

Exploring mechanisms of  
*C. elegans* microbiota-mediated  
protection

**DISSERTATION**

in fulfillment of the requirements for the degree of  
*Doctor rerum naturalium*  
of the Faculty of Mathematics and Natural Sciences  
at Kiel University

Submitted by Kohar Annie B. Kissoyan  
Department of Evolutionary Ecology and Genetics  
Zoological Institute, Kiel University

Kiel, 2021



# DECLARATION

I, Kohar Annie B. Kissoyan, declare that:

Apart from my supervisor's guidance, the content and design of the thesis is all my own work;

Specific aspects of my thesis were supported by colleagues; their contribution is specified in detail in the following section "Authors' Contributions";

The thesis has not already been submitted neither partially nor wholly as part of a doctoral degree to another examining body. Apart from the included published papers no other part of the thesis has been published nor submitted for publishing;

The thesis has been prepared subject to the Rules of Good Scientific Practice of the German Research Foundation (DFG).

Signature: \_\_\_\_\_



First referee: Prof. Dr. Hinrich Schulenburg

Second referee: Prof. Dr. Matthias Leippe

Date of oral examination: June 4<sup>th</sup>, 2021

Signature: \_\_\_\_\_



# Table of Contents

AUTHORS' CONTRIBUTIONS .....	7
SUMMARY .....	11
ZUSAMMENFASSUNG .....	13
INTRODUCTION .....	15
CHAPTER I	
Natural <i>C. elegans</i> microbiota protects against infection via the production of a cyclic lipopeptide of the viscosin group .....	53
CHAPTER II	
Protective <i>C. elegans</i> microbiota produce biofilm but are unable to inactivate pathogen toxin.....	75
CHAPTER III	
The functional repertoire contained within the native microbiota of the model nematode <i>Caenorhabditis elegans</i> .....	93
CHAPTER IV	
Transcriptome and proteome analysis of the <i>C. elegans</i> response to protective natural microbiota isolates.....	123
GENERAL DISCUSSION .....	169
LIST OF EQUIPMENT .....	189
ANNEX.....	191
Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms.....	191
ACKNOWLEDGMENTS .....	213
CURRICULUM VITAE.....	215





# AUTHORS' CONTRIBUTIONS

This Ph.D. thesis consists of four chapters, each represented by a published or an unpublished manuscript. Kohar Annie B. Kissoyan developed original ideas, performed the majority of the experiments, analyzed and interpreted results, and wrote the manuscripts with major contributions for Chapters I, II, and IV, and on a collaborative basis for chapters III and the manuscript listed in the Annex.

---

## CHAPTER I

### **Natural *C. elegans* microbiota protects against infection via the production of a cyclic lipopeptide of the viscosin group**

Kohar Annie B. Kissoyan, Moritz Drechsler, Eva-Lena Stange, Johannes Zimmermann, Christoph Kaleta, Helge Bode, Katja Dierking

K.D. conceived the study and secured funding. K.A.B.K. and K.D. designed the experiments, analyzed the data, and wrote the manuscript. M.D. and H.B.B. did the UPLC-HRMS, HPLC, and genomic analyses. J.Z. and C.K. analyzed the genomic data. E.-L.S. and K.A.B.K. conducted the PA14 experiments. K.A.B.K. conducted all other experiments.

---

## CHAPTER II

### **Protective *C. elegans* microbiota produce biofilm but are unable to inactivate pathogen toxin**

Kohar Annie B. Kissoyan, Christoph Giez, Guillaume Tetreau, Katja Dierking

K.D. secured funding and supervised the work. K.D. and K.A.B.K. conceived the study. K.A.B.K. and C.G. designed the experiments, performed, and analyzed the data. G.T. purified the BT679 toxins. K.D. and K.A.B.K. discussed and interpreted the data and wrote the manuscript.

---

## CHAPTER III

### **The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis elegans***

Johannes Zimmermann, Nancy Obeng, Wentao Yang, Barbara Pees, Carola Petersen, Silvio Waschina, Kohar Annie B. Kissoyan, Jack Aidley, Marc P. Höppner, Boyke Bunk, Cathrin Spröer, Matthias Leippe, Katja Dierking, Christoph Kaleta, Hinrich Schulenburg

H.S., C.K., J.Z., and N.O. conceived the study and wrote the original draft. H.S. and C.K. jointly supervised the work. J.Z., W.Y., S.W., M.P.H. analyzed the genomic data. N.O., B.P, C.P., K.A.B.K., and J.A. designed, performed, and analyzed experiments. B.B., C.S., M.L. K.D. C.K. and H.S. gave intellectual input. All authors discussed the data, read, and approved the final manuscript.

---

## CHAPTER IV

### **Transcriptome and proteome analysis of the *C. elegans* response to protective natural microbiota isolates**

Kohar Annie B. Kissoyan, Christian Treitz, Wentao Yang, Eva-Lena Stange, Bente Rackow, Andreas Tholey, Katja Dierking

K.D. secured funding and supervised the work. K.D. and K.A.B.K. conceived the study. K.A.B.K. designed, performed and analyzed experiments. ES and BR performed experiments. WY analyzed the transcriptomics data. CT and AT analyzed the proteomics data. K.D and K.A.B.K. discussed and interpreted the data.

---

## ANNEX

### **Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms.**

Philipp Rausch, Malte Rühlemann, Britt M. Hermes, Shauni Doms, Tal Dagan, Katja Dierking, Hanna Domin, Sebastian Fraune, Jakob von Frieling, Ute Hentschel, Femke-Anouska Heinsen, Marc Höppner, Martin Thomas Jahn, Cornelia Jaspers,

Kohar Annie B. Kissoyan, Daniela Langfeldt, Ateequr Rehman, Thorsten B. H. Reusch, Thomas Roeder, Ruth A. Schmitz, Hinrich Schulenburg, Ryszard Soluch, Felix Sommer, Eva Stukenbrock, Nancy Weiland-Bräuer, Philip Rosenstiel, Andre Franke, Thomas Bosch, John F. Baines.

P.R.a, P.R.o, A.F., T.B., and J.F.B. conceived and designed research. P.R.a and M.R. performed data analyses. P.R.a, M.R., B.H., S.D., and J.F.B. interpreted results and wrote the manuscript. P.R.a, M.R., T.D., K.D., H.D., S.D., S.F., J.F., U.H.H., F.A.H., B.H., M.H., M.J., C.J., K.A.B.K., D.L., A.R., T.B.H.R., T.R., R.A.S., H.S., R.S., F.S., E.S., N.W.B., P.R.o, A.F., T.B., and J.F.B generated and interpreted host specific data and gave intellectual input. All authors read and approved the final manuscript.

---

**Hiermit bestätige ich als Betreuer die obenstehenden Angaben.**

---



## SUMMARY

Animals and microorganisms live together, forming a network of associations between animal hosts and their resident microorganisms, the microbiota. One important beneficial microbiota function is protecting the host against pathogen infection, termed microbiota-mediated protection. Many studies describe microbiota-mediated protection in animals; however, the underlying mechanisms remain mostly elusive. In my Ph.D. thesis project, I investigated mechanisms of microbiota-mediated protection in the nematode model *Caenorhabditis elegans*.

To understand microbiota-mediated protection mechanisms, I first identified *C. elegans* natural microbiota isolates that protect the host against infection with *Bacillus thuringiensis*. Then, I explored the underlying mechanisms of protection by these microbiota isolates both from the bacterial and the host side. On the bacterial side, I showed that: 1) Some microbiota isolates can protect the host by directly inhibiting pathogen growth via massetolide E production, a lipopeptide of the viscosin group, while others protect the host without producing massetolide E. 2) Microbiota isolates can produce biofilm *in vitro*, suggesting biofilm formation as a potential protection mechanism, which merits *in vivo* investigation. 3) Microbiota isolates cannot protect against *B. thuringiensis* toxins (Cry21Aa3 and Cry14Aa2) and are rendered pathogenic under certain pathogen exposure conditions, revealing the context-dependent dual nature of microbiota protective function. Moreover, in collaboration with colleagues, we characterized a more extensive selection of natural microbiota isolates of *C. elegans* and elucidated their metabolic contribution to the host using whole-genome sequencing, mathematical modeling, and experimentation. We showed that *C. elegans* microbiota could synthesize all the essential vitamins and amino acids needed by the nematode, reinforcing its beneficial role to the host.

On the host side, I used a multi-omics approach to identify host genes and pathways required for microbiota-mediated protection. To this end, I obtained transcriptomic and proteomic datasets of the infected and uninfected worms treated or untreated with the protective microbiota isolates. I showed that: 1) Protective microbiota isolates influence the worm's innate immune response and cellular structural component functions. 2) Microbiota isolates induce changes in expression and abundances on both gene and protein levels, respectively, for galectins, C-type lectins, and lysozymes, suggesting their potential roles in microbiota-mediated host protection. Altogether, these findings indicate microbiota-mediated activation of innate immune response genes, suggesting that the microbiota isolates 'prime' the *C. elegans* immune response, increasing host preparedness for subsequent pathogen attack.

Overall, the findings of this thesis provide valuable insights into mechanisms of microbiota-mediated protection on the bacterial and host sides, expanding our understanding of host-microbe interactions.



# ZUSAMMENFASSUNG

Tiere und Mikroorganismen leben in einem engen Verhältnis und bilden dabei ein Assoziationsnetzwerk zwischen tierischen Wirten und ihren ansässigen Mikroorganismen, der Mikrobiota. Eine wichtige und Vorteile verschaffende Funktion der Mikrobiota ist dabei der vermittelte Schutz gegen Pathogeninfektionen. Obwohl Studien den durch Mikrobiota vermittelten Schutz bei Tieren beschreiben, sind die zugrunde liegenden Mechanismen größtenteils unbekannt. In meiner Doktorarbeit habe ich daher den durch Mikrobiota vermittelten Schutz gegen Pathogeninfektionen im Tiermodell *Caenorhabditis elegans* untersucht.

Um die durch Mikrobiota vermittelten Schutzmechanismen zu verstehen, identifizierte ich zunächst natürliche Mikrobiota-Isolate aus *C. elegans*, die den Wirt vor einer bakteriellen Infektion mit *Bacillus thuringiensis* schützen. Dann untersuchte ich die zugrunde liegenden Schutzmechanismen dieser Mikrobiota-Isolate, sowohl von Seiten des Wirtes, als auch von Seiten der Mikrobiota. Bezüglich der Bakterien zeigte ich, dass: 1) einige Mikrobiota-Isolate den Wirt über die Produktion von Massetolid E, ein Lipopeptid der Viskosingruppe, direkt schützen können, während andere Mikrobiota-Isolate den Wirt schützen ohne Massetolid E zu produzieren. 2) Mikrobiota-Isolate können *in vitro* Biofilme bilden und somit wahrscheinlich den Wirt schützen. Dieses Ergebnis verdient jedoch eine genauere Untersuchung *in vivo*. 3) Mikrobiota-Isolate können die *Bacillus thuringiensis* Toxine (Cry21Aa3 und Cry14Aa2) nicht inaktivieren und verstärken sogar die Pathogenität unter bestimmten Versuchsbedingungen. Dieses Ergebnis zeigt, dass die Mikrobiota-vermittelte Schutzfunktion Kontext abhängig und von dualer Natur ist. Darüber hinaus charakterisierte ich eine größere Auswahl natürlicher Mikrobiota-Isolate des Nematoden und untersuchte ihren metabolischen Beitrag mittels Ganzgenomsequenzierung des Wirtes, mathematischer Modellierung und Experimenten. Dabei zeigte sich die unterstützende Rolle der Mikrobiota, da sie alle für *C. elegans* essentiellen Vitamins und Aminosäuren synthetisieren kann.

Um für einen Mikrobiota-vermittelten Schutz benötigte Wirtsgene und -pathways zu identifizieren, verwendete ich einen *Multi-Omics*-Ansatz. Ich erhielt Transkriptomik- und Proteomik-Datensätze von infizierten und nicht infizierten Würmer, die mit den schützenden Mikrobiota-Isolaten oder einer nicht schützenden Kontrolle behandelt wurden. Ich zeigte, dass: 1) schützende Mikrobiota-Isolate die angeborene Immunantwort des Wurms und die Funktionen der zellulären Strukturkomponenten beeinflussen. 2) Mikrobiota-Isolate induzieren Veränderungen in der Gen- und Proteinexpression und -häufigkeit von Glaktinen, C-Typ-Lektinen und Lysozymen, was wiederum für den Mikrobiota-vermittelten Wirtsschutz von Bedeutung sein könnte. Insgesamt deuteten die Ergebnisse meiner Doktorarbeit auf eine durch Mikrobiota vermittelte Aktivierung angeborener Immunantwortgene hin, die das Immunsystem der Würmer auf eine Pathogeninfektion vorbereiten und somit vor zukünftigen Infektionen schützen könnten.

Insgesamt erweitert diese Arbeit das Wissen über die Mechanismen des Mikrobiota-vermittelten Schutzes, sowohl auf Seiten des Wirtes, als auch auf Seiten der Bakterien, und stellt somit eine Erweiterung unseres Wissens in Bezug auf Wirts-Mikroben-Interaktionen dar.



# INTRODUCTION

The existence of a microbial world, invisible to the naked eye, has fueled a fascinating journey through history, inciting us, humans, to strive and make the world of microbes visible and comprehensible. Antony Van Leeuwenhoek first saw microbes in the 17<sup>th</sup> century through the first microscope lens ever invented (Haller, 2018; Yong, 2016). However, humans have primarily ignored microbes until Louis Pasteur and Robert Koch developed the germ theory at the end of the 19<sup>th</sup> century (Haller, 2018), which revolutionized our view on disease prevention and treatment (Koch & McFall-Ngai, 2018). Microbes were identified as the notorious causative agents of plagues and diseases, thus needed to be eliminated with intensive hygiene measures, sanitation, and antibiotics (Blaser, 2015). The advent of novel sequencing technologies allows the thorough study of the microbial inhabitants of the deep sea (Hinze et al., 2019) to those of the forest soil (Baldrian, 2017) and the human gut (The Human Microbiome Project Consortium et al., 2012). As a result, nowadays, there is an emerging conceptual revolution on our view of microbes, particularly on the resident microbes of a host, the microbiota (Yong, 2016). The terminology, microbiota and microbiome, initially used interchangeably (Lederberg & McCray, 2001), is more recently defined as the host-associated microorganisms and their genetic content, respectively (Berg et al., 2020; Marchesi & Ravel, 2015). The latter definition of microbiota as host-associated microorganisms is adopted in this thesis.

The microbiota profoundly affects human health and can potentially be modulated to prevent or treat diseases (Haller, 2018). However, the promise of microbiota in solving current therapeutic problems is still unfulfilled since microbiota itself is mostly a black box. A black box in which the microbiota functions, and the underlying mechanisms of microbiota-mediated effects on the host, including the protective effects, remain elusive. The inherent complexity of microbiota and that of the human-microbiota-pathogen interactions make understanding microbiota-mediated protection more complicated (Ezcurra, 2018; Koch & McFall-Ngai, 2018; Zheng et al., 2020). To untangle these layers of complexity, I used a host-microbiota-pathogen tripartite model system. Thus, during my Ph.D. thesis, I explored microbiota-mediated protection against pathogen infection in the animal model *Caenorhabditis elegans* (*C. elegans*). I then investigated the underlying mechanisms of microbiota-mediated protection, both from the bacterial and the host sides.

In this introduction, I present the concept of microbiota-mediated protection, highlighting its significance. Then, I briefly introduce *C. elegans*, its pathogens, its immune defenses, and its recently identified natural microbiota. I also describe the *C. elegans*-microbiota system as a model to study microbiota-mediated protection. Finally, I conclude with an overview of the objectives and contents of this thesis chapters.

## **Microbiota-mediated protection**

Throughout evolution, animals and microorganisms have formed an intricate web of associations (Sampson & Mazmanian, 2015). In these associations, the microbiota affects the animal host's fitness and physiology (Douglas, 2019; Mcfall-Ngai et al., 2013). One of these associations' beneficial effects is microbiota-mediated protection (Masson & Lemaitre, 2017). Microbiota-mediated protection is achieved when microbiota mitigate or eliminate the pathogenic impact on the host (Chiu et al., 2017). Microbiota-mediated protection was demonstrated as early as 1907 when Elie Metchnikoff proposed that lactic acid-producing bacterial strains are beneficial to the host since they prevent the growth of harmful species within the colon (Sassone-Corsi & Raffatellu, 2015). Since Metchnikoff's self-experimentation with the ingestion of such bacteria, microbiota-mediated protection has been described in a myriad of organisms from humans (Buffie & Pamer, 2013) to plants (Vannier et al., 2019), where valuable insights are obtained particularly from animal models (Sassone-Corsi & Raffatellu, 2015).

Substantial evidence on microbiota-mediated protection comes from studies using germ-free animals (i.e., which completely lack microbiota) and antibiotic-treated animals (i.e., which have a disturbed microbiota). Germ-free mice are shown to be highly susceptible to pathogenic infections (e.g., diarrheal infections caused by *Salmonella enterica* serovars Typhimurium and Enteritidis), compared to mice harboring microbiota (Kennedy et al., 2018; Sassone-Corsi & Raffatellu, 2015; Vogt & Finlay, 2017). For example, a low dose of *Salmonella* (i.e., as low as ten colony forming units) is lethal for germ-free mice, whereas much higher doses (i.e.,  $10^3$ - $10^9$  colony forming units) are needed to kill 50% of mice with intact microbiota (Colins and Carter, 1978; Nardi et al., 1989; Vogt and Finlay, 2017). Also, animals with a dysbiotic microbiota (i.e., disturbed microbiota) are more susceptible to infection than their counterparts with an intact microbiota. For example,

antibiotic-induced alteration of gut microbiota composition in mice increases susceptibility to infection with pathogens, such as *S. Typhimurium* (Crowell et al., 2009) and *Clostridium difficile* (Lawley et al., 2009; Schubert et al., 2015). Moreover, microbiota is shown to be transferrable from healthy animals to germ-free or infected animals, thereby transferring microbiota-mediated protection against pathogen infection. Interestingly, fecal microbiota transplantation, from mice resistant to infection with the enteric bacterial pathogen *Citrobacter rodenticum* to susceptible mice, delayed pathogen colonization and mortality, protecting susceptible mice from infection (Willing et al., 2011). Thus, studies using animal experiments together with new technological advances have elucidated microbiota's complexity and its effects on host health (Sassone-Corsi & Raffatellu, 2015).

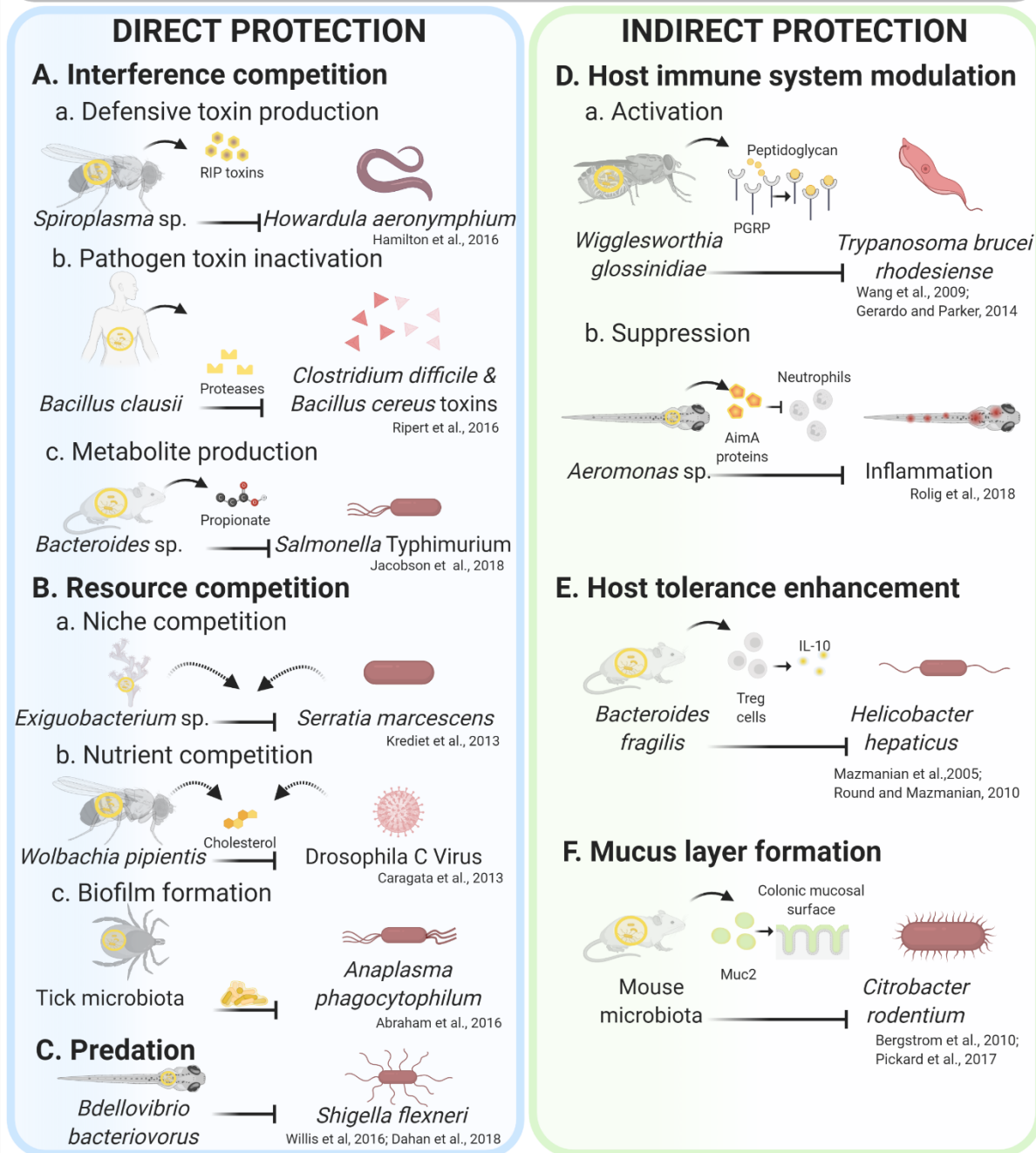
### *Classification of microbiota-mediated protection*

Microbiota-mediated protection, initially known as colonization resistance (van der Waaij et al., 1971), can be broadly classified as either direct or indirect protection (Buffie and Pamer, 2013; Pickard et al., 2017). Direct protection is defined as host-independent protection, where microbiota directly interacts with pathogens (Buffie and Pamer, 2013; Pickard et al., 2017). In contrast, indirect microbiota-mediated protection is defined as host-dependent protection, where microbiota activates host immune defenses, which in turn interact with the pathogen (Buffie and Pamer, 2013; Pickard et al., 2017). Based on various sources from the current literature (Douglas, 2019; Ford and King, 2016; Masson and Lemaitre, 2017), I will further classify direct and indirect microbiota-mediated protection into different mechanisms of protection. These different protection mechanisms are described further below, supplemented with examples of each that most relate to the contents of this thesis.

#### *Direct microbiota-mediated protection*

Direct microbiota-mediated protection may act through the following mechanisms to ensure host protection: A) interference competition, B) resource competition or C) predation (**Figure 1, blue box**) (Douglas, 2019; Ford and King, 2016; Masson and Lemaitre, 2017). These protection mechanisms are explained below.

## MICROBIOTA-MEDIATED PROTECTION MECHANISMS



**Figure 1: Microbiota-mediated protection mechanisms, classified into direct or indirect protection mechanisms.** The blue box presents mechanisms of direct microbiota-mediated protection. Direct protection can operate through A) interference competition, B) resource competition, or C) predation. The green box presents mechanisms of indirect microbiota-mediated protection. Indirect protection can operate through (D) host immune modulation, (E) tolerance enhancement, or (F) mucus development. Each mentioned mechanism is represented by an example discussed further in the text. The yellow circles represent the host microbiota, → indicate induction, and ⊥ indicate suppression. The figure was adapted from the mentioned references and (Douglas, 2019; Ford and King, 2016; Masson and Lemaitre, 2017), and was created with Biorender.com.

A) Interference competition: The protective microbe produces inhibitory compounds, such as toxins or metabolites, that inhibit or kill the pathogen (**Figure 1A**). Defensive toxin production occurs, for example, in *Drosophila neotestacea*, where the bacterial microbiota member, *Spiroplasma* sp., protects the host against the pathogenic nematode *Howardula aoronymphium*. *Spiroplasma* sp. produces the toxin ribosome-inactivating protein (RIP), which depurinates (removes purine) the pathogen's rRNA more than it depurinates the host's rRNA. Depurination of pathogen rRNA reduces infection intensity in the host; thus, the host tolerates the infection. However, the mechanism of RIP-induced specific depurination of pathogen rRNA is still unknown (**Figure 1Aa**) (Hamilton *et al.*, 2016; Masson and Lemaitre, 2017). Another example where interference competition is achieved is via the inactivation of pathogen-derived toxins in the human probiotic bacterium *Bacillus clausii*, which produces proteases that inhibit the cytotoxic effects of *Clostridium difficile* and *Bacillus cereus* toxins *in vitro* (Ripert *et al.*, 2016) (**Figure 1Ab**). Interference competition can also occur via metabolite production. An example comes from the mouse microbiota where *Bacteroides* sp. protects the host from the bacterial pathogen *S. Typhimurium* infection. *Bacteroides* sp. produces short-chain fatty acid propionate, which directly limits pathogen growth by disrupting intracellular pH homeostasis. *Bacteroides* sp. chemically increases intestinal propionate levels in mice and protects them from *S. Typhimurium* infection (Jacobson *et al.*, 2018) (**Figure 1Ac**).

B) Resource competition: The protective microbe and the pathogen compete for host resources through niche or nutrient competition or biofilm formation (**Figure 1B**). For example, niche competition occurs in corals, where colonization by microbiota member, *Exiguobacterium* sp. inhibits colonization by the pathogen *Serratia marcescens* (**Figure 1Ba**) (Krediet *et al.*, 2013a; Krediet *et al.*, 2013b; Masson and Lemaitre, 2017). *Exiguobacterium* sp. inhibits the induction of a range of catabolic enzymes (glucosidases), which are needed for the pathogen to colonize the coral mucus (Krediet *et al.*, 2013a). Nutrient competition occurs, for example, in *Drosophila melanogaster*, where microbiota member *Wolbachia pipientis* competes with the pathogen, *Drosophila C virus* (DCV), for cholesterol (**Figure 1Bb**) (Caragata *et al.*, 2013). Biofilm-mediated protection occurs in the tick *Ixodes scapularis*. Here, microbiota-produced biofilm and the strengthened peritrophic matrix (i.e., the glycoprotein-rich layer that separates the epithelial cells from the tick gut lumen) prevent colonization by the obligate intracellular bacterial pathogen *Anaplasma phagocytophilum*, thus protecting the host (**Figure 1Bc**) (Abraham *et al.*, 2017).

C) Predation: The protective microbe predated on the pathogen (**Figure 1C**). Predation occurs in Zebrafish (*Danio rerio*) larvae, where microbiota member *Bdellovibrio bacterivorus* predated on the pathogen *Shigella flexneri*, thus protects the host from infection (Willis et al., 2016; Dahan et al., 2018).

*Indirect microbiota-mediated protection*

Microbiota-mediated protection can also be achieved via indirect protection, which is host-dependent. Indirect microbiota-mediated protection may act through one of the following mechanisms: A) Host immune system modulation, B) Host tolerance enhancement or C) Mucus layer formation (**Figure 1, green box**) (Douglas, 2019; Ford and King, 2016; Masson and Lemaitre, 2017). These protection mechanisms are explained below.

A) Host immune system modulation: The protective microbe modulates the host's immune response in a manner that protects the host from pathogen infection (**Figure 1D**). Immune modulation can operate via immune response activation or suppression. Immune response activation occurs, for example, in tsetse flies (*Glossina morsitans*), where microbiota member *Wigglesworthia glossinidia* up-regulates peptidoglycan recognition protein (PGRP-LB). PGRP-LB scavenges *Wigglesworthia*-released peptidoglycan, a bacterial cell wall component, which activates host immune responses targeting the parasite African Trypanosome *Trypanosoma brucei rhodesiense*. Thus, activation of the immune responses protects the host from infection (**Figure 1Da**) (Wang et al., 2009; Gerardo and Parker, 2014). Whereas, immune response suppression occurs, for example, in Zebrafish larvae, where the microbiota member *Aeromonas* sp. secretes immunomodulatory protein AimA. AimA prevents excessive accumulation of neutrophils, innate immune cells that mediate a pro-inflammatory response. Thus, suppression of immune response protects the host from pathological inflammation, septic shock, and death (**Figure 1Db**) (Rolig et al., 2018).

B) Host tolerance enhancement: The protective microbe increases the fitness of its host during infection without reducing the fitness of the pathogen by enhancing host tolerance (e.g., via tissue damage prevention or repair) (**Figure 1E**). For example, microbiota member *Bacteroides fragilis* protects mice from experimentally induced colitis by *Helicobacter hepaticus* via enhancing host tolerance. *B. fragilis*-produced polysaccharide A induces and expands T regulatory cells through interaction with TLR2. T regulatory cells produce IL-10, anti-inflammatory cytokines, which limit

inflammation and promote tolerance of the infected mice to the pathogen (Mazmanian et al., 2005; Round and Mazmanian, 2010; Belkaid and Harrison, 2017).

C) Mucus layer formation: The protective microbe induces mucus layer formation, supporting the host's epithelial defenses (**Figure 1F**). Glycoprotein Muc2, the main component of mucus, controls pathogen load and intestinal damage in mice infected with *Citrobacter rodentium*. Mice microbiota activates Muc2, which controls pathogen load via limiting the bacteria associated with the colonic mucosal surface. Limited bacteria associated with colonic mucosal surface limits tissue damage and translocation of pathogens across the epithelium, thus protecting the host (Bergstrom et al., 2010; Pickard et al., 2017).

### *Significance of microbiota-mediated protection*

Microbiota-mediated protection is indispensable for many hosts as survival strategies, as shown in animal studies, which are followed by clinical studies and further highlight its significance on human health and disease. The significance of microbiota-mediated protection is particularly highlighted with the continuous search for novel treatment options for many inflammatory and chronic diseases, particularly with the continuous surge of antibiotic resistance and the emergence of new infectious and lifestyle diseases (Ventola, 2015). In humans, microbiota dysbiosis is linked to diseases and immune dysfunctions such as chronic inflammatory diseases; an example is the inflammatory bowel disease, Crohn's disease (Brown and Clarke, 2017). However, while such diseases are linked to microbiota dysbiosis, it often remains difficult to disentangle cause from effect (Øyri, Múzes, and Sipos, 2015). In other cases, treatments for infectious diseases can cause microbiota dysbiosis, determining the disease's outcome. For example, antibiotics can cause microbiota dysbiosis, which decreases resistance against toxin-producing *C. difficile*, and promotes *C. difficile* infection (Baohong Wang et al., 2017). Thus, one of the new treatment options includes using microbiota.

Microbiota therapeutics is commonly used to restore a dysbiotic microbiota. Such restoration can be done, for instance, using: a) probiotics, which is the ingestion of protective bacteria, b) prebiotics, which is the ingestion of nutrients that favor the growth of protective bacteria, c) fecal transplantation, which is the transplantation of fecal microbiota, i.e., the transfer of stool from a healthy donor to the gastrointestinal tract of an unhealthy recipient, commonly one suffering from

recurrent *C. difficile* infections (Brown, 2014; Fuentes et al., 2014). While these methods are currently used with varying success levels, there are some controversies about them and an overall ambiguity on the treatment efficacies. Such ambiguities arise since much is still unknown about these microbiota therapeutics, including a) the exact mechanisms of action, b) the therapeutic microbiota community composition (i.e., strain-level resolution) and function, c) therapeutic microbial numbers required, and d) factors affecting patient predisposition or responsiveness (e.g., genetic, diet, environmental stimuli, and age) to the treatment (Bojanova & Bordenstein, 2016; Goloshchapov et al., 2019; Khoruts, 2018; Wang et al., 2019). Besides, the specific description of healthy or normal gut microbiota is still elusive, partly since gut microbiota is complex, dynamic, and variable across healthy individuals (Aguirre De Cárcer, 2018; Almeida et al., 2019). Nevertheless, microbiota-mediated protection studies extend from the laboratory bench-side to the clinic bed-side, highlighting the necessity to understand the molecular bases of such protective effects.

Despite the growing interest in using direct and indirect microbiota-mediated protection to prevent or treat diseases, the underlying molecular mechanisms are still mostly unclear. The intricacies of mammalian systems (e.g., humans) and microbiota make studying the underlying molecular mechanisms more complicated in higher organisms. Thereby, a simple experimental model system that allows for manipulating both host and bacterial sides would decrease this complexity. One such simple yet robust experimental model is the nematode *Caenorhabditis elegans*, making it an ideal model to study microbiota-mediated protection. The benefits of using *C. elegans* to study microbiota-mediated protection are described in the next section.



## ***Caenorhabditis elegans*: an ideal model to study microbiota-mediated protection**

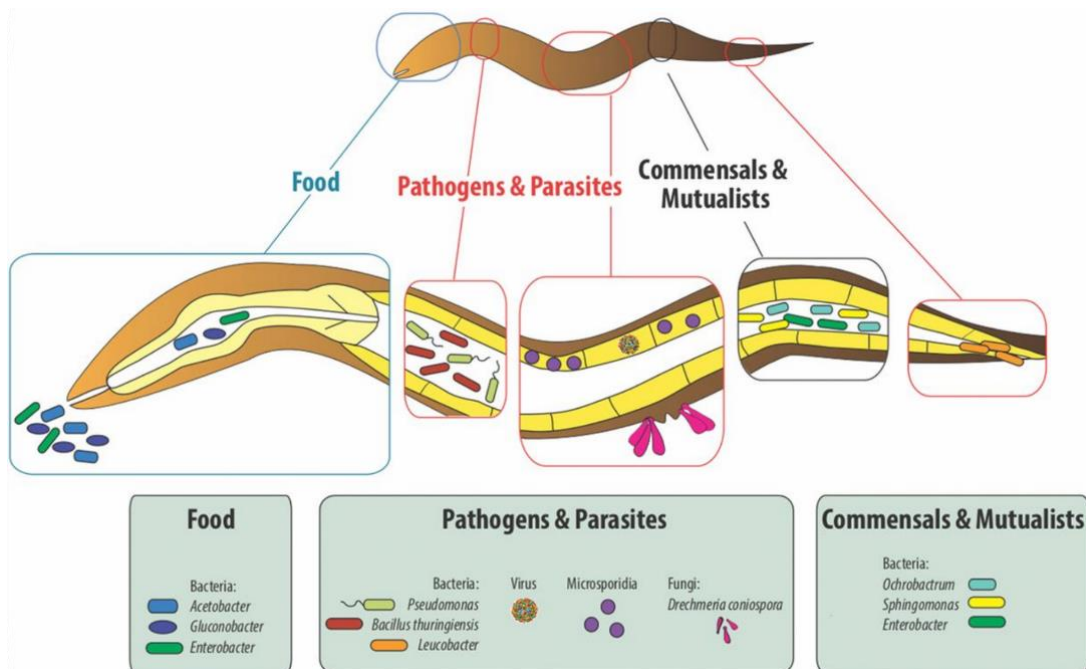
The nematode *C. elegans* is a widely used model organism in biological disciplines such as genetics, development, and neurobiology since its introduction by Sydney Brenner in the 1960s (Félix & Braendle, 2010). It has led to a myriad of significant discoveries in the field of life sciences (e.g., cell death machinery, RNA interference, Green Fluorescence Protein), some of which applicable across organisms (Apfeld & Alper, 2018). At the cellular and molecular level, evolutionarily conserved similarities in processes exist between *C. elegans* and higher organisms (Corsi et al., 2015). For instance, orthologs of human genes (60-80%) and genes associated with human diseases (40%) are reported in *C. elegans* (Corsi et al., 2015). Moreover, innate immune system architecture between the worm and human is largely conserved, with retained pathogenicity to various human pathogens, thus promoting the worm in the study of human pathogens (Aballay & Ausubel, 2002; Hoffmann et al., 1999; Mahajan-Miklos et al., 1999; Tan, et al., 1999). Such features make the apparently simple nematode a valuable model to obtain fundamental biological insights which may be transferable to higher organisms.

Furthermore, *C. elegans* has inherent advantages as an experimental model. For example, it has a small size of 1 mm for adult worms with a rapid life cycle and recovers from glycerol stocks frozen at -80°C. *C. elegans* is cheap to grow and easy to maintain since many nematodes can be grown in the limited space of a single petri dish containing a commonly available bacterial (*Escherichia coli* OP50) lawn. Other *C. elegans* features make it an ideal model to study host-microbe interactions. First, the presence of microorganisms can be controlled in *C. elegans* using a bleaching protocol. Bleaching, using a standardized procedure with alkaline hypochlorite solution, allows obtaining a gnotobiotic and age-synchronized population of nematodes (Stiernagle, 2006). Post-bleaching *C. elegans* can be grown in controlled and monoxenic conditions. Second, the nematode is transparent, allowing the visualization of bacterial colonization using microscopy in real-time and *in-vivo*. Third, several readouts that measure the effect of bacteria on *C. elegans* are well established: e.g., population growth, survival, lifespan, and fecundity assays. Fourth, *C. elegans* is genetically amenable, allowing the convenient use of various genetic tools on the nematode (e.g., RNA interference, mutant *C. elegans* strains, gene editing using CRISPR-cas9). Fifth, information on *C. elegans* genome and proteome are publicly available on “WormBase” (Stein et al., 2001). In

addition to WormExp, a tool developed specifically for *C. elegans* which allows the characterization of species-specific gene expression patterns (Yang et al., 2016). Sixth, naturally isolated *C. elegans* strains and their complete genomes are available on CeNDR (*C. elegans* Natural Diversity Resource). The CeNDR platform publicly offers the otherwise laborious dataset collection, which makes the study of genome-wide association (GWA) analyses much more feasible in *C. elegans* (Cook et al., 2017). Finally, *C. elegans* native microbiota is recently described (as discussed further below), based on which a *C. elegans* natural microbiota consortium (CeMBio) is now readily accessible for researchers (Dirksen et al., 2020). Thus, *C. elegans* is an ideal model to systematically study host-microbe interactions, such as microbiota-mediated protection against pathogen infection.

## *C. elegans* pathogen encounters

*C. elegans* is a free-living bacteriovore found in rotten plant material such as fruits (Félix & Braendle, 2010; Petersen, Saebelfeld, et al., 2015). Thus, *C. elegans* encounters and interacts with many microbes, including pathogens (Schulenburg & Félix, 2017) (**Figure 2**). *C. elegans*-pathogen interactions have driven the development of a defense system in the worm. However, many pathogens are still able to infect *C. elegans* (i.e., bacterial, fungal, and viral), using various infection routes (i.e., oral, intestinal, epidermal, vulval, and anal) (Troemel et al., 2008; Darby, 2005; Osman et al., 2018; Félix et al., 2011). In this Ph.D. thesis project, I mainly used two *C. elegans* pathogens: Gram-positive *Bacillus thuringiensis* and Gram-negative *Pseudomonas aeruginosa*; the former being the focal pathogen. Both pathogens are introduced further in the following sections.

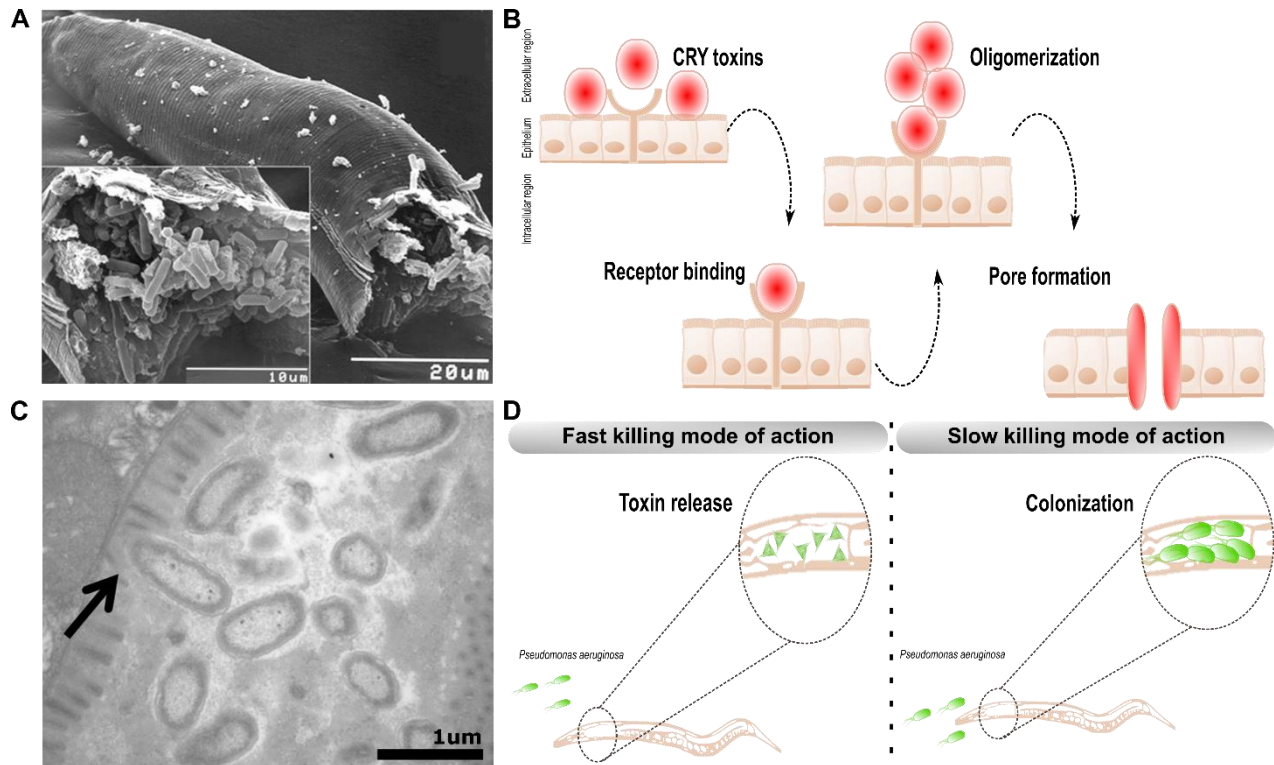


**Figure 2: Overview of biotic interactions of *C. elegans*.** The illustration shows examples of different interactions of *C. elegans* with microorganisms, such as those classified as *C. elegans* food (e.g., members of the genera *Acetobacter*, *Gluconobacter*, and *Enterobacter*), pathogens, and parasites (e.g., microsporidia, the Orsay virus, the fungus *D. coniospora*, and the bacteria *P. aeruginosa*, *B. thuringiensis*, and *Leucobacter* sp.), and commensals and mutualists (e.g., *Ochrobactrum*, *Sphingomonas*, and *Enterobacter*). Figure obtained from (Schulenburg & Félix, 2017).

*Bacillus thuringiensis*

*Bacillus thuringiensis* (Bt) is a Gram-positive, spore-forming, and soil-inhabiting bacterium. Bt coexists with *C. elegans* in its natural habitat, potentially making it a naturally relevant pathogen of the worm (Borgonie et al., 1995). Bt is used in agriculture as a natural insecticide and also to genetically modify crops to resist insect pests (Marroquin et al., 2000). That is because Bt sporulation yields crystalline parasporal inclusions, composed of one or more protoxins,  $\delta$ -endotoxins, which include nematocidal toxins (Borgonie et al., 1995; Broderick et al., 2006; Wei et al., 2003). Bt purified toxins, toxins expressed in *E. coli*, or Bt-sporulated cultures are different modes used to infect *C. elegans*. Oral uptake of the spore-toxin mixtures leads to host infection via the accumulation of Bt spores in the nematode gut and/or proliferation of Bt vegetative cells in the gut or throughout the whole body of the worm (Zárate-Potes et al., 2020). Solubilization, proteolytic activation, and binding of the toxins to intestinal receptors follow, which ultimately lead to pore formation and cell lysis in the host intestine (Griffitts et al., 2005)(**Figure 3A**). Thereby, Bt kills through pore forming toxins (PFTs), also called Cry toxins, because of their crystalline structure. Two mechanisms proposed to explain the death of infected insect hosts; A) the disruption of the midgut epithelium induces prolonged feeding cessation of the host. Thus, starvation and death follow. B) the extensive cell lysis provides spores access to a more favorable environment, where spores germinate and reproduce. Thus septicemia and death follow (Broderick et al., 2006). Bt may similarly exert its toxic effect in the nematode gut (**Figure 3B**).

Bt strains I used in this thesis project are the following: MYBt18247 (Bt247), which produces the nematocidal Cry toxins Cry6Ba2, and Cry6Ba3 (Hollensteiner et al., 2017); MYBt18679 (Bt679), which produces nematocidal Cry toxins Cry21Aa3, and Cry14Aa2 (Masri et al., 2015); MYBt407, which does not produce Cry toxins (Sheppard et al., 2013). This choice was based on the presence of previous work, which characterizes valuable aspects of *C. elegans*-Bt interactions, indicating variation in *C. elegans* behavioral and physiological defenses to Bt247 and Bt679, also evidenced in genetic and life history trait differences using experimental evolution (Hasshoff et al., 2007; Masri et al., 2015; Nakad et al., 2016; Schulenburg & Müller, 2004; Schulte et al., 2010). Moreover, phenotypic and genotypic evidence indicates a specificity for the *C. elegans* response to each of the strains Bt247 and Bt679 (Zárate-Potes et al., 2020), allowing to explore strain-specificity also in a microbiota-mediated protection context.



**Figure 3: *C. elegans* pathogens *B. thuringiensis* and *P. aeruginosa* and their infection mode of action.** (A) Scanning electron micrograph of a nematode infected with *B. thuringiensis*. The worm is broken open at the middle, shows internal disintegration and the presence of vegetative (non-infectious form) *B. thuringiensis*. Inset: higher magnification of the opened section. Photo from (Schulte et al., 2010). (B) An illustration of the mode of action for *B. thuringiensis* infection in *C. elegans*, adapted from (Griffitts & Aroian, 2005; Parker & Feil, 2005). (C) Transmission electron micrograph of the intestine cross-section of a nematode infected with *P. aeruginosa*. The arrow points to bacterial accumulation in the nematode lumen and shortened microvilli due to bacterial invasion. Photo from (Zhu et al., 2015). (D) An illustration of two possible modes of action for *P. aeruginosa* infection in *C. elegans* on solid agar: the fast killing and the slow killing modes of action. The fast killing mode of action occurs via toxin release in the nematode. The slow killing mode of action occurs via bacterial colonization in the nematode. This figure was adapted from (Kirienko et al., 2014).

### *Pseudomonas aeruginosa*

*Pseudomonas aeruginosa* is a Gram-negative, ubiquitous pathogen that thrives in soil and aquatic habitats and can colonize plants and animals (Klockgether & Tümmler, 2017; Mikkelsen et al., 2011). *P. aeruginosa* is an opportunistic pathogen, causing both acute and chronic infections in humans (Horcajada et al., 2019; Tümmler, 2019). Resulting infections range in severity, from benign to life-threatening. *P. aeruginosa* is most notorious for respiratory tract infections in intubated patients, such as those having pulmonary infections, patients with cystic fibrosis (CF), bronchiectasis, diffuse pan-bronchiolitis, and chronic obstructive pulmonary disease (Horcajada et

al., 2019; Tümmler, 2019). Moreover, it is the most widely studied *C. elegans* pathogen, particularly the strain PA14 (**Figure 3C**) (Kirienko et al., 2014). PA14, originally a clinically isolated strain, can kill *C. elegans* in different ways according to its growth conditions. When grown under low osmolarity conditions, PA14 kills *C. elegans* slowly, over several days, by colonizing and dividing in the nematode intestinal lumen (Tan et al., 1999). When grown under high osmolarity conditions, PA14 kills *C. elegans* fast, over several hours, via a low-molecular-weight diffusible toxin of the pyocyanin-phenazine class (**Figure 3D**) (Mahajan-Miklos et al., 1999). Additionally, when grown in liquid cultures, PA14 kills *C. elegans* rather slowly via a siderophore, disrupting host iron homeostasis (Kirienko et al., 2013).

In this thesis project, I used the *P. aeruginosa* PA14 strain, based on its intensive characterization as a model pathogen in humans and *C. elegans*. Also, PA14 contrasts with the Bt, which is a more naturally relevant pathogen for *C. elegans*. Thus, exploiting such pathogen differences between the Bt and PA14, could help investigate the spectrum of microbiota-mediated protection in *C. elegans*.

## ***C. elegans* defense responses**

*C. elegans*, being an invertebrate, lacks an adaptive immune response engaging professional immune cells (e.g., T and B lymphocytes), instead, the worm relies on its innate immune response. Many innate immune response mechanisms are known to be highly conserved at the molecular level among invertebrates and vertebrates (Hoffmann et al., 1999; Kimbrell & Beutler, 2001). *C. elegans* innate immune response consists of three complementary modes. These modes include behavioral responses, such as pathogen avoidance by *C. elegans* that decreases the chances of a worm-pathogen encounter. Another mode is the physical barriers in *C. elegans*, such as the worm cuticle that is a barrier against pathogen entry and the pharyngeal grinder that crushes bacteria. Finally, a third mode is the physiological innate immune responses, consisting of the following three components: recognition of a pathogen, activation of molecular signaling pathways, and the expression of immune effector molecules (Schulenburg & Müller, 2004). In the next section, I will briefly describe each of the three components of the physiological innate immune response in *C. elegans*.

Pathogen recognition is the first of three steps in an inducible immune response. Conserved pathogen structures, called microbe-associated molecular patterns (MAMPs), recognized by the immune system as “non-self”, bind to pattern recognition receptors (PRRs). The PRRs can also sense a danger signal, pathogen-induced damage on the host, the so-called danger-associated molecular patterns (DAMPs), which are recognized by the immune system as “modified self” (Dierking & Pita, 2020; Kim & Ewbank, 2018; Matzinger, 1994). Following pathogen recognition, the second step in the inducible immune response to the pathogen is activation of molecular signaling pathways. A molecular signaling cascade of pathways transfers the information to downstream targets, generally to a transcription factor or a network, which triggers a response. In *C. elegans*, there are several central signaling cascades, including three mitogen-activated protein kinase (MAPK) pathways (i.e., p38, the extracellular signal-regulated kinase (ERK), and the c-Jun-N-terminal kinase (JNK) MAPK pathways), a transforming growth factor- $\beta$ -like (TGF- $\beta$ ) pathway, and an insulin-like receptor (ILR/DAF-2) pathway (Dierking et al., 2016). The signaling pathways in *C. elegans* activate or repress downstream transcription factors. The third step in an inducible immune response is the expression of immune effector molecules, which finally destroy the pathogen or pathogenic determinants. Effector molecules rely on triggering protective cellular responses, such as autophagy, the production of reactive oxygen species (ROS), and the production

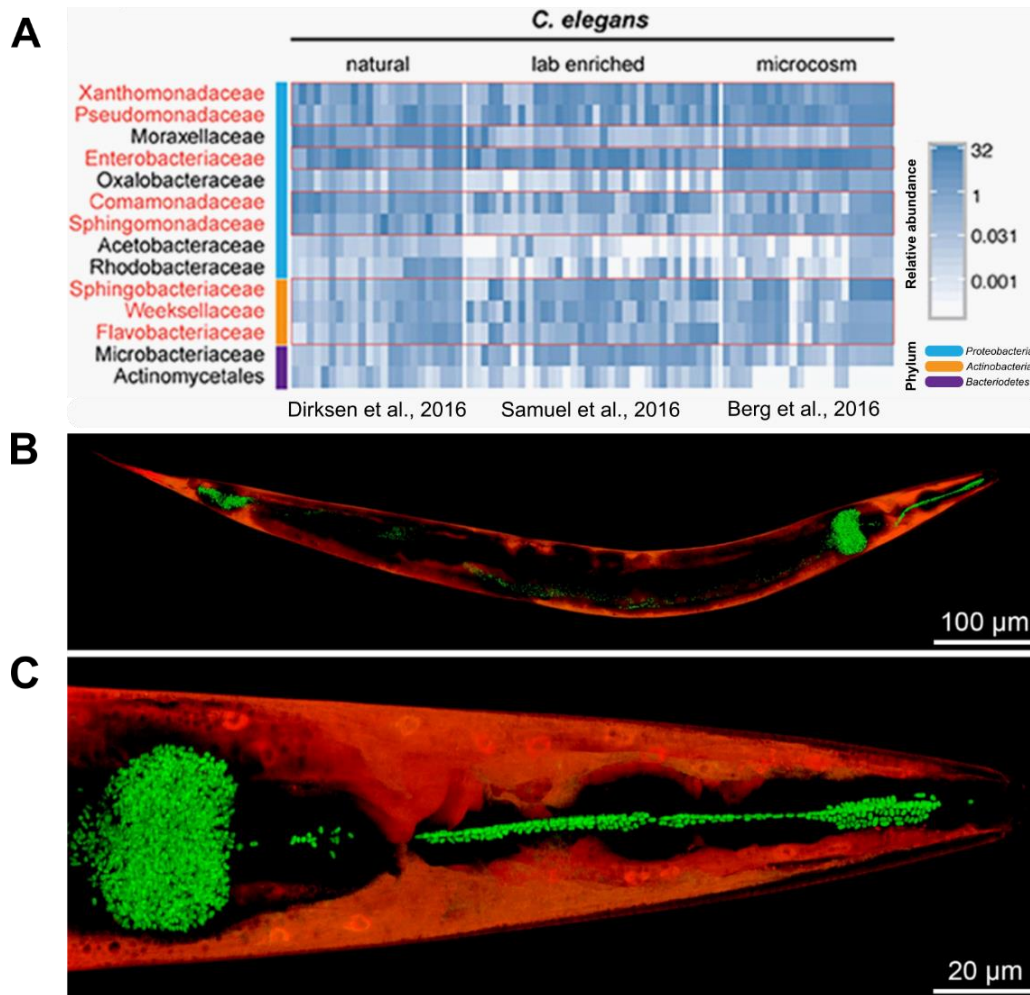
of antimicrobial proteins that aim to neutralize the pathogen (Dierking, Yang and Schulenburg, 2016). *C. elegans* also rely on putative immune effectors, which include the lysozymes, caenopores, caenacins, AntiBacterial Factor(ABF) peptides, and Neuropeptide-Like(NLP) peptides (Dierking et al., 2016; Roeder et al., 2010). *C. elegans* uses all three components in its physiological defense against pathogens.

