FROM DATA TO SCIENCE:
A MULTI-OMICS ANALYSIS OF THE PATHOBIOME

DISSERTATION

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

"DOCTOR RERUM NATURALIUM"

(DR. RER. NAT.)

SUBMITTED TO THE COUNCIL OF

THE FACULTY OF BIOLOGICAL SCIENCES

OF FRIEDRICH SCHILLER UNIVERSITY JENA

BY M. SC. BASTIAN SEELBINDER

BORN ON 19. JANUARY 1993 IN DESSAU

TABLE OF CONTENTS

SUMMARY

Humans represent a complex ecosystem colonized not only by our cells but trillions of other microbes such as bacteria, archaea, fungi, and viruses. This *microbiome* gains increasing interest due to its involvement in human health and disease. While we live in symbiosis with most of these travelers, dysbiosis can lead to the growth of pathogens. Pathobionts are commensal microbes and harmless in healthy individuals until specific circumstances occur. There is increasing interest in studying this pathobiome due to the rise in infections with high mortality rates and stagnant treatment options. Due to the complexity of possible interactions between the host and microbes, studies on microbial interactions are conducted at varying scales. In this thesis, we start to study interactions in small, well-controlled model systems *in vitro* and then at the community level *in vivo*. The key technology used to identify, quantify, and characterize microbes and study host-microbe interactions throughout my studies is whole-genome and transcriptome sequencing.

A standard tool to study interactions of and with the immune system is RNA-Seq. This method quantifies the expression of genes, the transcriptome. To study mechanisms of infection *in vitro*, dendritic cells (DCs) are commonly co-cultured with pathogens. *Aspergillus fumigatus* airway infections are associated with high mortality rates in immunocompromised patients. DCs are recognized as critical immune cells for the detection of and response to *A. fumigatus* infections. Historically, technical limitations allowed the simultaneous sequencing of two different organisms (dual) at most. We expanded on this concept by performing the first triple RNA-Seq ([1], Cell Reports). Thereby, we investigated triangular effects between two pathogens (*A. fumigatus* and *Cytomegalovirus* [CMV]) and DCs. Previous works suggested that CMV may increase the successful invasion of *Aspergillus*, but experimental evidence was lacking. Triple RNA-Seq allowed us to investigate the crosstalk of DC during co-infection of human DC. In contrast to expectations, the response of DCs was a mixture of fungi and viral defense but also included other reactions that were not observed during single infections. In the presence of CMV, *A. fumigatus* stopped expressing many genes, implying a potential to save energy when the virus already challenges DCs. In a follow-up study, we studied the genomes of 300 clinical and environmental isolates of *A. fumigatus* ([2], Nature Microbiology). We observed an underappreciated level of genomic diversity and showed how clinical isolates differ from environmental.

To date, the complex ecosystems present in the human body cannot be replicated in the lab. Hence, we used cultivation-free methods to study human microbiota-host interactions at the community level. We focused on the influence of the host's health status and environmental stressors, *e.g.*, antibiotics, on human gut bacteria and fungi. A common, non-invasive approach to studying the gut microbiota of living organisms is the extraction of nucleic acids such as DNA from host feces. Such samples serve as a proxy for the microbiome of the lower gastrointestinal tract. Gut bacteria are the most abundant and intensely studied part of the microbiome, but fungi were mostly neglected despite their possible roles in health and disease. In four studies, we performed community-level analyses of both bacteria and fungi concomitantly, with the focus on evaluating the risk of severe infections.

One major risk factor for overgrowth of opportunistic gut pathogens – such as the commensal fungus *Candida albicans* – is antibiotic drugs. Antibiotics are used to kill or stop the growth of bacterial pathogens. However, bacteria positively associated with the host's health are often targeted as well, which can allow other pathobionts to proliferate in turn. In two studies, we investigated the influence of antibiotics. First, we assessed the temporal dynamics of bacteria-fungal interactions in healthy subjects up to 90 days after treatment ([3], Microbiome). While some level of recovery was observed in the bacterial community, the effects on the fungal community seemed more stochastic and lasting. *Candida* abundance also increased shortly after treatment but was effectively inhibited later on. Overall, we describe several bacteria with the potential to inhibit or promote *Candida albicans in vivo*. In a follow-up study, we compared healthy against critically ill patients with or without antibiotic treatment ([4], Gut Microbes). Again, we found *Candida* to increase in abundance under antibiotic treatment in critically ill patients. More importantly, we found that antibiotic therapy in critically ill patients leads to an infection vulnerable microbiome composition characterized by extremely low levels of short-chain fatty acids. In the 5th study that focused on human infections ([5], *in preparation*), I studied the microbiome differences contributing to varying levels of *Candida* abundance in lung cancer patients undergoing anti-cancer antibody therapy. Using the knowledge acquired from the previous studies, we found a non-trivial connection between decreased gut anaerobes, increased lactate production by lactic acid bacteria, and high *Candida* abundance levels. A key component linking all the above was oxygen availability within the gut lumen. Our findings challenge the current perception that *Lactobacilli* and lactate always inhibit *Candida* species' overgrowth in the human gut.

In the last manuscript of my thesis ([6], PLoS One), I studied the gut microbiota associations with lung function recovery one year after tumor resection. Longitudinal data revealed associations between specific gut bacteria, fungi, and their metabolic pathways with the recovery of lung functions. Interestingly, an increase in VO_2 coincides with an increase in certain species and the GABA shunt pathway, suggesting that treatment outcomes might improve by enriching butyrate-producing species. Overall, our data suggested a link between loss of anaerobes and tumor recurrence. We contribute evidence to the hypothesis that anaerobes have beneficial effects on positive treatment outcomes.

In summary, while an extensive body of work has focused on understanding the virulence factors of common pathogens, such as *Aspergillus* and *Candida* species, very little work has been done on understanding the interplay of those pathogens with the host's symbionts or other pathogens at the start of my Ph.D. In my Ph.D. project, I used next-generation sequencing, advanced statistical approaches, and machine learning to significantly expanded our knowledge of the life of pathogens from an ecological point of view.

ZUSAMMENFASSUNG

Der Mensch stellt ein komplexes Ökosystem dar, das nicht nur von unseren Zellen, sondern auch von Billionen anderer Mikroben (Bakterien, Archaeen, Pilzen, und Viren) besiedelt wird. Dieses so genannte "Mikrobiom" gewinnt aufgrund seiner Beteiligung an der menschlichen Gesundheit und Krankheiten zunehmend an Interesse. Mit den meisten dieser Mikroben leben wir in Symbiose. Jedoch kann eine Dysbiose zum Wachstum von Krankheitserregern führen. Diese „Pathobionten" sind kommensale Mikroben, die für gesunde Menschen harmlos sind, bis bestimmte Umstände eintreten. Das Interesse diesen „Pathobiomen" nimmt durch die Zunahme von Infektionen mit hoher Sterblichkeitsrate und stagnierenden Behandlungsmöglichkeiten zu. Aufgrund der Komplexität der Interaktionen zwischen Wirt und Mikroben erfolgt die Erforschung mikrobieller Interaktionen auf verschiedenen Skalen. In der vorliegenden Dissertation beginnen wir mit der Untersuchung von Interaktionen in kleinen, gut kontrollierten Modellsystemen *in vitro* und dann auf der Gemeinschaftsebene *in vivo*. Die Schlüsseltechnologie zur Identifizierung, Quantifizierung und Charakterisierung von Mikroben und zur Untersuchung von Wirt-Mikroben-Interaktionen während meiner Studien sind Gesamtgenom und -transkriptom Sequenzierung.

Ein gängiges Instrument zur Untersuchung der Interaktionen von und mit dem Wirts-Immunsystem ist „RNA-Seq". Diese Methode quantifiziert die Expression von Genen, das „Transkriptom". Zur Untersuchung von Infektionen *in vitro* werden häufig dendritische Zellen (DCs) mit Krankheitserregern ko-kultiviert. Atemwegs-infektionen mit *Aspergillus fumigatus* sind bei immungeschwächten Patienten mit einer hohen Sterblichkeitsrate verbunden. DCs gelten als essenziell für die Erkennung von und Reaktion auf *A. fumigatus* Infektionen. In der Vergangenheit war es aus technischen Gründen nicht möglich mehr als zwei verschiedenen Organismen gleichzeitig zu sequenzieren (*dual*). Wir haben dieses Konzept erweitert, indem wir den ersten "*triple* RNA-Seq" durchgeführt haben ([1], *Cell Reports*). Dabei untersuchten wir trilaterale Effekte zwischen zwei Pathogenen (*A. fumigatus* und *Cytomegalovirus* [CMV]) und DCs. Man nahm bereits an, dass CMV den Erfolg von *A. fumigatus* Infektionen erhöhen könnte, aber experimentelle Beweise fehlten. Mit Hilfe des triple RNA-Seq konnten wir den Crosstalk von DCs während einer Koinfektion von menschlichen DCs untersuchen. Die Reaktion der DCs stellte sich als eine Mischung aus Pilz- und Virusabwehr und zusätzlichen Reaktionen heraus, die bei Einzelinfektionen nicht beobachtet wurden. In Gegenwart von CMV hat *A. fumigatus* die Expression einiger Gene eingestellt. Dies deutet darauf hin, dass der Pilz Energie sparen kann, wenn CMV die DCs attackieren. In einer Folgestudie haben wir die Genome von 300 klinischen und Umweltisolaten von *A. fumigatus* untersucht ([2], *Nature Microbiology*). Wir fanden ein unterschätztes Ausmaß an genomischer Vielfalt vor und zeigen, inwieweit sich klinische von Umweltisolaten unterscheiden.

Bislang lassen sich die komplexen Ökosysteme des menschlichen Körpers nicht im Labor nachbilden. Daher haben wir kultivierungsfreie Methoden verwendet, um die Inter-aktionen zwischen Mikrobiota und dem menschlichen Wirt auf der Gemeinschaftsstufe zu untersuchen. Wir konzentrierten uns dabei auf den Einfluss (a) des Gesundheitszustands des Wirts und (b) der Umweltstressoren, z. B. Antibiotika, auf das menschliche Darm-mikrobiom. Dabei quantifizierten wir die Nukleinsäuren, die wir aus Fäkalien extrahieren konnten. Die darin enthaltenen Mikroben repräsentieren das Mikrobiom des unteren gastro-intestinal Trakts. Bakterien sind der am häufigsten vorkommende und am meisten unter-suchte Teil des Darmmikrobioms, während Pilze trotz ihrer möglichen Rolle für Gesund-

heit und Krankheit weitgehend ignoriert wurden. In drei Studien haben wir Bakterien- und Pilzgemeinschaften analysiert, wobei der Schwerpunkt stets auf der Risikobewertung für schwere Infektionen lag. Ein Hauptrisikofaktor für das übermäßige Wachstum von opportunistischen Darmpathogenen - wie etwa dem kommensalen Pilz *Candida albicans* - sind Antibiotika. Antibiotika werden eingesetzt, um das Wachstum von bakteriellen Krankheitserregern zu bekämpfen. Häufig werden jedoch auch Bakterien attackiert, die mit der Gesundheit des Wirts in Verbindung gebracht werden, was wiederum die Vermehrung anderer Pathobionten ermöglichen kann. In zwei Studien haben wir daher den Einfluss von Antibiotika untersucht.

Zuerst untersuchten wir die zeitliche Dynamik der Interaktionen zwischen Bakterien und Pilzen bei gesunden Probanden bis zu 90 Tage nach der Behandlung ([3], *Microbiome*). Während sich die bakterielle Gemeinschaft in gewissem Umfang erholte, schienen die Auswirkungen auf die Pilzgemeinschaft eher stochastisch und dauerhaft zu sein. Die Abundanz von *Candida* nahm für kurze Zeit nach der Behandlung zu. Wir identifizierten mehrere Bakterien, die das Potenzial haben, das *Candida* Wachstum *in vivo* zu hemmen oder zu fördern. In einer Folgestudie verglichen wir gesunde mit schwerkranken Patienten mit und ohne Antibiotikabehandlung ([4], *Gut Microbes*). Erneut stellten wir fest, dass die Häufigkeit von *Candida* nach Antibiotikabehandlung bei kranken Patienten zunahm. Insbesondere fanden wir heraus, dass die Antibiotikabehandlung bei schwerkranken Patienten zu einer "infektionsanfälligen" Mikrobiomzusammensetzung führt, die durch extrem niedrige Mengen an kurzkettigen Fettsäuren gekennzeichnet war. In der fünften Studie über Infektionen im Menschen (Erstautor, *in Vorbereitung*) [5] wurden Lungenkrebspatienten mit Immunotherapie behandelt. Dabei untersuchte ich welche Unterschiede im Mikrobiom die verschiedenen Mengen an Darm *Candida* erklären. Unter Verwendung der Erkenntnisse der vorangegangenen Studien fanden wir einen wichtigen Zusammenhang zwischen der Verringerung von Anaeroben im Darm, erhöhten Laktat produzierenden Milchsäurebakterien, und erhöhter *Candida* Menge. Eine Schlüsselkomponente, die all diese Faktoren miteinander verbindet, ist die Menge an Sauerstoff im Darmlumen. Die Ergebnisse meiner letzten Studie stellen die gängige Auffassung in Frage, dass *Laktobazillen* und Laktat das Überwachstum von *Candida* spp. im menschlichen Darm grundsätzlich verringern.

Im letzten Manuskript meiner Dissertation (Erstautor, *PLoS One*) [6] untersuchte ich den Zusammenhang zwischen Darmmikrobiota und der Wiederherstellung der Lungenfunktion von Patienten einem Jahr nach Tumorresektion. Longitudinaldaten zeigten Zusammenhänge zwischen Darmbakterien, -pilzen, und deren Stoffwechselwegen, mit der Lungenfunktion. Interessanterweise ging ein Anstieg in VO_2 mit einem Anstieg bestimmter Spezies und des "GABA-Shunt"-Stoffwechselwegs einher, was darauf hindeutete, dass sich die Behandlung durch die Anreicherung von Butyrate-produzierender Spezies verbessern könnten. Unsere Daten deuten auf einen Zusammenhang zwischen dem Verlust von Anaeroben und dem Wiederauftreten von Tumoren hin, was die Hypothese einer positiven Wirkung von Anaeroben auf positive Behandlungsergebnisse untermauert.

Zusammenfassend lässt sich sagen, dass sich zwar zahlreiche Arbeiten auf das Verständnis der Virulenzfaktoren häufiger Krankheitserreger wie *Aspergillus*- und *Candida* Spezies konzentriert haben, dass aber zu Beginn meiner Doktorarbeit nur sehr wenig über das Zusammenspiel dieser Erreger mit den Symbionten des Wirts oder anderen Krankheitserregern bekannt war. In meiner Promotion habe ich Next-Generation Sequenzierung, fortschrittliche statistische Ansätze, und maschinelles Lernen eingesetzt, um unser Wissen über das Leben von Erregern aus ökologischer Sicht erheblich zu erweitern.

ABBREVIATIONS

Abbreviation	Full Name
ASV	amplicon sequencing variant
CAG	co-abundance cluster of genes
CLR	centered log-ratio
CMV	Cytomegalovirus
CNV	copy number variation
DA	differentially abundant
DC	dendritic cell
DNA	deoxynucleic acid
EC	Enzyme Commission
FDR	false-discovery rate
FEAST	Fast Expectation-maximization for microbial Source Tracking
FGS	first generation sequencing
GI	gastrointestinal system
GO	Gene Ontology
GRiD	Growth Rate InDex
GWAS	genome-wide association study
IA	Invasive Aspergillosis
ITS	internal transcribed spacers
KEGG	Kyoto Encyclopedia of Genes and Genomes
MGS	metagenomic species
NCBI	National Centre for Biotechnology Information
OTU	orthologous taxonomic units
PAMP	pathogen-associated molecular patterns
PCR	polymerase chain reaction
PERMANOVA	permutational analysis of variance
Pfam	protein family database
PRR	pattern recognition receptor
RNA	ribonucleic acid
SCFA	Short-chain fatty acid
SGS	second generation sequencing
SNP	single nucleotide polymorphism
SNV	Single-nucleotide variant
TGS	third generation sequencing
TSS	total-sum scaling
UniProt	universal database of proteins
UniRef	database of non-redundant proteins; based on UniProt
WMS	whole metagenome sequencing
WTS	whole meta-transcriptome sequencing

I. INTRODUCTION

1. The Human Host and its Bugs – The Holobiont

Microbes represent the most successful form of life. They are omnipresent in every environment and adapted to any niche, including water, soil, plants, and animals. Many microbes co-evolved with their host, thereby forming synergies by complementing metabolism and protecting each other from potentially dangerous pathogens. Formally, the term *microbiota* refers to the collection of all bacteria, fungi, archaea, and viruses within a habitat. *Microbiome* refers to their genes [7]. Microbial-host interactions are complex. Over the last two decades, research has found many associations between health, disease states, and microbiota. Many human-associated microbes are commensal, but some of them – *opportunistic pathogens* – may attack or invade their host under – often unknown – circumstances such as weakened host-immune defense. It was even suggested that every human disease could be related to the microbiome [8]. While this is likely an exaggeration, microbes are understood to be an integral part of our bodies - an additional organ. In the following sections, I will introduce terminologies and concepts used to characterize microbial communities.

1.1.   Essentials in Ecology Research

Ecology is the study of relationships between living organisms. While many concepts evolved from observing and counting visible organisms in relatively large ecosystems, many concepts were proven useful in microbial research.
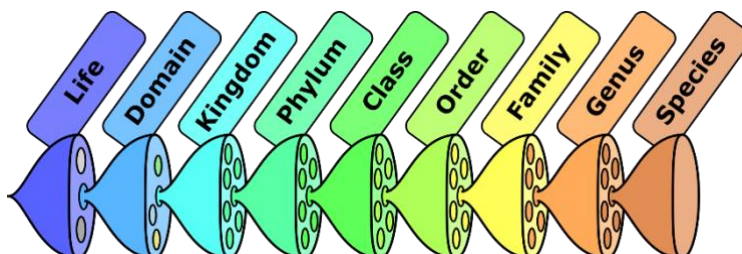
Biological Taxonomy

In biology, taxonomy is a scheme used to classify organisms into coherent units called "*taxa*" [9]. It is mostly based on morphological, physical, or genetic properties. "*Phylogeny*" is the evolutionary history of species assuming common ancestors of organisms and is commonly used to create taxonomies [9]. There are seven generally agreed taxonomic levels from broad to increasing precision: Kingdom, Phylum, Class, Order, Family, Genus, and Species (**Figure 1**). The catalogue of Life [9] currently separates Kingdoms into Bacteria, Archaea, Protozoa, Chromista, Plantae, Fungi and Animalia. Viruses are often considered as separate kingdom because viruses are not cells and may not share a common ancestor with cellular life.

Ecological Diversity

Diversity aims to describe the complexity of ecosystems. Diversity indices are often used in metagenomic research to compare habitats and as basic indicators of biotic and dysbiotic habitats. Diversity is defined in three general levels (**Figure 2**) [10]. Because diversity indices give complementary information, we chose them on a case-to-case basis. I used all the following indices in my work (**manuscripts III-VI**).

Alpha diversity can be measured by the number of distinct species detected (*richness*), by the distribution of taxon abundances (*Shannon*- and *Simpson*-indices), and by phylogeny-aware indices (*Faith's PD*) [10]. Assuming closely related organisms have similar functions or roles, phylogeny-aware diversity indices give similar diversity scores to taxa profiles with distinct but phylogenetically close organisms. Beta-diversity is commonly estimated as Bray -Curtis dissimilarity, even though its application to

sequencing data has been heavily criticized [11]. I expand on some technical challenges in **section 3.1**. One appropriate alternative is the log-ratio-based Aitchison index which compensates for compositionality bias [11]. Another index is UniFrac, which integrates phylogenetic relationships of taxa [11,12].

Figure 1. Taxonomic ranks of life from low resolution (left) to high resolution (right). Creative Commons license.
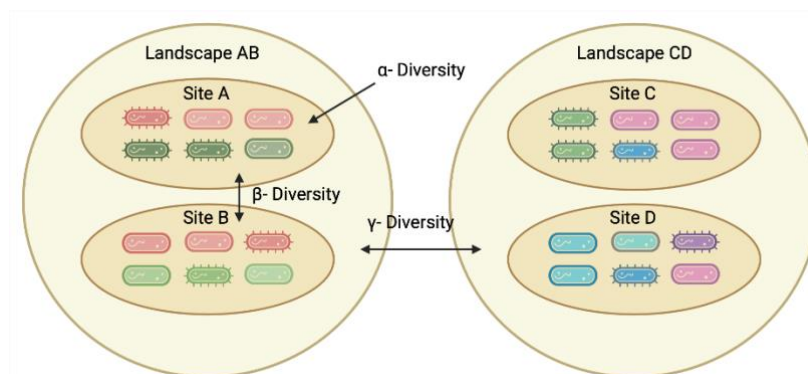
Figure 2. Three levels of ecological diversity. α-diversity describes diversity within a site (sample, habitat), β-diversity describes the diversity between two sites, and γ-diversity describes the diversity between two landscapes. Created with BioRender.com.

1.2. The Human Microbiome

The human body provides a nutrient-rich habitat for various microorganisms, many of which are essential for the host's homeostasis. The typical human body of a 70kg male is composed of $3 \cdot 10^{13}$ somatic cells and, according to current estimation, an extra $3.8 \cdot 10^{13}$ microbial cells [13]. While the human genome encodes roughly 20,000 protein-coding genes, the bacterial metagenome carries at least 100 times more genes and is therefore coined the "second genome" [14]. The gastrointestinal tract contains over 90% of these microbial cells, but there are still roughly 10^{12} on the skin and a modest amount in the lungs [13,15]. Together with these microbes, we form a *holobiont*, whose health depends on symbiotic interactions between the residents – microbes – and host [8].

Microbes are masters at sensing and responding to their environment. Host-associated microbes cover diverse roles. On epithelial surfaces, they act as physical barriers against foreign pathogens [16–18], degrade and modify compounds, and have roles in the maturation and function of the immune system. Resident strains have evolved mechanisms to break down antimicrobial peptides produced by the human host, utilize low-energy sources such as free lipids for growth, and improve adherence to specific surfaces. Low-

energy compounds are frequently broken down in ways that benefit us: especially gut microbiota convert indigestible food into nutrients, release vitamins, peptides, and immune regulatory compounds, to name a few. Microbes also create complex natural products such as macrolides and polyketides, many of which have antimicrobial and immunomodulatory activities [19].

The human gut represents a unique, complex, and heterogeneous environment. Because most of the manuscripts in this thesis are studies on the gut microbiota of the large intestine, I will briefly introduce the structure of the gastrointestinal tract and the conditions microbes encounter therein.
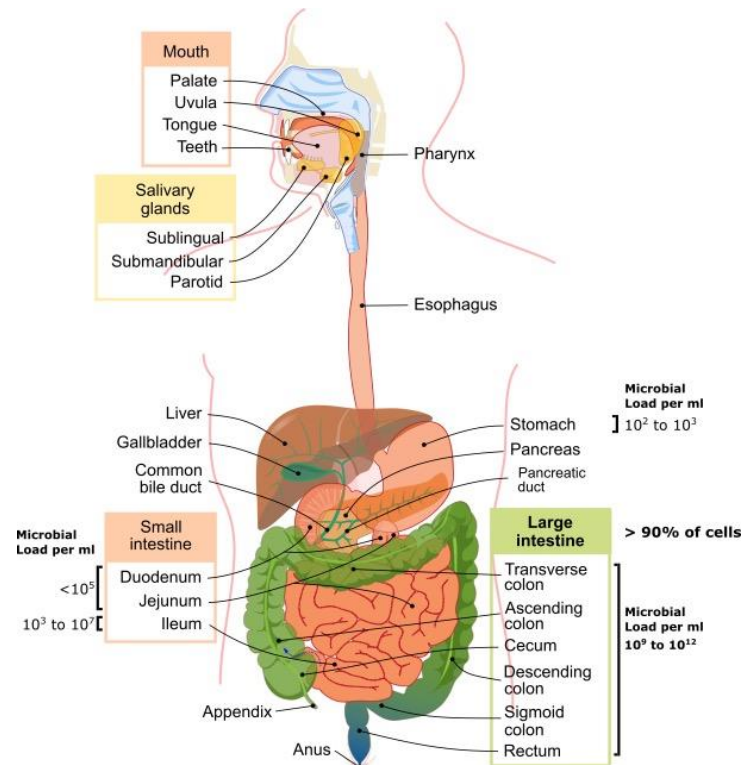
Figure 3. Basic anatomy of the human gastrointestinal system. The majority of microbial cells (>90%) is found in the large intestine [18]. Creative Commons license, Marina Ruiz.

The Human Gastrointestinal Tract

The human gastrointestinal system (GI) is a complex organ with many different sites and functions. While the overall role of the GI is the extraction of nutrients and removal of indigestible waste, it also represents a unique environment for microbes to thrive and the largest compartment of the immune system [18]. Thus, it is constantly exposed to antigens, commensal bacteria, and pathogens. The GI tract (**Figure 3**) begins in the mouth, with the esophagus carrying fluids and food to the stomach. The lower GI tract starts with the small intestine, connecting the stomach outlet and the large intestine. The small and large intestines form a tube of column-shaped epithelial cells and end in the anus.

The small intestine plays a role in dietary component digestion and absorption and the synthesis of antimicrobial peptides [18]. The large intestine, on the other hand, performs little digestion and is physiologically primarily responsible for absorbing water and eliminating undigested food [18]. Only a few immune cells are enriched, including macrophages, IgA+ plasma cells, and Treg cells [18]. However, because the gut's distal regions are also home to billions of commensal human bacteria, it's an important target for

research into host-microbe interactions [13,18]. The lower GI can directly influence microbes via crosstalk between colonocytes, immune cells, antimicrobial peptides, exchange of metabolites, limiting carbon sources, and gut movement. Analogously, microbes can interact with the GI by exchanging metabolites and binding mucosal surfaces.

Proximal regions of the GI are still challenging to access non-invasively. Instead, fecal samples were used as proxies for most studies on the large intestinal microbiota of humans and mice [10,20] and in the microbiome studies presented in this thesis (**manuscripts III-VI**).

The Human Gut Bacteriome

The bacteriome is the most abundant and frequently studied part of the gut microbiome [21,22]. Bacterial composition varies over the different parts of the GI, but most notably between upper and lower GI tracts [23]. Molecular techniques estimate 500-1000 different species in the GI of healthy adults, most of which are harmless [24,25]. Gut microbes have major roles in maintaining host homeostasis [26], extracting nutrients and energy from food [25,27], host immune function [22,28] and defense against pathobionts [26,29]. They also have important roles in several diseases such as obesity [30,31], malnutrition, inflammatory bowel disease [22,32], metabolic disorders [22], type 2 diabetes [30,33], neurological disorders and cancer treatment [25,34,35].

Despite large-scale projects such as the Human Microbiome Project [7,36], defining a "healthy" microbiome remains a challenging and ongoing topic. In healthy adults, 98% of bacterial species are from the phyla *Firmicutes*, *Bacteroides*, *Actinobacteria,* and *Proteobacteria* [26]. However, at increased resolution, the composition of gut communities and functions is highly personalized and dynamic over time [26]. An ensemble of factors with significant influence on the composition was identified, including age, gender, diet, alcohol and lifestyle, hyper immunity, and host genetics [26]. Furthermore, metabolic and structural patterns consistent across healthy human individuals were identified [37,38]. These structural patterns are often described in terms of *stability, resilience, perturbation, resistance, redundancy*, and differences in the abundance of specific taxonomic clades.

The Role of Gut Strict Anaerobes

While microbes produce a variety of compounds, one group of microbes consistently showed differences in abundance in the manuscripts in this thesis: *obligate anaerobes*[18]. These microbes are an exciting example of co-evolution between microbes and their host. Obligate anaerobes – microbes that cannot survive in the presence of oxygen – are the main colonizers of the large intestine [18]. Many anaerobes ferment non-digestible fiber from food, thereby creating small molecules called short-chain fatty acids (SCFA), such as acetate, propionate, lactate, and butyrate [22,39]. SCFAs are abundant in stool samples [40] and have antimicrobial, immunoregulatory, and homeostasis properties [22]. In addition, some SCFAs are critical to maintaining gut hypoxia: Colonocytes consume microbial butyrate as a major energy source [40] to perform aerobic respiration [41], using up most of the cells' oxygen. The resulting hypoxia, in turn, selects for obligate anaerobe bacteria, creating a positive feedback look selecting for SCFA-producing bacteria. However, some of these compounds can also be harmful in dysbiotic communities. For example, in autistic spectrum disorder, subjects show *increased* bacterial diversity alongside higher concentrations of SCFAs and ammonia [42].

The Human Gut Mycobiome

Fungi are the most abundant group of eukaryotes in the gut [43]. However, gut *mycobiota* remain understudied due to technical challenges, including their lower abundance, differences in the cell wall and DNA extraction, larger and more complex genomes, and lack of reference genomes required for comparative genomics [44]. In addition, while fungal cells and genomes are much larger than bacterial cells, fungi are vastly outnumbered by bacteria in stool samples (10^5 compared to 10^{11} [45,46]). I address some solutions to these issues in **section 2.3**. These notwithstanding, gut fungi were often associated with disease development [47]. Therefore, mycobiota were an important part of microbiomes to study during my Ph.D.

Mycobiota composition is even more dynamic than bacterial. Still, many fungal genera are commonly detected: around 50 fungal genera are typically reported in stool samples, although only about 10 explain most of the community compositions [48,49]. Commonly detected genera include *Saccharomyces, Candida, Malassezia, Cryptococcus, Penicillium, Fusarium,* and *Yarrowia* [47,49–51]. What factors and how they affect fungal communities are not well established. A recent study reported significant, systematic differences by host age and diet [47]. Interestingly, dairy consumption was positively correlated with *Saccharomyces* and negatively with *Candida*.

Little is known about the positive health impacts of fungi. *Saccharomyces* is generally considered harmless to humans, and some selected species (*S. cerevisiae*) are even used as probiotics. In addition, fungi are likely involved in conditioning the immune system, as evident by their involvement in inducing and preventing allergies [49]. Therefore, my work aimed to identify affecters of mycobiome diversity and composition.

1.3.   Pathobiosis

Environmental stressors can drive some commensal microbes from mutualistic or harmless states into pathogenic states [52]. Such microbes are called "*opportunistic pathogens*" [53]. The prime example of an opportunistic pathogen is *Escherichia coli*, one of the most prevalent gut bacteria [54]. *E. coli* is a facultative anaerobic bacterium. While prolonged exposure to oxygen is toxic, most strains can utilize aerobic respiration for short amounts of time, which can help them outcompete other microbes [54]. Incidentally, an expansion of facultative anaerobes in the gut was frequently observed in inflammatory diseases and dysbiotic communities [55–57]. Another concept is the induction or increase of virulence by horizontal gene transfer between microbes [53]. Thereby, commensal strains may turn pathogenic *on accident*, or already virulent strains spread their virulence factors. The exact mechanisms of pathogenesis are often complex and likely context-dependent (temporally and locally) [53]. It was thus suggested to use the term *pathobiome*, which describes a set of circumstances, microorganisms, and their interactions, that result in a diseased state of the host [8]. Pathobiome is often defined or identified by dysbiotic gut communities.

Markers of Dysbiosis

Dysbiosis of bacterial communities in the large intestine is not well-defined. Due to the complexity and dynamics of microbial communities, microbiomes from healthy subjects are typically used to describe a *balanced* or *biotic* microbiome [58,59]. In this sense, dysbiosis is identified by microbial communities sowing significant structural differences from those of healthy individuals [58,59]. This is possible because microbiomes of healthy individuals often demonstrate higher similarity than dysbiotic ones, following the idea: "There are many ways to be sick, but only a few ways to stay healthy". Typical dysbiosis markers include low taxonomic and functional diversity, reduced abundance of strict

anaerobes, presence of pathogens in host blood due to loss in the gut barrier integrity, or substantial changes in metabolic states [53,58].

Perturbation

Perturbation is an external event that causes a distinct selective pressure on the ecosystem [60]. Microbiota demonstrate a certain capacity to revert perturbations (*resilience*). However, resilience works in both directions: perturbations can result in resilient unhealthy states, some of which are associated with diseases such as obesity [60]. Perturbations were identified in most diseases linked with microbiota [19]. Many factors influence microbial composition in potentially harmful ways, including age, diet, treatment with antimicrobial agents, pharmaceutical proton-pump inhibitor drugs, xenobiotics, environmental toxicants, pet exposure, and birth delivery mode [19].

A powerful source of perturbation is *antibiotics*, a broad group of compounds with typically antibacterial effects [60]. Some diseases, such as antibiotic-induced colitis or inflammatory bowel disease, are directly linked to antibiotic treatment [19]. Even though bacteria are the main target of these drugs, there is evidence of a profound influence of gut fungi as well. Some studies on the gut of mice indicated a substantial increase in the colonization of pathogenic fungi such as *Candida* at multiple body sites in the early aftermath of antibiotics treatment [3,4,61–64]. However, how much this applies to humans is unclear and therefore addressed in **manuscripts III-IV**.

The fungal pathobiont C. albicans

C. albicans is a dimorphic, opportunistic, facultative anaerobic fungus [65] and one of the most prevalent in the human gut [66]. While harmless to healthy hosts, it often infects patients with immunocompromised immunity, causing an often lethal disease *systemic candidiasis* [66,67]. In its commensal form, it resides in ball-shaped *yeast* cells, but on contact with surfaces, it forms tube-shaped *hyphae* [65]. This morphological switch is considered a key virulence factor [65,66]. While the circumstances enhancing its virulence and allowing the translocation of the fungus through the epithelial lumen are not entirely clear, the intestinal microbiota are considered a key regulator to prevent fungal overgrowth [66].

1.4. Fungal Infection

The human body represents an attractive environment for many microbes, both beneficial and detrimental. While human pathogenic fungi are an increasing cause of dangerous diseases [68], the conditions that allow fungal infection are complex and not fully understood. Host immune dysfunction is a key requirement for fungal infection [69], and the bacteriome is likely to have substantial roles in promoting and inhibiting opportunistic fungi. Bacteria can limit fungal colonization by producing antifungal compounds, competing for nutrients, cellular contract, chemotaxis, or physiochemical changes to the local environment [70,71]. Likewise, bacteria may promote fungal growth by exchanging metabolites, creating biofilms, and adherence to the pathogen [70]. But findings for the latter are sparse due to the difficulty of studying microbes directly in their native environment.

Host Immunity

The host can adopt three main strategies to deal with microbes: *avoidance*, *resistance*, and *tolerance*. A simplified view of the human immune system describes three levels of

complementary systems: (**1**) anatomic and physiological barriers, (**2**) innate immunity, and (**3**) adaptive immune system [72].

Physiological barriers include intact skin, low stomach pH, and bacteriolytic enzymes in several body fluids [72]. In the intestine, a set of diverse microbes residing on the mucosal surface act as an integral part of host defense. On breaching this first defense, microbes encounter cells of the innate immune system [72]. These cells have invariant receptors that detect conserved components presented by pathogens [72]. The system lacks specificity but responds fast and will activate pro-inflammatory responses, including mobilizing the more specific but slower adaptive immune system [72]. One type of innate cell (dendritic cell) was of particular interest in my studies.

The innate immune system consists of a group of specialized, *"professional"* cells that perform phagocytosis with high efficiency [73]: Monocytes that differentiate into macrophages and dendritic cells (DC), granulocytes (neutrophils, eosinophils, basophils, mast cells) and specific lymphocytes (natural killer cells) [73,74]. Pattern recognition receptors (PRRs) in professional cells recognize pathogen-associated molecular patterns (PAMPs) that are present in microbial cell walls and foreign to the host [75]. While the overall mechanism is understood, the recognition mechanisms of PAMPs are mostly unknown [75].

Aspergillus fumigatus infections

On a daily basis, every human inhales thousands of airborne conidia of the fungus *Aspergillus fumigatus* [75]. While harmless to healthy human individuals, *A. fumigatus* frequently infects the respiratory systems of immunocompromised patients, especially after treatment with immune-suppressant drugs used during stem-cell or organ transplantation [75]. Upon infection, *A. fumigatus* causes *Invasive aspergillosis* (IA), a life-threatening disease associated with high mortality rates and insufficient treatment options [76]. Anti-fungal prophylaxis, including azole drug treatment, works only in a fraction of patients, and drug-resistant *A. fumigatus* strains were already detected [76,77]. In one of our studies, we investigated the potential origins of azole resistance across many *A. fumigatus* isolates (**manuscript II**).

A. fumigatus is a dimorphic fungus, *i.e.*, it exists in non-pathogenic states (conidia) but can also form hyphae (germ tubes) [75]. Infections start by conidia reaching lung surfaces such as alveoli which are protected by innate immune cells [75]. Without successful clearance, the fungus germinates hyphae and forms a colony. It degrades the surrounding tissue to acquire nutrients, which can obstruct lung function [75]. While innate immune cells are generally poor at recognizing conidia [75], antigen-presenting cells, especially DCs, recognize hyphae well [75]. Upon detection of the pathogen, DCs release pro- and anti-inflammatory cytokines, present antigens of the detected pathogens to T cells, and release cytokines to promote B cell activation [74,75]. Consequently, DCs act as an essential regulator of innate and adaptive immune systems [74,75].

The virulence of *A. fumigatus* may further be enhanced during simultaneous infection by other microbes or viruses [78]. To address the latter, I conducted a study on the transcriptome changes of a human herpes virus, *A. fumigatus*, and human dendritic cells during co-infection by both pathogens (**manuscript I**).

2. Quantification of Host and Microbes by Genome Sequencing

In the past, the study of microbiota was limited to the cultivation of microbes in Petri dishes. But many, if not most, gut microbes are strict anaerobes and challenging to cultivate [44]. But around the year 2000, things began to change. Advancements in genomics technology allowed faster and cheaper sequencing of large portions of genomic material from biological samples, resulting in a new approach: Comparative Genomics. Cultivation methods were replaced mainly by DNA sequencing of marker genes or entire metagenomes, revealing an enormous number of new organisms, microbial functions, and biodiversity [79]. The following section will describe sequencing technology and its use in genomic research, as used throughout my studies.

2.1. Next-Generation Sequencing

Genomics evolved around the concept of a *reference genome*, which is a roadmap for a typical individual of a species [80]. In its most simple form, it is a consensus nucleotide sequence and an annotation of functional regions [81]. Such references are used together with genomic sequences acquired from new samples. Many companies offer kits and services to extract nucleic acids (RNA or DNA) from biological samples such as stool, blood, and cell cultures and perform sequencing. However, it is still essential to understand how this quantification works, as it affects downstream analyses. For all current sequencing technologies, cells are first isolated from the sample, then lysed, and their nucleic acids extracted.

1st-generation sequencing (FGS) methods are based on Sanger sequencing [82]. Long nucleic acid molecules were broken down into many random fragments (*shotgun*). For RNA sequencing (RNA-Seq), RNA molecules were converted before or after fragmentation into complementary DNA (cDNA) by a reverse transcriptase enzyme [83]. DNA or cDNA were then amplified (cloned) *in vivo* using a host bacterium (*Escherichia coli*) or *in vitro* using polymerase chain reaction (PCR) to acquire enough material for the sequencing [82–84]. The final readout of cDNA sequences are *reads*, a term used for DNA readouts from newer technology as well.

Next-generation, or 2nd-generation sequencing (SGS), methods reduced cost and increased speed of sequencing. This was achieved by (**a**) combining amplification and sequencing and (**b**) the parallel sequencing of thousands to millions of *colonies* of single-DNA fragments [84]. Compared to FGS, the sequencing output increased by five orders of magnitude [85]. While the latest FGS technology produces sequences of up to 1000nt [86], modern Illumina machines can sequence fragments of up to 600nt [87]. While there are different commercial technologies, Illumina platforms became the *de facto* standard for SGS due to high–throughput, low-error readout, relatively low cost, and excellent protocols for DNA extraction and filtering [87].

Next-next-generation sequencing - 3rd generation - will enable the sequencing of *single* molecules of *even longer* length and *without* amplification [84]. Current technology allows sequencing of fragments up to 13kb [88] and some even 150kb [86]. Furthermore, RNA does not need to be converted to cDNA, removing some of the biased caused by the conversion process. However, while these technologies offer some advantages (which I discuss **on page 203**), they have higher base-calling error rates [86]. Especially technologies aiming for over 30kb may exhibit up to 20% [86]. This elevated error rate and the high volume of genomic material required for sequencing made this technology impractical for many quantitative analyses (*e.g.*, transcriptomics).

The choice of sequencing machines depends on the main objectives of the study. The following section will introduce the most important methods in the downstream processing of SGS data to quantify genes and microbes. A general overview is given in **Figure 4**. Again, only DNA fragments of close to 600nt can be sequenced with high quality [89], exceeding the average length of protein-coding genes and genomes. This major limitation was dealt with in two general ways: (**a**) extraction and amplification of small, specific genomic regions – *amplicons* or (**b**) deep metagenome sequencing with binning techniques.

Quality control, Genome assembly and Read mapping

The first step in processing read libraries involved *trimming* low-quality read segments and the removal of low-quality reads and contaminants (*e.g.*, PCR primers or foreign DNA) [90]. A position-wise quality score (*Phred Score*) indicates read quality [87]. Commonly, base-call error rates of 1 in 100bp (Q20; 99%) to 1 in 1000bp (Q30; 99.9%) are chosen [87]. I performed trimming-related tasks with the program `Trimmomatic` in all manuscripts included in this thesis [90]. The next typical steps tasks involve the reconstruction of genomes (*assembly*), identification of homologous sequences (*alignment*), or quantification of enriched segments (*mapping*).

Reads originating from the same chromosome may be identical, aside from a few mutations, and have significant degrees of overlapping regions. To reconstruct a genome, reads are *assembled* based on these overlapping sequence sections into progressively longer sequences (*contigs*) [84].

Sequence *alignment* is the process of matching two or more sequences to identify regions of high similarity. The alignment tool `BLAST` [91] is often employed to predict the functions of genes (query) by sequence homology and allows low sequence identity to the reference gene (<80%). Alternatively, the aligner `DIAMOND` gained popularity in metagenomic research [92]. `DIAMOND` is 1.000-10.000 times faster than `BLAST+` but has lower sensitivity and is limited to *amino acid* sequences [92]. Notably, the new ultrasensitive mode of `DIAMOND` achieves `BLAST`-level accuracy at an 80-fold faster speed [92].

Sequence *mapping* is a particular type of alignment but involves assigning billions of short sequences (*i.e.*, reads) against a few reference sequences. Mapping is used to quantify genomic regions (*e.g.*, genes), create coverage profiles, or identify short mutations (**section 2.7**). Mapping approaches are high-speed but require a high similarity between reads and reference segments. While RNA-Seq research uses mapping tools such as `HiSat2` with increased speed and accuracy [93] (**manuscript I**), metagenomic studies use older tools such as `BWA` or `Bowtie2` because these programs allow more mismatches between a read and a target region (**manuscripts III-VI**).

2.2.   Single and Dual RNA-Seq

While infection studies involve two or more organisms, transcriptome sequencing was limited to only one organism at a time [94]. Cells from different organisms had to be separated and sequenced independently, leading to substantial biases in gene quantification down the line [94,95]. But one decade ago, laboratory and *in silico* improvements allowed for the first *dual RNA-Seq* experiments of human cells and one pathogenic species, significantly decreasing sequencing costs and bias [96]. So far, mammalian host cells have been challenged with one viral, bacterial, fungal, or eukaryotic pathogen [94]. In this thesis, we go one step further to perform the first "triple RNA-Seq" to study triangular interactions between human cells and two pathogens (**manuscript I**).

2.3.   Fungal ITS Amplicon Sequencing

To date, most microbiome studies focused on characterizing community compositions (**Figure 4**) [97]. A cost-effective, high-throughput, and common approach involves sequencing of marker genes *amplicons*, *i.e.*, genes that can be used to delineate taxa [98]. Standard marker genes are the highly conserved ribosomal RNA (rRNA) genes (16S for bacteria; 18S and 23S for eukaryotes) or the variable regions between them [99]. A common technique to create amplicons uses polymerase chain reaction (PCR) and small DNA template molecules (*primers*) to define and extract the region of interest from a pool of DNA [99].
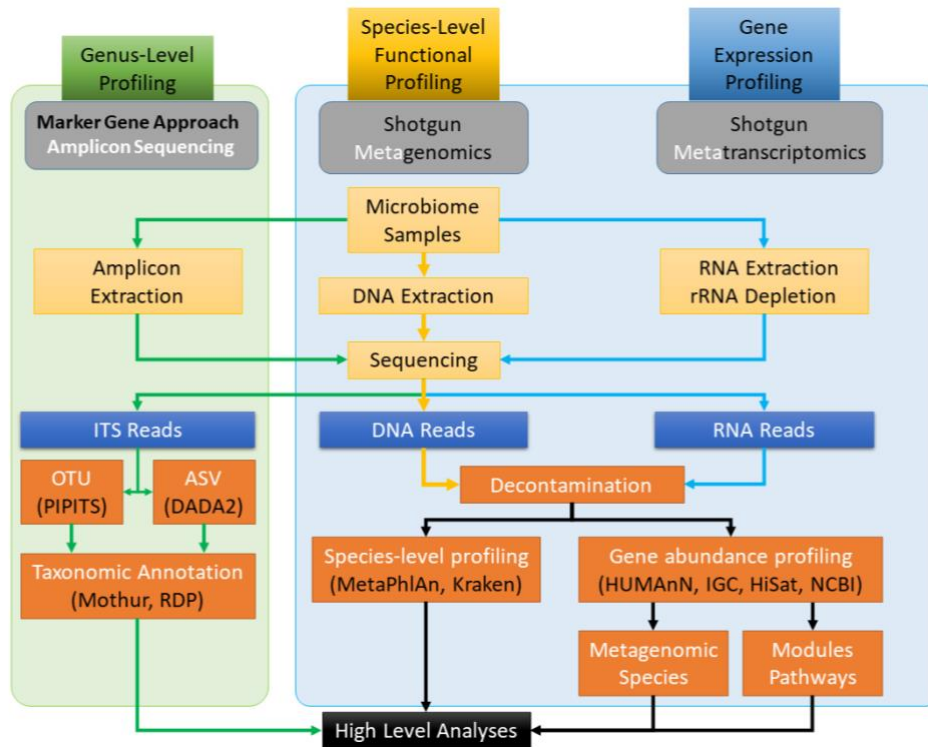
Figure 4. An overview of our microbiome analysis workflows. Amplicon-based approaches are mostly limited to taxonomic profiling. Whole shotgun genome sequencing approaches can be applied to derive bacterial species and functional profiles. Reads are decontaminated *in silico* (*e.g.*, host-reads). Metatranscriptomics requires additional filtering steps pre- and post-sequencing. High-level analyses such as diversity, differential abundance, or co-abundance network analysis are required to gain biological insights. This figure is based on figures from [10,137].
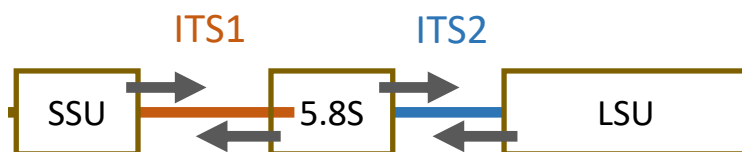
Figure 5. Internal transcript spacer regions (ITS) of the eukaryotic rRNA gene (small subunit – SSU; 5.8S; large subunit - LSU). Each ITS region can be fully sequenced by merging the overlapping regions of paired-end reads.

10

Fungal ITS Sequencing

The eukaryotic 23S rDNA region is over 1500 nt long and exceeds the effective length of SGS technology [44,89,100]. Instead, DNA fragments are extracted based on highly conserved regions within the rDNA gene (5.8S) and sub-sequent variable regions *internal transcribed spacers* (ITS) are sequenced (**Figure 5**) [44,89,99,101]. Thereby, conserved regions are used to capture as many species as possible, while variable regions discriminate taxa [89,102]. Choosing ITS1 or ITS2 is important due to amplification-specific bias across different species [44,103]. Sequencing both regions would be best in theory but is impossible with SGS technology [44,103]. In our studies, we used ITS2 as it is more generic [50,103], has a more consistent length (186 bp on average, but up to 730 bp) [50,103], and captures more diversity of human gut fungi [44].

Next, taxonomic abundance profiles are created from the amplicon reads. The short length of ITS2 fragments results in, on average, sequence overlaps between paired-end reads (**Figure 5**). Reads are merged based on these segments before or during the quantification process [104–106]. Due to mutations and sequencing errors, we expect some divergence between reference ribosomal DNA sequences and amplicons reads. Furthermore, we expect reads from unknown organisms. To resolve these, the following steps involve read clustering, taxonomic assignment of clusters, and cluster quantification. I used two evolved concepts in my work: orthologous taxonomic units (OTUs) or amplicon sequencing variants (ASVs), both of which were used in manuscripts of this thesis.

Amplicon Profiling

In OTU clustering, amplicon sequences of high similarity (*e.g.*, 97%) are clustered together to form a unit (OTU) [89,107]. A representative sequence is chosen for each OTU and used for taxonomic assignment. Additionally, reference sequences of known taxa can be included to improve accuracy and speed-up taxonomic assignment of clusters. These OTU picking procedures are thereby further divided based on the use of the reference sequence collection. In *open*-reference picking, amplicons are first clustered against the reference sequences. Clusters containing a reference sequence inherit the taxonomy of that reference. Amplicons without a hit to the reference are then clustered de novo. In *de novo* picking, the sequence with a median similarity between all cluster sequences is used as representative of the cluster. We used a complete framework for read merging, ITS extraction and open-reference OTU picking called `PIPITS` [105] in two **manuscripts (III, VI)**.

A considerable drawback of OTU clustering is the need for an (often arbitrarily) chosen similarity cut-off of 97% [107,108]. While justified historically, genomes from thousands of species revealed many clades requiring higher (>99%) or lower (<97%) cutoffs for delineation [102,107]. The new approaches amplicon sequencing variants (ASV) and zero-radius OTUs are *denoising* procedures that aim to detect and correct for sequencing errors so that each resulting ASV represents exactly one organism [89]. In contrast to OTU clustering, denoising approaches do not require use-defined cutoffs, offer superior sensitivity and precision, and can be directly compared across studies [108]. We used `DADA2` [104] to estimate ASVs in two **manuscripts (IV, V)**.

Taxonomic Assignment

Taxonomic assignment is performed using ITS-optimized reference sequence databases such as INSDC, Warcup ITS, and UNITE. UNITE [109] is the most favored and was also suggested in a recent review [50]. Different alignment software is proposed in the literature with large variations in prediction accuracy [110]. `BLAST+`, `RDP`, `SINTAX`, and `Mothur` and commonly used [110]. In our studies [3,6] and others [110], `Mothur` often had superior

taxonomic accuracy compared to the other common methods, but there is still considerable room for further improvement [110].

2.4.   Whole Metagenome Sequencing

While amplicon sequencing technology is cheap, it is limited to primarily genus-level resolution and cannot reliably be used to estimate functional information (**Table 1**; **Figure 4**) [89,111–113]. While some tools like PICRUSt [112] offer a rough estimation of the functional potential purely from a bacterial composition, they rely too heavily on assumptions of functions of representative species. A more comprehensive solution uses the entire collection of DNA within a sample, which allows for superior assessment of diversity, functional quantification, growth-rate estimation, accurate species-level quantification, and even strain identification [114]. However, this applies primarily to the study of bacteria. Of note, fungi remain largely undetected in whole shotgun metagenome samples due to low abundance and insufficient sequencing depth [44], making amplicon sequencing a better approach for studies of this kingdom.

Table 1 Pros and Cons of genomic analyses for microbiome research [10].

Methods	Advantages and Use	Disadvantages
Marker-Gene Sequencing	• Cheap and amenable to low-biomass and high-host DNA samples • Many public datasets	• Amplification bias and results vary heavily on primers and amplification targets • Requires *a priori* domain knowledge • Limited to mostly genus-level resolution • Very limited functional resolution
Whole Meta-genome Sequencing (WMS)	• Quantifies genes, functions, species, and strains • Allows population-wide gene *de novo* assembly and mining • Quantifies all kingdoms of life • Allows *in situ* growth rate estimation	• Expensive, laborious, and complex samples preparation • Host-DNA can influence downstream analysis • Identifying dead cells is computationally intense • Inaccuracies in genome assemblies
Whole meta-Transcriptome Sequencing (WTS)	• Identifies active species when paired with metagenomes • Captures microbial activity to treatment exposure the best	• Most complex sample preparation • RNA Degradation leads to strong bias in profiles • Host and microbe rRNA depletion is required • Biases towards organisms with high transcription rates

Formally, the *metagenome* is the collection of all genomes (or genes) present in a sample, including the host. But in practice, host DNA is depleted before and after sequencing to prevent hybrid assemblies, assembly errors, and the detection of spurious microbes [114]. Host-depletion and the removal of other potential contaminants are done *in silico* by filtering out reads that align with high identity to relevant sections of the human reference genome (**Figure 4**) [114,115]. Metagenomes are acquired through *whole meta-genome sequencing* (WMS) [10]. Analogously, the *metatranscriptome* is the collection of all transcribed RNA molecules derived from *whole meta-transcriptome sequencing* (WTS) [10]. Metatranscriptome studies of the microbiome gave vastly different results compared to metagenomic results [38,116,117] due to various factors. (**a**) Only a fraction of genes will be expressed by a microbe at any given time, (**b**) transcription rates vary across cells,

and (**c**) essential genes are sometimes expressed primarily by relatively low abundant species. In addition, it can be subject to additional sources of bias. RNA is also far less viable than DNA [118]. Unless deep-frozen, RNA can degrade within a few days, which leads to uneven degradation of fragments, amplification bias, and incorrect transcriptome profiles [118]. Furthermore, host and microbial rRNA is usually removed before sequencing (**Figure 4**). rRNA accounts for at least 90% of sequencing material without giving much information, resulting in wasted resources [10,117]. RNA sequences may further differ from their reference gene sequence due to post-transcriptional modifications. Still, transcriptome profiling can separate alive from dead cells, identify key microbiota members of otherwise low abundance, and quantify real-time microbiota response to treatment [10,38,119,120]. Furthermore, DNA and RNA profiles can be combined by normalizing RNA gene abundances with their DNA gene abundance, resulting in as *transcriptional activity* [3,38,120] (used in **manuscript III**).

Whole Metagenome Profiling

The next step is assigning or grouping high-quality reads into taxa, genes, and functions (**Figure 4**). This is not a trivial task considering that reads from metagenomic stool samples originate from hundreds to thousands of species [24]. In addition, the shortness of SGS technology comes at the cost of precision: assigning reads to specific genomes, organisms, or even genes, is challenging. Within the last two decades, a myriad of methods (**a**) defined marker genes to distinguish microbes, (**b**) created comprehensive references and pangenomes, (**c**) proposed new concepts of species and strains, or (**d**) focused on functional quantification and annotation. The following section describes different strategies for transforming read sequences into interpretable units. I will thereby focus on quantifying (**a**) taxonomic composition, (**b**) functional composition, and (**c**) joined taxonomic-functional composition.

Microbial Composition by Marker-gene Quantification

To estimate microbial composition directly from reads, three main strategies exist: DNA to marker (*e.g.*, `MetaPhlan2`, `mOTU2`), DNA to protein sequence (*e.g.*, `KAIJU`), or DNA to reference DNA (*e.g.*, `Kraken`) [121]. In most manuscripts (**manuscripts III-VI**), we used `MetaPhlAn` [122,123], a DNA-to-marker method that maps reads to a database of clade-specific marker genes and quantifies relative abundance at each taxonomic rank [121]. The marker genes were identified from reference genomes of many human gut samples [121]. While some "DNA to reference DNA" tools such as `Kraken` claimed superior performance, the choice of tools depends on a trade-off between sensitivity (*number of distinct species*) and specificity (*accuracy of their estimated abundance*) [121]. Furthermore, a recent review addressed the difference between estimating (**a**) the number of reads per taxon and (**b**) taxonomic abundance [121]. Taxonomic abundance estimation (*e.g.*, by `MetaPhlAn`) includes bias correction against varying genome sizes and marker-gene copy-numbers. At the time of writing, `MetaPhlAn` and `mOTU2` were found to be accurate estimators of taxonomic abundance with comparable resolution [121].

Reference-based approach: gene family catalogues

Identifying correlations between taxonomic clades and specific traits is insightful, but associations at this level are often difficult to confirm. It can therefore be more informative to look at genomic information. However, constructing complete and accurate genomes from complex biological samples is challenging because reads originate from a diverse set

of often closely-related organisms [124]. Instead, gut metagenomes are usually a collection of *genes* [84,125]. Therefore, genes with high sequence similarity (*e.g.*, 90%) are grouped into *gene families*. The resulting representative reference genomes and unified gene family catalogs can serve as a common ground for data analysis, interpretation, discussion, and cross-study comparison [120,125,126].

Gene catalogs can be built from reference genomes of public databases such as the National Centre for Biotechnology Information (NCBI) [127,128] or by merging all potentially relevant genes from *de novo* assemblies [125]. An example of the latter is the integrated gene catalog of the human gut microbiome (IGC). IGC comprises 9.8 million non-redundant genes reconstructed from thousands of *de novo* gut metagenome assemblies [125]. A more sophisticated procedure is presented by the HUMAnN pipeline [120,123]. First, MetaPhlAn identifies microbial species in the sample. Then, the species' pan-genomes (ChocoPhlAn) are merged and used as the reference genome for high-speed read mapping with bowtie2. Lastly, reads not assigned to genes of known species are mapped with DIAMOND to a universal database of protein-coding genes (UniRef90). A unique advantage of HUMAnN lies in quantifying microbial functions *per species* but is limited to (**a**) the knowledge of bacterial genomes and (**b**) requires more computational resources. We used HUMAnN in most studies included in this thesis (**manuscripts III-VI**).

2.5.   Growth Rate Estimation from metagenomic data

Metagenomics captures the DNA of living, stationery, and dead cells [129]. Therefore, the state and origin of many microbes (*e.g.*, microbes found in food) cannot be determined. However, the replication rate of microbial species can be estimated from metagenomic sequencing data to identify proliferating species [129]. The core idea relies on the assumption that bacterial replication starts bidirectionally from one region (*ori*) and proceeds until a terminal region (*ter*) [130]. Therefore, species replicating faster would have a stronger genome coverage at *ori* compared to *ter* [129]. This is reflected in the gene coverage, which can therefore be used to estimate the replication rate of microbes. We estimated bacterial growth rates using "Growth Rate index" (GRiD) [129] (**manuscripts II**, **III)** to verify that species with significant differential abundance were replicating despite antibiotic drugs' inhibitory effects.

2.6.   Metagenomic Species

Metagenomic species (MGS) is a concept produced by Nielson et al. in 2014 to identify novel microbial species in metagenomic samples [131] and was used to assess the functions of both known and unknown microbial species. Conceptually, genes are clustered by high linear co-abundance of genes (CAGs), assuming that genes have highly correlated DNA abundance across different samples if they originate from the same source (organism) [131]. Instead of computing all pair-wise gene abundance, the heuristic canopy clustering algorithm is used to make this computation feasible for dozens of millions of genes [131]. Original work defined CAG abundance as either (**a**) the median abundance of clustered genes [131] or (**b**) the sum of the 50 most-correlated genes within a cluster [132]. MGS are CAGs with many genes (at least 500-700) [131]. Taxonomic assignment is based on the consensus taxonomy assigned to each gene, *i.e.*, If at least 50% of genes in an MGS are assigned to one species, that MGS is assigned to that species [131,132]. Furthermore, MGS can be computed directly from gene family catalogs, as demonstrated in **manuscripts (III, IV**) and by others [56]. Furthermore, I enabled strain-level annotation of MGS utilizing the concept of *pan-genomes* as required for follow-up experiments (**manuscript III**).

2.7.   Strain Analysis

While genomes from different isolates of the same species are very similar (by definition), they represent only snapshots of continuously evolving organisms. To identify the genomic changes turning a commensal into a pathobiont or how pathogens adapted to humans, we used two methods to describe genetic variations in the manuscripts of this thesis: genome-wide association study (GWAS) and pan-genomics [80].

Coverage profiles of mapped reads [81] can be used to identify short insertions, deletions, and substitutions. Changes at a singular position are called *single nucleotide variants* (SNVs) [133]. An SNV found in at least 1% of the study population is a *single nucleotide polymorphism* (SNP). A common tool to analyze SNPs is GWAS. GWAS performs association tests between each SNP and a phenotype of interest [134]. However, due to the fast evolution time of microbes, phenotypic variations can also result from duplications or deletions of larger genomic regions, so-called *copy number variation*s (CNVs) [134]. Because of that, many methods use SNPs in gene coding regions [135] or selected marker genes [136].

Pan-genome analysis, in contrast, deals with the presence and absence of entire genes and genomic regions [80]. The *pan-genome* is the whole set of genes from all strains of a clade. It comprises a *core genome* – genes found in all or most genomes, a *shell genome* or *accessory genome* – genes found in at least two genomes, and *unique genes* found in only one isolate [80]. Importantly, this definition works well for prokaryotes but not eukaryotes. Especially for eukaryotes with large genomes (>500 Mb), exons explain only a fraction of genetic diversity. Instead, all genomic sequences should be considered, including intergenic ones [80]. Our studies used pan-genome analysis to expand our understanding of *A. fumigatus* diversity across different environments (**manuscript II**) and to annotate MGS at the strain level (**manuscript III**).

2.8.   Functional Annotation

To better understand the functions of genes, they can be grouped based on a higher functional context, such as enzymic classes, cellular compartments, metabolic reactions, and pathways. Functional categories of novel genes are often inferred *in silico*. Approaches are based on the sequence and structural information of genes with known or putative functions [137]. Annotation approaches fall into four major categories: homology-based (*e.g.*, alignment), motif-based (*e.g.*, Hidden Markov Models, neural networks), context-based and specialized approaches [137]. Public databases encompass our collective knowledge of verified and predicted functions for genes. The grouping of genes can be achieved through gene-set enrichment tests [138] or by accumulating gene profiles into functional profiles [90]. In the following, I will introduce the databases most important in my studies.

Enzyme Commission (EC) annotation is a strict classification hierarchy used to organize reactions catalyzed by enzymes [139] and is one of the most important resources for functional annotation. It was originally invented by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology in order to assign consistent naming schemes to enzyme functions [140]. Unlike other functional annotations, EC groups enzymes by common functions instead of sequence similarity. EC is maintained by ExPASY, which performs manual curation of protein functions [141].

T The Gene Ontology (GO) Consortium provides structured, controlled vocabularies of biological terms that can be used to summarize and analyze functional knowledge of gene products [142]. GO annotations are used because they describe a gene's role in a process, a location in a cell, or molecular function, even if the gene's activity is still being

investigated or changed in the future [142,143]. GO terms are thereby organized in a hierarchy of major domains: Molecular Functions, Biological Processes, and Cellular Components [142]. A gene can therefore be assigned to multiple GO terms with complementary information.

MetaCyc and KEGG are large metabolic pathway databases with similar aims and scope and are used across all projects involving sequencing data [144,145]. In contrast to GO, KEGG and MetaCyc provide precise information about the molecular function of genes and do not include genes whose function is not identified [146].

KEGG, founded in 1995, was one of the first public annotation resources, including tools to assign higher-order functionality to gene sequences. KEGG defined so-called KEGG orthologs (KOs), a set of manually curated orthologous gene groups based on public reference genomes. KOs form the base for multiple types of annotation within the database, including EC numbers, KEGG modules, KEGG pathways, and KEGG compounds.

MetaCyc [145] is considerably newer and more curated but did not supersede KEGG yet. A direct comparison between the two databases is difficult, but some concrete numbers were published in 2013 [144] and 2019 [145]. MetaCyc super pathways are roughly equivalent to KEGG pathway maps, while MetaCyc base pathways relate to KEGG modules [144]. At the time of writing, MetaCyc has fewer super pathways than KEGG (382 vs. 548) but substantially more base pathways (2,980 vs. 457) [147,148]. MetaCyc has more substantially more metabolic reactions (17,509 vs 11,741). KEGG still covers more compounds (18,905 vs. 17,490), but this difference has decreased significantly since the last major comparison in 2013 [144]. Importantly, MetaCyc contains more database attributes, which allows for finding more relations between compounds, reactions, pathways, and species. Both databases should be considered complimentary. KEGG has advantages in the number of compounds, a more robust user API, and a more robust functional hierarchy. Functional associations can then further be stratified with MetaCyc's more curated reactions. Soon, MetaCyc is likely to become the main resource the metagenomic research.

The Pfam database is a protein-centric database for classifying protein sequences into families and domains [149]. Pfam is useful to get first insights into protein-coding genes for which little to no functional annotation is present or if a particular class (*e.g.*, antibiotics) is of interest.

UniProt knowledge database is another essential resource used to identify and annotate protein families. It contains around 190 million protein sequences submitted from projects across the globe [142]. UniProt supplies several sub-databases with different levels of annotation, curation, and redundancy. To reduce the enormous amount of often redundant gene sequences, the "UniRef" databases were created and are routinely updated [142]. Therefore, proteins are clustered by their homology, starting with 100% (UniRef100) and further clustering by 90% (UniRef90) down to 50% (UniRef50). The UniRef90 database is frequently used as a reference to annotate *de novo* gene family catalogs. UniRef90 genes often contain EC, GO, Pfam, KEGG, and other annotations that can also serve as quick approximations of novel genes' functions. The HUMAnN pipeline uses UniRef90 as the primary reference for genome annotations, gene catalogs, and subsequent higher-level annotations.

16

3. Downstream Data Analysis

After quantifying genomic material into abundance profiles, the next step is the identification of associations between *features* (genes, microbes) and host traits (*e.g.*, age, host disease, treatment, genetics). Approaches used in my work can be divided roughly into (**a**) univariate methods, which infer associations of only one feature with one or more traits, and (**b**) multivariate methods, which associate multiple features with one or more traits. Examples of univariate methods are simple linear regression (including Pearson correlation) and generalized linear models (GLMs). Regression is often used to determine the differential abundance of one species between two genders (f ~ group). A typical example of multivariate estimation is a test for significant differences in β-diversity between groups. The standard method (Permutational analysis of variance – PERMANOVA) estimates grouping related differences from *pairs* of samples ($Sample_1$ + $Sample_2$ ~ group). Other examples include sophisticated machine learning frameworks, which often integrate the abundance of multiple features to make predictions. I will focus on the methodologies utilized in the included manuscripts, although I will address common alternatives briefly where merited.
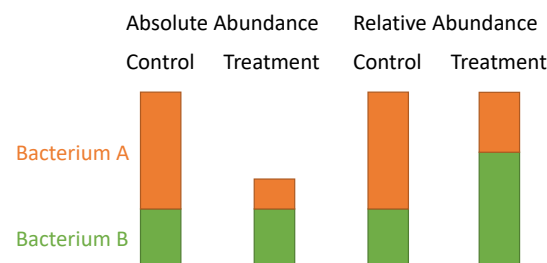
Figure 6. Compositionality bias. While bacterium A changes, B does not. However, due to sub-sampling and scaling effects, both species may appear differentially abundant. Figures based on Matchado et. al [150].

3.1. Bias in Genomic Data

Many different analyses methods, from simple regression models to entirely new algorithms, were developed in the past decade without reaching a clear consensus among the scientific community working on microbiome [128,150]. A good overview of the sources of bias is presented by Weiss [151] and Bharti and Grimm [137], and a comprehensive overview of processing pipelines was published by Breitwieser [128], and analysis methods by Matchado [150]. The three most significant sources of bias frequently mentioned in the literature are *compositionality*, *relative abundance*, and *sparsity*.

Protocols for sequencing library preparation are optimized towards maximum yield, regardless of the density of the original material [152]. While this maximizes sequencing success, the link between cell density and sequencing depth is lost, resulting in a proportional data structure [11]. One stark consequence is negative correlation bias, by which the abundance increase of one feature requires an equal decrease of the remaining features and *vice versa* (**Figure 6**) [11,152]. This "constraint sum problem" in high-throughput sequencing data needs to be addressed to avoid spurious estimates of statistical properties [11,153–155].

We rarely know the real abundance of genes, cells, or microbes in samples because (**a**) it is hard to precisely control the amount of DNA extracted from a sample and (**b**) due to amplification bias [156]. As a consequence, sequencing data estimates only *relative abundance* or *relative observed abundance* [156].

Zero-inflation (*sparsity*) bias comes from uncertainty about the meaning of *zero* and how to replace zeros before log-transformation. Zero abundance estimates have different sources, including insufficient sample material, other high abundance taxa, and true absence [157]. Workarounds involve (**a**) replacing zeros with a constant, positive value or (**b**) modeling zeros as a statistical process [150].

Of note, many of these problems can be mitigated when data is analyzed within the correct framework. In the following sections, I will introduce the most important *in silico* solutions to these problems used throughout my Ph.D., but I will further address promising technological advancements in the Discussion.

Normalization
Normalization aims to account for *technical*, non-biological effects that prevent count data from accurately reflecting abundance differences. Common factors include gene length, GC content, and sequencing depth [151]. Read length and GC-content are *within*-sample effects: they affect comparisons between feature abundances within a sample. On the other hand, sequencing depth is a *between*-sample effect: They can affect comparisons of *the same* feature *across* samples. Bias is mitigated or removed by a mathematical transformation that results in an *invariant*: a data property that is identical (within or between samples) after a specific transformation was applied.

3.2.   Differential Abundance Testing
We consider a feature differentially abundant if its' group-wise mean differs significantly between at least two conditions [158]. Significance relates to *P*-values, a statistical concept used to deal with randomness in scientific measurements. The *P*-value is defined as the probability of observing an outcome *at least* as high (or higher) than expected by chance assuming the Null-Hypothesis to be true [159]. Such a *P*-value is valid if we can make reasonable assumptions about the expected distribution of the measurement under experiment conditions. Indeed, we rarely know the true distribution of data in the real world, but we can often either approximate it or transform data to fit a distribution better.

We can generally differ between differential abundance estimators that make minimal assumptions about the distribution of data (*non-parametric*) and those that make additional assumptions in order to increase statistical power (*parametric*) [90]. Non-parametric tests such as the Wilcoxon rank-sum or Kruskal-Wallis test perform statistical inference on data transformed into *ranks* [90] (**manuscript IV**). In addition, they can often be applied to data without known distributions. While less common, non-parametric generalized linear models exist as well [160,161]. In contrast to many other non-parametric methods, these models support the control of additional covariates and confounders. I used such a model (*Rfit* [161]) in my work to control for additional covariates such as gender. Parametric methods, in contrast, are more diverse and their applicability context specific.

Parametric tools – RNA-Seq
RNA-Seq data generated by high-throughput sequencing machines follows a Poisson distribution with over-dispersion, also known as a negative binomial distribution [158]. This distribution is commonly assumed in generalized linear regression models fit to gene expression data [151,158,162]. This assumption is also made by the leading methods for estimating differential gene expression in bulk RNA-Seq experiments, `DESeq2` and `edgeR` [158]. However, in stark contrast to metagenome studies, only 4-6 biological replicates per group are sufficient to control the false-discovery rate (FDR) at the desired level (*e.g.*, 5%)

for typical case-control RNA-Seq experiments [158]. This power is achieved because **(a)** gene expression data of one organism is described well by negative binomial distributions and **(b)** by further assuming insignificant differences in gene expression between treatment groups for most genes. The latter assumption makes between-group normalization procedures very robust against violations of all the assumptions made. I used `DESeq2` and `edgeR` in **manuscript I**.

Parametric tools – Metagenomics
`DESeq2` and `edgeR` were also applied in metagenome investigations [90]. However, their applicability has been doubted due to statistical disparities in metagenomic data [151,156]. Furthermore, microbiome abundance profiles (metagenomics and metatranscriptomics) have substantially different properties: **(a)** Zero-inflation is usually high at genus and lower levels. **(b)** The constant-sum constraint is stronger because the number of species in a taxonomic composition is much smaller than in gene compositions. **(c)** The majority of species or genes– *or even all* –can be differentially abundant between traits [163], which violates the central assumption of most (RNA-Seq) scaling normalization methods [156].

Early solutions that addressed these problems involved **(a)** generalized linear mixed-models with zero-inflation components [164], **(b)** normalizing abundances by assessing library size factors derived from only some taxa across samples [156], and **(c)** by making further assumptions on parametric distributions of data [156]. For example, `Maaslin2` [165] is a general regression framework that applied a GLM to log-transforms total-sum scaled data (**manuscript V**). Another example is metagenomeSeq [156,164], which uses a log-normal mixture model with a zero-inflation component (**manuscript III**). While powerful, this tool was later shown to have inflated FDR with certain types of complex microbiome data, indicating the need for further methodological improvements [163]. Still, both methods yielded consistent results in my study but showed differences in statistical power between projects.

Enrichment Tests
Once a set of differentially abundant genes or microbes is determined, it can be helpful to identify their common functional processes or properties. This is often done using *overrepresentation analysis* (ORA) methods [138]. Thereby, a contingency table of *features* and their membership is used (*e.g.*, gene identifies to GO terms) together with a set of features (*e.g.*, DA or network clusters). Hypergeometric (Fisher) or chi-square tests are commonly used to determine significant enrichments in membership [138] (**manuscript I-VI**). For protein-protein interactions and enrichments, the STRING platform provides comprehensive analysis and visualization tools (**manuscript I**) [166]. For microbes, specialized tools like microbe-set enrichment analysis (MSEA; **manuscript V**) were developed [167]. MSEA categorized all published microbe-host-gene interactions and performs tests on enrichments for host-disease genes.

3.3.   Correlation Network Analysis
Network-based analytical approaches have proven useful to study complex systems with many interactions, such as gene-regulatory or gene abundance networks [168]. Given the complex interactions between thousands of individual species found in microbiome samples, such network analysis methods are also helpful in the microbiome field [169], including cross-kingdom associations. A key feature of network analysis is that structural elements of networks appear to be ubiquitous to most complex systems. Network biology

approaches range from correlation methods to complex graph-based modeling approaches [168].

Co-Abundance Methods

A network is formally a graph in which features (genes, microbes, traits) are nodes, and edges between these nodes represent their interactions. The correlation of feature abundances (*co-abundance*) is a common approach to generating hypotheses about feature dependencies and networks. Estimating naïve Pearson and Spearman correlation coefficients is popular in many microbiome fields [150]. However, Weiss et al. [15] and others [170,171] discussed the strengths and weaknesses of different approaches. Since naïve methods do not address sparsity and compositionality, they can create many *spurious* (false-positive) correlations. Matchado et al. compiled a comprehensive yet not exhaustive list of alternative approaches in metagenomics [150]. More sophisticated methods often rely on some form of sparsity assumption (*e.g.*, `SparCC`, `SPIEC-EASI`, `BAnOCC`) and infer quantities relating to the log-transformed unobserved counts (`SPIEC-EASI`, `BAnOCC`) [170]. The sparsity assumption increases statistical power when studying thousands of interactions from only dozens or hundreds of samples [171,172]. To mitigate bias, log-ratio transformations such as the centered log-ratio (CLR) [150] (**manuscript V**) or Median-by-ratio (MED) [162,163] (**manuscript I**) are applied to raw read counts as a pre-processing step. Alternatively, a Dirichlet multinomial model can be used to directly account for compositionality (`BAnOCC`) [150] (**manuscript III**) but at the cost of high runtime.

Most methods address only the compositionally or library size bias but not zero-inflation bias. Workarounds involve (**a**) replacing zeros with a constant, positive value or (**b**) modeling zeros as a statistical process. An example of the latter is `BAnOCC`, a Bayesian framework method to fit a log-normal prior and therefore does not require zero-replacement [170].

Ultimately, none of the existing methods is fully satisfying. Their applicability must be considered on a case-to-case basis, which was the case for manuscripts included in this thesis. Still, many tools yield complementary results, which could be combined for more robust conclusions [173].

Trans-Kingdom Correlation Analysis

Most network-based tools and models are designed for intra-kingdom interactions and may create spurious correlations in trans-kingdom estimations [150]. Regardless, sparsity-based methods such as `SparCC` were also applied in cross-kingdom studies [169,174]. However, an adaption of `SPIEC-EASI` enabled trans-kingdom correlation analysis [175]. It was thereby demonstrated that log-ratio transformations allow the study of bacterial-fungal correlation in lung microbiome data with a smaller bias [175]. Thereby, the composition of each kingdom is log-ratio transformed separately, allowing for a direct comparison of the resulting abundance ratios within and between samples. It is worth noting that `SPIEC-EASI` adds 1 to abundances to remove zeros, introducing considerable bias in data with high-sparsity or low abundance samples [11,176]. However, log-ratio transformations can be applied easily, and the resulting data can be analyzed using simpler or naïve correlation methods. For example, zeros can be replaced with Bayesian-multiplicative replacement (`BM`) [177] to replace zeros, the result log-ratio transformed with CLR, and then analyzed using Pearson or Spearman-based methods. For longitudinal data, I used a Bayesian framework (`BAnOCC`) due to its higher robustness (**manuscript III**). For data with

even greater dimensionality, `SparCC` (**manuscript IV**) and log-ratio-based approaches (**manuscripts V, IV**) were more suitable because these methods were fast and had sufficient statistical power.

3.4. Microbial Source Tracking

Microbial source tracking refers to statistical methods quantifying microbial contaminants in a metagenomic sample [178]. Thereby, they use the microbial abundance profiles from potential contaminant samples (*sources*) to estimate their species abundances in the observed microbial community (*sink*) [179]. However, the methodology is also helpful to quantify the divergence of microbial communities in longitudinal studies [180]. Fast Expectation-maximization for microbial Source Tracking (`FEAST`) [179] superseded previous approaches, including more robust estimations of *unknown* sources of taxa. `FEAST` was an integral part in **manuscript IV**.

3.5. Machine Learning

Machine learning is a vast area with a scope beyond this thesis. But since it was an essential part of the manuscripts of this thesis, I will briefly introduce the most important concepts. Machine learning is a form of artificial intelligence that uses statistical methods to identify predictors (*features*) of outcomes (*classes*). Learning can be *supervised*, in which samples are labeled (*e.g.*, treatment groups) and *unsupervised* (no labels) [181]. An example of supervised learning is predicting host disease from the corresponding microbiome composition. Examples of unsupervised learning are several clustering approaches (*e.g.*, k-nearest neighbor) and ordination techniques (*e.g.*, non-metric dimensional scaling). In the field of microbiome, machine learning is applied for a wide range of tasks, including the creation of patient-status classifiers [182], annotation of gene functions [183], and predicting metabolite profiles from microbiome compositions [184], just to name a few. One critical source of error is overfitting: A model that memorized the data but cannot make accurate predictions on new data. To assess the robustness and accuracy of machine learning models, the predictive power of models is evaluated using cross-validation [182]. Thereby, iteratively, a fixed number of samples is held out from both feature selection and model training and used only to make predictions.

II. OBJECTIVES OF THIS THESIS

Microbiome research emerged as an attractive field to study due to the involvement of microbiota in host homeostasis and disease. Research on host-pathogen interactions is essential for developing new diagnostics and therapeutics. A remarkably flexible tool for the study of microbiota is genome sequencing technology. In this thesis, I use several types of genome and transcriptome sequencing to assemble and quantify genes, estimate microbial abundance across distinct kingdoms, and apply advanced statistical methods to predict differences in microbial functions and interactions. Throughout this thesis, I focus on interactions between host, bacteria, fungi, and viruses and demonstrate how multi-omics technology can be used to create verifiable hypotheses.

The first two papers of my thesis are focused on one specific human pathogen, *Aspergillus fumigatus*. Since fungal infections are on the rise and mortality rates remain high despite improvements in antifungal therapy, understanding host-pathogen interactions between the human immune system and human pathogenic fungi is of major interest.

After breaching physiological barriers, innate immune cells are the first line of defense against invaders. Even though patients are often co-infected by human viruses in addition to *A. fumigatus*, previous studies were limited to the analysis of only one pathogen and one host cell. The latter is clinically relevant because some viruses, such as human cytomegalovirus (CMV), were hypothesized to enhance the infectious potential of *A. fumigatus*. However, hard evidence was lacking.

To address this hypothesis, we surpassed existing dual RNA-Seq approaches by developing the first triple RNA-Seq. It enabled the study of transcriptome changes between human monocyte-derived dendritic cells (moDCs) with simultaneous co-infection by two pathogens. Furthermore, we reconstructed and analyzed the pan-genomes of 300 *A. fumigatus* isolates to better comprehend strain diversity. My work addresses the following questions:

1. What mechanistic changes occur in the transcriptome of moDCs and each pathogen during co-infection by *A. fumigatus* and *CMV*?

2. Do clinical *A. fumigatus* strains differ from environmental strains, and are their consequences for infection studies?

In a broader context, and instead of focusing on specific pathogens, we also need to investigate the native environments of microbes such as the human gut. Many opportunistic microbes – "pathobionts" – such as certain members of *Enterococcus*, *Streptococcus*, *Escherichia*, and *Candida*, are commensal to the lower gastrointestinal tract of humans and harmless unless specific circumstances occur. It is still not sufficiently understood how the host, its' immune system, and commensal microbes control the pathogenesis of other microbes. However, a fair amount of research has identified several essential microbial mechanisms, host conditions, and drug interventions that are frequently associated with the promotion or inhibition of pathogenesis. In this regard, antibiotic administration, critical illness, and cancer are of strong interest.

Again, I used and adapted sophisticated pipelines to improve our understanding of microbial interactions. However, my focus here is on community-wide changes in human gut bacteriomes and mycobiomes. In particular, I address the following questions:

1. What changes in interactions occur between gut bacteria and fungi during oral antibiotic administration, and are the observed effects reversible?

2. Which findings from antibiotic drug administration studies on healthy humans translate to critically ill patients?

3. Are there microbial functions encoded in metagenomes that are resilient to various diseases and treatments, and what are they?

4. Can we identify bacteria that limit or promote the growth of common human pathogens such as *Candida* under various conditions in the gut of asymptomatic patients?

III. MANUSCRIPTS

Overview of Manuscripts

1. [co-first author] Seelbinder, Wallstabe, and Marischen *et al.* **Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model**. In: *Cell Rep*. 33 (2020), pp. 108389. DOI: 10.1016/J.CELREP.2020.108389

2. [co-author] Barber and Sae-Ong *et al.* **Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection**. In: *Nat Microbiol 2021 612*. 6 (2021), pp. 1526–36. DOI: 10.1038/s41564-021-00993-x

3. [co-first author] Seelbinder B, Chen J, et al. **Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria**. In: *Microbiome*. 8 (2020), pp. 1–20. 10.1186/S40168-020-00899-6

4. [co-author] Marfil-Sánchez and Zhang *et al.* **An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness**. In: *Gut Microbes*. 13 (2021). 10.1080/19490976.2021.1993598

5. [co-first author] Seelbinder and Lohinai *et al.* **Candida species explore alternative carbon sources to support growth in the human gut under dysbiotic conditions**. [*in preparation*] (2022).

6. [co-first author] Marfil-Sánchez and Seelbinder *et al.* **Gut microbiome functionality might be associated with exercise tolerance and recurrence of resected early-stage lung cancer patients**. In: *PLoS One*. 16 (2021), pp. e0259898. 10.1371/JOURNAL.PONE.0259898

25

I.   Manuscript:
Triple RNA-Seq to Study Host Co-Infection

FORM 1

Manuscript No. 1

Manuscript title: Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model

Authors: Seelbinder B., Wallstabe J., Marischen L., Weiss E., Wurster S., Page L., Löffler C., Bussemer L., Schmitt A. L., Wolf T., Linde J., Cicin-Sain L., Becker J., Kalinke U., Vogel J., Panagiotou G., Einsele H., Westermann A. J., Schäuble S., Loeffler J.

Bibliographic information: *Cell Reports* 33 (2020), pp. 108389, 10.1016/J.CELREP.2020.108389

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
Seelbinder	10%	90%		35%	
Wallstabe	10%		40%	20%	
Marischen	5%		40%	25%	
Westermann	25%			5%	30%
Schäuble	25%	10%		5%	30%
Loeffler J.	25%				30%
Others	0%	0%	20%	10%	10%
Total:	100%	100%	100%	100%	100%

Overview:

Together with my co-authors from Würzburg, we developed a new methodology termed "triple RNA-Seq" to investigate the transcriptome of both pathogens and host cells simultaneously. We studied the interactions between a fungal (*Aspergillus fumigatus*) and a viral pathogen (human cytomegalovirus) during co-infection of human monocyte-derived dendritic cells. I created the pipelines for RNA-Seq data processing and analysis, including differential gene expression and gene co-expression network analyses to decipher cross-kingdom communication. We deliver strong evidence for synergistic effects of both pathogens during co-infection and identified a surprisingly large, distinct gene expression cascade during co-infection compared to single infections.

Cell Reports

Article

Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model

Bastian Seelbinder,[1,11] Julia Wallstabe,[2,11] Lothar Marischen,[2,11] Esther Weiss,[2] Sebastian Wurster,[2,3] Lukas Page,[2] Claudia Löffler,[2] Lydia Bussemer,[2] Anna-Lena Schmitt,[2] Thomas Wolf,[1] Jörg Linde,[4] Luka Cicin-Sain,[5,6] Jennifer Becker,[7] Ulrich Kalinke,[7] Jörg Vogel,[8,9] Gianni Panagiotou,[1,10] Hermann Einsele,[2] Alexander J. Westermann,[8,9,11] Sascha Schäuble,[1,11] and Juergen Loeffler[2,11,12,*]
[1]Systems Biology and Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute (HKI), 07745 Jena, Germany
[2]University Hospital Würzburg, Medical Hospital II, WÜ4i, 97080 Würzburg, Germany
[3]The University of Texas MD Anderson Cancer Center, Department of Infectious Diseases, Infection Control and Employee Health, Houston, TX 77030, USA
[4]Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Bacterial Infections and Zoonoses, 07743 Jena, Germany
[5]Department of Vaccinology and Applied Microbiology, Helmholtz Centre for Infection Research, Hannover-Braunschweig Site, 38124 Braunschweig, Germany
[6]Cluster of Excellence RESIST (EXC 2155), Hannover Medical School (MHH) Braunschweig, 38124 Braunschweig, Germany
[7]Institute for Experimental Infection Research, TWINCORE–Centre for Experimental and Clinical Infection Research, a joint venture between the Hannover Medical School and the Helmholtz Centre for Infection Research, Cluster of Excellence RESIST (EXC 2155), Hannover Medical School (MHH), 30625 Hannover, Germany
[8]Institute of Molecular Infection Biology (IMIB), University of Würzburg, 97080 Würzburg, Germany
[9]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), 97080 Würzburg, Germany
[10]Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, Hong Kong S.A.R., China
[11]These authors contributed equally
[12]Lead Contact
*Correspondence: loeffler_j@ukw.de
https://doi.org/10.1016/j.celrep.2020.108389

SUMMARY

High-throughput RNA sequencing (RNA-seq) is routinely applied to study diverse biological processes; however, when performed separately on interacting organisms, systemic noise intrinsic to RNA extraction, library preparation, and sequencing hampers the identification of cross-species interaction nodes. Here, we develop triple RNA-seq to simultaneously detect transcriptomes of monocyte-derived dendritic cells (moDCs) infected with the frequently co-occurring pulmonary pathogens *Aspergillus fumigatus* and human cytomegalovirus (CMV). Comparing expression patterns after co-infection with those after single infections, our data reveal synergistic effects and mutual interferences between host responses to the two pathogens. For example, CMV attenuates the fungus-mediated activation of pro-inflammatory cytokines through NF-κB (nuclear factor κB) and NFAT (nuclear factor of activated T cells) cascades, while *A. fumigatus* impairs viral clearance by counteracting viral nucleic acid-induced activation of type I interferon signaling. Together, the analytical power of triple RNA-seq proposes molecular hubs in the differential moDC response to fungal/viral single infection or co-infection that contribute to our understanding of the etiology and, potentially, clearance of post-transplant infections.

INTRODUCTION

Allogenic stem cell transplantation (alloSCT) has advanced the therapy of hematological malignancies and is potentially curative for a spectrum of nonmalignant hematological disorders (Singh and McGuirk, 2016). The first successful solid organ transplantation (SOT) took place in 1954; today, transplant statistics are steadily increasing, with over 36,500 organ transplants in the United States in 2018 (based on OPTN data as of January 9, 2019; Harrison et al., 1956). Reduced-intensity conditioning regimes, novel therapeutic strategies to combat graft-versus-

host disease, and tailored supportive care have improved alloSCT and SOT outcomes. However, opportunistic infections are still a clinical challenge and a major source of post-transplant complications (Ullmann et al., 2016).

Invasive aspergillosis (IA), predominantly causing pulmonary infections, is responsible for significant post-transplant morbidity, mortality, and incremental cost burdens (Drgona et al., 2014). In addition, human cytomegalovirus (CMV)-associated infections, including CMV pneumonia, remain the most common infectious complications in alloSCT recipients (Camargo and Komanduri, 2017), and CMV viremia is associated

with increased early overall mortality after alloSCT (Green et al., 2016). CMV-disease-related mortality has significantly declined with improved prophylactic medication, PCR-based diagnostics, and pre-emptive antiviral treatment, yet indirect CMV effects continue to adversely impact alloSCT outcomes (de la Cámara, 2016; Duarte and Lyon, 2018). Notably, CMV infections pose an independent risk factor for development of IA in alloSCT recipients (Garcia-Vidal et al., 2008; Marr et al., 2002), and invasive mycoses are a frequent cause of mortality in patients surviving CMV disease (de la Cámara, 2016; Nichols et al., 2002).

Ample evidence indicates that CMV alters the human immune response to escape host surveillance and establish latent persistence (Cheung et al., 2009; Hahn et al., 1998; Kaminski and Fishman, 2016; Kotenko et al., 2000; Taylor-Wiedeman et al., 1991). Several proteins encoded by CMV broadly modulate the magnitude and quality of host immune cell functions (Miller-Kittrell and Sparer, 2009). For example, CMV-secreted immunosuppressive cytokine homologs inhibit dendritic cell maturation and survival as well as dendritic cell-mediated T-helper (Th) cell activation and Th1 differentiation (Chang et al., 2004; Raftery et al., 2004), which are crucial mechanisms for linking innate and adaptive immunity to fungal pathogens. Vice versa, *Aspergillus fumigatus*, the most frequent cause of IA, suppresses human T cell activation in response to CMV (Stanzani et al., 2005). However, the molecular events underlying the differential impact of CMV and *A. fumigatus* on human mononuclear cell functions and the reciprocal immune defense in co-infections are largely unexplored (Grow et al., 2002; Martino et al., 2009; Mikulska et al., 2009; Solak et al., 2013; Upton et al., 2007).

High-throughput RNA sequencing (RNA-seq) has greatly advanced our understanding of infections (Saliba et al., 2017; Colgan et al., 2017) and even allows simultaneous studies of host and pathogen transcriptomes (Westermann et al., 2012). These "dual RNA-seq" approaches concurrently isolate host and pathogen RNA, convert it into cDNA libraries for sequencing, and separate the transcriptomes at the computational level by mapping sequencing reads to the respective reference genomes. To date, dual RNA-seq has been applied to diverse infection models to study virulence mechanisms of—and mammalian immune responses to—viral (Juranic Lisnic et al., 2013; Wesolowska-Andersen et al., 2017), bacterial (Westermann et al., 2017), and fungal (Wolf et al., 2018) pathogens as well as eukaryotic parasites (Choi et al., 2014; Pittman et al., 2014). However, multi-organism RNA-seq has not previously been applied to co-infection settings.

Here, we advanced the concept of multi-organism RNA-seq by developing triple RNA-seq, which we applied to human monocyte-derived dendritic cells (moDCs) challenged with the two pulmonary pathogens CMV strain TB40 and *A. fumigatus*. The identified modulations of the host's immunological state upon single infection and co-infection were independently validated by flow cytometry and multiplex cytokine secretion assays. Our collective findings suggest unique interdependencies of CMV and *A. fumigatus* during co-infection that add to our molecular understanding of the synergy between CMV infection and the development of invasive mold infections in immunocompromised patients.

