitochondrial activity, except the preincubation with the metabolite was done for 24 h prior to the XTT assay and the initial *C. albicans* concentration was 1×10^4 cell/ml.

Deletion mutant screening

C. albicans mutants from the deletion mutant library[40] **(Tab. S4)** were cultivated in YPD broth in 96-well plates and incubated overnight at 30°C with shaking at 180 rpm. The overnight cultures were adjusted to an OD_{600} of 0.0025 in KBM. Then, the diluted overnight cultures were diluted 1:2 in KBM in 96-well plates with a confluent layer of differentiated intestinal epithelial cells (C2BBe1:HT29-MTX, 70:30). Damage was measured *via* the LDH activity assay (see Quantification of cytotoxicity (LDH)). Positive hits were validated with additional damage experiments where the mutants were cultivated in YPD broth in 25 ml Erlenmeyer flasks overnight at 30°C with shaking at 180 rpm. Fungal cells where then washed in PBS, counted, and adjusted to a concentration of 4×10^5 cells/ml in KBM. The epithelial cell monolayer was infected with 50 µl (96-well, transwell) or 250 µl (24-well). For growth curves, the diluted overnight cultures were diluted 1:2 in KBM in 96 well plates and incubated for 24 h at 37°C with 5% CO_2 in a microplate reader (Tecan Infinite M200; i-control software). Growth was monitored with OD_{600} measurements every 30 min over 24 h. Data was analysed using Graphpad prism version 8. Data from at least 3 biological replicates was analysed for statistical significance using a one-way ANOVA with multiple comparisons.

Data and code availability

Untargeted metabolomics data is available as supplementary data 1. Biomass objective function values and associated flux ranges for all reactions for all simulations and investigated media conditions are available as supplementary data 2. Simulation sets over different fractions of required objective function values are available as supplementary data 3. Transcriptomics data is available in the ArrayExpress database under accession code: E-MTAB-11090. The python code for the metabolic models and simulations is available at:
https://github.com/mohammadmirhakkak/Calbicans_LRhamnosus_EpithelialCells_Interaction
Source data are provided with this manuscript.

10. FIGURE LEGENDS

Fig. 1: Reduced C. albicans inocula cause significant damage compared to C. albicans in the presence of live L. rhamnosus that reduces pathogenicity contact-independently

(A) Fungal burden assessed by quantification of *C. albicans* CFUs and (B) the necrotic cell damage of IECs quantified by LDH activity at 24 hours post infection (hpi). Cells were infected with a *C. albicans* infection inoculum (4×10^5/ml) in the presence and absence of *L. rhamnosus,* or with a reduced infection inoculum (1×10^4/ml) in the absence of *L. rhamnosus*. (C, D) Necrotic cell damage of IECs quantified by LDH activity at 24 hpi, (C) in the presence and absence of *L. rhamnosus* colonization and antibiotic treatment with Gentamicin and Penicillin/Streptomycin at 4 hpi, or (D) in the presence and absence of *L. rhamnosus* colonization where *L. rhamnosus* was in direct contact with the cells or physically separated using a transwell insert with a 0.4 µm pore size. Bars represent the mean and standard deviation (SD) of n=3 independent experiments, dot plots show the mean of the technical replicates of the individual experiments, data were compared for significance using an unpaired t-test (two-tailed, one-sample), * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

Fig. 2: IECs metabolically foster L. rhamnosus growth

(A) Principal component analysis of metabolite composition assessed by untargeted metabolomics in supernatants collected at 6 and 12 hpi. (B) Differentially increased or decreased metabolites in comparison of the different conditions shown in proportional Venn diagrams. Data summarized from n=5 at 6 and 12 hpi. (C) Hierarchical clustering based on Euclidean distance of relative metabolite abundance in the supernatants at 6 hpi. (D) Growth of *L. rhamnosus* in KBM, on IECs, in transwells physically separated from IECs, or in supernatants of IECs (spent) assessed by counting CFUs on MRS agar. Data is shown as the mean and SD. (E) *In silico* prediction of *L. rhamnosus* biomass formation in KBM or supernatants of IECs. (F) Growth of *L. rhamnosus* assessed by CFUs after 48 h incubation in KBM supplemented with single metabolites or combinations of metabolites. Data of n=3 biological replicates are shown as the mean and standard deviation (SD). Data were tested for significance using a t-test (two tailed, one-sample) against growth in KBM, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$. (G) Comparison of metabolic pathway activity levels between different conditions as indicated. Relative pathway change was determined by identifying the number of pathway-specific reactions for which feasible flux ranges differ according to flux variability analysis.

Fig. 3: L. rhamnosus induces an unfavourable environment for C. albicans

(A,B) Representative images of *C. albicans* morphology following growth in presence of (A) different metabolites at 50 mM, at neutral or acidic pH, for 20 h at 37°C with 5% CO_2 (n=2 biological replicates) or (B) a combination of selected metabolites each at 5mM (cytosine, phenylpyruvate, 2-hydroxy-4-(methylthio)-butyric acid, 3-phenyllactic acid, 2-Hydroxyisocaproic acid and D-indole-3-lactic acid) and lactic acid at 15 mM, for 20 h at 37°C with 5% CO_2 (n=4 biological replicates). Yeast morphology is indicated with arrowheads. (C) Hyphal length of *C. albicans* grown for 4 h in KBM in the presence or absence of the combination of selected metabolites at 37°C with 5% CO_2 (n=4 biological replicates). (D) Necrotic damage of IECs measured by the LDH activity in the supernatant at 24 hpi with *C. albicans* in the presence or absence of the combination of selected metabolites. (E) *C. albicans*-induced necrotic damage of IECs measured by the LDH activity in the supernatant and (F) *C. albicans* translocation across the epithelial barrier assessed in the presence or absence of cytosine at 50 mM at 24 hpi. Bars represent the mean and SD of n=4 independent experiments, dot plots show the mean of the technical replicates of the individual experiments, boxplots represent the distribution of the total measurements (center line, median; box limits, upper and lower quartiles; whiskers, range). Data were compared for significance using paired t-tests (two-tailed, one-sample), * = $p \leq 0.05$, ** = $p \leq 0.01$. (G) Phenotypic microarray growth

experiments for wild-type *C. albicans* in presence of each metabolite as a carbon, nitrogen, or phosphorous source (left), metabolome data measured at 6 and 12 h and metabolic modelling predictions (right) are indicated for selected metabolites. For metabolic modelling, media was adapted from metabolome data derived from supernatants of IECs. Uptake or secretion was determined by identifying feasible flux ranges for metabolite-specific exchange reactions alongside optimization for biomass. Asterisks show statistical significance (ANOVA for phenotypic microarrays, Wilcoxon test for metabolomics, FDR adjusted $p \leq 0.05$). For the entire panel of metabolites see Fig. S5.

Fig. 4: C. albicans undergoes transcriptional changes during infection of L. rhamnosus-colonized IECs

(A) Principal component analysis of *C. albicans* gene expression at 6 and 24 h during *in vitro* infection of IECs in the presence and absence of *L. rhamnosus* colonization. (B) Volcano plots showing differentially regulated *C. albicans* genes at 6 and 24 hpi as a result of *L. rhamnosus* colonization prior to infection based on the criteria of a Log_2 fold change of > 1 or < -1 and a Bonferroni corrected *p*-value of < 0.05 (dark blue and dark red) and < 0.1 (light blue and light red). (C) Venn diagram analysis of the overlap in differentially expressed genes at 6 and 24 hpi. Data summarized from n=3 and n=4 independent experiments at 6 and 24 hpi respectively. (D) PCA of *C. albicans* gene expression at 24 hpi during *in vitro* infection of IECs in the presence and absence of *L. rhamnosus* colonization and in the presence and absence of antibiotics.

Fig. 5: C. albicans undergoes transcriptional metabolic adaptations when infecting L. rhamnosus colonized epithelium

(A) GO-term enrichment analysis of differentially regulated genes (Log_2 fold change > 1 or < -1 and $p < 0.1$) analysed with the GO-term finder on the *Candida* Genome Database website and reduced with the Revigo program (http://revigo.irb.hr/) (similarity: Tiny (0.4)). Significantly enriched GO-terms are plotted based on the $-Log_{10}$ *p*-value. Data summarized from n=3 at 24 hpi. (B) Heatmap highlighting the transcriptional regulation of *C. albicans* metabolic genes as a result of *L. rhamnosus* colonization at 6 and 24 hpi. Legend colour represents the Log_2 fold change of the regulation in presence *vs.* absence of *L. rhamnosus*. The asterisks (*) represent significance, based on the criteria of a Log_2 fold change of > 1 or < -1 and a Bonferroni corrected *p*-value of < 0.05.

Fig. 6: Central metabolism of C. albicans is altered by L. rhamnosus colonization.

Reactions associated to glycolysis, TCA cycle, oxidative phosphorylation are indicated, as well as relationships with additional metabolic pathways (pentose phosphate pathway, nitrogenated bases, sulfur metabolism, butanoate, propanoate, and panthenoate, CoA, β-Alanine and glutathione metabolism). Information on metabolome (12 h) and transcriptomic (24 h) data are combined with *in silico* genome-wide metabolic flux predictions (12 h). Dotted arrows represent several combined reactions. For the reaction abbreviations see Tab. S3.

Fig. 7: Screening of C. albicans deletion mutants

(A) Ability of deletion mutants to induce necrotic cell damage of IECs assessed by LDH activity in the supernatant at 24 hpi. Data is shown as the mean and SD with dots showing the individual replicates. Deletion mutants were compared to the wild-type control using a one-way ANOVA and Dunnett's Multiple Comparison post-hoc analysis. Mutants with a significantly increased or decreased damage potential (dark blue, p-value ≤ 0.05; light blue, p-value ≤ 0.1) are labelled. Horizontal lines correspond to the mean of the damage induced by the wild-type ± SD. (B) Growth rates of the *ptp3*Δ/Δ mutant with significantly reduced growth (blue line) compared to the parental strain (black line) in KBM. Lines represent the mean and SD of n=3 independent experiments and were compared for significance using an ANOVA with Bonferroni-adjusted post-hoc analysis, * = $p \leq 0.05$. (C) Representative images of the *ptp3*Δ/Δ deletion mutant and the parental strain morphologies, after 24 h incubation in KBM at 37°C with 5% CO_2.

11. BIBLIOGRAPHY

1. Neville, B. A., D'Enfert, C. & Bougnoux, M.-E. Candida albicans commensalism in the gastrointestinal tract. *FEMS Yeast Res.* **15**, fov081 (2015).
2. Nash, A. K. *et al.* The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome* **5**, 153 (2017).
3. Kumamoto, C. A., Gresnigt, M. S. & Hube, B. The gut, the bad and the harmless: Candida albicans as a commensal and opportunistic pathogen in the intestine. *Curr. Opin. Microbiol.* **56**, 7–15 (2020).
4. Zhai, B. *et al.* High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. *Nat. Med.* **26**, 59–64 (2020).
5. Miranda, L. N. *et al.* Candida colonisation as a source for candidaemia. *J. Hosp. Infect.* **72**, 9–16 (2009).
6. Schroeder, M. *et al.* Epidemiology, clinical characteristics, and outcome of candidemia in critically ill patients in Germany: a single-center retrospective 10-year analysis. *Ann. Intensive Care* **10**, 142 (2020).
7. Koh, A. Y., Köhler, J. R., Coggshall, K. T., Van Rooijen, N. & Pier, G. B. Mucosal Damage and Neutropenia Are Required for Candida albicans Dissemination. *PLoS Pathog.* **4**, e35 (2008).
8. Ost, K. S. *et al.* Adaptive immunity induces mutualism between commensal eukaryotes. *Nature* (2021) doi:10.1038/s41586-021-03722-w.
9. Matsubara, V. H., Bandara, H. M. H. N., Mayer, M. P. A. & Samaranayake, L. P. Probiotics as Antifungals in Mucosal Candidiasis. *Clin. Infect. Dis.* **62**, 1143–1153 (2016).
10. Förster, T. M. *et al.* Enemies and brothers in arms: Candida albicans and gram-positive bacteria. *Cell. Microbiol.* **18**, 1709–1715 (2016).
11. Millar, M. R., Bacon, C., Smith, S. L., Walker, V. & Hall, M. A. Enteral feeding of premature infants with Lactobacillus GG. *Arch. Dis. Child.* **69**, 483–487 (1993).
12. d'Enfert, C. *et al.* The impact of the Fungus-Host-Microbiota interplay upon Candida albicans infections: current knowledge and new perspectives. *FEMS Microbiol. Rev.* **45**, (2021).
13. De Gregorio, P. R., Silva, J. A., Marchesi, A. & Nader-Macías, M. E. F. Anti- Candida activity of beneficial vaginal lactobacilli in in vitro assays and in a murine experimental model. *FEMS Yeast Res.* **19**, (2019).
14. Jang, S. J., Lee, K., Kwon, B., You, H. J. & Ko, G. Vaginal lactobacilli inhibit growth and hyphae formation of Candida albicans. *Sci. Rep.* **9**, 8121 (2019).
15. Parolin, C. *et al.* Isolation of Vaginal Lactobacilli and Characterization of Anti-Candida Activity. *PLoS One* **10**, e0131220 (2015).
16. Allonsius, C. N. *et al.* Inhibition of Candida albicans morphogenesis by chitinase from Lactobacillus rhamnosus GG. *Sci. Rep.* **9**, 2900 (2019).
17. Allonsius, C. N. *et al.* Interplay between Lactobacillus rhamnosus GG and Candida and the involvement of exopolysaccharides. *Microb. Biotechnol.* **10**, 1753–1763 (2017).
18. Rossoni, R. D. *et al.* Clinical strains of Lactobacillus reduce the filamentation of Candida albicans and protect Galleria mellonella against experimental candidiasis. *Folia Microbiol. (Praha).* **63**, 307–314 (2018).
19. de Barros, P. P. *et al.* Lactobacillus paracasei 28.4 reduces in vitro hyphae formation of Candida albicans and prevents the filamentation in an experimental model of Caenorhabditis elegans. *Microb. Pathog.* **117**, 80–87 (2018).
20. Oliveira, V. M. C., Santos, S. S. F., Silva, C. R. G., Jorge, A. O. C. & Leão, M. V. P. Lactobacillus is able to alter the virulence and the sensitivity profile of Candida albicans. *J. Appl. Microbiol.*

121, 1737–1744 (2016).
21. MacAlpine, J. *et al.* A small molecule produced by Lactobacillus species blocks Candida albicans filamentation by inhibiting a DYRK1-family kinase. *Nat. Commun.* **12**, 6151 (2021).
22. Matsubara, V. H., Wang, Y., Bandara, H. M. H. N., Mayer, M. P. A. & Samaranayake, L. P. Probiotic lactobacilli inhibit early stages of Candida albicans biofilm development by reducing their growth, cell adhesion, and filamentation. *Appl. Microbiol. Biotechnol.* **100**, 6415–6426 (2016).
23. Matsuda, Y., Cho, O., Sugita, T., Ogishima, D. & Takeda, S. Culture Supernatants of Lactobacillus gasseri and L. crispatus Inhibit Candida albicans Biofilm Formation and Adhesion to HeLa Cells. *Mycopathologia* **183**, 691–700 (2018).
24. Kang, C.-H. *et al.* In vitro probiotic properties of vaginal Lactobacillus fermentum MG901 and Lactobacillus plantarum MG989 against Candida albicans. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **228**, 232–237 (2018).
25. Köhler, G. A., Assefa, S. & Reid, G. Probiotic Interference of Lactobacillus rhamnosus GR-1 and Lactobacillus reuteri RC-14 with the Opportunistic Fungal Pathogen Candida albicans. *Infect. Dis. Obstet. Gynecol.* **2012**, 1–14 (2012).
26. Graf, K. *et al.* Keeping Candida commensal : how lactobacilli antagonize pathogenicity of Candida albicans in an in vitro gut model. (2019) doi:10.1242/dmm.039719.
27. Maurer, M. *et al.* A three-dimensional immunocompetent intestine-on-chip model as in vitro platform for functional and microbial interaction studies. *Biomaterials* 119396 (2019) doi:10.1016/j.biomaterials.2019.119396.
28. Fitzsimmons, N. & Berry, D. R. Inhibition of Candida albicans by Lactobacillus acidophilus: evidence for the involvement of a peroxidase system. *Microbios* **80**, 125–33 (1994).
29. Lourenço, A., Pedro, N. A., Salazar, S. B. & Mira, N. P. Effect of Acetic Acid and Lactic Acid at Low pH in Growth and Azole Resistance of Candida albicans and Candida glabrata. *Front. Microbiol.* **9**, (2019).
30. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
31. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
32. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
33. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276 (2003).
34. Lipinska-Zubrycka, L. *et al.* Anticandidal activity of Lactobacillus spp. in the presence of galactosyl polyols. *Microbiol. Res.* **240**, 126540 (2020).
35. Niku-Paavola, M.-L., Laitila, A., Mattila-Sandholm, T. & Haikara, A. New types of antimicrobial compounds produced by Lactobacillus plantarum. *J. Appl. Microbiol.* **86**, 29–35 (1999).
36. Nieminen, M. T. *et al.* A Novel Antifungal Is Active against Candida albicans Biofilms and Inhibits Mutagenic Acetaldehyde Production In Vitro. *PLoS One* **9**, e97864 (2014).
37. Sakko, M. *et al.* 2-hydroxyisocaproic acid is fungicidal for Candida and Aspergillus species. *Mycoses* **57**, 214–221 (2014).
38. Radivojevic, J. *et al.* Polyhydroxyalkanoate-based 3-hydroxyoctanoic acid and its derivatives as a platform of bioactive compounds. *Appl. Microbiol. Biotechnol.* **100**, 161–172 (2016).
39. Mirhakkak, M. H. *et al.* Metabolic modeling predicts specific gut bacteria as key determinants for Candida albicans colonization levels. *ISME J.* **15**, 1257–1270 (2021).
40. Noble, S. M., French, S., Kohn, L. A., Chen, V. & Johnson, A. D. Systematic screens of a

Candida albicans homozygous deletion library decouple morphogenetic switching and pathogenicity. *Nat. Genet.* **42**, 590–598 (2010).
41. Siscar-Lewin, S., Hube, B. & Brunke, S. Antivirulence and avirulence genes in human pathogenic fungi. *Virulence* **10**, 935–947 (2019).
42. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
43. Markowiak-Kopeć, P. & Śliżewska, K. The Effect of Probiotics on the Production of Short-Chain Fatty Acids by Human Intestinal Microbiome. *Nutrients* **12**, 1107 (2020).
44. Prete, R. *et al.* Beneficial bile acid metabolism from Lactobacillus plantarum of food origin. *Sci. Rep.* **10**, 1165 (2020).
45. Kim, J. *et al.* Lactobacillus rhamnosus GG modifies the metabolome of pathobionts in gnotobiotic mice. *BMC Microbiol.* **21**, 165 (2021).
46. Gutierrez, D. *et al.* Antibiotic-induced gut metabolome and microbiome alterations increase the susceptibility to Candida albicans colonization in the gastrointestinal tract. 1–15 (2019) doi:10.1093/femsec/fiz187.
47. Guinan, J. & Thangamani, S. Antibiotic-induced alterations in taurocholic acid levels promote gastrointestinal colonization of Candida albicans. *FEMS Microbiol. Lett.* **365**, (2018).
48. Guinan, J., Wang, S., Hazbun, T. R., Yadav, H. & Thangamani, S. Antibiotic-induced decreases in the levels of microbial-derived short-chain fatty acids correlate with increased gastrointestinal colonization of Candida albicans. *Sci. Rep.* **9**, 8872 (2019).
49. Guinan, J., Villa, P. & Thangamani, S. Secondary bile acids inhibit Candida albicans growth and morphogenesis. doi:10.1093/femspd/fty038.
50. Mardinoglu, A. *et al.* An Integrated Understanding of the Rapid Metabolic Benefits of a Carbohydrate-Restricted Diet on Hepatic Steatosis in Humans. *Cell Metab.* **27**, 559-571.e5 (2018).
51. Schluter, J. & Foster, K. R. The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection. *PLoS Biol.* **10**, e1001424 (2012).
52. Lok, B. *et al.* The assimilation of different carbon sources in Candida albicans : Fitness and pathogenicity. *Med. Mycol.* **59**, 115–125 (2021).
53. Williams, R. B. & Lorenz, M. C. Multiple Alternative Carbon Pathways Combine To Promote Candida albicans Stress Resistance, Immune Interactions, and Virulence. *MBio* **11**, (2020).
54. Lindsay, A. K. *et al.* Analysis of Candida albicans Mutants Defective in the Cdk8 Module of Mediator Reveal Links between Metabolism and Biofilm Formation. *PLoS Genet.* **10**, e1004567 (2014).
55. Muñoz, J. F. *et al.* Coordinated host-pathogen transcriptional dynamics revealed using sorted subpopulations and single macrophages infected with Candida albicans. *Nat. Commun.* **10**, 1607 (2019).
56. Burgain, A. *et al.* A novel genetic circuitry governing hypoxic metabolic flexibility, commensalism and virulence in the fungal pathogen Candida albicans. *PLOS Pathog.* **15**, e1007823 (2019).
57. Sasse, A., Hamer, S. N., Amich, J., Binder, J. & Krappmann, S. Mutant characterization and in vivo conditional repression identify aromatic amino acid biosynthesis to be essential for Aspergillus fumigatus virulence. *Virulence* **7**, 56–62 (2016).
58. Scott, J. *et al.* Targeting Methionine Synthase in a Fungal Pathogen Causes a Metabolic Imbalance That Impacts Cell Energetics, Growth, and Virulence. *MBio* **11**, (2020).
59. Liboro, K. *et al.* Transcriptomic and Metabolomic Analysis Revealed Roles of Yck2 in Carbon Metabolism and Morphogenesis of Candida albicans. *Front. Cell. Infect. Microbiol.* **11**, (2021).

60. Askew, C. *et al.* Transcriptional Regulation of Carbohydrate Metabolism in the Human Pathogen Candida albicans. *PLoS Pathog.* **5**, e1000612 (2009).
61. Tucey, T. M. *et al.* Glucose Homeostasis Is Important for Immune Cell Viability during Candida Challenge and Host Survival of Systemic Fungal Infection. *Cell Metab.* **27**, 988-1006.e7 (2018).
62. Lagree, K. *et al.* Roles of Candida albicans Mig1 and Mig2 in glucose repression, pathogenicity traits, and SNF1 essentiality. *PLOS Genet.* **16**, e1008582 (2020).
63. MacCallum, D. M. *et al.* Different Consequences of ACE2 and SWI5 Gene Disruptions for Virulence of Pathogenic and Nonpathogenic Yeasts. *Infect. Immun.* **74**, 5244–5248 (2006).
64. Mulhern, S. M., Logue, M. E. & Butler, G. Candida albicans Transcription Factor Ace2 Regulates Metabolism and Is Required for Filamentation in Hypoxic Conditions. *Eukaryot. Cell* **5**, 2001–2013 (2006).
65. Black, B., Lee, C., Horianopoulos, L. C., Jung, W. H. & Kronstad, J. W. Respiring to infect: Emerging links between mitochondria, the electron transport chain, and fungal pathogenesis. *PLOS Pathog.* **17**, e1009661 (2021).
66. Song, Y. *et al.* ADH1 promotes Candida albicans pathogenicity by stimulating oxidative phosphorylation. *Int. J. Med. Microbiol.* **309**, 151330 (2019).
67. Su, C., Lu, Y. & Liu, H. Reduced TOR signaling sustains hyphal development in Candida albicans by lowering Hog1 basal activity. *Mol. Biol. Cell* **24**, 385–397 (2013).
68. Ruben, S. *et al.* Ahr1 and Tup1 Contribute to the Transcriptional Control of Virulence-Associated Genes in Candida albicans. *MBio* **11**, (2020).
69. Wang, H. *et al.* Candida albicans Zcf37, a zinc finger protein, is required for stabilization of the white state. *FEBS Lett.* **585**, 797–802 (2011).
70. Palanisamy, S. K. A., Ramirez, M. A., Lorenz, M. & Lee, S. A. Candida albicans PEP12 Is Required for Biofilm Integrity and In Vivo Virulence. *Eukaryot. Cell* **9**, 266–277 (2010).
71. Mailänder-Sánchez, D. *et al.* Antifungal defense of probiotic Lactobacillus rhamnosus GG is mediated by blocking adhesion and nutrient depletion. *PLoS One* **12**, e0184438 (2017).
72. James, K. M., MacDonald, K. W., Chanyi, R. M., Cadieux, P. A. & Burton, J. P. Inhibition of Candida albicans biofilm formation and modulation of gene expression by probiotic cells and supernatant. *J. Med. Microbiol.* **65**, 328–336 (2016).
73. Wang, S. *et al.* Antimicrobial Compounds Produced by Vaginal Lactobacillus crispatus Are Able to Strongly Inhibit Candida albicans Growth, Hyphal Formation and Regulate Virulence-related Gene Expressions. *Front. Microbiol.* **08**, (2017).
74. Chan, F. K.-M., Moriwaki, K. & De Rosa, M. J. Detection of Necrosis by Release of Lactate Dehydrogenase Activity. in 65–70 (2013). doi:10.1007/978-1-62703-290-2_7.
75. R Core Team (2020). R Foundation for Statistical Computing, Vienna, A. U. R: A language and environment for statistical computing. (2020).
76. Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.12. (2019).
77. Galili, T. dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. (2015) doi:DOI: 10.1093/bioinformatics/btv428.
78. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396 (2021).
79. Larsson, J. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.1.0 (2020).
80. Lüttich, A., Brunke, S. & Hube, B. Isolation and Amplification of Fungal RNA for Microarray Analysis from Host Samples. in 411–421 (2012). doi:10.1007/978-1-61779-539-8_28.
81. Naglik, J. R., Moyes, D. L., Wächtler, B. & Hube, B. Candida albicans interactions with epithelial cells and mucosal immunity. *Microbes Infect.* **13**, 963–976 (2011).

82. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (2016).
83. Micallef, L. & P., R. Application for Drawing 3 Set Area-Proportional Venn Diagrams. doi:http://www.eulerdiagrams.org/eulerAPE/.
84. Skrzypek, M. S. *et al.* The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.* **45**, D592–D596 (2017).
85. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**, e21800 (2011).
86. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).

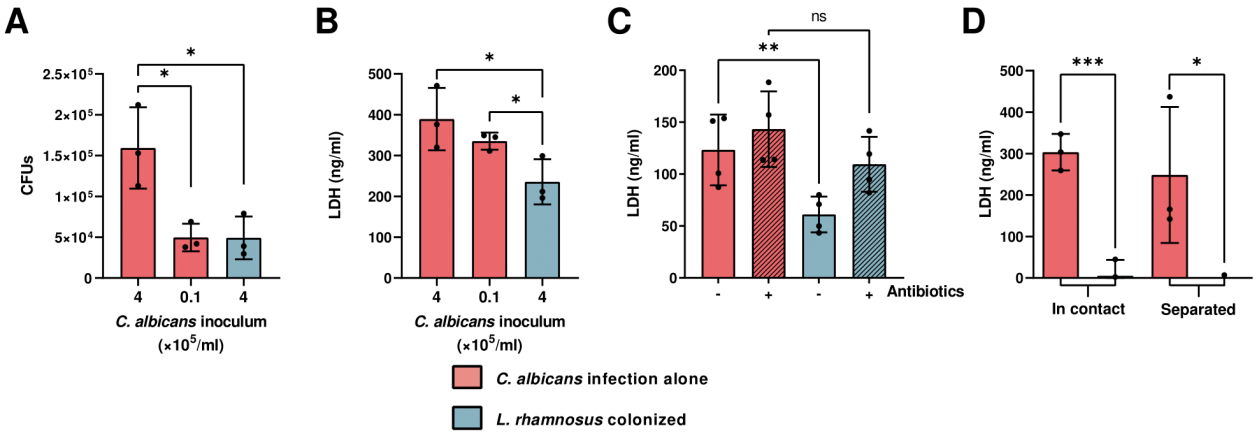

A
B
C
D
CFUs
LDH (ng/ml)
C. albicans inoculum
(×10^5/ml)
Antibiotics
In contact
Separated
ns
C. albicans infection alone
L. rhamnosus colonized

Figure 1

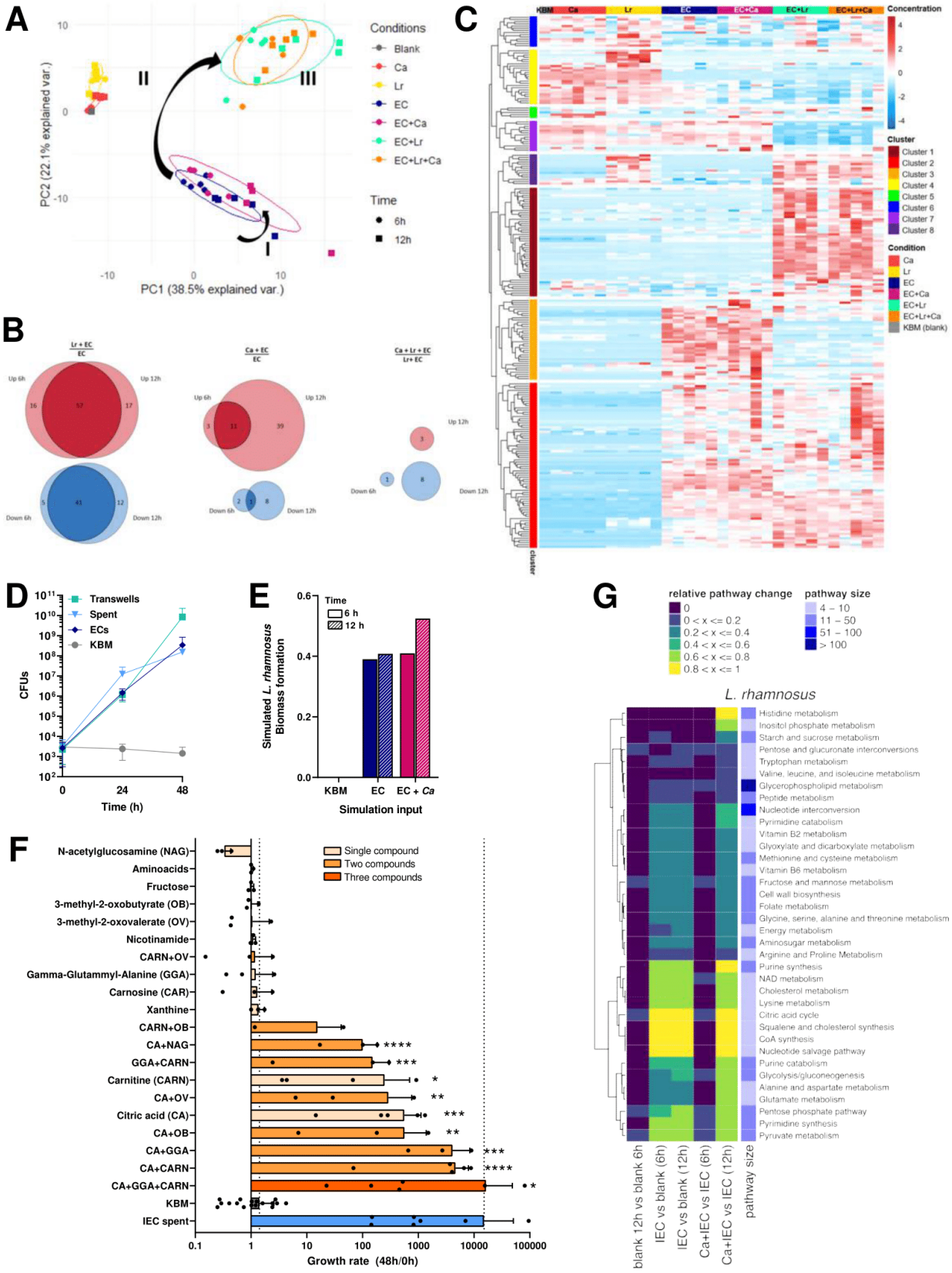

A
PC2 (22.1% explained var.)
PC1 (38.5% explained var.)
Conditions
Blank
Ca
Lr
EC
EC+Ca
EC+Lr
EC+Lr+Ca
Time
6h
12h
B
C
Concentration
Cluster
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Condition
KBM (blank)
D
Transwells
Spent
ECs
KBM
CFUs
Time (h)
E
Time
6 h
12 h
Simulated L. rhamnosus Biomass formation
KBM
EC
EC + Ca
Simulation input
F
Single compound
Two compounds
Three compounds
N-acetylglucosamine (NAG)
Aminoacids
Fructose
3-methyl-2-oxobutyrate (OB)
3-methyl-2-oxovalerate (OV)
Nicotinamide
CARN+OV
Gamma-Glutammyl-Alanine (GGA)
Carnosine (CAR)
Xanthine
CARN+OB
CA+NAG
GGA+CARN
Carnitine (CARN)
CA+OV
Citric acid (CA)
CA+OB
CA+GGA
CA+CARN
CA+GGA+CARN
KBM
IEC spent
Growth rate (48h/0h)
G
relative pathway change
pathway size
L. rhamnosus
Histidine metabolism
Inositol phosphate metabolism
Starch and sucrose metabolism
Pentose and glucuronate interconversions
Tryptophan metabolism
Valine, leucine, and isoleucine metabolism
Glycerophospholipid metabolism
Peptide metabolism
Nucleotide interconversion
Pyrimidine catabolism
Vitamin B2 metabolism
Glyoxylate and dicarboxylate metabolism
Methionine and cysteine metabolism
Vitamin B6 metabolism
Fructose and mannose metabolism
Cell wall biosynthesis
Folate metabolism
Glycine, serine, alanine and threonine metabolism
Energy metabolism
Aminosugar metabolism
Arginine and Proline Metabolism
Purine synthesis
NAD metabolism
Cholesterol metabolism
Lysine metabolism
Citric acid cycle
Squalene and cholesterol synthesis
CoA synthesis
Nucleotide salvage pathway
Purine catabolism
Glycolysis/gluconeogenesis
Alanine and aspartate metabolism
Glutamate metabolism
Pentose phosphate pathway
Pyrimidine synthesis
Pyruvate metabolism
blank 12h vs blank 6h
IEC vs blank (6h)
IEC vs blank (12h)
Ca+IEC vs IEC (6h)
Ca+IEC vs IEC (12h)
pathway size

Figure 2

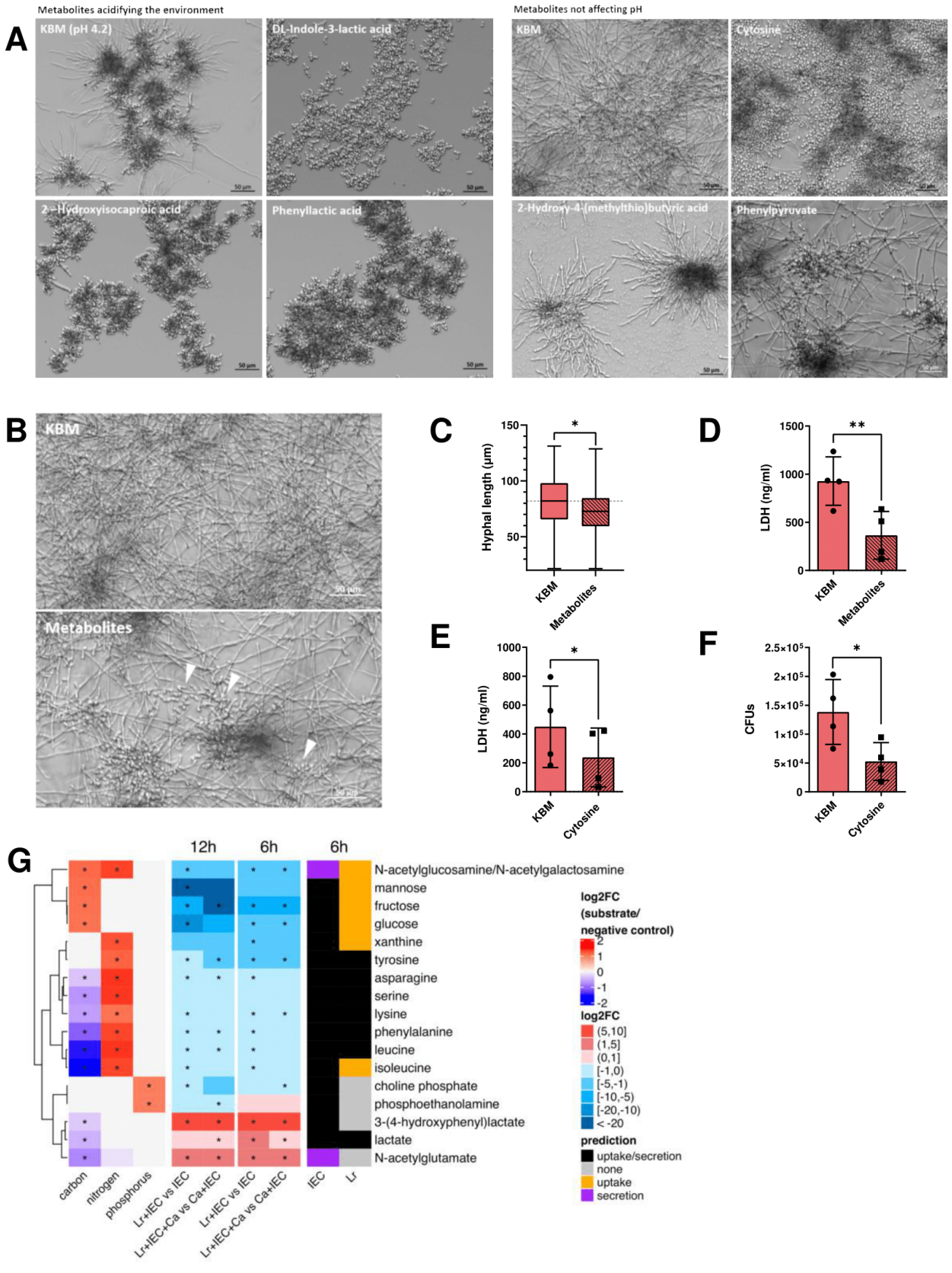

A
Metabolites acidifying the environment
KBM (pH 4.2)
DL-Indole-3-lactic acid
2-Hydroxyisocaproic acid
Phenyllactic acid
Metabolites not affecting pH
KBM
Cytosine
2-Hydroxy-4-(methylthio)butyric acid
Phenylpyruvate
B
KBM
Metabolites
C
Hyphal length (µm)
KBM
Metabolites
D
LDH (ng/ml)
KBM
Metabolites
E
LDH (ng/ml)
KBM
Cytosine
F
CFUs
KBM
Cytosine
G
12h
6h
6h
N-acetylglucosamine/N-acetylgalactosamine
mannose
fructose
glucose
xanthine
tyrosine
asparagine
serine
lysine
phenylalanine
leucine
isoleucine
choline phosphate
phosphoethanolamine
3-(4-hydroxyphenyl)lactate
lactate
N-acetylglutamate
carbon
nitrogen
phosphorus
Lr+IEC vs IEC
Lr+IEC+Ca vs Ca+IEC
Lr+IEC vs IEC
Lr+IEC+Ca vs Ca+IEC
IEC
Lr
log2FC (substrate/ negative control)
log2FC
(5,10]
(1,5]
(0,1]
[-1,0)
[-5,-1)
[-10,-5)
[-20,-10)
< -20
prediction
uptake/secretion
none
uptake
secretion

Figure 3

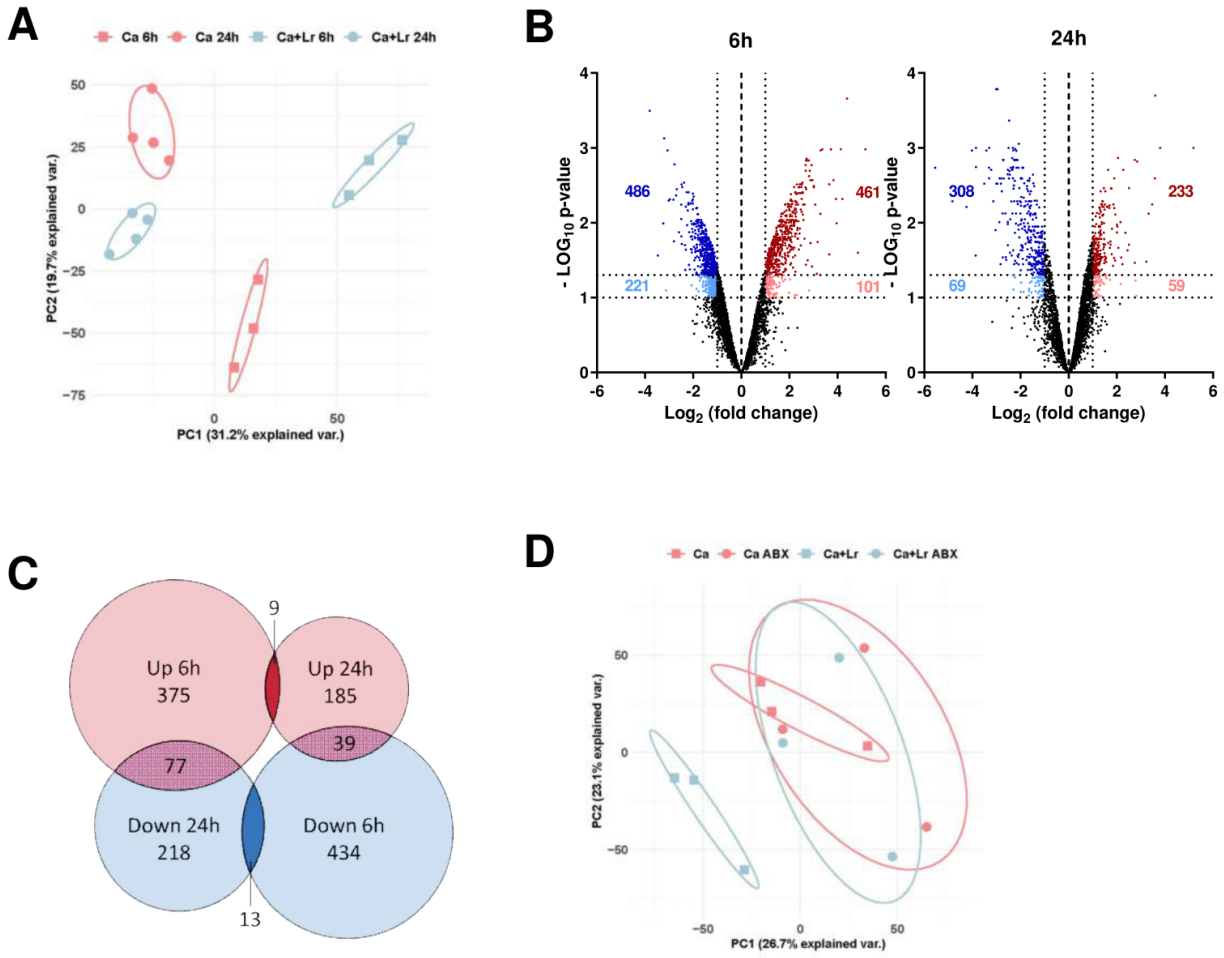

A
Ca 6h
Ca 24h
Ca+Lr 6h
Ca+Lr 24h
PC2 (19.7% explained var.)
PC1 (31.2% explained var.)
B
6h
24h
- LOG10 p-value
Log2 (fold change)
486
461
221
101
308
233
69
59
C
Up 6h
375
Up 24h
185
9
77
39
Down 24h
218
Down 6h
434
13
D
Ca
Ca ABX
Ca+Lr
Ca+Lr ABX
PC2 (23.1% explained var.)
PC1 (26.7% explained var.)