*C. elegans* immune defense described against the focal pathogen, Bt, used in this thesis, include the following characterized pathways: p38 (Huffman et al., 2004), JNK MAPK (Kao et al., 2011; Treitz et al., 2015; Yang et al., 2015), hypoxia pathway (Bellier et al., 2009), unfolded protein response (UPR) (Bischof et al., 2008), and autophagy (Chen et al., 2017). These pathways are required for defense against *Escherichia coli* expressing PFT Cry5B. DAF-2/ILR pathway (Hasshoff et al. 2007; Yang et al. 2015) and DAF-16 Forkhead Box O (FOXO) (Wang, Nakad, and Schulenburg, 2012) are described in *C. elegans* infected with Bt247. Additionally, effector proteins are involved in *C. elegans* defense against Bt PFTs. These effector proteins include saposin-like proteins (caenopores) (SPP-12) (Hoeckendorf & Leippe, 2012), lysozymes (*lys-2*, *lys-5*, *lys-7*) (Boehnisch et al., 2011), and galectins (LEC-6, LEC-8) (Ideo et al., 2009; Bing Wang et al., 2017). Moreover, the strain-specific response of *C. elegans* to Bt247 and Bt679 is shown to be via the GATA transcription factor ELT-2 (Zárate-Potes et al., 2020).



## ***C. elegans* natural microbiota**

Alongside its pathogen encounters, *C. elegans* is also characterized by interactions with its natural microbiota, which has been neglected until recently. Three pioneering studies on the *C. elegans* microbiome, published in 2016, showed that *C. elegans* has a defined, species-rich, and non-random microbiome (Berg et al., 2016a; Dirksen et al., 2016; Samuel et al., 2016) (**Figure 4A**). In these three studies, the taxonomic composition of *C. elegans* microbiome was described using slightly different experimental approaches. *C. elegans* microbiome was either sampled from the nematodes' associated natural habitat (Samuel et al., 2016), wild nematodes collected from rotting plant material (Dirksen et al., 2016), or worms grown in the lab, in compost microcosm conditions (Berg et al., 2016a). Despite variations in experimental designs, laboratory settings, and perturbances originating from a natural versus an artificial environment, these three studies were followed by a meta-analysis, which shows an overlap in certain bacterial taxa, indicating a core *C. elegans* natural microbiome (Zhang et al., 2017). The resulting *C. elegans* core microbiome, at the family level, is composed of: *Xanthomonadaceae*, *Pseudomonadaceae*, *Enterobacteriaceae*, *Comamonadaceae*, *Sphingomonadaceae*, *Sphingobacteriaceae*, *Weksellaceae* and *Flavobacteriaceae* (Zhang et al., 2017) (**Figure 4A**). Moreover, *C. elegans* microbiome was also described from wild worms in the Kiel Botanical Garden over many seasons, where variations, at the species and strain levels, in the microbiome community composition were also noted (Johnke et al., 2020). Similar observations were reported in the transcriptome data analysis of worms grown on the synthetic communities of previously isolated worm gut microbiome (Berg et al., 2019). Here, the authors also show the contribution of host genes (i.e., role of TGF $\beta$ /Bone Morphogenetic Protein (BMP) signaling), in controlling the abundances of *Enterobacter*, and thus affecting microbiome community composition (Berg et al., 2019). More recently, based on the current knowledge on *C. elegans* core microbiome, and in parallel to the research conducted in this thesis, a consortium of 12 of its members called CeMbio has been established, making it much more feasible to perform future research on the *C. elegans* microbiome (Dirksen et al., 2020).



**Figure 4: *C. elegans* has a core natural microbiome.** A) The heatmap shows 14 bacterial families present in 100 % of the natural worm microbiome (Dirksen et al., 2016) and their abundance across samples. Here the microbiome of wild *C. elegans* was used. Red highlighted bacterial families that are also abundant in lab-enriched microbiome and microcosm. In lab-enriched samples, the microbiome of *C. elegans* from its natural habitat was used (Samuel et al., 2016), while in the microcosm, the microbiome of *C. elegans* raised in soil emulating the natural worm habitat was used (Berg et al., 2016a). The column colors on the right of the heatmap (blue, yellow, and purple) indicate the different bacterial phyla (Blue: *Proteobacteria*, Yellow: *Bacteroidetes*, Purple: *Actinobacteria*). Figure adapted from (Zhang et al., 2017). B-C) Colonization of *C. elegans* with the native microbiota isolate *Pseudomonas fluorescens* MYb115 labeled with a GFP fluorescent marker. Accumulation of MYb115::sfGFP is observed, especially in the front and posterior regions of the *C. elegans* gut in (B) and the *C. elegans* pharynx and grinder in (C). A transgenic *C. elegans* strain that expresses dsRed under the control of an epidermal promoter (*pcol-12::dsRed*) was used. CSLM image is courtesy of Jan Michels and Lena Peters.

Moreover, *C. elegans* microbiota isolates (i.e., several *Proteobacteria*) colonize and persist in the worm gut; some (e.g., *Ochrobactrum* and *Pseudomonas*) persist even under stressful conditions (e.g., high temperature, high osmolarity, bacterial or fungal pathogens)(Dirksen et al., 2016; Kissoyan et al., 2019; Yang et al., 2019). Microbiota colonization and persistence indicates that

these bacteria are more than just food to the bacterivore worm (**Figure 4B&C**). How *C. elegans* microbiota affects host physiology in the absence or presence of different stressful conditions is not entirely clear. Few studies show *C. elegans* physiological features affected by its natural microbiota, including, for example, transcriptome and proteome studies with *C. elegans* microbiota *Ochrobactrum anthropi* MYb71 and *Ochrobactrum pituitosum* MYb237 indicate the microbiota-mediated changes on the nematode dietary response, development, fertility, immunity and energy metabolism (Cassidy et al., 2018; Yang et al., 2019). *C. elegans* microbiota is also shown to have neuromodulatory effects on the host. *Providencia alcalifaciens*-produced tyramine manipulates the nematode sensory decision via bypassing the biosynthesis of tyramine, affecting the worm behavior as well as food choice (O'Donnell et al., 2020). However, it remains unclear what the *C. elegans* natural microbiota can do and how relevant are the different microbiota interactions in a more natural context. This is further explored in **Chapter III** of this thesis, where we investigated the functional repertoires of *C. elegans* natural microbiota members.

Recent exploration of *C. elegans* in its naturally relevant environmental context has also provided new insights on worm immunity against pathogen infection. Bacterial isolates belonging to the different genera, i.e., *Enterobacter*, *Gluconobacter* and *Pseudomonas*, are shown to enhance host immune defenses against pathogens (Berg et al., 2016b; Dirksen et al., 2016; Kissoyan et al., 2019; Montalvo-Katz et al., 2013; Samuel et al., 2016). *Enterobacter cloacae* protects *C. elegans* against the pathogen *Enterococcus faecalis* via, so far, unknown mechanism of protection (Berg et al., 2016b). Similarly, *Gluconobacter* protects the worms against detrimental bacterial members of its natural environment (i.e., *Chryseobacterium* sp. JUb44, *Serratia* sp. JUb9, and *Pseudomonas* sp. GRb0427), whereby the mechanism of protection is unknown yet (Samuel et al., 2016). *Pseudomonas mendocina* protects the nematode against *P. aeruginosa* indirectly via the activation of the p38MAPK pathway (described further below)(Montalvo-Katz et al., 2013). *Pseudomonas* sp, including *P. lurida* MYb11, protects the worm against the fungal pathogen *Drechmeria coniospora*; however, its mechanism in fungal protection is still unknown (Dirksen et al., 2016). Intriguingly, *Pseudomonas lurida* (MYb11&MYb12) also protects the nematode against *P. aeruginosa* and *B. thuringiensis* (Kissoyan et al., 2019). This is presented in **Chapter I** of this thesis, where we explored and identified a direct protection mechanism against *B. thuringiensis* in *C. elegans*. Also, in **Chapter I**, other *Pseudomonas* sp. are identified that protect the worm against both pathogens,

however the mechanisms of protection thereof are still elusive, thus we have explored them further in **Chapters II** and **IV** of this thesis.

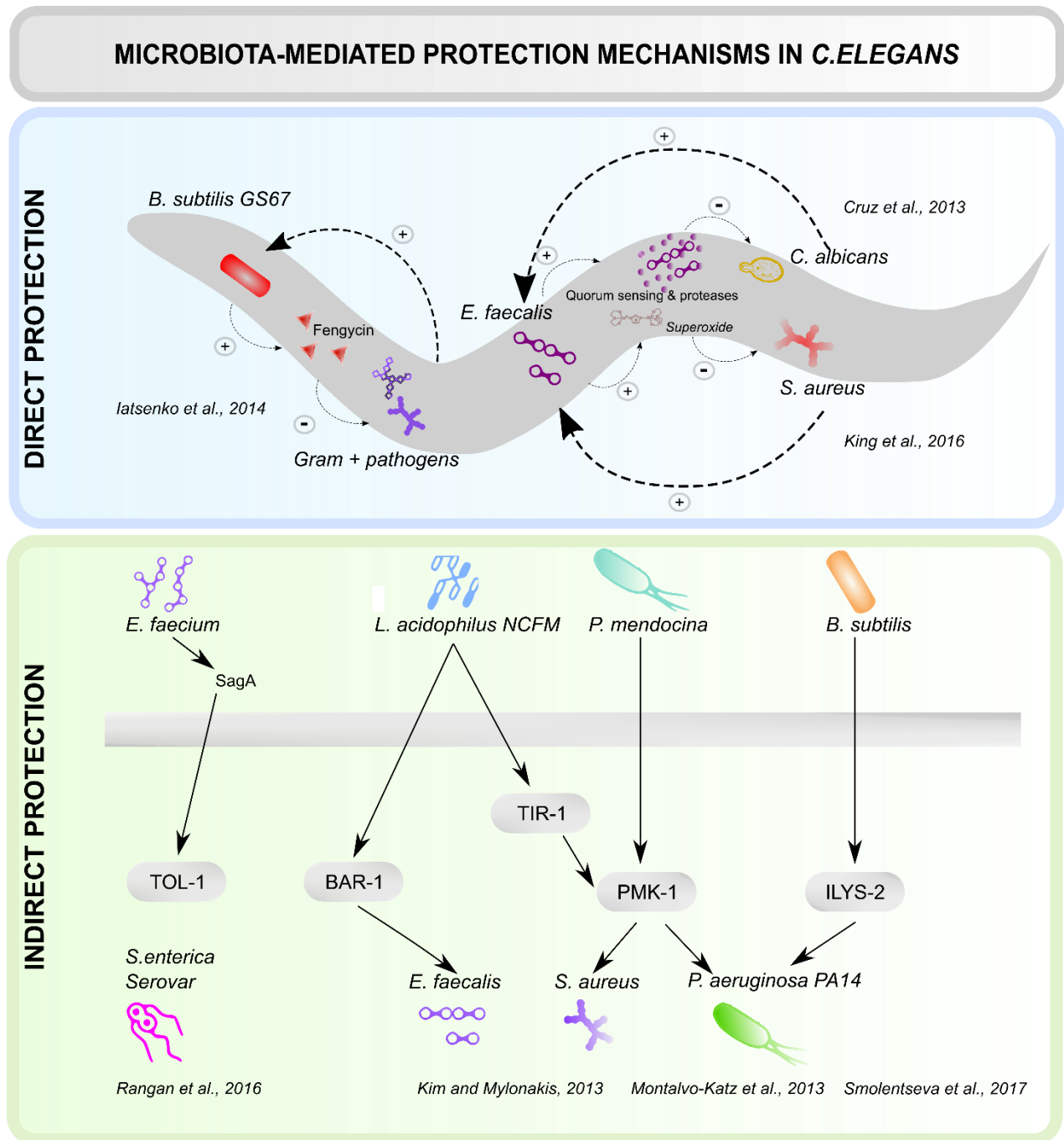
Further below, I describe a select set of studies most relevant to this thesis, where microbiota-mediated protection mechanism is described in more detail. In these examples, the microbiota or pathogen or both are relevant to the human rather than the nematode. Nevertheless, these studies provide insights, which can be helpful to elucidate the molecular mechanisms underlying *C. elegans*- natural microbiota interactions in pathogen-infected worms.

## Microbiota-mediated protection in *C. elegans*

Long before identifying *C. elegans* natural microbiota, *C. elegans* has been used to study host-microbe interactions. However, in these studies, the microbes used are mostly human microbiota or pathogen species, which are not necessarily relevant to *C. elegans*. Below are examples of such studies investigating microbiota-mediated protection mechanisms in *C. elegans*, classified here as direct or indirect protection.

### *Direct microbiota-mediated protection in C. elegans*

Microbiota isolates can protect *C. elegans* directly via interference competition with pathogenic bacteria. For example, *Bacillus subtilis* GS67 was shown to protect *C. elegans* from Gram-positive pathogens via an inhibitory cyclic lipopeptide fengycin. Fengycin-mediated microbial antagonism directly protects *C. elegans* from Gram-positive pathogens (**Figure 5**)(Iatsenko et al., 2014). Further, *Enterococcus faecalis* was shown to evolve into a protective microbe when *C. elegans* was challenged with *Staphylococcus aureus*, a more virulent pathogen (King et al., 2016). *E. faecalis*-induced protection was shown to be via antimicrobial superoxide production (King et al., 2016). Moreover, worms fed with *E. faecalis* showed protection against infection with fungus *Candida albicans*. Here, *E. faecalis*-induced protection was shown to be due to the quorum sensing *fsr* locus in *E. faecalis* and proteases GelE and SerE (**Figure 5**)(Cruz et al., 2013). Whether similar mechanisms are also involved with *C. elegans* natural microbiota isolates requires further exploration.



**Figure 5: Microbiota-mediated protection mechanisms in *C. elegans*, classified into direct or indirect protection.** The blue box presents the different microbiota-mediated protection examples (microbiota not native to *C. elegans*) that protect the host via a direct protection mechanism. These include: *Bacillus subtilis* GS67-produced inhibitory fengycin, which protects the worm against Gram-positive pathogens; *Enterococcus faecalis*-induced quorum sensing and superoxide production that protect the worm from *Staphylococcus aureus* and *Candida albicans* infections, respectively. The green box presents different examples of microbiota-mediated protection that are under the umbrella of indirect protection. These examples include: *Enterococcus faecium*-secreted antigen A (SagA), which generates muramyl-peptide fragments that induce pathogen tolerance in *C. elegans* via signaling through TOL-1; *L. acidophilus* NCFM -induced *Enterococcus faecalis* protection, which involves signaling through PMK-1/TIR-1 and BAR-1 pathways; *Pseudomonas mendocina*-induced *P. aeruginosa* PA14 protection involving signaling through PMK-1 pathway, and *Bacillus megaterium*-induced *P. aeruginosa* PA14 protection with unknown mechanism of action.

*Bacillus subtilis* protects *C. elegans* via ILYS-2-induced-biofilm production that inhibits *Pseudomonas aeruginosa* PA14 infection; (+) indicates induction, (-) indicates suppression.

### *Indirect microbiota-mediated protection in C. elegans*

Microbiota isolates can protect indirectly by modulating *C. elegans* immune defenses. For example, *Enterococcus faecium* secretes antigen A (SagA), which protects *C. elegans* against *Salmonella enterica* serovar Typhimurium pathogenesis via inducing pathogen tolerance. The bacterial NlpC/p60 peptidoglycan hydrolase activity of SagA generates muramyl-peptide fragments that induce pathogen tolerance in *C. elegans* via a *tol-1*-dependent manner. Moreover, SagA that was heterologously expressed induced protection against *Salmonella* pathogenesis in *C. elegans* and mice (**Figure 5**)(Rangan et al., 2016).

The probiotic strain *Lactobacillus acidophilus* NCFM is shown to modulate *C. elegans* immune response and protect *C. elegans* against Gram-positive pathogens. Worms fed with *L. acidophilus* show prolonged survival against the pathogens *Enterococcus faecalis* and *Staphylococcus aureus*. *L. acidophilus* activates key immune signaling pathways involved in *C. elegans* defenses against Gram-positive bacteria. Immune response investigations using the *C. elegans* mutant strains *tir-1*, *bar-1*, and *pmk-1* showed the involvement of the Toll/interleukin-1 receptor, the  $\beta$ -catenin signaling pathway, and the p38 MAPK pathway, respectively (Kim & Mylonakis, 2012). While the precise mechanism is yet unknown, it is likely to involve detecting MAMPs or DAMPs, which signal an upregulation of immune defense pathways (Clark & Hodgkin, 2014; Kim & Mylonakis, 2012) (**Figure 5**)

Moreover, *Pseudomonas mendocina*, a member of laboratory-established *C. elegans* microbiota, protects *C. elegans* against infection with PA14 (Montalvo-Katz et al., 2013). Here again, *P. mendocina*-mediated protection was shown to be via worm immune response activation. *P. mendocina*-induced protective effect was not observed in *C. elegans pmk-1* mutants, indicating the involvement of the p38 MAPK pathway (**Figure 5**)(Montalvo-Katz et al., 2013). In another study, *B. subtilis* was shown to protect *C. elegans* against *P. aeruginosa* PA14 via biofilm formation, induced by *C. elegans* innate immune response. Worms fed with *B. subtilis* formed protective biofilm, induced by *ilys-2* activation in the worm. *ilys-2* encodes an invertebrate lysozyme in the worm's innate immune response. *ilys-2* knockout mutant worms fed with *B. subtilis* did not form protective biofilm and were not protected from *P. aeruginosa*. Thus, *ilys-2* activity in *C. elegans*

was shown to be responsible for biofilm-mediated protection from *P. aeruginosa* (**Figure 5**)(Smolentseva et al., 2017).

Overall, it is crucial to highlight that most of the microbiota and pathogen combinations mentioned above are relevant to humans but irrelevant to *C. elegans* in its natural habitat. Investigating *C. elegans* instead with pathogens that are naturally relevant to the nematode might be more valuable. Natural microbes have likely molded the functional diversity of the nematode's gene repertoire, and thus their use may better describe the underlying protection mechanisms elicited under pathogenic stress (Schulenburg et al., 2008). Furthermore, the nematode may produce a more specific and efficient response to natural pathogens against which the nematode has had the time to evolve (Schulenburg & Müller, 2004). Therefore, I use the *C. elegans* natural microbiota to explore microbiota-mediated protection.

In conclusion, the functional characterization of *C. elegans* natural microbiota is still to be uncovered, including the protective functions that affect the worm in stressful conditions of pathogen infection. Consequently, the underlying mechanisms of microbiota-mediated protection in the nematode remain elusive. Therefore, I exploited the recently described *C. elegans* natural microbiota and the benefits of the *C. elegans* model system in a focused study on *C. elegans*-microbiota-pathogen interactions to address these gaps in the literature.



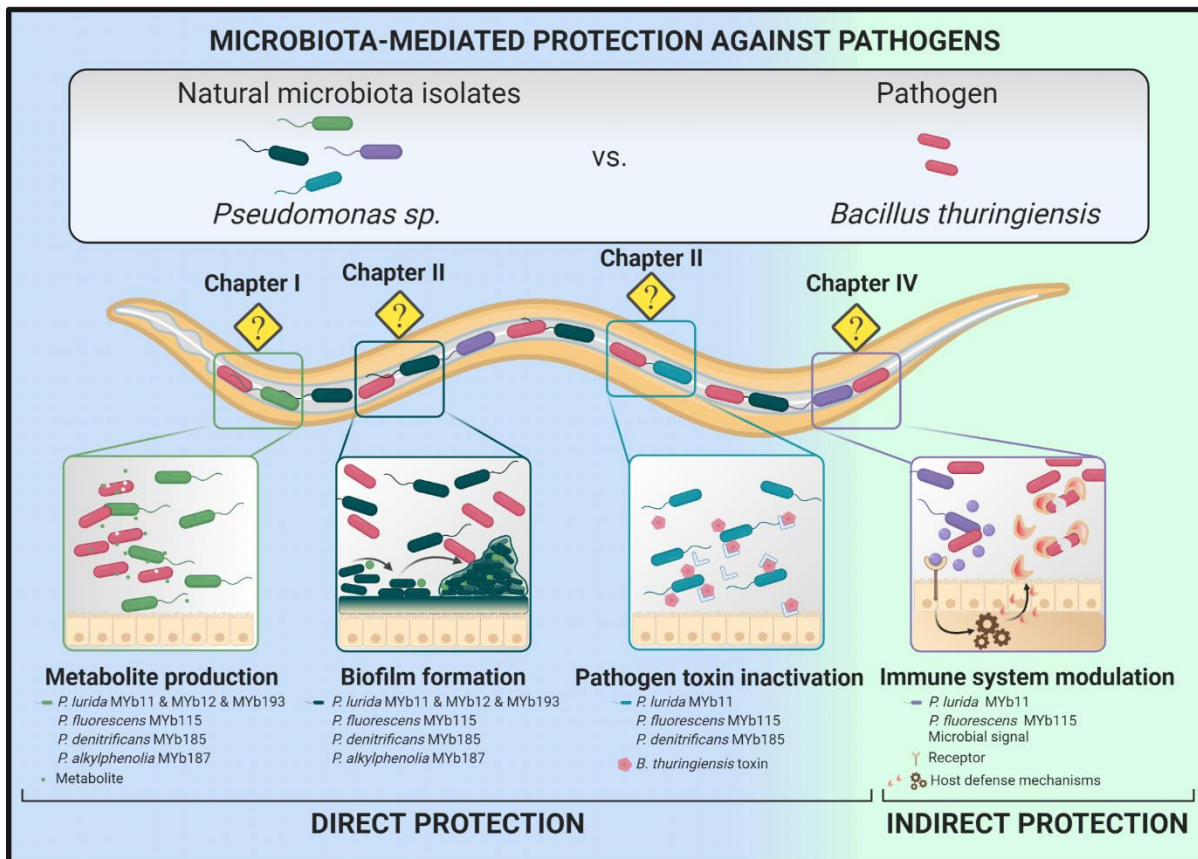
## Objectives and overview of this thesis

In this Ph.D. thesis project, I aimed to explore the role of *C. elegans*' natural microbiota in protecting the host against pathogen infection. Subsequently, I investigated the underlying interactions between host and microbiota, with a focus on mechanisms of microbiota-mediated protection, viewed both from the bacterial and the host perspectives. The following paragraphs briefly describe the aims and contents of the chapters in this thesis.

**Chapter I** aimed to explore the *C. elegans* natural microbiota protective potential and to investigate the mechanisms of direct microbiota mediated protection, focusing on the involvement of microbiota-derived metabolites (**Figure 6**). We identified natural *C. elegans* microbiota isolates that increase *C. elegans* resistance to pathogen infection. We showed that *C. elegans* natural microbiota isolates belonging to the genus *Pseudomonas* provide paramount protection to nematodes infected with the pathogen *Bacillus thuringiensis*. Moreover, we found that the *Pseudomonas* isolates *P. lurida* MYb11 and MYb12 protect the worm against *B. thuringiensis* by directly inhibiting pathogen growth, both *in vitro* and *in vivo*. *P. lurida*-induced direct protection was via the cyclic lipopeptide of the viscosin group, massetolide E. Other protective *Pseudomonas* isolates (i.e., *P. fluorescens* MYb115) induced microbiota-mediated protection indirectly, most likely through the activation of host immune defense mechanisms. Thus, we here provide new insights on the functional importance of the natural microbiota isolates of *C. elegans* and expand the knowledge on bacterial metabolites that influence host physiology upon pathogen infection. This chapter is a manuscript that was published in Current Biology on March 18, 2019.

**Chapter II** was aimed to build on our previous findings of microbiota-mediated protection and delve deeper into investigating two additional potential direct protection mechanisms: the capacity of the protective microbiota isolates a) to directly inactivate the *B. thuringiensis* toxins (Cry21Aa3 and Cry14Aa2), and b) to produce biofilm, acting as a protective barrier against *B. thuringiensis* infection (**Figure 6**). Here, we show that some *C. elegans* natural microbiota isolates *P. lurida* MYb11, *P. fluorescens* MYb115, and *Pseudomonas denitrificans* MYb185 protect the worm from infection with Bt spores but not from exposure to Bt toxins. Furthermore, protective microbiota isolates shift from being protective to harmful to the worm, depending on pathogen exposure conditions, revealing a context-dependent dual nature of *C. elegans* microbiota function. We also show that the protective *Pseudomonas* isolates MYb11, MYb12, *P. denitrificans* MYb185, and *P.*

*lurida* MYb193 produce biofilm *in vitro*, as an initial step in investigating biofilm-mediated protection. Overall, we here eliminate microbiota-mediated toxin-inactivation as a protection mechanism, while suggesting the involvement of biofilm, which merits further *in vivo* investigation. This chapter is a manuscript ready for submission.

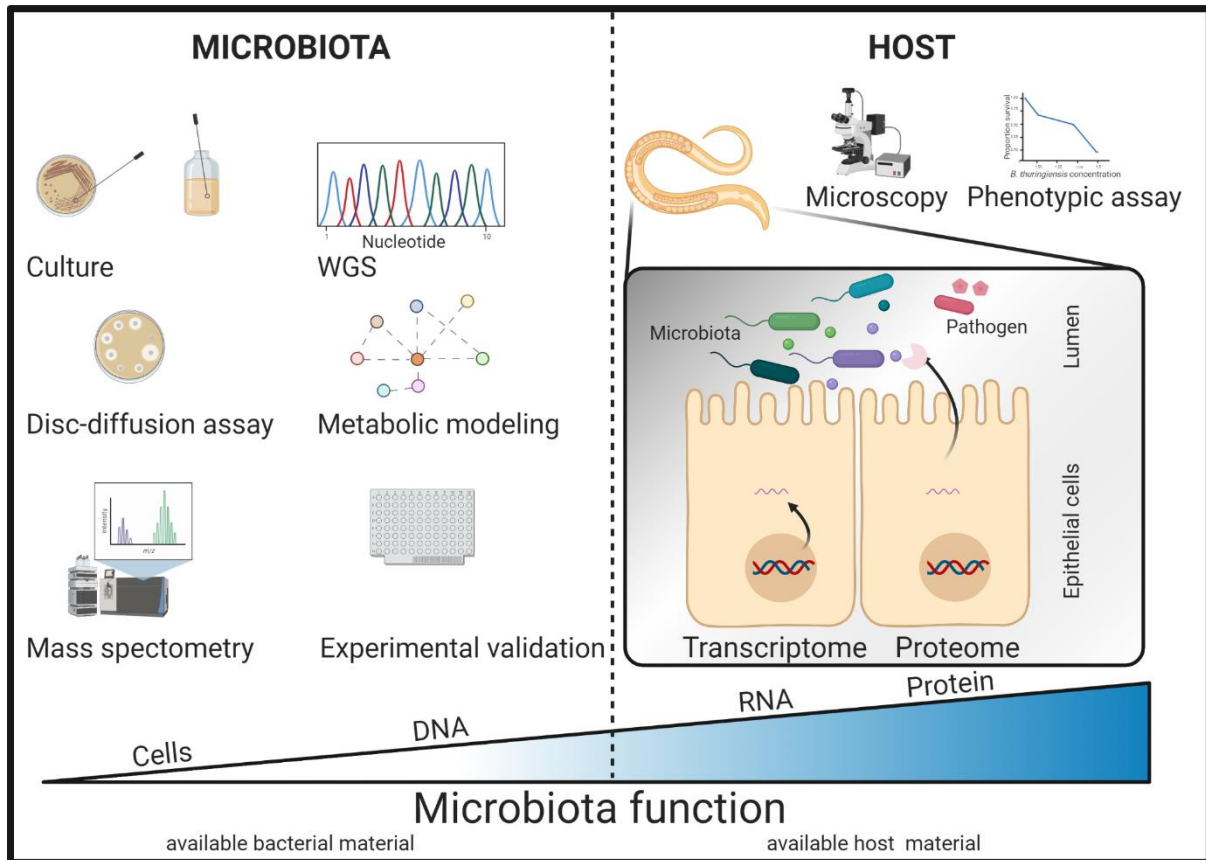


**Figure 6:** The focal objectives of each of Chapters I, II, and IV of this thesis were to explore how *C. elegans* natural microbiota isolates of the genus *Pseudomonas* sp. protect the host from infection with the pathogen *B. thuringiensis*. Microbiota-mediated protection in this thesis was investigated first on bacterial and then on the host side. On the bacterial side, in **Chapters I and II**, the focal objectives were to explore if microbiota-mediated protection of *C. elegans* natural microbiota member, *Pseudomonas lurida* MYb11, MYb12, and MYb193, *P. fluorescens* MYb115, *P. denitrificans* MYb185, and *P. alkylphenolia* MYb187 are achieved via direct protection. Further elucidating on direct protection mechanisms, the focus was on exploring the possibility of metabolite protection, biofilm production and pathogen toxin inactivation by the protective microbiota. On the host side, in **Chapter IV**, the focus was to explore if microbiota-mediated protection in *C. elegans* is achieved via host immune system modulation. This figure was created with Biorender.com.

**Chapter III** aimed to elucidate the functions of a more extensive list of *C. elegans* natural microbiota isolates. Here, in collaboration with my colleagues, we used whole-genome sequences of 77 of the natural microbiota isolates of *C. elegans* to perform metabolic modeling accompanied by experimental verification. We showed that *C. elegans* natural microbiota can synthesize essential nutrients for the worm. Moreover, we identified bacterial traits that influence the microbiota's ability to colonize the nematode and subsequently affect its fitness. Overall, we introduce a resource that improves our understanding of the nematode in a more natural setting and a resource that provides a novel framework to characterize microbiota-mediated functions in *C. elegans*. This chapter is a manuscript that was published in the ISME Journal on September 4, 2019.

**Chapter IV:** was aimed at exploring microbiota-mediated inducible responses in *C. elegans* upon pathogen infection. Here we used a multi-omics approach to identify host gene expression and protein abundance profiles upon microbiota-mediated protection in the worm (**Figure 6**). Our results indicate that protective microbiota isolates *Pseudomonas lurida* MYb11 and *Pseudomonas fluorescens* MYb115 influence the worm's innate immune response and cellular structural component functions. Comparing the transcriptome and proteome data revealed overlaps between differentially expressed genes and their corresponding protein abundances for some galectins, C-type lectins, and lysozymes, suggesting their potential roles in microbiota-mediated protection. Our results suggest microbiota-mediated activation of the innate immune response via priming the host to subsequent pathogen attack. Furthermore, our findings provide a framework at both the transcript and protein level and a list of candidate genes for further functional investigation of *C. elegans* host-microbiota-pathogen interactions. This chapter is a manuscript in preparation.

Overall, this thesis adds to the knowledge of the functions of natural microbiota isolates of *C. elegans*, particularly in host protection, identifying protective mechanisms on the bacterial and host side. An illustration, presented below, summarizes the overall approach used in this thesis (**Figure 7**). Finally, in the **General Discussion**, the findings of this thesis are further discussed in the light of overarching concepts, considering the current literature, discussing the limitations of this work, and proposing future directions and perspectives on microbiota-mediated protection research.



**Figure 7: Overall approach used in this thesis.** The approach used in this thesis to study microbiota-mediated protection starts with identifying the microbiota functional potential. For this, bacterial cultures on solid agar plates and in liquid nutrient broth, along with disc-diffusion and mass spectrometry techniques, were used (**Chapters I and II**). Cell-level data investigation on the microbiota isolates was followed with bacterial DNA isolation and whole-genome sequencing (WGS), metabolic modeling and experimental validation (**Chapter III**). Microbiota-mediated protection on the host i.e., *C. elegans* side was investigated using microscopy and phenotypic assays, in addition to the isolation of host RNA and transcriptome and proteome level analyses of *C. elegans* with and without microbiota and pathogen *B. thuringiensis* exposure conditions (**Chapter IV**). This figure was created with Biorender.com.

## REFERENCES

- Aballay, A., & Ausubel, F. M. (2002). *Caenorhabditis elegans* as a host for the study of host–pathogen interactions. *Current Opinion in Microbiology*, 5(1), 97–101.
- Abraham, N. M., Liu, L., Jutras, B. L., Yadav, A. K., Narasimhan, S., Gopalakrishnan, V., Ansari, J. M., Jefferson, K. K., Cava, F., Jacobs-Wagner, C., & Fikrig, E. (2017). Pathogen-mediated manipulation of arthropod microbiota to promote infection. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), E781–E790.
- Aguirre De Cárcer, D. (2018). The human gut pan-microbiome presents a compositional core formed by discrete phylogenetic units. *Scientific reports*, 8, 14069.
- Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D., & Finn, R. D. (2019). A new genomic blueprint of the human gut microbiota. *Nature*, 568(7753), 499–504.
- Apfeld, J. and Alper, S. (2018) What can we learn about human disease from the nematode *C. elegans*?, *Methods in Molecular Biology*. Johanna K. DiStefano (ed.), *Disease Gene Identification: Methods and Protocols, Methods in Molecular Biology*, vol. 1706, Humana Press Inc., 53–75.
- Baldrian, P. (2017). Forest microbiome: Diversity, complexity and dynamics. *FEMS Microbiology Reviews*, 41(2), 109–130.
- Belkaid, Y., & Harrison, J. O. (2017). Homeostatic immunity and the microbiota. *Immunity*, 46(4), 562-576.
- Bellier, A., Chen, C.-S., Kao, C.-Y., Cinar, H. N., & Aroian, R. V. (2009). Hypoxia and the hypoxic response pathway protect against pore-forming toxins in *C. elegans*. *PLOS Pathogens*, 5(12), e1000689.
- Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M. C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., ... Schloter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, 8(1).
- Berg, M., Monnin, D., Cho, J., Nelson, L., Crits-Christoph, A., & Shapira, M. (2019). TGF $\beta$ /BMP immune signaling affects abundance and function of *C. elegans* gut commensals. *Nature Communications*, 10(1), 604.
- Berg, M., Stenuit, B., Ho, J., Wang, A., Parke, C., Knight, M., Alvarez-Cohen, L., & Shapira, M. (2016a). Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *The ISME Journal*, 10(8), 1998–2009.
- Berg, M., Zhou, X. Y., & Shapira, M. (2016b). Host-specific functional significance of *Caenorhabditis* gut commensals. *Frontiers in Microbiology*, 7(1622).
- Bergstrom, K. S. B., Kisson-Singh, V., Gibson, D. L., Ma, C., Montero, M., Sham, H. P., Ryz, N., Huang, T., Velcich, A., Finlay, B. B., Chadee, K., & Vallance, B. A. (2010). Muc2 protects against lethal infectious colitis by disassociating pathogenic and commensal bacteria from the colonic mucosa. *PLOS Pathogens*, 6(5).
- Bischof, L. J., Kao, C.-Y., Los, F. C. O., Gonzalez, M. R., Shen, Z., Briggs, S. P., van der Goot, F. G., & Aroian, R. V. (2008). Activation of the Unfolded Protein Response is required for defenses against bacterial pore-forming toxin *in vivo*. *PLoS Pathogens*, 4(10), e1000176.
- Blaser, M. J. (2015). Missing microbes: How the overuse of antibiotics is fueling our modern plagues. New York: Picador.

- Boehnisch, C., Wong, D., Habig, M., Isermann, K., Michiels, N. K., Roeder, T., May, R. C., & Schulenburg, H. (2011). Protist-type lysozymes of the nematode *Caenorhabditis elegans* contribute to resistance against pathogenic *Bacillus thuringiensis*. *PLoS ONE*, 6(9), e24619.
- Bojanova, D. P., & Bordenstein, S. R. (2016). Fecal transplants: what is being transferred? *PLoS Biology*, 14(7), e1002503.
- Borgonie, G., Van Driessche, R., Leyns, F., Arnaut, G., De Waele, D., & Coomans, A. (1995). Germination of *Bacillus thuringiensis* spores in bacteriophagous nematodes (Nematoda: Rhabditida). *Journal of Invertebrate Pathology*, 65(1), 61–67.
- Broderick, N. A., Raffa, K. F., & Handelsman, J. (2006). Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(41), 15196–15199.
- Brown, R. L., & Clarke, T. B. (2017). The regulation of host defences to infection by the microbiota. *Immunology*, 150(1), 1–6.
- Brown, W. R. (2014). Fecal microbiota transplantation in treating *Clostridium difficile* infection. *Journal of Digestive Diseases*, 15(8), 405–408.
- Buffie, C. G., & Pamer, E. G. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. *Nature Reviews Immunology*, 13(11), 790–801.
- Caragata, E. P., Rancès, E., Hedges, L. M., Gofton, A. W., Johnson, K. N., O’Neill, S. L., & McGraw, E. A. (2013). Dietary cholesterol modulates pathogen blocking by *Wolbachia*. *PLOS Pathogens*, 9(6), e1003459.
- Cassidy, L., Petersen, C., Treitz, C., Dierking, K., Schulenburg, H., Leippe, M., & Tholey, A. (2018). The *Caenorhabditis elegans* proteome response to naturally associated microbiome members of the genus *Ochrobactrum*. *Proteomics*, 18(8), 1700426.
- Chen, H.-D., Kao, C.-Y., Liu, B.-Y., Huang, S.-W., Kuo, C.-J., Ruan, J.-W., Lin, Y.-H., Huang, C.-R., Chen, Y.-H., Wang, H.-D., Aroian, R. V., & Chen, C.-S. (2017). HLH-30/TFEB-mediated autophagy functions in a cell-autonomous manner for epithelium intrinsic cellular defense against bacterial pore-forming toxin in *C. elegans*. *Autophagy*, 13(2), 371–385.
- Chiu, L. et al. (2017). Protective microbiota: from localized to long-reaching co-immunity. *Frontiers in Immunology*, 8, 1678.
- Clark, L. C., & Hodgkin, J. (2014). Commensals, probiotics and pathogens in the *Caenorhabditis elegans* model. *Cellular Microbiology*, 16(1), 27–38.
- Collins, F. M., & Carter, P. B. (1978). Growth of salmonellae in orally infected germ-free mice. *Infection and Immunity*, 21(1), 41–47.
- Cook, D. E., Zdraljevic, S., Roberts, J. P., & Andersen, E. C. (2017). CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Research*, 45.
- Corsi, A. K., Wightman, B. and Chalfie, M. (2015). A transparent window into biology: A primer on *Caenorhabditis elegans*. *Genetics*, 200(2), 387–407.
- Croswell, A., Amir, E., Tegatz, P., Barman, M., & Salzman, N. H. (2009). Prolonged impact of antibiotics on intestinal microbial ecology and susceptibility to enteric Salmonella infection. *Infection and Immunity*, 77(7), 2741–2753.
- Cruz, M. R., Graham, C. E., Gagliano, B. C., Lorenz, M. C., & Garsin, D. A. (2013). *Enterococcus faecalis* inhibits hyphal morphogenesis and virulence of *Candida albicans*. *Infection and Immunity*, 81(1), 189–200.
- Dahan, D., Jude, B. A., Lamendella, R., Keesing, F., & Perron, G. G. (2018). Exposure to arsenic alters the microbiome of larval zebrafish. *Frontiers in Microbiology*, 20(159).
- Darby, C. (2005). Interactions with microbial pathogens. *WormBook*, ed. The *C. elegans* Research Community.

- Dierking, K. and Pita, L. (2020). Receptors mediating host-microbiota communication in the metaorganism: the invertebrate perspective, *Frontiers in Immunology*, *11*(1251).
- Dierking, K., Yang, W., & Schulenburg, H. (2016). Antimicrobial effectors in the nematode *Caenorhabditis elegans*: an outgroup to the Arthropoda. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1695), 20150299.
- Dirksen, P., Assié, A., Zimmermann, J., Zhang, F., Tietje, A.-M., Marsh, S. A., Félix, M.-A., Shapira, M., Kaleta, C., Schulenburg, H., & Samuel, B. S. (2020). CeMbio - The *Caenorhabditis elegans* microbiome resource. *G3: Genes, Genomes, Genetics*, *10*(8).
- Dirksen, P., Marsh, S. A., Braker, I., Heitland, N., Wagner, S., Nakad, R., Mader, S., Petersen, C., Kowallik, V., Rosenstiel, P., Félix, M.-A., & Schulenburg, H. (2016). The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biology*, *14*(38).
- Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, *17*, 764–775.
- Ezcurra, M. (2018). Dissecting cause and effect in host-microbiome interactions using the combined worm-bug model system. *Biogerontology*, *19*, 1–12.
- Félix, M.-A., Ashe, A., Piffaretti, J., Wu, G., Nuez, I., Bélicard, T., Jiang, Y., Zhao, G., Franz, C. J., Goldstein, L. D., Sanroman, M., Miska, E. A., & Wang, D. (2011). Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to Nodaviruses. *PLOS Biology*, *9*(1), e1000586.
- Félix, M.-A., & Braendle, C. (2010). The natural history of *Caenorhabditis elegans*. *Current Biology*, *20*(22), 965–969.
- Ford, S. A., & King, K. C. (2016). Harnessing the power of defensive microbes: Evolutionary implications in nature and disease control. *PLoS Pathogens*, *12*(4), e1005465.
- Fuentes, S., Van Nood, E., Tims, S., Heikamp-De Jong, I., Ter Braak, C. J. F., Keller, J. J., Zoetendal, E. G., & De Vos, W. M. (2014). Reset of a critically disturbed microbial ecosystem: Faecal transplant in recurrent *Clostridium difficile* infection. *The ISME Journal*, *8*(8), 1621–1633.
- Gerardo, N. M., & Parker, B. J. (2014). Mechanisms of symbiont-conferred protection against natural enemies: An ecological and evolutionary framework. *Current Opinion in Insect Science*, *4*, 8-14.
- Goloshchapov, O. V., Olekhovich, E. I., Sidorenko, S. V., Moiseev, I. S., Kucher, M. A., Fedorov, D. E., Pavlenko, A. V., Manolov, A. I., Gostev, V. V., Veselovsky, V. A., Klimina, K. M., Kostyukova, E. S., Bakin, E. A., Shvetcov, A. N., Gumbatova, E. D., Klementeva, R. V., Shcherbakov, A. A., Gorchakova, M. V., Egozcue, J. J., ... Afanasyev, B. V. (2019). Long-term impact of fecal transplantation in healthy volunteers. *BMC Microbiology*, *19*(1), 312.
- Griffitts, J. S., & Aroian, R. V. (2005). Many roads to resistance: How invertebrates adapt to Bt toxins. *BioEssays*, *27*(6), 614–624.
- Griffitts, J. S., Haslam, S. M., Yang, T., Garczynski, S. F., Mulloy, B., Morris, H., Cremer, P. S., Dell, A., Adang, M. J., & Aroian, R. V. (2005). Glycolipids as receptors for *Bacillus thuringiensis* crystal toxin. *Science*, *307*(5711), 922–925.
- Haller, D. (Ed.) (2018). *The Gut Microbiome in Health and Disease*. Cham, Switzerland: Springer International Publishing.
- Hasshoff, M., Böhnisch, C., Tonn, D., Hasert, B., & Schulenburg, H. (2007). The role of *Caenorhabditis elegans* insulin-like signaling in the behavioral avoidance of pathogenic *Bacillus thuringiensis*. *The FASEB Journal*, *21*(8), 1801–1812.

- Hinzke, T., Kleiner, M., Breusing, C., Felbeck, H., Häsler, R., Sievert, S. M., Schlüter, R., Rosenstiel, P., Reusch, T. B. H., Schweder, T., & Markert, S. (2019). Host-microbe interactions in the chemosynthetic riftia pachyptila symbiosis. *MBio*, *10*(6).
- Hoeckendorf, A., & Leippe, M. (2012). SPP-3, a saposin-like protein of *Caenorhabditis elegans*, displays antimicrobial and pore-forming activity and is located in the intestine and in one head neuron. *Developmental & Comparative Immunology*, *38*(1), 181–186.
- Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., & Ezekowitz, R. A. B. (1999). Phylogenetic perspectives in innate immunity. *Science*, *284*(5418), 1313–1318.
- Hollensteiner, J., Poehlein, A., Spröer, C., Bunk, B., Sheppard, A. E., Rosenstiel, P., Schulenburg, H., & Liesegang, H. (2017). Complete genome sequence of the nematocidal *Bacillus thuringiensis* MYBT18247. *Journal of Biotechnology*, *260*, 48–52.
- Horcajada, J. P., Montero, M., Oliver, A., Sorlí, L., Luque, S., Gómez-Zorrilla, S., Benito, N., & Grau, S. (2019). Epidemiology and treatment of multidrug-resistant and extensively drug-resistant *Pseudomonas aeruginosa* infections. *Clinical Microbiology Reviews*, *32*(4).
- Huffman, D. L., Abrami, L., Sasik, R., Corbeil, J., van der Goot, F. G., & Aroian, R. V. (2004). Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins. *Proceedings of the National Academy of Sciences*, *101*(30), 10995–11000.
- Iatsenko, I., Yim, J. J., & Schroeder, F. C. (2014). *B. subtilis* GS67 protects *C. elegans* from Gram-positive pathogens via fengycin-mediated microbial antagonism. *Current Biology*, *24*(22), 2720–2727.
- Ideo, H., Fukushima, K., Gengyo-Ando, K., Mitani, S., Dejima, K., Nomura, K., & Yamashita, K. (2009). A *Caenorhabditis elegans* glycolipid-binding galectin functions in host defense against bacterial infection. *The Journal of Biological Chemistry*, *284*(39), 26493–26501.
- Johnke, J., Dirksen, P., & Schulenburg, H. (2020). Community assembly of the native *C. elegans* microbiome is influenced by time, substrate and individual bacterial taxa. *Environmental Microbiology*, *22*(4), 1265–1279.
- Kao, C.-Y., Los, F. C. O., Huffman, D. L., Wachi, S., Kloft, N., Husmann, M., Karabrahimi, V., Schwartz, J.-L., Bellier, A., Ha, C., Sagong, Y., Fan, H., Ghosh, P., Hsieh, M., Hsu, C.-S., Chen, L., & Aroian, R. V. (2011). Global functional analyses of cellular responses to pore-forming toxins. *PLOS Pathogens*, *7*(3), e1001314.
- Kennedy, E. A., King, K. Y., & Baldridge, M. T. (2018). Mouse microbiota models: Comparing germ-free mice and antibiotics treatment as tools for modifying gut bacteria. *Frontiers in Physiology*, *9*(1534).
- Khoruts, A. (2018). Targeting the microbiome: From probiotics to fecal microbiota transplantation. *Genome Medicine*, *10*(80).
- Kim D.H. and Ewbank J.J. Signaling in the innate immune response (2018). *WormBook*, ed. The *C. elegans* Research Community, WormBook.
- Kim, Y., & Mylonakis, E. (2012). *Caenorhabditis elegans* immune conditioning with the probiotic bacterium *Lactobacillus acidophilus* strain NCFM enhances Gram-positive immune responses. *Infection and Immunity*, *80*(7), 2500–2508.
- Kimbrell, D. A., & Beutler, B. (2001). The evolution and genetics of innate immunity. *Nature Reviews Genetics*, *2*(4), 256–267.
- King, K. C., Brockhurst, M. A., Vasieva, O., Paterson, S., Betts, A., Ford, S. A., Frost, C. L., Horsburgh, M. J., Haldenby, S., & Hurst, G. D. (2016). Rapid evolution of microbe-mediated protection against pathogens in a worm host. *The ISME Journal*, *10*(8), 1915–1924.



- Kirienko, N. V., Ausubel, F. M., & Ruvkun, G. (2015). Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proceedings of the National Academy of Sciences*, *96*(2), 715–720.
- Kirienko N.V., Cezairliyan B.O., Ausubel F.M., Powell J.R. (2014) *Pseudomonas aeruginosa* PA14 pathogenesis in *Caenorhabditis elegans*. In Filloux A. & Ramos JL. (Eds.), *Pseudomonas Methods and Protocols. Methods in Molecular Biology (Methods and Protocols)*, 1149, 653-669. Humana Press, New York, NY.
- Kirienko, N. V, Kirienko, D. R., Larkins-Ford, J., Wählby, C., Ruvkun, G., & Ausubel, F. M. (2013). *Pseudomonas aeruginosa* disrupts *Caenorhabditis elegans* iron homeostasis, causing a hypoxic response and death. *Cell Host & Microbe*, *13*(4), 406–416.
- Kissoyan, K. A. B., Drechsler, M., Stange, E.-L., Zimmermann, J., Kaleta, C., Bode, H. B., & Dierking, K. (2019). Natural *C. elegans* microbiota protects against infection via production of a cyclic lipopeptide of the viscosin group. *Current Biology*, *29*(6), 1030–1037.
- Klockgether, J., & Tümmler, B. (2017). Recent advances in understanding *Pseudomonas aeruginosa* as a pathogen [version1, referees: 3 approved]. *F1000Research*, *6*(1261).
- Koch, E. J., & McFall-Ngai, M. (2018). Model systems for the study of how symbiotic associations between animals and extracellular bacterial partners are established and maintained. *Drug Discovery Today: Disease Models*, *28*, 3–12.
- Krediet, C. J., Ritchie, K. B., Alagely, A., & Teplitski, M. (2013a). Members of native coral microbiota inhibit glycosidases and thwart colonization of coral mucus by an opportunistic pathogen. *ISME Journal*, *7*(5), 980–990.
- Krediet, C. J., Ritchie, K. B., Paul, V. J., & Teplitski, M. (2013b). Coral-associated microorganisms and their roles in promoting coral health and thwarting diseases. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1755).
- Lawley, T. D., Clare, S., Walker, A. W., Goulding, D., Stabler, R. A., Croucher, N., Mastroeni, P., Scott, P., Raisen, C., Mottram, L., Fairweather, N. F., Wren, B. W., Parkhill, J., & Dougan, G. (2009). Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infection and Immunity*, *77*(9), 3661–3669.
- Lederberg, J., & McCray, A. (2001). `Ome Sweet `Omics--A Genealogical Treasury of Words. *The Scientist*, *15*(7), 8.
- Mahajan-Miklos, S., Tan, M.-W., Rahme, L. G., & Ausubel, F. M. (1999). Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa*–*Caenorhabditis elegans* pathogenesis model. *Cell*, *96*(1), 47–56.
- Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, *3*(1), 31.
- Marroquin, L. D., Elyassnia, D., Griffiths, J. S., Feitelson, J. S., & Aroian, R. V. (2000). *Bacillus thuringiensis* (Bt) toxin susceptibility and isolation of resistance mutants in the nematode *Caenorhabditis elegans*. *Genetics*, *155*(4), 1693–1699.
- Masri, L., Branca, A., Sheppard, A. E., Papkou, A., Laehnemann, D., Guenther, P. S., Prahl, S., Saebelfeld, M., Hollensteiner, J., Liesegang, H., Brzuszkiewicz, E., Daniel, R., Michiels, N. K., Schulte, R. D., Kurtz, J., Rosenstiel, P., Telschow, A., Bornberg-Bauer, E., & Schulenburg, H. (2015). *Bacillus thuringiensis* virulence and its cry toxin genes. *PLOS Biology*, *13*(6).
- Masson, F., & Lemaitre, B. (2017). Protection from within. *ELife*, *6*, e24111.
- Matzinger, P. (1994). Tolerance, danger, and the extended family. *Annual Review Immunology*, *12*, 991–1045.

- Mazmanian, S. K., Cui, H. L., Tzianabos, A. O., & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, *122*(1), 107–118.
- Mcfall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V, Domazet-Lo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Neelson, K., Pierce, N. E., ... Kremer, N. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3229–3236.
- Mikkelsen, H., Sivaneson, M., & Filloux, A. (2011). Key two-component regulatory systems that control biofilm formation in *Pseudomonas aeruginosa*. *Environmental Microbiology*, *13*(7), 1666–1681.
- Montalvo-Katz, S., Huang, H., Appel, M. D., Berg, M., & Shapira, M. (2013). Association with soil bacteria enhances p38-dependent infection resistance in *Caenorhabditis elegans*. *Infection and Immunity*, *81*(2), 514–520.
- Nakad, R. et al. (2016). Contrasting invertebrate immune defense behaviors caused by a single gene, the *Caenorhabditis elegans* neuropeptide receptor gene npr-1. *BMC Genomics*, *17*(280), 1–20.
- Nardi RM, Silva ME, Vieira EC, Bambirra EA, N. J. (1989). Intra-gastric infection of germfree and conventional mice with *Salmonella typhimurium*. *Brazilian Journal of Medical and Biological Research*, *22*(11), 1389–1392.
- O'Donnell, M. P., Fox, B. W., Chao, P.-H., Schroeder, F. C., & Sengupta, P. (2020). A neurotransmitter produced by gut bacteria modulates host sensory behaviour. *Nature*, *583*, 415–420.
- Osman, G. A., Fasseas, M. K., Koneru, S. L., Essmann, C. L., Kyrou, K., Srinivasan, M. A., Zhang, G., Sarkies, P., Félix, M. A., & Barkoulas, M. (2018). Natural infection of *C. elegans* by an oomycete reveals a new pathogen-specific immune response. *Current Biology*, *28*(4), 142–143.
- Øyri, S. F., Muzes, G., & Sipos, F. (2015). Dysbiotic gut microbiome: A key element of Crohn's disease. *Comparative Immunology, Microbiology and Infectious Diseases*, *43*, 36–49.
- Parker, M. W., & Feil, S. C. (2005). Pore-forming protein toxins: from structure to function. *Progress in Biophysics and Molecular Biology*, *88*(1), 91–142.
- Petersen, C., Saebelfeld, M., Barbosa, C., Pees, B., Hermann, R. J., Schalkowski, R., Strathmann, E. A., Dirksen, P., & Schulenburg, H. (2015). Ten years of life in compost: temporal and spatial variation of North German *Caenorhabditis elegans* populations. *Ecology and Evolution*, *5*(16), 3250–3263.
- Pickard, J. M., Zeng, M. Y., Caruso, R., & Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunological Reviews*, *279*(1), 70–89.
- Rangan, K. J., Pedicord, V. A., Wang, Y. C., Kim, B., Lu, Y., Shaham, S., Mucida, D., & Hang, H. C. (2016). A secreted bacterial peptidoglycan hydrolase enhances tolerance to enteric pathogens. *Science*, *353*(6306), 1434–1437.
- Ripert, G., Racedo, S. M., Elie, A.-M., Jacquot, C., Bressollier, P., & Urdaci, M. C. (2016). Secreted compounds of the probiotic *Bacillus clausii* strain o/c inhibit the cytotoxic effects induced by *Clostridium difficile* and *Bacillus cereus* toxins. *Antimicrobial Agents and Chemotherapy*, *60*(6), 3445–3454.