RESULTS

Triple RNA-Seq of Viral/Fungal Co-infections to Simultaneously Study Host and Pathogen Gene Expression

Infection assays of human immune cells with either CMV or *A. fumigatus* were previously established (Paijo et al., 2016; Lother et al., 2014). Here, we built upon these protocols and challenged moDCs with *A. fumigatus* germ tubes or CMV, either separately (single infection) or in combination (co-infection). Host cell viability (Figure S1A), infection rates (Figures S1B and S1C), and morphology (Figure S1D), analyzed by flow cytometry and fluorescence microscopy, demonstrated sustained infections with the pathogens in both single- and co-infection settings. No prominent changes in fungal morphology or infection rate were observed between the two infection settings (Figures S1C and S1D). In contrast, we observed an increased virus infection rate in the presence of *A. fumigatus* as compared to CMV single infection (Figure S1B).

As a prerequisite for multi-organism RNA-seq analysis, several parameters—including lysis conditions, multiplicities of infections (MOIs), and efficient removal of ribosomal RNA (rRNA)—need to be empirically determined for a given infection system. Here, we first established these parameters for single-infection settings of moDCs with either CMV or *A. fumigatus* (Figures S2A–S2C; Data S1). Application of dual RNA-seq to the resulting RNA samples led to the detection of the presumed infection-specific host expression patterns (Figures S2D and S2E), including an induction of pro-inflammatory host marker genes *IL1A/B*, *CCL3*, and *TLR2* upon *A. fumigatus* infection (Braedel et al., 2004; Lass-Flörl et al., 2013; Walsh et al., 2005) and *IFNB* and *CCL2* activation after CMV infection (Loewendorf and Benedict, 2010; McNab et al., 2015). In addition, altered fungal expression levels in the presence of moDCs (Figure S2D), e.g., of the gliotoxin mRNA (*gliF*) that encodes a mycotoxin known to be produced during host infection, as well as induced expression of immunomodulatory viral mRNAs (Figure S2E), further supported the reliability of our approach.

To characterize all three transcriptomes in sequential or simultaneous co-infection settings, we expanded the method toward triple RNA-seq (Figure 1A). MoDCs were either first infected with CMV and, after 4.5 h, additionally with *A. fumigatus*, or first infected with *A. fumigatus* followed by CMV challenge 2 h later, or simultaneously infected with both pathogens. Since dual RNA-seq data indicated that major gene expression changes occurred during the first hours after fungal and viral exposure, co-infection samples were harvested after 9 h and subjected to triple RNA-seq. Co-infection data were compared with transcriptome data from single infections (harvested in parallel with co-infections) and with data from uninfected moDCs or *A. fumigatus* mono-cultures.

We extrapolated sequencing depth requirements for triple RNA-seq to sufficiently cover all three transcriptomes. Since dual RNA-seq indicated host transcriptome coverage to be rate limiting (Figures S2A and S2B), we increased the sequencing depth for triple RNA-seq by ~5-fold, yielding ~15 million non-ribosomal human reads, a threshold above which further increases in sequencing depth have diminishing returns

Cell Reports
Article
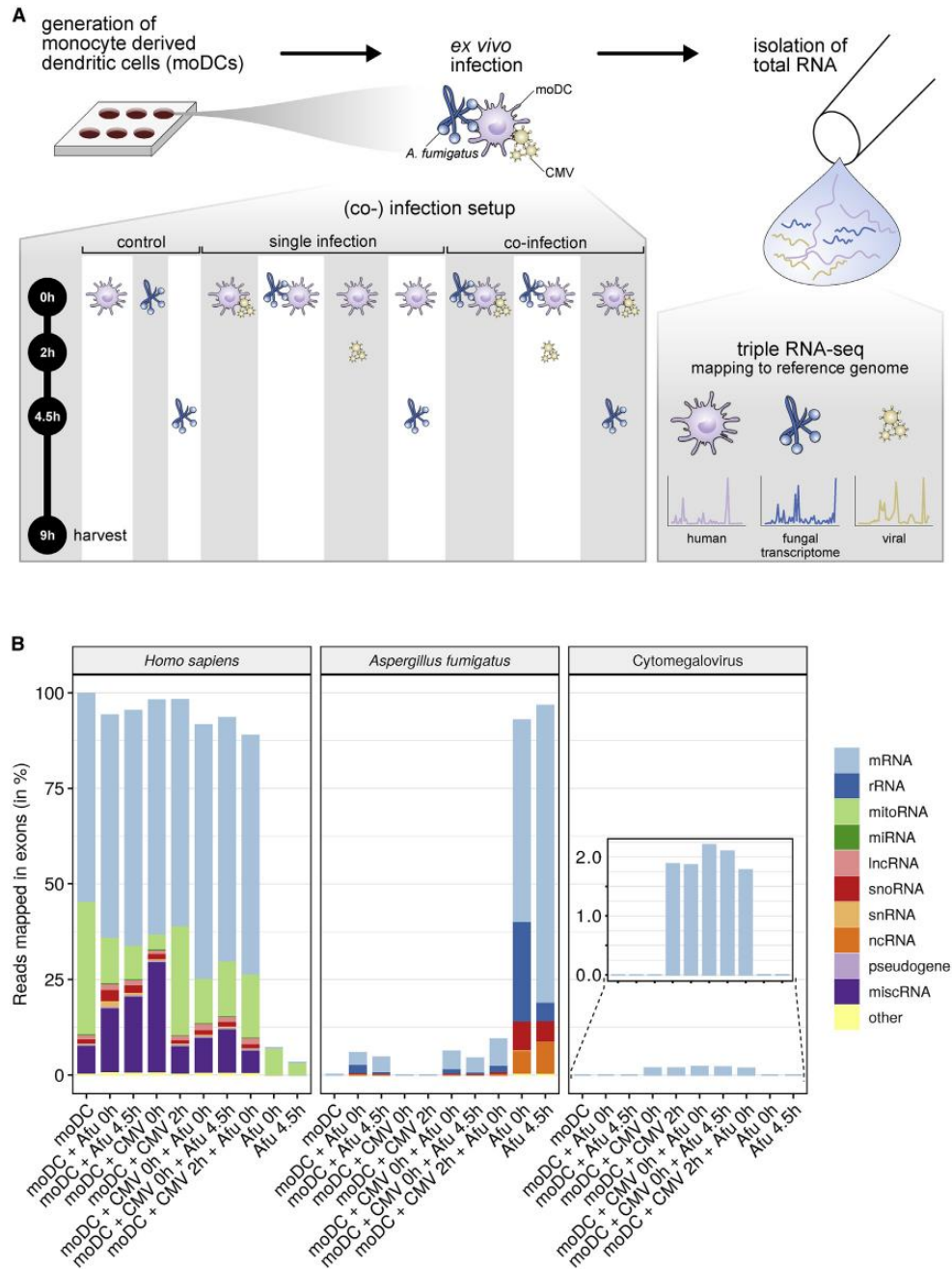
Figure 1. Triple RNA-Seq Outline

(A) Triple RNA-seq pipeline. Setup and time frame for controls, single infection, and co-infection are indicated; n = 4.
(B) Percentages of sequencing reads that mapped to the annotated reference genome of the three studied organisms. Reads per organism were further assigned to the indicated transcript classes. Afu, *Aspergillus fumigatus*; moDC, monocyte-derived dendritic cell; CMV, cytomegalovirus; mitoRNA, mitochondrial RNA; miRNA, microRNA; lncRNA, long noncoding RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; ncRNA, noncoding RNA; miscRNA, miscellaneous RNA.

([Ching et al., 2014](#); [Liu et al., 2014](#)). Between 67% and 95% of the quality-filtered reads ([Data S1](#); [Data S2](#)) were successfully aligned to the reference genomes. While the vast amount of reads mapped to the human genome, *A. fumigatus* contributed up to 24% and CMV contributed up to 1.6% of mapped reads in single-infection and co-infection samples ([Figure 1](#)B). As expected, the majority of human (55%) and fungal (55%–75%) reads derived from mRNA. Albeit that ribosomal depletion was less efficient for the fungal than for the human transcriptome, rRNA-derived reads did not exceed 25% in the *A. fumigatus* data subset, allowing for differential gene expression analyses. Other noncoding RNA classes were adequately represented in the two eukaryotic transcriptomes, whereas CMV-annotated genes all encode mRNAs ([Figure 1](#)B).

Collectively, these data confirmed the high technical quality of the triple RNA-seq data, with relative proportions of host-to-pathogen read ratios and assignment to major RNA classes matching the predictions extrapolated from the dual RNA-seq pilot ([Data S1](#)).

Transcriptomes during Co-infection Differ Globally from Single Infections

Although principal-component analysis (PCA) of moDC transcriptomes revealed no time-point-specific segregation ([Figure 2](#)A), distinct clusters were obtained for single-infection conditions with *A. fumigatus* or CMV and co-infection. Similarly, *A. fumigatus* samples failed to cluster according to infection time point yet globally differed between fungal mono-culture and (co-)infection samples ([Figure 2](#)B). Finally, CMV transcriptome profiles formed clusters for single infection and co-infection despite the compact size of the CMV genome (166 genes; [Figure 2](#)C). In the following, due to the absence of time-point-specific segregation, we disregard temporal information and analyzed infection samples based on etiology (i.e., uninfected, single viral or single fungal infection, and co-infection).

The triple RNA-seq approach potentially reduces systemic noise that may impede cross-species gene expression correlations when separately sampling host and pathogen transcriptomes ([Westermann et al., 2017](#)). To globally investigate interspecies co-expression, we calculated node betweenness and degree, two metrics to quantify co-expression network complexity and identify hub genes ([Figure 3](#); [Data S3](#)). Focusing on innate immunity-associated effects, we restricted the network analysis to human genes contained within InnateDB ([Breuer et al., 2013](#)). The first two networks depicted in [Figure 3](#) show host-pathogen gene-gene correlations present in the intersection of the respective single- and co-infection networks, whereas the remaining networks refer to correlations in the set differences between infection settings. Co-infection networks shared only few cross-species correlations with single-infection networks, whereas co-infection exclusive correlations exhibited a high degree of connectivity and were mostly based on positive correlations. Very few correlations were specific to single *A. fumigatus* infection, but fungus-host connectivity increased during co-infection, suggesting the fungus to adapt to—and maybe benefit from—the presence of CMV. In contrast, the single CMV infection network consisted of many unique, mostly positive, inter-species correlations. This implies a specialized

CMV response to human target cells largely ignoring the presence of the fungus.

Cross-species expression correlation analysis can pinpoint host factors with potential as future biomarkers or drug targets. *CXCL11*, for example, showed considerable network importance specifically during co-infection without any obvious correlations in single-infection settings ([Data S3](#)). Similarly, *TNF* showed high node degree and betweenness, specifically during co-infection, but occupied a hub position also in the intersection correlation network of fungal single- and co-infection ([Data S3](#)). This hints toward a generally important role for tumor necrosis factor (TNF) in the response to *A. fumigatus*, regardless of the additional presence of CMV. *RELA*, in contrast, possesses central network importance in the intersection of CMV single- and co-infection ([Data S3](#)), suggesting a particular relevance of this factor in the immune response against CMV.

Differential Gene Expression Profiles in the Three Interacting Organisms

Relative to uninfected moDCs, there was an increased number of differentially expressed human genes upon co-infection as compared to either single infection ([Figure S3](#)A). Moreover, we identified specific moDC gene sets with distinct expression patterns between, but low variance within, different infection etiologies ([Figure 4](#)A). For example, *TNF*, *IL1A*, and *CXCL8* were specifically upregulated upon the sensing of *A. fumigatus* as previously reported ([Balloy et al., 2008](#); [Caffrey-Carr et al., 2017](#); [Caffrey et al., 2015](#); [Cortez et al., 2006](#); [Mehrad et al., 1999](#); [Roilides et al., 1998](#)). On the other hand, induction of *IFNB* (encoding a first-line defense type I interferon [IFN] to CMV infection; [Marshall and Geballe, 2009](#)), *CXCL10*, and *CXCL11* (associated with immune clearance in CMV viremia; [Cheeran et al., 2003](#); [Knoblach et al., 2011](#); [Murayama et al., 2012](#)) was specific to CMV infection.

The majority of *A. fumigatus* genes differentially expressed in the presence of moDCs compared with the fungal mono-culture were shared between single infection and co-infection ([Figure S3](#)B). For example, the *cat1* mRNA that encodes a fungal catalase to break down host-derived hydrogen peroxide was equally highly expressed by *A. fumigatus* during single- and co-infection. This suggests that moDCs were the primary driver of fungal transcriptional reprogramming. This notwithstanding, a greater total number of fungal genes were regulated during single infection than during co-infection, thereby defining an *A. fumigatus* gene set whose regulation might be dispensable for infection in the presence of CMV. CMV expression analysis, on the other hand, revealed largely overlapping sets of differentially expressed viral genes during single infection and co-infection ([Figure S3](#)B).

The Host Response to Co-infection Suggests Synergy among *A. fumigatus* and CMV

Closer inspection of moDC expression data identified a subset of key immune pathways whose activity differed markedly between infection settings, including Toll-like receptor (TLR) signaling, nucleic acid sensing, and C-type lectin receptor signaling ([Figure S4](#)). In co-infected moDCs, expression levels of *AIM2* and *CCL5*, factors involved in cytosolic DNA sensing ([Figure S5](#)) and with known roles in the defense against both fungal and viral

Cell Reports
Article
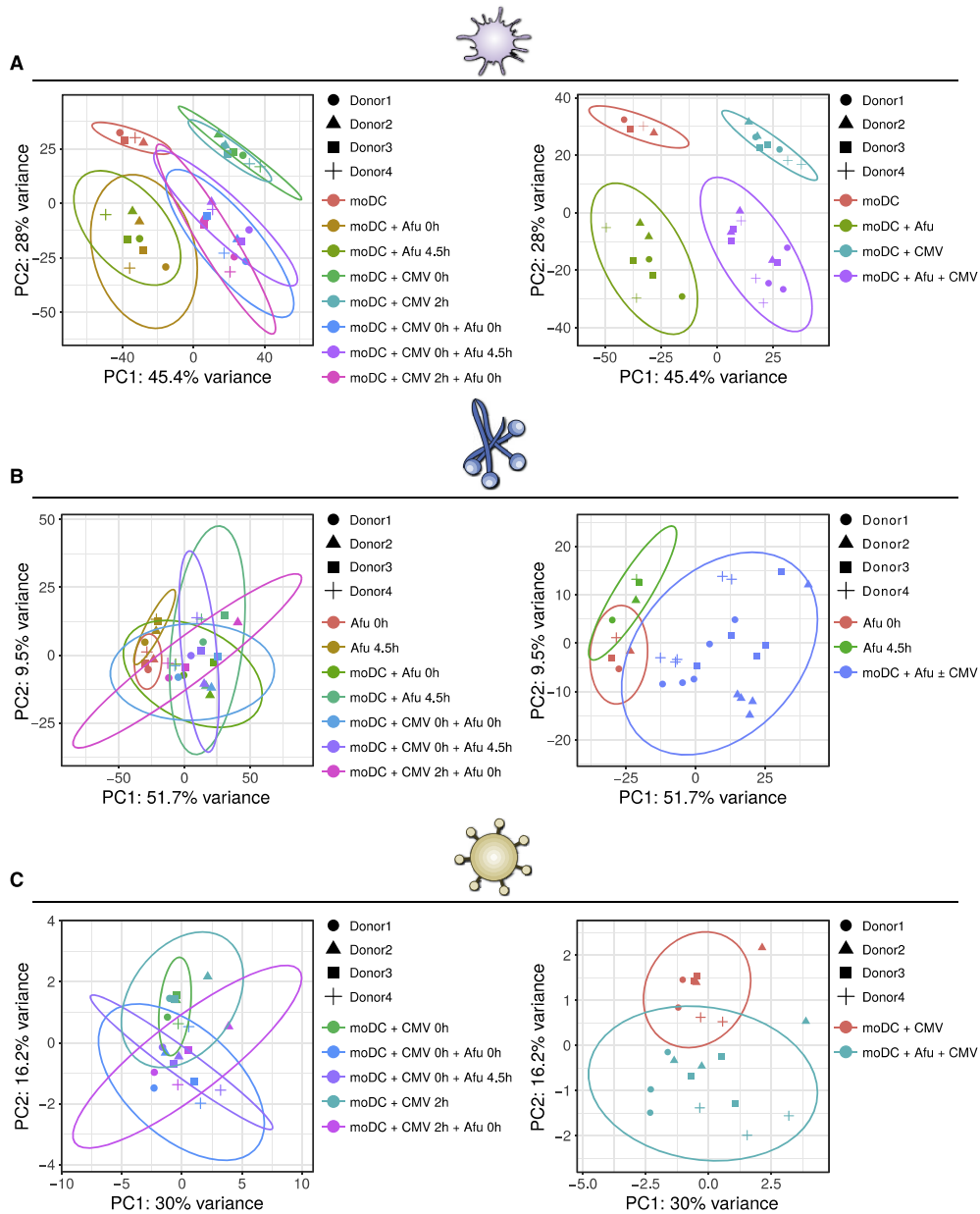
Figure 2. PCA Identifies Co-clustering Transcriptomes

(A–C) PCA for all three organisms after median-by-ratio normalization. Ellipses depict 95% confidence intervals for conditions. PCA for genes associated with (A) *Homo sapiens*, (B) *Aspergillus fumigatus*, or (C) cytomegalovirus (CMV) reference genome. Left: PCAs for all conditions. Right: (A) infection conditions grouped independent of infection time point, (B) data for single infection and co-infection grouped and compared to single-culture *A. fumigatus*, and (C) data for single infection versus co-infection independent of time point. Abbreviations are as explained in the Figure 1 legend.
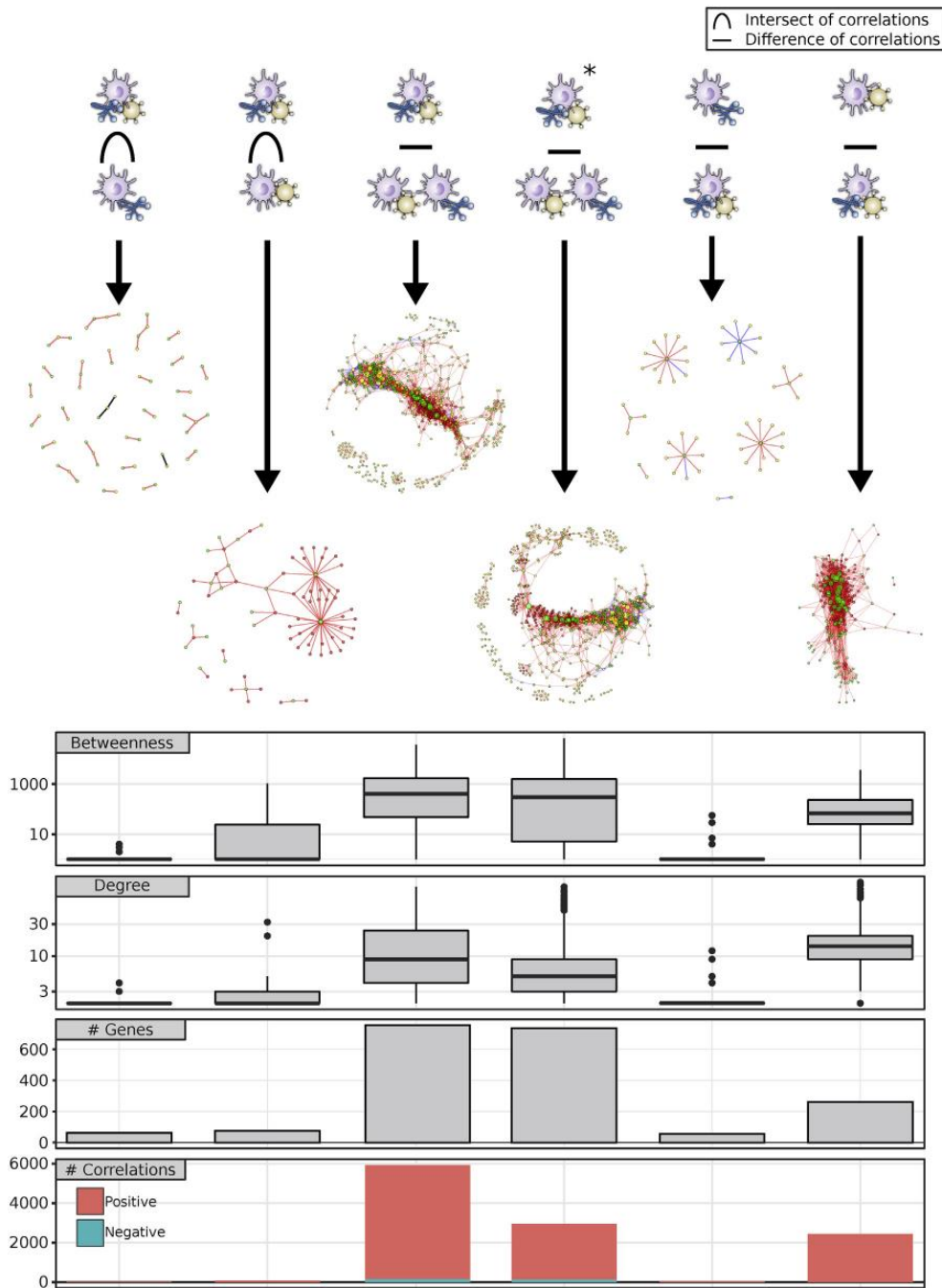
Figure 3. Interspecies Correlation Networks Pinpoint Synergies during Co-infection

Networks indicate significant correlations (edges: $p < 0.05$; $|\text{rho}| \geq 0.95$; red indicates positive correlation; and blue indicates negative correlation) of gene expression values (nodes: green represents *Homo sapiens*, yellow indicates *Aspergillus fumigatus*, and red indicates cytomegalovirus [CMV]) between

(legend continued on next page)

Cell Reports
Article

infections (Huang and Levitz, 2000; Man et al., 2016; Tyner et al., 2005), matched their cumulative induction after single infections (Figure 4A). However, a number of immune-related moDC genes did not display such additive expression patterns. For instance, the expression of nuclear factor κB (NF-κB)-dependent genes was mainly driven by *A. fumigatus* through TNF signaling (Figure 4B). Co-infection with CMV, however, reduced the expression of those genes as compared to fungal single infection. Similar expression patterns were detected for *IL10* and *IL1B*. Vice versa, RIG-I and ZBP1 signaling (previously associated with type I IFN responses; Onomoto et al., 2010; Yang et al., 2020), as well as expression of *IFNB*, *CXCL10*, and *TLR3*, displayed strong activation upon CMV single infection, which was counteracted by the additional presence of *A. fumigatus* (Figure 4B). The gene for apoptosis-associated speck-like protein (ASC), which is critical for host cell survival upon viral infection (Kumar et al., 2013), was highly expressed in uninfected and CMV-infected moDCs. In contrast, expression of *ASC* was downregulated in the presence of *A. fumigatus*, thereby mirroring the expression pattern of *Dectin-1*, encoding an important fungal infection sensor (Taylor et al., 2007). This points at a disparate role for ASC in response to viral and fungal infections, suggesting an inhibiting effect upon fungal infection that is dominant over the induction upon viral challenge. This hypothesis was further supported by the fact that *IL1B*, despite its functional connection to ASC (Martinon and Tschopp, 2007), was not co-expressed with *ASC*. Finally, *cGAS* and *STING*, both encoding receptors for foreign nucleic acids, and STING also functioning as an IFN-stimulating factor associated with the viral glycoprotein US9 (Choi et al., 2018), were induced upon single CMV infection but downregulated upon co-infection (Figure 4B).

Taken together, these expression data support the existence of two distinct host response patterns to *A. fumigatus* or CMV infection (Figure 4C). Expression of *IL1B*, *IL10*, and NF-κB-associated genes peaked in *A. fumigatus*-infected moDCs, whereas cGAS, STING, and RIG-I signaling and expression of *IFNB*, *CXCL10*, *TLR3*, and *ZBP1* showed maximal expression upon CMV single infection. Relative to their induction upon the respective single infections, expression levels for all those host genes dropped in co-infected cells, suggesting mutual interfering effects between the two host responses with possibly synergistic effects for the two pathogens.

Independent Validation of Expression Changes

Quantitative real-time PCR measurement of genes from the three organisms that were differentially expressed in the triple RNA-seq data supported the sequencing-derived expression changes (Figure 5A; Figure S5). For instance, differential expression of human *ZBP1*, fungal *cat1*, and the mRNA for the viral envelope glycoprotein UL4 in single infection and co-infection settings could be confirmed in this way (Figure 5A).

We next traced the expression of selected host factors, which were called as differentially expressed in the RNA-seq analysis,

on the protein level. For instance, as shown by flow cytometry of antibody-stained moDCs, expression of the surface marker CCR7 was highest after co-infection, which was in line with mRNA levels in the triple RNA-seq dataset (Figure S6). Additionally, altered secretion levels of key cytokines produced from genes that showed pathogen-specific expression patterns were confirmed by multiplex ELISA (Figure 5B; Figure S7). For example, the secreted levels of interleukin (IL)-1β and IFN-β differed between fungal and viral infections (Figure 5B), echoing the differential expression of their cognate mRNAs in the sequencing data.

Altogether, these results underpin the sequencing data, confirm some of the key transcriptomic changes in the infected host cells to extend to the protein level, and provide further support of the notion that pulmonary pathogens differentially affect their host at a global level during co-infection (Reese et al., 2016).

DISCUSSION

Cross-kingdom interactions in polymicrobial infections are increasingly recognized as crucial virulence determinants that shape the outcome of life-threatening infectious diseases (Arvanitis and Mylonakis, 2015; Bergeron et al., 2017). In addition to direct physical interaction and inter-kingdom signaling, altered immunopathology is considered to foster co-infections (Arvanitis and Mylonakis, 2015). Specifically, viral pathogens associated with long-term persistence such as CMV have evolved an armamentarium of counter-strategies to shape-shift the host environment, allowing for immune surveillance evasion and establishment of latent infection (Freeman, 2009; Picarda and Benedict, 2018). Although evidence is limited, previous studies reported broad alterations of host immunity elicited by CMV in immunocompromised patients, entailing the predisposition to subsequent opportunistic fungal diseases (Yong et al., 2018). To improve our understanding of this co-occurrence, we here set out to dissect the molecular interplay of *A. fumigatus*, CMV, and their shared host cells during co-infection.

Establishment of the Triple RNA-Seq Approach

To study the complex interplay of different pathogen classes with their human host and among each other, we implemented the previously proposed concept of multi-organism triple RNA-seq (Westermann et al., 2017). We selected moDCs—being at the border of innate and adaptive immunity—as a host model to establish this technology for a variety of both biological and technical aspects. Myeloid precursors form a reservoir for latent CMV infection and their differentiation into dendritic cells can trigger virus reactivation (Hahn et al., 1998; Jarvis and Nelson, 2002), while moDCs represent a well-studied surrogate model in the context of fungal infection (Hsieh et al., 2017; Mezger et al., 2008; Morton et al., 2011). Additionally, moDCs can be generated in large quantities by well-standardized protocols for the upfront optimization of infection and RNA processing

organisms. The asterisk indicates that correlations between CMV and *A. fumigatus* were removed. Node betweenness and degree are measures of network complexity. Higher degree indicates increased (local) connectivity. Higher betweenness indicates density. Corresponding boxplots show network-wise median (center line), confidence interval (boxes), and quantile (25% and 75%) values. # Genes and # Correlations display numbers of genes and correlation types, respectively.
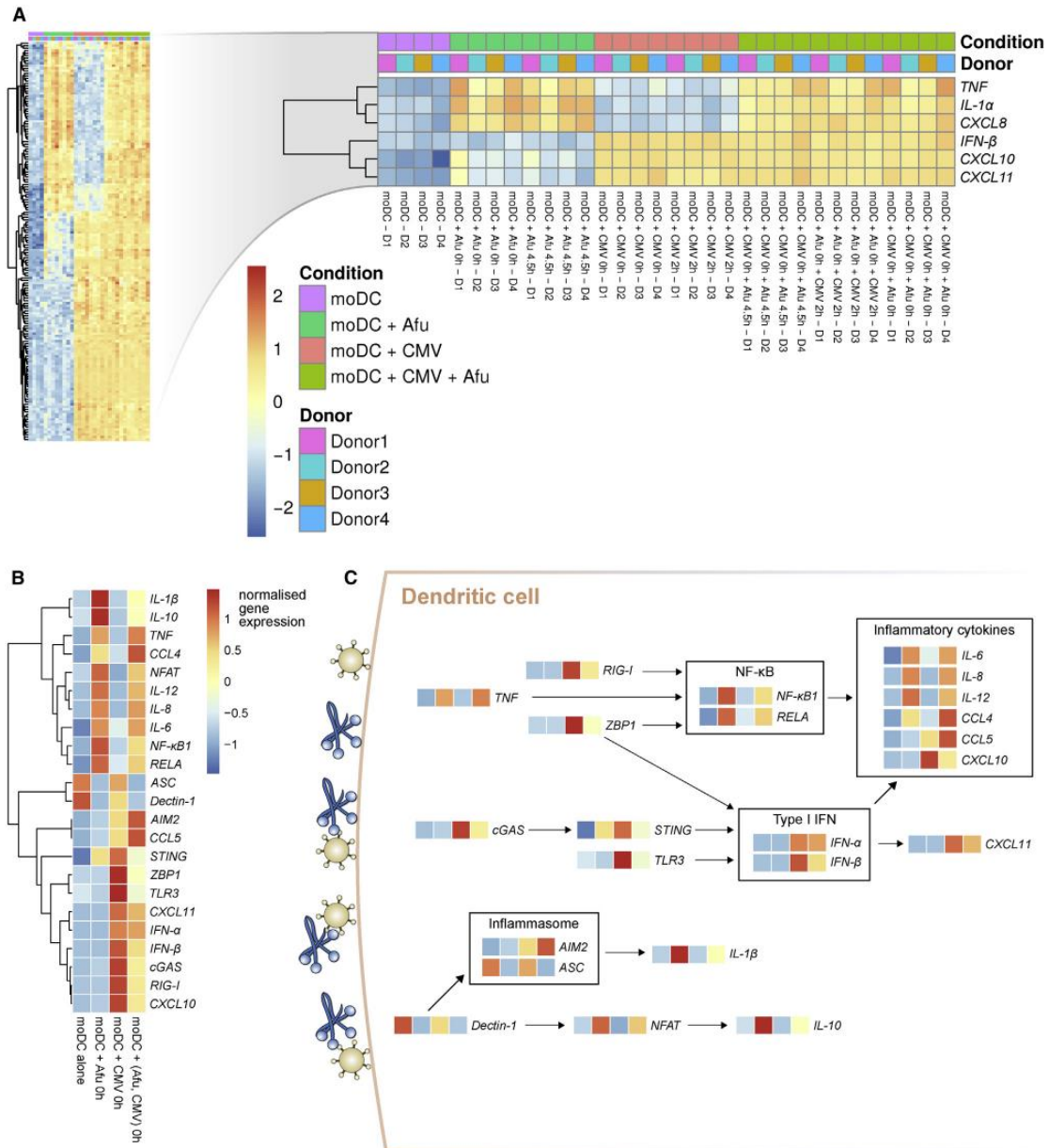
Figure 4. Detailed Differential Gene Expression Analysis of Monocyte-Derived Dendritic Cells

(A) Genes selected for high variance across all conditions and low variance within individual conditions regardless of infection time point. Left: differentially expressed genes (DEGs) with distinct expression patterns across experimental conditions (no stimulation, single infection, and co-infection). Right: examples of innate immune-relevant genes allowing for clear separation of infection types (inferred from innateDB; https://www.innatedb.com/).

(B and C) Selection of host factors involved in monocyte-derived dendritic cell response to cytomegalovirus (CMV), *Aspergillus fumigatus* infection, or co-infection is depicted based on the clustering of similar expression profiles (B) and in topological context (C). Topological analyses are based on Kyoto Encyclopedia of Genes and Genomes reference signaling maps (Figure S4). Abbreviations are as explained in the Figure 1 legend.
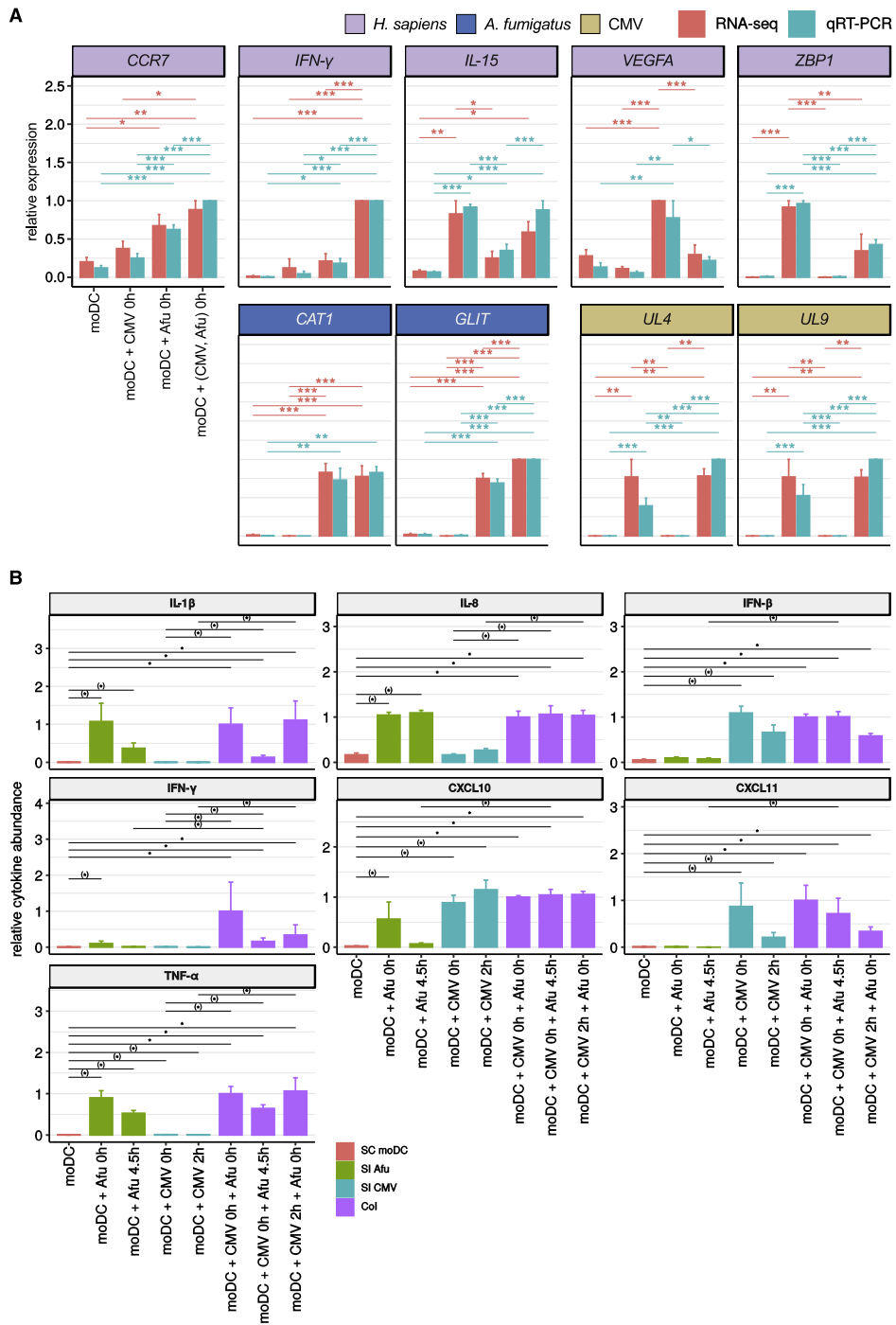
Cell Reports
Article

protocols and can be analyzed by a variety of phenotypic assays. Despite these advantages, the *ex vivo* transcriptional events described in our study harbor some limitations. These include the fact that a single cell type under well-defined infection conditions in cell culture cannot capture the complex and diverse microenvironments present in, e.g., blood or lung specimens. In consequence, part of the observed transcriptional response might be different in actually co-infected patients, depending again on a large variety of patient-specific factors.

However, the described triple RNA-seq approach potentially provides several advantages over conventional, single-species host or pathogen transcriptomics. Multi-organismal RNA-seq is more resource efficient, reducing costs for cDNA library preparation—the most expensive step in current RNA-seq protocols. Moreover, batch effects may occur during RNA isolation, rRNA depletion, reverse transcription, library generation, or sequencing and can divergently affect host and pathogen transcriptomes when they are treated separately. Batch effects, however, would affect all transcriptomes equally when they are processed in the same reaction. Therefore, subtle interspecies correlations in gene expression that may be masked by technical noise in conventional RNA-seq approaches might be uncovered in multi-organism RNA-seq. This analytical advantage, although not systematically evaluated in this study, was exemplified by the notable number of significant correlations in gene expression across species (Figure 3) and minimal intra-group variance for different infection conditions despite human donor variability (Figure 4A).

Molecular Aspects Underlying the Synergy between Viral and Fungal Pulmonary Infections

Triple RNA-seq recapitulated previously described functional alterations of CMV- and *A. fumigatus*-infected moDCs. Globally, our findings are in line with *in vivo* data showing *A. fumigatus* to drive dendritic cells directly (and indirectly, via polynuclear monocytes) toward a Th17-type immune response (Figure 6A) and showing CMV infection to result in cGAS-STING signaling, leading to an initial production of IFN-α/β, CXCL10, and CXCL11, as well as a subsequent Th1 response (Figures 6A and 6B) (Lio et al., 2016; Paijo et al., 2016; Shankar et al., 2018). Moreover, our results recapitulate the CMV-mediated induction of TLR3 signaling and type I IFNs (Mezger et al., 2009) and recapitulated the known upregulation of genes activated through *A. fumigatus*-induced pattern-recognition receptor (PRR) signaling, such as *TNF*, *IL8*, and *IL1B* (Figures 6A and 6C).

Besides, we found several novel mutual interferences within the host response to both pulmonary pathogens under co-infection settings. Specifically, CMV repressed critical PRRs and downstream effectors, including targets essential for mounting

a protective anti-*Aspergillus* immune response in high-risk patients (e.g., *CARD9*; Figure 6C). Further, CMV co-infection counteracted *A. fumigatus*-induced upregulation of NF-κB and NFAT (nuclear factor of activated T cells) cascades, resulting in lower induction levels of key proinflammatory cytokines such as IL1B (Figure 6C). This antagonizing effect was particularly pronounced in a scenario when CMV infection preceded *A. fumigatus* exposure (Figure 5B). Thus, whereas most previously described interferences between virus- and fungus-induced host responses were attributed to post-translational alterations caused by CMV proteins (Gredmark-Russ and Söderberg-Nauclér, 2012), our findings showcase the impact of viral co-infection on early transcriptional programs of fungus-exposed dendritic cells.

Conversely, *A. fumigatus* co-infection hampered the induction of viral nucleic-acid-sensing cascades (RIG-I, cGAS, and ZBP1), resulting in reduced expression of *IFNB* and *CXCL10* (Figure 6B), key mediators that are considered predictive for spontaneous CMV clearance in transplant recipients (Lisboa et al., 2015). This may contribute to the increased viral infection rate in *A. fumigatus* co-infected dendritic cells (Figure S1B). Collectively, these observations suggest that *A. fumigatus* and CMV cooperate by mutual interference with host inflammasome activation via downregulation of pathogen-associated molecular patterns by evolutionarily conserved host receptors and signaling cascades. Although not the primary focus of this study, the uncovered transcriptional interferences provide a foundation for guiding future targeted studies in transplant recipients aiming to functionally characterize the clinical implications of these interdependencies.

Exploiting the unique potential of multi-organism RNA-seq, our data simultaneously provide insight into the mutual impact of co-infection on gene expression of the two pathogens. We observed CMV-moDC co-expression networks that were exclusive to single-infection but absent from co-infection settings, and we report correlations between *A. fumigatus* and moDC gene expression that occurred selectively during co-infection (Figure 3). These analyses highlighted human innate immunity-associated genes such as *CXCL10*, *CXCL11*, *AIM2*, and *TLR3* that occupied hub-like positions in the resulting inter-species correlation networks (Figure 3; Data S3). These central positions are exemplified by elevated connectivity to further genes within the correlation networks, suggesting an important role of these genes in information transduction of the immune response to the introduced pathogens. Next to the pathogen-specific expression signature of these genes, the additional network relevance adds to their particular suitability as biomarker candidates to pinpoint viral or fungal presence, but requires further investigation.


Figure 5. Validation on RNA and Protein Level

(A) Comparison of RNA-seq-based and quantitative real-time PCR-based relative expression levels for selected genes. Quantitative real-time PCR-based abundance values normalized to delta-aminolevulinate synthase mRNA (see also Figure S5 for further quantitative real-time PCR validations). RNA-seq-based and quantitative real-time PCR-based gene expression values relative to maximum value per gene per replicate; n = 4.

(B) Relative cytokine concentrations in culture supernatants derived from multiplex ELISA (see also Figure S7 for further measured cytokines). Cytokine levels are relative to co-infection for 9 h. n = 4.

(A and B) Error bars indicate SEM. Asterisks indicate significance level after multiple test correction (false discovery rate; FDR): $^{(*)}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.
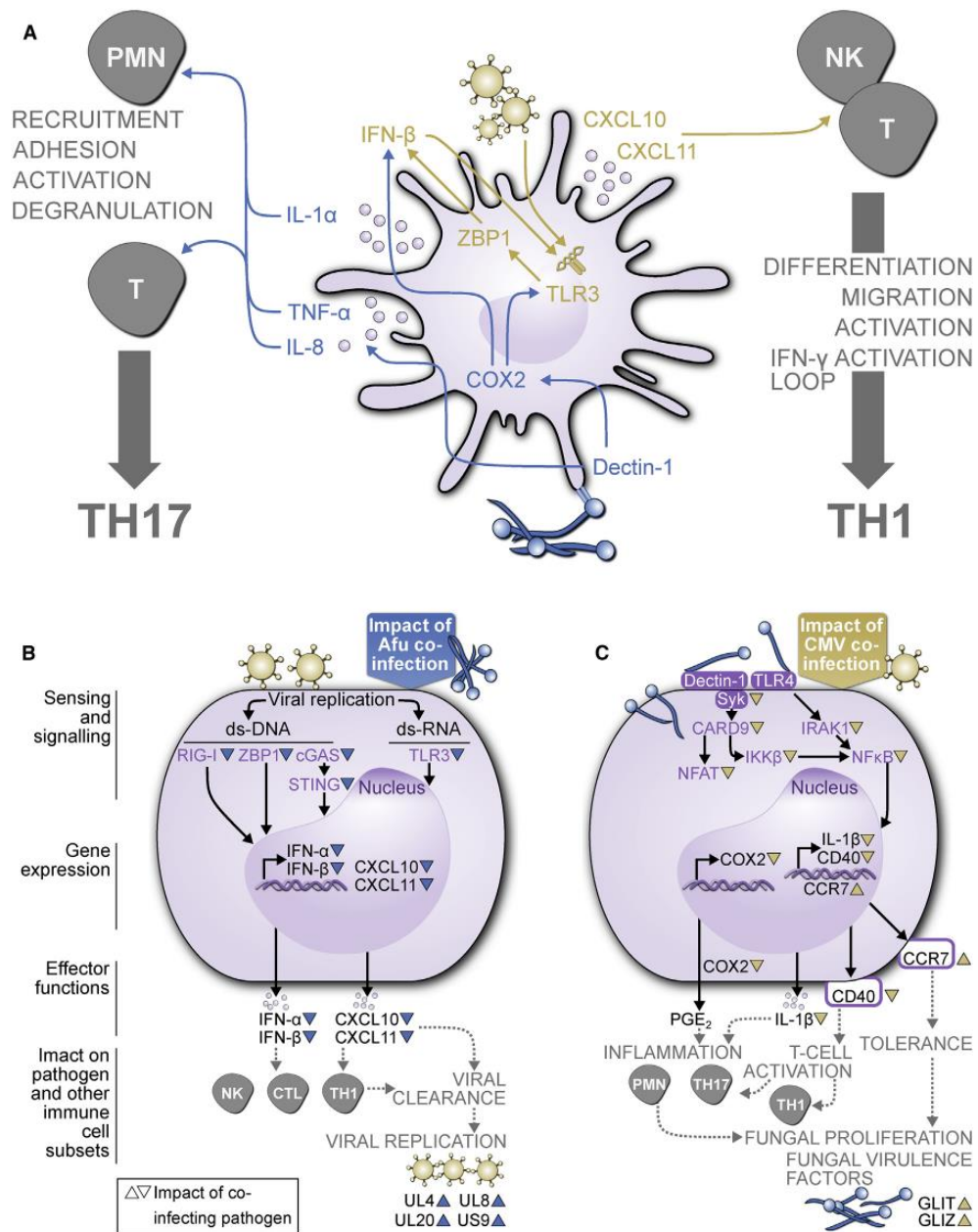
Figure 6. Concerted Defense and Mutual Interference during Co-infection of Monocyte-Derived Dendritic Cells with Fungal and Viral Pathogens

(A) moDC response to fungal and viral challenge. Blue arrows indicate *A. fumigatus*-affected host factors; yellow arrows indicate CMV-affected host factors.
(B) Influence of *A. fumigatus* co-infection on the host response to CMV.
(C) Influence of CMV co-infection on the response to *A. fumigatus*.

PMN, polymorphonuclear leukocytes; NK, natural killer cell; T, lymphocyte T cell; Th, T helper cell; CTL, cytotoxic T cell; CARD9, caspase recruitment domain family member 9; CCR7, C-C motif chemokine receptor 7; cGAS, cyclic GMP-AMP synthase; COX2, cyclooxygenase 2; CXCL, C-X-C motif chemokine ligand;

(legend continued on next page)

The presented triple RNA-seq data represent a rich resource for future exploration. That is, a systematic analysis of expression patterns of human non-immune genes might reveal novel interferences of pulmonary infections with host metabolism, cell cycle, or apoptosis. Moreover, a detailed dissection of differentially expressed fungal or viral genes may lead to the discovery of novel virulence mechanisms in each pathogen.

Perspective

State-of-the-art dual RNA-seq protocols use as few as 10,000 infected cells (Westermann et al., 2017) for ~100–200 ng total RNA, and a similar sensitivity can be expected for the described triple RNA-seq method. This would allow for substantial downscaling of cell samples in the future, allowing for expansion of the analysis to less abundant primary immune cell types such as macrophage and dendritic cell subsets, or alveolar macrophages from bronchoalveolar lavage fluid. Further feasible refinements combining multi-organism RNA-seq and cell sorting or tissue dissection could shift research from the bulk level to phenotypically defined cell populations, for example, to comparatively evaluate transcriptome signatures of infected and uninfected neighboring cells (Westermann and Vogel, 2018). Additionally, triple RNA-seq may be harnessed to investigate the influence of immunosuppressive drugs on co-infected cells or to test how co-infection affects the dendritic cell's ability to cross-present antigens from each pathogen to T cells. Finally, multi-organism RNA-seq holds great promise for expanding our understanding of immune dysregulation in other polymicrobial infections, including post-influenza aspergillosis (Lacoma et al., 2019; Vanderbeke et al., 2018), multifaceted microbial interactions in patients with cystic fibrosis (Sass et al., 2019), or—up-to-the-minute—invasive fungal infections of critically ill coronavirus disease 2019 (COVID-19) patients (Koehler et al., 2020).

In summary, this study established triple RNA-seq to profile transcriptional networks in *A. fumigatus* and CMV infection and to investigate routes of mutual host-pathogen interferences in a co-infection setting. Cost-effectiveness, reproducibility in spite of donor variability, depth, and accuracy propose triple RNA-seq as a powerful tool to probe the emerging concept of direct and immune cell-mediated cross-kingdom inter-dependencies of pathogens in high-risk patients.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead Contact
  - ○ Materials Availability
  - ○ Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Ethics Statement
- METHOD DETAILS
  - ○ Primary cell isolation and differentiation
  - ○ Pulmonary pathogens
  - ○ moDC infection assays
  - ○ Flow cytometry analysis
  - ○ Microscopy
  - ○ Establishment of optimized RNA-seq conditions for single infections
  - ○ RNA extraction, rRNA depletion, cDNA library preparation, and triple RNA-seq
  - ○ Quantitative reverse transcription PCR-based validation of differential gene expression
  - ○ Multiplex cytokine secretion assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ RNA-seq data processing
  - ○ Differential gene expression analysis
  - ○ Interspecies and intraspecies gene expression analyses

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108389.

AUTHOR CONTRIBUTIONS

Conceptualization: A.J.W., J. Loeffler, J.V., S.S., and U.K.; Bioinformatics: B.S., J. Linde, S.S., and T.W.; Methodology/Validation: A.-L.S., E.W., J.B., J.W., L.C.-S., L.B., and L.M.; Visualization: A.J.W., B.S., E.W., L.M., S.S., and S.W.; Resources: L.C.-S., J.B., U.K., C.L., and B.S.; Project Administration/Supervision: A.J.W., G.P., H.E., J. Loeffler, S.S., and U.K.; Funding Acquisition: A.J.W., J.V., J. Loeffler, J. Linde, S.S., and U.K.; Writing – Original Draft: A.J.W., B.S., J. Loeffler, J.W., L.M., L.P., S.S., and S.W.; Writing – Review & Editing: all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Dectin-1, C-type lectin domain containing 7A; GLIT/Z, gliotoxin biosynthetic gene cluster; IFN, interferon; IL, interleukin; IKKβ, inhibitor of nuclear factor kappa B kinase subunit beta; IRAK1, interleukin 1 receptor-associated kinase 1; NFAT, nuclear factor of activated T cells; NFκB, nuclear factor κB; PGE$_2$, prostaglandin E$_2$; RIG-I, retinoic acid inducible gene I; STING, stimulator of interferon genes; Syk, spleen-associated tyrosine kinase; TLR, Toll-like receptor; TNF-α, tumor necrosis factor alpha; UL, member of RL11 CMV gene family; US9, CMV glycoprotein; ZBP1, Z-DNA-binding protein 1.

Cell Reports
Article

 CellPress
OPEN ACCESS

REFERENCES

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

Arvanitis, M., and Mylonakis, E. (2015). Fungal-bacterial interactions and their relevance in health. Cell. Microbiol. *17*, 1442–1446.

Balloy, V., Sallenave, J.M., Wu, Y., Touqui, L., Latgé, J.P., Si-Tahar, M., and Chignard, M. (2008). Aspergillus fumigatus-induced interleukin-8 synthesis by respiratory epithelial cells is controlled by the phosphatidylinositol 3-kinase, p38 MAPK, and ERK1/2 pathways and not by the toll-like receptor-MyD88 pathway. J. Biol. Chem. *283*, 30513–30521.

Bergeron, A.C., Seman, B.G., Hammond, J.H., Archambault, L.S., Hogan, D.A., and Wheeler, R.T. (2017). Candida albicans and Pseudomonas aeruginosa Interact To Enhance Virulence of Mucosal Infection in Transparent Zebrafish. Infect. Immun. *85*, e00475-17.

Braedel, S., Radsak, M., Einsele, H., Latgé, J.P., Michan, A., Loeffler, J., Haddad, Z., Grigoleit, U., Schild, H., and Hebart, H. (2004). Aspergillus fumigatus antigens activate innate immune cells via toll-like receptors 2 and 4. Br. J. Haematol. *125*, 392–399.

Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., and Lynn, D.J. (2013). InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res. *41*, D1228–D1233.

Caffrey, A.K., Lehmann, M.M., Zickovich, J.M., Espinosa, V., Shepardson, K.M., Watschke, C.P., Hilmer, K.M., Thammahong, A., Barker, B.M., Rivera, A., et al. (2015). IL-1α signaling is critical for leukocyte recruitment after pulmonary Aspergillus fumigatus challenge. PLoS Pathog. *11*, e1004625.

Caffrey-Carr, A.K., Kowalski, C.H., Beattie, S.R., Blaseg, N.A., Upshaw, C.R., Thammahong, A., Lust, H.E., Tang, Y.W., Hohl, T.M., Cramer, R.A., and Obar, J.J. (2017). Interleukin 1α Is Critical for Resistance against Highly Virulent Aspergillus fumigatus Isolates. Infect. Immun. *85*, e00661-17.

Camargo, J.F., and Komanduri, K.V. (2017). Emerging concepts in cytomegalovirus infection following hematopoietic stem cell transplantation. Hematol. Oncol. Stem Cell Ther. *10*, 233–238.

Cerqueira, G.C., Arnaud, M.B., Inglis, D.O., Skrzypek, M.S., Binkley, G., Simison, M., Miyasato, S.R., Binkley, J., Orvis, J., Shah, P., et al. (2014). The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. Nucleic Acids Res. *42*, D705–D710.

Chang, W.L., Baumgarth, N., Yu, D., and Barry, P.A. (2004). Human cytomegalovirus-encoded interleukin-10 homolog inhibits maturation of dendritic cells and alters their functionality. J. Virol. *78*, 8720–8731.

Cheeran, M.C., Hu, S., Sheng, W.S., Peterson, P.K., and Lokensgard, J.R. (2003). CXCL10 production from cytomegalovirus-stimulated microglia is regulated by both human and viral interleukin-10. J. Virol. *77*, 4502–4515.

Cheung, A.K., Gottlieb, D.J., Plachter, B., Pepperl-Klindworth, S., Avdic, S., Cunningham, A.L., Abendroth, A., and Slobedman, B. (2009). The role of the human cytomegalovirus UL111A gene in down-regulating CD4+ T-cell recognition of latently infected cells: implications for virus elimination during latency. Blood *114*, 4128–4137.

Ching, T., Huang, S., and Garmire, L.X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. RNA *20*, 1684–1696.

Choi, Y.J., Aliota, M.T., Mayhew, G.F., Erickson, S.M., and Christensen, B.M. (2014). Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions. PLoS Negl. Trop. Dis. *8*, e2905.

Choi, H.J., Park, A., Kang, S., Lee, E., Lee, T.A., Ra, E.A., Lee, J., Lee, S., and Park, B. (2018). Human cytomegalovirus-encoded US9 targets MAVS and STING signaling to evade type I interferon immune responses. Nat. Commun. *9*, 125.

Colgan, A.M., Cameron, A.D., and Kröger, C. (2017). If it transcribes, we can sequence it: mining the complexities of host-pathogen-environment interactions using RNA-seq. Curr. Opin. Microbiol. *36*, 37–46.

Cortez, K.J., Lyman, C.A., Kottilil, S., Kim, H.S., Roilides, E., Yang, J., Fullmer, B., Lempicki, R., and Walsh, T.J. (2006). Functional genomics of innate host defense molecules in normal human monocytes in response to Aspergillus fumigatus. Infect. Immun. *74*, 2353–2365.

de la Cámara, R. (2016). CMV in Hematopoietic Stem Cell Transplantation. Mediterr. J. Hematol. Infect. Dis. *8*, e2016031.

Drgona, L., Khachatryan, A., Stephens, J., Charbonneau, C., Kantecki, M., Haider, S., and Barnes, R. (2014). Clinical and economic burden of invasive fungal diseases in Europe: focus on pre-emptive and empirical treatment of Aspergillus and Candida species. Eur. J. Clin. Microbiol. Infect. Dis. *33*, 7–21.

Duarte, R.F., and Lyon, S. (2018). Novel approaches to CMV after HCT: report from the 27th European Congress of Clinical Microbiology and Infectious Diseases, Vienna, Austria, 22-25 April 2017. Future Sci. OA *4*, FSO296.

Dunn, W., Chou, C., Li, H., Hai, R., Patterson, D., Stolc, V., Zhu, H., and Liu, F. (2003). Functional profiling of a human cytomegalovirus genome. Proc. Natl. Acad. Sci. USA *100*, 14223–14228.

Freeman, R.B., Jr. (2009). The 'indirect' effects of cytomegalovirus infection. Am. J. Transplant. *9*, 2453–2458.

Garcia-Vidal, C., Upton, A., Kirby, K.A., and Marr, K.A. (2008). Epidemiology of invasive mold infections in allogeneic stem cell transplant recipients: biological risk factors for infection according to time after transplantation. Clin. Infect. Dis. *47*, 1041–1050.

Gredmark-Russ, S., and Söderberg-Nauclér, C. (2012). Dendritic cell biology in human cytomegalovirus infection and the clinical consequences for host immunity and pathology. Virulence *3*, 621–634.

Green, M.L., Leisenring, W., Xie, H., Mast, T.C., Cui, Y., Sandmaier, B.M., Sorror, M.L., Goyal, S., Özkök, S., Yi, J., et al. (2016). Cytomegalovirus viral load and mortality after haemopoietic stem cell transplantation in the era of pre-emptive therapy: a retrospective cohort study. Lancet Haematol. *3*, e119–e127.

Grow, W.B., Moreb, J.S., Roque, D., Manion, K., Leather, H., Reddy, V., Khan, S.A., Finiewicz, K.J., Nguyen, H., Clancy, C.J., et al. (2002). Late onset of invasive aspergillus infection in bone marrow transplant patients at a university hospital. Bone Marrow Transplant. *29*, 15–19.

Hahn, G., Jores, R., and Mocarski, E.S. (1998). Cytomegalovirus remains latent in a common precursor of dendritic and myeloid cells. Proc. Natl. Acad. Sci. USA *95*, 3937–3942.

Harrison, J.H., Merrill, J.P., and Murray, J.E. (1956). Renal homotransplantation in identical twins. Surg. Forum *6*, 432–436.

Hsieh, S.H., Kurzai, O., and Brock, M. (2017). Persistence within dendritic cells marks an antifungal evasion and dissemination strategy of Aspergillus terreus. Sci. Rep *7*, 10590.

Huang, C., and Levitz, S.M. (2000). Stimulation of macrophage inflammatory protein-1alpha, macrophage inflammatory protein-1beta, and RANTES by Candida albicans and Cryptococcus neoformans in peripheral blood mononuclear cells from persons with and without human immunodeficiency virus infection. J. Infect. Dis. *181*, 791–794.

Jarvis, M.A., and Nelson, J.A. (2002). Mechanisms of human cytomegalovirus persistence and latency. Front. Biosci *7*, d1575–d1582.

Juranic Lisnic, V., Babic Cac, M., Lisnic, B., Trsan, T., Mefferd, A., Das Mukhopadhyay, C., Cook, C.H., Jonjic, S., and Trgovcich, J. (2013). Dual analysis of the murine cytomegalovirus and host cell transcriptomes reveal new aspects of the virus-host cell interface. PLoS Pathog. *9*, e1003611.

Kaminski, H., and Fishman, J.A. (2016). The Cell Biology of Cytomegalovirus: Implications for Transplantation. Am. J. Transplant. *16*, 2254–2269.

Kasmapour, B., Kubsch, T., Rand, U., Eiz-Vesper, B., Messerle, M., Vondran, F.W.R., Wiegmann, B., Haverich, A., and Cicin-Sain, L. (2017). Myeloid Dendritic Cells Repress Human Cytomegalovirus Gene Expression and Spread by Releasing Interferon-Unrelated Soluble Antiviral Factors. J. Virol. *92*, e01138-17.

Knoblach, T., Grandel, B., Seiler, J., Nevels, M., and Paulus, C. (2011). Human cytomegalovirus IE1 protein elicits a type II interferon-like host cell response that depends on activated STAT1 but not interferon-γ. PLoS Pathog. *7*, e1002016.

Koehler, P., Cornely, O.A., Böttiger, B.W., Dusse, F., Eichenauer, D.A., Fuchs, F., Hallek, M., Jung, N., Klein, F., Persigehl, T., et al. (2020). COVID-19 associated pulmonary aspergillosis. Mycoses *63*, 528–534.

Kotenko, S.V., Saccani, S., Izotova, L.S., Mirochnitchenko, O.V., and Pestka, S. (2000). Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). Proc. Natl. Acad. Sci. USA *97*, 1695–1700.

Kumar, M., Roe, K., Orillo, B., Muruve, D.A., Nerurkar, V.R., Gale, M., Jr., and Verma, S. (2013). Inflammasome adaptor protein Apoptosis-associated speck-like protein containing CARD (ASC) is critical for the immune response and survival in west Nile virus encephalitis. J. Virol. *87*, 3655–3667.

Lacoma, A., Mateo, L., Blanco, I., Méndez, M.J., Rodrigo, C., Latorre, I., Villar-Hernandez, R., Domínguez, J., and Prat, C. (2019). Impact of Host Genetics and Biological Response Modifiers on Respiratory Tract Infections. Front. Immunol. *10*, 1013.

Lass-Flörl, C., Roilides, E., Löffler, J., Wilflingseder, D., and Romani, L. (2013). Minireview: host defence in invasive aspergillosis. Mycoses *56*, 403–413.

Latgé, J.P. (1999). Aspergillus fumigatus and aspergillosis. Clin. Microbiol. Rev. *12*, 310–350.

Lio, C.W., McDonald, B., Takahashi, M., Dhanwani, R., Sharma, N., Huang, J., Pham, E., Benedict, C.A., and Sharma, S. (2016). cGAS-STING Signaling Regulates Initial Innate Control of Cytomegalovirus Infection. J. Virol. *90*, 7789–7797.

Lisboa, L.F., Egli, A., Fairbanks, J., O'Shea, D., Manuel, O., Husain, S., Kumar, D., and Humar, A. (2015). CCL8 and the Immune Control of Cytomegalovirus in Organ Transplant Recipients. Am. J. Transplant. *15*, 1882–1892.

Liu, Y., Zhou, J., and White, K.P. (2014). RNA-seq differential expression studies: more sequence or more replication? Bioinformatics *30*, 301–304.

Loewendorf, A., and Benedict, C.A. (2010). Modulation of host innate and adaptive immune defenses by cytomegalovirus: timing is everything. J. Intern. Med. *267*, 483–501.

Lother, J., Breitschopf, T., Krappmann, S., Morton, C.O., Bouzani, M., Kurzai, O., Gunzer, M., Hasenberg, M., Einsele, H., and Loeffler, J. (2014). Human dendritic cell subsets display distinct interactions with the pathogenic mould Aspergillus fumigatus. Int. J. Med. Microbiol. *304*, 1160–1168.

Man, S.M., Karki, R., and Kanneganti, T.D. (2016). AIM2 inflammasome in infection, cancer, and autoimmunity: Role in DNA sensing, inflammation, and innate immunity. Eur. J. Immunol. *46*, 269–280.

Marr, K.A., Carter, R.A., Boeckh, M., Martin, P., and Corey, L. (2002). Invasive aspergillosis in allogeneic stem cell transplant recipients: changes in epidemiology and risk factors. Blood *100*, 4358–4366.

Marshall, E.E., and Geballe, A.P. (2009). Multifaceted evasion of the interferon response by cytomegalovirus. J. Interferon Cytokine Res. *29*, 609–619.

Martino, R., Piñana, J.L., Parody, R., Valcarcel, D., Sureda, A., Brunet, S., Briones, J., Delgado, J., Sánchez, F., Rabella, N., and Sierra, J. (2009). Lower respiratory tract respiratory virus infections increase the risk of invasive aspergillosis after a reduced-intensity allogeneic hematopoietic SCT. Bone Marrow Transplant. *44*, 749–756.

Martinon, F., and Tschopp, J. (2007). Inflammatory caspases and inflammasomes: master switches of inflammation. Cell Death Differ. *14*, 10–22.

McNab, F., Mayer-Barber, K., Sher, A., Wack, A., and O'Garra, A. (2015). Type I interferons in infectious disease. Nat. Rev. Immunol. *15*, 87–103.

Mehrad, B., Strieter, R.M., and Standiford, T.J. (1999). Role of TNF-alpha in pulmonary host defense in murine invasive aspergillosis. J. Immunol. *162*, 1633–1640.

Mezger, M., Kneitz, S., Wozniok, I., Kurzai, O., Einsele, H., and Loeffler, J. (2008). Proinflammatory response of immature human dendritic cells is mediated by dectin-1 after exposure to Aspergillus fumigatus germ tubes. J. Infect. Dis. *197*, 924–931.

Mezger, M., Bonin, M., Kessler, T., Gebhardt, F., Einsele, H., and Loeffler, J. (2009). Toll-like receptor 3 has no critical role during early immune response of human monocyte-derived dendritic cells after infection with the human cytomegalovirus strain TB40E. Viral Immunol. *22*, 343–351.

Mikulska, M., Raiola, A.M., Bruno, B., Furfaro, E., Van Lint, M.T., Bregante, S., Ibatici, A., Del Bono, V., Bacigalupo, A., and Viscoli, C. (2009). Risk factors for invasive aspergillosis and related mortality in recipients of allogeneic SCT from alternative donors: an analysis of 306 patients. Bone Marrow Transplant. *44*, 361–370.

Miller-Kittrell, M., and Sparer, T.E. (2009). Feeling manipulated: cytomegalovirus immune manipulation. Virol. J. *6*, 4.

Morton, C.O., Varga, J.J., Hornbach, A., Mezger, M., Sennefelder, H., Kneitz, S., Kurzai, O., Krappmann, S., Einsele, H., Nierman, W.C., et al. (2011). The temporal dynamics of differential gene expression in Aspergillus fumigatus interacting with human immature dendritic cells in vitro. PLoS ONE *6*, e16016.

Moutaftsi, M., Mehl, A.M., Borysiewicz, L.K., and Tabi, Z. (2002). Human cytomegalovirus inhibits maturation and impairs function of monocyte-derived dendritic cells. Blood *99*, 2913–2921.

Murayama, T., Li, Y., Takahashi, T., Yamada, R., Matsubara, K., Tuchida, Y., Li, Z., and Sadanari, H. (2012). Anti-cytomegalovirus effects of tricin are dependent on CXCL11. Microbes Infect. *14*, 1086–1092.

Nichols, W.G., Corey, L., Gooley, T., Davis, C., and Boeckh, M. (2002). High risk of death due to bacterial and fungal infection among cytomegalovirus (CMV)-seronegative recipients of stem cell transplants from seropositive donors: evidence for indirect effects of primary CMV infection. J. Infect. Dis. *185*, 273–282.

Onomoto, K., Onoguchi, K., Takahasi, K., and Fujita, T. (2010). Type I interferon production induced by RIG-I-like receptors. J. Interferon Cytokine Res. *30*, 875–881.

Paijo, J., Döring, M., Spanier, J., Grabski, E., Nooruzzaman, M., Schmidt, T., Witte, G., Messerle, M., Hornung, V., Kaever, V., and Kalinke, U. (2016). cGAS Senses Human Cytomegalovirus and Induces Type I Interferon Responses in Human Monocyte-Derived Cells. PLoS Pathog. *12*, e1005546.

Picarda, G., and Benedict, C.A. (2018). Cytomegalovirus: Shape-Shifting the Immune System. J. Immunol. *200*, 3881–3889.

Pittman, K.J., Aliota, M.T., and Knoll, L.J. (2014). Dual transcriptional profiling of mice and Toxoplasma gondii during acute and chronic infection. BMC Genomics *15*, 806.

Raftery, M.J., Wieland, D., Gronewald, S., Kraus, A.A., Giese, T., and Schönrich, G. (2004). Shaping phenotype, function, and survival of dendritic cells by cytomegalovirus-encoded IL-10. J. Immunol. *173*, 3383–3391.

Reese, T.A., Bi, K., Kambal, A., Filali-Mouhim, A., Beura, L.K., Bürger, M.C., Pulendran, B., Sekaly, R.P., Jameson, S.C., Masopust, D., et al. (2016). Sequential Infection with Common Pathogens Promotes Human-like Immune Gene Expression and Altered Vaccine Response. Cell Host Microbe *19*, 713–719.

Roilides, E., Dimitriadou-Georgiadou, A., Sein, T., Kadiltsoglou, I., and Walsh, T.J. (1998). Tumor necrosis factor alpha enhances antifungal activities of polymorphonuclear and mononuclear phagocytes against Aspergillus fumigatus. Infect. Immun. *66*, 5999–6003.

Saliba, A.-E., Santos, S.C., and Vogel, J. (2017). New RNA-seq approaches for the study of bacterial pathogens. Curr. Opin. Microbiol. *35*, 78–87.

Sallusto, F., and Lanzavecchia, A. (1994). Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. J. Exp. Med. *179*, 1109–1118.

Sampaio, K.L., Weyell, A., Subramanian, N., Wu, Z., and Sinzger, C. (2017). A TB40/E-derived human cytomegalovirus genome with an intact US-gene region and a self-excisable BAC cassette for immunological research. Biotechniques *63*, 205–214.

Sass, G., Nazik, H., Penner, J., Shah, H., Ansari, S.R., Clemons, K.V., Groleau, M.C., Dietl, A.M., Visca, P., Haas, H., et al. (2019). Aspergillus-Pseudomonas

40

Cell Reports
Article

interaction, relevant to competition in airways. Med. Mycol. 57 (Suppl 2), S228–S232.

Seelbinder, B., Wolf, T., Priebe, S., McNamara, S., Gerber, S., Guthke, R., and Linde, J. (2019). GEO2RNAseq: An easy-to-use R pipeline for complete pre-processing of RNA-seq data. bioRxiv. https://doi.org/10.1101/771063.

Shankar, J., Cerqueira, G.C., Wortman, J.R., Clemons, K.V., and Stevens, D.A. (2018). RNA-Seq Profile Reveals Th-1 and Th-17-Type of Immune Responses in Mice Infected Systemically with Aspergillus fumigatus. Mycopathologia 183, 645–658.

Singh, A.K., and McGuirk, J.P. (2016). Allogeneic Stem Cell Transplantation: A Historical and Scientific Overview. Cancer Res. 76, 6445–6451.

Solak, Y., Biyik, Z., Cizmecioglu, A., Genc, N., Ozbek, O., Gaipov, A., and Yek-san, M. (2013). Cytomegalovirus and Aspergillus spp. coinfection in organ transplantation: a case report and review of the literature. CEN Case Rep. 2, 59–67.

Stanzani, M., Orciuolo, E., Lewis, R., Kontoyiannis, D.P., Martins, S.L., St John, L.S., and Komanduri, K.V. (2005). Aspergillus fumigatus suppresses the human cellular immune response via gliotoxin-mediated apoptosis of mono-cytes. Blood 105, 2258–2265.

Taylor, P.R., Tsoni, S.V., Willment, J.A., Dennehy, K.M., Rosas, M., Findon, H., Haynes, K., Steele, C., Botto, M., Gordon, S., and Brown, G.D. (2007). Dectin-1 is required for beta-glucan recognition and control of fungal infection. Nat. Immunol. 8, 31–38.

Taylor-Wiedeman, J., Sissons, J.G., Borysiewicz, L.K., and Sinclair, J.H. (1991). Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells. J. Gen. Virol. 72, 2059–2064.

Tyner, J.W., Uchida, O., Kajiwara, N., Kim, E.Y., Patel, A.C., O'Sullivan, M.P., Walter, M.J., Schwendener, R.A., Cook, D.N., Danoff, T.M., and Holtzman, M.J. (2005). CCL5-CCR5 interaction provides antiapoptotic signals for macro-phage survival during viral infection. Nat. Med. 11, 1180–1187.

Ullmann, A.J., Schmidt-Hieber, M., Bertz, H., Heinz, W.J., Kiehl, M., Krüger, W., Mousset, S., Neuburger, S., Neumann, S., Penack, O., et al.; Infectious Diseases Working Party of the German Society for Hematology and Medical Oncology (AGIHO/DGHO) and the DAG-KBT (German Working Group for Blood and Marrow Transplantation) (2016). Infectious diseases in allogeneic haematopoietic stem cell transplantation: prevention and prophylaxis strategy guidelines 2016. Ann. Hematol. 95, 1435–1455.

Upton, A., Kirby, K.A., Carpenter, P., Boeckh, M., and Marr, K.A. (2007). Inva-sive aspergillosis following hematopoietic cell transplantation: outcomes and prognostic factors associated with mortality. Clin. Infect. Dis. 44, 531–540.

Vanderbeke, L., Spriet, I., Breynaert, C., Rijnders, B.J.A., Verweij, P.E., and Wauters, J. (2018). Invasive pulmonary aspergillosis complicating severe influ-enza: epidemiology, diagnosis and treatment. Curr. Opin. Infect. Dis. 31, 471–480.

Walsh, T.J., Roilides, E., Cortez, K., Kottilil, S., Bailey, J., and Lyman, C.A. (2005). Control, immunoregulation, and expression of innate pulmonary host defenses against Aspergillus fumigatus. Med. Mycol. 43 (Suppl 1), S165–S172.

Wesolowska-Andersen, A., Everman, J.L., Davidson, R., Rios, C., Herrin, R., Eng, C., Janssen, W.J., Liu, A.H., Oh, S.S., Kumar, R., et al. (2017). Dual RNA-seq reveals viral infections in asthmatic children without respiratory illness which are associated with changes in the airway transcriptome. Genome Biol. 18, 12.

Westermann, A.J., and Vogel, J. (2018). Host-Pathogen Transcriptomics by Dual RNA-Seq. Methods Mol. Biol. 1737, 59–75.

Westermann, A.J., Gorski, S.A., and Vogel, J. (2012). Dual RNA-seq of path-ogen and host. Nat. Rev. Microbiol. 10, 618–630.

Westermann, A.J., Barquist, L., and Vogel, J. (2017). Resolving host-pathogen interactions by dual RNA-seq. PLoS Pathog. 13, e1006033.

Wiertz, E.J., Jones, T.R., Sun, L., Bogyo, M., Geuze, H.J., and Ploegh, H.L. (1996). The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. Cell 84, 769–779.

Wolf, T., Kämmer, P., Brunke, S., and Linde, J. (2018). Two's company: study-ing interspecies relationships with dual RNA-seq. Curr. Opin. Microbiol. 42, 7–12.

Yang, D., Liang, Y., Zhao, S., Ding, Y., Zhuang, Q., Shi, Q., Ai, T., Wu, S.Q., and Han, J. (2020). ZBP1 mediates interferon-induced necroptosis. Cell. Mol. Im-munol. 17, 356–368.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for poly-merase chain reaction. BMC Bioinformatics 13, 134.

Yong, M.K., Slavin, M.A., and Kontoyiannis, D.P. (2018). Invasive fungal dis-ease and cytomegalovirus infection: is there an association? Curr. Opin. Infect. Dis. 31, 481–489.

41

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-human CD40-PE-Vio®770 (recombinant human IgG1, REAfinity)	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-110-948, Clone: monoclonal REA733; RRID:AB_2658003
anti-human CD80-APC (recombinant human IgG1, REAfinity)	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-117-719, Clone: monoclonal REA661; RRID:AB_2751414
anti-human CD209-VioBlue® (recombinant human IgG1, REAfinity)	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-110-456, Clone: monoclonal REA617; RRID:AB_2656252
anti-human CCR7-VioBlue® (recombinant human IgG1, REAfinity)	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-117-353, Clone: monoclonal REA546; RRID:AB_2733933
anti-human TLR2-PE-Vio®770 (recombinant human IgG1, REAfinity)	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-099-022, Clone: monoclonal REA109; RRID:AB_2656975
anti-human TLR3-APC (mouse IgG1κ)	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-096-885, Clone: monoclonal TLR3.7; RRID:AB_2660004
Bacterial and Virus Strains		
human CMV strain TB40/E-mNeonGreen	Helmholtz Centre for Infection Research, Braunschweig, Germany Kasmapour et al., 2017	https://www.helmholtz-hzi.de
human CMV strain TB40/SE	University Freiburg, Medical Center, Institute of Virology, Freiburg, Germany; Sampaio et al., 2017	https://www.uniklinik-freiburg.de/de.html
Biological Samples		
Human peripheral venous blood from healthy adult donors for generation of monocyte-derived dendritic cells	Paijo et al., 2016; Sallusto and Lanzavecchia, 1994	N/A
Chemicals, Peptides, and Recombinant Proteins		
RNAprotect® Cell Reagent	QIAGEN, Hilden, Germany	Cat#76526
Viobility 405/520 Fixable Dye	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-109-814
iTaq Universal SYBR® Green Supermix	Bio-Rad Laboratories GmbH, Feldkirchen, Germany	Cat#1725124
Recombinant human IL-4, premium grade	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-093-922
Recombinant human GM-CSF, premium grade	Miltenyi Biotec, Bergisch Gladbach, Germany	Cat#130-093-866
Critical Commercial Assays		
RiboPure-Yeast Kit	Thermo Fisher Scientific, Waltham, MA, USA	Cat#AM1926
DNase I, RNase free (1.000 U)	Thermo Fisher Scientific, Waltham, MA, USA	Cat#EN0521
M-MLV reverse transcriptase	Invitrogen, Carlsbad, CA, USA	Cat#28025013
cDNA First Strand Synthesis Kit	Thermo Fisher Scientific, Waltham, MA, USA	Cat#K1612
Ribo-Zero Gold rRNA removal kit (human, mouse, rat) (24 reactions)	Illumina, San Diego, CA, USA	Cat#MRZG12324
ProcartaPlex 16-plex Immunoassay (IFN-α, IFN-β, IFN-γ, IL-1α, IL-1β, IL-2, IL-6, IL-8, IL-10, IL-12p70, IL-17A, IL-23, CXCL10, CXCL11, CCL5, TNF-α)	Thermo Fisher Scientific, Waltham, MA, USA	Cat#PPX-16, customized

Cell Reports
Article

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Triple RNA-seq, *A. fumigatus*, CMV and *H. sapiens*	https://www.ncbi.nlm.nih.gov/geo/	GSE134344
Dual RNA-seq *A. fumigatus* and *H. sapiens*	https://www.ncbi.nlm.nih.gov/geo/	GSE135450
Dual RNA-seq CMV and *H. sapiens*	https://www.ncbi.nlm.nih.gov/geo/	GSE136217
Experimental Models: Organisms/Strains		
Aspergillus fumigatus (wildtype)	ATCC	ATCC: 46645
Aspergillus fumigatus dTomato	ATCC; Lother et al., 2014	ATCC: 46645
Aspergillus fumigatus GFP	ATCC; Lother et al., 2014	ATCC: 46645
Oligonucleotides		
See Table S1 for primer sequences (ALAS1, IL15, CCR7, ZBP1, VEGFA, IFNG, TLR3, CD40, GliT, GliZ, Ilv3, Cat1, SrbA, UL4, UL8, UL20, US16, UL9, US9)	This paper	N/A
Software and Algorithms		
CellQuest Pro Software (Becton & Dickinson)	https://www.bd.com/en-us	Becton & Dickinson, Franklin Lakes, NJ, USA
FACSDiva Software	https://www.bd.com/en-us	Becton & Dickinson, Franklin Lakes, NJ, USA
FlowJo Software, v10	https://www.flowjo.com; https://www.bd.com/en-us	Treestar/ Becton & Dickinson, Ashland, OR, USA
FCS Express Software, v7	https://denovosoftware.com	De Novo Software, Pasedena, CA, USA
NIS Elements Imaging software, v5.02.00	https://www.microscope.healthcare.nikon.com	Nikon Instruments, Amsterdam, Netherlands
R	https://cran.r-project.org/	3.5.1
featureCounts	https://cran.r-project.org/	Rsubread 1.28.0
DESeq	https://bioconductor.org	1.30.0
DESeq2	https://bioconductor.org	1.18.1
limma voom	https://bioconductor.org	3.34.6
edgeR	https://bioconductor.org	3.20.7
geo2RNaseq	https://singularity-hub.org/accounts/login/?next=/collections/4387; https://bitbucket.org/Xentrics/geo2rnaseq/src/master/	0.9.12
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	0.11.8
Trimmomatic	http://www.usadellab.org/cms/?page=trimmomatic	0.36
HiSat2	https://daehwankimlab.github.io/hisat2/	2.1.0
SAMtools	https://www.htslib.org/	1.7
MultiQC	https://multiqc.info/	1.5
Jupyter	https://jupyter.org/	4.4.0
Tidyverse	https://cran.r-project.org/	1.2.1
Magrittr	https://cran.r-project.org/	1.5.0
Reshape2	https://cran.r-project.org/	1.4.4
ggplot2	https://cran.r-project.org/	3.1.1
Ggpubr	https://cran.r-project.org/	0.2.5
Ggsci	https://cran.r-project.org/	2.9
Dplyr	https://cran.r-project.org/	0.7.6
Tidyr	https://cran.r-project.org/	1.3.1
Stringr	https://cran.r-project.org/	1.3.0

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AnnotationDbi	https://bioconductor.org	1.44.0
org.Hs.eg.db	https://bioconductor.org	3.10.0
setRank	https://cran.r-project.org/	1.1.0
Pathview	https://bioconductor.org	1.26.0
Pheatmap	https://cran.r-project.org/	1.0.12
igraph	https://cran.r-project.org/	1.2.5
WriteXLS	https://cran.r-project.org/	5.0.0
corrplot	https://cran.r-project.org/	0.84
RcolorBrewer	https://cran.r-project.org/	1.1-2
enrichR	https://cran.r-project.org/	2.1.0
cowplot	https://cran.r-project.org/	1.1.0
Other		
H. sapiens reference genome	https://www.ncbi.nlm.nih.gov/	GRCH 38 v89
A. fumigatus Af293 reference genome	http://www.aspergillusgenome.org/	s03-m05-r09
CMV reference genome	https://www.ncbi.nlm.nih.gov/	EF999921.1

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Juergen Loeffler (loeffler_j@ukw.de).

Materials Availability
All primary material generated in this study will be made available upon request following publication. A completed Materials Transfer Agreement might be necessary, especially if there is potential for commercial application.

Data and Code Availability
All primary sequencing data and processed data described in this manuscript have been deposited in the NCBI Gene Expression Omnibus under the accession numbers GEO: GSE134344, GSE135450 and GSE136217.
   Code for preprocessing RNA sequencing data and analysis is deposited and available at https://github.com/SchSascha/manuscript_tripleRNAseq.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Healthy blood donors were exclusively in the age between 18-59 years and otherwise excluded from providing blood samples. Both sexes were included in the study equally and at random; the sex of the blood donors was anonymised and represents the normal distribution within this population.

Ethics Statement
The processing of human peripheral venous blood from healthy adult donors was approved by the Ethical Committee of the University Hospital Würzburg (#302/12).

METHOD DETAILS

Primary cell isolation and differentiation
To generate moDCs, monocytes were isolated from leukoreduction system chambers containing blood from healthy volunteers and standard density-gradient centrifugation followed by positive selection using magnetic-activated cell sorting (CD14 MicroBeads, human, Miltenyi Biotec). Monocytes were cultured in CellGenix GMP dendritic cell medium (serum-free, CellGenix) supplemented with 120 μg gentamicin (Refobacin, Merck) in 24-well plates with 1×10^6 cells/ml. Cells were differentiated for 6 days by addition of 1,000 U/ml granulocyte macrophage-colony stimulating factor (Miltenyi Biotec) and 1,000 U/ml interleukin (IL)-4 (Miltenyi Biotec) (Mezger et al., 2008; Paijo et al., 2016; Sallusto and Lanzavecchia, 1994).

Cell Reports
Article

Pulmonary pathogens

Reporter strain human CMV TB40/E-mNeonGreen and CMV TB40/SE (non-fluorescent) were generated according to published protocols (Kasmapour et al., 2017; Paijo et al., 2016; Sampaio et al., 2017). *A. fumigatus* (ATCC 46645) germ tubes were prepared overnight in RPMI (Invitrogen) and used for dual or triple RNA-seq. For flow cytometry and microscopy (see below), dTomato- or GFP-expressing *A. fumigatus* germ tubes were used.

moDC infection assays

Primary human cells were infected with CMV at a multiplicity of infection (MOI) 3 and with subsequent centrifugal enhancement at 300 x *g* for 30 min. *A. fumigatus* germ tubes, which represent an invasive immunogenic morphotype of the fungus, were added at MOI 0.5. For co-infections, pathogens were added simultaneously or subsequently (Figure 1A). After harvesting, cells were centrifuged and pellets collected in HBSS (Invitrogen) or RNA*protect* Cell Reagent (QIAGEN), and cell-free supernatants were stored at −80°C. For flow cytometry and microscopy, (see below) moDCs were infected with CMV for 24 h before adding germ tubes, because detection of NeonGreen fluorescence in infected cells requires sufficient replication of the virus.

Flow cytometry analysis

CMV and *A. fumigatus* infection rates were determined by flow cytometry, measuring fluorescent signals of moDCs positive for mNeonGreen (CMV) or GFP (*A. fumigatus*), respectively (Figures S1B and S1C). Viability of moDCs was determined by staining with Viobility 405/520 Fixable Dye (Miltenyi Biotec), defining negative populations as viable. Additionally, cells were stained with anti-CD40-PE-Vio770 (REA733, Miltenyi Biotec), anti-CD80-APC (REA661, Miltenyi Biotec), anti-CD209-VioBlue, (REA617, Miltenyi Biotec), anti-CCR7-VioBlue (REA546, Miltenyi Biotec), anti-TLR2-PE-Vio770 (REA109, Miltenyi Biotec) or anti-TLR3-APC (TLR3.7, Miltenyi Biotec). Mean fluorescence intensities of these surface markers were measured and isotypes subtracted. Data were acquired on a FACSCalibur using CellQuest Software (Becton & Dickinson) or FACSCanto II using FACSDiva software (Becton & Dickinson). Data analysis was done with FlowJo (Treestar/Becton & Dickinson, version 10) or FCS Express 7 (*De Novo* Software).

Microscopy

Morphology and fluorescence signals for CMV (mNeonGreen) or *A. fumigatus* (dTomato) were analyzed in 48-well cell culture plates using a Nikon Eclipse Ti microscope (Nikon) with an Okolab incubator set at 37°C. Images were obtained at 20-fold magnification and processed using NIS Elements Imaging software (Nikon, version 5.02.00).

Establishment of optimized RNA-seq conditions for single infections

As a prerequisite for multi-organism RNA-seq analysis in our infection model, lysis conditions should be sufficiently harsh to disrupt cellular membranes of all interacting organisms, but sufficiently mild to maintain high RNA integrity. Additionally, multiplicities of infection (MOIs) must be adjusted to ensure homogeneous coverage of individual transcriptomes in the resulting sequencing data. That is, the relative proportions of transcriptomes in isolated RNA samples should match the ratio of the respective genome sizes. Therefore, we first evaluated yield and quality of RNA isolated from moDCs infected with either *A. fumigatus* or CMV at different time points and MOIs, and observed high RNA integrity (RIN > 7) for all conditions (Figures S2A and S2B).

While highly abundant in any organism (> 90% of total cellular RNA), ribosomal RNA (rRNA) provides little informative value about cellular physiology. Therefore, ribosomal transcripts are typically depleted from sequencing libraries, either by active rRNA pull-out (e.g., Ribo-Zero technology) or enrichment of polyadenylated transcripts. In principle, both options are suitable for our infection model, since mRNAs of all three interacting organisms are polyadenylated. We found that both approaches efficiently depleted human ribosomal reads (Figure S2C). However, to retain potentially interesting non-polyadenylated transcripts (e.g., microRNAs [miRNAs], small nucle(ol)ar RNAs [snRNAs, snoRNAs] and polyA- long noncoding RNAs [lncRNAs]; Figure S2C), we employed Ribo-Zero technology for further RNA-seq experiments.

After rRNA removal from single-infection samples, cDNA libraries were prepared and sequenced to shallow depth (∼5-8 million reads/library) for initial quality assessment. Obtained sequencing reads aligned to their parental reference genome with little cross-mapping observed (Figures S2A and S2B). In fact, the vast majority of cross-mapped reads derived from mitochondrial genes present in both *Aspergillus* and human cells, and all of these reads were removed from further analyses. As expected, the fungal-to-human read ratio in *Aspergillus*-infected samples increased with MOI. In line with the high proportions of fungal reads (∼20% of total mapped reads), human exon coverage was the rate-limiting factor (Figure S2A), thus favoring the low-dose infections (MOI 0.5) for further experiments. While the vast majority of reads in CMV-infected moDCs mapped to the human genome, viral read proportion and exon coverage increased time-dependently, indicative of viral replication (Figure S2B).

Even with the low sequencing depth used in this pilot experiment, the induction of marker genes for human dendritic cell activation and pathogenicity-related fungal genes was detected (Figure S2D). For example, upregulation of *IL-1, CCL3,* and *TLR2* indicated activation of well-described pro-inflammatory cascades in moDCs infected with *Aspergillus* (Braedel et al., 2004; Lass-Flörl et al., 2013; Walsh et al., 2005). Fungal cells also showed elevated expression of genes for toxic molecules such as the gliotoxin GliF (Latgé, 1999). Similarly, expression profiles during virus infection reflected expected patterns (Figure S2E) as CMV-infected moDCs upregu-lated *IFN-γ* and *CCL2* (Loewendorf & Benedict, 2010; McNab et al., 2015), whereas viral gene expression was generally induced over

time (Figure S2E displays viral expression at two later stages relative to an early infection stage) indicating active proliferation (Dunn et al., 2003; Moutaftsi et al., 2002; Wiertz et al., 1996).

RNA extraction, rRNA depletion, cDNA library preparation, and triple RNA-seq

For extraction of human, fungal and viral RNA, RiboPure RNA Purification Kit yeast (Thermo Fisher Scientific) was used according to the manufacturer's instructions. Total RNA was treated with 0.13 U/µl DNase I (Thermo Fisher Scientific) for 30 min at 37°C to remove contaminating genomic DNA. The integrity of DNase-treated RNA was assessed on a bioanalyzer (Agilent). All samples had RNA integrity numbers (RIN) $\geq$ 7.0. Where explicitly indicated (as polyA+), total RNA was directly converted into first-strand cDNA using an oligo(dT)$_{25}$ primer, fragmented (four 30 s ultrasound pulses), and processed as outlined below. Otherwise, ribosomal transcripts were actively removed using Ribo-Zero Gold rRNA removal kits (human, mouse, rat) (Illumina) following manufacturer's instructions for 500 ng DNase-treated RNA as input for rRNA depletion.

The cDNA libraries for Illumina sequencing were generated by Vertis Biotechnologie AG, Freising-Weihenstephan, Germany, with rRNA-free RNA sheared via ultrasound sonication (four 30 s pulses, 4°C) to generate 200- to 400-nucleotide fragments, on average. Fragments < 20 nucleotides were removed using Agencourt RNAClean XP kits (Beckman Coulter Genomics), and Illumina TruSeq adapters were ligated to the 3′ ends of remaining fragments. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase (NEB) with 3′ adapters used as primer target sites. First-strand cDNA was purified, and 5′ Illumina TruSeq sequencing adapters were ligated to 3′ ends of antisense cDNA. Resulting cDNA was PCR-amplified to about 10 to 20 ng/µl using high-fidelity DNA polymerase. TruSeq barcode sequences were included in 5′ and 3′ TruSeq sequencing adapters. The cDNA libraries were purified using Agencourt AMPure XP kits (Beckman Coulter Genomics) and analyzed by capillary electrophoresis (Shimadzu MultiNA microchip).

For sequencing, cDNA libraries were pooled in approximately equimolar amounts. Pools were size-fractionated to 200-600 bp using differential cleanup with Agencourt AMPure kits (Beckman Coulter Genomics). Aliquots of cDNA pools were analyzed by capillary electrophoresis (Shimadzu MultiNA microchip). Sequencing was performed on a NextSeq 500 platform (Illumina) at Vertis Biotechnologie AG, Freising-Weihenstephan, Germany (single-end mode, 75 cycles).

Quantitative reverse transcription PCR-based validation of differential gene expression

RNA of untreated moDCs or moDCs after infection with CMV, *A. fumigatus* or both was reverse transcribed into cDNA using Thermo Fisher cDNA First Strand Synthesis Kits. Primers (Sigma-Aldrich) were designed using Primer-Blast (Ye et al., 2012) and The Aspergillus Genome Database (Cerqueira et al., 2014), avoiding the occurrence of target sequences in the other two organisms (Table S1). Quantitative reverse transcription (qRT)-PCR was conducted using SYBRGreen Master Mix from (BioRad) in a Step One System (Applied Biosystems). Primer specificity was confirmed by agarose (Roth) gel electrophoresis (Serva) of PCR amplicons using ethidium bromide (Thermo Fisher). Gel images were documented in a Multi-Image Light Cabinet (Alpha Innotech).

Multiplex cytokine secretion assays

Cell culture supernatants were analyzed by multiplex cytokine secretion assays according to the manufacturer's instructions using 16-plex ProcartaPlex Immunoassays (Thermo Fisher Scientific) including IFN-α, IFN-β, IFN-γ, IL-1α, IL-1β, IL-2, IL-6, IL-8, IL-10, IL-12p70, IL-17A, IL-23, CXCL10, CXCL11, CCL5, and TNF-α.