Figure 4

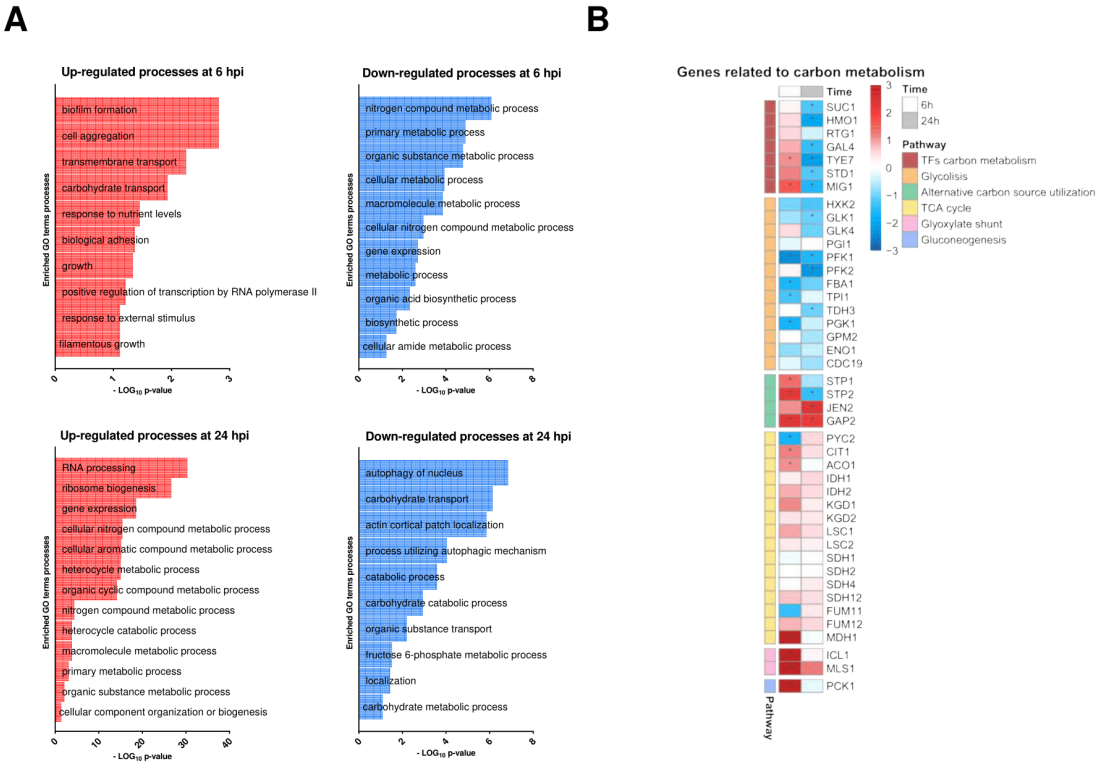

A
Up-regulated processes at 6 hpi
biofilm formation
cell aggregation
transmembrane transport
carbohydrate transport
response to nutrient levels
biological adhesion
growth
positive regulation of transcription by RNA polymerase II
response to external stimulus
filamentous growth
Enriched GO terms processes
- LOG10 p-value
Down-regulated processes at 6 hpi
nitrogen compound metabolic process
primary metabolic process
organic substance metabolic process
cellular metabolic process
macromolecule metabolic process
cellular nitrogen compound metabolic process
gene expression
metabolic process
organic acid biosynthetic process
biosynthetic process
cellular amide metabolic process
Up-regulated processes at 24 hpi
RNA processing
ribosome biogenesis
gene expression
cellular nitrogen compound metabolic process
cellular aromatic compound metabolic process
heterocycle metabolic process
organic cyclic compound metabolic process
nitrogen compound metabolic process
heterocycle catabolic process
macromolecule metabolic process
primary metabolic process
organic substance metabolic process
cellular component organization or biogenesis
Down-regulated processes at 24 hpi
autophagy of nucleus
carbohydrate transport
actin cortical patch localization
process utilizing autophagic mechanism
catabolic process
carbohydrate catabolic process
organic substance transport
fructose 6-phosphate metabolic process
localization
carbohydrate metabolic process
B
Genes related to carbon metabolism
Time
SUC1
HMO1
RTG1
GAL4
TYE7
STD1
MIG1
HXK2
GLK1
GLK4
PGI1
PFK1
PFK2
FBA1
TPI1
TDH3
PGK1
GPM2
ENO1
CDC19
STP1
STP2
JEN2
GAP2
PYC2
CIT1
ACO1
IDH1
IDH2
KGD1
KGD2
LSC1
LSC2
SDH1
SDH2
SDH4
SDH12
FUM11
FUM12
MDH1
ICL1
MLS1
PCK1
Pathway
Time
6h
24h
Pathway
TFs carbon metabolism
Glycolisis
Alternative carbon source utilization
TCA cycle
Glyoxylate shunt
Gluconeogenesis

Figure 5

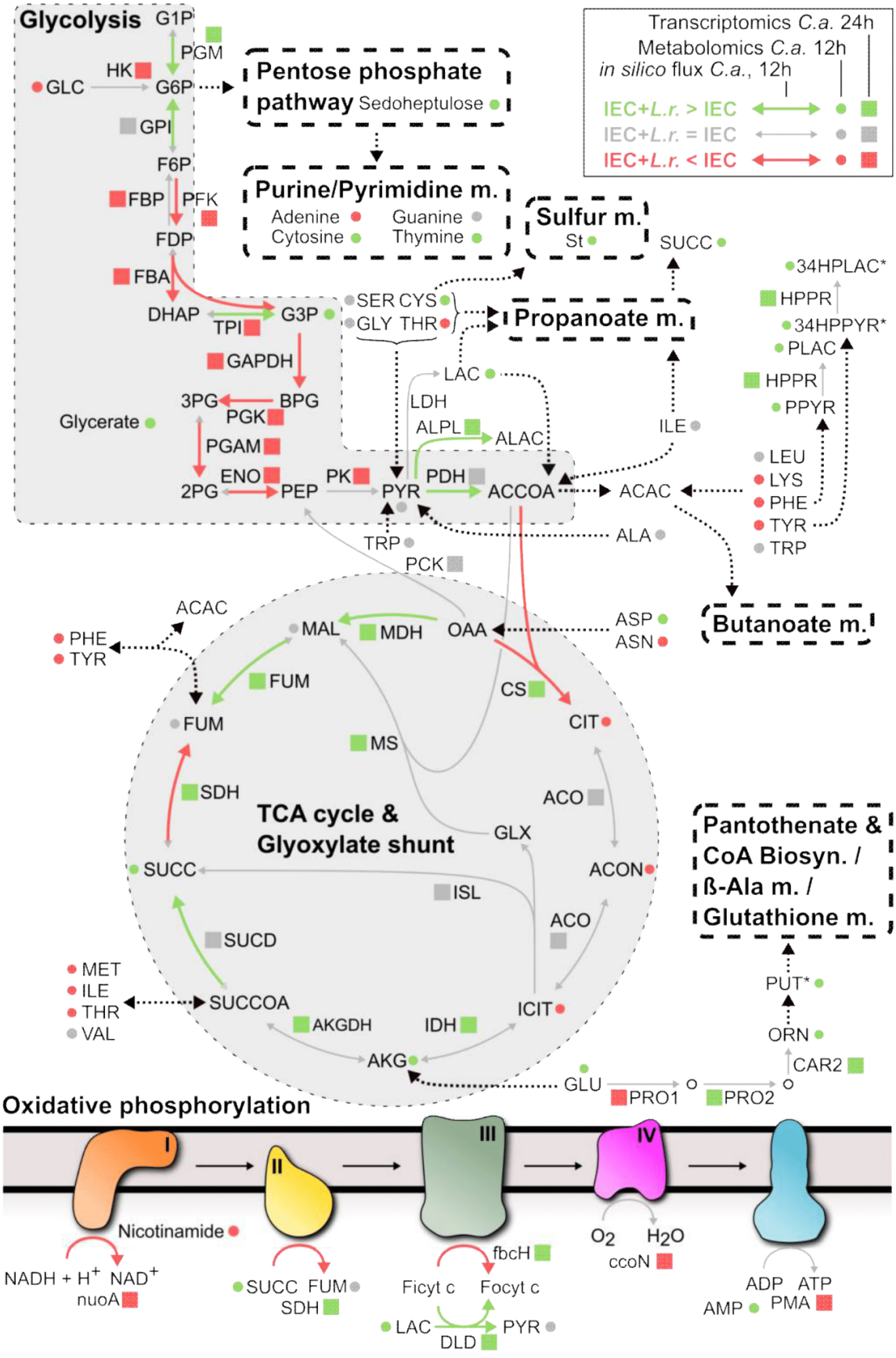

Glycolysis
G1P
PGM
HK
GLC
G6P
GPI
F6P
FBP
PFK
FDP
FBA
DHAP
TPI
G3P
GAPDH
3PG
BPG
PGK
Glycerate
PGAM
2PG
ENO
PEP
PK
PYR
PDH
ACCOA
Pentose phosphate pathway
Sedoheptulose
Purine/Pyrimidine m.
Adenine
Guanine
Cytosine
Thymine
Sulfur m.
St
SUCC
SER CYS
GLY THR
Propanoate m.
LAC
LDH
ALPL
ALAC
ILE
ACAC
LEU
LYS
PHE
TYR
TRP
ALA
TRP
PCK
34HPLAC*
HPPR
34HPPYR*
PLAC
HPPR
PPYR
Butanoate m.
ASP
ASN
MAL
MDH
OAA
CS
CIT
FUM
MS
ACO
GLX
ACON
ISL
SDH
TCA cycle &
Glyoxylate shunt
SUCC
SUCD
SUCCOA
AKGDH
IDH
ICIT
AKG
ACO
MET
ILE
THR
VAL
Pantothenate &
CoA Biosyn. /
ß-Ala m. /
Glutathione m.
PUT*
ORN
CAR2
GLU
PRO1
PRO2
Transcriptomics C.a. 24h
Metabolomics C.a. 12h
in silico flux C.a., 12h
IEC+L.r. > IEC
IEC+L.r. = IEC
IEC+L.r. < IEC
Oxidative phosphorylation
I
II
III
IV
Nicotinamide
NADH + H+ NAD+
nuoA
SUCC FUM
SDH
fbcH
Ficyt c
Focyt c
LAC
PYR
DLD
O2
H2O
ccoN
ADP ATP
AMP
PMA

Figure 6

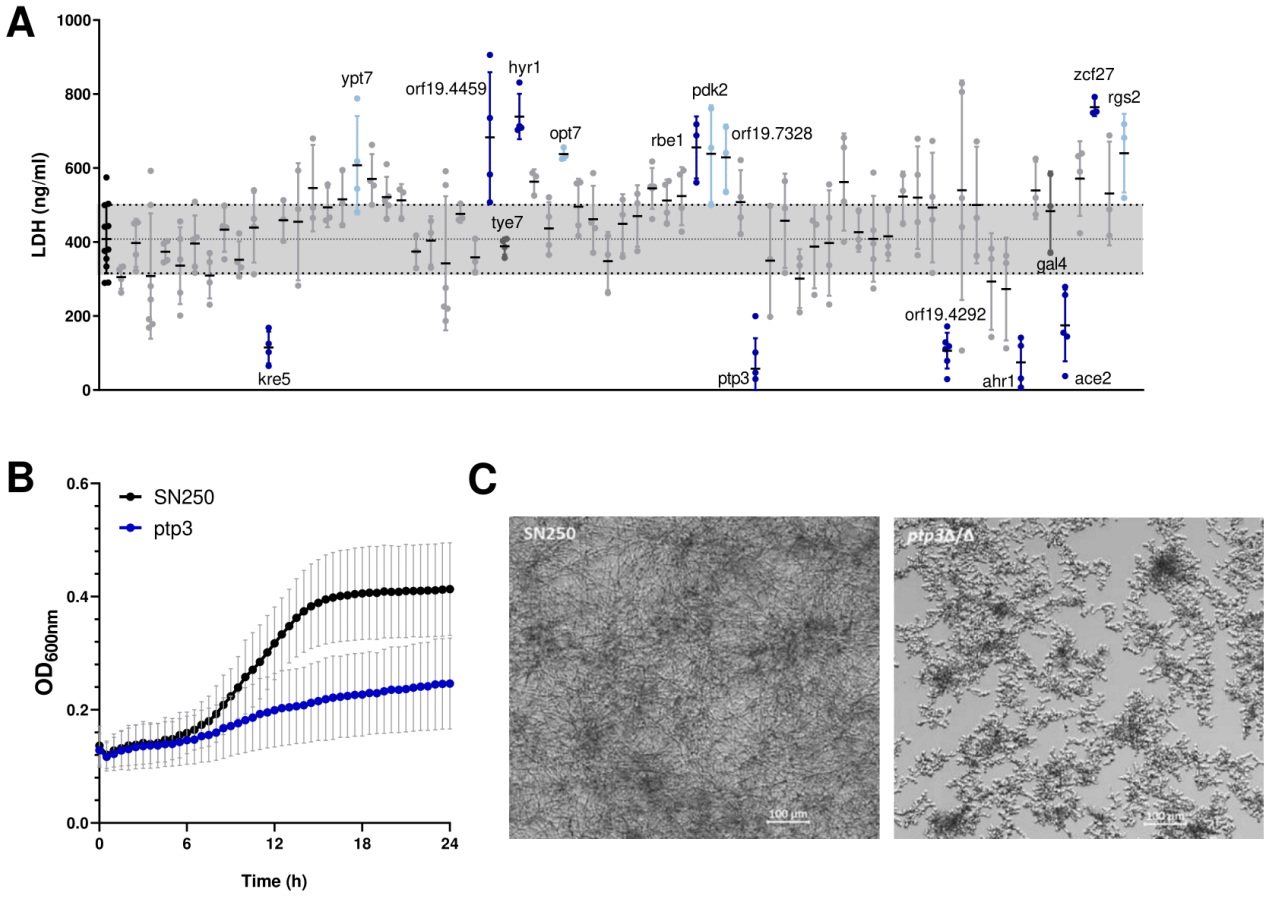

A
LDH (ng/ml)
1000
800
600
400
200
0
ypt7
orf19.4459
hyr1
opt7
tye7
rbe1
pdk2
orf19.7328
zcf27
rgs2
gal4
orf19.4292
kre5
ptp3
ahr1
ace2
B
SN250
ptp3
OD600nm
0.6
0.4
0.2
0.0
0
6
12
18
24
Time (h)
C
SN250
ptp3Δ/Δ
100 µm

Figure 7

Table 1: Metabolites tested to for their support of *L. rhamnosus* growth

Compound name	Experimental concentration	Company
Citric acid	5 mM	Roth
Carnitine	5 mM	Sigma
Gamma-glutamyl-alanine	5 mM	Sigma
Xanthine	0.05 mM	Sigma
3-methyl-2-oxovalerate	5 mM	Sigma
3-methyl-2-oxobutyrate	5 mM	Sigma
Non-Essential Amino Acids + L-Glutamine	1 mM	Gibco
Fructose	1 mM	Sigma
Nicotinamide	0.001 mM	Fluka
N-acetyl-glucosamine	1 mM	Sigma
Carnosine	5 mM	Sigma

Table 2: Tested metabolites to impair *C. albicans* filamentation

Compound name	Experimental concentration		Company
	alone	in combination	
2-deoxyinosine	50mM	-	Sigma
Allantoine	50mM	-	Roth
Alpha hydroxycaproate (HICA)	50mM	5mM	Sigma
Cytosine	50mM	5mM	Sigma
Histidine	50mM	-	Roth
Hydroxymethylbutyrate	50mM	5mM	Sigma
Indolelactate	50mM	5mM	Sigma
Sodium lactate	50mM	15mM	Sigma
Phenyllactic acid	50mM	5mM	Sigma
Phenylpyruvate	50mM	5mM	Sigma
Pipecolate	50mM	-	Sigma
Thymine	50mM	-	Sigma
Uridine	50mM	-	Roth

Supplementary material for

Lactobacillus rhamnosus colonisation antagonizes *Candida albicans* by forcing metabolic adaptations that compromise pathogenicity

Raquel Alonso-Roman[1], Antonia Last[1], Mohammad H. Mirhakkak[2], Jakob L. Sprague[1], Lars Möller[1], Peter Großmann[2], Katja Graf[1,3], Rena Gratz[1], Selene Mogavero[1], Slavena Vylkova[4], Gianni Panagiotou[2,5], Sascha Schäuble[2], Bernhard Hube[1,6], Mark S. Gresnigt[7]

Corresponding author: bernhard.hube@hki-jena.de

Affiliations:

1: Department of Microbial Pathogenicity Mechanisms, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Jena, Germany

2: Systems Biology and Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Jena, Germany

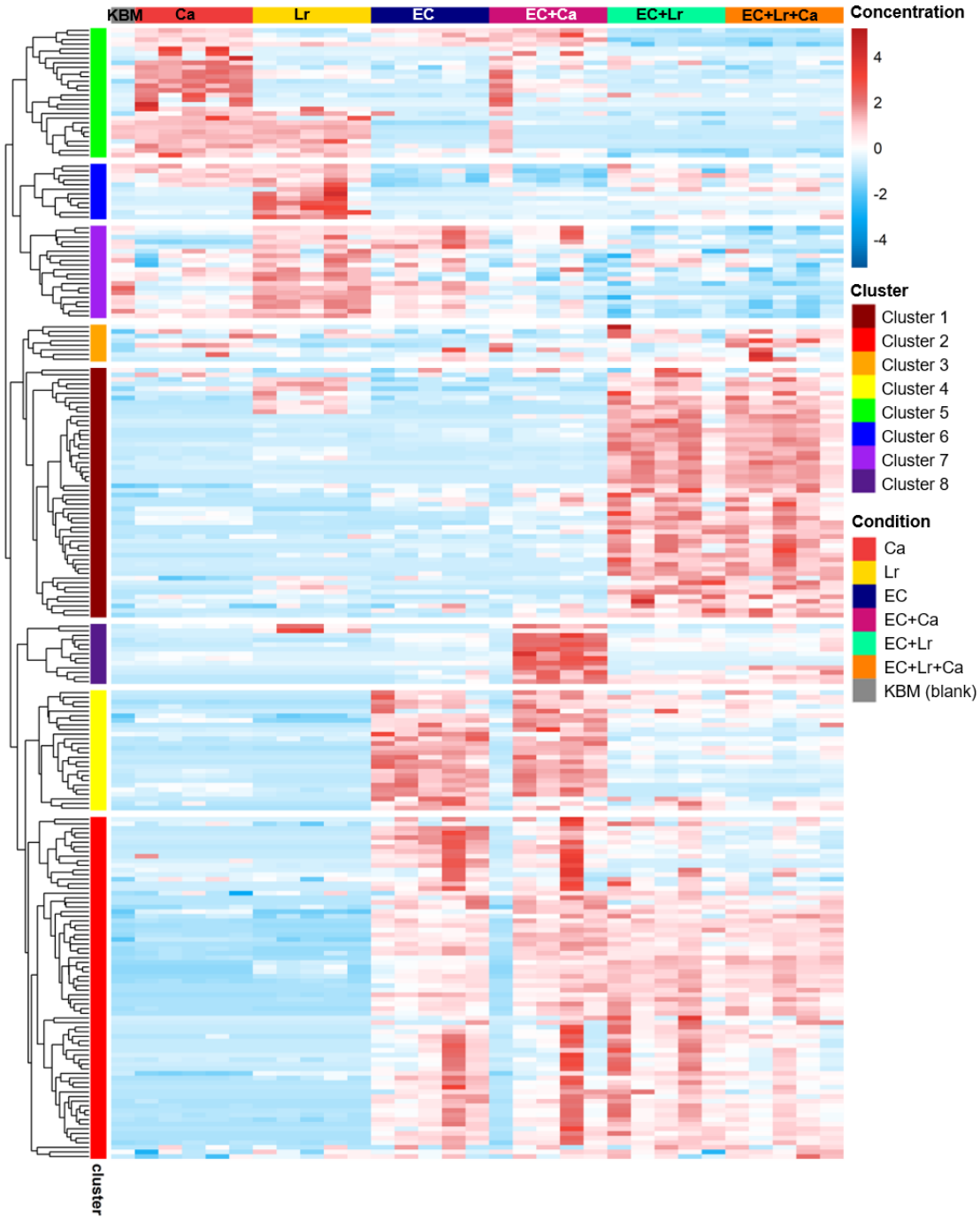
3: Dynamic42 GmbH, Jena, Germany

4: Septomics Research Centre, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knoell-Institute, Jena, Germany

5: Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, China

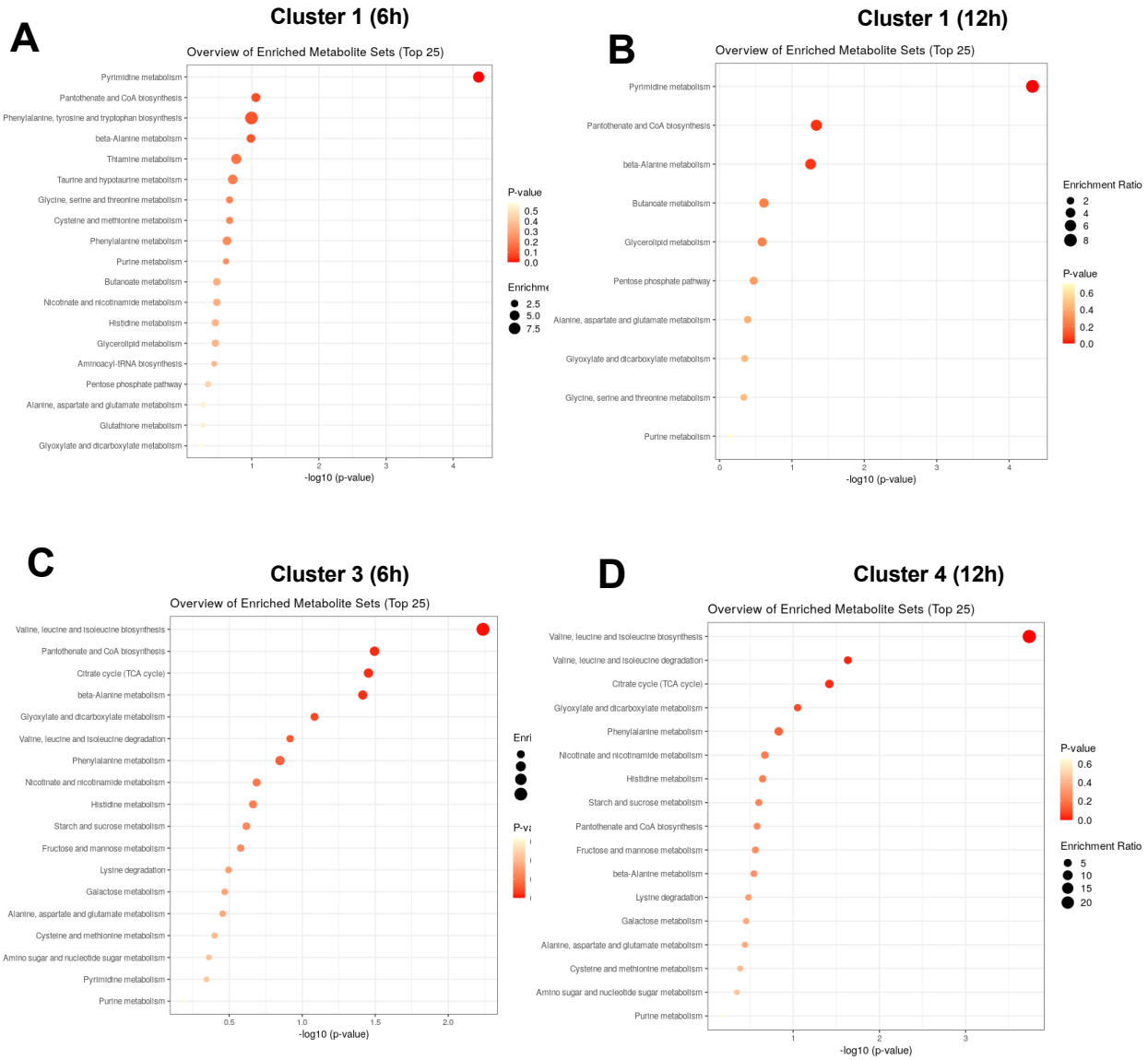
6: Institute of Microbiology, Friedrich Schiller University, Jena, Germany

7: Junior Research Group Adaptive Pathogenicity Strategies, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knoell-Institute, Jena, Germany



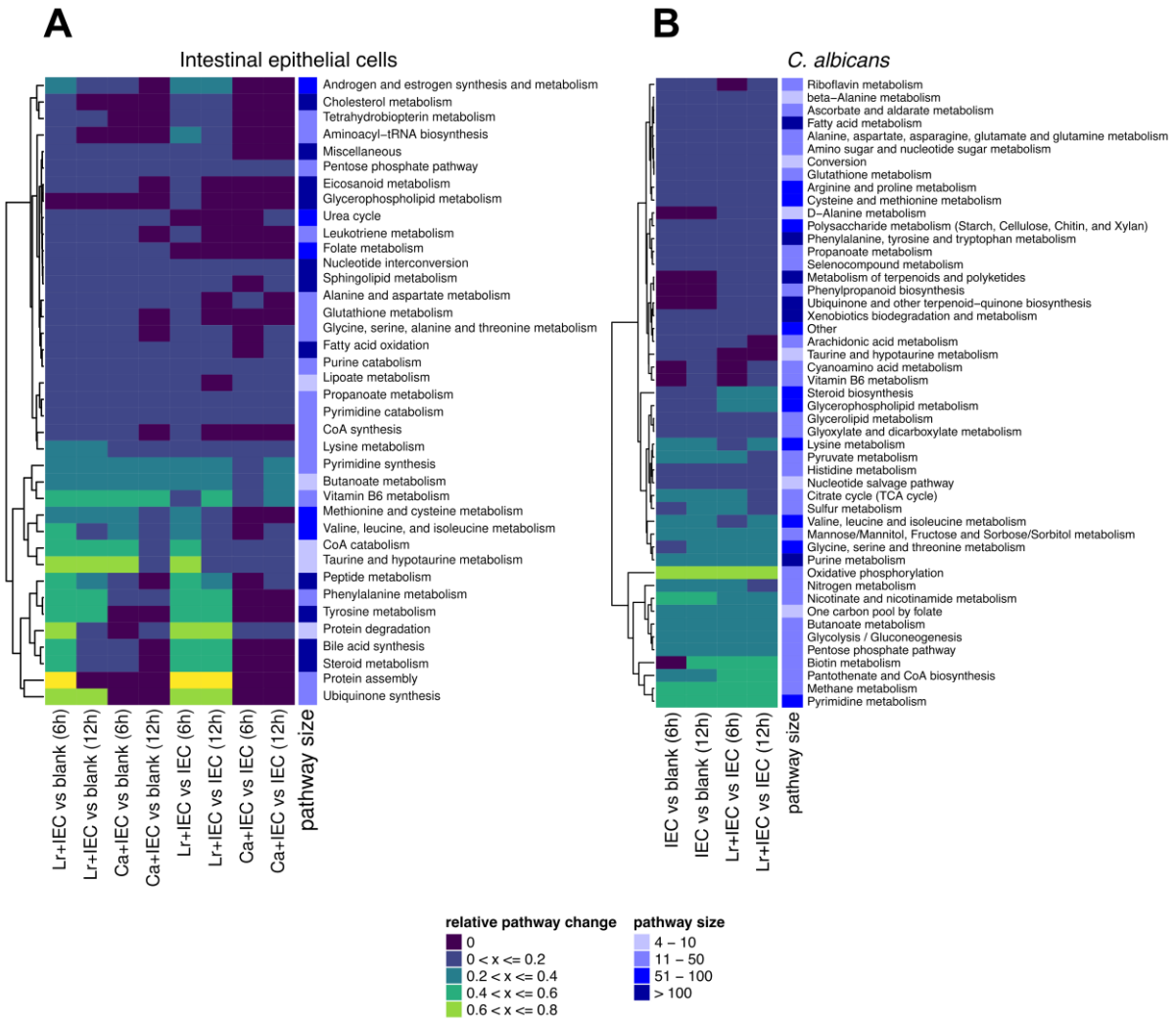
Supplementary Fig. 1: Untargeted metabolomics of in vitro intestinal epithelial L. rhamnosus colonization and C. albicans infection

Hierarchical clustering of the presence and absence of metabolites in the *in vitro* model at 12 hours post infection (hpi).



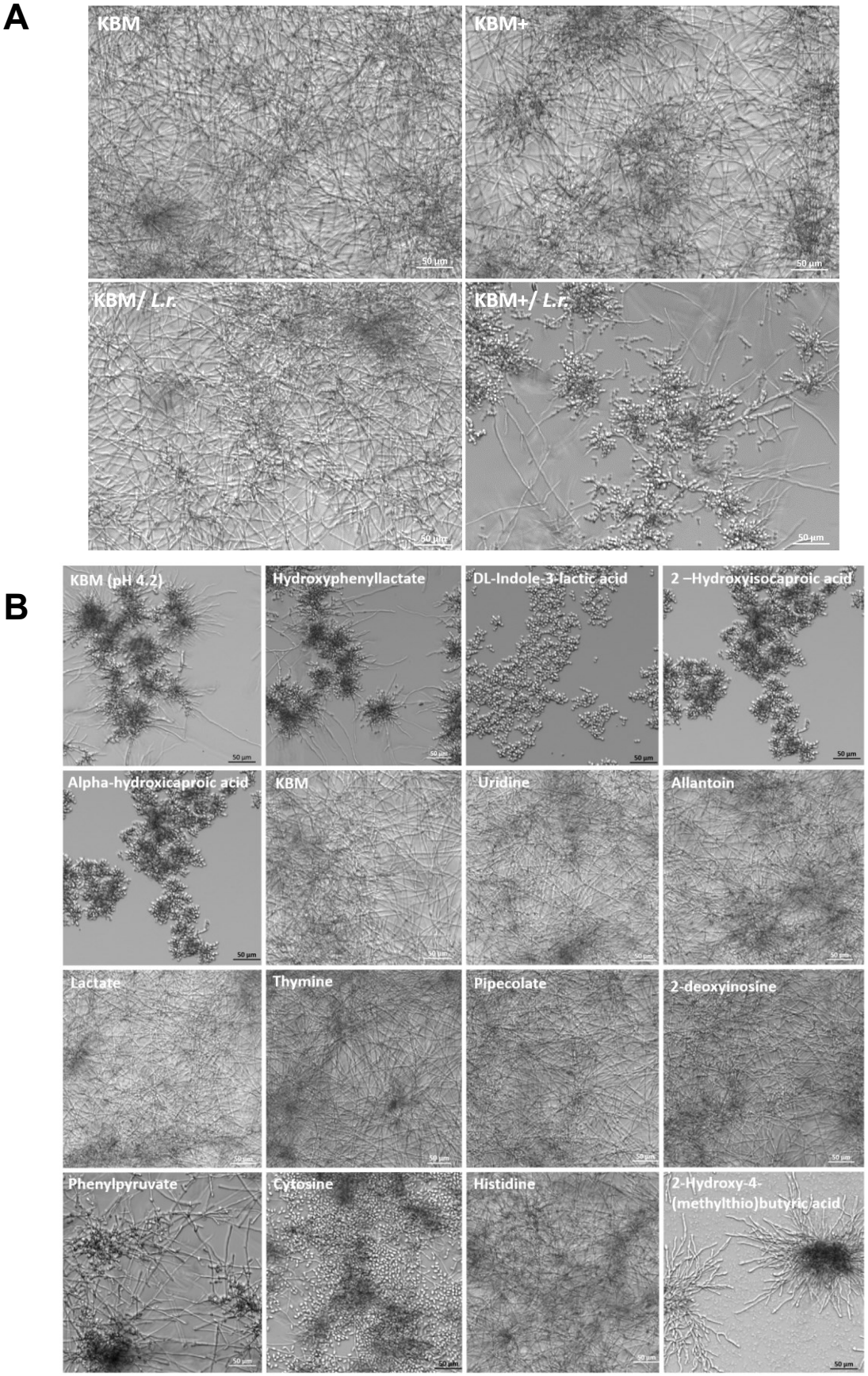
Supplementary Fig. 2: Enriched metabolic pathways

Overview of the top 25 enriched metabolic sets in the supernatants in the clusters of (A) metabolites produced by IECs and *L. rhamnosus* in synergy at 6h and (B) 12h; and (C) produced by IECs and consumed by *L. rhamnosus* at 6h and (D) 12h. Dot size represents enrichment ratio and colour of the dots represents *p*-value.



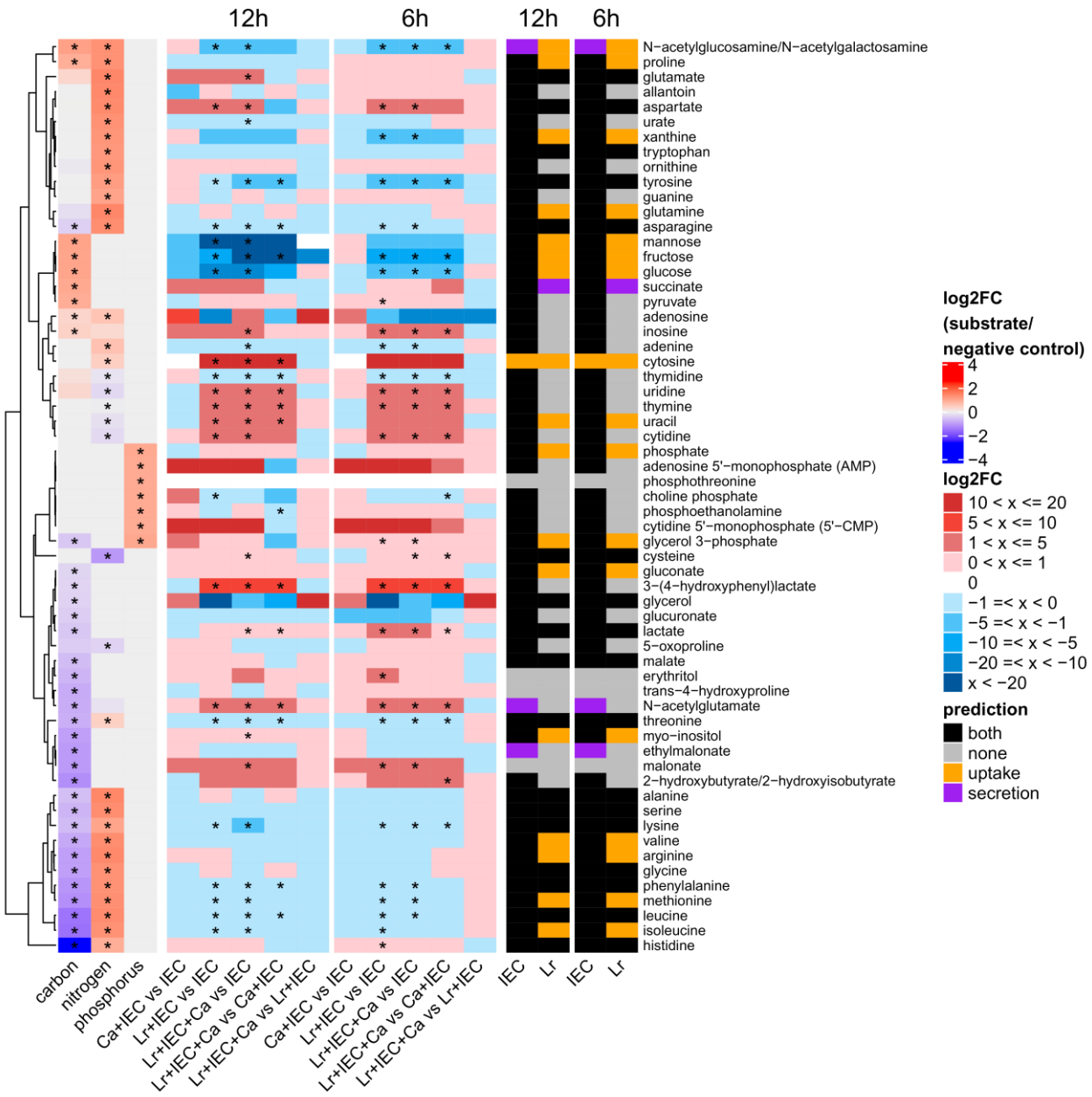
Supplementary Fig. 3: Pathway flux activity of IECs and C. albicans

Simulation of pathway activity levels in IECs (A) and in *C. albicans* (B) compared between different setups as indicated. Relative pathway change was determined by identifying the number of pathway-specific reactions for which feasible flux ranges differ according to flux variability analysis.



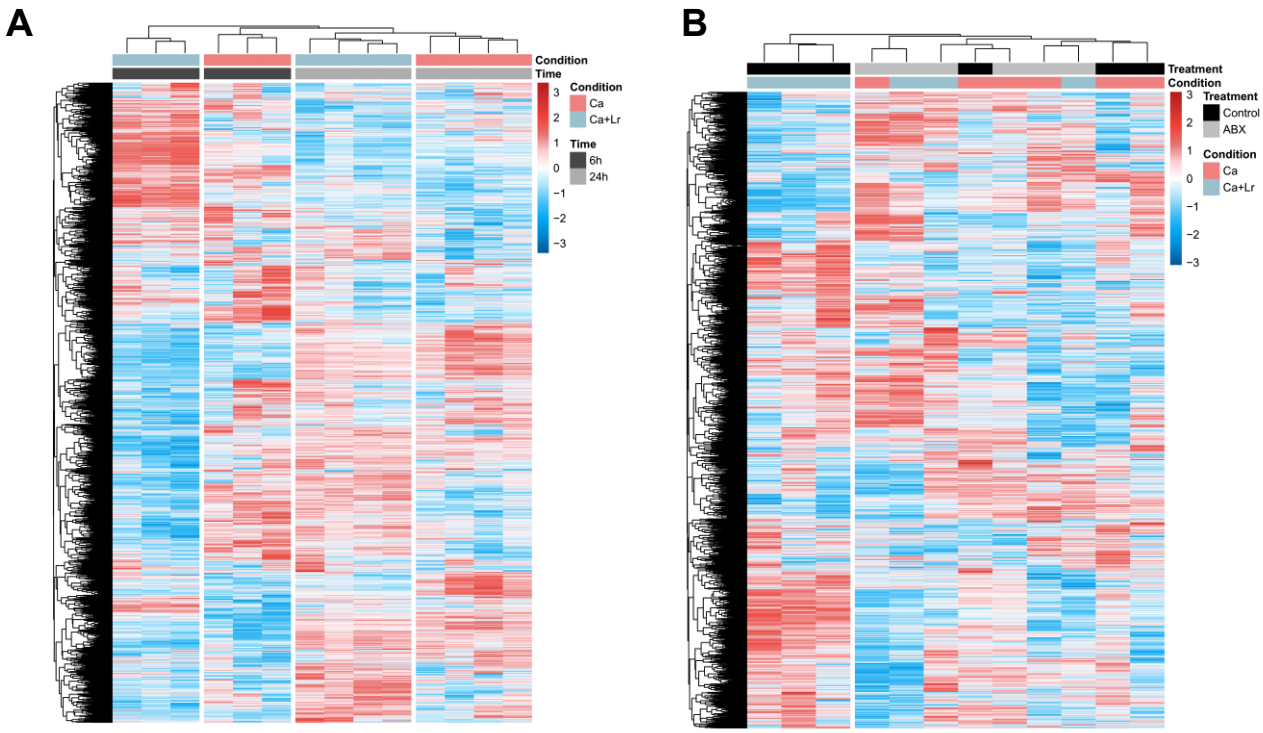
Supplementary Fig. 4: Impact of L. rhamnosus-colonization derived metabolites on C. albicans

(A) Representative images of *C. albicans* morphology changes grown in *L. rhamnosus*-conditioned or unconditioned media, after 20 h at 37°C with 5% CO_2 (n=3 biological replicates). (B) Representative images of *C. albicans* morphology grown in presence of all the metabolites tested, at neutral or acidic pH, after 20 h at 37°C with 5% CO_2 (n=2 biological replicates).



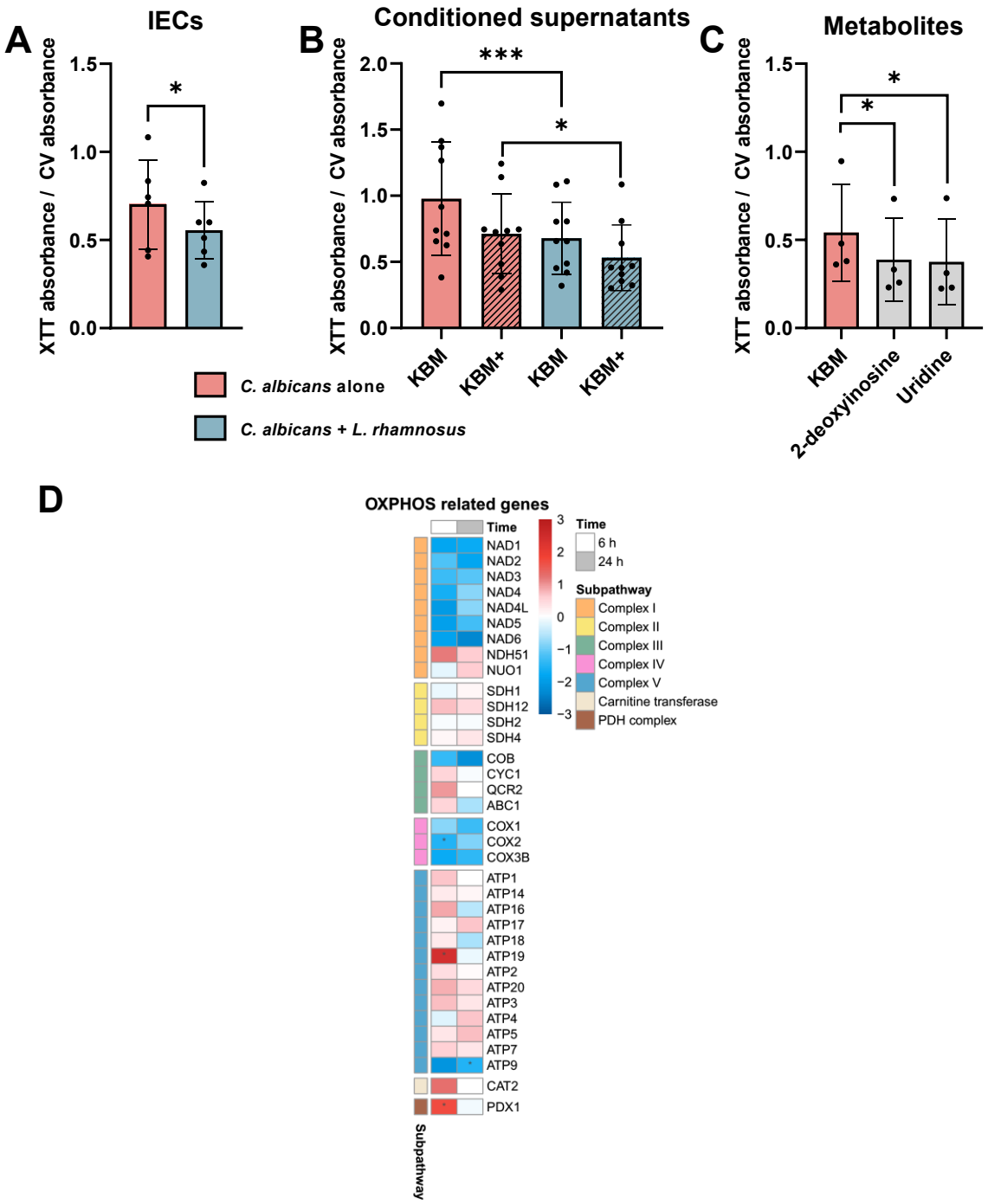
Supplementary Fig. 5: L. rhamnosus colonization creates an unfavourable metabolic environment for C. albicans

Phenotypic microarray growth experiments for wild-type *C. albicans* on a carbon, nitrogen or phosphorous source (left). Metabolome data measured at 6 and 12 h as well as metabolic modelling predictions (right) are indicated for the whole panel of metabolites. For metabolic modelling, media was adapted from metabolome data derived from IECs spent media or blank. Uptake or secretion was determined by identifying feasible flux ranges for metabolite-specific exchange reactions alongside optimization for biomass. * = $p_{adj} \leq 0.05$.



Supplementary Fig. 6: Changes in C. albicans gene expression upon infection of L. rhamnosus colonized IECs

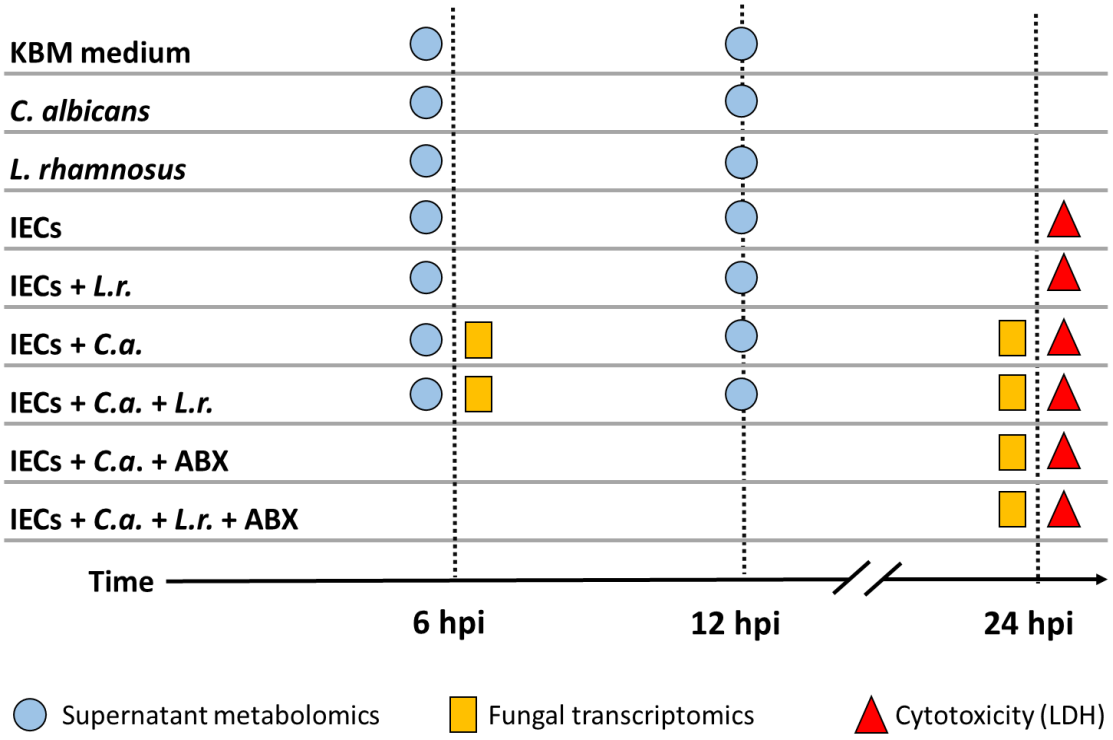
(A) Hierarchical clustering based on Euclidean distance of *C. albicans* gene expression at 6 and 24 h during *in vitro* infection of IECs in the presence and absence of *L. rhamnosus* colonization. Data summarized from n=3 and n=4 independent experiments at 6 and 24 hpi respectively. (B) Hierarchical clustering based on Euclidean distance of *C. albicans* gene expression during *in vitro* infection of IECs in the presence and absence of *L. rhamnosus* colonization and in the presence and absence of antibiotics. Data summarized from n=3 at 24 hpi.



Supplementary Fig. S7 L. rhamnosus modulates activity of C. albicans mitochondrial dehydrogenases

(A) Mitochondrial activity, represented as conversion of XTT by mitochondrial dehydrogenases (XTT absorbance) normalized with biomass (crystal violet (CV) absorbance, of *C. albicans* incubated in supernatants of *L. rhamnosus*-colonized or uncolonized IECS (1h, 37°C, 5% CO_2), or (B) in *L. rhamnosus*-conditioned or unconditioned KBM or KBM+ medium independently of host cells (1h, 37°C, 5% CO_2), or (C) in presence or absence of 50 mM uridine or 2-deoxyinosine (24h, 37°C, 5% CO_2). (D) Heatmap showing the transcriptional regulation of *C. albicans* OXPHOS genes as a result of *L. rhamnosus* colonization prior to infection at 6 and 24 hpi. Legend colour represents the Log_2 fold change of the regulation in presence *vs.* absence of *L. rhamnosus*. The asterisks (*)

represent significance, based on the criteria of a Log_2 fold change of > 1 or < -1 and a Bonferroni corrected *p*-value of < 0.05.



Supplementary Fig. 8: Study design

Overview of the samples, taken at different time points for transcriptional, metabolic, and cytotoxicity analysis in the *in vitro* model, in the presence of intestinal epithelial cells (IECs), *L. rhamnosus* (*L.r.*), *C. albicans* (*C.a.*) and antibiotics (ABX).

Supplementary Table 1: Metabolites of the clusters 1-8 shown in Fig. 2B (6 hpi).

Cluster 1 "Secreted by *L. rhamnosus* colonized IECs"
(N(1) + N(8))-acetylspermidine
1-carboxyethyltyrosine
1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)
2'-deoxycytidine
2'-deoxyinosine
2-hydroxy-3-methylvalerate
2-hydroxy-4-(methylthio)butanoic acid
2-hydroxybutyrate/2-hydroxyisobutyrate
2-hydroxyglutarate
2R,3R-dihydroxybutyrate
3-(4-hydroxyphenyl)lactate
3-formylindole
3-hydroxyoctanoate
3-indoleglyoxylic acid
4-methylthio-2-oxobutanoate
5-methyluridine (ribothymidine)
allantoin
alpha-hydroxyisocaproate
alpha-hydroxyisovalerate
benzoate
cysteine
cytidine
cytosine
dihydroorotate
glycerate
glycerophosphoglycerol
histidine
hypoxanthine
imidazole lactate
indolelactate
lactate
N-acetylglutamate
N-acetylproline
N-acetylthreonine
N-acetyltryptophan
N-acetyltyrosine
N-carbamoylaspartate
N-formylmethionine
nicotinamide riboside
nicotinate ribonucleoside
orotate
phenyllactate (PLA)
phenylpyruvate
phosphate
pipecolate
p-toluic acid
sedoheptulose
thioproline
thymine
trimethylamine N-oxide
uracil
uridine

Cluster 2 "Secreted by IECs"
1-carboxyethylisoleucine
1-carboxyethylleucine
1-carboxyethylphenylalanine
1-carboxyethylvaline
1-methylnicotinamide
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)
1-stearoyl-2-oleoyl-GPC (18:0/18:1)
2'-deoxycytidine
2'-deoxyuridine
2'-O-methylcytidine
3-hydroxy-3-methylglutarate
3-hydroxybutyrate (BHBA)

Cluster 2 "Secreted by IECs" Continued
3-hydroxyisobutyrate
4-acetamidobutanoate
4-imidazoleacetate
7-methylguanine
alanine
Allantoin
alpha-ketoglutarate
4-guanidinobutanoate
arabonate/xylonate
argininosuccinate
Benzoate
beta-alanine
Betaine
cholesterol
choline phosphate
creatine
creatinine
cystathionine
dimethylarginine (SDMA + ADMA)
erythritol
erythronate
ethylmalonate
fumarate
gamma-glutamyl-alpha-lysine
gamma-glutamylthreonine
glucuronate
glycerol 3-phosphate
glycerophosphoethanolamine
glycerophosphoglycerol
glycerophosphoinositol
glycerophosphorylcholine (GPC)
glycerophosphoserine
glycine
guanine
gulonate
homoarginine
hypotaurine
hypoxanthine
kynurenine
Lactate
malate
mannitol/sorbitol
methionine sulfone
N1-methyladenosine
N1-methylinosine
N6,N6,N6-trimethyllysine
N-acetylalanine
N-acetyl-isoputreanine
N-acetylleucine
N-acetylneuraminate
nicotinamide ribonucleotide (NMN)
nicotinamide riboside
N-monomethylarginine
ornithine
palmitoyl sphingomyelin (d18:1/16:0)
pantothenate
p-cresol sulfate
phenol red
phenol sulfate
phenylacetylglycine
proline
pseudouridine
pyridoxal
pyridoxamine
pyridoxate
pyridoxine (Vitamin B6)
ribonate

Cluster 2 "Secreted by IECs" Continued
succinate
sulfate
taurine
thiamin (Vitamin B1)
threonate
trans-4-hydroxyproline
Urate

Cluster 3 "Secreted by IECs consumed by *L. rhamnosus"*
1-ribosyl-imidazoleacetate
2'-O-methyluridine
3-methyl-2-oxobutyrate
3-ureidopropionate
4-methyl-2-oxopentanoate
5-methylthioadenosine (MTA)
acetylcarnitine (C2)
aconitate [cis or trans]
carnitine
carnosine
citrate
cysteinylglycine disulfide
deoxycarnitine
fructose
gamma-glutamylalanine
gamma-glutamylglutamine
gamma-glutamylhistidine
gamma-glutamylisoleucine
gamma-glutamylleucine
gamma-glutamylmethionine
gamma-glutamylserine
Hippurate
Isocitrate
N-acetylglucosamine/N-acetylgalactosamine
N-acetylmethionine
nicotinamide
S-1-pyrroline-5-carboxylate
Xanthine

Cluster 4 "Consumed by IECs"
2'-deoxyadenosine
3-sulfo-L-alanine
Adenosine
alpha-lipoate
Arginine
Aspartate
cysteine s-sulfate
cysteine sulfinic acid
Cystine
Gluconate
Glutamate
Glutamine
Inosine
Isoleucine
Leucine
mannonate
methionine
methionine sulfoxide
N-acetylarginine
N-acetylglutamine
phenylalanylhydroxyproline
phosphoethanolamine
phosphothreonine
Serine
Valine

Cluster 5 "Secreted by *C. albicans*"
2,3-dihydroxyisovalerate
4-hydroxyphenylpyruvate
folate
glutarate (C5-DC)
glycerol
kynurenate

Cluster 6 "Inconclusive"
2-oxoarginine
3-hydroxybutyroylglycine
5-oxoproline
adenosine 5'-monophosphate (AMP)
biotin
choline
cytidine 5'-monophosphate (5'-CMP)
myo-inositol
N-acetylasparagine
N-acetylaspartate (NAA)
N-acetylhistidine
pyroglutamine
tryptophan

cluster 7 "Consumed by *L. rhamnosus* colonized IECs"
adenine
arabitol/xylitol
asparagine
cysteine-glutathione disulfide
fructosyllysine
gamma-glutamylvaline
Glucose
Lysine
phenylalanine
Ribitol
riboflavin (Vitamin B2)
threonine
thymidine
Tyrosine

Cluster 8 "Secreted by *L. rhamnosus* alone"
alpha-ketoglutaramate
argininate
diacetylspermidine
glutamine conjugate of C6H10O2 (2)
malonate
methylsuccinate
mevalonate
mevalonolactone
N-acetylleucine
N-acetylisoleucine
N-acetylphenylalanine
N-acetylputrescine
N-acetylserine
pyruvate

Supplementary Table 2: Metabolites of the clusters 1-8 shown in Fig. S1 (12 hpi).