- Roeder, T., Stanisak, M., Gelhaus, C., Bruchhaus, I., Grötzinger, J., & Leippe, M. (2010). Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. *Developmental and Comparative Immunology*, *34*(2), 203–209.
- Rolig, A. S., Sweeney, E. G., Kaye, L. E., Desantis, M. D., Perkins, A., Banse, A. V., Hamilton, M. K., & Guillemin, K. (2018). A bacterial immunomodulatory protein with lipocalin-like domains facilitates host–bacteria mutualism in larval zebrafish. *ELife*, *7*.
- Round, J. L., & Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12204–12209.
- Sampson, T. R., & Mazmanian, S. K. (2015). Control of brain development, function, and behavior by the microbiome. *Cell Host and Microbe*, *17*(5), 565–576.
- Samuel, B. S., Rowedder, H., Braendle, C., Félix, M. A., & Ruvkun, G. (2016). *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), E3941–E3949.
- Sassone-Corsi, M., & Raffatellu, M. (2015). No vacancy: how beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *The Journal of Immunology*, *194*(9), 4081–4087.
- Schubert, A. M., Sinani, H., & Schloss, P. D. (2015). Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against *Clostridium difficile*. *MBio*, *6*(4).
- Schulenburg, H., & Müller, S. (2004). Natural variation in the response of *Caenorhabditis elegans* towards *Bacillus thuringiensis*. *Parasitology*, *128*(4), 433–443.
- Schulenburg, H., & Félix, M. A. (2017). The natural biotic environment of *Caenorhabditis elegans*. *Genetics*, *206*, 55–86.
- Schulenburg, H., Hoepfner, M. P., Weiner, J., & Bornberg-Bauer, E. (2008). Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. *Immunobiology*, *213*, 237–250.
- Schulte, R. D., Makus, C., Hasert, B., Michiels, N. K., & Schulenburg, H. (2010). Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7359–7364.
- Sheppard, A.E., Poehlein, A., Rosenstiel, P., Liesegang, H., and Schulenburg, H. (2013). Complete Genome Sequence of *Bacillus thuringiensis* Strain 407 Cry-. *Genome Announc*, *1*, e00158–12.
- Smolentseva, O., Gusarov, I., Gautier, L., Shamovsky, I., Defrancesco, A. S., Losick, R., & Nudler, E. (2017). Mechanism of biofilm-mediated stress resistance and lifespan extension in *C. elegans*. *Scientific Reports*, *7*(1).
- Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J., & Spieth, J. (2001). WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Research*, *29*(1), 82–86.
- Tan, M. W., Rahme, L. G., Sternberg, J. A., Tompkins, R. G., & Ausubel, F. M. (1999). *Pseudomonas aeruginosa* killing of *Caenorhabditis elegans* used to identify *P. aeruginosa* virulence factors. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(5), 2408–2413.
- The Human Microbiome Project Consortium, Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*, 207–214.

- Treitz, C., Cassidy, L., Höckendorf, A., Leippe, M., & Tholey, A. (2015). Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematicidal *Bacillus thuringiensis*. *Journal of Proteomics*, *113*, 337–350.
- Troemel 1, E. R., Félix, M.-A., Whiteman, N. K., Barrière, A., & Ausubel, F. M. (2008). Microsporidia are natural intracellular parasites of the nematode *Caenorhabditis elegans*. *PLOS Biology*, *6*(12), e309.
- Tümmler, B. (2019). Emerging therapies against infections with *Pseudomonas aeruginosa* [version 1; peer review: 2 approved]. *F1000Research*, *8*(1371).
- van der Waaij, D., Berghuis-de Vries, J. M., & Lekkerkerk, Lekkerkerk-v. (1971). Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *The Journal of Hygiene*, *69*(3), 405–411.
- Vannier, N., Agler, M. and Hacquard, S. (2019). Microbiota-mediated disease resistance in plants. *PLoS pathogens*, *15*(6), e1007740.
- Ventola, C. L. (2015). The antibiotic resistance crisis: causes and threats. *P & T Journal*, *40*(4), 277–283.
- Vogt, S. L., & Finlay, B. B. (2017). Gut microbiota-mediated protection against diarrheal infections. *Journal of travel medicine*, *24*(1), 39–43.
- Wang, Baohong, Yao, M., Lv, L., Ling, Z., & Li, L. (2017). The human microbiota in health and disease. *Engineering*, *3*, 71-82.
- Wang, Bing, Wang, H., Xiong, J., Zhou, Q., Wu, H., Xia, L., Li, L., & Yu, Z. (2017). A proteomic analysis provides novel insights into the stress responses of *Caenorhabditis elegans* towards nematicidal Cry6A toxin from *Bacillus thuringiensis*. *Scientific Reports*, *7*(1), 14170.
- Wang, J. W., Kuo, C. H., Kuo, F. C., Wang, Y. K., Hsu, W. H., Yu, F. J., Hu, H. M., Hsu, P. I., Wang, J. Y., & Wu, D. C. (2019). Fecal microbiota transplantation: Review and update. *Journal of the Formosan Medical Association*, *118*, 23–31.
- Wang, Jingwen, Wu, Y., Yang, G., & Aksoy, S. (2009). Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 12133–12138.
- Wang, Jun, Nakad, R., & Schulenburg, H. (2012). Activation of the *Caenorhabditis elegans* FOXO family transcription factor DAF-16 by pathogenic *Bacillus thuringiensis*. *Developmental & Comparative Immunology*, *37*(1), 193–201.
- Wei, J.-Z., Hale, K., Carta, L., Platzer, E., Wong, C., Fang, S.-C., & Aroian, R. V. (2003). *Bacillus thuringiensis* crystal proteins that target nematodes. *Proceedings of the National Academy of Sciences*, *100*(5), 2760–2765.
- Willing, B. P., Vacharaksa, A., Croxen, M., Thanachayanont, T., & Finlay, B. B. (2011). Altering host resistance to infections through microbial transplantation. *PLoS ONE*, *6*(10), e26988.
- Willis, A. R., Moore, C., Mazon-Moya, M., Krokowski, S., Lambert, C., Till, R., Mostowy, S., & Sockett, R. E. (2016). Injections of predatory bacteria work alongside host immune cells to treat shigella infection in zebrafish larvae. *Current Biology*, *26*(24), 3343–3351.
- World Health Organization, Food and Agriculture Organization of the United Nations & Joint FAO/WHO Expert Committee on Food Additives. Meeting (72nd : 2010 : Rome, Italy). (2011). Evaluation of certain contaminants in food: seventy-second [72nd] report of the Joint FAO/WHO Expert Committee on Food Additives. World Health Organization. *WHO Technical Report Series*, *959*, 55–64.

- Yang, W., Dierking, K., Esser, D., Tholey, A., Leippe, M., Rosenstiel, P., & Schulenburg, H. (2015). Overlapping and unique signatures in the proteomic and transcriptomic responses of the nematode *Caenorhabditis elegans* toward pathogenic *Bacillus thuringiensis*. *Developmental & Comparative Immunology*, *51*(1), 1–9.
- Yang, W., Petersen, C., Pees, B., Zimmermann, J., Waschina, S., Dirksen, P., Rosenstiel, P., Tholey, A., Leippe, M., Dierking, K., Kaleta, C., & Schulenburg, H. (2019). The inducible response of the nematode *Caenorhabditis elegans* to members of its natural microbiota across development and adult life. *Frontiers in Microbiology*, *10*, 1793.
- Yang, W., Dierking, K. and Schulenburg, H. (2016). WormExp: a web-based application for a *Caenorhabditis elegans* -specific gene expression enrichment analysis. *Bioinformatics*, *32*(6), 943–945
- Yong, E. (2016). *I Contain Multitudes: The Microbes Within Us and a Grander View of Life*. London: Penguin Random House.
- Zárate-Potes, A., Yang, W., Pees, B., Schalkowski, R., Segler, P., Andresen, B., Haase, D., Nakad, R., Rosenstiel, P., Tetreau, G., Colletier, J.-P., Schulenburg, H., & Dierking, K. (2020). The *C. elegans* GATA transcription factor *elt-2* mediates distinct transcriptional responses and opposite infection outcomes towards different *Bacillus thuringiensis* strains. *PLOS Pathogens*, *16*(9), e1008826.
- Zhang, F., Berg, M., Dierking, K., Félix, M.-A., Shapira, M., Samuel, B. S., & Schulenburg, H. (2017). *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, *8*(485).
- Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Research*, *30*(6), 492–506.
- Zhu, J., Cai, X., Harris, T. L., Gooyit, M., Wood, M., Lardy, M., & Janda, K. D. (2015). Disarming *Pseudomonas aeruginosa* virulence factor *lasB* by leveraging a *Caenorhabditis elegans* infection model. *Chemistry and Biology*, *22*(4), 483–491.



## **CHAPTER I**

### **Natural *C. elegans* microbiota protects against infection via the production of a cyclic lipopeptide of the viscosin group**

Kohar Annie B. Kissoyan, Moritz Drechsler, Eva-Lena Stange, Johannes Zimmermann, Christoph Kaleta, Helge Bode, Katja Dierking

**Published manuscript in Current Biology (2019), 29(6), 1030–1037**

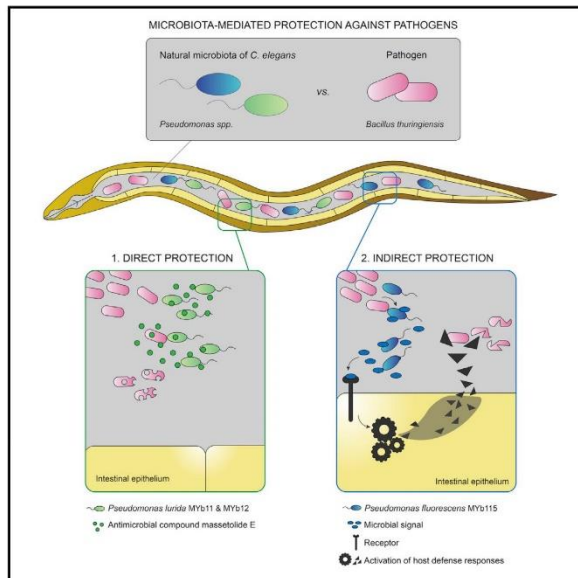




## Current Biology

### Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group

#### Graphical Abstract



#### Authors

Kohar A.B. Kissoyan, Moritz Drechsler, Eva-Lena Stange, Johannes Zimmermann, Christoph Kaleta, Helge B. Bode, Katja Dierking

#### Correspondence

kdierking@zoologie.uni-kiel.de

#### In Brief

Kissoyan et al. provide novel insights into the function of the native microbiota of the model nematode *C. elegans*. Their work highlights the role of microbes in supporting *C. elegans* defense responses and the diversity of immune-protective mechanisms, including the involvement of microbiota-derived metabolites.

#### Highlights

- Natural *C. elegans* microbiota confer protection against pathogen infection
- Different *Pseudomonas* isolates protect *C. elegans* through distinct mechanisms
- *P. lurida* isolates produce massetolide E and directly inhibit pathogen growth
- *P. fluorescens*-mediated protection may depend on indirect, host-mediated mechanisms

Kissoyan et al., 2019, Current Biology 29, 1–8  
March 18, 2019 © 2019 The Authors. Published by Elsevier Ltd.  
<https://doi.org/10.1016/j.cub.2019.01.050>

CellPress

Please cite this article in press as: Kissoyan et al., Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group, *Current Biology* (2019), <https://doi.org/10.1016/j.cub.2019.01.050>

Current Biology  
Report

CellPress

## Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group

Kohar A.B. Kissoyan,<sup>1</sup> Moritz Drechsler,<sup>2</sup> Eva-Lena Stange,<sup>1</sup> Johannes Zimmermann,<sup>4</sup> Christoph Kaleta,<sup>4</sup> Helge B. Bode,<sup>2,3</sup> and Katja Dierking<sup>1,5,\*</sup>

<sup>1</sup>Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany

<sup>2</sup>Fachbereich Biowissenschaften, Goethe Universität Frankfurt, 60438 Frankfurt am Main, Germany

<sup>3</sup>Buchmann Institute for Molecular Life Sciences, Goethe Universität Frankfurt, 60438 Frankfurt am Main, Germany

<sup>4</sup>Research Group Medical Systems Biology, Institute of Experimental Medicine, Christian-Albrechts-Universität zu Kiel, 24105 Kiel, Germany

<sup>5</sup>Lead Contact

\*Correspondence: [kdierking@zoologie.uni-kiel.de](mailto:kdierking@zoologie.uni-kiel.de)

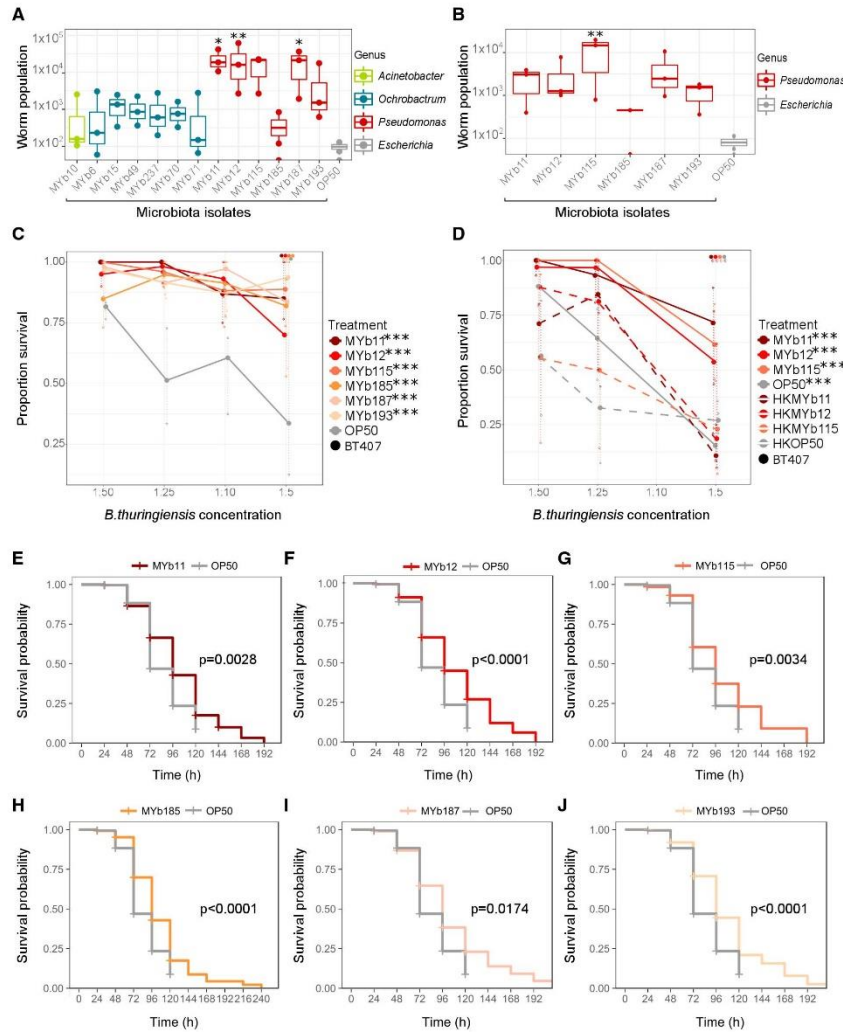
<https://doi.org/10.1016/j.cub.2019.01.050>

### SUMMARY

*Caenorhabditis elegans* is associated in nature with a species-rich, distinct microbiota, which was characterized only recently [1]. Thus, our understanding of the relevance of the microbiota for nematode fitness is still at its infancy. One major benefit that the intestinal microbiota can provide to its host is protection against pathogen infection [2]. However, the specific strains conferring the protection and the underlying mechanisms of microbiota-mediated protection are often unclear [3]. Here, we identify natural *C. elegans* microbiota isolates that increase *C. elegans* resistance to pathogen infection. We show that isolates of the *Pseudomonas fluorescens* subgroup provide paramount protection from infection with the natural pathogen *Bacillus thuringiensis* through distinct mechanisms. We found that the *P. lurida* isolates MYb11 and MYb12 (members of the *P. fluorescens* subgroup) protect *C. elegans* against *B. thuringiensis* infection by directly inhibiting growth of the pathogen both *in vitro* and *in vivo*. Using genomic and biochemical analyses, we further demonstrate that MYb11 and MYb12 produce massetolide E, a cyclic lipopeptide biosurfactant of the viscosin group [4, 5], which is active against pathogenic *B. thuringiensis*. In contrast to MYb11 and MYb12, *P. fluorescens* MYb115-mediated protection involves increased resistance without inhibition of pathogen growth and most likely depends on indirect, host-mediated mechanisms. This work provides new insight into the functional significance of the *C. elegans* natural microbiota and expands our knowledge of bacteria-derived compounds that can influence pathogen colonization in the intestine of an animal.

### RESULTS AND DISCUSSION

Surprisingly little is known about the ecology of the model nematode *Caenorhabditis elegans*. The natural microbiota of *C. elegans* was characterized only recently [1] and its effects on worm fitness remain to be uncovered. Here, we systematically tested which members of the *C. elegans* microbiota protect the worm from infection with its natural pathogen, *Bacillus thuringiensis*. We selected 13 isolates that comprise the most abundant genera of the *C. elegans* natural core microbiome, namely *Pseudomonas*, *Ochrobactrum*, and *Acinetobacter* [6], and that are able to enter and persist in the nematode gut [1]. First, we assessed population growth (as proxy for fitness) of the natural *C. elegans* isolate MY316, which was co-isolated with the microbiota isolates, 5 days after exposure to the nematocidal *B. thuringiensis* strains MYBT18247(BT247) mixed with the respective microbiota isolate or with the laboratory food bacterium *Escherichia coli* OP50 as control. The non-nematocidal *B. thuringiensis* strain BT407 served as non-pathogen control. We found that worms on pathogenic BT247 showed a clear increase in population growth when infected in the presence of five microbiota isolates—all of the genus *Pseudomonas*—compared to infected worms on *E. coli* OP50 (Figure 1A). Moreover, worms on the non-pathogenic BT407 also showed increased population growth in the presence of the *Pseudomonas* isolates (Figure S1A), indicating that these microbiota isolates also represent good growth environments for *C. elegans* under control conditions, confirming previous work [1]. We thus focused on the *Pseudomonas* microbiota isolates, which can be assigned to the following species based on complete 16S rRNA sequences: *P. lurida* MYb11, *P. lurida* MYb12, *P. fluorescens* MYb115, and *P. lurida* MYb193 (all of the *P. fluorescens* group), *P. denitrificans* MYb185 (*P. pertucinogena* group), and *P. alkylphenolia* MYb187 (*P. putida* group). We then asked if they also affect the fitness of the *C. elegans* reference laboratory strain N2. Indeed, BT247 infected N2 worms showed increased population growth in the presence of all six *Pseudomonas* isolates (MYb11, MYb12, MYb115, MYb185, MYb187, and MYb193), albeit statistically significant only for MYb115 (Figure 1B). N2 worms on the



**Figure 1. Natural Microbiota Isolates of the Genus *Pseudomonas* Protect *C. elegans* from Pathogen Infection**

(A and B) Population size measured in the presence of pathogenic *B. thuringiensis* BT247 of (A) the natural *C. elegans* isolate MY316 on one of 13 microbiota isolates and (B) the laboratory N2 reference strain on one of six *Pseudomonas* isolates. The food bacterium *E. coli* OP50 was used as control. Three L4 larvae were picked onto the infection plates. Population size was assessed after 5 days at 20°C. The boxplots indicate the median, range, and upper and lower quartiles of the worm population data ( $n = 3$  independent runs). Statistical analysis was performed using generalized linear model (GLM) [7] framework and the obtained p values were corrected using Bonferroni correction for multiple testing [8].

(C and D) Survival of *C. elegans* N2 on different concentrations of pathogenic BT247 mixed with either (C) one of six living *Pseudomonas* isolates or (D) one of the three *Pseudomonas* isolates—MYb11, MYb12, and MYb115—alive (continuous lines) or heat killed (dashed lines). *E. coli* OP50 was used as control and the *B. thuringiensis* strain BT407 as non-pathogen control. Error bars represent the range of the median of survival proportions of four technical replicates ( $n = 4$ ), representative of at least three independent runs. Statistical analyses were performed using GLM [7] framework and Bonferroni correction for multiple testing [8] with the OP50 control treatment group.

(legend continued on next page)

non-pathogenic BT407 also showed increased median population growth in the presence of four *Pseudomonas* isolates (Figure S1B). This might point to differences in nutritional qualities between the different *Pseudomonas* isolates, which was previously suggested to underlie the ability of different bacteria to affect *C. elegans* physiological processes such as development, growth, aging, and longevity [11].

We considered that increased resistance to BT infection may contribute to the observed increased fitness of infected worms on natural microbiota isolates. We thus challenged *C. elegans* N2 worms on the respective *Pseudomonas* isolates with pathogenic BT247 at different concentrations and assessed survival 24 h post infection (p.i.) as proxy for resistance. We found that worms on all tested *Pseudomonas* isolates showed significantly increased survival when compared to infected worms on *E. coli* OP50 (Figure 1C). Taken together, these results show that *C. elegans* microbiota isolates of the genus *Pseudomonas* provide paramount protection from infection with *B. thuringiensis* MYBT247. In previous work, the *Pseudomonas* isolates MYb11, MYb187, and MYb193 have been reported to inhibit fungal growth *in vitro* [1]. Importantly, MYb11 protected the nematode from infection by the fungal pathogen *Drechmeria coniospora* also *in vivo* through a hitherto unknown mechanism [1]. Our work thus extends our knowledge on the protective effect of MYb11 to bacterial infection in the intestine.

It was previously shown that *P. mendocina*, which is a member of laboratory-established *C. elegans* microbiota, protects the worm against infection with pathogenic *P. aeruginosa* PA14 [12]. While *P. aeruginosa* PA14 is not a natural *C. elegans* pathogen, it is one of the most extensively studied *C. elegans* pathogens [13, 14]. To investigate if the protection conferred by the *Pseudomonas* isolates is specific against infection with the Gram-positive pathogen *B. thuringiensis* or rather is a generic protection against a broader range of pathogens, we assessed survival of worms grown on the respective microbiota isolates or on *E. coli* OP50 from the L1 stage and infected with the Gram-negative *P. aeruginosa* PA14 at the L4 stage. We observed prolonged survival for worms on all six *Pseudomonas* isolates tested (Figures 1E–1J). This indicates that the *Pseudomonas* microbiota isolates confer resistance to PA14 infection, in addition to BT247 infection.

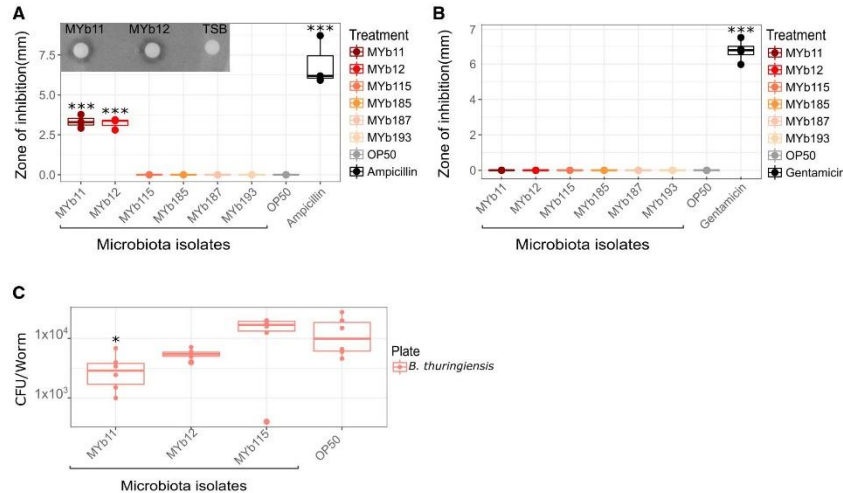
MYb11, MYb12, MYb115, MYb185, MYb187, and MYb193 confer resistance to Gram-positive, as well as Gram-negative, bacterial pathogens. This broader protective effect might be based on activation of general *C. elegans* defense responses. The aforementioned *P. mendocina* protects *C. elegans* against infection with PA14 likely through activation of p38 MAPK signaling [12], an innate immune signaling pathway central for *C. elegans* defense against several different pathogens [13]. However, apart from this one example, the mechanisms by which the *C. elegans* microbiota protects the worm from pathogen infection are completely unknown. Intestinal microbiota can confer protection against intestinal pathogens in two ways:

through indirect activation of the host immune response or through direct microbe-microbe competition [3, 15, 16]. In a first step toward exploring the molecular mechanism underlying the observed microbiota-mediated protection, we tested whether the protective effects require alive bacteria and assessed worm survival after BT247 infection in the presence of both alive and heat-inactivated microbiota isolates MYb11, MYb12, and MYb115, which strongly and consistently protected *C. elegans* from *B. thuringiensis* infection. Worms on heat-inactivated MYb11, MYb12, and MYb115 showed significantly decreased survival compared to worms on alive microbiota isolates (Figure 1D), indicating that bacterial metabolic activity is required for the protective effect.

We then asked whether the protective effect is provided through direct microbe-microbe antagonism—i.e., if the *Pseudomonas* isolates directly inhibit pathogen growth. To answer this question, we first tested the effect of cell-free supernatant of the six *Pseudomonas* isolates (MYb11, MYb12, MYb115, MYb185, MYb187, and MYb193) on pathogen growth in a disc-diffusion assay *in vitro*. While the supernatant of MYb11 and MYb12 strongly inhibited growth of *B. thuringiensis* BT247, supernatant of the other protective *Pseudomonas* isolates did not have an effect on *B. thuringiensis* growth (Figure 2A). In contrast, none of the *Pseudomonas* isolates inhibited growth of *P. aeruginosa* PA14 (Figure 2B). We are currently investigating how *Pseudomonas* microbiota isolates protect the worm from PA14 infection. Here, we focused on MYb11 and MYb12, the only *Pseudomonas* isolates showing antagonistic properties against BT247. When we investigated the range of antagonistic activity of MYb11 and MYb12 supernatant on several Gram-positive and Gram-negative bacteria, we found that growth of all tested *B. thuringiensis* strains and growth of *Staphylococcus aureus* SA113 was strongly inhibited (Figures S2A–S2D). In contrast, MYb11 and MYb12 supernatant did not inhibit growth of the Gram-positive natural isolates *Leucobacter tardus* MYb258 and *Rhodococcus erythropolis* MYb53 or any of the tested Gram-negative bacteria (Figures S2A–S2D). Taken together, our results suggest that the two *P. lurida* isolates MYb11 and MYb12 have antagonistic capacities and directly inhibit the growth of *B. thuringiensis* through the production of bacteriostatic and/or bactericidal compounds. In contrast, the other *Pseudomonas* isolates might protect the nematode from BT247 infection through distinct mechanisms.

To further explore the potentially distinct mechanisms underlying the observed microbiota-mediated protection from BT247 infection, we subsequently focused on the microbiota isolates MYb11, MYb12, and MYb115. All three consistently protected *C. elegans* from BT247 infection, while only MYb11 and MYb12 inhibited *B. thuringiensis* growth *in vitro*. Next, we investigated if MYb11, MYb12, and MYb115 inhibit growth of *B. thuringiensis* *in vivo* and thus infected *C. elegans* with BT247 mixed with MYb11, MYb12, MYb115, or *E. coli* OP50. We assessed pathogen load 24 h p.i. by counting colony-forming

(E–J) Survival of *C. elegans* N2 on pathogenic *P. aeruginosa* PA14 grown on any of the six *Pseudomonas* isolates, (E) MYb11, (F) MYb12, (G) MYb115, (H) MYb185, (I) MYb187, and (J) MYb193. Kaplan-Meier analysis was followed by log-rank test [9] and Bonferroni correction [10]. The graphs are based on pooled data of three independent runs, each with four technical replicates with at least 30 worms per plate and treatment condition. Worms were grown from L1 stage on the subsequent microbiota isolate and infected at the L4 stage on slow-killing PA14 plates. Alive worms were transferred daily to fresh PA14 infection plates. p values are considered significant and denoted with asterisks according to \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001. See also Figure S1.



**Figure 2. *Pseudomonas lurida* MYb11 and MYb12 Have Antagonistic Activity and Inhibit Pathogen Growth *In Vitro* and *In Vivo***

(A and B) Growth inhibition of (A) *B. thuringiensis* BT247 and (B) *P. aeruginosa* PA14 by cell-free supernatant of the six *Pseudomonas* microbiota isolates, *E. coli* OP50, or a respective antibiotic as positive control measured as diameter of inhibition zone (mm) in a disc diffusion assay. The data points represent the mean diameter of three technical replicates of three independent runs ( $n = 3$ ). The boxplots indicate the median, range, and upper and lower quartiles of the data. The diameters are standardized by subtracting the 6 mm diameter of the disc from the overall diameter measured. The inset in (A) is a representative image of the disc diffusion assay showing that supernatant of MYb11 and MYb12 inhibit growth of BT247 in contrast to the TSB medium control. Student's *t* test against the OP50 control, followed by Bonferroni correction, was performed [9].

(C) Colonization of *C. elegans* intestine by BT247 in the presence of the *Pseudomonas* isolates MYb11, MYb12, and MYb115 or the *E. coli* OP50 control. Pathogen load was measured in a CFU assay. L4 larvae were exposed to BT247 mixed with the respective microbiota isolate or *E. coli* OP50 for 24 h. The boxplots indicate the median, range, and upper and lower quartiles of five technical replicates ( $n = 5$ ). These data are representative of at least three independent runs. *p* values are considered significant and denoted with asterisks according to \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ . See also Figure S2.

units (CFUs). We observed that pathogen load was clearly decreased in infected worms on MYb11 and MYb12, but not in worms on MYb115, when compared to worms on *E. coli* OP50 (Figure 2C). These results suggest that MYb11 and MYb12 control pathogen load *in vivo*, while MYb115 protects the worm from pathogenic impact without directly reducing pathogen load.

Worms on the *Pseudomonas* isolates MYb11, MYb12, and MYb115 show clearly increased fitness and resistance to *B. thuringiensis* infection. While MYb11 and MYb12 seem to protect worms through directly inhibiting pathogen growth, worms on MYb115 are protected despite a high pathogen burden. Accordingly, MYb115 might provide protection by limiting pathogen-induced damage to the intestinal epithelium. We thus asked whether the damage caused by *B. thuringiensis* infection is reduced in worms on MYb115, as well as on MYb11 and MYb12. *B. thuringiensis* produces spores that are associated with pore-forming toxins called Cry toxins [17, 18]. The damage caused by these toxins can lead to loss of intestinal barrier function. To assess the severity of the damage caused by infection, we measured *C. elegans* intestinal barrier function using the “smurf” assay [19], in which worms are exposed to blue dye. As long as intestinal barrier function is intact, the blue dye is contained in the intestine of worms. If intestinal integrity is disrupted,

the dye leaks into the body cavity. We observed that the blue dye was indeed contained in the intestine of uninfected worms (Figure 3A, 3C, 3E, and 3G), while it leaked into the body cavity in control worms infected with BT247, thus displaying the smurf phenotype (arrows in Figure 3H). Notably, *B. thuringiensis*-induced disruption of intestinal barrier function was significantly reduced in BT247 infected worms on the *Pseudomonas* isolates MYb11, MYb12, and MYb115 (Figure 3B, 3D, 3F, and 3I). Together, our results indicate that MYb11 and MYb12 protect the host through direct antagonism, while MYb115 helps the host to cope with the infection despite a high pathogen burden. We are currently investigating if the effect of MYb115 is mediated by an indirect mechanism—i.e., through reinforcement of intestinal barriers, increased damage repair, or activation of other defense responses. However, we cannot exclude that MYb115 has a direct effect on the Cry toxins released by the BT spores—e.g., by degradation of the toxins—as was previously shown for a protease secreted by the probiotic *B. clausii* that inhibits the toxic effects of *B. cereus* toxins in a cytotoxicity assay using mammalian cell lines [20].

To identify the putative bacteriostatic and/or bactericidal compounds produced by MYb11 and MYb12, we analyzed extracts from bacterial cultures using ultra-performance liquid chromatography (UPLC) coupled with high-resolution mass



**Figure 3. *B. thuringiensis*-Induced Disruption of Intestinal Barrier Integrity Is Reduced in Worms on *Pseudomonas* Isolates MYb11, MYb12, and MYb115**

(A–H) Representative light-microscopy pictures of non-infected (panels on left side) and BT247-infected (panels on right side) worms on (A and B) MYb11, (C and D) MYb12, (E and F) MYb115, and (G and H) *E. coli* OP50. The blue dye is contained in the intestinal lumen of non-infected worms and worms on MYb11, MYb12, and MYb115 while it leaks into the body cavity of infected control worms (black arrows in H), showing the smurf phenotype.

(I) Quantification of the smurf phenotype. The percentage of smurf worms was calculated per treatment with at least 30 worms ( $n = 4$  replicates). The boxplots indicate the median, range, and upper and lower quartiles of the data. Student's pairwise *t* test and Bonferroni-based adjustment were used to correct for multiple testing with the OP50 control treatment group [8]. These data are representative of two independent runs.

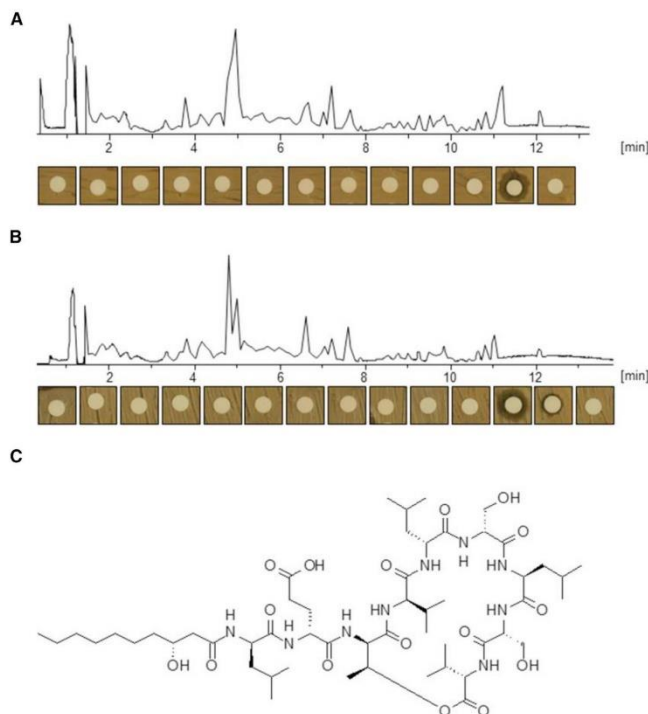
*p* values are considered significant and denoted with asterisks according to: \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ .

spectrometry (HRMS). In parallel, we used antiSMASH to search for known biosynthetic gene clusters (BGCs) [21] in the genomes of MYb11 and MYb12. Both methods lead to the discovery of viscosin ( $m/z$  1126.68 [M+H]<sup>+</sup>; RT 12.1 min) and the viscosin BGC [22], respectively. To confirm that this compound is indeed active against the pathogens BT247 and BT679, we fractionated the extracts of MYb11 and MYb12 using preparative high-performance liquid chromatography (HPLC) and tested each fraction against both pathogens (Figure 4). We could show that the active compound against the two pathogens was not viscosin but instead massetolide E ( $m/z$  1112.66 [M+H]<sup>+</sup>; RT 11.9 min) (Figures 4 and S3). Massetolide E is a viscosin derivative that contains a valine residue instead of the Leu/Ile residue (Figure S4).

Massetolide E and viscosin are cyclic lipopeptide biosurfactants of the viscosin group synthesized by nonribosomal peptide synthetases (NRPSs) in *Pseudomonas* spp. isolated from soil- and plant-associated environments, as well as marine habitats [4, 5]. Massetolides and viscosin lipopeptides are known to have both antifungal and antibacterial activities, preferentially against Gram-positive bacteria [4, 5, 23, 24], which is in line with the protective effect of MYb11 and MYb12 against fungi [1] and the Gram-positive *B. thuringiensis* (this study). Interestingly, *B. subtilis*, which is not a member of the *C. elegans* natural microbiota but has been used to study host-microbe interactions in the worm, was reported to protect the worm from Gram-positive pathogens through fengycin-mediated microbial antagonism [25]. Fengycins are also cyclic lipopeptides that are produced by *Bacillus* spp [5] and were recently shown to inhibit *S. aureus* quorum sensing, which in turn was shown to be crucial for intestinal colonization by *S. aureus* in mice [26]. Interestingly, while the fengycin-mediated inhibitory effect was clearly linked to interference with quorum sensing, the *Bacillus* isolates investigated by Piewngam et al. only show a minor growth inhibitory effect on *S. aureus* [26]. This is in contrast to the strong growth inhibitory effect we observed for MYb11 and MYb12 supernatant on *S. aureus* (Figures S2A, S2B, and S2D) and might suggest that the mechanisms of bacterial growth inhibition between fengycin and massetolide E are distinct. These and our findings thus emphasize the relevance of bacterial lipopeptide production for competitive interactions with coexisting bacteria. Furthermore, the findings reported here broaden our understanding of bacteria-derived compounds that might directly influence bacterial pathogen colonization in the animal gut. The mechanism of massetolide E-mediated pathogen growth inhibition has yet to be defined.

Our results indicate that direct microbial antagonism is important for the protective effect mediated by MYb11 and MYb12 observed in the context of BT247 infection. But interestingly, Tran et al. provide evidence of a cyclic lipopeptide of the viscosin group (massetolide A) stimulating defense responses in plants, thus potentially indirectly activating the host immune response: when the roots of the tomato plant were treated with massetolide A, the leaves showed enhanced resistance to infection with the oomycete *P. infestans* [27]. As massetolide E produced by MYb11 and MYb12 seems to have no activity against Gram-negative bacteria, yet MYb11 and MYb12 also protected worms from infection by PA14, protection against PA14 infection might be mediated through indirect mechanisms. Whether

Please cite this article in press as: Kissoyan et al., Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group, *Current Biology* (2019), <https://doi.org/10.1016/j.cub.2019.01.050>



**Figure 4. Identification of Massetolide E as Inhibitory Compound in Extracts from *Pseudomonas* Isolates MYb11 and MYb12**  
(A and B) Base peak chromatograms of extracts from (A) MYb11 and (B) MYb12 with images of the disc diffusion assay showing the bioactivity of 0.8-min fractions correlating to the respective portion of the chromatogram against *B. thuringiensis* BT679 (BT247 looked identical).  
(C) Structure of the antibiotic depsipeptide massetolide E. The stereochemistry of massetolide E is derived from domain organization of the NRPS and known massetolide structures.  
See also Figures S3 and S4.

MYb11 and MYb12 can also protect *C. elegans* from pathogen infection through an indirect mechanism and if massetolide E contributes to this effect remains to be determined.

### Conclusion

We identify a defined function for members of the native microbiota of the model nematode *C. elegans*: our work suggests that three microbiota isolates of the *P. fluorescens* species complex protect the worm against infection with the natural *C. elegans* pathogen *B. thuringiensis* through distinct mechanisms. The two isolates MYb11 and MYb12 directly inhibit pathogen growth, likely through the production of massetolide E, a cyclic lipopeptide of the viscosin group. In contrast, the isolate MYb115 limits pathogen-induced damage without inhibiting pathogen growth, likely through an indirect, host-defense-dependent mechanism. These findings provide novel insights into the function of the native microbiota of the model nematode *C. elegans*, highlighting the role of microbes in supporting the worm immune response. Moreover, the identification of massetolide E broadens our understanding of microbiota-derived compounds that might directly influence pathogen colonization in the animal gut. Thus, our study strengthens the potential of *C. elegans* as a model for the identification of immune-protective mechanisms, which is key for the development of microbiota-based therapeutic strategies.

6 *Current Biology* 29, 1–8, March 18, 2019

### STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Nematode strains and maintenance
  - Bacterial strains and maintenance
- METHOD DETAILS
  - Population growth assay
  - Survival assay
  - Pathogen load assay
  - Disc diffusion assay
  - Intestinal barrier function assay
  - Genomic analyses
  - UPLC- high resolution mass spectrometry
  - Preparative high-performance liquid chromatography
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at <https://doi.org/10.1016/j.cub.2019.01.050>.

**ACKNOWLEDGMENTS**

We thank Hinrich Schulenburg for generous intellectual input, advice, and support; Barbara Pees and Alejandra Zárate-Potes for helpful discussions; the Schulenburg group for valuable feedback; and the *Caenorhabditis* Genetics Center (University of Minnesota, Minneapolis, Minnesota, USA), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We also thank Susanne Landis (<https://www.scienstratation.com/>) for creating the graphical abstract. This work was funded by the German Science Foundation DFG (Collaborative Research Center CRC1182 Origin and Function of Metaorganisms, Project A1.2 to K.D. and Project INF to C.K.), Germany. Work in the Bode lab was supported by the LOEWE Center Translational Biodiversity Genomics (TBG) funded by the state of Hesse, Germany.

**AUTHOR CONTRIBUTIONS**

K.D. conceived the study and secured funding. K.A.B.K. and K.D. designed the experiments, analyzed the data, and wrote the manuscript. M.D. and H.B.B. did the UPLC-HRMS, HPLC, and genomic analyses. J.Z. and C.K. analyzed the genomic data. E.-L.S. and K.A.B.K. conducted the PA14 experiments. K.A.B.K. conducted all other experiments.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: November 20, 2018

Revised: December 21, 2018

Accepted: January 18, 2019

Published: February 28, 2019

**REFERENCES**

- Dirksen, P., Marsh, S.A., Braker, I., Heitland, N., Wagner, S., Nakad, R., Mader, S., Petersen, C., Kowalik, V., Rosenstiel, P., et al. (2016). The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biol.* *14*, 38.
- Round, J.L., and Mazmanian, S.K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* *9*, 313–323.
- Ubeda, C., Djukovic, A., and Isaac, S. (2017). Roles of the intestinal microbiota in pathogen protection. *Clin. Transl. Immunology* *6*, e128.
- Geudens, N., and Martins, J.C. (2018). Cyclic lipopeptides from *Pseudomonas* spp. – biological swiss-army knives. *Front. Microbiol.* *9*, 1867.
- Raaijmakers, J.M., De Bruijn, I., Nybroe, O., and Ongena, M. (2010). Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol. Rev.* *34*, 1037–1062.
- Zhang, F., Berg, M., Dierking, K., Félix, M.A., Shapira, M., Samuel, B.S., and Schulenburg, H. (2017). *Caenorhabditis elegans* as a model for microbiome research. *Front. Microbiol.* *8*, 485.
- Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A* *135*, 370–384.
- Dunn, O.J. (1961). Multiple comparisons among means. *J. Am. Stat. Assoc.* *56*, 52–64.
- Kaplan, E.L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* *53*, 457–481.
- Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Am. Stat. Assoc.* *50*, 1096–1121.
- Kim, D.H. (2013). Bacteria and the aging and longevity of *Caenorhabditis elegans*. *Annu. Rev. Genet.* *47*, 233–246.
- Montalvo-Katz, S., Huang, H., Appel, M.D., Berg, M., and Shapira, M. (2013). Association with soil bacteria enhances p38-dependent infection resistance in *Caenorhabditis elegans*. *Infect. Immun.* *81*, 514–520.
- Kim, D.H., and Ewbank, J.J. (2018). Signaling in the innate immune response. *Wormbook 2018*, 1–35.
- Tan, M.W., Rahme, L.G., Sternberg, J.A., Tompkins, R.G., and Ausubel, F.M. (1999). *Pseudomonas aeruginosa* killing of *Caenorhabditis elegans* used to identify *P. aeruginosa* virulence factors. *Proc. Natl. Acad. Sci. USA* *96*, 2408–2413.
- Chiu, L., Bazin, T., Truchetet, M.-E., Schaeveerbeke, T., Delhaes, L., and Pradeu, T. (2017). Protective microbiota: from localized to long-reaching co-immunity. *Front. Immunol.* *8*, 1678.
- Ford, S.A., and King, K.C. (2016). Harnessing the power of defensive microbes: evolutionary implications in nature and disease control. *PLoS Pathog.* *12*, e1005465.
- Hollensteiner, J., Poehlein, A., Spröer, C., Bunk, B., Sheppard, A.E., Rosenstiel, P., Schulenburg, H., and Liesegang, H. (2017). Complete genome sequence of the nematocidal *Bacillus thuringiensis* MYBT18247. *J. Biotechnol.* *260*, 48–52.
- Wang, J., Nakad, R., and Schulenburg, H. (2012). Activation of the *Caenorhabditis elegans* FOXO family transcription factor DAF-16 by pathogenic *Bacillus thuringiensis*. *Dev. Comp. Immunol.* *37*, 193–201.
- Gelino, S., Chang, J.T., Kumsta, C., She, X., Davis, A., Nguyen, C., Panowski, S., and Hansen, M. (2016). Correction: Intestinal Autophagy Improves Healthspan and Longevity in *C. elegans* During Dietary Restriction. *PLoS Genet.* *12*, e1006271.
- Ripert, G., Racedo, S.M., Elie, A.-M., Jacquot, C., Bressolier, P., and Urdaci, M.C. (2016). Secreted compounds of the probiotic *Bacillus clausii* strain o/c inhibit the cytotoxic effects induced by *Clostridium difficile* and *Bacillus cereus* toxins. *Antimicrob. Agents Chemother.* *60*, 3445–3454.
- Weber, T., Charusanti, P., Musiol-Kroll, E.M., Jiang, X., Tong, Y., Kim, H.U., and Lee, S.Y. (2015). Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. *Trends Biotechnol.* *33*, 15–26.
- de Bruijn, I., de Kock, M.J., Yang, M., de Waard, P., van Beek, T.A., and Raaijmakers, J.M. (2007). Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species. *Mol. Microbiol.* *63*, 417–428.
- Gerard, J., Lloyd, R., Barsby, T., Haden, P., Kelly, M.T., and Andersen, R.J. (1997). Massetolides A-H, antimycobacterial cyclic decapeptides produced by two *pseudomonads* isolated from marine habitats. *J. Nat. Prod.* *60*, 223–229.
- Reder-Christ, K., Schmidt, Y., Dörr, M., Sahl, H.-G., Josten, M., Raaijmakers, J.M., Gross, H., and Bendas, G. (2012). Model membrane studies for characterization of different antibiotic activities of lipopeptides from *Pseudomonas*. *Biochim. Biophys. Acta* *1818*, 566–573.
- Iatsenko, I., Yim, J.J., Schroeder, F.C., and Sommer, R.J. (2014). *B. subtilis* GS67 protects *C. elegans* from Gram-positive pathogens via fengycin-mediated microbial antagonism. *Curr. Biol.* *24*, 2720–2727.
- Piewngam, P., Zheng, Y., Nguyen, T.H., Dickey, S.W., Joo, H.-S., Villaruz, A.E., Glose, K.A., Fisher, E.L., Hunt, R.L., Li, B., et al. (2018). Pathogen elimination by probiotic *Bacillus* via signalling interference. *Nature* *562*, 532–537.
- Tran, H., Ficke, A., Asiwewe, T., Höfte, M., and Raaijmakers, J.M. (2007). Role of the cyclic lipopeptide massetolide A in biological control of *Phytophthora infestans* and in colonization of tomato plants by *Pseudomonas fluorescens*. *New Phytol.* *175*, 731–742.
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Brucoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohleben, W., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* *43* (W1), W237–43.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* *77*, 71–94.
- Borgonie, G., Van Driessche, R., Leyns, F., Arnaet, G., De Waele, D., and Coomans, A. (1995). Germination of *Bacillus thuringiensis* spores in



Please cite this article in press as: Kissoyan et al., Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group, *Current Biology* (2019), <https://doi.org/10.1016/j.cub.2019.01.050>

- bacteriophagous nematodes (Nematoda: Rhabditida). *J. Invertebr. Pathol.* 65, 61–67.
31. Nakad, R., Snoek, L.B., Yang, W., Ellendt, S., Schneider, F., Mohr, T.G., Rösingh, L., Masche, A.C., Rosenstiel, P.C., Dierking, K., et al. (2016). Contrasting invertebrate immune defense behaviors caused by a single gene, the *Caenorhabditis elegans* neuropeptide receptor gene *npr-1*. *BMC Genomics* 17, 280.
32. Kirienco, N.V., Cezairliyan, B.O., Ausubel, F.M., and Powell, J.R. (2014). *Pseudomonas aeruginosa* PA14 pathogenesis in *Caenorhabditis elegans*. *Methods Mol. Biol.* 1149, 653–669.
33. Tobias, N.J., Wolff, H., Djahanschiri, B., Grundmann, F., Kronenwerth, M., Shi, Y.M., Simonyi, S., Grün, P., Shapiro-Ilan, D., Pidot, S.J., et al. (2017). Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat. Microbiol.* 2, 1676–1685.

Please cite this article in press as: Kissoyan et al., Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group, *Current Biology* (2019), <https://doi.org/10.1016/j.cub.2019.01.050>

CellPress

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and Virus Strains</b>		
<i>Escherichia coli</i> (OP50)	Caenorhabditis Genetics Center	N/A
<i>Bacillus thuringiensis</i> MYBT18247 (BT247)	Agriculture Research Service Patent Culture Collection of Microorganisms and Cell Cultures-NRRL (United States Department of Agriculture, Peoria, Illinois, USA). Isolated initially from NRRL-BT18247.	N/A
<i>Bacillus thuringiensis</i> BT407	Christina Nielsen-LeRoux (INRA-France).	N/A
<i>Bacillus thuringiensis</i> MY18679 (BT679)	Agriculture Research Service Patent Culture Collection of Microorganisms and Cell Cultures -NRRL (United States Department of Agriculture, Peoria, Illinois, USA). Isolated initially from NRRL-BT18679.	N/A
<i>Pseudomonas aeruginosa</i> PA14	Jonathan Ewbank	N/A
MYb6 ( <i>Ochrobactrum</i> sp. BS30)	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
MYb10 ( <i>Acinetobacter</i> sp. 'LB BR 12338')	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
MYb15 ( <i>Ochrobactrum</i> sp. BS30)	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
MYb49 ( <i>Ochrobactrum</i> sp. LC497)	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
MYb70 ( <i>Ochrobactrum pseudogrignonense</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
MYb237 ( <i>Ochrobactrum</i> sp.)	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
MYb71 ( <i>Ochrobactrum</i> sp. R-26465)	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
MYb11 ( <i>Pseudomonas lurida</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
MYb12 ( <i>Pseudomonas lurida</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
MYb115 ( <i>Pseudomonas fluorescens</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
MYb185 ( <i>Pseudomonas denitrificans</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
MYb187 ( <i>Pseudomonas alkylphenolia</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
MYb193 ( <i>Pseudomonas lurida</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany; this paper.	N/A
<i>Staphylococcus aureus</i> (SA113)	Andreas Peschel	N/A
MYb258 ( <i>Leukobacter tardus</i> )	Natural isolate from Kiel. Isolated by Julia Johnke; this paper.	N/A
MYb53 ( <i>Rhodococcus erythropolis</i> )	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
<i>Serratia marcescens</i> (Db11)	Jonathan Ewbank	N/A
<i>Pseudomonas fluorescens</i>	Hinrich Schulenburg	N/A
<i>Photobacterium luminescens</i> (Jm12)	Hinrich Schulenburg	N/A
<i>Pseudomonas aeruginosa</i> (PA01)	Wolfgang Streit	N/A
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Tetramisole hydrochloride	Sigma-Aldrich	Lot#SLBN8309V
Triton X-100	Biochemica	Lot#8W002155
Erioglaucine disodium salt	Sigma-Aldrich	Lot#MKBX7318
Zirconia beads (1mm diameter)	BioSpec Products	Cat#11079110
Sodium azide	Roth	Art#K305.1

(Continued on next page)

Please cite this article in press as: Kissoyan et al., Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group, *Current Biology* (2019), <https://doi.org/10.1016/j.cub.2019.01.050>

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
MYb11 genome sequence	NCBI	GenBank: BioSample SAMN07581396
MYb12 genome sequence	NCBI	GenBank: BioSample SAMN07581345
Experimental Models: Organisms/Strains		
<i>C. elegans</i> : Strain wild-type N2	Caenorhabditis Genetics Center	WB Strain: N2
<i>C. elegans</i> : Natural isolate MY316	Natural isolate from Kiel, Schleswig-Holstein, Germany [1];	N/A
Software and Algorithms		
RStudio	GNU	Version 1.0.153
ggplot: Various R Programming Tools for Plotting Data.	R package	Version 1.0.153
SPSS	IBM®SPSS®STATISTICS	Version 24.0
Inkscape	GNU	Version 0.92
ImageJ	NIH image	RRID: SCR_003070
Image Lab	Biorad Laboratories	Version 5.2.1
AntiSMASH	[28]	Version 3.0
Other		
Ampicillin Natriumsalz	Roth	Art#K029.1
Gentamycin sulfat-Lösung	Roth	Art#HN09.2

#### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Katja Dierking ([kdierking@zoologie.uni-kiel.de](mailto:kdierking@zoologie.uni-kiel.de)).