Cytokines with concentration levels above measurement range were set to 1.05 times the maximum. Concentration levels below measurement range or negative concentrations were set to 0. Significant changes in the concentration of cytokines were determined using pairwise two-sided Wilcoxon rank-sum tests. P values for each test were adjusted for multiple testing using FDR. We rejected the null hypothesis for FDR < 0.1.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq data processing

Preprocessing of raw reads including quality control and gene abundance estimation was done with GEO2RNaseq pipeline version 0.9.12 in *R* version 3.5.1 (Seelbinder et al., 2019). Quality analysis was done with FastQC version 0.11.8 before and after trimming. Read-quality trimming was done with Trimmomatic version 0.36. Adaptor sequences were removed, window size trimming performed (15 nucleotides, average Q < 25) including 5′ and 3′ per-base trimming for Q < 3 and removal of sequences shorter than 30 nucleotides. The reference genome in FASTA format was created by combining references *Homo sapiens* (GRCH 38 v89), *A. fumigatus* Af293 (s03-m05-r09) and human herpesvirus 5 strain TB30/E clone TB40-BAC4 (EF999921.1). Reference annotation was created by extracting and combining exon features from corresponding annotation files. The reference genome was indexed with exon information using HiSat2 version 2.1.0. Paired-end read alignment used HiSat2 on the created reference genome. Only concordantly aligned pairs of reads were used. Mapping statistics per organism were calculated using the "calc_triple_mapping_stats" function of GEO2RNaseq. SAMtools version 1.7 with the "flagstat" subcommand was used to deduce alignment quality. Gene abundance estimation was done with featureCounts (*R* package Rsubread version 1.28.0) in paired-end mode with default parameters. MultiQC version 1.5 was used to summarize the output of FastQC, Trimmomatic, HiSat, featureCounts and SAMtools (Data S1). In addition to the count matrix with gene abundance for all three species, species-specific gene count matrices were extracted

Cell Reports
Article

from the complete count matrix. For nonstatistical analyses, count matrices were normalized using median-by-ratio normalization (MRN) as described before (Anders and Huber, 2010). Correlation analysis, principal component analysis, and clustering were performed per species based on MRN gene-abundance data. Clustering between samples used the complete linkage (farthest neighbor) method. Raw files are accessible under the Gene Expression Omnibus accession number GSE134344, GSE135450 and GSE136217.

Differential gene expression analysis

Differential gene expression was analyzed by GEO2RNaseq per species. Pairwise tests were performed between control, single-infection, and co-infection groups with and without consideration of infection time. Four statistical tools (DESeq 1.30.0, DESeq2 1.18.1, limma voom 3.34.6 and edgeR 3.20.7) were used, and p values were corrected for multiple testing using the false-discovery rate method q = FDR(p) for each tool. In addition, mean MRN, transcripts per kilobase million (TPKM) and reads per kilobase million (RPKM) values were computed per test per group including corresponding $\log_2$ fold-changes. Gene expression differences were considered significant if they were reported significant by all four tools (q < 0.01 and $|\log_2$ MRN$| \geq 1$ for *H. sapiens*; q < 0.05 for *A. fumigatus* and CMV).

To compare RNA-seq and qRT-PCR derived relative differences, expression values were scaled by the maximum value per replicate. To keep statistical analysis comparable, significance was assessed based on pairwise t tests. Resulting p values were corrected for multiple testing using FDR. Tests were performed individually for RNA-seq and qRT-PCR per gene (Figure 5; Figure S5).

Interspecies and intraspecies gene expression analyses

Abundances of genes with nonzero coverage from all three species and all samples were MRN normalized. Spearman's correlations between gene abundances were calculated per treatment group (moDC, moDC + *A. fumigatus* [Afu], moDC + CMV, moDC + CMV + Afu) for infection time 0 h. Only significant correlations with p < 0.01 and absolute correlation $\geq 30\%$ were used for further analysis. *R* package iGraph version 1.2.4.1 was used to create, compare, analyze (node degree and betweenness) and plot significant correlations as networks. Node degree centrality describes the number of edges incident to a node. Node betweenness centrality describes the number of shortest paths through a node for all pairs of nodes. For cross-species analysis, correlations between genes of the same species were ignored. For additional immune system relevant gene correlation analysis, genes from *H. sapiens* were retained only if they were present in the curated database InnateDB (https://www.innatedb.com/). moDC gene set analysis with distinct expression patterns between, but low variance within different infection etiologies yielded 160 genes that were further filtered for immune system-relevant genes (Figure 4).
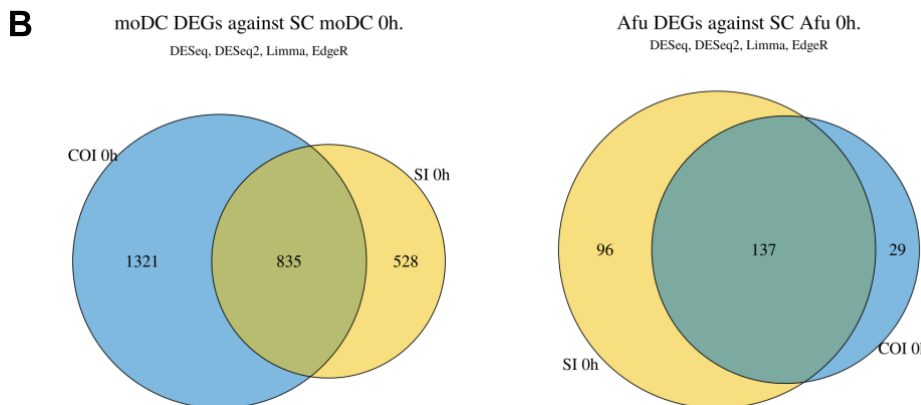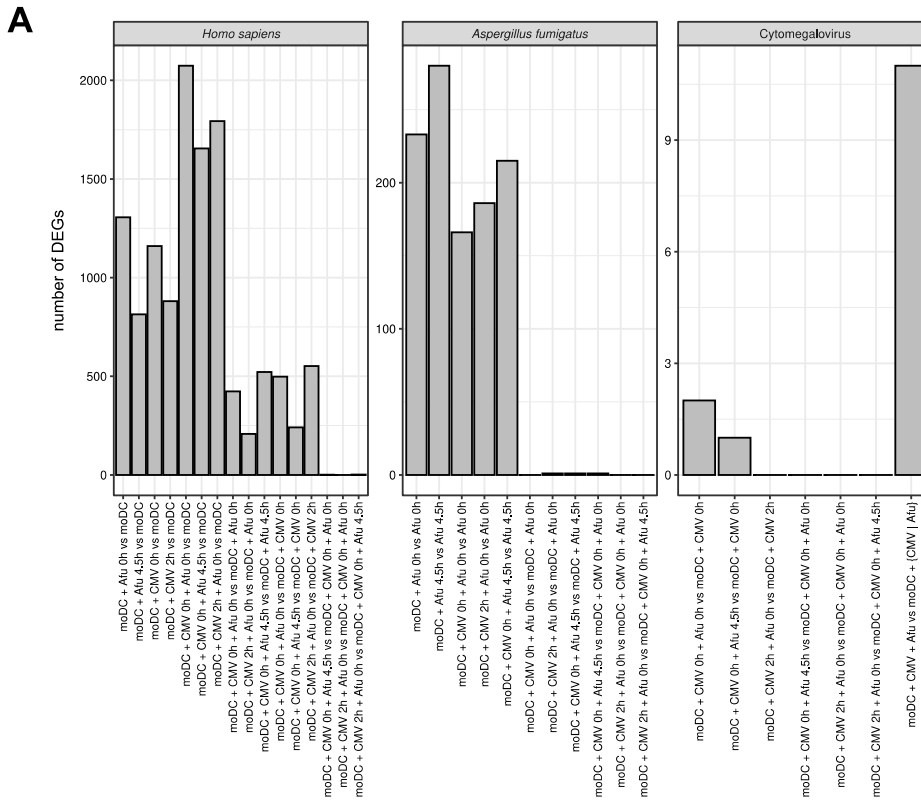
Supplement

Supplementary Figure S1
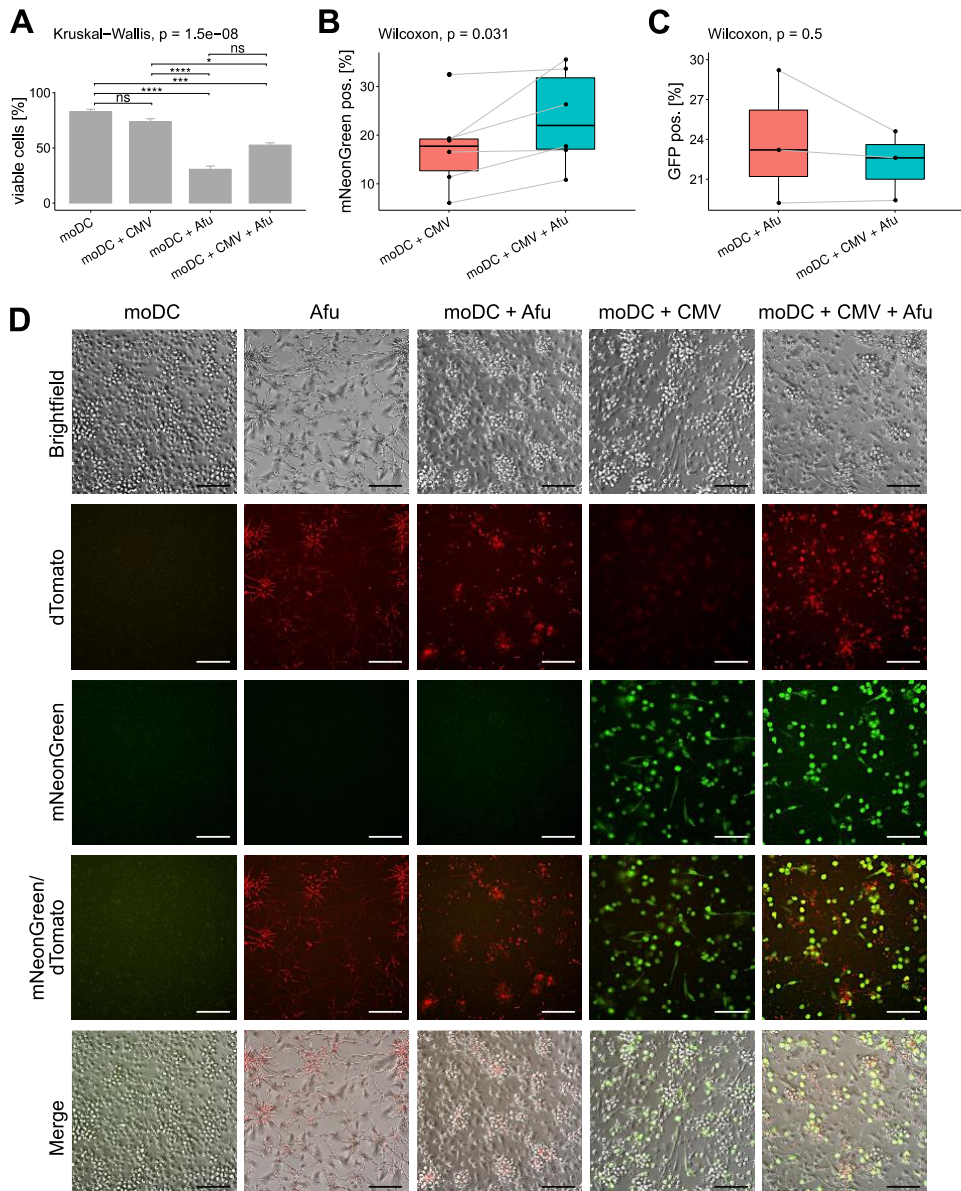
Supplementary Figure S1: Infection of monocyte-derived dendritic cells (moDCs) challenged with Aspergillus fumigatus (Afu) and cytomegalovirus (CMV) in single- or co- infection settings. (A) moDCs were stained with a live/dead marker and cells were analyzed by flow cytometry. Viability of moDCs is depicted as mean + SEM. n = 4. Asterisks indicate significance values after multiple test corrected (FDR) pairwise Dunns test. ****: p < 0.0001, ***: p < 0.001; *: p < 0.05; ns: not significant.. (B) Quantification of CMV-infected moDCs during single- or fungal co-infection. Virus-infected cells were distinguished from uninfected bystander cells by flow cytometric measurement (CMV constitutively expresses the fluorescence reporter mNeonGreen). Boxplots show median and interquartile range. n = 6. (C) moDC infection rate with A. fumigatus in presence or absence of CMV as determined using a non- fluorescent CMV strain and GFP-expressing A. fumigatus germ tubes. Phagocytosis of A. fumigatus was measured 3 h after addition of germ tubes by flow cytometry (quantification of cells positive for GFP). Boxplots show median and interquartile range. n = 3. (D) Infections of moDCs with CMV (mNeonGreen) and A. fumigatus (expressing dTomato) were analyzed by fluorescence microscopy at 9 h after addition of A. fumigatus germ tubes. Images illustrate one representative result out of three independent experiments. Scale bars indicate 100 μm. Related to Figure 1.

Supplementary Figure S3

Supplementary Figure S3: Differentially expressed genes (DEGs). (A): Number of DEGs for each experimental condition. (B): Common and unique number of DEGs for co-infection (COI) and single-infection (SI) compared with uninfected monocyte-derived dendritic cells (moDCs) in single-culture (SC) (left) or Aspergillus fumigatus (Afu) in single-culture (right) for 9 h. CMV: cytomegalovirus. Related to Figure 2.

Supplementary Figure S4

Supplementary Figure S4: Gene expression fold-changes from left to right in each rectangle: Co-infection vs. no stimulation, co-infection vs. single-infection with cytomegalovirus (CMV) at time point 0 h, co-infection vs. single-infection with Aspergillus fumigatus at time point 0 h. Red color indicates an upregulation in the left hand side condition of comparisons. To visualize also small fold-changes, color-code for expression changes beyond -1.5 and 1.5 were capped to the maximum intensity. A: Toll-like receptor signaling pathway, B: Cytosolic DNA-sensing pathway, C: C-type lectin receptor signaling pathway. Related to Figure 4.

Supplementary Figure S5

Supplementary Figure S5: Relative gene expression and quantitative qRT-PCR-derived values for selected genes. qRT-PCR-derived values were normalized against delta-aminolevulinate synthase mRNA. n = 4. Mean and SEM are shown. Afu: Aspergillus fumigatus, DC: monocyte-derived dendritic cell, CMV: cytomegalovirus. Asterisks indicate significance level after multiple test correction (FDR): ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$. Related to Figure 5.

Supplementary Figure S6
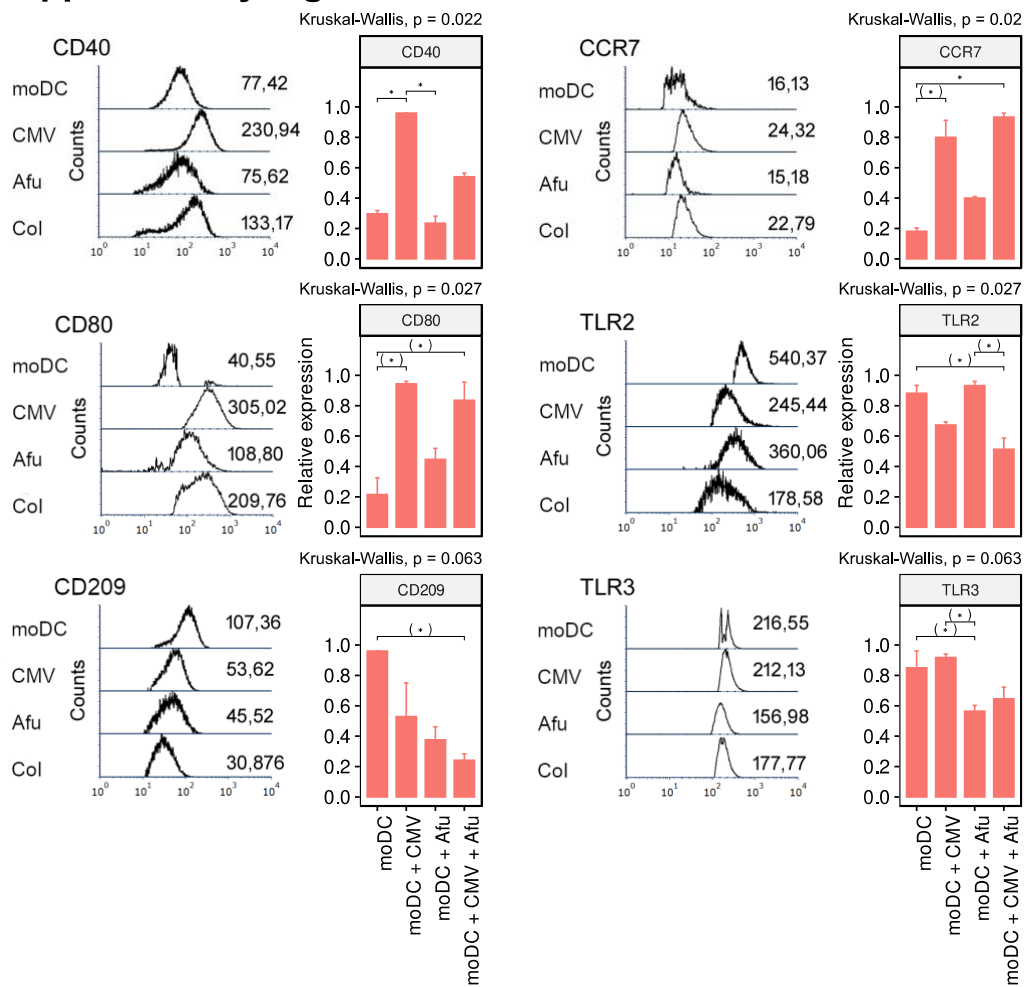
Supplementary Figure S6: Surface marker expression on monocyte-derived dendritic cells (moDCs). moDCs were infected with cytomegalovirus (CMV, mNeonGreen) for 24 h or left uninfected. Then, Aspergillus fumigatus (Afu) germ tubes were added for 9 h in a single- or co- infection setting. Cells were stained for CD40, CCR7, CD80, TLR2, CD209, and TLR3. Histograms show fluorescence signal intensities subtracted by isotype controls for one representative result out of three independent experiments. Bar plots refer to the mean and SEM of relative surface marker expression over the three replicates. Asterisks indicate significance after multiple test corrected (FDR) pairwise Dunns test: *: p < 0.05; (*): p < 0.01. Related to Figure 5.

Supplementary Figure S7
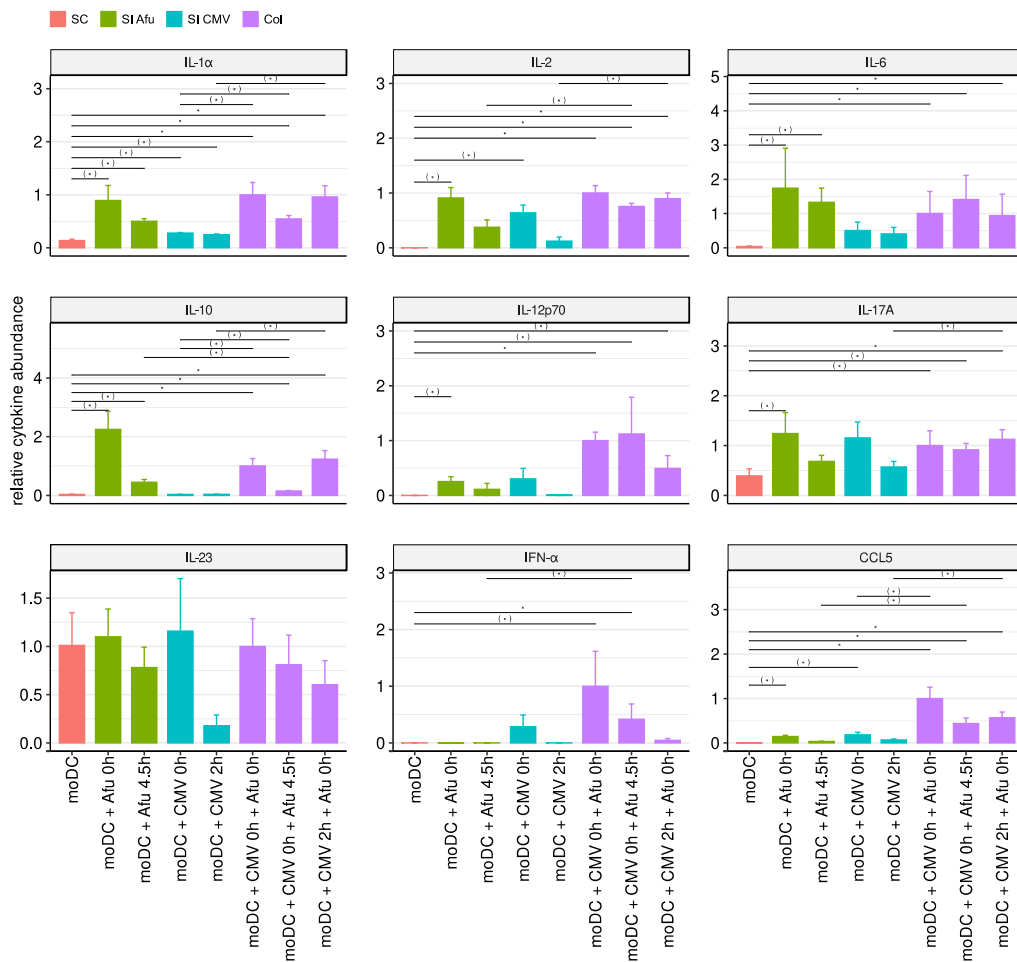
Supplementary Figure S7: Multiplex cytokine secretion assays. Relative concentrations of cytokines in culture supernatants were normalized to their mean in moDC + CMV 0h + Afu 0h. n = 4. Mean and SEM are given. Afu: Aspergillus fumigatus, DC: monocyte-derived dendritic cells, CMV: cytomegalovirus. Asterisks indicate significance level after multiple test correction (FDR): *: $p < 0.05$; (*): $p < 0.1$. Related to Figure 5.

II. Manuscript:
Aspergillus fumigatus Pan-Genomics

FORM 1

Manuscript No. 2

Manuscript title: Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection

Authors: Barber A., Sae-Ong T., Kang K., **Seelbinder B.**, Li Jun, Walther G., Panagiotou G., Kurzai O.

Bibliographic information: *Nature Microbiology* 612. 6 (2021), DOI: 10.1038/s41564-021-00993-x

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
Barber	30%	15%	80%	40%	
Sae-Ong	30%	80%		40%	
Seelbinder		5%		5%	
Panagiotou	20%			5%	50%
Kurzai	20%			5%	50%
Others		0%	20%	5%	
Total:	100%	100%	100%	100%	100%

Overview:

In a follow-up study, we employed next-generation sequencing and comparative genomics to investigate how genetic diversity contributes to virulence, and antifungal drug resistance in 300 *A. fumigatus* isolates from clinical and environmental sources. We found large, unappreciated diversity among *A. fumigatus* strains, including a group of clinical strains enriched for azole resistance. This study led to profound implications on the use of the standard Af293 strain and its' reference genome in infection studies (such as the one in manuscript I).

Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection

Amelia E. Barber[1,8,9], Tongta Sae-Ong[2,9], Kang Kang [2], Bastian Seelbinder [2], Jun Li[3,4], Grit Walther[5], Gianni Panagiotou[2,6 ✉] and Oliver Kurzai[1,5,7 ✉]

***Aspergillus fumigatus* is an environmental saprobe and opportunistic human fungal pathogen. Despite an estimated annual occurrence of more than 300,000 cases of invasive disease worldwide, a comprehensive survey of the genomic diversity present in *A. fumigatus*—including the relationship between clinical and environmental isolates and how this genetic diversity contributes to virulence and antifungal drug resistance—has been lacking. In this study we define the pan-genome of *A. fumigatus* using a collection of 300 globally sampled genomes (83 clinical and 217 environmental isolates). We found that 7,563 of the 10,907 unique orthogroups (69%) are core and present in all isolates and the remaining 3,344 show presence/absence of variation, representing 16–22% of the genome of each isolate. Using this large genomic dataset of environmental and clinical samples, we found an enrichment for clinical isolates in a genetic cluster whose genomes also contain more accessory genes, including genes coding for transmembrane transporters and proteins with iron-binding activity, and genes involved in both carbohydrate and amino-acid metabolism. Finally, we leverage the power of genome-wide association studies to identify genomic variation associated with clinical isolates and triazole resistance as well as characterize genetic variation in known virulence factors. This characterization of the genomic diversity of *A. fumigatus* allows us to move away from a single reference genome that does not necessarily represent the species as a whole and better understand its pathogenic versatility, ultimately leading to better management of these infections.**

Diseases caused by the mould *Aspergillus fumigatus* are a major cause of human morbidity and mortality[1,2]. Invasive aspergillosis is particularly problematic in immunocompromised patients, resulting in a mortality rate of up to 50%[3,4]. Treatment of infections caused by *A. fumigatus* relies on triazole antifungal drugs. However, resistance to these frontline therapies is increasing, and the mortality rate for resistant infections is 25% higher than susceptible infections[5,6]. Although the most frequently identified resistance mutations occur in the cellular target of the triazoles—that is, *cyp51a*—up to 30% of the resistant isolates have no identifiable resistance mechanisms[7], complicating the recognition and treatment of these problematic infections.

While the host immune status is an important determinant in the development of aspergillosis, the substantial phenotypic variability observed among *A. fumigatus* isolates indicates that intraspecies diversity also plays a role in the disease[8–14]. This includes marked differences in virulence in animal models[9,10,14], fitness under hypoxia[9], growth under chemical stress(es)[11], nutritional heterogeneity[12] and induction of host inflammatory mediators[8]. As an indicator of the genomic diversity underlying the phenotypic variability observed in *A. fumigatus*, genomic comparisons between the reference strains Af293 and A1163 reveal tracts of variable gene content

between the two[15], and 7% of Af293 genes are not present in A1163 (FungiDB). Despite this variation, previous studies of *A. fumigatus* have largely only analysed genomic information in the context of the reference genome and were limited to the genetic material present in Af293 due to the technical challenges of de novo eukaryotic genome analysis[16–19]. In addition, most of the isolates that have been sequenced to date are of clinical origin, thereby obscuring the genomic relationship between environmental isolates and those causing human disease.

In this study we constructed de novo genome assemblies of 300 *A. fumigatus* genomes ($n = 217$ environmental isolates and $n = 83$ clinical isolates) and used them to define the pan-genome of this important human fungal pathogen as well as the relationship between environmental and clinical isolates. We also leveraged the power of genome-wide association studies (GWAS) to identify genomic variation associated with human infection and triazole resistance, revealing a new range of therapeutic targets to combat these life-threatening infections.

Results

De novo assembly of 300 *A. fumigatus* genomes. In this study we used reference-guided and de novo assembly methods to analyse

[1]Research Group Fungal Septomics, Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany. [2]Research Group Systems Biology and Bioinformatics, Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany. [3]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China. [4]School of Data Science, City University of Hong Kong, Hong Kong, China. [5]National Reference Center for Invasive Fungal Infections (NRZMyk), Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany. [6]Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, Hong Kong, China. [7]Institute for Hygiene and Microbiology, University of Würzburg, Würzburg, Germany. [8]Present address: Junior Research Group Fungal Informatics, Leibniz Institute of Natural Product Research and Infection Biology–Hans Knöll Institute, Jena, Germany. [9]These authors contributed equally: Amelia E. Barber, Tongta Sae-Ong. ✉e-mail: gianni.panagiotou@leibniz-hki.de; okurzai@hygiene.uni-wuerzburg.de

the genomes of 300 *A. fumigatus* isolates, representing environmental and clinical isolates from different locations across the globe. Among these, 188 samples were novel environmental and clinical isolates from Germany that were sequenced as part of this study. The remaining 112 isolates, including 64 isolates that were sequenced by us in a previous study[20], were pulled from public data repositories as raw sequence data. Our overall dataset was comprised of 217 environmental isolates and 83 clinical isolates from Europe, Asia, North America, South America and the International Space Station (Supplementary Data 1). Forty-three of 294 isolates were resistant to one or more medical triazoles, as determined by European Committee on Antimicrobial Susceptibility Testing (EUCAST) broth microdilution[21]. Azole susceptibility data were not available for six isolates. We generated de novo genome assemblies of these 300 isolates using paired-end Illumina sequencing to facilitate the unrestricted analysis of genomic diversity in *A. fumigatus*. The mean number of contigs in our assemblies was 948 and the mean N50, a marker of genome contiguity representing the weighted median contig length, was 145,494 base pairs (bp; Supplementary Table 1 and Supplementary Data 1). The mean genome size of our assemblies was 28.6 Mb (range, 26.9–30.8 Mb), with an average of 9,408 open reading frames (ORFs) per isolate and a range of 9,169 to 11,231. Using BUSCO as a measure of genome completeness, we found that an average of 97% of the expected single-copy orthologues were found and present as single copies in our genome assemblies.

To perform population genomic analyses, we aligned reads against the Af293 reference genome. We observed an average of 78,692 single nucleotide variants (SNVs) per isolate (range, 23,029–149,537) or approximately three SNVs per kilobase (Supplementary Data 1). We also detected an average of 7,383 short insertions or deletions (indels) per isolate (range, 2,528–16,134). Of the 329,405 non-redundant SNVs identified among our isolates, 33% (107,779) were not described in FungiDB, release 39. Together, this reveals a pronounced level of genetic diversity in *A. fumigatus* at the nucleotide level and considerably extends the previously recognised diversity.

The *A. fumigatus* pan-genome contains 7,563 core and 3,344 accessory genes. To examine the full genomic diversity of *A. fumigatus*, we used our de novo genome assemblies to define and characterize its pan-genome. The pan-genome is the collective gene set of a species and is composed of core genes found in all individuals and accessory genes that are not shared between all members of the species. We identified a total of 12,798 gene clusters that condensed into 10,907 non-redundant orthogroups. The *A. fumigatus* pan-genome was composed of a core genome of 7,563 orthogroups in all 300 isolates (69% of the pan-genome), 935 softcore orthogroups in >95% of the isolates (9% of the pan-genome), 1,367 shell genes in 5–95% of the isolates (13% of the pan-genome) and a cloud genome of 1,043 genes present in less than 5% of the isolates (10% of the pan-genome; Fig. 1a). Each isolate contained an average of 9,199 orthogroups (range, 8,987–9,629) and an average of 1,636 orthologous accessory-gene clusters (range, 1,424–2,066), corresponding to 16–22% of the total genome of the isolate. The pan-genome was closed—that is, the number of pan-genes did not substantially increase after the addition of approximately 250 genomes (Fig. 1b). Gene association analysis identified 53 co-occurring gene modules containing 2–251 genes (Fig. 1c).

The protein sequences of the core genes were significantly longer than the softcore or accessory genomes. The geometric mean of the length of the core genes was 436 amino acids compared with 310 amino acids for the softcore genes and 191 for the shell/cloud genes (Fig. 1d). To examine the evolutionary forces working on the core and accessory genomes, we calculated the rate of non-synonymous-to-synonymous substitutions (d_N/d_S). The geometric mean of the d_N/d_S ratio among all 10,907 pan-genes was

0.53, with significant differences between the genome compartments. The core genome showed the strongest evidence of negative or purifying selection ($d_N/d_S = 0.49$), whereas the softcore and accessory genomes had d_N/d_S ratios of 0.68 and 0.69, respectively (Fig. 1e). The lower d_N/d_S values for the core genes relative to the accessory genes indicate that they are under a higher degree of purifying selection—although neither genome compartment is evolving neutrally, as indicated by ratios of less than one.

The core genome contained a higher proportion of proteins with annotated domains, as 85% of the core genes contained at least one annotated InterPro domain compared with 71% of the softcore genes and 51% of the accessory genes (Fig. 1f). The core genome was enriched for 3,140 Pfam domains—including protein kinase domains, transcription factor domains and ABC transporters—whereas the accessory genome was enriched for 546 Pfam domains—including short-chain dehydrogenases and cytochrome P450 enzymes (Extended Data Fig. 1a). For Gene Ontology (GO) annotations, the core genome was enriched for protein binding, ATP binding, carbohydrate metabolic functions, signal transduction and 1,497 total annotations (Supplementary Data 2). The accessory genome was enriched for haem binding, response to oxidative stress and 244 total GO annotations (Supplementary Data 2).

Many of the shell and cloud genes were located on the subtelomeric ends of chromosomes 1 and 7, as measured by their position in Af293 (Extended Data Fig. 1b). Of the 10,907 orthologous gene clusters (homologous genes identified in different isolates) identified in the *A. fumigatus* pan-genome, 87% were present in Af293 (Supplementary Data 3). Overall, we identified an average of 494 genes per isolate that were absent in Af293 and a cumulative 1,934 unique ORFs were not present in Af293. In summary, the core genome of *A. fumigatus* represented 69% of the total identified orthogroups and was distinct from the accessory genome in length, function and the strength of purifying selection.

Chronic disease isolates are more genetically diverse than isolates from invasive disease and the environment. We examined the population genomics of isolates from the environment, invasive disease and chronic aspergillosis. Due to the lower number of isolates in the chronic disease group, the environmental and clinical samples were downsampled to match the number of chronic disease isolates ($n = 19$). Interestingly, the isolates from chronic disease group were significantly more diverse at the nucleotide level than isolates from invasive disease or the environment, as measured by the nucleotide diversity (π) calculated across overlapping 5 kb windows (Extended Data Fig. 2). In contrast, isolates from the invasive disease group showed less nucleotide diversity than isolates from the environment and chronic disease. The geometric mean of the genome-wide nucleotide diversity was 1.3×10^{-5} for the isolates from the chronic disease group, 8.3×10^{-6} for the environmental isolates and 6.9×10^{-6} for the isolates from the invasive disease group.

The Af293-containing genetic cluster is enriched for clinical isolates. In a phylogeny built from the coding nucleotide sequences of 5,380 single-copy orthologues, all 300 isolates formed a monophyletic group that was clearly distinct from the related outgroups of *Aspergillus oerlinghausenensis* and *Aspergillus fischeri* (Fig. 2 and Extended Data Fig. 3a). Isolates from Germany, collected and sequenced by us, intermixed with the globally sampled isolates from publicly available repositories, with no strong geographic clustering observed. We also found a high degree of congruence between the phylogeny built from the core genome sequence from de novo genome assemblies and phylogenies built using reference-guided SNV data from whole-genome SNVs and neutral loci (Extended Data Fig. 3b–d). Based on genome coverage at the MAT locus, we found an equal split of isolates of both mating types ($n = 148$ MAT1-1 isolates and $n = 149$ MAT1-2 isolates; Fig. 2).
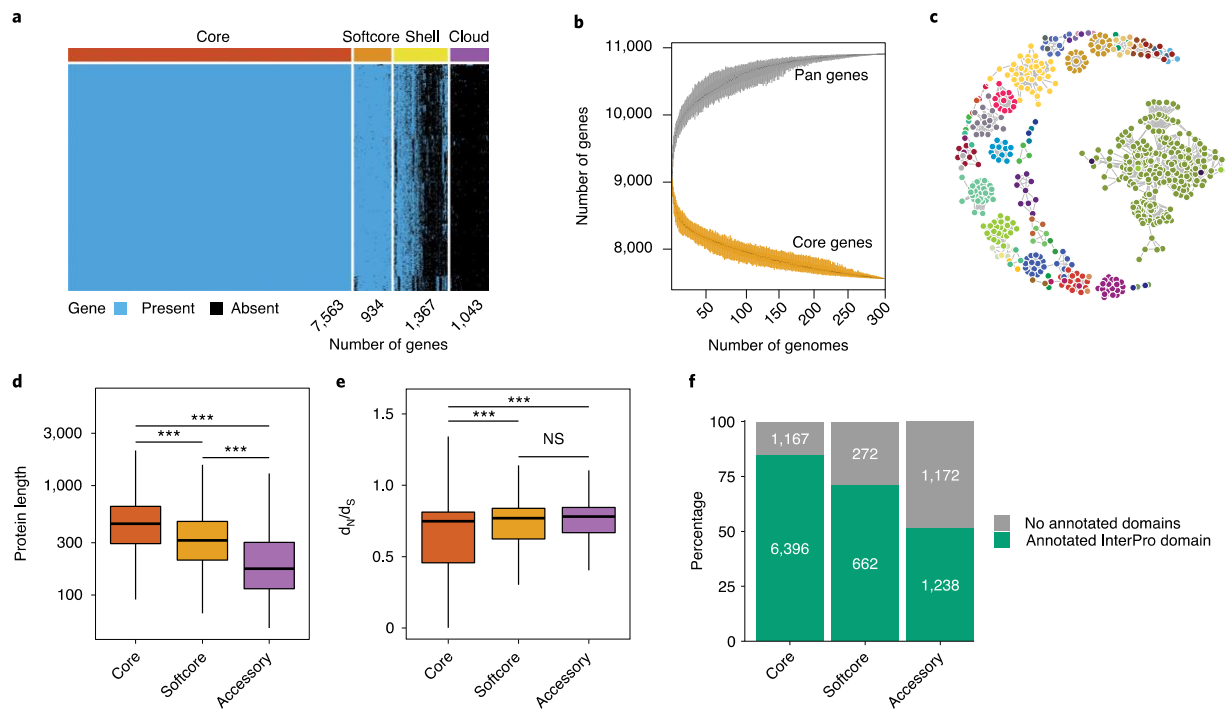
Fig. 1 | The pan-genome of *A. fumigatus*. a, Presence/absence matrix of 10,907 orthologous gene clusters identified from 300 *A. fumigatus* genomes. The pan-genome is subdivided into core (orthogroups present in all isolates), softcore (orthogroups present in >95% of the isolates), shell (orthogroups present in 5–95% of the isolates) and cloud (orthogroups present in less than 5% of the isolates) genomes. **b**, Pan and accessory (softcore, cloud and shell) genome size as the number of genomes included increases. Darker hues represent the 25th and 75 percentiles, while the lighter hues represent the range. **c**, Co-occurring gene modules ($n = 53$) of the *A. fumigatus* accessory genome. Each circle indicates a gene and its association with other genes indicated by edges (lines). The module significance was identified using two-sided binomial exact tests with Bonferroni's correction ($P < 0.05$). Only positive associations are illustrated. The colour indicates module membership. **d**, Amino-acid-sequence lengths of core, softcore and accessory (cloud and shell) genes. The exact P values were $<2 \times 10^{-16}$ for all of the indicated comparisons. **e**, Ratio of the non-synonymous substitutions to synonymous substitutions in core, softcore and accessory genes. Genes with ratios greater than one are under positive selection, whereas genes with ratios less than one are under purifying selection. The exact P values were: core versus softcore, $P = 1.2 \times 10^{-7}$; core versus accessory, $P < 2 \times 10^{-16}$; and softcore versus accessory, $P = 0.15$. **c–e**, $n = 7,563$ core, 935 softcore and 2,410 accessory orthogroups. **d,e**, In the box-and-whisker plots, the horizontal line in the box indicates the 50th percentile and the box extends from the 25th to the 75th percentile. The whiskers encompass the lowest and highest values within 1.5× the interquartile range. Statistical significance was determined using a two-sided Mann–Whitney U-test with Bonferroni's correction; ***$P < 0.001$ and NS, not significant. **f**, Number (indicated in the bars) and fraction of core, softcore and accessory genes containing an annotated InterPro domain.

To look for evidence of genomic recombination in *A. fumigatus*, we performed a neighbour-net analysis, a phylogenetic method that allows for the representation of conflicting genetic signals that result from sexual recombination or gene conversion. The neighbour-net tree built from core genes had a highly reticulated centre, which indicates a marked degree of conflicting genetic information in the phylogenetic network and is suggestive of abundant genetic recombination in the species (Fig. 3a).

Discriminant analysis of principle components[22] was used to identify seven as the best supported number of genetic clusters in our dataset based on our de novo, reference and pan-gene count-based approaches (Extended Data Fig. 4a–c). Cluster 6 had the largest number of isolates ($n = 80$), followed by cluster 2 ($n = 53$), cluster 5 ($n = 48$), cluster 7 ($n = 43$), cluster 3 ($n = 35$), cluster 4 ($n = 22$) and cluster 1 ($n = 19$; Fig. 4a). Interestingly, cluster 5 was enriched for clinical isolates (Fisher's exact test with Benjamini–Hochberg correction, $P = 0.02$). This cluster also contained the reference strain Af293, which is a clinical isolate from a patient who died of invasive aspergillosis[23]. Together, we observed an enrichment for clinical

isolates in one cluster as well as evidence of abundant genetic recombination in *A. fumigatus*.

Genetic cluster 5 contains more accessory genes and a distinct genomic profile. As genetic cluster 5 was statistically enriched for clinical isolates, we examined the genomes of each cluster to identify differences that might predispose the genetic background of cluster 5 towards human infection as well as characterize potential functional differences between the genetic clusters. Interestingly, clusters 5 and 2 contained significantly more accessory genes than the other clusters (Fig. 4a). The median number of accessory genes for cluster 5 was 1,965 compared with 1,895, 1,842, 1,882, 1,814 and 1,790 for clusters 2, 3, 1, 7 and 6, respectively (Fig. 4a). Cluster 4 had the smallest number of accessory genes, with a median of 1,749.

To predict the functional differences between the clusters, we calculated the abundance of Pfam domains and the frequency of GO annotations in the different clusters and compared the variance between clusters. A total of 170 GO annotations showed significant variation in their relative frequency between clusters (Fig. 4b and

ARTICLES
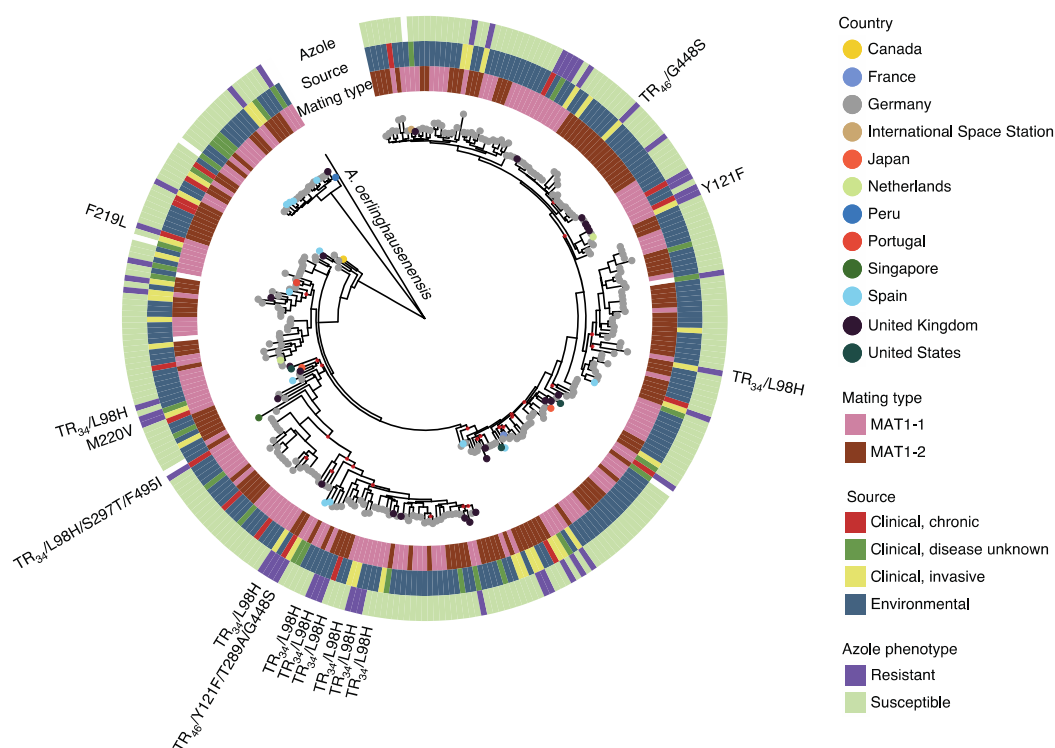
Fig. 2 | Whole-genome phylogeny of environmental and clinical *A. fumigatus*. Phylogenetic tree constructed from coding nucleotide sequences of 5,380 single-copy orthologues shared by *A. fumigatus*, *A. fischeri* and *A. oerlinghausenensis*. The phylogeny is rooted with *A. oerlinghausenensis* and the branch length was shortened for illustration. The coloured symbols at the end of branches represent the country where the sample was isolated. The red dots in the tree structure indicate nodes with ultrafast bootstrap values of less than 0.96. The metadata rings on the outside of the tree indicate the Azole phenotype (where resistance is defined as a minimum inhibitory concentration above the EUCAST breakpoint for one or more triazoles), source of the isolate and the mating type. Mutations in the *cyp51a* gene relative to the Af293 reference genome are also indicated on the outside of the tree.

Supplementary Data 4). Among these were an increased frequency of genes involved in oxidation–reduction processes, iron-ion binding, carbohydrate metabolic processes and proteolysis in cluster 5 (Fig. 4c and Supplementary Data 4). Significant variation between clusters was observed for the abundance of 269 Pfam domains (Supplementary Data 4). Cluster 5 had an increased abundance of major facilitator superfamily transporters, amino-acid permeases and chitin-recognition proteins (Fig. 4d and Supplementary Data 4).

For the GO categories and Pfam domains that did not show a significant difference in copy number between the genetic clusters, we reasoned that there could still be functional differences due to the presence of high-impact variants such as frameshifts or the gain/loss of stop codons. To examine this, we calculated the fraction of genes containing a high-impact variant(s) for each functional annotation and compared the incidence across the clusters. A total of 945 GO annotations contained significant differences in the incidence of high-impact variants between the clusters (Supplementary Data 4). Among these were a reduced number of high-impact variants in chromatin organization and mismatch repair-annotated genes in clusters 5 and 2 (Fig. 4e). We also quantified the incidence of high-impact variants in Pfam domain-containing genes and identified 482 domains with significant differences between the clusters (Supplementary Data 4). These included a reduced number of high-impact variants in cytochrome P450 enzymes and bZIP transcription factors in clusters 2 and 5. In summary, we observed distinct genomic profiles between the genetic clusters of *A. fumigatus*,

including a larger number of accessory genes in clusters 2 and 5 in addition to copy-number variation and incidence of high-impact variants in functional annotations such as Pfam domains and GO categories.

***A. fumigatus* exhibits variation in virulence-associated genes.** Using a database of 360 virulence- or fitness-associated genes for *A. fumigatus*[19,24–28] (Supplementary Data 5), we examined our 300 genomes for the presence/absence of these genes, changes in copy number relative to Af293 and the incidence of high-impact variants (for example, frameshifts and nonsense mutations). This list includes genes involved in metabolism, signalling, cell-wall biology, secondary metabolism, stress responses and antifungal drug resistance. Overall, these virulence-associated factors were well conserved. No variation in copy number, presence/absence of genes or genetic alterations anticipated to have a high functional impact was detected in 57% (205/360) of the genes. The remaining 155 virulence-associated genes had some degree of genetic variation expected to affect gene function among our 300 genomes, which can be visualized in Fig. 5 (full summary in Supplementary Data 5). Underscoring the fundamental role of these genes for the fitness of *A. fumigatus* in the environment and the human host, most cases of gene loss or high-impact genetic variation were uncommon and observed in less than 5% of the isolates (*n* = 121 genes). However, the remaining 34 genes displayed more pervasive genetic variation, including 76% of the isolates (229/300) showing frameshifts in the

Fig. 3 | Neighbour-net trees from 70 _A. fumigatus_ isolates. a, Phylogenetic network of the coding nucleotide sequence of 5,380 single-copy orthologous genes (whole-genome sequencing). The reticulated core indicates abundant conflicting genetic information among isolates, suggestive of recombination in the species. **b,c,** Phylogenetic networks of the coding nucleotide sequences of _cyp51a_ (**b**) and _cyp51b_ (**c**) plus 1,000 bp up- and downstream of the genes. Splits (parallel bands) indicate conflicting nucleotide patterns and their length is proportional to the number of bases supporting the split. The _cyp51a_ genotypes of the isolates are indicated at the branch tips. Red-hued labels indicate TR$_{34}$-containing alleles. Blue-hued labels denote TR$_{46}$ lineage-associated alleles. All additional _cyp51_ polymorphisms are indicated with black labels. Dashed circles in **c** denote the point on the phylogenetic network where all the indicated alleles localize. **a–c,** _n_ = 10 isolates from each of the genetic clusters identified in the dataset. Scale bar indicates nucleotide substitutions per site. **d,** Ratio of non-synonymous-to-synonymous substitutions in _cyp51a_ and _cyp51b_. The ratios were calculated from the genomes of 300 _A. fumigatus_ isolates. Both genes have ratios of less than one and are under purifying selection. Statistical significance was determined using a two-sided Mann–Whitney _U_-test; ***_P_ < 0.001; the exact _P_ value was _P_ < 2.2 × 10^{-16}. The bold horizontal line indicates the 50th percentile and the filled box extends from the 25th to the 75th percentile. The whiskers encompass the lowest and highest values within 1.5× the interquartile range.

serine protease _pr1_ (_Afu7g04930_) and 71% of the isolates (213/300) showing high-impact variants in the putative sensor histidine kinase _tcsB_ (_Afu2g00660_).

Overall, secondary metabolism genes showed the highest variability among the virulence-associated genes, with 59 genes either being absent or showing a predicted loss of function among the 300 isolates. Interestingly, 97% of the isolates (292/300) possessed extensively degraded copies of the non-ribosomal peptide synthetase _nrps8_ (also known as _pes3_ or _Afu5g12730_), a gene whose deletion showed increased virulence in a murine model of invasive aspergillosis[29]. We also observed variability in the biosynthetic gene cluster encoding fumagillin, including absence of the fumagillin tailoring enzyme _fmaG_ in 89% of the isolates (267/300), the absence of _fumR_ in 49% of the isolates (146/300) and a complete loss of the cluster in three isolates (1%). Finally, we observed variants predicted to impact the biosynthesis of the immunosuppressive virulence factor gliotoxin in 6% of the isolates (17/300). These included high-impact

variants in _gliZ_ (_n_ = 11 isolates); _gliA_ (_n_ = 5 isolates); _gliP_ and _gliF_ (_n_ = 3 isolates each); _gliI_, _gliT_ and _gliJ_ (_n_ = 2 isolates each); and _glicC_ and _gliG_ (_n_ = 1 isolate each).

In addition to cases of gene loss, we observed cases of gene amplification in virulence-associated genes relative to Af293 and A1163. A total of 53 genes showed gene amplification, including 5% of the isolates (_n_ = 16) with increased copy number of the putative catalase-peroxidase _cat2_ (_Afu8g01670_), which is upregulated in response to neutrophils[30]. In addition, 5% of the isolates (_n_ = 14) had increased copy numbers of the zinc transporter _zrfC_ (_Afu4g09560_) and 3% (_n_ = 10) had increased copy numbers of the putative ABC multidrug transporter _Afu5g12720_. In summary, although roughly half of the virulence-associated genes described to date were well conserved among the 300 genomes examined, we observed high-impact genetic variation in many virulence-associated genes, which could perhaps explain the wide range in virulence observed among _A. fumigatus_ isolates.

GWAS-identified fungal genetic variation associated with clinical isolates. To better understand how the environmental saprobe *A. fumigatus* can cause disease in the non-native niche of the human lung, we performed a GWAS study to identify fungal variants associated with clinical isolates relative to environmental isolates as well as fungal variants associated with the specific disease states of invasive and chronic disease (Extended Data Fig. 5a). Using a linear mixed model and a minor allele frequency (MAF) > 0.05, we identified 68 genomic positions with genetic variants associated with clinical isolates relative to environmental isolates (Supplementary Data 6). These variants included hits in 27 protein-coding genes, comprising both genes with established roles in virulence as well as uncharacterized ORFs (Supplementary Table 2). Among the genes previously implicated in the virulence of *A. fumigatus* were the sterol regulatory element binding protein *srbA*, which is involved in both growth in hypoxia and iron homeostasis[31,32], the global transcriptional regulator *pacC* required for fungal invasion during pulmonary infection[33,34] and the transcription factor *acuK* that regulates gluconeogenesis and iron acquisition[35]. The analysis also identified variants in genes whose role in virulence is less established, including a microtubule spindle protein (*Afu2g16260*), a heat shock-responsive protein (*Afu4g04680*), a putative polyketide synthase (*Afu6g13930*) and histone H1 (*Afu3g06070*), which is upregulated in conidia exposed to neutrophils (AspDB).

The virulence potential of *A. fumigatus* is influenced by the host and its underlying disease status. The factors critical for the establishment of invasive infection in a neutropenic lung are probably not the same as those required for long-term survival in the human lung, as in the case of chronic diseases such as cystic fibrosis and allergic bronchopulmonary aspergillosis. We thus performed association analysis for genetic variants associated with isolates from both invasive (acute) and chronic aspergillosis. There was a high degree of overlap between the genetic variants identified in this analysis and those from the analysis of all clinical isolates, regardless of the disease status of the host, but fewer variants and genes were identified for each underlying clinical disease (Extended Data Fig. 5b). We identified 21 genomic positions with SNVs and short indels significantly associated with invasive aspergillosis (Supplementary Table 2 and Supplementary Data 6). Nine of the ten variants located in coding genes that were associated with invasive disease were shared with isolates from all clinical origins and included the transcription factors *acuK* and *pacC* as well as the tubulin beta-2 subunit *tub2*. Chronic disease had variants at five genomic positions, two of which were within coding genes: *Afu2g03540*, an orthologue of GPI-anchored cell protein *cspA* (*Afu3g08990*) and a L-cytosine transmembrane transporter (*Afu6g14530*; Supplementary Table 2 and Supplementary Data 6).

Triazole target genes display distinct phylogenetic networks and imbalanced levels of stabilizing selection. The paralogous genes *cyp51a* (*Afu4g06890*) and *cyp51b* (*Afu7g03740*) encode the molecular targets of the triazoles. Despite this, most resistance mutations

and mechanisms have been described in *cyp51a*. Triazole-resistant isolates were distributed throughout the phylogeny (Fig. 2). However, most isolates carrying the TR$_{34}$/L98H allele of *cyp51a* were clustered near each other. The close genetic relationship between isolates carrying TR$_{34}$/L98H is in agreement with previous work suggesting a single origin of this allele[36].

To investigate the evolutionary features of the triazole targets, we built neighbour-net trees from the coding sequence of *cyp51a* and *cyp51b* plus 1,000 bp of the up- and downstream flanking sequences. A phylogenetic network built from *cyp51a* sequences showed multiple splits (parallel bands), indicating conflicting genetic information among our isolates that could arise from recombination (Fig. 3b). Genetic recombination by isolates carrying the TR$_{34}$/L98H allele is supported by its presence in isolates of both mating types (Fig. 2). By comparison, a neighbour-net tree of *cyp51b*, which is located on a different chromosome, did not show any conflicting genetic information, as demonstrated by the lack of reticulation in the phylogenetic network (Fig. 3c). We also observed a reassortment of *cyp51a* genotypes in the tree constructed from *cyp51b* sequences relative to that constructed from *cyp51a* (Fig. 3b,c). In the tree constructed from *cyp51a* sequences, isolates carrying the TR$_{34}$/L98H allele of *cyp51a* were located at five distinct points on the phylogenetic network at positions that did not overlap with the positions of other *cyp51a* mutant alleles. In the network of *cyp51b* sequences, strains carrying the TR$_{34}$/L98H allele were found only at two positions in the network that also contained other *cyp51a* mutant alleles.

To assess the selective forces working on *cyp51a* and *cyp51b*, we examined the d_N/d_S ratios of each gene. The d_N/d_S ratios of *cyp51b* were significantly lower than *cyp51a* (mean value of 0.01 and 0.27 for *cyp51b* and *cyp51*, respectively), indicating that *cyp51b* is under a stronger degree of stabilization selection than *cyp51a* (Fig. 3d). Together, our results demonstrate higher levels of genetic disagreement in the isolate sequences of *cyp51a* compared with *cyp51b* and that *cyp51a* is under less stabilizing selection than *cyp51b*.

GWAS-identified genetic changes associated with triazole resistance. We subsequently performed variant-based GWAS to identify genomic changes associated with triazole resistance. Among the 294 samples with available susceptibility data, 44 were resistant to one or more triazole. Of these, 15 contained mutations in *cyp51a* that have been previously shown to confer triazole resistance (for example, TR$_{34}$/L98H and TR$_{46}$/Y121F/T289A/G448S) and 29 were resistant by unknown mechanisms. When we performed a linear mixed-model GWAS using a MAF > 0.01, we identified 16 genomic positions associated with triazole resistance (Supplementary Table 3 and Supplementary Data 6). These included the known TR$_{34}$ and L98H variants in the triazole target enzyme *cyp51a*. However, we repeated our analysis using a MAF > 0.05 to give a more robust variant list with fewer false positives given that association studies with smaller datasets such as ours are underpowered for the detection of true associations with rare variants (Extended Data Fig. 5c). Using this more stringent criterion, we condensed our variant list

Fig. 4 | Pan-genomic differences between the clusters of *A. fumigatus*. a, Number of accessory genes (right) present in the genomes of isolates belonging to each genetic cluster (left). Statistical significance was determined using a one-way analysis of variance and Tukey's honest significance test (one-sided). The letters denote significances as a compact letter display where groups that are not significantly different from each other are indicated with the same alphabet letter; *P* < 0.05. The bold line in the box-and-whisker plot indicates the 50th percentile, and the box extends from the 25th to the 75th percentiles. The whiskers denote the lowest and highest values within 1.5× the interquartile range. **b**, Heatmap showing the normalized abundance of GO annotations exhibiting significant variance in frequency between the clusters (bottom; *n* = 127 GO annotations). Statistical significance was determined using one-way analysis of variance with Bonferroni's correction (*P* < 0.05). The mean number of genes containing each GO annotation across the 300 genomes is shown (top). Note the graph is on a log$_{10}$ scale but the *y*-axis shows actual values. **c**, Genome copy number for select GO annotations from **b** across the clusters. **d**, Genome copy number for select Pfam annotations across the clusters. **e**, The incidence of high-impact variants (for example, frameshift and loss of start) relative to Af293 was analysed for GO annotations that did not contain significant copy-number variation between the clusters. A selected subset of GO categories with significant variation in the incidence of high-impact variants between the genetic clusters is shown. **c–e**, The boxes denote the mean (crossbar) ± s.e.m. for the isolates of each cluster.

to variants in three protein-coding genes (Extended Data Fig. 5d and Supplementary Table 3). These included a microtubule bundle protein (*Afu2g16260*), a FGGY-family kinase induced by heat shock (*Afu4g04680*) and its adjacent, uncharacterized ORF *Afu4g04690*. The role of these genes in triazole resistance is an exciting area to follow up on.

ARTICLES
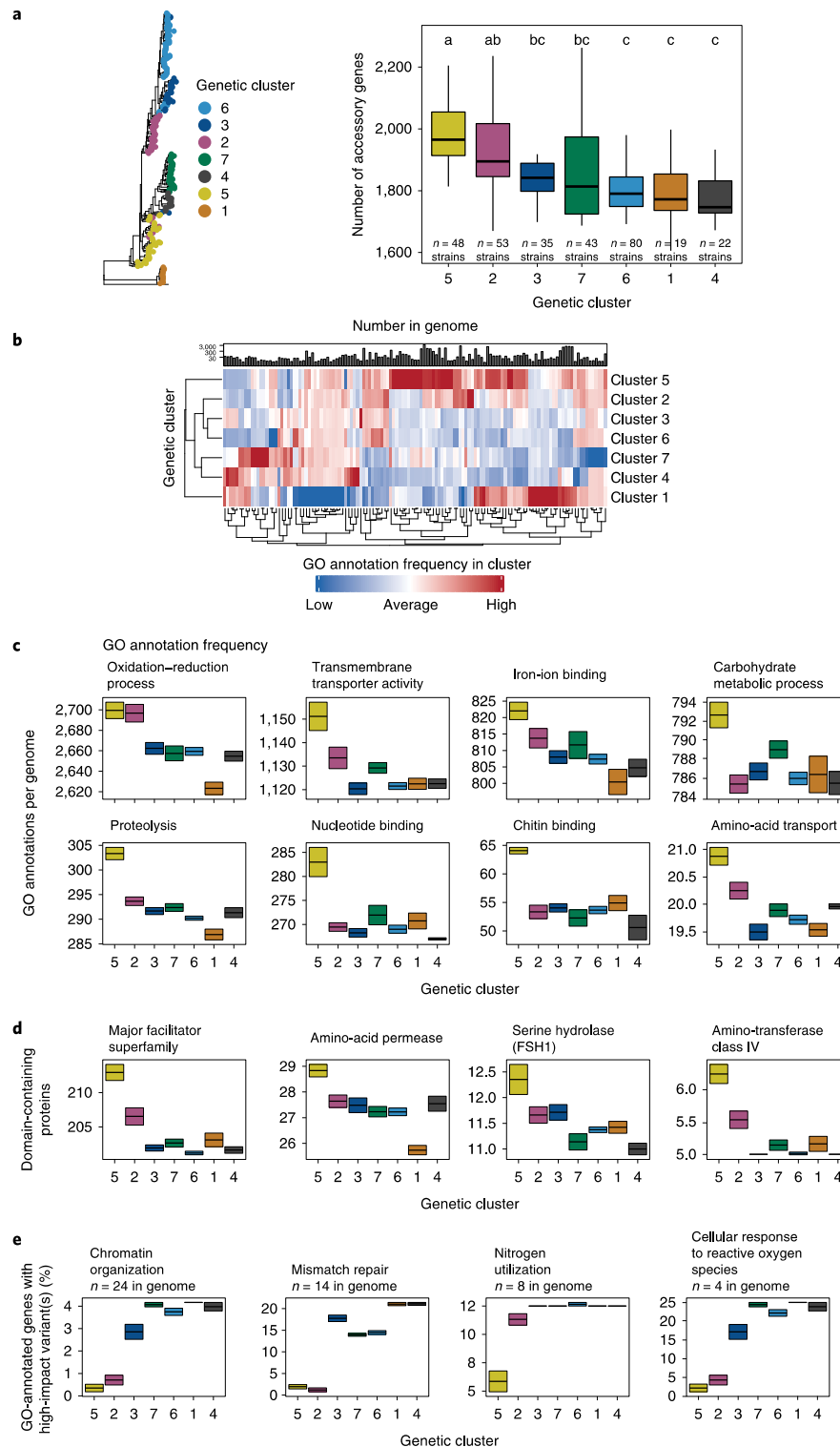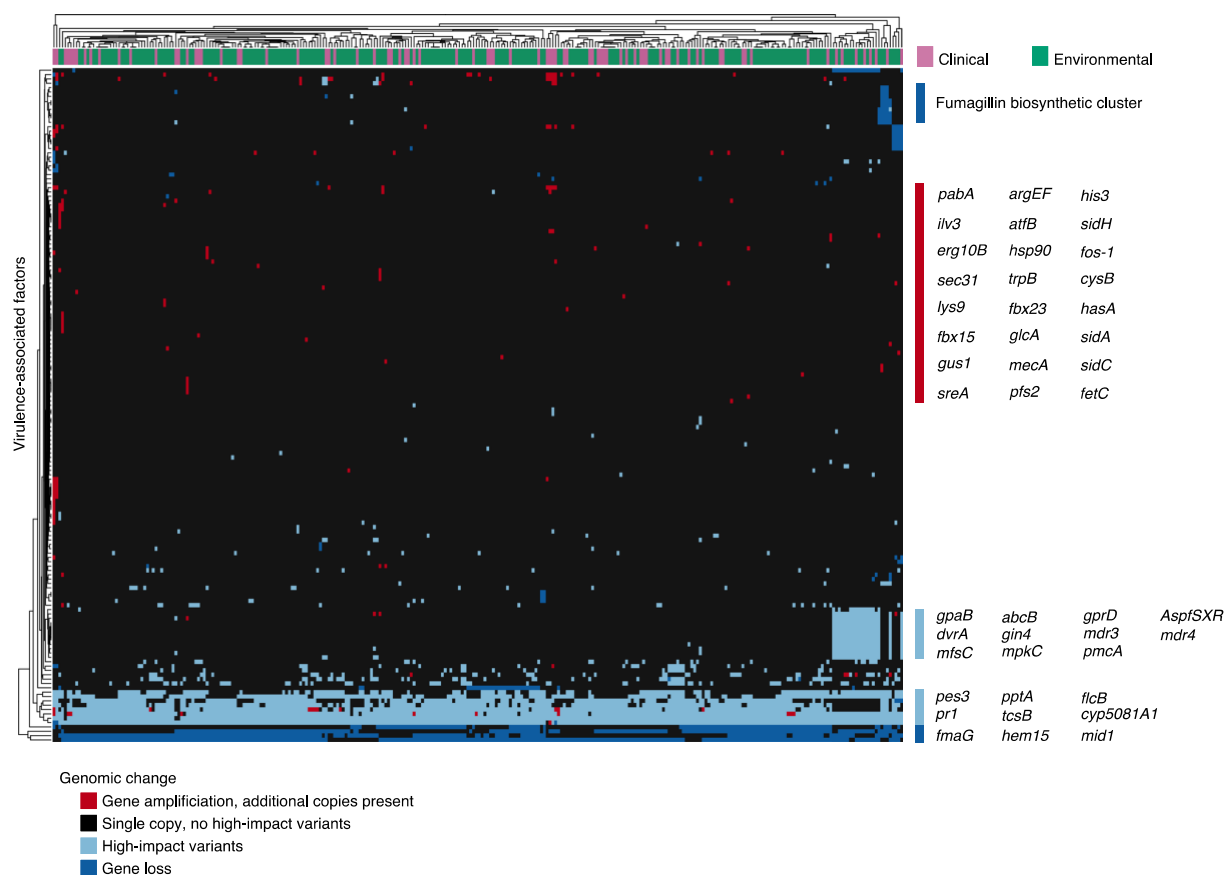
Fig. 5 | Genomic variation among known *A. fumigatus* virulence-associated factors. Heatmap of 155 virulence-associated genes where variation in copy number (gene loss or amplification) or the presence of high-impact variants (for example frameshift, loss or gain of stop codon) was observed relative to Af293 (bottom). The source of the isolate is indicated (top). Gene names for select virulence-associated factors are annotated (right).

Discussion

In this study we defined the pan-genome of *A. fumigatus* using 300 genomes, including a large number of environmental isolates largely absent from previous analyses[15–19]. Compared with *A. fumigatus*, the human commensal and opportunistic pathogen *Candida albicans* was shown to have a lower level of pan-genomic diversity, with 91% of pan-genes present in all isolates[37]. In the same study, a proof-of-concept pan-genome for *A. fumigatus* was also built using genomic data from 12 isolates and 83% of the pan-genome was found to be conserved in all isolates[37]. By contrast, our findings indicate that *A. fumigatus* has a much larger pan-genome and only 69% of the genes identified are present in all isolates; this discrepancy in results is likely to be due to the limited number of isolates included in the former study. In addition, the average BUSCO genome completeness of the assemblies used for their analysis was below 85%, suggesting that notable genetic content was unaccounted for[37]. Future work utilizing chromosome-level assemblies of *A. fumigatus* isolates will allow for a finalized pan-genome of the species with additional information on the evolutionary dynamics of chromosomal organization.

Through pan-genomic analyses we discovered notable genetic variation in virulence factors that have largely only been studied in one or two reference strains. Although most of these cases were infrequent and observed in fewer than 5% of the isolates, some, such as pseudogenization of the non-ribosomal peptide synthetase

nrps8 (or *pes3*), was observed in 97% of the isolates. The largest virulence-associated genetic variation was in secondary metabolism genes, an observation in line with a previous study of 66 isolates[38]. Both studies observed low-incidence variation in the gliotoxin and fumagillin/pseurotin biosynthetic gene clusters as well as high-incidence variation in the fumigermin biosynthetic gene cluster. In addition, although our analysis quantified high-impact genetic changes in virulence determinants, there is almost certainly additional genetic variation that impacts fungal virulence that is difficult to predict on the global scale. The genomes generated here provide a valuable resource for addressing how intraspecies variability in virulence determinants affects infection.

We observed an enrichment for clinical isolates in genetic cluster 5, suggesting that this genetic background might be more fit in the human environment. However, clinical isolates were distributed throughout the phylogeny, highlighting the overall fitness of *A. fumigatus*. In addition, this organism can take advantage of numerous, diverging clinical diseases to establish an infection. We thus performed a genome-wide association study (GWAS) to identify fungal variants associated with clinical disease in general as well as acute and chronic disease, and identified largely overlapping gene sets. However, information on the underlying clinical disease was not available for all samples and the isolates from chronic disease represented a small fraction of the dataset. Future genomic analyses including additional samples from specific underlying diseases will

further illuminate the complex interplay between *A. fumigatus* and specific host disease environments.

The rising incidence of resistance to first-choice antifungals, the triazoles, is a major challenge for the management of *A. fumigatus* infections. This problem is further complicated by up to 30% of the isolates having no identifiable resistance mechanism. We performed GWAS and identified 12 genes associated with triazole resistance. These hits included previously identified variants in the triazole target gene *cyp51a* as well as genes that had not been previously linked to triazole resistance. As a caveat, association studies are underpowered at detecting associations with rare variants. Accordingly, we only screened for association of genetic variants present in at least 1% and 5% of samples. Thus, there are potentially additional resistance-associated variants that were not identified by our analysis. This is perhaps the case for the HMG-CoA demethylase *hmg1*. Clinically observed mutations in this gene conferred triazole resistance to *A. fumigatus* following reconstruction in an isogenic background[39]. Although our analysis did not identify any variants of this gene associated with triazole resistance, manual examination uncovered three triazole-resistant isolates containing single non-synonymous substitutions in *hmg1* (E306D, P309L and C369R). These variants were not considered in the GWAS due to their low prevalence in the dataset. However, the exact role these substitutions play in the resistance of this isolate is unclear, particularly for one isolate that also contained *cyp51a* alterations associated with resistance (TR$_{46}$/G448S). No variants were observed in *hapE* and *cyp51b*, two additional genes linked to triazole resistance.

In summary, this study provides a comprehensive view of the genetic diversity in this important human fungal pathogen. Characterization of the intraspecies diversity and moving away from a single reference genome that does not necessarily represent *A. fumigatus* as a whole will ultimately help us understand its metabolic and pathogenic versatility.

Methods

A. fumigatus isolates analysed in this study. Of the 300 isolates analysed, 188 (49 clinical and 139 environmental isolates) were newly sequenced as part of this study. The 49 clinical isolates sequenced were collected by the German National Reference Center for Invasive Fungal Infections between 2014 and 2018. Bronchial alveolar lavage was the most frequent form of sample collection, representing 31% (15/49) of clinical isolates. The remaining clinical samples were isolated from other pulmonary sources, such as sputum or bronchial secretions, and the exact site of isolation was unavailable for 10% (5/49) of the samples. The 139 environmental isolates sequenced as part of this study were obtained from soil sampling of 11 farms in Germany between 2016 and 2018. Sixty-four of the remaining 112 isolates had been previously sequenced by us as part of a previous study[20] (BioProject PRJNA595552), while 48 had been previously sequenced by other groups and data downloaded from the NCBI Sequence Read Archive. In total, the dataset was comprised of 213 environmental isolates and 87 clinical isolates from Europe, Asia, North America, South America and the International Space Station. A detailed list of the isolates and their metadata characteristics can be found in Extended Data Fig. 1.

Antifungal susceptibility testing. The 188 novel isolates included in this study were screened for azole resistance using the agar-based VIPcheck Assay (Mediaproducts BV) based on EUCAST E.DEF 10.1 following the manufacturer's protocol. Isolates that showed distinguishable germination and hyphal growth on any of the azole-containing wells were subjected to EUCAST broth microdilution (protocol E.DEF 9.3.2; ref. [21]) to define the minimum inhibitory concentrations. Antifungal susceptibility in the clinical isolates was also assessed following EUCAST protocol E.DEF 9.3.2 and resistance was defined in both isolate sets using the EUCAST-established clinical breakpoints. Antifungal susceptibility data from published isolates were obtained from the source publications detailed in Supplementary Data 1.

Genome sequencing and quality assessment. Genomic DNA was extracted from isolates (cultured in Sabouraud Glucose broth at 37 °C with shaking) using a Quick-DNA fungal/bacterial miniprep kit (Zymo Research) following the manufacturer's suggested protocol. Library preparation and Illumina 2 × 150 bp paired-end sequencing were performed on a NextSeq 500 v.2 by LGC Genomics (environmental isolates) and GeneWiz (clinical isolates). Raw FASTQ files were filtered for quality using the following steps: adaptor sequences were removed, bases with an overall quality score of <20 were trimmed and reads shorter than 30 bp were removed. The remaining sequences were verified for quality using FastQC v.0.11.5 (Babraham Institute).

Reference-guided genome analysis. High-quality sequencing reads were aligned to the *A. fumigatus* Af293 reference genome v.2015-09-27 using BWA-MEM v.0.7.8-r779-dirty[40]. PCR duplicates were marked using MarkDuplicate from Picard v.2.18.25. Variant calling to detect SNVs and short indels was performed using the GATK Toolkit (v.4.1.0.0)[41]. Briefly, before variant calling, BAM files were recalibrated using GATK BaseRecalibrator, ApplyBSQ and an in-house dataset of known SNVs generated from the Af293 reference genome and SNVs present in FungiDB, release 39, with ≥80% read frequency and base call ≥ 20. Variant detection was performed using HaplotypeCaller and high-quality variants were identified using GATK best practices (SNP: QD < 2.0 ‖ MQ < 40.0 ‖ FS > 60.0 ‖ MQ RankSum < −12.5 ‖ ReadPosRankSum < −8.0; indel: QD < 2.0 ‖ FS > 200.0 ‖ ReadPosRankSum < −20.0). For downstream analyses, individual VCF files were combined into a single file using bcftools v.01.1.1. Variant function was predicted using SnpEff v.4.3t[42] and 1,000 bp as the cutoff for upstream and downstream flanking of the ORFs. To balance the analysis of high-quality variants with the potential bias introduced by true variants being discarded due to insufficient support, individual variants that failed the quality filter in a sample were included in the variant dataset if at least 95% of the total samples with a variant at that position passed the quality control. This hybrid-filtered variant dataset was used as the input for GWAS and genetic diversity analyses.

A. fumigatus has two mating types (MAT1-1 and MAT1-2), which are encoded within idiomorphic loci on chromosome 3 (refs. [43,44]). The mating type was assigned by calculating the genomic coverage at *Afu3g06160* and *Afu3g06170* using the knowledge that MAT1-2 isolates, including the reference strain Af293, contain a truncated copy of the HMG box mating-type transcription factor (*Afu3g06170*) and an additional gene (*MAT1-2-4*; also known as *Afu3g06160*) that are absent in MAT1-1 isolates. Isolates showing zero coverage in the genomic region of *Afu3g06160* following alignment to Af293 were assigned the mating type MAT1-1. Samples that were not assigned the mating type MAT1-1 were confirmed to be the mating type MAT1-2 through calculation of the genomic coverage at *MAT1-2-1* and *MAT1-2-4*. The ratio of coverage for *MAT1-2-1* and *MAT1-2-4* relative to the genome-wide depth of coverage was between 0.75 and 1.25 for all samples that were assigned the mating type MAT1-2.

Analysis of genomic diversity. Genomic diversity statistics were calculated based on SNV data generated as described earlier. The nucleotide diversity (π) was also calculated using VCFtools[45] with a window size of 5,000 bp and a 500 bp step size. To ensure that differences in sample sizes between the isolate populations did not skew the results, environmental and clinical samples from the acute disease group were downsampled to match the number of isolates from chronic disease in the dataset.

De novo genome assembly and annotation. Genomes were assembled de novo using IDBA-hybrid v.1.1.3 with the Af293 reference genome as a guide[46]. The quality of the genome assembly was assessed using QUAST v.5.0.2 (ref. [47]). Contigs that were shorter than 500 bp or possessing >95% identity and coverage overlap with other contigs were removed. Gene prediction and functional annotation were performed using Funannotate pipeline v.1.5.2-4cfc7f8 (ref. [48]), integrating the following steps. Assemblies were masked for repetitive elements using RepeatMasker (v.4.0.8)[49] using Dfam and RepBase repeat libraries[50]. Gene prediction was performed using EvidenceModeler v.1.1.1 (ref. [51]), incorporating evidence data generated using GeneMark-ES[52] (minimum gene length, 120 bp; and maximum intron length, 3,000 bp) and Augustus[53] (training set, *A. fumigatus*). Gene models predicted to encode peptides shorter than 50 amino acids or transposable elements, or to include span gaps were removed. Transfer-RNA prediction was performed using tRNAscan-SE v.2.0 (ref. [54]). Functional annotation was predicted using PFAM v.43 (ref. [55]), MEROPS v.12 (ref. [56]), dbCAN2 release 7.0 (ref. [57]) and BUSCO v.4.1.4 (ref. [58]). KofamScan v.1.2.0-0 (ref. [59]) was used to assign Kyoto Encyclopedia of Genes and Genomes orthologues to predicted protein sequences and InterProScan v.5.19 (ref. [60]) was used to identify the protein families.

Pan-genome analysis. OrthoFinder was used to identify and cluster orthologous genes[61]. Clustering was performed on the protein sequences of the 300 *A. fumigatus* genomes analysed in this study. In addition, protein sequences from the reference strains Af293 and A1163 were added to improve the identification of the cluster functions. Orthologous gene clusters were assigned a gene identifier from Af293 if they grouped with a single sequence of Af293. If a cluster was not assigned a Af293 gene identifier, but a single A1163 sequence was present, the cluster was assigned the gene identifier from A1163. Orthologous clusters that could not be grouped with a single Af293 or A1163 gene were queried against the NCBI RefSeq non-redundant protein database using DIAMOND using the following criteria: E-value cutoff of 1 × 10⁵, percent identity > 70%, minimum query coverage > 50% and minimum subject coverage > 50%. If at least 70% of the protein sequences in the cluster were assigned to any protein in the NCBI non-redundant protein database, the cluster name was assigned to the name of the RefSeq with the highest contribution. If only 50–70% of the protein sequences in a cluster were assigned to the same protein, the matching sequences were assigned the name of the RefSeq match and the

remaining sequences were left unassigned. The remaining clusters without a match in Af293, A1163 or the non-redundant database were considered novel clusters and had putative functions assigned based on their Funannotate (KofamScan and InterProScan) prediction. For these clusters that were not present in Af293 or the non-redundant database, only clusters present in at least 5% of samples were included to limit false gene predictions. The pan-genome was defined based on gene presence/absence variation in the approved cluster meeting the above criteria. Enrichment analysis was performed using a Fisher's exact test with Bonferroni's correction.

Whole-genome phylogeny. The core genome phylogeny (Fig. 2) was inferred from 5,380 single-copy orthologous genes shared by the two reference strains Af293 and A1163, the 300 *A. fumigatus* genomes analysed in this study, the related species *A. oerlinghausenensis* and *A. fischeri*, which was used to root the tree. Orthologues were identified and clustered using OrthoFinder[61]. Cluster peptide sequences were aligned using MUSCLE v.3.8.1551 (ref. [62]). The resulting peptide alignment was back-translated to a nucleotide sequence using PAL2NAL[63] and concatenated. The phylogeny was inferred from this core nucleotide alignment using IQ-TREE 2 (ref. [64]). The ModelFinder Plus module of IQ-TREE 2 was used to identify GTR+F+R8 as the best fitting substitution and site heterogeneity models for phylogeny construction. Branch support was computed using UFBoot2 ultrafast bootstraps[65]. ClonalFrameML[66] was then used to account for recombination in the phylogeny and rescale branch lengths accordingly.

The SNV-based phylogenies (Extended Data Fig. 4c,d) were constructed by first filtering out loci that showed zero coverage in any sample. For the phylogeny constructed from neutral loci, fourfold degenerate sites were used. For both the non-zero coverage and neutral loci phylogenies, SNVs were concatenated and used as the input for IQ-TREE 2. As with the core nucleotide phylogeny, ModelFinder was employed and identified GTR+F+ASC+R8 as the best fitting model for the non-zero coverage phylogeny and TVM+F+ASC+R8 as the best fitting model for the neutral loci phylogeny. Branch supports were calculated using UFBoot2.

Genetic clusters were identified using discriminant analysis of principle components[22]. To create phylogenetic network trees with clearly visible branches and network structure, the genomes were downsampled by randomly selecting ten genomes per cluster, resulting in a total of 70 samples. Neighbour-net phylogenies were inferred and visualized using the R package phangorn (v.2.5.5)[67] based on a similarity matrix of core nucleotide sequences for the whole-genome network phylogeny and nucleotide sequence alignment for the *cyp51a* and *cyp51b* genes with network phylogenies. The phylogenies were visualized using the R package Ggtree[68].

Estimation of d_N/d_S. The protein-coding sequences of each gene cluster were aligned using MUSCLE v.3.8.1551 (ref. [62]). PAL2NAL[63] was then used to convert the resulting amino-acid alignment to a nucleotide alignment that records whether a base-pair substitution resulted in a synonymous or non-synonymous change. Finally, the CODEML package of PAML[69] was used to calculate the d_N/d_S value of each orthogroup. Median values were used for comparison.

Gene co-occurrence in the pan-genome. Gene co-occurrence networks were computed using Coinfinder[70] using a presence/absence matrix of the pan-genome and a significance cutoff of 0.05 by binomial exact test with Bonferroni's correction. Networks were visualized using the R package igraph.

SNV-based GWAS and pan-GWAS. Before analysis, variant classes were assigned as follows: C, SNVs; G, insertions; D, deletions; and A, reference base. VCF files were converted to plink format using VCFtools[45] and filtered using a MAF of 0.05, which resulted in 352,306 SNVs and 24,726 indels for analysis. Positions with a missingness, or the number of individuals where there was SNV information was available, of >1% were removed from the analysis. The GWAS was performed using the EMMA eXpedited (EMMAX) software package[71], applying a linear mixed model with azole resistance (susceptible/resistant), source (environmental/clinical) or clinical disease (chronic/acute infection) as the phenotypic traits. The GEMMA, treeWAS and ECAT software packages were also tested in the framework of this project. EMMAX was ultimately selected over these tools because it accounted for sample structure the best, providing the least-inflated Q–Q plots (Extended Data Fig. 5a,c). Significant variants were determined using a cutoff of $P < 0.01$ with false-discovery-rate correction. The pan-GWAS was performed using a presence/absence matrix of the orthologous gene clusters, where zero denoted absent gene clusters and one represented gene clusters that were present in the genome. Associations between pan-gene presence/absence, isolate source and azole resistance were calculated using Scoary v.1.6.16 (ref. [72]).

Availability of isolates. The isolates that were sequenced in this study were submitted to and are publicly available in the Jena Microbial Resource Collection.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Raw FASTQ files for the isolates sequenced in this study were uploaded to the NCBI Sequence Read Archive and are publicly available under BioProject PRJNA697844. The accession numbers for the publicly available sequence data are listed in Extended Data Fig. 1. Annotated genome assemblies for sequence data generated in this study and for 64 isolates sequenced by us in a previous study[20] were submitted to NCBI GenBank and are available under the NCBI BioSample numbers listed in Extended Data Fig. 1. Datasets from FungiDB, release 39, are available at https://fungidb.org/fungidb/app/downloads/release-39/. The NCBI RefSeq non-redundant protein database v.22.01.08 is accessible at https://ftp.ncbi.nlm.nih.gov/blast/db/cloud/2018-01-22/. Source data are provided with this paper.

References
1. Latgé, J. P. and Chamilos, G. *Aspergillus fumigatus* and Aspergillosis in 2019. *Clin. Microbiol. Rev.* https://doi.org/10.1128/CMR.00140-18 (2019).
2. *Invasive Aspergillosis. LIFE* http://www.life-worldwide.org/fungal-diseases/invasive-aspergillosis (2020).
3. Harrison, N. et al. Incidence and characteristics of invasive fungal diseases in allogeneic hematopoietic stem cell transplant recipients: a retrospective cohort study. *BMC Infect. Dis.* **15**, 584 (2015).
4. Kuster, S. et al. Incidence and outcome of invasive fungal diseases after allogeneic hematopoietic stem cell transplantation: a Swiss transplant cohort study. *Transpl. Infect. Dis.* **20**, e12981 (2018).
5. Heo, S. T. et al. Changes in in vitro susceptibility patterns of *Aspergillus* to triazoles and correlation with aspergillosis outcome in a tertiary care cancer center, 1999–2015. *Clin. Infect. Dis.* **65**, 216–225 (2017).
6. Lestrade, P. P. et al. Voriconazole resistance and mortality in invasive aspergillosis: a multicenter retrospective cohort study. *Clin. Infect. Dis.* **68**, 1463–1471 (2019).
7. Snelders, E. et al. Emergence of azole resistance in *Aspergillus fumigatus* and spread of a single resistance mechanism. *PLoS Med.* **5**, 1629–1637 (2008).
8. Rizzetto, L. et al. Strain dependent variation of immune responses to *A. fumigatus*: definition of pathogenic species. *PLoS ONE* **8**, 2–14 (2013).
9. Kowalski, C. H. et al. Heterogeneity among isolates reveals that fitness in low oxygen correlates with *Aspergillus fumigatus* virulence. *mBio* **7**, e01515-16 (2016).
10. Alshareef, F. & Robson, G. D. Genetic and virulence variation in an environmental population of the opportunistic pathogen *Aspergillus fumigatus*. *Microbiology* **160**, 742–751 (2014).
11. Knox, B. P. et al. Characterization of *Aspergillus fumigatus* isolates from air and surfaces of the International Space Station. *mSphere* **1**, e00227-16.
12. Ries, L. N. A. et al. Nutritional heterogeneity among *Aspergillus fumigatus* strains has consequences for virulence in a strain- and host-dependent manner. *Front. Microbiol.* **10**, 854 (2019).
13. Steenwyk, J. L. et al. Variation among biosynthetic gene clusters, secondary metabolite profiles, and cards of virulence across *Aspergillus* species. *Genetics* **216**, 481–497 (2020).
14. Dos Santos, R. A. C. et al. Genomic and phenotypic heterogeneity of clinical isolates of the human pathogens *Aspergillus fumigatus*, *Aspergillus lentulus*, and *Aspergillus fumigatiaffinis*. *Front. Genet.* **11**, 459 (2020).
15. Fedorova, N. D. et al. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet.* **4**, e1000046 (2008).
16. Abdolrasouli, A. et al. Genomic context of azole-resistance mutations in *Aspergillus fumigatus* using whole-genome sequencing. *mBio* **6**, e00536 (2015).
17. Garcia-Rubio, R. et al. Genome-wide comparative analysis of *Aspergillus fumigatus* strains: the reference genome as a matter of concern. *Genes* **9**, 363 (2018).
18. Fan, Y., Wang Y. and Xu, J. Comparative genome sequence analyses of geographic samples of *Aspergillus fumigatus*—relevance for amphotericin B resistance. *Microorganisms* **8**, 1673 (2020).
19. Puértolas-Balint, F. et al. Revealing the virulence potential of clinical and environmental *Aspergillus fumigatus* isolates using whole-genome sequencing. *Front. Microbiol.* **10**, 1970 (2019).
20. Barber, A. E. et al. Effects of agricultural fungicide use on *Aspergillus fumigatus* abundance, antifungal susceptibility, and population structure. *mBio* **11**, e02213-20 (2020).
21. Arendrup, M. C. et al. Method for the determination of broth dilution minimum inhibitory concentrations of antifungal agents for conidia forming moulds. E.DEF 9.3.2. *EUCAST* https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/AFST/Files/EUCAST_E_Def_9.3.2_Mould_testing_definitive_revised_2020.pdf (2020).
22. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
23. Nierman, W. C. et al. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature* **438**, 1151–1156 (2005).
24. Steenwyk, J. L. et al. Genomic and phenotypic analysis of COVID-19-associated pulmonary aspergillosis isolates of *Aspergillus fumigatus*. *Microbiol. Spectr.* **9**, e0001021 (2021).

25. Abad, A. et al. What makes *Aspergillus fumigatus* a successful pathogen? Genes and molecules involved in invasive aspergillosis. *Rev. Iberoam. Micol.* **27**, 155–82 (2010).

26. Bignell, E., et al. Secondary metabolite arsenal of an opportunistic pathogenic fungus. *Philos. Trans. R Soc. B* **371**, 20160023 (2016).

27. Kjaerbolling, I. et al. Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. *Proc. Natl Acad. Sci. USA* **115**, E753–E761 (2018).

28. Urban, M. et al. PHI-base: the pathogen-host interactions database. *Nucleic Acids Res.* **48**, D613–D620 (2020).

29. O'Hanlon, K. A. et al. Targeted disruption of nonribosomal peptide synthetase pes3 augments the virulence of *Aspergillus fumigatus*. *Infect. Immun.* **79**, 3978–3992 (2011).

30. Sugui, J. A. et al. Genes differentially expressed in conidia and hyphae of *Aspergillus fumigatus* upon exposure to human neutrophils. *PLoS ONE* **3**, e2655 (2008).

31. Willger, S. D. et al. A sterol-regulatory element binding protein is required for cell polarity, hypoxia adaptation, azole drug resistance, and virulence in *Aspergillus fumigatus*. *PLoS Pathog.* **4**, e1000200 (2008).

32. Blatzer, M., et al. SREBP coordinates iron and ergosterol homeostasis to mediate triazole drug and hypoxia responses in the human fungal pathogen *Aspergillus fumigatus*. *PLoS Genet.* **7**, e1002374 (2011).

33. Bertuzzi, M. et al. The pH-responsive PacC transcription factor of *Aspergillus fumigatus* governs epithelial entry and tissue invasion during pulmonary aspergillosis. *PLoS Pathog.* **10**, e1004413 (2014).

34. Bignell, E. et al. The *Aspergillus* pH-responsive transcription factor PacC regulates virulence. *Mol. Microbiol.* **55**, 1072–1084 (2005).

35. Pongpom, M. et al. Divergent targets of *Aspergillus fumigatus* AcuK and AcuM transcription factors during growth in vitro versus invasive disease. *Infect. Immun.* **83**, 923–933 (2015).

36. Camps, S. M. T. et al. Molecular epidemiology of *Aspergillus fumigatus* isolates harboring the TR34/L98H azole resistance mechanism. *J. Clin. Microbiol.* **50**, 2674–2680 (2012).

37. McCarthy, C. G. P. & Fitzpatrick, D. A. Pan-genome analyses of model fungal species. *Microb. Genom.* **5**, e000243 (2019).

38. Lind, A. L. et al. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biol.* **15**, e2003583 (2017).

39. Rybak, J. Mutations in *hmg1*, challenging the paradigm of clinical triazole resistance in *Aspergillus fumigatus*. *mBio* **10**, e00437-19 (2019).

40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

41. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

42. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

43. Varga, J. Mating type gene homologues in *Aspergillus fumigatus*. *Microbiology* **149**, 816–819 (2003).

44. Paoletti, M. et al. Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Curr. Biol.* **15**, 1242–1248 (2005).

45. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

46. Peng, Y., et al. IDBA—a practical iterative de Bruijn graph de novo assembler. In *Proc. Research in Computational Molecular Biology. RECOMB 2010.* (ed. Berger, B.) 426–440 (Springer, 2010).

47. Gurevich, A. et al. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

48. Palmer, J & Stajich J. Funannotate v.1.5.3. *Zenodo* https://zenodo.org/record/2604804 (2019).

49. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. (2013–2015); http://www.repeatmasker.org/

50. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).

51. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).

52. Ter-Hovhannisyan, V. et al. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).

53. Stanke, M. et al. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

54. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57 (2016).

55. Finn, R. D. et al. Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).

56. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **40**, D343–D350 (2012).

57. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).

58. Simao, F. A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

59. Aramaki, T. et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).

60. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

61. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

62. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**, 113 (2004).

63. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).

64. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

65. Hoang, D. T. et al. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

66. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).

67. Schliep, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

68. Yu, G. et al. Two methods for mapping and visualizing associated data on phylogeny using Ggtree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).

69. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

70. Whelan, F. J., Rusilowicz, M. and McInerney, J. O. Coinfinder: detecting significant associations and dissociations in pangenomes. *Microb. Genom.* **6**, e000338 (2020).

71. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).

72. Brynildsrud, O. et al. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **17**, 238 (2016).

Author contributions

A.E.B., G.P. and O.K. conceptualized and designed the study. The experimental work was performed by A.E.B. and G.W. A.E.B., B.S., G.P., K.K., J.L., O.K., T.S.-O. and G.W. analysed the data and interpreted results. A.E.B. wrote the primary manuscript and all of the listed authors were involved in the editing and review of the manuscript. O.K. acquired the primary funding for this work.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-021-00993-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-021-00993-x.