Cluster 1 "Secreted by *L. rhamnosus* colonized IECs"
(N(1) + N(8))-acetylspermidine
1-carboxyethyltyrosine
2'-deoxyinosine
2-hydroxy-3-methylvalerate
2-hydroxy-4-(methylthio)butanoic acid
2-hydroxybutyrate/2-hydroxyisobutyrate
2-hydroxyglutarate
2R,3R-dihydroxybutyrate
3-(4-hydroxyphenyl)lactate
3-formylindole
3-hydroxyoctanoate
3-indoleglyoxylic acid
3-ureidopropionate
5-methyluridine (ribothymidine)
alpha-hydroxyisocaproate
alpha-hydroxyisovalerate
cytidine
cytosine
dihydroorotate
glycerate
indolelactate
mevalonate
mevalonolactone
N-acetylglutamate
N-acetylthreonine
N-acetyltryptophan
N-acetyltyrosine
N-carbamoylaspartate
nicotinate ribonucleoside
phenyllactate (PLA)
pipecolate
sedoheptulose
thymine
uracil
uridine

Cluster 2 "Secreted by IECs"
1-carboxyethylisoleucine
1-carboxyethylleucine
1-carboxyethylphenylalanine
1-carboxyethylvaline
1-methylnicotinamide
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)
1-stearoyl-2-oleoyl-GPC (18:0/18:1)
2'-deoxycytidine
2'-deoxyuridine
2'-O-methylcytidine
3-hydroxy-3-methylglutarate
3-hydroxybutyrate (BHBA)
3-hydroxyisobutyrate
4-imidazoleacetate
7-methylguanine
Alanine
Allantoin
alpha-ketoglutarate
arabonate/xylonate
argininosuccinate
Benzoate
beta-alanine
betaine

Cluster 2 "Secreted by IECs" continued
cholesterol
choline phosphate
creatine
creatinine
cystathionine
dimethylarginine (SDMA + ADMA)
erythritol
erythronate
ethylmalonate
fumarate
gamma-glutamylthreonine
glucuronate
glycerol 3-phosphate
glycerophosphoethanolamine
Glycerophosphoglycerol
Glycerophosphoinositol
glycerophosphorylcholine (GPC)
Glycerophosphoserine
Gulonate
Homoarginine
Hypotaurine
Hypoxanthine
Lactate
malate
mannitol/sorbitol
methionine sulfone
N1-methyladenosine
N1-methylinosine
N6,N6,N6-trimethyllysine
N-acetylalanine
N-acetyl-isoputreanine
N-acetylleucine
N-acetylneuraminate
nicotinamide riboside
N-monomethylarginine
ornithine
Palmitoyl sphingomyelin (d18:1/16:0)
pantothenate
p-cresol sulfate
phenol sulfate
pseudouridine
pyridoxal
pyridoxamine
pyridoxate
ribonate
Sulfate
taurine
trans-4-hydroxyproline
Urate

Cluster 3 "Inconclusive"
1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)
3-hydroxybutyroylglycine
benzoate
cholesterol
imidazole lactate
methylsuccinate
N-acetylaspartate (NAA)
N-acetylproline

Cluster 4 "Secreted by IECs and consumed by *L. rhamnosus*"
1-ribosyl-imidazoleacetate
3-methyl-2-oxobutyrate
3-methyl-2-oxovalerate
4-methyl-2-oxopentanoate
5-methylthioadenosine (MTA)
acetylcarnitine (C2)
aconitate [cis or trans]
carnitine
carnosine
citrate
deoxycarnitine
fructose
gamma-glutamylalanine
gamma-glutamylglutamine
gamma-glutamylhistidine
gamma-glutamylisoleucine
gamma-glutamylleucine
gamma-glutamylmethionine
gamma-glutamylserine
gamma-glutamylvaline
hippurate
isocitrate
N-acetylglucosamine/N-acetylgalactosamine
N-acetylmethionine
nicotinamide
S-1-pyrroline-5-carboxylate
xanthine

Cluster 5 "Secreted by *C. albicans*"
2,3-dihydroxyisovalerate
2'-deoxyadenosine
3-sulfo-L-alanine
4-methylthio-2-oxobutanoate
adenosine
alpha-lipoate
arabitol/xylitol
aspartate
benzoate
cysteine s-sulfate
glycerol
imidazole lactate
kynurenate
N-acetylleucine
N-acetylputrescine
phenylalanylhydroxyproline
ribitol

Cluster 6 "Secreted by *L. rhamnosus* alone"
adenosine
argininate
diacetylspermidine
gluconate
glutamine conjugate of C6H10O2 (2)
Inosine
malonate
N-acetylisoleucine
N-acetylphenylalanine
N-acetylserine
phosphothreonine

cluster 7 "Consumed by *L. rhamnosus* colonized IECs"
adenine
arginine
asparagine
biotin
cystine
fructosyllysine
glucose
glutamine
HEPES
isoleucine
leucine
lysine
mannitol/sorbitol
mannose
methionine
phenylalanine
serine
Threonine
tryptophan
valine

Cluster 8 "Secreted by *C. albicans* infected IECs"
cysteine-glutathione disulfide
indoleacetate
isovalerate (i5:0)
nicotinamide ribonucleotide (NMN)

Supplementary Table 3: Abbreviations used in Fig.6

Abbreviation	Full name
Reactions	
ACO	Aconitate Hydratase
AKGDH	alpha-Ketoglutarate Dehydrogenase
ALPL	2-Acetolactate Pyruvate-Lyase
CAR2	Ornithine Aminotransferase
ccoN	Cytochrome c Oxidase

CS	Citrate Synthase
DLD	D-lactate Dehydrogenase
ENO	Enolase
FBA	Fructose-Bisphosphate Aldolase
fbcH	Ubiquinol-Cytochrome c Reductase
FBP	Fructose-1,6-Bisphosphatase
FUM	Fumarate Hydratase
G6PD	Glucose-6-phosphate 1-Dehydrogenase
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GPI	Glucose-6-Phosphate Isomerase
HK	Hexokinase
HPPR	Hydroxyphenylpyruvate Reductase
IDH	Isocitrate Dehydrogenase
ISL	Isocitrate Lyase
LDH	L-Lactate Dehydrogenase
MDH	Malate Dehydrogenase
MS	Malate Synthase
nuoA	NADH-Quinone Oxidoreductase
PCK	Phosphoenolpyruvate Carboxykinase
PDH	Pyruvate Dehydrogenase
PFK	6-Phosphofructokinase
PGAM	Phosphoglycerate Mutase
PGD	6-Phosphogluconate Dehydrogenase
PGK	Phosphoglycerate Kinase
PGL	6-Phosphogluconolactonase
PGM	Phosphoglucomutase
PK	Pyruvate Kinase
PMA	Plasma Membrane ATPase
PRO1	gamma-glutamyl kinase
PRO2	gamma-glutamyl phosphate reductase
RPE	Ribulose-Phosphate 3-Epimerase
RPI	Ribose 5-Phosphate Isomerase
SDH	Succinate Dehydrogenase
SUCD	Succinyl-CoA Synthetase
TAL	Transaldolase
TK	Transketolase
TPI	Triosephosphate Isomerase
Metabolites	
2PG	2-Phospho-D-glycerate
34HPLAC*	3,4-Dihydroxyphenylpyruvate (mapped to Phenylpyruvate measurements)
34HPPYR*	3,4-Dihydroxyphenyllactate (mapped to Phenyllactate measurements)
3PG	3-Phospho-D-glycerate
ACAC	Acetoacetate
ACCOA	Acetyl-CoA
ADP	Adenosine 5'-diphosphate
ALA	Alanine
ALAC	2-Acetolactate
AMP	Adenosine 5'-monophosphate
ATP	Adenosine 5'-triphosphate
BPG	D-Glycerate 1,3-diphosphate
DHAP	Dihydroxyacetone phosphate
F6P	beta-D-Fructose 6-phosphate
FDP	beta-D-Fructose 1,6-bisphosphate
Ficyt c	Ferricytochrome c
Focyt c	Ferrocytochrome c

G1P	alpha-D-Glucose 1-phosphate
G3P	D-Glycerate 3-phosphate
G6P	alpha-D-Glucose 6-phosphate
GLC	Glucose
GLU	Glutamate
H20	Water
H+	Hydrogen
ILE	Isoleucine
LAC	Lactate
LEU	Leucin
LYS	Lysine
MET	Methionine
NADH	Nicotinamide adenine dinucleotide (reduced)
NAD+	Nicotinamide adenine dinucleotide
O2	Oxygen
ORN	Ornithine
PEP	Phosphoenolpyruvate
PHE	Phenylalanine
PLAC	Phenyllactate
PPYR	Phenylpyruvate
PRCOA	Propanoyl-CoA
PRPP	5-Phosphoribosyl diphosphate
PUT*	Putrescine (mapped to N-acetylputrescine)
PYR	Pyruvate
St	Sulfate
SUCC	Succinate
THR	Threonine
TRP	Tryptophan
TYR	Tyrosine
VAL	Valine

Supplementary Table 4: Overview of screened deletion mutants for damage potential
Damage phenotype of deletion mutants from the Noble collection (Noble *et al.* 2010 *Nat Genetics*) corresponding to differentially regulated genes during infection of *L. rhamnosus* colonized IECs. Mutant phenotype is classified as hypovirulent or hypervirulent when the damage caused to IECs was significantly lower or higher than the wild type, respectively, and non-significant (ns) when it was not.

KO gene	Mutant phenotype	KO gene	Mutant phenotype
ace2	Hypovirulent	orf19.5565	ns
ahr1	Hypovirulent	orf19.6874	ns
aox1	ns	orf19.7328	Hypervirulent
aro80	ns	orf19.7370	ns
asm3	ns	orf19.7516	ns
cfl2	ns	orf19.7554	ns
cht2	ns	pdk2	Hypervirulent
cip1	ns	pex4	ns
cyb5	ns	pga13	ns
fcr1	ns	pga45	ns
gal4	ns	pga7	ns
gis2	ns	pra1	ns
het1	ns	prk1	ns
hgt6	ns	prn2	ns
hyr1	Hypervirulent	prn4	ns
kre5	Hypovirulent	ptp3	Hypovirulent
mal2	ns	pwp1	ns
mec3	ns	rbe1	Hypervirulent
mid1	ns	rbt4	ns
nik1	ns	rgs2	Hypervirulent
opt4	ns	rim13	ns
opt7	Hypervirulent	sef1	ns
orf19.1314	ns	snq2	ns
orf19.173	ns	sod6	ns
orf19.215	ns	swi4	ns
orf19.3156	ns	tye7	ns
orf19.3395	ns	uga32	ns
orf19.3720	ns	vid27	ns
orf19.3982	ns	wor2	ns
orf19.4014	ns	ycp4	ns
orf19.4292	Hypovirulent	yhb1	ns
orf19.4445	ns	ypt7	Hypervirulent
orf19.4459	Hypervirulent	ywp1	ns
orf19.449	ns	zcf27	Hypervirulent
orf19.4843	ns		

No 6. *Candida* expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions

FORM 1

Manuscript No. 6

Manuscript title: Candida expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions

Authors: Bastian Seelbinder, Zoltan Lohinai, Ruben Vazquez-Uribe, Xiuqiang Chen, Silvia Lopez, **Mohammad H. Mirhakkak**, Anja Wellejus, Glen J. Weiss, Morten Sommer, Gianni Panagiotou

Bibliographic information:

The candidate is

☐ First author, ☐ Co-first author, ☐ Corresponding author, ☒ Co-author.

Status: In preparation

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Seelbinder, B.	30%	90%		80%	
Lohinai, Z.	5%		70%	5%	70%
Panagiotou, G.	60%			15%	20%
Mirhakkak, M.		5%			
Others	5%	5%	30%		10%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the faculty)

FORM 2

Manuscript No. 6

Short reference Seelbinder and Lohinai et al. (2021), *In preparation*

Contribution of the doctoral candidate

My contribution to the manuscript was genome-scale metabolic modelling analyses.

Figures #	X	Approximate contribution of the doctoral candidate to the figures: 5%

Candida expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions

Bastian Seelbinder[1, †], Zoltan Lohinai[2, †], Ruben Vazquez-Uribe[3], Xiuqiang Chen[1], Silvia Lopez[4], Mohammad Mirhakkak[1], Anja Wellejus[4], Glen J. Weiss[5], Morten O. A. Sommer[3], Gianni Panagiotou[1,6,7,*]

Affiliations

[1] Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Systems Biology & Bioinformatics, Beutenbergstraße 11A, 07745 Jena, Germany
[2] National Koranyi Institute of Pulmonology, Pihenő út 1-3., H-1121, Budapest, Hungary
[3] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, DK-2800, Lyngby, Denmark
[4] Chr. Hansen A/S, Human Health Innovation, Hoersholm, Denmark
[5] MiRanostics Consulting, 1846 E. Innovation Park Dr., Ste 100, Oro Valley, AZ, 85755, United States
[6] The State Key Laboratory of Pharmaceutical Biotechnology, the University of Hong Kong, Hong Kong SAR, China
[7] Department of Medicine, the University of Hong Kong, Hong Kong SAR, China

[†] These authors contributed equally
Correspondence: Gianni.Panagiotou@leibniz-hki.de (G.P.)

12. Abstract

The overgrowth of *Candida* species in the human gut is considered a prerequisite for invasive candidiasis [1]. However, the reason that many individuals with high levels of gastrointestinal *Candida* do not develop systemic candidiasis is unclear. Positive and negative interactions have been observed between individual *Candida* species and gut bacteria. These observations are from studies that were mainly conducted in mice or *in vitro.* Few large-scale human studies aimed to identify intestinal ecological signatures associated with *Candida* genus expansion in the gut. We integrated mycobiome and shotgun metagenomics data from 75 patients with lung cancer to determine the role of gut bacteria in shaping mycobiome composition. In addition, we developed machine learning models that used only bacterial taxa or functional relative abundances to predict the levels of *Candida* genus in an external validation cohort with an area under the curve of 78.6-81.1%. Last, we proposed an intriguing mechanism for *Candida* species overgrowth based on a decrease in short-chain fatty acid producing-bacteria resulting from increased oxygen levels. These conditions favour the growth of oxygen-tolerant lactic acid-producing bacteria, creating a metabolic niche for *Candida* species to use lactate as a carbon source and overtake their fungal competitors (especially *Saccharomyces*) in the human gut. We experimentally demonstrate that lactate supports the overgrowth of *Candida* species and show that lactate producing-bacteria also have a positive impact on gut barrier integrity. These observations emphasise the complex ecological interaction between multiple microbiome-gut epithelium factors that are involved in *Candida* species overgrowth and dissemination.

13. Introduction

Candida species, predominantly *C. albicans, C. glabrata, C. tropicalis,* and *C. parapsilosis,* are among the most common causes of bloodstream infections. These infections result in high rates of mortality for patients in intensive care units or who have a dysfunctional epithelial barrier or compromised immunity [2,3]. Although the pathogenicity of different *Candida* species has been extensively studied [4–9], only a handful of studies focused on understanding the commensal lifestyle of the fungus in the human gut. We recently performed a systematic evaluation of the interactions between human gut bacteria and *C. albicans* using genome-scale modelling and pairwise growth simulations [10]. We showed that 81% of *C. albicans* interactions with approximately 900 gut bacteria species were mutualistic (positive growth effects for both *C. albicans* and bacteria) or parasitic (negative growth effect on *C. albicans*, positive growth effect on bacteria), with only a few examples of parasitism, in which *C. albicans* exerted negative effects on gut bacteria. *C. albicans* is a common gut member in the majority of the human population. Therefore, our findings support the hypothesis that the colonisation success of *C. albicans* is the result of adapting to life in the intestine and avoiding competitive interactions with other gut microbes.

Most attempts to identify specific species of gut bacteria that inhibit or promote the growth of *Candida* species were conducted in murine models. However, in contrast to human gut communities, adult murine gut communities naturally prevent *Candida* species colonisation [11], showing substantial differences with humans in immune system regulation [12] and microbial composition [13–15], and challenging the translatability of the findings to humans. Gnotobiotic mouse models overcome some of these challenges [16] and were applied to study colonisation by *Candida* species [17]. Yet, limitations persist. A recent study demonstrated that colonisation is still incomplete, and some key bacteria genera (*Faecalibacterium*, *Bifidobacterium*) did not engraft at all [18].

Nevertheless, the realisation that the gastrointestinal tract is a major source of systemic candidiasis [1] has propelled efforts beyond animal models to identify predisposing factors that may lead to microbiome engineering strategies aimed at preventing candidiasis. This shift in focus has been further supported by evidence from human studies that gut bacterial dysbiosis triggered by broad-spectrum antibiotics is associated with increased colonisation of *Candida* species in the gut by [19]. However, antibiotics are not the only drugs associated with an elevated risk of *Candida* species overgrowth. Initial findings in animal models and humans suggest that chemotherapeutic agents lead to a reduced total number of gut bacteria and alterations in gut microbiota composition [20–22], which may contribute to the increased risk of systemic candidiasis in cancer patients. While most studies on systemic candidiasis and cancer have focused on haematological malignancies, recent epidemiological studies suggest that the risk for patients with solid tumours, such as head and neck and lung cancer, is equally high [23].

Recently, an analysis of a small cohort of allogeneic haematopoietic cell transplantation patients that included 11 candidiasis patients and 7 controls indicated that an expansion of *Candida* species in the gut occurs before bloodstream infection [24]. However, gut mycobiome analyses of both healthy individuals [25] and individuals with a variety of diseases [10,26] revealed that *Candida* species can also be the dominant fungi of the mycobiome without the host showing any signs of systemic infection. Therefore, overgrowth and systemic infection may be independent processes. Elucidating the role of *Candida* species as commensals and revealing the intestinal ecological context that leads to their expansion in the human gut is critical to designing prophylactic strategies for life-threatening systemic candidiasis. Therefore, we performed an integrative analysis of the mycobiome, microbiome, and phageome of 75 lung cancer patients to determine an intestinal ecological signature associated with *Candida* species overgrowth (Figure 1a), which we confirmed in an independent cohort of 11 individuals. We further provided experimental evidence for a competitive advantage of *Candida* over *Saccharomyces* species*,* the other main fungal residents in the human gut, while

exploring alternative carbon sources under dysbiotic conditions characterised by increased oxygen availability.

14. **Results**

High variability of *Candida* levels among infection-free lung cancer patients

We recruited 75 patients at the National Koranyi Institute of Pulmonology (Budapest, Hungary) and County Hospital of Pulmonology (Torokbalint, Hungary) with advanced-stage lung cancer (Table 1; adenocarcinoma n=40, squamous cell carcinoma n=28, others n=7). No patients had any signs of fungal infection during the recruitment period. Faecal samples were collected for analysis of fungal, bacterial, and viral biomes after the initiation of single-agent anti-PD-1 antibody immunotherapy (nivolumab, n=44; pembrolizumab, n=31). The majority of patients received chemotherapy prior to immunotherapy (n=59). We built ribosomal DNA internal transcribed spacer 2 (ITS2) libraries for estimating fungal genera and species relative abundance in all 75 patients. On average, we generated 78,332 (mean absolute deviation [MAD] 43,056) high-quality, non-chimeric reads per sample. Amplicon sequencing variants (ASVs) were estimated using DADA2 [27] resulting in 76 fungal genera (18 ± 10 per sample) and 113 fungal species (25 ± 15 per sample) (Figure 1a). Investigating genus-level fungal profiles showed that *Candida* and *Saccharomyces* were the highest contributing genera in 17 and 44 samples, respectively (Figure 1b). Fungal co-abundance networks revealed a strong, significant negative correlation between *Candida* and *Saccharomyces* at the genus (Spearman's coefficient; $P<0.05$; $|r|>0.25$; Figure S1) and species level (Spearman's; $P<0.05$; $|r|>0.25$; Figure S2).

We subsequently investigated the fungal genera that were the main drivers of variation in composition-aware mycobiome beta diversity (Aitchison distance; Figure 1c-e). Stepwise distance-based redundancy analysis (dbRDA) revealed that a large fraction (robust R^2=18.5%) of non-redundant fungal species diversity was explained by the two dominating fungal genera, *Saccharomyces* and *Candida* (Figure 1c), while other fungal genera explain an additional 20%. We subsequently examined anthropometric and lifestyle characteristics among the patients for significant correlation to ordination axes, including age, gender, body mass index (BMI,) diet, and antibiotic use ($P<0.05$; Figure 1e). Interestingly, 'antibiotic use' prior to anti-cancer treatment (3-6 months before stool sampling) correlated significantly with ordination results ($P<0.05$; Figure 1e), consistent with our previous findings in healthy individuals that antibiotic use can have a longer-lasting impact on the mycobiome compared to the microbiome [25]. Antibiotic use was also weakly correlated with higher levels of the *Candida* genus, similar to our previous observation [26].

We compared different mycobiome normalisation methods and observed high correlations between the normalised abundance estimates of the *Candida* genus (Pearson $r \geq 0.87$; $P<0.001$; Figure S3). To properly account for compositional data, all downstream analyses used fungal abundances normalised by the centred log-ratio (CLR) [28–30]. For *Candida* CLR abundances, the median separated *Candida* abundance symmetrically (Figure S3). Therefore, we grouped patients in two clusters: high-*Candida* (**HC**, n=38) and low-*Candida* (**LC**, n=37) for above or below the median *Candida* CLR normalised abundance (Figure S3). This grouping correlated significantly with high and low *Candida* ITS reads (Rfit $P<0.001$; Figure 1f), but not with the number of total ITS copies (Rfit $P>0.05$; Figure 1g) indicating that the sequencing depth did not affect the grouping. We further confirmed the grouping by testing for significant differences in fungal relative abundance between the two groups (Table S2). As expected, *Candida* genus abundance was increased in the **HC** group ($\log_2$ fold change [$\log_2$FC]=5.1; q=2E-12), whereas at the species level, *C. albicans* ($\log_2$FC=5.5; q=1E-7) and *C. tropicalis* ($\log_2$FC=2.1; q=0.04) drove the observed genus abundance differences. Although the beta diversity was significantly different between the **HC** and **LC** groups (PERMANOVA P=0.001; R^2=5.7%; Figure 1h), we did not find significant differences in the fungal

genera and species alpha diversities (by Shannon, Simpson, Pilou's Evenness indices; $P>0.05$; Table S1). We further examined if the classification of patients into **HC** and **LC** groups was explained by differences in basic patient characteristics such as gender, age, BMI, antibiotic use, alcohol consumption, tumour histology, chronic obstructive pulmonary disease, or anti-cancer treatment drug. None of these factors except for BMI was significantly different between the groups. Despite recent findings in mice and humans that *Candida* species may promote weight gain [31,32], we found BMI significantly decreased in our cohort in the **HC** group (Table 1; U-Test $P=0.006$). Therefore, in subsequent statistical comparisons between the **HC** and **LC** groups, we adjusted for differences in BMI.

Table 1 Anthropometric, clinical and lifestyle data for High (N=39) and Low (N=37) *Candida* groups.

Variable	N	High N = 38[1]	Low N = 37[1]	p-value[2]	Variable	N	High N = 38[1]	Low N = 37[1]	p-value[2]
Gender	75			0.7	**Alcohol**	75			0.5
Female		17 (45%)	14 (38%)		Never		19 (50%)	16 (43%)	
Male		21 (55%)	23 (62%)		Current		3 (7.9%)	6 (16%)	
Age	75	69 (64, 74)	65 (63, 69)	0.2	Former		2 (5.3%)	4 (11%)	
Body Mass Index	75	24.1 (21.4, 28.3)	28.7 (25.9, 32.8)	0.006	Occasionally		14 (37%)	11 (30%)	
Antibiotic Use Before therapy	74			0.3	**Histology**	75			0.7
No		34 (92%)	30 (81%)		Adenocarcinoma		22 (58%)	18 (49%)	
Yes		3 (8.1%)	7 (19%)		Squamous		13 (34%)	15 (41%)	
Unknown		1	0		Other		3 (7.9%)	4 (11%)	
Antibiotic Use After therapy	72			0.14	**Immunotherapy Drug**	75			0.7
No		28 (76%)	32 (91%)		nivolumab		21 (55%)	23 (62%)	
Yes		9 (24%)	3 (8.6%)		pembrolizumab		17 (45%)	14 (38%)	
Unknown		1	2		**Responder**	75			0.3
Line of Treatment	75			0.8	Yes		33 (87%)	28 (76%)	
1		9 (24%)	7 (19%)		No		5 (13%)	9 (24%)	
2		19 (50%)	23 (62%)		**COPD**	75			0.7
3		9 (24%)	6 (16%)		Without		16 (42%)	18 (49%)	
4		1 (2.6%)	1 (2.7%)		With		22 (58%)	19 (51%)	

[1]Statistics presented: n (%); Median (IQR)

[2]Statistical tests performed: chi-square test of independence; Wilcoxon rank-sum test; Fisher's exact test

Distinct microbiome signature associated with the high *Candida* group

We then shifted our focus on the gut bacterial community, specifically taxonomic and functional properties that might be pivotal in supporting the growth of *Candida* species in the human gut. We performed whole-metagenomic sequencing on the same stool samples used for the mycobiome analysis, generating an average of 26,106,952 (MAD 3,876,437) reads per sample. We used HUMAnN2 [33] to compute bacterial species and function abundance profiles. After applying a 10% prevalence filter, we estimated the relative abundance of 234 bacterial species (98 ± 18 per sample), 84 bacterial genera (46 ± 7), 394 MetaCyc pathways (339 ± 21), 1688 Enzyme Commission (EC) numbers (1361 ± 83), 5120 KEGG orthology (KO) terms (3943 ± 340), and 155 KEGG pathways (120 ± 3) (Figure 1a). To ensure that only bacterial information was used in further analyses, we removed features with unknown or non-bacterial origin from functional abundance profiles using the species-stratified output of HUMAnN2.

Procrustes analysis revealed a significant correlation between beta diversity for fungal and bacterial species ($P=0.046$, $r=0.53$; Table S1). Bacterial species explained around 13% of fungal species beta diversity and dbRDA using bacterial species suggested *Clostridium*, *Lactobacillus*, *Eubacterium*, and *Citrobacter* species had the highest explanatory power for mycobiome variation (Figure 1d).

Interestingly, a biplot of bacteria species abundances onto fungal species diversity indicated positive correlations between higher *Candida* genus abundance and several *Lactobacillus* species (Figure 1e), which was further confirmed in a crosskingdom network of fungal genera and bacterial species (Spearman's $P<0.05$; $r < -0.25$; Figure S2). In contrast, *Citrobacter* and *Eubacterium* species correlated negatively with *Candida* abundance.

We also observed significant separation between the **HC** and **LC** groups in bacterial species and functional beta diversity (Figure 2a-c; PERMANOVA; $P<0.05$; Table S1). For explaining variance, the functional properties of the bacteria community (MetaCyc, KOs, and ECs) showed the largest between-group differences (Figure S1, $R^2=2.5\%\pm0.5\%$). Bacterial species alpha diversity (by Shannon, Simpson, Pilou's Evenness, and Chao1 indices) was not significantly different between the **HC** and **LC** groups (Rfit $P>0.05$; Table S1). Surprisingly, functional alpha diversity (by MetaCyc pathways) showed higher diversity in the **HC** than the **LC** group (Figure 2e; $P=0.02$). In contrast, the contributional diversity of bacterial species to pathways (contributional alpha diversity) was significantly lower in many **HC** enriched pathways (Rfit; n=23; $P<0.05$; Figure 2f) and significantly lower overall (Rfit; $P<0.05$; Figure 2g). Together, these findings implied that the bacterial community in the **HC** group had greater metabolic potential, but the **LC** group had greater functional redundancy, a property exhibited by robust microbiota [34].

We then stratified bacteria based on their metabolic tolerance to oxygen. Species capable of growing under low oxygen level, including facultative anaerobes, were labelled 'aerobes'. We found significantly fewer obligate anaerobes in **HC** compared to **LC** (Figure 2h; $\Delta R^2=8\%$, $P=0.017$) and a trend for an increased aerobe/anaerobe ratio in the **HC** group (Figure 2i; $\Delta R^2=5\%$, $P=0.058$). This result was consistent with a comparison of 8 patients with candidemia and 7 controls where the expansion of *Candida* species was associated with a substantial loss of anaerobes diversity [24] as well as a previous study in mice in which antibiotic treatment with sufficient depletion of anaerobic bacteria was related to increased *Candida* species colonisation [11]. An increase in aerobes in the **HC** group, with their aerobic respiration, might explain the observed increase in metabolic diversity.

To complement our study on the ecological context associated with *Candida* species expansion in the human gut, we also quantified phage abundance using the recent release of the Metagenomic Gut Virus (MGV) catalogue. We used quasi-mapping for fast estimation of phage contig and viral operational taxonomic unit (vOTU) relative abundance using Salmon in metagenomic mode [35,36]). On average, 2.4% of metagenomic reads were assigned to prevalent viral contigs, with some samples reaching 4.7% (Figure S5). We did not observe a significant difference in the percentage of assigned phage reads between the **HC** and **LC** groups (Figure S5) or vOTU beta diversity (Figure 2e; $P>0.10$). However, a closer look into diversity-generating retroelements (DGRs) [37,38] revealed a substantial, significant reduction in DGRs phage genes in the **HC** group (two-sided Fisher test; $P=7e-5$; odds ratio=0.3; Figure 2j). DGR elements use error-prone reverse transcriptase to induce random mutations into the genomes of their host at specific target genes, creating population-wide hypervariability [39,40]. Since DGRs have beneficial effects (e.g., adaption advantage) on their targeted host [39,40], the enrichment in these phages may imply a more robust microbiome in **LC**.

Abundance of lactic acid bacteria and SCFA producers accurately predicted *Candida* levels

We performed supervised machine learning (ML) to investigate if members of our cohort could be classified as **HC** or **LC** solely based on bacterial taxonomic or functional relative abundances. We applied SIAMCAT for model training and evaluation [41] but adopted data augmentation training to the default algorithm [42]. We tested our models in an additional validation cohort of 11 immunotherapy-treated lung cancer patients. In the validation cohort, **HC** and **LC** were defined using the same abundance thresholds used for the main cohort. Bacterial species abundance classified patients as **HC** or **LC** with high accuracy in both our main cohort (Figure 3a;

crossvalidation area under the receiver operating characteristic [CV auROC]=77.9%) and validation cohort (Figure 3a; auROC=78.6%). With bacterial functional abundances (by EC), we achieve slightly higher accuracy (Figure 3a; CV auROC=80.4%; additional auROC=82.1%). We further investigated if we could predict high vs. low abundance levels for the species *C. albicans*, *C. sake*, and *C. glabrata,* which were the most prevalent and abundant species in our cohort. High and low abundance groups for each of the species were formed based on mean species CLR abundances and ML models were built analogous to the **HC** vs. **LC** genus models. The identified microbiome signatures classified patients as having high or low abundance well for *C. albicans*, *C. glabrata*, and *C. sake* and in both, our main cohort (Figure 3a; CV auROC: *C. albicans*=77.5%, *C. sake*: 77.0%, *C. glabrata*: 73.9%) and test cohorts (Figure 3a; Test auROC: *C. albicans*=86.7%, *C. sake:* 86.7%, *C. glabrata*: 79.2%). Interestingly, phage vOTU abundance showed a potential to predict *Candida* genus and species levels in the main cohort (Figure 3b; CV auROC: 73%±1%) but had less potential for the independent cohort (Figure 3b; Test auROC: 53%-75%). *C. albicans* levels were predicted robustly in both the main and validation cohorts (auROC=73%) using solely phage composition. This result was interesting considering recent evidence of inhibition of *C. albicans* by bacteriophages [43].

We then inspected bacterial species predictive of *Candida* genus levels and crosschecked those species with results from differential abundance analysis (by Maaslin2). We found that many of the bacterial species predictive of **LC** with high robustness (at least 80%; $P<0.05$; false discovery rate [FDR]<0.2; Figure 3c) were short-chain fatty acid (SCFA) producers [44–49], including *Actinomyces odontolyticus*, *Bifidobacterium adolescentis, Eubacterium rectale, Anaerotruncus colihominis*, *Alistipes ihumii AP11*, several *Lachnospiraceae* species, *Pseudoflavonifractor capillosus*, and *Odoribacter splanchnicus*. We retrieved genome-scale metabolic models of the bacteria species enriched in the **LC** group from the AGORA repository [50] and simulated growth on different diets using flux balance analysis (FBA; Table S3). We monitored the potential to produce SCFAs and confirmed that many of the bacterial species enriched in the **LC** group can secrete at least one of acetate, propionate, or butyrate at varying levels (see Methods and Table S3). The importance of SCFAs in suppressing *C. albicans* colonisation has been reported by us [25,26] and others [11]. However, the suppressive function of SCFAs appears to be towards all *Candida* species in general. Several mechanisms by which propionate and butyrate suppress *Candida* species colonisation have been suggested, including regulation of the immune system [51] and direct inhibition [52,53]. However, SCFAs also have a major impact on oxygen availability [54,55]. Therefore, a decrease in SCFA producers in the **HC** group should be accompanied by an expansion of facultative aerobes. Incidentally, we noticed an increase in oxygen-tolerant bacteria in the **HC** group (Figure 3c). Interestingly, several lactic acid bacteria were significantly higher in the **HC** group and consistently selected as top features in the ML models (for example, *Lactobacillus gasseri* and *Lactococcus lactis*) (Figure 3c). We also observed an increased abundance of *Enterobacteriaceae* species (*Escherichia* species*, Klebsiella pneumoniae*). A crossdomain correlation analysis between fungal genera and bacterial species abundance confirmed positive correlations between *Candida* genus, *Lactobacillus*, *Lactococcus*, *Klebsiella* and *Escherichia* species in our study cohort (Figure S2). Similar to the **LC**-enriched species, FBA analysis suggested that in addition to the lactic acid bacteria enriched in the **HC** group, *K. pneumoniae* and *E. coli* also secrete lactate (Table S3).

Microbial set enrichment analysis (MSEA) [56] revealed that bacterial genera that were significantly increased in the **HC** group ($P<0.05$; *Lactobacillus*, *Lactococcus*, *Streptococcus*, *Bacteroides* and *Odoribacter*; Table S4) were associated with multiple human disease genes (n=280) suggesting a dysbiotic microbiome. In contrast, bacteria highly abundant in the **LC** patients showed no enrichment in disease genes despite covering more genera (n=12). These results were also qualitatively the same at $P<0.10$. Further functional enrichment analysis of the 280 human disease genes based on the KEGG pathway database, indirectly linked **HC**-associated bacterial species to

cytokine and chemokine responses (Figure S6; Table S4). IL-17 signalling was of special interest because it is associated with gut inflammation and *Candida* species colonisation [57]. IL-17 also has a role in immune cell recruitment after bacterial invasion [58] and is a key role in stimulating host immunity upon *Candida* infection [58].

We then examined bacterial metabolic functions by analysing MetaCyc pathway abundance (Figure 3d; Table S2). In total, 78 pathways were more abundant in **HC** compared to only 11 in **LC** ($P<0.05$; $q<0.15$), matching the observation of increased functional alpha diversity in **HC**. One of the abundant pathways in the **HC** group produces lactate from hexitols (P461-PWY). In agreement with the observation of decreased anaerobes, aerobic respiration (fatty acid and beta oxidation pathways, TCA-bypass, TCA cycle II) and synthesis pathways for compounds in cell membranes of aerobic bacteria (menaquinol and ubiquinone synthesis) increased in the **HC** group. Notably, bby slightly relaxing the *P*-value to 0.07, we obtained 5 additional TCA cycle pathways with higher abundance in the **HC** group (TCA cycle I, IV, V, VII and partial TCA cycle in obligate autotrophs), effectively covering 6 of 9 TCA pathways in the MetaCyc database (Figure 3d; Table S2). Gene-set enrichment analysis confirmed the elevated levels of aerobic respiration in the **HC** group ($q<0.05$; Table S2). Based on the increased abundance of lactate producers, we examined functions related to lactate utilisation. We found D-lactate dehydrogenase (D-LDH) was significantly increased in the **HC** group by gene family (UniRef90; $P=0.029$; Figure 3e) and EC levels (EC:1.1.1.28; $\log_2FC=1.09$; $P=0.053$; $q=0.23$; Table S2). We also found a significant increase in (S)-2-hydroxy-acid oxidase (EC:1.1.3.15; $\log_2FC=0.99$; $P=0.048$; $q=0.23$) which reduces aliphatic hydroxy acids, including lactate, using flavin mononucleotide and oxygen. In contrast, L-LDH genes and enzymes did not show significant changes ($P>0.10$). We used the Metabolic Analysis of Metagenomes using fBA and Optimization (MAMBO) algorithm [59] to predict the metabolic flux of the complete bacterial community and found a significant increase in D-lactate secretion in the **HC** compared to the **LC** group ($P=0.051$; Figure 3f).

In summary, we identified a distinct gut microbial signature predictive of high *Candida* genus abundance in infection-free lung cancer patients. This signature describes a dysbiotic gut microbiome state characterized by a systematic decrease in SCFA producers, which results in increased oxygen-tolerant microbes, including certain lactic acid-producing bacteria.

C. albicans has a competitive advantage using lactate under microaerobic conditions

Our data suggested a possible link between microbial lactate production, higher availability of molecular oxygen and increased *Candida* species abundance in the gut. To test the hypothesis that *Candida* species may outgrow other fungi, specifically the *Saccharomyces* species that are their main competitors in our study cohort, we selected *C. albicans* and *S. cerevisiae* var. *boulardii* for further experiments. Since our ITS2 data could not reliably separate species from the *Saccharomyces in situ stricto* group [60], we selected *S. boulardii* as the best proxy. A fluorescently labelled strain available in our laboratory facilitated competition experiments with *C. albicans.* We used two strains of *C. albicans* (BWP17, RM1000) that showed significant alterations in morphology, drug resistance, and host-cell stimulation depending on the main carbon source [58]. In individual *in vitro* growth assays with *C. albicans* BWP17 and RM1000 and *S. boulardii-GFP* on different carbon sources (lactate, glucose, or lactate + glucose) with varying oxygen levels (anaerobic, microaerobic, or aerobic) (Figure 4a-b; Table S5) only *C. albicans* was able to use lactate as a sole carbon source in aerobic and microaerobic conditions. All three yeast were unable to grow on lactate under fully anaerobic conditions (Figure 4b). The most significant growth of *C. albicans* compared to other oxygen levels and carbon sources was on lactate as the sole carbon source in microaerobic conditions (t-test; $P<0.05$; Figure 4a-b). In fully aerobic conditions, growth on solely lactate was strain dependent but generally lower than in combined media (Figure 4a-b). These observed differences in growth on lactate were more pronounced for BWP17 than RM1000. In microaerobic conditions, all species

showed significant reductions in growth rate in glucose+lactate compared to growth on at least one sole carbon source (t-test; $P<0.05$). When glucose was present, *S. boulardii* growth was unaffected by additional lactate in microaerobic and aerobic conditions (t-test; $P<0.05$; Figure 4a-b). However, both *C. albicans* strains showed a slight growth reduction in glucose+lactate (t-test; $P<0.05$; Figure 4a).

Together, these findings demonstrate the importance of oxygen for *C. albicans* growth in lactate as a carbon source. To evaluate a direct competition between *Candida* and *Saccharomyces* species, we co-cultured them under the same conditions as individual growth assays (Figure 4c). The growth of both *C. albicans* strains was consistently higher than *S. boulardii-GFP* on lactate as the sole carbon source in microaerobic or aerobic conditions. Since oxygen levels in the lower gastrointestinal tract are unlikely to be normoxic (aerobic), the results in microaerobic conditions are more representative of the true competitive landscape in the human gut for the two yeasts, at least during gut inflammation.

Last, we evaluated if the main lactic acid producing-bacteria identified in our study could be involved in the translocation of *Candida* species through intestinal walls. We focused on the impact of *Lactobacillus* and *Lactococcus* species on gut integrity, which is a critical factor in systemic candidiasis [61]. We evaluated the impact of *L. gasseri* and *L. lactis* by measuring transepithelial electrical resistance (TEER; Figure 4d). As positive controls, we used *L. rhamnosus* (LGG®) a well-known probiotic, and *B. adolescentis* and *O. splanchnicus*, which were enriched in the **LC** group. *B. adolescentis* had the largest negative fold-change when comparing **LC** vs. **HC** groups (Figure 3c). It is an important microbe associated with health [62], and enhances gut barrier function [63]. *O. splanchnicus* produces an array of diverse SCFAs [64] and its absence is associated with various inflammatory diseases [46,65]. Both *L. gasseri* and *L. lactis* showed protective effects (as increased area under the curve in TEER), comparable to LGG®, higher than *O. splanchnicus* but lower than *B. adolescentis* (Figure 4d).

We found strong evidence that *C. albicans* strains have growth advantages compared to *Saccharomyces* species in utilizing lactate, and this effect is dependent on the levels of molecular oxygen available in the environment. However, although the lactate producers supported *Candida* species growth by creating a metabolic niche, they might also contribute positively to the integrity of the gut barrier. These findings align with our initial hypothesis that microbiome factors that support overgrowth are not necessarily involved in the dissemination of *Candida*.

15. Discussion

Colonisation resistance is a crucial function of the gut microbiota in infectious disease. In addition to protecting the host from external pathogens, local microbiota also prevent expansion and invasion of intestinal pathobionts [66]. Microbial resistance in gastrointestinal infections have both direct and indirect mechanisms. Infections can be limited directly via metabolic by-products (bacteriocins, acids, peptides) of the gut microbiota [67], or by outcompeting pathogens for space, metabolites, and nutrients [68]. Intestinal pathogens can also be inhibited indirectly when the local microbiota calibrate host immune responses to them [69] or induce the formation of a protective mucin layer that covers the gut epithelium. Perturbation of the resident microbiota is thus a risk for infection by a pathobiont infection that is ordinarily held at bay by these mechanisms. *Candida* species are gut symbionts that can become an aggressive pathogen under specific circumstances. Despite *Candida* species being the fourth most common cause of nosocomial bloodstream infections [70] the number of human studies investigating their interplay with the gut microbiota is surprisingly low compared to studies on bacterial pathogens such as *Clostridium, Enterococcus, Salmonella* and *Enterobacteriaceae* [71].

Most of the work to identify specific bacterial promoters or inhibitors of *Candida* species colonisation has been performed in mouse models. Fan *et al.* [11] demonstrated how *Bacteroides*

thetaiotamicron can protect mice from *C. albicans* colonisation by activating innate immune effectors and the antimicrobial peptide LL-37. The bacteria taxonomic annotation was based on 16S rRNA, therefore the changes in the metabolic capacity of the gut microbiome associated with the *C. albicans* colonisation remained unclear. Tan *et al.* [7] used mice to show that several gram-positive bacteria, including *Staphylococcus aureus,* shed peptidoglycan units that trigger hyphae formation of *C. albicans.* The only human study to concomitantly examine the mycobiome and microbiome to find common colonisation patterns between *Candida* species and the gut microbiome was presented recently by Zhai *et al.* [24]. The authors concluded that systemic candidiasis begins with expansion of *Candida* species in the human gut. Using 16S rRNA, they observed a reduction in the levels of anaerobes in patients with systemic candidiasis compared to non-infected cohorts. However, the small number of patients (8) in this study and the lack of functional characterisation of the microbiome left many questions unanswered.

Using shotgun metagenomics of stool samples from 75 lung cancer patients combined with ITS sequencing, we substantially expanded our knowledge on the ecology of *Candida* species inhibition and overgrowth. Some studies suggest that immunotherapy limits *Candida* overgrowth. Although all lung cancer patients in our study received immunotherapy, we observed high variation in *Candida* genus levels. The mycobiome of several patients was completely dominated by *Candida* species. In some other samples, these fungi were virtually absent. No patients, including those with extremely high levels of the *Candida* in the gut, showed any sign of infection. However, we do not know if any of those patients were diagnosed with systemic candidiasis after the completion of our study.

Our relatively large study cohort allowed us to develop a machine learning model based solely on bacterial taxa or functional abundance. The model had high accuracy in classifying patients from an external cohort into groups with high abundance or low gut abundance of the *Candida* genus. Some bacterial species or functions are suggested to affect the growth of individual *Candida* species [11,72]. We mostly focused on *C. albicans* and demonstrated general properties of the gut microbiome that are associated with the successful colonisation of *Candida* species. We also developed ML models to predict the levels of individual *Candida* species. These models showed high accuracy but did not significantly improve classification over the genus-based model. A phage-based machine learning model showed a high predictive power for *C. albicans* and requires further attention in the future in light of recent evidence [43].