#### EXPERIMENTAL MODEL AND SUBJECT DETAILS

##### Nematode strains and maintenance

###### Maintenance of *C. elegans* strains

*Caenorhabditis elegans* N2 strain and *C. elegans* MY316 strain were grown and maintained on nematode growth media (NGM) seeded with the *Escherichia coli* strain OP50 at 20°C. The NGM-OP50 lawn was prepared by streaking *E. coli* OP50 from a frozen stock onto Tryptic Soy Agar (TSA) plates, followed by an overnight incubation at 37°C. A single fresh colony was then picked, inoculated in Tryptic Soy Broth (TSB), and grown overnight in a shaker-incubator at 37°C. 500 µl of the culture was seeded on NGM plates and stored at 20°C until use [29].

###### Synchronization of *C. elegans* populations

Worm populations were synchronized by bleaching prior to each experiment: Worms were grown on NGM-OP50 plates. Plates with gravid hermaphrodites and eggs were washed with M9 buffer, and bleached using alkaline hypochlorite solution (1 ml of 5M NaOH and 1 ml of NaClO 12% bleach), followed by multiple washes with M9 buffer. L1s were allowed to hatch in M9 buffer overnight at 20°C. Next, L1 larvae were transferred to fresh NGM-OP50 plates, and incubated at 20°C. Fourth instar larval (L4) stage hermaphrodite nematodes were used in all infection assays (described in detail below). One day adult hermaphrodite nematodes were used for the intestinal barrier function assay (described in detail below).

##### Bacterial strains and maintenance

###### *C. elegans* microbiota isolates and *E. coli* OP50

Natural *C. elegans* microbiota isolates from Northern Germany, originating from the *C. elegans* wild isolate MY316 or its substrate, were used [1]. The natural microbiota isolates and the *E. coli* OP50 control were grown on TSA plates at 25°C and cultured in TSB in a shaker-incubator overnight at 28°C.

###### *Bacillus thuringiensis* strains

Spore-enriched cultures of *Bacillus thuringiensis* strains (MYBT247, MYBT679, and MYBT407) were obtained following a protocol established by Borgonie et al. [30] with some modifications as described in detail in the following: *B. thuringiensis* (BT) strains were thawed from frozen stocks and streaked onto Luria-Bertani (LB) plates, followed by an overnight incubation at 25°C. A sterile pipette tip was used to scrape off bacteria from the plate and ejected into LB medium for an overnight incubation at 28°C. 1 mL of overnight *B. thuringiensis* culture was then transferred into 1000 mL of sterile BT medium (Bacto-peptone (Sigma) (7.5 g/l), glucose

(5.56 mM),  $\text{KH}_2\text{PO}_4$  (22.06 mM),  $\text{K}_2\text{HPO}_4$  (22.99 mM), pH adjusted to 7.2) with 5 mL of filter-sterilized salt solution ( $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  (100 mM),  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$  (2.37 mM),  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  (9.76 mM),  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  (14.39 mM)) and 1250  $\mu\text{l}$  of  $\text{CaCl}_2$  (1M). Cultures were incubated in a shaker-incubator at 120 rpm and 28°C for seven days. On day 4 of incubation 5 mL of salt solution and 1250  $\mu\text{l}$  of 1M  $\text{CaCl}_2$  was again added to the culture medium. The spore-crystal solution was harvested by centrifugation of the culture in 50 mL Falcon tubes (SARSTEDT) at 3000 rpm for 10 min. The pelleted particles were resuspended in Phosphate Buffered Saline (PBS), and the number of particles were counted using the Neubauer counting chamber method. The aliquots were stored in 1 mL tubes, at -20°C, with a concentration range of  $10^9$ - $10^{10}$  particles/mL for BT247 and BT679, and a concentration range of  $10^3$ - $10^4$  particles/mL for BT407. BT spore solution aliquots were freshly thawed before every infection experiment.

#### ***Pseudomonas aeruginosa* PA14**

The *Pseudomonas aeruginosa* PA14 strain was grown on LB Agar (LBA) plates and cultured overnight at 37°C in LB broth prior to the infection assay, which is described in detail below.

### **METHOD DETAILS**

#### **Population growth assay**

The population growth assay was used as a proxy to measure worm fitness by extrapolating the offspring produced by a single worm after two generations. Peptone-free medium (PFM) plates were used in all assays with BT to avoid the germination of spores on plates. PFM plates were seeded with a 500  $\mu\text{l}$ -lawn a BT spore solution of BT407, BT247 or BT679 mixed with an overnight culture of each of the tested microbiota isolate, adjusted to an  $\text{OD}_{600}$  of 10 with PBS. The concentration of the BT-microbiota mixes were 1:50 for the non-pathogenic control BT407 and a concentration of 1:200 for BT247 and BT679. The lawns were allowed to dry overnight at a temperature of 20°C. Three worms at the L4 stage were picked using a platinum wire onto the respective lawns, and incubated at 20°C for 5 days. The produced worm populations were then washed off the plates using 2 mL M9T (M9 buffer and 0.1% Triton) and frozen in 2 mL tubes at -20°C until counting of worms. The tubes were thawed and weighed prior to counting of worms, to correct for volumes lost during the washing procedure. From each tube a 10  $\mu\text{l}$  aliquot was used to count the number of worms, and the average of three independent aliquot counts from each sample was used to extrapolate the offspring produced from an individual worm. A control treatment group with OP50 mixed with each of the three BT strains was included in every run. The population growth assay was repeated for at least three independent runs with each of the two worm strains N2 and MY316. All treatments were randomized by initiating random codes on plates, to prevent experimenter bias.

The protective effect of the microbiota isolates on the population growth was analyzed in triplicates using a linear model framework (GLM), where the population growth of the worms after 5 days was incorporated as the dependent variable and the microbiota isolates and the pathogen strains as fixed factors independently. The post hoc analysis was done using the Tukey test and Bonferroni based adjustment was used to correct for multiple comparison testing [8].

#### **Survival assay**

##### ***Bacillus thuringiensis* survival assay**

*B. thuringiensis* survival assays were done according to a standardized protocol [31]. Accordingly, 6 cm PFM plates were inoculated with 75  $\mu\text{l}$  of each of the microbiota isolates or OP50, adjusted to  $\text{OD}_{600}$  of 10, mixed with different dilutions (1:5, 1:10, 1:25, 1:50) of the BT247 spore solution. For the non-pathogenic BT407 control the highest dilution (1:5) was used, mixed with each of the microbiota isolates or OP50. The plates were left to dry overnight at 20°C. On the infection day, 30 L4 nematodes were placed on each of the previously randomized plates, followed by incubation at 20°C. Worms were assessed as alive or dead using gentle touching with a sterile platinum wire. Alive and dead worms were counted 24h post-infection. Escaped worms were censored. For experiments with heat-inactivated bacteria, overnight cultures of the bacteria, prepared as described above, were adjusted to an  $\text{OD}_{600}$  of 10, and afterward were heat-inactivated by incubation at 80°C for a duration of 2h.

All BT survival assays were done in at least three independent runs with four technical replicates per treatment group in each run.

We used GLM analysis to evaluate the effects of the microbiota isolates on the survival of worms when challenged with each of the BTs across the concentration range of each of the different BT strains, using worm survival as dependent variable and each of the microbiota isolates and the BT concentration as fixed factors. The post hoc analysis was done using the Tukey test and corrections for multiple comparison testing with the OP50 control treatment group were done by Bonferroni based adjustment.

##### ***Pseudomonas aeruginosa* survival assay**

*P. aeruginosa* PA14 survival assay were done following the protocol described in Kirienko et al. [32] with minor modifications, as described in the following: Worms were bleached and maintained on the respective microbiota or OP50 lawn from the L1 stage until the L4 stage, as described above. Then worms were washed off with M9 and 30 worms were transferred to each slow killing infection (SK) plate that were inoculated with 10  $\mu\text{l}$  of overnight culture of either PA14 or OP50, followed by a 24 h incubation at 37°C and another 24 h incubation at 25°C. SK plates are enriched NGM plates (containing 0.35% instead of 0.25% peptone), which ensure efficient "slow killing" of worms [14].

Alive and dead worms were scored every 24 h, while the surviving worms were transferred to new SK plates. Survival assays were performed in at least three independent runs with four technical replicates per treatment group in each run. The data of the runs were pooled and analyzed by Kaplan-Meier and log rank test, and corrected for multiple testing with Bonferroni, using the platform RStudio (Version 1.0.153).

**Pathogen load assay**

Pathogen load assay was done by counting BT247 colony forming units (CFUs) as follows: Worms were grown on NGM-OP50 plates until the L4 stage and then exposed to each of the microbiota isolates or *E. coli* OP50 at an OD<sub>600</sub> of 10 mixed with BT247 spores at a concentration of 1:25. 24h post-infection ten adult worms were placed into tubes containing M9 buffer and 25 mM Tetramisole (TM buffer) using a hair picker. The supernatant was removed and the worms were washed twice with TMG buffer containing gentamicin (100 µg/mL) in TM buffer, and twice with M9 buffer. The last M9 washing buffer was transferred to another tube to serve as a washing control. Finally, 100 µL of PBST (PBS with 1% Triton X-100) was added to the tubes containing the worms. Sterile zirconia beads (1 mm in diameter) were then added to the tubes containing both the worms and the control supernatant and the samples were homogenized using the GenoGrinder200. The resulting suspension was serially diluted (10<sup>-1</sup>-10<sup>-6</sup>) with PBS and plated on TSA plates, for the subsequent growing and counting of BT colonies, after an incubation at 25°C for 24-48h.

The CFU score of the washing control was subtracted from the CFU score of the worm samples and divided by the number of worms recorded in the last washing step. The experiment was done in three independent runs with six technical replicates each. Students pairwise t test followed by Bonferroni correction for multiple testing was performed [8]. Results were considered significant when *P*-value ≤ 0.05. All treatments were randomized initiating random codes on plates and tubes, to prevent experimenter bias.

**Disc diffusion assay**

Kirby-Bauer disk diffusion method was performed in accordance with the guidelines of the Clinical and Laboratory Standards Institute (CLSI) and with the following modifications. Microbiota isolates, *E. coli* OP50 and BT247 or PA14 were grown in TSB in a shaker incubator overnight at 28°C. The bacterial suspensions were adjusted to an OD<sub>600</sub> of 10 with PBS. Culture supernatants were filtered with a 0.2-micron filter (SARSTEDT) to obtain cell-free supernatants. Mueller Hinton agar plates (9 cm in diameter) having a depth of 4mm of the medium were inoculated with each of the three BTs or PA14, using a cotton swab to evenly distribute the inoculums on the plate. Sterile Whatman (6mm in diameter) discs (GE healthcare) were inoculated with 15µl of the microbiota isolates' or *E. coli* OP50 filtered supernatant and placed on each of the BT or PA14 inoculated plates. 15 µl ampicillin (50 mg/mL) was used as positive control for BT and 15 µl gentamicin (20 mg/mL) was used as positive control for PA14. The plates were incubated for 48 h at 25°C. Images of each plate were taken using the ChemiDocTouch Imaging system (GelDoc Camera, BioRAD) and analyzed with subsequent Image lab software (Version 5.2.1). Zones of inhibition were measured with ImageJ (1.50e, NIH, USA). The assays were performed in at least three independent runs for each pathogen, with three technical replicates for each treatment group in each run.

The effects of MYb11 and MYb12 non-filtered supernatants on growth of six Gram-positive and 5 Gram-negative bacteria (Figure S2) was tested as described above. The bacterial suspensions were adjusted to an OD<sub>600</sub> of 1 with PBS. TSB was used as negative control.

Student's t test was used to evaluate the differences in means of the diameter of the zones of inhibition produced in the three runs (each run having three technical replicates), in comparison to the control OP50 treatment group. A Bonferroni based adjustment was used to correct for multiple comparison testing with the control. All treatments were randomized initiating random codes on plates, to prevent experimenter bias.

**Intestinal barrier function assay**

Nematodes were raised and infected as described in the pathogen load assay above and treated according to a published protocol by Gelino et al. [19] as described in detail in the following: Worms were washed off the infection plates with 5 mL S buffer (NaCl (100 mM), K<sub>2</sub>HPO<sub>4</sub> (6.5 mM), and KH<sub>2</sub>PO<sub>4</sub> (43.5 mM)) 24 h post-infection and suspended in a liquid suspension (1:1) of S buffer mixed with blue food dye (Erioglaucine disodium salt, Sigma-Aldrich, 5.0% wt/vol in water) for 3 h.

Animals were then washed three times with S buffer, centrifuged at 3500 rpm for 1 min, and 20 µl of the pellet was mounted on microscopic slides padded with 2% Agarose. 10 µl Sodium azide (20 mM) was added to paralyze the worms. The worms were visually analyzed for the presence and/or the absence of leaked blue dye outside the intestinal lumen and throughout the body cavity using a Leica dissecting microscope (LEICA M205 FA). The percentage of "smurf" worms was calculated per treatment, each plate containing at least 30 worms. The assay was performed in two independent runs with four technical replicates each. Student's pairwise t test and Bonferroni based adjustment was used to correct for multiple testing with the OP50 control treatment group. All treatment groups were randomized to avoid observer bias.

**Genomic analyses**

MYb11 and MYb12 sequences are available from NCBI GenBank, Bioproject PRJNA400855. antiSMASH [28] was used to search for known biosynthetic gene clusters (BGCs) in both genomes.

**UPLC- high resolution mass spectrometry**

The Ultra performance liquid chromatography high resolution mass spectrometry (UPLC-HRMS) analyses were performed with an UltiMate 3000 system (Thermo Fisher) coupled to an Impact II qToF mass spectrometer (Bruker) as described previously [33]. ACN with 0.1% formic acid in H<sub>2</sub>O was used as solvent. The flow rate was set at 0.4 mL min<sup>-1</sup> with a gradient from 5% ACN to 95% ACN over 15 min. The mass spectrometer was calibrated using 10 mM sodium formate before data acquisition. The MS method used for data acquisition also included an internal calibrant window before the data acquisition of each biological sample where 10 mM sodium formate were injected. The internal calibrant was used by Bruker DataAnalysis to correct the acquired mass data. The following

Please cite this article in press as: Kissoyan et al., *Natural C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group, *Current Biology* (2019), <https://doi.org/10.1016/j.cub.2019.01.050>

CellPress

MS settings were used for data acquisition: source settings: capillary voltage 4500 V, nebulizer nitrogen gas pressure 3 bar, ion source temperature 200°C, dry gas flow 8 L min<sup>-1</sup>; scan settings: ion polarity positive, mass range 90 to 1500, spectra rate 3 Hz; tune parameters: transfer funnel 1 RF 300 Vpp, funnel 2 RF 300 Vpp, isCID off, hexapole RF 60 Vpp; stepping settings: (i) 0 ms, collision RF 500, transfer time 82.5 μs, collision energy 65%; (ii) 25 ms, collision RF 500, transfer time 82.5 μs, collision energy 85%; (iii) 45 ms, collision RF 500, transfer time 82.5 μs, collision energy 100%; (iv) 75 ms, collision RF 500, transfer time 82.5 μs, collision energy 130%; MS/MS settings: 8 precursor ions, threshold 1000 counts (absolute), activated active exclusion after 3 spectra and 0.5 min release time, active precursor reconsidering factor 4, smart exclusion 2 times.

#### Preparative high-performance liquid chromatography

Methanolic XAD-16 extracts from 1 L cultures of MYb11 and MYb12 in M9 medium were used for fractionation. Preparative high-performance liquid chromatography (HPLC) was performed using an Agilent LC 1260 Infinity II Preparative LC/MSD System with an Agilent prep C18 column (10 μm, 30 × 250 mm) with a flow of 40 mL/min and a gradient from 5% to 100% ACN in 18 min. The ACN concentration remained at 100% for 3 min (MYb11) or 5.5 min (MYb12), respectively. 13 and 14 0.8 min-fractions were collected over time from 1.5 min to the end of the respective run. The fractions were dried and stored at -20°C. 50 μg of each fraction were used in disc diffusion assays as described above.

#### QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were done using RStudio (Version 1.0.153) and SPSS Statistics (Version 24.0). Inkscape was used for the graphical illustrations. When applicable normality of the data was examined using quantile plots and the Shapiro-Wilk normality test. Levene's test was used to inquire about the homogeneity of the variance, and proceed with the appropriate test accordingly. All pairwise comparisons were followed with Bonferroni correction for multiple testing. The detailed statistical tests used for each experiment are mentioned in the figure legends and in the method details section above, and included in the supplementary file. Significance was determined when the *P*-value ≤ 0.05. To avoid experimenter bias, we initiated random number codes for the treatments, incubated samples in a spatially randomized manner and evaluated the samples in a randomized order to minimize the influence of random environmental effects and observer bias.

#### DATA AND SOFTWARE AVAILABILITY

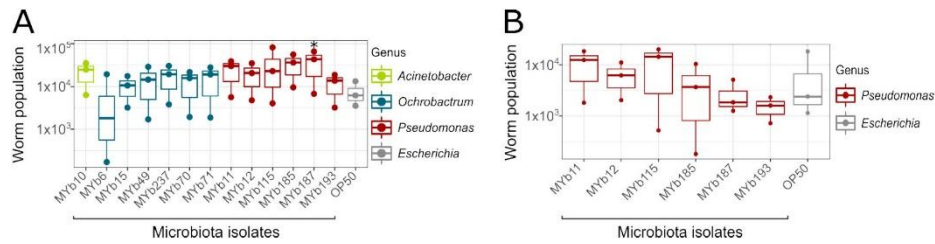
All analyses were performed with SPSS Statistics (Version 24.0) or RStudio (Version 1.0.153) as indicated in the specific experiment sections. Graphs were plotted using the ggplot function in R, and edited in Inkscape (Version 0.92). All raw data and statistical analyses are available on Mendeley: <https://data.mendeley.com/datasets/3464534ccd/draft?a=482aa37c-2ebd-4990-9b14-69383fafa3a>. All custom R scripts for analyses are available upon request.

Current Biology, Volume 29

**Supplemental Information**

**Natural *C. elegans* Microbiota Protects  
against Infection via Production  
of a Cyclic Lipopeptide of the Viscosin Group**

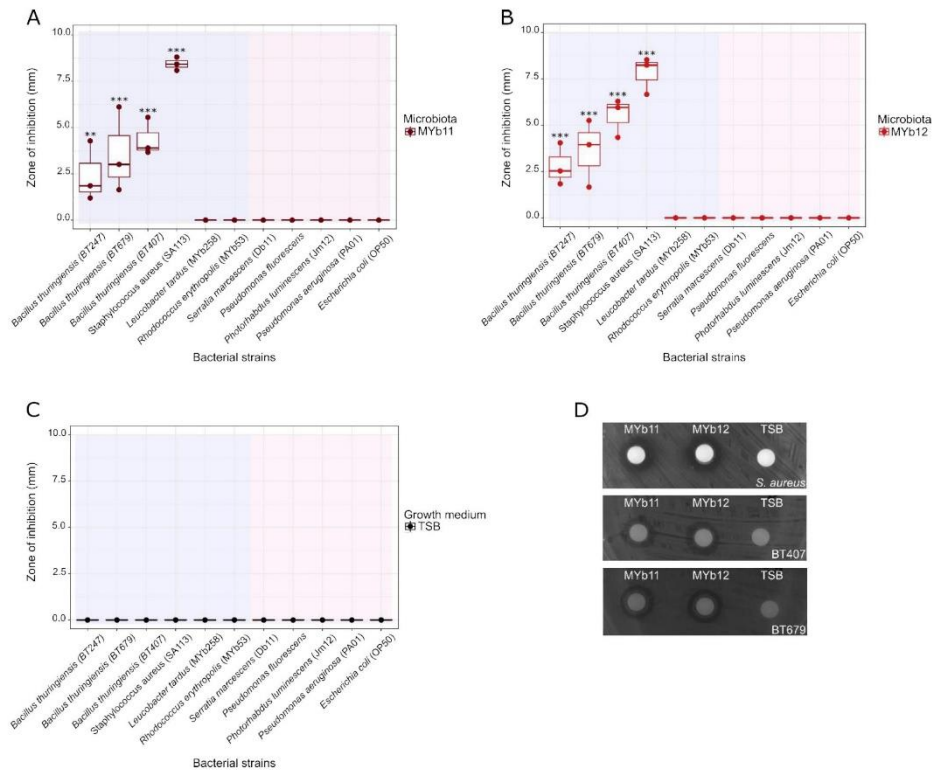
**Kohar A.B. Kissoyan, Moritz Drechsler, Eva-Lena Stange, Johannes  
Zimmermann, Christoph Kaleta, Helge B. Bode, and Katja Dierking**



**Figure S1. Natural microbiota isolates of the genus *Pseudomonas* increase *C. elegans* fitness in the presence of the non-pathogenic *B. thuringiensis* strain BT407. Related to Figure 1.**

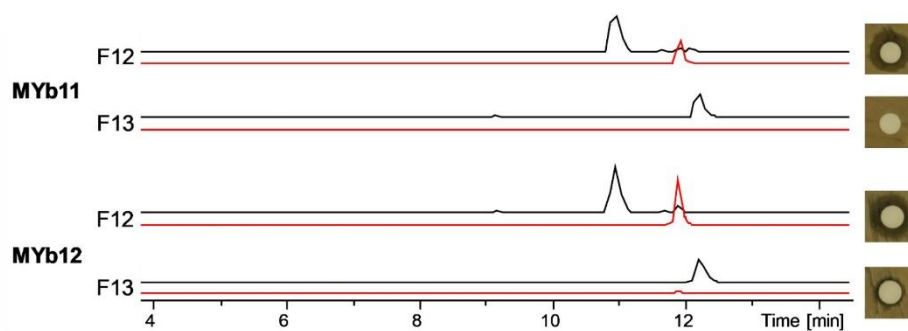
(A, B) Population size measured in the presence of non-pathogenic *B. thuringiensis* (BT407) of (A) the natural *C. elegans* isolate MY316 on one of 13 microbiota isolates and (B) the *C. elegans* N2 strain on one of 6 *Pseudomonas* isolates. *E. coli* OP50 was used as control. Three L4 larvae were picked onto the infection plates. Population size was assessed after 5 days at 20°C. The boxplots indicate the median, range and upper and lower quartiles of the worm population data ( $n = 3$  independent runs). Statistical analysis were performed using GLM [S1] model framework and the obtained  $P$  values were corrected using Bonferroni correction for multiple testing [S2].  $P$  values are considered significant and denoted with asterisks according to:  $*P \leq 0.05$ ,  $**P \leq 0.01$ ,  $***P \leq 0.001$ .



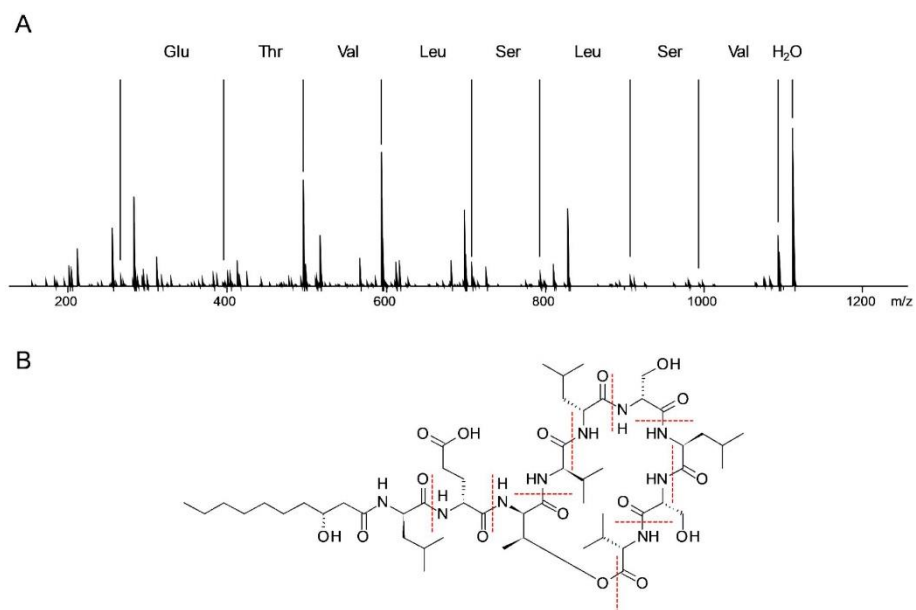


**Figure S2. *Pseudomonas lurida* MYb11 and MYb12 have antagonistic activity and inhibit growth of the Gram-positive bacterial strains BT407, BT679, and *S. aureus* SA113. Related to Figure 2.**

Growth inhibition of different Gram-positive and Gram-negative bacteria by supernatant of (A) MYb11, (B) MYb12, and (C) the TSB growth medium control, measured as diameter of inhibition zone (mm) in a disc diffusion assay. The data points represent the mean diameter of three technical replicates of three independent runs ( $n=3$ ). The boxplots indicate the median, range and upper and lower quartiles of the data. The diameters are standardized by subtracting the 6 mm diameter of the disc from the overall diameter measured. Student's pairwise t-test was performed against the TSB, followed by Bonferroni correction for multiple testing. Purple shade denotes Gram-positive bacteria and pink shade denotes Gram-negative bacteria. (D) Representative images of the disc diffusion assay showing that supernatant of MYb11 and MYb12 inhibit growth of *S. aureus*, BT407, and BT679.  $P$  values are considered significant and denoted with asterisks according to:  $*P \leq 0.05$ ,  $**P \leq 0.01$ ,  $***P \leq 0.001$ .



**Figure S3. Base peak chromatograms (BPCs, black) and extracted ion chromatograms (EICs, red) ( $m/z$  1112.66) and bioactivity of fractions 12 and 13 from *Pseudomonas* isolates MYb11 and MYb12. Related to Figure 4. Normalized BPCs and normalized EICs of fractions 12 and 13 of MYb11 and MYb12 extracts are shown. Images of the disc diffusion assay showing the bioactivity of the respective fraction against *B. thuringiensis* BT679 (BT247 looked identical) are shown next to the BPCs and EICs.**



**References**

- [S1] Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. Gen.* 135, 370.
- [S2] Dunn, O.J. (1961). Multiple comparisons among means. *J. Am. Stat. Assoc.* 56, 52–64.

## **CHAPTER II**

### **Protective *C. elegans* microbiota produce biofilm but are unable to inactivate pathogen toxin**

Kohar Annie B. Kissoyan, Christoph Giez, Guillaume Tetreau, Katja Dierking

**Manuscript ready for submission as a brief report**



## SUMMARY

The *Caenorhabditis elegans* natural microbiota was described only recently. As a result, our understanding of its effects on nematode physiology, such as microbiota-mediated protection against pathogen infection, is at its infancy. We previously showed that the protective *C. elegans* natural microbiota isolates *Pseudomonas lurida* MYb11, and MYb12 produce massetolide E, which directly inhibits growth of the nematode pathogen *Bacillus thuringiensis*(Bt). However, the protection mechanisms of other *C. elegans* microbiota isolates, such as *Pseudomonas fluorescens* MYb115 and *Pseudomonas denitrificans* MYb185, are still unclear. Here we investigate two additional potential protection mechanisms: The capacity of microbiota isolates to inactivate Bt toxins Cry21Aa3 and Cry14Aa2 and to form a protective biofilm. We find that *C. elegans* microbiota isolates MYb11, MYb115, and MYb185 protect the worm from infection with Bt spores but not from exposure to Bt toxins. In contrast, the microbiota isolates had a negative effect on worm survival upon toxin exposure. We also show that the protective *Pseudomonas* isolates MYb11, MYb12, MYb185, and *P. lurida* MYb193 produce biofilm *in vitro*, as a first step in investigating biofilm-mediated protection. In sum, we here show that microbiota-mediated protection is not achieved via Bt toxin inactivation and that protective microbiota produce biofilm *in vitro*, suggesting biofilm formation as a potential mechanism of protection, which merits *in vivo* investigation. Overall, this work highlights the function of microbiota as a double-edged sword, being protective or pathogenic for the host, according to specific pathogen exposure conditions

**KEYWORDS:** *C. elegans*, *Pseudomonas*, *B. thuringiensis*, microbiota, biofilm, toxin.

## INTRODUCTION

Throughout evolution, metazoans and microbes have shaped intricate host-microbe symbiotic associations (Sampson & Mazmanian, 2015). Nowadays, the host and its resident microbes, microbiota, are viewed as one entity, the metaorganism (Bosch & McFall-Ngai, 2011). One host-microbe interaction within the metaorganism is microbiota-mediated protection against infection (Masson & Lemaitre, 2017). Investigation of microbiota-mediated protection using simple model organisms can be very efficient, as microbiota composition is less complex and can be easily controlled (Douglas, 2019).

One such simple organism is the nematode *Caenorhabditis elegans*, a powerful model to investigate host-microbe interactions (Clark & Hodgkin, 2014; Douglas, 2019). The recent characterization of the *C. elegans* native microbiota opened up the possibility to use the advantages of the worm as an experimental model system in host-microbiota research (Dirksen et al., 2016; Samuel et al., 2016; Berg et al., 2016a; Zhang et al., 2017). Previously, we have identified six *C. elegans* microbiota isolates that protect the worm against infection with the Gram-positive bacterium *Bacillus thuringiensis* MYBt247(Bt247): *Pseudomonas lurida* MYb11, MYb12, and MYb193, *Pseudomonas fluorescens* MYb115, *Pseudomonas denitrificans* MYb185, and *Pseudomonas alkylphenolia* MYb187 (Kissoyan et al., 2019). Moreover, we have shown that two of the protective isolates, *P. lurida* MYb11 and MYb12, produce massetolide E, an antimicrobial cyclic lipopeptide of the viscosin group, which directly inhibits the growth of *B. thuringiensis* Bt247 and also another Bt strain, MYBt679(Bt679) *in vitro* (Kissoyan et al., 2019). The other protective *Pseudomonas* isolates MYb115, MYb185, MYb187, and MYb193 do not seem to inhibit pathogen growth directly, and the mechanism of protection is still unknown (Kissoyan et al., 2019). Bt produces crystal pore-forming toxins (Cry PFTs) that are important virulence factors involved in Bt pathogenesis. Cry toxins induce damage in the nematode via penetrating the target cell membrane, introducing pores, and eventually leading to cell death. Different strains of Bt produce distinct Cry PFTs and vary in virulence (Hollensteiner et al., 2017; Masri et al., 2015; Papkou et al., 2019; Schulte et al., 2010; Zárate-Potes et al., 2020). For example, while Bt247 produces the nematocidal toxin Cry6Ba (Papkou *et al.*, 2019), Bt679 produces the nematocidal toxins Cry21Aa3 and Cry14Aa2 (Masri et al., 2015; Zárate-Potes et al., 2020). *C. elegans* is killed by purified Bt679 toxins, in contrast to the purified Bt247 toxin that alone is not able to kill worms (Zárate-Potes et



al., 2020). *C. elegans* is shown to generate a differentiated response to the Bt247 and Bt679 strains of the same pathogen species (Zárate-Potes et al., 2020). Thus, we here investigate microbiota-mediated protection against Bt679 in *C. elegans*.

In host defense against spore and toxin-producing pathogens such as Bt, several mechanisms of direct (i.e., defined as host-independent protection, where microbiota interacts directly with the pathogen) microbiota-mediated protection have been described, in addition to the production of inhibitory compounds. One possible alternative direct protection mechanism is the inactivation of pathogen-produced toxins by the microbiota or antagonization of the toxic effect. For example, *Bacillus clausii*, a sporulating bacterium that constitutes part of the probiotic drug Enterogermina (a commercialized probiotic used for the prevention and treatment of acute diarrhea), produces proteases that inhibit the cytotoxic effects of *Clostridium difficile* and *Bacillus cereus* toxins *in vitro* (Ripert et al., 2016). Similarly, *Pseudomonas melophthora*, the symbiont of the apple maggot *Rhagoletis pomonella* produces esterases, which degrade six different insecticidal toxins hydrolytically, thus protecting the host (Boush & Matsumura, 1967). Thus, we here investigate if protective *C. elegans* microbiota isolates can inactivate Bt toxins.

Another possible mechanism for microbiota-mediated protection is biofilm-mediated protection. For example, in the tick *Ixodes scapularis* microbiota-produced biofilm and the strengthened peritrophic matrix (i.e., the glycoprotein-rich layer that separates the epithelial cells from the tick gut lumen) prevent colonization by the obligate intracellular bacterial pathogen *Anaplasma phagocytophilum*, thus protecting the host (Abraham et al., 2017). In *C. elegans* biofilm formation by *B. subtilis* prolongs worm lifespan and protects it from pathogen infection by *Pseudomonas aeruginosa* (Donato et al., 2017; Smolentseva et al., 2017). Thus, we here investigate if protective *C. elegans* microbiota isolates can form biofilm.

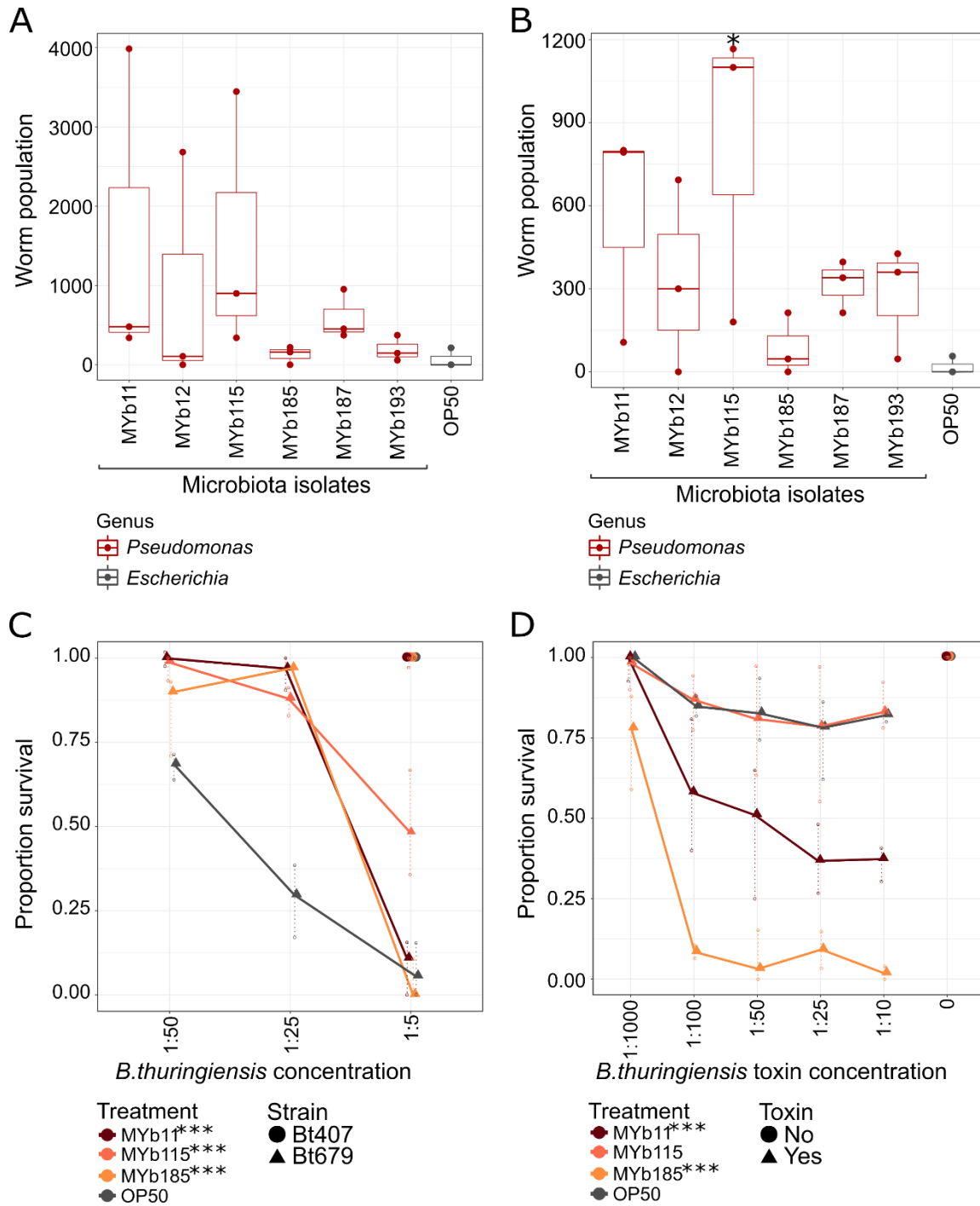
## RESULTS AND DISCUSSION

### *Natural C. elegans microbiota isolates increase worm fitness and resistance to B. thuringiensis MYBt18679(Bt679)*

Previously, we reported that several *C. elegans* natural microbiota isolates of the genus *Pseudomonas* increase worm fitness and resistance in the presence of the nematocidal Bt strain MYBt18247 (Bt247) (Kissoyan et al., 2019). Here, we explore if the previously observed protective effect of the microbiota isolates *Pseudomonas lurida* MYb11, MYb12, and MYb193, *Pseudomonas fluorescens* MYb115, *Pseudomonas denitrificans* MYb185, and *Pseudomonas alkylphenolia* MYb187 against Bt247 also occurs when we infect worms with Bt679.

First, we assessed worm population growth (as a proxy for fitness) of the natural *C. elegans* strain MY316 (**Figure 1A**) and the *C. elegans* laboratory strain N2 (**Figure 1B**) infected with Bt679 spore solution on the six microbiota isolates. We found that worms infected with Bt679 showed an increase in population growth in the presence of all tested microbiota isolates, compared to infected worms on the laboratory food control *E. coli* OP50, albeit statistically significant only for N2 on MYb115. However, these results align with our previous data on the protective effect of these microbiota isolates against infection with Bt247.

Next, we assessed worm survival (as a proxy for resistance) of Bt679 spore-infected N2 worms in the presence of three representative protective microbiota isolates MYb11, MYb115, and MYb185. Worms on the microbiota isolates showed significantly increased survival compared to infected worms on *E. coli* OP50 (**Figure 1C**). This result also concurs with our previous results on *C. elegans* microbiota-mediated protection against Bt247 (Kissoyan et al., 2019). Thereby, natural microbiota isolates of *C. elegans* enhance worm fitness and resistance and protect the host not only against Bt247 but also against Bt679.



**Figure 1: Natural microbiota isolates, *Pseudomonas* sp., protect *C. elegans* from infection with *B. thuringiensis* Bt679 spores, but not from purified Bt679 toxins.** (A and B) Population growth measured in the presence of pathogenic *B. thuringiensis* Bt679 of (A) the natural *C. elegans* isolate MY316 and (B) the laboratory N2 reference strain on one of six microbiota isolates. We used *E. coli* OP50 as control. Three L4 larvae were picked onto the infection plates. Population size was assessed after five days at 20°C. The boxplots indicate the median, range, and upper and lower quartiles of the worm population data ( $n = 3$  independent runs). Statistical analysis was

performed using a generalized linear model (GLM) framework, and the obtained *p*-values were corrected using Bonferroni correction for multiple testing. (C and D) Survival proportion of *C. elegans* N2 on different concentrations of *B. thuringiensis* Bt679 (C) spore solution and (D) purified toxins (containing Cry21Aa3 and Cry14Aa2) mixed with each of the microbiota isolates (MYb11, MYb115, and MYb185), and OP50 as a control. *B. thuringiensis* strain Bt407 (mixed with each of the microbiota isolates or OP50) was used as a non-pathogen control in the infection experiments with Bt spore mix. Error bars denote the range of the median of survival proportions of four technical replicates (*n* = 4), each containing at least 30 worms per treatment condition. Survival experiments were performed in at least two independent runs (for the Bt679 spores) and three independent runs (for the Bt679 toxins). Statistical analyses were performed using the GLM framework and Bonferroni correction for multiple testing with the OP50 control treatment group. Significance codes: 0 [\*\*\*], 0.001[\*\*], 0.01 [\*]

### *Microbiota isolates P. lurida MYb11, P. fluorescens MYb115, and P. denitrificans MYb185 do not protect against B. thuringiensis Bt679 toxins*

Our previous data and the data presented here demonstrate that *Pseudomonas* isolates MYb11, MYb12, MYb115, MYb185, MYb187, and MYb193 protect the worm from infection against both Bt247 and Bt679 (Kissoyan et al., 2019). Moreover, we have previously shown that the protective *C. elegans* microbiota isolates MYb11 and MYb12 produce the antimicrobial compound massetolide E and directly inhibit the growth of both Bt247 and Bt679 *in vitro* (Kissoyan et al., 2019). However, there is no evidence to show that the production of massetolide E is the exclusive protection mechanism of MYb11 and MYb12. Also, other protective microbiota isolates (e.g., MYb115) do not produce massetolide E, yet protect the worms against *B. thuringiensis* infection (Kissoyan et al., 2019). Microbiota-mediated direct protection could be achieved via interference competition (e.g., toxin inactivation), niche or nutrient competition and predation (reviewed elsewhere (Douglas, 2019; Ford & King, 2016; Masson & Lemaitre, 2017)). Microbiota-mediated toxin inactivation could for instance be achieved via the production of proteases that inhibit the cytotoxic effects of Bt toxins, similar to *Bacillus clausii*-produced proteases that inhibit the cytotoxic effects of *Clostridium difficile* and *Bacillus cereus* toxins *in vitro* (Ripert et al., 2016). Thereby, here we first investigated if *C. elegans* natural microbiota isolates show the capacity to inactivate Bt toxins. Thus, we exposed *C. elegans* to a mix of purified Bt679 toxins (containing Cry21Aa3 and Cry14Aa2) in the presence or absence of the three selected microbiota isolates MYb11, MYb115, and MYb185. Intriguingly, we found that worms on MYb11 and MYb185 showed significantly increased susceptibility to the Bt679 toxins (**Figure 1D**), contrary to the protective effect we observed when infecting worms with Bt679 spores (**Figure 1C**). While the protective microbiota isolates MYb11 and MYb185 interfere with Cry PFTs action when the worms

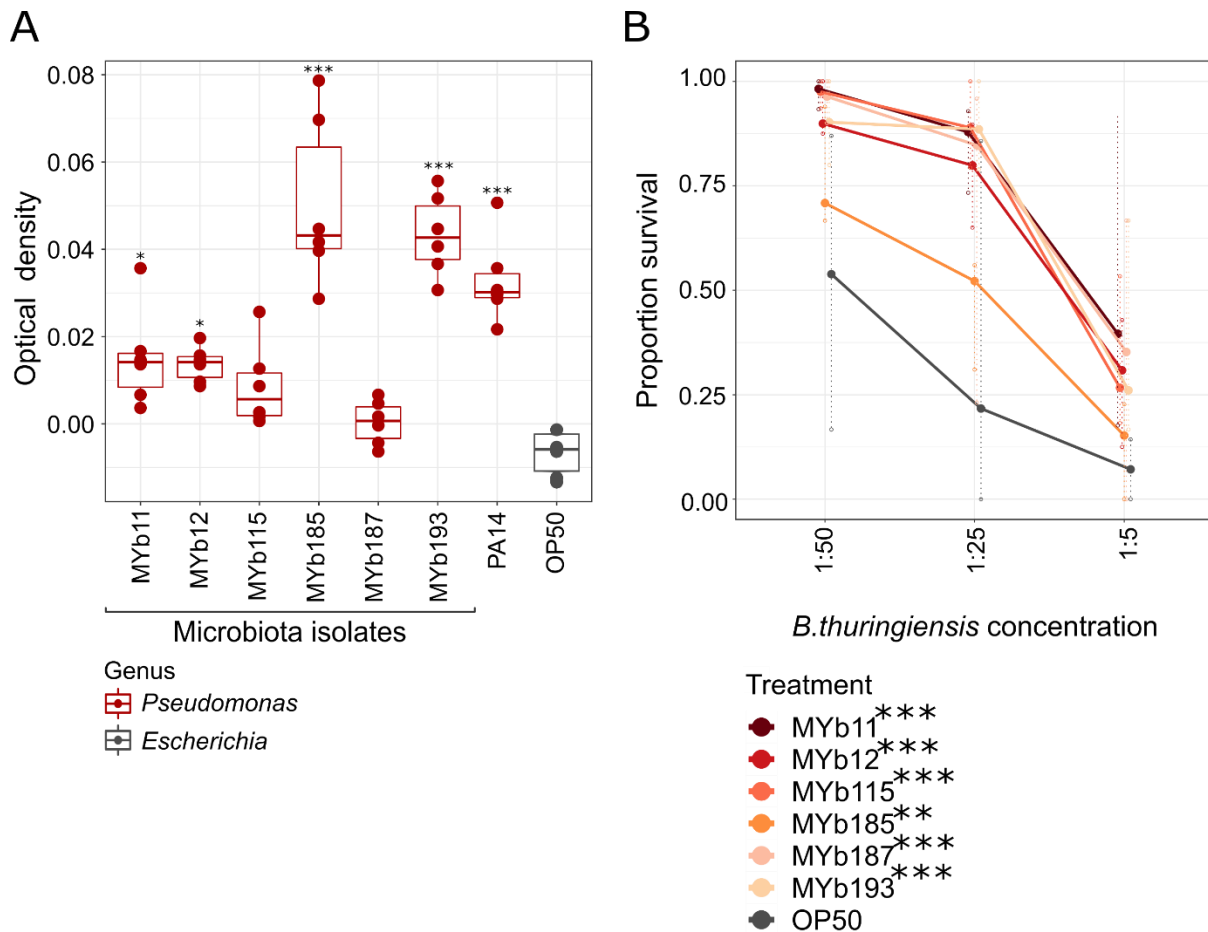
are infected with Bt679, the microbiota are rendered pathogenic when worms are exposed to purified Bt679 toxins. This might imply that the microbiota-mediated protective effect is probably via an interaction of the microbiota with the Bt spores, via the inhibition or limiting of Bt679 spore germination into vegetative cells, which otherwise replicate inside the worm. Furthermore, there was no difference between worms on MYb115 and worms on the OP50 control (in two out of the three independent runs, see *supplementary Figure S1*). However, as the Bt679 toxins kill worms much less efficiently than Bt679 spores and worm survival is relatively high for worms on both MYb115 and the OP50 control, it is challenging to detect protective effects. It thus remains unclear if MYb115 has a protective effect on worm survival upon exposure to the Bt679 toxins.

MYb11 and MYb185 do not protect the worm against the Bt679 toxins; on the contrary, both microbiota isolates increase *C. elegans* susceptibility to the Bt toxins, revealing a double-edged sword nature of *C. elegans* microbiota. These results are in line with previous reports on the effect of the microbiota on Bt pathogenicity in insects. In the gypsy moth *Lymantria dispar* L. Bt killing requires the presence of the host resident microbiota member *Enterobacter* sp. in the insect hemolymph (Broderick et al., 2006). Similarly, the Bt toxin Cry1Ca drastically alters gut microbiota in the lepidopteran host *Spodoptera littoralis*, resulting in the predominance of *Serratia* and *Clostridium* species and rendering the resident microbiota pathogenic (Caccia et al., 2016). Thus, these data highlight a potential role of the microbiota in aggravating the effect of Bt toxins. The mechanisms of microbiota dual nature in pathogen protection could be the extra-intestinal colonization of microbiota in the host, following the Bt toxin-induced epithelial damage (i.e. intestinal pore formation), as was the case in the moths (Broderick et al., 2006; Caccia et al., 2016). It needs to be determined in the future how MYb11 and MYb185 increase *C. elegans* susceptibility to Bt679 toxins, and whether extra-intestinal colonization of microbiota or synergism of toxins with microbiota could explain this dual nature.

*Microbiota isolates P. lurida MYb11, P. lurida MYb12, P. denitrificans MYb185, and P. lurida MYb193 produce biofilm in vitro*

As a first step to explore the possibility that the microbiota isolates MYb11, MYb12, MYb115, MYb185, MYb187, and MYb193 protect the host against pathogen infection via biofilm formation, we assessed their capacity to form biofilm. To this end, we measured *in vitro* biofilm production

using a microtiter plate assay (O’toole, 2003; Kissoyan et al., 2016) (**Figure 2A**). We used the *C. elegans* laboratory food bacterium *E.coli* OP50, which is known to be incapable of biofilm production, as the negative control (Arata et al., 2020; Donato et al., 2017; Tan & Darby, 2004). We found that biofilm production was significantly higher in four out of the six protective *Pseudomonas* isolates (MYb11, MYb12, MYb185, and MYb193) compared to the control (OP50) (**Figure 2A**). Moreover, the biofilm levels produced by the microbiota isolates were comparable to that of *P. aeruginosa* PA14, which is a notorious biofilm former (Wei & Ma, 2013; Haney et al., 2018). Intriguingly, the biofilm-producing bacteria also significantly protected the worms against Bt679 spore-induced infection (**Figure 2B**). Thus, the protective microbiota isolates produce biofilm *in vitro*. If biofilm formation is involved in microbiota-mediated protection against pathogen infection *in vivo* needs to be investigated in the future.

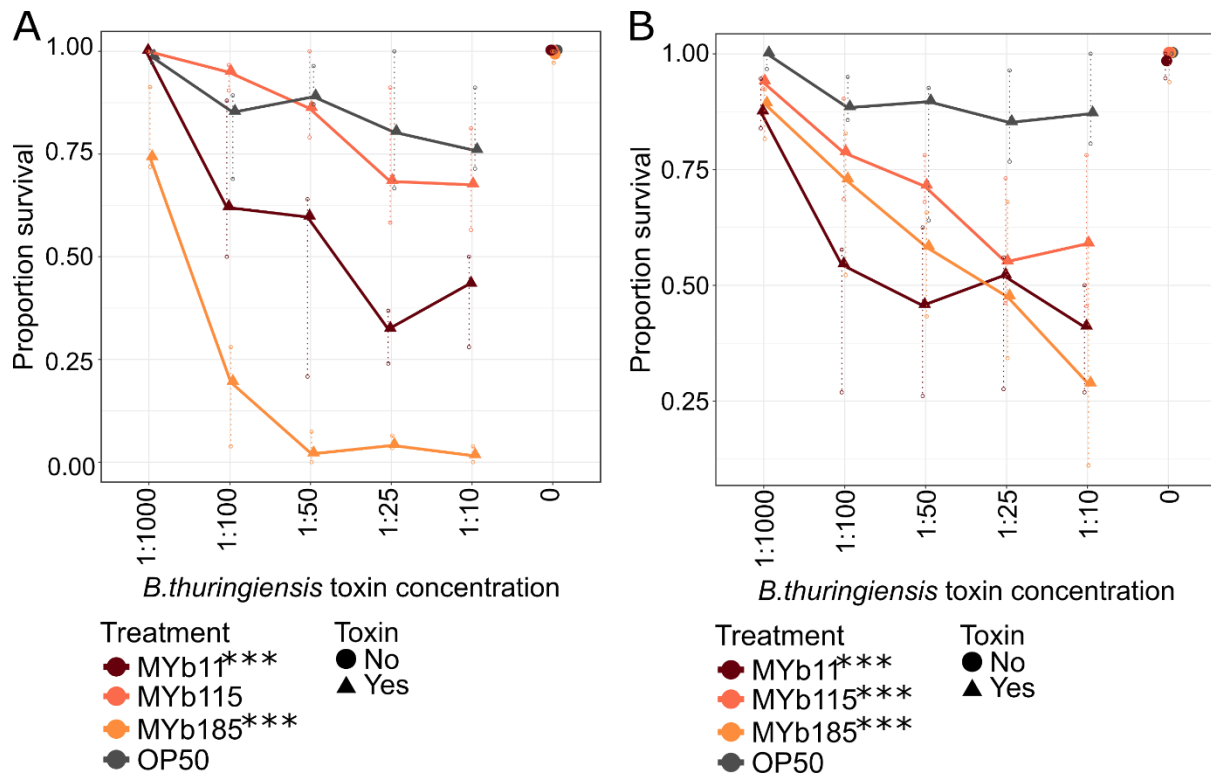


**Figure 2: Natural *C. elegans* microbiota isolates of the genus *Pseudomonas* produce biofilm *in vitro* and protect the nematode against *B. thuringiensis* infection.** (A) Quantification of biofilm formation by each of the six microbiota isolates, the negative control *E. coli* OP50, and the positive control *P.*

*aeruginosa* PA14. Optical density (OD) directly correlates with biofilm production. This is a representative of at least three independent runs, each with six technical replicates. The measurements were standardized using wells containing only the nutrient broth (TSB). A paired *t-test* compared to the control OP50, followed by Bonferroni correction for multiple testing was performed. (B) Survival proportion of *C. elegans* N2 on different concentrations of *B. thuringiensis* Bt679 spore solution. Spore solution was mixed with each of the OD600 adjusted (OD =10) microbiota isolates (MYb11, MYb12, MYb115, MYb185, MYb187, and MYb193) and OP50 as a control. Error bars denote the range of the median of survival proportions of four technical replicates (n = 4), each containing at least 30 worms per treatment condition. Statistical analyses were performed using the GLM framework and Bonferroni correction for multiple testing with the OP50 control treatment group. Significance codes: 0 [\*\*\*], 0.001[\*\*], 0.01 [\*].

In conclusion, we show that the protective *C. elegans* microbiota isolates *Pseudomonas lurida* MYb11, MYb12, and MYb115, *Pseudomonas denitrificans* MYb185, *Pseudomonas alkylphenolia* MYb187, and *Pseudomonas lurida* MYb193 enhance worm fitness and resistance in the presence of Bt679 spores, which produce the nematocidal Cry PFTs Cry21Aa3 and Cry14Aa2. Moreover, we provide initial evidence that the *C. elegans* protective microbiota isolates MYb11 and MYb185 do interfere with Cry PFTs action, but in contrast are rendered pathogenic when worms are exposed to purified Bt679 toxins. Thus, we exclude toxin inactivation as a microbiota-mediated protection mechanism. Furthermore, we show that *C. elegans* protective microbiota isolates MYb11, MYb12, MYb185, and MYb193 can produce biofilm *in vitro*. How far biofilm formation by the microbiota contributes to protecting the host from *B. thuringiensis* infection *in vivo* and whether it requires the synergistic effect of other mechanisms of protection remains to be determined.