Correspondence and requests for materials should be addressed to Gianni Panagiotou or Oliver Kurzai.

Peer review information *Nature Microbiology* thanks Nancy Keller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

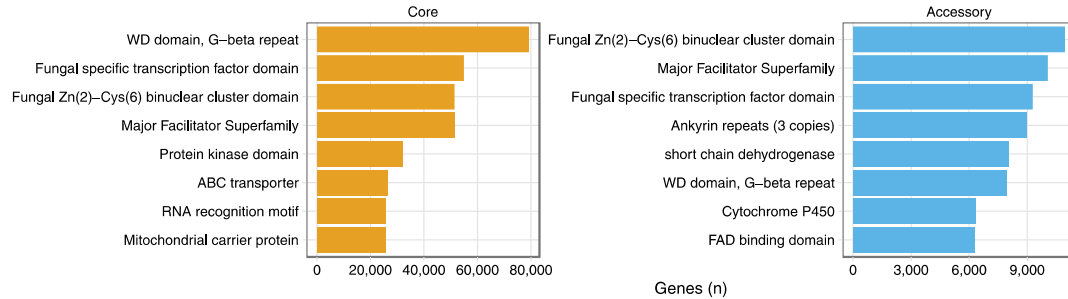**Reprints and permissions information** is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
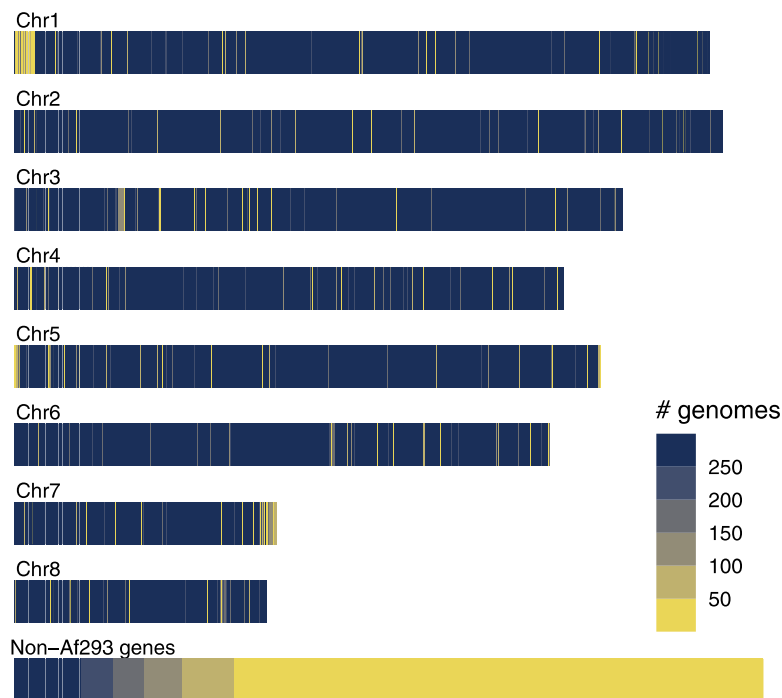
Supplement

a

**Top Pfam domains among core and accessory proteins**

b

Extended Data Fig. 1 | The pan-genome of *A. fumigatus*. (A) Most frequently occurring Pfam domains among the core and accessory genomes. Values represent the total sum of domain-containing proteins among all 300 genomes. (B) Conservation of Af923 genes in the *A. fumigatus* pan-genome, arranged by chromosomal location in Af293. Each gene in Af293 is represented by a uniform-sized band that is coloured according to its prevalence among the 300 isolates analysed. Genes not in Af293 and their relative frequency are depicted at the bottom.
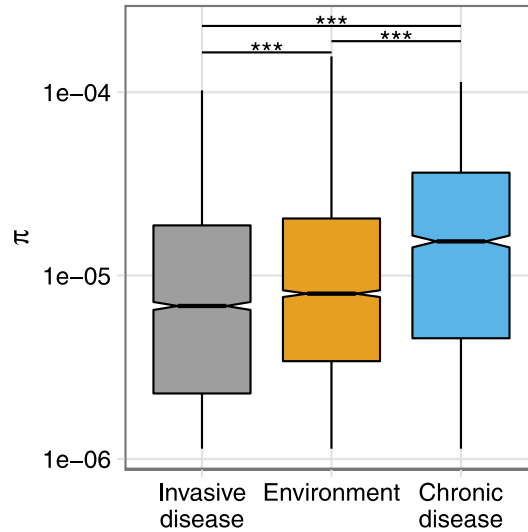
Extended Data Fig. 2 | Nucleotide diversity (π) of *A. fumigatus* isolates from the environment, invasive disease and chronic disease. π was calculated using 5 kb sliding windows across with genome with a 500 bp step size. Due to the underrepresentation of isolates from chronic disease in the dataset, isolates from the environment and invasive disease were downsampled to match the number of isolates from chronic disease (n = 19 isolates per group). The bold line in the box-and-whisker plot indicates the 50th percentile, and the box extends from the 25th to the 75th percentiles. The whiskers denote the lowest and highest values within 1.5 interquartile range. Statistical significance determined by two-sided Mann–Whitney *U* test with Bonferroni correction. *** represents *P* < 0.001. Exact P-values are: chronic vs. environmental: *P* = 97e-39; chronic vs invasive: *P* = 1.6e-78; invasive vs. environmental: P = 5.6e-14.
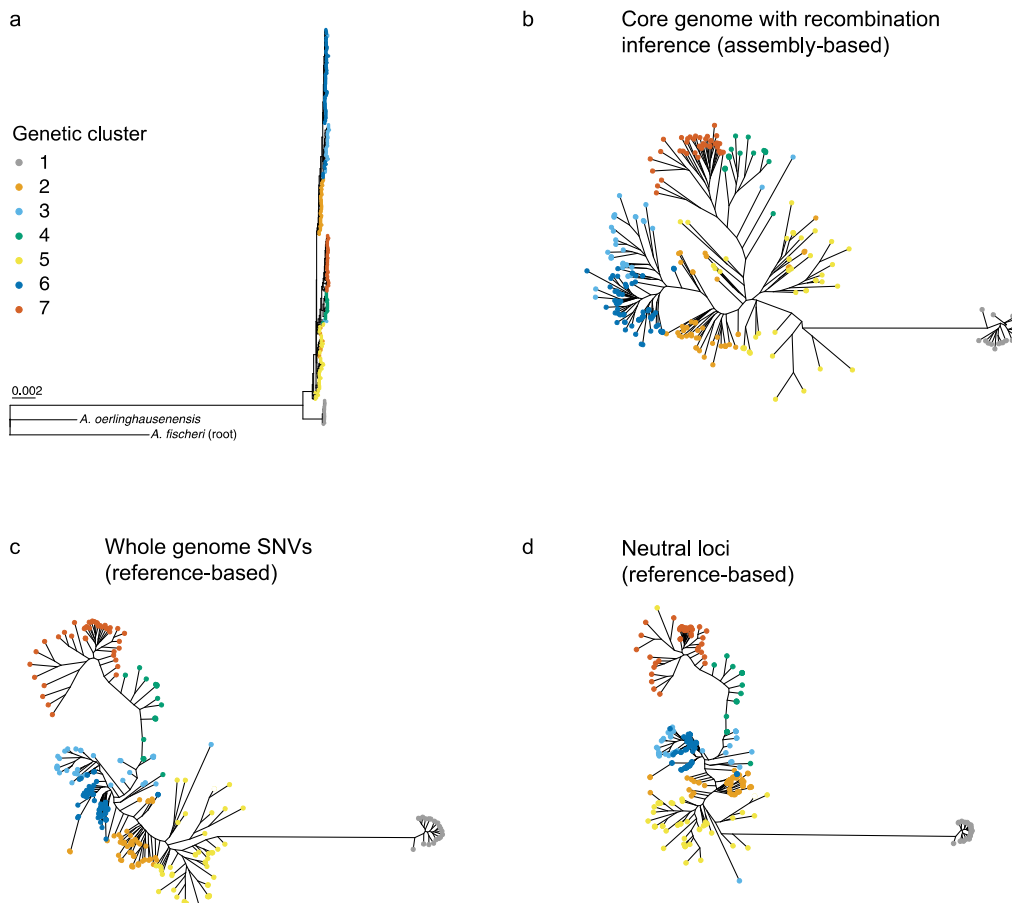
Extended Data Fig. 3 | Phylogenies constructed from the genomes of 300 *A. fumigatus* using de novo assembled genomes and reference-base analyses.
(a-b) Core genome phylogeny built from nucleotide coding sequence of 5,380 single-copy orthologous genes shared by all 300 *A. fumigatus* isolates, *A. oerlinghausenensis* and *A. fischeri* (alignment length = 9,178,893 bp). Panel a shows the phylogeny rooted with *A. fischeri* and depicts the scaled relationship between the two outgroups and the *A. fumigatus* samples. Panel b depicts this phylogeny unrooted and with outgroups removed for comparison to the other phylogenies. (c) Phylogeny from concatenated SNVs following read alignment to Af293 and variant calling (n = 341,031 base pair). Genomic positions with zero coverage in any sample were removed from the alignment. (d) SNV-based phylogeny constructed from 4-fold degenerate (neutral) loci (n = 35,052 base pair).
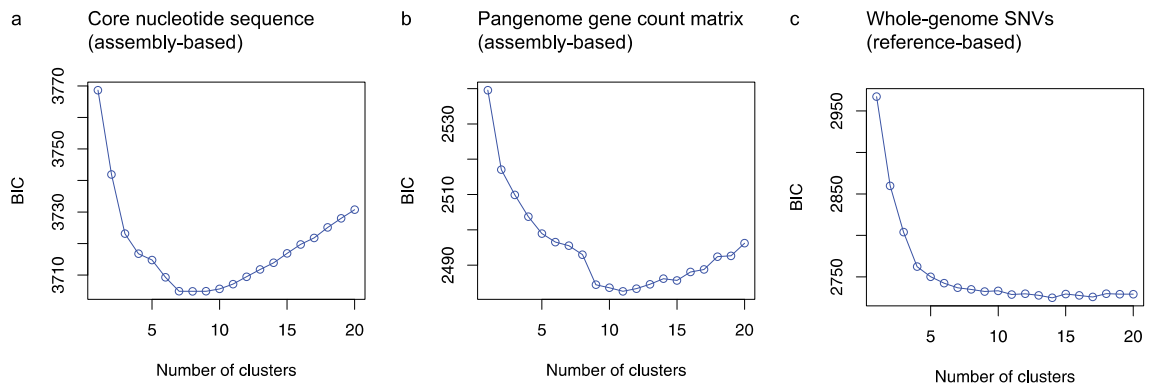
Extended Data Fig. 4 | Discriminant analysis of principle components of 300 *A. fumigatus* isolates. (a-c) Number of clusters vs. Bayesian information criteria (BIC) was used to assess the best supported number of genetic clusters for the dataset. The input for analysis was either (a) a non-gapped core nucleotide alignment from 5,830 single-copy orthologous genes (b) a gene count matrix from orthogroup-based clustering of pansequences or (c) or whole-genome SNV data with of positions with zero genomic coverage in any isolate in the dataset excluded.
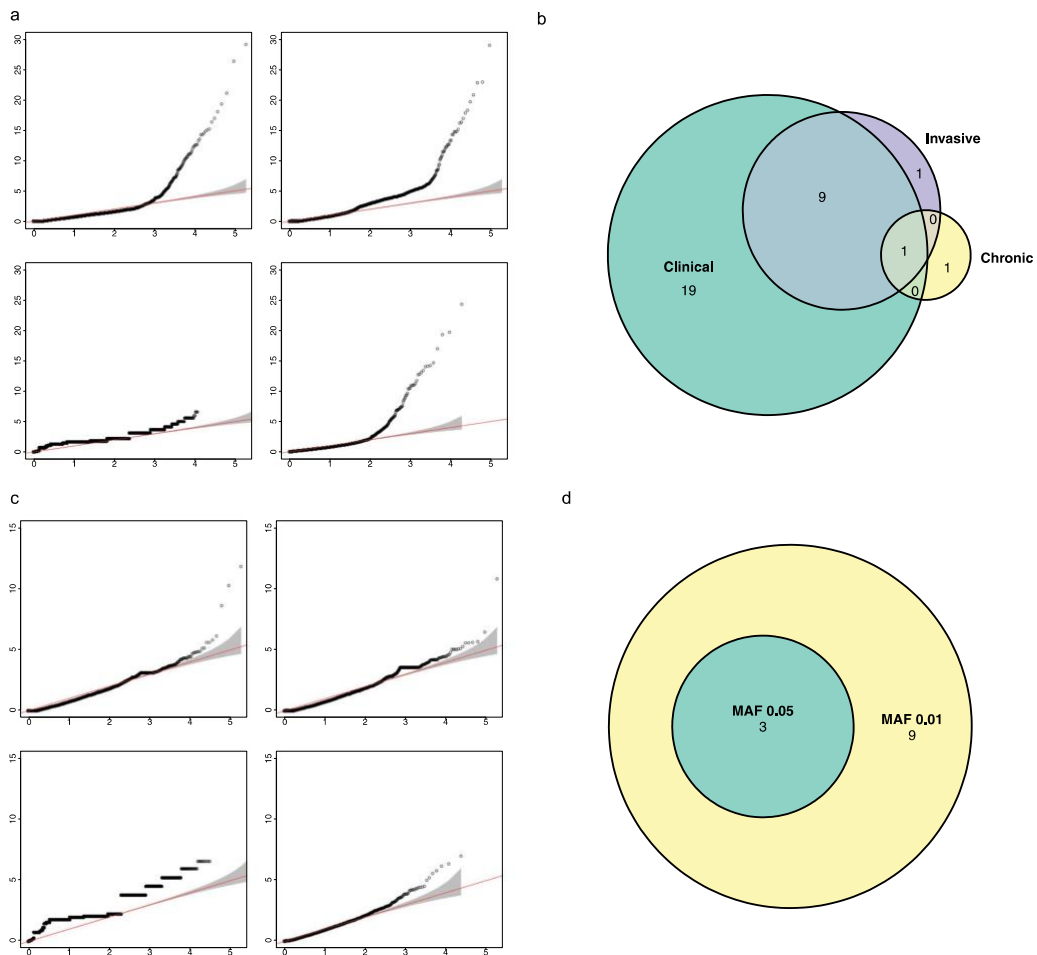
Extended Data Fig. 5 | GWAS for variants associated with clinical isolates and triazole resistance. (a & c) Q–Q plots for association with isolate source (a; clinical vs. environmental) and triazole resistance (c; resistance to one or more triazole vs. susceptible to all examined; c). Four software were utilized: EMMAX (top, left), GEMMA (top, right), treeWAS (bottom, left) and ECAT (bottom, right). The resulting Q–Q plots were used to identify the tool that produced outputs where the expected p-value distribution (x-axis) best matched the observed p-values (y-axis). (b) Venn diagram showing the gene overlap between GWAS for all clinical strains relative to environmental and significant genes specific to acute and chronic disease. (d) Venn diagram showing the gene overlap for association with triazole resistance when minor allele frequencies (MAF) of 0.01 and MAF 0.05 were used.

III.   Manuscript:
Temporal Gut Microbiota Changes during Antibiotics

FORM 1

Manuscript No. 3

Manuscript title: Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact

Authors: Seelbinder B., Chen J., Brunke S., Vazquez-Uribe R., Meyer A. C., De Oliveira Lino F. S., Chan Ka Fai, Imamovic L., Tsang C. C., Woo P. C. Y., Sommer M., Panagiotou G.

Bibliographic information: *Microbiome*. 8 (2020), pp. 1–20, 10.1186/ S40168-020-00899-6

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
Seelbinder	30%	70%		60%	
Chen	30%	20%		20%	
Brunke			20%		
V.-Uribe			20%		
Meyer			20%		
Imamovic			20%		
Woo					20%
Panagiotou	30%			5%	60%
Others	10%	10%	20%	5%	20%
Total:	100%	100%	100%	100%	100%

Overview:
Here, I studied within- and cross-kingdom interactions between gut bacteria and fungi in healthy human individuals longitudinally. I investigated how these interactions are shaped under external perturbation by oral antibiotic administration over 4 different time points. A combination of metagenomics, metatranscriptomics, ITS sequencing, and advanced statistical approaches (MGS, comparative genomics, co-abundance network analysis) allowed me to assess the resilience of the microbiome in terms of functional potential and expression. Strict anaerobe bacteria suffered the most from treatment. However, while some level of recovery was observed in the bacterial community, the effects on the fungal community seemed more stochastic and lasting. The fungal opportunist *Candida* abundance increased shortly after treatment but was effectively inhibited later.

Microbiome

Open Access

Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria

Bastian Seelbinder[1†], Jiarui Chen[1,2†], Sascha Brunke[3], Ruben Vazquez-Uribe[4], Rakesh Santhaman[1], Anne-Christin Meyer[3], Felipe Senne de Oliveira Lino[4], Ka-Fai Chan[5], Daniel Loos[1], Lejla Imamovic[4], Chi-Ching Tsang[5], Rex Pui-kin Lam[6], Siddharth Sridhar[5,7,8], Kang Kang[1], Bernhard Hube[4], Patrick Chiu-yat Woo[5,7,8], Morten Otto Alexander Sommer[4*] and Gianni Panagiotou[1,2*] (iD)

Abstract

Background: Antibiotic treatment has a well-established detrimental effect on the gut bacterial composition, but effects on the fungal community are less clear. Bacteria in the lumen of the gastrointestinal tract may limit fungal colonization and invasion. Antibiotic drugs targeting bacteria are therefore seen as an important risk factor for fungal infections and induced allergies. However, antibiotic effects on gut bacterial-fungal interactions, including disruption and resilience of fungal community compositions, were not investigated in humans. We analysed stool samples collected from 14 healthy human participants over 3 months following a 6-day antibiotic administration. We integrated data from shotgun metagenomics, metatranscriptomics, metabolomics, and fungal ITS2 sequencing.

Results: While the bacterial community recovered mostly over 3 months post treatment, the fungal community was shifted from mutualism at baseline to competition. Half of the bacterial-fungal interactions present before drug intervention had disappeared 3 months later. During treatment, fungal abundances were associated with the expression of bacterial genes with functions for cell growth and repair. By extending the metagenomic species approach, we revealed bacterial strains inhibiting the opportunistic fungal pathogen *Candida albicans*. We demonstrated in vitro how *C. albicans* pathogenicity and host cell damage might be controlled naturally in the human gut by bacterial metabolites such as propionate or 5-dodecenoate.

Conclusions: We demonstrated that antibacterial drugs have long-term influence on the human gut mycobiome. While bacterial communities recovered mostly 30-days post antibacterial treatment, the fungal community was shifted from mutualism towards competition.

---

* Correspondence: msom@bio.dtu.dk; Gianni.Panagiotou@hki-jena.de
†Bastian Seelbinder and Jiarui Chen contributed equally to this work.
⁴Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, DK-2800 Lyngby, Denmark
¹Leibniz Institute for Natural Product Research and Infection Biology—Systems Biology and Bioinformatics, Hans Knöll Institute, Adolf-Reichwein-Straße 23, 07745 Jena, Germany
Full list of author information is available at the end of the article

Background

The human gut microbiome is a complex ecosystem of bacteria, fungi, archaea, and phages [1]. The majority of research has focused on the bacterial part of the gut microbiome and their role in health and disease [2–4]. However, the critical role of fungi in host homeostasis remains is less well studied. Fungal dysbiosis may increase symptoms of inflammation, especially in the gut lumen [5]. Treating mice with fluconazole, an antifungal drug, increases the immune response and severity of experimentally induced colitis [6] but also induced allergic airway disease [7]. Fluconazole seems to substantially impact only certain types of fungi such as *Candida*, but not *Aspergillus* species [6].

Antibiotic treatment has a well-established detrimental effect on the composition of gut bacteria [8–11], but the effect on the fungal community is less clear. Nevertheless, antibiotic use is linked to overgrowth of particular fungi at multiple body sites [7, 8, 12]. Noverr et al. used a murine model to induce development of airway allergies by enriching for *Candida* and *Aspergillus* species in the gut followed by antibiotic treatment [10]. Theoretically, commensal bacteria may limit fungal colonization by production of antifungal compounds [13], competition for available nutrients, cellular contact, chemotaxis, or physiochemical changes to the local environment [14, 15]. Fungi defend themselves by secreting molecules, forming biofilms or forming mutualistic bonds with other bacteria. *Candida albicans*, for example, secretes the metabolite farnesol which interferes with the quorum-sensing of *Pseudomonas aeruginosa* [14, 15]. However, *C. albicans* can also enhance biofilm formation by *Staphylococcus aureus* in vitro. *Pseudomonas fluorescens* promotes the growth of the mycorrhizal fungus *Laccaria bicolor* in soil. Which bacterial-fungal interactions take place in the gastrointestinal tract of humans remains to be investigated. To date, the complex community of gut microbes is thought to be coevolved to maintain relative homeostasis in healthy humans [16].

Defining gut fungal consortia and their stability, resilience, and dynamics may reveal cause-effect relationships with bacteria. Although evidence is available on bacterial-fungal interactions in the gut at the taxonomic level [13–15], we do not have a comprehensive understanding of how bacterial functions influence the growth of particular fungi. Bacterial microbiome studies were often performed by amplifying the DNA of the ribosomal 16S gene. However, metagenome shotgun sequencing allows species- and sometimes even strain-level taxonomic classification, as well as the estimation of gene functions [17–19]. Furthermore, gene expression in microbial communities is not strictly matched with metagenomic potential [20]. Often, studies neglect the high transcriptional activity of some less-abundant species to metabolic functions.

In order to better understand the microbiome, we provide data to follow both, the bacterial and fungal communities of the lower human gastrointestinal tract over 3 months after antibiotic treatment concomitantly. We provide an overview of how the mycobiome and its interactions with the bacterial microbiome change and we reveal dependencies of specific fungal species from bacterial functions at DNA and RNA levels.
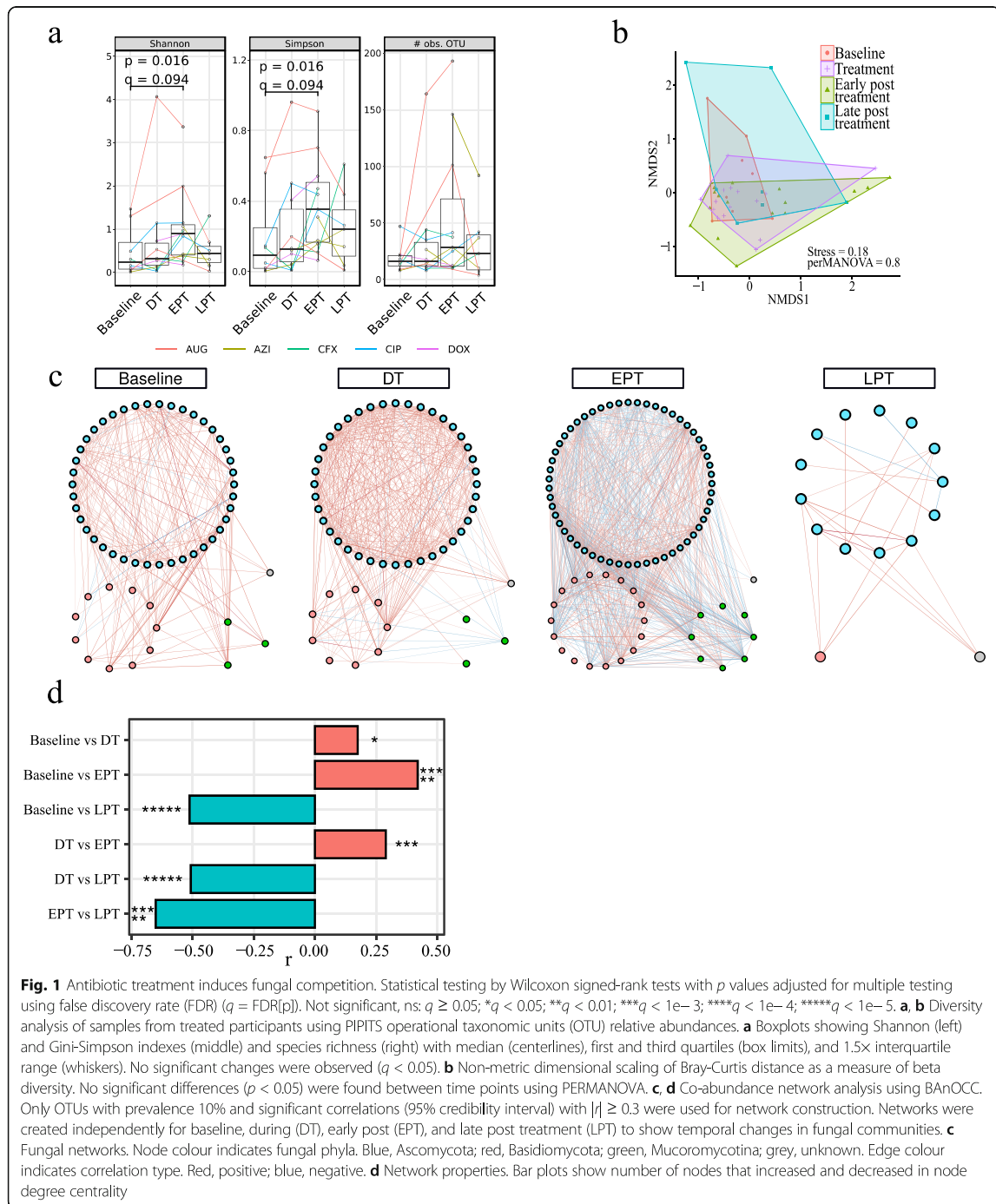
Results

Antibiotic treatment triggers long-lasting dynamics at fungal species level

We included 14 healthy human participants, 12 receiving the antibiotic intervention and 2 controls. Stool samples were collected at 4–6 time points per participants. We used 5 different antibiotics (one for each pair of treated participants). Samples were collected 15 days before administration of antibiotics (baseline), at 4 and 6 days of treatment (during treatment [DT]), 15 and 30 days after (early post treatment [EPT]), and 90 days after treatment (late post treatment [LPT]). We built high-quality libraries for ITS2 sequencing for 59 of 74 available stool samples. We estimated the fungal relative abundance using the PIPITS pipeline [21]. ITS sequences were clustered into operational taxonomic units (OTU) and taxonomically annotated using Mothur [22]. Antibiotic treatment led to a significant increase in species-level fungal alpha diversity during early post treatment compared to baseline (Fig. 1a; two-sided Wilcoxon rank-sum test, $p = 0.016$, $q = 0.094$). Controls showed a considerable increase as well, although statistical significance could not be estimated due to the number of subjects ($n = 2$). At the level of individual antibiotic drugs (Suppl. Fig. 1), Augmentin and ciprofloxacin more than doubled baseline diversity. In contrast, changes for doxycycline and azithromycin were mild. Beta diversity using Bray-Curtis was not significantly different between time points in treated samples (Fig. 1b; PERMANOVA, $p > 0.05$).

We subsequently investigated differences in fungal genera relative abundance over time. *Candida* genus increased 15-fold from baseline to treatment ($q = 0.004$; Suppl. Table 1). *Candida* increase was observed for all antibiotics except Augmentin (Suppl. Fig. 2). At the species level, results were more distinct (Suppl. Table 2) and for this analysis, we considered only prevalent fungal species (defined as present in 15% of samples). Comparing relative abundance changes from baseline to during treatment, only *Saccharomyces sydowii* decreased significantly ($q < 0.05$). However, the opportunistic pathogen *Candida albicans* tended to increase 7-fold ($q < 0.07$) and was affected the most by Augmentin and doxycycline (Suppl. Fig. 3). Furthermore, *C. albicans* was

Fig. 1 Antibiotic treatment induces fungal competition. Statistical testing by Wilcoxon signed-rank tests with *p* values adjusted for multiple testing using false discovery rate (FDR) (*q* = FDR[p]). Not significant, ns: *q* ≥ 0.05; **q < 0.05; ***q < 0.01; ****q < 1e− 3; *****q < 1e− 4; ******q < 1e− 5. **a**, **b** Diversity analysis of samples from treated participants using PIPITS operational taxonomic units (OTU) relative abundances. **a** Boxplots showing Shannon (left) and Gini-Simpson indexes (middle) and species richness (right) with median (centerlines), first and third quartiles (box limits), and 1.5× interquartile range (whiskers). No significant changes were observed (*q* < 0.05). **b** Non-metric dimensional scaling of Bray-Curtis distance as a measure of beta diversity. No significant differences (*p* < 0.05) were found between time points using PERMANOVA. **c**, **d** Co-abundance network analysis using BAnOCC. Only OTUs with prevalence 10% and significant correlations (95% credibility interval) with |*r*| ≥ 0.3 were used for network construction. Networks were created independently for baseline, during (DT), early post (EPT), and late post treatment (LPT) to show temporal changes in fungal communities. **c** Fungal networks. Node colour indicates fungal phyla. Blue, Ascomycota; red, Basidiomycota; green, Mucoromycotina; grey, unknown. Edge colour indicates correlation type. Red, positive; blue, negative. **d** Network properties. Bar plots show number of nodes that increased and decreased in node degree centrality

detected in nine participants after treatment even though in only five at baseline.

Twenty-three species changed significantly from treatment to early post treatment periods (*q* < 0.05), many of which were not present before or after treatment. Many common fungi like *Saccharomyces* spp., opportunistic pathogens such as *C. albicans*, *C. parapsilosis*, and *Malassezia restricta*—a fungus recently connected to

pancreatic cancer [23]—decreased in abundance, whereas less common fungi such as *Candida boidinii* increased in abundance. A minor decrease in abundance of *C. albicans* was also observed in controls, but not nearly as much.

To test for long-lasting changes, we compared relative abundances at baseline to late post treatment and found six species with significant changes. We further noticed that only 14 fungal species passed the prevalence filter at baseline and late post treatment, whereas up to 44 were observed during and early post treatment, suggesting that antibiotics temporarily created a niche for less common fungal species. In summary, the number of detected, prevalent species more than doubled during treatment and early post treatment, but these species had not successfully colonized 3 months later. Most changes were found within the first month after treatment, implying a delayed response of the fungal community to the treatment. Over one third of the fungal species present before treatment showed significant changes even 90 days after treatment.

Antibiotic treatment increases co-exclusion in fungal communities

We evaluated changes induced in the mycobiome from antibiotic administration by creating co-abundance networks based on ITS abundances. Networks were created for baseline and for during, early, and late post treatment periods (Fig. 1d). Only significant edges (95% credibility) with absolute correlation of at least 0.3 were retained. Generally, we found significant correlations within and between *Ascomycota*, *Basidiomycota*, and *Mucoromycotina* species. At baseline, we found mostly positive correlations (240 positive and 10 negative) among 57 fungal species. During treatment, the number of correlations almost doubled (406 positive, 17 negative), whereas at early post treatment, correlation numbers doubled again. In contrast to the previous networks, more than half of the significant correlations were negative (399 negative, 550 positive), implying a major switch from mutualistic relationships at baseline and during treatment to competition between fungal species as they try to re-establish a stable community. We also observed these negative correlation patterns within and between fungal phyla. At late post treatment, this conflict persisted. Most co-abundance patterns had disappeared—only 25 correlations among 15 fungal species remained. We confirmed these trends by testing for significant changes in node degree centrality (Fig. 1e; Suppl. Table 3).

In conclusion, based on diversity, abundance, and network analysis, we observed that gut fungal communities started to change alongside the bacterial communities during treatment. Many fungi failed to colonize successful and competition emerged during early post treatment, leading to changes that lasted 90 days after treatment. The human mycobiome became more stochastic, leading to strikingly less co-abundance patterns among fungal species. These findings indicated that the gut mycobiome was not resilient enough to recover from the influence of antibiotics within 3 months.

Changes in functional metagenomic diversity from antibiotics are not strictly followed by changes in metatranscriptomic diversity

We characterized the subjects' microbiomes at baseline (Suppl. Fig. 4) and found that bacterial communities were dominated by bacteria from the *Bacteroidetes* and *Firmicutes* phyla but with strong variation in ratio, as expected in healthy individuals [16, 24]. In line with previous studies, we observed a significant decrease in bacterial species alpha diversity (Suppl. Fig. 5). Ciprofloxacin had the strongest (− 40%) and cefuroxime the weakest negative effect (− 5%), whereas controls only an insignificant increase (2%) (Suppl. Fig. 6). Beta diversity was significantly different during antibiotic treatment, but not in controls. In addition, we found that antibiotic treatment had the strongest influence on moderately abundant bacterial species (Suppl. Fig 5). We then estimated bacterial growth using GRiD [25] (Suppl. Table 4). In antibiotic-treated subjects, median growth of species decreased significantly during treatment compared to baseline as expected ($p = 0.009$, $r = − 0.56$, Suppl. Fig. 7). Interestingly, the number of species with growth rate greater 1 increased significantly ($p = 0.0016$, $r = 0.68$).

We subsequently investigated functional changes based on bacterial gene family abundance in the metatranscriptome and the metagenome. Alpha diversity of relative DNA gene family abundance was significantly reduced during treatment compared to baseline (Fig. 2a; $q < 0.05$), but not in controls (Suppl. Fig. 8). Despite the changes at the DNA level, the alpha diversity for relative RNA gene family abundance did not change significantly between time points (Fig. 2b; $q > 0.05$).

We investigated differences in beta diversity of gene family abundances based on Bray-Curtis dissimilarity (Fig. 2c, d). We performed ordination and statistical testing using distance-based redundancy analysis (dbRDA) using "subject id" as a constrained variable and "sample time" as an independent variable. In treated subjects, both DNA and RNA functional abundances showed significant differences in centroids between timepoints (Fig. 2c, d; DNA: $F = 2$, $p = 0.002$, Df = 5; RNA: $F = 1.6$, $p = 0.037$, Df = 5). This was not observed in controls (Suppl. Fig. 9). Pairwise dbRDA revealed a significant difference from baseline to treatment in DNA functional abundance ($F = 3.13$, $q = 0.014$, Df = 1; Suppl. Table 5). For RNA abundance, we observed only a trend ($F = 1.6$, $p = 0.085$, Df = 1). Overall,

Fig. 2 Metagenomic contributional alpha diversity of metabolic function is severely reduced by antibiotic treatment. Diversity analysis of metagenomic and metatranscriptomic samples from participants using HUMAnN2 relative abundances. **a**, **b** Alpha diversity of gene family relative abundances using **a** metagenomic and **b** metatranscriptomic data. Boxplots show species richness (left), Shannon (middle), and Gini-Simpson indexes (right) at 15 days before treatment (baseline), during (DT), and 30 days (EPT) and 90 days post treatment (LPT). Median (centerlines), first and third quartiles (box limits), and 1.5× interquartile range (whiskers) are shown. Lines between boxes connect same-donor samples. Statistical testing was by Wilcoxon signed-rank test with p values adjusted for multiple testing using false discovery rate (q; *0.01 ≤ p < 0.05). **c**, **d** Constrained ordination of Bray-Curtis dissimilarity based on gene family abundances measured using principle coordinate analysis (PCoA). We used distance-based redundancy analysis to show the explained variance by sample time points while accounting for participant-specific influence. **e**–**g** Contributional Shannon diversity of MetaCyc pathways of baseline and treatment samples. **e** Top, metagenomic and bottom, metatranscriptomic contributions. Mean (solid lines) and first and third quartiles (transparent ribbons) are shown. **f** Mean contributional diversity per participant per time point for DNA (top) and RNA (bottom). **g** Species with significantly increased (blue, +) and decreased (red, −) contribution to pathways based on two-sided Wilcoxon signed-rank test adjusted for compositionality (q < 0.1)

these findings implied that the genetic potential of the bacterial community was reduced during treatment as expected. However, gene expression changes were considerably less compared to the metagenome and not as consistent among participants. Similarly, antibiotic treatment had no significant effect on the transcriptional activity of the core and variable metabolic pathways (as defined in [20]; Suppl. Fig. 10). In agreement with previous findings [20], the metatranscriptome was much more dynamic than the metagenome.

Diversity of bacterial contribution to metabolic pathways is systematically reduced by antibiotics

We investigated if the contribution of species to a given pathway changed significantly over time [26] (Fig. 2g, h; Suppl. Fig. 11). By DNA, the median contributional alpha diversity of antibiotic-treated participants decreased significantly from baseline to treatment (Shannon: $\log_2$ fold-change [lf2] = − 0.4; q = 0.015; Simpson: lf2 = − 0.24; q = 0.015). Controls showed no significant changes ($q > 0.05$). In contrast, we did not observe significant changes in alpha diversity measures for RNA ($q > 0.05$). We further investigated if the contribution of single bacterial species to metabolic pathways changed significantly between time points. We implemented a compositionality test as described in Palleja et al. [27], considering all pathways, and found 9 bacterial species whose contribution significantly changed ($q < 0.1$; Fig. 2i). Important gut commensal bacteria including *Akkermansia muciniphila*, *Faecalibacterium prausnitzii*, *Odoribacter splanchnicus*, and *Bifidobacterium adolescentis* contributed less during treatment. A decline of such butyrate-producing species following antibiotic treatment has been observed before [11, 27]. In contrast, the multiantibiotic-resistant bacterium *Clostridium bolteae* [28] contributed more.

Antibiotic treatment lastingly reduced bacterial-fungi interactions

We increased the functional resolution of bacterial species using the metagenomic species (MGS) concept [29], which allows identification of taxonomically unidentified bacterial species. We further improved the method to identify some bacteria at the strain level based on their genetic potential. In contrast to previous studies, we used HUMAnN2 [19] gene family profiles as references in accordance with a published protocol [30]. HUMAnN2-derived profiles allowed us to retrieve MGS with high purity (i.e. more than 95% of genes in an MGS group originated from the same species; Suppl. Table 6). We then identified 26 MGS with significant change in relative abundance during treatment compared to baseline (Suppl. Fig. 12; Suppl. Table 7), which was not observed in controls. Six of these had species-level annotation and were consistently decreased independent of the antibiotic drug used (*Ruminococcus lactaris*, *Dialister invisus*, *Odoribacter splanchnicus*, *Bacteroidetes bacterium ph8*, *Akkermansia muciniphila*, *Bifidobacterium adolescentis*; full list in Suppl. Figs. 13 and 14).

We combined MGS and ITS relative abundance data and used BAnOCC [31] to infer intra- and cross-kingdom associations among bacterial and fungal species. We created co-abundance networks at the species level for baseline and for during, early post and late post treatment periods independently as described above (Suppl. Fig. 15). In order to find significant changes in the structure of co-abundance networks, we compared differences in node degree. The degree of a node is defined by the number of significant correlations with that node. Hence, an increase in node degree implies an increase of potential interactions, i.e. an increase of potentially relevant effects. To study changes in bacterial-fungal interactions, we tested for significant differences in node degree centrality considering only cross-kingdom correlation (Fig. 3a, c; Suppl. Table 8). We observed a temporal increase in node degree during treatment compared to baseline (q = 0.055). From during to early post treatment, this degree dropped (q = 0.0185) and decreased further at late post treatment (q = 0.0185). To find lasting changes, we compared baseline against late post treatment and found significantly reduced degree (q = 0.00134). Considering these results in addition to the loss of correlations observed in the fungal network, we conclude that antibiotic treatment was a triggering event for disturbances in bacterial-fungal interactions. These disturbances ultimately drove gut bacteria and fungi towards independence.

We looked more closely at co-abundance patterns involving bacterial species with significant changes in abundance or pathway contribution during treatment (Fig. 3c; all significant correlation in Suppl. Table 9). *C. bolteae* increased in relative abundance during treatment and correlated positively with many fungal species during treatment, such as the opportunistic pathogen *C. albicans*, or the mycotoxin producers *Aspergillus penicillioides* and *Penicillium glandicola*. In contrast, *O. splanchnicus* was persistently negatively correlated with opportunistic pathogens from the genera *Candida*, *Aspergillus*, and *Alterna*. *O. splanchnicus* is part of the healthy gut community but rarely investigated in terms of its role. *Roseburia inulinivorans* was negatively associated to opportunistic pathogens *C. albicans*, *C. sake*, and *P. glandicola*. Low *Roseburia* abundance was associated with higher glucose levels and ulcerative colitis [32, 33]. Likewise, *Eubacterium rectale* was negatively associated with *C. albicans* and *P. glandicola*. Depending on the diet, *Eubacterium rectale* decreased glucose and insulin levels [34]. Notably, butyrate-producing species were negatively associated with at least one opportunistic fungal pathogen.

At last, we considered bacterial-fungal correlations together with MGS abundance changes during treatment and fungal abundance changes in early post treatment. Bacterial species with decreased relative abundance and negative correlation to a fungus that showed an increased abundance were considered competitors. For example, *O. splanchnicus* was decreased during treatment, and showed negative correlation to *C. albicans*. A list of possible bacterial-fungal competitors is shown in Suppl. Table 10.
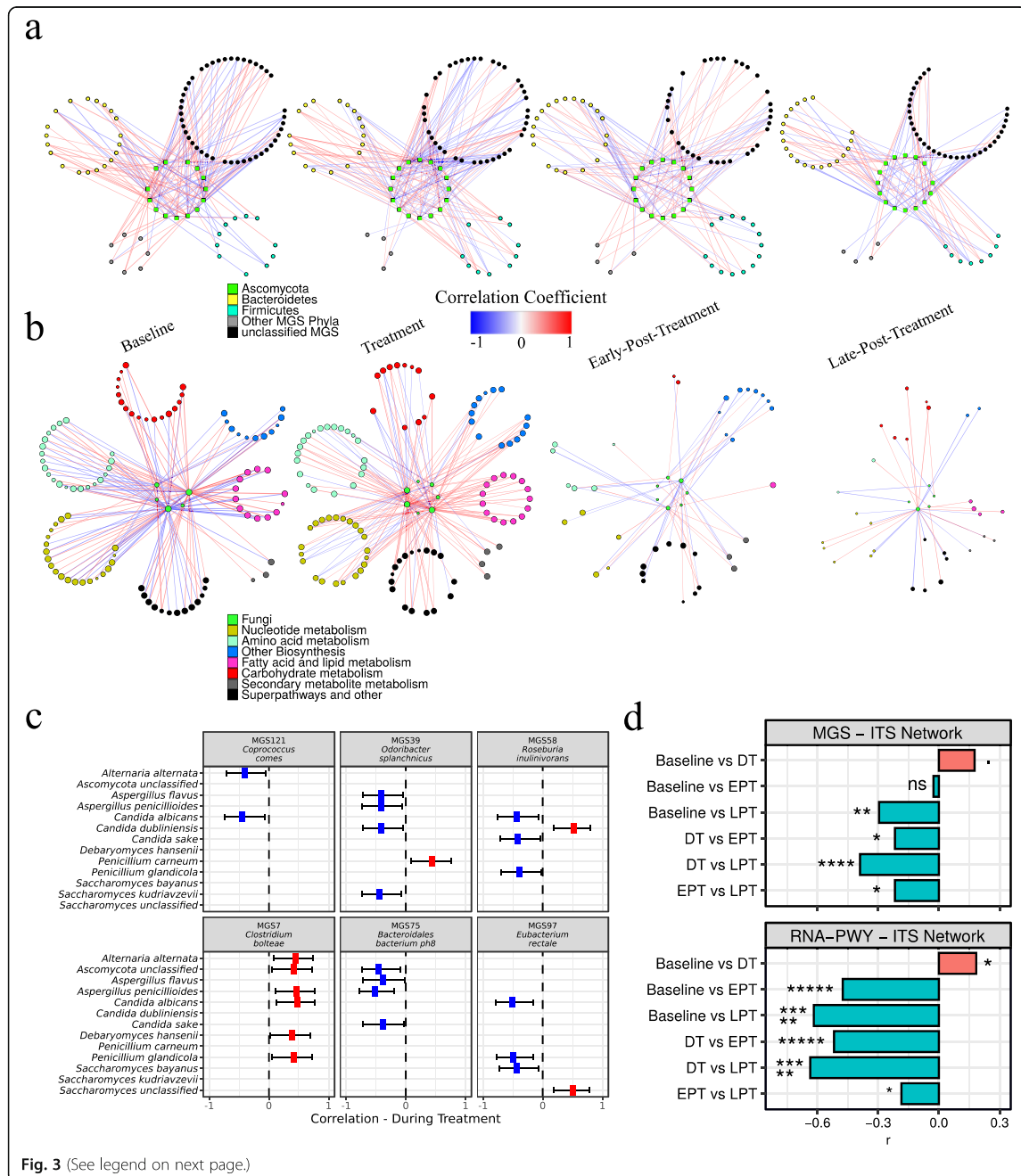
Fig. 3 (See legend on next page.)

Manuscript III

Seelbinder *et al. Microbiome*   (2020) 8:133   Page 8 of 20

(See figure on previous page.)
Fig. 3 Cross-kingdom interactions among fungi, bacteria species, and pathway expression. **a**, **b** Co-abundance networks at indicated time points using BAnOCC with **a** 25% and **b** 50% prevalence filter. Only significant edges are shown (based on 95% credibility interval) with $|r| \geq 0.3$. Negative correlations (blue), positive correlations (red). Networks are left (baseline) to right (late post treatment). **a** Correlations among fungal and bacterial species based on metagenomic species (MGS) and internal transcribed spacer (ITS) relative abundance. **b** Correlations among fungal species and pathway expression based on HUMAnN2 RNA pathway and internal transcribed sequence relative abundance. Superpathways and other pathways that did not fit into the six major categories were grouped as "other". **c** Estimated correlation between bacterial and fungal species during treatment. Positive (red), negative (blue). Error lines show 95% confidence intervals. **d** Effect size of node degree change. *r* values change from − 1 (100% decrease) to 1 (100% increase). (Top) MGS and ITS relative abundances. (Bottom) RNA-PWY and ITS relative abundances. Statistical testing for significant changes in node degree was performed using a two-sided Wilcox signed-rank test. *P* values were adjusted for multiple testing using FDR. Node degree was determined independently for baseline, during (DT), early post (EPT), and late post treatment (LPT). Significance is indicated by symbols (ns, $q \geq 0.05$; *$q < 0.05$; **$q < 0.01$; ***$q < 1e- 3$; ****$q < 1e- 4$; *****$q < 1e- 5$)

Prevalent fungi correlated with pathway expression during treatment

We investigated relationships among metabolic pathway expression levels (MetaCyc database—PWY; metatranscriptomic abundance) and fungal ITS abundance (Fig. 3b) by creating co-abundance networks analogous to the bacteria-fungi network. We tested for significant changes in node degree considering only correlations between fungal OTUs and pathway expression. From baseline to treatment, node degree increased ($q = 0.0185$; Fig. 3d). Most correlations were positive (146 of 189) during treatment in contrast to baseline (82 of 153). Hence, and despite the increase of variance of metatranscriptome diversity during treatment, we still observed co-abundance with fungal species during treatment. This observation suggested a mutual influence between the fungal community and expression of bacterial metabolic pathways. About one third of correlations at baseline involved *C. albicans* and one third involved *Saccharomyces*. After treatment, node degree dropped significantly to below baseline levels (during vs. early post: $q = 4e- 11$; early vs. late post: $q = 0.0185$; baseline vs. late post: $q = 3e- 15$). Almost all *C. albicans* co-abundance patterns were lost at 90 days post treatment, with *Saccharomyces* genus accounting for over 70% of remaining correlations (24 of 31). Overall, *Saccharomyces* appears to be more resilient with respect to bacterial metabolic pathways expression than other prevalent fungi.

We then increased our resolution by focusing on correlations between fungal OTU abundance and pathways in broader functional categories (Suppl. Fig. 16; Suppl. Table 11). We observed a significant increase in node degree from baseline to treatment for pathway functions in nucleotide metabolism ($q = 0.026$) and biosynthetic pathways (e.g. for vitamins, tetrapyrroles, NAD) ($q = 0.043$). Almost all correlations were positive. We found no significant changes in remaining categories ($q < 0.1$; metabolism of amino acids, carbohydrates, fatty acids and lipids, secondary metabolites). Node degree dropped significantly after treatment in all categories except secondary metabolite metabolism ($q < 0.05$).

Since our treatment targeted bacteria, we expected the bacterial community to respond to the selective pressure with strong, directed changes in pathway expression. Most metatranscriptomic changes appeared to be stochastic. Yet we still observed mostly positive co-abundance patterns between fungal abundance and bacterial pathway expression during treatment, especially with functions required for bacterial growth. Even though correlations do not imply causations, when performed on multiple different levels, they can still offer significant insights. The results suggested a common origin for changes in the mycobiome and pathway expression: if gut fungi generally take advantage of reduced complexity in the bacterial community, we would expect an increase in fungal diversity. However, we observed no systematic change. Highly abundant and adapted fungi may still overgrow. In both scenarios, we would expect an increase in negative correlations between fungal and bacterial abundances during treatment but found mostly positive correlations. Generally, antibiotics drove the mycobiome alongside the microbiome, leading to a temporal increase in fungal richness, but also increased fungal competition subsequently. On the long run, antibiotic treatment broke down most of the inferred relationships between bacteria and fungi, as shown by diverging mycobiomes 3 months after treatment.

Key bacterial species and molecular mediators of *Candida albicans* colonization

Our ITS data showed at least one *C. albicans* read per participant over 112 days but with varying relative abundance from 0 to 42%. *C. albicans* was detected during treatment even if it was not detected at baseline, as in other studies [12, 35], confirming antibiotic treatment as risk factor for colonization and overgrowth of this fungus. Furthermore, *C. albicans* significantly decreased 2 weeks after treatment, implying the indirect impact on its growth by the microbiome. We searched for metabolites detected in the human gut that may inhibit or promote *C. albicans* growth. We performed metabolomics analyses on a subset of 15 stool samples and calculated Spearman's correlations for the relative abundance of

each metabolite and relative abundance of *C. albicans* by ITS (Suppl. Fig. 17). Based on these findings, we performed *C. albicans* growth assays in defined medium containing serial dilutions of selected metabolites. With several metabolites, including 4-hydroxybenzoic acid and 8,11,14-eicosatrienoic acid, we observed only minor growth reductions at the highest concentrations (Suppl. Fig. 18). More pronounced growth reduction occurred with adipic acid, aminoadipic acid, and ornithine, but fungi still grew with high concentrations of these metabolites. In contrast, propionic acid, acetic acid, and cis-5-dodecenoic acid fully inhibited growth at a range of concentrations. We then tested if the substances directly damaged human cells. Using a human vaginal cell line (A431) without *C. albicans*, the bile acid lithocholate (LCA) and cis-5-dodecenoic acid showed limited cytotoxicity (Suppl. Fig. 19). No other substances caused detectable cell damage.

Next, we assayed the effect on human cells by *C. albicans* in presence of the same metabolites. At higher concentrations, when in vitro fungal growth was reduced, human cell damage decreased with the short-chain fatty acids (SCFAs) propionic ($p < 0.05$) and acetic acid ($p < 0.001$). Acetic or cis-5-docenoic acid ($p < 0.01$) almost fully abolished cell damage by *C. albicans*. Benzoic acid reduced damage to a lesser extent ($p = 0.051$). Since the morphology (yeast or hyphal cells) is critical for its damaging potential, we investigated if the substances also led to morphological changes in *C. albicans* (Fig. 4e). On high concentrations, hyphae formation and growth were almost completely suppressed by 5-dodecenoic and acetic acid. 5-dodecenoate also reduced hyphae formation under growth-permitting concentration. Glutathione only allowed formation of chains elongated yeasts resembling pseudohyphae. LCA partially suppressed hyphal growth at the high concentration, resulting in high numbers of pseudohyphae and yeast cells.

These metabolites that affect *C. albicans* growth negatively may also promote the growth of its fungal competitors, such as *Saccharomyces* spp., *Penicillium* spp. and *Aspergillus* spp. Therefore, we repeated the correlation analysis with the corresponding OTUs (Suppl. Table 12). For each fungal species, we found several metabolites with positive correlation. Considering metabolites negatively affecting *C. albicans*, only 2-methyl butanoic acid and 3-hydroxy butyric acid were found to be significantly positively correlated with *Penicillium spinulosum* and LCA with *Aspergillus flavus*. Still, promotive effects on other fungal species need to be verified in future work.

Bacterial supernatant inhibits C. albicans growth

We investigated which gut bacteria might be the main direct or indirect producers or contributors of these compounds in our human participants. We correlated metabolite concentrations with MGS relative abundances (Suppl. Fig. 20) and focused on positive associations. We looked at species that correlated with multiple, different metabolites. *Bacteroides coprophilus* correlated with aminoadipic acid and acetate; *C. comes* with 4-hydroxybenzoic acid, 5-dodecenoate, and glutathione; *F. prausnitzii* with 4-hydroxybenzoic acid; *E. lenta* with 5-dodecenoate and eicosatrienoic acid; *B. eggerthii* with 5-dodecenoate and eicosatrienoic acid; and *O. splanchnicus* with acetate. All six species correlated with LCA or its derivates.

Our correlation methods helped us to pinpoint bacteria that may promote or inhibit *C. albicans* growth (Fig. 4a). For testing these predicted associations in vitro, we selected bacterial strains based on sufficient confidence in our strain-level inference in addition to significant correlation to *C. albicans*. We performed the strain identification directly from the MGS profiling. Instead of strain detection methods using single-nucleotide polymorphisms (e.g. StrainPhlAn [18], metaSNV [36], ConStrains [37]), we adopted a strategy based on gene content as in PanPhlAn [38]. We therefore analysed reads corresponding to a specific MGS. For example, gene coverage for *O. splanchnicus* strains for two participants (N, E; Fig. 4b; Suppl. Fig. 21) showed that both subjects had the highest coverage for strain DSM 20712, so we selected DSM 20712 for in vitro assays. In the end, we were interested in bacterial strains for which we found significant correlation with inhibitory metabolites, significant correlation with *C. albicans*, and high confidence from the strain inference. Based on these results, we selected *Bacteroides eggerthii* and *Odoribacter splanchnicus* for further in vitro experiments.

We determined the antifungal effect of metabolites produced by selected bacterial strains using their sterilized culture supernatants as growth medium for *C. albicans*. We measured *C. albicans* growth using 100% or 50% supernatant diluted in modified Gifu anaerobic media (mGAM) (Fig. 4c). Percentage inhibition was compared to optimal growth conditions in fresh medium. *C. albicans* growth was significantly inhibited by supernatants from *B. eggerthii* (50% growth) or *O. splanchnicus* (40%). Using 100% bacterial supernatants had stronger effects, showing that inhibition was proportional to supernatant dilution (Suppl. Fig. 22). We tested two additional *C. albicans* strains to exclude that observed effects were strain specific but saw no differences (Suppl. Fig. 22). *B. eggerthii* and *O. splanchnicus* also inhibited *C. albicans* growth in pairwise in vitro coculturing experiments (Suppl. Fig. 23).

Finally, we analysed the supernatant of these species to characterize their metabolic capabilities that may relate to *C. albicans* growth (Suppl. Fig. 24-25). We included the supernatant from *Ruminococcus* [Blautia] *torques* as positive control, since this species was shown to have

Fig. 4 *Candida albicans* growth promotors and inhibitors. **a** Bacterial species co-abundant with *C. albicans*. Line colours and type indicate correlation coefficients. **b** Example using *Odoribacter splanchnicus* for genomic strain inference from metagenomic species (MGS) reads. Strains were inferred for each time point (*x*-axis) from number of genes with 0.5 reads per base (*y*-axis) per-reference genome. Data are from the antibiotic-treated participant N and untreated subject E. The number of genes coverage for each tested *O. splanchnicus* strains are shown. DSM 20712 was identical with another strain labelled NCTC10825. **c** Growth rate inhibition of *C. albicans* strain SC5314 cultivated with 50% and 100% bacterial supernatant (from *Bacteroides eggerthii* and *Odoribacter splanchnicus*) compared to control of medium (mGAM) only. **d** Damage of human vaginal epithelial cells (A431) based on release of lactate dehydrogenase (LDH) with metabolites at inhibitory concentrations. Grey lines, zero effect. Positive values imply cell damage. (Top) Human cells cultured without *C. albicans*. Effect compared to untreated cells. (Bottom) Cells co-cultured with *C. albicans*. Values are relative to damage caused by *C. albicans* without additional metabolites. Negative values imply less cell damage. **e** Composition of morphology of *C. albicans* cultures quantified by concentration of morphological types. Some metabolites cause atypical formation of hyphae–like structures ("Pseudohyphae/elongated yeast chains"). Some inhibited the formation of filaments or disrupted growth in general ("Yeast/no growth")

positive effect on *C. albicans* growth previously (Mirhak-kak et al., 2020, under review) and correlated positively with *C. albicans* in our study. Compared to quality control samples, *O. splanchnicus* supernatant contained high concentrations of butyric acid (7-fold relative conc.),

which suppresses *C. albicans* growth in vitro [39]. But we also measured elevated levels of the growth suppressing metabolites adipic and aminoadipic acid, and ornithine. In contrast, *B. eggerthii* supernatant contained elevated levels of acetic acid (1.6-fold), formic acid (3-

fold), and hexanoic acid (2-fold). The full growth-inhibiting metabolites 2-methyl-propanoic acid and propanoic acid were also found in supernatants of *O. splanchnicus* and *B. eggerthii*, but roughly 3-times higher in *B. eggerthii*. In contrast, *R. torques* produced only formic acid in higher abundance (1.25-fold), but almost none of the strong inhibitory SCFA. Altogether, the supernatant analysis shows that propionate, ornithine, and benzoic acid are effective inhibitors of *C. albicans* growth, and these compounds were likely produced by *B. eggerthii* and *O. splanchnicus* also in the human gut.

Discussion

Mouse models can offer some advantages for studying competitive relationships between gut bacteria and fungi. Previous studies have shown that antibiotics induce fungal overgrowth in the murine gut lumen [12, 35, 40]. However, antibiotic doses used in mice experiments create an almost germ-free environment after treatment, which is unlikely to apply to the human gut with clinical use of antibiotics. Furthermore, the mice gut microbiome and human gut microbiome differ considerably [41–43]. For example, many *Firmicutes* spp., which represent major colonizers of the human gut, cannot efficiently colonize the murine gut. Sovran et al. showed that *Enterobacteriaceae* play an important role for bacteria-fungi interactions in the murine gut [44]. In their study, *Enterobacteriaceae* accounted for 40 to 65% of reads in Vancomycin treated mice. We investigated the relative abundance of *Enterobacteriaceae* spp. in our human subjects. However, the accumulated relative abundance of *Enterobacteriaceae* spp. for most samples was below 1% before, during and after treatment (median 0.02%; except for Augmentin with 13%), making it difficult to assess whether *Enterobacteriaceae* were relevant for bacterial-fungal interactions in the human gut (Suppl. Fig. 26).

In this study, we investigated if fungal overgrowth was induced in the human gastrointestinal tract under physiological conditions. We present evidence that changes on fungal abundance at the species level are highly dynamic in the lower human gastrointestinal tract. Even though gut bacteria and fungi successfully prevented several temporarily detected fungi from colonization the lumen lastingly, we found significant alternations to the relative abundance of several fungi even 90 days after antibiotic treatment.

We used 5 different broad-spectrum antibiotics which are commonly used to treat human diseases [45]. Recent work by Maier et al. [46] addresses the issue that most knowledge of antibiotic drugs and their bacterial targets is based on pathogens and not the commensal microbiome. In a large screening of 144 different antibiotics and the 40 most common gut microbial strains, most antibiotics inhibited growth of all tested bacterial strains. Only *Clostridium* showed consistent resistance to many drugs. Indeed, some bacterial species are stronger or less affected depending on the antibiotic used. We investigated how much these expected differences apply to our data. Effect sizes varied, but overall, most of the significant changes were independent of specific antibiotic drugs. Because of our small cohort size, we cannot assess if the differences in effect sizes are due to differences in baseline communities or differential inhibition of the drugs. More work is required by using bigger cohorts as well as other antibiotic drugs with narrower targets.

Co-abundance networks inferred mutual relationships between fungal species at baseline and during treatment. Post treatment, however, competition emerged. Furthermore, we observed far fewer co-abundance patterns between fungi and bacteria in early and late post treatment periods, indicating profound decline in bacterial-fungal interactions. Overall, we found the fungal community to be less resilient than the bacterial. Based on these data, we hypothesize that the dominant gut fungi of healthy individuals were in balance with gut bacteria. Antibiotic administration induced profound changes to gut bacteria that translated into changes in fungal abundance that lasted until the end of our study period. Indeed, these results must be considered with caution, as we did not perform quantitative estimations of bacterial and fungal abundances. In most cases, relative abundance estimation does not allow inference of true direction of change [47]. For quantification, bacterial cells are counted by flow cytometry in addition to DNA sequencing or qPCR [47, 48]. However, broad-spectrum antibiotics decrease bacterial cell counts by 3 orders of magnitude [47]. We estimated bacterial growth in situ to show that bacterial growth was significantly impaired at the community level. Hence, significant decrease in relative abundance of species will likely be reflected in true abundance as well. In future work, increasing the number of study subjects will help to increase certainty in and resolution of the findings.

One of the largest knowledge gaps about the basic biology of gut microbial balance is the lack of comprehensive functional analyses. Metatranscriptome studies have found both more [20] and less [49] uniformity in individual participants' profiles compared to respective metagenomes. Despite minor changes in beta diversity, we found no significant changes induced by antibiotic treatment in gene family alpha diversity, species contribution, or transcriptional activity of metabolic pathways. This result was most likely due to high variability in the metatranscriptome, as observed previously in healthy humans [20]. However, fungal abundance and bacterial growth may have influenced one another because mutual relations between fungal abundances and expression

of bacterial functions for growth were inferred, especially during treatment. Because these patterns were not as pronounced before and after treatment, we identified antibiotic administration as the main driver of this change.

Understanding and finding microbial mediators of fungal pathogens may help to improve antifungal treatments. We focused our study on *C. albicans*, testing in vitro if growth was affected by compounds produced by two bacterial species, *B. eggerthii* and *O. splanchnicus*. Although the supernatant of each bacterium was used in combination with optimal *C. albicans* growth medium, the supernatants inhibited *C. albicans* growth considerably. Such a condition is plausible for the lower human intestine, because we expect most easily metabolizable compounds, e.g. carbohydrates, to be absorbed by the small intestine. Furthermore, the two species may be physically separated in the gut lumen. Some of the metabolites with clear growth reduction to *C. albicans* were found in bacterial supernatants. However, we cannot exclude potential promoting effects of other bacteria that could occur in the same vicinity.

A decline in gut bile acids and SCFA is linked to disease states [50, 51], but cause-effect mechanisms are less understood. We identified five metabolites that naturally occur in the human gut to effectively inhibit growth and/or lower hyphae formation, a key attribute of *C. albicans* virulence [52]. Acetate is a prototypical SCFA that dampens the immune response at higher concentrations [53]. The SCFA propionate plays an important role in immune regulation [54]. Lithocholate is a secondary bile acid and such secondary bile acids may inhibit *C. albicans* growth [50]. Glutathione is an antioxidant that dampens cell damage [55]. Cis-5-dodecenoic acid suppressed hyphae formation entirely. A similar compound, cis-2-dodecenoic acid, is produced by *Burkholderia cenocepacia* and strongly interferes with *C. albicans* growth [56, 57]. In contrast to previous studies [39, 51] we also show that acetate, 5-dodecenoic acid, and propionate also significantly reduced *C. albicans*-mediated host cell damage in vitro. These compounds could also support the growth of fungal *C. albicans* competitors. However, a correlation analysis between these metabolites and multiple different *Saccharomyces*, *Penicillium*, and *Aspergillus* spp. did not indicate that. Nevertheless, this needs to be experimentally verified in future work.

Several limitations should be highlighted. Observing gut bacterial and fungi concomitantly is difficult as long as bacterial and fungal abundances are estimated using two independent sequencing technologies. Improvements in correlation methods mitigate some of the resulting problems. Still, our correlation results regarding inter-kingdom species-species correlation could be improved in the future. Estimating cell counts per kingdom would further help to improve correlation estimates. Our findings are further limited to just the 5 antibiotic drugs used. Even though many significant findings seemed consistent across the drugs, increasing the number of patients for each drug would help to get more differentiated results. When studying mechanistic effects with respect to *C. albicans* growth, we could not simulate the complexity of the gut community. We aimed to find metabolic regulators, but the growth of fungi and bacteria in the gut is certainly based on a combination of several metabolic factors and environmental conditions. We looked at a variety of aspects from host cell damage to morphology, but these were still in vitro findings.

Our results indicated that antibiotic treatment has a longer-lasting impact on gut fungi than bacteria, driving fungal communities from mutualism to competition. This work also advanced MGS methods for resolving microbiome compositions and interactions. Of potential clinical relevance, we demonstrate how particular SCFAs and bile acids produced by gut bacteria restricted human cell damage from *C. albicans* but also show other compounds with considerable effects.

Conclusions

Theoretically, bacteria and fungi compete for resources available on the gut lumen, but they may also support one another. In this study, we investigated the temporal, concomitant changes of gut bacteria and fungi in humans. We demonstrate that antibacterial drugs have long-term influence on the human gut mycobiome, driving fungal communities from mutualism to competition. We further show how metabolites produced by bacteria such as cis-5-dodecenoic acid may actively suppress pathogenicity of opportunistic fungi such as *C. albicans*. We thereby show that gut bacterial-fungal interactions are an important consideration for antibacterial treatment.

Methods

Study design

Human participants

Stool samples were gathered from 14 healthy adults, aged 18–65 years, from Denmark and Hong Kong. Samples were collected over 3–4 months. The Danish study was approved by the local ethics committee in Region Zealand, Denmark (SJ-383), and the Hong Kong study was approved by the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 17-042). All work was performed in accordance with the Good Clinical Practice principles and the Helsinki Declaration. Written informed consent was obtained from all participants. Patient characteristics are described in (Suppl. Table 13). Subjects with any of the exclusion criteria below were not eligible for entry

into the present study: (i) history of taking antibiotics over the last 6 months, (ii) receiving systemic antifungals/antifungal mouthwashes or probiotics concurrently, (iii) patients suffering from immunosuppressive conditions or taking immunosuppressants, and (iv) severe medical comorbidities requiring frequent hospitalization. Another cohort of six healthy, untreated individuals from Canada was acquired from a previous study from Raymond et al. [58].

Treatment
Of the participants, 12 were treated for 6 days with 1 antibiotic drug out of 5: doxycycline (tetracycline class), azithromycin (macrolide class), Augmentin (β-lactam class), ciprofloxacin (quinolone class), and cefuroxime (β-lactam class). Two untreated participants were used as controls.

Sampling
From each participant in the clinical study in Denmark, 6 stool samples were obtained: one 15 days before treatment (± 1 day), two during treatment (days 3 and 5 of treatment ± 1 day), and three at 15, 30, and 90 days after treatment (± 1 day). From each participant in the clinical study in Hong Kong, four stool samples were obtained at 7 days before treatment (± 1 day), day 6 of treatment, and 30 and 90 days after treatment. Collected samples were aliquoted and stored at − 80° immediately after collection until DNA extraction. Stool samples of control patients treated with placebo [58] were acquired before, 7 days, and 90 days after treatment.

Metagenomics and metatranscriptomics sequencing
For participants in the clinical study in Denmark, bacterial metagenomics and metatranscriptomics raw data were obtained from Kang et al. (*in preparation*). Briefly, DNA was extracted using a MO BIO PowerMax Soil DNA Extraction Kit (MO BIO Laboratories, Inc) and purified with PowerClean Pro DNA Clean-Up Kits (MO BIO Laboratories, Inc.) according to the manufacturer's protocol. For RNA, rRNA was depleted using a Ribo-Zero Gold rRNA removal kit—Epidemiology (Illumina). The remaining total RNA was extracted using a MO BIO PowerMicrobiome™ RNA Isolation Kit (MO BIO Laboratories, Inc.). RNA and DNA sequencing were performed on an Illumina HiSeq 2000 (PE125). For participants in the clinical study in Hong Kong, bacterial DNA and RNA were extracted from 200 mg aliquots of frozen stool by Beijing Genome Institute (BGI). DNA was extracted using an E.Z.N.A.® Stool DNA Kit according to the manufacturer's protocol. For RNA, rRNA was depleted using a Ribo-Zero™ Magnetic Kit. The remaining total RNA was extracted using a RiboPure-Yeast Kit. All samples were sequenced on an Illumina HiSeq 4000

platform (Illumina, San Diego, California, USA; paired-end, insert size 350 bp, read length 150 bp for DNA and 100 bp for RNA).

Internal transcribed spacer sequencing
All stool samples from both cohorts were processed by Novogene for internal transcribed spacer (ITS) sequencing. 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 micro-centrifuge 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). The resulting fungal sequences were amplified using ITS2-F: 5′ GCATCGATGAAGAACGCAGC-3′ and ITS2-R: 5′ TCCTCCGCTTATTGATATGC-3′ primers [59, 60]. ITS2 amplicons were generated in three steps by PCR with 38 cycles: 98 °C 10s, 59 °C 10s, and 72 °C 30s followed by sequencing on the Illumina HiSeq platform (2 × 250 bp, Novogen, China).

Metabolomics
For 4 participants, bile acid profiles and MicrobioMET profiles were assessed by Metabo-Profile (Shanghai, China) using aliquots of frozen stool. For bile acid profiles, bile acid-free matrix (BAFM) was obtained using the charcoal-stripping protocol. Calibrators and quality controls were prepared for the BAFM and processed as for extraction of bile acids from stool samples. About 10 mg prechilled zirconium oxide beads were added to 10 mg stool with 15 μl ultrapure water. To each sample, a

200-μl aliquot of prechilled acetonitrile/methanol containing 10 internal standards was added for homogenization. After centrifugation at 13,500 rpm and 4 °C for 20 min, 50 μl supernatant was transferred to 96-well plates. Acetonitrile/water (150 μl) was added for gentle shaking for 5 min before injection into an ultra-performance liquid chromatography column coupled to tandem mass spectrometry (UPLC-MS/MS) system to quantitate bile acids.

MicrobioMET profiles including aromatic phenols and indoles, phenolic acids, short-chain fatty acids and branched-chain amino acids, amino acids, and organic acids were quantitated using gas chromatography coupled to time-of-flight mass spectrometer (GC-TOFMS). Stool aliquots (50 mg) were homogenized with 300 μl NaOH (1 M) solution using a homogenizer and centrifuged at 13,500 rpm and 4 °C for 20 min. Supernatants (200 μl) were transferred into autosampler vials and residue extracted with 200 μl cold methanol. After a second homogenization and centrifugation, 167 μl supernatant was combined with the first supernatant in the autosampler vial. Extracts were capped and used for automated sample derivatization by a robotic multipurpose sample MPS2 with dual heads (Gerstel, Muehlheim, Germany). Samples pre-treated with sodium sulfate were shaken at 1500 rpm and 4 °C for 20 min and transferred to capped empty autosampler vials for the GC-TOFMS.

Data processing

Quality control of sequence data

Quality control of raw reads (DNA, RNA) used a previously described pipeline [61]. Adapter sequences, low-quality bases ($Q < 20$), duplicated reads, reads shorter than 75 bp and reads mapping to the human genome with 95% coverage were filtered out. Computational scripts are at https://github.com/TingtZHENG/VirMiner/.

In situ bacterial growth rate estimation

Quality controlled FASTQ samples were sub-sampled to 2 million reads per sample. GRiD version 1.3 [25] was used with the corresponding stool database on sub-sampled samples to assess the growth bacterial strains. Default parameters were used but with minimum coverage threshold of 0.5 in order to investigate growth rates for different thresholds. After investigating the results, and as suggested by the GRiD authors, we continued with the growth estimates for strains with coverage 1.0 or higher. Statistical testing of (a) median growth rates and (b) the number of growing species was performed with a Wilcoxon signed-rank test. Normalized effect size r was estimated using the R package "rcompanion" and its function "wilcoxonPairedR".

Abundance profiling

HUMAnN2 [19] version 0.11.1 was used to estimate gene family abundances in metagenomic DNA and RNA samples. Resulting reads per kilo-base (RPK) for gene family abundances at species level (including unclassified taxa) were further normalized by counts per million (CPM), resulting in a transcripts per kilo-base million (TPKM) like normalization.

PIPITS pipeline [21] version 1.4.5 was used for ITS with default parameters including quality filtering, read-pair merging, ITS2 filtering, and chimaera removal. Remaining reads were binned based on 97% similarity as operational taxonomic unit and aligned to the UNITE fungi database using Mothur classifier [22]. For further downstream analysis, all samples were normalized by cumulative sum scaling using MetagenomeSeq [62].

For bile acid profiles, raw data from UPLC-MS/MS were processed using QuanMET software (v1.0, Metabo-Profile) for peak integration, calibration and quantitation for each bile acid. The analyte concentration of unknown bile acid was calculated using a calibration curve.

For MicrobioMET profiles, raw data from the GC-TOFMS were processed using proprietary software XploreMET (v2.0, Metabo-Profile) for automatic baseline denosing, smoothing, peak picking, and peak signal alignment. MS-based quantitative metabolomics determined the concentration of unknown metabolites by comparing the unknown to a calibration curve. Abundance of MirobioMET profiles was calculated to minimize large individual variations in metabolites.

Metagenomic sequences from HUMAnN2 profiles

TPKM-normalized gene family abundances from DNA were clustered using mgs-canopy version 1.0 software (https://anaconda.org/bioconda/mgs-canopy). We used standard parameters except for a Pearson correlation coefficient cut-off of 0.95 instead of the default 0.9. Gene family clusters with at least 700 genes were considered metagenomic sequences (MGS). Taxonomic annotation of MGS used species annotation information available for each gene family. We calculated contributions of each species to an MGS (including unclassified taxa). An MGS was annotated to species level using the largest gene family distribution if the gene contribution of that species was at least 51% and the second largest species (a) was "unclassified" or (b) contributed at most 10%. MGS with more than 90% gene contribution from the same species were considered "pure" or "unambiguous". Using a more stringent species assignment than the original method [29], from a total of 213 MGS, we obtained 80 with species-level assignment (Suppl. Table 6).

Genomic strains from MGS

MGS with species assignments were processed independently. Reads that (1) contributed to the abundance of an MGS, and (2) mapped to the inferred species (based on ChocoPhlAn reference [19]) were extracted. We used PanPhlAn [38] version 1.2.1.3 to create species-specific pangenomes based on reference genomes from the National Center for Biotechnology Information (Suppl. Table 14), mapped reads against the species pangenome, and calculated per-gene per-reference profiles. Gene abundance was normalized to reads per base. A gene was covered sufficiently if it had at least 0.5 reads per base. We accepted a strain reference if: (1) at least 90% of genes in the MGS were found to have sufficient coverage, and (2) the reference had the highest number of covered genes. For experimental verification, we considered using a commercially available strain if the number of covered genes was at most 1% less than the best-fitting strain.

Diversity analysis

Diversity analysis of fungal and bacterial communities was performed in R version 3.6.1 using the package vegan [63] version 2.5-5. Testing for significant differences in alpha diversity between time points was performed using a two-sided Wilcoxon signed-rank test. Resulting p values were adjusted for multiple testing using FDR. UniFrac metrics measured beta diversity by accounting for phylogenetic similarities of different species. Weighted UniFrac gives the most importance to dominant species. Unweighted UniFrac does not consider abundance. Generalized UniFrac with $\alpha = 50\%$ gives the most weight to moderately abundant species [64] and the generalized UniFrac with $\alpha = 75\%$ to species with abundance between median and dominant levels.

Transcriptional activity

Relative abundances using DNA and RNA were normalized to transcripts per million. Let f denote a gene or pathway. The transcriptional activity of f is defined as the TPKM-normalized RNA abundance of f divided by the TPKM-normalized DNA abundance of f.

Core metatranscriptome

The core metatranscriptome was described in [20]. Briefly, we used MetaCyc pathway relative abundances as generated by HUMANn2 for both DNA and RNA. We calculated transcriptional activity for each pathway. The core metatranscriptome was defined as the set of pathways with a sample prevalence of at least 80% with variable metatranscriptome having prevalence of 30 to 80%. Pathways with less than 30% prevalence were ignored.

Contributional alpha diversity

We followed the procedure in [26], with some exceptions. For each MetaCyc pathway (PWY), the contribution of species to the pathway was determined. Ecological alpha diversity measures (Shannon and Simpson) were applied per sample and separately using DNA and RNA data. Mean diversity per sample was used to test for significant differences between time points using pairwise two-sided signed Wilcoxon tests. Resulting p values were corrected for multiple testing using false discovery rate (FDR).

Statistics for MGS and ITS abundance

We used MetagenomeSeq [62] version 1.22.0 with a zero-inflated Gaussian mixture model. Following the MetagenomeSeq vignette, CSS normalization was applied on relative abundance data. All possible pairwise tests between the different sampling time points were performed (baseline, DT, EPT, and LPT time points). We controlled for patient-wise differences when possible. For MGS, D2 and D4 were excluded to improve signal quality. A 15% prevalence filter was used for each test independently. Controlling for multiple testing was performed on p values using FDR.

Two-way PERMANOVA testing

Stool samples from the same participant were statistically dependent. To test for significant differences in means of beta diversity between different time points, two-way permutational analyses of variance (PERMANOVA) were performed using "subject id" as covariate and "sample time point" as second independent variable. We performed tests on beta diversity matrices using the function "adonis" as implemented in R package vegan with 9999 permutations. We reported F values, R^2, and p values for "sample time point". P values from pairwise PERMANOVA tests were corrected for multiple testing using FDR.

Compositionality tests

We implemented a compositionality test from Palleja et al. [27]. Briefly, we used the test to address the issue of false-positive and false-negative findings in compositional data [65]. We accepted significant findings for a species based on relative abundance only if they would still be significant if other species were removed from the abundance table. Therefore, if one species was removed, the data were total-sum normalized and p values calculated. The procedure was repeated for all species. The final p value for a species was determined using the highest calculated p value. Thus, a species could not become significant because of depletion or inflation of another dominant species. Since this test was very

conservative, we used a higher q value of 0.1 to decide significance to avoid overlooking potential findings.

Correlation analyses using stool metabolite abundance

To identify metabolites with a potential effect on *Candida*, *Saccharomyces*, *Penicillium*, and *Aspergillus* spp., we calculated Spearman's correlations for total-sum scaling (TSS) ITS abundance and both bile acid and MicroMET profiles. To account for zero-inflation, we considered only samples with nonzero abundance of *Candida albicans* (5 samples). We then considered all significant correlations ($p < 0.05$) with an absolute correlation of at least 60%.

To identify direct or indict bacterial producers of the metabolites, total-sum scaled MGS abundances were correlated with log2 transformed metabolites abundances. Correlation was inferred using sparse partial least squared analysis (sPLS) by utilizing relevance vectors (R package mixOmics [30]).

Co-abundance networks

Co-abundance networks were created based on total-sum-normalized data using BAnOCC [31]. Significance of an edge was determined as described [20]. For posterior inference, we used the 95% credible interval. An edge was therefore considered significant if the corresponding 95% credible interval did not contain zero. Only significant correlations with an absolute estimated coefficient of at least 30% were used for analysis. Significant changes in network structure between any two time points were determined using Wilcoxon signed-rank tests on node degree. Effect sizes are reported in terms of a standardized effect size analogous to the one used for the Mann-Whitney test, $r = z/\sqrt{n}$, where z is the z-statistic of the paired test and n is the number of observations. r values are analogous to Pearson correlation coefficients. Hence, r ranges from -1 (100% decrease) to 1 (100% increase). Formula and implementation can be found in the R package "rcompanion".

Fungal species co-abundance network

TSS-normalized operational taxonomic unit (OTU) abundances based on ITS2 data were used. OTUs detected in less than 10% of samples were removed. BAnOCC was executed with 5 chains, 5000 iterations, and 1000 warmup cycles to reach convergence. BAnOCC was used as described above.

MGS-ITS network with BAnOCC

MGS and ITS relative abundances were independently total-sum normalized. Only species measured in 25% of samples were used further. Abundances of less prevalent species were summed per sample into a group called "other" to maintain library sizes. MGS and ITS features

abundances were combined and analysed using BAnOCC as described above.

RNA-PWY-ITS network with BAnOCC

RNA abundances of PWY and ITS were independently total-sum normalized. A 50% samples prevalence filter was applied to make this computation feasible and decrease false-positive rate. Abundances of less prevalent features were summed per sample into a group called "other". BAnOCC was used as described above.

Supernatant experiments

Strains and culture conditions

Odoribacter splanchnicus (DSM20712), *Bacteroides eggerthii* (DSM20697), *C. albicans* (SC5314/ ATCC MYA-2876), *C. albicans* (ATCC 10231), and *C. albicans* (ATCC 18804) were grown at 37 °C under anaerobic conditions (anaerobic gas mixture, 95% N_2, and 5% H_2) in pre-reduced modified Gifu anaerobic media (mGAM; Nissui Pharmaceutical Co. Ltd.) broth for liquid cultures or mGAM broth supplemented with agar (Nissui Pharmaceutical Co. Ltd.) for growth on plates.

Sterile bacterial supernatants

Bacterial strains grown for 48 h in mGAM broth were subcultured 1:50 in fresh mGAM broth and grown for 48 h in anaerobic conditions at 37 °C. Bacterial cultures were spun down at 11,000×g for 5 min. Supernatants were carefully removed and filtered through 0.2-μM syringe filters to remove bacteria in suspension.

Supernatant growth inhibition assays

C. albicans growth rates were analysed in 200 μl liquid mGAM with 50% or 100% sterile bacterial supernatant added. *C. albicans* inoculations were at 1:1000 from an overnight culture grown in aerobic conditions at 37 °C. Cultures were in 96-well microtiter plates at 37 °C with orbital shaking 365 cpm (2 mm). 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 the log OD measurements in log phase and calculating slopes for timepoints in log phase where r^2 was closest to 1, using at least 12 time points (2 h apart).

Supernatant metabolite assays

Analysis of SCFA in samples was carried out by MS-Omics as follows. Samples were acidified using hydrochloride acid, and deuterium labelled internal standards where added. All samples were analysed in a randomized order. Analysis was performed using a high polarity column (Zebron™ ZB-FFAP, GC Cap. Column 30 m × 0.25 mm × 0.25 μm) installed in a GC (7890B, Agilent) coupled with a quadropole 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 [68].

Other compounds such as bile acids were analysed using MS/MS. 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 and positive ionization mode. The UPLC was performed using a slightly modified version of the protocol described by Catalin et al. (UPLC/MS Monitoring of Water-Soluble Vitamin Bs in Cell Culture Media in Minutes, Water Application note 2011, 720004042en). Peak areas were extracted using Compound Discoverer 2.0 (Thermo Scientific). Identification of compounds were performed at four levels: level 1—identification by retention times (compared against in-house authentic standards), accurate mass (with an accepted deviation of 3 ppm), and MS/MS spectra; level 2a—identification by retention times (compared against in-house authentic standards), accurate mass (with an accepted deviation of 3 ppm); level 2b—identification by accurate mass (with an accepted deviation of 3 ppm), and MS/MS spectra; level 3—identification by accurate mass alone (with an accepted deviation of 3 ppm).

C. albicans growth inhibition by metabolites
Metabolites were acquired from the companies Sigma-Aldrich, Merck KGaA, and Roth. More specific details can be found in Suppl. Table 15.

C. albicans growth curves
Dilution series of metabolites in water were started at concentrations approximately 10-fold below maximum solubility in water where applicable (Suppl. Table 16). Dilutions were in synthetic SD medium ($1\times$ yeast nitrogen base, 2% glucose, 0.5% NH_4SO_4). *C. albicans* was grown overnight in YPD (1% yeast extract, 2 % peptone, 2 % glucose), washed $3\times$ in sterile water, and inoculated at 1:100 ($OD_{600} \approx 0.2$). Absorbance was measured every 15 min with an infinite M200pro microwell plate reader (Tecan, Austria) set to 30 °C with intermittent shaking (10 s orbital shaking before each measurement). Generation times were calculated from the obtained triplicate growth curves.

Host cell damage assays
To determine the influence of metabolites on the general capacity of *C. albicans* to cause host cell damage, we used an established epithelial cell model based on the vaginal epithelial cell line A431. A431 were grown in RPMI media containing 10% foetal bovine serum (FBS), and 200 μl cells at 10^5 cells/ml were seeded into 96-well plates and incubated at 37 °C, 5% CO_2. After 48 h, cells were washed with $1\times$ PBS, and 100 μl compound at indicated concentrations was added, followed by 100 μl *Candida* cells at multiplicity of infection 1. Incubation continued under the same conditions for 24 h. Basal lactate dehydrogenase (LDH) release (low control) was determined with 200 μl RPMI, and maximum LDH release (high control) determined by addition of 100 μl 0.5 % Triton X-100 to cells in 100 μl RPMI. Plate were centrifuged at $250\times g$ for 10 min and supernatants were removed and diluted 1:10 and mixed with 100 μl freshly prepared LDH assay mix (Roche). After 25 min at room temperature in the dark, LDH activity was determined with a microplate reader (Tecan infinite M200) as absorbance (A) at 492 nm, with 660 nm as a reference. Damage was calculated as $(A_{sample} - A_{low})/(A_{high} - A_{low})$.

C. albicans morphology
The effect of metabolites on *C. albicans* morphology was tested at all concentrations used in cell damage assays. Metabolites were diluted in 250 μl RPMI medium with 10% FBS and added to 250 μl *C. albicans* in RPMI in 24-well plates to indicated concentrations. Plates were incubated at 37 °C and 5% CO_2 for 4 h to induce hyphae formation. Medium was removed and cells fixed with Histofix 4% formaldehyde solution. Morphology was evaluated using an inverse microscope (Axio Zeiss Vert. A1) to differentiate yeasts, hyphae, and pseudohyphae.

Pairwise co-cultivation experiments
Interactions between *C. albicans* and *B. eggerthii* and *O. splanchnicus* were assayed via pairwise cultivations. *C. albicans* cell counts were compared to control conditions of cultivation without bacteria.

Fungal and bacterial cells were grown anaerobically at 37 °C for up to 48 h in mGAM and used as inocula for pairwise experiments. Inocula biomasses were estimated via OD600 and adjusted to 1.0 by diluting in appropriate media. Inocula were transferred to microplates containing the same media to a final OD600 of 0.01. Ratios of fungal to bacteria cells were 1:1. Microplates were incubated at 37 °C statically under anaerobic conditions. Cell counts from inocula were resolved, prior to the co-cultivation experiments, via flow cytometry (BD LSRFortessa, BD Biosciences, Franklin Lakes, NJ, USA).

Five microplates were prepared using the same inoculum. Microplates were removed from the anaerobic chamber every 5 h (0, 5, 10, 15, and 20 h cultivation). Cells were immediately fixed in 2% formaldehyde for 15 min at room temperature by mixing an equal amount of sample volume and 4% formaldehyde (Sigma-Aldrich, Saint Louis, MI, USA) [66, 67]. After fixing, total *C.*

albicans cells were counted via flow cytometry (BD LSRFortessa, BD Biosciences, Franklin Lakes, NJ, USA). Experiments were performed in triplicate.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40168-020-00899-6.

Additional file 1: Supplementary Figures.

Additional file 2. Statistical results for differences in fungal relative abundance between time points at genus level.

Additional file 3. Statistical results for differences in fungal relative abundance between time points at species level.

Additional file 4. Fungal network properties and statistical results for differences in node degree between time points.

Additional file 5. Growth rate indices of bacterial strains per sample as estimated by GRiD. GRiD values of 1 imply no growth.

Additional file 6. Statistical results for differences between group-centroids between different time points using distance-based redundancy analysis.

Additional file 7. Information on taxonomic annotation of co-abundant gene clusters (CAG). For each CAG, the bacterial species contributing the highest–and second highest–number of genes is shown. For the most contributing species, the percentage of genes assigned to that species is shown in "PercentageOfTaxonContribution". The column "trusted" indicates if the corresponding cluster passed our annotation criteria.

Additional file 8. Statistical results for differences in MGS relative abundance between time points.

Additional file 9. MGS network properties and statistical results for differences in node degree between time points.

Additional file 10. Cross-kingdom correlations representing the MGS-ITS correlation network. Correlations are listed for each time point.

Additional file 11. List of bacterial species with predicted positive or negative correlation to selected fungal species. The table further contains information on the corresponding MGS, time-related fold-changes, and for which antibiotic drugs the observation was made.

Additional file 12. Statistical results for differences in node degree between time points in RNA-MetaCyc network. Results are given for the entire network as well as differences based on funcational categories.

Additional file 13. Significant correlations between fungal species and stool metabolites.

Additional file 14. Patient information.

Additional file 15. List of published genome assemblies available on NCBI. A subset of these were used to infer strain information on several MGS clusters.

Additional file 16. Metabolite company information.

Additional file 17. Dilution series of metabolites for which we tested their effect on *Candida albicans* growth, morphology, and host cell damage.

Authors' contributions
Conceptualizations: M.O.A.S. and G.P. Funding acquisitions: M.O.A.S. and G.P. Investigations: R.V.U., F.L., P.C.Y.W., K.-F.C., C.-C.T., R.P.K.L., S.S., J.C., S.B., A.M., and B.S. Methodology: B.S., J.C., R.V.U., A.M., R.S., F.L., S.B., D.L., and K.K. Project administration: G.P. Resources: G.P., B.H., M.O.A.S., and P.C.Y.W. Supervision: G.P. Validation: all authors. Visualization: B.S., J.C., and R.S.. Writing—original draft: B.S. and G.P. Writing—review and editing: B.S, M.O.A.S., B.H., and G.P. The authors read and approved the final manuscript.

Availability of data and materials
Metagenomic and transcriptomic data generated and analysed during the current study are available in the NCBI SRA repository as Bio Projects PRJNA573821, PRJNA573905, and PRJNA579284. Data from previous work analysed in the current study is available in the NCBI SRA repository as Bio Project PRJINA588313.
Metabolite profiles were uploaded to MetaboLights (https://www.ebi.ac.uk/metabolights/) with accession MTBLS1846.
Code used for non-standard statistical tests are available as a git repository on https://bitbucket.org/Xentrics/antibiotic-gut-fungi/src/master/.

Ethics approval and consent to participate
The Danish study was approved by the local ethics committee in Region Zealand, Denmark (SJ-383). The Hong Kong study was approved by the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 17-042).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have not competing interests.

Author details
[1]Leibniz Institute for Natural Product Research and Infection Biology—Systems Biology and Bioinformatics, Hans Knöll Institute, Adolf-Reichwein-Straße 23, 07745 Jena, Germany. [2]Department of Medicine, State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, SAR, China. [3]Leibniz Institute for Natural Product Research and Infection Biology—Microbial Pathogenicity Mechanisms, Hans Knöll Institute, Adolf-Reichwein-Straße 23, 07745 Jena, Germany. [4]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, DK-2800 Lyngby, Denmark. [5]Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong. [6]Emergency Medicine Unit, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong. [7]State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Pokfulam, Hong Kong. [8]Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The University of Hong Kong, Pokfulam, Hong Kong.

References
1. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature [Internet]. 2007 [cited 2019 May 10];449:804–10. Available from: http://www.nature.com/articles/nature06244.
2. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Curr Opin Gastroenterol [Internet]. NIH Public Access; 2015 [cited 2019 May 16];31:69–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25394236.
3. Chakraborty A, Ghosh S, Chowdhary G, Maulik U, Chakrabarti S. DBETH: A database of bacterial exotoxins for human. Nucleic Acids Res [Internet]. Narnia; 2012 [cited 2019 May 10];40:D615–20. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr942.

4. Forster SC, Browne HP, NK. HPMCD: the database of human microbial communities from metagenomic datasets and microbial reference genomes. Nucleic Acids Res. 2016.

5. Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, et al. Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. Science (80-). American Association for the Advancement of Science; 2012;336:1314–7.

6. Wheeler ML, Limon JJ, Bar AS, Leal CA, Gargus M, Tang J, et al. Immunological consequences of intestinal fungal dysbiosis. Cell Host MicrobeElsevier. 2016;19:865–73.

7. Li X, Leonardi I, Semon A, Doron I, Gao IH, Putzel GG, et al. Response to fungal dysbiosis by gut-resident CX3CR1+ mononuclear phagocytes aggravates allergic airway disease. Cell Host Microbe [Internet]. Cell Press; 2018 [cited 2019 Sep 4];24:847-856.e4. Available from: https://www.sciencedirect.com/science/article/abs/pii/S1931312818305602.