What we found particularly intriguing in our human study was the enrichment of several potential lactic-acid producing bacteria in the **HC** group. *Lactobacillus* species such as *L. gasseri* are particularly interesting as recent *in vitro* experiments show they prevent hyphae formation without reducing the growth of *C. albicans* [73]. The exact role of lactate on *Candida* species growth is unclear. For *C. albicans*, the change from glucose to lactate as the main carbon source is tightly linked to changes in cell wall composition. Ene *et al.*, and Ballou *et al*., independently demonstrated *in vitro* that a lactate-rich environment assists *C. albicans* in hiding from the innate immune system by stimulating interleukins [58] or by beta-glucan masking [74]. In contrast, Gutierrez *et al.* found lactate inhibited the growth of *C. albicans* at higher concentrations [75], while MacAlpine *et al.* reported no impact on growth at physiological levels of lactate [73]. Notably, the oxygen status of the experiments is unclear and *C. albicans* was grown at 30°C or 42°C, which are not physiologically relevant in the gut. However, the ecology of the human gut is more complex. The growth response of *Candida* species to increased levels of lactate depends on both the growth of other fungi and on environmental conditions. We used human metagenomics data and *in vitro* competition experiments to demonstrate that under low oxygen conditions, induced by a reduction of SCFA producers [55], *Candida* species may gain a competitive advantage due to growth on lactate as the primary carbon source in the gut. Our observations bring to light another possible mechanism by which SCFA producers inhibit *Candida* overgrowth in addition to mechanisms in the literature such as direct inhibition [52] and immune regulation [51,66,76]. However, although lactate producers may promote

Candida species growth in the human gut under microaerobic conditions, they also increase protection of the human host from systemic candidiasis by increasing gut barrier integrity. Limitations of our study are that we did not monitor oxygen levels in the gut of patients and provide only indirect evidence in the form of signs of dysbiosis, increased ratios of aerobes to anaerobes and microbial functional pathways related to aerobic respiration in the **HC** group. No patient in this group developed systemic candidiasis during our study. However, a longer follow up would be necessary to delineate the ecological context associated with overgrowth and dissemination of *Candida* species. Nevertheless, human studies with many participants like ours are needed to evaluate which findings from *in vitro* experimentation are relevant to the human gut and to design prophylactic, microbiome-driven strategies for patients at high risk of candidiasis.

Acknowledgements

We would like to thank the patients and their families for participating in the stool collection studies. We would also like to thank Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) CRC/Transregio 124 'Pathogenic fungi and their human host: Networks of interaction', subproject INF and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Germany's Excellence Strategy - EXC 2051 – Project-ID 390713860 for financial support. R.V.U and M.O.A.S acknowledge support from The Novo Nordisk Foundation, Challenge programme, CAMiT under grant agreement: NNF17CO0028232 and grant number: NNF20CC0035580. We also thank Dr S. Brunke for sharing the *C. albicans* strains, Dr T. Zheng for her initial work on phage annotation, and Dr R. Santhanam for pilot work on the mycobiome of cancer samples.

Author contributions

Conceptualisations: B.S., Z.L., R.V.U., M.O.A.S., and G.P. **Funding acquisitions**: G.P., M.O.A.S. **Investigations**: R.V.U., Z.L., G.J.W., X.C., M.M, A.W., S.L., and B.S. **Methodology**: B.S., Z.L., G.J.W., M.M, X.C., and R.V.U. **Project administration**: G.P. **Resources**: G.P., M.O.A.S. **Supervision**: G.P., M.O.A.S. **Validation**: all authors. **Visualisation**: B.S., R.V.U. **Writing**—original draft: B.S. and G.P. **Writing**—review and editing: all authors.

Competing interests

GJW is an employee of SOTIO Biotech Inc., a former employee of Unum Therapeutics; reports personal fees from Spring Bank Pharmaceuticals, Imaging Endpoints II, MiRanostics Consulting, Gossamer Bio, Paradigm, International Genomics Consortium, Angiex, IBEX Medical Analytics, GLG Council, Guidepoint Global, Genomic Health, Oncacare, Rafael Pharmaceuticals, and SPARC-all outside this submitted work; has ownership interest in Unum Therapeutics (now Cogent Biosciences), MiRanostics Consulting, Exact Sciences, Moderna, Agenus, Aurinia Pharmaceuticals, and Circulogene-outside the submitted work; and has issued patents: PCT/US2008/072787, PCT/US2010/043777, PCT/US2011/020612, and PCT/US2011/037616-all outside the submitted work.

The remaining authors declare they have no competing interests.

Data availability statement

Fungal internal transcript spacer (ITS) and whole metagenomic sequencing (WMS) data generated in this study were uploaded to the NCBI sequence read archive (SRA) under Bio Project **PRJNA811494**.

16. Tables

Table S1. Test results for significant differences in alpha and beta diversity between groups.

Table S2. Differential abundance test results of bacteria, fungal, MetaCyc pathways, Enzyme commission (EC) and viral operational taxonomic unit (vOTU) abundances.
Table S3. Genome-scale metabolic modelling results for short-chain fatty acids from different diets and oxygen levels.
Table S4. Microbial set enrichment results (MSEA).
Table S5. Growth rates, colony-forming units (CFU) and competition experiments.

17. Figures

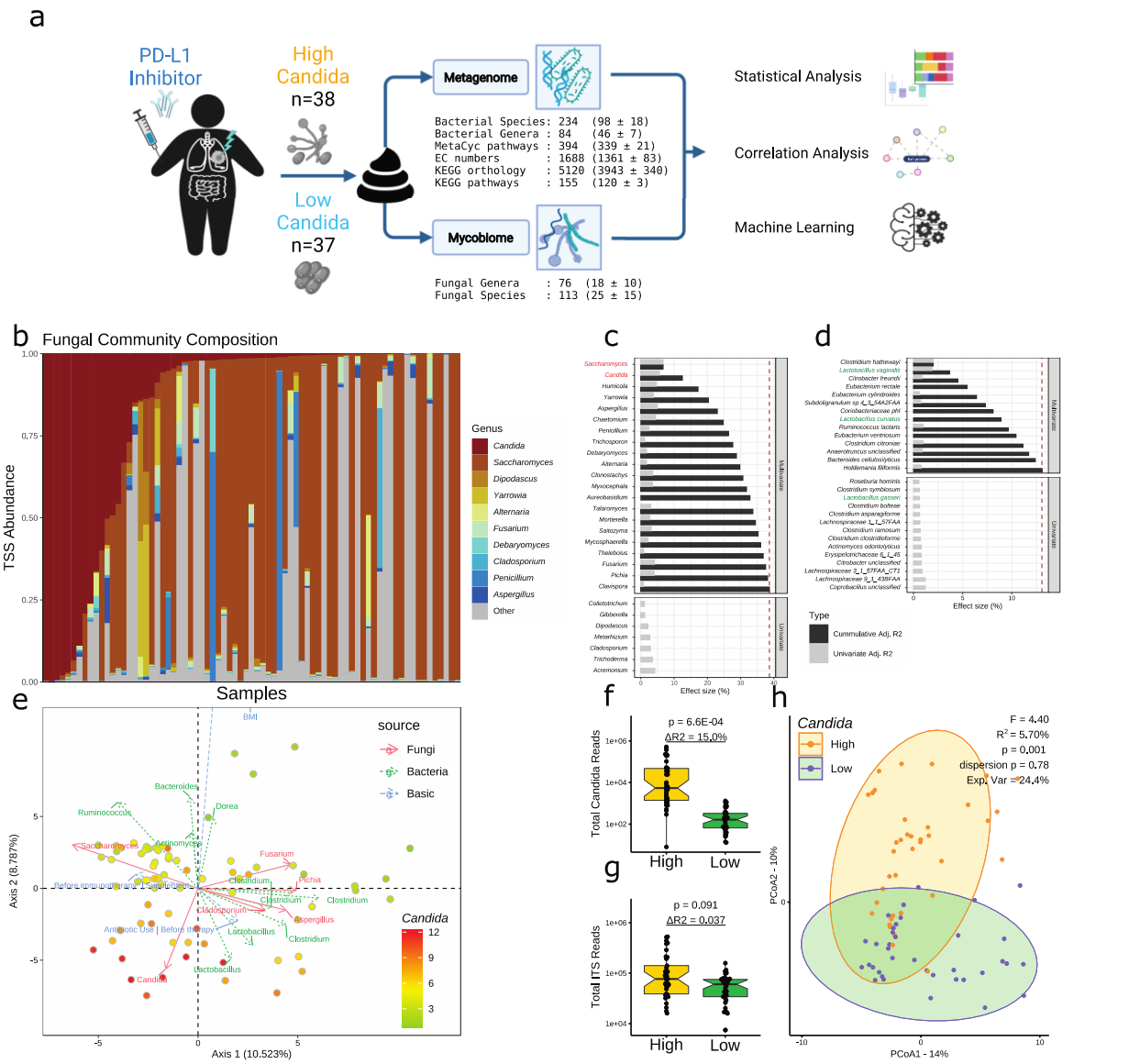


Figure 1. Explanatory factors of mycobiome diversity. (a) Study design. Indicated are the total number of taxonomic and functional features annotated in our study and in parentheses, their average per sample. **(b)** Fungal genus relative abundance profile (by total-sum scaling, TSS). The top 10 genera by highest median abundance are indicated by colour, with grey indicating abundance of the remaining fungi. **(c-d)** Distance-based redundancy analysis (dbRDA) of fungal species beta diversity (Aitchison distance). Explanatory factors are **(c)** fungal genera and **(d)** bacterial species relative abundance. Only significant terms are shown (P<0.05). Displayed are cumulative explained variances of full, non-redundant models (black) and single-term statistics (grey). Important fungal and bacterial taxa are coloured red or green, respectively. **(e)** Principal coordinate analysis (PCoA) biplot of fungal species beta diversity (Aitchison). Samples are coloured by *Candida* centre log ratio-normalised abundance. Top features by ordination correlation (t-test; P<0.05) are bacterial (green), fungal (red) and patient characteristic (blue) arrows indicating the direction of covariance between feature abundances and the first two ordination axes. Samples are in a gradient from high (red) to low (green) *Candida* abundance. Axes show explained variance. **(f-g)** Notched boxplots showing **(f)** total number of *Candida* internal transcribed spacer (ITS) reads and **(g)** the total number of ITS reads. Significance was assessed using non-parametric generalized linear models (Rfit) controlled for body mass index and gender. **(h)** PCoA of fungal genera beta diversity. *Candida* grouping (High vs. Low) shows significant separation (P=0.001; PERMANOVA). Within-group diversity differences were insignificant (dispersion P>0.05). Circles indicate 95% confidence interval of within-group diversity.

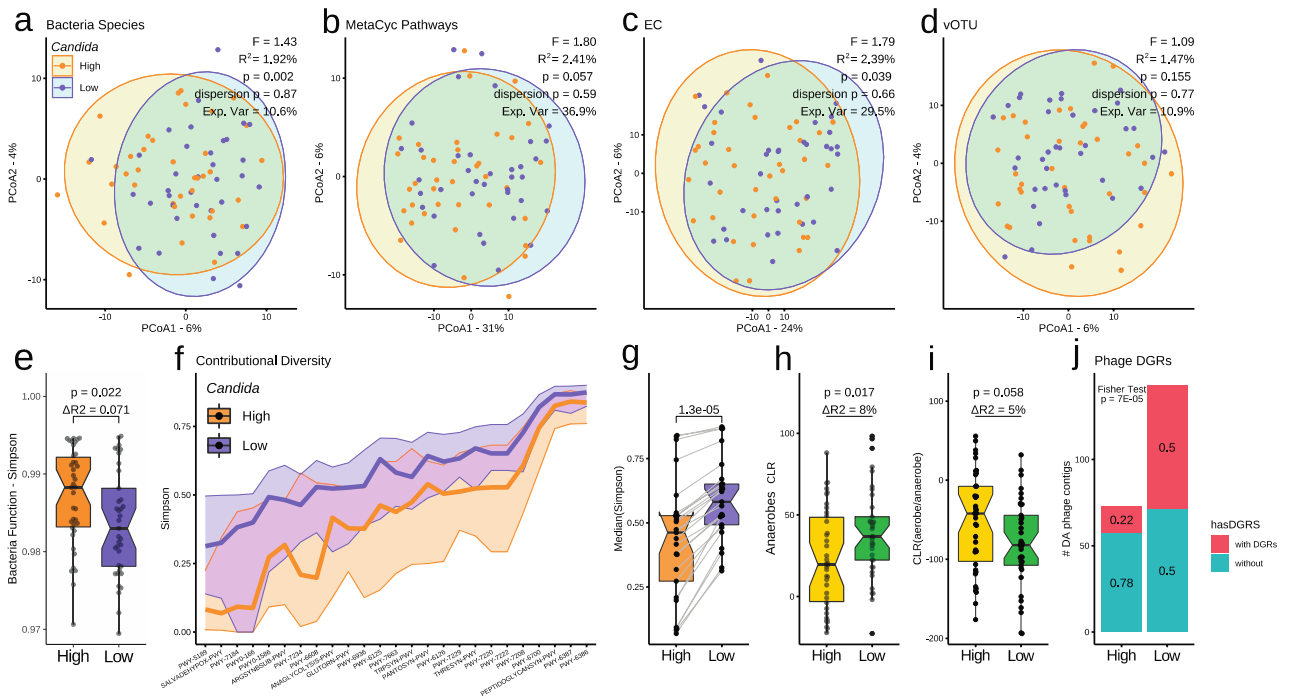


Figure 2. Microbiome contribution to high and low *Candida* groups. (a-d) Principal coordinate analysis (PCoA) of beta diversity (Aitchison distance) using **(a)** bacterial species, **(b)** bacterial metabolic pathways, **(c)** bacterial enzyme functions (EC, Enzyme Commission) or **(d)** viral orthologous taxonomic unit (vOTU) abundance profiles. Bacterial function profiles were stratified for bacteria-only abundances. Diversity was significantly different between *Candida* groups (High vs. Low; $p<0.05$; PERMANOVA; betadisper $P>0.05$). **(e)** Notched boxplots of bacterial function (MetaCyc pathway) alpha diversity (Simpson). **(f)** MetaCyc pathways with significant contributional Simpson diversity (Rfit drop-test; $P<0.05$). Centre lines indicate the median Simpson diversity (y-axis) of a pathway (x-axis). Ribbons indicate the 25% and 75% quantiles. Colours indicate High or Low *Candida* group. **(g)** Notched boxplots show the median contributional diversity per pathway (P=1e-5; paired Wilcoxon test). For each pathway (point), grey lines indicate change in diversity from High to Low group. **(h-i)** Notched boxplots summarizing the abundance of bacterial taxa stratified by tolerance to molecular oxygen. Only strict and obligate anaerobes were considered anaerobes. Significance was assessed using non-parametric generalised linear models (Rfit) controlled for body mass index and gender. Effect sizes are indicated as ΔR^2. **(j)** Number of phages contigs with (blue) or without (red) diversity-generating retro-elements (DGRs). DGRs were significantly enriched in samples of the low *Candida* group ($P<0.05$; two-sided Fisher test).

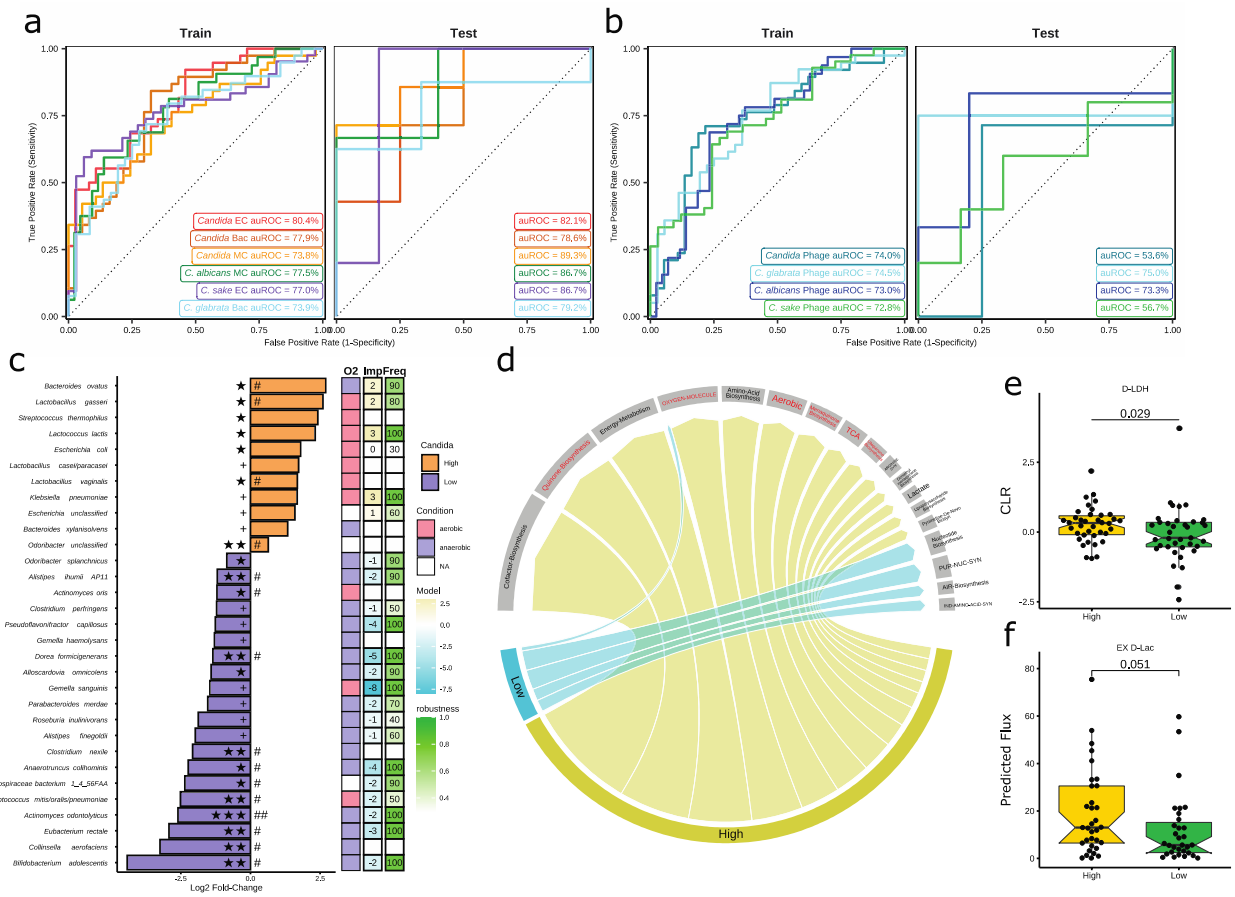


Figure 3. Gut bacterial signatures predictive of *Candida* levels. (a-b) Machine learning performance for predicting High and Low levels of all *Candida* genus and species (*C. albicans*, *C. sake*, *C. glabrata*). Performance is shown as area under the receiver operator characteristic (auROC) using **(a)** bacterial feature abundance or **(b)** viral operational taxonomic unit abundance (vOTU). auROC in the main cohort (**train**) was assessed using prediction on hold-out samples during 10-fold cross validations. The resulting model was used to predict *Candida* High vs. Low in an independent set of patients (**test**; n=11). For *Candida* genus, three models are shown (red, brown, orange). **(c)** Differential abundance of bacterial species. Significance was assessed using MaasLin2 (default) controlled for body mass index and gender. Significance is indicated as ^{+}P<0.10, $^{\star}P$<0.05, $^{\star\star}P$<0.01, $^{\star\star\star}P$<0.001 and $^{\#}FDR$(p)<0.2, $^{\#\#}FDR$(p)<0.1 (FDR, false-discovery rate). Right, species tolerance to oxygen (O_2), machine learning importance (Imp) and robustness (Freq). Importance indicates the contribution of a species towards predicting High (positive; yellow) or Low (negative; blue) *Candida*. Robustness is the number of times a feature was included in a model during cross validation. **(d)** Significantly enriched functional classes of significant MetaCyc pathways (Gene Set Enrichment test). Metabolic functions related to aerobe metabolism are red. **(e-f)** Boxplots summarising **(e)** centre log ratio-normalized gene family abundance (UniRef90) of D-lactate dehydrogenase (D-LDH) and **(f)** metabolite (D-lactate) flux predicted by MAMBO from bacterial profiles. Significance was assessed using non-parametric generalised linear models (Rfit) controlled for body mass index and gender.

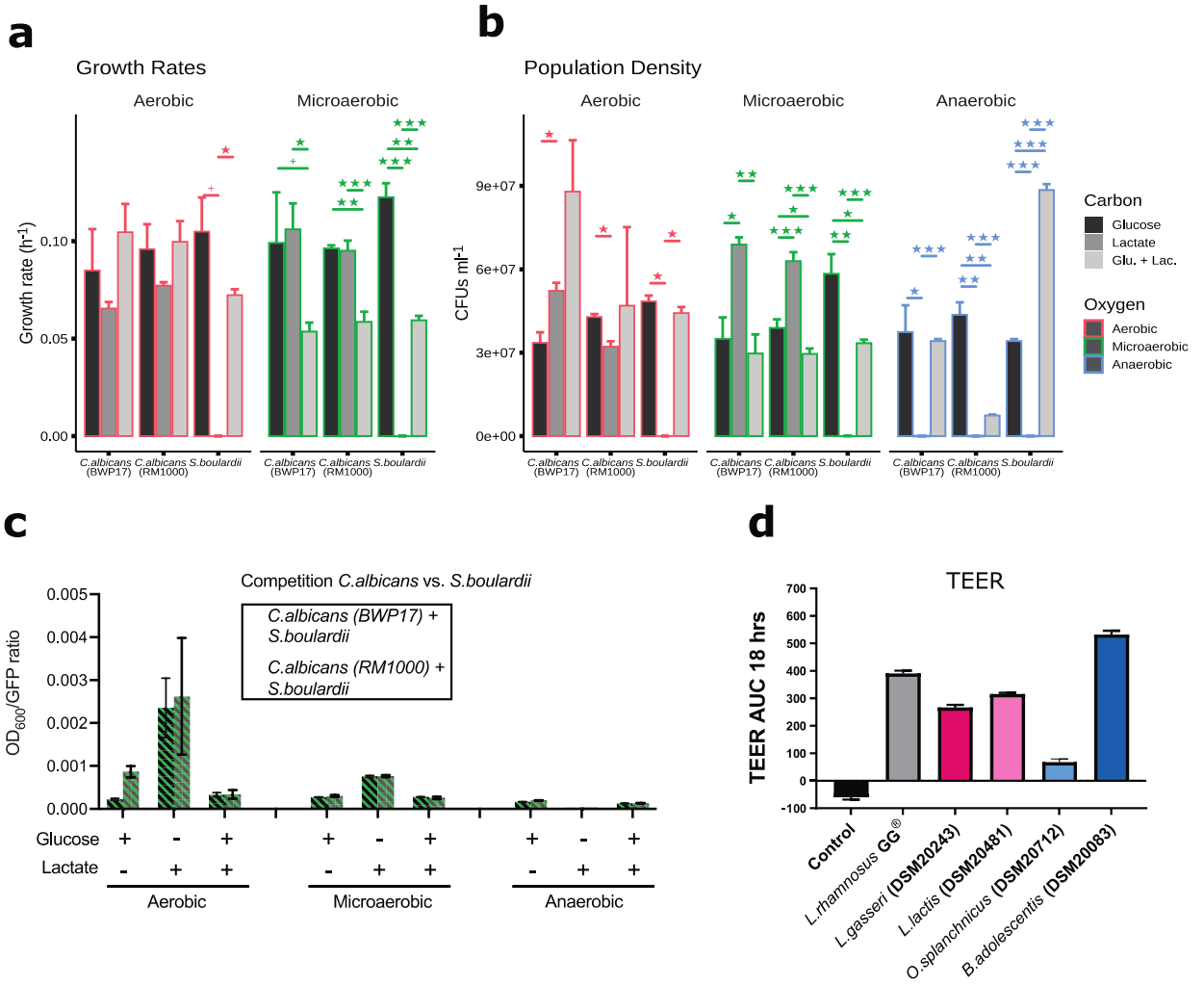


Figure 4. Bacterial transepithelial electrical resistance and yeast growth under varying levels of oxygen and carbon sources. (a-d) Histograms showing means and standard deviations. Significance was assessed using two-sided t-tests and indicated as $^{+}P<0.10$, $^{\star}P<0.05$, $^{\star\star}P<0.01$, $^{\star\star\star}P<0.001$. **(a)** Growth rates of yeast (*Candida albicans* BWP17, *C. albicans* RM1000, *Saccharomyces boulardii*) under carbon sources (glucose, lactate, or both) and oxygen levels (0%, 1%, and 21%). **(b)** Colony-forming units (CFU)·ml^{-1} were determined at the end of growth experiments. **(c)** Co-culture competition of *C. albicans* against *S. boulardii*. Species were added 1:1 and growth tracked for 24 hours. Green fluorescent protein (GFP)-labelled *S. boulardii* were used to differentiate *S. boulardii* from *C. albicans* cells. The ratio of final optical density (OD) to GFP (*S. boulardii*) indicates *C. albicans* abundance. Higher values indicate more *C. albicans*. **(d)** Transepithelial electrical resistance (TEER) to assess protective effects of bacteria. Results are area under the curve (AUC) of TEER curves over time. Values greater than 0 imply protection compared to control (no bacteria).

Figure S1. Crosskingdom co-abundance networks of bacterial species and fungal genera abundances. Correlations were assessed using Spearman's coefficients on centre log ratio (CLR)-normalized (left) or total-sum scaled (TSS)-normalized abundance data (right). Only significant correlations with $P<0.05$ and $|r|>10\%$ are shown. Red, solid edges show positive correlations; blue-dashed edges show negative correlations. Colour strength indicates estimated correlation strength. **(a)** CLR-based networks capture microbial abundance correlations better, especially when comparing features between distinct compositions. However, strong changes in observed relative abundance (TSS) were lost. **(b)** TSS networks revealed correlations between dominant species but were less reliable for low-abundance species. **(a-b)** *Lactobacillus* species such as *L. gasseri* and *L. lactis* and *Escherichia* species correlated positively with *Candida* regardless of normalization, and short-chain fatty acid producers like *Lachnospiraceae* and *Actinomyces odontolyticus* were found to correlate negatively.

Figure S2. *Candida albicans* correlated negatively with *Yarrowia lipolytica* and *Saccharomyces paradoxus* abundance. **(a)** Fungal co-abundance networks of significant correlations ($P<0.05$; $|r|>0.10$; Spearman's coefficient) with *C. albicans*. Solid, red lines indicate positive correlations. Dashed, blue lines indicate negative correlations. **(b)** Fungal species abundance profile indicating the top 10 most abundant fungal species based on total-sum scaling. Species are shown in colours. Other fungi are grey.

Figure S3. Normalised *Candida* abundance correlates across methods. In normalised space **(a)**, all methods except rarefaction showed acceptable linear correlation (cumulative sum scaling [CSS]; centralized log-ratio [CLR]; balance-based CLR [bal-CLR]). Rarefaction differs were resolved using a simple rank transformation **(b)**, implying at least a non-linear relationship among all methods. The strongest agreement was seen for CSS and CLR. Rarefaction was a poor normalisation choice, as it accounted only for differences in sequencing depth and not for compositionality effects in relative abundance data. **(c)** *Candida* CLR-normalised abundance (y-axis) per sample (x-axis). Samples are ordered from low to high CLR abundance. Dashed lines indicate median sample and median abundance thresholds.

Figure S4. **Principal coordinate analysis (PCoA) of beta diversity for bacterial species and functional, fungal genus and species, and viral operational taxonomic unit abundances (vOTU)**. R^2 describes estimated explained variance by *Candida* High vs. Low grouping (Adonis2; PERMANOVA). MetaCyc and KEGG Orthology had the largest effect sizes.

Figure S5. Metagenomic gut virus (MGV) Pphage mapping rate. (a) Percentage of reads per sample that passed (blue) or did not pass (red) a 10% prevalence filter. More than 30% of reads were assigned to low prevalence contigs in a few samples. **(b)** Percentage of reads assigned to phage contigs by *Candida* High or Low abundance. Statistical significance was assessed by unpaired Wilcoxon rank-sum tests. The median 2.5% of reads were assigned to prevalent viral contigs.

Figure S6. Species enriched in the High *Candida* (HC) group were related to human-disease genes. Microbial Set Enrichment Analysis revealed that bacterial genera with significant differential abundance in the HC group (bottom) were frequently associated with human disease genes (n=280) in contrast to the low *Candida* (LC) group (n=0). Green nodes indicate human disease-associated genes that were tested for significant enrichment in KEGG pathways (red nodes). Grey edges indicate an association between either (1) bacteria and host genes or (2) host genes and pathways. Red edges indicate associations with interleukins and chemokines. IL-17 signalling is of interest because of strong roles in immune cell recruitment after bacterial invasion and in tumour response.

18. Materials and methods

Ethics statement

This study was conducted according to the guidelines of the Helsinki Declaration of the World Medical Association. The national-level ethics committee (Hungarian Scientific and Research Ethics Committee of the Medical Research Council, ETTTUKEB/EKU) approved the study. Patients gave consent to participate in this study. After clinical information was collected, patient identifiers were removed so patients cannot be identified directly or indirectly.

Study participants and faecal collection

Faecal samples from 75 lung cancer patients were collected for analysis of fungal, viral, and bacterial microbiomes. All faecal samples were collected at the National Koranyi Institute of Pulmonology, Budapest, Hungary, and the County

Hospital of Pulmonology, Torokbalint, Hungary. Patients received a combination therapy of chemotherapy and immunotherapy. An additional cohort of 11 patients from the same hospital was recruited to provide stool samples for validation of the machine learning models. These patients were treated with immunotherapy drugs only.

Faecal DNA extraction

Stool samples were kept frozen at –80°C before being sent to Novogene (UK) for DNA extraction and sequencing. Stool samples were thoroughly mixed with 900 µL CTAB lysis buffer. All samples were incubated at 65°C for 60 min before being centrifuged at 12 000 g for 5 min at 4°C. Supernatants were transferred to fresh 2 mL microcentrifuge tubes and 900 µL phenol: chloroform: isoamyl alcohol (25:24:1, pH = 6.7; Sigma-Aldrich) was added for extraction. Samples were mixed thoroughly before incubation at room temperature for 10 min. Phase separation occurred by centrifugation at 12 000 g for 15 min at 4°C, and the upper aqueous phase was re-extracted with 900 µL phenol:chloroform:isoamyl alcohol. Samples were then centrifuged at 12 000 g for 10 min at 4°C, and the upper aqueous phases were transferred to 2 mL microcentrifuge tubes. Final extraction was with 900 µL chloroform: isoamyl alcohol (24:1) and centrifugation at 12 000 g for 15 min at 4°C. DNA was precipitated by adding the upper phase to 450 µL isopropanol (Sigma-Aldrich) containing 50 µL 7.5 M ammonium acetate (Fisher). Samples were incubated at −20°C overnight. Shorter incubations (1 h) produced lower DNA yields. Samples were centrifuged at 7500 g for 10 min at 4°C, and DNA pellets were washed three times in 1 mL 70% (v/v) ethanol (Fisher). Pellets were air-dried and re-suspended in 200 µL 75 mM TE buffer (pH = 8.0; Sigma-Aldrich).

Fungal ITS2 sequencing

The concentration of extracted genomic DNA was determined by Qubit 2.0 and DNA quality was checked using gel electrophoresis. PCR reactions used 200 ng DNA with primer sets specific to hypervariable regions (ITS3-2024F, ITS4-2409R). Primer sets had unique barcodes. PCR products was separated on gels and fragment with the proper amplification size were extracted and purified. Purified PCR product was used as a template for library preparation. PCR products were pooled in equal amounts and then end polished, A-tailed, and ligated with adapters. Fragments were filtered with beads. After 1 PCR cycle (to make the library double-stranded), libraries were analysed for size distribution and quantified using real-time PCR. Paired-end sequencing of the library was performed on an Illumina Hiseq2500 (2x250bp).

Fungal ITS2 annotation

ITS2 raw reads were quality controlled, merged and filtered for chimeric reads using DADA2 in R 4.1, BioConductor 3.13 [77]. The median, high-quality, non-chimeric 78,332 ITS2 reads were extracted. From these, amplicon sequencing variants (ASVs) were estimated. Representative sequences of ASVs were aligned to the fungal UNITE database (2017-12-01) [78,79] using the Mothur classifier [80] to improve classification accuracy. ASV counts were summed at species and genus levels and a 10% prevalence filter applied. Normalisation was by applying Bayesian zero-replacement and subsequent centred log ratio (CLR) transformation.

Whole metagenomics sequencing

After DNA extraction, a sequencing library was generated based on Illumina technology and following manufacturers' recommendations. Index codes were added to each sample. Briefly, genomic DNA was randomly fragmented to 350 bp and fragments narrow-size-selected with sample purification beads. Fragments were polished, A-tailed, and ligated to adapters, filtered with beads, and amplified by PCR. Libraries were analysed for size distribution and quantified using real-time PCR. Paired-end sequencing of the library was on an Illumina HiSeq2500 (2x150bp).

Bacterial taxonomic and functional profiling

Quality control of raw reads was performed using Sunbeam [81], including removal of low-quality reads (Phred score > 25 over 10 nucleotides), adaptors, and human-host related reads. Bacterial taxonomic profiling was performed using MetaPhlAn2 [82] with default parameters. For bacterial functional annotation, the HUMAnN2 pipeline was employed [33] with default parameters. Since HUMAnN2 does not directly support paired-end reads, paired-end sample mates were merged as suggested by the HUMAnN2 authors. HUMAnN2 assigns pathways and gene families based on MetaCyc [83] and UniRef90 [84] databases, respectively. Abundances of gene families in each sample were reported in reads per kilobase (RPK) and further normalised by copies per million (CPM), effectively yielding transcript per kilobase million (TPKM). To remove fungi-related abundances, we kept only functional abundances that were directly assigned to bacterial taxa (i.e., excluding other kingdoms and unclassified abundances). The resulting bacteria-only abundances were summed for each feature category (MetaCyc pathways, Enzyme Commission (EC), KEGG orthology terms (KO), KEGG pathways). A 10% prevalence filter was applied to bacterial species, bacterial genera, MetaCyc pathways, EC, KO, and KEGG pathways. Summaries are reported as the median ± median absolute deviation (MAD).

Co-abundance networks

Spearman's correlation analyses of microbial abundances within and between kingdoms were performed by abundance profiles normalized (TSS or CLR) for each kingdom separately. Only significant correlations were considered further

(P<0.05; $|r|$>0.10). Networks were transformed and analysed using the R package TidyGraph [85] and visualised using GGraph [86].

Statistical analysis

To assess the significance of univariate measurements while correcting for confounders such as body mass index (BMI) and gender, we used a rank-based generalised linear model (GLM) as implemented in Rfit [87,88]. In contrast to simple linear models or GLMs, the rank-based GLMs assess unspecific, non-linear trends in data. This test was applied to taxonomic and functional alpha diversity, and contributional alpha diversity. To test for significant differences in bacterial and functional abundances, we used Maaslin2 [89]. In addition to the default settings (total-sum scaling, log-transform, GLM), we set a minimum prevalence filter of 20% (15 samples) with 1e-6 minimum abundance. Tests are corrected for differences in BMI and gender. *P*-values were adjusted for multiple testing using false discovery rate (FDR).

Cohort stratification

Samples were stratified into two groups (High and Low *Candida*) based on median CLR-normalised *Candida* genus abundance. CLR normalisation was performed by (i) removing features with less than 10% sample prevalence but summing counts of low prevalence taxa into a new 'LOW_PREV' feature to preserve the proportions of the remaining features, (ii) Bayesian zero replacement to maintain log-ratios [90,91] and (iii) applying CLR transform as suggested for compositional data [30,90–92]. Samples higher than median *Candida* levels were classified as the High group; others as the Low group.

Diversity analysis

Diversity calculations were performed using the R package vegan [93]. The alpha diversity indices of bacterial and fungal communities were calculated using Shannon, Simpson, observed amplicon sequencing variants, and Pilou's evenness index [93]. For beta diversity, analyses were performed using the Aitchison index and Bayesian-zero replacement as suggested [28,29,91,94] to overcome compositionally related biases. Between-group significance was estimated using PERMANOVA as implemented in the Adonis (vegan [95]) using 1000 permutations. Beta-dispersion (within-group diversity) was assessed using betadisper (vegan).

Aerobe and anaerobe annotation

Culture conditions of bacterial species annotated with MetaPhlAn2 were manually searched in DSMZ (https://www.dsmz.de/collection/catalogue/microorganisms/catalogue/bacteria) and ATTC bacterial collection (https://www.lgcstandards-atcc.org/Products/Cells_and_Microorganisms/Bacteria). Strict or obligate anaerobes were annotated as anaerobes. Facultative anaerobes and obligate aerobes were classified as aerobes. Uncultured bacteria were not annotated.

Machine learning

We performed logistic regression based on GLMnet models as implemented in the R package SIAMCAT [41]. We screened several model normalisations (rank-unit, log-unit, CLR), feature selection (receiver operating curve (ROC)-based with n=20-100) and model (elastic Net, Lasso, Ridge, Lasso-LL, Ridge-LL) settings. Zeros were imputed by dividing the smallest non-zero abundances value by 10. Models were trained with feature abundances from only one category at a time (bacterial taxa, MetaCyc pathways, KEGG orthology terms, KEGG pathways). A 30% prevalence filter was applied in the main cohort of this study. The test cohort was filtered for features prevalent in the main cohort. BMI and gender were included as additional, fixed covariates in the models.
Data augmentation was applied to each training slot of the SIAMCAT model analogous to previous work [42]. Specifically, train-test splits were initiated in the SIAMCAT model object. For each fold, new samples were generated from the training samples per class so that any class imbalance was removed after adding augmented samples. Therefore, training samples were used to fit a negative binomial distribution and the fitted distribution used to sample a fixed number of samples (between 10 and 40). The resulting training cohort had both original and generated samples; test samples remained unchanged. Class labels of generated samples were never used for performance evaluation (i.e., prediction performance) and were relevant only for feature selection and model fitting.

Transepithelial electrical Resistance (TEER)

Bacterial growth conditions

The following strains were assessed: probiotic strain *Lactocaseibacillus rhamnosus* GG (DSM 33156, formerly known as *Lactobacillus rhamnosus* GG, referred to by use of the trademark LGG®), provided by Chr. Hansen A/S, Denmark and four type strains; *Lactobacillus gasseri* (DSM20243), *Lactococcus lactis* subsp. *lactis* (DSM20481), *Bifidobacterium adolescentis* (DSM20083) and *Odoribacter splanchnicus* (DSM20712). LGG® (DSM33156), *L. gasseri* (DSM20243)

and *B. adolescentis* (DSM20083) were inoculated from frozen stock and cultured overnight at 37°C in De Man, Rogosa and Sharpe broth (MRS) broth, pH 6.5 (Difco™, 288110) with 0.05% cysteine hydrochloride monohydrate under anaerobic conditions in 7L anaerobic boxes with 2 sachets of 3.5L AnaeroGen™ Sachets (Thermo Scientific™ AN0035). *L. lactis* (DSM20481) was grown in M17 broth (Biokar BK088HA) supplemented with 0.5% lactose aerobically at 30°C. Ten-fold dilution series were prepared from the overnight cultures of these 4 strains and incubated under the same conditions as mentioned above. Late exponential/early stationary phase were selected based on measures of optical density at 600nm (OD_{600nm}). *O. splanchnicus* (DSM20712) was inoculated from freeze-dried stocks and cultured overnight at 37°C in Gifu Anaerobic Medium broth (GAM; HyServe Code 05422) anaerobically using sealed N_2-gas flushed Hungate tubes (Glasgerätebau Ochs Laborfachhandel, Art. No. 1020471). *O. splanchnicus* (DSM20712) was off-gassed twice per day as the strain is a gas producer. Overnight cultures were subcultured once, and the overnight culture was used for the experiment.

Caco-2 cell Culture

The human colon adenocarcinoma Caco-2 cell line (DSMZ ACC 169, Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany) was cultured in T75 Nunc™ EasYFlask™ (ThermoFisher 156499) in Dulbecco′s Modified Eagle′s Medium (DMEM; Gibco™ 21885-025) supplemented with 20% Fetal Bovine Serum (FBS; Gibco™ 10270) and 1% MEM Non-essential amino acids (Biowest; X0557-100), and 1% Pen-Strep-Amp B (Biological Industries, Israel, 03-033-1B) in 5% CO_2 at 37°C.

Transepithelial electrical resistance (TEER)

When the cells were approximately 80% confluent the medium was removed and cells were washed twice in Hanks balanced salt solution (HBSS; Gibco, 14175). Cells (passage no.22) were dissociated by adding 2mL of TrypLE Express Enzyme (Gibco, 12604) and left for 4 min in the incubator at 37°C. Approximately 10 mL of medium was added to dissociated cells and 10^5 cells/well were seeded onto 12-well, 12 mm Transwell® inserts with 0.4 µm pore polyester membrane insert (Corning; 3460). Growth media was renewed twice per week and cells were cultured for 21 days to obtain a monolayer. The day prior to the TEER experiment, transwells were transferred to the CellZscope2 (Nanoanalytics, Munster, Germany) and the medium was changed to antibiotic-free medium (DMEM supplemented with 20% FBS and 1% MEM non-essential amino acids). The CellZscope2 was maintained in a humidified atmosphere at 37°C with 5% CO_2 and in order to stabilize electrical resistance, overnight measurements of TEER before Caco-2 cells stimulation allowed determination of a baseline and served as a quality control.

Each of the bacterial cultures was washed twice in Phosphate Buffered Saline pH 7.4 (PBS; Sigma 806552) before adjusting the OD_{600nm} to 4.0 in antibiotic-free medium. When adding the bacterial solutions to the transwells the bacterial suspensions were diluted 8-fold resulting in a final OD_{600nm} of 0.5. To stimulate the Caco-2 cells with bacteria, TEER measurements were paused, the CellZscope2 was removed from the CO_2 incubator, and 100 µL of apical medium was replaced by bacterial suspension or controls. LGG® and DMEM medium served as positive and negative control, respectively. The CellZscope2 was transferred back to the CO_2 incubator, and the TEER measurements were resumed and continued overnight for a total of 18 hours. Changes in TEER during bacterial stimulation were calculated relative to the latest value recorded immediately prior to the stimulation (baseline measurement, set to 100%) and mean area under curve (AUC) was calculated for each condition.

Genome-scale metabolic modelling

We detected and downloaded matched bacterial Genome-scale metabolic models (GEMs) from the AGORA (https://vmh.life) [50] and CarveMe collections (https://github.com/cdanielmachado/embl_gems/tree/master/models) [96]. As participants were not restricted to a specific diet, we downloaded 12 different diets (https://vmh.life/#nutrition) for simulations. These were DACH (a healthy diet for people between 19 and 51 years old), EU average, gluten-free, high-protein, ketogenic, low-carbohydrate, Mediterranean (with abundant fresh plant foods, minimally processed food, and olive oil as the principal fat source), type 2 diabetes, unhealthy (very low amounts of fibers and very high sugars and fats), vegan, vegetarian, and high fiber diets.

All analyses were done in COBRApy (v0.17.1) using Python (v3.6.8) and optimisation solvers provided by IBM CPLEX (v12.8.0.0). To make models viable for each diet, essential metabolite concentrations were determined by running COBRApy's minimum_medium function. Simulations were done in the presence of very limited oxygen (1 unit) to mimic the gut environment. Flux variability analysis (FVA) for short-chain fatty acids and lactic acid exchange reactions supporting growth were determined by running flux_variability_analysis functions. These functions were run in loopless mode (loopless=True), allowing 10% deviation from the optimal objective function value (biomass) and minimum overall flux (fraction_of_optimum=0.9, pfba_factor=1.1).

The Metabolic Analysis of Metagenomes using FBA and Optimization (MAMBO) algorithm was used to associate the most likely metabolite abundance profile with our metagenomic samples [97]. MAMBO is based on semi-Markov chains that optimize for a high correlation between a metagenomic relative abundance profile and a predicted metabolic profile. Prediction was based on bacterial GEMs associated with the given metagenomic sample. IBM ILOG CPLEX [98] was used as a solver and run in a Python environment (v3.7).

Fungal co-culture and *in vitro* assays

Candida albicans-RM1000, *C. albicans*-BWP17, and *Saccharomyces boulardii-GFP* were cultured in yeast extract-peptone-dextrose (YPD) agar plates containing 10 $g \cdot L^{-1}$ yeast extract, 20 $g \cdot L^{-1}$ casein peptone and 20 $g \cdot L^{-1}$ glucose at 37°C in aerobic conditions. Overnight liquid cultures were prepared by picking a single colony from a YPD agar plate and inoculating synthetic complete (SC) minimal media containing 6.7 $g \cdot L^{-1}$ yeast nitrogen base supplemented with 1% glucose. For aerobic growth conditions, overnight liquid cultures were incubated at 37°C and 200 RPM for ~16 hours. For microaerobic (1% oxygen, 94% N_2 and 5% H_2) or anaerobic (95% N_2 and 5% H_2) growth conditions, liquid broth was prereduced in anaerobic conditions 24 hours before inoculation. Next, strains were inoculated from YPD agar plates into prereduced SC minimal media supplemented with glucose, and cultures were incubated at 37°C in microaerobic or anaerobic conditions for ~16 hours without shaking.

Growth and competition experiments of *C. albicans* and *S. boulardii* in different carbon sources and oxygen availability

To determine the effect of L-lactate and glucose on *C. albicans* and *S. boulardii* growth, overnight cultures were diluted to optical density $(OD)_{600}$ of 0.1 in 100 µL fresh SC medium supplemented with 1% carbon source (1% glucose, 1% L-lactate, or glucose plus L-lactate each at 0.5%), and incubated in aerobic, microaerobic, or anaerobic conditions, respectively. Cultures were performed in flat-bottom, 96-well plates, sealed to prevent evaporation using BreathEasy film (Sigma-Aldrich). Plates were incubated for 24 hours at 37°C, with continuous orbital shaking at 900 rpm. The ttemperatures used (37°C) did not induce notable hyphae formation in growth assays and therefore did not perturb growth measurements. Cell densities were measured every 10 min at OD_{600} using a microtiter plate reader (BioTek Synergy H1). Growth rates were calculated by plotting log of OD_{600} in log phase and calculating the slope of time points where r^2 was closest to 1, using at least 24 time points over a 4-hour log phase period. The plate reader could not be adapted to strict anaerobic conditions; therefore, determining growth rates of anaerobic experiments was not possible and only final time points were determined for these experiments.

Competition growth experiments were performed by diluting overnight cultures to OD_{600} of 0.1 and inoculating both at 1:1 in a final volume of 100 µL fresh SC medium supplemented with the corresponding 1% carbon source. Cultures were incubated for 24 hours in 96-well plates using the same growth conditions as described for the single-strain cultures. OD_{600} and GFP signals (excitation λ 485, emission λ 510) were measured at intervals of 10 minutes. Since GFP requires oxygen to emit a fluorescent signal, anaerobic samples were incubated aerobically for 1 hour on ice before measuring the final OD_{600} and GFP. GFP signal from anaerobic samples was measured as a single time point after 24 hours.

Colony-forming units $(CFU) \cdot ml^{-1}$ were determined at the end of growth experiments. Cultures were serially diluted 10-fold in an 8-step dilution series and 5 µL of the dilution series were spotted on YPD agar plates and incubated aerobically for 48 hours at 37°C. To determine $CFUs \cdot ml^{-1}$, single colonies were counted from the highest dilutions, divided by spot volume (5 µL) and multiplied by the dilution factor. Technical replicates were made of all experiments.