## SUPPLEMENTARY FIGURES



**Supplementary Figure 1: Natural *C. elegans* microbiota isolates *Pseudomonas lurida* MYb11, *Pseudomonas fluorescens* MYb115, and *Pseudomonas denitrificans* MYb185 do not protect against *B. thuringiensis* purified Cry toxins.** (A) and (B) are each independent runs of the survival assay performed with the *B. thuringiensis* Bt679 purified toxins (containing Cry21Aa3 and Cry14Aa2), mixed with different concentrations of OD adjusted microbiota isolates and the *E. coli* OP50 control. Error bars denote the range of the median of survival proportions of four technical replicates ( $n = 4$ ). Statistical analyses were performed using the GLM framework and Bonferroni correction for multiple testing with the OP50 control treatment group. Significance codes: 0 [\*\*\*], 0.001[\*\*], 0.01 [\*]



## SUPPORTING INFORMATION

### MATERIALS AND METHODS

#### *Worm strain and maintenance*

*C. elegans* N2 strains were grown and maintained on nematode growth media (NGM) seeded with the *Escherichia coli* strain OP50 at 20°C, according to the routine maintenance protocol (Stiernagle, 2006). Worm populations were synchronized and incubated at 20°C. L4 hermaphrodites were used in all infection assays (described in detail below).

#### *Bacterial strains and maintenance*

Natural *C. elegans* microbiota isolates: *Pseudomonas lurida* MYb11, *Pseudomonas lurida* MYb12, *Pseudomonas fluorescens* MYb115, *Pseudomonas denitrificans* MYb185, *Pseudomonas alkylphenolia* MYb187, and *Pseudomonas lurida* MYb193 were used. The microbiota isolates were previously isolated from the *C. elegans* wild strain MY316 or its substrate (Dirksen et al., 2016). The microbiota isolates and the *E. coli* OP50 control were grown on Tryptic Soy Agar (TSA) plates at 25°C, and liquid bacterial cultures were grown in Tryptic Soy Broth (TSB) in a shaker-incubator overnight at 28°C.

*Bacillus thuringiensis* (MYBt247, MYBt679, and MYBt407) spore aliquots were obtained following a previously established protocol (Borgonie et al., 1995) with some modifications described in detail earlier (Kissoyan et al., 2019). The Bt spore aliquots were prepared in bulk and stored at -20°C, with the following concentration ranges:  $10^9$ - $10^{10}$  particles/ml for Bt247 and Bt679, and  $10^3$ - $10^4$  particles/ml for Bt407. Before every assay, new Bt aliquots were thawed and used.

*Pseudomonas aeruginosa* (PA14) was grown on Luria-Bertani Agar (LBA) plates and cultured at 37°C in LB broth overnight.

#### *B. thuringiensis* MYBt679 (Bt679) toxin purification

Spores from Bt679 were resuspended in LB liquid medium and isolated on LBA. A preculture launched overnight by inoculating 5 mL LB liquid medium with one isolated colony was spread on the T3 sporulation medium. T3 plates were incubated at 30°C for four days until full sporulation of the bacteria and release of crystals. The mixture of spores and crystals was collected using a cell scraper, resuspended in ultrapure water, and centrifuged once at 10,000 g for 45 min. The pellet was

resuspended in water. Spores and crystals were sonicated on ice for a total of 18 min, consisting of 3 cycles of 4 min of sonication at 70% power and 2 min of relaxing in ice. Sonication was performed in a cold room using an Omni Sonic Ruptor 400 (Omni International). Crystals were separated from spores by a discontinuous sucrose gradient (67-72-79%) using an SW-32 Ti swinging-bucket rotor on an Optima XPN ultracentrifuge (Beckman Coulter). After 16h of ultracentrifugation (23,000g, 4°C), the crystals were recovered from both the 67-72% and 72-79% interphases, and the pellet was discarded. Crystals were then washed by performing several rounds of centrifugation and resuspension in water to discard as much sucrose as possible. To analyze the toxin content, purified crystals suspension was verified on SDS-PAGE, as reported earlier (Zárate-Potes et al., 2020).

### *Population growth assay*

The population growth assay was done to measure the nematode fitness (by proxy) through extrapolating the offspring produced by a single worm after two generations according to a standardized protocol (Dirksen et al., 2016; Kissoyan et al., 2019). Population growth assay was performed at least three times in independent runs with each of the two worm strains N2 and MY316. We used GLM analysis to evaluate the effects of the microbiota isolates on the population growth assay, as described previously (Kissoyan et al., 2019).

### *Survival assay*

*B. thuringiensis* survival assays were done per the standardized protocol (Nakad et al., 2016; Kissoyan et al., 2019). Worms were assessed 24 hpi as alive or dead using gentle touching with a sterile platinum wire. Similarly, for the survival assay with the *B. thuringiensis* purified toxin, the microbiota isolates with the adjusted OD were mixed with different dilutions (1:10, 1:50, 1:100, 1:1000) of the purified toxin or no toxin (0). Bt survival assays were done in at least two independent runs, each having four technical replicates per treatment group. We used GLM analysis to evaluate the effects of the microbiota isolates on the survival of worms when challenged with each of the Bt strains across the concentration range, as described previously (Kissoyan et al., 2019).

### *Microtiter plate assay*

Bacterial isolates MYb11, MYb12, MYb115, MYb185, MYb193, OP50, and PA14, were grown in a microtiter plate assay to assess their ability of biofilm formation *in vitro*. We used the protocol described earlier with minor modifications (O'toole, 2003; Kissoyan *et al.*, 2016). Briefly, 100 µl of each of the bacterial isolate suspensions, adjusted to an OD of 0.001 with TSB, were transferred

to the microtiter plate in a randomized format. We had a minimum of triplicates for each condition. TSB-filled with similar amounts were considered the untreated controls. Following the incubation period, the planktonic bacteria were discarded. Crystal Violet was used to stain the biofilm formed, and after dissolving with ethanol, a 150- $\mu$ l aliquot from each well was transferred into a new polystyrene microtiter plate. The optical density was then measured at 630 nm using a plate reader. A paired *t-test* compared to the control OP50, followed by Bonferroni correction for multiple testing, was performed.

### *Statistical analyses*

Statistical analyses were performed by RStudio (Version 3.6.1). The statistical test used for each assay is mentioned in the respective sections above and in the figure legends. Graphs were plotted using the ggplot in R and were edited in Inkscape (Version 0.92). All treatments were randomized using random codes on plates to prevent experimenter bias.

## **ACKNOWLEDGMENTS**

We are grateful for the comments and valuable discussions from the Schulenburg group. We thank the Caenorhabditis Genetics Center (University of Minnesota, Minneapolis, Minnesota, USA), which is funded by the NIH Office of Research Infrastructure Programs (P40OD010440). This work was funded by the German Science Foundation DFG (Collaborative Research Center CRC1182 Origin and Function of Metaorganisms, Project A1.2), Germany. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

## **AUTHOR CONTRIBUTIONS**

K.D. secured funding and supervised the work. K.D. and K.A.B.K. conceived the study. K.A.B.K. and C.G. designed the experiments, performed, and analyzed the data. G.T. purified the BT679 toxins. K.D. and K.A.B.K. discussed and interpreted the data and wrote the manuscript.

## **DECLARATION OF INTERESTS**

The authors declare no competing interests.

## REFERENCES

- Abraham, N. M., Liu, L., Jutras, B. L., Yadav, A. K., Narasimhan, S., Gopalakrishnan, V., Ansari, J. M., Jefferson, K. K., Cava, F., Jacobs-Wagner, C., & Fikrig, E. (2017). Pathogen-mediated manipulation of arthropod microbiota to promote infection. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(5), E781–E790.
- Arata, Y., Oshima, T., Ikeda, Y., Kimura, H., & Sako, Y. (2020). OP50, a bacterial strain conventionally used as food for laboratory maintenance of *C. elegans*, is a biofilm formation defective mutant. *MicroPublication Biology*.
- Berg, M., Stenuit, B., Ho, J., Wang, A., Parke, C., Knight, M., Alvarez-Cohen, L., & Shapira, M. (2016a). Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *The ISME Journal*, *10*(8), 1998–2009.
- Berg, M., Zhou, X. Y., & Shapira, M. (2016b). Host-specific functional significance of *Caenorhabditis* gut commensals. *Frontiers in Microbiology*, *7*(1622).
- Borgonie, G., Van Driessche, R., Leyns, F., Arnaut, G., De Waele, D., & Coomans, A. (1995). Germination of *Bacillus thuringiensis* spores in bacteriophagous nematodes (Nematoda: Rhabditida). *Journal of Invertebrate Pathology*, *65*(1), 61–67.
- Bosch, T. C. G., & McFall-Ngai, M. J. (2011). Metaorganisms as the new frontier. *Zoology*, *114*(9), 185–190.
- Boush, G. M., & Matsumura, F. (1967). Insecticidal degradation by *Pseudomonas melophthora*, the bacterial symbiote of the apple maggot. *Journal of Entomology*, *60*(4), 918–920.
- Broderick, N. A., Raffa, K. F., & Handelsman, J. (2006). Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. *Proceedings of the National Academy of Sciences*, *103*(41), 15196–15199.
- Caccia, S., Di Lelio, I., La Storia, A., Marinelli, A., Varricchio, P., Franzetti, E., Banyuls, N., Tettamanti, G., Casartelli, M., Giordana, B., Ferré, J., Gigliotti, S., Ercolini, D., & Pennacchio, F. (2016). Midgut microbiota and host immunocompetence underlie *Bacillus thuringiensis* killing mechanism. *Proceedings of the National Academy of Sciences*, *113*(34), 9486–9491.
- Clark, L. C., & Hodgkin, J. (2014). Commensals, probiotics and pathogens in the *Caenorhabditis elegans* model. *Cellular Microbiology*, *16*(1), 27–38.
- Dahan, D., Preston, G. M., Sealey, J., & King, K. C. (2020). Impacts of a novel defensive symbiosis on the nematode host microbiome. *BMC Microbiology*, *20*(1), 159.
- Dirksen, P., Assié, A., Zimmermann, J., Zhang, F., Tietje, A.-M., Marsh, S. A., Félix, M.-A., Shapira, M., Kaleta, C., Schulenburg, H., & Samuel, B. S. (2020). CeMbio - The *Caenorhabditis elegans* microbiome resource. *G3: Genes, Genomes, Genetics*, *10*(8).
- Dirksen, P., Marsh, S. A., Braker, I., Heitland, N., Wagner, S., Nakad, R., Mader, S., Petersen, C., Kowallik, V., Rosenstiel, P., Félix, M.-A., & Schulenburg, H. (2016). The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biology*, *14*(38).
- Donato, V., Ayala, F. R., Cogliati, S., Bauman, C., Costa, J. G., Leñini, C., & Grau, R. (2017). *Bacillus subtilis* biofilm extends *Caenorhabditis elegans* longevity through downregulation of the insulin-like signalling pathway. *Nature Communications*, *8*, 14332.
- Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, *17*, 764–775.

- Dubois, N. R., & Dean, D. H. (1995). Synergism between cryia insecticidal crystal proteins and spores of *Bacillus thuringiensis*, other bacterial spores, and vegetative cells against *Lymantria dispar* (Lepidoptera: Lymantriidae) Larvae. *Environmental Entomology*, 24(6), 1741–1747.
- Haney, E. F., Trimble, M. J., Cheng, J. T., Vallé, Q., & Hancock, R. E. W. (2018). Critical assessment of methods to quantify biofilm growth and evaluate antibiofilm activity of host defence peptides. *Biomolecules*, 8(2).
- Hollensteiner, J., Poehlein, A., Spröer, C., Bunk, B., Sheppard, A. E., Rosenstiel, P., Schulenburg, H., & Liesegang, H. (2017). Complete genome sequence of the nematocidal *Bacillus thuringiensis* MYBT18247. *Journal of Biotechnology*, 260, 48–52.
- Kissoyan, K.A.B., Bazzi, W., Hadi, U., & Matar, G. M. (2016). The inhibition of *Pseudomonas aeruginosa* biofilm formation by micafungin and the enhancement of antimicrobial agent effectiveness in BALB/c mice. *Biofouling*, 32(7).
- Kissoyan, K.A.B., Drechsler, M., Stange, E.-L., Zimmermann, J., Kaleta, C., Bode, H. B., & Dierking, K. (2019). Natural *C. elegans* microbiota protects against infection via production of a cyclic lipopeptide of the viscosin group. *Current Biology*, 29(6), 1030–1037.
- Masri, L., Branca, A., Sheppard, A. E., Papkou, A., Laehnemann, D., Guenther, P. S., Prahl, S., Saebelfeld, M., Hollensteiner, J., Liesegang, H., Brzuszkiewicz, E., Daniel, R., Michiels, N. K., Schulte, R. D., Kurtz, J., Rosenstiel, P., Telschow, A., Bornberg-Bauer, E., & Schulenburg, H. (2015). *Bacillus thuringiensis* virulence and its cry toxin genes. *PLOS Biology*, 13(6).
- Masson, F., & Lemaitre, B. (2017). Protection from within. *eLife*, 6, e24111.
- Nakad, R., Snoek, L. B., Yang, W., Ellendt, S., Schneider, F., Mohr, T. G., Rösingh, L., Masche, A. C., Rosenstiel, P. C., Dierking, K., Kammenga, J. E., & Schulenburg, H. (2016). Contrasting invertebrate immune defense behaviors caused by a single gene, the *Caenorhabditis elegans* neuropeptide receptor gene *npr-1*. *BMC Genomics*, 17(280), 1–20.
- O’toole, G. A. (2003). To build a biofilm. *Journal of Bacteriology*, 185(9), 2687–2689.
- Papkou, A., Guzella, T., Yang, W., Koepfer, S., Pees, B., Schalkowski, R., Barg, M.-C., Rosenstiel, P. C., Teotónio, H., & Schulenburg, H. (2019). The genomic basis of Red Queen dynamics during rapid reciprocal host-pathogen coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, 116(3), 923–928.
- Ripert, G., Racedo, S. M., Elie, A.-M., Jacquot, C., Bressollier, P., & Urdaci, M. C. (2016). Secreted compounds of the probiotic *Bacillus clausii* strain o/c inhibit the cytotoxic effects induced by *Clostridium difficile* and *Bacillus cereus* toxins. *Antimicrobial agents and chemotherapy*, 60(6), 3445–3454.
- Sampson, T. R., & Mazmanian, S. K. (2015). Control of brain development, function, and behavior by the microbiome. *Cell Host and Microbe*, 17(5), 565–576.
- Samuel, B. S., Rowedder, H., Braendle, C., Félix, M. A., & Ruvkun, G. (2016). *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), E3941–E3949.
- Schulte, R. D., Makus, C., Hasert, B., Michiels, N. K., & Schulenburg, H. (2010). Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), 7359–7364.
- Smolentseva, O., Gusarov, I., Gautier, L., Shamovsky, I., Defrancesco, A. S., Losick, R., & Nudler, E. (2017). Mechanism of biofilm-mediated stress resistance and lifespan extension in *C. elegans*. *Scientific Reports*, 7(1).

- Stiernagle, T. (2006). Maintenance of *C. elegans*. *WormBook: The Online Review of C. elegans Biology*.
- Tan, L., & Darby, C. (2004). A movable surface: Formation of *Yersinia* sp. biofilms on motile *Caenorhabditis elegans*. *Journal of Bacteriology*, 186(15), 5087–5092.
- Wei, Q., & Ma, L. Z. (2013). Biofilm matrix and its regulation in *Pseudomonas aeruginosa*. *International Journal of Molecular Sciences*, 14(10), 20983–21005.
- Weldon, S. R., Russell, J. A., & Oliver, K. M. (2020). More is not always better: Coinfections with defensive symbionts generate highly variable outcomes. *Applied and Environmental Microbiology*, 86(5), 1–14.
- Zárate-Potes, A., Yang, W., Pees, B., Schalkowski, R., Segler, P., Andresen, B., Haase, D., Nakad, R., Rosenstiel, P., Tetreau, G., Colletier, J.-P., Schulenburg, H., & Dierking, K. (2020). The *C. elegans* GATA transcription factor *elt-2* mediates distinct transcriptional responses and opposite infection outcomes towards different *Bacillus thuringiensis* strains. *PLOS Pathogens*, 16(9), e1008826.
- Zhang, F., Berg, M., Dierking, K., Félix, M.-A., Shapira, M., Samuel, B. S., & Schulenburg, H. (2017). *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, 8(485).

## **CHAPTER III**

### **The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis elegans***

Johannes Zimmermann, Nancy Obeng, Wentao Yang, Barbara Pees,  
Carola Petersen, Silvio Waschina, Kohar Annie Kissoyan, Jack Aidley,  
Marc P. Hoepfner, Boyke Bunk, Cathrin Spröer, Matthias Leippe, Katja  
Dierking, Christoph Kaleta, Hinrich Schulenburg







## The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis elegans*

Johannes Zimmermann<sup>1</sup> · Nancy Obeng<sup>2</sup> · Wentao Yang<sup>2</sup> · Barbara Pees<sup>3</sup> · Carola Petersen<sup>2,3</sup> · Silvio Waschina<sup>1</sup> · Kohar A. Kissoyan<sup>2</sup> · Jack Aidley<sup>2</sup> · Marc P. Hoepfner<sup>4</sup> · Boyke Bunk<sup>5</sup> · Cathrin Spröer<sup>5</sup> · Matthias Leippe<sup>3</sup> · Katja Dierking<sup>2</sup> · Christoph Kaleta<sup>1</sup> · Hinrich Schulenburg<sup>1,2,6</sup>

Received: 15 February 2019 / Revised: 11 June 2019 / Accepted: 17 July 2019  
© The Author(s) 2019. This article is published with open access

### Abstract

The microbiota is generally assumed to have a substantial influence on the biology of multicellular organisms. The exact functional contributions of the microbes are often unclear and cannot be inferred easily from 16S rRNA genotyping, which is commonly used for taxonomic characterization of bacterial associates. In order to bridge this knowledge gap, we here analyzed the metabolic competences of the native microbiota of the model nematode *Caenorhabditis elegans*. We integrated whole-genome sequences of 77 bacterial microbiota members with metabolic modeling and experimental characterization of bacterial physiology. We found that, as a community, the microbiota can synthesize all essential nutrients for *C. elegans*. Both metabolic models and experimental analyses revealed that nutrient context can influence how bacteria interact within the microbiota. We identified key bacterial traits that are likely to influence the microbe's ability to colonize *C. elegans* (i.e., the ability of bacteria for pyruvate fermentation to acetoin) and affect nematode fitness (i.e., bacterial competence for hydroxyproline degradation). Considering that the microbiota is usually neglected in *C. elegans* research, the resource presented here will help our understanding of this nematode's biology in a more natural context. Our integrative approach moreover provides a novel, general framework to characterize microbiota-mediated functions.

### Introduction

Multicellular organisms are continuously associated with microbial communities. The ongoing interactions have likely influenced evolution of the involved microbes and hosts, affecting bacterial growth characteristics or host development, metabolism, immunity, and even behavior [1]. Host organisms and their associated microorganisms (i.e., the microbiota) are thus widely assumed to form a functional unit, the metaorganism, where microbial traits expand host biology [2]. To date, most microbiota studies focus on describing bacterial taxonomic composition, using 16S

These authors contributed equally: Johannes Zimmermann, Nancy Obeng

These authors jointly supervised this work: Christoph Kaleta, Hinrich Schulenburg

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-019-0504-y>) contains supplementary material, which is available to authorized users.

✉ Christoph Kaleta  
c.kaleta@iem.uni-kiel.de

✉ Hinrich Schulenburg  
hschulenburg@zoologie.uni-kiel.de

<sup>1</sup> Research Group Medical Systems Biology, Institute of Experimental Medicine, Christian-Albrechts University, Kiel, Germany

<sup>2</sup> Research Group of Evolutionary Ecology and Genetics, Zoological Institute, Christian-Albrechts University, Kiel, Germany

<sup>3</sup> Research Group of Comparative Immunobiology, Zoological Institute, Christian-Albrechts University, Kiel, Germany

<sup>4</sup> Institute of Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany

<sup>5</sup> Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

<sup>6</sup> Max-Planck Institute for Evolutionary Biology, Ploen, Germany

Published online: 04 September 2019

SPRINGER NATURE

rRNA amplicon sequencing [3]. These studies revealed that specific taxa reliably associate with certain hosts, for example Bacteroidetes and Firmicutes with humans, *Snodgrassella* and *Gilliamella* with honeybees, or *Lactobacillus* and *Acetobacter* with *Drosophila* [4–6]. 16S profiling, however, is insufficient to identify bacterial functions relevant for the interaction [7]. More insights can be obtained from bacterial genome sequences. For example, genomic analysis of bee microbiota members revealed complementary functions in carbohydrate metabolism, suggesting syntrophic interactions among bacteria [8]. Further, the systems biology approach of constraint-based modeling permits inference of genome-scale metabolic models and prediction of microbial phenotypes [9], as demonstrated for whiteflies and their endosymbionts [10, 11] and also hosts with complex microbiotas [12, 13].

The nematode *Caenorhabditis elegans* is an important model organism in biomedical research. Yet, almost all *C. elegans* research has been without microbiota. In fact, the nematode's microbiota was only characterized recently, consisting mostly of Gammaproteobacteria (*Enterobacteriaceae*, *Pseudomonaceae*, *Xanthomonadaceae*) and Bacteroidetes (*Sphingobacteriaceae*, *Weeksellaceae*, *Flavobacteriaceae*) [14–17], some of which persist in the worm intestine [15, 18, 19]. The microbiota composition is influenced by both host genotype and environment, and appears similar across geographic regions ([14, 15]; see meta-analysis in [17]). The few studies on microbiota functions highlight an influence on *C. elegans* fitness, stress resistance, and pathogen protection [15]. Previous studies also combined *C. elegans* with soil bacteria, revealing that these can provide specific nutrients [20–24]. Bacterial metabolism can also affect the worm's response to drugs against cancer and diabetes [25–28]. To date, the functions of the native microbiota have not been systematically explored.

Our aim was to establish the natural *C. elegans* microbiota as a model for studying microbiota functions. We extended previous 16S rRNA data [15] by sequencing whole genomes for 77 bacteria, which are associated with *C. elegans* in nature, most likely as part of the intestinal microbiota, and also *Escherichia coli* OP50, the nematode's standard laboratory food. We reconstructed metabolic networks from the genomes to explore the bacteria's metabolic competences and microbe–microbe interactions. We additionally characterized bacterial physiology and assessed which bacterial traits shape colonization ability and influence *C. elegans* fitness.

## Materials and methods

### Materials

Microbiota strains were previously isolated from natural *C. elegans* isolates or corresponding substrates in Northern

Germany ([15]; Supplementary Table S1). Briefly, bacteria from worms were obtained after nematodes were thoroughly washed and broken up by vortexing with Zirconium beads. Most bacteria are likely from the intestines, yet an association with the nematode cuticle cannot be excluded [15]. A representative set of 77 strains was chosen for genome sequencing, covering 79.5% of the diversity and abundance of the top ten genera and still 54.6% of that of the top 20 genera from the native *C. elegans* microbiota (some common microbiota members such as Flavobacteria could not yet be isolated [15]). For physiological analysis, bacteria were cultured in tryptic soy broth (TSB) at 28 °C. We performed experiments with the main *C. elegans* laboratory strain N2 (see below), thus allowing us to use previous literature and concurrent experiments with the standard food *E. coli* OP50 as a reference. For these experiments, bacterial TSB cultures (500 µl at OD<sub>600</sub> = 10) were spread onto peptone-free medium (PFM). Maintenance and bleaching, to obtain gnotobiotic, age-synchronized worms, followed standard methods [29].

### Genome sequencing

Total bacterial DNA was isolated using a cetyl-trimethylammonium-bromid (CTAB) approach [30]. Sequencing was based on Illumina HiSeq and in a subset of nine strains additionally PacBio (Supplementary Table S1). For PacBio, SMRTbell™ template library was prepared following manufacturer's instructions (Pacific Biosciences, US; Protocol for Greater Than 10 kb Template Preparation). SMRT sequencing was performed on the PacBio RSII (Pacific Biosciences, US), applying 240-min movie lengths. PacBio data was assembled using the RS\_HGAP\_Assembly.3 protocol (SMRT Portal version 2.3.0). Chromosomes and chromids were circularized, unusual redundancies at contig ends and artificial contigs were removed. Error correction was performed by Illumina reads mapping onto genomes using BWA [31] with subsequent variant calling using VarScan [32]. QV60 consensus concordances were confirmed for all genomes. Annotations were obtained with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). For samples with only Illumina data, low-quality reads and adaptors were trimmed with Trimmomatic v0.36 [33]. Genomes were assembled using SPAdes v3.8.0 [34] and contigs greater than 1000 bp annotated with PGAP and Prokka v1.11 [35]. Genomes were compared with BRIG [36]. BUSCO [37] analysis revealed high-genome completeness, irrespective of sequencing technology (mean completeness of 96.81%; Supplementary Figs. S12 and S13). As assembly quality of the available OP50 genome was low (NCBI project PRJNA41499; >2900 contigs, 73% completeness), we sequenced and assembled it again (218 contigs, 98.7% completeness).

The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis*...

All sequences are available from NCBI Genbank, Bio-project PRJNA400855.

### Reconstruction of metabolic networks

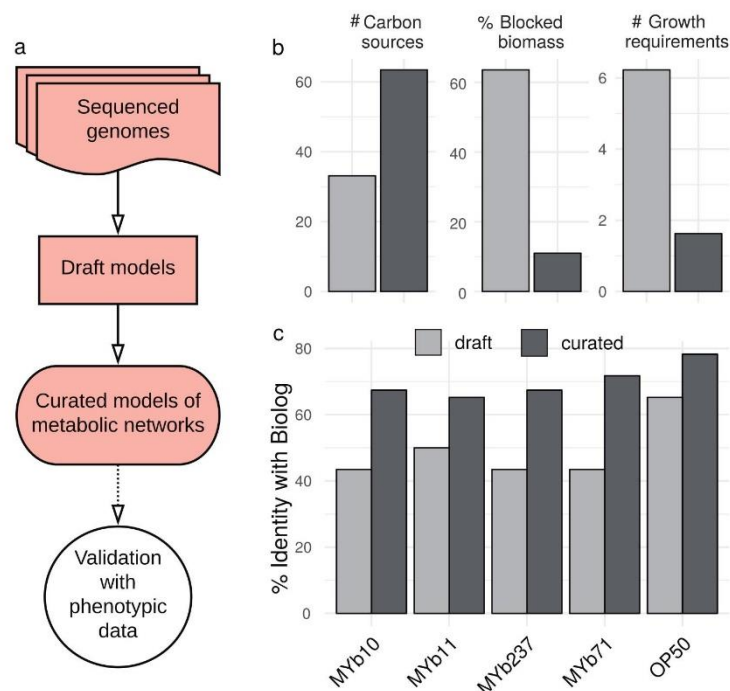
Metabolic networks were reconstructed following two steps (Fig. 1a). First, genomes were used to create draft metabolic models, using ModelSEED version 2.0 [38] and associated SEED reaction database. Second, we corrected errors and extended drafts by (i) finding futile cycles, (ii) allowing growth with the isolation medium (TSB), (iii) improving biosynthesis of biomass components, (iv) extending capacities to use different carbon sources, and (v) checking for additional fermentation products. Curation was based on combining topological- and sequenced-based gap-filling using gapseq (version 0.9; <https://github.com/jotech/gapseq>), pathway definitions of MetaCyc release 22 [39], and UniProt [40]. The presence of enzymatic reactions was inferred by BLAST with bitscore of  $\geq 250$  ( $\geq 150$  for more conservative estimation), and 75% minimum query coverage. Moreover, reactions were assumed to be present if overall pathway completeness was  $>75\%$  or if it was  $>66\%$  and key pathway enzymes were present. Host-microbe

interaction genes were identified with the virulence factor database [41]. The resulting curated models (Supplementary data S1) were used for further metabolic network analysis. Genome incompleteness did not have a large effect on pathway reconstruction (Supplementary Fig. S14). Computations were done with GNU parallel [42].

### Phylogenetic correlation and clustering of metabolic pathways

We assessed the correlation between metabolic pathway similarity and phylogenetic relationships, using pairwise comparisons of bacteria. We specifically focused on 16S rRNA sequences to calculate phylogenetic similarities, in order to enable comparisons with the standard microbiota analysis approach, based on 16S amplicon sequencing. 16S similarity was scored as percent identity with biostings [43], using data from the SILVA database [44] based on best hits of extracted genomic 16S rRNA using RNAmmer [45]. To determine overall metabolic distances between isolates, metabolic networks were treated as vectors, clustered horizontally, followed by computation of Euclidean distances between vectors. Cluster similarity was estimated

**Fig. 1** Genomes of bacterial isolates, reconstruction and validation of metabolic networks. **a** Pipeline for metabolic network reconstruction. Sequenced genomes were used to create draft metabolic models. Draft models were curated using topological- and sequenced-based gap-filling. The resulting models were validated with physiological data (BIOLOG GN2; see Fig. 3); these models represent the metabolic networks of microbiome isolates and were used for functional inference. **b** Model improvements by curation, leading to an increase in accurate prediction of uptake of carbon sources, and decreases in the prediction of non-producible biomass components and the number of components needed for growth. **c** Model curation improved agreement with experimental data, as for example the BIOLOG results



SPRINGER NATURE

by average linkage and assessed via multi-scale bootstrapping (10,000 replications) using pvclust [46].

### BIOLOG experiments

We used BIOLOG GN2 plates to assess metabolic competences of selected bacteria, including MYb10, MYb11, MYb71, MYb237, and OP50. Bacterial cultures were washed thrice using phosphate-buffered saline (PBS) and density adjusted to  $OD_{600} = 1.150 \mu\text{l}$  bacterial suspension per well of BIOLOG plate was incubated at 28 °C for 46 h. Tetrazolium dye absorption ( $OD_{595}$ ) was measured every 30 min (three replicates per strain). Substrate reduction was inferred from fold-change in tetrazolium absorbance:

$$\text{Foldchange} = \frac{OD_{t46} - OD_{t0}}{OD_{t0}} - OD_{\text{control}}$$

Fold-changes in water were subtracted as background. Hierarchical clustering of strains was based on average fold-change profiles (Ward's clustering; Euclidean distance) and bootstrapping ( $n = 100$ ). We analyzed metabolic specialization by  $k$ -means clustering of substrates ( $k = 7$ ,  $n = 10^3$ ; [47]) (Supplementary Fig. S1). Statistical analyses were performed in R version 3.3.1 [48] and ggplot2 [49].

### Bacterial growth experiments

To validate BIOLOG results, we assessed growth of MYb11, MYb71, and their co-culture in defined media with either alpha-D-glucose or D-(+)-sucrose as carbon sources. We focused on these two isolates, because they are members of two common taxa of the native microbiota of *C. elegans* [15] and because detailed information is available on the interaction of these two isolates with *C. elegans*, including their ability to colonize the nematode gut, persist under stressful conditions, influence nematode population growth, and provide protection against pathogens [15, 18]. Our defined medium is related to S medium [29]: 0.3% NaCl, 1 mM  $\text{MgSO}_4$ , 1 mM  $\text{CaCl}_2$ , 25 mM  $\text{KPO}_4$ , 0.1%  $\text{NH}_4\text{NO}_3$ , 0.05 mM EDTA, 0.025 mM  $\text{FeSO}_4$ , 0.01 mM  $\text{MnCl}_2$ , 0.01 mM  $\text{ZnSO}_4$ , 0.01 mM  $\text{CuSO}_4$ , and 1% carbon source. Medium without carbon source served as negative and TSB as positive control. Overnight cultures were washed and adjusted to  $3.94 \times 10^7$  CFUs for growth experiments. Microtiter plates were incubated as BIOLOG plates above.  $OD_{600}$  was measured every 30 min, and cultures plated after 48 h. Selective plating of MYb71 using kanamycin (10  $\mu\text{g}/\text{ml}$ ) allowed to quantify MYb11/MYb71 proportions in co-culture. Three independent runs with technical replicates were assessed with Mann–Whitney  $U$ -tests and  $P$ -value adjustment by false discovery rate (fdr, Benjamini Yoav et al. [50]).

### Simulation of bacterial in silico growth

We used the curated models to simulate growth of MYb11 and MYb71 with sucrose as carbon source using BacArena [51]. Sucrose invertases were identified with gapseq (<https://github.com/jotech/gapseq>) and secreted peptides with SignalP 4.1 [52]. The MYb71 extracellular sucrose invertase was modeled as independent species with a single sucrose invertase reaction and exchange reactions for sucrose, glucose, and fructose. Carbon source utilization and metabolic by-products were predicted using flux balance and variability analysis in R with sybil [53]. Flux balance analysis is a constrained-based method to estimate intra-cellular reaction activities by linear optimization [54], permitting inference of bacterial growth. A carbon source was assumed utilizable if the minimal solution of the corresponding exchange was negative and a byproduct producible if the maximal solution of exchange positive.

### Simulation of ecological interactions

We assessed possible interactions among bacteria using joined models, assuming a common compartment for metabolite exchange between microbes. Activity of individual reactions (i.e., fluxes) was linearly coupled to biomass production to prevent unrealistic exchange fluxes, such as those that solely benefit the partner but not the producer [55]. The objective function was set to maximize the sum of fluxes through both biomass reactions. Two growth media were used for simulations, TSB (Supplementary Table S2) and a glucose minimal medium with thiamine and traces (0.001 mM) of sucrose and methionine to allow initial bacterial growth (Supplementary Table S3). Joined growth rates ( $j_1$ ,  $j_2$ ) were compared to single growth rates ( $s_1$ ,  $s_2$ ). Mutualism was defined as  $j_1 > s_1$  and  $j_2 > s_2$ , competition as  $j_1 < s_1$  and simultaneously  $j_2 < s_2$ , parasitism as  $j_1 < s_1$  and simultaneously  $j_2 > s_2$  (or  $j_2 < s_2$  and  $j_1 > s_1$ ), and commensalism as  $j_1 = s_1$  and  $j_2 > s_2$  (or  $j_2 = s_2$  and  $j_1 > s_1$ ).

### Experimental analysis of bacterial colonization and bacterial effects on *C. elegans* population growth

We examined bacterial colonization of *C. elegans* (i.e., bacteria attached to worms after the washing protocol, thus mainly consisting of intestinal bacteria) by quantifying CFUs extracted from young adults exposed to bacteria for 24 h. In detail, L4 larvae were raised on OP50 lawns and placed on each of the considered bacteria (500  $\mu\text{l}$ ,  $OD_{600} = 10$ ; only one bacterium present). After 24 h, they were washed in a series of buffers (2  $\times$  M9 buffer with 25 mM tetramisole, 2  $\times$  M9 with 25 mM tetramisole and 100  $\mu\text{g}/\text{ml}$  gentamicin, 1  $\times$  PBS with 0.025% Triton-X) to remove

The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis*...

bacteria from nematode surfaces, and homogenized in the GenoGrinder 2000 using 1 mm zirconia beads (1200 strokes/min, 3 min). Worm homogenate and supernatant control were plated onto tryptic soy agar for quantification.

We further measured worm population growth as a proxy for worm fitness. Three L4, raised on OP50 lawns, were transferred onto lawns with the considered bacteria and total worm numbers counted after five days at 20 °C.

### Regression models

We analyzed the association between phenotypic measurements (i.e., bacterial colonization, worm fitness) and metabolic or virulence characteristics using Spearman rank correlation and random forest regression analysis. Significance of correlations was assessed with permutation tests (100 randomly generated features, FDR-adjusted *P*-values). For random forest regression, the R package VSURF served to select features via permutation-based score of importance [56] and otherwise default settings (ntree = 2000, ntry =  $p/3$ ).

### Adaptive strategies

According to the universal adaptive strategy theory (UAST) [57, 58], heterotrophic bacteria follow one of three strategies: (i) rapid growth and thus good competitor, (ii) high resistance and thus stress-tolerator, or (iii) fast niche occupation and thus ruderal. We categorized bacterial isolates using published UAST criteria for genomic data [58], which are based on three scores, inferred from genome sequences and metabolic models. In detail, the components of a competitive strategy were large genome size, antibiotics production (presence of pathways belonging to “Antibiotic-Biosynthesis” category in MetaCyc), high-catabolic diversity (Metacyc: “Energy-Metabolism”), and siderophore biosynthesis (Metacyc: “Siderophores-Biosynthesis”). The criteria for stress-tolerators were auxotrophies, slow growth rates in TSB, few rRNA copies, and exopolysaccharides production (MetaCyc pathways: PWY-6773, PWY-6655, PWY-6658, PWY-1001, PWY-6068, PWY-6082, PWY-6073). Those for a ruderal strategy were fast growth in TSB, multiple rRNA copies, and low catabolic diversity (Metacyc: “Energy-Metabolism”). The characteristics of each isolate were related to the other bacteria, yielding a relative score, thereby assuming that different strategies are present in the microbial community as a whole. For each isolate, we assessed whether the inferred value belonged to the lower or upper quantile of this criterium (in case of growth rates we used the mean instead). The total adaptive score per strategy was scaled by the number of features considered per strategy. An isolate was assumed to follow the strategy, for which it produced the highest score. If two strategies had

the same score, then we assumed a mixed strategy. The UAST classification remained stable with the same qualitative order, irrespective of genome completeness or sequencing technology (Supplementary Figs. S15 and S16).

## Results

### Genomes of bacterial isolates, reconstruction and validation of metabolic networks

We obtained whole genome sequences for 77 bacterial isolates of the *C. elegans* microbiota (Table 1). Of these, nine were sequenced with PacBio technology, allowing their full assembly, yielding either a single-circular chromosome (four strains) or three chromosomes/chromids in case of the five isolates of the genus *Ochrobactrum*, which is known to have more than one chromosome [59] (Supplementary Table S1). The remaining isolates were sequenced with Illumina only, resulting in assemblies with 11 up to 243 contigs. For four genera (*Ochrobactrum*, *Pseudomonas*, *Arthrobacter*, *Microbacterium*), we included more than five strains and identified substantial intra-generic genome variation (Supplementary Fig. S2).

To study the microbiota’s functional repertoire, we reconstructed genome-scale metabolic models (Fig. 1a and Supplementary Data S1). The initial metabolic models were curated by screening for transporter proteins and filling of missing reactions (gap-filling). Curation increased model quality, including doubling of utilizable carbon sources, reduced absence of essential biosynthesis pathways (e.g., for nucleotides or amino acids) from 60% to below 10%, and reduction in the required additional compounds for growth on defined media from on average six to one (Fig. 1b). To validate our metabolic models, we experimentally quantified the ability of five selected bacteria to utilize 46 carbon sources using the BIOLOG approach. The results produced 49.6% overlap with the initial and 70% overlap with the curated models (Fig. 1c and Supplementary Fig. S9). A 70% overlap is generally consistent with previous studies with model organisms like *Salmonella enterica*, *E. coli*, *Bacillus subtilis*, or *Pseudomonas putida* [60–62]. Notably, the models in these studies were manually reconstructed, highlighting the quality of our automated reconstructions.

### Metabolic diversity within the microbiome of *C. elegans*

We used the curated metabolic networks to assess the relationship between metabolic and phylogenetic similarities and the bacteria’s metabolic potential. For phylogenetic relationships, we specifically focused on 16S rRNA

**Table 1** Overview of bacterial isolates from the natural microbiota of *C. elegans* included in this study

Phylum	Order	Genus/Family	Isolate
Proteobacteria	Xanthomonadales	<i>Stenotrophomonas</i>	MYb238, <b>MYb57</b>
Proteobacteria	Pseudomonadales	<i>Pseudomonas</i>	MYb1, MYb114, MYb115, MYb117, MYb12, MYb13, MYb16, MYb17, MYb184, MYb185, MYb2, MYb22, MYb3, MYb60, MYb75, <b>MYb11</b> , MYb187, <b>MYb193</b>
Proteobacteria	Pseudomonadales	<i>Acinetobacter</i>	MYb10
Proteobacteria	Enterobacterales	<i>Erwinia</i>	MYb121
Proteobacteria	Enterobacterales	<i>Escherichia</i>	MYb137, MYb5, OP50
Terrabacteria group	Actinobacteria	<i>Micrococcaceae</i>	MYb211, MYb213, MYb214, MYb216, MYb221, MYb222, MYb224, MYb227, MYb229, MYb23, MYb51
Terrabacteria group	Actinobacteria	<i>Microbacteriaceae</i>	MYb24, MYb32, MYb40, MYb43, MYb45, MYb50, MYb54, MYb62, MYb64, MYb66, MYb72
FCB group	Bacteroidetes	<i>Flavobacteriales</i>	MYb25, MYb44, MYb7
Proteobacteria	Caulobacterales	<i>Brevundimonas</i>	MYb31, MYb33, MYb46, MYb52
Terrabacteria group	Bacilli	<i>Paenibacillaceae</i>	MYb63
Proteobacteria	Rhizobiales	<i>Ochrobactrum</i>	<b>MYb6</b> , MYb14, <b>MYb15</b> , MYb18, MYb19, MYb29, <b>MYb49</b> , <b>MYb58</b> , MYb68, <b>MYb71</b> , MYb237
Proteobacteria	Burkholderiales	<i>Achromobacter</i>	MYb9, <b>MYb73</b>
Terrabacteria group	Bacilli	<i>Bacillaceae</i>	MYb48, MYb56, MYb67, MYb78, MYb209, MYb212, MYb220
Bacteroidetes	Sphingobacteriales	<i>Sphingobacterium</i>	MYb181
Actinobacteria	Actinomycetales	<i>Rhodococcus</i>	MYb53

Strains with PacBio sequencing data are given in bold

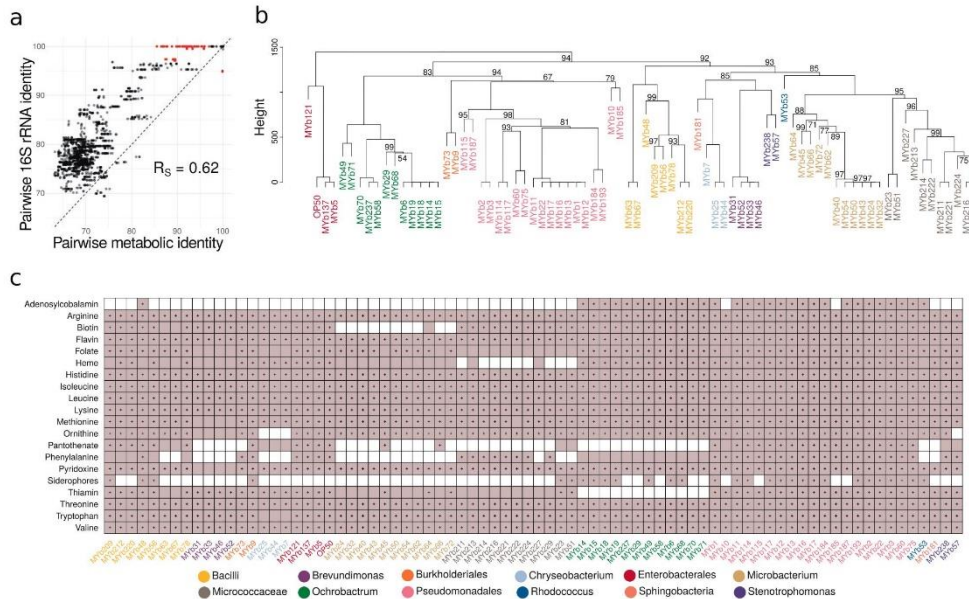
sequences, as they are most commonly used to characterize microbiota communities [3]. We found that pairwise 16S phylogenetic relationships are generally indicative of metabolic network similarities (Fig. 2a; Spearman rank correlation,  $R_S = 0.6199$ ,  $P < 0.0001$ ). Phylogenetic similarities appeared to be larger than metabolic similarities, suggesting some variation in metabolic competences within taxa. Such variation even occurred among isolates with 16S identity above 97%, often used as a species cut-off. This was confirmed through hierarchical clustering of inferred metabolic networks (Fig. 2b), for example for the genus *Pseudomonas* with three clearly separated clusters (see similar patterns for *Enterobacter*, *Ochrobactrum*, or *Microbacterium*). We conclude that variation in metabolic competences is generally related to bacterial phylogeny albeit some variation being present within genera.

We next assessed the bacteria's metabolic competences (Supplementary Table S4). In general, the inferred metabolic competences are consistent with the aerobic and heterotrophic lifestyle of the *C. elegans* host. The glycolysis, at least the partial pentose phosphate pathway, the tricarboxylic acid cycle, and enzymes enabling oxidative phosphorylation (cytochrome oxidases) were present in all genomes. Almost all isolates possessed enzymes enabling tolerance to microaerobic conditions (e.g., cytochrome bd oxidase). Some Bacilli, *Pseudomonas*, and *Ochrobactrum* showed competences for chemolithotrophic lifestyle (nitrite and formate oxidation) and anaerobic respiration (nitrate,

arsenate reduction). Pathways related to CO<sub>2</sub> fixation (reductive TCA or anaplerosis) were found in several *Pseudomonas*, *Microbacterium*, or Bacilli. Two Bacillales strains showed capacity to degrade polysaccharides, such as starch, cellulose, mannan, rhamnogalacturonan (e.g., *Paenibacillus* MYb63, *Bacillus* MYb67). The microbiota members are able to produce all essential substances required for *C. elegans* growth, which the nematode cannot synthesize on its own (i.e., all essential amino acids and vitamins; Fig. 2c). Most variation among isolates was observed in the biosynthetic pathways of B12, pantothenate, phenylalanine, and siderophores (Fig. 2c). Simulation of in silico growth (Supplementary Fig. S9) suggests that all bacteria can use simple sugars, such as glucose, ribose, or arabinose, while only some can degrade lactose, maltodextrin, or sucrose. Short-chain fatty acids can be generated by all bacteria (Supplementary Fig. S9), while they vary in succinate, cysteine, and valine production. Several microbes possess potential virulence genes, especially *Pseudomonas* and *Escherichia* isolates (Supplementary Table S5).

We subsequently focused on *Ochrobactrum* and *Pseudomonas* isolates. These two genera are enriched in the native microbiota of *C. elegans*, comprising 10–20% of the associated bacteria, they are particularly well able to colonize the nematode gut [15], and some isolates can protect *C. elegans* from pathogens [15, 18]. Most *Pseudomonas* isolates can provide all required substances for nematode growth.

The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis*...



**Fig. 2** Metabolic network clustering and distribution of important pathways. **a** Correlation between pairwise similarities in 16S rRNA sequences and metabolic networks is shown. Red indicates pairs with a 16S rRNA identity above 97% and metabolic identity below 97% and vice versa. **b** Hierarchical clustering of metabolic networks based on pathway prediction. *P*-values were calculated via multi-scale bootstrap resampling. In case of full support (i.e., *P* = 100), *P*-values are not shown (For a complete list of different unbiased *P*-values and

bootstrap values see Supplementary Fig. S11). **c** Prediction of bacterial capacity to produce metabolites favoring *C. elegans* growth. Filled squares in light purple indicate that the metabolic networks predict the presence of the biosynthetic pathway required to produce essential amino acids and co-factors. Black dots within the filled squares indicate that pathway presence is supported by more conservative parameters (BLAST bitscore  $\geq 150$ ). Different bacterial genera in **b**, **c** are indicated by different colors of the strain names (Table 1)

*Ochrobactrum* isolates can produce vitamin B12, like *Pseudomonas* isolates, but unlike most other microbiota members (Fig. 2c). Moreover, the *Ochrobactrum* isolates vary from other microbiota members in degradation pathways, energy-metabolism, vitamin biosynthesis, and potential virulence factors (Supplementary Table S6). They apparently lack thiamine and pantothenate vitamin biosynthetic pathways, essential for *C. elegans*. They possess a unique Brucella-like putatively immune-modulating LPS (Supplementary Table S5).

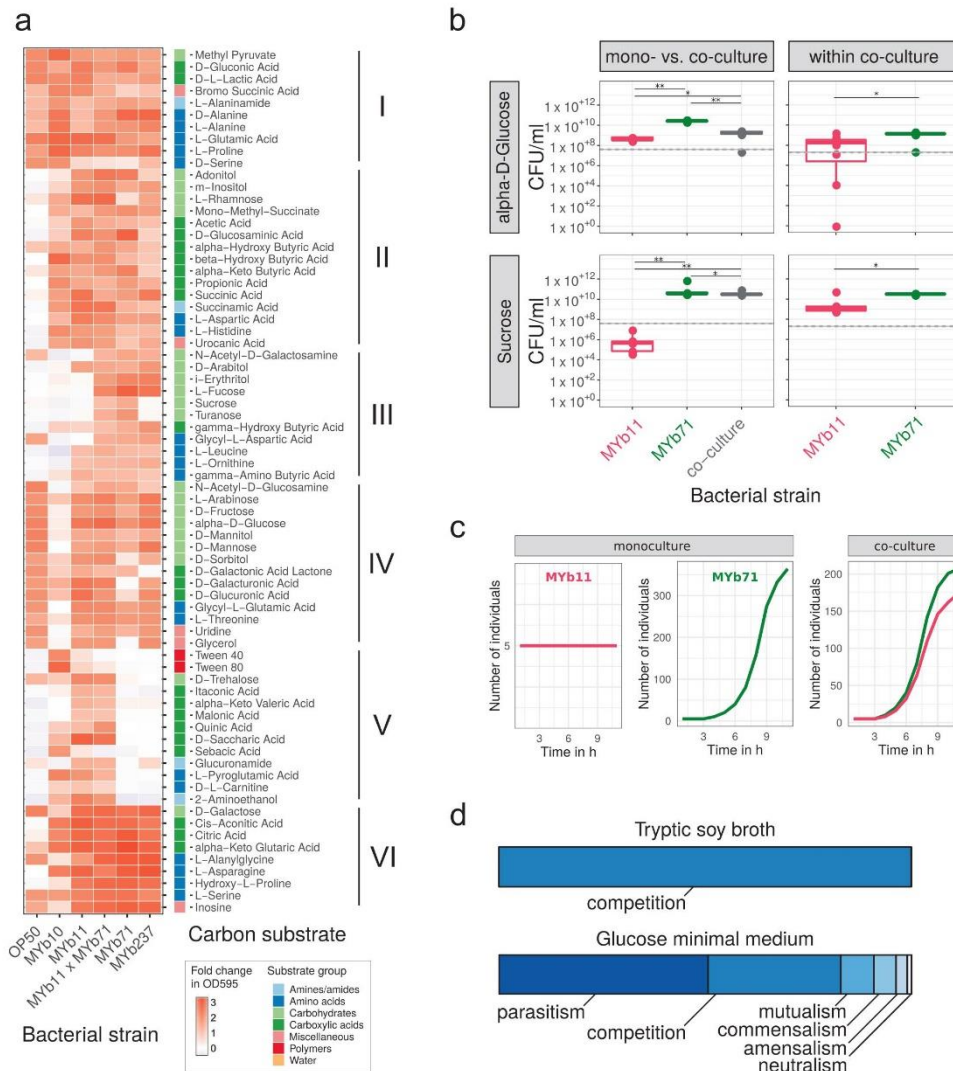
In summary, *C. elegans* harbors a microbial community with diverse metabolic competences, which can supply all essential nutrients for *C. elegans* and includes several *Ochrobactrum* and *Pseudomonas* isolates capable of producing important vitamins such as B12.

**Nutrient context influences ecological interactions within the microbiota**

To study how metabolic repertoires affect bacterial growth and interactions within the microbiota, we characterized

carbon source utilization of selected isolates and tested growth in different nutrient environments in vitro and in silico. Using the BIOLOG approach, we focused on prominent *C. elegans* microbiota members that colonize worms and affect host fitness, including MYb71, MYb237 (both *Ochrobactrum*), MYb10 (*Acinetobacter*), MYb11 (*Pseudomonas lurida*), and *E. coli* OP50 as control (Supplementary Fig. S3; ref. [15]). For a first insight into bacterial interactions, we additionally included a MYb11-MYb71 mixture (two strains that can co-exist in *C. elegans* [15]).

Metabolic repertoires of the strains differ and the four microbiota isolates deviate from OP50 in carboxylic and amino acid metabolism (Fig. 3a, cluster II; and Supplementary Fig. S4). MYb10 was least versatile at using carboxylic acids and sugar alcohols (Fig. 3a, cluster IV), while MYb11 and both *Ochrobactrum* could additionally metabolize unique sets of carboxylic acids and sugar alcohols, respectively (Fig. 3a, cluster V and III). Notably, the disaccharides sucrose and turanose were only metabolized by MYb71 (Fig. 3a, cluster III), although sucrose invertases were present in the genomes of both MYb71 and MYb11



**Fig. 3** Realized carbon metabolism and growth. **a** Profiles of carbon substrate use of *Acinetobacter* sp. (MYb10), *Pseudomonas lurida* (MYb11), *Ochrobactrum* sp. (MYb71), *Ochrobactrum* sp. (MYb237), and *E. coli* OP50 in BIOLOG GN2 plates over 46 h. The fold-change in indicator dye absorption from 0 to 46 h indicates that the particular compound is metabolized. *k*-means clustering ( $k = 7$ ) of substrates by fold-change highlights metabolic differences between strains. See Supplementary Fig. S5 for cluster VII with substrates used poorly across most strains. **b** Colony-forming units per ml (CFU/ml) of MYb11 and MYb71 in mono- and co-culture at 48 h in alpha-D-

glucose and sucrose-containing minimal media. The horizontal and dashed lines indicate mean and SD of CFU/ml at inoculation. Statistical differences were determined using Mann-Whitney *U*-tests and corrected for multiple testing using *fdr*, where appropriate. Significant differences are indicated by stars (\*\* for  $P < 0.01$ ; \* for  $P < 0.05$ ). Data from three independent experiments is shown. **c** In silico growth of MYb11 and MYb71 in mono- and co-culture in sucrose-thiamine medium using BacArena with an arena of  $20 \times 20$  and five initial cells per species. **d** Bacterial interaction types observed during in silico co-cultures of all combinations of the 77 microbiota isolates and OP50



The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis*...

(cf. pathway: sucrose degradation I, Supplementary Table S4). In co-culture, the metabolic repertoires of MYb11 and MYb71 appeared additive.