8. Kim Y-G, Udayanga KGS, Totsuka N, Weinberg JB, Núñez G, Shibuya A. Gut dysbiosis promotes M2 macrophage polarization and allergic airway inflammation via fungi-induced PGE2. Cell Host MicrobeElsevier. 2014;15:95–102.

9. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A [Internet]. National Academy of Sciences; 2011 [cited 2019 May 10];108 Suppl 1:4554–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20847294.

10. Noverr MC, Noggle RM, Toews GB, Huffnagle GB. Role of antibiotics and fungal microbiota in driving pulmonary allergic responses. Infect Immun Am Soc Microbiol. 2004;72:4996–5003.

11. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat MicrobiolNature Publishing Group. 2018;1.

12. Dollive S, Chen Y-Y, Grunberg S, Bittinger K, Hoffmann C, Vandivier L, et al. Fungi of the murine gut: episodic variation and proliferation during antibiotic treatment. PLoS One [Internet]. Public Library of Science; 2013 [cited 2019 May 16];8:e71806. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23977147.

13. 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. Shared Science Publishers; 2018;5:249.

14. Peleg AY, Hogan DA, Mylonakis E. Medically important bacterial–fungal interactions. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2010 [cited 2019 May 10];8:340–9. Available from: http://www.nature.com/articles/nrmicro2313.

15. Frey-Klett P, Burlinson P, Deveau A, Barret M, Tarkka M, Sarniguet A. Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. Microbiol Mol Biol Rev [Internet]. American Society for Microbiology; 2011 [cited 2019 May 10];75:583–609. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22126995, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3232736.

16. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–14.

17. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods [Internet]. Nature Publishing Group; 2012 [cited 2019 May 27];9:811–4. Available from: http://www.nature.com/articles/nmeth.2066.

18. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res [Internet]. Cold Spring Harbor Laboratory Press; 2017 [cited 2019 May 10];27:626–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28167665.

19. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat MethodsNature Publishing Group. 2018;15:962.

20. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat MicrobiolNature Publishing Group. 2018;3:356.

21. Gweon HS, Oliver A, Taylor J, Booth T, Gibbs M, Read DS, et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. Methods Ecol EvolWiley Online Library. 2015;6:973–80.

22. 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 [Internet]. American Society for Microbiology; 2009 [cited 2019 May 13];75:7537–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19801464.

23. Aykut B, Pushalkar S, Chen R, Li Q, Abengozar R, Kim JI, et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature [Internet]. Nature Publishing Group; 2019 [cited 2019 Nov 4];574:264–7. Available from: http://www.nature.com/articles/s41586-019-1608-2.

24. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. NatureNature Publishing Group. 2010;464:59–65.

25. Emiola A, Oh J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. Nat Commun [Internet]. Springer US; 2018;9. Available from: http://dx.doi.org/https://doi.org/10.1038/s41467-018-07240-8.

26. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat MicrobiolNature Publishing Group. 2018;3:337.

27. Palleja A, Kashani A, Allin KH, Nielsen T, Zhang C, Li Y, et al. Roux-en-Y gastric bypass surgery of morbidly obese patients induces swift and persistent changes of the individual gut microbiota. Genome Med [Internet]. 2016;8:67. Available from: https://doi.org/https://doi.org/10.1186/s13073-016-0312-1.

28. Dehoux P, Marvaud JC, Abouelleil A, Earl AM, Lambert T, Dauga C. Comparative genomics of *Clostridium bolteae* and *Clostridium clostridioforme* reveals species-specific genomic properties and numerous putative antibiotic resistance determinants. BMC Genomics [Internet]. BioMed Central; 2016 [cited 2019 Sep 23];17:819. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3152-x.

29. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat BiotechnolNature Publishing Group. 2014;32:822.

30. Pedersen HK, Forslund SK, Gudmundsdottir V, Petersen AØ, Hildebrand F, Hyötyläinen T, et al. A computational framework to integrate high-throughput '-omics' datasets for the identification of potential mechanistic links. Nat Protoc [Internet]. Nature Publishing Group; 2018 [cited 2019 May 17];13:2781–800. Available from: http://www.nature.com/articles/s41596-018-0064-z.

31. Schwager E, Mallick H, Ventz S, Huttenhower C. A Bayesian method for detecting pairwise associations in compositional data. PLoS Comput BiolPublic Library of Science. 2017;13:e1005852.

32. Tamanai-Shacoori Z, Smida I, Bousarghin L, Loreal O, Meuric V, Fong SB, et al. *Roseburia* spp.: a marker of health? Future Microbiol. Future Medicine Ltd.; 2017. p. 157–170.

33. Machiels K, Joossens M, Sabino J, De Preter V, Arijs I, Eeckhaut V, et al. A decrease of the butyrate-producing species roseburia hominis and faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. GutBMJ Publishing Group. 2014;63:1275–83.

34. Martínez I, Lattimer JM, Hubach KL, Case JA, Yang J, Weber CG, et al. Gut microbiome composition is linked to whole grain-induced immunological improvements. ISME JNature Publishing Group. 2013;7:269–80.

35. Shibuya A, Shibuya K. Exploring the gut fungi-lung allergy axis. Cell Host MicrobeElsevier. 2018;24:755–7.

36. Costea PI, Munch R, Coelho LP, Paoli L, Sunagawa S, Bork P. metaSNV: A tool for metagenomic strain level analysis. Wang K, editor. PLoS One [Internet]. Public Library of Science; 2017 [cited 2019 May 10];12:e0182392. Available from: https://dx.plos.org/https://doi.org/10.1371/journal.pone.0182392.

37. Luo C, Knight R, Siljander H, Knip M, Xavier RJ, Gevers D. ConStrains identifies microbial strains in metagenomic datasets. Nat Biotechnol [Internet]. Nature Publishing Group; 2015 [cited 2019 May 10];33:1045–52. Available from: http://www.nature.com/articles/nbt.3319.

38. Albanese D, Donati C. Strain profiling and epidemiology of bacterial species from metagenomic sequencing. Nat CommunNature Publishing Group. 2017;8:2260.

39. 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.

40. Mason KL, Downward JRE, Falkowski NR, Young VB, Kao JY, Huffnagle GB. Interplay between the gastric bacterial microbiota and *Candida albicans* during postantibiotic recolonization and gastritis. Infect Immun. 2012;80:150–8.

41. 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 JNature Publishing Group. 2017;11:676–90.

42. 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. MicrobiomeBioMed Central. 2017;5:87.

43. Fouladi F, Glenny EM, Matthew ECB, Sioda M, Thomas SA, Wang Y, et al. Sequence variant analysis reveals poor correlations in microbial taxonomic abundance between humans and mice after gnotobiotic transfer. ISME J [Internet]. Springer US; 2020; Available from: http://dx.doi.org/https://doi.org/10.1038/s41396-020-0645-z.

44. 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 [Internet]. BioMed Central Ltd.; y [cited 2020 Apr 14];6: 152. Available from: https://microbiomejournal.biomedcentral.com/articles/1 0.1186/s40168-018-0538-9.

45. World Health Organization (WHO). World health organization model list of essential medicines. Ment Holist Heal Some Int Perspect [Internet]. 2019; 119–34. Available from: https://www.who.int/medicines/publications/essentialmedicines/en/.

46. Maier L, Goemans CV, Pruteanu M, Wirbel J, Kuhn M, Cacace E, et al. Dissecting the collateral damage of antibiotics on gut microbes. bioRxivCold Spring Harbor Laboratory. 2020;2020(01):09.893560.

47. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature [Internet]. Nature Publishing Group; 2017 [cited 2020 Apr 21];551:507–11. Available from: http://dx.doi.org/https://doi.org/10.1038/nature24460.

48. Roager HM, Hansen LBS, Bahl MI, Frandsen HL, Carvalho V, Gøbel RJ, et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. Nat MicrobiolNature Publishing Group. 2016;1:1–9.

49. Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, Pérez-Cobas AE, et al. Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS OnePublic Library of Science. 2011; 6:e17447.

50. Guinan J, Villa P, Thangamani S. Secondary bile acids inhibit *Candida albicans* growth and morphogenesis. Pathog Dis Oxford University Press. 2018;76.

51. 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 [Internet]. Nature Publishing Group; 2019 [cited 2019 Oct 23];9:8872. Available from: http://www.nature.com/articles/s41598-019-45467-7.

52. Witchley JN, Penumetcha P, Abon N V., Woolford CA, Mitchell AP, Noble SM. *Candida albicans* morphogenesis programs control the balance between gut commensalism and invasive infection. Cell Host Microbe [Internet]. Cell Press; 2019 [cited 2019 Mar 22];25:432-443.e6. Available from: https://www.sciencedirect.com/science/article/pii/S1931312819301040.

53. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2014 [cited 2019 Jul 2];12:661–72. Available from: http://www.nature.com/articles/nrmicro3344.

54. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes. Taylor and Francis Inc.; 2016. p. 189–200.

55. Pompella A, Corti A. Editorial: the changing faces of glutathione, a cellular protagonist. Front Pharmacol [Internet]. Frontiers; 2015 [cited 2019 Aug 9];6: 98. Available from: http://www.frontiersin.org/Experimental_Pharmacology_and_Drug_Discovery/10.3389/fphar.2015.00098/full.

56. Zhang Y, Cai C, Yang Y, Weng L, Wang L. Blocking of Candida albicans biofilm formation by cis-2-dodecenoic acid and trans-2-dodecenoic acid. J Med Microbiol [Internet]. Microbiology Society; 2011 [cited 2019 Jul 2];60: 1643–50. Available from: http://jmm.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.029058-0.

57. Yang DL, Zhang YQ, Hu YL, Weng LX, Zeng GS, Wang LH. Protective effects of cis-2-dodecenoic acid in an experimental mouse model of vaginal candidiasis. Biomed Environ Sci [Internet]. Biomedical and Environmental Sciences; 2018 [cited 2019 Jul 2];31:816–28. Available from: https://www.sciencedirect.com/science/article/pii/S0895398819300534.

58. Raymond F, Ouameur AA, Déraspe M, Iqbal N, Gingras H, Dridi B, et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. ISME JNature Publishing Group. 2016;10:707–20.

59. White TJ, Bruns T, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR protocols: a guide to methods and applications. PCR Protoc a Guid to methods Appl. San Diego, CA: Academic Press; 1990. p. 315–22.

60. Zuo T, Wong SH, Cheung CP, Lam K, Lui R, Cheung K, et al. Gut fungal dysbiosis correlates with reduced efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. Nat Commun [Internet]. Nature Publishing Group; 2018 [cited 2019 Aug 5];9:3663. Available from: http://www.nature.com/articles/s41467-018-06103-6.

61. Kang K, Ni Y, Li J, Imamovic L, Sarkar C, Kobler MD, et al. The environmental exposures and inner- and intercity traffic flows of the metro system may contribute to the skin microbiome and resistome. Cell Rep [Internet]. Cell Press; 2018 [cited 2019 May 21];24:1190-1202.e5. Available from: https://www.sciencedirect.com/science/article/pii/S2211124718310519.

62. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat MethodsNature Publishing Group. 2013; 10:1200.

63. Dixon P. VEGAN, a package of R functions for community ecology. J Veg SciWiley Online Library. 2003;14:927–30.

64. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. BioinformaticsOxford University Press. 2012; 28:2106–13.

65. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nat Rev MicrobiolNature Publishing Group. 2018;1.

66. Chao Y, Zhang T. Optimization of fixation methods for observation of bacterial cell morphology and surface ultrastructures by atomic force microscopy. Appl Microbiol BiotechnolSpringer. 2011;92:381–92.

67. Moloney M, McDonnell L, O'Shea H. Atomic force microscopy of BHK-21 cells: an investigation of cell fixation techniques. UltramicroscopyNorth-Holland. 2004;100:153–61.

68. Lea G. Johnsen, Peter B. Skou, Bekzod Khakimov, Rasmus Bro. Gas chromatography – mass spectrometry data processing made easy. J Chromatogr A. 2017;1503:57–64.

Publisher's Note

Supplement

Suppl. Fig. 1: Fungal species alpha diversity per antibiotic. Lines show average values. Error lines show standard error. Alpha diversity was measured by Shannon index, Simpson 1-D, and the number of observed OTUs. We focus on Shannon index in the following. Baseline diversity was similar between subjects, ranging from 0.2-0.5 (one outlier with 1.4). In many cases, diversity increased already during treatment (DT). Augmentin showed high variation between subjects and Cefotaxime a slight decrease. In most cases, we observed a spiked increase 30d post treatment (EPT), including controls. Doxycycline exhibited an increase, but overall flatter response. Augmentin showed the strongest peak increase by far. Diversity at 90d post treatment (LPT) was diverse, but indices fell within the range of controls. Overall, Augmentin and Cefotaxime induced a considerable gain in fungal diversity compared to controls. Ciprofloxacin induced higher than control change in one patient as well. Only Azithromycin and Doxycycline showed variation in alpha diversity within the same range as observed for controls.
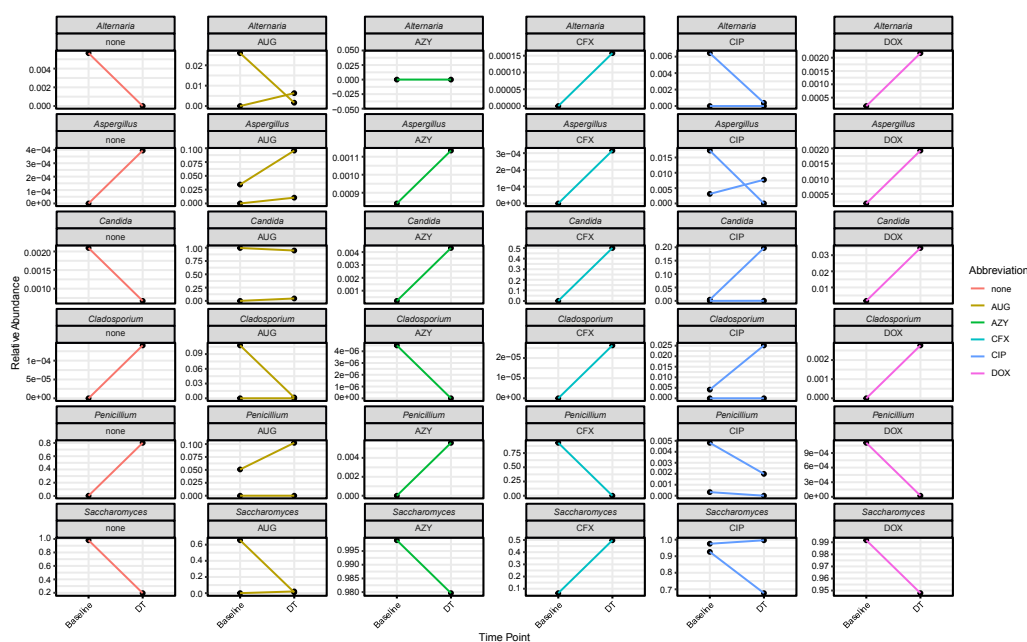
Suppl. Fig. 2: Mean relative abundance per antibiotic of fungal genera for Baseline and DT. Only Candida had significant increase in relative abundance (15-fold). At the level of individual antibiotic drugs, we also observed increased relative abundance except for AUG (almost no change).
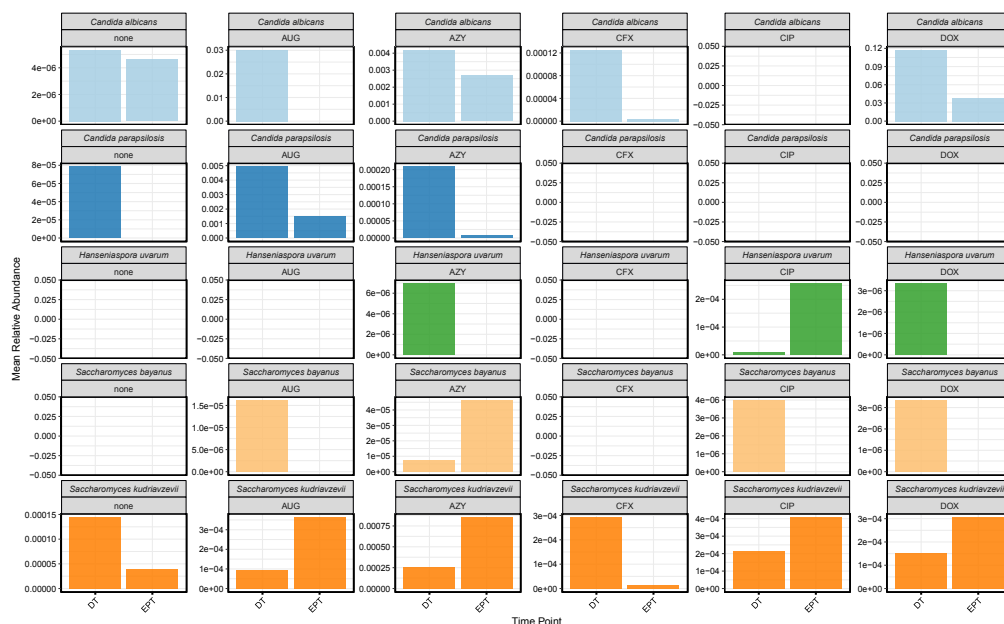
Suppl. Fig. 3: Mean relative abundance per antibiotic of fungal species with significant change from DT to EPT. Displayed are five fungal species. *Candida albicans* was measured in each group except in CIP. In the remaining cases, its abundance decreased consistently. AUG and DOX had the strongest effect, especially with respect to the initiate abundance levels > 10%. A minor reduction in growth was also observed in controls, but not as much. *Candida parapsiloses* was measured in at most halve of the patients. In decreased profoundly in relative abundance at EPT in both, treated and untreated patients. The other 3 fungi had less consistent patterns at the level of individual antibiotic drugs.
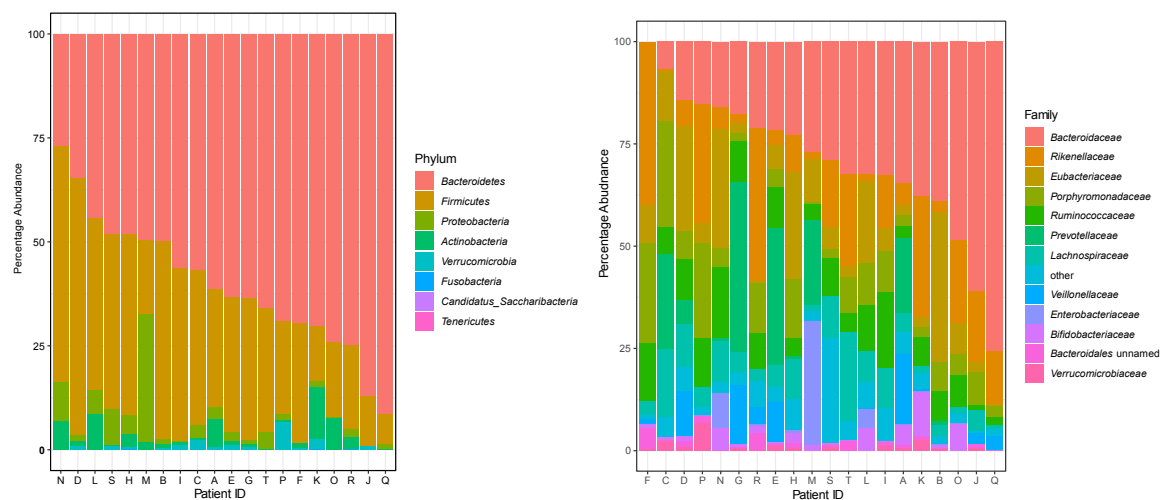
Suppl. Fig. 4 (left): Bacterial phylum composition of baseline samples. Patients (x-axis) were ordered by the level of the highest contributing phylum (Bacteroides). Phyla were ordered by their average contribution across samples. Most patients were dominated by Bacteroides and Firmicutes spp. These two phyla together accounted for 75-98% of sample-wise abundances. In patient M, we observed an unusual strong contribution by Proteobacteria (30%), which are otherwise the 3rd most abundant phylum on average. Remaining contributions were from Actinobacteria, Verrucomicrobia and Fusobacteria. An insignificant fraction was contributed by Candidatus Saccharibacteria and Tenericutes.

Suppl. Fig. 4 (right): Bacterial family composition of baseline samples. Patients (x-axis) were ordered by the level of the highest contributing phylum (*Bacteroidaceae*). The 12 most abundant families across samples are shown. Abundance from other families were summed up as "other". Families were ordered by their average contribution across samples. Overall, we observed a strong difference between *Bacteroidaceae* spp. (from <1% to 75%) and the remaining bacterial spp. contributions. *Bacteroidaceae, Rikenellaceae, Eubacteriaceae, Porphyromonoadaceae, Ruminococcaceae, Prevotellaceae* and *Lachnospiraceae* accounted for most contributions across samples. *Enterobacteriaceae* accounted for only 0.02% on average, except patients M (30%), L (5%) and N (9%), which are all control patients.
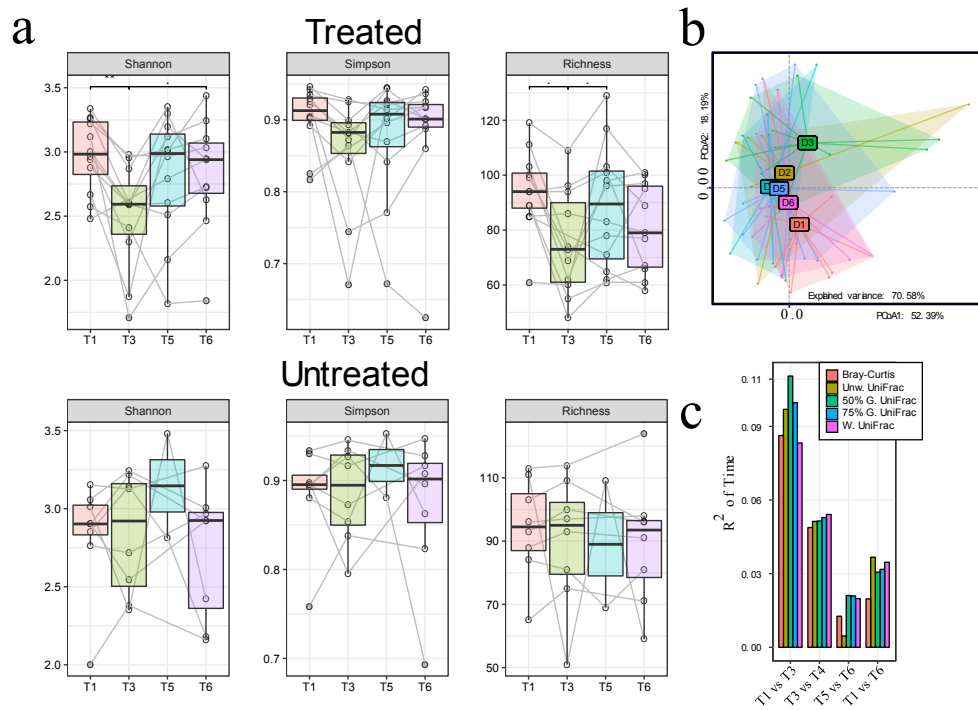
Suppl. Fig. 5: Antibiotics induced severe changes in bacterial community 6 days after treatment. (a-c) Diversity analysis of samples from subjects using MetaPhlAn2 relative abundances. **(a)** Boxplots showing Species Richness (right), Shannon (left) and Gini-Simpson Index (middle). First row shows diversity for antibiotic treated samples. Second row show alpha diversity for control samples. The median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers) are shown. Lines between boxes connect same-donor samples. Statistical testing was performed using a Wilcoxon signed-rank test and p values were adjusted for multiple testing using FDR (q). Signs indicates significance level (**: q<0.01; .: 0.05<q<0.1). **(b)** Principle coordinate analysis of generalized UniFrac distance (a = 0.5) as a measure of beta diversity. **(b-c)** We tested for differences between times in treated subjects while controlling for subjects using a pairwise two-way PERMANOVA. **(c)** R² values of covariate "Time" from pairwise PERMANOVA. Five different measures of beta diversity were tested independently. R² values for consecutive samples are shown.
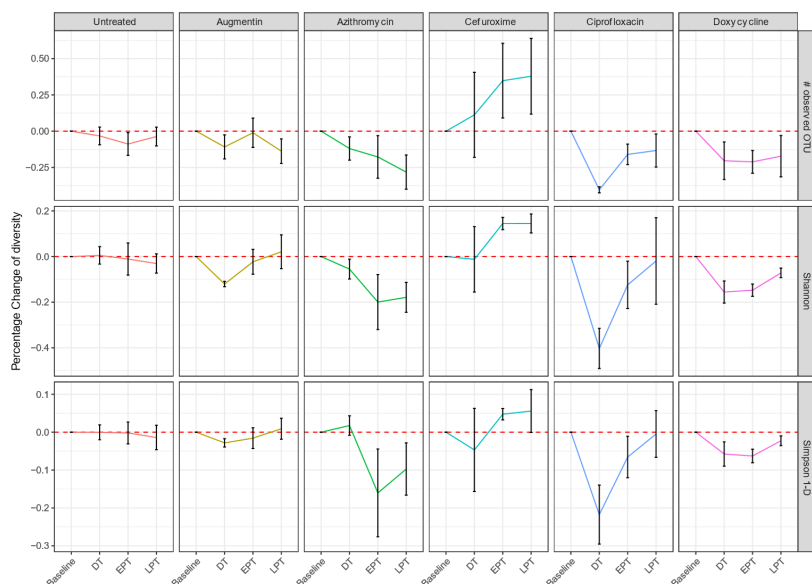
Suppl. Fig. 6: Proportional change of bacterial species alpha diversity compared to Baseline. Proportional change of bacterial alpha diversity at species level based on MetaPhlAn2 OTU profiling. In untreated patients, we observed a drop in species richness during after treatment. However, Shannon and Simpson diversity remained mostly unchanged, implying changes due to sequencing depth or other technical artifacts. In antibiotic treated patients, alpha diversity changes the most for Shannon diversity. It decreased by 5%-40% during treatment, except for CFX treated patients. One CFX treated patient showed a monotonous increase in diversity compared to baseline. The strongest, negative impact was observed using CIP. In most patients, alpha diversity was increased 90d post treatment compared to their respective DT and EPT time points. This implies that most of the original diversity was regained after treatment, but not all. Furthermore, the difference between LPT and Baseline alpha diversity range from +18% to -21%. Some of difference in late-post treatment diversity might be explained with natural variation we observed in controls (left-most panel; roughly +/- 5%).
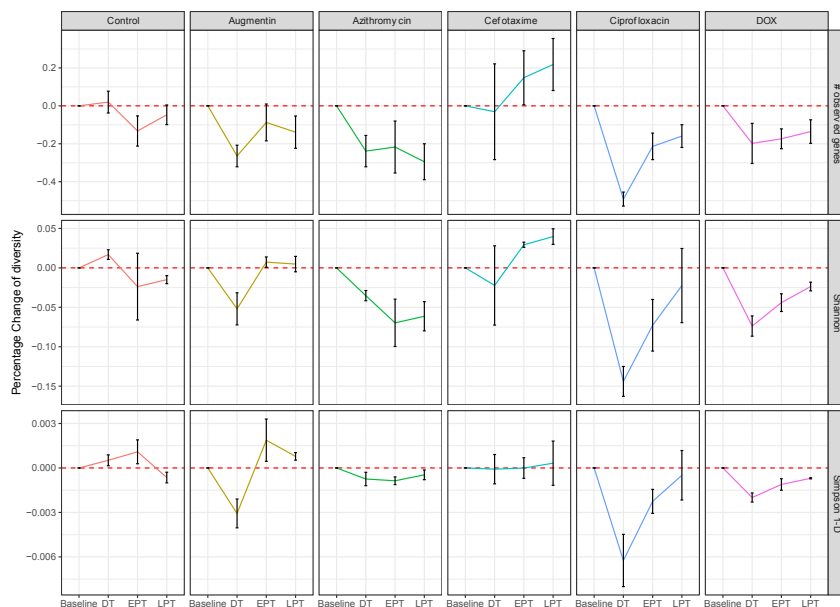
Suppl. Fig. 7: Proportional change of DNA gene family alpha diversity compared to Baseline. Gene family abundance was estimated by HUMAnN2 pipeline. In untreated patients, we observed a drop in gene richness after treatment (EPT). However, Shannon and Simpson had contrasting differences to that (mixed for Shannon, increase for Simpson), implying changes due to sequencing depth or other technical artifacts. In antibiotic treated patients, richness decreased by 19%-49% during treatment, except for CFX treated patients. One CFX treated patient showed a monotonous increase in diversity compared to baseline. Differences in Shannon diversity were less severe, ranging from 0.02% to 0.18%. These two findings together imply that many genes lost during treatment (as measured by richness) were in relatively low abundance (and hence did not affect Shannon diversity much). So overall, the changes observed for bacterial species diversity were qualitatively like those for gene family

diversity. The major exceptions were the control samples, which exhibited bigger variance in diversity at the functional level compared to the species level
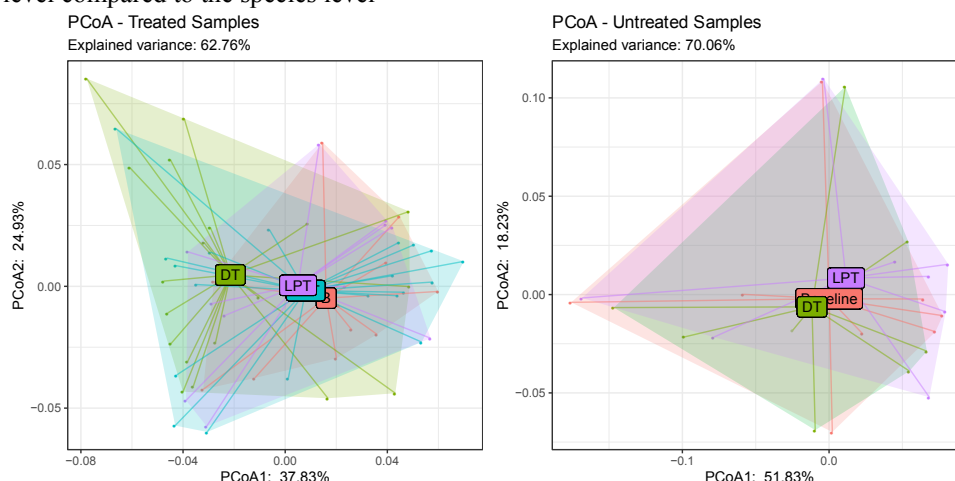
Suppl. Fig. 8: Principle coordinate analysis of DNA bacterial function beta diversity between time points. Beta-diversity was measured as Bray-Curtis index. We separate between treated samples (left) and untreated controls (right). Time points were pre-treatment (Baseline, B; red), treatment (DT; green), 30d post treatment (EPT; blue) and 90d post treatment (LPT; violette). In treated samples, we observe a significant difference between Baseline and DT, q=0.014). Control samples did not show significance differences across time points.
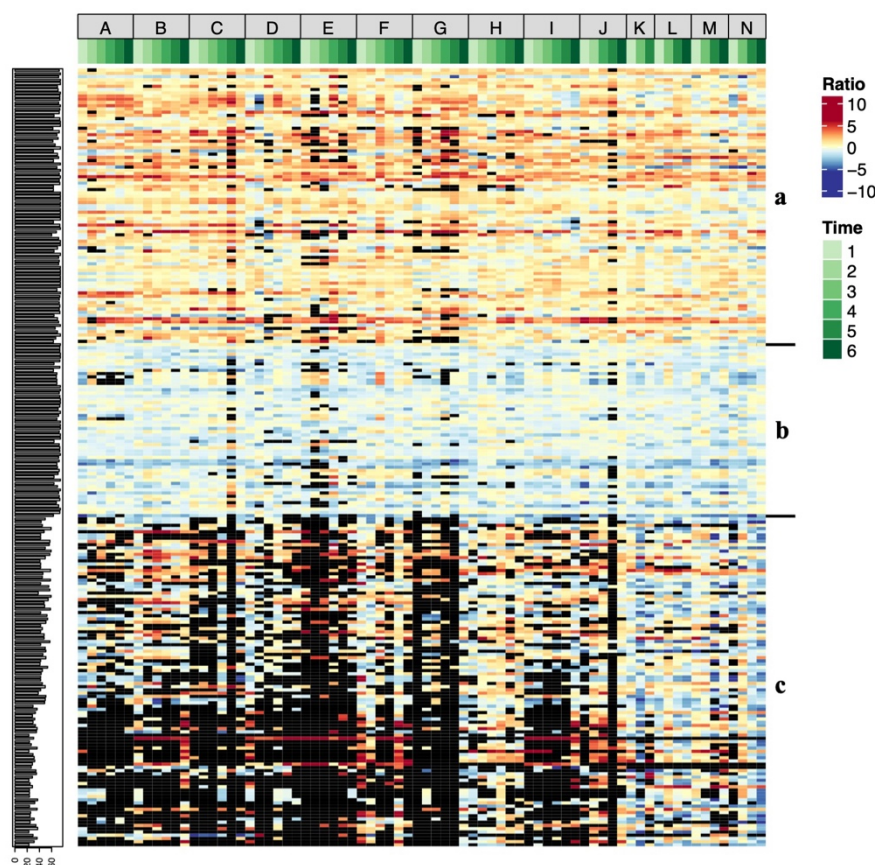
Suppl. Fig. 9: Core- and Variable Metatranscriptome of MetaCyc pathways (PWY). (a) Up-regulated core with RNA/DNA > 1 and prevalence >= 80%. **(b)** Down-regulated core with RNA/DNA < 1 and prevalence >= 80%. **(c)** Variable metatranscriptome with prevalence between 30% and 80%, ordered by mean RNA/DNA ratio. There were no significant differences in RNA/DNA ratio of PWYs between time points.
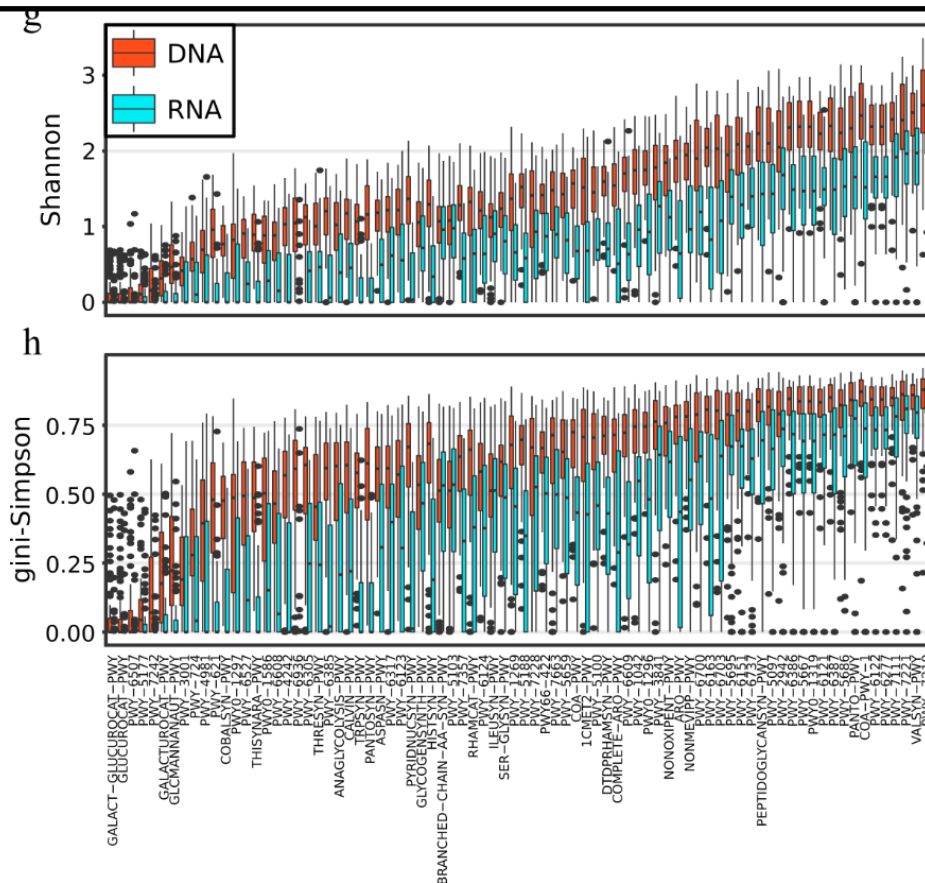
Suppl. Fig. 10: Contributional alpha diversity of MetaCyc pathways using Simpson (**g**; 1-D) and Shannon (**h**) diversity indices. Pathways are ordered by the sum of the mean DNA and mean RNA Shannon diversity. RNA diversity was generally lower than DNA diversity. Shannon diversity dropped more gradual compared to Simpson. A Shannon diversity of 2 implies that only 4 species are truly relevant for the contribution to the corresponding pathway. A Simpson diversity of 75% implies that (in addition) 1 species accounts for 75%.
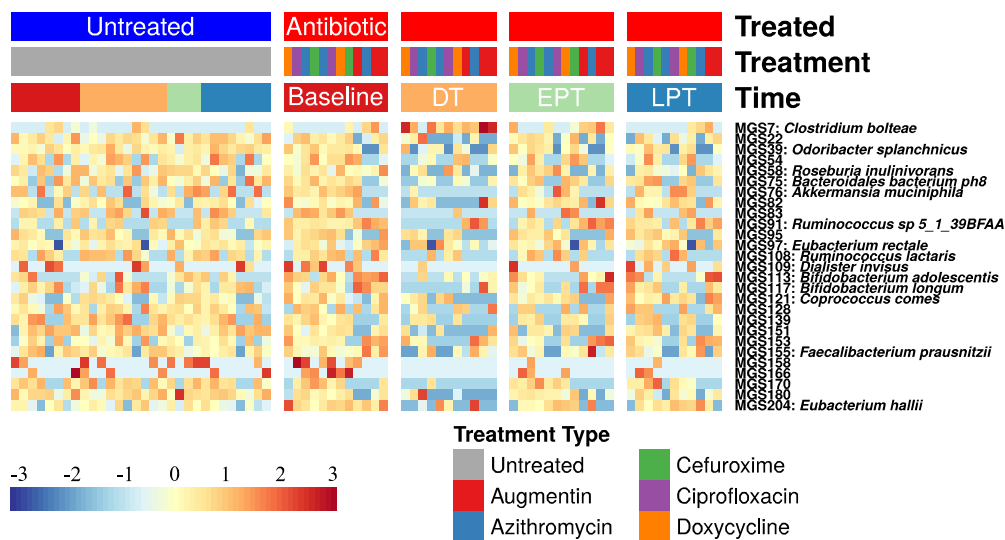
Suppl. Fig. 11: 26 MGS were sign. diff. abundant during treatment. Heatmap of z-transformed, cumulative sum scaled relative abundances of significantly differentially abundant MGS between baseline and treatment time (ZIG model; Time and Patient as covariate; $q_{Time} < 0.05$).
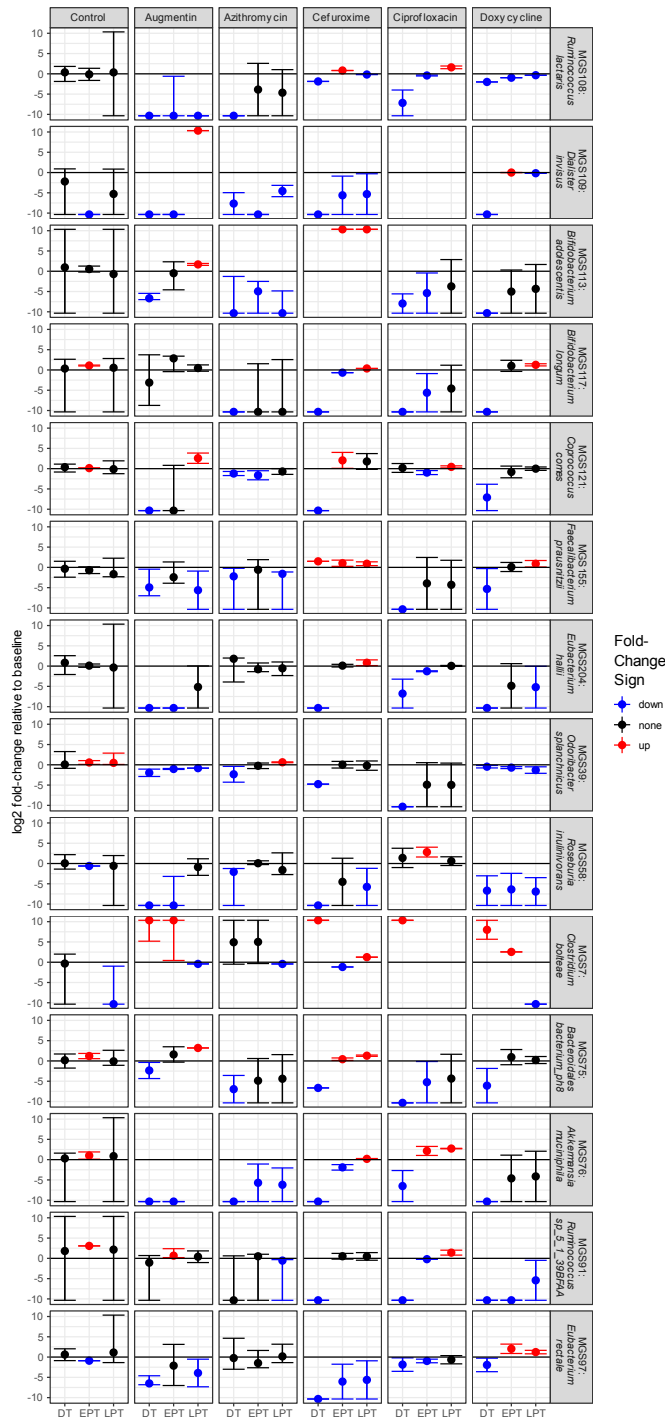
Suppl. Fig. 12: Fold-change of named MGS. Out of 26 MGS with sign. change during treatment, 14 had species level annotation. Here, log₂ fold-changes relative to baseline levels are shown per treatment group. Observation with 0 counts before and after treatment were ignored. Fold-changes of +/- infinity were set to 110% of the strongest non-infinite value to presence/absent observations. Whiskers show minimum and maximum values. Colours indicate if fold-changes were positive (red), negative (blue) or both (black) at a given time in each treatment group. 6 of these were consistently decreased independent of the antibiotic drug used: *Ruminococcus lactaris, Dialister invisus, Odoribacter splanchnicus, Bacteroidetes bacterium ph8, Akkermansia muciniphila, Bifidobacterium adolescentis*. Among the other species (8), 7 had negative fold change in 4 out of 5 drugs: *Bifidobacterium longum, Coprococcus comes, F. prausnitzii A2-165, Eubacterium hallii, Roseburia inulinivorans, Ruminococcus sp. 5_1_39BFAA, Eubacterium rectale*. *Clostridium bolteae* had positive fold change in 4 out of 5 drugs.
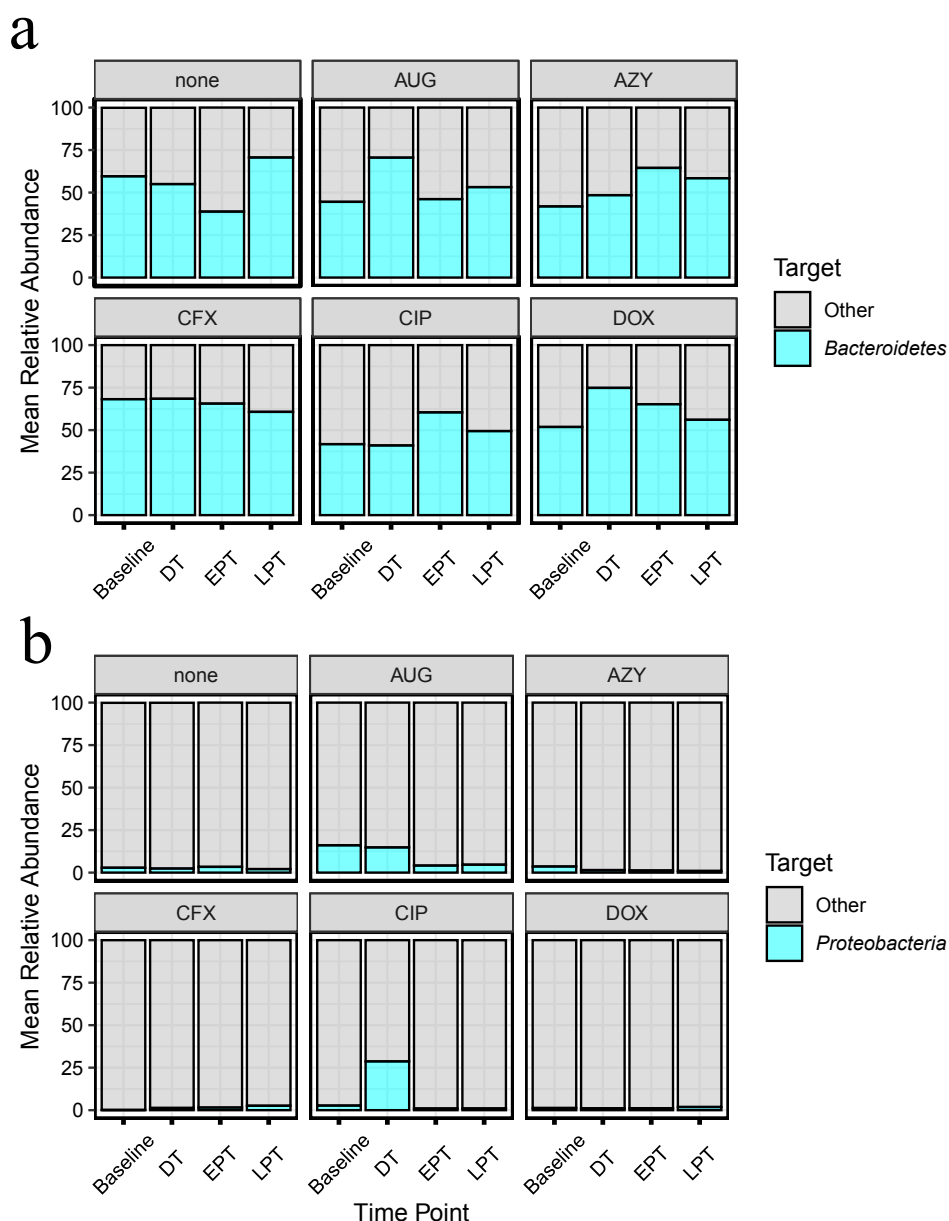
Suppl. Fig. 13: Mean relative abundance of species with possible differential effect by specific antibiotic classes. Maier et al. (https://doi.org/10.1101/2020.01.09.893560) reported that broad-spectrum antibiotics targeted all measured - the 40 most common - gut bacteria. However, beta-lactams may have a different influence on *Bacteroides* (a) compared to other species. Here, this would apply to Augmentin (AUG) and Cefuroxime (CFX). In all treated samples, *Bacteroidetes* increased in relative proportion during (DT) or early-post treatment (EPT). Despite the expectation, the effect seems to be rather low in CFX treated samples, which had comparatively high levels of *Bacteroidetes* at baseline. Likewise, Doxycycline (tetracycline class) showed an effect equal to Augmentin. So overall, we only noticed a slight delay in response for CIP and AZY, but not qualitative   difference.

Analogous, macrolides may have a different influence on *Proteobacteria*. Here, this would apply to Azithromycin. In most subjects, *Proteobacteria* have less than 1% contribution, making it hard to assess if relative abundance changes are due to colonization difficulties or antibiotic effect. We noticed that Augmentin lead to a severe decline in abundance for this phylum even though it belongs to an entirely different class. In contrast, CIP lead to an increase. Since we have relative abundance, we cannot assess if abundance increase relates to an actual increase in abundance. An increase can be the result of a severe decrease of the remaining community. Such could be the case   for   CIP.

So all together, we cannot find a selective difference in targets by the different antibiotics used.

a

b
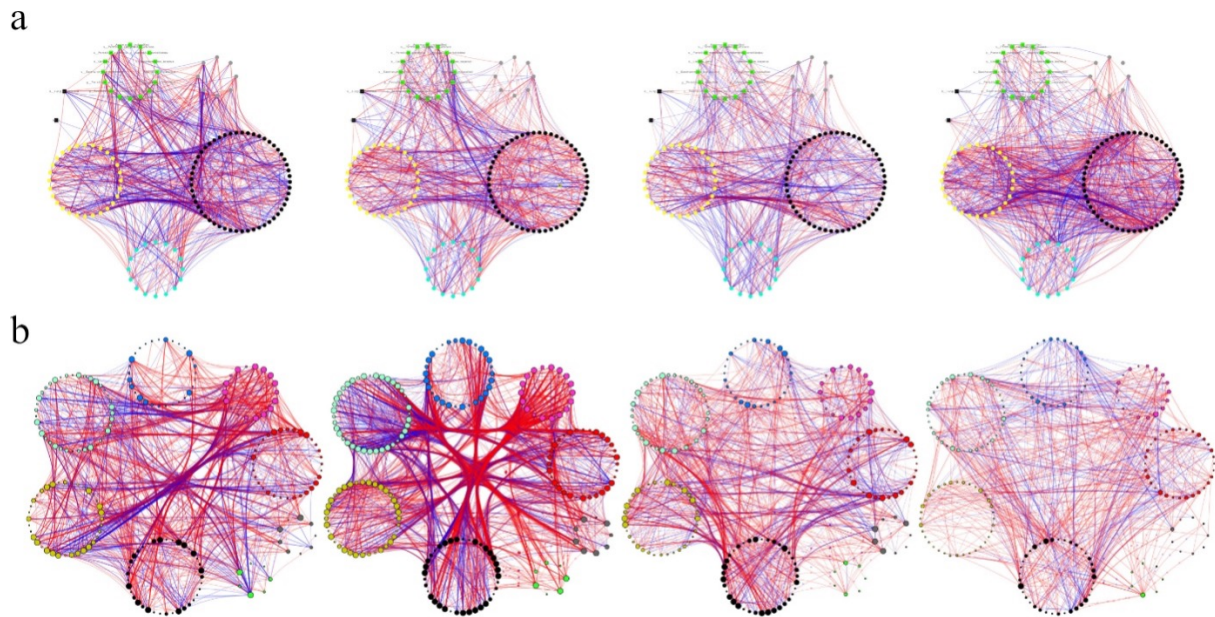
Suppl. Fig. 14: interactions between fungi, bacteria species and pathway expression. (**a-b**) Co-abundance networks of different time points using BAnOCC with (a) 25% and (**b**) 50% prevalence filter. Only significant edges (based on 95% credibility interval) with |r| >= 0.3 are shown. Negative correlations (blue), positive correlations (red). Networks are ordered from left (Baseline) to right (Late-Post-Treatment). (**a**) Correlations between fungal and bacterial species based on MGS and ITS relative abundances. Node colors indicate phyla. Unclassified MGS are black. (**b**) Correlations between fungal species and pathway expression based on HUMAnN2 RNA pathway and ITS relative abundances. Node colors indicate fungi (green) and functional groups. Superpathways and other pathways which did not fit into the six major categories were grouped into "other".
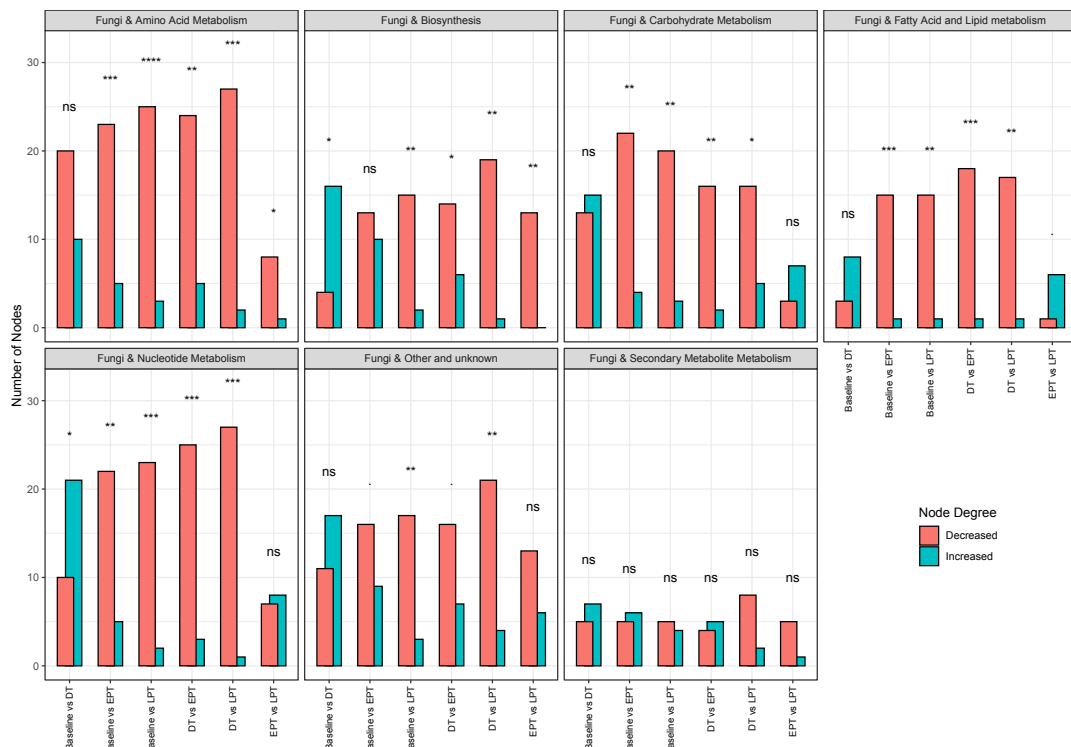
Suppl. Fig. 15: Node degree centrality of RNA-PWY-ITS network based on correlations between fungal species and MetaCyc pathway groups. Bar plots showing the number of nodes which increased and decreased in centrality between time points. Statistical testing for significant changes in centrality was performed using a two-sided Wilcox signed-rank test. P values were adjusted for multiple testing. Significance is indicated by symbols (ns: q ≥ 0.05; *: q < 0.05; **: q < 0.01; ***: q < 1e-3; ****: q < 1e-4; *****: q <1e-5).
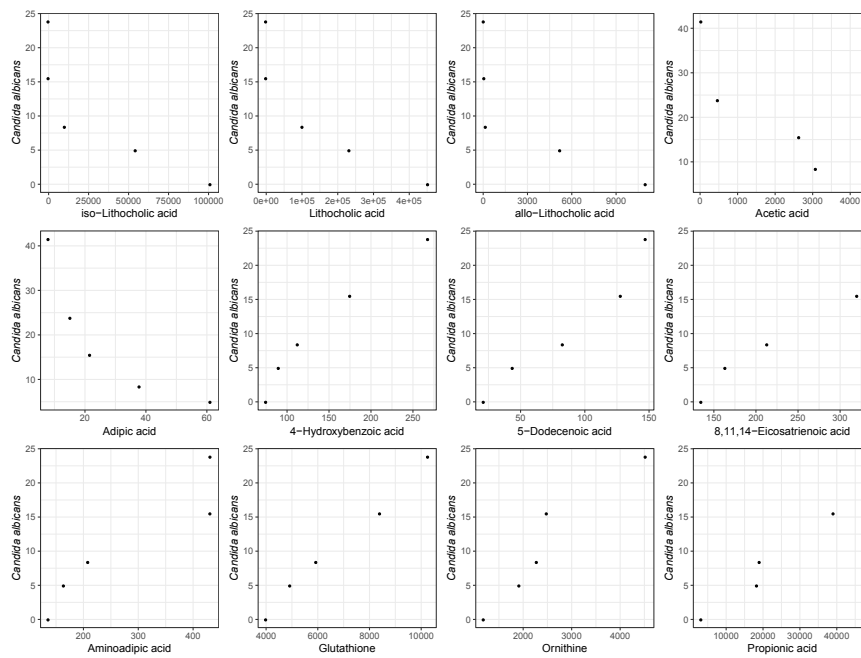
Suppl. Fig. 16: Metabolites significantly correlated with C. albicans OTU abundance levels (Spearman |rho| > 0.3; p < 0.05).
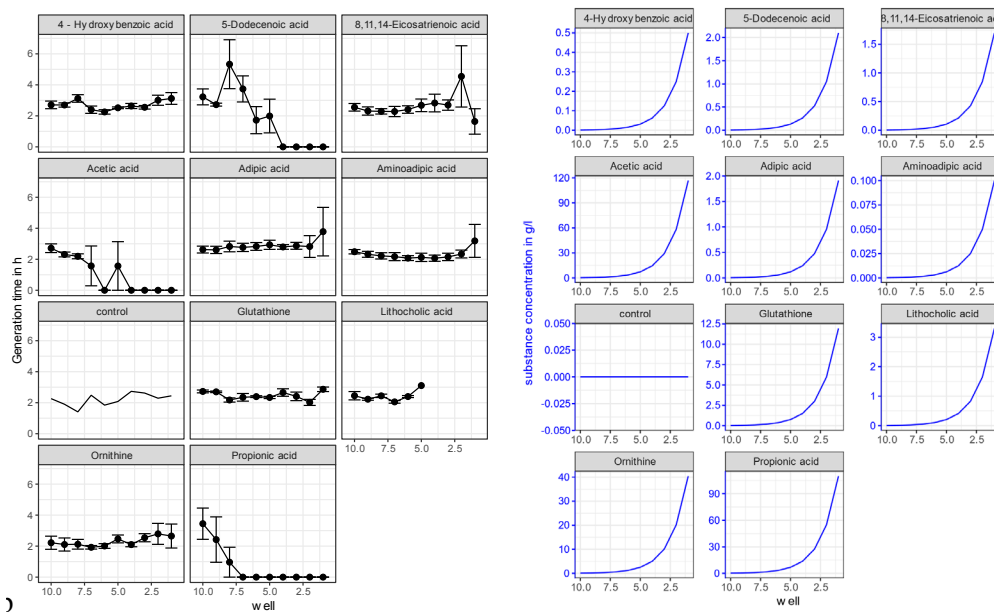
Suppl. Fig. 17: *C. albicans* **growth inhibition by metabolites.** (left) Generation times of C. albicans in h over varying degrees of substance concentrations. Substances were diluted (from left to right). For some compounds, measurements failed at higher concentrations. (right) Substance concentration per well in g/l.
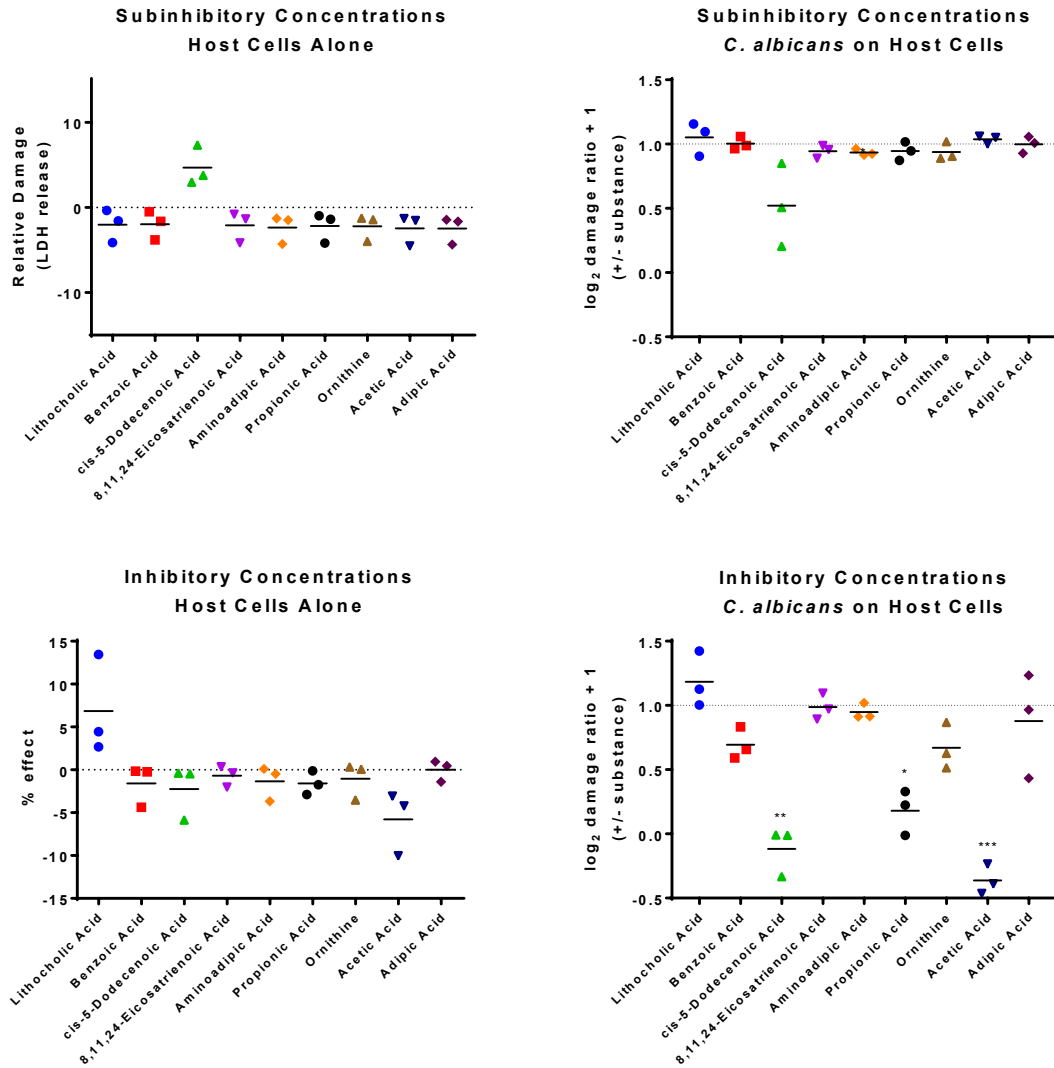
Suppl. Fig. 18: Inhibition of *C. albicans* growth and host cell damage. We tested the effect of the metabolites on a human vaginal cell line (A431) at two concentrations each: the lowest concentration where *C. albicans* growth was significantly inhibited and the highest without any significant effect on growth (table 1; top-left; bottom-left). There was some limited cytotoxicity observed for lithocholic acid and cis-5-dodecenoic acid, but all other substances did not elicit any detectable host cell damage. Glutathione, however, interfered with our assay and was therefore excluded from these studies. Next, we assayed the effect of the same concentration on cell damage by *C. albicans*. None of the substances affected the host-pathogen interactions at the lower concentration (top-right). The notable exception is cis-5-dodecenoic acid, with severely reduced the damage by *C. albicans* to the host cells, albeit not at a statistically significant level. At the higher concentrations (bottom-right), where fungal growth was reduced in vitro, we also observed lower host cell damage in presence of the short-chain fatty acids propionic and acetic acid, where damage by *C. albicans* was nearly fully abolished, and again of cis-5-docenoic acid. To lesser extent and not statistically significant there was a tendency to lower damage also with ornithine (p=0.087) and benzoic acid (p=0.051) in the medium.

Suppl. Table 1

[µg/µl]	Lithocholic Acid	Benzoic Acid	cis-5-Dodecenoic Acid	8,11,24-Eicosa-trienoic Acid	Aminoadipic Acid	Propionic Acid	Ornithine	Acetic Acid	Adipic Acid
Subinhibitory	0.103	0.25	0.263	0.85	0.05	0.0004	1.27	0.0004	0.95
Inhibitory	0.413	1.00	1.05	3.40	0.20	1.70	5.05	7.29	3.82
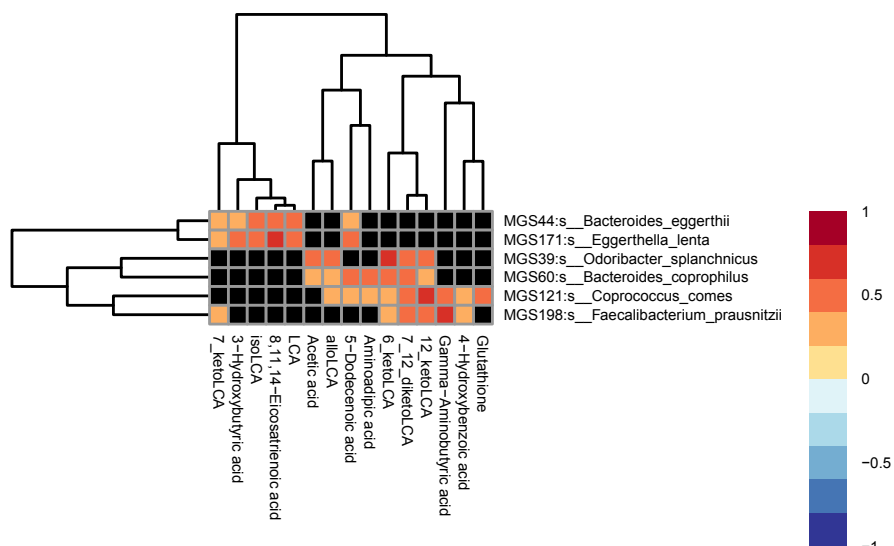
Suppl. Fig. 19: Correlations of bile acid and metabolite abundance with MGS abundance. Cell colour indicates correlation strength (blue: negative; red: positive). Absolute correlation below 0.3 and insignificant correlations (p > 0.05) are coloured white.
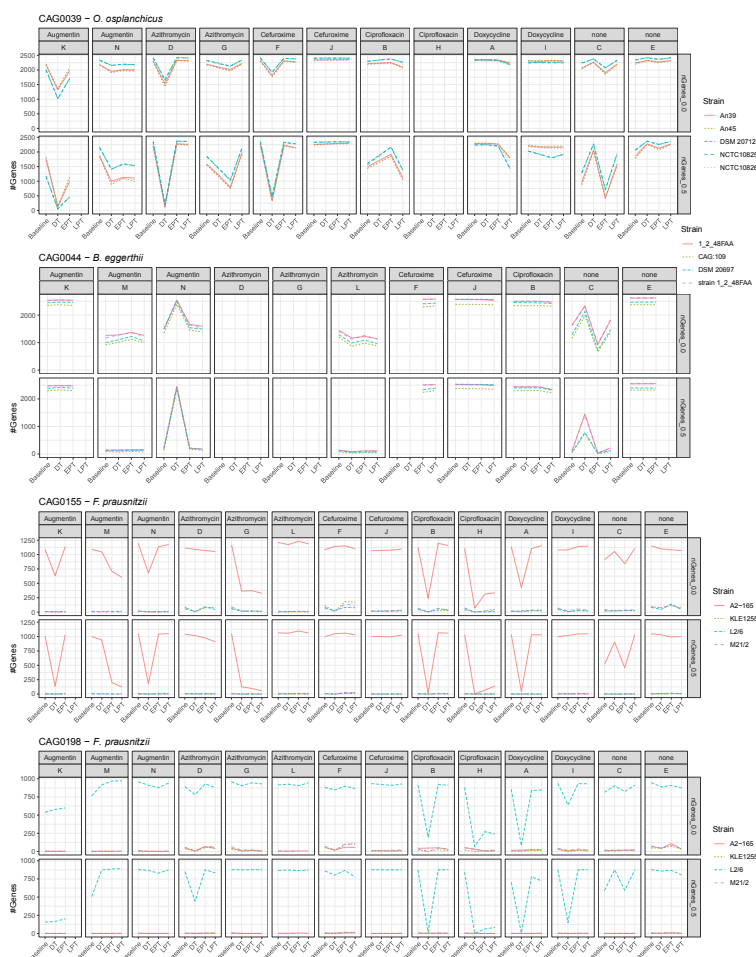
Suppl. Fig. 20: Number of genes contributing to each MGS. For each sample, the number of genes with at least 1 gene (first row of each panel) and at least 50% coverage (2. Row) per strain. Different strains are indicated by colour and line type. For Faecalibacterium prausnitzii, we have 2 MGS with exactly 1 matching strain for each (CAG0155 with A2-165 and CAG0198 with L2/6)
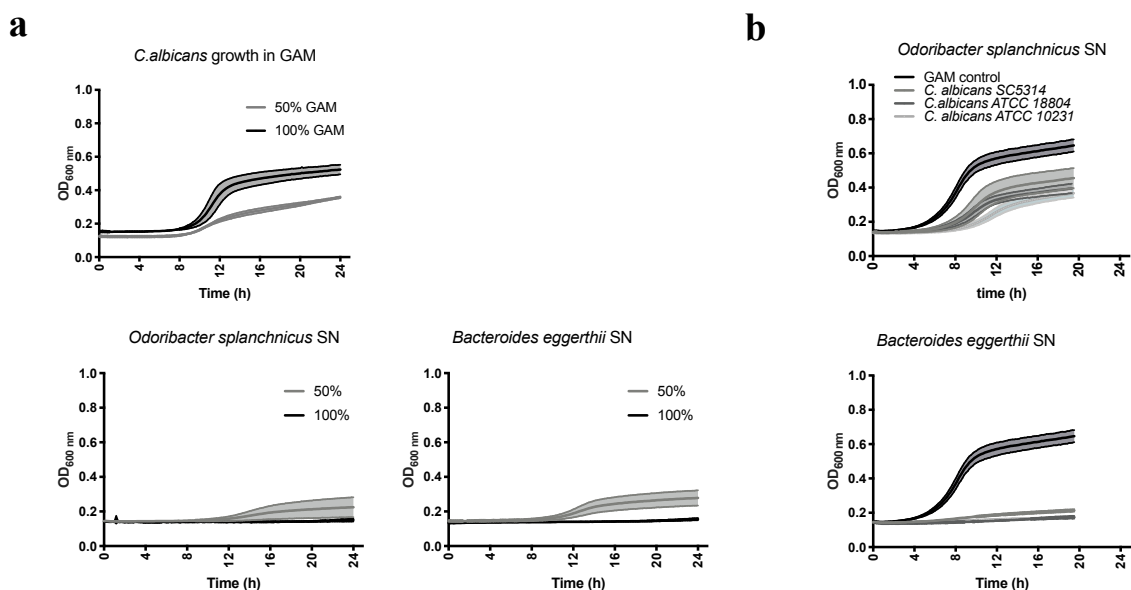
Suppl. Fig. 21: *C. albcians* **growth rate on bacterial supernatant. (a)** Growth curves for *C.albicans* strain SC5314 mGAM media with 50% or 100% sterile bacterial supernatant added. **(b)** Growth curves for three *C.albicans* strains; SC5314, ATCC10231 and ATCC 18804 in mGAM media with 100% sterile bacterial supernatant added.
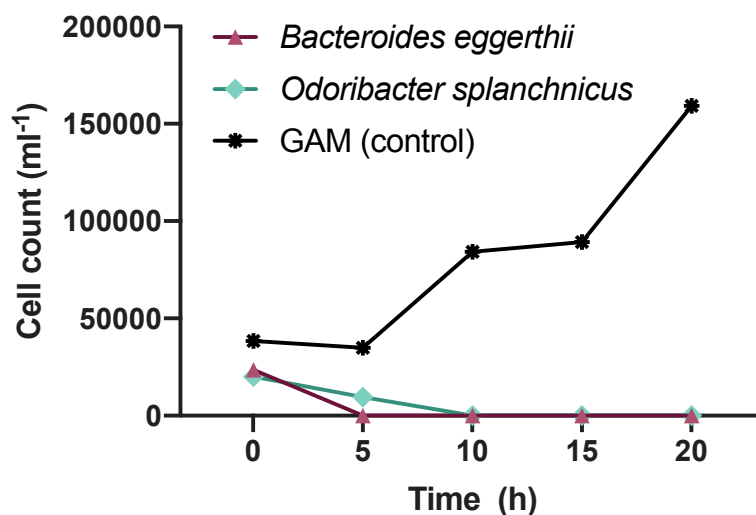
Suppl. Fig. 22: Co-cultivation of *C. albicans* with bacterial species. Cell counts of *C. albicans* was measured using FACS. When co-cultured with B. eggerthii or O. splanchnicus, no growth occurred.
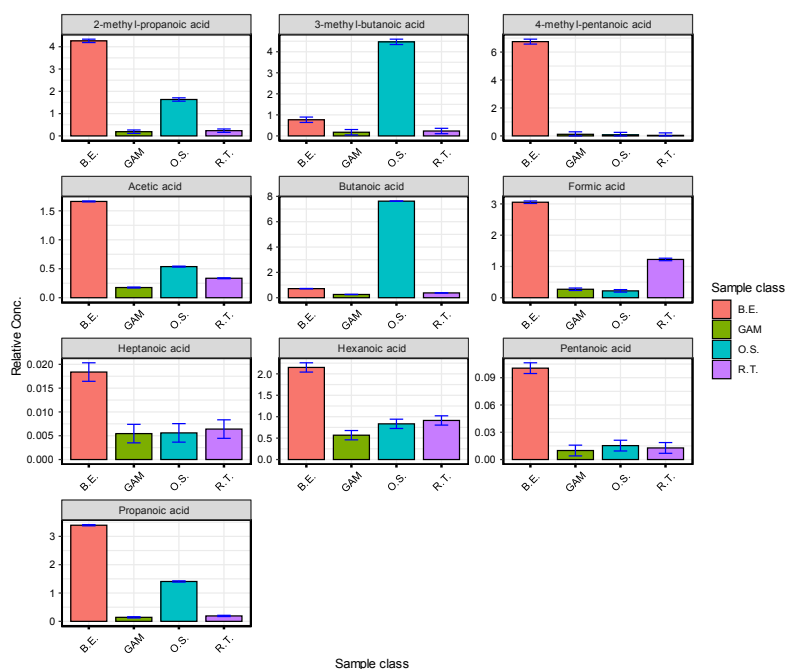
Suppl. Fig. 23: Relative concentration of short-chain-fatty-acids (SCFA) in bacterial supernatants. Here, we display all SCFA with significant difference to control samples (fold-change greater 2). Control samples were created by random pooling of sample material (see main manuscript methods for details). We measured the supernatants of *Bacteroides eggerthii* (B.E.), *Odoribacter splanchnicus* (O.S.) and *Ruminococcus [Blautia] torques* (R.T.). OS and BE were the main producer of most of these compounds. OS was by far the strongest producer of butyrate (butanoic acid & 3-methyl butanoic acid). RT produced Formic acid, but not as much BE.
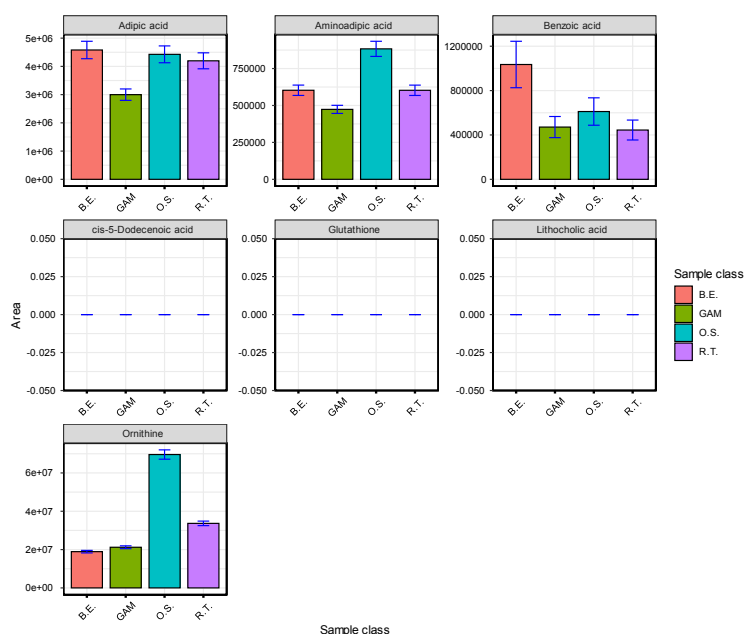
Suppl. Fig. 24: Area of MS-MS measured metabolites in bacterial supernatants. Here, we display all non-SCFA metabolites related to the *C. albicans* growth inhibition experiments. Control samples were created by random pooling of sample material (see main manuscript methods for details). We measured the supernatants of *Bacteroides eggerthii* (B.E.), *Odoribacter splanchnicus* (O.S.) and *Ruminococcus [Blautia] torques* (R.T.). Bile acids (Lithocholic acid, cis-5-Dodecenoic acid) were below detection threshold, implying that these species do either (a) not produce such compounds or (b) possible due to missing components in the growth medium used. RT produced Adipic acid is levels comparable to OS and BE. OS produced higher levels of Aminoadipic acid and BE produced higher levels of Benzoic acid. Ornithine was primarily produced by OS, but also a little by RT.
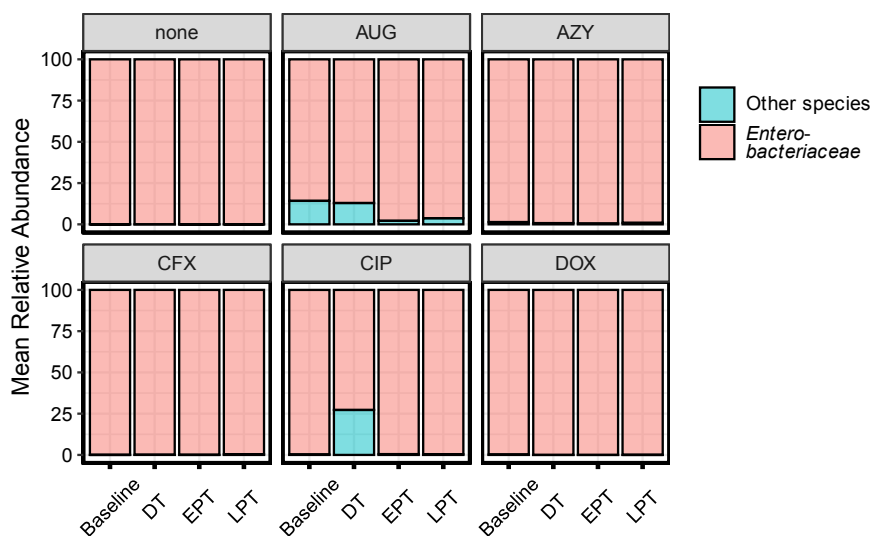
Suppl. Fig. 25: Mean relative abundance of *Enterobacteriaceae* spp. per ATB. We investigated the relative abundance of *Enterobacteriaceae* spp. and the impact of antibiotics [ATB] treatment. In most samples, the accumulated relative abundance of *Enterobacteriaceae* spp. (blue) at baseline was below 1%, implying only a minor role for the overall community. The exception were patients treated with AUG, which had roughly 14% *Enterobacteriaceae*. *Enterobacteriaceae* relative abundance in AUG treated patients decreased profoundly at EPT and also LPT compared to baseline levels. In CIP, we observed a strong increase of up to 27% during treatment, but far below 1% for time points before and after treatment.

IV. Manuscript:
Gut Microbiota of Critically Ill and Antibiotics

FORM 1

Manuscript No. 4

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

Authors: Marfil-Sánchez A., Zhang Lu, Alonso-Pernas Pol, Mirhakkak M., Mueller M., **Seelbinder B.**, Ni Y., Santhanam R., Busch A., Beemelmanns C., Ermolaeva M., Bauer M., Panagiotou G.

Bibliographic information: *Gut Microbes*. 13 (2021), 10.1080/19490976.2021.1993598

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
M.-Sánchez	30%	45%		40%	
Zhang	30%	45%		40%	
A.-Pernas			20%		
Mueller			20%		
Seelbinder		5%		5%	
Busch			20%		
Bauer	20%		20%		60%
Panagiotou	20%			10%	30%
Others		5%	20%	5%	10%
Total:	100%	100%	100%	100%	100%

Overview:

In this follow-up, we used a similar framework to study tri-partite communication between bacteria, fungi, and the human host during critical illness. We elucidated how antibiotic treatment could induce detrimental effects on the health of the patients in addition to the effects caused by infection. Again, we found *Candida* to increase in abundance under antibiotic treatment in critically ill patients. More importantly, we found that antibiotic therapy in critically ill patients leads to an "infection vulnerable" microbiome composition characterized by extremely low levels of short-chain fatty acids. Together with Manuscript III, our studies deliver consistent evidence for the growth of opportunistic fungi such as *Candida* in the human gut during oral antibiotic administration and generated several further hypotheses for control mechanisms of *Candida* by gut microbiota.

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.