19. References

1. Miranda LN, van der Heijden IM, Costa SF, Sousa API, Sienra RA, Gobara S, et al. Candida colonisation as a source for candidaemia. J Hosp Infect. W.B. Saunders; 2009;72:9–16.
2. Pappas PG, Lionakis MS, Arendrup MC, Ostrosky-Zeichner L, Kullberg BJ. Invasive candidiasis. Nat Rev Dis Prim 2018 41. Nature Publishing Group; 2018;4:1–20.
3. Cesaro S, Tridello G, Blijlevens N, Ljungman P, Craddock C, Michallet M, et al. Incidence, Risk Factors, and Long-term Outcome of Acute Leukemia Patients With Early Candidemia After Allogeneic Stem Cell Transplantation: A Study by the Acute Leukemia and Infectious Diseases Working Parties of European Society for Blood and Marrow Transplantation. Clin Infect Dis. Oxford Academic; 2018;67:564–72.
4. Mayer FL, Wilson D, Hube B. Candida albicans pathogenicity mechanisms. https://doi.org/104161/viru22913. Taylor & Francis; 2013;4:119–28.
5. Gunsalus KTW, Tornberg-Belanger SN, Matthan NR, Lichtenstein AH, Kumamoto CA. Manipulation of Host Diet To Reduce Gastrointestinal Colonization by the Opportunistic Pathogen Candida albicans. mSphere. American Society for Microbiology; 2016;1.
6. Asbeck EC van, Clemons K V., Stevens DA. Candida parapsilosis: a review of its epidemiology, pathogenesis, clinical aspects, typing and antimicrobial susceptibility. https://doi.org/103109/10408410903213393. Taylor & Francis; 2009;35:283–309.
7. Tan CT, Xu X, Qiao Y, Wang Y. A peptidoglycan storm caused by β-lactam antibiotic's action on host microbiota drives Candida albicans infection. Nat Commun. Springer US; 2021;12.
8. Lopes JP, Stylianou M, Backman E, Holmberg S, Jass J, Claesson R, et al. Evasion of immune surveillance in low oxygen environments enhances candida albicans virulence. MBio. American Society for Microbiology; 2018;9.
9. Ueno K, Matsumoto Y, Uno J, Sasamoto K, Sekimizu K, Kinjo Y, et al. Intestinal Resident Yeast Candida glabrata Requires Cyb2p-Mediated Lactate Assimilation to Adapt in Mouse Intestine. PLoS One. Public Library of Science; 2011;6:e24759.
10. Mirhakkak MH, Schäuble S, Klassert TE, Brunke S, Brandt P, Loos D, et al. Metabolic modeling predicts specific gut bacteria as key determinants for Candida albicans colonization levels. ISME J 2020 155. Nature Publishing Group; 2020;15:1257–70.
11. Fan D, Coughlin LA, Neubauer MM, Kim J, Kim MS, Zhan X, et al. Activation of HIF-1α and LL-37 by commensal bacteria inhibits Candida albicans colonization. Nat Med 2015 217. Nature Publishing Group; 2015;21:808–14.
12. Sellers RS. Translating Mouse Models: Immune Variation and Efficacy Testing. Toxicol Pathol. SAGE Publications Inc.;

2017;45:134–45.
13. Sovran B, Planchais J, Jegou S, Straube M, Lamas B, Natividad JM, et al. Enterobacteriaceae are essential for the modulation of colitis severity by fungi. Microbiome. BioMed Central Ltd.; 2018;6:152.
14. Zhang L, Bahl MI, Roager HM, Fonvig CE, Hellgren LI, Frandsen HL, et al. Environmental spread of microbes impacts the development of metabolic phenotypes in mice transplanted with microbial communities from humans. ISME J. Nature Publishing Group; 2017;11:676–90.
15. Staley C, Kaiser T, Beura LK, Hamilton MJ, Weingarden AR, Bobr A, et al. Stable engraftment of human microbiota into mice with a single oral gavage following antibiotic conditioning. Microbiome. BioMed Central; 2017;5:87.
16. Shultz LD, Brehm MA, Victor Garcia-Martinez J, Greiner DL. Humanized mice for immune system investigation: progress, promise and challenges. Nat Rev Immunol 2012 1211. Nature Publishing Group; 2012;12:786–98.
17. Bratburd JR, Keller C, Vivas E, Gemperline E, Li L, Rey FE, et al. Gut microbial and metabolic responses to salmonella enterica serovar typhimurium and candida albicans. MBio. American Society for Microbiology; 2018;9.
18. Lundberg R, Toft MF, Metzdorff SB, Hansen CHF, Licht TR, Bahl MI, et al. Human microbiota-transplanted C57BL/6 mice and offspring display reduced establishment of key bacteria and reduced immune stimulation compared to mouse microbiota-transplantation. Sci Reports 2020 101. Nature Publishing Group; 2020;10:1–16.
19. Zaborin A, Smith D, Garfield K, Quensen J, Shakhsheer B, Kade M, et al. Membership and behavior of ultra-low-diversity pathogen communities present in the gut of humans during prolonged critical illness. MBio. American Society for Microbiology; 2014;5.
20. van Vliet MJ, Harmsen HJM, de Bont ESJM, Tissing WJE. The Role of Intestinal Microbiota in the Development and Severity of Chemotherapy-Induced Mucositis. PLOS Pathog. Public Library of Science; 2010;6:e1000879.
21. Zwielehner J, Lassl C, Hippe B, Pointner A, Switzeny OJ, Remely M, et al. Changes in Human Fecal Microbiota Due to Chemotherapy Analyzed by TaqMan-PCR, 454 Sequencing and PCR-DGGE Fingerprinting. PLoS One. Public Library of Science; 2011;6:e28654.
22. Lin XB, Dieleman LA, Ketabi A, Bibova I, Sawyer MB, Xue H, et al. Irinotecan (CPT-11) Chemotherapy Alters Intestinal Microbiota in Tumour Bearing Rats. PLoS One. Public Library of Science; 2012;7:e39764.
23. Tang HJ, Liu WL, Lin HL, Lai CC. Epidemiology and Prognostic Factors of Candidemia in Cancer Patients. PLoS One. Public Library of Science; 2014;9:e99103.
24. Zhai B, Ola M, Rolling T, Tosini NL, Joshowitz S, Littmann ER, et al. High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. Nat Med. Springer US; 2020;26:59–64.
25. Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer AC, et al. Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. Microbiome. BioMed Central Ltd; 2020;8:1–20.
26. Marfil-Sánchez A, Zhang L, Alonso-Pernas P, Mirhakkak M, Mueller M, Seelbinder B, et al. An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness. Gut Microbes. Taylor & Francis; 2021;13.
27. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. Nature Publishing Group; 2016;13:581.
28. Luz Calle M. Statistical analysis of metagenomics data. Genomics and Informatics. Korea Genome Organization; 2019.
29. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML. Balances: a New Perspective for Microbiome Analysis. mSystems. American Society for Microbiology; 2018;3:53–71.
30. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. Frontiers; 2017;8:2224.
31. García-Gamboa R, Kirchmayr MR, Gradilla-Hernández MS, Pérez-Brocal V, Moya A, González-Avila M. The intestinal mycobiota and its relationship with overweight, obesity and nutritional aspects. J Hum Nutr Diet. John Wiley & Sons, Ltd; 2021;34:645–55.
32. Sun S, Sun L, Wang K, Qiao S, Zhao X, Hu X, et al. The gut commensal fungus, Candida parapsilosis, promotes high fat-diet induced obesity in mice. Commun Biol 2021 41. Nature Publishing Group; 2021;4:1–11.
33. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. Springer US; 2018;15:962–8.
34. Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol. Elsevier Current Trends; 2016;24:402–13.
35. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017 144. Nature Publishing Group; 2017;14:417–9.
36. Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. Sci Reports 2020 101. Nature Publishing Group; 2020;10:1–15.
37. Ye Y. Identification of Diversity-Generating Retroelements in Human Microbiomes. Int J Mol Sci 2014, Vol 15, Pages 14234-14246. Multidisciplinary Digital Publishing Institute; 2014;15:14234–46.
38. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, et al. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science (80-). 2002;295:2091–4.
39. Medhekar B, Miller JF. Diversity-generating retroelements. Curr Opin Microbiol. Elsevier Current Trends; 2007;10:388–95.
40. Wu L, Gingery M, Abebe M, Arambula D, Czornyj E, Handa S, et al. Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. Nucleic Acids Res. Oxford Academic; 2018;46:11–24.
41. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol 2021 221. BioMed Central; 2021;22:1–27.
42. Lo C, Marculescu R. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. BMC Bioinforma 2019 2012. BioMed Central; 2019;20:1–14.
43. Nazik H, Joubert LM, Secor PR, Sweere JM, Bollyky PL, Sass G, et al. Pseudomonas phage inhibition of Candida albicans. Microbiol (United Kingdom). 2017;163:1568–77.

44. Eeckhaut V, van Immerseel F, Croubels S, de Baere S, Haesebrouck F, Ducatelle R, et al. Butyrate production in phylogenetically diverse Firmicutes isolated from the chicken caecum. Microb Biotechnol. Wiley-Blackwell; 2011;4:503.
45. Gotoh A, Nara M, Sugiyama Y, Sakanaka M, Yachi H, Kitakata A, et al. Use of Gifu Anaerobic Medium for culturing 32 dominant species of human gut microbes and its evaluation based on short-chain fatty acids fermentation profiles. Biosci Biotechnol Biochem. Oxford Academic; 2017;81:2009–17.
46. Hiippala K, Barreto G, Burrello C, Diaz-Basabe A, Suutarinen M, Kainulainen V, et al. Novel Odoribacter splanchnicus Strain and Its Outer Membrane Vesicles Exert Immunoregulatory Effects in vitro. Front Microbiol. Frontiers Media S.A.; 2020;11.
47. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. MBio. American Society for Microbiology; 2019;10.
48. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Front Immunol. 2020;11:906.
49. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett. Oxford Academic; 2009;294:1–8.
50. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 2016 351. Nature Publishing Group; 2016;35:81–9.
51. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic T reg cell homeostasis. Science (80-). American Association for the Advancement of Science; 2013;341:569–73.
52. Guinan J, Wang S, Hazbun TR, Yadav H, Thangamani S. Antibiotic-induced decreases in the levels of microbial-derived short-chain fatty acids correlate with increased gastrointestinal colonization of Candida albicans. Sci Rep. Nature Publishing Group; 2019;9:8872.
53. Nguyen LN, Lopes LCL, Cordero RJB, Nosanchuk JD. Sodium butyrate inhibits pathogenic yeast growth and enhances the functions of macrophages. J Antimicrob Chemother. 2011;66:2573–80.
54. Rigottier-Gois L. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. ISME J 2013 77. Nature Publishing Group; 2013;7:1256–61.
55. Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science (80-). NIH Public Access; 2018;362.
56. Kou Y, Xu X, Zhu Z, Dai L, Tan Y. Microbe-set enrichment analysis facilitates functional interpretation of microbiome profiling data. Sci Rep. Nature Research; 2020;10:1–12.
57. Kumamoto CA. Inflammation and gastrointestinal Candida colonization. Curr Opin Microbiol. NIH Public Access; 2011;14:386.
58. Ene I V., Cheng S-C, Netea MG, Brown AJP. Growth of Candida albicans Cells on the Physiologically Relevant Carbon Source Lactate Affects Their Recognition and Phagocytosis by Immune Cells. Deepe GS, editor. Infect Immun. 2013;81:238–48.
59. Chowdhury S, Fong SS. Computational Modeling of the Human Microbiome. Microorganisms. Multidisciplinary Digital Publishing Institute; 2020;8:197.
60. Stewart GG. Saccharomyces: Introduction. Encycl Food Microbiol Second Ed. 2014;3:297–301.
61. Zangl I, Pap IJ, Aspöck C, Schüller C. The role of Lactobacillus species in the control of Candida via biotrophic interactions. Microb Cell. Shared Science Publishers; 2020;7:1.
62. Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L. Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. Front. Microbiol. Frontiers Research Foundation; 2016. p. 979.
63. Krumbeck JA, Rasmussen HE, Hutkins RW, Clarke J, Shawron K, Keshavarzian A, et al. Probiotic Bifidobacterium strains and galactooligosaccharides improve intestinal barrier function in obese adults but show no synergism when used together as synbiotics. Microbiome. BioMed Central Ltd.; 2018;6:1–16.
64. Oh BS, Choi WJ, Kim JS, Ryu SW, Yu SY, Lee JS, et al. Cell-Free Supernatant of Odoribacter splanchnicus Isolated From Human Feces Exhibits Anti-colorectal Cancer Activity. Front Microbiol. Frontiers Media S.A.; 2021;12:3454.
65. Xing C, Wang M, Ajibade AA, Tan P, Fu C, Chen L, et al. Microbiota regulate innate immune signaling and protective immunity against cancer. Cell Host Microbe. Elsevier Inc.; 2021;29:959-974.e7.
66. Buffie CG, Pamer EG. Microbiota-mediated colonization resistance against intestinal pathogens. Nat Rev Immunol 2013 1311. Nature Publishing Group; 2013;13:790–801.
67. Rea MC, Sit CS, Clayton E, O'Connor PM, Whittal RM, Zheng J, et al. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. Proc Natl Acad Sci U S A. National Academy of Sciences; 2010;107:9352–7.
68. Abt MC, Pamer EG. Commensal bacteria mediated defenses against pathogens. Curr Opin Immunol. Elsevier Ltd; 2014;29:16–22.
69. Kamada N, Núñez G. Regulation of the Immune System by the Resident Intestinal Bacteria. Gastroenterology. W.B. Saunders; 2014;146:1477–88.
70. Koehler P, Stecher M, Cornely OA, Koehler D, Vehreschild MJGT, Bohlius J, et al. Morbidity and mortality of candidaemia in Europe: an epidemiologic meta-analysis. Clin Microbiol Infect. Elsevier B.V.; 2019;25:1200–12.
71. Harris VC, Haak BW, Boele van Hensbroek M, Wiersinga WJ. The Intestinal Microbiome in Infectious Diseases: The Clinical Relevance of a Rapidly Emerging Field. Open Forum Infect Dis. Oxford Academic; 2017;4.
72. Pradhan A, Avelar GM, Bain JM, Childers DS, Larcombe DE, Netea MG, et al. Hypoxia promotes immune evasion by triggering β-glucan masking on the candida albicans cell surface via mitochondrial and cAMP-protein kinase A signaling. MBio. American Society for Microbiology; 2018;9.
73. MacAlpine J, Daniel-Ivad M, Liu Z, Yano J, Revie NM, Todd RT, et al. A small molecule produced by Lactobacillus species blocks Candida albicans filamentation by inhibiting a DYRK1-family kinase. Nat Commun 2021 121. Nature Publishing Group; 2021;12:1–16.
74. Ballou ER, Avelar GM, Childers DS, Mackie J, Bain JM, Wagener J, et al. Lactate signalling regulates fungal β-glucan masking and immune evasion. Nat Microbiol 2016 22. Nature Publishing Group; 2016;2:1–9.

75. Gutierrez D, Weinstock A, Antharam VC, Gu H, Jasbi P, Shi X, et al. Antibiotic-induced gut metabolome and microbiome alterations increase the susceptibility to Candida albicans colonization in the gastrointestinal tract. FEMS Microbiol Ecol. Oxford Academic; 2020;96.
76. Bhaskaran N, Quigley C, Paw C, Butala S, Schneider E, Pandiyan P. Role of short chain fatty acids in controlling Tregs and immunopathology during mucosal infection. Front Microbiol. Frontiers Media S.A.; 2018;9:1995.
77. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004 510. BioMed Central; 2004;5:1–16.
78. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. Oxford Academic; 2019;47:D259–64.
79. Abarenkov K, Nilsson RH, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, et al. The UNITE database for molecular identification of fungi – recent updates and future perspectives. New Phytol. John Wiley & Sons, Ltd; 2010;186:281–5.
80. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high- throughput community sequencing data. Nat Publ Gr. Nature Publishing Group; 2010;7:335–6.
81. Clarke EL, Taylor LJ, Zhao C, Connell A, Lee JJ, Fett B, et al. Sunbeam: An extensible pipeline for analyzing metagenomic sequencing experiments. Microbiome. Microbiome; 2019;7:1–13.
82. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015;12:902–3.
83. Kothari A, Kubo A, Fulcher CA, Weaver DS, Weerasinghe D, Foerster H, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2013;42:D459–71.
84. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, Consortium U. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2014/11/13. Oxford University Press; 2015;31:926–32.
85. Thomas M, Pedersen L. Package 'tidygraph' - A Tidy API for Graph Manipulation. 2020;
86. Thomas Lin Pedersen. Package 'ggraph' - An Implementation of Grammar of Graphics for Graphs and Networks. 2021;
87. Robust Nonparametric Statistical Methods - Thomas P. Hettmansperger, Joseph W. McKean - Google Books [Internet]. [cited 2021 Apr 30]. Available from: https://books.google.de/books?hl=de&lr=&id=6w3LBQAAQBAJ&oi=fnd&pg=PP1&dq=Hettmansperger,+T.P.+and+McKean+J.W.+(2011),+Robust+Nonparametric+Statistical+Methods,+2nd+ed.,+New+York:+Chapman-Hall.&ots=r7W8u49KFE&sig=h1iuYnC752P5ej75VcLaNYBG-PU#v=onepage&q&f=false
88. CRAN - Package Rfit [Internet]. [cited 2021 Apr 30]. Available from: https://cran.r-project.org/web/packages/Rfit/index.html
89. Tipton L, Müller CL, Kurtz ZD, Huang L, Kleerup E, Morris A, et al. Fungi stabilize connectivity in the lung and skin microbial ecosystems. Microbiome 2018 61. BioMed Central; 2018;6:1–14.
90. Schwager E, Mallick H, Ventz S, Huttenhower C. A Bayesian method for detecting pairwise associations in compositional data. PLoS Comput Biol. Public Library of Science; 2017;13:e1005852.
91. Martín-Fernández J-A, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in compositional data sets. Stat Model An Int J. SAGE Publications Ltd; 2015;15:134–58.
92. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. 2015;
93. Oksanen J, Blanchet G, Kindt R, Legendre P, O'Hara RB, Simpson G, et al. vegan: Community Ecology Package. 2011.
94. Kaul A, Mandal S, Davidov O, Peddada SD. Analysis of microbiome data in the presence of excess zeros. Front Microbiol. Frontiers Media S.A.; 2017;8:2114.
95. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. Wiley Online Library; 2003;14:927–30.
96. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. Oxford Academic; 2018;46:7542–53.
97. Garza DR, Van Verk MC, Huynen MA, Dutilh BE. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. Nat Microbiol 2018 34. Nature Publishing Group; 2018;3:456–60.
98. Cplex IBMI. V12. 1: User's Manual for CPLEX. Int Bus Mach Corp. 2009;46:157.

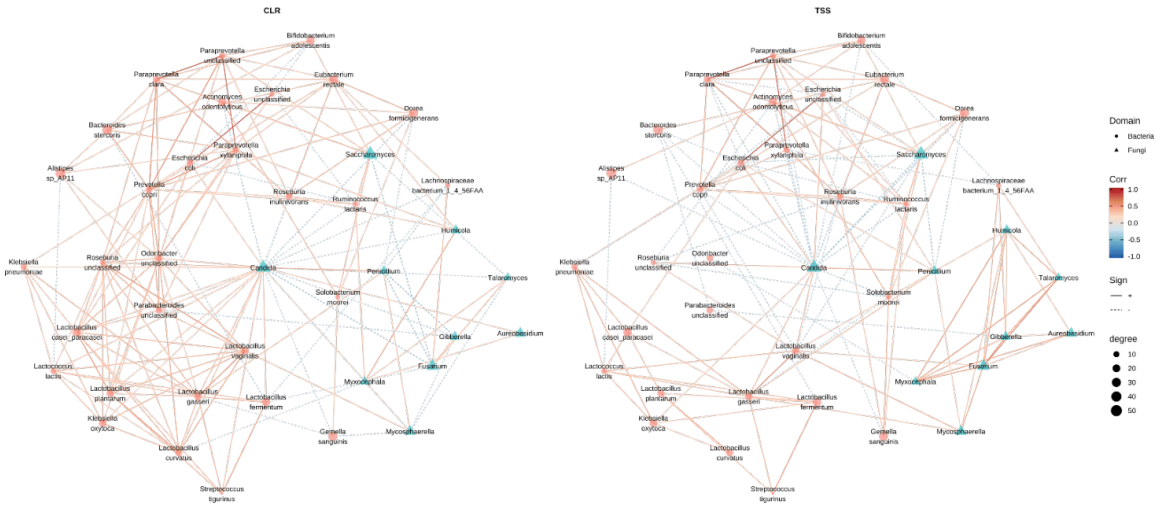
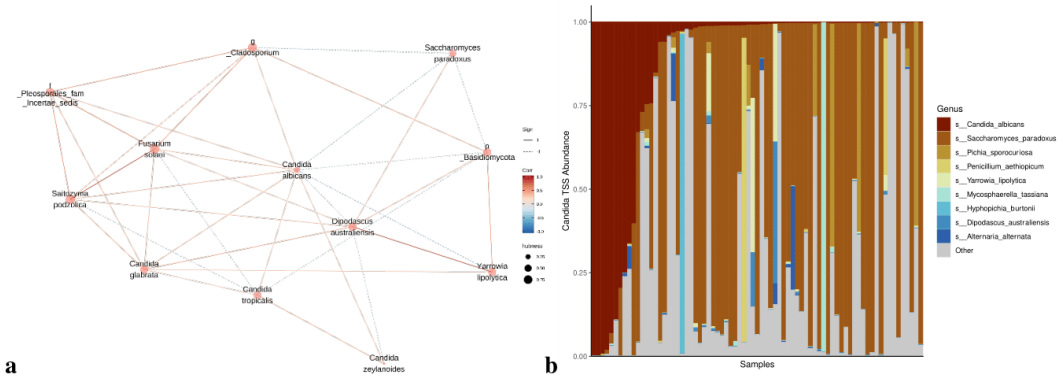


Figure S1. Cross-kingdom co-abundance networks of bacterial species and fungal genera abundances. Correlation was assessed using Spearman on CLR-normalized (left) or TSS-normalized (right) abundance data. Only significant correlations with $p<0.05$ and $|r|>10\%$ are shown. Red, solid edges show positive correlation; blue-dashed edges show negative correlation. Colour strength indicates estimated correlation strength. **(a)** CLR-based networks capture microbial abundance differences changes better, especially when comparing features between distinct compositions. However, strong changes in abundance were lost. **(b)** TSS networks can reveal correlations between dominant species but are less reliable for low-abundance species. **(a-b)** *Lactobacillus spp.* such as *L. gasseri* and *L. lactis*, and *Escherichia spp.* were found to correlate positively with *Candida* regardless of normalization, and SCFA producers like *Lachnospiraceae* and *Actinomyces odontolyticus* were found to correlate negatively.



a b

Figure S2. *Candida albicans* correlated negatively with *Yarrowia lipolytica* and *Saccharomyces paradoxus* abundance. **(a)** Fungal co-abundance networks of significant correlations ($P<0.05$; $|r|>0.10$; Spearman) with *C. albicans*. Solid, red lines indicate positive correlations. Dashed, blue lines indicate negative correlations. **(b)** Fungal species abundance profile indicating the top 10 most abundant fungal species based on total-sum scaling. Species are shown in colours. Other fungi are shown in Gray.

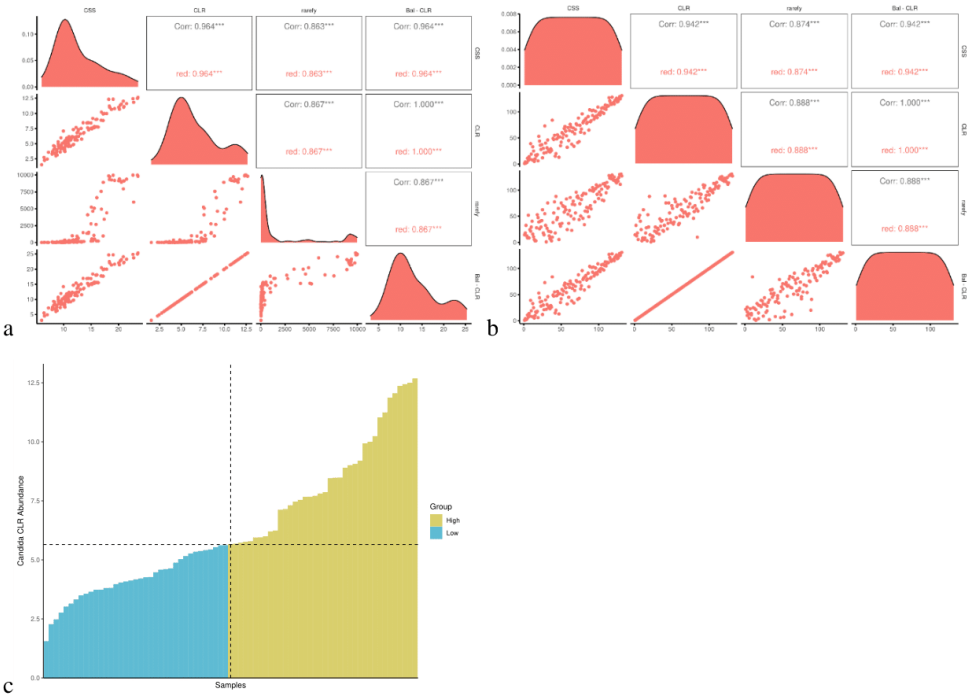


Figure S3. Normalized Candida abundance correlates across different methods. In normalized space **(a)**, we see decent linear correlation between all methods (cumulative sum scaling [CSS]; centralized log-ratio [CLR]; Balance-based CLR [bal-CLR]) except for rarefaction. The latter can be resolved using a simple rank transformation **(b)**, implying, at least, a non-linear relationship between all methods. The strongest agreement can be observed between CSS and CLR. Notably, rarefaction would be a poor choice of normalization here, as it only accounts for differences in sequencing depth but not for compositionality effects in relative abundance data. **(c)** *Candida* CLR normalized abundance (y-axis) per samples (x-axis). Samples were ordered from low to high CLR abundance. Dashed lines indicate the median sample and median abundance thresholds.

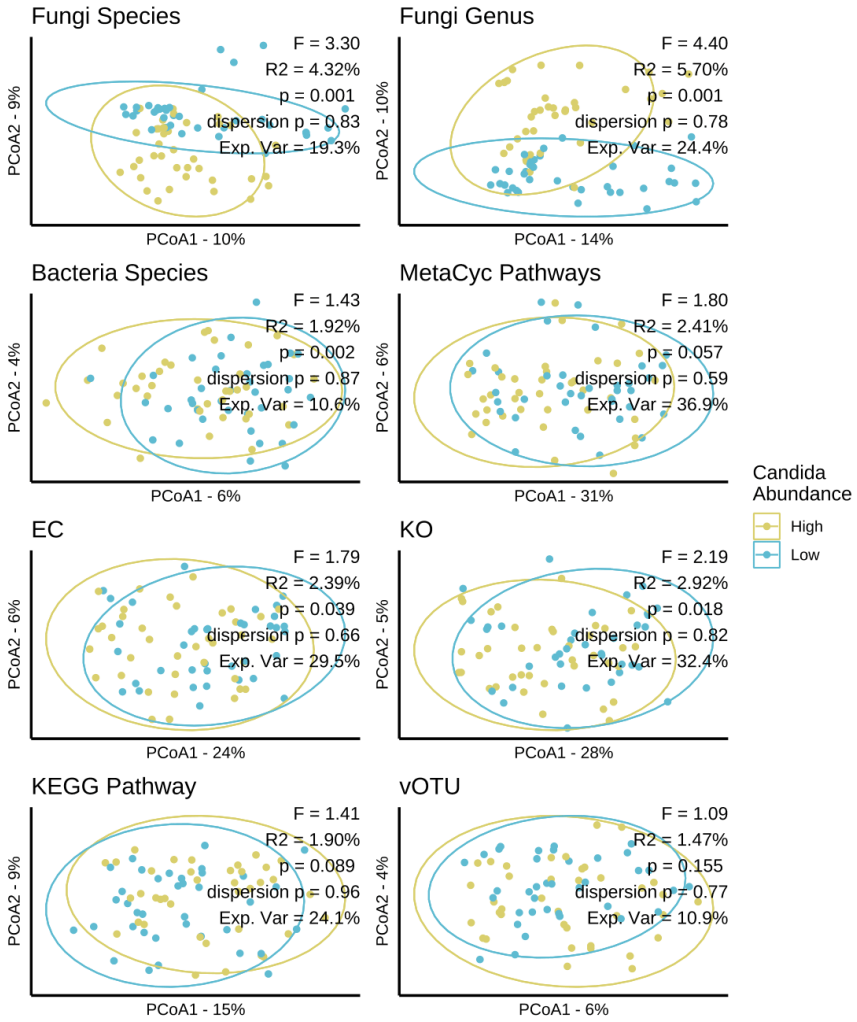


Figure S4. **Principal Coordinate Analysis (PCoA) of beta-diversity for bacterial species and functional, fungal genus and species, and viral OTU abundances**. R^2 describes the estimated explained variance by *Candida* High vs Low grouping (Adonis2; PERMANOVA). MetaCyc has the strongest effect size (together with KEGG Orthology).

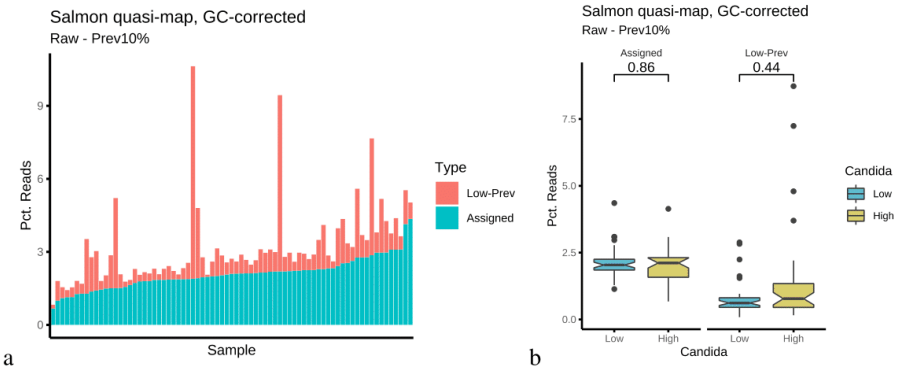


Figure S5. MGV Phage mapping rate. (a) Percentage of reads per sample that passed a 10% prevalence filter (blue) or not (red). More than 30% of reads were assigned to low prevalence contigs in a few samples. **(b)** Percentage of reads assigned to phage contigs by Candida High or Low abundance. Statistical significance was assessed by unpaired Wilcoxon rank-sum tests. On median, 2.5% of reads were assigned to prevalent viral contigs.

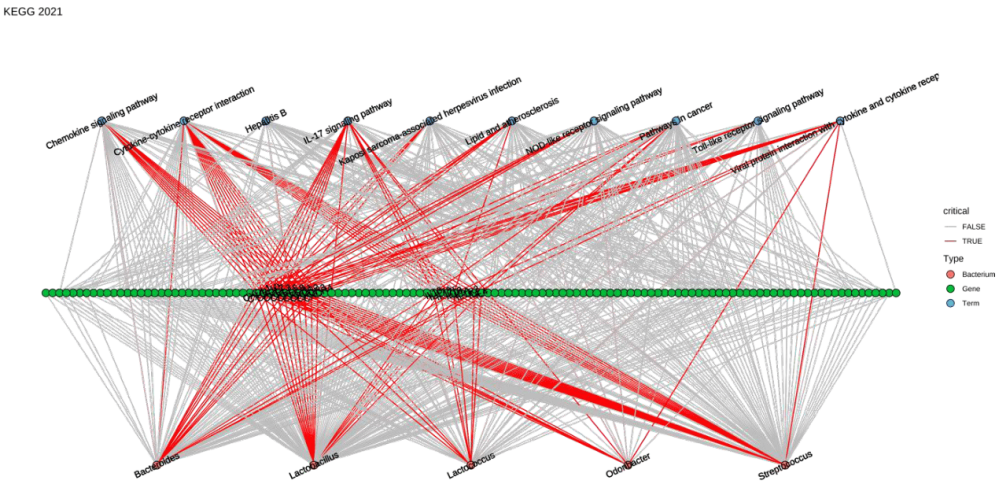


Figure S6. Species enriched in HC group were related to human-disease genes. Microbial Set Enrichment Analysis (MSEA) revealed that bacterial genera with significant differential abundance in the HC group (bottom) were frequently associated with human disease genes (n=280) in stark contrast to LC (n=0). Green nodes indicate human disease-associated genes, which in turn were tested for significant enrichment in KEGG pathways (red nodes). Grey edges indicate an association between either (1) bacteria and host genes or (2) host genes and pathways. Red edges indicate associations with interleukins and chemokines. Especially IL-17 signalling is of interest because it has strong roles in immune cell recruitment after bacterial invasion and in tumour response.

No 7. A pan-genome resembling genome-scale metabolic model platform of 252 *Aspergillus fumigatus* strains reveals growth dependencies from the lung microbiome

FORM 1

Manuscript No. 7

Manuscript title: A pan-genome resembling genome-scale metabolic model platform of 252 *Aspergillus fumigatus* strains reveals growth dependencies from the lung microbiome

Authors: Mohammad H. Mirhakkak, Xiuqiang Chen, Tongta Sae-Ong, Lin Lin Xu, Thorsten Heinekamp, Oliver Kurzai, Amelia Barber, Axel Brakhage, Sebastien Boutin, Sascha Schäubl, Gianni Panagiotou

Bibliographic information (if published or accepted for publication: Citation):

The candidate is (Please tick the appropriate box.)

☐ First author, ☒ Co-first author, ☐ Corresponding author, ☐ Co-author.

Status: in preparation

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Mirhakkak, M.H.	25%	40%		25%	
Chen, X.	25%	40%		25%	
Sae-Ong, T.		10%		5%	
Xu, L.		10%		5%	
Schäuble, S.	25%			20%	20%
Panagiotou, G.	25%			20%	40%
Others			100%		40%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

FORM 2

Manuscript No. 7

Short reference Mirhakkak *et al*., in preparation

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures: My contribution to the publication is reconstruction of pan- and 252 strain GEMs of *A. fumigatus*, following GEM analysis to identify the GEMs differences, and performing the related *in silico* simulations linked to the lung microbiome data.

Figures # 1, S1	☒	**100%**
Figure # 2	☒	Approximate contribution: 85**%** (All the figure parts except F)
Figure # 4	☒	Approximate contribution: 25**%** (Figure part B)
Figures # 3, S2	☐	**0%**

(Add more table boxes depending on the number of figures)

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

A pan-genome resembling genome-scale metabolic model platform of 252 *Aspergillus fumigatus* strains reveals growth dependencies from the lung microbiome

Mohammad H. Mirhakkak[*,1], Xiuqiang Chen[*,1], Tongta Sae-Ong[1], Lin Lin Xu[1], Thorsten Heinekamp[2], Oliver Kurzai[3,4,5], Amelia Barber[6], Axel Brakhage[2,7], Sebastien Boutin[8,9], Sascha Schäuble[#,1], Gianni Panagiotou[#,1,10]

Affiliations

[1] Systems Biology & Bioinformatics Unit, Leibniz Institute for Natural Product Research and Infection Biology–Hans Knöll Institute, 07745, Jena, Germany

[2] Department of Molecular and Applied Microbiology, Leibniz Institute for Natural Product Research and Infection Biology, Jena 07745, Germany

[3] Institute for Hygiene and Microbiology, University of Würzburg, Würzburg, Germany

[4] Research Group Fungal Septomics, Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany

[5] National Reference Center for Invasive Fungal Infections (NRZMyk), Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany

[6] Junior Research Group Fungal Informatics, Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany

[7] Institute of Microbiology, Friedrich Schiller University Jena, Jena 07745, Germany

[8] Department of Infectious Diseases, Medical Microbiology and Hygiene, University Hospital Heidelberg, Im Neuenheimer Feld 324, 69120, Heidelberg, Germany

[9] Translational Lung Research Center Heidelberg (TLRC), German Center for Lung Research (DZL), University of Heidelberg, Heidelberg, Germany

[10] Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, Hong Kong, China

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG - www.dfg.de) CRC/Transregio 124 "'Pathogenic fungi and their human host: Networks of interaction" (DFG project number 210879364, subproject INF to MM, SS and GP) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2051 – Project-ID 390713860 (AB, LR, AB and GP). We thank Yueqiong Ni for fruitful discussions and proofreading our manuscript.

*These authors have contributed equally

#Correspondence to Sascha.Schaeuble@leibniz-hki.de and Gianni.Panagiotou@leibniz-hki.de

Working title: Aspergillus fumigatus strain-GEM collection

Abstract

The saprotrophic fungus *Aspergillus fumigatus* is an opportunistic human fungal pathogen and one of the most common causes of infectious death in immunocompromised patients. Here, we present a collection of 252 strain-specific, genome-scale metabolic models (GEMs) of this important fungal pathogen to study and better understand the metabolic component of its pathogenic versatility. Metabolism showed a notable accessory reactome of 22.7%, which was mainly associated to amino acid, but also nucleotide, and nitrogen metabolism. Presence of reactions and feasible reaction fluxes supporting fungal growth were sufficient to differentiate environmental from clinical strain origin. In addition, shotgun metagenomics of sputum from 40 cystic fibrosis patients before and after they were diagnosed with an *A. fumigatus* infection suggests that the fungus shapes the lung microbiome towards a more beneficial fungal growth environment associated with glycine/serine biosynthesis and the shikimate pathway. Taken together, the here presented first collection of *A. fumigatus* strain-GEMs highlights metabolic differences between different strains of environmental and clinical origin and improves our understanding of fungal survival in the non-native environment of the human lung. These may serve as starting points for the development of alternative clinical intervention strategies targeting the fungal metabolic needs for survival and colonization by diet modification or microbiome intervention..

Introduction

Fungal infections are an emerging public concern for both human health care and economics (*1*, *2*). *Aspergillus fumigatus* is a globally occurring environmental saprotrophic mold that poses a serious threat to hospitalized, particularly immunocompromised patients (*3*). It affects more than 1 million people annually with invasive aspergillosis (IA) and 3 million with chronic pulmonary aspergillosis, both of which show high mortality rates especially in vulnerable cohorts, while diagnostics remain challenging (https://gaffi.org/why/fungal-disease-frequency/, April 2022). On top, the prevalence of chronic obstructive pulmonary disease (COPD) appears to be much higher than estimated with IA contributing substantially to fatal disease progression (*4*), while *A. fumigatus* is related to as many as half of the worldwide cystic fibrosis cases (*5*).

It remains largely unknown to which extent environmental and clinical *A. fumigatus* isolates possess distinct characteristics to cope with external stresses or accessible nutrient profiles in challenging environments such as the human lung. Recently, we explored the genetic diversity of *A. fumigatus* to reveal a remarkably low fraction of core genes shared by all members of the species (69% of the total genes identified) (*6*). However, how the genetic diversity of *A. fumigatus* influences phenotypic and metabolic heterogeneity, particularly in their ability to thrive in the non-native niche of the human lung, has not been addressed yet.

One promising approach to study metabolic capabilities and growth dependencies of pathogens is the application of genome-scale metabolic model (GEM) reconstruction and analysis(*7*). We have previously applied GEM analysis to reveal gut microbiome species that influence colonization levels of the opportunistic fungal pathogen *Candida albicans* (*8*). Given the exponentially increasing number of available genome sequences, the reconstruction of multi-strain genome-scale metabolic models is now possible. The first multi-strain-GEM collection of *Escherichia coli* enabled the definition of strain-specific adaptation to nutrition availability and the prediction of nutritional auxotrophies in some strains (*9*). Protocols and databases were consequently updated to allow for bacterial GEM reconstruction at strain resolution (*10*, *11*), while reconstructions of multi-strain-GEMs remain to be explored in eukaryotes.

In this study we provide the first non-bacterial, multi-strain-GEM reconstruction using *A. fumigatus* as a fungal model organism. Defining metabolic differences between 252 environmental and clinical strain-specific GEMs allowed us to identify metabolic reactions that differ between the two populations. Subsequently, we performed shotgun metagenomics on sputum from 40 cystic fibrosis patients before and after they were diagnosed with

A. fumigatus infection. By computationally defining the metabolic output of the lung microbiome, we propose that the presence of *A. fumigatus* shapes the metabolic landscape of the lung microbiome in a manner favorable for fungal growth. Resolving the impact of genetic diversity on *A. fumigatus* metabolism appears important to extend our understanding of adaptation mechanisms particularly with respect to aromatic amino acid metabolism involving the Shikimate pathway that can ultimately guide the development of new antifungal therapies.

Results

Reconstruction of a most comprehensive *Aspergillus fumigatus* pan-GEM

To create a template for subsequent strain-specific GEM design, we first derived a comprehensive pan-GEM for *A. fumigatus* metabolism (Fig. 1A). To start, we combined available draft reconstructions for *A. fumigatus* with seven automatically derived draft reconstructions for different *Aspergillus spp.* (see Methods for details) (*12*, *13*). This approach allowed us to acquire as many as possible *Aspergillus*-associated reactions in the core metabolism of *A. fumigatus* (i.e. metabolic reactions present in all strains). It also allowed us to acquire a more comprehensive catalogue of optional accessory metabolic reactions by defining strain subset diversity enabling the subsequent strain-specific gap filling curation steps (Fig. 1A). In total, this first draft model was comprised of 7,606 reactions (of which 3,233 are responsible for metabolite exchange with simulated environment) and 3,578 metabolites.

Next, we adapted 62 metabolic components based on fungal and particularly *A. fumigatus* specific literature information to create the biomass objective function essential to simulate *A. fumigatus* growth rates (see Methods) (*14*). The largest fractions of the derived biomass function included carbohydrates and proteins (42.8% and 30%, respectively). Additional essential components included lipids, DNA, and energetic co-factors (Fig. 1B, Supplementary Table S1).

Subsequently, we screened available *A. fumigatus* gene information relevant for metabolism and added 1,453 genes and 2,003 corresponding gene to reaction rules for metabolic reactions, as defined by KEGG (https://www.kegg.jp/) or MetaCyc (https://metacyc.org/, Methods). The remaining 2,370 metabolic reactions (excluding exchange reactions) could not be mapped to any gene in our pan-GEM draft model and were removed from the generic pan-GEM accordingly. However, these reactions were retained for

subsequent strain-specific refinement steps, which require accessory information including gap-filling of fragmented metabolic pathways (Fig. 1A). Concurrently, we incorporated reaction-to-pathway association information from both KEGG and MetaCyc. The broadest pathway categories included amino acids and carbohydrates (Fig. 1C). For the 2,003 metabolic reactions with gene annotation, we predicted nine compartments in our pan-GEM using WoLF PSORT (Fig. 1D, see Methods) (*15*). In parallel, we identified and resolved erroneous energy-generating cycles (*16*) by correcting or removing thermodynamically implausible reactions, such as cases where free energy dissipation was diminished.

For the final curation of our pan-GEM, we generated phenotypic growth data for *A. fumigatus* wild type (Af293 strain) and five mutant strains affecting nitrogen or carbon metabolic components, and considered publicly available gene essentiality information (*17*) (Supplementary Table S2). The initial agreement of our pan-GEM to our metabolite specific growth data was already good (Fig. 1E). To optimize the simulation accuracy of our pan-GEM, we manually resolved any incompatibility between our growth data, available gene essentiality data, and our *in silico* model predictions. These curation efforts improved growth simulation accuracy from 58% to 84% for all tested carbon sources and improved nitrogen growth simulation accuracy from 55% to 85% (Fig. 1E). The pan-GEM achieved 79% and 65% compatibility for the tested phosphorus and sulfur sources respectively and reached 83% if we neglected sulfur source growth data for *ΔniaD* and *ΔlysF* (Fig. 1E) (see Methods). This final model also reached 75% agreement with the available gene essentiality data (Fig. 1F). Altogether, our final pan-GEM of *A. fumigatus* was comprised of 1,453 genes, 3,882 reactions and 4,170 metabolites distributed across 9 compartments. Of these, 3,051 metabolic reactions and 1,957 metabolites were unique across all compartments.

A. fumigatus strains show notable accessory reaction content

Using a genomic dataset of 252 *A. fumigatus* strains from Germany (203 environmental and 49 clinical strains) that we generated previously (*6*), we mapped strain-specific gene profiles to the reference pan-GEM and subsequently derived strain-specific GEMs (Supplementary Table S3). For all strain-specific GEMs, we ensured viable growth was predicted in minimal media with glucose as the carbon source by identifying and resolving minimal sets of essential reactions (*18*) and crosschecking against blocked reactions with FASTCC (see

Methods) (*19*). Model size varied in the different *A. fumigatus* strains from 1,366 to 1,455 reactions (mean 1,413).

Although all strain-specific GEMs are derived from *A. fumigatus*, we found a strikingly low number of core metabolic components shared by all GEMs. In line with the considerably high genome diversity of this organism (*6*), only 984 metabolic genes (69.8%) and 1,150 metabolic reactions (77.3%) were shared by all strain-specific GEMs, resulting in a large degree of metabolic variation across all GEMs (426 accessory genes and 338 accessory reactions). Most accessory content was involved in nucleotide, energy (including oxidative phosphorylation and nitrogen metabolism) and amino acid metabolic pathways (Fig. 2A). Only 56%, 63% and 68% of all reactions included in the strain-GEMs for these pathways, respectively, were conserved across all strain models, demonstrating considerable metabolic pathway variation between strains (Fig. 2A). The majority of the accessory content (70% of accessory genes, 77% of accessory reactions, Table 1, Fig. 2B) was shared by more than 80% of all strain-GEMs. We previously observed that one genetic lineage of *A. fumigatus* possessed significantly fewer accessory genes than the other lineages, including notably fewer metabolic accessory genes (*6*) (Supplementary Fig. S1). In contrast, metabolic reaction content in the strain-GEMs did not show a reduced number of metabolic reactions in this lineage, demonstrating the presence of redundancy among metabolic accessory genes (Fig. 2B). Finally, a small, but notable amount of reactions appeared in at most 40% of all strain-GEMs (Table 1, Fig. 2A) comprising mostly reactions of amino acid metabolism, but also of lipid and energy metabolism including but not limited to nitrogen dependent chorismate pyruvate-lyase or nicotinamidase and acyl-CoA dependent acyltransferases.

Taken together, our generated 252 strains showed notable accessory content and therefore potential metabolic diversity among the strains as well as metabolic robustness despite reduced accessory metabolically relevant genes.

Metabolic activity of 25 reactions allows differentiation between environmental and clinical strain-GEMs

When calculating the pairwise Jaccard distance, we found that strain-specific GEMs differed by at most 15% (Fig. 2C). Neither accessory reaction information nor Jaccard distance allowed discriminating metabolic capabilities between environmental and clinical strains (Fig. 2B, C). However, we identified eight metabolic reactions present primarily in either environmental or clinical strain-GEMs that, when taken together, were able to significantly

differentiate the two populations (exact Fisher-test, $p<0.05$, Fig. 2D). In agreement with the statistical significance of these eight reactions, decision tree machine learning (ML) using the presence or absence of these metabolic reactions, as well as the capability of the strains to grow on different minimal media compositions required only a few steps to correctly categorize 216 out of 252 strains (86%, Fig. 2E). Four of these reactions were chorismate dependent and involved chorismate lyase activity that generates 4-hydroxybenzoate and pyruvate from chorismate. Chorismate lyase activity is linked to differential activity in the shikimate pathway, which has been associated with virulence in *A. fumigatus* (*20*, *21*). Interestingly, the ability to convert chorismate and glutamine to anthranilate, pyruvate and glutamate, as well as, of utilizing carbon metabolites including methionine, acetate, succinate and also thioredoxin was sufficient for the strain origin classification and yielded complementary metabolic discriminators to the sole presence/absence statistical analysis of metabolic reactions in our strain-GEM collection (Fig. 2C, D).

Given that only a few metabolic reactions were sufficient to differentiate strains from clinical and environmental origin by statistical and decision tree analysis, we further explored whether reaction fluxes between strain-GEMs could be used to further improve differentiation of strain origin. We analyzed feasible reaction flux ranges for all strain-GEMs by simulating each on minimal media including glucose as a carbon source and calculating growth supporting flux ranges using flux variability analysis (FVA) (*22*). The derived flux ranges were subsequently used as the input for ML-based classification (see Methods). Classifying environmental from clinical strains achieved an accuracy of 0.80 (AUC = 0.72) with information from only 25 reactions (Fig. 2F, Supplementary Table S4). In addition to previously highlighted chorismate-associated reactions, the ML-model also selected features associated with amino acid reactions, such as homoserine succinate-lyase or L-methionine:oxidized-thioredoxin S-oxidoreductase or cystathionine gamma-lyase, suggesting aromatic amino acid metabolism as a differentiating factor of clinical and environmental *A. fumigatus* strains.

Taken together, we did not observe major differences in strain origin given the strain's accessory gene or reaction content (Fig. 2B, Supplementary Fig. S1) or complete metabolic reaction presence (Fig. 2C). In contrast, we identified a small defined set of reactions that mainly associate to amino acid and chorismate metabolic acitivity which are sufficient to differentiate clinical from environmental origin to a large extend (Fig. 2D-F).