We next assessed whether the differences in MYb11 and MYb71 metabolic competences shape bacterial interactions in growth media with only a single carbon source. We focused on these two isolates as a model and proof-of-principle, because their interaction with *C. elegans* has been characterized in detail, including efficient colonization of nematodes, persistence under stressful conditions, an effect on nematode fitness, and protection against pathogens [15, 18]. We considered growth in the presence of two sugars, which are characteristic for the *C. elegans* natural habitat (e.g., rotting fruits and plant matter). We did not observe any growth in a control medium without carbon source, and thus the tested bacteria are not chemoautotrophic (Supplementary Fig. S6). In minimal medium with alpha-D-glucose, both MYb11 and MYb71 grew, yet exhibited distinct growth dynamics (Fig. 3b and Supplementary Fig. S6). MYb71 produced more CFUs than MYb11 in co-culture (Fig. 3b), suggesting that MYb71 has a growth advantage and/or interferes with MYb11 in some other way. In agreement with the BIOLOG results, a medium including only sucrose supported growth of MYb71 but not MYb11 in monoculture (Fig. 3b and Supplementary Fig. S6). Surprisingly, MYb11 grew in co-culture, indicating parasitic growth (Fig. 3b). Thus, the presence of different carbon sources can change the interaction type between two isolates.

We subsequently assessed the basis for co-growth of MYb11 and MYb71 in sucrose medium, using genome sequence information and in silico growth simulations. Interestingly, we found a secreted sucrose invertase in the genome of MYb71 but not MYb11 (Supplementary Fig. S10). In silico simulations demonstrated that MYb71 can grow in sucrose medium, MYb11 alone does not, while both grow in co-culture (Fig. 3c), confirming our in vitro results. Genome sequence information strongly suggests that growth of both in co-culture is mediated by a secreted enzyme from MYb71.

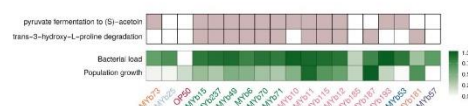
Taking a more global perspective, we next investigated in silico the potential ecological interactions among bacteria. We compared bacterial growth characteristics in monoculture and co-culture in different nutrient environments. In rich medium (TSB), the exclusive interaction type was competition, indicated by lower growth rates in co- vs. monoculture (Fig. 3d). This changed completely in glucose minimal medium: 50% interactions were parasitic (i.e., the growth rate for one isolate was higher in co-culture than in monoculture, while this pattern was opposite for the other isolate of a pair), 30% were competitive, and 8% mutualistic (i.e., growth rates for both isolates higher in co-culture than the monocultures; Fig. 3d). Under these minimal medium

conditions, the most frequently exchanged metabolites across bacteria were glyceraldehyde, acetate, and ethanol (Supplementary Fig. S7). We conclude that the nutrient context modulates bacterial growth, consistently identified in silico and in vitro, and thereby shapes bacteria–bacteria interactions.

### Specific metabolic competences predict bacterial colonization ability and bacterial effects on nematode fitness

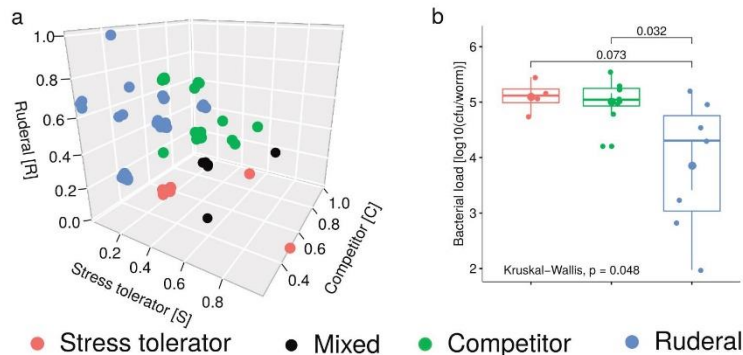
We next characterized traits involved in the interaction between bacteria and *C. elegans*, especially the bacteria's colonization ability and their effects on worm fitness. We focused on 18 microbiota isolates based on (i) their abundance in the *C. elegans* microbiota, (ii) enrichment in worms, and (iii) effects on worm population growth [15, 63]. OP50 was included as control. The bacteria varied substantially in their ability to colonize *C. elegans* and their effects on nematode fitness (Fig. 4; Supplementary Fig. S3; and Supplementary Table S7). Importantly, these two microbiota characteristics were significantly related with certain metabolic competences. Pyruvate fermentation to (S)-acetoin was significantly associated with bacterial load and the degradation of trans-3-hydroxyproline with nematode population growth (Fig. 4 and Supplementary Table S8).

To further explore the potential behavior of all microbiota isolates in an ecological context, we interpreted their genomic and metabolic traits in light of the universal adaptive strategy theory [57, 58]. Twenty-six isolates were associated with a competitive, nine with a stress-tolerating, and 37 with a ruderal (fast niche occupiers) strategy (Fig. 5a and Supplementary Table S9). The remaining six isolates showed a mixed strategy (same score for competition and stress-tolerance). Interestingly, isolates with different adaptive strategies also varied in their colonization ability (Fig. 5b): bacteria with competitive or stress-tolerance



**Fig. 4** Relationship of bacterial metabolic competences with their colonization ability and their effects on nematode fitness. Presence of metabolic traits (light purple color), which were found to be associated with the bacteria's ability to colonize *C. elegans* or affect nematode population growth as a proxy for worm fitness (green color). Regression models suggested that the pathway of pyruvate fermentation to acetoin influences bacterial load while the presence of hydroxyproline degradation is associated with *C. elegans* population growth. Colonization and population growth data was normalized; darker colors indicate increased capacities. Different bacterial genera are indicated by the different colors of the strain names (Table 1)

SPRINGER NATURE



**Fig. 5** Different adaptive strategies within the microbiota and their relationship to worm colonization. We applied the universal adaptive strategy theory proposed for soil bacteria [58] to categorize the bacterial isolates. **a** Based on genomic and metabolic features, each isolate obtained a score for the competitive (C), stress-tolerating (S), and ruderal (R) strategy, which is represented in the 3D-coordinate system.

**b** Bacterial colonization behavior in comparison to adaptive strategies. Isolates that were categorized as ruderal produced the lowest bacterial load, whereas stress-tolerator and competitors had the highest values. The difference in bacterial load between ruderal and other strategies was significant (Wilcoxon rank-sum test,  $P = 0.01$ )

strategies showed higher bacterial load than those with ruderal strategy (Wilcoxon rank-sum test,  $P = 0.01$ ). Moreover, for the competitive and stress-tolerance isolates, bacterial load was significantly correlated with the inferred score (Spearman,  $R_s = 0.37$ ,  $P = 0.1$ ; Supplementary Fig. S8). Taken together, the competitive and stress-tolerating strategies are most prevalent within the *C. elegans* microbiota and relate to bacterial colonization capacity.

## Discussion

We here present the first overview of the functional repertoire contained within the native microbiota of *C. elegans* and provide a metabolic framework for functional analysis of host-associated microbial communities. Whole-genome sequences were used to reconstruct the metabolic network of 77 microbiota members. We found that bacterial metabolic competences vary and that the community as a whole can produce nutrients essential for *C. elegans* growth. We identified a significant correlation between metabolic similarities and phylogenetic relationships inferred from 16S rRNA sequences, which are commonly used for bacterial classification. Metabolic variation was larger than evident from 16S data alone, suggesting that metabolic competences can be derived to only limited extent from 16S sequences and should ideally be reconstructed from whole-genome information. For selected bacteria, we validated the model predictions using physiological analyses. Moreover, both in vitro and in silico approaches demonstrated that the nutrient environment can modulate bacterial interactions,

for example, from competition to mutualism. We further identified specific metabolic modules that appear to shape the interaction with the host. Finally, we considered a combination of genomic, metabolic, and cellular traits to infer bacterial life history strategies according to the universal adaptive strategy theory [57, 58], finding that bacterial colonization ability is associated with a competitive or stress-tolerant strategy. In the following, we will discuss in more detail (i) the diversity of metabolic competences in the microbiota and possible implications for *C. elegans* biology, (ii) how metabolic networks shape bacteria–bacteria interactions, and (iii) how bacterial traits affect colonization and *C. elegans* fitness.

Our analysis revealed that the microbiota members are jointly able to synthesize all essential nutrients required by *C. elegans*. The considered isolates varied in their capacity to produce vitamins essential to *C. elegans*, such as folate, thiamine, and vitamin B12, which are known to affect nematode physiology and life history [21–23, 25, 64, 65]. For example, vitamin B12 influences propionate breakdown, it can accelerate development, and reduce fertility [21, 65]. Of the characterized bacteria, only *Pseudomonas* and *Ochrobactrum* strains had the pathways to produce vitamin B12. Their enrichment in the microbiota should therefore affect the metabolic state and fitness of *C. elegans*.

Our study demonstrated that the nutrient environment can change bacterial interactions. In our simulations, competitive interactions dominated in rich medium, while parasitic and mutualistic interactions in minimal medium. Interactions between *Pseudomonas lurida* MYb11 and *Ochrobactrum* MYb71 shifted from parasitic to competitive in a sucrose- vs. glucose-supplemented medium. We

The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis*...

detected a secreted sucrose invertase in the MYb71 genome, which otherwise lacks sucrose transporters. Thus, we propose that MYb71 breaks down sucrose extracellularly, and the monosaccharides glucose and fructose become exploitable by MYb11. While a similar phenomenon was described for yeast with engineered auxotrophies [66, 67], it was here observed for naturally coexisting host-associated bacteria. This emphasizes the relevance of nutrient context in host microbiota interactions. Importantly, no single growth medium might reliably predict all possible interaction types. It is therefore essential to consider the nutrient context to fully understand bacterial interactions within the microbiota (e.g., ref. [68]).

Our analysis further identified two bacterial traits that appear to influence the interaction with the host. Colonization ability was associated with pyruvate fermentation to (S)-acetoin. This fermentation pathway includes the ketone diacetyl as an intermediate, whose buttery odor attracts *C. elegans* and promotes feeding behavior [69]. In detail, diacetyl binds the transmembrane odor receptor *odr-10* and affects odortaxis [69–71]. As a result, worms are more attracted to bacterial lawns with this particular smell [69]. Indeed, lactic acid bacteria in rotting citrus fruits were more attractive to worms when releasing diacetyl [72]. Similarly, entomopathogenic *Steinernema* nematodes were more attracted to insect cadavers infected with the diacetyl-producing bacterial symbionts of the nematode [73]. Thus, if worms are attracted to diacetyl-producing bacteria, they should spend more time in their presence. This alone could increase bacterial uptake and colonization.

We also found that trans-3-hydroxyproline degradation in bacteria is associated with increased nematode fitness. In worms, hydroxyproline is present in collagen type IV, a major component of the extracellular matrix in the pharynx, intestine, and cuticle [74–76]. The breakdown of hydroxyproline can generate reactive oxygen species [77]. These may act as signaling molecules, which could affect cellular proliferation [78] and *C. elegans* reproduction [79]. Whether ROS in the gut increases brood size is unknown. Alternatively, bacteria with the degradation pathway may utilize the amino acid as a carbon source, consistent with the “microbiome on the leash” hypothesis, characterized by host-selection of beneficial bacterial traits [80].

In conclusion, our study provides a resource of naturally associated bacteria, their whole-genome sequences, and reconstructed metabolic competences that can be exploited to study and understand *C. elegans* in an ecologically meaningful context. This resource may help to further establish *C. elegans* as a model for studying host–microbe interactions.

**Acknowledgements** We thank Simone Severitt and Nicole Heyer for technical assistance regarding PacBio genome sequencing, Jolantha

Swiderski for long-read genome assemblies, Peter Deines for advice on BIOLOG assays, the Kiel BiMo/LMB for access to their core facilities, and the CRC 1182 and the Schulenburg group for general advice.

**Funding** German Science Foundation within the Collaborative Research Center CRC 1182 on Origin and Function of Metaorganisms, projects A1 (KD, ML, HS), A4 (HS), and INF (MPH, CK) and under Germany’s Excellence Strategy – EXC 22167-390884018 (Precision Medicine in Chronic Inflammation; CK, HS); the Competence Center for Genome Analysis Kiel (CCGA Kiel; CK, HS); the Max-Planck Society (Fellowship to HS); and the International Max-Planck Research School for Evolutionary Biology (NO).

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

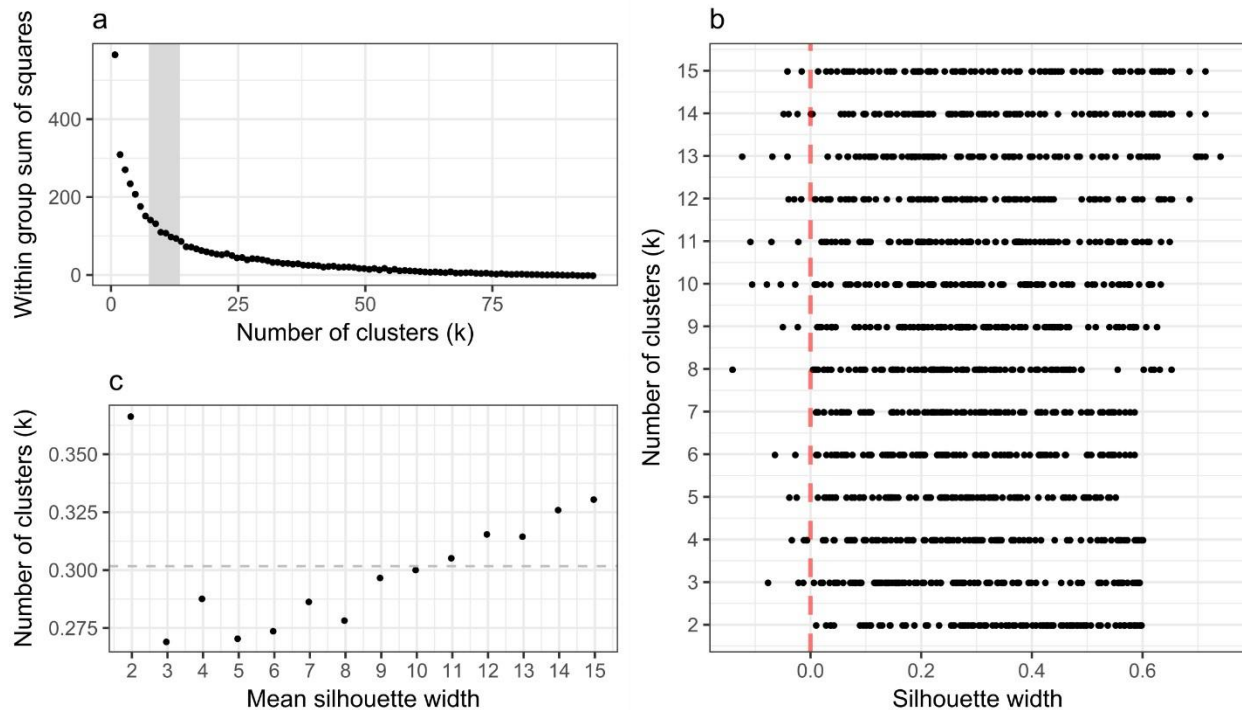
### References

- McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA*. 2013;110:3229–36.
- Bosch TCG, Miller DJ. The holobiont imperative. Vienna: Springer; 2016. <https://doi.org/10.1007/978-3-7091-1896-2>
- Pascoe EL, Hauffe HC, Marchesi JR, Perkins SE. Network analysis of gut microbiota literature: an overview of the research landscape in non-human animal studies. *ISMEJ*. 2017;11:2644–51.
- Wong CNA, Ng P, Douglas AE. Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ Microbiol*. 2011;13:1889–900.
- Consortium THMP. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207–14.
- Moran NA, Hansen AK, Powell JE, Sabree ZL. Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS ONE*. 2012;7:e36393.
- Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science*. 2016;353:1272–7.
- Kwong WK, Engel P, Koch H, Moran NA. Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc Natl Acad Sci USA*. 2014;111:11509–14.

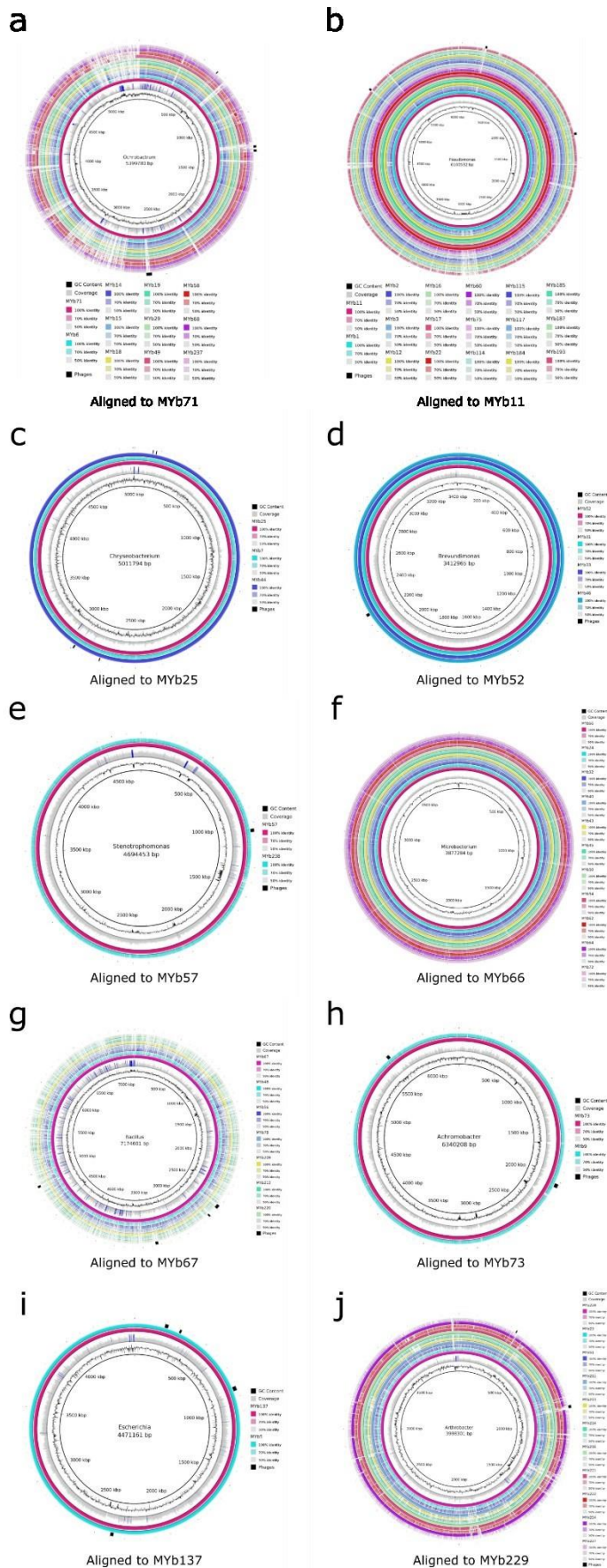
9. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet.* 2014;15:107–20.
10. Luan J-B, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, Ling K-S, et al. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol.* 2015;7:2635–47.
11. Ankrah NYD, Luan J, Douglas AE. Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. *J Bacteriol.* 2017;199:e00872–16.
12. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol.* 2017;35:81–9.
13. Bauer E, Laczny CC, Magnúsdóttir S, Wilmes P, Thiele I. Phenotypic differentiation of gastrointestinal microbes is reflected in their encoded metabolic repertoires. *Microbiome.* 2015;3:55.
14. Berg M, Stenuit B, Ho J, Wang A, Parke C, Knight M, et al. Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *ISME J.* 2016;10:1998–2009.
15. Dirksen P, Marsh SA, Braker I, Heitland N, Wagner S, Nakad R, et al. The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biol.* 2016;14:38.
16. Samuel BS, Rowedder H, Braendle C, Félix M-A, Ruvkun G. *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proc Natl Acad Sci USA.* 2016;113:E3941–9.
17. Zhang F, Berg M, Dierking K, Félix M-A, Shapira M, Samuel BS, et al. *Caenorhabditis elegans* as a model for microbiome. *Res Front Microbiol.* 2017;8:485.
18. Kissonyan KAB, Drechsler M, Stange E-L, Zimmermann J, Kaleta C, Bode HB, et al. Natural *C. elegans* microbiota protects against infection via production of a cyclic lipopeptide of the viscosin group. *Curr Biol.* 2019;29:1030–1037.e5.
19. Berg M, Monnin D, Cho J, Nelson L, Crits-Christoph A, Shapira M. TGF $\beta$ /BMP immune signaling affects abundance and function of *C. elegans* gut commensals. *Nat Commun.* 2019;10:604.
20. MacNeil LT, Watson E, Arda HE, Zhu LJ, Walhout AJM. Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell.* 2013;153:240–52.
21. Watson E, MacNeil LT, Ritter AD, Yilmaz LS, Rosebrock AP, Caudy AA, et al. Interspecies systems biology uncovers metabolites affecting *C. elegans* gene expression and life history traits. *Cell.* 2014;156:1336–7.
22. Chaudhari SN, Mukherjee M, Vagasi AS, Bi G, Rahman MM, Nguyen CQ, et al. Bacterial folates provide an exogenous signal for *C. elegans* germline stem cell proliferation. *Dev Cell.* 2016;38:33–46.
23. Virk B, Jia J, Maynard CA, Raimundo A, Lefebvre J, Richards SA, et al. Folate acts in *E. coli* to accelerate *C. elegans* aging independently of bacterial biosynthesis. *Cell Rep.* 2016;14:1611–20.
24. Shapira M. Host-microbiota interactions in *Caenorhabditis elegans* and their significance. *Curr Opin Microbiol.* 2017;38:142–7.
25. Cabreiro F, Au C, Leung K-Y, Vergara-Irigaray N, Cochemé HM, Noori T, et al. Metformin retards aging in *C. elegans* by altering microbial folate and methionine metabolism. *Cell.* 2013;153:228–39.
26. Scott TA, Quintaneiro LM, Norvaisas P, Lui PP, Wilson MP, Leung K-Y, et al. Host-microbe co-metabolism dictates cancer drug efficacy in *C. elegans*. *Cell.* 2017;169:442–456.e18.
27. García-González AP, Ritter AD, Shrestha S, Andersen EC, Yilmaz LS, Walhout AJM. Bacterial metabolism affects the *C. elegans* response to cancer chemotherapeutics. *Cell.* 2017;169:431–41.e8.
28. Norvaisas P, Cabreiro F. Pharmacology in the age of the holobiont. *Curr Opin Syst Biol.* 10:34–42.
29. Stiermagle T. Maintenance of *C. elegans*. In: *WormBook*, edited by The *C. elegans* Research Community, WormBook, 2006. <https://doi.org/10.1895/wormbook.1.101.1>, <http://www.wormbook.org>.
30. von der Schulenburg JH, Hancock JM, Pagnamenta A, Sloggett JJ, Majerus ME, Hurst GD. Extreme length and length variation in the first ribosomal internal transcribed spacer of ladybird beetles (Coleoptera: Coccinellidae). *Mol Biol Evol.* 2001;18:648–60.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
32. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22:568–76.
33. Bolger AM, Lohse BU. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77.
35. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
36. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genom.* 2011;12:402.
37. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: user guide. *Bioinformatics.* 2015;31:3210–2.
38. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol.* 2010;28:977–82.
39. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucl Acids Res.* 2018;46:D633–9.
40. UniProt Consortium. Activities at the universal protein resource (UniProt). *Nucl Acids Res.* 2014;42:D191–8.
41. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucl Acids Res.* 2016;44:D694–7.
42. Tange O. GNU parallel—the command-line power tool. *login. USENIX Magazine.* 2011;36:42–7.
43. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: efficient manipulation of biological strings. R package version 2.50.1, Vienna, Austria; 2018. <https://bioconductor.org/packages/release/bioc/html/Biostrings.html>.
44. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl Acids Res.* 2013;41:D590–6.
45. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl Acids Res.* 2007;35:3100–8.
46. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics.* 2006;22:1540–2.
47. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl Stat.* 1979;28:100–108.
48. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2013. <http://www.R-project.org/>.
49. Wickham H. ggplot: elegant graphics for data analysis. Springer-Verlag, New York; 2016. <https://ggplot2.tidyverse.org>.

50. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57:289–300.
51. Bauer E, Zimmermann J, Baldini F, Thiele I, Kaleta C. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Comput Biol*. 2017;13:e1005544.
52. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–6.
53. Gelius-Dietrich G, Desouki AA, Fritzsche CJ, Lercher MJ. Sybil—efficient constraint-based modelling in R. *BMC Syst Biol*. 2013;7:125.
54. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28:245–8.
55. Heinken A, Sahoo S, Fleming RMT, Thiele I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes*. 2013;4:28–40.
56. Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. *The R Journal*. 2015;7:19–33. <https://doi.org/10.32614/RJ-2015-018>.
57. Grime JP. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am Nat*. 1977;111:1169–1194.
58. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol*. 2017;15:579–90.
59. Chain PSG, Lang DM, Comerci DJ, Malfatti SA, Vergez LM, Shin M, et al. Genome of *Ochrobactrum anthropi* ATCC 49188T, a versatile opportunistic pathogen and symbiont of several eukaryotic hosts. *J Bacteriol*. 2011;193:4274–5.
60. AbuOun M, Suthers PF, Jones GI, Carter BR, Saunders MP, Maranas CD, et al. Genome scale reconstruction of a *Salmonella* metabolic model: comparison of similarity and differences with a commensal *Escherichia coli* strain. *J Biol Chem*. 2009;284:29480–8.
61. Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS Comput Biol*. 2009;5:e1000308.
62. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*. 2007;282:28791–9.
63. Sieber M, Pita L, Weilan-Bräuer N, Dirksen P, Wang J, Mortzfeld B, et al. The neutral metaorganism. *bioRxiv*. 2018;17:367243.
64. Bito T, Matsunaga Y, Yabuta Y, Kawano T, Watanabe F. Vitamin B12 deficiency in *Caenorhabditis elegans* results in loss of fertility, extended life cycle, and reduced lifespan. *FEBS Open Bio*. 2013;3:112–7.
65. Watson E, Olin-Sandoval V, Hoy MJ, Li C-H, Lousse T, Yao V, et al. Metabolic network rewiring of propionate flux compensates vitamin B12 deficiency in *C. elegans*. *eLife*. 2016;5:e17670.
66. Hoek TA, Axelrod K, Biancalani T, Yurtsev EA, Liu J, Gore J. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biol*. 2016;14:e1002540.
67. Zomorodi AR, Segrè D. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nat Commun*. 2017;8:1563.
68. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science*. 2015;350:663–6.
69. Ryan DA, Miller RM, Lee K, Neal SJ, Fagan KA, Sengupta P, et al. Sex, age, and hunger regulate behavioral prioritization through dynamic modulation of chemoreceptor expression. *Curr Biol*. 2014;24:2509–17.
70. Zhang Y, Chou JH, Bradley J, Bargmann CI, Zinn K. The *Caenorhabditis elegans* seven-transmembrane protein ODR-10 functions as an odorant receptor in mammalian cells. *Proc Natl Acad Sci USA*. 1997;94:12162–7.
71. Sengupta P, Chou JH, Bargmann CI. odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell*. 1996;84:899–909.
72. Choi JI, Yoon K-H, Subbammal Kalichamy S, Yoon S-S, II Lee J. A natural odor attraction between lactic acid bacteria and the nematode *Caenorhabditis elegans*. *ISMEJ*. 2016;10:558–67.
73. Baiocchi T, Lee G, Choe D-H, Dillman AR. Host seeking parasitic nematodes use specific odors to assess host resources. *Sci Rep*. 2007;7:6270.
74. Page AP, Johnstone IL. The cuticle. *Wormb: online review of *Caenorhabditis elegans* biology*; 1–15. <http://www.ncbi.nlm.nih.gov/pubmed/18050497>
75. Hutter H, Vogel BE, Plenefisch JD, Norris CR, Proenca RB, Spieth J, et al. Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science*. 2000;287:989–94.
76. Graham PL, Johnson JJ, Wang S, Sibley MH, Gupta MC, Kramer JM. Type IV collagen is detectable in most, but not all, basement membranes of *Caenorhabditis elegans* and assembles on tissues that do not express it. *J Cell Biol*. 1997;137:1171–83.
77. Cooper SK, Pandhare J, Donald SP, Phang JM. A novel function for hydroxyproline oxidase in apoptosis through generation of reactive oxygen species. *J Biol Chem*. 2008;283:10485–92.
78. Halliwell B. Biochemistry of oxidative stress. *Biochem Soc Trans*. 2007;35:1147–50.
79. De Henau S, Tilleman L, Vangheel M, Luyckx E, Trashin S, Pauwels M, et al. A redox signalling globin is essential for reproduction in *Caenorhabditis elegans*. *Nat Commun*. 2015;6:8782.
80. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. *Nature*. 2017;548:43–51.

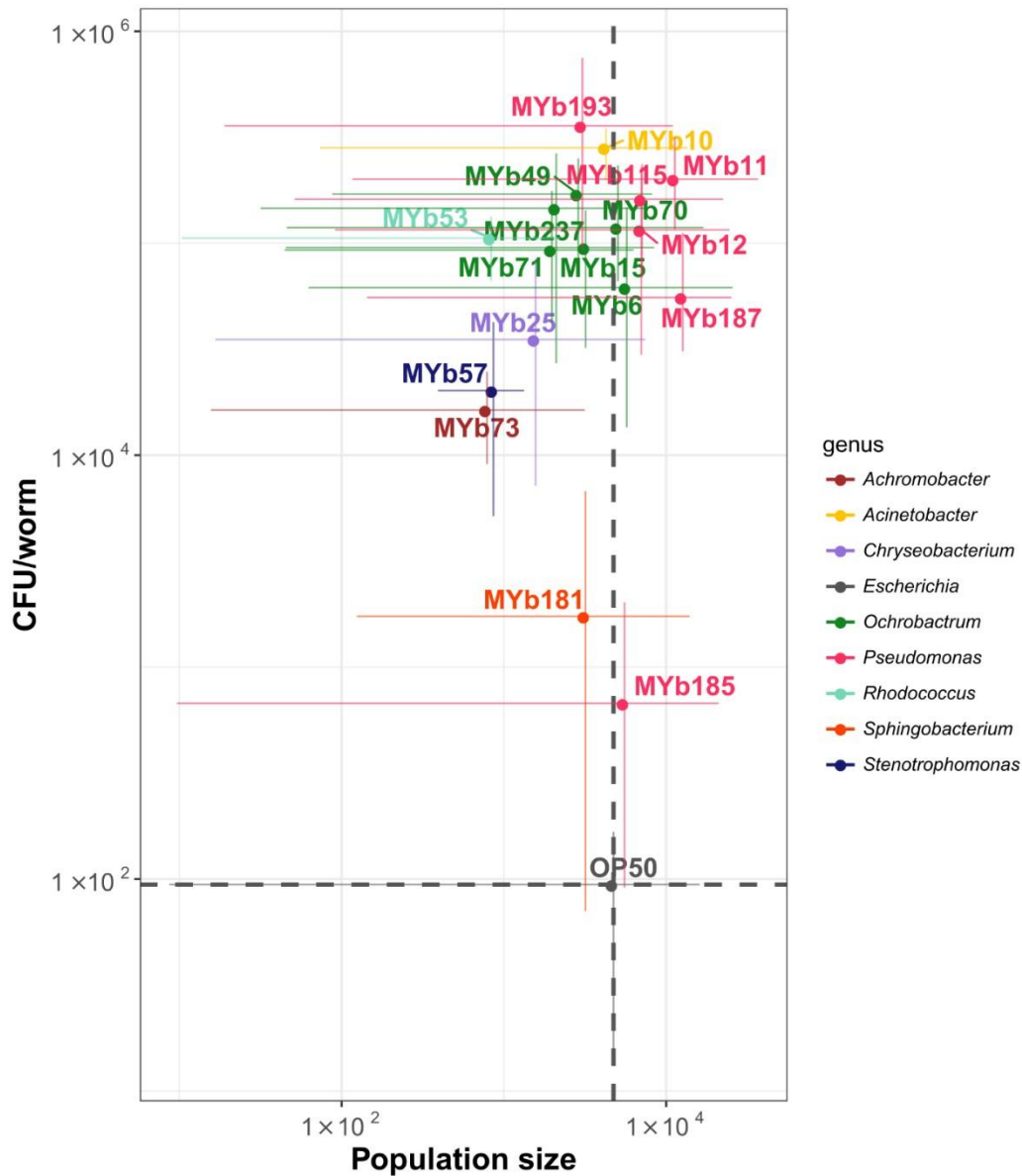
## Supplementary figures



**Supplementary Fig. S1. Diagnostic plots of k-means clustering.** (a) Within group sum of squares (SS) across clustering with different  $k$ . Grey shaded region highlights number of clusters where within group SS starts decreasing less rapidly with an increase in  $k$ . (b) Silhouette width of clustering with different  $k$ . A silhouette width = 0 indicates points of several clusters overlapping, while a silhouette width = 1 implies a point being exclusive to one cluster. (c) Mean silhouette width across clustering with different  $k$  100 times. Dashed grey line indicates overall mean silhouette width observed.

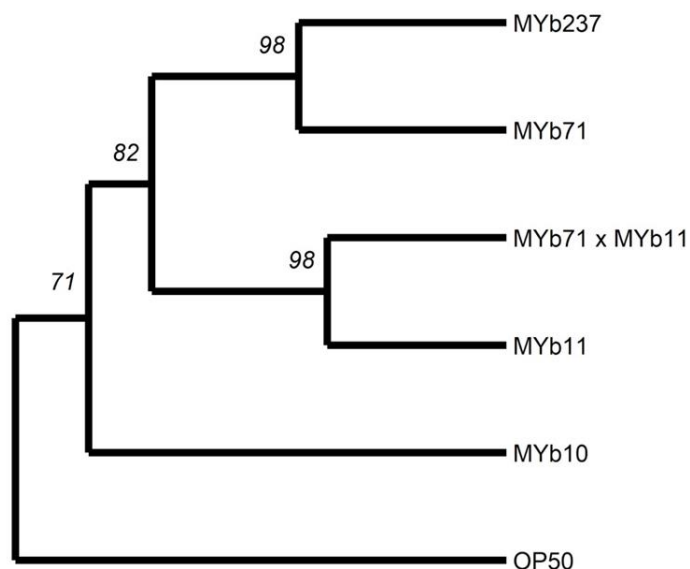


**Supplementary Fig. S2. Genome comparison of selected bacterial taxa using circular plots.** Selected taxa include: (a) *Ochrobactrum*, (b) *Pseudomonas*, (c) *Chryseobacterium*, (d) *Brevundimonas*, (e) *Stenotrophomonas*, (f) *Microbacterium*, (g) *Bacillus*, (h) *Achromobacter*, (i) *Escherichia*, and (j) *Arthrobacter*. For each taxon, the genome sequence with highest quality (i.e., genome sequence with the smallest number of contigs) was identified and then used as a reference for alignment of the remaining genomes. GC content, coverage, and predicted bacterial phage information are given for the reference genome of each taxon. The color intensity in each ring indicates BLAST match identity. The genomes within a particular taxon are usually highly similar, except in the case of *Bacillus*, which produces higher diversity across the included isolates. Predicted phage regions tend to be highly diverse across isolates within a taxon, except for *Chryseobacterium*, *Brevundimonas*, *Stenotrophomonas*, *Achromobacter*, and *Escherichia*, which may be due to the small number of species in these taxa.

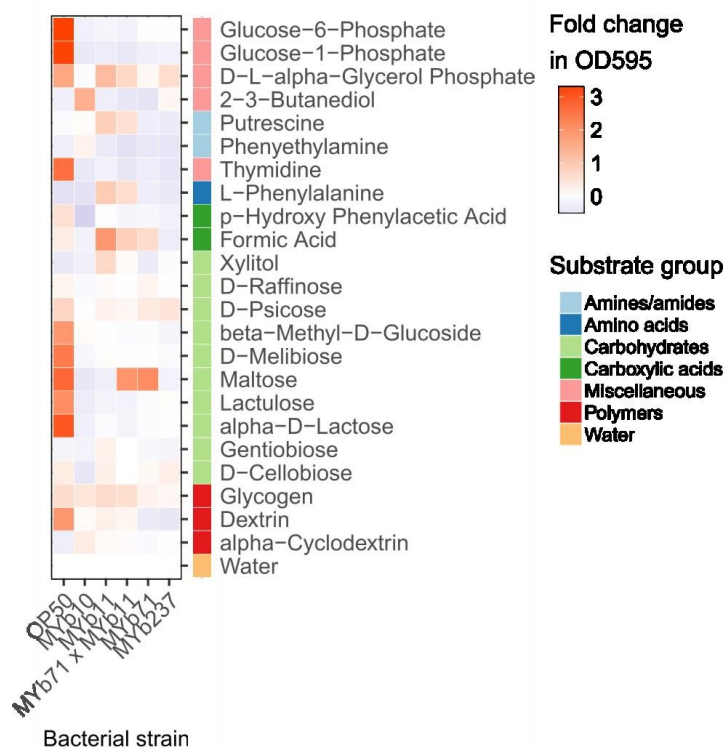


**Supplementary Fig S3. Variation in population growth and colonization levels of *C. elegans* in mono-association with natural microbiota isolates.** Three L4 larvae were exposed to bacterial lawns on PFM plates for five days, and F2 population sizes quantified ( $n = 3-6$ ). To count colony forming units per worm, L4 larvae were transferred from NGM plates with OP50, exposed to microbiota lawns for 24 h, washed and the associated bacteria extracted ( $n = 5$ ). Dashed lines show the mean population size and bacterial load of the canonical food bacteria *E. coli* OP50.

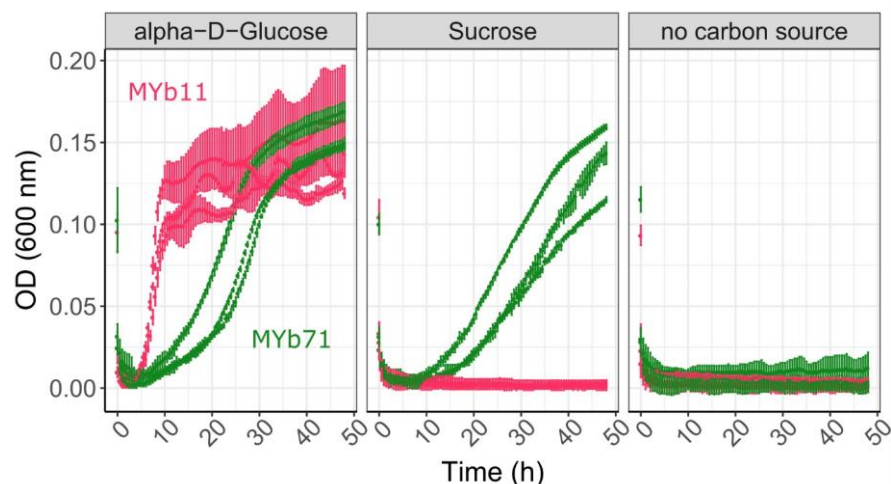




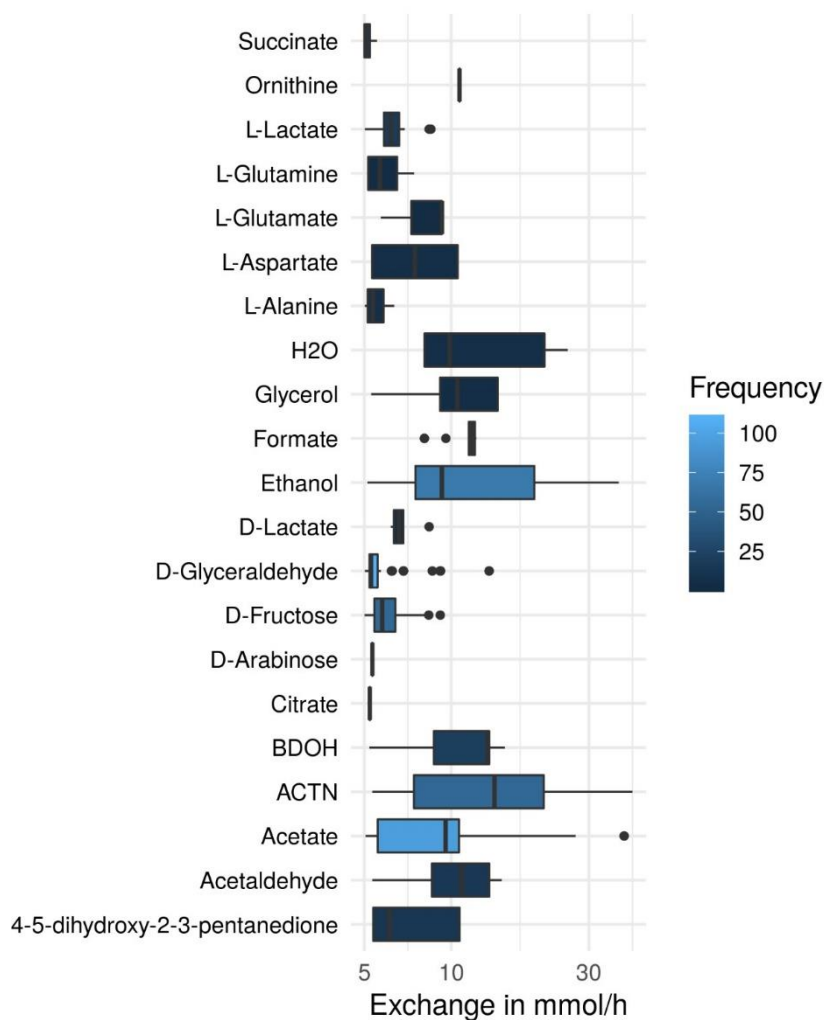
**Supplementary Fig. S4. Hierarchical clustering of strains based on BIOLOG profiles.** Clustering is based on Ward's algorithm and Euclidean distance measures. Bootstrap support (nboot = 1000) is shown on nodes.



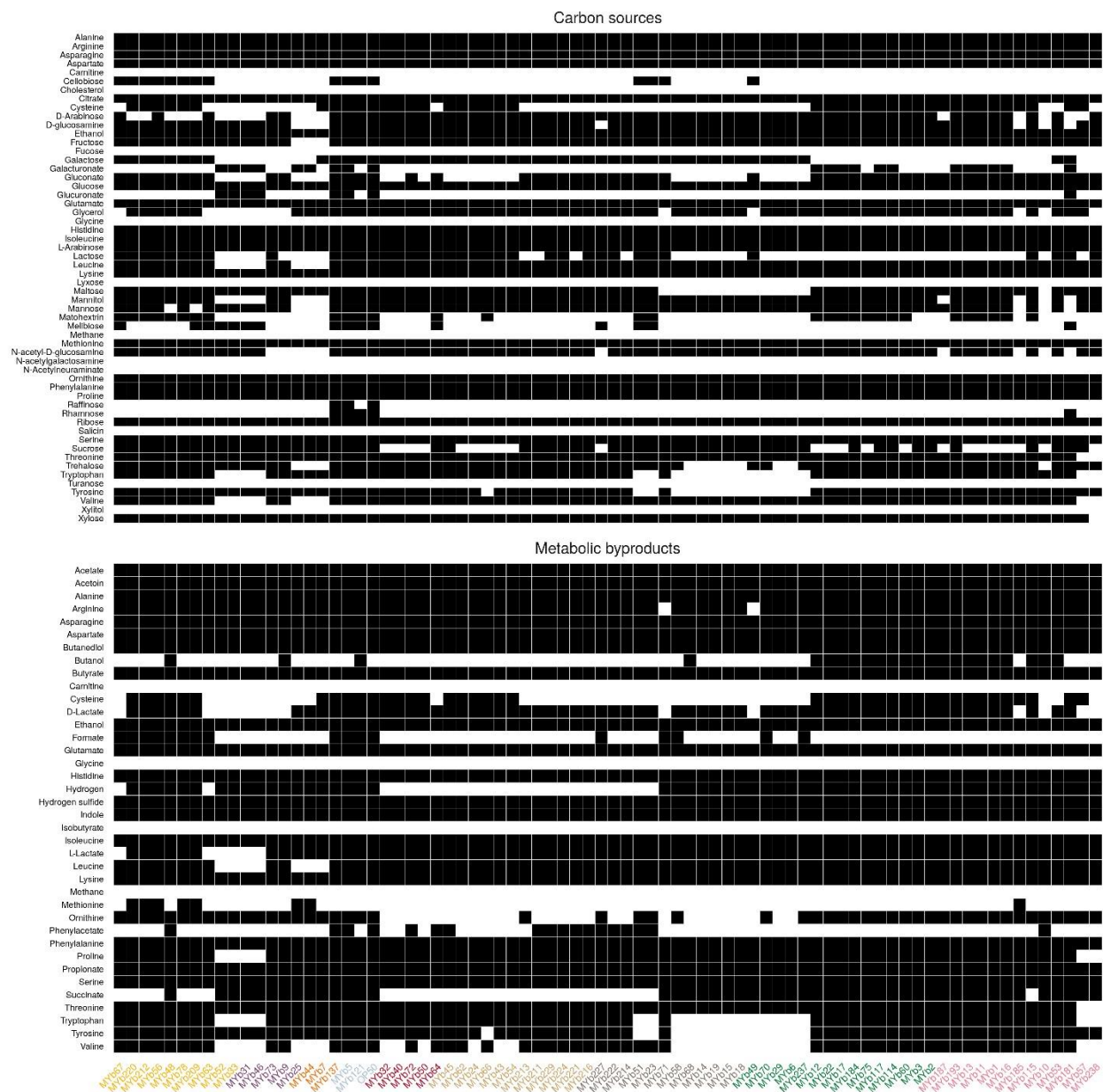
**Supplementary Fig. S5. Cluster 7 of BIOLOG profiling.** Profiles of carbon substrate use of *Acinetobacter* sp. (MYb10), *Pseudomonas lurida* (MYb11), *Ochrobactrum* sp. (MYb71), *Ochrobactrum* sp. (MYb237), and *E. coli* OP50 in BIOLOG GN2 plates over 46 h. The fold-change in indicator dye absorption from 0 to 46 h indicates that the indicated compound is metabolized. K-means clustering ( $k = 7$ ) of substrates by fold-change highlights metabolic differences between strains. Clusters I - VI are shown in Figure 3 in the main text.



**Supplementary Fig. S6. Culture of MYb11 and MYb71 in defined media with single carbon substrates.** Growth of MYb11 and MYb71 in chemically defined media with a single carbon source (i.e., alpha-D-glucose or sucrose), or no-carbon control (over 46 h in 96-well plates).



**Supplementary Fig. S7. Metabolites exchanged in *in silico* pairwise interactions on a minimal medium.** Substances which were exchanged between bacterial isolates in simulation of ecological interactions. In a minimal medium, metabolic byproducts could influence the growth rates of organisms. The figure shows metabolites, which were predicted to be exchanged most frequently and in highest quantity.



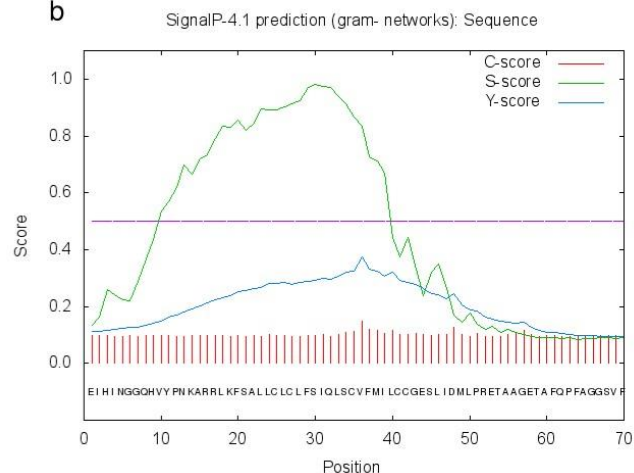
**Supplementary Fig. S9. Prediction of carbon sources and metabolic byproducts.** In simulation of bacteria metabolism, we predicted substances which could be used as carbon sources (left panel) and substances that may be secreted as metabolic byproducts during for example fermentative processes (right panel). Microbiota isolates are given along the x-axis and compounds along the y-axis.

a

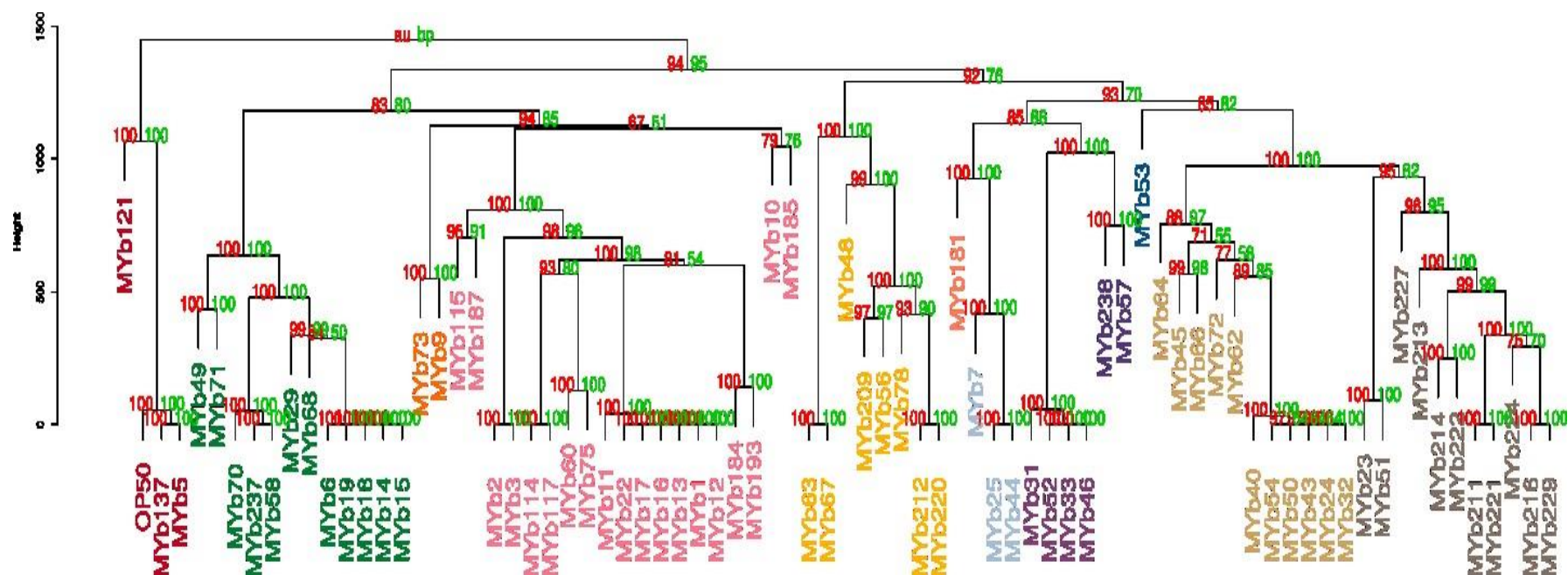
**>Sucrose invertase in MYb71 genome**

EIHINGGQHV\*----YPNKARRLKFSALLCLCLF\*SI\*QLSCV  
 FMILCCGESLIDMLPRETAAG--ETAFQPFAGGVSFNTAIA  
 LGRLDVPTGFFSGISSDFFGEVLRDNLARSNVDYSFAAIS  
 DRPTT-LAFVRL-VDGQARYAFYDENTAGRMLTESDMPY-  
 VDDAIDAMLFGCISLISEPCGSVYEALMT-REAPRRVMFL  
 DPNIRAGFITDREKHLHRMKRMIALADIVKLSDEDLAWF  
 GEKGSHEIAAEWLKLGPKLVVITKGAHGADAYTAKATV  
 RVPGVKVDVVDTVGAGDTVNAGILASLHNQGLLDKDAL  
 VELTEDQIHSVAVALGVRAAAVTVSRAGANPPW

b

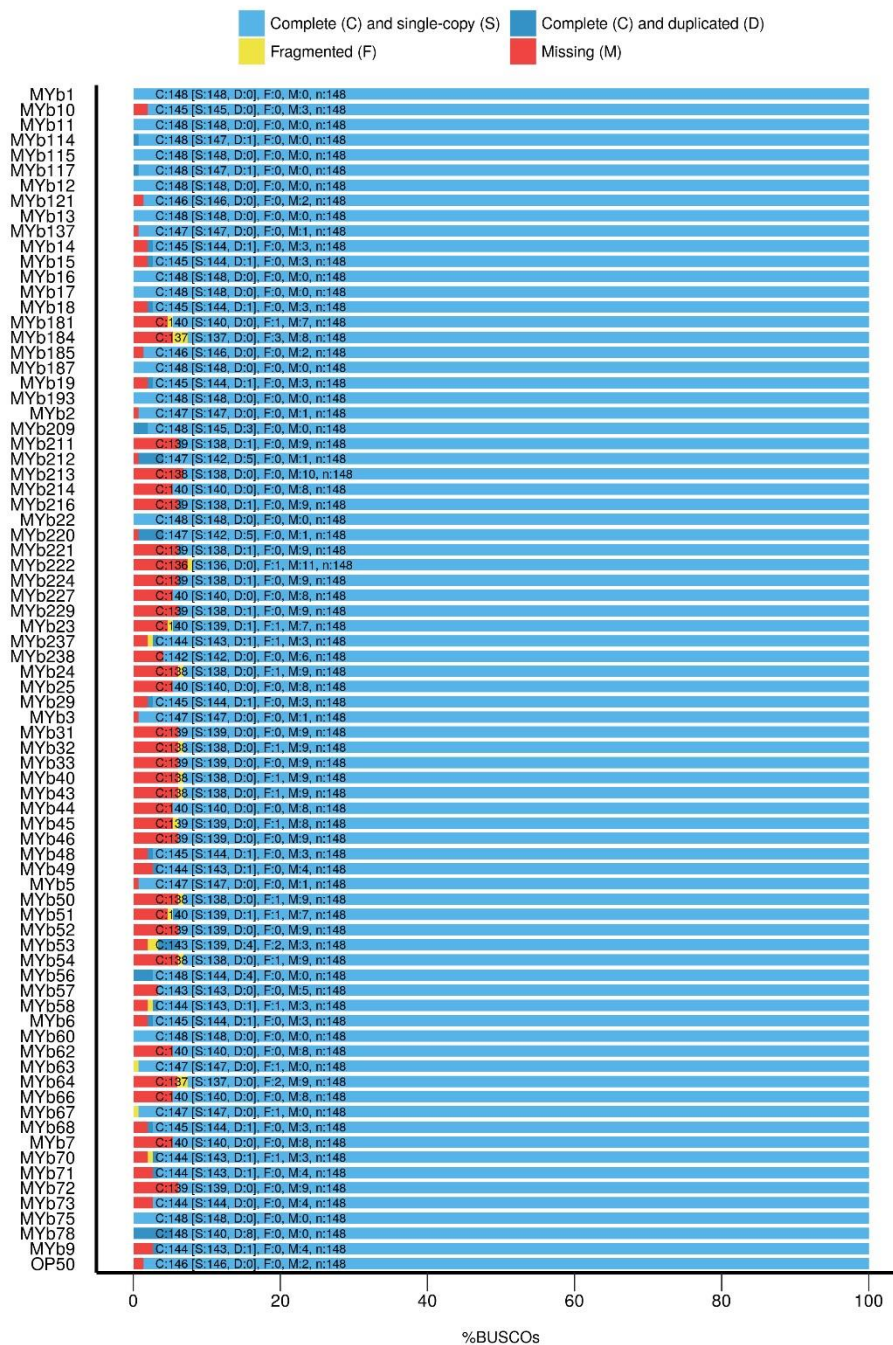


**Supplementary Fig. S10. External sucrose invertase in MYb71.** (a) Amino acid sequence of the sucrose invertase found in the MYb71 genome. (b) We used SignalP to check for secretory signatures. We found that a discrimination score of  $D = 0.518$  (i.e. weighted average of the mean  $S$  and the max.  $Y$  scores) which indicates a signal peptide (signalP: Name=Sequence SP='YES' Cleavage site between pos. 35 and 36: LSC-VF  $D=0.518$   $D$ -cutoff=0.420 Networks=SignalP-noTM).