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

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
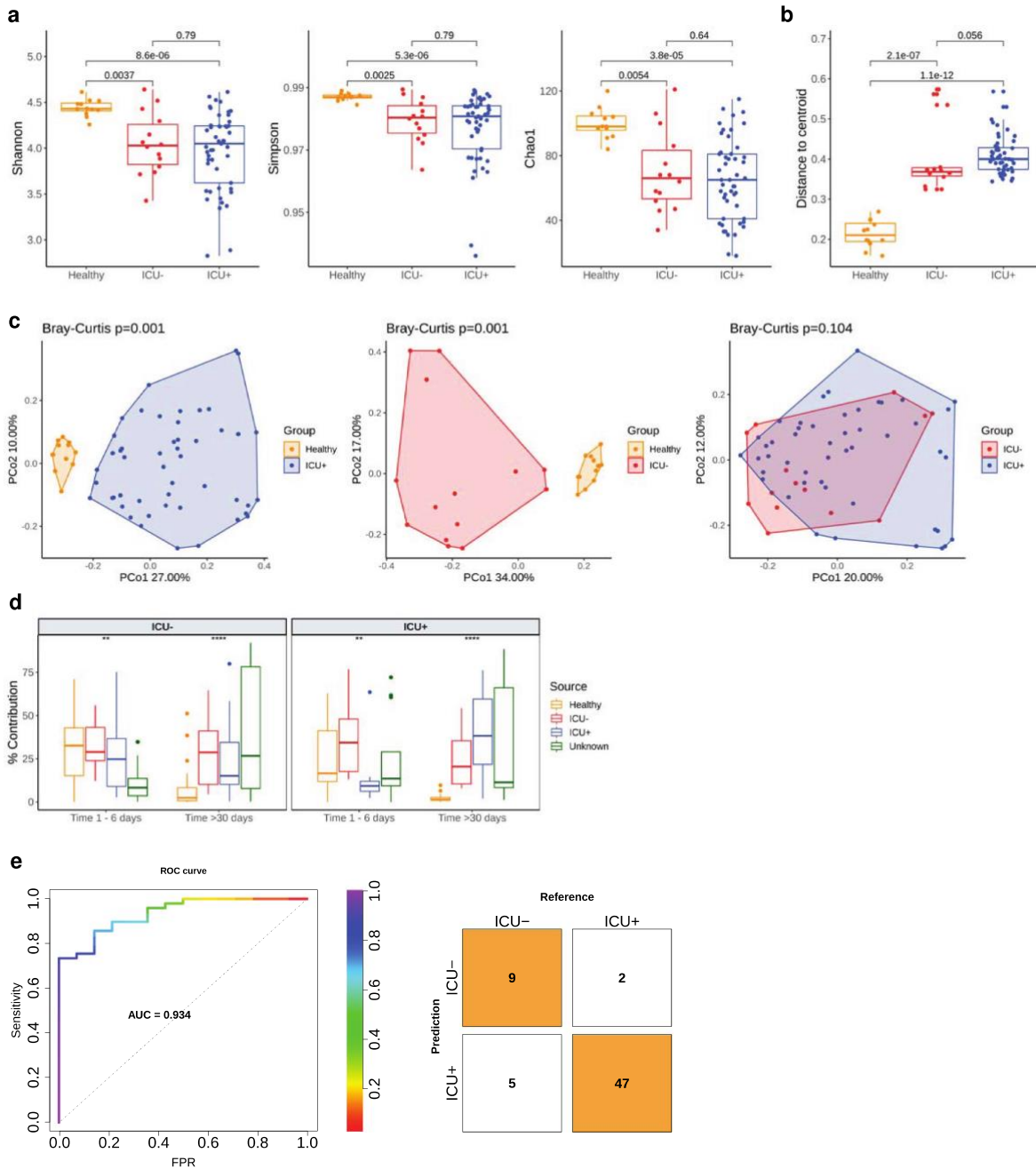
Figure 1. Distinct gut microbiota signatures in ICU patients. (a − b, d) Box plots showing the median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers) measurements. A comparison was considered significant if $P < .05$.(a) Alpha diversity of bacterial species 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.
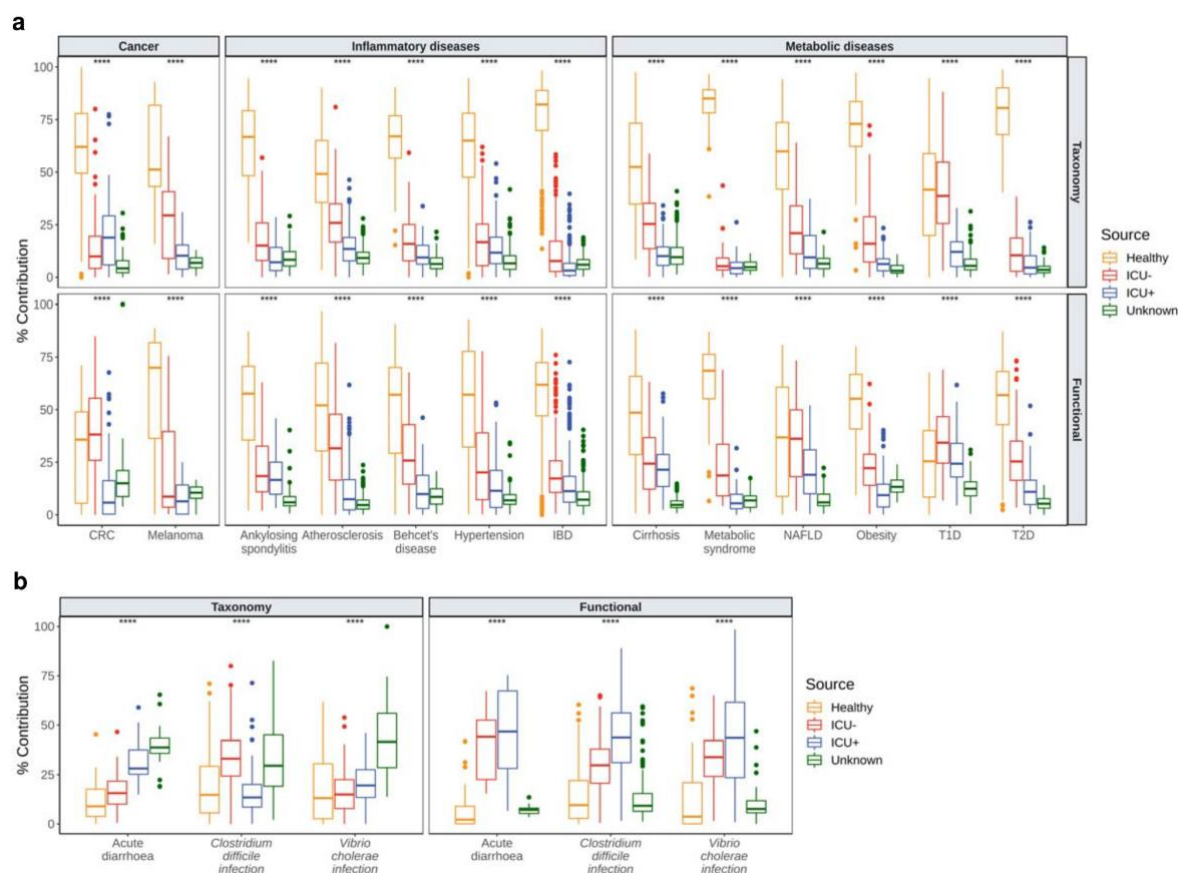
Figure 2. Comparative analysis of the microbiome of critically ill patients with other diseases. (a – b) FEAST [43] estimation of microbial source contribution for each "sink". Here, sinks are taxonomic and functional compositions from a panel of diseases. Source are taxonomic and functional compositions 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
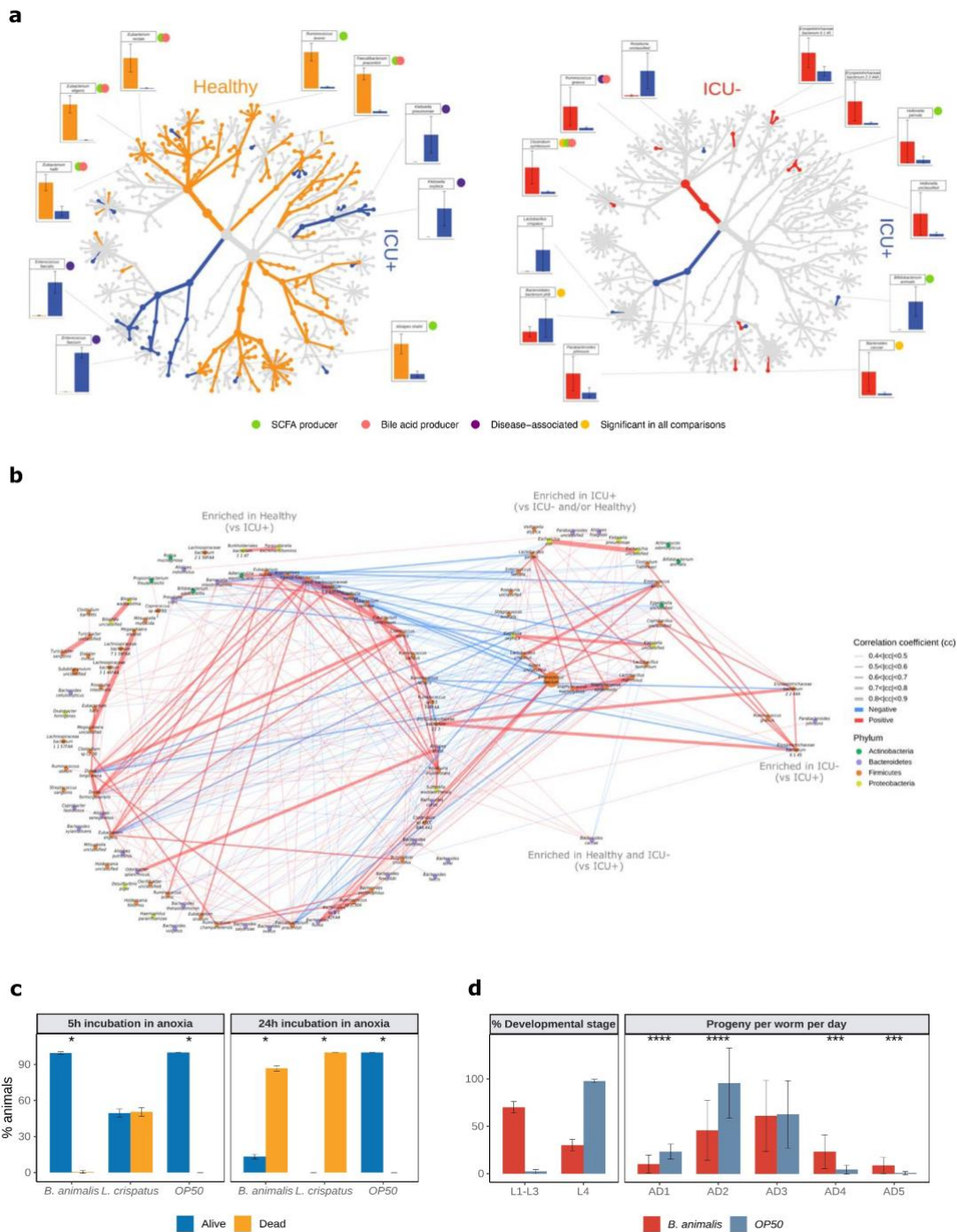
Figure 3. Differential abundance analysis reveals changes in bacteria with important functional properties in critically ill. (a) Taxonomic trees visualized using R package metacoder.[47] Only taxa differentially abundant between Healthy, ICU$^+$ or ICU$^-$ ($P < .05$, Wilcoxon rank-sum test) are highlighted 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-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
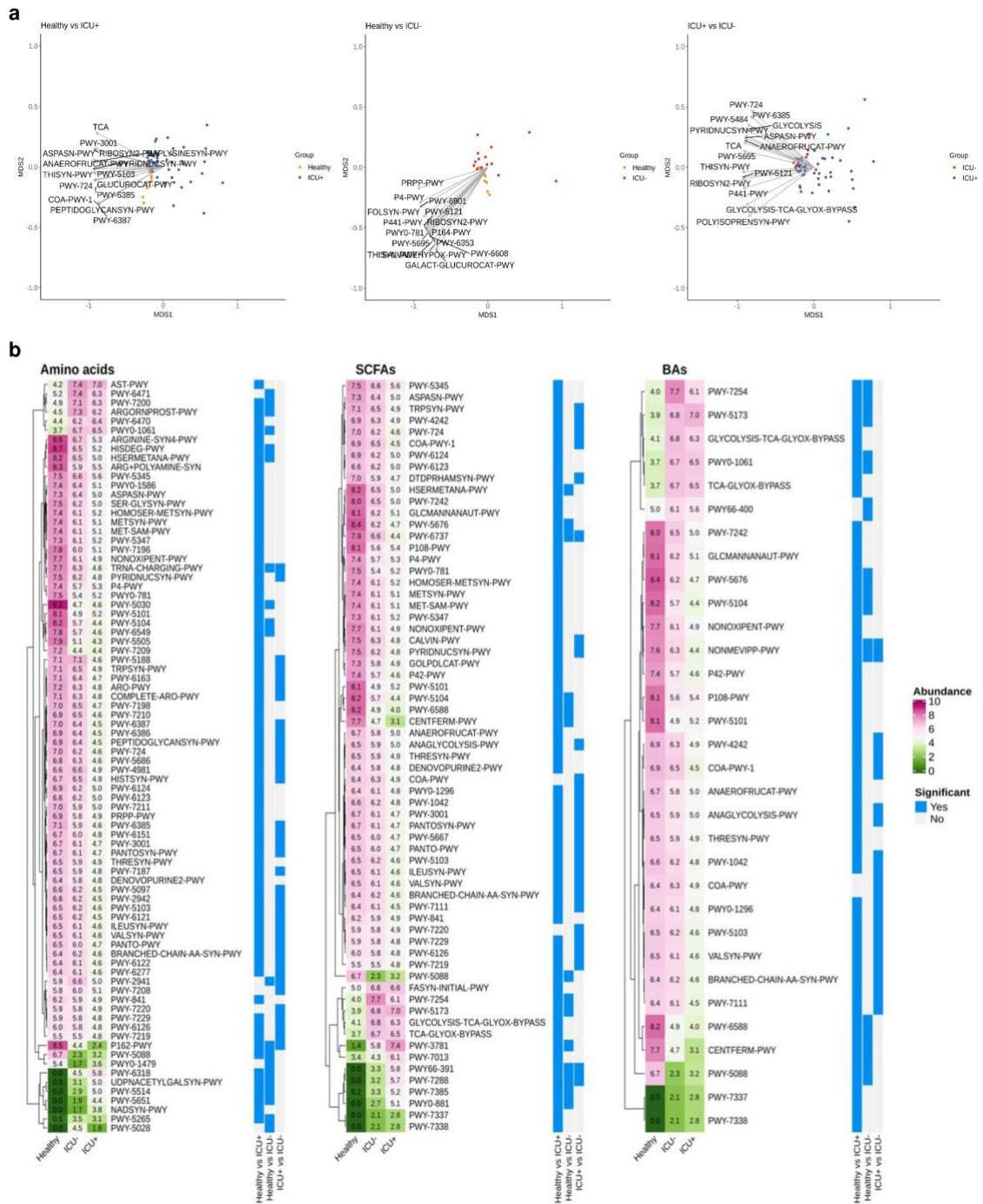
Figure 4. Functional shifts contributed by the microbiome. (a) Multidimensional scaling (MDS) plot of the samples based on the pathway abundances. The top 15 pathways with the strongest significant correlation with the overall ordination (FDR<0.05 with function envfit from the R package vegan [75]) are 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; keto-lithocholic acid, deoxycholic acid, glycolithocholic acid, hyodeoxycholic acid, isolithocholic acid, litho-cholic 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
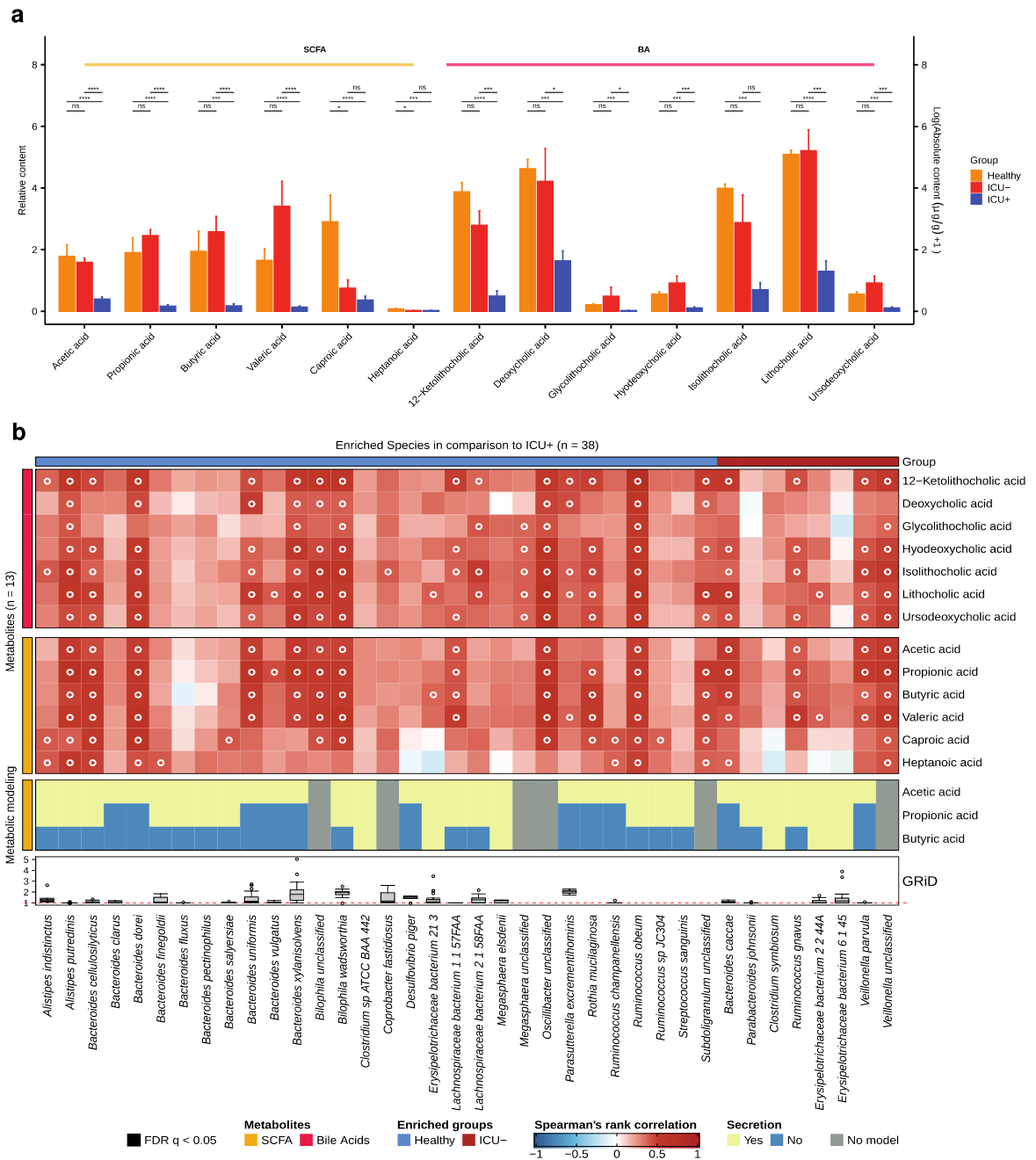
Figure 5. Associations between gut microbiome and short-chain fatty acids and bile acids. (a) Content of SCFAs and BAs (log$_e$) in three groups are plotted as bar plots. Wilcoxon rank-sum test, *P < .05, **P < .01, ***P < .001, ****P < .0001, ns: not significant.(b) Heatmap showing Spearman's correlations 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.

administered 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, $P = 5.97e-8$, 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
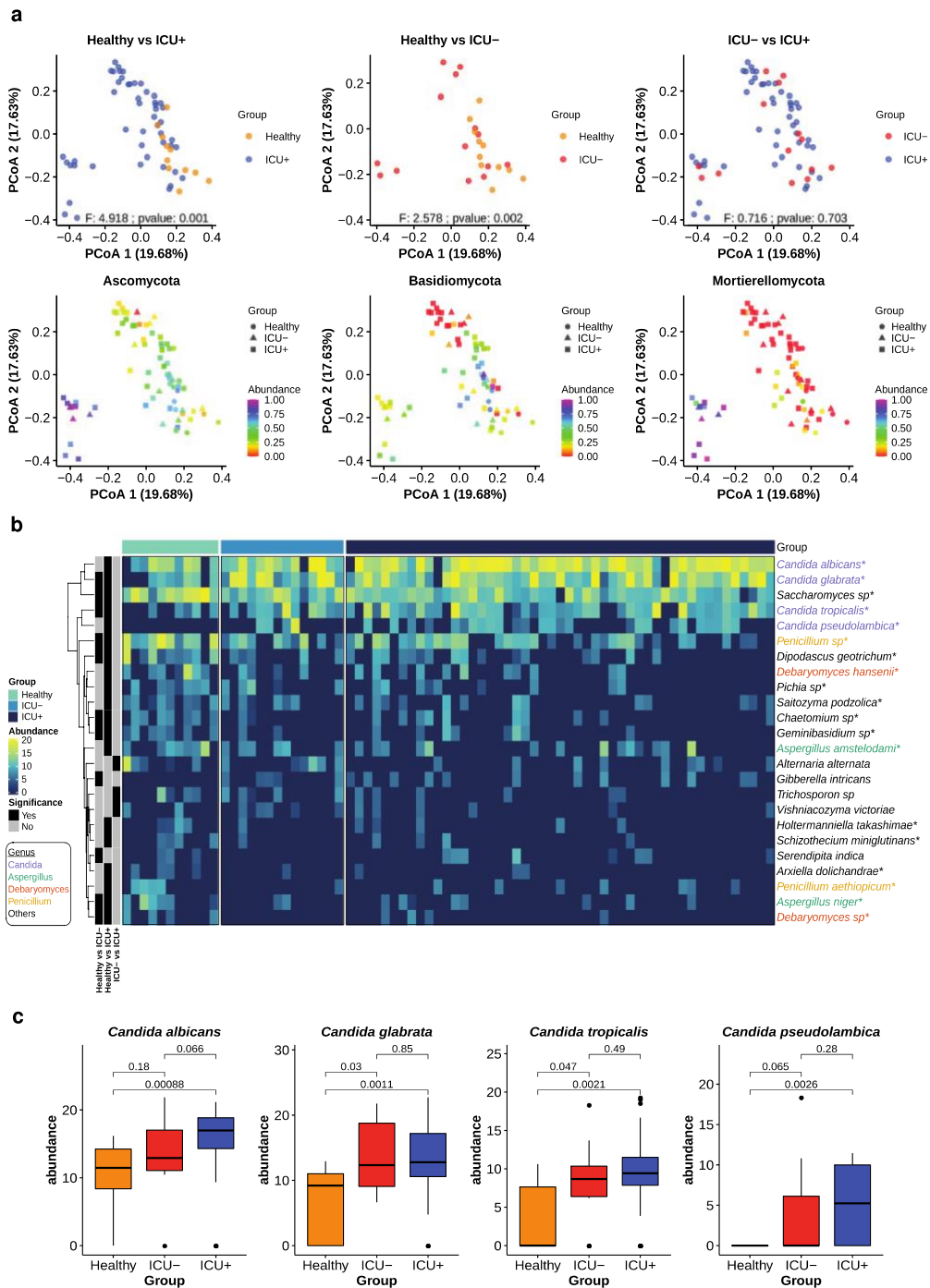
Figure 6. ICU patients have distinct mycobiome profiles. (a) Upper – PCoA plots of the gut mycobiome based on species-level Bray-Curtis distances. Color indicates the groups. Significant differences were determined using PERMANOVA ($P < .05$). Lower – the same PCoA plots as upper panel highlighting the 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,

120

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/ernaeh rung/fresubin-original) was considered. Metabolic composition of complex products such as fish oil was described by vmh diet designer (https://www.vmh.life/#nutrition/dietde signer). 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

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

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 microbiomes. 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 palaeo-geographical 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.

Supplement

Supplementary File 1

Characteristic	ICU+ group, N = 54	ICU- group, N = 16	Healthy group, N = 9	P-value (ICU+ vs ICU-)	P-value (Healthy vs ICU)
male sex	38 (70%)	11 (69%)	5 (56%)	1.0000	0.6197
Age [years]	66 (58, 76)	72 (65, 78)	63 (57, 70)	0.1321	0.1786
BMI [kg/m^2]	27.5 (25.9, 30.1)	26.8 (24.2, 29.1)	25.2 (21.9, 26.5)	0.3088	0.0705
type of admission					
medical	20 (37%)	7 (44%)		0.8476	
elective surgery	19 (35%)	6 (38%)		1.0000	
surgical emergency	15 (28%)	3 (19%)		0.6891	
type of surgery					
none	20 (37%)	7 (44%)		0.8476	
cardiac surgery	17 (31%)	6 (38%)		0.8830	
neurosurgery	8 (15%)	2 (12%)		1.0000	
abdominal surgery	5 (9.3%)	0 (0%)		0.4774	
trauma surgery	1 (1.9%)	0 (0%)		1.0000	
other	3 (5.6%)	1 (6.2%)		1.0000	
COPD	6 (11%)	2 (12%)		1.0000	
cardiovascular disease	35 (65%)	10 (62%)		1.0000	
renal disease	13 (24%)	3 (19%)		0.9152	
diabetes mellitus	17 (31%)	8 (50%)		0.2888	
focus of infection					
none	0 (0%)	16 (100%)		9.93E-16	
respiratory tract	37 (69%)	0 (0%)		5.70E-06	
Abdominal	6 (11%)	0 (0%)		0.3756	
bones/soft tissue	3 (5.6%)	0 (0%)		0.7941	
Chest	2 (3.7%)	0 (0%)		1.0000	
catheter associated infection	1 (1.9%)	0 (0%)		1.0000	
urogenital	1 (1.9%)	0 (0%)		1.0000	
unknown	4 (7.4%)	0 (0%)		0.6114	
origin of infection					
none	0 (0%)	16 (100%)		9.93E-16	
nosocomial (ICU)	40 (74%)	0 (0%)		6.66E-07	
nosocomial (non-ICU)	9 (17%)	0 (0%)		0.1855	
community acquired	5 (9.3%)	0 (0%)		0.4774	
antibiotics					
piperacillin/tazobactam	32 (59%)	0 (0%)		0.0001	
meropenem	22 (41%)	0 (0%)		0.0055	
none	0 (0%)	16 (100%)		9.93E-16	
mechanical ventilation	31 (57%)	6 (38%)		0.2644	
vasopressor therapy	37 (69%)	5 (31%)		0.0172	
dialysis	18 (33%)	0 (0%)		0.0186	
ICU length of stay [days]	16 (10, 27)	21 (14, 29)		0.3073	
ICU mortality	14 (26%)	2 (12%)		0.4328	
hospital length of stay [days]	31 (24, 45)	29 (19, 47)		0.7873	
hospital mortality	16 (30%)	2 (12%)		0.2931	

Continuous data are expressed as median and interquartile range. BMI: body mass index, COPD: chronic obstructive disease, ICU: intensive care unit. Continuous data were compared by the t-test, dichotomous variables by the chi-squared test.

Supplementary Fig. 1
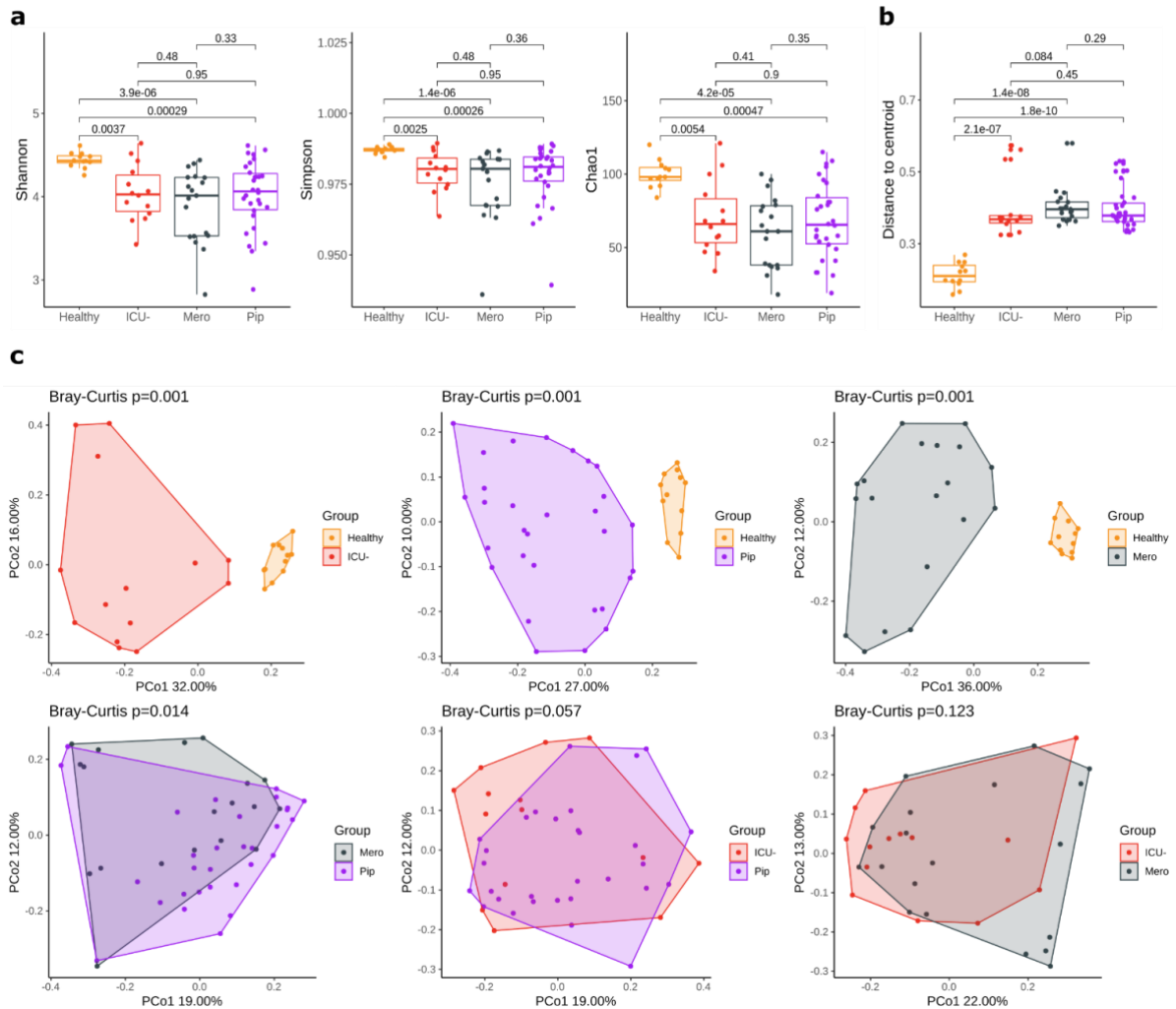
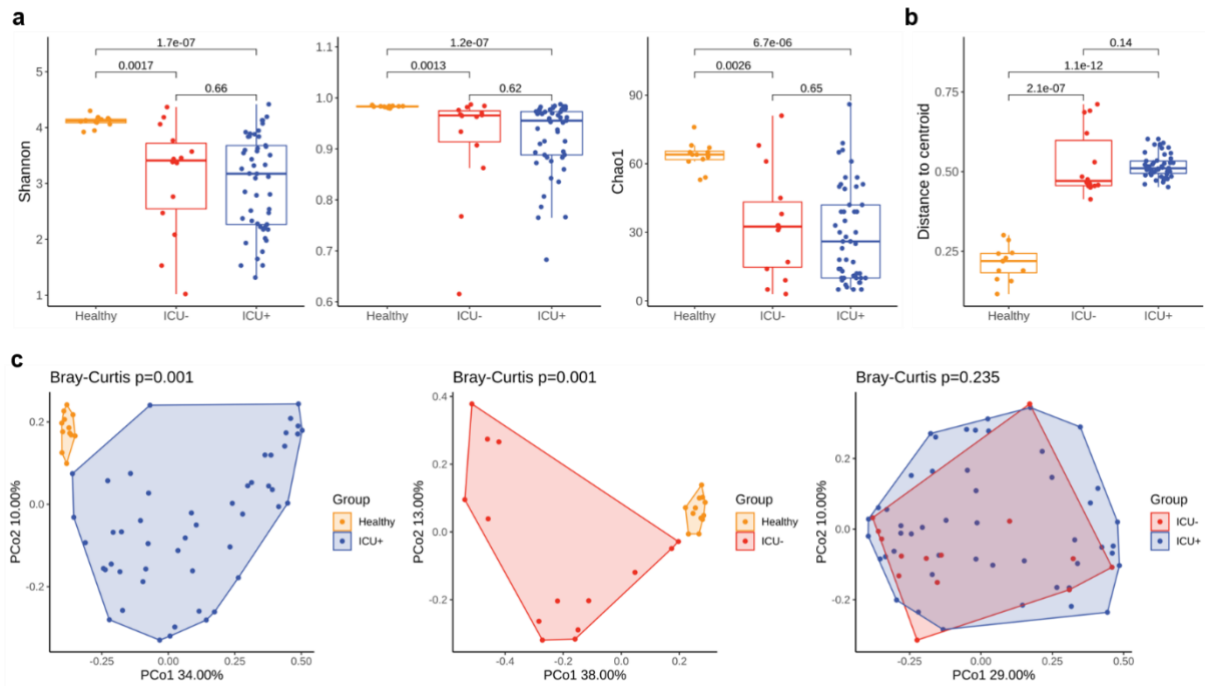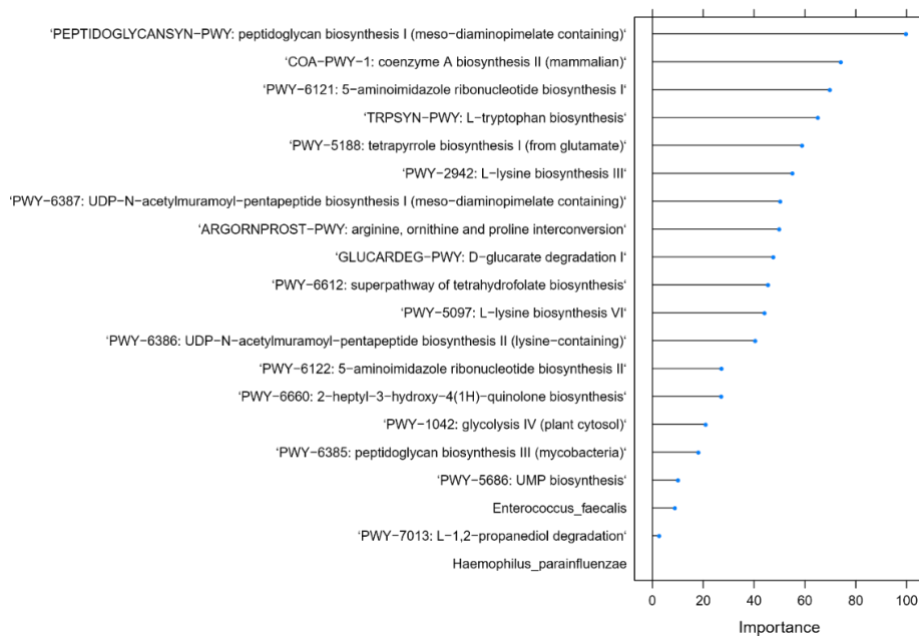**Figure S1. Distinct gut microbiota composition in ICU patients**

(A – B) Box plots showing the median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers) measurements. Significant differences were determined using Wilcoxon rank-sum test. A comparison was considered significant if P<0.05.

(A) Alpha diversity of bacterial species using Shannon (left), Simpson (middle), and Chao1 (right) indices.

(B) Beta dispersion of bacterial species measured as the distance of the samples from one group to the group *centroid* in multivariate space.

(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<0.05.

Supplementary Fig. 2

Figure S2. Distinct gut microbiota composition in ICU patients

(A - B) Box plots showing the median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers) measurements. Significant differences were determined using Wilcoxon rank-sum test. A comparison was considered significant if P<0.05.

(A) Alpha diversity of Metagenomic Species using Shannon (left), Simpson (middle), and Chao1 (right) indices.

(B) Beta dispersion of Metagenomic Species measured as the distance of the samples from one group to the group *centroid* in multivariate space.

(C) Principal component analysis (PCoA) of Bray-Curtis dissimilarity between Metagenomic Species abundance profiles. Significant differences were determined using PERMANOVA and were considered significant if P<0.05.

Supplementary Fig. 3

Figure S3. Most important features for ICU⁺ and ICU⁻ classification. Plot showing the top 20 most important features at the taxonomic and functional level for the classification of ICU⁺ and ICU⁻, based on the Random Forest classifier.

Supplementary Fig. 4
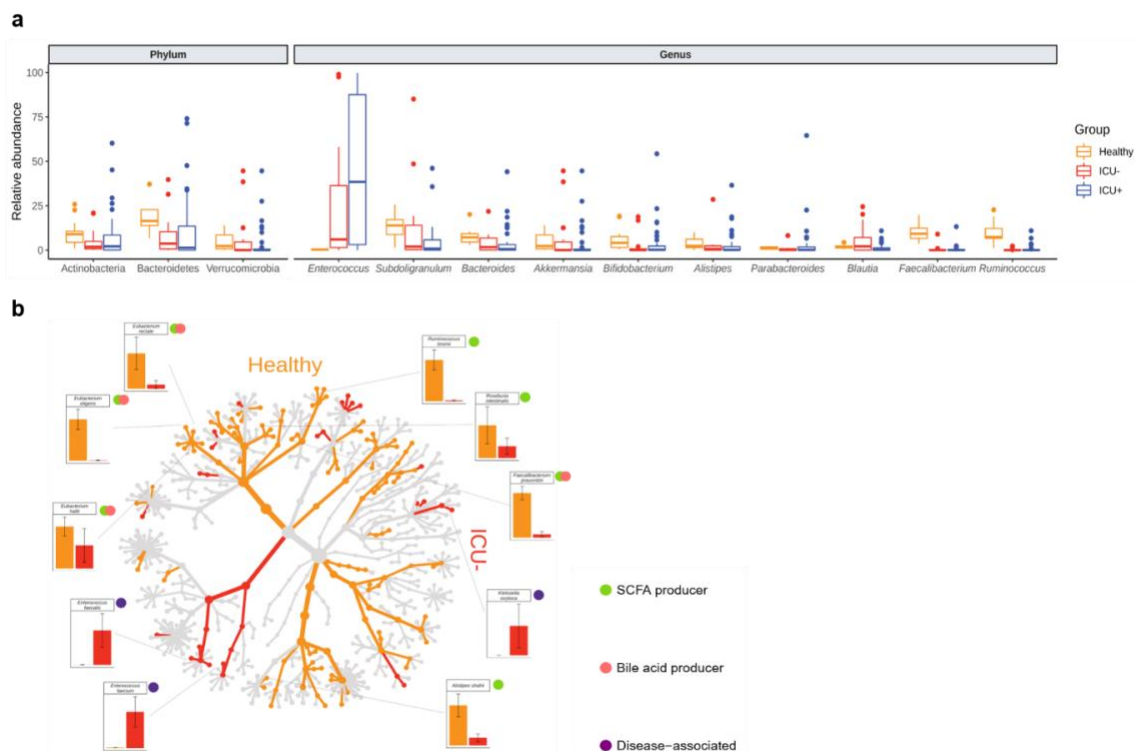
Figure S4. ICU patients have a distinct taxonomic composition
(A) Box plots showing the median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), of the relative abundance of significantly different phyla and the most

abundant significant genera. Significant differences were determined using Kruskal-Wallis test. A comparison was considered significant if P<0.05.

(B) Taxonomic tree visualized using R package metacoder [1]. Only taxa differentially abundant between Healthy and ICU⁻ (P<0.05, 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) or disease-associated species (purple circles).

Supplementary Fig. 5
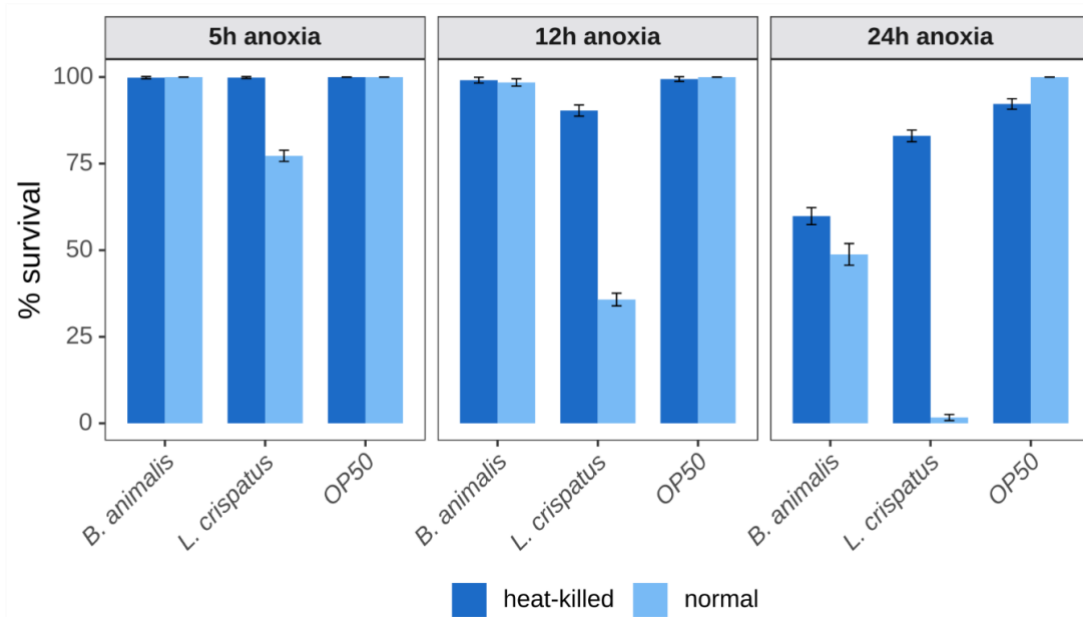
Figure S5. The distinct impact of live bacteria on the host toxicity of L. cripatus and B. animalis. Germ-free L1 larval stage C. elegans worms were populated with depicted bacterial strains (live and heat-killed) in the anoxic chamber for indicated times, followed by transfer to normoxia and UV-killed OP50 E. coli diet. Host survival was measured after 24 h of normoxic culture.

Supplementary Fig. 6
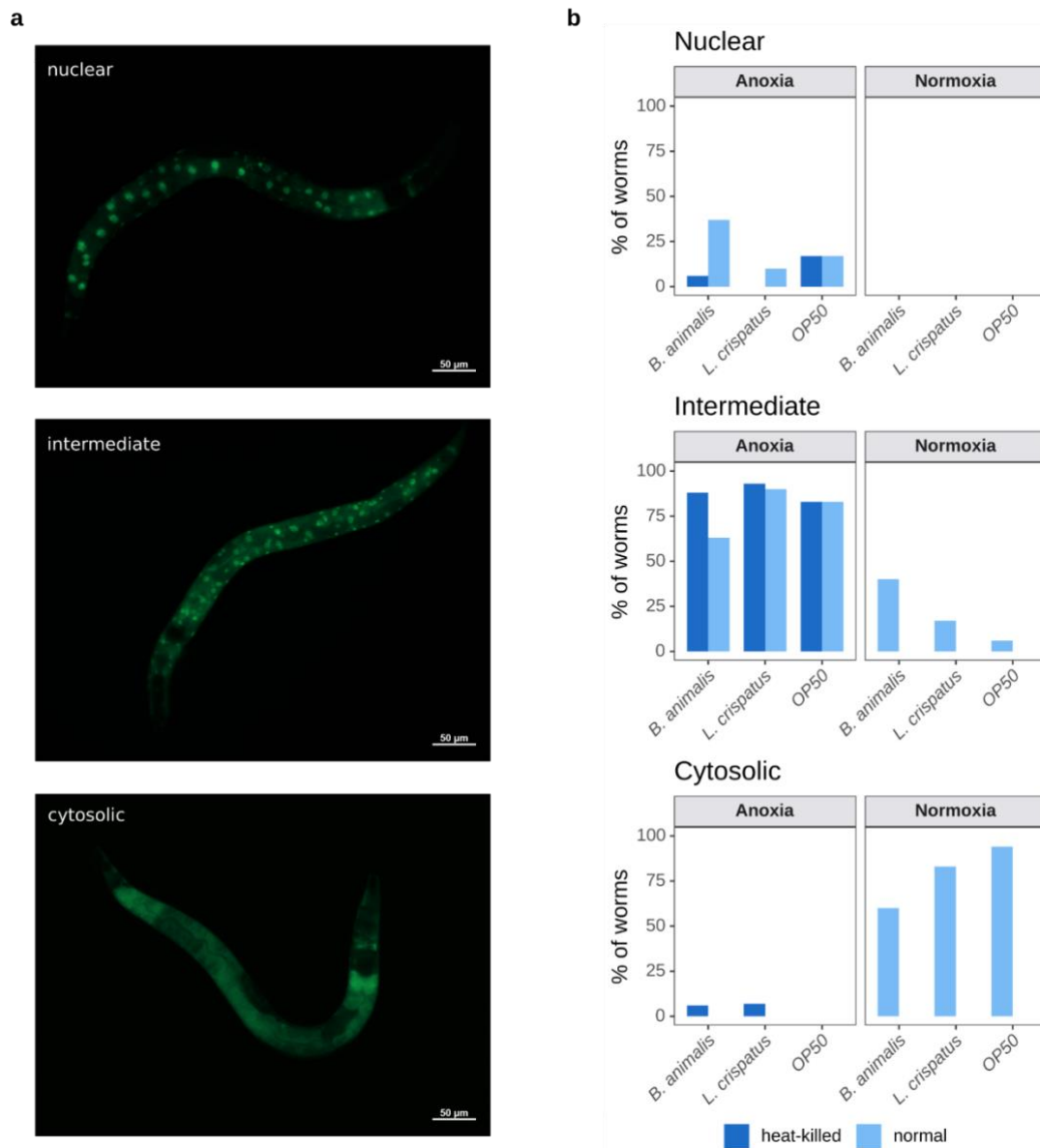

Figure S6. Live B. animalis is the strongest inducer of the protective DAF-16/FOXO nuclear translocation in the host.
Age-synchronized germ-free transgenic worms were grown until L4 stage on UV-killed OP50. L4 worms were populated with depicted bacterial strains (live and heat-killed) and incubated for 5h in the anoxic chamber. DAF-16::GFP localization was assessed microscopically with 30 animals measured for each condition. Representative images of three localization types are shown in (A) and quantification is provided in (B).

Supplementary Fig. 7
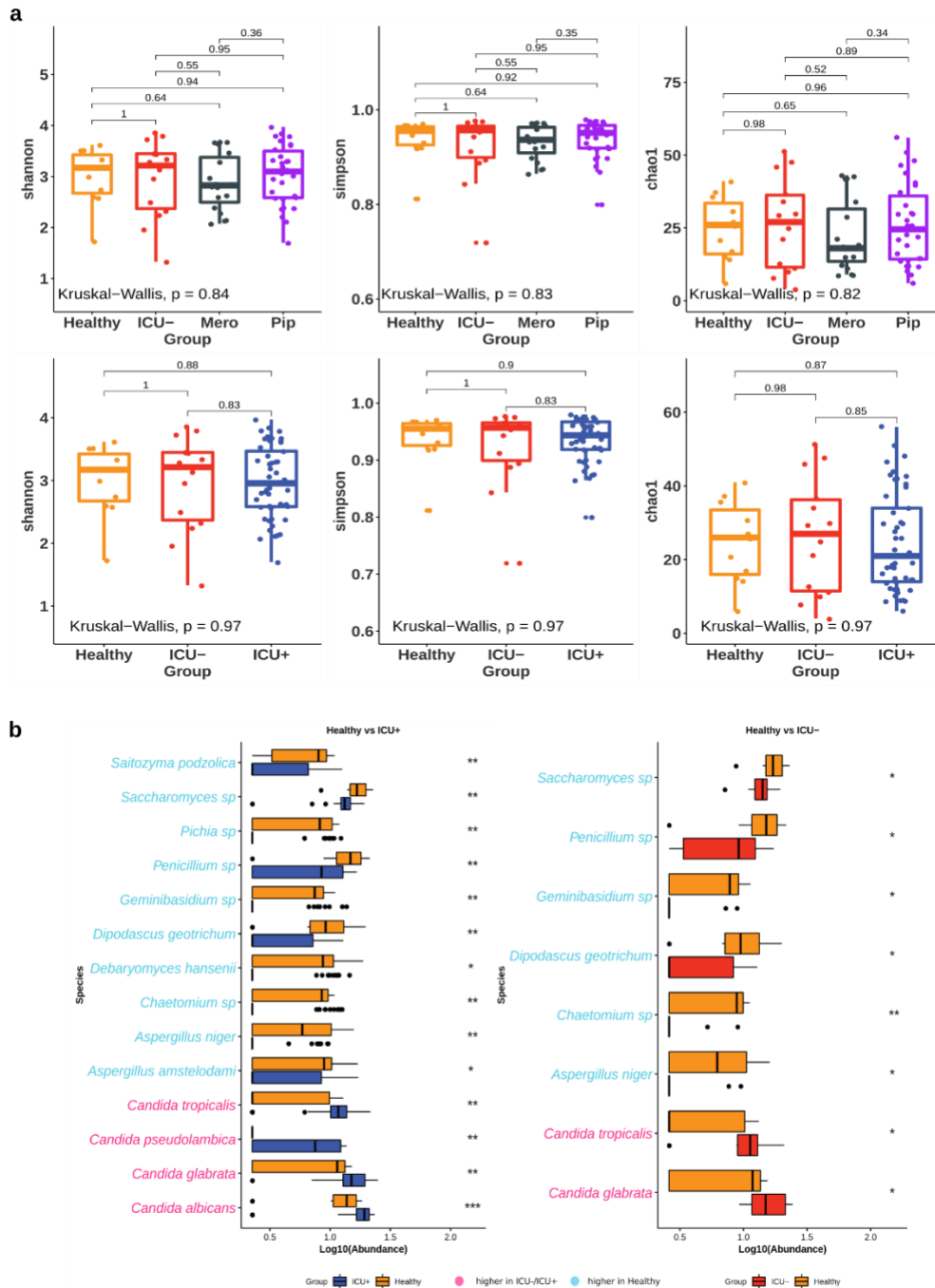
Figure S7: Comparison of gut mycobiome alpha diversity and species abundances.
(A) Box plots, with median (centerlines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), showing mycobiome alpha diversity (indices of Shannon, Simpson, and Chao1) in different categories. Accordingly, significant differences were measured using the Kruskal-Wallis or Wilcoxon rank-sum test.
(B) Comparison of fungal species' abundance. Significant differences were measured using Wilcoxon rank-sum test (*P<0.05; **P<0.01, ***P<0.001). The color of species name indicates whether its median abundance is higher in the ICU⁻/ICU⁺ group (pink) or the Healthy group (blue)

Supplementary Fig. 8
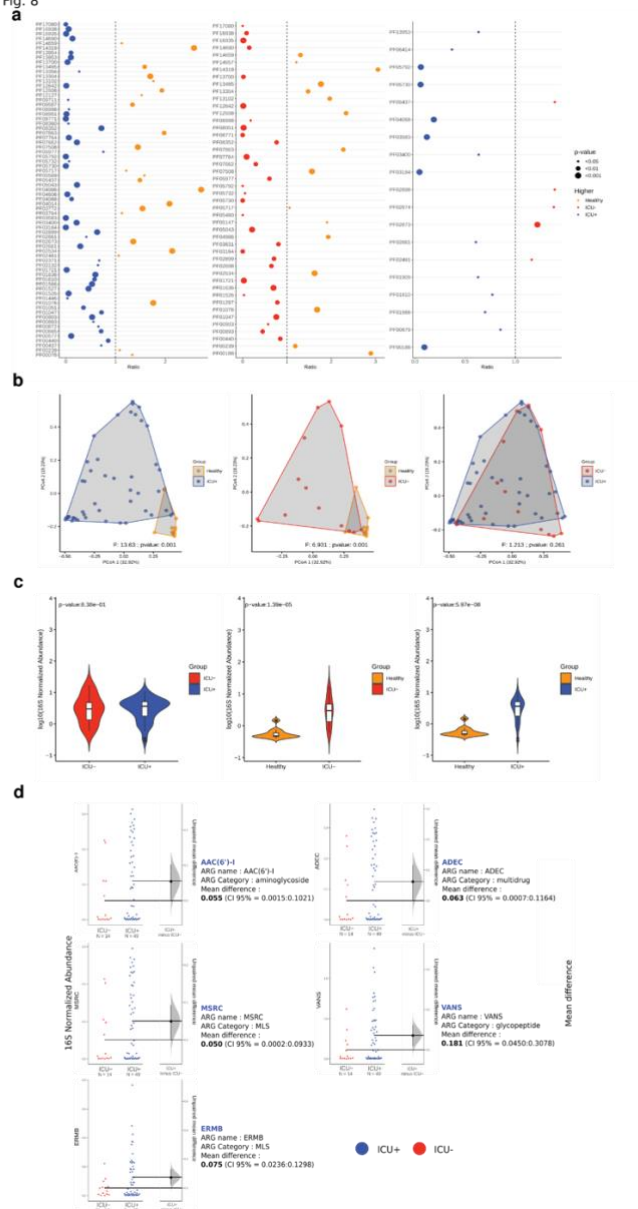
Figure S8: ICU patients harbor a very different resistome compared to healthy individuals.
(A) Dot plot showing the differentially abundant bacterial Pfams (P<0.05, Wilcoxon rank-sum test) related to the resistome and/or mobilome. The size of the dots represents the strength of statistical significance. The color of the dots reflects the group with the higher abundance.
(B) PCoA plots based on Bray-Curtis distances for antibiotic resistance genes (ARGs) in each pairwise comparison. Significant differences were determined using PERMANOVA.
(C) Comparison between the 3 groups of total cumulative relative ARG abundance for ARGs (Wilcoxon rank-sum test). In the corresponding samples for individuals from different classes (ICU$^+$, ICU$^-$ and Healthy), the relative abundance of ARGs was standardized by the content of 16S rRNA.
(D) Representative significantly different ARGs between ICU$^+$ and ICU$^-$ (dabestr, 95% confidence interval (95% CI)). ARGs with the top difference in ICU$^-$ vs ICU$^+$ comparisons are included. All data points are plotted. The mean difference (the effect size) and its 95% confidence interval are displayed as a point estimate and vertical bar respectively, on a separate but aligned axis. Adjacent to the plots are thorough annotations of the ARG term, category, mean difference and its 95% confidence interval.
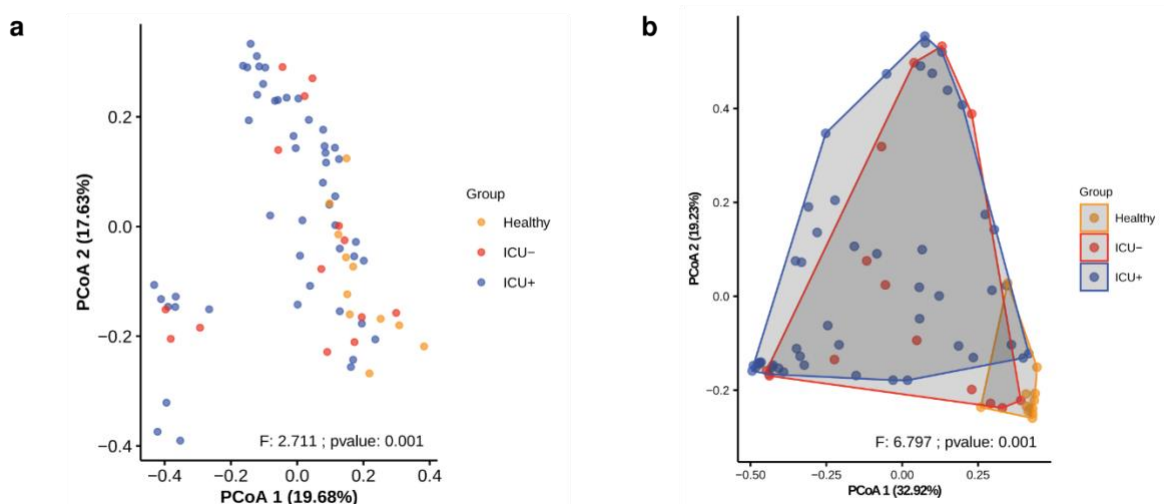
Supplementary Fig. 9
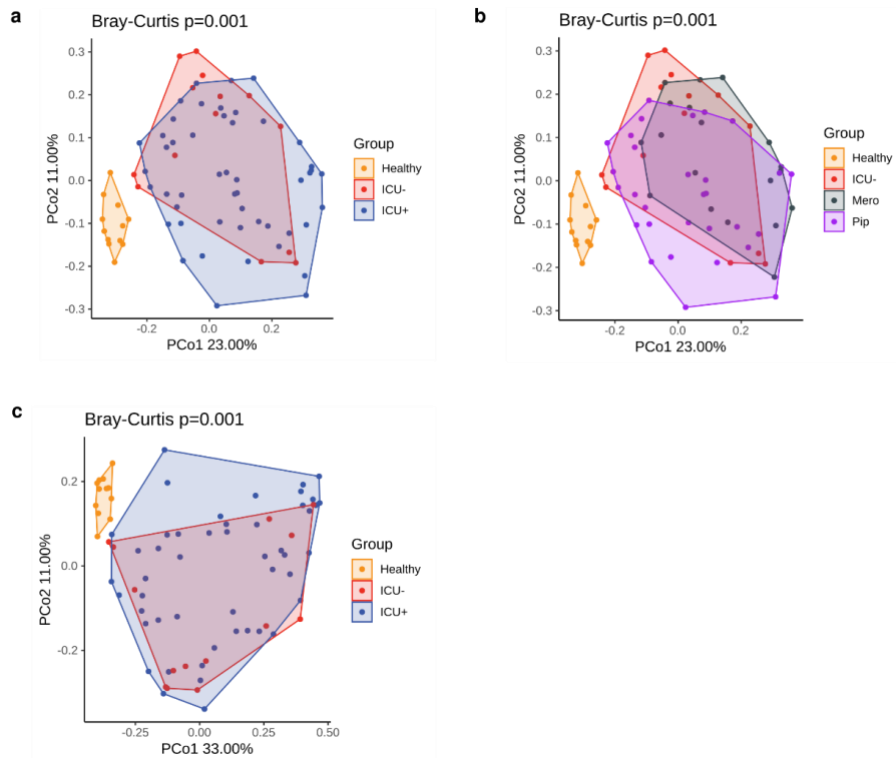
Figure S9. Distinct gut microbiota signatures in ICU patients.
Principal Component Analysis (PcoA) of Bray-Curtis dissimilarity between (A-B) bacterial species abundance profiles, and (C) Metagenomic Species (MGS) abundance profiles. Significant differences were determined using PERMANOVA and were considered significant if P<0.05.

Supplementary Fig. 10

Figure S10. Distinct mycobiome and ARGs in ICU patients.
Principal Component Analysis (PcoA) of Bray-Curtis dissimilarity between (A) gut mycobiome species abundance profiles, and (B) ARGs. Significant differences were determined using PERMANOVA and were considered significant if P<0.05.

V.   Manuscript:
Gut Microbiota control commensal *Candida*

FORM 1

Manuscript No. 5

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

Authors: Seelbinder B., Lohinai Z., Vazquez-Uribe R., Chen X., Lopez S., Mirhakkak M., Wellejus A., Weiss G., Sommer M., Panagiotou G.

Bibliographic information:

The candidate is

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

Status: *in preparation to be submitted*

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

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

Overview:

We performed metagenomics and ITS sequencing on stool samples from lung cancer patients undergoing anti-cancer immunotherapy. Through machine learning, I identified a robust microbial signature explaining varying levels of *Candida* abundance. We also found a non-trivial connection between decreased anaerobes, *Lactobacillus*, gut oxygen levels, and *Candida* abundance levels. Finally, through experiments conducted by our collaborators, we deliver further evidence for a competitive advantage of *Candida* over other fungal residents in the human gut due to alternative carbon sources present under dysbiotic conditions characterized by increased oxygen availability.

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.)

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.

1

Introduction

41

42 *Candida* species, predominantly *C. albicans, C. glabrata, C. tropicalis,* and *C. parapsilosis,* are among
43 the most common causes of bloodstream infections. These infections result in high rates of mortality
44 for patients in intensive care units or who have a dysfunctional epithelial barrier or compromised
45 immunity [2,3]. Although the pathogenicity of different *Candida* species has been extensively studied
46 [4–9], only a handful of studies focused on understanding the commensal lifestyle of the fungus in the
47 human gut. We recently performed a systematic evaluation of the interactions between human gut
48 bacteria and *C. albicans* using genome-scale modelling and pairwise growth simulations [10]. We
49 showed that 81% of *C. albicans* interactions with approximately 900 gut bacteria species were
50 mutualistic (positive growth effects for both *C. albicans* and bacteria) or parasitic (negative growth
51 effect on *C. albicans*, positive growth effect on bacteria), with only a few examples of parasitism, in
52 which *C. albicans* exerted negative effects on gut bacteria. *C. albicans* is a common gut member in
53 the majority of the human population. Therefore, our findings support the hypothesis that the
54 colonisation success of *C. albicans* is the result of adapting to life in the intestine and avoiding
55 competitive interactions with other gut microbes.

56 Most attempts to identify specific species of gut bacteria that inhibit or promote the growth of
57 *Candida* species were conducted in murine models. However, in contrast to human gut communities,
58 adult murine gut communities naturally prevent *Candida* species colonisation [11], showing
59 substantial differences with humans in immune system regulation [12] and microbial composition [13–
60 15], and challenging the translatability of the findings to humans. Gnotobiotic mouse models overcome
61 some of these challenges [16] and were applied to study colonisation by *Candida* species [17]. Yet,
62 limitations persist. A recent study demonstrated that colonisation is still incomplete, and some key
63 bacteria genera (*Faecalibacterium, Bifidobacterium*) did not engraft at all [18]. Nevertheless, the
64 realisation that the gastrointestinal tract is a major source of systemic candidiasis [1] has propelled
65 efforts beyond animal models to identify predisposing factors that may lead to microbiome engineering
66 strategies aimed at preventing candidiasis. This shift in focus has been further supported by evidence
67 from human studies that gut bacterial dysbiosis triggered by broad-spectrum antibiotics is associated
68 with increased colonisation of *Candida* species in the gut by [19]. However, antibiotics are not the
69 only drugs associated with an elevated risk of *Candida* species overgrowth. Initial findings in animal
70 models and humans suggest that chemotherapeutic agents lead to a reduced total number of gut bacteria
71 and alterations in gut microbiota composition [20–22], which may contribute to the increased risk of
72 systemic candidiasis in cancer patients. While most studies on systemic candidiasis and cancer have
73 focused on haematological malignancies, recent epidemiological studies suggest that the risk for
74 patients with solid tumours, such as head and neck and lung cancer, is equally high [23].

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

2

144

86   individuals. We further provided experimental evidence for a competitive advantage of *Candida* over
87   *Saccharomyces* species*,* the other main fungal residents in the human gut, while exploring alternative
88   carbon sources under dysbiotic conditions characterised by increased oxygen availability.

89

Results

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

92   We recruited 75 patients at the National Koranyi Institute of Pulmonology (Budapest, Hungary) and
93   County Hospital of Pulmonology (Torokbalint, Hungary) with advanced-stage lung cancer (Table 1;
94   adenocarcinoma n=40, squamous cell carcinoma n=28, others n=7). No patients had any signs of
95   fungal infection during the recruitment period. Faecal samples were collected for analysis of fungal,
96   bacterial, and viral biomes after the initiation of single-agent anti-PD-1 antibody immunotherapy
97   (nivolumab, n=44; pembrolizumab, n=31). The majority of patients received chemotherapy prior to
98   immunotherapy (n=59). We built ribosomal DNA internal transcribed spacer 2 (ITS2) libraries for
99   estimating fungal genera and species relative abundance in all 75 patients. On average, we generated
100   78,332 (mean absolute deviation [MAD] 43,056) high-quality, non-chimeric reads per sample.
101   Amplicon sequencing variants (ASVs) were estimated using DADA2 [27] resulting in 76 fungal
102   genera (18 ± 10 per sample) and 113 fungal species (25 ± 15 per sample) (Figure 1a). Investigating
103   genus-level fungal profiles showed that *Candida* and *Saccharomyces* were the highest contributing
104   genera in 17 and 44 samples, respectively (Figure 1b). Fungal co-abundance networks revealed a
105   strong, significant negative correlation between *Candida* and *Saccharomyces* at the genus (Spearman's
106   coefficient; $P<0.05$; $|r|>0.25$; Figure S1) and species level (Spearman's; $P<0.05$; $|r|>0.25$; Figure S2).

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

119   We compared different mycobiome normalisation methods and observed high correlations between
120   the normalised abundance estimates of the *Candida* genus (Pearson $r \geq 0.87$; $P<0.001$; Figure S3). To
121   properly account for compositional data, all downstream analyses used fungal abundances normalised
122   by the centred log-ratio (CLR) [28–30]. For *Candida* CLR abundances, the median separated *Candida*
123   abundance symmetrically (Figure S3). Therefore, we grouped patients in two clusters: high-*Candida*
124   (**HC**, n=38) and low-*Candida* (**LC**, n=37) for above or below the median *Candida* CLR normalised
125   abundance (Figure S3). This grouping correlated significantly with high and low *Candida* ITS reads
126   (Rfit $P<0.001$; Figure 1f), but not with the number of total ITS copies (Rfit $P>0.05$; Figure 1g)
127   indicating that the sequencing depth did not affect the grouping. We further confirmed the grouping
128   by testing for significant differences in fungal relative abundance between the two groups (Table S2).

3

129 As expected, *Candida* genus abundance was increased in the **HC** group (log$_2$ fold change
130 [log$_2$FC]=5.1; *q*=2E-12), whereas at the species level, *C. albicans* (log$_2$FC=5.5; *q*=1E-7) and
131 *C. tropicalis* (log$_2$FC=2.1; *q*=0.04) drove the observed genus abundance differences.

132 Although the beta diversity was significantly different between the **HC** and **LC** groups
133 (PERMANOVA *P*=0.001; *R*^{2}=5.7%; Figure 1h), we did not find significant differences in the fungal
134 genera and species alpha diversities (by Shannon, Simpson, Pilou's Evenness indices; *P*>0.05;
135 Table S1). We further examined if the classification of patients into **HC** and **LC** groups was explained
136 by differences in basic patient characteristics such as gender, age, BMI, antibiotic use, alcohol
137 consumption, tumour histology, chronic obstructive pulmonary disease, or anti-cancer treatment drug.
138 None of these factors except for BMI was significantly different between the groups. Despite recent
139 findings in mice and humans that *Candida* species may promote weight gain [31,32], we found BMI
140 significantly decreased in our cohort in the **HC** group (Table 1; U-Test *P*=0.006). Therefore, in
141 subsequent statistical comparisons between the **HC** and **LC** groups, we adjusted for differences in
142 BMI.

143

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	14		Never		19	16	
Male			21	23		Current		3	6	
Age		75	69	65	0.2	Former		2	4	
Body Mass Index		75	24.1 (21.4,	28.7 (25.9,	0.006	Occasionally		14 (37%)	11 (30%)	
Antibiotic Use Before therapy	Use	74			0.3	**Histology**	75			0.7
No			34	30		Adenocarcinoma		22	18	
Yes			3	7		Squamous		13	15	
Unknown			1	0		Other		3	4	
Antibiotic Use After therapy	Use	72			0.14	**Immunotherapy Drug**	75			0.7
No			28	32		nivolumab		21	23	
Yes			9	3		pembrolizumab		17	14	
Unknown			1	2		**Responder**	75			0.3
Line of Treatment		75			0.8	Yes		33	28	
1			9	7		No		5	9	
2			19	23		**COPD**	75			0.7
3			9	6		Without		16	18	
4			1	1		With		22	19	

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

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

144

Distinct microbiome signature associated with the high *Candida* group

146 We then shifted our focus on the gut bacterial community, specifically taxonomic and functional
147 properties that might be pivotal in supporting the growth of *Candida* species in the human gut. We
148 performed whole-metagenomic sequencing on the same stool samples used for the mycobiome
149 analysis, generating an average of 26,106,952 (MAD 3,876,437) reads per sample. We used
150 HUMAnN2 [33] to compute bacterial species and function abundance profiles. After applying a 10%
151 prevalence filter, we estimated the relative abundance of 234 bacterial species (98 ± 18 per sample),
152 84 bacterial genera (46 ± 7), 394 MetaCyc pathways (339 ± 21), 1688 Enzyme Commission (EC)

146

153   numbers (1361 ± 83), 5120 KEGG orthology (KO) terms (3943 ± 340), and 155 KEGG pathways (120
154   ± 3) (Figure 1a). To ensure that only bacterial information was used in further analyses, we removed
155   features with unknown or non-bacterial origin from functional abundance profiles using the species-
156   stratified output of HUMAnN2.

157   Procrustes analysis revealed a significant correlation between beta diversity for fungal and bacterial
158   species (P=0.046, r=0.53; Table S1). Bacterial species explained around 13% of fungal species beta
159   diversity and dbRDA using bacterial species suggested *Clostridium*, *Lactobacillus*, *Eubacterium*, and
160   *Citrobacter* species had the highest explanatory power for mycobiome variation (Figure 1d).
161   Interestingly, a biplot of bacteria species abundances onto fungal species diversity indicated positive
162   correlations between higher *Candida* genus abundance and several *Lactobacillus* species (Figure 1e),
163   which was further confirmed in a crosskingdom network of fungal genera and bacterial species
164   (Spearman's P<0.05; r < -0.25; Figure S2). In contrast, *Citrobacter* and *Eubacterium* species
165   correlated negatively with *Candida* abundance.

166   We also observed significant separation between the **HC** and **LC** groups in bacterial species and
167   functional beta diversity (Figure 2a-c; PERMANOVA; P<0.05; Table S1). For explaining variance,
168   the functional properties of the bacteria community (MetaCyc, KOs, and ECs) showed the largest
169   between-group differences (Figure S1, R^2=2.5%±0.5%). Bacterial species alpha diversity (by
170   Shannon, Simpson, Pilou's Evenness, and Chao1 indices) was not significantly different between the
171   **HC** and **LC** groups (Rfit P>0.05; Table S1). Surprisingly, functional alpha diversity (by MetaCyc
172   pathways) showed higher diversity in the **HC** than the **LC** group (Figure 2e; P=0.02). In contrast, the
173   contributional diversity of bacterial species to pathways (contributional alpha diversity) was
174   significantly lower in many **HC** enriched pathways (Rfit; n=23; P<0.05; Figure 2f) and significantly
175   lower overall (Rfit; P<0.05; Figure 2g). Together, these findings implied that the bacterial community
176   in the **HC** group had greater metabolic potential, but the **LC** group had greater functional redundancy,
177   a property exhibited by robust microbiota [34].

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

187   To complement our study on the ecological context associated with *Candida* species expansion in
188   the human gut, we also quantified phage abundance using the recent release of the Metagenomic Gut
189   Virus (MGV) catalogue. We used quasi-mapping for fast estimation of phage contig and viral
190   operational taxonomic unit (vOTU) relative abundance using Salmon in metagenomic mode [35,36]).
191   On average, 2.4% of metagenomic reads were assigned to prevalent viral contigs, with some samples
192   reaching 4.7% (Figure S5). We did not observe a significant difference in the percentage of assigned
193   phage reads between the **HC** and **LC** groups (Figure S5) or vOTU beta diversity (Figure 2e; P>0.10).
194   However, a closer look into diversity-generating retroelements (DGRs) [37,38] revealed a substantial,
195   significant reduction in DGRs phage genes in the **HC** group (two-sided Fisher test; P=7e-5; odds
196   ratio=0.3; Figure 2j). DGR elements use error-prone reverse transcriptase to induce random mutations

5

197 into the genomes of their host at specific target genes, creating population-wide hypervariability
198 [39,40]. Since DGRs have beneficial effects (e.g., adaption advantage) on their targeted host [39,40],
199 the enrichment in these phages may imply a more robust microbiome in **LC**.

200

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

203 We performed supervised machine learning (ML) to investigate if members of our cohort could be
204 classified as **HC** or **LC** solely based on bacterial taxonomic or functional relative abundances. We
205 applied SIAMCAT for model training and evaluation [41] but adopted data augmentation training to
206 the default algorithm [42]. We tested our models in an additional validation cohort of 11
207 immunotherapy-treated lung cancer patients. In the validation cohort, **HC** and **LC** were defined using
208 the same abundance thresholds used for the main cohort. Bacterial species abundance classified
209 patients as **HC** or **LC** with high accuracy in both our main cohort (Figure 3a; crossvalidation area
210 under the receiver operating characteristic [CV auROC]=77.9%) and validation cohort (Figure
211 3a; auROC=78.6%). With bacterial functional abundances (by EC), we achieve slightly higher
212 accuracy (Figure 3a; CV auROC=80.4%; additional auROC=82.1%). We further investigated if we
213 could predict high vs. low abundance levels for the species *C. albicans*, *C. sake*, and *C. glabrata,*
214 which were the most prevalent and abundant species in our cohort. High and low abundance groups
215 for each of the species were formed based on mean species CLR abundances and ML models were
216 built analogous to the **HC** vs. **LC** genus models. The identified microbiome signatures classified
217 patients as having high or low abundance well for *C. albicans*, *C. glabrata*, and *C. sake* and in both,
218 our main cohort (Figure 3a; CV auROC: *C. albicans*=77.5%, *C. sake*: 77.0%, *C. glabrata*: 73.9%) and
219 test cohorts (Figure 3a; Test auROC: *C. albicans*=86.7%, *C. sake:* 86.7%, *C. glabrata*: 79.2%).
220 Interestingly, phage vOTU abundance showed a potential to predict *Candida* genus and species levels
221 in the main cohort (Figure 3b; CV auROC: 73%±1%) but had less potential for the independent cohort
222 (Figure 3b; Test auROC: 53%-75%). *C. albicans* levels were predicted robustly in both the main and
223 validation cohorts (auROC=73%) using solely phage composition. This result was interesting
224 considering recent evidence of inhibition of *C. albicans* by bacteriophages [43].

225 We then inspected bacterial species predictive of *Candida* genus levels and crosschecked those
226 species with results from differential abundance analysis (by Maaslin2). We found that many of the
227 bacterial species predictive of **LC** with high robustness (at least 80%; *P*<0.05; false discovery rate
228 [FDR]<0.2; Figure 3c) were short-chain fatty acid (SCFA) producers [44–49], including *Actinomyces*
229 *odontolyticus*, *Bifidobacterium adolescentis, Eubacterium rectale, Anaerotruncus colihominis*,
230 *Alistipes ihumii AP11*, several *Lachnospiraceae* species, *Pseudoflavonifractor capillosus*, and
231 *Odoribacter splanchnicus*. We retrieved genome-scale metabolic models of the bacteria species
232 enriched in the **LC** group from the AGORA repository [50] and simulated growth on different diets
233 using flux balance analysis (FBA; Table S3). We monitored the potential to produce SCFAs and
234 confirmed that many of the bacterial species enriched in the **LC** group can secrete at least one of
235 acetate, propionate, or butyrate at varying levels (see Methods and Table S3). The importance of
236 SCFAs in suppressing *C. albicans* colonisation has been reported by us [25,26] and others [11].
237 However, the suppressive function of SCFAs appears to be towards all *Candida* species in general.
238 Several mechanisms by which propionate and butyrate suppress *Candida* species colonisation have
239 been suggested, including regulation of the immune system [51] and direct inhibition [52,53].
240 However, SCFAs also have a major impact on oxygen availability [54,55]. Therefore, a decrease in
241 SCFA producers in the **HC** group should be accompanied by an expansion of facultative aerobes.

6

148

242   Incidentally, we noticed an increase in oxygen-tolerant bacteria in the **HC** group (Figure 3c).
243   Interestingly, several lactic acid bacteria were significantly higher in the **HC** group and consistently
244   selected as top features in the ML models (for example, *Lactobacillus gasseri* and *Lactococcus lactis*)
245   (Figure 3c). We also observed an increased abundance of *Enterobacteriaceae* species (*Escherichia*
246   species*, Klebsiella pneumoniae*). A crossdomain correlation analysis between fungal genera and
247   bacterial species abundance confirmed positive correlations between *Candida* genus, *Lactobacillus*,
248   *Lactococcus*, *Klebsiella* and *Escherichia* species in our study cohort (Figure S2). Similar to the **LC**-
249   enriched species, FBA analysis suggested that in addition to the lactic acid bacteria enriched in the **HC**
250   group, *K. pneumoniae* and *E. coli* also secrete lactate (Table S3).

251   Microbial set enrichment analysis (MSEA) [56] revealed that bacterial genera that were
252   significantly increased in the **HC** group (*P*<0.05; *Lactobacillus*, *Lactococcus*, *Streptococcus*,
253   *Bacteroides* and *Odoribacter*; Table S4) were associated with multiple human disease genes (n=280)
254   suggesting a dysbiotic microbiome. In contrast, bacteria highly abundant in the **LC** patients showed
255   no enrichment in disease genes despite covering more genera (n=12). These results were also
256   qualitatively the same at *P*<0.10. Further functional enrichment analysis of the 280 human disease
257   genes based on the KEGG pathway database, indirectly linked **HC**-associated bacterial species to
258   cytokine and chemokine responses (Figure S6; Table S4). IL-17 signalling was of special interest
259   because it is associated with gut inflammation and *Candida* species colonisation [57]. IL-17 also has
260   a role in immune cell recruitment after bacterial invasion [58] and is a key role in stimulating host
261   immunity upon *Candida* infection [58].

262   We then examined bacterial metabolic functions by analysing MetaCyc pathway abundance (Figure
263   3d; Table S2). In total, 78 pathways were more abundant in **HC** compared to only 11 in **LC** (*P*<0.05;
264   *q*<0.15), matching the observation of increased functional alpha diversity in **HC**. One of the abundant
265   pathways in the **HC** group produces lactate from hexitols (P461-PWY). In agreement with the
266   observation of decreased anaerobes, aerobic respiration (fatty acid and beta oxidation pathways, TCA-
267   bypass, TCA cycle II) and synthesis pathways for compounds in cell membranes of aerobic bacteria
268   (menaquinol and ubiquinone synthesis) increased in the **HC** group. Notably, bby slightly relaxing the
269   *P*-value to 0.07, we obtained 5 additional TCA cycle pathways with higher abundance in the **HC** group
270   (TCA cycle I, IV, V, VII and partial TCA cycle in obligate autotrophs), effectively covering 6 of 9
271   TCA pathways in the MetaCyc database (Figure 3d; Table S2). Gene-set enrichment analysis
272   confirmed the elevated levels of aerobic respiration in the **HC** group (*q*<0.05; Table S2).

273   Based on the increased abundance of lactate producers, we examined functions related to lactate
274   utilisation. We found D-lactate dehydrogenase (D-LDH) was significantly increased in the **HC** group
275   by gene family (UniRef90; *P*=0.029; Figure 3e) and EC levels (EC:1.1.1.28; $\log_2$FC=1.09; *P*=0.053;
276   *q*=0.23; Table S2). We also found a significant increase in (S)-2-hydroxy-acid oxidase (EC:1.1.3.15;
277    $\log_2$FC=0.99; *P*=0.048; *q*=0.23) which reduces aliphatic hydroxy acids, including lactate, using flavin
278   mononucleotide and oxygen. In contrast, L-LDH genes and enzymes did not show significant changes
279   (*P*>0.10). We used the Metabolic Analysis of Metagenomes using fBA and Optimization (MAMBO)
280   algorithm [59] to predict the metabolic flux of the complete bacterial community and found a
281   significant increase in D-lactate secretion in the **HC** compared to the **LC** group (*P*=0.051; Figure 3f).

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

7

286

***C. albicans* has a competitive advantage using lactate under microaerobic conditions**
287
288 Our data suggested a possible link between microbial lactate production, higher availability of
289 molecular oxygen and increased *Candida* species abundance in the gut. To test the hypothesis that
290 *Candida* species may outgrow other fungi, specifically the *Saccharomyces* species that are their main
291 competitors in our study cohort, we selected *C. albicans* and *S. cerevisiae* var. *boulardii* for further
292 experiments. Since our ITS2 data could not reliably separate species from the *Saccharomyces in situ*
293 *stricto* group [60], we selected *S. boulardii* as the best proxy. A fluorescently labelled strain available
294 in our laboratory facilitated competition experiments with *C. albicans*. We used two strains of *C.*
295 *albicans* (BWP17, RM1000) that showed significant alterations in morphology, drug resistance, and
296 host-cell stimulation depending on the main carbon source [58]. In individual *in vitro* growth assays
297 with *C. albicans* BWP17 and RM1000 and *S. boulardii-GFP* on different carbon sources (lactate,
298 glucose, or lactate + glucose) with varying oxygen levels (anaerobic, microaerobic, or aerobic)
299 (Figure 4a-b; Table S5) only *C. albicans* was able to use lactate as a sole carbon source in aerobic and
300 microaerobic conditions. All three yeast were unable to grow on lactate under fully anaerobic
301 conditions (Figure 4b). The most significant growth of *C. albicans* compared to other oxygen levels
302 and carbon sources was on lactate as the sole carbon source in microaerobic conditions (t-test; $P<0.05$;
303 Figure 4a-b). In fully aerobic conditions, growth on solely lactate was strain dependent but generally
304 lower than in combined media (Figure 4a-b). These observed differences in growth on lactate were
305 more pronounced for BWP17 than RM1000. In microaerobic conditions, all species showed significant
306 reductions in growth rate in glucose+lactate compared to growth on at least one sole carbon source (t-
307 test; $P<0.05$). When glucose was present, *S. boulardii* growth was unaffected by additional lactate in
308 microaerobic and aerobic conditions (t-test; $P<0.05$; Figure 4a-b). However, both *C. albicans* strains
309 showed a slight growth reduction in glucose+lactate (t-test; $P<0.05$; Figure 4a).

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

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

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

335

Discussion

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

350   Most of the work to identify specific bacterial promoters or inhibitors of *Candida* species
351 colonisation has been performed in mouse models. Fan *et al.* [11] demonstrated how *Bacteroides*
352 *thetaiotamicron* can protect mice from *C. albicans* colonisation by activating innate immune effectors
353 and the antimicrobial peptide LL-37. The bacteria taxonomic annotation was based on 16S rRNA,
354 therefore the changes in the metabolic capacity of the gut microbiome associated with the *C. albicans*
355 colonisation remained unclear. Tan *et al.* [7] used mice to show that several gram-positive bacteria,
356 including *Staphylococcus aureus,* shed peptidoglycan units that trigger hyphae formation of *C.*
357 *albicans*. The only human study to concomitantly examine the mycobiome and microbiome to find
358 common colonisation patterns between *Candida* species and the gut microbiome was presented
359 recently by Zhai *et al.* [24]. The authors concluded that systemic candidiasis begins with expansion of
360 *Candida* species in the human gut. Using 16S rRNA, they observed a reduction in the levels of
361 anaerobes in patients with systemic candidiasis compared to non-infected cohorts. However, the small
362 number of patients (8) in this study and the lack of functional characterisation of the microbiome left
363 many questions unanswered.

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

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

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

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

411

10

422 Author contributions
423 **Conceptualisations**: B.S., Z.L., R.V.U., M.O.A.S., and G.P. **Funding acquisitions**: G.P., M.O.A.S.
424 **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.,
425 G.J.W., M.M, X.C., and R.V.U. **Project administration**: G.P. **Resources**: G.P., M.O.A.S.
426 **Supervision**: G.P., M.O.A.S. **Validation**: all authors. **Visualisation**: B.S., R.V.U. **Writing**—original
427 draft: B.S. and G.P. **Writing**—review and editing: all authors.

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

438 The remaining authors declare they have no competing interests.

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

442

443 Tables
444 **Table S1.** Test results for significant differences in alpha and beta diversity between groups.

445 **Table S2.** Differential abundance test results of bacteria, fungal, MetaCyc pathways, Enzyme
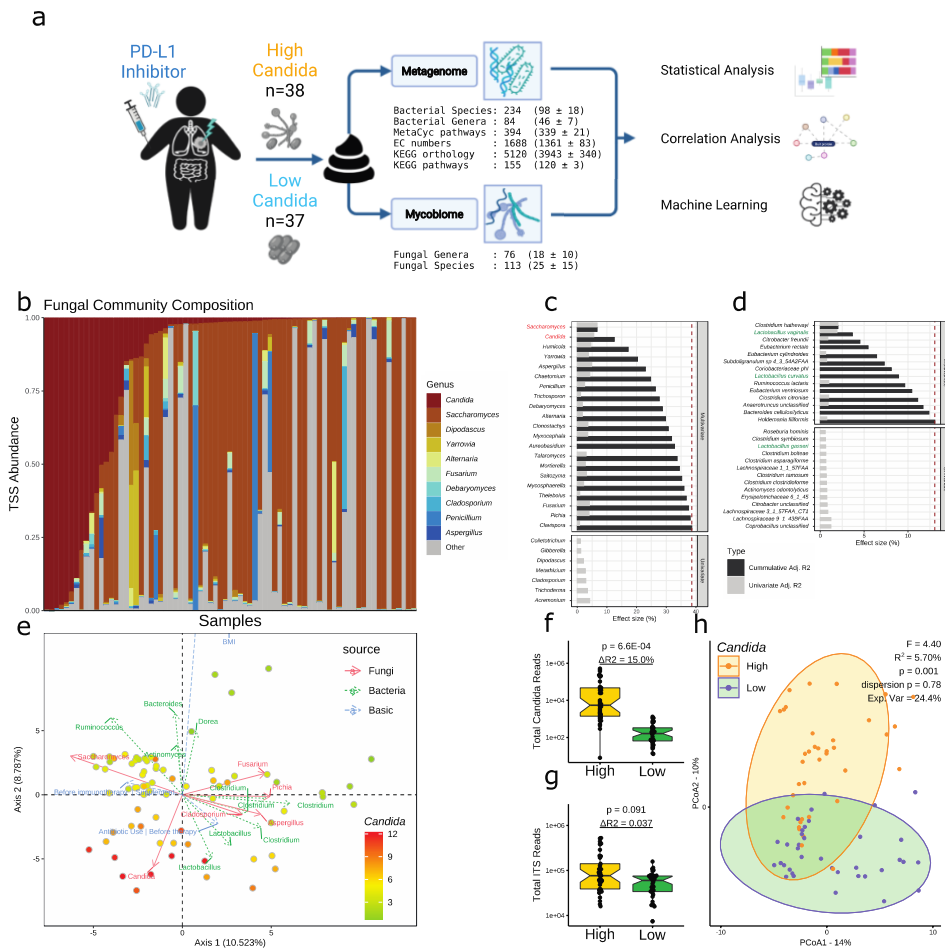446 commission (EC) and viral operational taxonomic unit (vOTU) abundances.

447 **Table S3.** Genome-scale metabolic modelling results for short-chain fatty acids from different diets
448 and oxygen levels.

449 **Table S4.** Microbial set enrichment results (MSEA).

450 **Table S5.** Growth rates, colony-forming units (CFU) and competition experiments.

451 Figures
452

11

453

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.

12

154
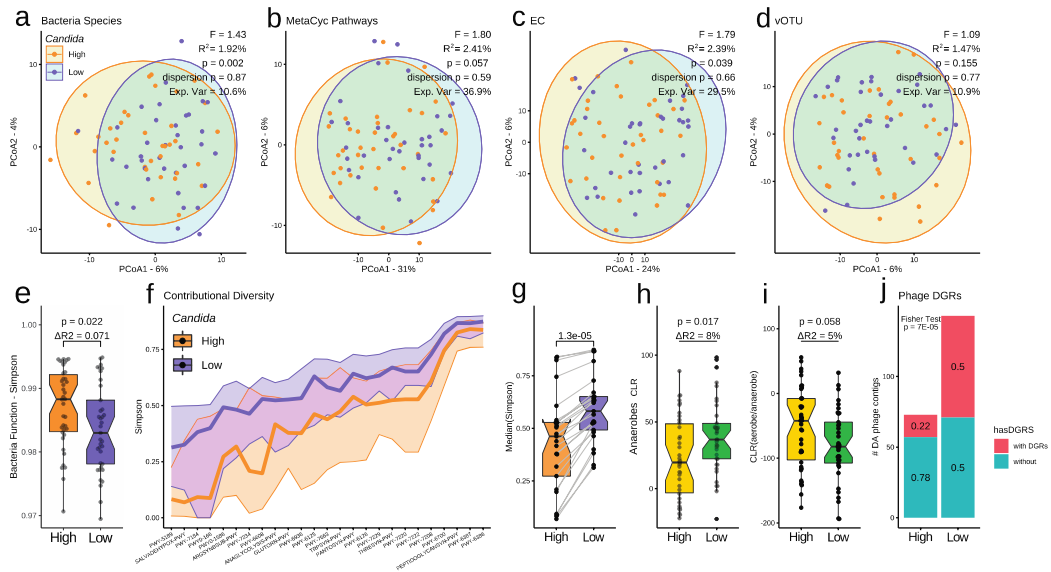
Figure 2. Microbiome contribution to high and low *Candida* groups. (a-d) Principal coordinate analysis (PCoA) of beta diversity (Aitchison distance) using **(a)** bacterial species, **(b)** bacterial metabolic pathways, **(c)** bacterial enzyme functions (EC, Enzyme Commission) or **(d)** viral orthologous 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).

490
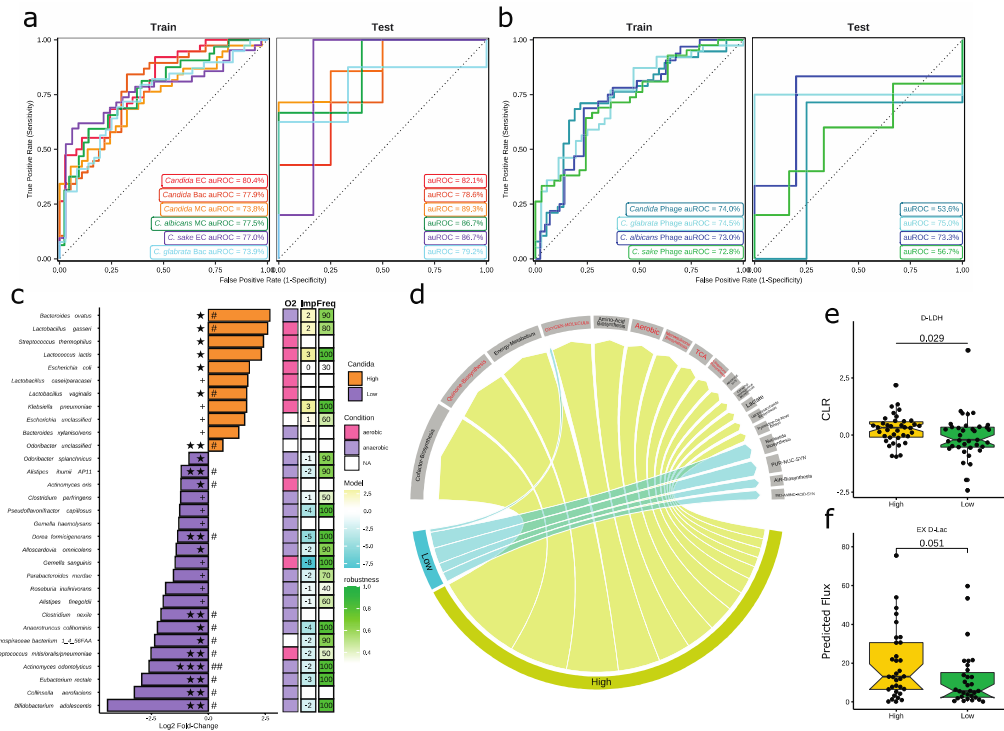
Figure 3. Gut bacterial signatures predictive of *Candida* levels. (a-b) Machine learning performance for predicting High and Low levels of all *Candida* genus and species (*C. albicans*, *C. sake*, *C. glabrata*). Performance is shown as area under the receiver operator characteristic (auROC) using **(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, [★]*P*<0.05, [★★]*P*<0.01, [★★★]*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.

508

14

509
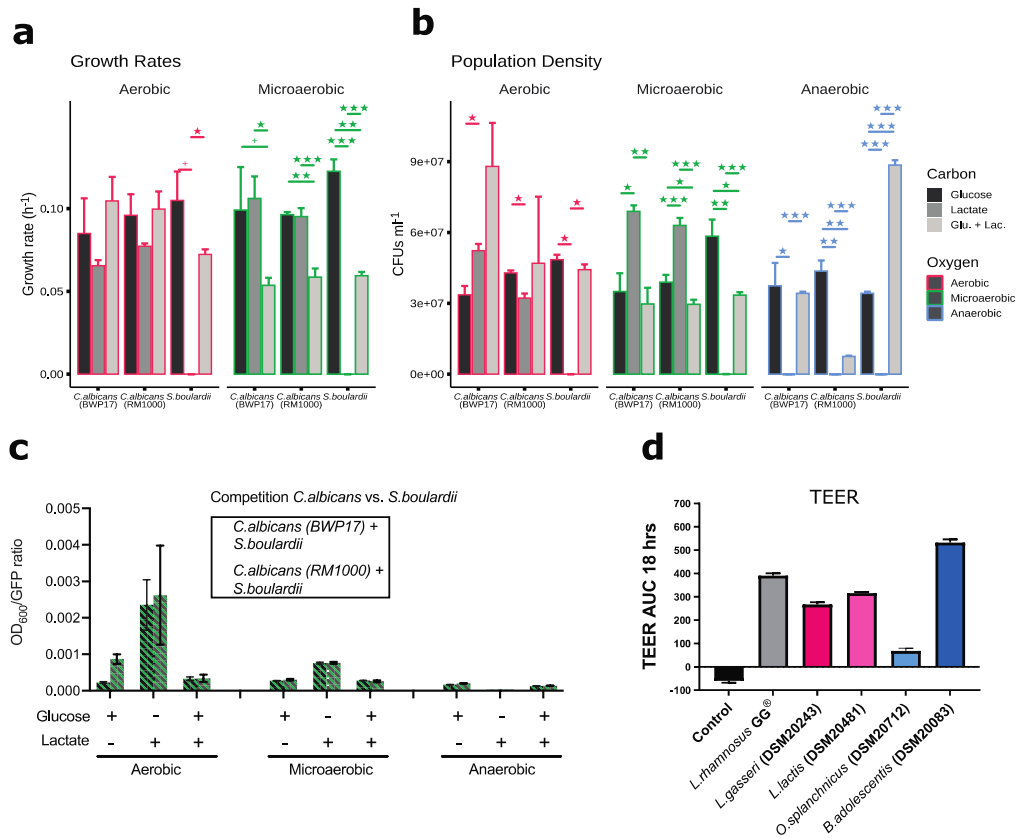
Figure 4. Bacterial transepithelial electrical resistance and yeast growth under varying levels of oxygen and carbon sources. (a-d) Histograms showing means and standard deviations. Significance was assessed using two-sided t-tests and indicated as $^{+}P<0.10$, $\star P<0.05$, $\star\star P<0.01$, $\star\star\star P<0.001$. (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).

521

15

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

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

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

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

550

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

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

564

565 Materials and methods

566 Ethics statement

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

571 Study participants and faecal collection

572 Faecal samples from 75 lung cancer patients were collected for analysis of fungal, viral, and bacterial microbiomes. All
573 faecal samples were collected at the National Koranyi Institute of Pulmonology, Budapest, Hungary, and the County
574 Hospital of Pulmonology, Torokbalint, Hungary. Patients received a combination therapy of chemotherapy and
575 immunotherapy. An additional cohort of 11 patients from the same hospital was recruited to provide stool samples for
576 validation of the machine learning models. These patients were treated with immunotherapy drugs only.

577 Faecal DNA extraction

578 Stool samples were kept frozen at −80°C before being sent to Novogene (UK) for DNA extraction and sequencing. Stool
579 samples were thoroughly mixed with 900 μL CTAB lysis buffer. All samples were incubated at 65°C for 60 min before
580 being centrifuged at 12 000 g for 5 min at 4°C. Supernatants were transferred to fresh 2 mL microcentrifuge tubes and 900
581 μL phenol: chloroform: isoamyl alcohol (25:24:1, pH = 6.7; Sigma-Aldrich) was added for extraction. Samples were mixed
582 thoroughly before incubation at room temperature for 10 min. Phase separation occurred by centrifugation at 12 000 g for
583 15 min at 4°C, and the upper aqueous phase was re-extracted with 900 μL phenol:chloroform:isoamyl alcohol. Samples
584 were then centrifuged at 12 000 g for 10 min at 4°C, and the upper aqueous phases were transferred to 2 mL microcentrifuge
585 tubes. Final extraction was with 900 μL chloroform: isoamyl alcohol (24:1) and centrifugation at 12 000 g for 15 min at
586 4°C. DNA was precipitated by adding the upper phase to 450 μL isopropanol (Sigma-Aldrich) containing 50 μL 7.5 M
587 ammonium acetate (Fisher). Samples were incubated at −20°C overnight. Shorter incubations (1 h) produced lower DNA
588 yields. Samples were centrifuged at 7500 g for 10 min at 4°C, and DNA pellets were washed three times in 1 mL 70%
589 (v/v) ethanol (Fisher). Pellets were air-dried and re-suspended in 200 μL 75 mM TE buffer (pH = 8.0; Sigma-Aldrich).

590 Fungal ITS2 sequencing

591 The concentration of extracted genomic DNA was determined by Qubit 2.0 and DNA quality was checked using gel
592 electrophoresis. PCR reactions used 200 ng DNA with primer sets specific to hypervariable regions (ITS3-2024F, ITS4-
593 2409R). Primer sets had unique barcodes. PCR products was separated on gels and fragment with the proper amplification
594 size were extracted and purified. Purified PCR product was used as a template for library preparation. PCR products were
595 pooled in equal amounts and then end polished, A-tailed, and ligated with adapters. Fragments were filtered with beads.
596 After 1 PCR cycle (to make the library double-stranded), libraries were analysed for size distribution and quantified using
597 real-time PCR. Paired-end sequencing of the library was performed on an Illumina Hiseq2500 (2x250bp).

598 Fungal ITS2 annotation

599 ITS2 raw reads were quality controlled, merged and filtered for chimeric reads using DADA2 in R 4.1, BioConductor 3.13
600 [77]. The median, high-quality, non-chimeric 78,332 ITS2 reads were extracted. From these, amplicon sequencing variants
601 (ASVs) were estimated. Representative sequences of ASVs were aligned to the fungal UNITE database (2017-12-01)
602 [78,79] using the Mothur classifier [80] to improve classification accuracy. ASV counts were summed at species and genus
603 levels and a 10% prevalence filter applied. Normalisation was by applying Bayesian zero-replacement and subsequent
604 centred log ratio (CLR) transformation.

605 Whole metagenomics sequencing

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

611 Bacterial taxonomic and functional profiling

612 Quality control of raw reads was performed using Sunbeam [81], including removal of low-quality reads (Phred score >
613 25 over 10 nucleotides), adaptors, and human-host related reads. Bacterial taxonomic profiling was performed using
614 MetaPhlAn2 [82] with default parameters. For bacterial functional annotation, the HUMAnN2 pipeline was employed [33]

17

615 with default parameters. Since HUMAnN2 does not directly support paired-end reads, paired-end sample mates were
616 merged as suggested by the HUMAnN2 authors. HUMAnN2 assigns pathways and gene families based on MetaCyc [83]
617 and UniRef90 [84] databases, respectively. Abundances of gene families in each sample were reported in reads per kilobase
618 (RPK) and further normalised by copies per million (CPM), effectively yielding transcript per kilobase million (TPKM).
619 To remove fungi-related abundances, we kept only functional abundances that were directly assigned to bacterial taxa (i.e.,
620 excluding other kingdoms and unclassified abundances). The resulting bacteria-only abundances were summed for each
621 feature category (MetaCyc pathways, Enzyme Commission (EC), KEGG orthology terms (KO), KEGG pathways). A 10%
622 prevalence filter was applied to bacterial species, bacterial genera, MetaCyc pathways, EC, KO, and KEGG pathways.
623 Summaries are reported as the median ± median absolute deviation (MAD).

624 Co-abundance networks
625 Spearman's correlation analyses of microbial abundances within and between kingdoms were performed by abundance
626 profiles normalized (TSS or CLR) for each kingdom separately. Only significant correlations were considered further
627 ($P<0.05$; $|r|>0.10$). Networks were transformed and analysed using the R package TidyGraph [85] and visualised using
628 GGraph [86].

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

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

643 Diversity analysis
644 Diversity calculations were performed using the R package vegan [93]. The alpha diversity indices of bacterial and fungal
645 communities were calculated using Shannon, Simpson, observed amplicon sequencing variants, and Pilou's evenness index
646 [93]. For beta diversity, analyses were performed using the Aitchison index and Bayesian-zero replacement as suggested
647 [28,29,91,94] to overcome compositionally related biases. Between-group significance was estimated using
648 PERMANOVA as implemented in the Adonis (vegan [95]) using 1000 permutations. Beta-dispersion (within-group
649 diversity) was assessed using betadisper (vegan).

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

656 Machine learning
657 We performed logistic regression based on GLMnet models as implemented in the R package SIAMCAT [41]. We screened
658 several model normalisations (rank-unit, log-unit, CLR), feature selection (receiver operating curve (ROC)-based with
659 n=20-100) and model (elastic Net, Lasso, Ridge, Lasso-LL, Ridge-LL) settings. Zeros were imputed by dividing the
660 smallest non-zero abundances value by 10. Models were trained with feature abundances from only one category at a time
661 (bacterial taxa, MetaCyc pathways, KEGG orthology terms, KEGG pathways). A 30% prevalence filter was applied in the
662 main cohort of this study. The test cohort was filtered for features prevalent in the main cohort. BMI and gender were
663 included as additional, fixed covariates in the models.

18

664 Data augmentation was applied to each training slot of the SIAMCAT model analogous to previous work [42]. Specifically,
665 train-test splits were initiated in the SIAMCAT model object. For each fold, new samples were generated from the training
666 samples per class so that any class imbalance was removed after adding augmented samples. Therefore, training samples
667 were used to fit a negative binomial distribution and the fitted distribution used to sample a fixed number of samples
668 (between 10 and 40). The resulting training cohort had both original and generated samples; test samples remained
669 unchanged. Class labels of generated samples were never used for performance evaluation (i.e., prediction performance)
670 and were relevant only for feature selection and model fitting.

671 ## Transepithelial electrical Resistance (TEER)

672 ### Bacterial growth conditions
673 The following strains were assessed: probiotic strain *Lactocaseibacillus rhamnosus* GG (DSM 33156, formerly known as
674 *Lactobacillus rhamnosus* GG, referred to by use of the trademark LGG), provided by Chr. Hansen A/S, Denmark and
675 four type strains; *Lactobacillus gasseri* (DSM20243), *Lactococcus lactis* subsp. *lactis* (DSM20481), *Bifidobacterium*
676 *adolescentis* (DSM20083) and *Odoribacter splanchnicus* (DSM20712). LGG (DSM33156), *L. gasseri* (DSM20243) and
677 *B. adolescentis* (DSM20083) were inoculated from frozen stock and cultured overnight at 37°C in De Man, Rogosa and
678 Sharpe broth (MRS) broth, pH 6.5 (Difco™, 288110) with 0.05% cysteine hydrochloride monohydrate under anaerobic
679 conditions in 7L anaerobic boxes with 2 sachets of 3.5L AnaeroGen™ Sachets (Thermo Scientific™ AN0035). *L. lactis*
680 (DSM20481) was grown in M17 broth (Biokar BK088HA) supplemented with 0.5% lactose aerobically at 30°C. Ten-fold
681 dilution series were prepared from the overnight cultures of these 4 strains and incubated under the same conditions as
682 mentioned above. Late exponential/early stationary phase were selected based on measures of optical density at 600nm
683 (OD_{600nm}). *O. splanchnicus* (DSM20712) was inoculated from freeze-dried stocks and cultured overnight at 37°C in Gifu
684 Anaerobic Medium broth (GAM; HyServe Code 05422) anaerobically using sealed N_2-gas flushed Hungate tubes
685 (Glasgerätebau Ochs Laborfachhandel, Art. No. 1020471). *O. splanchnicus* (DSM20712) was off-gassed twice per day as
686 the strain is a gas producer. Overnight cultures were subcultured once, and the overnight culture was used for the
687 experiment.