The structure of the lung microbiome changes upon *A. fumigatus* colonization To investigate the applicability of the GEMs from clinical origin for the prediction of metabolic components supporting *A. fumigatus* growth in the human lung, we analyzed sputum samples from 40 cystic fibrosis patients in Germany (*cf.* Methods for cohort description). For all patients, we had an initially culture-negative sample and a subsequent sample that was positive for *A. fumigatus* growth. To investigate the changes to the lung microbiota after *A. fumigatus* colonization, we performed shotgun metagenomic sequencing for all 80 sputum samples (*A. fumigatus*-negative and positive), generating an average of 5.59 Gbp of sequencing data per sample (*s.d*. 0.80 Gbp). Using Kraken for taxonomic profiling, we identified 228 genera and 598 species from all samples using a relative abundance cut-off of 0.1%. Despite the differences in patient cohort, starting biomaterial, and sequencing method, the taxonomic annotation of the top 10 most abundant genera (Supplementary Table S5) showed striking similarities to two recent studies, where the lung microbiome of *A. fumigatus* infected and control patients was investigated using either sputum samples and or bronchoalveolar lavage and 16S rRNA sequencing (*23*, *24*).

The prevalence of the top abundant genera was consistently high (Fig. 3A). Notably, from the top 10 abundant genera, *Sphingomonas*, *Burkholderia, Stenotrophomonas* and *Pseudomonas* were detected as highly abundant (within the 10 most abundant genera) in 8, 12, 17 and 36 out of 80 samples, (10%, 15%, 21%, and 45%, respectively) showing an uneven distribution in the population (Supplementary Table S5). Similarly, the most prevalent species, making up between 2.5% and 51.3% relative abundance were present in most samples (70%) with the exception of *Burkholderia multivorans* and *Sphingomonas sp. FARSPH*, which had a prevalence of 38.8% and 41.3%, respectively (Supplementary Table S5). Intriguingly, *Pseudomonas aeruginosa* was among the top10 most abundant species in 16 samples before and only 11 samples after *A. fumigatus* infection in the same patients, although this species has been described to commonly outgrow in cystic fibrosis patients and co-occur frequently with *A. fumigatus* colonization (*25*).

Since we did not find statistically different alpha- and beta-diversity (Supplementary Fig. 2A, B), we analyzed species co-abundance networks to further examine the compositional changes of the lung microbiome following *A. fumigatus* colonization. Using differential gene correlation analysis (DGCA), we generated networks from differentially correlated microbial pairs in *A. fumigatus* negative versus *A. fumigatus* positive patients' paired samples (Fig. 3B). We then analyzed the resulting networks using MEGENA (*26*) and identified two notable modules in the global network that contained four differentially

abundant species (Metagenomseq, zero-inflated gaussian mixture model $p<0.05$) between the *A. fumigatus* negative and subsequently positive patients' samples. The interactions of these six species — *Schaalia meyeri*, *Abiotrophia defectiva*, *Pseudomonas fulva, Pseudomonas resinovorans, Pseudomonas sp. S1-A32-2,* and *Haemophilus parahaemolyticus* — were highlighted to emphasize discussion (Fig. 3B). Existing edges (class +/0 in Fig. 3B) of *Pseudomonas sp. S1-A32-2, Abiotrophia defectiva and Haemophilus parainfluenzae* with *Clostridium intestinale, Actinomyces sp. Oral taxon 171* and *Streptococcus sp. oral taxon 431* in *A. fumigatus*-negative samples were lost upon colonization with *A. fumigatus*. Similar patterns were observed in the negative association (-/0) between *Schaalia meyeri* with *Clostridium intestinale, Rhizobium leguminosarum* and *Pseudomonas sp. DY-1*, and also *for Pseudomonas resinovorans, Pseudomonas fulva* and *Pseudomonas sp. S1-A32-2* with *Haemophilus influenzae, Streptococcus sp. Oral taxon 064* and *Streptococcus sp. Oral taxon 431* (magenta edges, Fig. 3E). The associations of *Pseudomonas resinovorans* with *Streptococcus salivarius, and Schaalia odontolytica, Schaalia meyeri, Haemophilus parainfluenzae, Streptococcus intermedius* with *Capnocytophaga endodontalis, Streptococcus sp. oral taxon 064, Veillonella dispar* and *Schaalia meyeri*, respectively, also changed direction in the presence of *A. fumigatus* (0/-, -/0, respectively).

To evaluate the functional implications of microbiome restructuring following *A. fumigatus* colonization, we performed KEGG orthology (KO) enrichment analysis in the identified four modules in our co-abundance networks (see details in Methods). Interestingly, both module 1 and 2 were enriched in amino acid metabolism (e.g. phenylalanine, tyrosine, and tryptophan, but also valine and (iso-)leucine). Further enrichments included propanoate and butanoate metabolism (module 1), folate biosynthesis (module 2), glycan biosynthesis and fatty acid metabolism (module 3) and cyanoamino acid metabolism (module 4, Supplementary Table S5).

In summary, albeit not significantly different with alpha or beta diversity we identified a distinct set of co-abundance differences in the lung microbiome upon *A. fumigatus* colonization. The associated enriched metabolic functions pinpointed again towards amino acid, particularly aromatic amino acid pathways, but also fatty acid, nitrogen and sulfur metabolic pathways, suggesting that lung microbiome metabolic activity is reshaped in the presence of *A. fumigatus*.

A. fumigatus is predicted to contribute to the shaping of the lung microbiome and its metabolic activity to support its own growth

We subsequently investigated whether the changes in the lung microbial community triggered by the presence of *A. fumigatus* were accompanied by changes in the metabolic output of the microbiome. To integrate not only the metabolic output of the microbiome, but also the host's, the pathogen's, as well as additional factors such as dietary molecules, we opted for *in silico* prediction. Towards this aim we derived the most likely lung microbiome metabolic profile supporting the relative abundances of our metagenomics species by growth rate using the MAMBO algorithm (*27*). Briefly, MAMBO iteratively calculates growth rates of bacterial models corresponding to a samples' metagenomic profile and infers a metabolome profile. We found significant differences in the beta-diversity of derived metabolite profiles between patient samples before and after *A. fumigatus* infection (Euclidean distance; PERMANOVA, p=0.03, Fig. 4A, Supplementary Table S6).

We next quantified how the changes in the metabolic output of the lung microbiome following *A. fumigatus* colonization might alter the predicted growth of the *A. fumigatus* clinical strain-GEMs. Using the MAMBO-derived metabolite profiles present after *A. fumigatus* colonization, we observed that the GEMs of the 49 clinical strains showed a significant increase in the predicted growth rate compared to GEMs simulated on the metabolic outputs from before *A. fumigatus* colonization (14% increase, Wilcoxon signed rank test, p=3.55e-15, Fig. 4B) suggesting that the changes induced by *A. fumigatus* in the lung microbiome led to a nutritional profile supporting its own growth.

To explore next whether we can identify a connection between the altered lung microbiome and the metabolic capacity of *A. fumigatus* we analyzed feasible flux ranges of reactions that were associated to enriched metabolic subsystems, which we identified before in the *A. fumigatus* affected CF lung microbiome (Fig. 3B, Supplementary Table S4). We identified 54 metabolic reactions across all *A. fumigatus* clinical GEMs that showed significantly altered lower or upper flux ranges to support fungal growth simulated with FVA on MAMBO derived media before compared to after *A. fumigatus* confirmed colonization (FDR corrected paired Wilcoxon test, p≤0.05, Fig. 4C, Supplementary Table S6). Most filtered reactions showed significant differences in the upper range, which suggests increased metabolic activity of *A. fumigatus* (Fig. 4C). Affected pathways mainly included (aromatic) amino acid metabolism, but also nitrogen, sulfur, butanoate or steroid metabolic pathways (Supplementary Table S6). Although predicted flux ranges overlapped between *A. fumigatus* negative and positive samples, the change of direction is mostly consistent on a per-strain-

GEM level (Fig. 4D). Interestingly, the reactions Tryptamine:oxygen oxidoreductase (EC 1.4.3.4) and Chorismate pyruvate-lyase (EC 4.1.3.27) were identified before as major discriminators between environmental and clinical strains simulated on minimal media (Fig. 2E). Only 17 reactions like showed significant differences in both, lower and upper flux bounds (Supplementary Table S6) with NADPH:oxidized-thioredoxin oxidoreductase (EC 1.8.1.9) showing notably constrained flux bound variability across all simulated strain GEMs.

Finally, reinvestigating our phenotypic microarray data for the clinical strain Af293 wild type we found also a positive growth effect of most amino acids tested as carbon source including the aromatic amino acids phenylalanine, tyrosine and tryptophan (Supplementary Table S2). Intermediates such as chorismate, anthralinate and cholines are not part of commercial microarray platforms, but appeared in multiple minimal media based and cystic fibrosis associated ML classification models. These metabolites pinpoint to an elevated role of the Shikimate pathway and warrant further investigation.

Discussion

In this study, we built the first suite of *A. fumigatus* genome-scale strain-specific metabolic reconstructions originating from 252 environmental and clinical isolates from Germany (*6*). We (i) reconstructed a comprehensive pan-GEM of *A. fumigatus* metabolism in a data-driven manner, which we validated against phenotypic microarray and gene essentiality data; (ii) derived 252 strain-specific GEM models by considering respective genome assemblies and manually curating the strain-specific GEMs towards growth feasibility and minimal fractioned network topologies; and (iii) determined metabolic differences differentiating clinical from environmental strains, such as metabolic reactions involving several amino acids, particularly aromatic amino acids as well as chorismate or thioredoxin. Chorismate is an important precursor for aromatic amino acids and formed in the Shikimate pathway. This seven step pathway is not present in animals and enables the synthesis of aromatic amino acids tyrosine, phenylalanine and tryptophan. Thioredoxin is an important factor for DNA synthase metabolism and was associated to *A. fumigatus* virulence before (*28*, *29*).

Multi-strain-GEMs have been utilized previously to elucidate the metabolic diversity of human-pathogenic bacteria. For example, they have defined the pan metabolic capabilities of *Pseudomonas putida* (*30*), loss of fitness relevant pathways for survival in the gastrointestinal environment in extraintestinal *Salmonella spp.* (*31*), and strain-specific metabolic capabilities in *Staphylococcus aureus* linked to pathogenic traits and virulence acquisitions (*32*). Here we bring this strategy to exploring metabolic diversity in a eukaryotic

fungal pathogen for the first time. This strain-specific *A. fumigatus* GEMs platform is publicly available (Biomodels repository) as a platform for investigating the metabolic diversity influencing growth rate capabilities, metabolic adaptation and pathogenicity in this important human fungal pathogen. As a proof-of-concept for the applicability of our fungal GEM collection, we investigated sputum samples from a cohort of 40 cystic fibrosis patients for which samples from before and after confirmed *A. fumigatus* infection were collected. Clinical isolate specific simulations and analysis interestingly showed significantly increased growth rates in the patient samples after a confirmed *A. fumigatus* infection suggesting that the fungus influences the lung microbiome composition towards a more favorable fungal growth. Given these findings, our analyses of strain-resolution GEMs showed that we can recapitulate metabolic cues important for *A. fumigatus* growth that were reported before. Here, we showed that particularly fungal metabolic activity associated to aromatic amino acid metabolism and the Shikimate pathway are not only also important for discriminating environmental from clinical strains, but also to differentiate metabolic activity in clinical *A. fumigatus* strains given metabolite profiles shaped by the lung microbiome in cystic fibrosis patients. Our data-driven analysis highlighted 54 metabolic reactions, for which we predicted significant different flux ranges after *A. fumigatus* colonization of the lung. Our insights suggest that *A. fumigatus* influences its microbiome environment towards a more favorable growth environment. In addition, these reactions do not only appear in aromatic amino acid metabolism, but also sulfur, nitrogen and lipid metabolic pathways, highlighting the advantage of including topological pathway information when analyzing metabolic activity. Together with the 25 metabolic reactions, which we identified as important features of our ML driven classification to differentiate environmental from clinical strains, these reactions represent primarily novel metabolic targets for *A. fumigatus* growth modulation, which need to be investigated further to confirm their potential as biomarker, diagnostic or treatment target with respect to *A. fumigatus* colonization.

As a potential caveat to our study, there may be genomic differences between the clinical strain collection used to build the strain-specific GEMs and the clinical strains present in the cystic fibrosis patients. Though both datasets originate from Germany, the majority of the clinical strains in our GEM collection were from a different patient cohort (invasive aspergillosis). Although genetic diversity differs between strains from invasive and chronic *A. fumigatus* associated disease, we have also shown in our previous study that genomic similarities of clinical strains are relatively high even when the strains originate from different countries (*6*). We used the Af293 *A. fumigatus* genome annotation as a reference, which

precludes potentially metabolically relevant genes present in further *A. fumigatus* annotation. In lack of an existing model system we used MAMBO, which relies on taxonomic species annotation to infer the contribution of the lung microbiome to the most likely nutritional profile in the human lung.

Altogether, the presented analyses demonstrated that strain level fungal genome-scale metabolic modeling is feasible and contributes towards our mechanistic understanding of the genome diversity impact on phenotype of *A. fumigatus*. Moreover we could show a pronounced impact of the lung microbiome profile on available nutrition, which appeared to foster *A. fumigatus* colonization levels. Targeting towards patient stewardships involving diets with particularly suboptimal fungal growth compositions and drugging against fungal specific metabolic routes in the context of (aromatic) amino acid biosynthesis and also the Shikimate pathway (*33*), which is unavailable to the human host, appear promising targets that warrant further analyses in patients suffering from pulmonary diseases involving *A. fumigatus* colonization.

Methods

Biomass formulation

We adapted a specific *A. fumigatus* biomass composition according to several literature sources. First, we assigned proportions of main biomass components as described before (*34*) to 38.8% carbohydrates, 9.9% lipids, 30% proteins as well as 0.6% DNA and 3.7% RNA. Since this resource neglected polyols we added 4% polyols as were reported before for *Aspergillus oryzae* (*35*) to a total of 42.8% carbohydrates. After adding a fraction of 6.6% co-factors these main components made up the total biomass composition together with reported 6.4% ash fraction (*34*). Next, we screened the literature to further specify fractions of subcategories for carbohydrates, e.g. glucans or trehalose (*36*–*38*), lipids, including sterols, phospholipid, neutral lipid and free fatty acid compositions (*39*, *40*), amino acid composition of the protein content (*41*, *42*) and co-factor content including energy carriers, such as NADH or vitamins like riboflavin (*43*, *44*). After calculating the mmol/g content for each fraction, we added the ATP demand according to prior developed models from *Saccharomyces cerevisae* and *Aspergillus niger* (*45*, *46*) as well as added the non-growth associated ATP maintenance value as reported for the curated *Saccharomyces cerevisae* GEM (Supplementary Table S1)(*46*). Finally, we modified the proportion of all components to resemble 1g dry weight (Supplementary Table S1).

Pan-GEM reconstruction

All reconstruction and analysis efforts were done with COBRApy (v0.17.1)(*47*) in python 3.6.8 and the academic version of the IBM CPLEX solver (v12.8.0.0).

We gathered and combined information from automatically generated draft reconstructions based on the CoRoCo pipeline (*12*). We downloaded the Aspergillus CoReCo model for *A. fumigatus* (Biomodels ID MODEL1604280029) and further Aspergillus models from the CoReCo repository including *Aspergillus oryzae* (Biomodels ID MODEL1604280012), *Aspergillus nidulans* (Biomodels ID MODEL1604280008), *Aspergillus niger* (Biomodels ID MODEL1604280021), *Aspergillus clavatus* (Biomodels ID MODEL1604280016), *Aspergillus terreus* (Biomodels ID MODEL1604280019), and *Aspergillus gossypii* (Biomodels ID MODEL1604280044) from the BioModels repository (https://www.ebi.ac.uk/biomodels/). In addition we adapted metabolite and reaction information from a recently published *A. fumigatus* central metabolism model (*13*). Combined together, this yielded a base model consisting of 7,606 reactions of which 3,233 were

exchange reactions and 3,578 metabolites. All subsequent curation efforts were tailored towards keeping only reactions, for which annotation information was available or which were necessary to keep the model feasible. By filtering duplicate reactions and metabolites we reduced the model by 73 reactions and 201 metabolites. The biomass formation was modified to given literature on *Aspergillus fumigatus* metabolism and enriched with information from closely related species when we did not find *A. fumigatus* specific information as described in the *Biomass formulation* section. Next, we screened the KEGG (https://www.kegg.jp/) and MetaCyc (https://metacyc.org/) database for gene annotation to *A. fumigatus* metabolism and added 1,453 genes to the model. Whenever available we adopted AND and OR relationships of the genes for metabolic reaction encoding and cross-checked with gene to reaction encoding in the yeast consensus model (*46*). During this step 2,370 reactions could not be mapped to any annotated gene and were therefore removed from the template model. Instead these reactions were kept aside for subsequent gap-filling procedures. Further curation efforts were run in parallel, since any modification influenced different aspects of the curation efforts. This step included compartmentalization, resolving erroneous energy generating cycles (EGCs) (*16*) and gene essentiality information (*17*) as well as adaptation to phenotypic growth assays (cf. Methods section *Biolog phenotypic microarray*, Supplementary Table S2). To add compartment information for all reactions we applied WoLF PSORT subcellular localization prediction (*15*). A reaction was allocated to a particular compartment if more than 50% of the associated genes are predicted to be located in that compartment with more than 50% probability. Reactions, including exchange reaction, were associated to nine compartments accordingly. These included cytoplasm, mitochondrion, nucleus, peroxisome, endoplasmic reticulum, lipid particles, vacuole, golgi and extracellular space. In cases where the prediction was ambiguous or precluded a viable model as measured by biomass production based on defined minimal media we adapted concurrent alternative compartment localization as either predicted by WoLF PSORT or included in the curated *S. cerevisae* GEM (*46*). Compartment-connecting transport reactions were adapted from the yeast consensus model (*46*). A minimal set of additional necessary transport reactions were added by using gap-filling functionality as provided by COBRApy in order to allow biomass precursor production based on minimal media with glucose.

In parallel we resolved again (EGCs)(*16*) and adapted our GEM model to publicly available gene essentiality data (*17*). EGCs are metabolic reactions running in a potentially non-trivial circle without a net-flux except for generating energy carriers. ATP, CTP, GTP, UTP, ITP, NADH, NADPH, FADH2, FMNH2, Acetyl-CoA, L-Glutamate, ubiquinol-8,

demethylmenaquinol-8, menaquinol-8 were part in at least one EGC (Supplementary Table S7). The directions of 44 reactions were refined considering the reaction directionality according to the BiGG (*48*) and BRENDA (*49*) databases and Gibbs free energy of the reactions stated in the MetaCyc database (*50*). Incompatible gene essentiality information, were resolved by either correcting feasible thermodynamically reaction direction or removal of erroneously present reactions without gene annotation.

Finally, we ran several phenotypic microarrays with the *A. fumigatus* reference strain Af293 and five mutant strains (detailed out in the next section) and identified essential carbon, nitrogen, sulfur and phosphorus components (Supplementary Table S2). This step included again plausible correction of thermodynamically feasible reaction directions and removal of present reactions without gene annotation. In case incompatibilities could not be resolved in this way we screened our catalogue of initially removed reactions without gene annotation using gap-filling procedures using the COBRApy gap-fill functionality. Resolving growth compatibility for two of our mutants (*ΔniaD* and *ΔlysF*, cf. *Biolog phenotypic microarray*) on sulfur would have caused a notable performance drop in the overall growth prediction for all investigated growth media and gene essentiality performance. Since growth accuracy on sulfur was very good for the remaining wild type and two mutant strains and because optimizing growth on carbon and nitrogen sources was very good over all mutant data, we refrained from resolving *ΔniaD* and *ΔlysF* sulfur growth accuracy (Fig. 1B).

Strain-GEM reconstruction and curation

Recently, the pan-genome of *A. fumigatus* was derived for 300 environmental and clinical strains from a global distribution (*6*). Mapping the genomes for 252 of these strains to the Af293 *A. fumigatus* reference genome annotation, we identified metabolically relevant genes by requiring at least 95% sequence identity (small deviations from that threshold did not change the results) under the rational that high sequence identity preserves metabolic function. To ensure that all strain-specific GEMs were showing non-zero growth capabilities based on minimal media with glucose as carbon source we identified and resolved minimal sets of essential reactions that needed to operate in adaptation to the minimal cut set concept (*18*). Finally, we guaranteed a consistent network property by identifying and discarding blocked reactions per isolate with FASTCC (*19*).

Biolog phenotypic microarray

Fungal strains were grown at 25°C for 7 days prior to experimental assay on Malt agar supplemented with 5 mM uracil. Mature conidia were harvested by rubbing plates with sterile distilled water and the resulting solution filtered through a 30 µm cell strainer to remove mycelial fragments. The spore solutions were then adjusted to a transmittance of 75%. Phenotypic microarrays were performed using Biolog Phenotypic Microarray plates PM1, PM2, PM3, and PM4 and plates prepared following the manufacture's protocol for filamentous fungi with the modification of 0.16 ml of Biolog Redox Dye D added to the master mix of each plate for the quantification of metabolic activity. The plates were incubated at 37°C for three days and the metabolic activity measured colorimetrically using an OmniLog microplate reader with readings taken every 15 minutes. Experiments were performed in biological duplicates or triplicates (Supplementary Table S2). The phenotypic microarray results were analyzed in R, and statistical comparison was done using Dunnett-type comparison of growth signals of negative control against all the other wells in one plate. All the wells with greater signals than the negative control and p-value < 0.05 were considered as growth cases.

Cystic Fibrosis sample acquisition

This study was approved by the ethics committee of the University of Heidelberg and written informed consent was obtained from all patients or their parents/legal guardians (S-370/2011). Patients were treated according to standard of care (*51*). The diagnosis of cystic fibrosis was verified by established diagnostic criteria (*52*, *53*). Spontaneously expectorated sputum was collected during visits at the Cystic Fibrosis Center at the University Hospital Heidelberg and frozen in liquid nitrogen on the day of visit. Pulmonary function testing was performed on the same day of sputum collection according to ATS/ERS (European Respiratory Society) guidelines (*54*, *55*) and FEV1 (Forced expiratory value in 1s) values were normalized according to the global lung function initiative (*56*).

Metagenomics and subsequent MAMBO analysis

Sputum samples of 40 cystic fibrosis patients were collected before and after they had positive *A. fumigatus* colonization. The cohort comprised 15 females and 25 males (80 samples in total) with age=23.6±4.96 (mean±standard deviation) before *A. fumigatus* infection.

Trimmomatic was used to clip adapter and low-quality bases (v0.36, ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:1:TRUE, LEADING:3, TRAILING:3,

SLIDINGWINDOW:4:15, MINLEN:30). Remaining reads with less than 30 base pairs length were discarded. BWA (v07.17) was used to align quality filtered reads to the human reference genome (hg38). From originally 1.9e+07±2.7e+06 (mean±standard deviation) metagenomic reads 7.8e+05±8.9e+05 remained after preprocessing per sample. To estimate the taxonomic composition of the non-human reads, Kraken2 (v2.0.7,, default parameters) was used with its standard database as reference. Low abundance species were removed at a cut-off=0.1% (Supplementary Table S5). For the functional composition annotation, the MG-RAST (v4.0.3) pipeline was used to assign non-human reads to KEGG pathways. R packages vegan (v2.5) and picante (v1.8.2) were used to calculate alpha diversity with Shannon and Phylogenetic diversity index for each samples on the read counts of species, Statistical differences between samples before (*A. fumigatus* -) and after (*A. fumigatus* -) infection with *A. fumigatus* were obtained by Wilcoxon signed rank test. For beta diversity, R package coda.base (v0.3.1) was used to calculate the pairwise Aitchison distance for samples on the relative abundance of species. Statistical difference between samples before (*A. fumigatus* -) and after (*A. fumigatus* -) infection with *A. fumigatus* was calculated by PERMANOVA.

The abundance network was constructed based on relative abundance values of all detected species (prevalence filter: 10%, abundance filter: 0.1%). DGCA (v2.0.0) was applied to construct the network from differentially correlated microbial pairs in paired cystic fibrosis samples before compared to after *A. fumigatus* infection (empirical p value<0.05). Subsequently, MEGENA (v1.3.7) was used to identify co-expressed modules in the constructed network using significant differing microbial pairs (module p value<0.05). To identify molecular functions, we investigated enrichment of KEGG pathway information (https://www.genome.jp/kegg/pathway.html) by permutation testing to determine whether correlations between modules and KOs were possible by chance or not (*57*). Firstly, for a given module all correlation values and p values between a particular KEGG Orthology and all species in this module were obtained using the spearman correlation methodThe sum of absolute correlation values in this module was then calculated. Following that, the same number of species in the module were chosen at random 1000 times from all species, and the sum of absolute values of every correlation was calculated for each set. Finally, the sum of significant correlation values in a given module was evaluated whether it was higher than in 95% of the sums of significant correlation values in the repeated random selected species.

To associate the most likely metabolite abundance profile to our metagenomic samples we applied the MAMBO algorithm (*27*). In brief, MAMBO optimizes a high correlating

metabolic profile to a given metagenomic relative abundance profile based on bacterial GEMs associated to a given metagenomic sample. We opted for using only GEMs associated to species from the metagenomics profile at an abundance threshold of 0.5% to ensure that the abstracted media is associated to dominant species in the lung. 53 bacteria species were present beyond this more rigid threshold for which we found and downloaded 51 matching bacterial GEMs from the AGORA (https://vmh.life)(*58*) and CarveMe collection (https://github.com/cdanielmachado/embl_gems/tree/master/models) (*59*). Optimizations were run in a python environment (v3.7) using a HPC (192 cores, 1TB RAM). After removing metabolites that appeared in less than 80% of the samples, missing metabolite abundance values in any remaining sample were imputed with MICE (miceRanger, v1.4.0 with m=1 and maxiter = 50) resulting in a final list of 357 metabolites (Supplementary Table S6). Metabolites differing significantly in the MAMBO associated media for samples before compared to after *A. fumigatus* infected were identified in three steps. Firstly, candidate metabolites were selected using $p<0.2$ as cut-off from a Wilcoxon signed rank test (*60*). Secondly, the identified metabolites in the first step were investigated with an adaptive Lasso statistical design using R package glmnet (v4.1) to identify important metabolites for group differentiation (family="binomial", type.measure="class")(*61*). Finally, we used a fixed Lasso design using R package selectiveInference (v1.2.5) as post-selection inference method to identify significance for each of the important metabolites ($p\leq0.05$) (*62*).

Machine learning approach

Unless otherwise noted we used the following machine learning methodology. In cases where the group sizes were unbalanced (e.g. environmental and clinical origin labels) we randomly sampled 50% of the majority group and oversampled samples of the minority group using ADASYN implemented in R package imbalance (v1.0.2.1). Subsequently, feature selection was performed using Boruta (v7.0.0), VSURF (v1.1.0), MUVR (v0.0.973) and sPLS-DA (mixOmics, v6.16.0). These steps were repeated 50 times and selected features as well as their selection frequency recorded. Finally, the Extra Trees algorithm from PyCaret (v2.3.2) was run for different feature sets scanning different frequency cut-offs to optimize the best cut-off value for ML performance. The best hyperparameters of the Extra Trees model was automatically selected by scikit-optimize (v0.8.1, bayesian optimization).

References

1. M. C. Fisher, D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw, S. J. Gurr, Emerging fungal threats to animal, plant and ecosystem health. *Nature*. **484**, 186–194 (2012).
2. M. C. Fisher, S. J. Gurr, C. A. Cuomo, D. S. Blehert, H. Jin, E. H. Stukenbrock, J. E. Stajich, R. Kahmann, C. Boone, D. W. Denning, N. A. R. Gow, B. S. Klein, J. W. Kronstad, D. C. Sheppard, J. W. Taylor, G. D. Wright, J. Heitman, A. Casadevall, L. E. Cowen, Threats posed by the fungal kingdom to humans, wildlife, and agriculture. *MBio*. **11** (2020), doi:10.1128/MBIO.00449-20.
3. A. Arastehfar, A. Carvalho, J. Houbraken, L. Lombardi, R. Garcia-Rubio, J. D. Jenks, O. Rivero-Menendez, R. Aljohani, I. D. Jacobsen, J. Berman, N. Osherov, M. T. Hedayati, M. Ilkit, D. James-Armstrong, T. Gabaldón, J. Meletiadis, M. Kostrzewa, W. Pan, C. Lass-Flörl, D. S. Perlin, M. Hoenigl, Aspergillus fumigatus and aspergillosis: From basics to clinics. *Stud. Mycol.* **100**, 100115 (2021).
4. E. E. Hammond, C. S. McDonald, J. Vestbo, D. W. Denning, The global impact of Aspergillus infection on COPD. *BMC Pulm. Med.* **20**, 1–10 (2020).
5. J. Armstead, J. Morris, D. W. Denning, Multi-Country Estimate of Different Manifestations of Aspergillosis in Cystic Fibrosis. *PLoS One*. **9**, e98502 (2014).
6. A. E. Barber, T. Sae-Ong, K. Kang, B. Seelbinder, J. Li, G. Walther, G. Panagiotou, O. Kurzai, Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection. *Nat. Microbiol. 2021 612*. **6**, 1526–1536 (2021).
7. A. Bordbar, J. M. Monk, Z. A. King, B. O. Palsson, Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet. 2014 152*. **15**, 107–120 (2014).
8. M. H. Mirhakkak, S. Schäuble, T. E. Klassert, S. Brunke, P. Brandt, D. Loos, R. V. Uribe, F. Senne de Oliveira Lino, Y. Ni, S. Vylkova, H. Slevogt, B. Hube, G. J. Weiss, M. O. A. Sommer, G. Panagiotou, Metabolic modeling predicts specific gut bacteria as key determinants for Candida albicans colonization levels. *ISME J.* **15**, 1257–1270 (2021).
9. J. M. Monk, P. Charusanti, R. K. Aziz, J. A. Lerman, N. Premyodhin, J. D. Orth, A. M. Feist, B. Ø. Palsson, Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20338–43 (2013).
10. C. J. Norsigian, N. Pusarla, J. L. McConn, J. T. Yurkovich, A. Dräger, B. O. Palsson, Z. King, BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **48**, D402–D406 (2020).
11. C. J. Norsigian, X. Fang, Y. Seif, J. M. Monk, B. O. Palsson, A workflow for generating multi-strain genome-scale metabolic models of prokaryotes. *Nat. Protoc.* **15**, 1 (2020).
12. S. Castillo, D. Barth, M. Arvas, T. M. Pakula, E. Pitkänen, P. Blomberg, T. Seppanen-Laakso, H. Nygren, D. Sivasiddarthan, M. Penttilä, M. Oja, Whole-genome metabolic model of Trichoderma reesei built by comparative reconstruction. *Biotechnol. Biofuels*. **9**, 252 (2016).
13. M. Srivastava, E. Bencurova, S. K. Gupta, E. Weiss, J. Löffler, T. Dandekar, Aspergillus fumigatus Challenged by Human Dendritic Cells: Metabolic and Regulatory Pathway Responses Testify a Tight Battle. *Front. Cell. Infect. Microbiol.* **0**, 168 (2019).
14. A. M. Feist, B. O. Palsson, The biomass objective function. *Curr. Opin. Microbiol.* **13**, 344–349 (2010).
15. P. Horton, K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, K.

Nakai, WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585 (2007).
16. C. J. Fritzemeier, D. Hartleb, B. Szappanos, B. Papp, M. J. Lercher, Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLOS Comput. Biol.* **13**, e1005494 (2017).
17. W. Hu, S. Sillaots, S. Lemieux, J. Davison, S. Kauffman, A. Breton, A. Linteau, C. Xin, J. Bowman, J. Becker, B. Jiang, T. Roemer, Essential Gene Identification and Drug Target Prioritization in Aspergillus fumigatus. *PLoS Pathog.* **3** (2007), doi:10.1371/JOURNAL.PPAT.0030024.
18. S. Klamt, E. D. Gilles, Minimal cut sets in biochemical reaction networks. *Bioinformatics*. **20**, 226–234 (2004).
19. N. Vlassis, M. P. Pacheco, T. Sauter, Fast Reconstruction of Compact Context-Specific Metabolic Network Models. **10**, e1003424 (2014).
20. A. Sasse, S. N. Hamer, J. Amich, J. Binder, S. Krappmann, Mutant characterization and in vivo conditional repression identify aromatic amino acid biosynthesis to be essential for Aspergillus fumigatus virulence. *Virulence*. **7**, 56 (2016).
21. J. S. Brown, A. Aufauvre-Brown, J. Brown, J. M. Jennings, H. Arst, D. W. Holden, H. DW, Signature-tagged and directed mutagenesis identify PABA synthetase as essential for Aspergillus fumigatus pathogenicity. *Mol. Microbiol.* **36**, 1371–1380 (2000).
22. R. Mahadevan, C. H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–76 (2003).
23. A. Hérivaux, J. R. Willis, T. Mercier, K. Lagrou, S. M. Goncąlves, R. A. Goncąles, J. Maertens, A. Carvalho, T. Gabaldón, C. Cunha, Lung microbiota predict invasive pulmonary aspergillosis and its outcome in immunocompromised patients. *Thorax* (2021), doi:10.1136/THORAXJNL-2020-216179.
24. O. G. G. de Almeida, C. P. da C. Capizzani, L. Tonani, P. H. Grizante Barião, A. F. da Cunha, E. C. P. De Martinis, L. A. G. M. M. Torres, M. R. von Zeska Kress, The Lung Microbiome of Three Young Brazilian Patients With Cystic Fibrosis Colonized by Fungi. *Front. Cell. Infect. Microbiol.* **10**, 668 (2020).
25. S. M. Gonçalves, K. Lagrou, C. Duarte-Oliveira, J. A. Maertens, C. Cunha, A. Carvalho, The microbiome-metabolome crosstalk in the pathogenesis of respiratory fungal diseases. *Virulence*. **8**, 673–684 (2017).
26. W. M. Song, B. Zhang, Multiscale Embedded Gene Co-expression Network Analysis. *PLOS Comput. Biol.* **11**, e1004574 (2015).
27. D. R. Garza, M. C. Verk, M. A. Huynen, B. E. Dutilh, Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nat. Microbiol. 2018 34*. **3**, 456 (2018).
28. A. C. Marshall, S. E. Kidd, S. J. Lamont-Friedrich, G. Arentz, P. Hoffmann, B. R. Coad, J. B. Bruning, Structure, mechanism, and inhibition of aspergillus fumigatus thioredoxin reductase. *Antimicrob. Agents Chemother.* **63** (2019), doi:10.1128/AAC.02281-18/SUPPL_FILE/AAC.02281-18-S0001.PDF.
29. J. Binder, Y. Shadkchan, N. Osherov, S. Krappmann, The Essential Thioredoxin Reductase of the Human Pathogenic Mold Aspergillus fumigatus Is a Promising Antifungal Target. *Front. Microbiol.* **11**, 1383 (2020).
30. J. Nogales, J. Mueller, S. Gudmundsson, F. J. Canalejo, E. Duque, J. Monk, A. M. Feist, J. L. Ramos, W. Niu, B. O. Palsson, High-quality genome-scale metabolic modelling of Pseudomonas putida highlights its broad metabolic capabilities. *Environ. Microbiol.* **22**, 255–269 (2020).
31. Y. Seif, E. Kavvas, J.-C. Lachance, J. T. Yurkovich, S.-P. Nuccio, X. Fang, E. Catoiu,

M. Raffatellu, B. O. Palsson, J. M. Monk, Genome-scale metabolic reconstructions of multiple Salmonella strains reveal serovar-specific metabolic traits. *Nat. Commun.* **9**, 3771 (2018).
32. E. Bosi, J. M. Monk, R. K. Aziz, M. Fondi, V. Nizet, B. Ø. Palsson, Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci.* **113**, E3801–E3809 (2016).
33. K. Jastrzębowska, I. Gabriel, Inhibitors of amino acids biosynthesis as antifungal agents. *Amino Acids*. **47**, 227–249 (2015).
34. T. Tsukahara, Changes in Chemical Composition of Conidia of Aspergillus fumigatus during Maturation and Germination. *Microbiol. Immunol.* **24**, 747–751 (1980).
35. H. Pedersen, M. Carlsen, J. Nielsen, Identification of Enzymes and Quantification of Metabolic Fluxes in the Wild Type and in a Recombinant Aspergillus oryzae Strain. *Appl. Environ. Microbiol.* **65**, 11 (1999).
36. D. Maubon, S. Park, M. Tanguy, M. Huerre, C. Schmitt, M. C. Prévost, D. S. Perlin, J. P. Latgé, A. Beauvais, AGS3, an α(1–3)glucan synthase gene family member of Aspergillus fumigatus, modulates mycelium growth in the lung of experimentally infected mice. *Fungal Genet. Biol.* **43**, 366–375 (2006).
37. M. J. Lee, D. C. Sheppard, in *The mycota : a comprehensive treatise on fungi as experimental systems for basic and applied research*, K. Esser, D. Hoffmeister, Eds. (Cham: Springer, ed. 3, 2016).
38. D. Hagiwara, S. Suzuki, K. Kamei, T. Gonoi, S. Kawamoto, The role of AtfA and HOG MAPK pathway in stress tolerance in conidia of Aspergillus fumigatus. *Fungal Genet. Biol.* **73**, 138–149 (2014).
39. B. Ghfir, J. L. Fonvieille, Y. Koulali, R. Ecalle, R. Dargent, Effect of essential oil ofHyssopus officinalis on the lipid composition ofAspergillus fumigatus. *Mycopathol. 1994 1263*. **126**, 163–167 (1994).
40. L. Alcazar-Fuoli, E. Mellado, G. Garcia-Effron, J. F. Lopez, J. O. Grimalt, J. M. Cuenca-Estrella, J. L. Rodriguez-Tudela, Ergosterol biosynthesis pathway in Aspergillus fumigatus. *Steroids*. **73**, 339–347 (2008).
41. M. Schrettl, N. Beckmann, J. Varga, T. Heinekamp, I. D. Jacobsen, C. Jöchl, T. A. Moussa, S. Wang, F. Gsaller, M. Blatzer, E. R. Werner, W. C. Niermann, A. A. Brakhage, H. Haas, HapX-Mediated Adaption to Iron Starvation Is Crucial for Virulence of Aspergillus fumigatus. *PLOS Pathog.* **6**, e1001124 (2010).
42. E. Barreto-Bergter, P. A.J. Gorin, L. R. Travassos, Cell constituents of mycelia and conidia of Aspergillus fumigatus. *Carbohydr. Res.* **95**, 205–217 (1981).
43. A.-M. Dietl, Z. Meir, Y. Shadkchan, N. Osherov, H. Haas, Riboflavin and pantothenic acid biosynthesis are crucial for iron homeostasis and virulence in the pathogenic mold Aspergillus fumigatus. *https://doi.org/10.1080/21505594.2018.1482181*. **9**, 1036–1049 (2018).
44. S. W. Chocklett, P. Sobrado, Aspergillus fumigatus SidA Is a Highly Specific Ornithine Hydroxylase with Bound Flavin Cofactor. *Biochemistry*. **49**, 6777–6783 (2010).
45. M. R. Andersen, M. L. Nielsen, J. Nielsen, Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. *Mol. Syst. Biol.* **4**, 178 (2008).
46. H. Lu, F. Li, B. J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marcišauskas, P. M. Anton, D. Lappa, C. Lieven, M. E. Beber, N. Sonnenschein, E. J. Kerkhoven, J. Nielsen, A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun. 2019 101*. **10**, 1–13

(2019).
47. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
48. Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, N. E. Lewis, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* (2015), doi:10.1093/nar/gkv1049.
49. S. Placzek, I. Schomburg, A. Chang, L. Jeske, M. Ulbrich, J. Tillack, D. Schomburg, BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res.* **45**, D380–D388 (2017).
50. R. Caspi, R. Billington, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, P. E. Midford, Q. Ong, W. K. Ong, S. Paley, P. Subhraveti, P. D. Karp, The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **46**, D633–D639 (2018).
51. C. Castellani, A. J. A. Duff, S. C. Bell, H. G. M. Heijerman, A. Munck, F. Ratjen, I. Sermet-Gaudelus, K. W. Southern, J. Barben, P. A. Flume, P. Hodková, N. Kashirskaya, M. N. Kirszenbaum, S. Madge, H. Oxley, B. Plant, S. J. Schwarzenberg, A. R. Smyth, G. Taccetti, T. O. F. Wagner, S. P. Wolfe, P. Drevinek, ECFS best practice guidelines: the 2018 revision. *J. Cyst. Fibros.* **17**, 153–178 (2018).
52. K. De Boeck, N. Derichs, I. Fajac, H. R. de Jonge, I. Bronsveld, I. Sermet, F. Vermeulen, D. N. Sheppard, H. Cuppens, M. Hug, P. Melotti, P. G. Middleton, M. Wilschanski, New clinical diagnostic procedures for cystic fibrosis in Europe. *J. Cyst. Fibros.* **10 Suppl 2** (2011), doi:10.1016/S1569-1993(11)60009-X.
53. P. M. Farrell, T. B. White, C. L. Ren, S. E. Hempstead, F. Accurso, N. Derichs, M. Howenstine, S. A. McColley, M. Rock, M. Rosenfeld, I. Sermet-Gaudelus, K. W. Southern, B. C. Marshall, P. R. Sosnay, Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. *J. Pediatr.* **181S**, S4-S15.e1 (2017).
54. P. D. Wagner, The physiological basis of pulmonary gas exchange: implications for clinical interpretation of arterial blood gases. *Eur. Respir. J.* **45**, 227–243 (2015).
55. M. R. Miller, Defining airflow obstruction. *Eur. Respir. J.* **45**, 560 (2015).
56. P. H. Quanjer, S. Stanojevic, T. J. Cole, X. Baur, G. L. Hall, B. H. Culver, P. L. Enright, J. L. Hankinson, M. S. M. Ip, J. Zheng, J. Stocks, C. Schindler, Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur. Respir. J.* **40**, 1324–1343 (2012).
57. Permutation, Parametric and Bootstrap Tests of Hypotheses. *Permut. Parametr. Bootstrap Tests Hypotheses* (2005), doi:10.1007/B138696.
58. S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, I. Thiele, Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2016).
59. D. Machado, S. Andrejev, M. Tramontano, K. R. Patil, Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
60. D. W. Hosmer, S. Lemeshow, R. X. Sturdivant, Applied Logistic Regression: Third Edition. *Appl. Logist. Regres. Third Ed.*, 1–510 (2013).
61. H. Zou, The Adaptive Lasso and Its Oracle Properties. *https://doi.org/10.1198/016214506000000735*. **101**, 1418–1429 (2012).
62. J. D. Lee, D. L. Sun, Y. Sun, J. E. Taylor, Exact post-selection inference, with application to the lasso. *Ann. Stat.* **44**, 907–927 (2013).

Figure captions

Figure 1: General reconstruction workflow and *A. fumigatus* pan-genome-scale metabolic model (GEM) statistics. (**A)** Workflow towards *A. fumigatus* strain-specific GEM reconstructions. Colors indicate different strains and associated metabolic models. (**B-F)** Characteristics of pan-GEM reconstruction for *A. fumigatus*. **(B)** Contribution of macromolecules in comprising one unit of biomass (Supplementary Table S1). **(C)** Distribution of pan-GEM reactions across major pathway categories (Supplementary Table S7). **(D)** Distribution of pan-GEM reactions across nine compartments (Supplementary Table S7). **(E)** Growth prediction accuracy of pan-GEM for *A. fumigatus* wild-type (Af293) and five mutant strains using phenotypic microarray data. Growth accuracy for phenotypic microarrays on sulfur (S) are also indicated for neglecting *ΔniaD* and *ΔlysF* mutants (see Results for details, Supplementary Table S2). C: carbon, N: nitrogen, P: phosphor, S: sulfur. **(F)** Confusion matrix of pan-GEM accuracy in predicting the essentiality of 20 genes according to the literature (see Results and Methods).

Figure 2: Core and accessory metabolic capabilities of all *A. fumigatus* strain-specific genome-scale metabolic models (GEMs). The core and accessory metabolic content was determined for 252 unique *A. fumigatus* strains with environmental and clinical origin. (A) Summary of the core and accessory reactome across higher level metabolic pathway categories. Pathway categories are according to the KEGG pathway definition (https://www.kegg.jp/kegg/pathway.html). **(B)** The distribution of the accessory reactome across all isolate models. Indicated percentage ranges correspond to accessory reaction presence across all strain-GEMs. **(C)** Heatmap with pairwise Jaccard distance values for isolate GEM pairs based on presence or absence of metabolic reactions. **(D)** Fisher-test based most statistically significant reactions enriched in the indicated isolate subsets. Presence frequency indicates the fraction of reaction presence over all investigated models (color associates to sample origin). **(E)** Decision tree optimized towards showing best separation into clinical and environmental isolate origin. The decision tree is based on absence/presence of metabolic reactions and growth capability on different nutrients across all isolate GEMs. **(F)** Machine learning mean AUC performance based on FVA derived flux ranges for all reactions (objective function: biomass) of strain models with clinical vs. environmental origin.

Figure 3: Metagenomics sequencing of 80 paired sputum samples from cystic fibrosis (N=40) patients before (*A. fumigatus* -) and after *A. fumigatus* infection (*A. fumigatus* +). (A) Relative abundances of the top 10 genera and species over all 80 samples. X axis is ordered by patient sample. **(B)** Differential correlation analysis of species in *A. fumigatus*+ relative to *A. fumigatus*- to reveal changes in the interactome of the lung microbiome upon *A. fumigatus* colonization. Edge colors and different class information indicate direction of correlation in *A. fumigatus*-/*A. fumigatus*+. The associated count indicates the number of species pairs in the network exhibiting this pattern of change. Only species pairs with significant differential correlations were included (permutation test, $p<0.05$), Species with orange background indicate significant differentially abundant species between *A. fumigatus*- vs *A. fumigatus*+ samples (metagenomeSeq, zero inflated gaussian mixture model, $p<0.05$). Labels: *A. fumigatus*-: samples before infection; *A. fumigatus*+: samples after infection.

Figure 4: Statistics for MAMBO-derived metabolite profiles in cystic fibrosis patient samples. (A) Beta diversity (Euclidean distance) of MAMBO derived media. PERMANOVA was used to assess the statistical significance of beta diversity comparisons. **(B)** Growth rate differences of genome-scale metabolic models corresponding to clinical *A. fumigatus* strains based on MAMBO derived media compositions associated to cystic fibrosis samples before and after *A. fumigatus* infection. **(C)** Significantly different **f**lux ranges (either in lower or upper bound) of clinical strain models simulated with FVA on MAMBO derived media before and after *A. fumigatus* infection. Significance was tested according to paired Wilcoxon signed rank test and adjusted by FDR. **(D)** Three selected enzymatic reactions with significant flux bound differences in either lower or upper bound as displayed in **(C)**. Both bounds are indicated. EC1.4.3.4: Tryptamine:oxygen oxidoreductase; EC4.1.3.27: Chorismate pyruvate-lyase; EC1.8.1.9: NADPH:oxidized-thioredoxin oxidoreductase**.** *A.f.*-/+: *A. fumigatus* positive and negative samples, respectively.

Tables

Table 1: Number of accessory genes and reactions across all isolate GEMs. The occurrence column refers to the number of reactions occurring in a certain fraction range of isolate specific GEMs based on 338 accessory reactions in total.

	Genes		Reactions	
Occurrence (in %)	Mean	Standard deviation	Mean	Standard deviation
[1,20]	1.9	1.8	4.7	7.7
[21,40]	6.9	2.5	1.3	1.2
[41,60]	16.7	3.2	17.4	6.0
[61,80]	17.8	2.4	20.2	6.3
[81,99]	296.9	8.1	219.3	6.8

Supplementary material

Supplementary Table S1 – Biomass composition

Supplementary Table S2 – Phenotypic growth information

Supplementary Table S3 – *A. fumigatus* strain metadata

Supplementary Table S4 – ML model using minimal media

Supplementary Table S5 – Metagenomics

Supplementary Table S6 – Metabolite analysis

Supplementary Table S7 – Detailed GEM information

Supplementary Figure S1 – Accessory genome conservation among 252 Aspergillus fumigatus strains. Indicated percentage ranges denote accessory gene presence across the genomes of all strain-GEMs. Cluster with gray background denotes genetic lineage of *A. fumigatus* with significantly fewer accessory genes than other lineages as previously published (6). Associates to Fig. 2B.