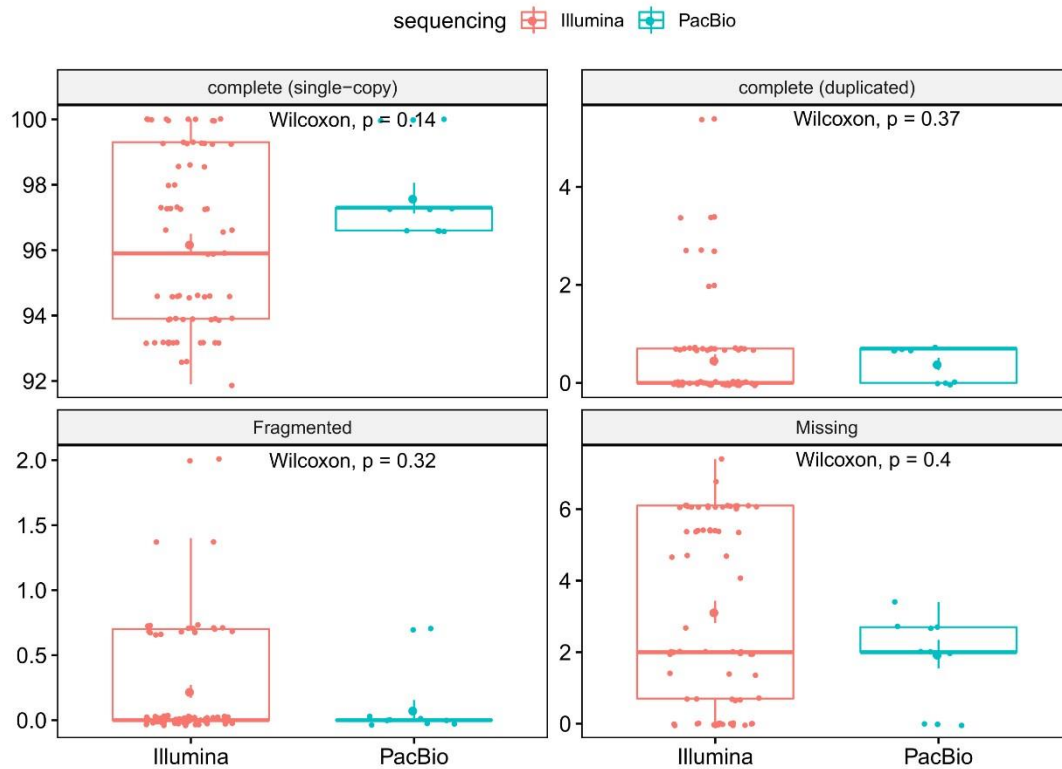


**Supplementary Fig. S11. Hierarchical clustering of metabolic networks based on pathway prediction.** Metabolic networks were clustered according to their pathway completeness score by Euclidean distances and similarity of clusters was estimated by average linkage. The quality of the clustering was tested by multiscale bootstrap resampling. Green values next to branches indicate the bootstrap probability (number of 90 means e.g. that the cluster exists in 90 of 100 runs). In addition to this, the approximately unbiased p-value from multiscale bootstrap is shown in red (see (1) for details).

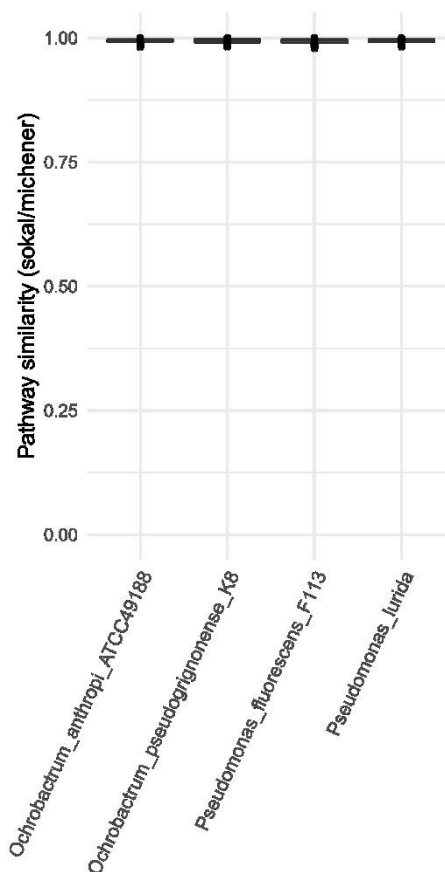
## BUSCO Assessment Results



**Supplementary Fig. S12. Genomic completeness as assessed with BUSCO.** Overview of complete, fragmented and missing matches that were found for all sequenced genomes. The following categories are considered by BUSCO (2): i) complete and single copy: high scoring (i.e.: 90% of the minimum bitscore from an HMM search) and large alignment length matches with one copy; ii) complete and duplicated: high scoring and large alignment length matches with several copies; iii) fragmented: high scoring but shorter alignment length matches; and iv) missing: low scoring matches. The mean completeness level was 96.81% with a standard deviation of 2.65%. For further details, see the BUSCO user guide ([http://gitlab.com/ezlab/busco/raw/master/BUSCO\\_v3\\_userguide.pdf](http://gitlab.com/ezlab/busco/raw/master/BUSCO_v3_userguide.pdf)).

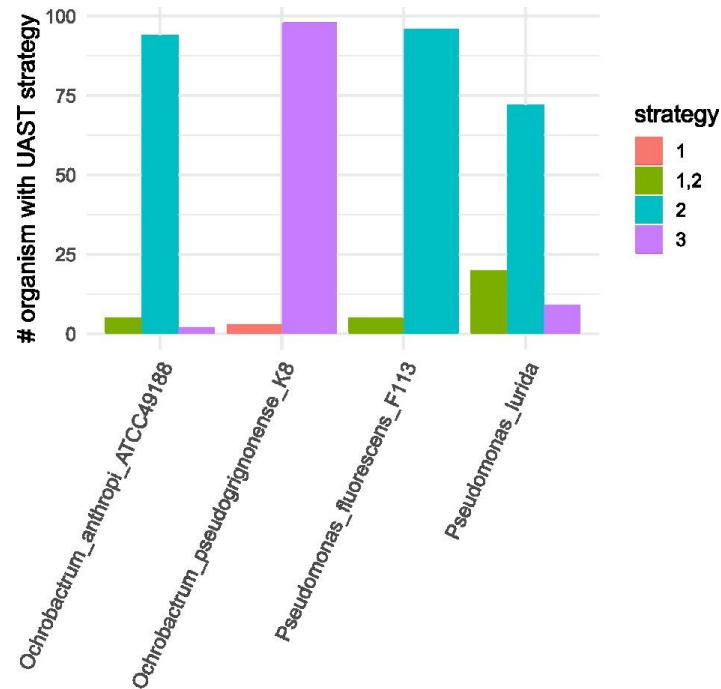


**Supplementary Fig. S13. Comparison of genome completeness for different sequencing techniques.** Complete, fragmented and missing matches found by BUSCO were compared between genomes reconstructed with either Illumina or PacBio technology, in order to assess to what extent sequencing technique influenced draft genome quality. The four BUSCO categories are the same as those explained in the legend to Supplementary Figure S12. The difference in inferred values between Illumina- and PacBio-sequenced genomes was tested with a Wilcoxon rank sum test. None of the comparisons was significant.

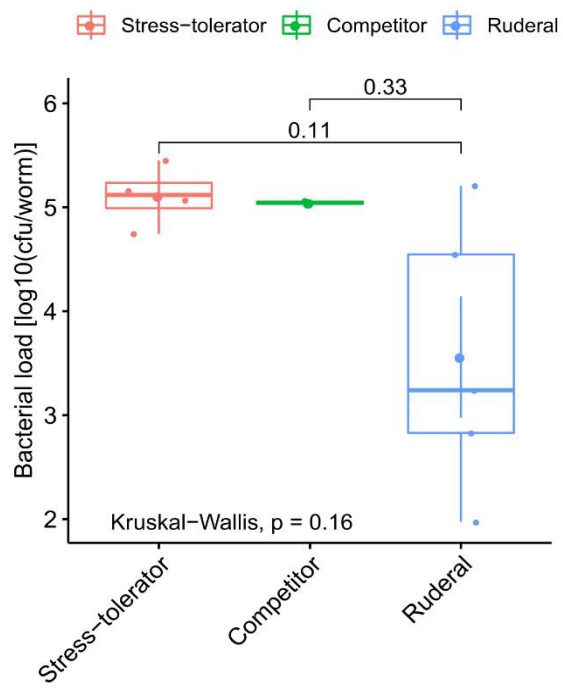


**Supplementary Fig. S14. Pathway similarity of incomplete genomes.** To assess the influence of incomplete genome coverage on inferred metabolic pathways, we downloaded four high quality genomes of our two focal genera (*Pseudomonas* and *Ochrobactrum*) from NCBI. We used the following, published, high quality genomes for this analysis: i) *Ochrobactrum anthropi* ATCC49188 (RefSeq assembly accession numbers: GCF\_001652485.1\_ASM165248v1); ii) *Ochrobactrum pseudogrignonense* K8 (GCF\_000017405.1\_ASM1740v1); iii) *Pseudomonas fluorescens* F113 (GCF\_000237065.1\_ASM23706v1); and iv) *Pseudomonas lurida* (GCF\_002966835.1\_ASM296683v1). We used these genomes to randomly remove sequence chunks. For this, we simulated similar levels of genome completeness as in our data by drawing from a normally distributed sample parameterized by the results of our BUSCO genome analysis (mean = 96.81%, sd = 2.65%; Supplementary Figure S12). The presence of metabolic pathways for randomly shortened genomes was compared to the prediction for the original reference genomes. The overlap ranged from 97.8 to 100%. The overall mean pathway similarity between 400 randomly shortened genomes and their complete counterparts was 99.4%.





**Supplementary Fig. S15. Stability of adaptive strategy predictions. The robustness of UAST classification was tested with randomly shortened representative genomes.** High quality reference genomes were randomly shortened to the same extent as found by BUSCO for the genomes of this study. See legend to Supplementary Figure S14 for more details. For each taxon, 100 shortened genomes were generated and then reclassified according to the universal adaptive strategies (UAST) as done for all other genomes of this study. The bars show how many times a certain strategy was predicted for the incomplete genomes of a particular taxon. The analysis identified the same strategy in at least 94% of shortened genomes for *Ochrobactrum anthropi* ATCC49188, *Ochrobactrum pseudogrignonense* K8, and also *Pseudomonas fluorescens* F113. For *Pseudomonas lurida*, 72% of shortened genomes were consistent with strategy 2 and 20% still with both strategies 1 and 2. A description of the strategies and their definitions is given in Fierer et al. (3).



**Supplementary Fig. S16. Bacterial colonization behavior in comparison to adaptive strategies for Illumina genomes.**

Since sequencing technique could impact rRNA copy number detection, which is used for adaptive strategies classification, we repeated the analysis regarding the colonization potential for Illumina genomes only (69 isolates). Here we found the same qualitative results that competitive and stress-tolerating strategies were associated with higher colonization phenotypic data.

## Supplementary tables

All supplementary tables can be found at <https://www.nature.com/articles/s41396-019-0504-y> and only title legends are reprinted here for reference.

**Table S1: Genome characteristics.** Genomic overview of all bacterial isolates. Based on sequence identity, the closest related species/strains are shown in column B. Genome and assembly statistics (contigs, N50, GC, ...) can be found in columns G:M. The corresponding bioproject to access data via NCBI is shown in column N.

**Table S2: *In silico* TSB-based medium.** Compounds and maximal uptake rates for Tryptic soy broth (TSB) medium used to simulate bacterial growth.

**Table S3: *In silico* glucose minimal medium with thiamine.** Compounds and maximal uptake rates for glucose minimal medium with thiamine used to simulate bacterial growth.

**Table S4: Predicted pathways.** For each bacterial isolate (column A) the predicted presence (column D and column E for with more conservative bitscore cutoff) for all considered metabolic pathways (column C) is given. Here, a number of 1 means present and 0 means not available. In column G the hierarchy (i.e. subsystem) of the pathway is shown.

**Table S5: Predicted virulence.** In this table the presence of virulence traits based on homology with the virulence factor database is shown. In column A the isolates are given and columns B:AV show the presence of virulence factors (1 for true and 0 for false).

**Table S6: Traits with differences in *Ochrobactrum*.** Table consists of metabolic pathways and virulence factors which showed to be significantly changed in isolates belonging to the *Ochrobactrum* genus (based on a Wilcoxon signed rank test). For each trait (column B) a FDR corrected P-value (column C) and the mean for *Ochrobactrum* (column D) and all other isolates (column E) is given.

**Table S7: Experimental phenotypic data: bacterial load and *C. elegans* population growth.** Sheet 1 shows the number of colony forming units per worm (column D) shown across microbiota isolates (column A) and experimental repetitions of the analysis (i.e., runs; column C). Sheet 2 shows the mean number of worms counted (3 samples counted per population) per population of worms (column C) and standard deviation (column D) on the respective microbiota isolates (column A).

**Table S8: Regression analysis to infer metabolic competences associated with bacterial colonization and host fitness.** Two regression approaches were used to find traits which were associated with experimental data of bacterial load in *C. elegans* and the fitness of *C. elegans* when the worm was grown together with the bacterial isolates. In this table, the significant/important traits of both approaches are listed.

**Table S9: Traits and scores used to categorize isolates according to adaptive strategies.** Table of metabolic traits and model features associated with stress-tolerating, competitive, or ruderal strategies. For each isolate the scores and classification for each strategy are listed.

## Supplementary data

Supplementary data can be found at <https://www.nature.com/articles/s41396-019-0504-y> and only file names are reprinted here for reference.

**Supplementary data S1. Genome-scale metabolic models of the microbiota of *C. elegans*.** Zip archive of metabolic models for each isolate in Systems biology markup 222 language (SBML) format. In addition to this, an R file with a list of all models is provided.

## **CHAPTER IV**

### **Transcriptome and proteome analysis of the *C. elegans* response to protective natural microbiota isolates**

Kohar Annie B. Kissoyan, Christian Treitz, Wentao Yang, Eva-Lena Stange, Bente Rackow, Andreas Tholey, Katja Dierking

**Manuscript in preparation**



## ABSTRACT

The natural *C. elegans* microbiota affects many aspects of nematode physiology, including pathogen defenses. However, underlying molecular mechanisms of microbiota-mediated protection against pathogen infection remain elusive. To fill this knowledge gap, we used a multi-omics approach to assess the *C. elegans* response to the protective natural microbiota isolates *Pseudomonas lurida* MYb11 and *P. fluorescens* MYb115 on transcript and protein level, in the presence and absence of the pathogen *Bacillus thuringiensis*.

Our results indicate that the protective microbiota isolates influence the worm's innate immune response and cellular structural component functions. Our findings suggest that the natural microbiota isolates modulate *C. elegans* defense response, even in the absence of a pathogen, preparing the host for subsequent pathogen attacks. Enrichment analyses on co-regulated genes indicate an involvement of the insulin/insulin-like growth factor and p38 and JNK MAPK pathways in microbiota-mediated protection. A comparison of the transcriptome and proteome data revealed overlaps between differentially expressed genes and their corresponding protein abundances for some galectins, C-type lectins, and lysozymes, suggesting their potential roles in microbiota-mediated protection.

Overall, our study reports the inducible responses of *C. elegans* to its protective microbiota, suggesting microbiota-mediated activation of the innate immune response via priming the host. Our findings also provide a framework at both the transcript and protein level and a list of candidate genes for further functional investigation of *C. elegans* host-microbiota-pathogen interactions.

**KEYWORDS:** *C. elegans*, multi-omics, microbiota, transcriptomics, proteomics, RNA-Seq

## INTRODUCTION

Multicellular hosts are closely associated with their resident microorganisms, the microbiota, which plays a vital role in governing host health and disease states (Ezcurra, 2018; Haller, 2018). Due to such close associations, the host and its microbiota can be seen as one functional entity, the metaorganism (Bosch & McFall-Ngai, 2011). In the metaorganism, the microbiota affects host biological functions ranging from digestion and metabolism to development, reproduction, aging, and even behavior (Cabreiro & Gems, 2013; Nicholson et al., 2012; Sampson & Mazmanian, 2015; Sison-Mangus et al., 2015). Another essential function of the microbiota is protecting the host from pathogen infection (Caragata et al., 2013; Hamilton et al., 2016; Krediet et al., 2013; Ripert et al., 2016). Strong evidence of microbiota-mediated protection against pathogen infection comes from observations in germ-free mice (i.e., completely lacking gut microbiota) and antibiotic-treated mice (i.e., with disturbed gut microbiota), which are more susceptible to pathogen infection compared to their conventionally raised (i.e., with intact microbiota) counterparts (Kennedy et al., 2018; Sassone-Corsi & Raffatellu, 2015; Vogt & Finlay, 2017). Microbiota-mediated protection is of paramount interest in medical applications as a therapeutic agent used to restore a dysbiotic microbiota, particularly in treating diseases associated with altered gut microbiota (Khoruts, 2018; Rolhion & Chassaing, 2016). However, the molecular mechanisms underlying microbiota-mediated protection are often unclear, thus hindering its optimal use in therapeutics (Khoruts, 2018; Lemon et al., 2012; Vogt & Finlay, 2017).

Microbiota studies using animal experiments and new technological advances have elucidated microbiota's complexity, highlighting the benefits of choosing an approach and a model organism that allows untangling such complexity (Sassone-Corsi & Raffatellu, 2015). So far, less complicated experimental designs that use a reductionist approach, with binary (i.e., one host and one microbe) and gnotobiotic conditions, are most successful in advancing our understanding of the underlying processes of specific host functions (Fischbach, 2018). Moreover, simple animal models aid in further untangling the complex host-microbiota interactions since they support protocols allowing for the reproducible manipulation of microbiota and assigning function to individual taxa (Douglas, 2019). One such simple yet powerful and cost-effective model is the nematode *Caenorhabditis elegans*. *C. elegans* is a tiny, transparent, and genetically amenable worm, in which standardized protocols allow the manipulation of the host and its bacteria, maintaining a gnotobiotic host (Sterken et al., 2015).



Thus, *C. elegans* serves as an ideal model for host-microbiota-pathogen interaction studies. *C. elegans*, like other animal models, has been used in gene expression profiling studies (with microarrays or RNA-seq), where gene expression patterns of the worms are compared in various sex, age, development, tissue, diet, or stress-related treatments (Baugh et al., 2009, Bulcha et al., 2019; Deng et al., 2011; Reinke et al., 2000; Schwarz et al., 2012). Moreover, the *C. elegans* natural microbiota was characterized recently, allowing the study of microbiota function (Berg et al., 2016a; Dirksen et al., 2016, 2020; Johnke et al., 2020; Samuel et al., 2016; Zhang et al., 2017).

Recent exploration of *C. elegans* in its natural habitat has provided new insights into microbiota-mediated protection against pathogen infection. Bacterial isolates belonging to the genera *Enterobacter*, *Gluconobacter*, and *Pseudomonas* have been shown to enhance *C. elegans* immune defenses against pathogens (Berg et al., 2016b; Dirksen et al., 2016; Kissoyan et al., 2019; Montalvo-Katz et al., 2013; Samuel et al., 2016). Microbiota-mediated protection can be achieved via direct (host-independent) or indirect (host-dependent) processes (Buffie and Pamer, 2013; Pickard et al., 2017). We have previously shown that *C. elegans* natural microbiota isolates *Pseudomonas lurida* MYb11, and *Pseudomonas fluorescens* MYb115 protect the worm against *Bacillus thuringiensis* infection. *P. lurida* MYb11 protects worms from *B. thuringiensis* infection directly via the production of the antimicrobial compound massetolide E (Kissoyan et al., 2019). However, there is no evidence to show that the production of massetolide E is the exclusive protection mechanism of MYb11. Also, MYb115 does not produce massetolide E, yet it protects the worms against *B. thuringiensis* infection (Kissoyan et al., 2019). The exact mechanism of MYb115-mediated protection in the worms is not yet known. Thereby, how the worm is affected during microbiota-mediated protection against a pathogen infection with Bt is not yet entirely clear.

Here, we used omics analysis to study microbiota-mediated indirect protection. Previous studies using omics approaches with *C. elegans* have shown microbiota-induced effects on the worm; A transcriptome and proteome analysis of the *C. elegans* response to natural microbiota isolates of the genus *Ochrobactrum* revealed an effect of the microbiota on dietary response, development, fertility, immunity and energy metabolism at both transcript and protein levels (Cassidy et al., 2018; Yang et al., 2019). *Stenotrophomonas maltophilia* strains, both pathogenic and non-pathogenic, induce distinct changes in the nematode transcriptional responses, revealing strain-specific molecular responses (Radeke & Herman, 2020). In the

context of microbiota-mediated protection, using transcriptomics, *B. subtilis* was shown to differentially regulate metabolic pathways, such as sphingolipid metabolism (Goya et al., 2020). Further, the authors show that *Bacillus subtilis* indirectly protects *C. elegans* from the pathological accumulation of  $\alpha$ -Synuclein via the insulin/insulin-like growth factor-1 (IGF-1) signaling pathway and the modulation of the host sphingolipid metabolism (Goya et al., 2020).

The current study's objectives were to identify host genes and proteins that are a) induced in *C. elegans* upon microbiota exposure, b) involved in microbiota-mediated protection against pathogen infection. To this end, we focused on the two protective microbiota isolates, *P. lurida* MYb11 and *P. fluorescens* MYb115 and the pathogen *B. thuringiensis*. To assess the influence of microbiota isolates on host gene expression and protein abundance both in the presence and absence of pathogen infection, we performed whole-genome transcriptome and proteome analysis after microbiota and pathogen exposure. We performed enrichment analyses of the transcriptome data using the Gene Ontology and the *C. elegans* specific gene-expression database WormExp to pinpoint the underlying biological functions involved in microbiota-mediated protection. We also assessed the overrepresented transcription factor binding sites to characterize better the involved regulatory processes. Finally, we complemented transcriptome with proteome dataset to identify overlaps. Based on this combined dataset, we used the connectivity within a gene network model to rank candidate genes involved in the microbiota-mediated protection for guiding future functional analyses (**Figure 1**).

## MATERIALS AND METHODS

### *C. elegans and bacterial strains*

Worm strains used in this study were the *C. elegans* N2 strain and the mutant *C. elegans* strains *pmk-1(km25)* and *mek-1(ks25)*. Worm strains were initially provided by the CGC (Caenorhabditis Genetics Center), funded by NIH Office of Research Infrastructure Programs (P40 OD010440). All worm strains were thawed, grown, bleached, and maintained on nematode growth medium at 20°C according to the routine protocol (Stiernagle, 2006). Before the transcriptome and proteome experiments, worms were maintained for at least two generations on Peptone-free nematode growth media (PFM), with lawns prepared from each of the food bacteria MYb11, MYb115, and OP50 adjusted at an OD<sub>600</sub>=10.

Bacterial strains used in this study were *Pseudomonas lurida* MYb11 and *Pseudomonas fluorescens* MYb115, which are members of the native microbiota of *C. elegans* (Dirksen et al., 2016; Kissoyan et al., 2019), and the standard laboratory *C. elegans* laboratory food bacterium *Escherichia. coli* OP50. All bacterial strains were grown and maintained on Tryptic Soy Agar (TSA) plates at 25°C, and in Tryptic Soy Broth (TSB) at 28°C, according to a previously published protocol (Kissoyan et al., 2019).

*Bacillus thuringiensis* (MYBt247 and MYBt407) spore aliquots were obtained following a previously established protocol (Borgonie et al., 1995). The Bt spore aliquots were prepared in advance in huge quantities and stored at -20°C, with the following concentration ranges: 10<sup>9</sup>-10<sup>10</sup> particles/ml for the pathogenic Bt247, and 10<sup>3</sup>-10<sup>4</sup> particles/ml for non-pathogenic Bt407. Before every assay, new Bt aliquots were thawed and used.

### *Survival assay*

A survival assay was performed to confirm the microbiota-mediated protective phenotype in the collected samples parallel to the omics experiment sample collections. Before the survival assay, *C. elegans* was freshly thawed and grown with either MYb11, MYb115, or OP50, to allow worms to adapt to the growth conditions. Worms were then bleached, synchronized, and L1s grown on the specific bacterial lawn, from where at least 30 L4s were then transferred to the respective infection plates. The survival assays with the pathogen *B. thuringiensis* were done according to a standardized protocol described previously, with four technical replicates in each treatment (Kissoyan et al., 2019). Briefly, we used different concentrations of spore-

toxin mixtures of *Bacillus thuringiensis* to prepare lawns and infect worms as described earlier. We used the non-pathogenic Gram-positive *B. thuringiensis* strain Bt407 and the pathogenic strain MYBt18247 (Bt247) mixed with an OD of 10 of each of OP50, MYb11 or MYb115 at concentrations 1:100, 1:200, and no Bt (i.e., where worms were transferred to plates containing only the respective bacterial treatment; OP50, MYb11 or MYb115). Worms were assessed 24 hpi as alive or dead using gentle touching with a sterile platinum wire. We performed survival assays to confirm the protective effect of MYb11 and MYb115, based on which we chose the samples with which we proceeded with RNA-sequencing analyses.

Similarly, survival assays were performed for each of the mutant worm strains *pmk-1(km25)* and *mek-1(ks25)*.

### *Transcriptome analysis using RNA-seq*

Worms were grown on 9cm PFM plates with 600 $\mu$ l of bacterial lawn (OD<sub>600</sub>=10) of MYb11, MYb115, or OP50. For each replicate, a total of 1000 to 1500 synchronized hermaphrodites at the first larval stage (L1) were pipetted onto the bacterial lawns. To ensure enough worms are available even after the worm washing steps, at least 20 plates of each condition were prepared, and worms were pooled at the transfer and harvesting time-points. Worms were grown and maintained at 20°C and transferred two days later to the infection plates containing the respective lawns (i.e., MYb11, MYb115, and OP50) they were initially grown on, mixed with either the nematicidal Bt247 or the non-nematicidal Bt407 at a concentration of 1:200, 1:100, or none. The worms were harvested at two time-points, 12 hpi and 24 hpi. We prepared four completely independent replicates and included three of them for further analysis.

Worms were washed from the plates with M9T (M9 and 0.02% Triton) and centrifuged to obtain a pellet, followed by three washing steps. The supernatant was discarded, and 700 $\mu$ l TRIzol (Thermo Fisher Scientific, Waltham, MA, United States) was added. Samples were snap-frozen in liquid nitrogen by four freeze (liquid nitrogen)-thaw cycles (46°C thermo-shaker) to break open the nematode cuticle. Next, the samples were frozen and stored at -80°C until total RNA extraction, using the kit NucleoSpin RNA kit (Macherey-Nagel, Düren, Germany). All assays were performed without knowledge of bacterial strains; we used randomized codes to avoid possible experimenter bias during laboratory work.

RNA libraries were prepared for sequencing analysis using standard protocols (TruSeq RNA v2; catalog number RS-122-2001). Illumina HiSeq 2000 was used as a sequencing machine

with a paired-end strategy at a read length of 100 nucleotides. All raw data is available from the GEO database under the GSE number GSE136942.

The transcriptome data analysis was performed according to the in-house RNA libraries and pipeline (Yang et al., 2019). RNA-Seq reads were firstly trimmed for adaptor sequence, masked for low-quality sequence via Trimmomatic (Bolger et al., 2014) and then mapped to the *C. elegans* genome (Wormbase version WS235; www.wormbase.org) by STAR 2.5.3a (Dobin et al., 2013) under the default setting. Transcription abundance (read counts per gene) is extracted via HTSeq (Anders et al., 2015). Differential expression analysis was performed by aFold from ABSSeq (Yang et al., 2016a). Genes with a significant change between different conditions (adjusted  $p$ -value  $< 0.01$ ) were treated as a signature for each comparison. The  $\log_2$  transformed fold-changes were taken as input for k-means cluster analysis using cluster 3.0 (de Hoon et al., 2004). A heat map was generated by TreeView version 1.1.4r3 (Saldanha, 2004).

#### *Gene ontology, Gene set, and Motif enrichment analysis*

Gene Ontology (GO) analysis was performed with the DAVID Bioinformatics Resources 6.8 (Huang et al., 2009). Genes were assigned to one or more GO terms, categorized into Biological Process, Molecular Function, and Cellular Component. Genes are in other cases assigned to INTERPRO database terms. Significant enrichment of the GO terms was calculated using Fisher's exact test. FDR correction was used to correct for multiple testing. GO terms with FDR adjusted  $p$ -value  $< 0.05$  were considered significant.

Gene set taxon-specific enrichment was performed with WormExp (Yang et al., 2016b). WormExp is a web-based analysis tool specifically for *C. elegans*, thereby allowing the characterization of species-specific gene expression patterns. Gene sets with FDR adjusted  $p$ -value  $< 0.05$  were considered significant.

Motif enrichment analysis was performed on the promoter regions, from -600bp to 250 bp, relative to the transcription start sites, of the genes in each cluster. AMD was used to perform the *De novo* motif analysis (Shi et al., 2011).

#### *Proteome analysis using 1D/LCMS*

Worms were prepared similarly to the transcriptome experimental design described above. However, here we used four completely independent replicates for each treatment group and harvested the worms at one time-point, 12 hpi. Moreover, for harvesting, worms were washed

off the plates with M9, using a Steriflip system with a 20  $\mu\text{m}$  mesh filter (Merck, Darmstadt, Germany) (Cassidy et al., 2018). Worms were centrifuged to obtain a pellet, which was then transferred to a cryotube and snap-frozen with liquid nitrogen, and then stored in the -80 freezer until further processing.

200  $\mu\text{l}$  protein lysis buffer (100 mM Triethylammonium bicarbonate TEAB, 2% SDS, 5M Guanidinium chloride 2mM Dithiothreitol DTT; 2x Complete protease inhibitor) and 200  $\mu\text{l}$  of acid-washed glass beads were added to the samples for protein extraction. The samples were homogenized using a Bioruptor pico for 20 cycles of 30 s sonication and 30 s cooling at four  $^{\circ}\text{C}$ . Then, the protein concentration was determined by the BCA assay. The samples were reduced with ten mM DTT for one h at 60  $^{\circ}\text{C}$  and alkylated with 25 mM Chloroacetamide at 20  $^{\circ}\text{C}$  for 20 min. The samples were centrifuged for 10 min at 10,000 g, and aliquots of 100  $\mu\text{g}$  were prepared following the SP3 protocol (Hughes et al., 2019).

Digested peptides were analyzed using 1D LC-MS (1Dimensional Liquid Chromatography-Mass Spectrometry). A Dionex U3000 HPLC system was coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The digested samples were injected on a C18 PepMap 100 precolumn (column dimensions: 300  $\mu\text{m}$  i.d. x 5 mm; Thermo Scientific) with a flow rate of 30  $\mu\text{L}/\text{min}$ , trapped and desalted for 2 min and then separated on an Acclaim PepMap RSLC column (column dimension: 75  $\mu\text{m}$  i.d. x 50 cm; Thermo Scientific) over a gradient of eluent A (0.05% aqueous FA) and eluent B (80% ACN, 0.04% FA) with a flow rate of 0.3  $\mu\text{L}/\text{min}$ .

Following the injection of 5  $\mu\text{L}$  of the sample, peptides were eluted over a gradient from 5 % eluent B to 20 % eluent B in 27.5 min, then to 40 % B in 35 min, followed by an increase to 95 % eluent B in 20 min. After isocratic elution at 95 % eluent B for 18 min, the column was equilibrated for 10 min with 5 % eluent B. After each sample LC run, the column was washed using a blank run, injecting 5  $\mu\text{L}$  of loading buffer.

Full MS scans were acquired from 2 min to 200 min in positive ion mode with a resolution of 120,000, the AGC target was set to 4e5, the maximum injection time was 50 ms, and the scan range was from 400 m/z to 1100 m/z. Data-dependent FT-MS/MS spectra of the most intense precursor ions were acquired for 3 s with a resolution of 7,500; The isolation window was set to 1.2 m/z, the normalized collision energy was 30, the AGC target was 1e5 and a maximum injection time of 50 ms. Precursors with charge states of 2-7 were accepted, and isotopes were excluded. Dynamic exclusion was set to 60 s with a mass tolerance of 10 ppm. Apex detection

properties were set to an expected peak width (TWHM) of 25 s and the desired apex window of 35 %.

For the database search, the resulting spectral data were searched against a combined database downloaded from UniProt (08.07.2019), which included all the proteins of the different bacteria used on culture plates and those from *C. elegans*. MS raw files were searched against protein databases of target organisms and common contaminants using the Sequest search algorithm and the Proteome discoverer software (Thermo Scientific).

Label-free quantification was done using Minora Feature Detector with the following parameters; the minimum trace length was set to 4, the minimum number of Isotopes was set to 2 peaks, and the maximum  $\Delta$  RT of isotope pattern multiplets was set to 0.2 min. Unique and razor peptides were used for quantification. The precursor abundance was based on the peak area and normalized to the sum of all intensity values. The normalized abundance values were scaled so that the average abundance is 1, meaning that the sum of all valid values is equal to the sum of intensity values acquired for any protein.

### *Transcriptome and proteome combined dataset analysis*

We combined the transcriptome dataset with the corresponding proteome dataset for each condition by filtering the significant fold changes (FDR *p-value* < 0.01) of the different treatments in the proteome dataset and matching it with the respective transcriptome dataset based on the WormBase IDs. We only considered genes that correspond to proteins with significantly differential abundances, compared to the control worms. Thus, the combined dataset includes the log<sub>2</sub> transformed fold change of each treatment condition (OP50, MYb11 and MYb115 and Bt combinations) compared to the control worms (i.e., worms grown on only OP50), at the time-point 12 hpi. We maintained the same clustering (six clusters) as the transcriptome level. The heatmap from the combined dataset was visualized using a heatmap.2(gplots) function in R (version 3.6.1). Gene Ontology(GO) analysis was performed on each cluster with the DAVID Bioinformatics Resources 6.8 (Huang et al., 2009). Cluster 2 genes were then extracted from the combined list (51 genes) and queried in WormNetv2, a probabilistic functional gene network model, to rank the genes (Lee et al., 2010). Gene ranking was based on score, calculated from the number of connections and the strength of evidence from these connections. Cytoscape 3.8.2 was used to visualize the gene network (Shannon et al., 2003).

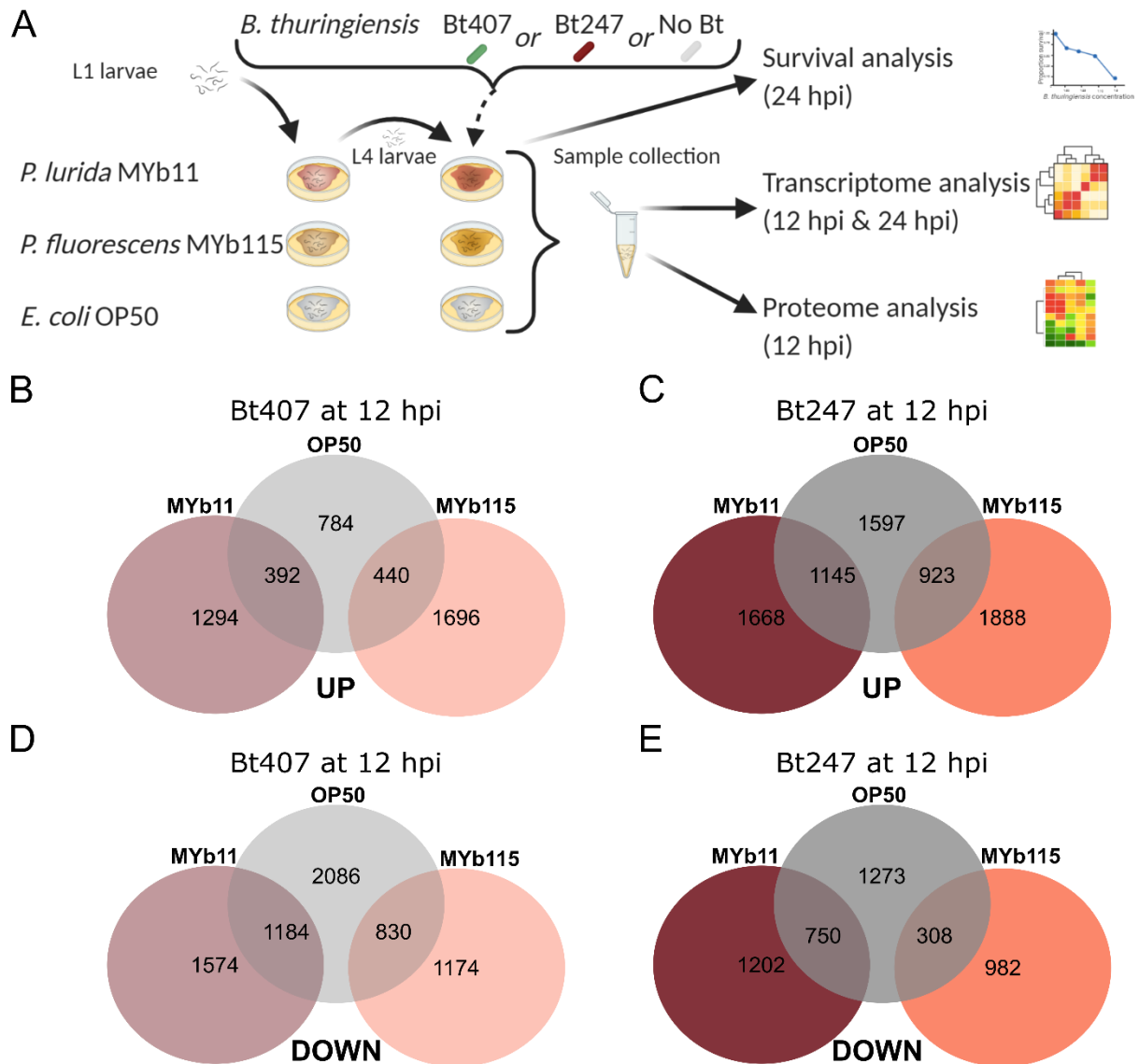
## RESULTS AND DISCUSSION

### *Microbiota isolates P. lurida MYb11 and P. fluorescens MYb115, and pathogen exposure determine C. elegans transcriptional profile*

We used RNA-Seq to assess the *C. elegans* transcriptome response to the protective microbiota isolates *P. lurida* MYb11 and *P. fluorescens* MYb115 in both the presence and absence of the Gram-positive bacterial pathogen *Bacillus thuringiensis* MYBt18247 (Bt247) 12- and 24-hours post-infection (hpi) (**Figure 1A**). We used the standard laboratory food bacterium *Escherichia coli* strain OP50 and the non-pathogenic *B. thuringiensis* strain MYBt407(Bt407) as control groups for the microbiota and pathogen treatment, respectively. In parallel, we performed survival assays and confirmed that MYb11 and MYb115 protect the worm against Bt247 infection (Kissoyan et al., 2019) (**Supplementary Figure 1**).

First, we used a t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis to explore transcriptomic variation across treatments (**Supplementary Figure 2**). We found that time and infection are factors with strong effects on the transcriptional response. Most samples cluster together according to time-points (12 hpi versus 24 hpi). Also, most samples from worms infected with Bt247 cluster together and are clearly distinct from samples from worms treated with non-pathogenic Bt407 or the *E. coli* OP50 control. The exceptions are samples from Bt247-infected worms on MYb115, which surprisingly cluster together with samples from uninfected worms on the control treatments (OP50 and Bt407) and are thus clearly distinct from infected worms on OP50 or MYb11. There are three possible explanations for such a distinct cluster of Bt247-infected worms on MYb115: First, worms grown on MYb115 are less infected than those on MYb11 and OP50. However, this suggestion is unlikely since we previously showed that pathogen load in Bt247-infected worms grown on MYb115 is as high as in worms on OP50 (Kissoyan et al., 2019). Second, MYb115 affects pathogen avoidance behavior by increasing lawn leaving or decreasing oral bacterial uptake; however, this requires further experimental validation. Third, the transcriptomic response to infection is repressed in infected worms on MYb115; below, we describe further indications for this response.





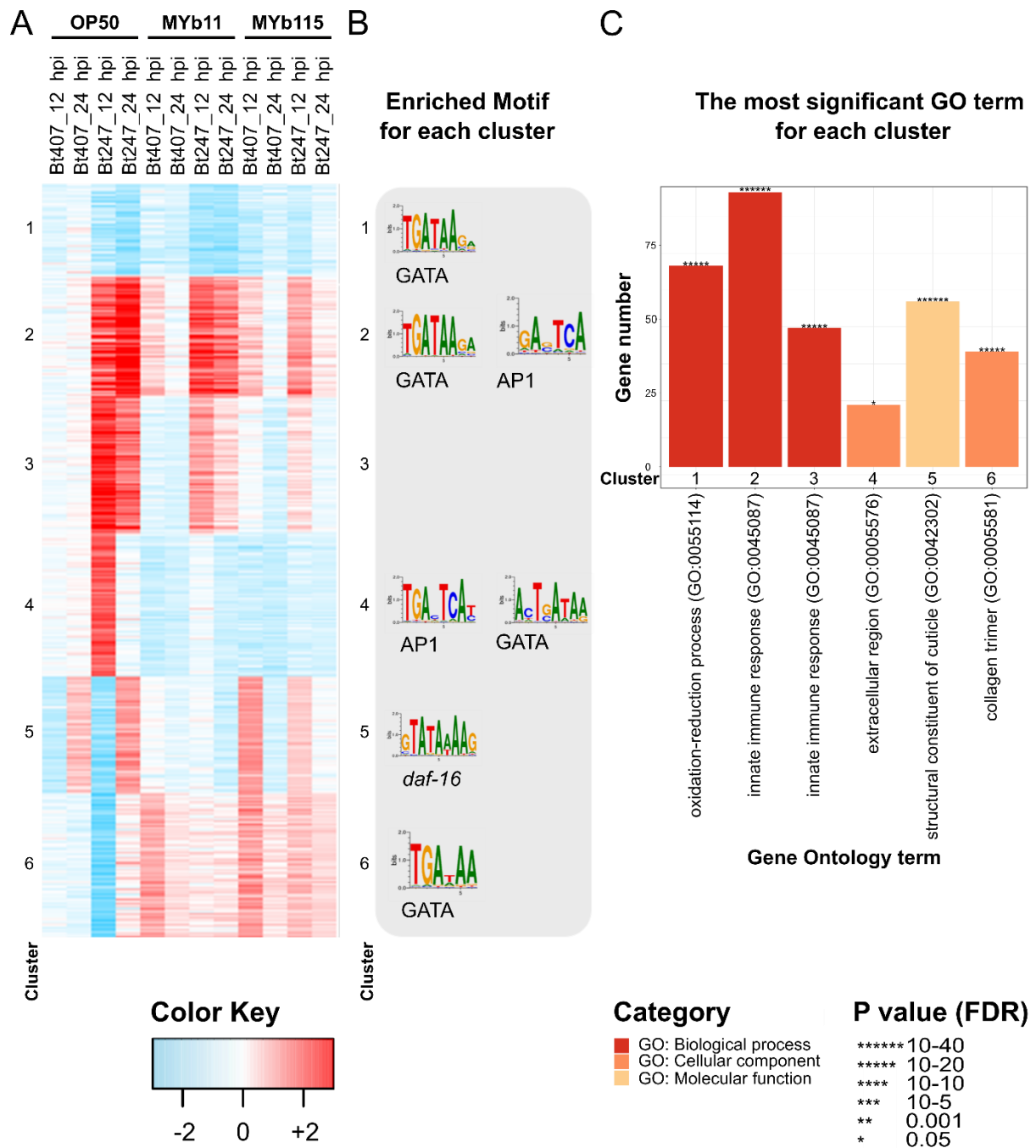
**Figure 1: Workflow of the multi-omics approach to study microbiota-mediated protection in *C. elegans* and the number of up-regulated, down-regulated, and co-regulated genes in each treatment condition compared to the control worms at 12 hpi.**

(A) Worms were grown on the microbiota isolates *P. lurida* MYb11 or *P. fluorescens* MYb115 or the standard laboratory food bacterium *E. coli* OP50 as a control, from the L1 larval stage to L4 larval stage. Worms were then infected at the L4 stage with pathogenic Bt247 or exposed to the non-pathogenic Bt407 or transferred to fresh plates with the respective bacteria (MYb11, MYb115 or OP50 with no Bt). We did survival analyses at 24 hpi and collected samples for transcriptome (12 and 24 hpi) and proteome analyses (12 hpi). This figure was created with Biorender.com. (B-E) Venn diagrams show the numbers of significantly up-regulated (A and B) and down-regulated (C and D) genes in worms grown on MYb11, MYb115, and OP50 and exposed to the non-pathogenic Bt407 (A and D) or infected with the pathogenic Bt247 (B and D) at 12 hpi. All conditions are compared to worms grown only on OP50 (no Bt). (See *Supplementary Figure 3* for the Venn diagrams at 24 hpi.)

Next, we assessed the transcriptional response in each treatment group (i.e., worms grown on each of OP50, MYb11 or MYb115 and exposed to non-pathogenic Bt407 or infected with the pathogenic Bt247) versus the non-infected control group (i.e., worms exposed only to OP50). We found that a total of 2870 genes were significantly differentially expressed in at least one of the treatment groups compared to the control (false discovery rate (FDR) adjusted p-value < 0.01). We observed a higher number of up-regulated genes in worms grown on the microbiota isolates MYb11 and MYb115, upon early exposure (12 hpi) to the non-pathogenic Bt407 (**Figure 1B**) and early infection (12 hpi) with the pathogenic Bt247 (**Figure 1C**) when compared to worms grown on OP50. Subsequently, the down-regulated gene numbers were higher in worms grown on OP50 and exposed to the non-pathogenic Bt407 or infected with the pathogenic Bt247 at 12 hpi, compared to the Bt407-exposed or Bt247-infected worms grown on the microbiota isolates (**Figure 1D and 1E**). Conversely, at the later time-point (24 hpi), the up-regulated gene numbers were higher in the worms grown on OP50 and exposed to the non-pathogenic Bt407 or infected with the pathogenic Bt247 compared to the Bt407-exposed or Bt247-infected worms grown on the microbiota isolates MYb11 and MYb115 (**Supplementary Figure 3**). This profile indicates an early onset activation of gene expression, even in the absence of pathogen, on worms grown on microbiota isolates suggesting microbiota-mediated heightened preparedness against current and potential pathogen attacks.

We then used K-means clustering on the differentially regulated gene-sets and identified six clusters of co-regulated genes (**Figure 2A**). Of the six clusters, we focused on clusters 2, 5 and 6. Cluster 2 contains genes that are strongly up-regulated by Bt247 infection in control worms on *E. coli* OP50 and by exposure to MYb11 and MYb115 at the early time point (12 hpi) already in the absence of a pathogen. Cluster 5 contains genes that are up-regulated early (12 hpi) in worms grown on MYb115, up-regulated late in worms grown on OP50, and not differentially expressed in worms on MYb11, thus showing a distinct response in worms to MYb115. Cluster 6 contains up-regulated genes in worms on both microbiota isolates, most strongly early (12 hpi), while they are not up-regulated in worms on OP50. The gene expression pattern in cluster 6, i.e., upregulation on worms grown on the microbiota isolates, implies that the microbiota isolates induced a panel of genes, which otherwise were not induced in the OP50-grown worms, and that this effect was independent of external pathogenic stress. Overall, the observed differential gene expression patterns across the focal clusters (i.e., 2, 5 and 6) show that the microbiota isolates MYb11 and MYb115 induce different *C. elegans*

transcriptional response compared to OP50. Moreover, each of the microbiota isolates MYb11 and MYb115 induce slightly different gene expression patterns in the focal clusters, suggesting that the two microbiota isolates activate different protection mechanisms against pathogen infection.



**Figure 2: Differential gene expression and enrichment analyses in *C. elegans* grown on *P. lurida* MYb11, *P. fluorescens* MYb115, or *E. coli* OP50 in the absence or presence of *B. thuringiensis*.** (A) Heatmap representing differentially expressed genes in worms grown on bacterial lawns of *P. lurida* MYb11, *P. fluorescens* MYb115, or *E. coli* OP50 followed by infection with pathogenic *B. thuringiensis* Bt247 or exposure to the non-pathogenic Bt407 at 12 hpi and 24 hpi. Co-regulated genes are divided into 6 clusters

via K-means clustering (de Hoon et al., 2004). Red and blue colors indicate up- and down-regulated genes, respectively. (B) Enriched transcription factor binding site (Motif) per cluster, detected by AMD (Shi et al., 2011) at the gene promoter region. (C) The most significant gene ontology (GO) term of co-regulated genes for each cluster. Gene Ontology (GO) analysis was performed with the DAVID Bioinformatics Resources 6.8 (Huang et al., 2009), where genes are assigned to one or more GO terms, which are also categorized into Biological Process, Molecular Function, and Cellular Component. The gene number ( $x$ -axis) refers to the *number of genes* annotated to the input list's specific GO term. Cluster identifies the respective heatmap cluster number in which the GO term was enriched. The statistical significance of GO enrichments was calculated using Fisher's exact test followed by FDR correction for multiple testing. Here the significance value denotes the FDR corrected  $p$ -values. A larger selection of GO terms for each cluster and the complete list of genes and the statistical significance values of each GO term are listed in **Table S1**.

*Promoter region motif enrichment analysis suggests a role for GATA, AP-1 and DAF-16 transcription factors in microbiota-mediated protection in C. elegans*

We then wanted to identify transcription factor binding sites that govern the gene expression changes in the different clusters using motif enrichment analysis on the promoter region (Shi et al., 2011). We identified enrichment of the GATA motif in the promoters of genes in cluster 2 and 6, of the Activator Protein-1 (AP-1) motif in the promoters of genes in cluster 2, and the Forkhead FOXO family *daf-16* motif in the promoters of genes in cluster 5 (**Figure 2B**). Our findings align with previous reports showing over-representation of the GATA motif in various intestine-specific and microbe-induced gene sets and in the promoters of innate immune system regulators (McGhee et al., 2009; Shapira et al., 2006; Yang et al., 2015). Also, in Cluster 2 the strong up-regulation of genes upon infection align with the fact that AP-1 transcription is activated in *C. elegans* via the JNK-like MAPK pathway, which is essential for the worm defense against the Bt Cry5B toxin (Kao et al., 2011) and Bt247 infection (Zárate-Potes et al., 2020). Insulin-like signaling via FOXO/DAF-16 activates pathogen defenses via antimicrobial genes, e.g., *lys-7* (Evans et al., 2008; Garsin et al., 2003), mitigates protein and cellular damage by reactive oxygen species (ROS) (Mohri-Shiomi & Garsin, 2008), and enhances behavioral avoidance to Bt (Hasshoff et al., 2007). Overall, our promoter region motif analysis suggests an involvement of the GATA, AP-1, and DAF-16 transcription factors in microbiota-mediated Bt defense in *C. elegans*; however, this merits empirical evidence.

*C. elegans* microbiota induces differential expression of genes involved in innate immune response and cellular structural components

*Innate immune response*

To explore the potential function of the differentially expressed genes, we used gene ontology (GO) enrichment analysis on the different clusters (**Figure 2C**). Various GO terms were significantly enriched (**Table 1 and Table S1 for the complete gene list**). The most significantly enriched GO term in cluster 2 is the immune response (GO:0045087) (e.g., *clec-41*, *lec-11*, *lys-1*, and *lys-2*). This term is over-represented upon infection and constitutively, in the absence of a pathogen, on the microbiota isolates in cluster 2. For some of these genes, their function in *C. elegans* pathogen defense is already described. For instance, *clec-41* encodes a protein with antimicrobial activity and is required for *C. elegans* defense against Bt247 (Pees et al., 2021). The overexpression of the lysozyme encoding gene *lys-1* increases worm resistance against Gram-positive pathogen *S. marcescens* (Mallo et al., 2002) and LYS-2 contributes to resistance against Bt (Boehnisch et al., 2011). Interestingly, CLEC-41, LYS-1, and LYS-2, and are primarily active in the intestine (Mallo et al., 2002; McGhee et al., 2009) (**Table 1**), and since MYb11 and MYb115 colonize and persist in the worm intestine, microbiota may have the advantage of time (extended stay) and space (close vicinity) to activate these lectins or lysozymes directly. Together, the up-regulation of immune response genes in cluster 2 implies that the microbiota isolates MYb11 and MYb115 may activate infection response genes early upon infection and even in the absence of a pathogen, 'priming' the defense response of the worm. Similarly, it has been shown that microbiota isolate *Pseudomonas mendocina* primes *C. elegans* (i.e., providing heightened preparedness) via the p38 MAPK immune pathway and protects the worm from subsequent infection with the pathogen *Pseudomonas aeruginosa* PA14 (Berg et al., 2019; Montalvo-Katz et al., 2013).