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

694 ### Transepithelial electrical resistance (TEER)
695 When the cells were approximately 80% confluent the medium was removed and cells were washed twice in Hanks
696 balanced salt solution (HBSS; Gibco, 14175). Cells (passage no.22) were dissociated by adding 2mL of TrypLE Express
697 Enzyme (Gibco, 12604) and left for 4 min in the incubator at 37°C. Approximately 10 mL of medium was added to
698 dissociated cells and 10^5 cells/well were seeded onto 12-well, 12 mm Transwell® inserts with 0.4 μm pore polyester
699 membrane insert (Corning; 3460). Growth media was renewed twice per week and cells were cultured for 21 days to obtain
700 a monolayer. The day prior to the TEER experiment, transwells were transferred to the CellZscope2 (Nanoanalytics,
701 Munster, Germany) and the medium was changed to antibiotic-free medium (DMEM supplemented with 20% FBS and
702 1% MEM non-essential amino acids). The CellZscope2 was maintained in a humidified atmosphere at 37°C with 5% CO_2
703 and in order to stabilize electrical resistance, overnight measurements of TEER before Caco-2 cells stimulation allowed
704 determination of a baseline and served as a quality control.

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

19

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

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

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

734 Fungal co-culture and *in vitro* assays
735 *Candida albicans*-RM1000, *C. albicans*-BWP17, and *Saccharomyces boulardii-GFP* were cultured in yeast extract-
736 peptone-dextrose (YPD) agar plates containing 10 g·L^{-1} yeast extract, 20 g·L^{-1} casein peptone and 20 g·L^{-1} glucose at 37°C
737 in aerobic conditions. Overnight liquid cultures were prepared by picking a single colony from a YPD agar plate and
738 inoculating synthetic complete (SC) minimal media containing 6.7 g·L^{-1} yeast nitrogen base supplemented with 1%
739 glucose. For aerobic growth conditions, overnight liquid cultures were incubated at 37°C and 200 RPM for ~16 hours. For
740 microaerobic (1% oxygen, 94% N2 and 5% H$_2$) or anaerobic (95% N2 and 5% H$_2$) growth conditions, liquid broth was
741 prereduced in anaerobic conditions 24 hours before inoculation. Next, strains were inoculated from YPD agar plates into
742 prereduced SC minimal media supplemented with glucose, and cultures were incubated at 37°C in microaerobic or
743 anaerobic conditions for ~16 hours without shaking.

744 Growth and competition experiments of *C. albicans* and *S. boulardii* in different carbon sources and
745 oxygen availability
746 To determine the effect of L-lactate and glucose on *C. albicans* and *S. boulardii* growth, overnight cultures were diluted to
747 optical density (OD)$_{600}$ of 0.1 in 100 µL fresh SC medium supplemented with 1% carbon source (1% glucose, 1% L-lactate,
748 or glucose plus L-lactate each at 0.5%), and incubated in aerobic, microaerobic, or anaerobic conditions, respectively.
749 Cultures were performed in flat-bottom, 96-well plates, sealed to prevent evaporation using BreathEasy film (Sigma-
750 Aldrich). Plates were incubated for 24 hours at 37°C, with continuous orbital shaking at 900 rpm. The ttemperatures used
751 (37°C) did not induce notable hyphae formation in growth assays and therefore did not perturb growth measurements. Cell
752 densities were measured every 10 min at OD$_{600}$ using a microtiter plate reader (BioTek Synergy H1). Growth rates were
753 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
754 least 24 time points over a 4-hour log phase period. The plate reader could not be adapted to strict anaerobic conditions;
755 therefore, determining growth rates of anaerobic experiments was not possible and only final time points were determined
756 for these experiments.

757 Competition growth experiments were performed by diluting overnight cultures to OD$_{600}$ of 0.1 and inoculating both
758 at 1:1 in a final volume of 100 µL fresh SC medium supplemented with the corresponding 1% carbon source. Cultures
759 were incubated for 24 hours in 96-well plates using the same growth conditions as described for the single-strain cultures.
760 OD$_{600}$ and GFP signals (excitation λ 485, emission λ 510) were measured at intervals of 10 minutes. Since GFP requires
761 oxygen to emit a fluorescent signal, anaerobic samples were incubated aerobically for 1 hour on ice before measuring the
762 final OD$_{600}$ and GFP. GFP signal from anaerobic samples was measured as a single time point after 24 hours.

763   Colony-forming units (CFU)·ml^{-1} were determined at the end of growth experiments. Cultures were serially diluted 10-
764 fold in an 8-step dilution series and 5 µL of the dilution series were spotted on YPD agar plates and incubated aerobically
765 for 48 hours at 37°C. To determine CFUs·ml^{-1}, single colonies were counted from the highest dilutions, divided by spot
766 volume (5 µL) and multiplied by the dilution factor. Technical replicates were made of all experiments.

References

768 1. Miranda LN, van der Heijden IM, Costa SF, Sousa API, Sienra RA, Gobara S, et al. Candida colonisation as a source for candidaemia.
769 J Hosp Infect. W.B. Saunders; 2009;72:9–16.
770 2. Pappas PG, Lionakis MS, Arendrup MC, Ostrosky-Zeichner L, Kullberg BJ. Invasive candidiasis. Nat Rev Dis Prim 2018 41. Nature
771 Publishing Group; 2018;4:1–20.
772 3. Cesaro S, Tridello G, Blijlevens N, Ljungman P, Craddock C, Michallet M, et al. Incidence, Risk Factors, and Long-term Outcome
773 of Acute Leukemia Patients With Early Candidemia After Allogeneic Stem Cell Transplantation: A Study by the Acute Leukemia and
774 Infectious Diseases Working Parties of European Society for Blood and Marrow Transplantation. Clin Infect Dis. Oxford Academic;
775 2018;67:564–72.
776 4. Mayer FL, Wilson D, Hube B. Candida albicans pathogenicity mechanisms. https://doi.org/104161/viru22913. Taylor & Francis;
777 2013;4:119–28.
778 5. Gunsalus KTW, Tornberg-Belanger SN, Matthan NR, Lichtenstein AH, Kumamoto CA. Manipulation of Host Diet To Reduce
779 Gastrointestinal Colonization by the Opportunistic Pathogen Candida albicans. mSphere. American Society for Microbiology; 2016;1.
780 6. Asbeck EC van, Clemons K V., Stevens DA. Candida parapsilosis: a review of its epidemiology, pathogenesis, clinical aspects, typing
781 and antimicrobial susceptibility. https://doi.org/103109/10408410903213393. Taylor & Francis; 2009;35:283–309.
782 7. Tan CT, Xu X, Qiao Y, Wang Y. A peptidoglycan storm caused by β-lactam antibiotic's action on host microbiota drives Candida
783 albicans infection. Nat Commun. Springer US; 2021;12.
784 8. Lopes JP, Stylianou M, Backman E, Holmberg S, Jass J, Claesson R, et al. Evasion of immune surveillance in low oxygen
785 environments enhances candida albicans virulence. MBio. American Society for Microbiology; 2018;9.
786 9. Ueno K, Matsumoto Y, Uno J, Sasamoto K, Sekimizu K, Kinjo Y, e t al. Intestinal Resident Yeast Candida glabrata Require s Cyb2p-
787 Mediated Lactate Assimilation to Adapt in Mouse Intestine. PLoS One. Public Library of Science; 2011;6:e24759.
788 10. Mirhakkak MH, Schäuble S, Klassert TE, Brunke S, Brandt P, Loos D, et al. Metabolic modeling predicts specific gut bacteria as
789 key determinants for Candida albicans colonization levels. ISME J 2020 155. Nature Publishing Group; 2020;15:1257 –70.
790 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
791 Candida albicans colonization. Nat Med 2015 217. Nature Publishing Group; 2015;21:808–14.
792 12. Sellers RS. Translating Mouse Models: Immune Variation and Efficacy Testing. Toxicol Pathol. SAGE Publications Inc.;
793 2017;45:134–45.
794 13. Sovran B, Planchais J, Jegou S, Straube M, Lamas B, Natividad JM, et al. Enterobacteriaceae are essential for the modulation of
795 colitis severity by fungi. Microbiome. BioMed Central Ltd.; 2018;6:152.
796 14. Zhang L, Bahl MI, Roager HM, Fonvig CE, Hellgren LI, Frandsen HL, et al. Environmental spread of microbes impacts the
797 development of metabolic phenotypes in mice transplanted with microbial communities from humans. ISME J. Nature Publishing Group;
798 2017;11:676–90.
799 15. Staley C, Kaiser T, Beura LK, Hamilton MJ, Weingarden AR, Bobr A, et al. Stable engraftment of human microbiota into mice with
800 a single oral gavage following antibiotic conditioning. Microbiome. BioMed Central; 2017;5:87.
801 16. Shultz LD, Brehm MA, Victor Garcia-Martinez J, Greiner DL. Humanized mice for immune system investigation: progress, promise
802 and challenges. Nat Rev Immunol 2012 1211. Nature Publishing Group; 2012;12:786–98.
803 17. Bratburd JR, Keller C, Vivas E, Gemperline E, Li L, Rey FE, et al. Gut microbial and metabolic responses to salmonella enterica
804 serovar typhimurium and candida albicans. MBio. American Society for Microbiology; 2018;9.
805 18. Lundberg R, Toft MF, Metzdorff SB, Hansen CHF, Licht TR, Bahl MI, et al. Human microbiota-transplanted C57BL/6 mice and
806 offspring display reduced establishment of key bacteria and reduced immune stimulation compared to mouse microbiota-transplantation.
807 Sci Reports 2020 101. Nature Publishing Group; 2020;10:1–16.
808 19. Zaborin A, Smith D, Garfield K, Quensen J, Shakhsheer B, Kade M, et al. Membership and behavior of ultra-low-diversity pathogen
809 communities present in the gut of humans during prolonged critical illness. MBio. American Society for Microbiology; 2014;5.
810 20. van Vliet MJ, Harmsen HJM, de Bont ESJM, Tissing WJE. The Role of Intestinal Microbiota in the Development and Severity of
811 Chemotherapy-Induced Mucositis. PLOS Pathog. Public Library of Science; 2010;6:e1000879.
812 21. Zwielehner J, Lassl C, Hippe B, Pointner A, Switzeny OJ, Remely M, et al. Changes in Human Fecal Microbiota Due to
813 Chemotherapy Analyzed by TaqMan-PCR, 454 Sequencing and PCR-DGGE Fingerprinting. PLoS One. Public Library of Science;
814 2011;6:e28654.
815 22. Lin XB, Dieleman LA, Ketabi A, Bibova I, Sawyer MB, Xue H, et al. Irinotecan (CPT-11) Chemotherapy Alters Intestinal Microbiota
816 in Tumour Bearing Rats. PLoS One. Public Library of Science; 2012;7:e39764.
817 23. Tang HJ, Liu WL, Lin HL, Lai CC. Epidemiology and Prognostic Factors of Candidemia in Cancer Patients. PLoS One. Public
818 Library of Science; 2014;9:e99103.
819 24. Zhai B, Ola M, Rolling T, Tosini NL, Joshowitz S, Littmann ER, et al. High-resolution mycobiota analysis reveals dynamic intestinal
820 translocation preceding invasive candidiasis. Nat Med. Springer US; 2020;26:59 –64.
821 25. Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer AC, et al. Antibiotics create a shift from mutualism to
822 competition in human gut communities with a longer-lasting impact on fungi than bacteria. Microbiome. BioMed Central Ltd; 2020;8:1–
823 20.
824 26. Marfil-Sánchez A, Zhang L, Alonso-Pernas P, Mirhakkak M, Mueller M, Seelbinder B, et al. An integrative understanding of the
825 large metabolic shifts induced by antibiotics in critical illness. Gut Microbes. Taylor & Francis; 2021;13.
826 27. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina

21

827 amplicon data. Nat Methods. Nature Publishing Group; 2016;13:581.

828 28. Luz Calle M. Statistical analysis of metagenomics data. Genomics and Informatics. Korea Genome Organization; 2019.

829 29. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML. Balances: a New Perspective for
830 Microbiome Analysis. mSystems. American Society for Microbiology; 2018;3:53–71.

831 30. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front
832 Microbiol. Frontiers; 2017;8:222 4.

833 31. García-Gamboa R, Kirchmayr MR, Gradilla-Hernández MS, Pérez-Brocal V, Moya A, González-Avila M. The intestinal mycobiota
834 and its relationship with overweight, obesity and nutritional aspects. J Hum Nutr Diet. John Wiley & Sons, Ltd; 2021;34:645 –55.

835 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
836 obesity in mice. Commun Biol 2021 41. Nature Publishing Group; 2021;4:1–11.

837 33. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of
838 metagenomes and metatranscriptomes. Nat Methods. Springer US; 2018;15:962–8.

839 34. Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol.
840 Elsevier Current Trends; 2016;24:402–13.

841 35. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression.
842 Nat Methods 2017 144. Nature Publishing Group; 2017;14:417–9.

843 36. Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo FJ. Systematic comparison and assessment of
844 RNA-seq procedures for gene expression quantitative analysis. Sci Reports 2020 101. Nature Publishing Group; 2020;10:1–15.

845 37. Ye Y. Identification of Diversity-Generating Retroelements in Human Microbiomes. Int J Mol Sci 2014, Vol 15, Pages 14234-
846 14246. Multidisciplinary Digital Publishing Institute; 2014;15:14234 –46.

847 38. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, et al. Reverse transcriptase-mediated tropism switching in
848 Bordetella bacteriophage. Science (80-). 2002;295:2091–4.

849 39. Medhekar B, Miller JF. Diversity-generating retroelements. Curr Opin Microbiol. Elsevier Current Trends; 2007;10:388 –95.

850 40. Wu L, Gingery M, Abebe M, Arambula D, Czornyj E, Handa S, et al. Diversity-generating retroelements: natural variation,
851 classification and evolution inferred from a large-scale genomic survey. Nucleic Acids Res. Oxford Academic; 2018;46:11–24.

852 41. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. Microbiome meta-analysis and cross-disease comparison enabled
853 by the SIAMCAT machine learning toolbox. Genome Biol 2021 221. BioMed Central; 2021;22:1–27.

854 42. Lo C, Marculescu R. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. BMC
855 Bioinforma 2019 2012. BioMed Central; 2019;20:1–14.

856 43. Nazik H, Joubert LM, Secor PR, Sweere JM, Bollyky PL, Sass G, et al. Pseudomonas phage inhibition of Candida albicans. Microbiol
857 (United Kingdom). 2017;163:1568–77.

858 44. Eeckhaut V, van Immerseel F, Croubels S, de Baere S, Haesebrouck F, Ducatelle R, et al. Butyrate production in phylogenetically
859 diverse Firmicutes isolated from the chicken caecum. Microb Biotechnol. Wiley-Blackwell; 2011;4:503.

860 45. Gotoh A, Nara M, Sugiyama Y, Sakanaka M, Yachi H, Kitakata A, et al. Use of Gifu Anaerobic Medium for culturing 32 dominant
861 species of human gut microbes and its evaluation based on short-chain fatty acids fermentation profiles. Biosci Biotechnol Biochem.
862 Oxford Academic; 2017;81:2009–17.

863 46. Hiippala K, Barreto G, Burrello C, Diaz-Basabe A, Suutarinen M, Kainulainen V, et al. Novel Odoribacter splanchnicus Strain and
864 Its Outer Membrane Vesicles Exert Immunoregulatory Effects in vitro. Front Microbiol. Frontiers Media S.A.; 2020;11.

865 47. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of human gut microbiota and short-chain
866 fatty acids in response to dietary interventions with three fermentable fibers. MBio. American Society for Microbiology; 2019;10.

867 48. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The Genus Alistipes: Gut Bacteria With Emerging Implications to
868 Inflammation, Cancer, and Mental Health. Front Immunol. 2020;11:906.

869 49. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS
870 Microbiol Lett. Oxford Academic; 2009;294:1–8.

871 50. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions
872 for 773 members of the human gut microbiota. Nat Biotechnol 2016 351. Nature Publishing Group; 2016;35:81–9.

873 51. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, et al. The microbial metabolites, short-chain fatty acids,
874 regulate colonic T reg cell homeostasis. Science (80-). American Association for the Advancement of Science; 2013;341:569–73.

875 52. Guinan J, Wang S, Hazbun TR, Yadav H, Thangamani S. Antibiotic-induced decreases in the levels of microbial-derived short-chain
876 fatty acids correlate with increased gastrointestinal colonization of Candida albicans. Sci Rep. Nature Publishing Group; 2019;9:8872.

877 53. Nguyen LN, Lopes LCL, Cordero RJB, Nosanchuk JD. Sodium butyrate inhibits pathogenic yeast growth and enhances the functions
878 of macrophages. J Antimicrob Chemother. 2011;66:2573 –80.

879 54. Rigottier-Gois L. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. ISME J 2013 77. Nature Publishing Group;
880 2013;7:1256–61.

881 55. Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science (80-). NIH Public Access;
882 2018;362.

883 56. Kou Y, Xu X, Zhu Z, Dai L, Tan Y. Microbe-set enrichment analysis facilitates functional interpretation of microbiome profiling
884 data. Sci Rep. Nature Research; 2020;10:1–12.

885 57. Kumamoto CA. Inflammation and gastrointestinal Candida colonization. Curr Opin Microbiol. NIH Public Access; 2011;14:386.

886 58. Ene I V., Cheng S-C, Netea MG, Brown AJP. Growth of Candida albicans Cells on the Physiologically Relevant Carbon Source
887 Lactate Affects Their Recognition and Phagocytosis by Immune Cells. Deepe GS, editor. Infect Immun. 2013;81:238–48.

888 59. Chowdhury S, Fong SS. Computational Modeling of the Human Microbiome. Microorganisms. Multidisciplinary Digital Publishing
889 Institute; 2020;8:197.

890 60. Stewart GG. Saccharomyces: Introduction. Encycl Food Microbiol Second Ed. 2014;3:297–301.

891 61. Zangl I, Pap IJ, Aspöck C, Schüller C. The role of Lactobacillus species in the control of Candida via biotrophic interactions. Microb
892 Cell. Shared Science Publishers; 2020;7:1.

893 62. Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L. Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies

22

894 for their stimulation in the human gut. Front. Microbiol. Frontiers Research Foundation; 2016. p. 979.

895 63. Krumbeck JA, Rasmussen HE, Hutkins RW, Clarke J, Shawron K, Keshavarzian A, et al. Probiotic Bifidobacterium strains and
896 galactooligosaccharides improve intestinal barrier function in obese adults but show no synergism when used together as synbiotics.
897 Microbiome. BioMed Central Ltd.; 2018;6:1–16.

898 64. Oh BS, Choi WJ, Kim JS, Ryu SW, Yu SY, Lee JS, et al. Cell-Free Supernatant of Odoribacter splanchnicus Isolated From Human
899 Feces Exhibits Anti-colorectal Cancer Activity. Front Microbiol. Frontiers Media S.A.; 2021;12:3454.

900 65. Xing C, Wang M, Ajibade AA, Tan P, Fu C, Chen L, et al. Microbiota regulate innate immune signaling and protective immunity
901 against cancer. Cell Host Microbe. Elsevier Inc.; 2021;29:959-974.e7.

902 66. Buffie CG, Pamer EG. Microbiota-mediated colonization resistance against intestinal pathogens. Nat Rev Immunol 2013 1311.
903 Nature Publishing Group; 2013;13:790–801.

904 67. Rea MC, Sit CS, Clayton E, O'Connor PM, Whittal RM, Zheng J, et al. Thuricin CD, a posttranslationally modified bacteriocin with
905 a narrow spectrum of activity against Clostridium difficile. Proc Natl Acad Sci U S A. National Academy of Sciences; 2010;107:9352–
906 7.

907 68. Abt MC, Pamer EG. Commensal bacteria mediated defenses against pathogens. Curr Opin Immunol. Elsevier Ltd; 2014;29:16–22.

908 69. Kamada N, Núñez G. Regulation of the Immune System by the Resident Intestinal Bacteria. Gastroenterology. W.B. Saunders;
909 2014;146:1477–88.

910 70. Koehler P, Stecher M, Cornely OA, Koehler D, Vehreschild MJGT, Bohlius J, et al. Morbidity and mortality of candidaemia in
911 Europe: an epidemiologic meta-analysis. Clin Microbiol Infect. Elsevier B.V.; 2019;25:1200–12.

912 71. Harris VC, Haak BW, Boele van Hensbroek M, Wiersinga WJ. The Intestinal Microbiome in Infectious Diseases: The Clinical
913 Relevance of a Rapidly Emerging Field. Open Forum Infect Dis. Oxford Academic; 2017;4.

914 72. Pradhan A, Avelar GM, Bain JM, Childers DS, Larcombe DE, Netea MG, et al. Hypoxia promotes immune evasion by triggering β-
915 glucan masking on the candida albicans cell surface via mitochondrial and cAMP-protein kinase A signaling. MBio. American Society
916 for Microbiology; 2018;9.

917 73. MacAlpine J, Daniel-Ivad M, Liu Z, Yano J, Revie NM, Todd RT, et al. A small molecule produced by Lactobacillus species blocks
918 Candida albicans filamentation by inhibiting a DYRK1-family kinase. Nat Commun 2021 121. Nature Publishing Group; 2021;12:1–
919 16.

920 74. Ballou ER, Avelar GM, Childers DS, Mackie J, Bain JM, Wagener J, et al. Lactate signalling regulates fungal β-glucan masking and
921 immune evasion. Nat Microbiol 2016 22. Nature Publishing Group; 2016;2:1–9.

922 75. Gutierrez D, Weinstock A, Antharam VC, Gu H, Jasbi P, Shi X, et al. Antibiotic-induced gut metabolome and microbiome alterations
923 increase the susceptibility to Candida albicans colonization in the gastrointestinal tract. FEMS Microbiol Ecol. Oxford Academic;
924 2020;96.

925 76. Bhaskaran N, Quigley C, Paw C, Butala S, Schneider E, Pandiyan P. Role of short chain fatty acids in controlling Tregs and
926 immunopathology during mucosal infection. Front Microbiol. Frontiers Media S.A.; 2018;9:1995.

927 77. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for
928 computational biology and bioinformatics. Genome Biol 2004 510. BioMed Central; 2004;5:1–16.

929 78. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The UNITE database for molecular
930 identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. Oxford Academic; 2019;47:D259–
931 64.

932 79. Abarenkov K, Nilsson RH, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, et al. The UNITE database for molecular identification
933 of fungi – recent updates and future perspectives. New Phytol. John Wiley & Sons, Ltd; 2010;186:281–5.

934 80. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput
935 community sequencing data. Nat Publ Gr. Nature Publishing Group; 2010;7:335–6.

936 81. Clarke EL, Taylor LJ, Zhao C, Connell A, Lee JJ, Fett B, et al. Sunbeam: An extensible pipeline for analyzing metagenomic
937 sequencing experiments. Microbiome. Microbiome; 2019;7:1–13.

938 82. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic
939 profiling. Nat Methods. 2015;12:902–3.

940 83. Kothari A, Kubo A, Fulcher CA, Weaver DS, Weerasinghe D, Foerster H, et al. The MetaCyc database of metabolic pathways and
941 enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 2013;42:D459–71.

942 84. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, Consortium U. UniRef clusters: a comprehensive and scalable alternative for
943 improving sequence similarity searches. Bioinformatics. 2014/11/13. Oxford University Press; 2015;31:926–32.

944 85. Thomas M, Pedersen L. Package 'tidygraph' - A Tidy API for Graph Manipulation. 2020;

945 86. Thomas Lin Pedersen. Package 'ggraph' - An Implementation of Grammar of Graphics for Graphs and Networks. 2021;

946 87. Robust Nonparametric Statistical Methods - Thomas P. Hettmansperger, Joseph W. McKean - Google Books [Internet]. [cited 2021
947 Apr 30]. Available from:
948 https://books.google.de/books?hl=de&lr=&id=6w3LBQAAQBAJ&oi=fnd&pg=PP1&dq=Hettmansperger,+T.P.+and+McKean+J.W.+
949 (2011),+Robust+Nonparametric+Statistical+Methods,+2nd+ed.,+New+York:+Chapman-
950 Hall.&ots=r7W8u49KFE&sig=h1iuYnC752P5ej75VcLaNYBG-PU#v=onepage&q&f=false

951 88. CRAN - Package Rfit [Internet]. [cited 2021 Apr 30]. Available from: https://cran.r-project.org/web/packages/Rfit/index.html

952 89. Tipton L, Müller CL, Kurtz ZD, Huang L, Kleerup E, Morris A, et al. Fungi stabilize connectivity in the lung and skin microbial
953 ecosystems. Microbiome 2018 61. BioMed Central; 2018;6:1–14.

954 90. Schwager E, Mallick H, Ventz S, Huttenhower C. A Bayesian method for detecting pairwise associations in compositional data.
955 PLoS Comput Biol. Public Library of Science; 2017;13:e1005852.

956 91. Martín-Fernández J-A, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in
957 compositional data sets. Stat Model An Int J. SAGE Publications Ltd; 2015;15:134–58.

958 92. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and Compositionally Robust Inference of Microbial
959 Ecological Networks. 2015;

960 93. Oksanen J, Blanchet G, Kindt R, Legendre P, O'Hara RB, Simpson G, et al. vegan: Community Ecology Package. 2011.

23

961   94. Kaul A, Mandal S, Davidov O, Peddada SD. Analysis of microbiome data in the presence of excess zeros. Front Microbiol. Frontiers
962   Media S.A.; 2017;8:2114.
963   95. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. Wiley Online Library; 2003;14:927 –30.
964   96. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome -scale metabolic models for microbial
965   species and communities. Nucleic Acids Res. Oxford Aca demic; 2018;46:7542 –53.
966   97. Garza DR, Van Verk MC, Huynen MA, Dutilh BE. Towards predicting the environmental metabolome from metagenomics with a
967   mechanistic model. Nat Microbiol 2018 34. Nature Publishing Group ; 2018;3:456–60.
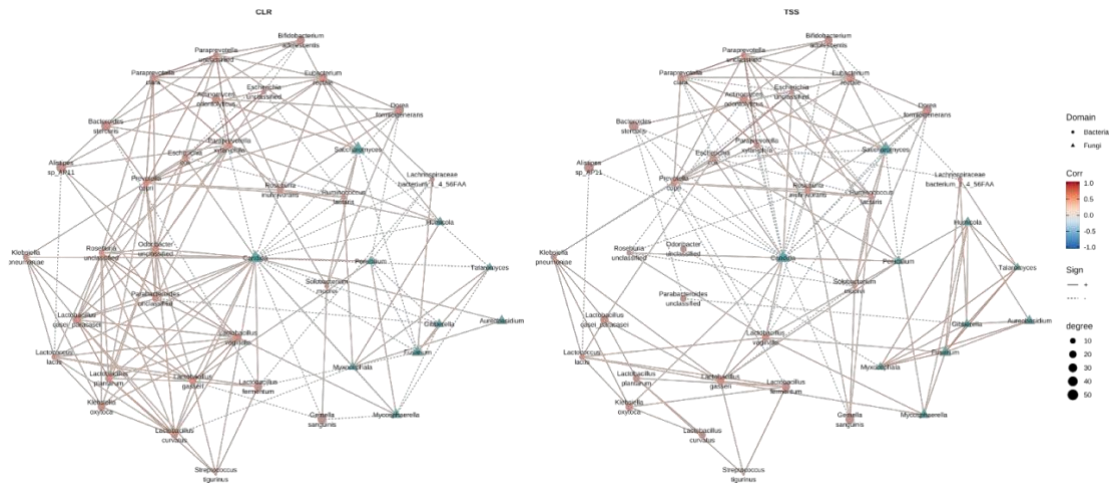968   98. Cplex IBMI. V12. 1: User's Manual for CPLEX. Int Bus Mach Corp. 2009;46:157.
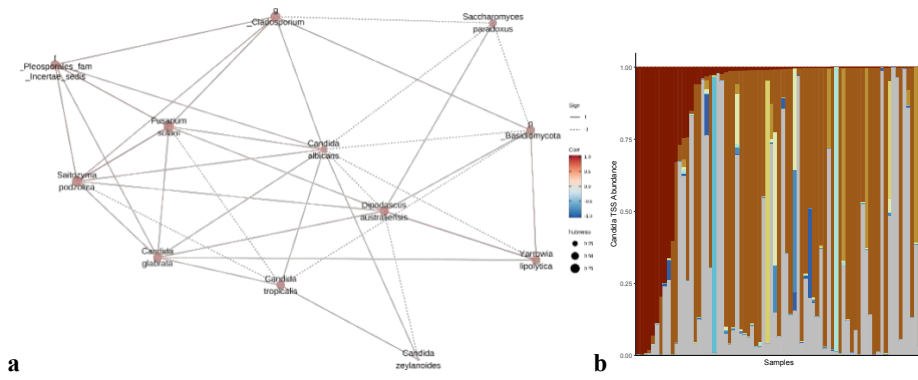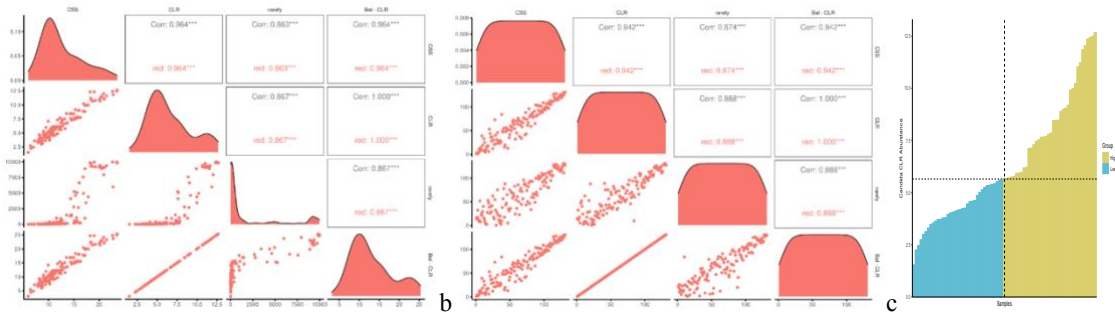969

Figure S1.

Figure S2.
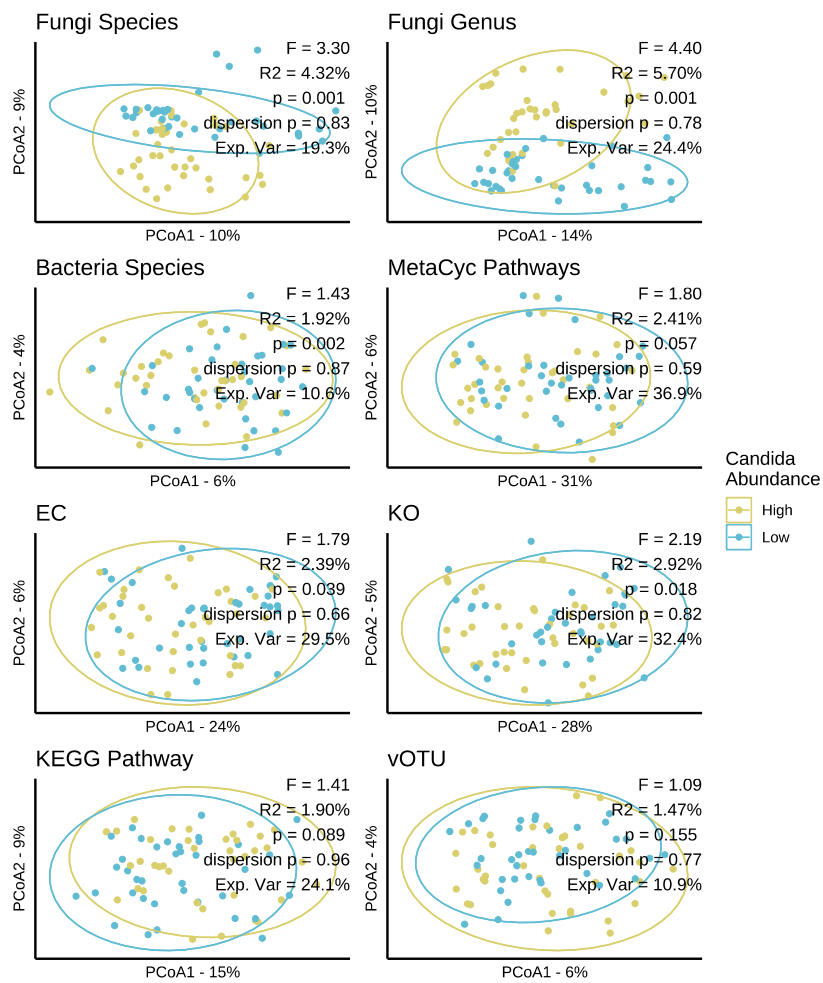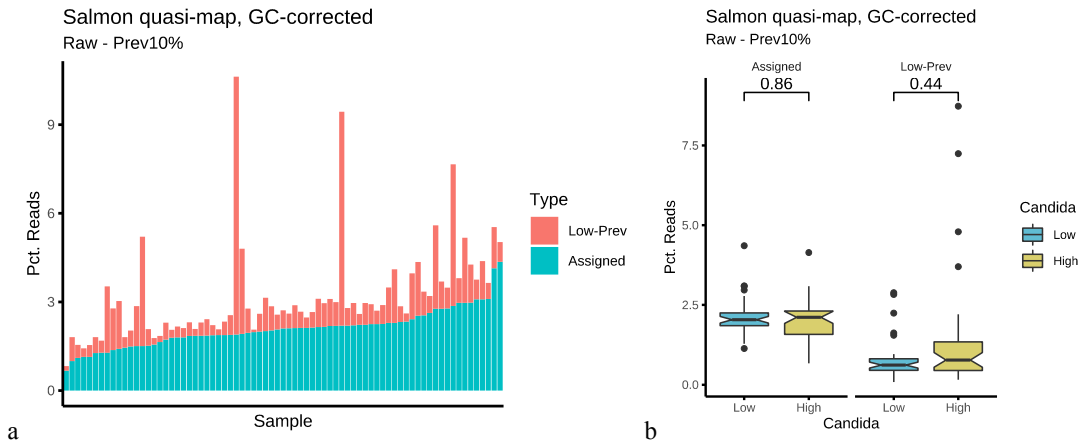
Figure S3.

Figure S4.
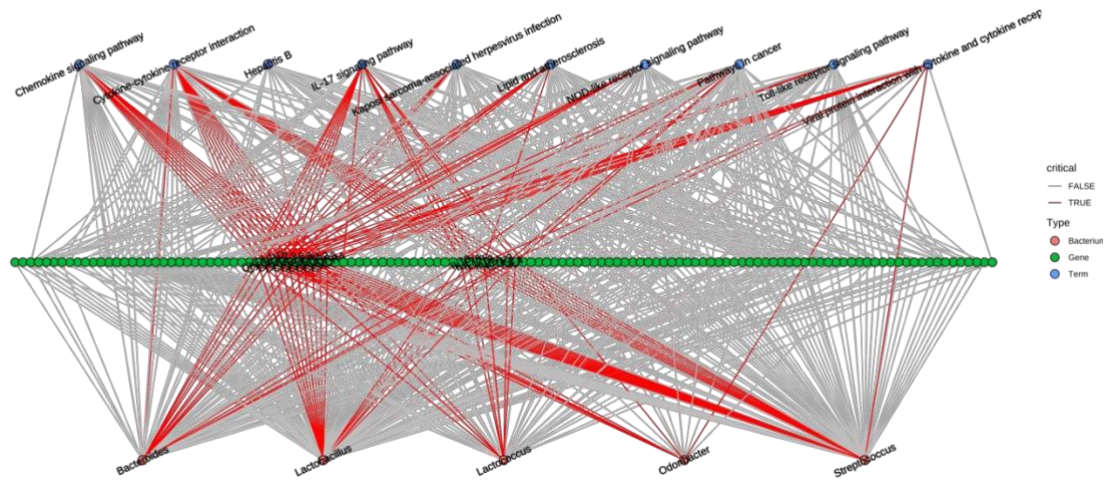
Figure S5.

Figure S6.

VI.   Manuscript:
Gut Microbiota and Lung Recovery

FORM 1

Manuscript No. 6

Manuscript title: Gut microbiome functionality might be associated with exercise tolerance and recurrence of resected early-stage lung cancer patients

Authors: Marfil-Sánchez A., **Seelbinder B.**, Varga J., Berta J., Hollosi V., Dome B., Megyesfalvi Z., Dulka E., Galffy G., Panagiotou G., Lohinai Z.

Bibliographic information: *PLoS One*. 16 (2021), pp. e0259898, 10.1371/JOURNAL.PONE.0259898

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
M.-Sánchez	30%	60%		55%	
Seelbinder	30%	40%		20%	
Panagiotou	20%			10%	50%
Lohinai	10%			10%	50%
Others	10%	0%	100%	5%	0%
Total:	100%	100%	100%	100%	100%

Overview:

I continued studying the role of bacteriomes and mycobiomes in human diseases. Here, we investigated gut microbiota associations with the lung function recovery of the host one year after tumour resection. Longitudinal data revealed associations between specific gut bacteria, fungi, and their metabolic pathways with the recovery of lung functions. Interestingly, an increase in VO_2 coincides with an increase in certain species and the "GABA shunt" pathway, which suggests that treatment outcomes might improve by enriching butyrate-producing species. Here, we revealed associations between specific gut bacteria, fungi, and their metabolic pathways with the recovery of lung function and exercise capacity.
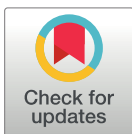
PLOS ONE

Gut microbiome functionality might be associated with exercise tolerance and recurrence of resected early-stage lung cancer patients

Andrea Marfil-Sánchez[1], Bastian Seelbinder[1], Yueqiong Ni[1], Janos Varga[2], Judit Berta[2], Virag Hollosi[2], Balazs Dome[2,3,4], Zsolt Megyesfalvi[2,3,4], Edit Dulka[5], Gabriella Galffy[5], Glen J. Weiss[6‡], Gianni Panagiotou[1‡*], Zoltan Lohinai[2‡]

1 Systems Biology and Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Jena, Germany, 2 National Koranyi Institute of Pulmonology, Budapest, Hungary, 3 Department of Thoracic Surgery, National Institute of Oncology-Semmelweis University, Budapest, Hungary, 4 Division of Thoracic Surgery, Department of Surgery, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria, 5 County Hospital of Torokbalint, Torokbalint, Hungary, 6 MiRanostics Consulting, Oro Valley, Arizona, United States of America

☯ These authors contributed equally to this work.
‡ GJW, GP and ZL are joint senior authors on this work.
* gianni.panagiotou@leibniz-hki.de

Abstract

Impaired exercise tolerance and lung function is a marker for increased mortality in lung cancer patients undergoing lung resection surgery. Recent data suggest that the gut-lung axis regulates systemic metabolic and immune functions, and microbiota might alter exercise tolerance. Here, we aimed to evaluate the associations between gut microbiota and outcomes in lung cancer patients who underwent lung resection surgery. We analysed stool samples, from 15 early-stage lung cancer patients, collected before and after surgical resection using shotgun metagenomic and Internal Transcribed Spacer (ITS) sequencing. We analysed microbiome and mycobiome associations with post-surgery lung function and cardiopulmonary exercise testing (CPET) to assess the maximum level of work achieved. There was a significant difference, between pre- and post-surgical resection samples, in microbial community functional profiles and several species from *Alistipes* and *Bacteroides* genus, associated with the production of SCFAs, increased significantly in abundance. Interestingly, an increase in VO_2 coincides with an increase in certain species and the "GABA shunt" pathway, suggesting that treatment outcome might improve by enriching butyrate-producing species. Here, we revealed associations between specific gut bacteria, fungi, and their metabolic pathways with the recovery of lung function and exercise capacity.

Introduction

The ethology or progression of lung cancer has been linked to numerous factors, including the gut microflora [1]. The microbiome can control epithelial cell proliferation/ differentiation, nutrition, detoxification, metabolism, and hormonal homeostasis. Gut microbiota has been implicated in tumour development in cancers [2, 3]. It has been shown that several chronic lung diseases, including cancer may be linked to a dysbiotic airway microbiota and frequently occur in conjunction with gastrointestinal disorders [4].

Previous studies showed that exercise training can alter the abundance of butyrate in the gut through an increase in the relative abundance of butyrate-producing species [5, 6]. Butyrate is a short-chain fatty acid (SCFA) that can regulate proliferation of epithelial cells in the gut, improve the integrity of the gut barrier, and alter the immune system and gene expression of the host [6]. A decrease in the abundance of Firmicutes has been repeatedly reported in lung cancer patients [2, 7] and has been previously associated with dysbiosis of gut metabolism, resulting in lower SCFAs concentration [2]. In contrast, a higher ratio of Firmicutes to Bacteroidetes bacteria has been show to correlate with maximal oxygen uptake during exercise (VO_2) [8]. Barton *et al.* [9] showed that athletes have greater faecal SCFA concentrations and altered abundance of bacterial pathways related to the biosynthesis of amino acids and the metabolism of carbohydrates. Others found that exercise training can alter the gut microbiota, at the taxonomic and functional level, in obese and non-obese humans [10] and reduce the expression of genes associated to metabolism of fructose and amino acids [11]. Moreover, exercise training may affect the gut mucus layer integrity, which is involved in the prevention of microbial adherence to the gut epithelium and acts as a substrate for bacteria of the mucosa-associated microflora. Increased heat shock protein response that prevents tight junction breakdown between epithelial cells [12] and improves resilience of the gut barrier [13, 14] has been observed in trained athletes. Others showed that physical activity decreases the risk for colorectal cancer by 24% [15].

Cardiopulmonary exercise testing (CPET) is a non-invasive, physiological test that provides a comprehensive overview of the pulmonary response to exercise and allows for the evaluation of the body's metabolic state, functional capacity, and impairment, through the assessment of submaximal and peak exercise responses. The metabolic changes that characterize health status and tumour recurrence have a stronger correlation with exercise tolerance, as measured by CPET, than with resting pulmonary and cardiac function testing, thus resulting in an increase in the use of CPET in patient management. Among many parameters, CPET measures the oxygen uptake (VO_2) peak that is used in predicting postoperative overall condition and pulmonary complications [16, 17].

In this study, our primary aim was to characterize the taxonomy and functionality of the gut microbiomes' metabolic interaction in lung cancer patients who underwent lung resection surgery. Our secondary aim was to describe the taxonomy changes in the same setting according to postoperative disease recurrence. We focus on the alterations of microbiome functionality to understand the complex gut-lung axis' metabolic interactions.

Materials and methods

Study population

From 2018 to 2019, we initially screened a cohort of 98 lung cancer patients from the National Koranyi Institute of Pulmonology and County Hospital of Torokbalint. From these, 20 were in early-stage and had their tumours surgically resected. Ultimately, 15 cases were suitable for the CPET testing and underwent pulmonary rehabilitation. A flow chart of study participants and

excluded patients is shown in S1 Fig. Baseline and followed-up samples were collected from those patients. Clinicopathological data collected included gender, age, smoking history, chemo- and radiotherapy treatments, type of operation, and overall survival (OS) or disease-free survival. Tumour, node, metastases stage according to the Union for International Cancer Control (8th edition [18]), and age at the time of diagnosis were recorded. The study and all treatments were conducted in accordance with the guidelines of the National Comprehensive Cancer Network and the Helsinki Declaration of the World Medical Association. The study was approved by the national level ethics committee (Hungarian Scientific and Research Ethics Committee of the Medical Research Council (ETTTUKEB- 50302-2/2017/EKU)). Informed consent was obtained for all patients. Stool samples were collected before (prior to surgery within 14 days) and 12 months after lung resection surgery. To assess a patient's overall condition, we performed a CPET test one year post-surgical resection to coincide with the time of follow-up stool sample collection. Major characteristics of the patient cohort are displayed in S1 Table. Patients received no neo-adjuvant therapy, and selected cases were treated with postoperative systemic chemotherapy with a platinum-based doublet regimen therapy (S1 Table).

Pulmonary function (PF) and CPET testing

We used spirometry at rest one year after surgery to assess key lung function parameters including {percentage of reference value} (Forced Expiratory Volume (FEV1) {%}), total Lung Capacity (TLC{%}), residual volume (RV{%}).

We performed CPET that begins with mild followed by intense exercise on an upright bicycle. Patients breathing through a mouthpiece. We measured the ventilation and respiratory gas parameters during exercise using oxygen and carbon dioxide gas analysers. Respiratory volumes were computed by integrating the air flow signals over the time of inspiration and expiration. The CPET test lasted until the maximal effort workload achieved for approximately 10 minutes.

We performed an incremental exercise test on an electronically braked cycle ergometer (Ergoline-900, Marquette) that begin with 3 minutes rest and 3 minutes constant pedalling at 20 W, work rate was increased from 5, 10, or 15 W/min in ramp profile. Patient pedalling speed was 60 rpm. Pulmonary ventilation (VE) and gas exchange (VO_2 and carbon dioxide output (VCO_2)) were measured breath-by-breath by a mass flow sensor and exercise metabolic measurement system (Vmax 29c, SensorMedics). During the test, we monitored heart rate (HR) using a 12-lead ECG (Cardiosoft, SensorMedics) and oxygen saturation by pulse oximetry (SatTrak, SensorMedics). We measured CPET parameters and lung function parameters including forced expiratory volume in one-second workload (Watt), maximal ventilation during exercise (VE max), ventilatory equivalents for oxygen (VE/VO_2), Oxygen pulse (O_2/HR), ventilatory equivalents for carbon dioxide (VE/VCO_2).

DNA extraction from stool samples

All samples were processed by Novogene. DNA extraction, library preparation and sequencing was done as described previously [19]. 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

173

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 7.5M ammonium acetate (Fisher). Samples were incubated at–20°C overnight. 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 75mM TE buffer (pH = 8.0; Sigma-Aldrich).

Library preparation and sequencing

Sequencing library was generated based on Illumina technologies and followed manufacturer's 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 analysed 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.

The concentration of genomic DNA for ITS2 sequencing 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, the library was analysed for size distribution and quantified using real-time PCR. The library was then to be sequenced on Hiseq2500.

Quality control

Quality control of raw reads was performed using the Sunbeam 2.1 pipeline [20]. First, cutadapt [21] version 2.8 was used to remove universal adapter sequences. Next, trimmomatic [22] version 0.36 was used to perform Illumina-specific adapter trimming, window quality trimming (Q5 over 25 nt), and 3' and 5' clipping (Q<6). Resulting reads shorter than 36nt were removed. Decontamination of human reads was performed by mapping quality-controlled reads with BWA [23] version 0.7.17 against a masked human reference genome (GRCh38-89). Masking of low entropy regions was performed using BBmask. Reads with 99% coverage and >97% identity to the human reference were removed.

Taxonomic and functional annotation

Taxonomic annotation was performed using MetaPhlAn2 [24] version 2.7.7 with default settings, generating taxonomic relative abundances.

The PIPITS pipeline [25] version 2.4 was used for taxonomy annotation of fungal Internal Transcribed Spacer (ITS) with default parameters including quality filtering, read-pair merging, ITS2 extraction and chimera removal. Remaining reads were binned based on 97% similarity as operational taxonomic units and aligned to the UNITE fungi database using Mothur classifier [26].

Functional annotation was performed using HUMAnN2 [27] version 0.11.2. In the pipeline, the reads were mapped to the MetaCyc database for pathway annotation, and the UniRef90 database to estimate gene family abundances. These abundances were aggregated into MetaCyc pathway abundances using MinPath based on the MetaCyc database.

Microbial community diversity analysis

Alpha diversity indices detailing microbial community composition within samples were calculated using the R packages vegan [28] and fossil [29]. For estimating beta diversity reflecting community dissimilarities, Bray-Curtis distances were calculated using R package vegan [28].

Co-abundance network

For bacterial co-abundance network reconstruction, the species relative abundance table was split into before and after surgical resection samples, and they were processed independently with SparCC [30].

Correlation analyses with lung function parameters

Partial Spearman's correlations adjusted for Chronic obstructive pulmonary disease (COPD) and cancer type were determined between significantly different bacterial features (species and pathways), fungal species abundances, and lung function parameters. For visualization only bacterial features with significant correlations ($P<0.05$) and fungal species with significant correlations ($P<0.05$) and absolute correlation coefficient > 0.65 were selected.

Prediction of VO_2, recurrence and overall survival (OS)

Prediction of VO_2, recurrence and OS was performed by calculating microbial balances [31]. A working implementation is available https://github.com/UVic-omics/selbal. A microbial balance is a special kind of log-contrast. Briefly, let X = (X_1,X_2,...,X_k) be a composition with k components. For two disjoint subsets of k_+ and k_- parts, respectively, the balance B of the component sets X_+ and X_- is determined by the formula:

$$B\left(X_+, X_-\right) \propto \frac{1}{k_+} \sum_{i \in I_+} \log X_i - \frac{1}{k_-} \sum_{j \in I_-} \log X_j$$

Feature selection for X_+ and X_- was performed using the function "selbal.cv", which implements an iterative cross-validation procedure to (i) identify the optimal number of components (C_{opt}) to be included in the balance and (ii) estimate the importance of the selected components. Only features (bacterial species and MetaCyc pathways) that changed significantly after surgery were used for feature selection.

Since "selbal.cv" is a forward selection process where components are included sequentially at every step, we have a sequence of balances, ranging from C = 2 to C = 20 components. Once C_{opt} has been determined, we apply "selbal.cv" to the whole data set, with the number of taxa C_{opt}, and obtain the global balance.

In order to estimate the robustness of the global balance and the importance of selected components, all balances with C_{opt} components obtained in the cross-validation process are retrieved and compared to the global balance, obtaining the relative frequencies of the different balances and the proportion of times that each component was included in a balance. The more often a component was selected, the higher its importance.

175

Statistical analysis

Testing for significant differences in alpha diversity between before and after surgical resection was performed using Wilcoxon signed-rank test. To test for significant differences in the microbial composition between before and after surgical resection, beta-diversity was analysed using permutational multivariate analysis of variance (PERMANOVA) as implemented in the function adonis from R package vegan [28].

Bacterial species and MetaCyc pathways were filtered by 10% prevalence across all samples and their relative abundances were used for statistical comparisons of before vs. after surgery. Differentially abundant features were identified by the Wilcoxon signed-rank test and were considered significantly different if $P < 0.05$ after FDR correction using the Benjamini-Hochberg procedure.

Species-species correlation coefficients were estimated as the average of 20 inference iterations and 100 permutations were used for the pseudo P-value calculation. A correlation was considered significant if pseudo $P < 0.05$.

Partial Spearman's correlations between lung function and CPET parameters with bacterial and fungal features, adjusted for Chronic obstructive pulmonary disease (COPD) and cancer type, were performed using R package ppcor [32] and were considered significant if $P < 0.05$.

Results

Community taxonomic and functional diversity before and one year after surgical resection

We analysed data from 15 lung cancer patients who underwent lung resection surgery. Our cohort includes a balanced set of men (n = 7) and women (n = 8), and patients with stage I (n = 4), stage II (n = 8) and stage IIIA (n = 3) disease. Tumours recurred in six patients. Detailed patients' characteristics are shown in S1 Table.

These subjects had an overall good performance status (Eastern Cooperative Group Performance Status 0) at diagnosis.

The structure and function of the gut microbiome were assessed using shotgun metagenomic sequencing. In total, 337 bacterial species were detected. First, we compared the overall composition of the gut microbiome communities pre- versus post-surgical resection. There were no significant changes in bacterial species alpha diversity over time (Shannon $P = 0.98$, Simpson $P = 0.56$, Chao1 $P = 0.53$; Wilcoxon signed-rank test) (Fig 1A). Bacterial species beta diversity did not show significant differences either, but indicated a trend (Bray-Curtis dissimilarity, $P = 0.074$, $R^2 = 4.8\%$, PERMANOVA) (Fig 1B). We further compared the species' beta diversity taking into account the different stages of disease and chemotherapy treatment and found no significant differences (S2A and S2B Fig). Alpha and beta diversity comparisons of recurrent and non-recurrent patients showed no significant differences either, but indicated a trend with non-recurrent patients having higher alpha diversity at both, pre- and post-surgical resection (S2C and S2D Fig).

We used shotgun metagenomic sequencing to further examine the variation of gut bacterial functions. In total, 498 MetaCyc pathways were retrieved. We did not observe significant differences in MetaCyc alpha diversity evenness and richness after surgical resection (Shannon $P = 0.28$, Simpson $P = 0.6$, Chao1 $P = 0.49$; Wilcoxon signed-rank test) (Fig 1C). However, we found a significant difference in bacterial functional profiles (Bray-Curtis dissimilarity) from post-surgical samples compared with pre-resection samples ($P = 0.047$, $R^2 = 6.4\%$, PERMANOVA) (Fig 1D).

176
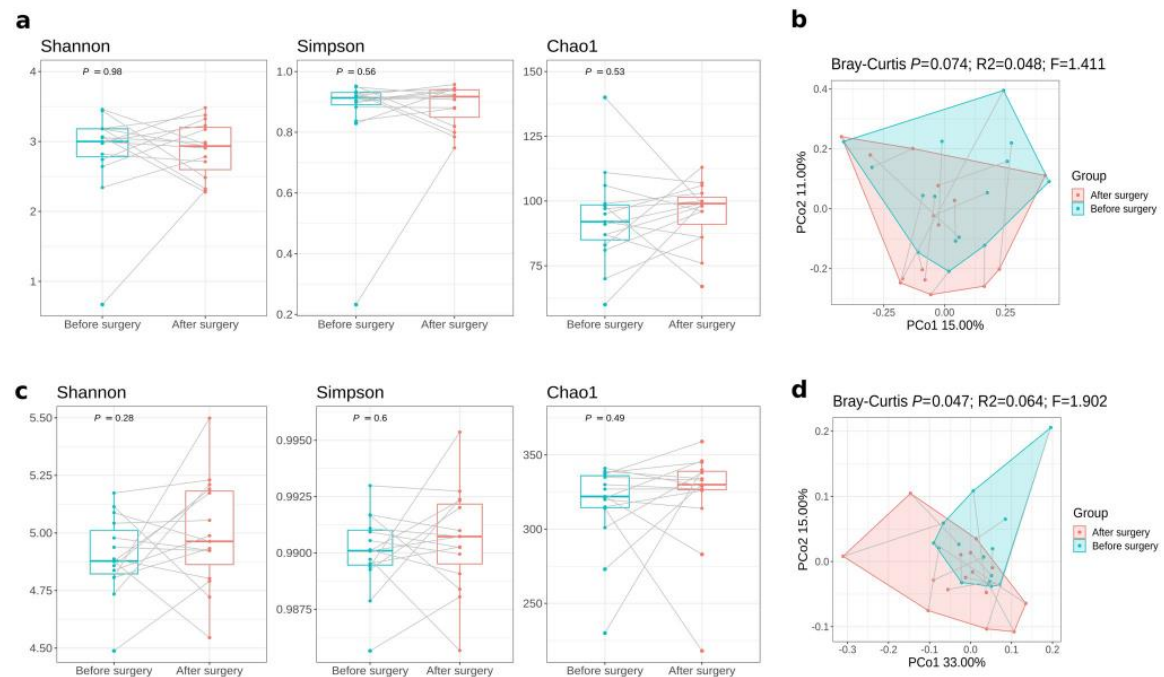
Fig 1. Comparison of the gut microbiota composition before and after surgical resection. (A, C) Boxplots, with median (centrelines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), showing alpha diversity Shannon, Simpson, and Chao1 indices at (A) bacterial species and (C) bacterial MetaCyc pathways. Gray lines connect samples of the same patient from before and after surgical resection. (B, D) Principal Coordinate Analysis plot based on Bray-Curtis distances before and after surgical resection at (B) bacterial species and (D) bacterial MetaCyc pathways. Gray lines connect the measurement of a patient before and after surgical resection.

https://doi.org/10.1371/journal.pone.0259898.g001

Alterations of gut microbial species, co-abundance network and functions one year after surgical resection

Next, we focused on taxonomic changes at the species level. We investigated which prevalent species changed significantly in relative abundance post-surgical resection. Thirty-two bacterial species showed significant changes after surgical resection compared to pre-surgical resection ($P<0.05$, Wilcoxon signed-rank test) (S2 Table). Of these, 15 were enriched pre-surgical resection, and 17 were enriched post-surgical resection (Fig 2A). We observed a significant decrease in the abundance of atypical opportunistic pathogens, such as *Klebsiella pneumoniae* [33], and *Odoribacter splanchnicus* [34]. These species are usually harmless within the gut of their host, but cause infection outside this niche. Other opportunistic pathogens like *Sutterella wadsworthensis* [35] decreased in abundance. Several species from *Alistipes* and *Bacteroides* genus, related to a healthy microbiome, increased significantly in abundance. These species have been associated with the production of SCFAs [36, 37].

We estimated relationships among gut microbes by constructing co-abundance networks based on bacterial species relative abundance (using SparCC [30]) for pre- and post-surgical resection groups, respectively. For evaluating the impact of surgery on the microbial community structure, we focused only on correlations between significant species that changed either from positive to negative or vice versa (Fig 2B). Overall, co-occurrence relationships were observed within and between phylum Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. We found more correlations changing from negative to positive post-surgical

Gut microbiome and exercise tolerance of resected early-stage lung cancer patients
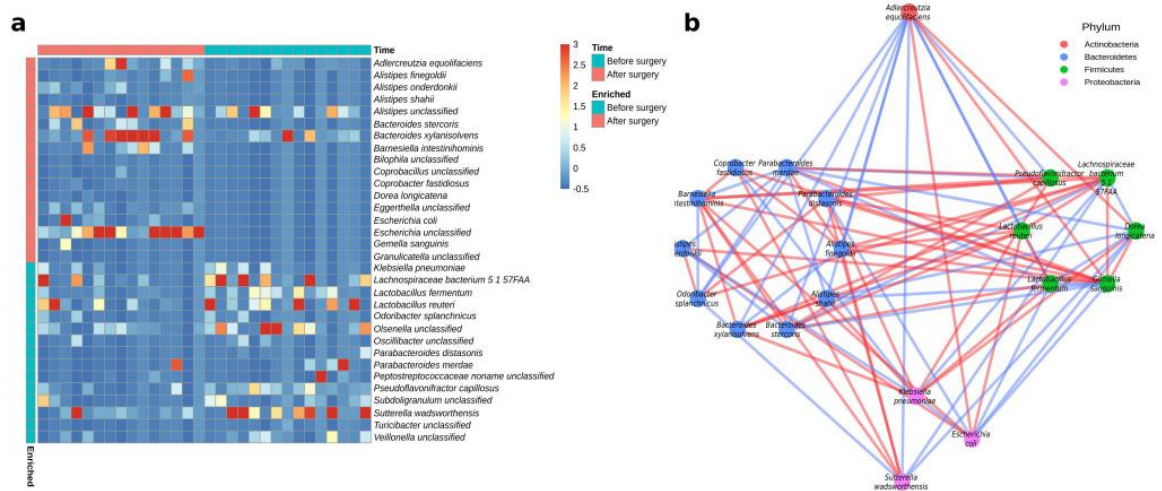
Fig 2. Taxonomic analysis of the gut microbiome in response to surgical resection. (A) Heatmap of differentially abundant bacterial species (*P*<0.05, Wilcoxon signed-rank test) before and after surgical resection. Red and blue in the far-left column indicates increased and decreased relative abundance, respectively. (B) Co-abundance network of bacterial species using SparCC [30]. Only correlations between differentially abundant bacterial species (*P*<0.05, Wilcoxon signed-rank test) that changed direction were used for network construction. The nodes are coloured based on their affiliated phyla. Edge colour indicates either correlations that changed from positive to negative (blue) or from negative to positive (red).

https://doi.org/10.1371/journal.pone.0259898.g002

resection (47) than from positive to negative (37). The species that showed the most changing correlations were *Gemella sanguinis*, *Adlercreutzia equolifaciens*, *Lachnospiraceae bacterium 5 1 57FAA*, *Parabacteroides merdae*, *K. pneumoniae* and *Barnesiella intestinihominis*.

Furthermore, 78 MetaCyc pathways differed significantly post-surgical resection compared to pre-surgical resection (*P*<0.05, Wilcoxon signed-rank test) (S3 Table) with 22 pathways enriched pre-surgical resection and 56 pathways post-surgical resection. The pathways with the highest increase were mostly related to the generation of precursor metabolites and energy and the biosynthesis of cofactors, electron carriers, and vitamins, whereas the pathways with the highest decrease were mostly related to core bacterial functions such as nucleoside and nucleotide biosynthesis (Fig 3).

Associations of bacteria and their function with lung function and CPET parameters

One year post-surgery, we investigated the significant associations between differentially abundant taxonomic and functional features with lung function and CPET parameters (FEV1%, TLC%, RV%, workload (Watt), VE, VO_2, O_2/HR, VE/VCO_2, VE/VO_2) adjusting for cofounders such as COPD and cancer type. In total, 12 bacterial species were significantly correlated (*P*<0.05) with lung function and CPET parameters. From these species, six were positively correlated with parameters: *Veillonella unclassified* (O_2/HR), *P. merdae* (RV%), *L. bacterium 5 1 57FAA* (TLC% and RV%), *Granulicatella unclassified* (VO_2), *G. sanguinis* (VO_2) and *Eggerthella unclassified* (RV%), whereas four species showed negative correlations: *Dorea longicatena* (Specific Airway Conductance (sGaw%)), *Coprobacillus unclassified* (VE and O_2/HR), *Bilophila unclassified* (VE and O_2/HR), *and B. intestinihominis* (workload and O_2/HR). *Parabacteroides distasonis* was positively correlated with VE/VO_2 and negatively correlated with workload while *Alistipes onderdonkii* was positively correlated with sGaw% and VE/VCO_2 and negatively correlated with workload (Fig 4A). We also investigated the relations between

Gut microbiome and exercise tolerance of resected early-stage lung cancer patients
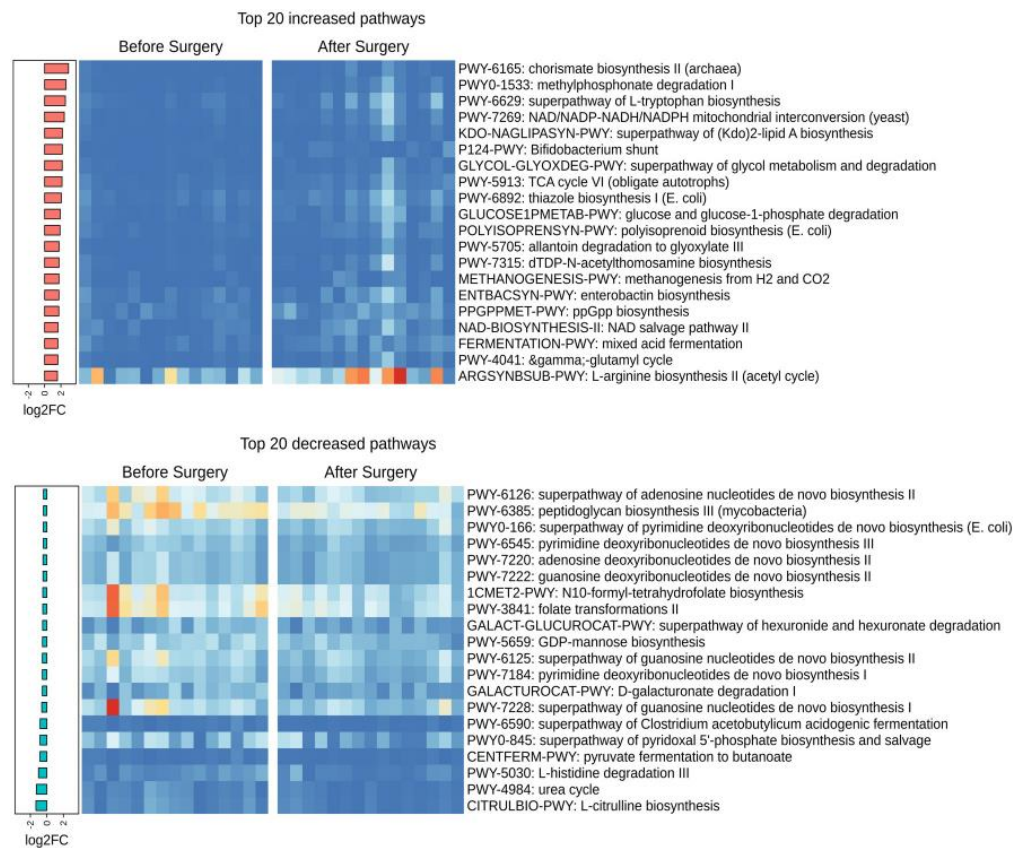
Fig 3. Functional analysis of the gut microbiome in response to surgical resection. Heatmaps of the top 20 increased (top) and decreased (bottom) differentially abundant bacterial MetaCyc pathways ($P<0.05$, Wilcoxon signed-rank test). Histogram (left panel) shows the pathways ranked by variation (log2 fold-change after surgical resection).

post-surgery changes in the abundance of species with post-surgery changes in lung function parameters (FEV, FEV1% and FEV/FVC). We identified 10 bacterial species, including species correlated with CPET parameters, such as *P. distasonis* and *B. intestinihominis* significantly correlated with changes in lung function parameters (S3A Fig).

Of the significantly differentially abundant MetaCyc pathways, four were positively associated with at least one of TLC%, sGaw%, VE, O_2/HR, VE/VCO_2, and VE/VO_2, including pathways for generation of precursor metabolites and energy, nucleoside and nucleotide degradation, and tetrapyrrole biosynthesis. Eighteen significantly differentially abundant MetaCyc pathways showed a negative correlation with lung function parameters (RV%, sGaw %, workload, and VE/VCO_2), including pathways responsible for compound degradation, utilization, and assimilation (N = 3), cofactor, electron carrier, and vitamin biosynthesis (N = 4), amino acid biosynthesis (N = 2), compound biosynthesis (N = 4), and generation of precursor metabolites and energy (N = 5) (Fig 4A) (S4 Table). Post-surgery changes in the abundance of 4 pathways, related to amino acid biosynthesis, cofactor, electron carrier, and vitamin biosynthesis, and compound degradation, utilization and assimilation, were positively associated with post-surgery changes in lung function parameters (S3A Fig). In contrast, post-surgery

Gut microbiome and exercise tolerance of resected early-stage lung cancer patients
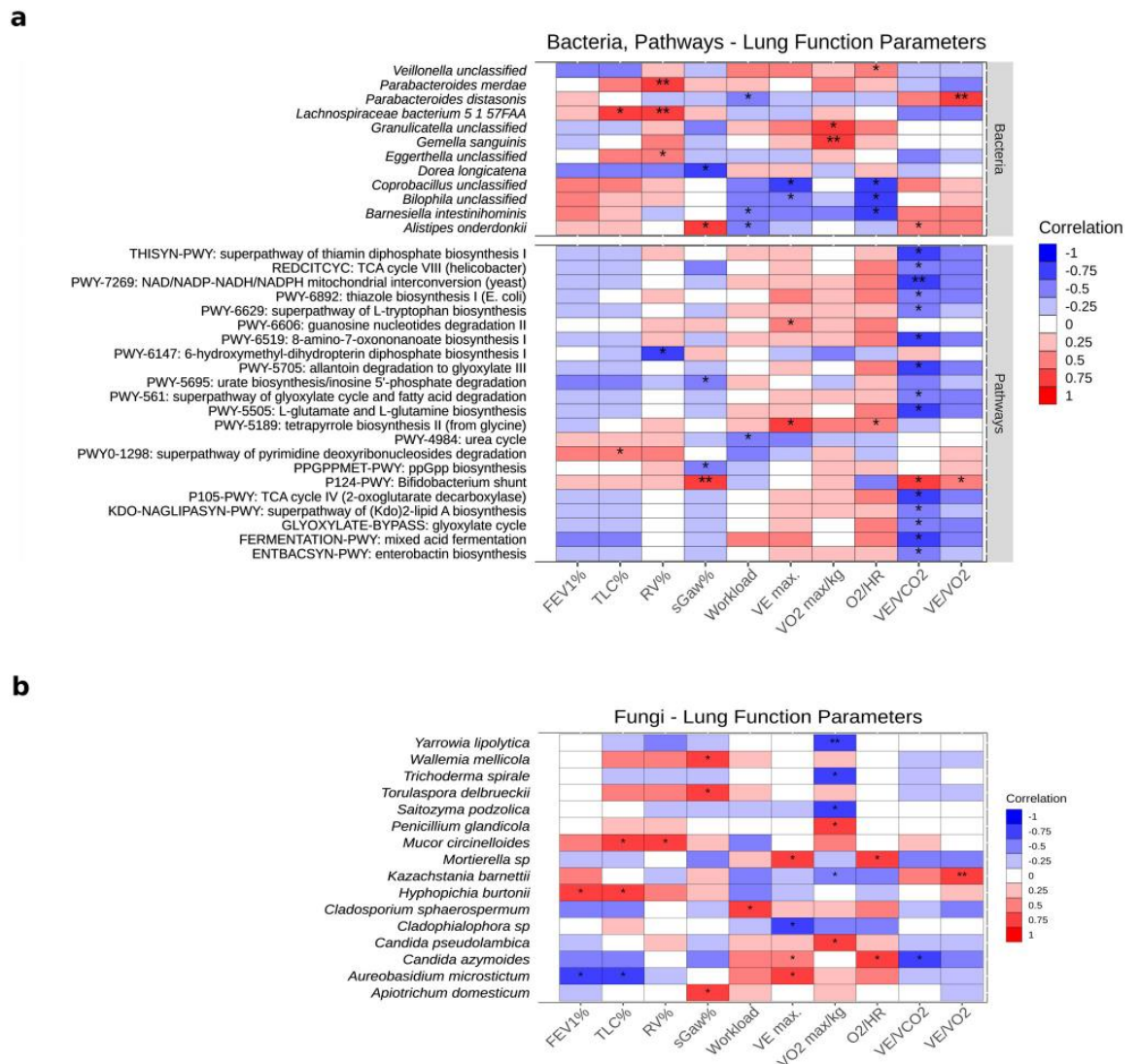
Fig 4. Correlations between taxonomic and functional profiles and CPET. (A) Heatmap of partial Spearman's rank correlation analysis between bacterial species and bacterial MetaCyc pathways versus lung function parameters adjusting for COPD and cancer type. Only differentially abundant species and pathways ($P<0.05$, Wilcoxon signed-rank test) were used. (B) Heatmap of partial Spearman's rank correlation analysis between fungal species versus lung function parameters adjusting for COPD and cancer type. (A, B) Cell colour indicates either negative correlation (blue) or positive correlation (red). Only species and pathways with significant correlations ((A) $P<0.05$; (B) $P<0.05$, absolute correlation coefficient > 0.65) are shown (*$P<0.05$, **$P<0.01$, ***$P<0.001$).

https://doi.org/10.1371/journal.pone.0259898.g004

changes in the abundance of two pathways were negatively associated with post-surgery changes in lung function parameters (S3A Fig).

We compared the abundance of aerobic and anaerobic species before and after surgery and found a strong, significant decrease in the abundance of anaerobes after surgical resection (r = 68.9%, P = 0.0054, Wilcoxon signed-rank test) (S4A Fig). When stratifying for recurrent and non-recurrent patients separately, recurrent patients showed a very strong, significant

Gut microbiome and exercise tolerance of resected early-stage lung cancer patients

decrease in the abundance of anaerobes (r = 89.9%, P = 0.031, Wilcoxon signed-rank test) (S4B Fig). Non-recurrent showed only a trend (r = 54.5%, P = 0.15, Wilcoxon signed-rank test). Given the difference in effect, these findings imply a potential role of anaerobe bacteria in tumour suppression.

At last, we tried to predict VO_2, tumour recurrence, and overall survival (OS) from either (a) bacterial or (b) MetaCyc pathway relative abundances (S5 Fig, S5 Table). We only looked at features with significant change post-surgical resection. We selected features whose microbial balance [31] has a high linear correlation to target response. For VO_2, the balance of only six bacterial species positively correlates with a Pearson correlation coefficient of $r = 0.78$ (P = 6E-4). Here, the balance was termed to be the ratio of *A. equolifaciens* and *Alistipes* abundances (both increased significantly after surgery) to *P. distasonis*, *P. merdae* (both decreased significantly after surgery) and *G. sanguinis* abundances. The balance of 7 MetaCyc pathways correlates even better with $r = 0.85$ (P = 7E-5). Functions for carbohydrate utilization, fatty acid oxidation and arginine production were selected. Interesting, an increase in VO_2 coincides with an increase in the pathway "GABA shunt", in which 4-aminobutyrate (GABA) is produced from glutamate and further metabolized to succinate. Improvements in exercise tolerance and treatment outcome were associated with butyrate producing species. Succinate is involved in several metabolic processes. Importantly, it is involved in the control of reactive oxygen species and tumorigenesis [38]. In contrast, functions for metabolizing glucose and fucose may have a negative impact on VO_2. In contrast to previous findings [8], the F/B ratio was not correlated or predictive of VO_2. However, the ratio of Proteobacteria against Euryarchaeota and Actinobacteria correlated with $r = 0.56$ (p = 0.032), implying a potential link between members of these phyla and VO_2.

For prediction of tumour recurrence vs non-recurrence, patients can be accurately classified by a balance of three bacteria (CV-AUC = 95%, Acc: 100%). *O. splanchnicus* is hereby associated with tumour free outcome, whereas *G. sanguinis* and *Olsenella* were associated with tumour recurrence. A balance of two MetaCyc functions was also predictive of recurrence (CV-AUC: 90%, Acc: 100%). Hereby, L-alanine biosynthesis was associated with positive outcome. Arginine and polyamine biosynthesis were predicted to increase tumour recurrence.

For OS, a balance of 5 MetaCyc pathways correlated well ($r = 87\%$, P = 2e-5). Interestingly, guanosine nucleotide de novo synthesis pathways were associated with improved OS, while adenosine and pyrimidine nucleotide de novo synthesis to decreased OS.

Associations of fungal species and lung function and CPET parameters

Besides the bacterial composition, we also attempted to associate the mycobiome with the lung function parameters. We built high-quality libraries for ITS2 sequencing for samples collected post-surgical resection (N = 15) and estimated the relative fungal abundance using the PIPITS pipeline [25]. In total, we identified 124 genera and 189 species. From these, 16 fungal species showed strong significant correlations ($P<0.05$, absolute correlation coefficient > 0.65) with lung function and CPET parameters (FEV1%, TLC%, sGaw%, Workload, VE, VO_2, and O_2/ HR) (Fig 4B). *Penicillium glandicola* and *Candida pseudolambica* were positively correlated with VO_2, whereas *Yarrowia lipolytica*, *Trichoderma spirale*, and *Saitozyma podzolica* showed negative correlations (S5 Table).

Discussion

CPET is a non-invasive method to test the overall condition of lung cancer patients who underwent lung resection surgery. CPET provides a comprehensive assessment of the exercise response and reflects the metabolic interactions of different organ systems. Shotgun

181

metagenomic sequencing allows taxonomic and functional annotation of the microbiome, thus it is the most comprehensive method for microbiome characterization.

Early metabolic changes might predict postoperative physical condition and outcome earlier than radiographic changes. The gut-lung axis plays a critical role in metabolic functional changes, and therefore, we hypothesized that the gut microbiome might be associated with different workloads and other CPET parameters.

Previous studies have shown that colorectal cancer patients have a disturbed gut microbial composition with a low abundance of species producing butyrate, such as species from the *Roseburia* and *Lachnospiraceae* genera, which may alter gene expression in healthy and cancerous cells [39]. In colorectal cancer cells, a dysfunction of the mitochondria contributes to an accumulation of butyrate in the cytosol and inhibition of histone deacetylases, resulting in the downregulation of proliferation and apoptosis pathways [39]. These changes may contribute to a decrease in tumour size and the probability of metastasis.

In our study, we observed post-surgical increases in (a) SCFA producing bacterial species and (b) pathways involved in carbohydrate, alcohol metabolism, and vitamin B production. These findings suggest a possible relationship between bacterial communities capable of metabolizing carbohydrates and alcohols into chemical compounds. The compounds involved in these pathways are considered important (a) to improve cell membrane barrier, (b) to act as precursor metabolites for human hormones such as serotonin and melatonin, and (c) for vitamin B_1 and B_3 production. Furthermore, we predicted a negative association between *O. splanchnicus* and tumor recurrence. *O. splanchnicus* is a potent butyrate producing species [40] by fermentation of L-Lysing (BioCyc P163-PWY [41]) and integral part of a healthy gut microbiome [42]. To improve CPET VO_2 and supress tumour recurrence, enriching the host microbiome with butyrate producing species and decreasing monosaccharide metabolizing species may improve treatment outcome.

From our study, we conclude that *A. equolifaciens*, *Alistipes*, *L. bacterium 5 1 57FAA* and *P. merdae* may be beneficial species for improved recovery of overall physical condition and lung capacity. These species were positively correlated with several lung function parameters or predictive of VO_2. In contrast, *G. sanguinis* was associated with decreased VO_2 and increased tumor recurrence. *D. longicatena* was associated with a decrease in lung function parameters, suggesting a negative impact of these two bacterial species on the recovery of lung capacity. Both *L. bacterium 5 1 57FAA* and *P. merdae* were negatively correlated with *D. longicatena*, implying that these two species may compete with *D. longicatena* for gut colonization and could therefore improve health post-surgery. While *O. splanchnicus* was not directly associated to improved lung function, a potential effect on tumour suppression should be worth further investigation. We observed a significant decrease in relative abundance of this bacterial species after resection. Like other SCFA producers, high fibber diets are required for the fermentation, proliferation and hence SCFA production. In the future, it should be worth to investigate if SCFA and other metabolites produced by *O. splanchnicus* are taken up by their human host and how these may influence tumour tissue.

Our study's limitation includes that in the real-life setting even though initial case numbers of recruited patients were high, the final numbers of patients with overall good clinical condition and available samples were low. To our best knowledge, the sample size is comparable to other studies of comparably unique conditions. Accordingly, this study did not allow us to draw robust conclusions on outcome-related biomarkers; however, it enabled us to meet our primary aim to study metabolic interactions. Another limitation is that the causality of metabolic and microbiome changes is not clear and follow-up studies with larger cohorts are needed.

We conclude that there are bacterial metabolic pathways that might be associated to increased oxygen uptake and exercise tolerance in patients one year after lung resection surgery. To our knowledge, this is the first study on microbiome functionality correlations with CPET, a highly specialized stress test with parameters that provide a unique detailed opportunity to study not only taxa but also metabolic interactions. Also, restoring specific bacteria might provide future therapeutic targets. Additionally, our findings highlight the potential of examining dynamic function parameters compared with traditional static metrics such as basic lung function and radiographic assessment in order to assess different organ systems' metabolic interactions. In this unique setting, the gut microbiota provided useful information on associations of exercise tolerance. We hypothesize that outcomes after lung resection surgery might be associated with distinct metabolic pathways that need to be confirmed in larger datasets. We offer critical postoperative metabolic and microbiome taxa and functional changes that are associated with distinct patient physical conditions and hopefully provide a reasonable basis for future studies aiming to increase patient outcomes.

Future studies in similar unique datasets are needed to confirm our findings and the modulation of gut microbiota, including butyrate producing taxa to increase long-term benefit from lung resection surgery.

Supporting information

S1 Fig. Study design of surgically resected lung cancer patient cohort.
(PDF)

S2 Fig. Comparison of the gut microbiota composition pre- and post-surgery. (A, B) Principal Coordinate Analysis plot based on Bray-Curtis distances of (A) the different stages and (B) chemotherapy treatment. (C) Boxplots, with median (centrelines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), showing alpha diversity Shannon, Simpson, and Chao1 indices of recurrent and non-recurrent patients. (D) Principal Coordinate Analysis plot based on Bray-Curtis distances pre- and post-surgery of recurrent and non-recurrent patients.
(PDF)

S3 Fig. Correlations between taxonomic and functional profiles and CPET. (A) Heatmap of partial Spearman's rank correlation analysis between the fold-change of bacterial species and bacterial MetaCyc pathways versus the fold-change lung function parameters adjusting for COPD and cancer type. Only differentially abundant species and pathways ($P<0.05$, Wilcoxon signed-rank test) were used. (B) Heatmap of partial Spearman's rank correlation analysis between fungal species versus the fold-change of lung function parameters adjusting for COPD and cancer type. (A-B) Cell color indicates either negative correlation (blue) or positive correlation (red). Only species and pathways with significant correlations ($P<0.05$) are shown (*$P<0.05$, **$P<0.01$, ***$P<0.001$).
(PDF)

S4 Fig. Prediction of VO2, tumour recurrence, and overall survival (OS). Prediction of VO_2, tumour recurrence, and overall survival (OS) from bacterial species (left), bacterial phyla (middle) or MetaCyc pathways (right) relative abundances.
(PDF)

S5 Fig. Abundance of aerobic and anaerobic species. (A, B) Boxplots, with median (centrerelines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), showing the abundance of aerobic and anaerobic species. Gray lines connect samples of the same

patient from before and after surgical resection.
(PDF)

S1 Table. Major patients' characteristics.
(XLSX)

S2 Table. Species Wilcoxon signed-rank test results.
(TSV)

S3 Table. Pathways Wilcoxon signed-rank test results.
(TSV)

S4 Table. Partial Spearman's correlation results of bacterial species and pathways with lung function and CPET parameters.
(TSV)

S5 Table. VO$_2$, recurrence and overall survival predictions.
(XLSX)

S6 Table. Partial Spearman's correlation results of fungal species with lung function and CPET parameters.
(TSV)

Author Contributions

Conceptualization: Glen J. Weiss, Gianni Panagiotou, Zoltan Lohinai.

Data curation: Judit Berta, Virag Hollosi, Balazs Dome, Edit Dulka, Gabriella Galffy.

Formal analysis: Andrea Marfil-Sánchez, Bastian Seelbinder, Yueqiong Ni.

Investigation: Andrea Marfil-Sánchez, Bastian Seelbinder, Janos Varga, Zsolt Megyesfalvi.

Resources: Judit Berta, Virag Hollosi, Balazs Dome, Edit Dulka, Gabriella Galffy.

Visualization: Andrea Marfil-Sánchez, Bastian Seelbinder.

Writing – original draft: Andrea Marfil-Sánchez, Bastian Seelbinder, Yueqiong Ni.

Writing – review & editing: Glen J. Weiss, Gianni Panagiotou, Zoltan Lohinai.

References

1. Zhuang H, Cheng L, Wang Y, Zhang Y-K, Zhao M-F, Liang G-D, et al. Dysbiosis of the Gut Microbiome in Lung Cancer. Front Cell Infect Microbiol. 2019; 9. https://doi.org/10.3389/fcimb.2019.00112 PMID: 31065547

2. Liu F, Li J, Guan Y, Lou Y, Chen H, Xu M, et al. Dysbiosis of the gut microbiome is associated with tumor biomarkers in lung cancer. Int J Biol Sci. 2019; 15: 2381–2392. https://doi.org/10.7150/ijbs.35980 PMID: 31595156

3. Zitvogel L, Galluzzi L, Viaud S, Vétizou M, Daillère R, Merad M, et al. Cancer and the gut microbiota: An unexpected link. Sci Transl Med. 2015; 7: 271ps1. https://doi.org/10.1126/scitranslmed.3010473 PMID: 25609166

4. Bingula R, Filaire M, Radosevic-robin N, Bey M, Berthon J, Bernalier-donadille A, et al. Desired Turbu-lence? Gut-Lung Axis, Immunity, and Lung Cancer. 2017; 2017. https://doi.org/10.1155/2017/5035371 PMID: 29075294

5. MATSUMOTO M, INOUE R, TSUKAHARA T, USHIDA K, CHIJI H, MATSUBARA N, et al. Voluntary Running Exercise Alters Microbiota Composition and Increases n-Butyrate Concentration in the Rat Cecum. Biosci Biotechnol Biochem. 2008; 72: 572–576. https://doi.org/10.1271/bbb.70474 PMID: 18256465

184

Gut microbiome and exercise tolerance of resected early-stage lung cancer patients

6. Peng L, Li Z-R, Green RS, Holzman IR, Lin J. Butyrate Enhances the Intestinal Barrier by Facilitating Tight Junction Assembly via Activation of AMP-Activated Protein Kinase in Caco-2 Cell Monolayers. J Nutr. 2009; 139: 1619–1625. https://doi.org/10.3945/jn.109.104638 PMID: 19625695

7. Zhang WQ, Zhao SK, Luo JW, Dong XP, Hao YT, Li H, et al. Alterations of fecal bacterial communities in patients with lung cancer. Am J Transl Res. 2018; 10: 3171–3185. PMID: 30416659

8. Durk RP, Castillo E, Márquez-Magaña L, Grosicki GJ, Bolter ND, Lee CM, et al. Gut Microbiota Composition Is Related to Cardiorespiratory Fitness in Healthy Young Adults. Int J Sport Nutr Exerc Metab. 2019; 29: 249–253. https://doi.org/10.1123/ijsnem.2018-0024 PMID: 29989465

9. Barton W, Penney NC, Cronin O, Garcia-Perez I, Molloy MG, Holmes E, et al. The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. Gut. 2017; gutjnl-2016-313627. https://doi.org/10.1136/gutjnl-2016-313627 PMID: 28360096

10. ALLEN JM, MAILING LJ, NIEMIRO GM, MOORE R, COOK MD, WHITE BA, et al. Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans. Med Sci Sport Exerc. 2018; 50: 747–757. https://doi.org/10.1249/MSS.0000000000001495 PMID: 29166320

11. Munukka E, Ahtiainen JP, Puigbó P, Jalkanen S, Pahkala K, Keskitalo A, et al. Six-Week Endurance Exercise Alters Gut Metagenome That Is not Reflected in Systemic Metabolism in Over-weight Women. Front Microbiol. 2018; 9. https://doi.org/10.3389/fmicb.2018.02323 PMID: 30337914

12. Dokladny K, Moseley PL, Ma TY. Physiologically relevant increase in temperature causes an increase in intestinal epithelial tight junction permeability. Am J Physiol Liver Physiol. 2006; 290: G204–G212. https://doi.org/10.1152/ajpgi.00401.2005 PMID: 16407590

13. Lira FS, Rosa JC, Pimentel GD, Souza HA, Caperuto EC, Carnevali LC, et al. Endotoxin levels correlate positively with a sedentary lifestyle and negatively with highly trained subjects. Lipids Health Dis. 2010; 9: 82. https://doi.org/10.1186/1476-511X-9-82 PMID: 20684772

14. Fehrenbach E, Niess AM, Schlotz E, Passek F, Dickhuth HH, Northoff H. Transcriptional and translational regulation of heat shock proteins in leukocytes of endurance runners. J Appl Physiol. 2000; 89: 704–10. https://doi.org/10.1152/jappl.2000.89.2.704 PMID: 10926657

15. Wolin KY, Yan Y, Colditz GA, Lee I-M. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer. 2009; 100: 611–616. https://doi.org/10.1038/sj.bjc.6604917 PMID: 19209175

16. Bolliger CT, Jordan P, Solèr M, Stulz P, Grädel E, Skarvan K, et al. Exercise capacity as a predictor of postoperative complications in lung resection candidates. Am J Respir Crit Care Med. 1995; 151: 1472–1480. https://doi.org/10.1164/ajrccm.151.5.7735602 PMID: 7735602

17. Morice RC, Walsh GL, Ali MK, Roth JA. Redefining the lowest exercise peak oxygen consumption acceptable for lung resection of high risk patients. Chest. 1996; 110: 161S.