Supplementary Figure S2 – Alpha and beta diversity of CF lung microbiome. **(A)** Statistical significance according to Shannon and Chao diversity index for alpha diversity). **(B)** Beta diversity (Aitchison distance) based on Kraken derived taxonomic profiles. Wilcoxon signed rank test was used for alpha diversity comparisons; PERMANOVA was used to assess the statistical significance of beta diversity comparisons.

Figure 1

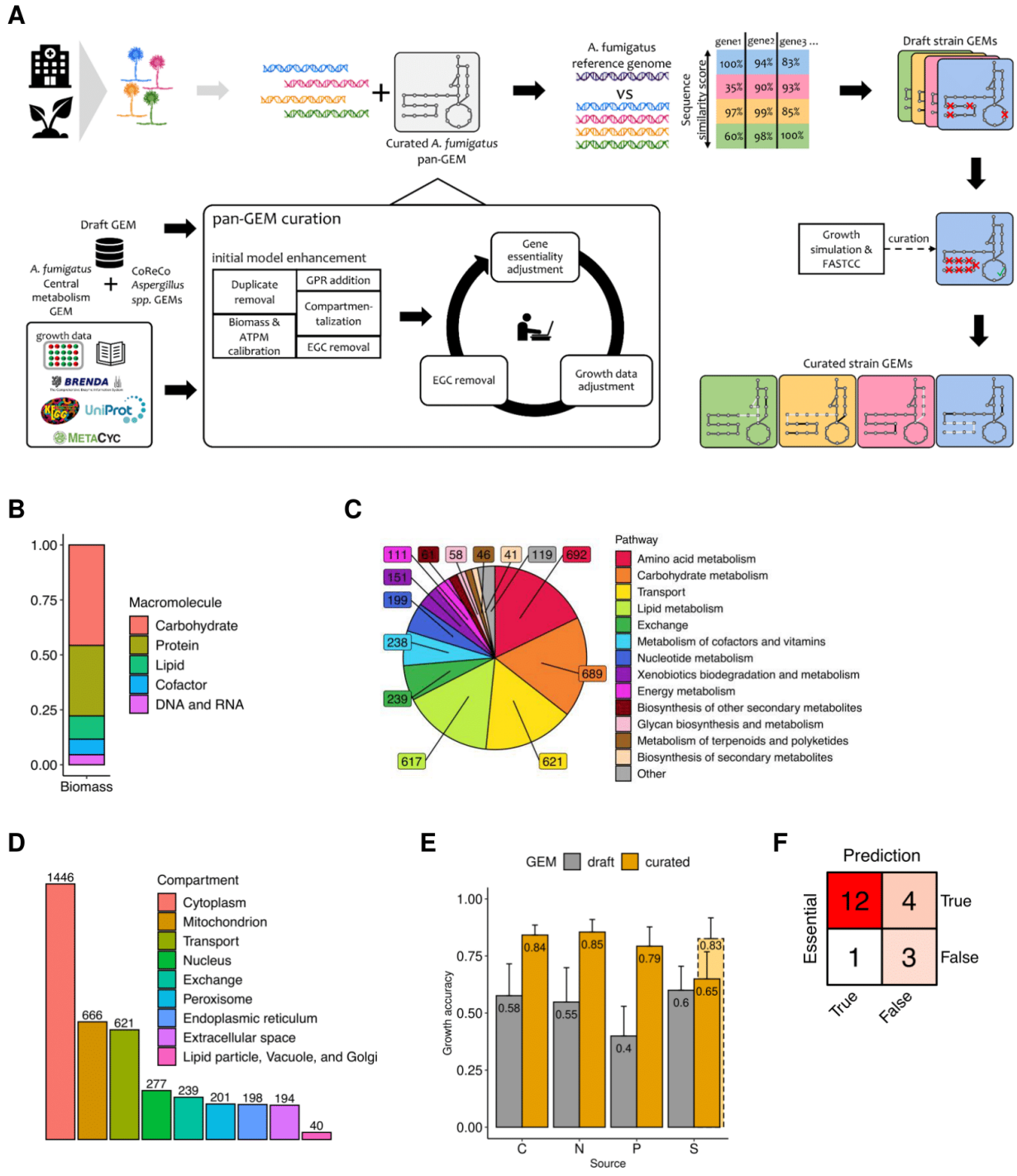

A
A. fumigatus reference genome
VS
Curated A. fumigatus pan-GEM
Sequence similarity score
gene1 gene2 gene3 ...
100% 94% 83%
35% 90% 93%
97% 99% 85%
60% 98% 100%
Draft strain GEMs
Draft GEM
A. fumigatus Central metabolism GEM
CoReCo Aspergillus spp. GEMs
growth data
BRENDA
KEGG
UniProt
METACYC
pan-GEM curation
initial model enhancement
Duplicate removal
GPR addition
Biomass & ATPM calibration
Compartmen-talization
EGC removal
Gene essentiality adjustment
Growth data adjustment
EGC removal
Growth simulation & FASTCC
curation
Curated strain GEMs
B
1.00
0.75
0.50
0.25
0.00
Biomass
Macromolecule
Carbohydrate
Protein
Lipid
Cofactor
DNA and RNA
C
111 61 58 46 41 119 692
151
199
238
239
617
621
689
Pathway
Amino acid metabolism
Carbohydrate metabolism
Transport
Lipid metabolism
Exchange
Metabolism of cofactors and vitamins
Nucleotide metabolism
Xenobiotics biodegradation and metabolism
Energy metabolism
Biosynthesis of other secondary metabolites
Glycan biosynthesis and metabolism
Metabolism of terpenoids and polyketides
Biosynthesis of secondary metabolites
Other
D
1446
666
621
277
239
201
198
194
40
Compartment
Cytoplasm
Mitochondrion
Transport
Nucleus
Exchange
Peroxisome
Endoplasmic reticulum
Extracellular space
Lipid particle, Vacuole, and Golgi
E
GEM draft curated
Growth accuracy
1.00
0.75
0.50
0.25
0.00
0.58 0.84
0.55 0.85
0.4 0.79
0.6 0.65 0.83
C N P S
Source
F
Prediction
Essential
12 4
1 3
True
False
True
False

Figure 2

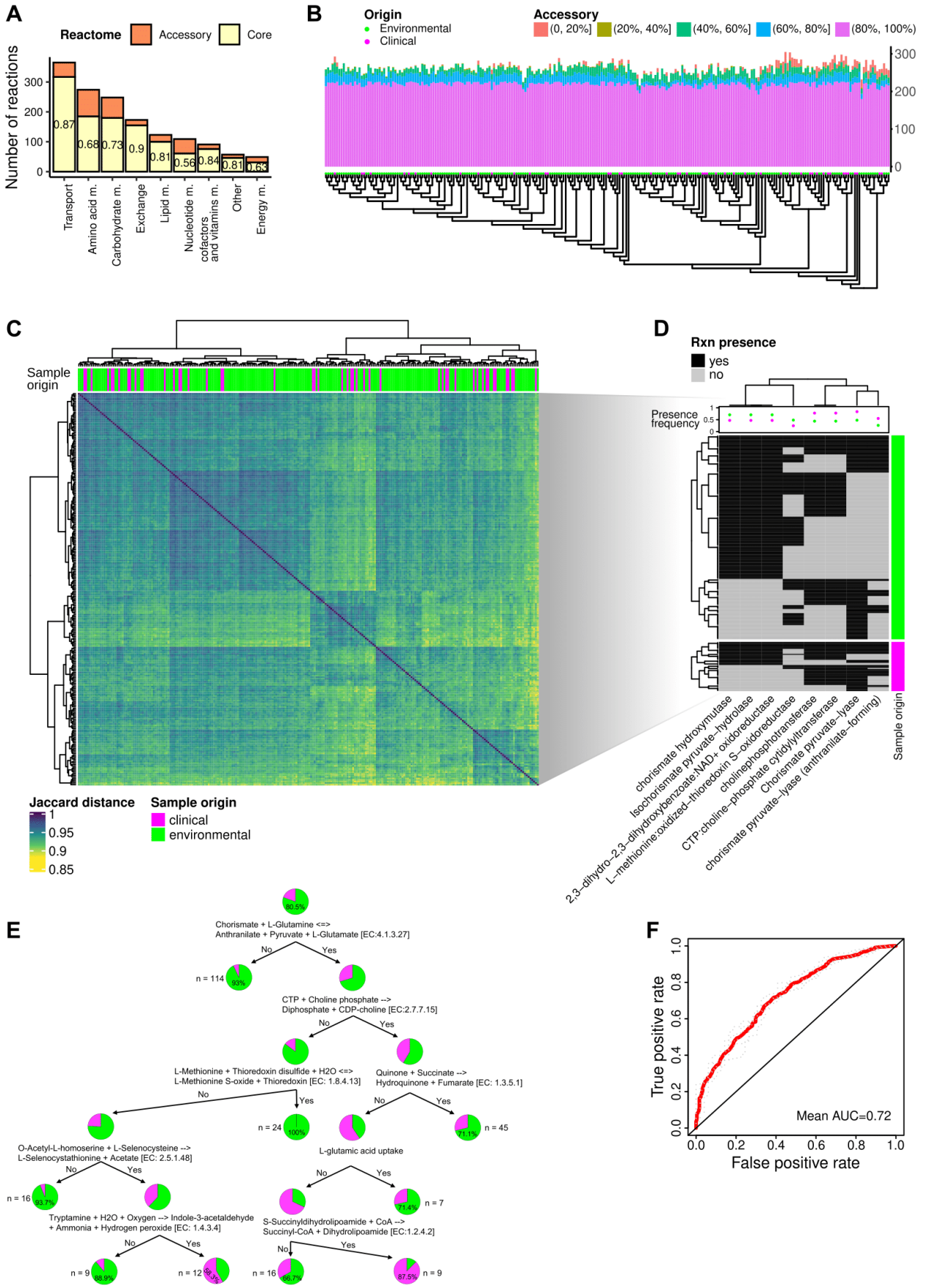

A
Reactome
Accessory
Core
Number of reactions
Transport
Amino acid m.
Carbohydrate m.
Exchange
Lipid m.
Nucleotide m.
cofactors
and vitamins m.
Other
Energy m.
B
Origin
Environmental
Clinical
Accessory
(0, 20%]
(20%, 40%]
(40%, 60%]
(60%, 80%]
(80%, 100%)
C
Sample origin
Jaccard distance
Sample origin
clinical
environmental
D
Rxn presence
yes
no
Presence frequency
chorismate hydroxymutase
Isochorismate pyruvate-hydrolase
2,3-dihydro-2,3-dihydroxybenzoate:NAD+ oxidoreductase
L-methionine:oxidized-thioredoxin S-oxidoreductase
cholinephosphotransferase
CTP:choline-phosphate cytidylyltransferase
Chorismate pyruvate-lyase
chorismate pyruvate-lyase (anthranilate-forming)
Sample origin
E
Chorismate + L-Glutamine <=>
Anthranilate + Pyruvate + L-Glutamate [EC:4.1.3.27]
CTP + Choline phosphate -->
Diphosphate + CDP-choline [EC:2.7.7.15]
L-Methionine + Thioredoxin disulfide + H2O <=>
L-Methionine S-oxide + Thioredoxin [EC: 1.8.4.13]
Quinone + Succinate -->
Hydroquinone + Fumarate [EC: 1.3.5.1]
O-Acetyl-L-homoserine + L-Selenocysteine -->
L-Selenocystathionine + Acetate [EC: 2.5.1.48]
L-glutamic acid uptake
Tryptamine + H2O + Oxygen --> Indole-3-acetaldehyde
+ Ammonia + Hydrogen peroxide [EC: 1.4.3.4]
S-Succinyldihydrolipoamide + CoA -->
Succinyl-CoA + Dihydrolipoamide [EC:1.2.4.2]
F
True positive rate
False positive rate
Mean AUC=0.72

Figure 3

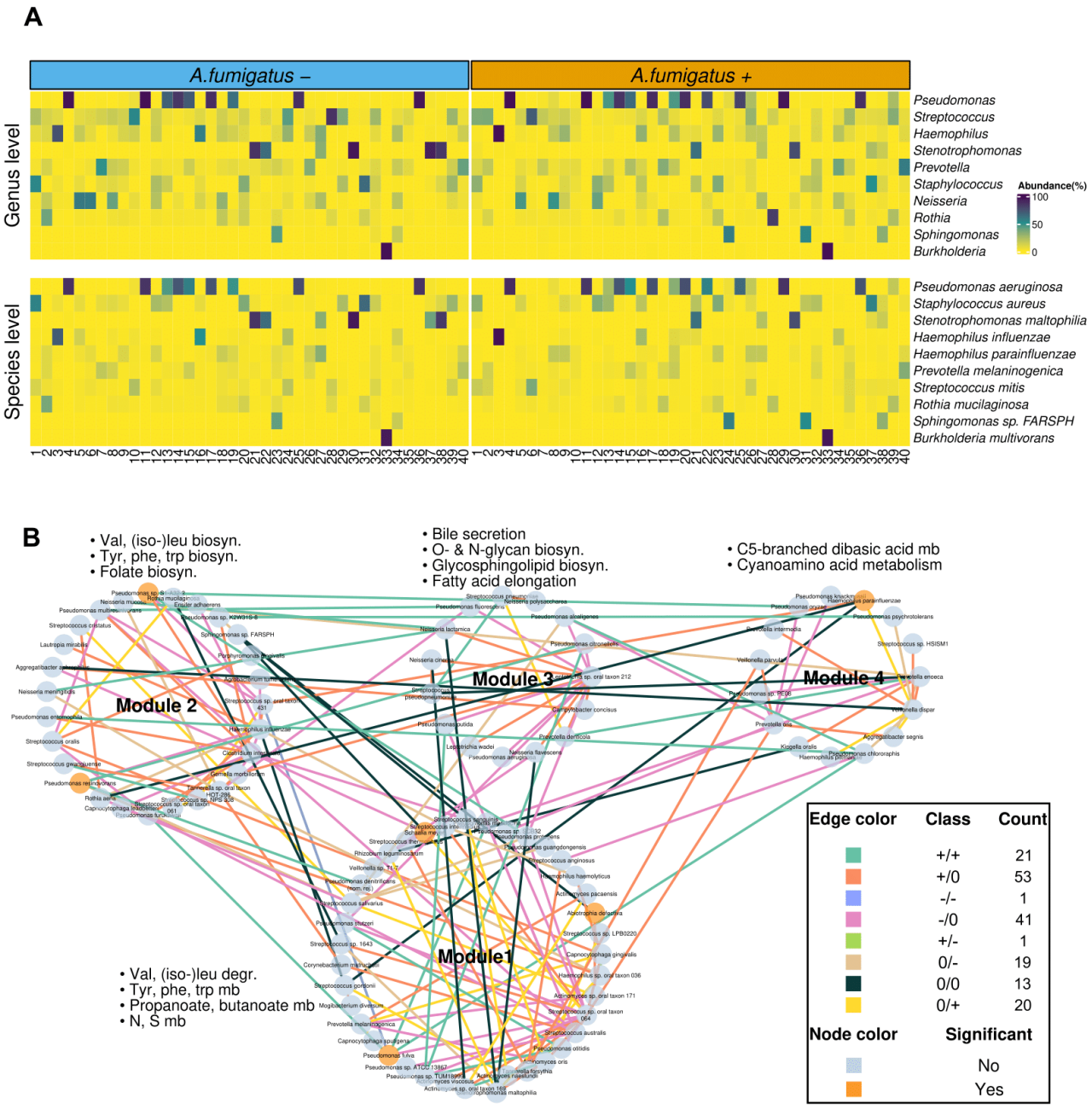

A
A.fumigatus –
A.fumigatus +
Genus level
Pseudomonas
Streptococcus
Haemophilus
Stenotrophomonas
Prevotella
Staphylococcus
Neisseria
Rothia
Sphingomonas
Burkholderia
Abundance(%)
100
50
0
Species level
Pseudomonas aeruginosa
Staphylococcus aureus
Stenotrophomonas maltophilia
Haemophilus influenzae
Haemophilus parainfluenzae
Prevotella melaninogenica
Streptococcus mitis
Rothia mucilaginosa
Sphingomonas sp. FARSPH
Burkholderia multivorans
B
• Val, (iso-)leu biosyn.
• Tyr, phe, trp biosyn.
• Folate biosyn.
• Bile secretion
• O- & N-glycan biosyn.
• Glycosphingolipid biosyn.
• Fatty acid elongation
• C5-branched dibasic acid mb
• Cyanoamino acid metabolism
Module 2
Module 3
Module 4
Module1
• Val, (iso-)leu degr.
• Tyr, phe, trp mb
• Propanoate, butanoate mb
• N, S mb
Edge color
Class
Count
+/+ 21
+/0 53
-/- 1
-/0 41
+/- 1
0/- 19
0/0 13
0/+ 20
Node color
Significant
No
Yes

Figure 4

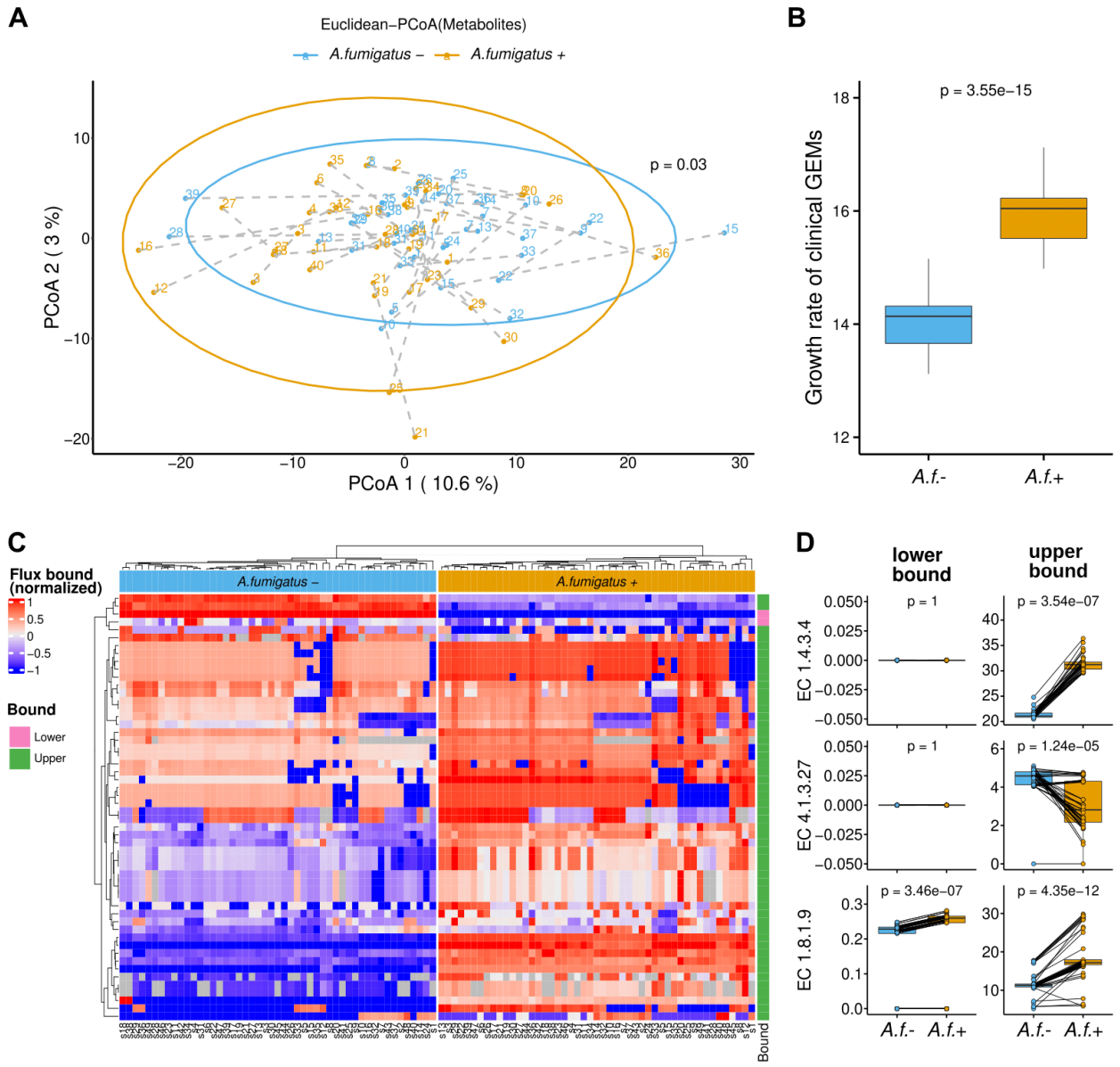

A
Euclidean−PCoA(Metabolites)
A.fumigatus −
A.fumigatus +
p = 0.03
PCoA 2 (3 %)
PCoA 1 (10.6 %)
B
p = 3.55e−15
Growth rate of clinical GEMs
A.f.-
A.f.+
C
Flux bound (normalized)
Bound
Lower
Upper
A.fumigatus −
A.fumigatus +
Bound
D
lower bound
upper bound
EC 1.4.3.4
EC 4.1.3.27
EC 1.8.1.9
p = 1
p = 3.54e−07
p = 1
p = 1.24e−05
p = 3.46e−07
p = 4.35e−12
A.f.-
A.f.+

Figure S1

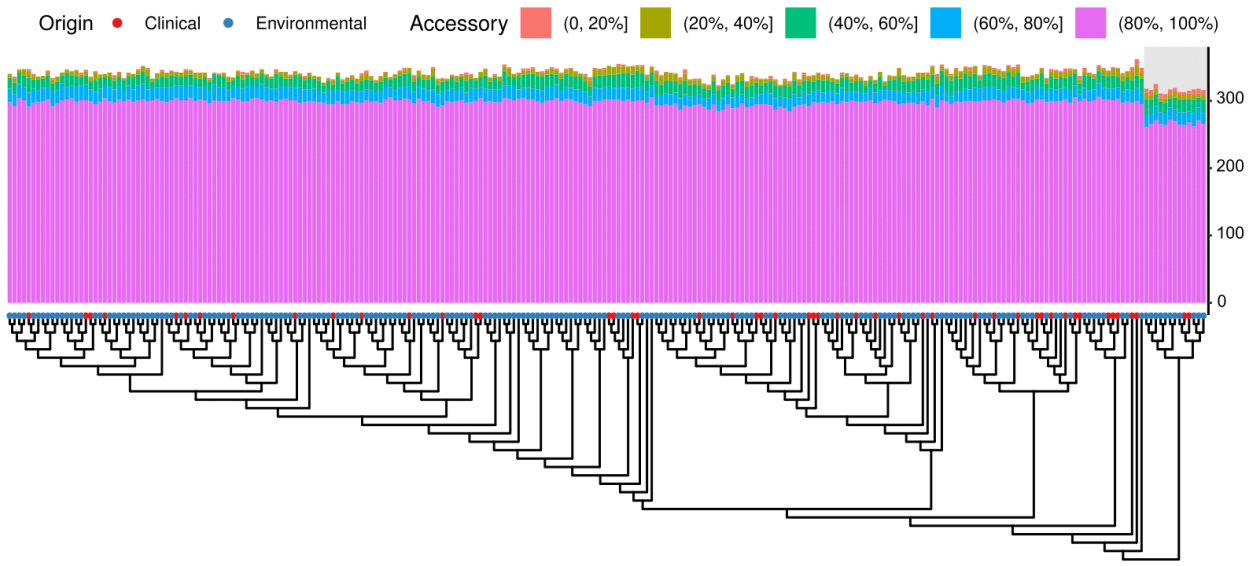
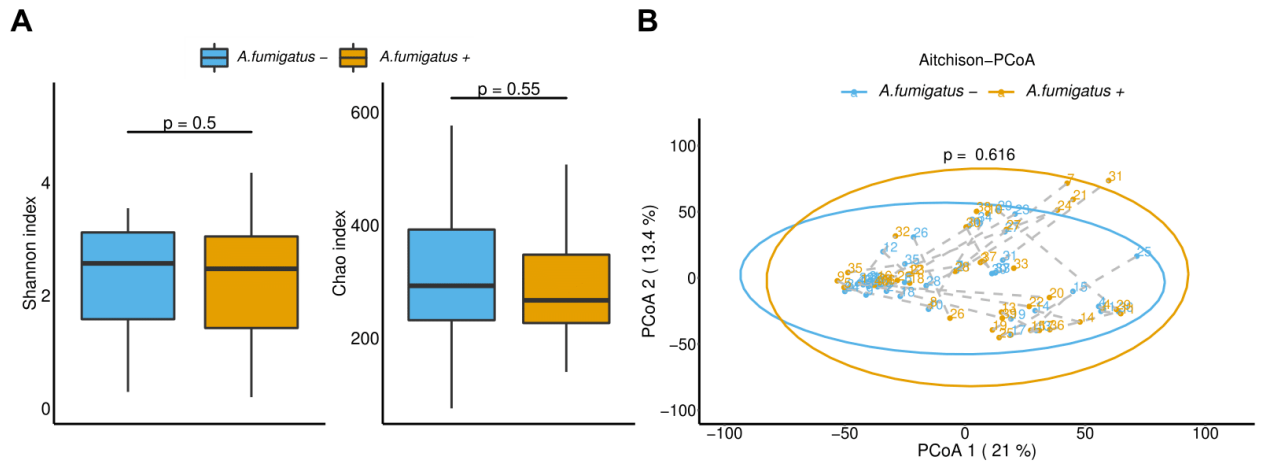

Origin
Clinical
Environmental
Accessory
(0, 20%]
(20%, 40%]
(40%, 60%]
(60%, 80%]
(80%, 100%)
300
200
100
0

Figure S2


A
A.fumigatus −
A.fumigatus +
p = 0.5
Shannon index
0
2
4
p = 0.55
Chao index
200
400
600
B
Aitchison−PCoA
A.fumigatus −
A.fumigatus +
p = 0.616
PCoA 2 (13.4 %)
−100
−50
0
50
100
PCoA 1 (21 %)

DISCUSSION

Human fungal pathogens cause a serious problem in medical health care. The increasing number of immunocompromised patients and high mortality rate indicates an underestimation of the problem by the medical health authorities. The opportunistic pathogen *Candida albicans* and the airborne fungus *Aspergillus fumigatus* are the most prevalent fungal pathogens causing infectious diseases. This cumulative dissertation includes six manuscripts (Table 1) applying computational systems biology and bioinformatics approaches. They indicated the advantage of using computational approaches to study the phenotypes involved in infection and revealed the role of metabolism in developing the infection by *C. albicans* and *A. fumigatus*. As a promising approach to predict the phenotypes of fungal pathogens in the microbial community and their metabolic interactions, GEMs appeared vital. Therefore, major efforts went into the reconstruction and analysis of fungal GEMs around *C. albicans* and *A. fumigatus*. The manuscripts belong to two topics: (i) role of computational systems biology for studying the phenotypes involved in infection, (ii) metabolic systems biology of fungal infection.

The first topic was addressed by two publications representing two case studies where the applicability of computational systems biology approaches to uncover phenotypes involved in infection was shown. The first one (Table 1) compares the host response against Rhinovirus and influenza through a systematic transcriptomics analysis. Gene set enrichment analysis was used to identify perturbed pathways by the viral infection using identified differentially expressed genes as input. The study revealed that Rhinovirus induced a more delayed and blunted host response than influenza virus infection. The network pathway enrichment analysis showed that steroid-related pathways are specifically enriched for Rhinovirus. Interferon response and other innate immune response predominated for both Rhinovirus and influenza virus infection.

The prerequisite for performing the gene set enrichment analysis is generation of transcriptomics data which may not truthfully reflect the metabolic activity as it does not necessarily correlate with protein concentration and enzyme activity (Maier et al., 2009). On the contrary, genome-scale metabolic modeling predicts the metabolic phenotype of the target organism as the endpoint of biological processes in a quantitative manner and without any requirement for producing large-scale transcriptomics data. Hence, GEM analysis is a way powerful approach than gene set enrichment analysis to reflect the metabolic behavior of the cell. For the future studies, a joint GEM of human and virus can be reconstructed and analyzed to obtain in-depth insights into the phenotypes involved in infection. It can be further enhanced by preparing metabolomics data and integrating it into the GEM analysis. Such

analysis can assist to identify anti-viral targets as recently shown for the SARS-CoV-2 (Renz et al., 2021).

Table 2. Overview of the manuscripts.

	Title	Journal	Status	Topic
1	Comparative Transcriptomic Analysis of Rhinovirus and Influenza Virus Infection (co-author)	Frontiers in Microbiology	Published	i
2	An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness (co-author)	Gut Microbes	Published	i
3	Metabolic modeling predicts specific gut bacteria as key determinants for *Candida albicans* colonization levels (first author)	The ISME Journal	Published	ii
4	*Candida albicans* SR-like protein kinases regulate different cellular processes: Sky1 is involved in control of ion homeostasis, while Sky2 is important for dipeptide utilization (co-author)	Frontiers in Cellular and Infection Microbiology	Accepted for publication	ii
5	*Lactobacillus rhamnosus* colonisation antagonizes *Candida albicans* by forcing metabolic adaptations that compromise pathogenicity (co-author)	Nature Communications	Revised version submitted	ii
6	*Candida* expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions (co-author)		In preparation	ii
7	A pan-genome resembling genome-scale metabolic platform of 252 *Aspergillus fumigatus* strains reveals growth dependencies from the lung microbiome (first author)		In preparation	ii

The second publication (Table 1) addresses the applicability of the GEM analysis by investigating the impact of antibiotic consumption on the metabolic and compositional state of the microbiome of ICU patients. It takes advantage of genome-scale metabolic modeling to confirm the findings from the metagenomics and metabolomics analysis. According to what statistical and correlation analysis suggested, bacterial species seemingly responsible for SCFA production were simulated using flux variability analysis (FVA). The predictions confirmed the contribution of bacterial species such as *Alistipes putredinis* to SCFA production. In conclusion, the study suggested that the loss of a handful of species, that are highly connected with the production of SCFAs and BAs, during antibiotic administration in the ICU allows the expansion of pathogenic species, which exhibit the potential to cause direct hindrance to host homeostasis.

In the FVA analysis, it was assumed that each bacterial species is growing independently of the metabolic effect of other bacterial species. In order to gain a more realistic flux analysis, a universal GEM out of the microbial community members can be

generated (Chan et al., 2017). This incorporates relative abundance values of bacterial species in each sample by which each bacterial species is enforced to grow proportionally and takes into account the possible crosstalk among microbiome members.

The diet fed to the ICU patients was retrieved and used in the GEM analysis. The additional compounds potentially secreted from the host were identified by FVA using the Recon 3D GEM (Brunk et al., 2018). Yet, this GEM is a generic reconstruction of human metabolism. Generating a context-specific GEM for the relevant human cells such as Intestinal epithelial cells can enhance the predictions for identifying the potential secretions by the host. Furthermore, selecting a proper objective function for the human cells is a challenge that needs to be addressed by an in-depth investigation of the cell phenotypes in accordance with the specific biological condition of the human cells in the study. For instance, the generation of metabolomics data in *in vitro* cultured human cells can reveal metabolic secretions that are potentially be considered an objective function.

Given the applicability of GEM analysis in microbiome studies, the second topic discusses metabolic systems biology of fungal infection by developing and applying genome-scale metabolic models of fungal pathogens, *C. albicans*, and *A. fumigatus*. The second topic involves five manuscripts (Table 1) addressing promoting and inhibiting factors on the growth of the aforementioned fungal pathogens from different aspects. To this aim, it emphasizes generating and analyzing GEMs and phenotypic microarray data analysis as an additional layer of analysis. It allowed identifying functional interactions in the human gut impacting the colonization of *C. albicans* (third manuscript), uncovering functions of protein kinases having a critical role in colonizing the *C. albicans* (fourth manuscript), revealing the role of Intestinal epithelial cells as the host effect on a *C. albicans* – *Lactobacillus rhamnosus* interaction (fifth manuscript), and, as a complementary analysis, revealing ecological signature of *Candida* expansion (sixth manuscript). Lastly, in the seventh manuscript, the phenotypic diversity of 252 *A. fumigatus* strains isolated from hospitals and agricultural lands in Germany was investigated. In this study, 252 genome-scale metabolic reconstructions were generated and analyzed to address the diversity in the metabolic capability of *A. fumigatus* strains. The clinical GEMs were further analyzed to explain how the fungus contributes to the shaping of the lung microbiome in a favorable manner to colonize better in the lung.

The next section discusses the main results and outlines perspectives on experimental validation, clinical relevance, and future model development in metabolic systems biology of fungal infection.

Metabolic systems biology of fungal infection

The opportunistic fungal pathogen, *C. albicans*, can overgrow in the human gut and develop infection upon given opportunity. From this point of view, it is crucial to identify functional interactions of the fungus with the bacterial community leading to inhibition or promotion of growth. In the third manuscript (Table 1), a GEM for the *C. albicans* was generated, validated, and used to simulate the interaction of the fungus with individual bacterial species in case of GEM availability in the literature. In total, simulation of over 900 pairwise interactions suggested bacterial species impacting *C. albicans* growth positively, negatively, or neutrally in three different media. Selected metabolite experiments and shotgun metagenomics sequencing supported the *in-silico* modeling concept based on pairwise metabolic interaction simulations.

Further studies of the predictions for major metabolic pathways, for example for carbon compounds and amino acids, may elucidate the specific mechanisms of these influences. Selected metabolite measurements in defined media could be used further for accurate predictions of potential metabolite candidates that are preferentially used by e.g., gut microbes over *C. albicans* and can hint towards bacterial species that specifically secrete *C. albicans*-inhibiting metabolites. Taken together, the findings of this publication support that specific gut bacteria influence gut colonization by *C. albicans*. Moreover, the analysis indicates that it may be possible to design synthetic communities with only a few bacterial species that could then influence essential metabolic activities of *C. albicans* and prevent fungal overgrowth.

The *C. albicans* GEM generation started from a draft reconstruction generated by an automated comparative pipeline called CoReCo for constructing the gapless metabolic networks for fungal species (Pitkänen et al., 2014). This comparative approach is particularly useful in scenarios where the quality of available sequence data is lacking. The pipeline employs a greedy algorithm to identify the missing reactions that can connect the nutrients to the end products, fulfilling the connectivity of carbon atoms in the network. Evaluating the draft reconstruction using the phenotypic microarray data demonstrated the high ability of the GEM to resemble true positive cases of growth prediction for consuming the carbon sources by the wild-type *C. albicans*. However, further assessment of the metabolic network revealed that this greedy gap-filling algorithm derives false positive cases for growth predictions of the GEM using carbon sources. On the other hand, relatively fewer true positives were identified for the non-carbon nutrient sources as the network connectivity through other atoms such as

nitrogen is overlooked in the original publication. Also, overestimation of reaction presence in the network resulted in Erroneous Energy-Generating Cycles (EGCs). The quality of the reconstruction for predicting the phenotypes was considerably increased by resolving such artifacts in the network.

Phenotypic microarray data played an important role in increasing the quality of the metabolic network by testing and correcting through classical iterative systems biology cycle for the GEM development. The data obtained from wild-type and mutant strains was analyzed to identify growth/no-growth conditions. Given the applicability of this technology, it was also used to reveal the phenotypes of Sky1 and Sky2 protein kinases of *C. albicans* in the fourth manuscript (Table 1). Adequate data analysis of phenotypic microarray data revealed the functions of Sky1 and Sky2 protein kinases. For future studies, metabolomics data of the two knocked-out genes can be generated to obtain specific metabolic cues linked to their functionality. Integration of metabolomics data into the *C. albicans* GEM can add another layer to uncovering the detailed phenotypes of the genes under different growth conditions.

The predictions by the current *C. albicans* GEM can be further improved by further refinements in order to generate the second version. For instance, allocating the metabolites and reactions in separate compartments (subcellular locations) can avoid unrealistic interconversion of metabolites by reactions supposed not to occur in the same compartment. Besides, a comprehensive single-knockout study can show the accuracy of the GEM for gene essentiality predictions and accordingly be used to eliminate errors from it. The biomass reaction in this model was adapted from the iMM904 GEM (Mo et al., 2009) of *Saccharomyces cerevisiae*. This adaptation is acceptable due to the closeness of two species in the phylogenetic tree of the fungi kingdom (Choi & Kim, 2017). Yet, adjusting the biomass reaction and corresponding pathways of biomass precursor production based on experimentally measured biomass composition of the fungus can considerably enhance the accuracy of predictions. In particular, it would be more accurate to include specific biomass reactions for different morphologies of the fungus to predict its metabolic behavior while developing in different formats such as yeast, pseudohyphae, and hyphae cells. This is especially important for pairwise interactions. They mimic the gut environment possessing a specific temperature and pH value which are the key determinants of *C. albicans* morphology during growth and development (Berman, 2006).

The current GEM was built based on available genome information in the databases which mainly belong to the SC5314 wild-type strain. Yet, the generation of specific GEMs at

strain resolution can capture the phenotypic diversity of *C. albicans* leading to a better understanding of fungal infection development through metabolism.

Additional diets and growth media compositions beyond the three used in this study may be tested, since e.g., GAM compositions may vary, therefore influence in silico predictions and in general, might not reflect in vivo conditions as well as high fiber, western, or other common human diets. Pairwise simulations showed different interaction distribution when using different diet compositions, e.g., GAM media vs. western or high fiber diets. This indicates that diet intervention might shape the microbial composition beneficial or detrimental for the infection development. A systematic study of diet effect on microbial composition may open the doors for designing personalized diets that can hamper predisposition to infection by the fungus. To this end, computational approaches incorporating GEMs, e.g., the recently published CODY framework (Geng et al., 2021), can be applied. Besides, additional *in silico* simulations of multiple interactions need to be addressed, to potentially extend our understanding of the intricate relationship between *C. albicans* and the gut microbiota and its effect on *C. albicans* levels (This is of major interest for *A. fumigatus* addressed later). Finally, tools that incorporate spatial information (Bauer et al., 2017; Biggs & Papin, 2013) could determine the impact of niche colonization by gut fungal and bacterial species.

In order to avoid biologically implausible solutions, coupling constraints were applied to pairwise simulations. They associate the fluxes through network reactions of one model to its biomass formation flux, inhibiting the benefit of one microbe from the paired one without producing any biomass. Thus, the biomass formations in both GEMs become dependent on each other. A Pareto optimality analysis can illustrate this dependency, in which one biomass formation value is fixed at different values and the other one is minimized and maximized to obtain the possible flux interval, and vice versa. This analysis can give an in-depth insight into the trade-off between two microbes. In this project, it was assumed that maximum summation of the growth rates is achieved when two microbes are paired.

This publication specifically focused on the intricate relationship between (gut) bacteria and *C. albicans* to elucidate their relationship independent of host factors in order to keep free parameters in a feasible range. Host factors are key modulators of fungal–bacterial or fungal–bacterial–host interactions as could be shown in other studies (Fan et al., 2015; Lakhani et al., 2017). Though adding considerable complexity to the setup, an extension of this conceptual approach to study specifically metabolic modes of fungal–bacterial interaction with host factors might play an important complementary role as long as the predictive

capacity can be supported by sufficient data around the triangle host, bacteria and fungus. To this end, the triangle of intestinal epithelial cells (host), *C. albicans*, and *Lactobacillus rhamnosus* as a special antagonizing case against the fungus was studied in the fifth manuscript (Table 1).

In this study, metabolic and molecular aspects of *L. rhamnosus*-mediated protection against *C. albicans* pathogenicity supported by omics data were investigated. Alternations of the metabolic environment due to the presence of *L. rhamnosus* forced *C. albicans* to transcriptionally reprogram its metabolism. Transcriptome and *in silico* analysis revealed changes in key metabolic pathways and regulator genes of *C. albicans*. In summary, the combined metabolome data, *in silico* metabolic modeling, transcriptome and mutant screening, as well as *in vitro* validations, provide fundamental contextualized insights into how *C. albicans* pathogenicity can be controlled or prevented.

Availability of the metabolomics of the shared media among the cells at different time points facilitated the tuning of the GEMs. In fact, it simplified the *in silico* simulations without further demand for performing pairwise or triplewise simulations. The network flux ranges obtained from metabolome-calibrated GEMs through flux variability analysis (FVA) predicted the metabolic pathways regulations when three species are interacting with each other. Maximization of growth rate was the objective function for each GEM. This reflects the flux changes towards maximization of the biomass which might not be necessarily the objective of the cells over time due to the intricate relationship among them. Hence, an in-depth investigation for selecting the proper objective function individually for each cell can be done for future studies.

The human Recon 3D GEM (Brunk et al., 2018) was used for simulating the metabolic behavior of the human epithelial cells. This is a generalized reconstruction containing all possible metabolic genes and reactions in humans. A contextualized GEM specific for epithelial cells can be made to increase the accuracy of the predictions. To this end, the challenge is to gain knowledge about the reactions certainly occurring in the desired tissue, epithelial cells in this case, or the availability of consistent omics data in replicates (Cho et al., 2019). The tissue-specific network can considerably increase the accuracy of the predictions through the GEM analysis.

In summary, the strategy of studying fungal–bacterial relationships in the gut using an *in silico*, a metabolism-driven approach already yielded promising results. The use of *in silico* predictions added another useful layer that can contribute to stratifying the identification of potentially clinically relevant gut bacteria in face of the steadily growing amount of high

throughput data. Ultimately, including metabolic *in silico* analysis could promote additional systems-biology and systems-medicine studies that focus on fungal infections and their often lethal implications to humans.

In the sixth manuscript, the overgrowth of Candida species was explained by integrating mycobiome and shotgun metagenomics data from 75 lung cancer patients. It demonstrated the critical role of gut bacteria in shaping microbiome composition. The adequate metagenomics and statistical analysis divided the samples into Low- and High-Candida groups. Many bacterial SCFA producers were found in the Low-Candida group. As a complementary analysis to confirm the findings, GEM analysis was performed to inspect whether the bacterial species enriched in the Low-Candida group can produce SCFAs.

To this end, FVA of exchange reactions for SCFAs such as acetic acid and propionic acid was performed alongside maximization of growth under 12 different diets. The same analysis investigated whether the bacterial species enriched in the High-Candida group can produce lactate. The analysis was done in microaerobic conditions to mimic the gut environment. In the simulations, oxygen was allowed to be consumed not more than one mmol/gDW/hour. As it is an arbitrary choice, a comprehensive sensitivity analysis can be performed to identify an oxygen uptake interval where the growth of all the GEMs is oxygen-dependent and inspect how metabolite secretions are altered.

The analysis was done without considering the role of Candida metabolism in interaction with the microbiome. Recent advances in GEM analysis of microbiome have allowed simulating a microbial community where the individual members are considered different compartments of the system (Chan et al., 2017; Koch et al., 2019). Individual community reconstructions for Candida and bacterial species can be generated using the available GEMs. A pairwise interaction simulation between these two communities might provide a high-resolution understanding of the role of individual bacterial species in secreting the related metabolites as the interspecies communications are considered in this type of analysis.

In the last manuscript (Table 1), the first suite of *A. fumigatus* genome-scale strain-specific metabolic reconstructions originating from 252 different isolates spanning different locations in Germany (Barber et al., 2021) was built. This platform was built upon comprehensive genome information of the Af293 strain as a reference genome. Yet despite the comprehensiveness of genome information for this specific strain, complementary genome information specific to the other *A. fumigatus* strains (Garcia-Rubio et al., 2018) can add more resolution to generated strain-specific GEMs.

Multi-strain GEMs were used before to elucidate the metabolic diversity in human health-relevant bacterial strains and revealed e.g. pan metabolic capabilities of *Pseudomonas putida* (Nogales et al., 2020), loss of fitness relevant pathways for survival in the gastrointestinal environment in extraintestinal *Salmonella spp.* (Seif et al., 2018), and *Staphylococcus aureus* strain-specific metabolic capabilities linked to pathogenic traits and virulence acquisitions (Bosi et al., 2016). The developed strain-specific *A. fumigatus* GEMs can be served as a platform for investigating the metabolic diversity influencing growth rate capabilities, metabolic adaptation, and pathogenicity in this important fungal pathogen for the human host.

Compared to the *C. albicans* GEM reconstruction described in the third manuscript (Table 1), more curations were carried out for *A. fumigatus* pan-GEM generation, which led to high accuracy in predicting both growth capabilities and gene essentiality of the fungus. Yet, the gene essentiality data was very limited. A bigger collection of gene knock-out experiments can be carried out for future GEM development to evaluate the accuracy more extensively. Additionally, the capability of the GEM to produce known secretion products, the presence of blocked reactions, known incapabilities of the *A. fumigatus*, and predicted growth speeds in different growth conditions can be tested, and the pan-GEM can be refined accordingly.

There may be genomic differences between the clinical strain collection that was used to build the strain-specific GEMs and the clinical strains present in the cystic fibrosis patients even though both were isolated in hospitals in Germany. However, as has been shown in the previous study the genomic similarities of clinical strains are relatively high even when the isolates originate from different countries (Barber et al., 2021). This study can be expanded to include isolates from different continents to obtain a better understanding of the phenotype diversity of the fungus at the global level.

One of the goals of this study was to identify the metabolic differences between clinical and environmental strains using machine learning approaches. However, the number of *A. fumigatus* strains isolated from hospitals is around one-fourth of the ones isolated from farms. The imbalance dataset typically poses a challenge for classification using machine learning approaches since they are usually designed to deal with an equal number of samples in each group. The classification was done considering down- and up-sampling to cope with the negative effect of the imbalance dataset to some extent. Yet, including more *A. fumigatus* isolates from hospitals can increase the resolution to reveal how the fungus probably adjusts the phenotypes through genetic mutations to colonize better in the host body.

It is acknowledged that the MAMBO-derived metabolic output of the lung microbiome is an approximation relying merely on the taxonomic annotation of species. The effect of the host factors on the microbiome and metabolome composition of the lung is not neglectable. Therefore, a more precise assessment of the contribution of the lung microbiome to the nutritional profile in the human lung can be done in case a model system capable of that is available. Integrating such data into the reconstructed strain GEMs can expand our understanding of how the fungus regulates its metabolism in response to both the microbiome and the human factors to colonize better in the lung. Additionally, the study can be further expanded to reveal the contribution of the fungus to development of the other lung microbiome diseases such as allergic bronchopulmonary aspergillosis (ABPA), chronic obstructive pulmonary disease (COPD), and invasive pulmonary aspergillosis (IPA) (Kolwijck & van de Veerdonk, 2014).

SUMMARY

In an ever-growing world-wide population, human infectious diseases are an increasingly serious problem for public health. In recent decades, advances in medical sciences and hygiene improvement within social communities have been improving respective diagnostics and treatments. However, global reports indicate that the infection rate and the consequent morbidity and mortality rates are still considerably high, showing that the problem remains challenging to address. Drug-resistance development of human pathogens through evolutionary stages and emerging new pathogenic species and strains (as the recent pandemic triggered by SARS-CoV-2 resulting in COVID-19 dramatically revealed) are the main obstacles to efficient treatments. Hence, developing new efficient therapeutic approaches is a serious challenge that requires more effective research.

In particular, more than a million deaths and millions of infectious diseases cases per year caused by fungal pathogens have been reported globally in recent years. It indicates that the infection threat caused by human fungal pathogens has been underestimated. Hence, more investments must be put into fungal research to overcome the problem. The opportunistic pathogen *Candida albicans* and the airborne *Aspergillus fumigatus* are the two most prevalent fungal pathogens causing serious issues in medical care units. *C. albicans* is a commensal fungus in human microbiota and does not cause diseases in healthy individuals. However, upon a given opportunity, it can switch from commensal to pathogenic state in immunocompromised patients and overgrow in some organs such as skin and mucosal surfaces of e. g. the gut and cause infection ranging from superficial to severe systematic candidiasis. Unlike *C. albicans*, the original niche of *A. fumigatus* is soil. However, it distributes its conidia into the air, which humans and animals daily inhale. Similar to *C. albicans*, it does not cause serious issues in healthy individuals. However, in immunocompromised patients, it can initiate or intensify lung diseases such as chronic asthma and cystic fibrosis.

Despite the recent advances in fungal research, there is a little knowledge about the role of fungal metabolism in developing the infection when coexisting within the human body with microbial community members in different organs. This dissertation applied computational tools and implemented systems biology approaches to uncover key factors in the colonization of the pathogens, especially *C. albicans* and *A. fumigatus*, from a systems biology perspective and unseen by wet-lab experiments alone. Next to multi-omics data analysis, a

major effort was put into genome-scale metabolic models (GEMs) generation and analysis as a promising approach to shed light on the role of metabolism to develop the infection.

As a proof of principle, the thesis starts with two studies showing the computational approaches' applicability to understand better or confirm findings obtained by analyzing omics data. It includes using GEMs analysis to study the role of metabolism to predispose infection in the human microbiome. Next, metabolic systems biology approaches using GEMs reconstruction and analysis of human fungal pathogens were employed to study the role of metabolism in developing fungal infections. In this regard, four manuscripts shed light on key factors leading to the inhibition or promotion of *C. albicans* growth. This especially includes the first available GEM reconstruction of this fungus to theoretically study the intricate interaction of the fungus with the human host and the microbial community members. Lastly, a platform of 252 *A. fumigatus* GEMs at the strain resolution was generated. It revealed the phenotypic diversity of *A. fumigatus* strains isolated from different hospitals and farms in Germany and explained the contribution of the fungus to the shaping of the metabolic landscape of the lung microbiome in a favorable manner for the fungal growth.

In summary, in this thesis I demonstrated that computational approaches and analyses can extend our understanding of fungal pathogenicity and might proof pivotal for development of future early diagnostics and therapies of fungal infectious diseases.

ZUSAMMENFASSUNG

In einer stets wachsenden Weltbevölkerung stellen Infektionskrankheiten ein zunehmend ernstes Problem für die öffentliche Gesundheit dar. In den letzten Jahrzehnten haben Fortschritte in den medizinischen Wissenschaften und die Verbesserung der Hygiene in verschiedenen Lebensgemeinschaften zu einer Verbesserung von Diagnostik und Behandlung geführt. Weltweite Berichte zeigen jedoch, dass die Infektionsrate und die daraus resultierenden Morbiditäts- und Mortalitätsraten immer noch beträchtlich sind, was darauf hindeutet, dass es nach wie vor schwierig ist, dieses Problem zu lösen. Die Entwicklung von Arzneimittelresistenzen bei humanen Krankheitserregern im Laufe der Evolution und das Auftreten neuer pathogener Arten und Stämme (wie die durch SARS-CoV-2 ausgelöste Pandemie resultierend in COVID-19 kürzlich dramatisch offenbarte) sind die größten Hindernisse für eine wirksame Behandlung. Daher ist die Entwicklung neuer wirksamer therapeutischer Ansätze eine ernsthafte Herausforderung und erfordert eine effektivere und gezielte Forschung.

Insbesondere wurden über die letzten Jahren weltweit mehr als eine Million Todesfälle und Millionen Fälle durch Pilzerreger verursachte Infektionskrankheiten pro Jahr gemeldet. Dies deutet darauf hin, dass die von Pilzerregern ausgehende Infektionsgefahr für den Menschen bisher unterschätzt wurde. Daher müssen mehr Investitionen in die Pilzforschung getätigt werden, um das Problem zu lösen. Der opportunistische Erreger *Candida albicans* und der über die Luft übertragene *Aspergillus fumigatus* sind die beiden häufigsten Pilzerreger, die in ernsthafte Probleme in medizinischen Einrichtungen verursachen. *C. albicans* ist ein kommensaler Pilz in der menschlichen Mikrobiota und verursacht bei gesunden Menschen keine Krankheiten. Bei immungeschwächten Patienten kann er jedoch vom kommensalen in den pathogenen Zustand übergehen und in einigen Organen wie der Haut oder auf den Schleimhautoberflächen z. B. des Darms überwuchern und Infektionen verursachen. Diese können von oberflächlicher bis zu schwerer systematischer Candidose reichen. Im Gegensatz zu *C. albicans* ist die ursprüngliche Nische von *A. fumigatus* der Boden. Er verbreitet jedoch seine Konidien über die Luft, die Menschen und Tiere täglich einatmen. Ähnlich zu *C. albicans* verursacht er bei gesunden Menschen keine ernsthaften Probleme. Bei immungeschwächten Patienten kann es jedoch Lungenkrankheiten wie invasive Aspergillose, chronisches Asthma oder Mukoviszidose auslösen oder verstärken.

Trotz der jüngsten Fortschritte in der Pilzforschung gibt es nur wenige Erkenntnisse über die Rolle des Pilzstoffwechsels in der Entwicklung einer Infektion, wenn er im menschlichen

Körper mit Mitgliedern des Mikrobioms verschiedener Organe koexistiert. In dieser Dissertation wurden computergestützte und systembiologische Ansätze implementiert, um Schlüsselfaktoren bei der Besiedlung von Krankheitserregern, insbesondere von *C. albicans* und *A. fumigatus*, aus einer systembiologischen Perspektive aufzudecken, welche durch Wet-lab Experimente allein nicht studierbar sind. Neben der Analyse von Multi-OMICS-Daten, wurde viel Arbeit in die Rekonstruktion und Analyse von genomskaligen Stoffwechselmodellen (GEMs) aufgewendet. Die Untersuchung von GEMs stellt einen vielversprechenden Ansatz dar, um die Rolle des Stoffwechsels bei der Entwicklung der Infektion zu beleuchten.

Zur Demonstration des Prinzips beginnt die Arbeit mit zwei Studien, die die Eignung von Berechnungsansätzen um wissenschaftliche Erkenntnisse, die durch die Analyse von OMICS-Daten gewonnen wurden, besser zu verstehen oder zu bestätigen. Dazu gehört die Verwendung der GEM-Analyse zur Untersuchung der Rolle des Stoffwechsels bei der Prädisposition für Infektionen im menschlichen Mikrobiom. Als Nächstes wurden metabolische systembiologische Ansätze unter Verwendung der GEM-Rekonstruktion und -Analyse menschlicher Pilzerreger eingesetzt, um die Rolle des Stoffwechsels bei der Entstehung von Pilzinfektionen zu untersuchen. In diesem Zusammenhang beleuchten vier Manuskripte die Schlüsselfaktoren, die zur Hemmung oder Förderung des Wachstums von C. albicans führen. Dazu gehört insbesondere die erste verfügbare GEM-Rekonstruktion dieses Pilzes, um die komplizierte Interaktion des Pilzes mit dem menschlichen Wirt und den Mitgliedern der mikrobiellen Gemeinschaft theoretisch zu untersuchen. Schließlich wurde eine Plattform mit 252 verschiedener A. fumigatus-GEMs erstellt. Diese offenbarte die phänotypische Vielfalt von A. fumigatus-Stämmen, die aus verschiedenen Krankenhäusern und landwirtschaftlichen Betrieben in Deutschland isoliert wurden und beleuchtete den Beitrag des Pilzes zur Manipulation der metabolischen Landschaft des Lungenmikrobioms zugunsten des eigenen Wachstums.

Zusammenfassend habe ich in dieser Arbeit gezeigt, dass computergestützte Ansätze und Analysen unser Verständnis der Pathogenität von Pilzen erweitern können. Diese Ansätze mögen sich als entscheidend für die Entwicklung künftiger Frühdiagnosen und Therapien von Pilzerkrankungen erweisen.

BIBLIOGRAPHY

Aderem, A., Adkins, J. N., Ansong, C., Galagan, J., Kaiser, S., Korth, M. J., Law, G. L., McDermott, J. G., Proll, S. C., Rosenberger, C., Schoolnik, G., & Katze, M. G. (2011). A systems biology approach to infectious disease research: Innovating the pathogen-host research paradigm. *MBio*, *2*(1). https://doi.org/10.1128/mBio.00325-10

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., & Nielsen, J. (2013). The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLoS Computational Biology*, *9*(3). https://doi.org/10.1371/journal.pcbi.1002980

Ahangarkani, F., Puts, Y., Nabili, M., Khodavaisy, S., Moazeni, M., Salehi, Z., Laal Kargar, M., Badali, H., & Meis, J. F. (2020). First azole-resistant Aspergillus fumigatus isolates with the environmental TR46/Y121F/T289A mutation in Iran. *Mycoses*, *63*(5), 430–436. https://doi.org/10.1111/myc.13064

Allonsius, C. N., Vandenheuvel, D., Oerlemans, E. F. M., Petrova, M. I., Donders, G. G. G., Cos, P., Delputte, P., & Lebeer, S. (2019). Inhibition of Candida albicans morphogenesis by chitinase from Lactobacillus rhamnosus GG. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-39625-0

Alonso-Hearn, M., Canive, M., Blanco-Vazquez, C., Torremocha, R., Balseiro, A., Amado, J., Varela-Martinez, E., Ramos, R., Jugo, B. M., & Casais, R. (2019). RNA-Seq analysis of ileocecal valve and peripheral blood from Holstein cattle infected with Mycobacterium avium subsp. paratuberculosis revealed dysregulation of the CXCL8/IL8 signaling pathway. *Scientific Reports*, *9*(1), 1–17. https://doi.org/10.1038/s41598-019-51328-0

An, G., Mi, Q., Dutta-Moscato, J., & Vodovotz, Y. (2009). Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *1*(2), 159–171. https://doi.org/10.1002/wsbm.45

Arakawa, K., Yamada, Y., Shinoda, K., Nakayama, Y., & Tomita, M. (2006). GEM system: Automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics*, *7*, 1–11. https://doi.org/10.1186/1471-2105-7-168

Arastehfar, A., Carvalho, A., van de Veerdonk, F. L., Jenks, J. D., Koehler, P., Krause, R., Cornely, O. A., Perlin, D. S., Lass-Flörl, C., & Hoenigl, M. (2020). COVID-19 associated pulmonary aspergillosis (CAPA)—from immunology to treatment. *Journal of Fungi*, *6*(2), 1–17. https://doi.org/10.3390/jof6020091

Armstrong, G. L., Conn, L. A., & Pinner, R. W. (1999). Trends in infectious disease mortality in the United States during the 20th century. *Journal of the American Medical Association*, *281*(1), 61–66. https://doi.org/10.1001/jama.281.1.61

Arnold, A., & Nikoloski, Z. (2014). Bottom-up metabolic reconstruction of arabidopsis and its application to determining the metabolic costs of enzyme production. *Plant Physiology*, *165*(3), 1380–1391. https://doi.org/10.1104/pp.114.235358

Barber, A. E., Sae-Ong, T., Kang, K., Seelbinder, B., Li, J., Walther, G., Panagiotou, G., & Kurzai, O. (2021). Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection. *Nature Microbiology*, *6*(12), 1526–1536. https://doi.org/10.1038/s41564-021-00993-x

Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., & Kaleta, C. (2017). BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Computational Biology*, *13*(5), 1–22. https://doi.org/10.1371/journal.pcbi.1005544

Beldomenico, P. M., & Begon, M. (2010). Disease spread, susceptibility and infection

intensity: vicious circles? *Trends in Ecology and Evolution*, *25*(1), 21–27. https://doi.org/10.1016/j.tree.2009.06.015

Berman, J. (2006). Morphogenesis and cell cycle progression in Candida albicans. *Current Opinion in Microbiology*, *9*(6), 595–601. https://doi.org/10.1016/j.mib.2006.10.007

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

Berthenet, E., Yahara, K., Thorell, K., Pascoe, B., Meric, G., Mikhail, J. M., Engstrand, L., Enroth, H., Burette, A., Megraud, F., Varon, C., Atherton, J. C., Smith, S., Wilkinson, T. S., Hitchings, M. D., Falush, D., & Sheppard, S. K. (2018). A GWAS on Helicobacter pylori strains points to genetic variants associated with gastric cancer risk. *BMC Biology*, *16*(1), 1–11. https://doi.org/10.1186/s12915-018-0550-3

Biggs, M. B., & Papin, J. A. (2013). Novel Multiscale Modeling Tool Applied to Pseudomonas aeruginosa Biofilm Formation. *PLoS ONE*, *8*(10), 1–8. https://doi.org/10.1371/journal.pone.0078011

Bochner, B. R., Gadzinski, P., & Panomitros, E. (2001). Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research*, *11*(7), 1246–1255. https://doi.org/10.1101/gr.186501

Bongomin, F., Gago, S., Oladele, R. O., & Denning, D. W. (2017). Global and multi-national prevalence of fungal diseases—estimate precision. *Journal of Fungi*, *3*(4). https://doi.org/10.3390/jof3040057

Bordbar, A., Yurkovich, J. T., Paglia, G., Rolfsson, O., Sigurjónsson, Ó. E., & Palsson, B. O. (2017). Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. *Scientific Reports*, *7*(March), 1–12. https://doi.org/10.1038/srep46249

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., & Palsson, B. Ø. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences*, *113*(26), E3801–E3809. https://doi.org/10.1073/pnas.1523199113

Brakhage, A. (2005). Systemic Fungal Infections Caused by Aspergillus Species: Epidemiology, Infection Process and Virulence Determinants. *Current Drug Targets*, *6*(8), 875–886. https://doi.org/10.2174/138945005774912717

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., & White, T. C. (2012). Hidden killers: Human fungal infections. *Science Translational Medicine*, *4*(165), 1–10. https://doi.org/10.1126/scitranslmed.3004404

Brown, G. D., Denning, D. W., & Levitz, S. M. (2012). Tackling human fungal infections. *Science*, *336*(6082), 647. https://doi.org/10.1126/science.1222236

Bruggeman, F. J., & Westerhoff, H. V. (2007). The nature of systems biology. *Trends in Microbiology*, *15*(1), 45–50. https://doi.org/10.1016/j.tim.2006.11.003

Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G. A., Aurich, M. K., Prlic, A., Sastry, A., Danielsdottir, A. D., Heinken, A., Noronha, A., Rose, P. W., Burley, S. K., Fleming, R. M. T., Nielsen, J., … Palsson, B. O. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology*, *36*(3), 272–281. https://doi.org/10.1038/nbt.4072

Çakir, T., Patil, K. R., Önsan, Z. I., Ülgen, K. Ö., Kirdar, B., & Nielsen, J. (2006). Integration of metabolome data with metabolic networks reveals reporter reactions. *Molecular Systems Biology*, *2*, 1–11. https://doi.org/10.1038/msb4100085

Calderone, R. a, & Fonzi, W. a. (2001). Virulence factors of Candida albicans. *TRENDS in*

Microbiology, *9*(7), 327–335. http://tim.trends.com
Cano-Gamez, E., & Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics*, *11*(May), 1–21. https://doi.org/10.3389/fgene.2020.00424
Chan, S. H. J., Simons, M. N., & Maranas, C. D. (2017). SteadyCom: Predicting microbial abundances while ensuring community stability. *PLoS Computational Biology*, *13*(5), 1–25. https://doi.org/10.1371/journal.pcbi.1005539
Chen, Y., Dong, F., Zhao, J., Fan, H., Qin, C., Li, R., Verweij, P. E., Zheng, Y., & Han, L. (2020). High azole resistance in aspergillus fumigatus isolates from Strawberry Fields, China, 2018. *Emerging Infectious Diseases*, *26*(1), 81–89. https://doi.org/10.3201/eid2601.190885
Cho, J. S., Gu, C., Han, T. H., Ryu, J. Y., & Lee, S. Y. (2019). Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring. *Current Opinion in Systems Biology*, *15*(March), 1–11. https://doi.org/10.1016/j.coisb.2019.02.009
Choi, J. J., & Kim, S. H. (2017). A genome Tree of Life for the Fungi kingdom. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(35), 9391–9396. https://doi.org/10.1073/pnas.1711939114
Cohen, M. L. (2000). *Changing patterns of infectious disease*. *406*(August). https://www.nature.com/articles/35021206
Conway, T., & Schoolnik, G. K. (2003). Microarray expression profiling: Capturing a genome-wide portrait of the transcriptome. *Molecular Microbiology*, *47*(4), 879–889. https://doi.org/10.1046/j.1365-2958.2003.03338.x
da Silva Dantas, A., Lee, K. K., Raziunaite, I., Schaefer, K., Wagener, J., Yadav, B., & Gow, N. A. (2016). Cell biology of Candida albicans–host interactions. *Current Opinion in Microbiology*, *34*, 111–118. https://doi.org/10.1016/j.mib.2016.08.006
De Lorenzo, V., & Galperin, M. (2009). Microbial systems biology: Bottom up and top down. *FEMS Microbiology Reviews*, *33*(1), 1–2. https://doi.org/10.1111/j.1574-6976.2008.00147.x
de Vries, D. H., Matzaraki, V., Bakker, O. B., Brugge, H., Westra, H. J., Netea, M. G., Franke, L., Kumar, V., & van der Wijst, M. G. P. (2020). Integrating GWAS with bulk and single-cell RNA-sequencing reveals a role for LY86 in the anti-Candida host response. *PLoS Pathogens*, *16*(4), 1–17. https://doi.org/10.1371/journal.ppat.1008408
Denning, D. W., Kneale, M., Sobel, J. D., & Rautemaa-Richardson, R. (2018). Global burden of recurrent vulvovaginal candidiasis: a systematic review. *The Lancet Infectious Diseases*, *18*(11), e339–e347. https://doi.org/10.1016/S1473-3099(18)30103-8
Dethlefsen, L., Huse, S., Sogin, M. L., & Relman, D. A. (2008). The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biology*, *6*(11). https://doi.org/10.1371/journal.pbio.0060280
Devoid, S., Overbeek, R., DeJongh, M., Vonstein, V., Best, A. A., & Henry, C. (2013). Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED. *Systems Metabolic Engineering*, *985*, 17–45.
Dias, O., Rocha, M., Ferreira, E. C., & Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Research*, *43*(8), 3899–3910. https://doi.org/10.1093/nar/gkv294
Durand, G. A., Raoult, D., & Dubourg, G. (2019). Antibiotic discovery: history, methods and perspectives. *International Journal of Antimicrobial Agents*, *53*(4), 371–382. https://doi.org/10.1016/j.ijantimicag.2018.11.010
El-Semman, I. E., Karlsson, F. H., Shoaie, S., Nookaew, I., Soliman, T. H., & Nielsen, J.

(2014). Genome-scale metabolic reconstructions of Bifidobacterium adolescentis L2-32 and Faecalibacterium prausnitzii A2-165 and their interaction. *BMC Systems Biology*, *8*(1), 1–11. https://doi.org/10.1186/1752-0509-8-41

Ene, I. V., Brunke, S., Brown, A. J. P., & Hube, B. (2014). Metabolism in fungal pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, *4*(12), 1–22. https://doi.org/10.1101/cshperspect.a019695

Fan, D., Coughlin, L. A., Neubauer, M. M., Kim, J., Kim, M. S., Zhan, X., Simms-Waldrip, T. R., Xie, Y., Hooper, L. V., & Koh, A. Y. (2015). Activation of HIF-1α and LL-37 by commensal bacteria inhibits Candida albicans colonization. *Nature Medicine*, *21*(7), 808–814. https://doi.org/10.1038/nm.3871

Faria, J. P., Rocha, M., Rocha, I., & Henry, C. S. (2018). Methods for automated genome-scale metabolic model reconstruction. *Biochemical Society Transactions*, *46*(4), 931–936. https://doi.org/10.1042/BST20170246

Fausto, A., Rodrigues, M. L., & Coelho, C. (2019). The still underestimated problem of fungal diseases worldwide. *Frontiers in Microbiology*, *10*(FEB), 1–5. https://doi.org/10.3389/fmicb.2019.00214

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., & Palsson, B. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, *3*(121), 1–18. https://doi.org/10.1038/msb4100155

Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, *13*(3), 344–349. https://doi.org/10.1016/j.mib.2010.03.003

Förster, T. M., Mogavero, S., Dräger, A., Graf, K., Polke, M., Jacobsen, I. D., & Hube, B. (2016). Enemies and brothers in arms: Candida albicans and gram-positive bacteria. *Cellular Microbiology*, *18*(12), 1709–1715. https://doi.org/10.1111/cmi.12657

Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B., & Lercher, M. J. (2017). Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Computational Biology*, *13*(4), 1–14. https://doi.org/10.1371/journal.pcbi.1005494

Garcia-Rubio, R., Monzon, S., Alcazar-Fuoli, L., Cuesta, I., & Mellado, E. (2018). Genome-wide comparative analysis of Aspergillus fumigatus strains: The reference genome as a matter of concern. *Genes*, *9*(7). https://doi.org/10.3390/genes9070363

Geiser, D. M. (2009). Sexual structures in Aspergillus: Morphology, importance and genomics. *Medical Mycology*, *47*(SUPPL. 1), 20–26. https://doi.org/10.1080/13693780802139859

Geng, J., Ji, B., Li, G., López-Isunza, F., & Nielsen, J. (2021). CODY enables quantitatively spatiotemporal predictions on in vivo gut microbial variability induced by diet intervention. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(13). https://doi.org/10.1073/pnas.2019336118

Gould, A. L., Zhang, V., Lamberti, L., Jones, E. W., Obadia, B., Korasidis, N., Carlson, J. M., Beerenwinkel, N., & Ludington, W. B. (2017). *Title : High-dimensional microbiome interactions shape host fitness Authors : Abstract : keywords : Significance : Introduction :* 1–72.

Graf, K., Last, A., Gratz, R., Allert, S., Linde, S., Westermann, M., Gröger, M., Mosig, A. S., Gresnigt, M. S., & Hube, B. (2019). Keeping Candida commensal: How lactobacilli antagonize pathogenicity of Candida albicans in an in vitro gut model. *DMM Disease Models and Mechanisms*, *12*(9). https://doi.org/10.1242/dmm.039719

Granillo, A. R., Canales, M. G. M., Espíndola, M. E. S., Rivera, M. A. M., De Lucio, V. M. B., & Tovar, A. V. R. (2015). Antibiosis interaction of Staphylococccus aureus on

Aspergillus fumigatus assessed in vitro by mixed biofilm formation. *BMC Microbiology*, *15*(1), 1–15. https://doi.org/10.1186/s12866-015-0363-2

Grassly, N. C., & Fraser, C. (2008). Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, *6*(6), 477–487. https://doi.org/10.1038/nrmicro1845

Haas, B. J., & Zody, M. C. (2010). Advancing RNA-Seq analysis. *Nature Biotechnology*, *28*(5), 421–423. https://doi.org/10.1038/nbt0510-421

Harbers, M., & Carninci, P. (2005). Tag-based approaches for transcriptome research and genome annotation. *Nature Methods*, *2*(7), 495–502. https://doi.org/10.1038/nmeth768

Heinken, A., & Thiele, I. (2015). Anoxic conditions promote species-specific mutualism between gut microbes In Silico. *Applied and Environmental Microbiology*, *81*(12), 4049–4061. https://doi.org/10.1128/AEM.00101-15

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, *35*(SUPPL.2), 585–587. https://doi.org/10.1093/nar/gkm259

Huffnagle, G. B., & Noverr, M. C. (2013). The emerging world of the fungal microbiome. *Trends in Microbiology*, *21*(7), 334–341. https://doi.org/10.1016/j.tim.2013.04.002

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Jamshidi, N., & Raghunathan, A. (2015). Cell scale host-pathogen modeling: Another branch in the evolution of constraint-based methods. *Frontiers in Microbiology*, *6*(OCT), 1–16. https://doi.org/10.3389/fmicb.2015.01032

Janbon, G., Quintin, J., Lanternier, F., & d'Enfert, C. (2019). Studying fungal pathogens of humans and fungal infections: fungal diversity and diversity of approaches. *Microbes and Infection*, *21*(5–6), 237–245. https://doi.org/10.1016/j.micinf.2019.06.011

Jensen, C. S., Norsigian, C. J., Fang, X., Nielsen, X. C., Christensen, J. J., Palsson, B. O., & Monk, J. M. (2020). Reconstruction and Validation of a Genome-Scale Metabolic Model of Streptococcus oralis (iCJ415), a Human Commensal and Opportunistic Pathogen. *Frontiers in Genetics*, *11*(March), 1–11. https://doi.org/10.3389/fgene.2020.00116

Kadosh, D., Najvar, L. K., Bocanegra, R., Olivo, M., Kirkpatrick, W. R., Wiederhold, N. P., & Patterson, T. F. (2016). Effect of antifungal treatment in a diet-based murine model of disseminated candidiasis acquired via the gastrointestinal tract. *Antimicrobial Agents and Chemotherapy*, *60*(11), 6703–6708. https://doi.org/10.1128/AAC.01144-16

Karp, P. D., Latendresse, M., Paley, S. M., Krummenacker, M., Ong, Q. D., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., Spaulding, A., Fulcher, C., Keseler, I. M., & Caspi, R. (2016). Pathway tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, *17*(5), 877–890. https://doi.org/10.1093/bib/bbv079

Khanna, S., Pardi, D. S., Kelly, C. R., Kraft, C. S., Dhere, T., Henn, M. R., Lombardo, M. J., Vulic, M., Ohsumi, T., Winkler, J., Pindar, C., McGovern, B. H., Pomerantz, R. J., Aunins, J. G., Cook, D. N., & Hohmann, E. L. (2016). A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent Clostridium difficile Infection. *Journal of Infectious Diseases*, *214*(2), 173–181. https://doi.org/10.1093/infdis/jiv766

Koch, S., Kohrs, F., Lahmann, P., Bissinger, T., Wendschuh, S., Benndorf, D., Reichl, U., & Klamt, S. (2019). Redcom: A strategy for reduced metabolic modeling of complex

microbial communities and its application for analyzing experimental datasets from anaerobic digestion. In *PLoS Computational Biology* (Vol. 15, Issue 2). https://doi.org/10.1371/journal.pcbi.1006759

Köhler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, S. M., Boerkoel, C. F., Boycott, K. M., Brudno, M., Buske, O. J., Chinnery, P. F., Cipriani, V., Connell, L. E., Dawkins, H. J. S., DeMare, L. E., Devereau, A. D., De Vries, B. B. A., … Robinson, P. N. (2017). The human phenotype ontology in 2017. *Nucleic Acids Research*, *45*(D1), D865–D876. https://doi.org/10.1093/nar/gkw1039

Kolwijck, E., & van de Veerdonk, F. L. (2014). The potential impact of the pulmonary microbiome on immunopathogenesis of Aspergillus-related lung disease. *European Journal of Immunology*, *44*(11), 3156–3165. https://doi.org/10.1002/eji.201344404

Korgaonkar, A., Trivedi, U., Rumbaugh, K. P., & Whiteley, M. (2013). Community surveillance enhances Pseudomonas aeruginosa virulence during polymicrobial infection. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 1059–1064. https://doi.org/10.1073/pnas.1214550110

Kumamoto, C. A., Gresnigt, M. S., & Hube, B. (2020). The gut, the bad and the harmless: Candida albicans as a commensal and opportunistic pathogen in the intestine. *Current Opinion in Microbiology*, *56*, 7–15. https://doi.org/10.1016/j.mib.2020.05.006

Lakhani, C. M., Tierney, B. T., Manrai, A. K., Yang, J., Visscher, P. M., & Patel, C. J. (2017). Hemizygosity Enables a Mutational Transition Governing Fungal Virulence and Commensalism. *Physiology & Behavior*, *176*(3), 139–148. https://doi.org/10.1016/j.chom.2019.01.005.Hemizygosity

Latgé, J.-P., & Chamilos, G. (2020). Aspergillus fumigatus and Aspergillosis in 2019. *Clinical Microbiology Reviews*, *30*.

Liao, Y. C., Huang, T. W., Chen, F. C., Charusanti, P., Hong, J. S. J., Chang, H. Y., Tsai, S. F., Palsson, B. O., & Hsiung, C. A. (2011). An experimentally validated genome-scale metabolic reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228. *Journal of Bacteriology*, *193*(7), 1710–1717. https://doi.org/10.1128/JB.01218-10

Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, *46*(15), 7542–7553. https://doi.org/10.1093/nar/gky537

Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R. M. T., & Thiele, I. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, *35*(1), 81–89. https://doi.org/10.1038/nbt.3703

Mahadevan, R., & Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, *5*(4), 264–276. https://doi.org/10.1016/j.ymben.2003.09.002

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters*, *583*(24), 3966–3973. https://doi.org/10.1016/j.febslet.2009.10.036

Máiz, L., Vendrell, M., Olveira, C., Girón, R., Nieto, R., & Martínez-García, M. Á. (2015). Prevalence and factors associated with isolation of aspergillus and candida from sputum in patients with non-cystic fibrosis bronchiectasis. *Respiration*, *89*(5), 396–403. https://doi.org/10.1159/000381289

Marr, K. A., Carter, R. A., Crippa, F., Wald, A., & Corey, L. (2002). Epidemiology and outcome of mould infections in hematopoietic stem cell transplant recipients. *Clinical Infectious Diseases*, *34*(7), 909–917. https://doi.org/10.1086/339202

Matsubara, V. H., Bandara, H. M. H. N., Mayer, M. P. A., & Samaranayake, L. P. (2016). Probiotics as Antifungals in Mucosal Candidiasis. *Clinical Infectious Diseases*, *62*(9), 1143–1153. https://doi.org/10.1093/cid/ciw038

Medlock, G. L., Carey, M. A., McDuffie, D. G., Mundy, M. B., Giallourou, N., Swann, J. R., Kolling, G. L., & Papin, J. A. (2018). Inferring Metabolic Mechanisms of Interaction within a Defined Gut Microbiota. *Cell Systems*, *7*(3), 245-257.e7. https://doi.org/10.1016/j.cels.2018.08.003

Miramón, P., Pountain, A., van Hoof, A., & Lorenz, M. C. (2020). *crossm The Paralogous Transcription Factors Stp1 and Stp2 of Acquisition and Host Interaction*. *September 2019*, 1–18.

Mishra, A. A., & Koh, A. Y. (2018). Adaptation of Candida albicans During Gastrointestinal Tract Colonization. *Current Clinical Microbiology Reports*, *5*(3), 165–172. https://doi.org/10.1007/s40588-018-0096-8

Mo, M. L., Palsson, B., & Herrgård, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*, *3*, 1–17. https://doi.org/10.1186/1752-0509-3-37

Molina-Mora, J. A., Chinchilla-Montero, D., Chavarría-Azofeifa, M., Ulloa-Morales, A. J., Campos-Sánchez, R., Mora-Rodríguez, R., Shi, L., & García, F. (2020). Transcriptomic determinants of the response of ST-111 Pseudomonas aeruginosa AG1 to ciprofloxacin identified by a top-down systems biology approach. *Scientific Reports*, *10*(1), 1–23. https://doi.org/10.1038/s41598-020-70581-2

Monk, J. M., Charusanti, P., Aziz, R. K., Lerman, J. A., Premyodhin, N., Orth, J. D., Feist, A. M., & Palsson, B. (2013). Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50), 20338–20343. https://doi.org/10.1073/pnas.1307797110

Mortensen, K. L., Johansen, H. K., Fuursted, K., Knudsen, J. D., Gahrn-Hansen, B., Jensen, R. H., Howard, S. J., & Arendrup, M. C. (2011). A prospective survey of Aspergillus spp. in respiratory tract samples: Prevalence, clinical impact and antifungal susceptibility. *European Journal of Clinical Microbiology and Infectious Diseases*, *30*(11), 1355–1363. https://doi.org/10.1007/s10096-011-1229-7

Mowat, E., Rajendran, R., Williams, C., McCulloch, E., Jones, B., Lang, S., & Ramage, G. (2010). Pseudomonas aeruginosa and their small diffusible extracellular molecules inhibit Aspergillus fumigatus biofilm formation. *FEMS Microbiology Letters*, *313*(2), 96–102. https://doi.org/10.1111/j.1574-6968.2010.02130.x

Nash, A. K., Auchtung, T. A., Wong, M. C., Smith, D. P., Gesell, J. R., Ross, M. C., Stewart, C. J., Metcalf, G. A., Muzny, D. M., Gibbs, R. A., Ajami, N. J., & Petrosino, J. F. (2017). The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome*, *5*(1), 153. https://doi.org/10.1186/s40168-017-0373-4

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

Norsigian, C. J., Kavvas, E., Seif, Y., Palsson, B. O., & Monk, J. M. (2018). iCN718, an updated and improved genome-scale metabolic network reconstruction of Acinetobacter baumannii AYE. *Frontiers in Genetics*, *9*(APR), 1–9. https://doi.org/10.3389/fgene.2018.00121

Notebaart, R. A., van Enckevort, F. H. J., Francke, C., Siezen, R. J., & Teusink, B. (2006). Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics*, *7*, 1–10. https://doi.org/10.1186/1471-2105-7-296

O'Malley, M. A., & Dupré, J. (2005). Fundamental issues in systems biology. *BioEssays*, *27*(12), 1270–1276. https://doi.org/10.1002/bies.20323

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

Palsson, B. O. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press.

Panpetch, W., Somboonna, N., Palasuk, M., Hiengrach, P., Finkelman, M., Tumwasorn, S., & Leelahavanichkul, A. (2019). Oral Candida administration in a Clostridium difficile mouse model worsens disease severity but is attenuated by Bifidobacterium. *PLoS ONE*, *14*(1), 1–18. https://doi.org/10.1371/journal.pone.0210798

Pfaller, M. A., & Diekema, D. J. (2007). Epidemiology of invasive candidiasis: A persistent public health problem. *Clinical Microbiology Reviews*, *20*(1), 133–163. https://doi.org/10.1128/CMR.00029-06

Pinney, J. W., Shirley, M. W., McConkey, G. A., & Westhead, D. R. (2005). metaSHARK: Software for automated metabolic network prediction from DNA sequence and its application to the genomes of Plasmodium falciparum and Eimeria tenella. *Nucleic Acids Research*, *33*(4), 1399–1409. https://doi.org/10.1093/nar/gki285

Pitkänen, E., Jouhten, P., Hou, J., Syed, M. F., Blomberg, P., Kludas, J., Oja, M., Holm, L., Penttilä, M., Rousu, J., & Arvas, M. (2014). Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Computational Biology*, *10*(2). https://doi.org/10.1371/journal.pcbi.1003465

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., … Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Ramírez-Zavala, B., Mottola, A., Haubenreißer, J., Schneider, S., Allert, S., Brunke, S., Ohlsen, K., Hube, B., & Morschhäuser, J. (2017). The Snf1-activating kinase Sak1 is a key regulator of metabolic adaptation and in vivo fitness of Candida albicans. *Molecular Microbiology*, *104*(6), 989–1007. https://doi.org/10.1111/mmi.13674

Rao, C., Coyte, K. Z., Bainter, W., Geha, R. S., Martin, C. R., & Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*, *591*(7851), 633–638. https://doi.org/10.1038/s41586-021-03241-8

Renz, A., Widerspick, L., & Dräger, A. (2021). Genome-scale metabolic model of infection with SARS-CoV-2 mutants confirms guanylate kinase as robust potential antiviral target. *Genes*, *12*(6). https://doi.org/10.3390/genes12060796

Richardson, M. D., & Warnock, D. W. (2012). Fungal Infection: Diagnosis and Management, Fourth Edition. In *Fungal Infection: Diagnosis and Management, Fourth Edition*. Wiley. https://doi.org/10.1002/9781118321492

Rossoni, R. D., dos Santos Velloso, M., Figueiredo, L. M. A., Martins, C. P., Jorge, A. O. C., & Junqueira, J. C. (2018). Clinical strains of Lactobacillus reduce the filamentation of Candida albicans and protect Galleria mellonella against experimental candidiasis. *Folia Microbiologica*, *63*(3), 307–314. https://doi.org/10.1007/s12223-017-0569-9

Seelbinder, B., Chen, J., Brunke, S., Vazquez-Uribe, R., Santhaman, R., Meyer, A. C., De Oliveira Lino, F. S., Chan, K. F., Loos, D., Imamovic, L., Tsang, C. C., Lam, R. P. K., Sridhar, S., Kang, K., Hube, B., Woo, P. C. Y., Sommer, M. O. A., & Panagiotou, G. (2020). Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. *Microbiome*, *8*(1), 1–20. https://doi.org/10.1186/s40168-020-00899-6

Seif, Y., Kavvas, E., Lachance, J.-C., Yurkovich, J. T., Nuccio, S.-P., Fang, X., Catoiu, E.,

Raffatellu, M., Palsson, B. O., & Monk, J. M. (2018). Genome-scale metabolic reconstructions of multiple Salmonella strains reveal serovar-specific metabolic traits. *Nature Communications*, *9*(1), 3771. https://doi.org/10.1038/s41467-018-06112-5

Shi, N., Li, N., Duan, X., & Niu, H. (2017). Interaction between the gut microbiome and mucosal immune system. *Military Medical Research*, *4*(1), 1–7. https://doi.org/10.1186/s40779-017-0122-9

Shirtliff, M. E., Peters, B. M., & Jabra-Rizk, M. A. (2009). Cross-kingdom interactions: Candida albicans and bacteria. *FEMS Microbiology Letters*, *299*(1), 1–8. https://doi.org/10.1111/j.1574-6968.2009.01668.x

Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., & Ruppin, E. (2011). Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. *PLoS Computational Biology*, *7*(3), 1–8. https://doi.org/10.1371/journal.pcbi.1002018

Shoaie, S., Ghaffari, P., Kovatcheva-Datchary, P., Mardinoglu, A., Sen, P., Pujos-Guillot, E., De Wouters, T., Juste, C., Rizkalla, S., Chilloux, J., Hoyles, L., Nicholson, J. K., Dore, J., Dumas, M. E., Clement, K., Bäckhed, F., & Nielsen, J. (2015). Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metabolism*, *22*(2), 320–331. https://doi.org/10.1016/j.cmet.2015.07.001

Simeonidis, E., & Price, N. D. (2015). Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology and Biotechnology*, *42*(3), 327–338. https://doi.org/10.1007/s10295-014-1576-3

Singh, D., & Lercher, M. J. (2020). Network reduction methods for genome-scale metabolic models. *Cellular and Molecular Life Sciences*, *77*(3), 481–488. https://doi.org/10.1007/s00018-019-03383-z

Srivastava, V., Singla, R. K., & Dubey, A. K. (2018). Emerging Virulence, Drug Resistance and Future Anti-fungal Drugs for Candida Pathogens. *Current Topics in Medicinal Chemistry*, *18*(9), 759–778. https://doi.org/10.2174/1568026618666180528121707

Suhre, K., Shin, S. Y., Petersen, A. K., Mohney, R. P., Meredith, D., Wägele, B., Altmaier, E., Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C. J., De Angelis, M. H., Kastenmüller, G., Köttgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H. W., … Gieger, C. (2011). Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, *477*(7362), 54–62. https://doi.org/10.1038/nature10354

Sun, Y., Zuo, T., Cheung, C. P., Gu, W., Wan, Y., Zhang, F., Chen, N., Zhan, H., Yeoh, Y. K., Niu, J., Du, Y., Zhang, F., Wen, Y., Yu, J., Sung, J. J. Y., Chan, P. K. S., Chan, F. K. L., Wang, K., Ng, S. C., & Miao, Y. (2021). Population-Level Configurations of Gut Mycobiome Across 6 Ethnicities in Urban and Rural China. *Gastroenterology*, *160*(1), 272-286.e11. https://doi.org/10.1053/j.gastro.2020.09.014

Swainston, N., Smallbone, K., Mendes, P., Kell, D., & Paton, N. (2011). The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *Journal of Integrative Bioinformatics*, *8*(2), 186. https://doi.org/10.1515/jib-2011-186

Thein, Z. M., Samaranayake, Y. H., & Samaranayake, L. P. (2006). Effect of oral bacteria on growth and survival of Candida albicans biofilms. *Archives of Oral Biology*, *51*(8), 672–680. https://doi.org/10.1016/j.archoralbio.2006.02.005

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

To, K. K. W., Lu, L., Fong, C. H. Y., Wu, A. K. L., Mok, K. Y., Yip, C. C. Y., Ke, Y. H., Sze, K. H., Lau, S. K. P., Hung, I. F. N., & Yuen, K. Y. (2018). Rhinovirus respiratory tract infection in hospitalized adult patients is associated with TH2 response irrespective of

asthma. *Journal of Infection*, *76*(5), 465–474. https://doi.org/10.1016/j.jinf.2018.02.005
Urzua, B., Hermosilla, G., Gamonal, J., Morales-Bozo, I., Canals, M., Barahona, S., Coccola, C., & Cifuentes, V. (2008). Yeast diversity in the oral microbiota of subjects with periodontitis: Candida albicans and Candida dubliniensis colonize the periodontal pockets. *Medical Mycology*, *46*(8), 783–793. https://doi.org/10.1080/13693780802060899
Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M., & Klenk, H. P. (2012). Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0034846
Van Rensburg, J. J., Lin, H., Gao, X., Toh, E., Fortney, K. R., Ellinger, S., Zwickl, B., Janowicz, D. M., Katz, B. P., Nelson, D. E., Dong, Q., & Spinolaa, S. M. (2015). The human skin microbiome associates with the outcome of and is influenced by bacterial infection. *MBio*, *6*(5). https://doi.org/10.1128/mBio.01315-15
Van Wey, A. S., Lovatt, S. J., Roy, N. C., & Shorten, P. R. (2016). Determination of potential metabolic pathways of human intestinal bacteria by modeling growth kinetics from cross-feeding dynamics. *Food Research International*, *88*, 207–216. https://doi.org/10.1016/j.foodres.2016.02.004
Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R. Y., Joossens, M., & Raes, J. (2016). Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*, *65*(1), 57–62. https://doi.org/10.1136/gutjnl-2015-309618
Vautier, S., Drummond, R. A., Redelinghuys, P., Murray, G. I., MacCallum, D. M., & Brown, G. D. (2012). Dectin-1 is not required for controlling Candida albicans colonization of the gastrointestinal tract. *Infection and Immunity*, *80*(12), 4216–4222. https://doi.org/10.1128/IAI.00559-12
Vehkala, M., Shubin, M., Connor, T. R., Thomson, N. R., & Corander, J. (2015). Novel R pipeline for analyzing biolog phenotypic microarray data. *PLoS ONE*, *10*(3), 1–14. https://doi.org/10.1371/journal.pone.0118392
Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine and International Health*, *25*(3), 278–280. https://doi.org/10.1111/tmi.13383
Vila, T., Sultan, A. S., Montelongo-Jauregui, D., & Jabra-Rizk, M. A. (2020). Oral candidiasis: A disease of opportunity. *Journal of Fungi*, *6*(1), 1–28. https://doi.org/10.3390/jof6010015
Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J., & Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME Journal*, *5*(2), 220–230. https://doi.org/10.1038/ismej.2010.118
Wang, B., Ning, Q., Hao, T., Yu, A., & Sun, J. (2016). Reconstruction and analysis of a genome-scale metabolic model for Eriocheir sinensis eyestalks. *Molecular BioSystems*, *12*(1), 246–252. https://doi.org/10.1039/c5mb00571j
Wang, Y., Lupiani, B., Reddy, S. M., Lamont, S. J., & Zhou, H. (2014). RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. *Poultry Science*, *93*(2), 485–493. https://doi.org/10.3382/ps.2013-03557
White, S. J., Rosenbach, A., Lephart, P., Nguyen, D., Benjamin, A., Tzipori, S., Whiteway, M., Mecsas, J., & Kumamoto, C. A. (2007). Self-regulation of Candida albicans population size during GI colonization. *PLoS Pathogens*, *3*(12), 1866–1878. https://doi.org/10.1371/journal.ppat.0030184
Wilson, J. A. C., Prow, N. A., Schroder, W. A., Ellis, J. J., Cumming, H. E., Gearing, L. J.,

Poo, Y. S., Taylor, A., Hertzog, P. J., Di Giallonardo, F., Hueston, L., Le Grand, R., Tang, B., Le, T. T., Gardner, J., Mahalingam, S., Roques, P., Bird, P. I., & Suhrbier, A. (2017). RNA-Seq analysis of chikungunya virus infection and identification of granzyme A as a major promoter of arthritic inflammation. In *PLoS Pathogens* (Vol. 13, Issue 2). https://doi.org/10.1371/journal.ppat.1006155

Wilson, L. S., Reyes, C. M., Stolpman, M., Speckman, J., Allen, K., & Beney, J. (2002). The direct cost and incidence of systemic fungal infections. *Value in Health*, *5*(1), 26–34. https://doi.org/10.1046/j.1524-4733.2002.51108.x

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M. A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., … Querengesser, L. (2007). HMDB: The human metabolome database. *Nucleic Acids Research*, *35*(SUPPL. 1), 521–526. https://doi.org/10.1093/nar/gkl923

Yin, X., Altman, T., Rutherford, E., West, K. A., Wu, Y., Choi, J., Beck, P. L., Kaplan, G. G., Dabbagh, K., DeSantis, T. Z., & Iwai, S. (2020). A Comparative Evaluation of Tools to Predict Metabolite Profiles From Microbiome Sequencing Data. *Frontiers in Microbiology*, *11*(December). https://doi.org/10.3389/fmicb.2020.595910

Zhang, Jianhua, Debets, A. J. M., Verweij, P. E., & Snelders, E. (2021). Azole-resistance development; how the Aspergillus fumigatus lifecycle defines the potential for adaptation. *Journal of Fungi*, *7*(8). https://doi.org/10.3390/jof7080599

Zhang, Jiaqin, & Biswas, I. (2009). A Phenotypic microarray analysis of Streptococcus mutans liaS mutant. *Microbiology (Reading, England)*, *155*(Pt. 1), 61. https://doi.org/10.1099/mic.0.023077-0.A

Zhou, L., Lei, X. H., Bochner, B. R., & Wanner, B. L. (2003). Phenotype MicroArray analysis of Escherichia coli K-12 mutants with deletions of all two-component systems. *Journal of Bacteriology*, *185*(16), 4956–4972. https://doi.org/10.1128/JB.185.16.4956-4972.2003

ACKNOWLEDGMENTS

The long way to completing this work was associated with many beautiful memories and strenuous phases. In this way, I was accompanied and supported by some special people, to whom I would like to express my heartfelt thanks. My special thanks go to Gianni Panagiotou since he made it possible to do my doctorate in the Systems Biology and Bioinformatics (SBI) group. During his supervision, he managed to find time to develop new ideas and approaches to solutions and overcome content-related difficulties. By leading the working group, he created the basis for uncomplicated and helpful interaction within the working group. He also allowed me to participate in numerous workshops and symposia and enhance my scientific knowledge.

I would like to have special gratitude for Sascha Schaeuble for training me in Constraint-based metabolic modeling and paving my way to graduation. Besides being a teacher, he is also a fantastic friend who helped me with many private matters.

I would like to thank the current and former employees of the working groups SBI and Applied Systems Biology (ASB) for the good working atmosphere. In particular, the scientific conversations during the breaks helped me learn more about Bioinformatic concepts and solve some scientific problems. I would also like to thank the excellent collegial and cooperative atmosphere at the Hans-Knöll Institute and our associated partners within the CRC/Transregio 124 FungiNet and all around the world for their precious collaborations in our research projects. I would also like to thank the Graduate School for Microbial Communications (JSMC) for providing great soft skill courses.

I would like to thank my German, international, and Iranian friends for creating beautiful moments during my stay in Germany that made dealing with the difficulties of living abroad way easier.

Last but not least, I would like to thank my family, my parents, and my two lovely sisters, Shima and Maryam. I am grateful for having a good relationship with all my family members, which has reassured me in different life stages. You made this path possible for me not only through your financial support during my studies but also through your loving upbringing, with which you gave me much more than just the bare minimum on my way. Lastly, I would like to thank my lovely fiancé, Farzaneh, for her patience and moral support in our long-distance relationship and her optimism about our future.

DECLARATION OF HONOUR

I hereby declare that I am familiar with the applicable doctoral regulations of the Faculty of Biological Sciences and that I have adhered to these regulations to the best of my knowledge. I have prepared this dissertation independently and only using the aids, data and sources specified. I only received support during my scientific work and for the preparation of the present dissertation text from the named co-authors and persons named in the acknowledgment. I have not sought the help of any outside placement or counseling services. No one has received direct or indirect monetary benefits for work related to the content of the submitted dissertation. The submitted dissertation has not previously been submitted as an examination paper for another academic examination. Specifically, I have not submitted them to any other college to obtain an academic degree.

Jena, 26 April, 2022

_____

Mohammadhassan Mirhakkak Esfahani