18. Mirsadraee S. The 7th lung cancer TNM classification and staging system: Review of the changes and implications. World J Radiol. 2012; 4: 0. https://doi.org/10.4329/wjr.v4.i4.128 PMID: 22590666

19. Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer A-C, et al. Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. Microbiome. 2020; 8: 133. https://doi.org/10.1186/s40168-020-00899-6 PMID: 32919472

20. Clarke EL, Taylor LJ, Zhao C, Connell A, Lee JJ, Fett B, et al. Sunbeam: An extensible pipeline for analyzing metagenomic sequencing experiments. Microbiome. 2019; 7: 1–13. https://doi.org/10.1186/s40168-018-0604-3 PMID: 30606251

21. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011; 17: 10. https://doi.org/10.14806/ej.17.1.200

22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

24. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015; 12: 902–903. https://doi.org/10.1038/nmeth.3589 PMID: 26418763

25. Gweon HS, Oliver A, Taylor J, Booth T, Gibbs M, Read DS, et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. Bunce M, editor. Methods Ecol Evol. 2015; 6: 973–980. https://doi.org/10.1111/2041-210X.12399 PMID: 27570615

26. 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: 7537–7541. https://doi.org/10.1128/AEM.01541-09 PMID: 19801464

27. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018; 15: 962–968. https://doi.org/10.1038/s41592-018-0176-y PMID: 30377376

28. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003; 14: 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

29. Vavrek MJ. fossil: Palaeoecological and palaeogeographical analysis tools. Palaeontol Electron. 2011; 14.

30. Friedman J, Alm EJ. Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol. 2012; 8: 1–11. https://doi.org/10.1371/journal.pcbi.1002462 PMID: 22629235

31. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML. Balances: a New Perspective for Microbiome Analysis. mSystems. 2018; 3: 53–71. https://doi.org/10.1128/mSystems.00053-18 PMID: 30035234

32. Kim S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. Commun Stat Appl Methods. 2015; 22: 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665 PMID: 26688802

33. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A. 2015; 112: E3574–81. https://doi.org/10.1073/pnas.1501049112 PMID: 26100894

34. Göker M, Gronow S, Zeytun A, Nolan M, Lucas S, Lapidus A, et al. Complete genome sequence of Odoribacter splanchnicus type strain (1651/6). Stand Genomic Sci. 2011; 4: 200–9. https://doi.org/10.4056/sigs.1714269 PMID: 21677857

35. Brook I, Wexler HM, Goldstein EJC. Antianaerobic antimicrobials: spectrum and susceptibility testing. Clin Microbiol Rev. 2013; 26: 526–46. https://doi.org/10.1128/CMR.00086-12 PMID: 23824372

36. 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. https://doi.org/10.3389/fimmu.2020.00906 PMID: 32582143

37. Rios-Covian D, Salazar N, Gueimonde M, de Los Reyes-Gavilan CG. Shaping the Metabolism of Intestinal Bacteroides Population through Diet to Improve Human Health. Front Microbiol. 2017; 8: 376. https://doi.org/10.3389/fmicb.2017.00376 PMID: 28326076

38. Tretter L, Patocs A, Chinopoulos C. Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis. Biochim Biophys Acta—Bioenerg. 2016; 1857: 1086–1101. https://doi.org/10.1016/j.bbabio.2016.03.012 PMID: 26971832

39. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 2012; 6: 320–329. https://doi.org/10.1038/ismej.2011.109 PMID: 21850056

40. Gomez-Arango LF, Barrett HL, McIntyre HD, Callaway LK, Morrison M, Dekker Nitert M. Increased Systolic and Diastolic Blood Pressure is Associated with Altered Gut Microbiota Composition and Butyrate Production in Early Pregnancy. Hypertension. 2016; 68: 974–981. https://doi.org/10.1161/HYPERTENSIONAHA.116.07910 PMID: 27528065

41. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform. 2019; 20: 1085–1093. https://doi.org/10.1093/bib/bbx085 PMID: 29447345

42. Nagai F, Morotomi M, Watanabe Y, Sakon H, Tanaka R. Alistipes indistinctus sp. nov. and Odoribacter laneus sp. nov., common members of the human intestinal microbiota isolated from faeces. Int J Syst Evol Microbiol. 2010; 60: 1296–1302. https://doi.org/10.1099/ijs.0.014571-0 PMID: 19667375

Supplements
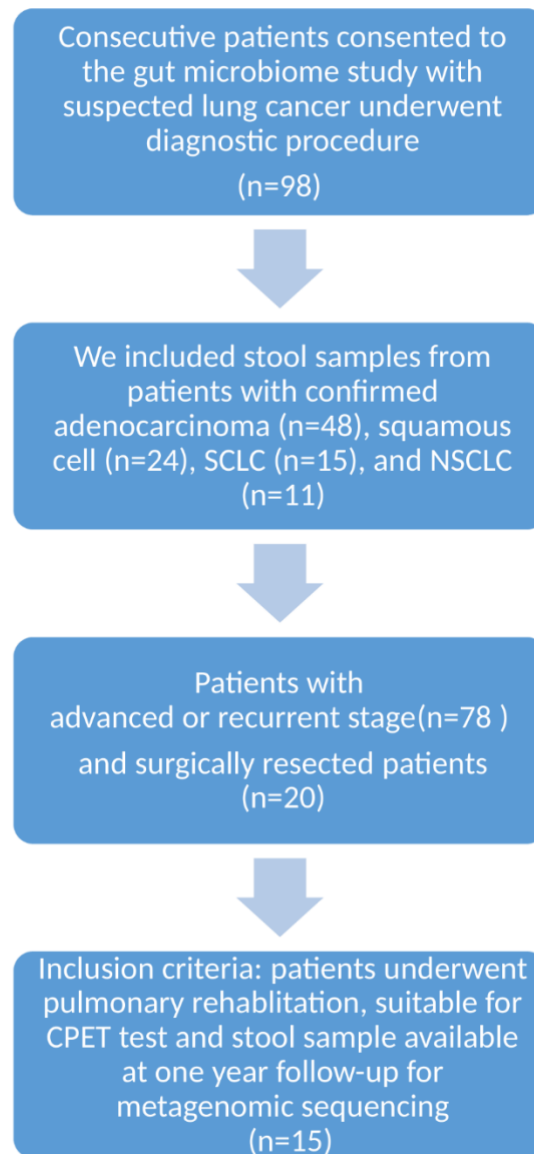
S1 Fig. Study design of surgically resected lung cancer patient cohort.
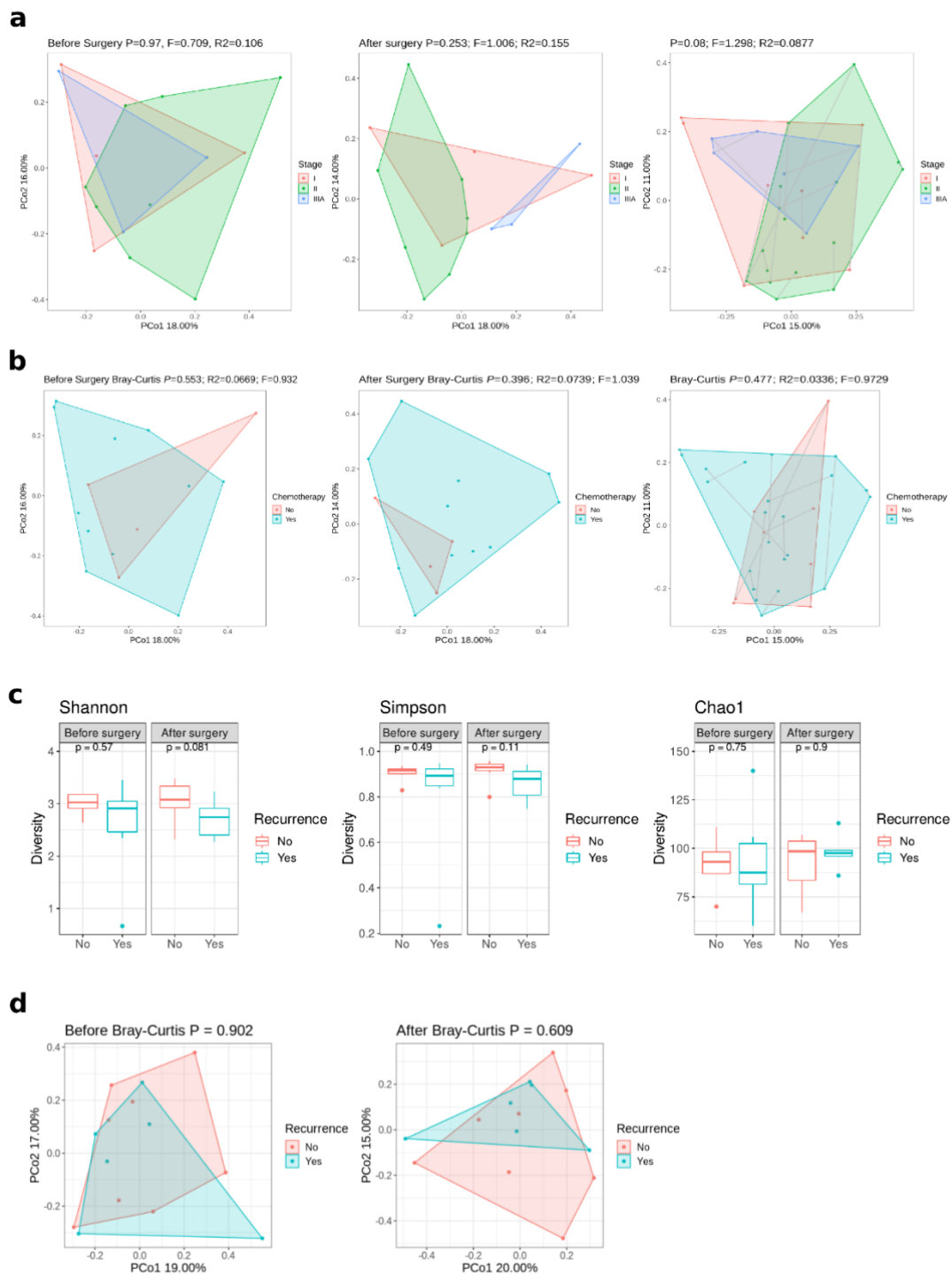
S2 Fig. Comparison of the gut microbiota composition pre- and post-surgery. (A, B) Principal Coordinate Analysis plot based on Bray-Curtis distances of (A) the different stages and (B) chemotherapy treatment. (C) Boxplots, with median (centrelines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), showing alpha diversity Shannon, Simpson, and Chao1 indices of recurrent and non-recurrent patients. (D) Principal Coordinate Analysis plot based on Bray-Curtis distances pre- and post-surgery of recurrent and non-recurrent patients.

S3 Fig. Correlations between taxonomic and functional profiles and CPET. (A) Heatmap of partial Spearman's rank correlation analysis between the fold-change of bacterial species and bacterial MetaCyc pathways versus the fold-change lung function parameters adjusting for COPD and cancer type. Only differentially abundant species and pathways (P<0.05, Wilcoxon signed-rank test) were used. (B) Heatmap of partial Spearman's rank correlation analysis between fungal species versus the fold-change of lung function parameters adjusting for COPD and cancer type. (A-B) Cell color indicates either negative correlation (blue) or positive correlation (red). Only species and pathways with significant correlations (P<0.05) are shown (*P<0.05, **P<0.01, ***P<0.001).

S4 Fig. Prediction of VO2, tumour recurrence, and overall survival (OS). Prediction of VO2, tumour recurrence, and overall survival (OS) from bacterial species (left), bacterial phyla (middle) or MetaCyc pathways (right) relative abundances.
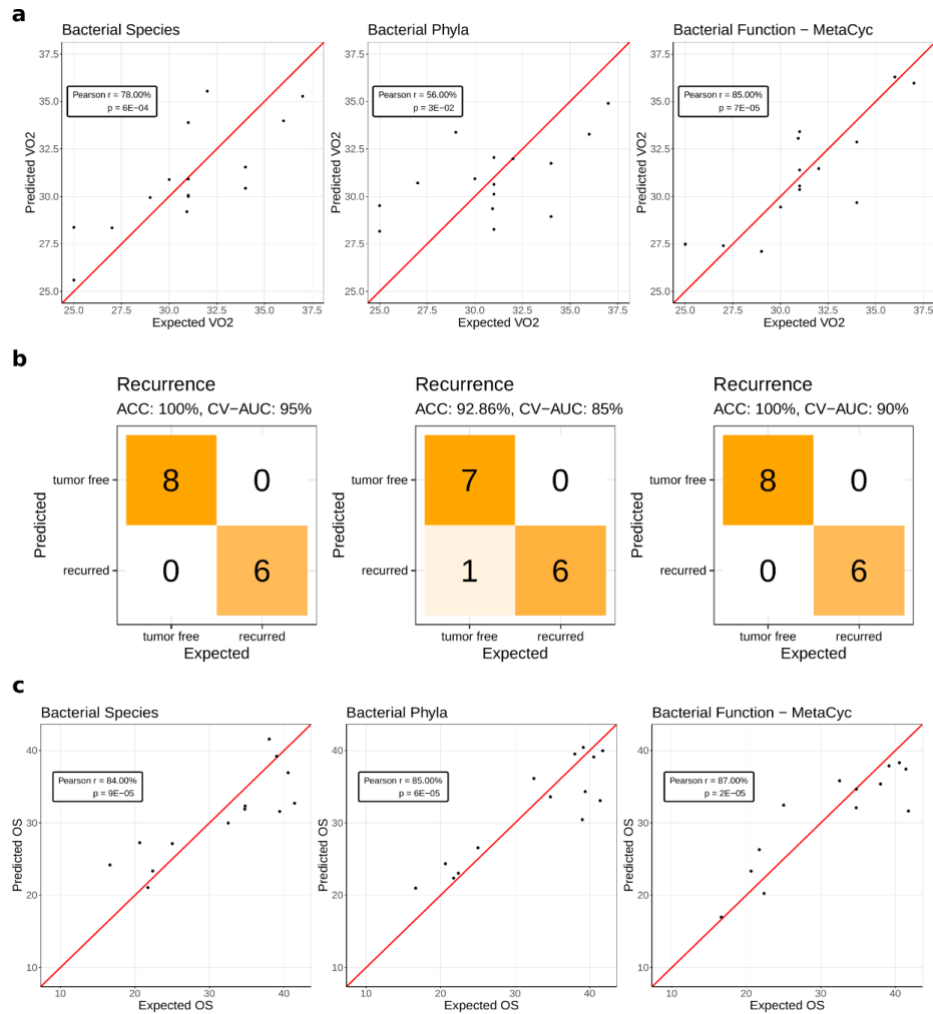
S5 Fig. Abundance of aerobic and anaerobic species. (A, B) Boxplots, with median (centrerelines), first and third quartiles (box limits) and 1.5x interquartile range (whiskers), showing the abundance of aerobic and anaerobic species. Gray lines connect samples of the same patient from before and after surgical resection.

IV. DISCUSSION

Not long ago, we believed that determining the DNA sequence of the entire human genome would be the key to understanding human biology [72]. However, our genomes are far more complex than anticipated and are insufficient to explain many common human conditions, including the development of immune system, diseases, and the success of drug treatments. Human microbiota is now recognized as a key factor in health and disease development and a critical expansion to the human system. To understand the role of microbiota, we need to identify and comprehend the conditions and mechanisms by which certain microbes switch from commensalism or mutualism to parasitism – the pathobiome. Within the last two decades, the number of described microbial species that live in and on humans increased by orders of magnitude. While uncommon, studies on archaea and bacteriophages are emerging, too. Since all biological life is based on DNA or RNA, all organisms can be detected, quantified, and analyzed using the same underlying technology.

The technologies used throughout my Ph.D. to study biological systems involve a combination of different sequencing technologies, including metagenomics, transcriptomics, amplicon-based sequencing, but also other OMICS data such as metabolomics. With the manuscripts included in this thesis, I have demonstrated the use of several complementary frameworks for the analysis of high-dimensional sequencing data to address different biological scientific research questions. Non-linear regression, machine learning, and co-abundance network analyses were frameworks that I applied consistently throughout my Ph.D. to identify possible interactions between microbes and their host, and how these interactions change over time or due to environmental stimuli.

This cumulative dissertation encompasses six manuscripts studying pathogens, microbial communities, and the host with increasing complexity. The first part of the discussion is about my research on the virulence mechanisms of one specific pathogen, *A. fumigatus* (**manuscripts I, II**). The second part of the discussion is about the intestinal ecology of gut microbiota, especially the community-level interactions between bacteria and fungi (**manuscripts III-VI**). In all my research work, even when the focus was on understanding the virulence and drug resistance of a specific pathogen, I considered the ecological context.

Aspergillus fumigatus Infection and Diversity

A. fumigatus is a human pathogenic fungus that causes a life-threatening infection in immunocompromised patients called *invasive aspergillosis* (IA). IA is one of the most common post-transplantation complications and therefore subject to intense research worldwide. In addition, IA treatment with azole drugs is limited due to mostly unknown but increasing azole resistance mechanisms [185]. In two **manuscripts (I, II)**, I addressed general questions concerning host cell immune defense against *A. fumigatus* and *A. fumigatus* diversity.

Adapting RNA-Seq to the simultaneous sequencing of three distinct organisms

The adaption of pathogens to the host, and the defense against invasion, involves a cascade of altered gene expression pathways in the interacting organisms. Before SGS technology, these investigations were limited to relatively few virulence factors. Then, microarrays enabled the quantification of transcriptomes, leading to new insights into virulence factors and the detection of novel pathogen-associated molecular patterns (PAMPs). In principle, this microarray technology could have been adapted to the simultaneous quantification of two organisms but was technically difficult and too expensive [95]. Soon after, RNA-Seq quickly overturned micro-array technology due to higher precision, greater resolution, flexibility and sensitivity, and ultimately lower cost. Still, *dual*-organism RNA-Seq approaches remained challenging. Technical limitations, especially in cell lysis and RNA extraction, required the separation of all organisms prior to sequencing, leading to considerable bias in organism-wise gene abundance profiles from cell isolation, DNA extraction, and sequencing [94,95]. But ultimately, *dual* RNA-Seq experiments were successful and applied to gain a deeper understanding of infection processes by creating single-infection models. So far, dual RNA-Seq was used to study mammalian host cells challenged with viral, bacterial, fungal, or eukaryotic pathogens [94]. Furthermore, *dual*-organism RNA-Seqs reduced experimental costs considerably because extraction and sequencing could be done together. Still, further technological improvements are required.

In clinical settings, humans are likely exposed to more than one pathogen at any given time [78]. During *A. fumigatus* infection, human CMV was suspected to frequently co-infect patients as well. To address this hypothesis, we developed the first *triple*-organism RNA-Seq that enabled the identification of crosstalk mechanisms between human monocyte-derived DCs and the two pathogens during single- and co-infections (**manuscript I**). This setup required a fine-tuned investigation of infection parameters (wet-lab side) and sequencing-related parameters (bioinformatics side).

At the wet-lab side, the amounts of pathogenic and host cells, as well as infection durations had to be identified. The latter was additionally addressed by sampling at varying infection durations. Sequencing parameters included the selection of RNA extraction and rRNA depletion protocols, sequencing depth requirements, and optimizing reference genomes for triple RNA-Seq mapping. Furthermore, I established the bioinformatic processing pipeline, including read processing, gene expression quantification, general visualizations, and differential abundance estimation. One of the primary results of the established pipeline is one RNA gene abundance matrix of all organisms that can be used for trans-species correlation networks and differential abundance testing. The profile can also be separated by species to investigate within-species effects. In both cases, subsequent analyses involved data ordination (*e.g.*, principal component analyses), differential gene expression analyses, functional enrichments, and network analyses.

Triple RNA-Seq enables a tri-partite view on DC, CMV, and A. fumigatus

The triple RNA-Seq (**manuscript I**) allowed the concurrent investigation of the transcriptome of each pathogen and host cells under varying conditions. We used dendritic cells (DCs) in my work because DCs are understood as the main bridge between (**a**) recognition of pathogens by innate immune cells and (**b**) adequate adaptive immunity response (clearance of pathogen) [75]. DCs are also considered a key element in successful defense against IA because they are exceptionally good at detecting *A. fumigatus* hyphae [75]. It was known that DC response to fungi differs substantially from those to viruses, and our work agreed largely with those expectations. Upon detecting proteins exposed in the cell wall of *Aspergillus* (especially their hyphae), DCs release predominantly Interleukins and TNF-α. This signaling cascade mobilizes the adaptive immune system, by which naïve T cells differential into TH17 cells. In stark contrast, detection of CMV antigens leads to the release of chemokines and the differentiation of naïve T cells into TH1 cells.

Interestingly, our study revealed that the co-infection phenotype of DCs is vastly distinct from either of the single infections and is likely the result of synergistic effects by both pathogens. A greater number of genes was significantly expressed during co-infection compared to single-infections. We also noticed that some PPRs critical for *Aspergillus* detection by DCs (and the subsequent anti-fungal response) were repressed in the presence of CMV. This might be because human CMV evolved mechanisms to modulate host-cell responses to increase the change for its' successful invasion. *Vice versa*, *A. fumigatus* hampered viral sensing cascades, leading to an overall greater rate of virus-infected DCs. Overall, we give evidence for the "synergy during co-infection" hypothesis and demonstrate that the knowledge gained from single-infection models may not translate well to co-infection conditions.

This new technology also allowed us to switch the perspective towards each pathogen. Under co-infection, we found that *A. fumigatus* had fewer differentially expressed genes compared to single infections. Most of the differentially expressed genes were the same between co-infection and single infection. This lower response could originate from several factors, including decreased host defense by DCs or lower colonization requirements. CMV, in contrast, did not show many differences between single- and co-infection. Since CMV is a virus with only around 280 genes, it may not have evolved mechanisms to change its behavior in the presence of other pathogens.

Gene co-expression networks to compare infections

The triple RNA-Seq approach potentially reduces systemic noise that may impede cross-species gene expression correlations when separately sampling host and pathogen transcriptomes [94]. I computed inter-species gene co-expression networks and investigated differences and commonalities in network topologies. In particular, I used set operations on nodes or edges to identify shared and unique correlations.

Co-infection networks shared only few cross-species with single-infection networks. Interestingly, very few correlations were specific to single *A. fumigatus* infection, but many CMV induced many unique cross-species correlations. Again, *A. fumigatus* seemed to benefit from CMV infection, but CMV did not benefit from *A. fumigatus* infection.

Perspectives – Clinical Relevance

The triple RNA-Seq (**manuscript I**) was used to study a clinically relevant co-infection and delivered hard evidence for the existence of synergistic effects between two human pathogens. However, our *in vitro* observations need further experimental confirmation in more complex models and *in vivo*. In patients, DCs and pathogens would be surrounded by a complex mixture of tissue (*e.g.*, alveoli), metabolites, peptides, and other immune cells, all of which may influence the behavior of host cells and pathogens.

Studying *A. fumigatus* lung infections under physiological conditions in humans and mice *in vivo* is very challenging due to variations in (**a**) the source of immunosuppression, (**b**) pathogen exposure, (**c**) co-infection status, (**d**) host genetics, and (**e**) treatment drug dosage doing and timing [185]. While studies on humanized mouse models are common, translating results from mouse models to humans requires caution [185]. Instead, we used an *in vitro* model system in which host cells were directly exposed to pathogens. While this setup is easier to replicate and insightful, host and pathogen cells may behave differently without other host-cells or their native conditions. An interesting next step would be *ex vivo* models on alveolar or other lung tissue. Such lung-on-a-chip models of human cells are being developed [186] could enable the study of human lung infection under more controlled infection parameters.

Perspectives - Further Improvements

The computational part of dual and triple RNA-Seq approaches is based on classical mapping approaches. It is worth noting that some recently developed quasi-mapping tools (*e.g.*, `Kallisto`) offer faster quantification and higher accuracy with more redundant reference genomes [187]. However, their accuracy in multi-organisms RNA-Seq was not evaluated thus far but will be interesting to investigate in future studies.

Aspergillus fumigatus demonstrates underappreciated diversity with potential consequences for future infection studies

Using one singular reference genome for comparative genomics in infections is common, but this approach is blind to the true genomic diversity observed in microbes. Due to low generation times, beneficial genomic changes (such as drug resistance) are acquired and distributed fast over microbial populations. While drug resistant *A. fumigatus* were observed in clinical settings, it is unclear if these are mechanisms are present in type-strains or wild-type strains.

Phenotypic variations in *A. fumigations* are expected to have consequences on disease development [188] and likely also for azole resistance [185]. However, thorough comparative genome analyses on isolates of different environments were lacking. To address this, we used whole-genome sequencing to assemble, annotate, and compare the genomes of 300 *A. fumigatus* isolates from different environmental and hospital sources (**manuscript II**). We have shown that *A. fumigatus* genomes are much more diverse than previously anticipated, with a core genome comprised of only 69% of pan-genes. I contributed to this work primarily by performing cluster enrichment tests. *A. fumigatus* did not cluster by country, which could result from frequent genomic exchange between fungal populations. This is quite likely because we had observed a high degree of genomic recombination events.

Importantly, clinical isolates formed a single cluster together with the Af293 type strain based on genetic variation in single-copy orthologous genes. This implies that (**a**) it could be the most representative reference for studying clinical infections under laboratory

conditions and (**b**) that human-associated *A. fumigatus* may require additional but common genomic adaptations to survive. However, clinical isolates were distributed over the entire phylogeny, reaffirming the ability of *A. fumigatus* to adapt to distinct environments.

Our study thereby indicates the need for further considerations in selecting specific *A. fumigatus* strains for infection studies. For the study of infection, *A. fumigatus* strains were primary selected based on their observed virulence. To identify and understand drug-resistance mechanisms, it will be crucial to study the structural and functional differences among *different A. fumigatus* strains in future studies. The triple RNA-Seq study (**manuscript I**) was based on the standard *A. fumigatus* Af293 reference genome, which was also used as comparator in our pan-genome study (**manuscript II**). While this genome represented clinical isolates well (**manuscript II**), it contained only 87% of pan-genes, indicating that analyses using Af293 could be blind to many genes and strain diversity. In addition, our infection study (**manuscript I**) involved *A. fumigatus* strain ATCC 46645, a strain very similar to Af293 but with differences in pan-genes.

A solution for future studies could be the creation and use of a general-purpose *A. fumigatus* pan-genome reference, as is already the case for *homo sapiens* [80].

Gut Microbiota in Health and Disease

In this thesis, I have shown multiple ways in which metagenomics can be used to investigate complex microbial systems and their interactions with the host. Our initial understanding of microbiota was largely supported by large-scale projects such as the Human Microbiome Project [7,36], which reported high temporal and spatial variation in gut microbiomes of humans and raised questions like: *How can this variation be explained*? *What constitutes a healthy microbiome*? A large body of studies associated changes in microbiota with many common diseases, including type 2 diabetes, inflammatory bowel diseases, heart disease, cancer, and autism spectrum disorder [22,25,30–35]. Genome sequencing technology enabled the investigations of complex microbiota and will likely continue to be a key technology for microbiome studies.

Defining gut fungal consortia and their stability, resilience, and dynamics may reveal cause-effect relationships with bacteria. Although evidence is available on bacteria-fungal interactions in the gut at the taxonomic level, we do not have a comprehensive understanding of how bacterial functions influence the growth of particular fungi. To better understand the entire microbiome, we provided data to follow both the bacterial and fungal communities of the lower human gastrointestinal tract (**manuscripts III-VI**). In the manuscripts included in this thesis, I analyzed large amounts of sequencing data, metabolomics, and phenotypes. I presented sophisticated pipelines to quantify whole metagenome, whole metatranscriptome, and ITS2 rDNA amplicon sequencing data to profile bacterial species, functional potential, functional expression, and fungal compositions. I presented methods to efficiently integrate this compositional data to improve our understanding of the underlying biological processes, which I will discuss further below.

Next-Generation Sequencing enables the study of microbial interactions

Most studies investigating the influence of gut microbiota used 16S rDNA amplicon sequencing to characterize bacterial communities [89]. While these studies generated many exciting hypotheses, the taxonomic resolution was limited to family and genus level abundances and could not elucidate microbial functions from the sequencing data. Higher taxonomic and functional resolutions are required to identify and understand microbiota-associated diseases and drug-resistance mechanisms (introduced in **sections 2.4ff**).

There are many ways in which genomic sequence data was analyzed, starting with analyzing diversity indices and structural differences in compositions (**manuscripts III-VI**). Species and functional information were combined into microbe-wise functional abundance profiles to estimate contributional alpha diversity [38,120] (**manuscripts III, VI**). Other concepts included transcriptional activity (**manuscript III**), machine learning (**manuscripts IV-VI**), and co-abundance network analysis (**manuscripts III-VI**).

Antibiotics drove fungal communities from mutualism to competition

Since the discovery of Penicillin in 1928 by Alexander Fleming, antibiotic treatment has emerged as a universal solution to the treatment and prevention of bacterial infections. However, the off-target effects on health-associated gut microbiota are still an ongoing debate and may not have a straightforward answer. Opportunistic bacteria are becoming increasingly resistant to antibiotics as a result of their widespread – and often pre-emptive – administration [189]. This resistance is a growing concern in the treatment of infectious

diseases and invasive procedures. Both intravenous and oral antibiotic administration have an impact on the largest reservoir of genetic diversity in the human body: the gut microbiome. In this thesis, I presented two studies (**manuscripts III, IV**) in which we studied the effects of antibiotic treatment on microbiota using joined quantification of gut bacteria and fungi. We identified several bacterial mechanisms promoting and inhibiting the overgrowth of fungal pathogens.

While the dynamics of bacterial abundance were documented well, fungal stability and resilience are mostly unclear [48]. Previous studies using mice models suggested a link between oral use of antibiotics and the overgrowth of opportunistic gut fungi, in particular, *Candida* [64,190,191]. This appears to be consistent with many, but not all, types of antibiotics [190,191]. However, it has yet to be understood if these effects are indirect – the changes in gut microbes result in the changes of the mycobiome – or if antibiotics can directly affect gut fungi. Microbiota can control the levels of other microbes in multiple ways, including the release of metabolic by-products such as bacteriocins, acids, and peptides [192] or by outcompeting other microbes for space, metabolites, and nutrients [193]. Intestinal microbes may also modulate host immune responses against other pathobionts [194] or induce the formation of the protective mucin layer that covers the gut epithelium.

In **manuscript III**, we followed healthy human subjects up to 90 days post treatment to assess community changes before, during, and after oral antibiotic administration. The short-term effects on the bacterial community were as expected. I observed a strong reduction in (**a**) bacterial taxonomic and functional alpha-diversity and (**b**) health-associated short-chain fatty acid (SCFA)-producing bacteria [195,196], but with variations dependent on the antibiotic class. I introduced the role of such anaerobes in **section 1.2 on page 4**. My findings were consistent with previous studies on human gut microbiota of similar design but different antibiotic drugs (Raymond *et al.* [197]) or the most potent antibiotic drug mix (Palleja *et al.* [56]). But unlike these two studies, we also characterized (**a**) microbial gene expression and (b) fungal composition.

Fungal communities exhibited a much higher degree of taxonomic variance compared to bacterial. However, we were able to attribute a significant part of that variation to antibiotic treatment. First, the number of observed fungal species almost doubled short-term (within one month post treatment). Second, by analyzing the topology of fungal co-abundance networks, I observed a shift from stable communities (mostly positive correlations) before treatment to a large increase in negative associations in the first month after treatment. At 90 days post treatment, most correlations disappeared. These observations support the idea of indirect antibiotic effects: since many bacterial taxa were effectively inhibited, several fungal species gained a chance to colonize the gut. We observed this as an increase in species numbers. However, during the recovery of gut bacteria, competition for space and nutrients reemerged, which puts more pressure on fungal species than bacterial species.

Gut Candida increased after antibiotic administration in healthy and diseased humans
Confirming the observation from mouse models [64,190,191], human gut levels of *Candida* increased in relative abundance shortly after antibiotic administration (**manuscript III**). However, *Candida* did not prevail in high abundance by the end of our study (90d),

confirming that the gut microbiota of healthy humans were resilient against pathogenic fungal overgrowth.

While microbiota of healthy patients showed resilience against adverse antibiotic effects, diseased peoples' may not, and treatment may do more harm than good. We thus investigated the robustness of gut microbiota in critically ill patients with and without antibiotic treatment and compared them to healthy subjects (**manuscript IV**). Critical illness revealed microbiota that differed significantly from those of healthy people, with some alterations accentuated by antibiotic therapy. In critically ill patients, antibiotic administration led to an "infection vulnerable" microbiome composition characterized by extremely low levels of SCFAs.

In both studies (**manuscripts III, IV**), we observed reduced relative abundance in short-chain fatty acid- and secondary bile acid-producing bacterial species during antibiotic administration, species essential for maintaining host homeostasis [195,196]. On the fungal side, *Candida* species' relative abundance increased in the critically ill but even further under additional antibiotic administration (**manuscript IV**). Together, the results of both **manuscripts (III, IV)** pointed to systematic and potentially detrimental effects on human gut microbiota as a direct consequence of antibiotic administration and support the idea that a decrease in anaerobe gut bacteria could allow enhanced colonization of gut fungi.

Perspectives - Counter-acting adverse antibiotic effects

The solution to (**a**) the increased incidence of microbial infections [198] and (**b**) stagnating progress in the development of new antibiotic drugs [199] could be the identification of new drug types and targets. Antimicrobial peptides [198], for example, are proteins produced by commensal microbes to combat pathogens. Gut microbiota, especially bacteria, are believed to be a rich reservoir for such compounds because of (**a**) their competitiveness against constantly invading microbes and (**b**) a large number of genes without known functions [198]. Another approach could be bacteriophage therapy [200]. Therefore, phages with high specificity to a pathogen are selected (and modified) to kill specific pathogens. The recent release of a metagenomic gut virus catalog indicates that over 90% of assembled viruses show no similarity to existing phylogenies [201], showing just how much more room for future studies is left.

Microbial Growth Rate Estimations separate dead wate from proliferators

One intrinsic issue of current metagenomic research is our inability of DNA-based abundance data to differentiate between living, stationary or dead cells [126]. Non-replicating cells can substantially influence relative abundance profiles [126]. This needs to be considered in studies using antibiotics. If the drug kills bacteria or inhibits their growth (as expected), the DNA of bacteria affected by the drug can still be abundant. To mitigate this issue, we have used *in silico* growth rate estimation (GRiD) in **manuscripts III** and **IV** (introduced in **section 2.5**). Due to high runtime requirements, we applied it only to species that showed differential abundance changes after antibiotic treatment. GRiD estimates were then used to (**a**) exclude species with low estimated growth from further analyses and (**b**) identify bacteria proliferating after treatment. It should be stressed that such GRiD can only make approximations of bacterial growth and does not present strain-level resolution. Furthermore, the high runtime of GRiD may hinder its widespread use. Still, *in silico* growth rate estimations are helpful in cases where microbial growth cannot easily be assessed, such as *in vivo* gut microbiota studies.

Subspecies quantification using MGS

Early work on whole metagenome studies was challenged by the huge number of novel genes originating from uncultured organisms and new strains. Several approaches to accommodate the unknown were introduced, from comprehensive protein-coding gene family reference libraries such as the IGC or `HUMAnN` (**on page 13**) to new species concepts such as MGS (introduced in **section 2.6 on page 14**). While the `HUMAnN` offers quantification of gene family abundances per species, it is not a complete solution. `HUMAnN` is based on a fixed pool of non-redundant pan-gene catalogs, which is difficult to expand with additional genes from *de novo* genome assemblies or specialized databases. For example, identifying and annotating antibiotic resistance genes (ARG) [183] was crucial to investigate the effects of antibiotic drugs on gut microbiota (**manuscripts III**, **IV**). Likewise, most gene family references offer very limited taxonomic information.

To identify the microbial species possessing ARG genes, we grouped genes by their co-abundance into MGS [131] (**manuscripts III**, **IV**). While the annotation of MGS usually stops at the species level [131,202], I demonstrated in **manuscript III** that an MGS could, in some cases, be annotated to the subspecies level by adapting the concept of pan-genomes [123]. Thereby, species were delineated based on MGS genes' presence/absence pattern compared to a pan-genome reference of the species assigned to the MGS. The approach could identify if an MGS represented a species *core* (genes shared by all strains), a *strain* (genes found in exactly one reference strain out of multiple), or a *sub-species* (genes present in some strains but not all).

Metatranscriptomes of microbes during antibiotic perturbation

Previous studies on human gut microbe gene *expression* found substantial differences in taxonomic and genomic profiles compared to those derived from metagenomics [38,203]. While a common narrative considers highly abundant species to have high impacts on microbial communities, it does not explain how species of low DNA abundance (*e.g.*, many SCFA producers) could have significant roles in host homeostasis. It was suggested that only a subset of species within a gut community might express essential genes [38,117,204]. In addition, metatranscriptomes of human gut bacteria were not investigated during antibiotic administration. To address some of these ideas, I have characterized and analyzed microbial the temporal properties of metatranscriptomes during antibiotic administration (**manuscript III**). On several occasions, I observed striking differences between metagenomic and metatranscriptomic data. For example, functional bacterial diversity (estimated from DNA) was reduced during antibiotic treatment, whereas transcriptomic diversity of bacterial function showed overall greater variance and greater resilience. In addition, my analysis of metabolic core pathways indicated many pathways whose transcriptional activity was virtually unchanged even under antibiotic treatment.

Perspectives - Resolution of metatranscriptomics needs further improvements

Studies using transcriptional activity resulted in conclusions that were different from results obtained using both types of sequencing data separately. However, it is vital to consider the *qualitative* differences between RNA and DNA data. In my study (**manuscript III**), RNA reads did not yield the same species-level resolution as DNA. While 60-80% of DNA reads were assigned to species-level contributions per sample, only 10-20% of RNA reads were. This could have been the result of three major factors: (**a**) post-transcriptional modifications, which lead to larger divergence between reference gene sequences and their reads, and (**b**) nucleotide degradation, which is often non-linear and leads to major distortions in profiling unless accounted for [118], and (**c**) insufficient resolution of the

`HUMAnN2` reference. Point (**b**) could potentially be addressed using RNA integrity number (RIN) as a covariate for linear regression, which was suggested to correct the resulting bias in single organism RNA-Seq [118]. Point (**c**) could be verified in future projects using updated gene-family catalogs such as the recent `HUMAnN3` release, which contains more than double the number of reference genes [123].

Perspectives – Co-abundance network analyses

The integration of joined bacterial-fungal abundance data is not solved with satisfaction, and methodological progress in trans-kingdom correlation analysis using relative abundance data analysis is scarce [150]. In theory, sequencing all organisms together at sufficient depth could enable the use of simpler and existing approaches with less biased outcomes [169]. But practically, DNA extraction protocols are often optimized and thus biased towards either bacteria or fungi [46,205,206]. Some tools quantifying multiple kingdoms from metagenomic data were published in recent years [126,207,208], and the resulting abundance profiles were used for cross-domain correlation analyses [169]. However, the accuracy of multi-kingdom quantification is debatable [209]. In my studies, including unpublished pilot work, the abundance profiles created by multi-kingdom estimators contradicted amplicon-based results and reported several unreasonable fungal taxa. Therefore, it may be necessary to research the accuracy of these tools on sufficiently complex but well-known model communities. In my work, we quantified bacterial and fungal compositions using two independent sequencing approaches (ITS2 and WMS). I consider both (holistic and separate) profiling approaches justified until more sophisticated approaches of these methodological approaches are available.

In my studies (**manuscripts III, V**), I performed trans-kingdom co-abundance network analyses using either (**a**) simple Spearman correlation on *per-domain* log-ratio transformed abundance data (*e.g.*, with CLR), or (**b**) more complex models that assume sparsity to increase statistical power. In the case of the latter, I applied `BAnOCC` on joined bacterial-fungi abundance data (**manuscript III**). The approach was robust against varying degrees of prevalence filtering. However, the number of features useable for analysis is very limited to, at most, a few hundred due to (**a**) quadratic growth in computational runtime and (**b**) very limited parallel computation options. Simpler approaches like Spearman correlations have the advantage that software optimized for speed and memory efficiency already exist. With these, correlations for thousands of features are easily estimated. Many sources of bias (**section 3**) were mitigated using data normalization such as log-ratio transformations (*e.g.*, CLR in **manuscript V**).

There are more approaches worth considering for future studies as well, most of which were summarized by Matchado *et al.* [150]. For example, `FlashWeave` was designed for studies with thousands of samples and features. It allows controlling for additional covariates, which is challenging with most existing correlation approaches [150]. Another method is an extension of sparse partial least square analyses, "`DIABLO`" [150]. `DIABLO` was used for multi-omics integration of microbiome, mRNA, and metabolomics data [210]. While an approach like `DIABLO` could have higher statistical power, the method requires a high degree of manual parameter tuning and considerations on data normalization, which again requires solutions to compositionality and zero-inflation bias. Another solution could be shifting from *relative* to *absolute* observed abundance through spike-in methods that I discuss **on page 204**.

Gut oxygen levels are important to understand Candida colonization

One group of microbes consistently showed differences in abundance in the **manuscripts (III-VI)** of this thesis: obligate anaerobes [18]. Many of these precise ferment undigestible food into SCFAs, some of which are consumed by colonocytes and therefore believed to be the main drivers of gut hypoxia [22] [40] [41]. These microbes are also very susceptible to a wide range of common antibiotics, as shown by us (**manuscripts III, IV**) and others [202,211]. In addition, we associated a decrease in SCFA producers with lung tumor recurrence (**manuscript VI**) and an increase in *Candida* species abundance (**manuscripts III-V**). The latter is quite interesting given the different settings employed across these manuscripts.

Theoretically, while bacteria and fungi compete for resources available in the gut lumen, they may also support one another. In **manuscript V**, we investigated gut microbial signatures explaining varying levels of gut *Candida* abundance. Again, we observed a systematic reduction in strict anaerobe abundance in samples with higher *Candida* levels but also an increase in several oxygen-tolerant bacteria. We proposed that increased gut oxygen (caused by a reduction in SCFAs) would allow *Candida* species to grow on lactate. Lactate would be produced by facultative anaerobe lactic acid bacteria such as *Lactobacillus*, many of which increased in relative abundance in samples with relatively high *Candida* levels.

While a connection between gut oxygen levels and pathobiosis was hypothesized [59], the idea remains mostly unaddressed. We demonstrated in further experiments that *Candida albicans* grows efficiently on sole-lactate under microaerobic (1% O_2) conditions, while *Lactobacillus* species enhance gut integrity. Together, our findings could explain one way in which *Candida* abundance could increase in the gut without becoming pathogenic.

Identification of microbial balance using machine learning

In **manuscript VI**, we shifted our attention to associations between treatment outcomes and microbiota. We used a longitudinal study design to characterize microbial changes related to lung function recovery one year after lung tumor resection. An intense testing procedure (cardiopulmonary exercise testing) was employed to assess several lung function parameters, alongside other patient characteristics such as blood CO_2 volume. Our data suggested a link between a reduction in anaerobe abundance and tumor recurrence, contributing evidence to the hypothesis of beneficial effects of strict anaerobes. While our analysis strategy was similar to the studies presented thus far, we used an additional method for the identification of small sets of microbes (or functions) associated with host traits (*e.g.*, VO_2) – *microbial balances* [212].

To identify robust predictors of lung function, I calculated "balances" [156,212]. Conceptually, balances are based on the idea that the abundance of one or more specific species can be used to calibrate the abundance of the other taxa. However, gut microbiome data is sparse, of high variance, and localized. Hence, multiple such balances may exist across samples simultaneously. In **manuscript VI**, I used the machine learning tool `selbal`, which aims to identify a set of microbes whose balance is predictive of a trait [156,212]. Here, a balance is a real-valued number calculated as the log-ratio of the sum of nominator and sum of denominator species [156,212]. A generalized linear model is trained with these balances. I used this procedure to identify correlations between balances of microbial species or functions with lung function parameters.

Perspectives - Quantification of fungal compositions: ASVs vs OTUs
Even with current sequencing technology, the quantification of fungal composition is still challenging and primarily achieved through ITS amplicon sequencing approaches (**section 2.3**) [50]. While ASVs are, in theory, superior to OTUs in every possible way [108], my experiences were both positive and negative. For sequencing data generated by older Illumina sequencing machines, ASVs performed best (**manuscripts IV, V**). However, some Illumina machines, especially newer models, perform additional quality binning to reduce file size[1], which broke the error model of DADA2 and directly affected variant calling. Since the DADA2 ASV pipeline is easy to use, this problem can be missed because the tool does not report problems with the error model and relies on the visual inspection of the user. I did find a way to correct the error model to my satisfaction. However, I could confirm that denoising approaches are better than OTU clustering if the underlying error models are specified correctly.

Perspectives - On the switch between transcriptomics and genomics
Careful readers may have noticed that studies on a single or few organisms are often based on the *transcriptomes* of organisms (**manuscript I**). In contrast, microbiota studies are based on meta*genomes* of organisms (**manuscripts III-VI**). While large-scale reference genomes for gut microbiota became available in recent years, DNA-Seq-based analyses prevail in microbiota research. I can only speculate why this is the case, given my experience throughout my Ph.D.

For rRNA marker-gene studies, RNA-Seq would indeed make little sense. The rate at which rRNA molecules are expressed by microbiota is highly variable and rarely of interest. It thus makes sense that DNA-Seq has become the *de facto* standard for SSU, LSU, and ITS amplicon-sequencing studies.

But for studies on the genes and genomes of organisms, we should reevaluate our goals. The study of genomics of an organism starts with DNA-Seq to reconstruct the genomes of isolates. This is common practice and useful to study the functional potential of microbes (**manuscripts II-VI**). But once high-quality genomes are available, transcriptomes (and proteomes) are of gain greater interest because they give us more information about the *actual* processes happening within the organism (**manuscript I, III**). One key issue in my studies was the (expected) discrepancy between RNA-derived reads and reference genomes due to post-transcriptional modifications. While a few mismatches between sequences can be handled with non-redundant reference genomes (as in **manuscript I**), this is not the case when the reference describes hundreds of species or millions of reference genes. In my studies (**manuscript III**), including unpublished pilot studies, metatranscriptome accuracy was accurate only at the gene family level but had limited accuracy for species-level deconvolution. Still, metatranscriptome studies may gain popularity with the use of 3rd generation sequencing technology. The ability to quantify entire transcripts may improve the resolution problem in the near future.

Perspectives - 3rd generation sequencing to improve amplicon-based studies
3rd generation sequencing machines offer the sequencing of much longer DNA fragments of up to 150kb length but at the cost of increased sequencing errors. Long-read sequencing was predominantly used for improved reconstruction of genomes. However, advancements in error control are making this technology suitable for meta-barcoding, paving the way for

---

[1] https://www.illumina.com/documents/products/whitepapers/whitepaper_datacompression.pdf

full-length amplicon sequencing studies. Single-molecule real-time (SMRT) sequencing (PacBio) can sequence 13kb fragments with 99.8% accuracy, almost matching the quality of SGS with much greater reads lengths [88]. This applies to all types of rRNA gene amplicons, including 16S (bacteria), 18S (eukaryotes), and full size ITS (eukaryotes), all of which are less than 5kb long. Hence, it is becoming a suitable technology for whole metagenomic sequencing. This notwithstanding, the analysis of long reads with existing gene catalog approaches will require substantial changes in read quantification procedures. For example, algorithms need to handle (**a**) differences in lengths between consensus genes and query reads and (**b**) greater numbers of mismatches. But if successful, it will significantly improve the accuracy and resolution of gene abundance and species estimates.

Perspectives - Microbial Spike-in to estimate the microbial load

The fact that sequencing depth cannot give us meaningful information about samples densities and cell counts creates a considerable problem in studies using samples with largely varying microbial loads [152]. This especially includes studies with low biomass samples (tissue metagenomics, lung microbiome, skin) and studies using microbe-targeting drugs (*e.g.*, antimicrobials) in which the microbial population density changes between pre- and post-treatment sampling points. One solution involves estimating bacterial load using qPCR of 16S genes [213]. While it is one of the simplest solutions, it has several drawbacks. In a complex ecosystem, qPCR is argued to introduce bias through the extraction, purification, and amplification of DNA, as well as variation in the 16S rRNA copy numbers and replication rates [214]. Furthermore, only one kingdom can be quantified in one process. Directly counting cells using flow cytometry is another option [214] but requires sufficient left-over material because cells used in this process cannot be used for sequencing. Ideally, cell count or density estimations are tied with the quantification of microbes. One emerging concept is the spiking of a fixed number of bacterial (or fungal) cells from one or more species that aren't present in the samples. After sequencing, the proportion of spike-in bacteria can be compared to their expected numbers to derive a reasonable estimate of the total number of cells sequenced from the sample. Each spiked-in bacterium serves as internal control – cell number estimates derived from either bacterium should be similar or indicate extraction bias otherwise. The latter is important because spike-in methods are still evolving, and some work from 2018 found striking differences in estimated cell numbers depending on the spike-in kit [213]. This should not discourage the use of this technology: it merely implies that experienced companies should do spike-ins until robust protocols are established.

Perspectives - Spatial-aware Microbiome

While stool samples are a non-invasive and accessible medium to characterize gut microbiota, they are not the best to capture gut microbiota. Due to the nature of our digestive tract, stool samples represent a mixture of microbiota from different parts of the gut. In addition, many disease-related differences (*e.g.*, those related to bile acid metabolism) start in the upper parts of the GI, such as the small intestine. These differences are unlikely to be captured by stool samples due to (**a**) a much smaller number of microbes compared to distal parts and (**b**) all the metabolic processing that occurs in the stool on the way through the GI. In addition, differences in stool consistency can lead to profound differences in the extracted material and potentially misleading conclusions about their microbiota.

One solution could be using small, programmable, collecting devices that are given to study subjects in pill form. One such technology is IntelliCap© which can take a small

sample of intestinal content when exposed to a well-defined combination of factors, including heat, pH, or reaching a specific position [215]. Such a device could substantially improve the way we study gut microbiota *in vivo*. While uncommon, I expect several devices to reach the market and academia in the following years.

Concluding Remark

In this thesis, several hypotheses on microbial interactions with themselves and the host were generated using next-generation sequencing technology that may serve as a basis for future experimental studies. Both RNA-Seq and metagenomic pipelines can be used to study microbial communication across varying scales. In the future, I expect a shift towards 3rd generation sequencing, advanced tissue models, and RNA-based microbiome studies, all of which are fascinating technologies to develop, use, and improve.

BIBLIOGRAPHY

[1] Seelbinder B, Wallstabe J, Marischen L, et al. **Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model**. In: *Cell Rep*. 33 (2020), pp. 108389.

[2] Barber AE, Sae-Ong T, Kang K, et al. **Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection**. In: *Nat Microbiol 2021 612*. 6 (2021), pp. 1526–36.

[3] Seelbinder B, Chen J, Brunke S, et al. **Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria**. In: *Microbiome*. 8 (2020), pp. 1–20.

[4] Marfil-Sánchez A, Zhang L, Alonso-Pernas P, et al. **An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness**. In: *Gut Microbes*. 13 (2021).

[5] Seelbinder B, Lohinai Z, Vazquez-Uribe R, et al. **Candida expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions**. [in prepar (2022).

[6] Marfil-Sá nchez A, Seelbinder B, Ni Y, et al. **Gut microbiome functionality might be associated with exercise tolerance and recurrence of resected early-stage lung cancer patients**. In: *PLoS One*. 16 (2021), pp. e0259898.

[7] Morgan XC**,** Segata N**,** Huttenhower C. **Biodiversity and functional genomics in the human microbiome**. In: *Trends Genet*. 29 (2013), pp. 51–8.

[8] Pitlik SD**,** Koren O. **How holobionts get sick-toward a unifying scheme of disease**. In: *Microbiome*. 5 (2017), pp. 64.

[9] Ruggiero MA, Gordon DP, Orrell TM, et al. **A Higher Level Classification of All Living Organisms**. In: *PLoS One*. 10 (2015), pp. e0119248.

[10] Knight R, Vrbanac A, Taylor BC, et al. **Best practices for analysing microbiomes**. In: *Nat Rev Microbiol* (2018), pp. 1.

[11] Gloor GB, Macklaim JM, Pawlowsky-Glahn V, et al. **Microbiome datasets are compositional: and this is not optional**. In: *Front Microbiol*. 8 (2017), pp. 2224.

[12] Chen J, Bittinger K, Charlson ES, et al. **Associating microbiome composition with environmental covariates using generalized UniFrac distances**. In: *Bioinformatics*. 28 (2012), pp. 2106–13.

[13] Sender R**,** Fuchs S**,** Milo R. **Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans**. In: *Cell*. 164 (2016), pp. 337–40.

[14] Zhu B**,** Wang X**,** Li L. **Human gut microbiome: the second genome of human body**. In: *Protein Cell*. 1 (2010), pp. 718–25.

[15] Sender R**,** Fuchs S**,** Milo R. **Revised Estimates for the Number of Human and Bacteria Cells in the Body**. In: *PLOS Biol*. 14 (2016), pp. e1002533.

[16] Byrd AL**,** Belkaid Y**,** Segre JA. **The human skin microbiome**. In: *Nat Rev Microbiol 2018 163*. 16 (2018), pp. 143–55.

[17] Sulaiman I**,** Schuster S**,** Segal LN. **Perspectives in lung microbiome research**. In: *Curr Opin Microbiol*. 56 (2020), pp. 24–9.

[18] Mowat AM**,** Agace WW. **Regional specialization within the intestinal immune system**. In: *Nat Rev Immunol 2014 1410*. 14 (2014), pp. 667–85.

[19] Durack J**,** Lynch S V. **The gut microbiome: Relationships with disease and opportunities for therapy**. In: *J Exp Med*. 216 (2019), pp. 20–40.

[20] Choo JM**,** Leong LEX**,** Rogers GB. **Sample storage conditions significantly influence faecal microbiome profiles**. In: *Sci Reports 2015 51*. 5 (2015), pp. 1–10.

[21] von Martels JZH, Sadaghian Sadabad M, Bourgonje AR, et al. **The role of gut microbiota in health and disease: In vitro modeling of host-microbe interactions at the aerobe-anaerobe interphase of the human gut**. In: *Anaerobe*. 44 (2017), pp. 3–12.

[22] Ivanov II**,** Honda K. **Intestinal Commensal Microbes as Immune Modulators**. In: *Cell Host Microbe*. 12 (2012), pp. 496–508.

[23] Vasapolli R, Schütte K, Schulz C, et al. **Analysis of Transcriptionally Active Bacteria Throughout the Gastrointestinal Tract of Healthy Individuals**. In: *Gastroenterology*. 157 (2019), pp. 1081-1092.e3.

[24] Manson JM**,** Rauch M**,** Gilmore MS. **The Commensal Microbiology of the Gastrointestinal Tract**. In: *Adv Exp Med Biol*. 635 (2008), pp. 15–28.

[25] Lozupone CA, Stombaugh JI, Gordon JI, et al. **Diversity, stability and resilience of the human gut microbiota**. In: *Nat 2012 4897415*. 489 (2012), pp. 220–30.

[26] DAS B**,** Nair GB. **Homeostasis and dysbiosis of the gut microbiome in health and disease**. In: *J*

Biosci. 44 (2019), pp. 1–8.

[27]   Yatsunenko T, Rey FE, Manary MJ, et al. **Human gut microbiome viewed across age and geography**. In: *Nat 2012 4867402*. 486 (2012), pp. 222–7.

[28]   Olszak T, An D, Zeissig S, et al. **Microbial exposure during early life has persistent effects on natural killer T cell function**. In: *Science (80-)*. 336 (2012), pp. 489–93.

[29]   Candela M, Perna F, Carnevali P, et al. **Interaction of probiotic Lactobacillus and Bifidobacterium strains with human intestinal epithelial cells: Adhesion properties, competition against enteropathogens and modulation of IL-8 production**. In: *Int J Food Microbiol*. 125 (2008), pp. 286–92.

[30]   Thingholm LB, Rühlemann MC, Koch M, et al. **Obese Individuals with and without Type 2 Diabetes Show Different Gut Microbial Functional Capacity and Composition**. In: *Cell Host Microbe*. 26 (2019), pp. 252- 264.e10.

[31]   ALLEN JM, MAILING LJ, NIEMIRO GM, et al. **Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans**. In: *Med Sci Sport Exerc*. 50 (2018), pp. 747–57.

[32]   Mirsepasi-Lauridsen HC, Vallance BA, Krogfelt KA, et al. **Escherichia coli pathobionts associated with inflammatory bowel disease**. In: *Clin Microbiol Rev*. 32 (2019).

[33]   Zhao L, Zhang F, Ding X, et al. **Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes**. In: *Science (80-)*. 359 (2018), pp. 1151–6.

[34]   Shaikh FY, Gills JJ, Sears CL. **Impact of the microbiome on checkpoint inhibitor treatment in patients with non-small cell lung cancer and melanoma**. In: *EBioMedicine*. 48 (2019), pp. 642–7.

[35]   Cruet-Hennequart S, Glynn MT, Murillo LS, et al. **Enhanced DNA-PK-mediated RPA2 hyperphosphorylation in DNA polymerase η-deficient human cells treated with cisplatin and oxaliplatin**. In: *DNA Repair (Amst)*. 7 (2008), pp. 582–96.

[36]   Huttenhower C, Gevers D, Knight R, et al. **Structure, function and diversity of the healthy human microbiome**. In: *Nat 2012 4867402*. 486 (2012), pp. 207–14.

[37]   Moya A, Ferrer M. **Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance**. In: *Trends Microbiol*. 24 (2016), pp. 402–13.

[38]   Abu-Ali GS, Mehta RS, Lloyd-Price J, et al. **Metatranscriptome of human faecal microbial communities in a cohort of adult men**. In: *Nat Microbiol*. 3 (2018), pp. 356.

[39]   Neish AS. **Microbes in Gastrointestinal Health and Disease**. In: *Gastroenterology*. 136 (2009), pp. 65–80.

[40]   Bindels LB, Porporato P, Dewulf EM, et al. **Gut microbiota-derived propionate reduces cancer cell proliferation in the liver**. In: *Br J Cancer 2012 1078*. 107 (2012), pp. 1337–44.

[41]   Litvak Y, Byndloss MX, Bäumler AJ. **Colonocyte metabolism shapes the gut microbiota**. In: *Science (80-)*. 362 (2018).

[42]   Wang L, Christophersen CT, Sorich MJ, et al. **Elevated fecal short chain fatty acid and ammonia concentrations in children with autism spectrum disorder**. In: *Dig Dis Sci*. 57 (2012), pp. 2096–102.

[43]   Hamad I, Raoult D, Bittar F. **Repertory of eukaryotes (eukaryome) in the human gastrointestinal tract: taxonomy and detection methods**. In: *Parasite Immunol*. 38 (2016), pp. 12–36.

[44]   Hamad I, Ranque S, Azhar EI, et al. **Culturomics and Amplicon-based Metagenomic Approaches for the Study of Fungal Population in Human Gut Microbiota**. In: *Sci Rep*. 7 (2017), pp. 1–8.

[45]   Richard ML, Sokol H. **The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases**. In: *Nat Rev Gastroenterol Hepatol 2019 166*. 16 (2019), pp. 331–45.

[46]   Yang F, Sun J, Luo H, et al. **Assessment of fecal DNA extraction protocols for metagenomic studies**. In: *Gigascience*. 9 (2020).

[47]   Shuai M, Fu Y, Zhong H-L, et al. **Mapping the human gut mycobiome in middle-aged and elderly adults: multiomics insights and implications for host metabolic health.** In: *Gut* (2022), pp. 1–9.

[48]   Hallen-Adams HE, Suhr MJ. **Fungi in the healthy human gastrointestinal tract**. In: *Virulence*. 8 (2017), pp. 352–8.

[49]   Paterson MJ, Oh S, Underhill DM. **Host–microbe interactions: commensal fungi in the gut**. In: *Curr Opin Microbiol*. 40 (2017), pp. 131–7.

[50]   Nilsson RH, Anslan S, Bahram M, et al. **Mycobiome diversity: high-throughput sequencing and identification of fungi**. In: *Nat Rev Microbiol*. 17 (2019).

[51]   Seed PC. **The Human Mycobiome**. In: *Cold Spring Harb Perspect Med*. 5 (2015), pp. a019810.

[52]   Overstreet RM, Lotz JM, Overstreet RM, et al. **Host–Symbiont Relationships: Understanding the Change from Guest to Pest** (2016), pp. 27–64.

[53]   Bass D, Stentiford GD, Wang HC, et al. **The Pathobiome in Animal and Plant Diseases**. In: *Trends Ecol Evol*. 34 (2019), pp. 996–1008.

REFERENCES

[54]   Tenaillon O, Skurnik D, Picard B, et al. **The population genetics of commensal Escherichia coli**. In: *Nat Rev Microbiol 2010 83*. 8 (2010), pp. 207–17.

[55]   Sovran B, Planchais J, Jegou S, et al. **Enterobacteriaceae are essential for the modulation of colitis severity by fungi**. In: *Microbiome*. 6 (2018), pp. 152.

[56]   Palleja A, Mikkelsen KH, Forslund SK, et al. **Recovery of gut microbiota of healthy adults following antibiotic exposure**. In: *Nat Microbiol*. 3 (2018), pp. 1255–65.

[57]   Zaborin A, Smith D, Garfield K, et al. **Membership and behavior of ultra-low-diversity pathogen communities present in the gut of humans during prolonged critical illness**. In: *MBio*. 5 (2014).

[58]   Levy M, Kolodziejczyk AA, Thaiss CA, et al. **Dysbiosis and the immune system**. In: *Nat Rev Immunol 2017 174*. 17 (2017), pp. 219–32.

[59]   Rigottier-Gois L. **Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis**. In: *ISME J 2013 77*. 7 (2013), pp. 1256–61.

[60]   Fassarella M, Blaak EE, Penders J, et al. **Gut microbiome stability and resilience: elucidating the response to perturbations in order to modulate gut health**. In: *Gut*. 70 (2021), pp. 595–605.

[61]   Li X, Leonardi I, Semon A, et al. **Response to Fungal Dysbiosis by Gut-Resident CX3CR1+ Mononuclear Phagocytes Aggravates Allergic Airway Disease**. In: *Cell Host Microbe*. 24 (2018), pp. 847- 856.e4.

[62]   Kim YG, Udayanga KGS, Totsuka N, et al. **Gut dysbiosis promotes M2 macrophage polarization and allergic airway inflammation via fungi-induced PGE2**. In: *Cell Host Microbe*. 15 (2014), pp. 95–102.

[63]   Dollive S, Chen Y-Y, Grunberg S, et al. **Fungi of the Murine Gut: Episodic Variation and Proliferation during Antibiotic Treatment**. In: *PLoS One*. 8 (2013), pp. e71806.

[64]   Noverr MC, Noggle RM, Toews GB, et al. **Role of antibiotics and fungal microbiota in driving pulmonary allergic responses**. In: *Infect Immun*. 72 (2004), pp. 4996–5003.

[65]   Mayer FL**,** Wilson D**,** Hube B. **Candida albicans pathogenicity mechanisms**. In: *Https://DoiOrg/104161/Viru22913*. 4 (2013), pp. 119–28.

[66]   Kumamoto CA**,** Gresnigt MS**,** Hube B. **The gut, the bad and the harmless: Candida albicans as a commensal and opportunistic pathogen in the intestine**. In: *Curr Opin Microbiol*. 56 (2020), pp. 7–15.

[67]   Pappas PG, Lionakis MS, Arendrup MC, et al. **Invasive candidiasis**. In: *Nat Rev Dis Prim 2018 41*. 4 (2018), pp. 1–20.

[68]   Pappas PG, Kauffman CA, Andes DR, et al. **Clinical Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious Diseases Society of America**. In: *Clin Infect Dis*. 62 (2016), pp. e1–50.

[69]   Zhai B, Ola M, Rolling T, et al. **High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis**. In: *Nat Med*. 26 (2020), pp. 59–64.

[70]   Peleg AY**,** Hogan DA**,** Mylonakis E. **Medically important bacterial–fungal interactions**. In: *Nat Rev Microbiol*. 8 (2010), pp. 340–9.

[71]   Frey-Klett P, Burlinson P, Deveau A, et al. **Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists.** In: *Microbiol Mol Biol Rev*. 75 (2011), pp. 583–609.

[72]   Turvey SE, Broide DH, Vancouver C, et al. **Innate immunity**. In: *J Allergy Clin Immunol*. 125 (2010), pp. S24–32.

[73]   Uribe-Querol E**,** Rosales C. **Phagocytosis: Our Current Understanding of a Universal Biological Process**. In: *Front Immunol*. 11 (2020), pp. 1066.

[74]   Germic N, Frangez Z, Yousefi S, et al. **Regulation of the innate immune system by autophagy: monocytes, macrophages, dendritic cells and antigen presentation**. In: *Cell Death Differ 2019 264*. 26 (2019), pp. 715–27.

[75]   Van De Veerdonk FL, Gresnigt MS, Romani L, et al. **Aspergillus fumigatus morphology and dynamic host interactions**. In: *Nat Rev Microbiol 2017 1511*. 15 (2017), pp. 661–74.

[76]   Kuster S, Stampf S, Gerber B, et al. **Incidence and outcome of invasive fungal diseases after allogeneic hematopoietic stem cell transplantation: A Swiss transplant cohort study**. In: *Transpl Infect Dis*. 20 (2018), pp. e12981.

[77]   Lestrade PP, Bentvelsen RG, Schauwvlieghe AFAD, et al. **Voriconazole Resistance and Mortality in Invasive Aspergillosis: A Multicenter Retrospective Cohort Study**. In: *Clin Infect Dis*. 68 (2019), pp. 1463–71.

[78]   Upton A, Kirby KA, Carpenter P, et al. **Invasive aspergillosis following hematopoietic cell transplantation: Outcomes and prognostic factors associated with mortality**. In: *Clin Infect Dis*. 44 (2007), pp. 531–40.

[79]   Huttenhower C, Gevers D, Knight R, et al. **Structure, function and diversity of the healthy human

microbiome. In: *Nature*. 486 (2012), pp. 207–14.

[80] Sherman RM**,** Salzberg SL. **Pan-genomics in the human genome era**. In: *Nat Rev Genet 2020 214*. 21 (2020), pp. 243–54.

[81] Carriço JA, Rossi M, Moran-Gilad J, et al. **A primer on microbial bioinformatics for nonbioinformaticians**. In: *Clin Microbiol Infect*. 24 (2018), pp. 342–9.

[82] Sanger F**,** Nicklen S**,** Coulson AR. **DNA sequencing with chain-terminating inhibitors**. In: *Proc Natl Acad Sci*. 74 (1977), pp. 5463–7.

[83] Dorado G, Gálvez S, Rosales TE, et al. **Analyzing Modern Biomolecules: The Revolution of Nucleic-Acid Sequencing – Review**. In: *Biomol 2021, Vol 11, Page 1111*. 11 (2021), pp. 1111.

[84] Wooley JC**,** Godzik A**,** Friedberg I. **A Primer on Metagenomics**. In: *PLoS Comput Biol*. 6 (2010).

[85] McGinn S**,** Gut IG. **DNA sequencing – spanning the generations**. In: *N Biotechnol*. 30 (2013), pp. 366–72.

[86] Bleidorn C. **Third generation sequencing: technology and its potential impact on evolutionary biodiversity research**. In: *Http://DxDoiOrg/101080/1477200020151099575*. 14 (2015), pp. 1–8.

[87] Tan G, Opitz L, Schlapbach R, et al. **Long fragments achieve lower base quality in Illumina paired-end sequencing**. In: *Sci Reports 2019 91*. 9 (2019), pp. 1–7.

[88] Wenger AM, Peluso P, Rowell WJ, et al. **Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome**. In: *Nat Biotechnol 2019 3710*. 37 (2019), pp. 1155–62.

[89] Abellan-Schneyder I, Matchado MS, Reitmeier S, et al. **Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing**. In: *MSphere*. 6 (2021).

[90] Boulund F, Pereira MB, Jonsson V, et al. Computational and statistical considerations in the analysis of metagenomic data. Metagenomics, *Elsevier*; (2018), p. 81–102.

[91] Chen Y, Ye W, Zhang Y, et al. **High speed BLASTN: an accelerated MegaBLAST search tool**. In: *Nucleic Acids Res*. 43 (2015), pp. 7762–8.

[92] Buchfink B**,** Reuter K**,** Drost H-G. **Sensitive protein alignments at tree-of-life scale using DIAMOND**. In: *Nat Methods 2021 184*. 18 (2021), pp. 366–8.

[93] Ashlee M Benjamin Marshall Nichols TWBGSGJEL. **Comparing reference-based RNA-Seq mapping methods for non-human primate data**. In: *BMC Genomics* (2014).

[94] Westermann AJ**,** Barquist L**,** Vogel J. **Resolving host–pathogen interactions by dual RNA-seq**. In: *PLOS Pathog*. 13 (2017), pp. e1006033.

[95] Westermann AJ**,** Gorski SA**,** Vogel J. **Dual RNA-seq of pathogen and host**. In: *Nat Rev Microbiol 2012 109*. 10 (2012), pp. 618–30.

[96] Asfaha Y, Schrenk C, Alves Avelar LA, et al. **Novel alkoxyamide-based histone deacetylase inhibitors reverse cisplatin resistance in chemoresistant cancer cells**. In: *Bioorganic Med Chem*. 28 (2020), pp. 115108.

[97] Thijs S, De Beeck MO, Beckers B, et al. **Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys**. In: *Front Microbiol*. 8 (2017), pp. 494.

[98] Milanese A, Mende DR, Paoli L, et al. **Microbial abundance, activity and population genomic profiling with mOTUs2**. In: *Nat Commun*. 10 (2019).

[99] Lundberg DS, Yourstone S, Mieczkowski P, et al. **Practical innovations for high-throughput amplicon sequencing**. In: *Nat Methods 2013 1010*. 10 (2013), pp. 999–1002.

[100] Johnson JS, Spakowicz DJ, Hong BY, et al. **Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis**. In: *Nat Commun*. 10 (2019), pp. 1–11.

[101] Usyk M, Zolnik CP, Patel H, et al. **Novel ITS1 Fungal Primers for Characterization of the Mycobiome**. In: *MSphere*. 2 (2017), pp. 1–11.

[102] Nilsson RH, Tedersoo L, Ryberg M, et al. **A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts**. In: *Microbes Environ*. 30 (2015), pp. 145.

[103] Yang RH, Su JH, Shang JJ, et al. **Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing**. In: *PLoS One*. 13 (2018), pp. e0206428.

[104] Callahan BJ, McMurdie PJ, Rosen MJ, et al. **DADA2: high-resolution sample inference from Illumina amplicon data**. In: *Nat Methods*. 13 (2016), pp. 581.

[105] Gweon HS, Oliver A, Taylor J, et al. **PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform**. In: *Methods Ecol Evol*. 6 (2015), pp. 973–80.

[106] Allali I, Arnold JW, Roach J, et al. **A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome**. In: *BMC Microbiol*. 17 (2017), pp. 1–16.

[107] Hugerth LW, Andersson AF. **Analysing microbial community composition through amplicon sequencing: From sampling to hypothesis testing**. In: *Front Microbiol*. 8 (2017), pp. 1561.

[108] Callahan BJ, McMurdie PJ, Holmes SP. **Exact sequence variants should replace operational taxonomic units in marker-gene data analysis**. In: *ISME J 2017 1112*. 11 (2017), pp. 2639–43.

[109] Abarenkov K, Nilsson RH, Larsson K-H, et al. **The UNITE database for molecular identification of fungi – recent updates and future perspectives**. In: *New Phytol*. 186 (2010), pp. 281–5.

[110] Anslan S, Nilsson RH, Wurzbacher C, et al. **Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding**. In: *MycoKeys*. 39 (2018), pp. 29.

[111] Mallick H, Franzosa EA, McIver LJ, et al. **Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences**. In: *Nat Commun 2019 101*. 10 (2019), pp. 1–11.

[112] Langille MGI, Zaneveld J, Caporaso JG, et al. **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences**. In: *Nat Biotechnol 2013 319*. 31 (2013), pp. 814–21.

[113] Edgar RC. **Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences**. In: *PeerJ*. 2018 (2018), pp. e4652.

[114] Pereira-Marques J, Hout A, Ferreira RM, et al. **Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis**. In: *Front Microbiol*. 10 (2019), pp. 1277.

[115] Schmieder R, Edwards R. **Fast identification and removal of sequence contamination from genomic and metagenomic datasets**. In: *PLoS One*. 6 (2011).

[116] Tanjung ZA, Aditama R, Sudania WM, et al. **Metatranscriptomics workflow analysis from environmental sample of fungi**. In: *AIP Conf Proc*. 1744 (2016).

[117] Gosalbes MJ, Durbán A, Pignatelli M, et al. **Metatranscriptomic approach to analyze the functional human gut microbiota**. In: *PLoS One*. 6 (2011), pp. 1–9.

[118] Gallego Romero I, Pai AA, Tung J, et al. **RNA-seq: Impact of RNA degradation on transcript quantification**. In: *BMC Biol*. 12 (2014), pp. 1–13.

[119] Gosalbes MJ, Durbán A, Pignatelli M, et al. **Metatranscriptomic approach to analyze the functional human gut microbiota**. In: *PLoS One*. 6 (2011), pp. e17447.

[120] Franzosa EA, McIver LJ, Rahnavard G, et al. **Species-level functional profiling of metagenomes and metatranscriptomes**. In: *Nat Methods*. 15 (2018), pp. 962–8.

[121] Sun Z, Huang S, Zhang M, et al. **Challenges in benchmarking metagenomic profilers**. In: *Nat Methods* (2021), pp. 1–9.

[122] Segata N, Waldron L, Ballarini A, et al. **Metagenomic microbial community profiling using unique clade-specific marker genes**. In: *Nat Methods*. 9 (2012), pp. 811–4.

[123] Beghini F, McIver LJ, Blanco-Míguez A, et al. **Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3** n.d.

[124] Pop M. **Genome assembly reborn: recent computational challenges**. In: *Brief Bioinform*. 10 (2009), pp. 354–66.

[125] Junhua Li Huijue Jia et. al. **An integrated catalog of reference genes in the human gut microbiome**. In: *Nat Biotechnol* (2014).

[126] Lind AL, Pollard KS. **Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing**. In: *Microbiome*. 9 (2021), pp. 58.

[127] O'Leary NA, Wright MW, Brister JR, et al. **Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation**. In: *Nucleic Acids Res*. 44 (2016), pp. D733–45.

[128] Breitwieser FP, Lu J, Salzberg SL. **A review of methods and databases for metagenomic classification and assembly**. In: *Brief Bioinform*. 20 (2018), pp. 1125–39.

[129] Emiola A, Oh J. **High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage**. In: *Nat Commun*. 9 (2018).

[130] Wang JD, Levin PA. **Metabolism, cell growth and the bacterial cell cycle**. In: *Nat Rev Microbiol 2009 711*. 7 (2009), pp. 822–7.

[131] Nielsen HB, Almeida M, Juncker AS, et al. **Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes**. In: *Nat Biotechnol*. 32 (2014), pp. 822.

[132] Vieira-Silva S, Falony G, Belda E, et al. **Statin therapy is associated with lower prevalence of gut microbiota dysbiosis**. In: *Nat 2020 5817808*. 581 (2020), pp. 310–5.

[133] Ku CS, Loy EY, Salim A, et al. **The discovery of human genetic variations and their use as disease markers: past, present and future**. In: *J Hum Genet 2010 557*. 55 (2010), pp. 403–15.

[134] Power RA, Parkhill J, De Oliveira T. **Microbial genome-wide association studies: lessons from human GWAS**. In: *Nat Rev Genet 2016 181*. 18 (2016), pp. 41–50.

[135] Sheynkman GM, Shortreed MR, Frey BL, et al. **Large-Scale Mass Spectrometric Detection of Variant Peptides Resulting from Nonsynonymous Nucleotide Differences**. In: *J Proteome Res*. 13 (2014).

[136] Truong DT, Tett A, Pasolli E, et al. **Microbial strain-level population structure and genetic diversity from metagenomes.** In: *Genome Res*. 27 (2017), pp. 626–38.

[137] Bharti R, Grimm DG. **Current challenges and best-practice protocols for microbiome analysis**. In: *Brief Bioinform*. 22 (2021), pp. 178–93.

[138] Tarca AL, Bhatti G, Romero R. **A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity**. In: *PLoS One*. 8 (2013), pp. e79217.

[139] Bairoch A. **The ENZYME database in 2000**. In: *Nucleic Acids Res*. 28 (2000), pp. 304–5.

[140] Bairoch A. **The ENZYME data bank**. In: *Nucleic Acids Res*. 22 (1994), pp. 3626–7.

[141] Duvaud S, Gabella C, Lisacek F, et al. **Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users**. In: *Nucleic Acids Res*. 49 (2021), pp. W216–27.

[142] Ashburner M, Ball CA, Blake JA, et al. **Gene Ontology: tool for the unification of biology**. In: *Nat Genet*. 25 (2000), pp. 25.

[143] Lovering RC, Roncaglia P, Howe DG, et al. **Improving Interpretation of Cardiac phenotypes and enhancing discovery with expanded knowledge in the gene ontology**. In: *Circ Cardiovasc Genet*. 11 (2018), pp. e001813.

[144] Altman T, Travers M, Kothari A, et al. **A systematic comparison of the MetaCyc and KEGG pathway databases**. In: *BMC Bioinformatics*. 14 (2013), pp. 1–15.

[145] Caspi R, Billington R, Keseler IM, et al. **The MetaCyc database of metabolic pathways and enzymes - a 2019 update**. In: *Nucleic Acids Res*. 48 (2020), pp. D455–D453.

[146] Kramarz B, Lovering RC. **Gene Ontology: A Resource for Analysis and Interpretation of Alzheimer's Disease Data**. In: *Alzheimer's Dis* (2019), pp. 23–36.

[147] **KEGG - Current Statistics**. In: *2022-02-23* n.d. https://www.kegg.jp/kegg/docs/statistics.html (accessed February 22, 2022).

[148] **Summary of MetaCyc, version 25.5** n.d. https://metacyc.org/organism-summary (accessed February 22, 2022).

[149] Mistry J, Chuguransky S, Williams L, et al. **Pfam: The protein families database in 2021**. In: *Nucleic Acids Res*. 49 (2021), pp. D412–9.

[150] Matchado MS, Lauber M, Reitmeier S, et al. **Network analysis methods for studying microbial communities: A mini review**. In: *Comput Struct Biotechnol J*. 19 (2021), pp. 2687–98.

[151] Weiss S, Xu ZZ, Peddada S, et al. **Normalization and microbial differential abundance strategies depend upon data characteristics**. In: *Microbiome*. 5 (2017).

[152] Lloréns-Rico V, Vieira-Silva S, Gonçalves PJ, et al. **Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases**. In: *Nat Commun 2021 121*. 12 (2021), pp. 1–12.

[153] Fernandes AD, Macklaim JM, Linn TG, et al. **ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq**. In: *PLoS One*. 8 (2013), pp. e67019.

[154] Hirano H, Takemoto K. **Difficulty in inferring microbial community structure based on co-occurrence network approaches**. In: *BMC Bioinforma 2019 201*. 20 (2019), pp. 1–14.

[155] Weiss S, Van Treuren W, Lozupone C, et al. **Correlation detection strategies in microbial data sets vary widely in sensitivity and precision**. In: *ISME J*. 10 (2016), pp. 1669.

[156] Lin H, Peddada S Das. **Analysis of microbial compositions: a review of normalization and differential abundance analysis**. In: *Npj Biofilms Microbiomes 2020 61*. 6 (2020), pp. 1–13.

[157] Kaul A, Mandal S, Davidov O, et al. **Analysis of microbiome data in the presence of excess zeros**. In: *Front Microbiol*. 8 (2017), pp. 2114.

[158] NICHOLAS J. SCHURCH PIETÀ SCHOFIELD et. al. **How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?**. In: *Rnajournal* (2016).

[159] Sullivan GM, Feinn R. **Using effect size—or why the P value is not enough**. In: *J Grad Med Educ*. 4 (2012), pp. 279–82.

[160] Hettmansperger TP, McKean JW. **Robust nonparametric statistical methods**. *CRC Press*; (2010).

[161] Kloke JD, McKean JW. **Rfit: Rank-based estimation for linear models**. In: *R J*. 4 (2012), pp. 57–64.

[162] Love MI, Huber W, Anders S. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. In: *Genome Biol*. 15 (2014), pp. 1–21.

[163] Lin H, Peddada S Das. **Analysis of compositions of microbiomes with bias correction**. In: *Nat*

Commun. 11 (2020), pp. 1–11.

[164] Paulson JN, Stine OC, Bravo HC, et al. **Differential abundance analysis for microbial marker-gene surveys**. In: *Nat Methods*. 10 (2013), pp. 1200.

[165] Mallick H, Rahnavard A, McIver LJ, et al. **Multivariable association discovery in population-scale meta-omics studies**. In: *PLOS Comput Biol*. 17 (2021), pp. e1009442.

[166] Szklarczyk D, Gable AL, Lyon D, et al. **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets**. In: *Nucleic Acids Res*. 47 (2019), pp. D607–13.

[167] Kou Y, Xu X, Zhu Z, et al. **Microbe-set enrichment analysis facilitates functional interpretation of microbiome profiling data**. In: *Sci Reports 2020 101*. 10 (2020), pp. 1–12.

[168] Layeghifard M, Hwang DM, Guttman DS. **Disentangling Interactions in the Microbiome: A Network Perspective**. In: *Trends Microbiol*. 25 (2017), pp. 217–28.

[169] Liu N-N, Jiao N, Tan J-C, et al. **Multi-kingdom microbiota analyses identify bacterial–fungal interactions and biomarkers of colorectal cancer across cohorts**. In: *Nat Microbiol 2022* (2022), pp. 1–13.

[170] Schwager E, Mallick H, Ventz S, et al. **A Bayesian method for detecting pairwise associations in compositional data**. In: *PLoS Comput Biol*. 13 (2017), pp. e1005852.

[171] Kurtz ZD, Müller CL, Miraldi ER, et al. **Sparse and Compositionally Robust Inference of Microbial Ecological Networks** (2015).

[172] Friedman J, Alm EJ. **Inferring Correlation Networks from Genomic Survey Data**. In: *PLoS Comput Biol*. 8 (2012), pp. 1–11.

[173] Li C, Lim KMK, Chng KR, et al. **Predicting microbial interactions through computational approaches**. In: *Methods*. 102 (2016), pp. 12–9.

[174] Rao C, Coyte KZ, Bainter W, et al. **Multi-kingdom ecological drivers of microbiota assembly in preterm infants**. In: *Nature*. 591 (2021), pp. 633–8.

[175] Tipton L, Müller CL, Kurtz ZD, et al. **Fungi stabilize connectivity in the lung and skin microbial ecosystems**. In: *Microbiome 2018 61*. 6 (2018), pp. 1–14.

[176] Ha MJ, Kim J, Galloway-Peña J, et al. **Compositional zero-inflated network estimation for microbiome data**. In: *BMC Bioinformatics*. 21 (2020), pp. 1–20.

[177] Martín-Fernández J-A, Hron K, Templ M, et al. **Bayesian-multiplicative treatment of count zeros in compositional data sets**. In: *Stat Model An Int J*. 15 (2015), pp. 134–58.

[178] Knights D, Kuczynski J, Charlson ES, et al. **Bayesian community-wide culture-independent microbial source tracking**. In: *Nat Methods*. 8 (2011), pp. 761–5.

[179] Shenhav L, Thompson M, Joseph TA, et al. **FEAST: fast expectation-maximization for microbial source tracking**. In: *Nat Methods*. 16 (2019), pp. 627–32.

[180] McDonald D, Ackermann G, Khailova L, et al. **Extreme Dysbiosis of the Microbiome in Critical Illness**. In: *MSphere*. 1 (2016).

[181] Tsiamis G, Badger J, Yildirim S, et al. **Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment**. In: *Georg Papoutsoglou*. 17 n.d.

[182] Wirbel J, Zych K, Essex M, et al. **Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox**. In: *Genome Biol 2021 221*. 22 (2021), pp. 1–27.

[183] Arango-Argoty G, Garner E, Pruden A, et al. **DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data**. In: *Microbiome*. 6 (2018), pp. 1–15.

[184] Morton JT, Aksenov AA, Nothias LF, et al. **Learning representations of microbe–metabolite interactions**. In: *Nat Methods*. 16 (2019), pp. 1306–14.

[185] Lewis RE, Verweij PE. **Animal Models for Studying Triazole Resistance in Aspergillus fumigatus**. In: *J Infect Dis*. 216 (2017), pp. S466–73.

[186] Shrestha J, Razavi Bazaz S, Aboulkheyr Es H, et al. **Lung-on-a-chip: the future of respiratory disease models and pharmacological studies**. In: *Https://DoiOrg/101080/0738855120191710458*. 40 (2020), pp. 213–30.

[187] Patro R, Duggal G, Love MI, et al. **Salmon provides fast and bias-aware quantification of transcript expression**. In: *Nat Methods*. 14 (2017), pp. 417.

[188] Rizzetto L, Giovannini G, Bromley M, et al. **Strain Dependent Variation of Immune Responses to A. fumigatus: Definition of Pathogenic Species**. In: *PLoS One*. 8 (2013), pp. e56651.

[189] Magill SS, Dumyati G, Ray SM, et al. **Evaluating Epidemiology and Improving Surveillance of Infections Associated with Health Care, United States**. In: *Emerg Infect Dis*. 21 (2015), pp. 1537.

[190] Azevedo MM, Teixeira-Santos R, Silva AP, et al. **The effect of antibacterial and non-antibacterial compounds alone or associated with antifugals upon fungi**. In: *Front Microbiol*. 6 (2015), pp. 669.

[191] Fan D, Coughlin LA, Neubauer MM, et al. **Activation of HIF-1α and LL-37 by commensal bacteria inhibits Candida albicans colonization**. In: *Nat Med 2015 217*. 21 (2015), pp. 808–14.

[192] Rea MC, Sit CS, Clayton E, et al. **Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile**. In: *Proc Natl Acad Sci U S A*. 107 (2010), pp. 9352–7.

[193] Abt MC, Pamer EG. **Commensal bacteria mediated defenses against pathogens**. In: *Curr Opin Immunol*. 29 (2014), pp. 16–22.

[194] Kamada N, Núñez G. **Regulation of the Immune System by the Resident Intestinal Bacteria**. In: *Gastroenterology*. 146 (2014), pp. 1477–88.

[195] Parada Venegas D, De la Fuente MK, Landskron G, et al. **Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases**. In: *Front Immunol*. 0 (2019), pp. 277.

[196] Guzior D V., Quinn RA. **Review: microbial transformations of human bile acids**. In: *Microbiome*. 9 (2021), pp. 1–13.

[197] Raymond F, Ouameur AA, Déraspe M, et al. **The initial state of the human gut microbiome determines its reshaping by antibiotics**. In: *ISME J*. 10 (2016), pp. 707–20.

[198] Ma Y, Guo Z, Xia B, et al. **Identification of antimicrobial peptides from the human gut microbiome using deep learning**. In: *Nat Biotechnol 2022* (2022), pp. 1–11.

[199] Plackett B. **Why big pharma has abandoned antibiotics**. In: *Nature*. 586 (2020), pp. S50–2.

[200] Brives C, Pourraz J. **Phage therapy as a potential solution in the fight against AMR: obstacles and possible futures**. In: *Palgrave Commun 2020 61*. 6 (2020), pp. 1–11.

[201] Nayfach S, Páez-Espino D, Call L, et al. **Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome**. In: *Nat Microbiol* (2021), pp. 1–11.

[202] Palleja A, Mikkelsen KH, Forslund SK, et al. **Recovery of gut microbiota of healthy adults following antibiotic exposure**. In: *Nat Microbiol* (2018), pp. 1.

[203] Mehta RS, Abu-ali GS, Drew DA, et al. **Stability of the human faecal microbiome in a cohort of adult men**. In: *Nat Microbiol*. 3 (2018), pp. 347–55.

[204] Schirmer M, Franzosa EA, Lloyd-Price J, et al. **Dynamics of metatranscription in the inflammatory bowel disease gut microbiome**. In: *Nat Microbiol*. 3 (2018), pp. 337–46.

[205] Wesolowska-Andersen A, Bahl MI, Carvalho V, et al. **Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis**. In: *Microbiome*. 2 (2014), pp. 1–11.

[206] Angebault C, Payen M, Woerther PL, et al. **Combined bacterial and fungal targeted amplicon sequencing of respiratory samples: Does the DNA extraction method matter?**. In: *PLoS One*. 15 (2020), pp. e0232215.

[207] Soverini M, Turroni S, Biagi E, et al. **HumanMycobiomeScan: A new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples**. In: *BMC Genomics*. 20 (2019), pp. 496.

[208] Donovan PD, Gonzalez G, Higgins DG, et al. **Identification of fungi in shotgun metagenomics datasets**. In: *PLoS One*. 13 (2018), pp. e0192898.

[209] R. Marcelino V, Holmes EC, Sorrell TC. **The use of taxon-specific reference databases compromises metagenomic classification**. In: *BMC Genomics*. 21 (2020), pp. 1–5.

[210] Clos-Garcia M, Andrés-Marin N, Fernández-Eulate G, et al. **Gut microbiome and serum metabolome analyses identify molecular biomarkers and altered glutamate metabolism in fibromyalgia**. In: *EBioMedicine*. 46 (2019), pp. 499–511.

[211] Gutierrez D, Weinstock A, Antharam VC, et al. **Antibiotic-induced gut metabolome and microbiome alterations increase the susceptibility to Candida albicans colonization in the gastrointestinal tract**. In: *FEMS Microbiol Ecol*. 96 (2020).

[212] Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, et al. **Balances: a New Perspective for Microbiome Analysis**. In: *MSystems*. 3 (2018), pp. 53–71.

[213] Frau A, Kenny JG, Lenzi L, et al. **DNA extraction and amplicon production strategies deeply influence the outcome of gut mycobiome studies**. In: *Sci Reports 2019 91*. 9 (2019), pp. 1–17.

[214] Vandeputte D, Kathagen G, D'Hoe K, et al. **Quantitative microbiome profiling links gut community variation to microbial load**. In: *Nature*. 551 (2017), pp. 507–11.

[215] Vadlapatla R, Wong EY, Gayakwad SG. **Electronic drug delivery systems: An overview** (2017).

CURRICULUM VITAE

Personal Data

Full Name:	Bastian Seelbinder
Date of Birth:	19.01.1993
E-Mail:	bastiseelbinder@web.de
ORCID:	https://orcid.org/0000-0002-7458-5289

Education

2018 - today	PhD Student, Leibniz -Institute for Natural Product Research and Infection Biology - Hans-Knöll-Institute (Leibniz-HKI), Systems Biology and Bioinformatics (SBI), Jena, Germany
2015 – 2017	M.Sc. in Bioinformatics, Friedrich-Schiller University, Jena
2012 – 2015	B. Sc. in Bioinformatics, Friedrich-Schiller University, Jena

Publications

Seelbinder B. and Wolf T. et al. **GEO2RNAseq: An easy-to-use R pipeline for complete pre-processing of RNA-seq data**. In *bioRxiv* (2019). https://doi.org/10.1101/771063

Zoran T. and Weber M. et al. **Treatment with etanercept and low monocyte concentration contribute to the risk of invasive aspergillosis in patients post allogeneic stem cell transplantation.** In *Scientific Reports* (2019). https://doi.org/10.1038/s41598-019-53504-8

Seelbinder B., Wallstabe J. and Marischen L. et al. **Triple RNA-Seq Reveals Synergy in a Human Virus Fungus Co-infection Model**. In: *Cell Reports* 33 (2020), pp. 108389. https://www.doi.org/10.1016/J.CELREP.2020.108389

Seelbinder B. and Chen J., et al. **Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria**. In: *Microbiome*. 8 (2020), pp. 1–20. https://www.doi.org/10.1186/S40168-020-00899-6

Marfil-Sá nchez A. and Seelbinder B., et al. **Gut microbiome functionality might be associated with exercise tolerance and recurrence of resected early-stage lung cancer patients**. In: *PLoS One*. 16 (2021). https://www.doi.org/10.1371/JOURNAL.PONE.0259898

Marfil-Sánchez A. and Zhang L., et al. **An integrative understanding of the large metabolic shifts induced by antibiotics in critical illness**. In: *Gut Microbes*. 13 (2021). https://www.doi.org/10.1080/19490976.2021.1993598

Barber A.E. and Sae-Ong T., et al. **Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection**. In: *Nature Microbiology 2021 612*. 6 (2021), pp. 1526–36. https://www.doi.org/10.1038/s41564-021-00993-x

Zoran T. et al. **Molecular Profiling Reveals Characteristic and Decisive Signatures in Patients after Allogeneic Stem Cell Transplantation Suffering from Invasive Pulmonary Aspergillosis.** In *Journal of Fungi* (2022). https://doi.org/10.3390/jof8020171

Seelbinder B. and Lohinai Z., et al. **Candida expansion in the human gut is associated with an ecological signature that supports growth under dysbiotic conditions**. *To be submitted*. (2022)

EHRENWÖRTLICHE ERKLÄRUNG

Hiermit erkläre ich, dass mir die geltende Promotionsordnung der Fakultät für Biowissenschaften bekannt ist und ich mich mit bestem Wissen an diese Ordnung gehalten habe. Die vorliegende Dissertation habe ich selbstständig und nur unter Verwendung der angegebenen Hilfsmittel, Daten und quellen angefertigt. Unterstützung während meiner wissenschaftlichen Arbeit und zur Erstellung der vorliegenden Dissertation habe ich nur von den genannten Koautoren und in der Danksagung erwähnten Personen erhalten. Ich habe keine Hilfe von Promotionsberatern oder externen Beratungsdiensten in Anspruch genommen. Niemand hat von mir mittelbare oder unmittelbare geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Eine gleiche, in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule oder anderen Fakultät, habe ich nicht eingereicht. Die vorgelegte Dissertation wurde nicht bereits zuvor als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Im speziellen habe ich sie an keiner anderen Hochschule eingereicht, um einen akademischen Grad zu erhalten.

Jena, April 2022

_____
Bastian Seelbinder

APPENDIX

<u>**FORM 2**</u>

Manuscript No. 1

Short reference [Seelbinder, Wallstabe, and Marischen et al. (2020), Cell Rep.]

Contribution of the doctoral candidate

My main work concerned was the creation and implementation of the *in silico* protocol for triple RNA-Seq data, especially the downstream processing and analysis of RNA-Seq data. Further analyses included testing for differential gene abundance and co-abundance network analyses.

Figure(s) # 1	X	Approximate contribution of the doctoral candidate to the figure: 20%
Figure(s) # 2	X	Approximate contribution of the doctoral candidate to the figure: 70%
Figure(s) # 3-4	X	Approximate contribution of the doctoral candidate to the figure: 90%
Figure(s) # 5-6	X	Approximate contribution of the doctoral candidate to the figure: 30%

<u>**FORM 2**</u>

Manuscript No. 2

Short reference [Barber and Sae-Ong et al. (2021), Nat. Micro.]

Contribution of the doctoral candidate

My contribution concerned methodological considerations, especially for cluster and functional enrichment tests.

Figure(s) # all	X	Approximate contribution of the doctoral candidate to the figure: 5%

FORM 2

Manuscript No. 3

Short reference [Seelbinder and Chen et al. (2020), Cell Rep.]

Contribution of the doctoral candidate

My work concerned the analysis of bacteriome data (including MGS analyses), trans-kingdom network analyses, and manuscript writing.

Figure(s) # 1	X	Approximate contribution of the doctoral candidate to the figure: 60%
Figure(s) # 2-3	X	Approximate contribution of the doctoral candidate to the figure: 85%
Figure(s) # 4	X	Approximate contribution of the doctoral candidate to the figure: 50%

FORM 2

Manuscript No. 4

Short reference [Marfil-Sánchez and Zhang et al. (2021), Gut Microbes]

Contribution of the doctoral candidate

My main work concerned was pilot work on microbial source tracking with FEAST and methodological considerations in network analyses.

Figure(s) # all	X	Approximate contribution of the doctoral candidate to the figure: 5%

D

FORM 2

Manuscript No. 5

Short reference [Seelbinder and Lohinai et al. (2021), *in preparation*]

Contribution of the doctoral candidate

My main work concerned the establishment of study groups, microbiome analyses, manuscript writing and figure creation. In particular, I performed all bioinformatic analyses with the exception of genome-scale metabolic modelling analyses (GSMM, MAMBO).

Figure(s) # all   X   Approximate contribution of the doctoral candidate to the figure: 95%

FORM 2

Manuscript No. 6

Short reference [Marfil-Sánchez and Seelbinder et al. (2021), PLoS One]

Contribution of the doctoral candidate

My main work concerned machine learning and manuscript writing.

Figure(s) # all   X   Approximate contribution of the doctoral candidate to the figure: 40%