In addition to the GO term enrichment analysis, we performed a complementary enrichment analysis using WormExp (Yang et al., 2016b), which revealed an enrichment of genes related to *C. elegans* pathogen defenses in cluster 2. For example, genes previously reported to be up-regulated by infection with *B. thuringiensis* Bt247, *P. aeruginosa* PA14, *S. marcescens*, and *P. luminescens* (**Table S2**) were enriched in cluster 2. WormExp enrichment analysis revealed enrichment of p38 MAPK and insulin/insulin-like growth factor-1 (IGF-1) signaling pathway target genes. Both pathways are essential regulators of *C. elegans* immune defenses (Kim et al., 2002; Kim & Ewbank, 2018; Schulenburg & Müller, 2004; Troemel et al., 2006). Thus,

WormExp analysis underscores the link between microbiota-mediated protection and the nematode's innate immune response.

#### *Cellular structural components*

Structure-related GO terms were enriched in cluster 5, structural constituent of the cuticle (GO:0042302) (e.g., *dpy-13*, *col-12*, *col-92*), and cluster 6, collagen trimer (GO:0005581) (e.g., *col-92*) (**Figure 2C**). In cluster 5, structure-related genes are up-regulated early on MYb115 only, while on OP50, these genes are up-regulated late, upon Bt247 infection and Bt407 exposure. In cluster 6, these structure-related genes are up-regulated in worms on the microbiota isolates early and constitutively, in contrast, to the worms on OP50, which do not show up-regulation of genes. These structural genes may curtail the major difference in the underlying protective processes distinguishing MYb11 from MYb115, shedding light on the involvement of MYb115 in enhancing the physical barriers of the worm, such as the cuticle. Indeed, cuticle collagens and basement membrane collagens are vital components of the nematode's physical defense strategy (Taffoni & Pujol, 2015). Cuticular collagen expression was also shown to respond to environmental stimuli (i.e., diet, early developmental arrest and population density) and is regulated by the GATA transcription factor ELT-3 (Mesbahi et al., 2020). Cuticular collagens are also shown to be up-regulated in response to pathogenic *P. aeruginosa* (Sellegounder et al., 2019); while other collagens such as *col-92* are shown to be required for the resistance to *B. thuringiensis* (Iatsenko et al., 2013), and collagens expressed in non-intestinal tissues are shown to be up-regulated in response to Cry5B toxins (Gallotta et al., 2020). However, the microbiota and pathogens (e.g., Bt) are localized in the *C. elegans* intestine and are thus not directly in contact with the worm cuticle. The upregulation of collagen expression in response to intestinal pathogens may be part of a general defense reaction and might imply a connective network of signals between the worm gut and the cuticle. Intriguingly, it has been shown that *C. elegans* cuticle structure changes dynamically in response to *P. aeruginosa* infection and that worm cuticle barrier defense is regulated by the worm nervous system (Sellegounder et al., 2019). However, how far gut microbiota or Bt infection exchange signals with the worm cuticle, and whether it involves a microbiota-nervous system axis, similar to the so-called gut-brain axis, merits further investigation.

The late up-regulation of structural gene expression in worms on OP50 in cluster 5 may also result from the developmental transition in worms at the later time-point (24 hpi) from L4 to the adult stage (Yang et al., 2019). It could alternatively indicate a slight developmental

difference between the worms grown on the different bacteria. During worm development, worms synthesize a cuticle five times, starting in the embryo stage, and at the four larval stages before each of the four molting cycles (Singh & Sulston, 1978). Thus, we expect cuticular genes to be expressed temporally, showing a peak to indicate the molting cycle and the cuticle's synthesis for the next cycle (Johnstone, 2000). However, the abundance of all cuticular collagen genes does not appear simultaneously peaking before each cycle. The expression of *dpy-13*, causing the dumpy phenotype, with a short body length of worms compared to the wild-type, is shown to peak in between the expected temporal cycle (Johnstone & Barry, 1996). Also, *col-12* expression peaks only before the L1 and L2 stages, almost absent before the adult stage, where *dpy-7* is shown to peak (Johnstone, 2000; Johnstone & Barry, 1996). Such patterns, showing the late up-regulation of structural genes on OP50 and their early up-regulation on MYb115, suggest that MYb115 accelerates worm development, similar to a *Comamonas* DA1877 diet for worms (Macneil et al., 2013); however, this would require further experimental validation.

Interestingly, we also observed enrichment of several genes involved in intermediate filament regulation (GO: 0005882) in cluster 2 (e.g., *ifb-2*), albeit statistically not a significant enrichment (**Table S1**). IFB-2, a structural component of the intestinal terminal web, together with other intermediate filaments, have been shown to play an essential role in the enhancement of barrier function and strengthening of epithelial cell integrity of the nematode in defense against the Bt toxin Cry 5B (Geisler & Leube, 2016). Recently, IFB-2, an intermediate filament expressed exclusively in the intestine, was shown to be essential for the MYb115-mediated protection in *C. elegans*. Mutant *ifb2(kc14)* worms, which lack the endotube (a dense intermediate filament-rich layer of the worm intestine cells), were not protected by MYb115 against Bt247 infection (Peters, 2020, Master thesis, unpublished). Thus, the involvement of intermediate intestinal filaments in microbiota-mediated protection could be worth further investigation.

Taken together, both microbiota isolates MYb11 and MYb115 induce differential expression of genes involved in *C. elegans* innate immune responses and structural components.

**Table 1: A concise description of selected genes, differentially regulated during microbiota-mediated protection against *B. thuringiensis*.**

Gene	Cluster	Concise Description <sup>a</sup>
<i>lys-1</i> <sup>i</sup>	2	defense response to G-ve and G+ve bacteria; innate immune response; expressed in ciliated neurons, head neurons, and intestine.
<i>lys-2</i>	2	predicted lysozyme activity; defense response to G-ve and G+ve bacteria; innate immune response
<i>clcc-4</i>	2	predicted to have carbohydrate-binding activity
<i>lec-8</i> <sup>i</sup>	2	an ortholog of galectin 3; exhibits glycolipid binding activity; response to toxic substance; expressed in intestine, pharyngeal-intestinal valve, and tail
<i>lec-9</i>	2	an ortholog of galectin 3; exhibits glycolipid binding activity
<i>lec-10</i> <sup>i</sup>	2	exhibits glycolipid binding activity; expressed in the intestine
<i>clcc-41</i>	2	predicted to have carbohydrate-binding activity; involved in positive regulation of chemotaxis
<i>ifb-2</i> <sup>i</sup>	2	an ortholog of lamin B2; expressed in the intestine
<i>ifc-2</i>	2	ortholog of keratin 71, 73, and 77; expressed in hypodermis, intestinal cell, and pharynx
<i>ifd-2</i>	2	an ortholog of keratin 85, lamin A/C, and lamin B1; predicted to encode a protein with the following domains: Phosphorylation site, Intermediate filament protein, Intermediate filament, rod domain
<i>ifp-1</i> <sup>i</sup>	2	expressed in head, head neurons, intestine; pharynx, and ventral nerve cord; predicted to encode a protein with the following domains: Phosphorylation site, Intermediate filament protein, Lamin tail domain superfamily, Intermediate filament, rod domain
<i>dpy-13</i>	5	predicted to be a structural constituent of the cuticle, involved in cuticle development involved in collagen and cuticulin-based cuticle molting cycle
<i>col-12</i>	5	predicted to be a structural constituent of cuticle
<i>unc-54</i>	6	an ortholog of myosin heavy chain 1, 6 and 7; exhibits actin filament binding activity and microfilament motor activity; involved in muscle contraction; oviposition; and skeletal muscle myosin thick filament assembly; expressed in anal depressor muscle, body wall musculature, head, vulval muscle, and in male
<i>ilys-5</i>	6	predicted to have lysozyme activity
<i>sod-1</i> <sup>i</sup>	6	an ortholog of superoxide dismutase 1; copper ion binding activity; protein homodimerization activity; and superoxide dismutase activity; involved in the regulation of brood size, vulval development, and removal of superoxide radicals; expressed in several structures, including intestinal cell, somatic gonad, and somatic nervous system
<i>fat-7</i>	6	an ortholog of stearyl-CoA desaturase 5 and 9 activity; involved in the fatty acid biosynthetic process and multicellular organism development
<i>lam-1</i> <sup>i</sup>	6	an ortholog of laminin subunit beta 1 and 2; involved in pharynx development and positive regulation of locomotion; expressed in the intestine
<i>lam-2</i>	6	an ortholog of laminin subunit gamma 1; involved in positive regulation of locomotion
<i>act-1</i>	6	an ortholog of actin beta; predicted to be a structural constituent of cytoskeleton; involved in cortical actin cytoskeleton organization; mitotic cytokinesis; muscle organ development; expressed in body wall musculature and gonad
<i>act-2</i>	6	an ortholog of actin beta predicted to have ATP binding activity, involved in cortical actin cytoskeleton organization; cytoskeleton-dependent cytokinesis, and locomotion; expressed in body wall musculature; gonad; hypodermis; and neurons

<sup>a</sup>based on Wormbase version WS277s

<sup>i</sup> known to be expressed in the intestine

G-ve: gram-negative; G+ve: gram-positive



*Proteome level changes observed upon microbiota-mediated protection in C. elegans*

To assess the *C. elegans* proteome response to the protective microbiota isolates MYb11 and MYb115 in the presence of pathogenic Bt247 and non-pathogenic Bt407, we performed a quantitative proteome analysis using LC-MS. We quantified a total of 4,506 proteins with FDR  $p$ -value  $< 0.01$  restriction (**Supplementary Figure 4S**). The taxonomic distribution of identified proteins revealed that most proteins (i.e., 3991 proteins) belong to *C. elegans*, followed by the bacterial isolates used in the different treatment groups: *E. coli* (323 proteins), *Pseudomonas* (96 proteins), and *B. thuringiensis* (88 proteins). On average, around 2500 *C. elegans* proteins were identified in each of the nine different treatment groups (**Supplementary Figure 4S**).

*Proteome dataset confirms the differential abundance of lectins and lysozymes upon microbiota-mediated protection*

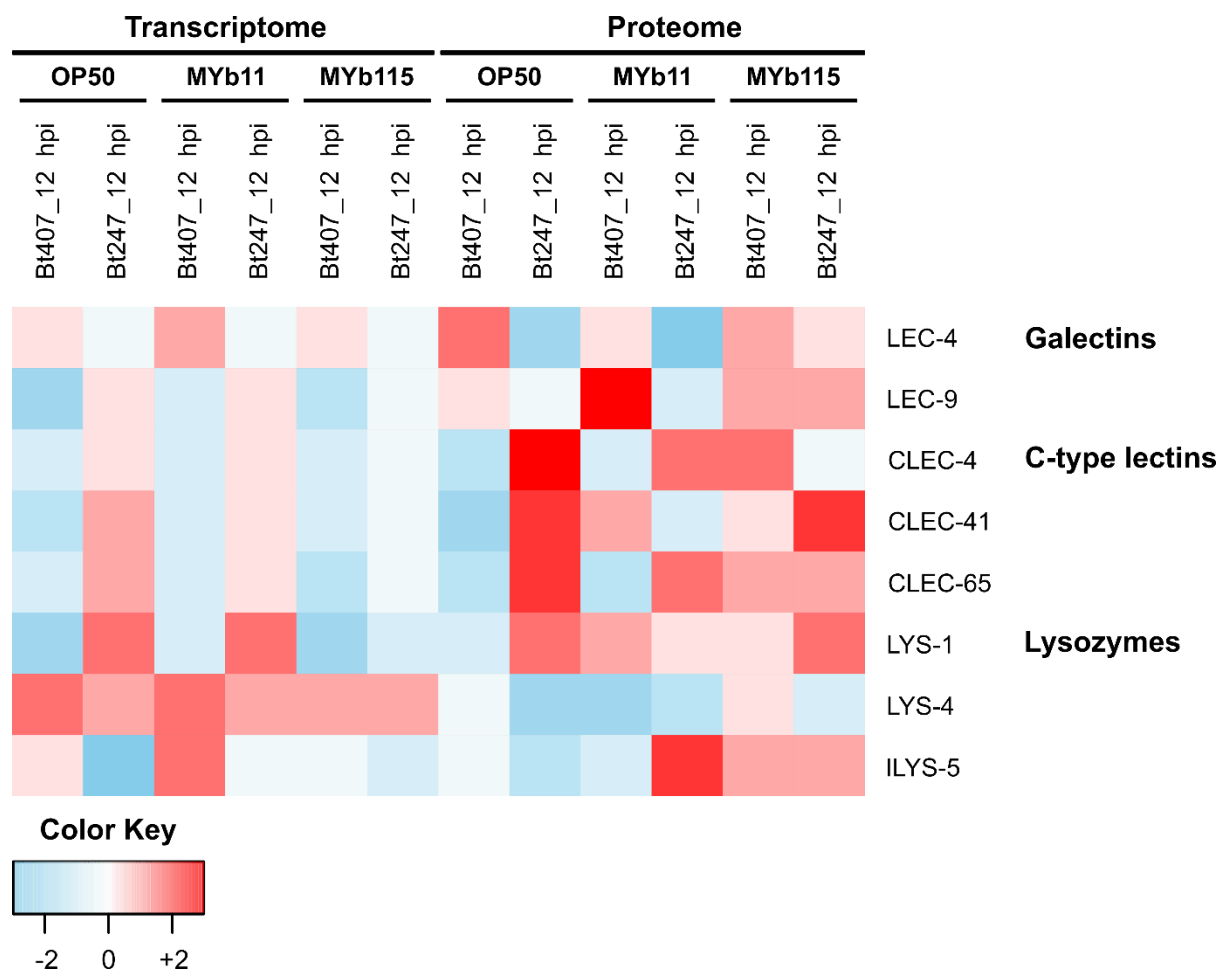
Next, we combined the transcriptome dataset with the corresponding proteome dataset for each treatment condition by filtering the significant fold changes (FDR  $p$ -value  $< 0.01$ ) of the different treatments in the proteome and matching it with the transcriptome dataset. The combined dataset revealed 365 differentially abundant proteins, which correspond to differentially regulated transcripts, across the different treatments compared to worms grown on only OP50. We then performed GO enrichment analysis on the combined dataset, which confirmed the GO terms obtained at the transcriptome level, such as the GO terms related to worm innate immune response, and cellular structural components (**Supplementary Figure 5S, Table S3**). Then, we focused on proteins of interest, such as all the lectins (galectins and C-type lectins) and lysozymes in the combined dataset (**Figure 3**).

Some of the proteins were directionally (i.e., simultaneous increase or decrease in abundance) consistent with their respective gene expression at the transcript level for specific treatments. Consistent patterns were observed, for instance, in Bt247-infected worms on MYb11 having increased abundance of *clec-4*, *clec-65*, *lys-1* and decreased abundance of *lec-4*. Bt247-infected worms on OP50 showed consistently increased abundance of *clec-4*, *clec-41*, *clec-65*, *lys-1* and decreased abundance of *lec-4* and *ilys-5*. Such patterns might suggest that the microbiota isolates have a targeted induction of a limited set of genes, rather than the general activation of more genes. However, a discordance of abundance at transcript and protein level was also

observed for other genes. This discordance could be explained by the potential post-translational modifications (e.g., proteolysis, glycosylation, phosphorylation) that change protein function (Hasin et al., 2017; Manzoni et al., 2018) and lead such a discordance (Hasin et al., 2017).

Compared to the Bt247-infected worms, Bt407-exposed worms mainly showed a decreased abundance of genes at the transcript level. However, at the proteome level, Bt407-exposed worms showed an increased abundance of proteins in worms grown on microbiota isolates compared to those on OP50. Moreover, MYb115-grown worms showed a strong increase in the abundance of all proteins in the absence of a pathogen. This pattern is yet another indication that microbiota, particularly MYb115, primes the host response in the absence of a pathogen.

Taken together, the overlap of specific genes differentially expressed in the combined dataset with their respective protein abundances suggests a robust response to microbiota-mediated protection against pathogen infection with different response patterns according to the microbiota isolate (MYb11 or MYb115) and the Bt strain (Bt407 or Bt247), which might reveal different microbiota-mediated protection mechanisms in the nematode on each of MYb11 and MYb115. Moreover, the increased abundance of proteins on microbiota-grown worms indicates the microbiota-mediated priming of host immune defenses.

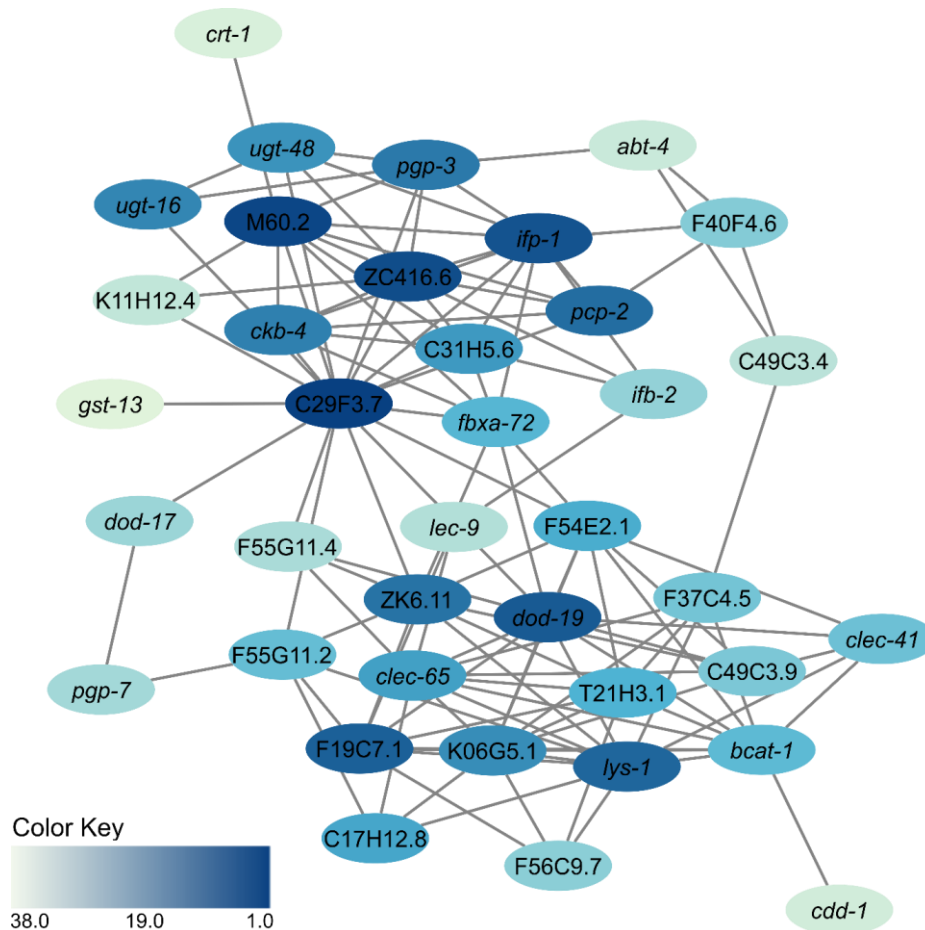


**Figure 3: Differential abundance of proteins and their respective gene expression at transcript level for galectins, C-type-lectins, and lysozymes in the combined dataset.** The heat map shows gene expression level and their respective protein abundances for a selected list of protein families: galectins (LEC-4 and -9), C-type lectins (i.e., CLEC-4, -41, and -65), lysozymes (i.e., LYS-1 and -4 and ILYS-5). Protein members of each family are presented on the right. Heatmap scale bars indicate fold changes of each treatment to the control worms, i.e., grown on OP50 with no Bt exposure at the log<sub>2</sub> scale. The results show the averages for the replicates of each of the treatments. Red and blue colors indicate up-and-down-regulation, respectively.

### *Gene network analysis to prioritize immune response genes in cluster 2 of the combined dataset*

We next wanted to have a list of prioritized candidate genes to direct future functional analyses. For this, we used WormNet, a probabilistic gene network model (Lee et al., 2010) and focused on the combined dataset cluster 2 (derived from cluster 2 of the transcriptome). This approach has been shown to identify functionally essential host genes induced by bacteria (Radeke & Herman, 2020). Here, cluster 2 was chosen since it showed enriched GO terms related to worm immune response and strongly up-regulated genes upon microbiota exposure both in the

presence and absence of pathogen (**Figure 2 & 5S**). Of the 51 genes in the combined dataset cluster 2, 38 genes were connected within the gene network with an AUC (Area Under the Curve) of 0.8348 ( $p$ -value= 8.8729E-27) (**Figure 6**). Here, *ifp-1* and *lys-1* are among the top 10 ranked genes, followed by *clec-65*, *clec-41*, and *ifb-2* (**Figure 6**). This prioritized list of candidate genes could direct future functional analyses in further investigating the mechanisms of microbiota-mediated protection.



**Figure 4: Gene network analysis to prioritize immune response genes in cluster 2 of the combined dataset.** Gene network analysis of genes from the combined dataset cluster 2 genes. We used WormNetv2 (Lee et al., 2010), a probabilistic functional gene network model, where 38 out of the queried 51 genes showed to be connected (AUC= 0.8348,  $p$ -value= 8.8729E-27). AUC is the area under the curve (receiver operating characteristic curve), and it gives a measure for the true-positive genes. A random network would have an AUC of 0.5, but a more connected network representing the perfect prediction would have an AUC value of 1. Thus, a value of AUC 0.8348 suggests a high predictive power of connected genes. Cytoscape 3.8.2 (Shannon et al., 2003) was used to visualize the gene network. Blue ellipses represent individual genes, and the grey lines the connections between these genes, based on both physical evidence on predictions of the model. Genes are colored according to their rank; higher-ranked genes are colored darker blue. Rank was determined according to score, which is calculated based on the number of connections and the strength of evidence from these connections (See **Supplementary Table S4** for the complete list of 38 genes, their scores and the connected genes).

*Initial in vivo experiments suggest the involvement of JNK MAPK in microbiota-mediated protection*

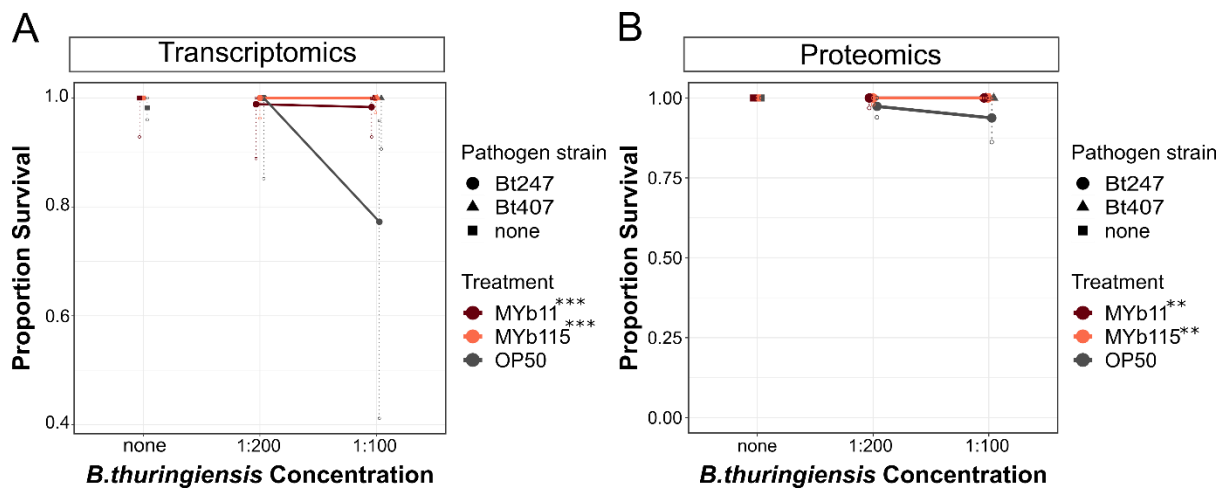
In the transcriptome and proteome analyses, *C. elegans* immune response emerged as a common denominator. WormExp and promoter enrichment analyses pointed to an involvement of the p38 and JNK MAPK pathways. p38 MAPK has already been shown to have an essential role in *P. mendocina*-mediated protection of the worms via priming them to subsequent *P. aeruginosa* infection (Montalvo-Katz et al., 2013). Thus we used *pmk-1(km25)* and *mek-1(ks25)* mutants with disrupted p38 MAPK and JNK MAPKK function, respectively, to assess their survival following Bt247 infection in the presence or absence of the microbiota isolates. *mek-1(ks25)* mutants grown on either MYb11 or MYb115 were significantly less protected than wild-type N2 worms on MYb11 and MYb115. In contrast, *pmk-1(km25)* worms were as protected as wild-type N2 worms on the microbiota isolates (**Supplementary Figure S6**). This data suggests that JNK MAPK may be required for microbiota-mediated protection against Bt247, but further confirmation is necessary. Concerning p38 MAPK, on the one hand, *pmk-1(km25)* mutants are not protected against *P. aeruginosa* infection when grown on the protective *C. elegans* microbiota member *Pseudomonas mendocina* (Montalvo-Katz et al., 2013). The p38 MAPK pathway is also linked to *C. elegans* defense against pathogen infection (Kim et al., 2002; Visvikis et al., 2014; Yang et al., 2015), and also to *C. elegans* response to Bt PFTs Cry 5B and Cry 21A (Bischof et al., 2008; Huffman et al., 2004) and the Bt679 strain (Zarate-Potes et al., 2020). On the other hand, the p38 MAPK pathway was shown to be dispensable for the defense against Bt247 (Zarate-Potes et al., 2020). Thus, further experiments are needed to confirm the necessity of p38 and JNK MAPK pathways in microbiota-mediated protection.

## **Conclusion**

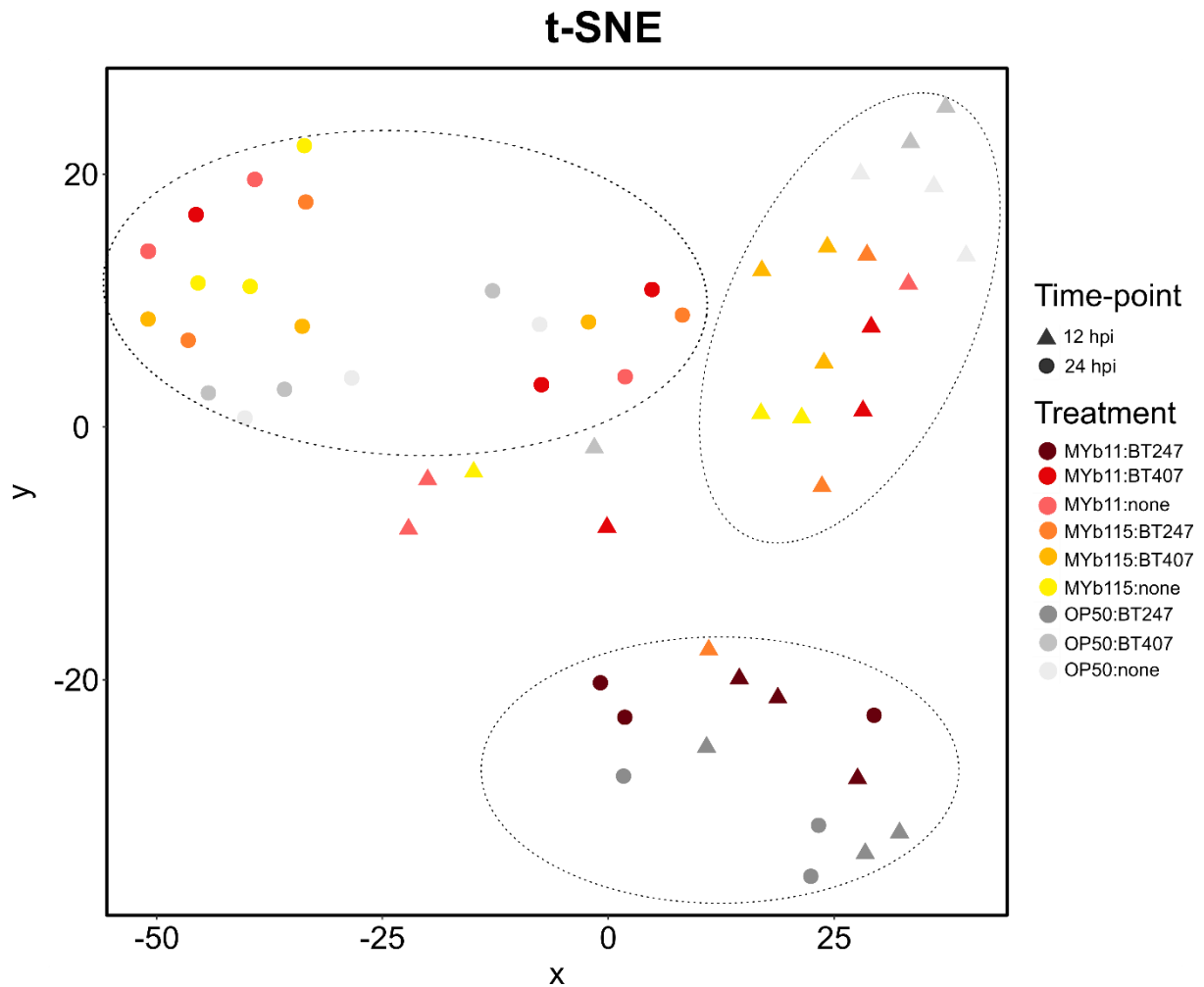
We used a multi-omics approach to explore the *C. elegans* response to protective microbiota isolates. We found that the protective microbiota isolates affect *C. elegans* innate immune response and cellular structural component functions. The multi-omics (i.e., transcriptome and proteome) analysis confirmed the differential abundance of three groups of proteins on microbiota-grown worms: galectins, C-type lectins, and lysozymes, suggesting their potential

roles in microbiota-mediated protection. Moreover, microbiota-induced increase in the abundance of genes and proteins that are shown to be involved in *C. elegans* pathogen defense, both in the absence of a pathogen, suggest that microbiota isolates MYb11 and particularly MYb115 prime the host defenses in preparation for a subsequent pathogen attack. Overall, we present the inducible responses of *C. elegans* to protective microbiota isolates and provide a framework at both the transcript and protein level, in addition to candidate genes, to facilitate further functional investigation of *C. elegans*-microbiota-pathogen tripartite interactions.

## SUPPLEMENTARY MATERIAL

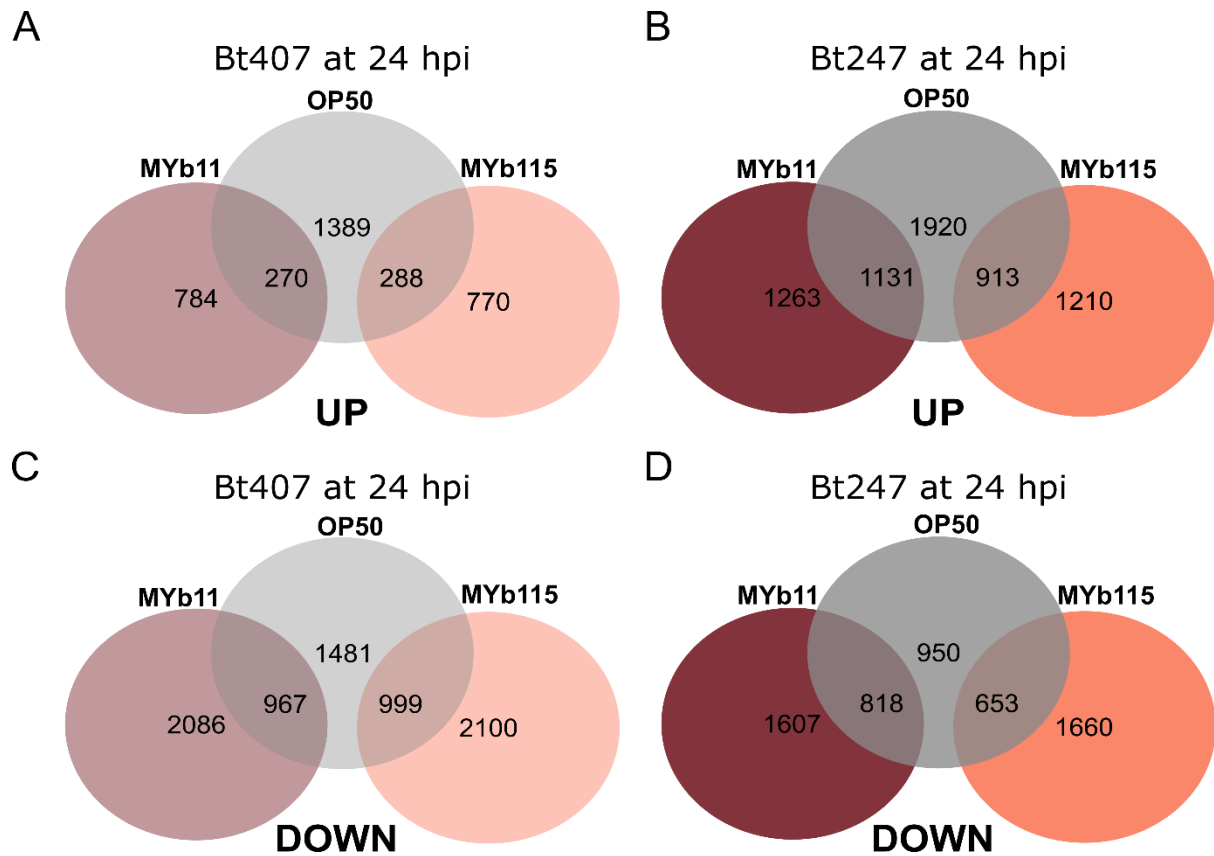


**Figure S1: Survival assays performed parallel to each of the omics infection assays (sample collection for transcriptomics and proteomics) indicate microbiota-mediated protection in the collected and processed samples.** Survival of *C. elegans* N2 on different concentrations of *B. thuringiensis* Bt247 spore solution parallel to (A) transcriptomics and (B) proteomics sample collections. Spore solution was mixed with OD600 adjusted (OD =10) microbiota isolates *P. lurida* MYb11 and *P. fluorescens* MYb115 or *E. coli* OP50 as a control. *B. thuringiensis* strain Bt407 is used as a non-pathogen control. *E. coli* OP50, which is not protective, was used as a negative control. Error bars show the median survival range proportions of four technical replicates (n = 4). Statistical analyses were performed using the GLM framework and Bonferroni correction for multiple testing with the OP50 control treatment group. Significance codes: 0 [\*\*\*], 0.001[\*\*], 0.01 [\*].

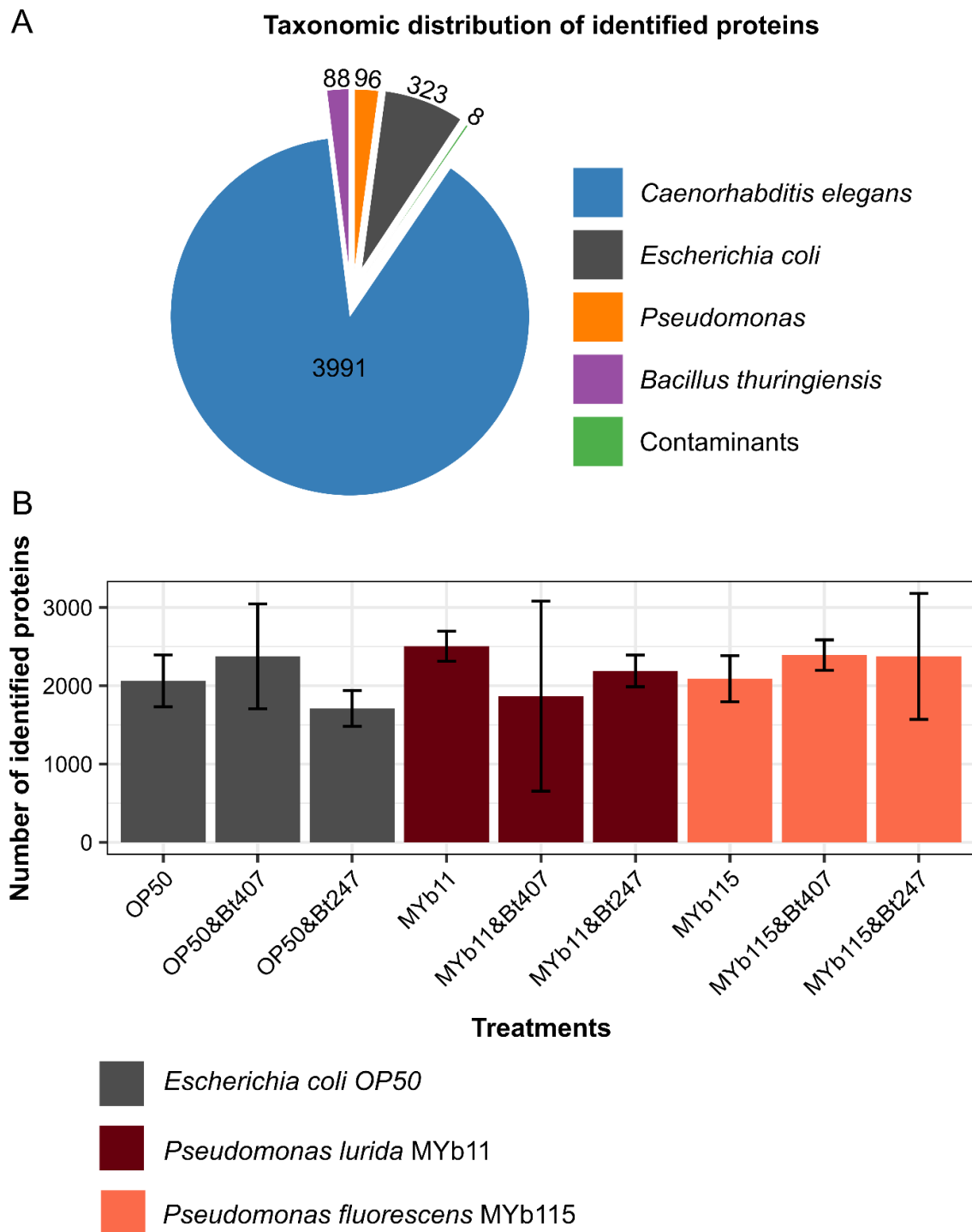


**Figure S2: t-distributed Stochastic Neighbor Embedding (t-SNE) analysis of transcriptomic variation.** t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis of transcriptomics variation. Variation is assessed for the different worm treatments: grown on the microbiota isolates *P. lurida* MYb11, *P. fluorescens* MYb115, or the *E. coli* OP50 control, then exposed to pathogenic Bt247, non-pathogenic Bt407 or no Bt. The samples were obtained at two exposure time-points, 12 hpi (represented by triangles) and 24 hpi (represented by circles). The different colors represent the different treatment groups.

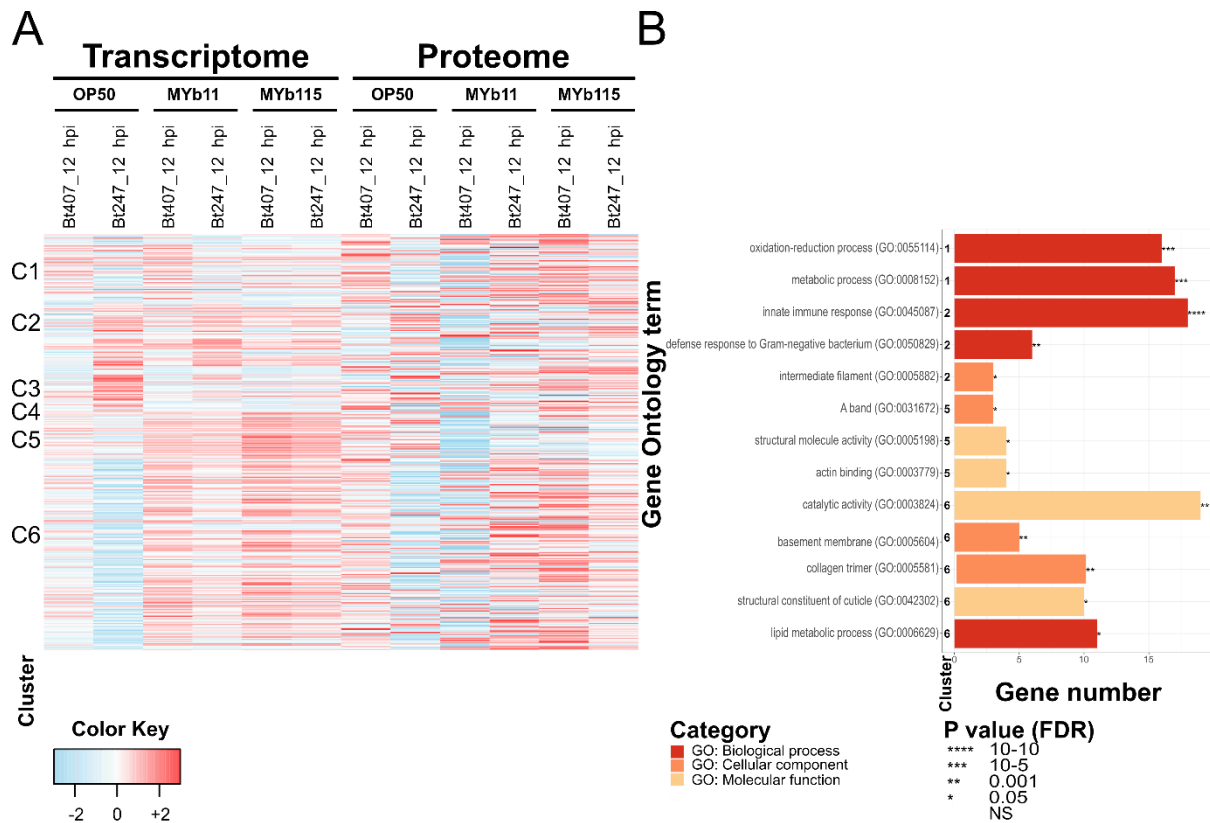




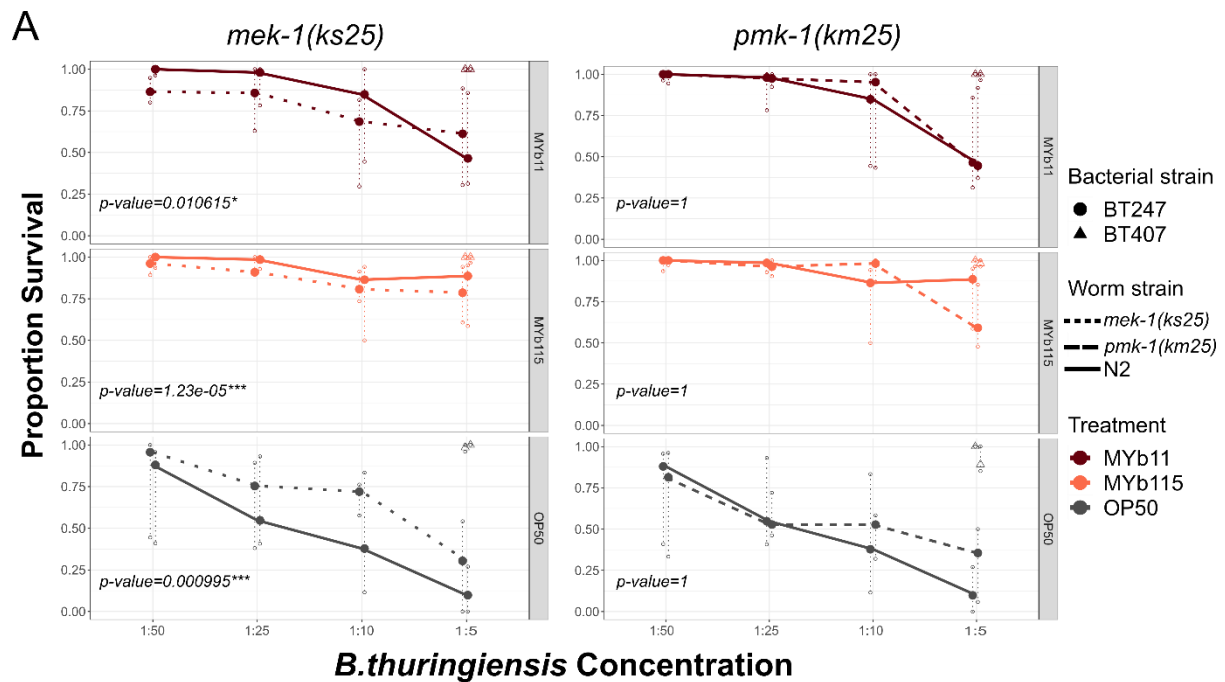
**Figure S3: The number of up-regulated, down-regulated, and co-regulated genes in each treatment condition compared to the respective treatment on *E. coli* OP50 at 24 hpi.** (A-D) Venn diagrams show the numbers of statistically significant up-regulated (A and B), down-regulated (C and D) and co-regulated (overlapping) genes in each of the treatments: worms grown on the bacterial isolates (i.e., MYb11, MYb115, and OP50) and exposed to Bt407 (A and C) or infected with Bt247 (B and D) at 24 hpi. All conditions are compared to the worms exposed to only OP50.



**Figure S4: Overall taxonomic distribution of identified proteins and their average numbers in each treatment group.** (A) Taxonomic distribution of proteins identified in the nine treatment conditions. The identified proteins belonged in decreasing order to the following taxa: *C. elegans*, *E. coli*, *Pseudomonas*, *B. thuringiensis*, and contaminants. (B) Bar graph of the average number of identified proteins in each of the nine treatment conditions observed in 4 biological replicates in each treatment group. The error bars present the standard deviation. The different colors present the food bacteria the worms were grown on: Grey: *E. coli* OP50, dark red: *P. lurida* MYb11, and orange: *P. fluorescens* MYb115.



**Figure S5: Comparison of differential gene expression at the transcriptome and proteome levels and selected enriched Gene Ontology (GO) terms on the significantly differential gene expression in the combined dataset.** (A) Heatmap shows expression changes of genes at transcript and protein levels 12 hpi in worms grown on each food bacteria *E. coli* OP50, *P. lurida* MYb11, and *P. fluorescens* MYb115, exposed to Bt407 or Bt247. Data are shown for genes with significant differential expression at the protein level. Red refers to up-regulated, and blue refers to down-regulated genes and proteins. Heatmap indicates the fold changes at the log<sub>2</sub> scale of each treatment condition compared to the worms grown on only OP50. (B) Selected and non-redundant gene ontology (GO) terms of co-regulated genes from each of the six clusters, classified into three categories: biological process (GO term), cellular compartment (GO term), and molecular function (GO term). GO selection criteria were based on the most significantly enriched term (when available) in each cluster; in addition to GO terms that potentially could be relevant to microbiota-mediated protection, according to current literature. The gene number refers to the *number* of genes annotated to the specific *term* from the input list. Cluster identifies the cluster in which the GO term was enriched. The significance value denotes the FDR corrected p-values. The complete list of genes and the statistical significance value of each selected enriched GO terms are listed in **Table S3**.



**Figure S6: The p38-MAPK pathway does not affect microbiota-mediated protection, while JNK-MAPK pathway does.** (A) Survival assay of the *pmk-1(km25)* and *mek-1(ks25)* mutants and the N2 strain, grown on microbiota isolates *P. lurida* MYb11 and *P. fluorescens* MYb115 or the *E. coli* OP50 control, followed by pathogen challenge with Bt247 or the pathogen control Bt407. Both mutants grown on either of the microbiota isolates MYb11 or MYb115 were not protected from the Bt infection. *pmk-1(km25)*, *mek-1(ks25)*, and the N2 strains are represented respectively by dashed, dotted, or solid lines. The different colors represent the bacterial lawn conditions on which the worms were grown on. Error bars represent the range of the median of survival proportions of four technical replicates (n = 4). Statistical analyses were performed using the GLM framework and Bonferroni correction for multiple testing with the OP50 control treatment group. This assay was performed for one run. Significance codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1. (B) Diagram of the *C. elegans* p38 and JNK MAPK immune response pathways.

**Table S1: A selection of enriched GO terms, gene counts, statistical significance, and the list of all the genes in the respective enrichment (on the transcriptome dataset).**

Cluster	Term	Count	FDR	All genes
1	GO:0055114~oxidation-reduction process	68	5.39299E-34	<i>cyp-32B1,cyp-25A1,T08H10.1,fmo-5,gpd-2,cyp-33D3,stdh-2,F49E12.10,fin-1,F41E6.5,dhs-20,dhs-23,cyp-13A6,cyp-13A8,hacd-1,acox-1,F17A9.5,fmo-2,fmo-1,cyp-13A10,C06E4.6,R05D8.7,cyp-33E1,F32D8.12,R08F11.7,ZK550.6,R04B5.5,ctl-2,F58F9.7,dhs-2,alh-12,F59F4.1,cyp-25A3,gbh-2,fat-5,hpd-1,alh-10,fat-6,ddo-2,cyp-37B1,dao-3,nnt-1,cyp-34A2,K09H11.1,Y71F9B.9,cyp-35A2,daao-1,ech-8,F12E12.11,gpx-6,F35C8.5,cyp-37A1,F55E10.6,cyp-13A12,F53C11.3,cyp-33E2,cyp-33E3,cyp-34A9,hgo-1,gpx-7,Y48A6B.9,cyp-33C2,fat-3,dhs-28,Y38F1A.6,W06H8.2,F54F3.4,F25D1.5</i>
2	GO:0045087~innate immune response	93	1.12473E-77	<i>F59B1.8,C49C3.9,clec-67,K04A8.1,irg-1,F49H6.3,cpr-3,pgp-3,valv-1,ZK6.11,clec-66,ZK1055.7,dod-19,F54D5.4,comt-2,F54E2.1,dod-17,irg-3,dod-20,hpo-6,F44G3.10,gst-38,F35E12.8,B0024.4,F35E12.6,F39G3.5,clec-74,dct-17,F38A1.9,clec-165,F35E12.10,F13A7.11,T24C4.8,C29F3.7,F22H10.2,C49C8.5,C30E1.8,F49H6.13,F08G2.5,F55G11.7,clec-187,clec-186,F55G11.2,F35E12.5,M60.2,gst-13,F55G11.8,Y46D2A.2,lec-11,T24C4.4,F20G2.5,clec-258,T27C5.8,ugt-44,F08G5.6,F11D11.3,lys-2,lys-1,C34H4.1,C14C6.5,C34H4.2,sox-4,F49F1.1,F35E12.2,F19C7.1,ins-11,clec-41,T25D10.1,K11H12.4,his-11,K09D9.1,PDB1.1,F01D5.5,C49G7.10,C17H12.6,zip-10,F01D5.1,C17H12.8,C18H7.11,F53C11.1,H20E11.3,C32H11.4,H20E11.2,C29F9.3,H20E11.1,T24B8.5,clc-1,Y17G7B.8,K11H12.3,K08D8.4,K08D8.5,clec-85,M04C3.2</i>
2	GO:0030246~carbohydrate-binding	25	3.50215E-06	<i>clec-258,clec-265,clec-67,clec-50,clec-190,clec-173,crt-1,clec-66,clec-64,clec-143,clec-65,clec-41,lec-6,clec-74,clec-45,clec-165,clec-4,W04B5.3,clec-187,clec-186,lec-8,lec-9,lec-10,lec-11,clec-85</i>

2	IPR001304: C-type lectin	19	0.004390259	<i>clec-265,clec-74,clec-258,clec-165,clec-45,clec-67,clec-50,F40F4.6,clec-173,clec-4,clec-187,clec-66,clec-186,F28B4.3,clec-143,clec-64,clec-65,clec-41,clec-85</i>
2	GO:0005882~ intermediate filament	4	3.670562267	<i>ifb-2,ifc-2,ifp-1,ifd-2</i>
3	GO:0045087~ innate immune response	47	7.13136E-23	<i>tsp-1,skr-3,C42D8.1,C25D7.5,C10C5.2,C08E8.4,irg-2,pqm-1,clec-62,tag-244,F56D2.5,C17H1.6,vab-9,Y41D4B.15,C33H5.13,Y41D4B.17,Y47G6A.5,F35E12.9,fbxa-30,clec-72,C04G6.5,T21F4.1,F54B8.4,C39H7.4,fbxa-59,Y46D2A.1,crn-2,F53A9.6,C50F4.9,spp-18,dod-22,T19D12.4,clec-83,tir-1,C18E9.9,elt-2,F16H6.10,nhr-115,D2007.1,F41C3.8,C14C11.4,klf-3,clec-84,nhr-112,fbxa-105,fbxa-150,F20C5.4</i>
3	GO:0043565~ sequence-specific DNA binding	25	0.02844573	<i>ets-4,nhr-231,nhr-7,nhr-54,nhr-8,nhr-98,nhr-203,zip-5,nhr-201,Y17G7B.22,nhr-202,nhr-123,nhr-184,nhr-178,ceh-37,jun-1,ets-9,nhr-214,nhr-63,elt-2,nhr-115,nhr-195,nhr-112,nhr-135,nhr-50</i>
4	GO:0005576~ extracellular region	21	0.028804501	<i>ins-20,ZK669.3,C03G6.5,ttr-5,nas-25,nas-37,cwn-2,tep-1,sod-4,nlp-28,nlp-29,dpy-31,asm-1,B0478.3,nlp-31,thn-1,F07F6.8,lbp-1,C27B7.7,ins-4,lea-1</i>
5	GO:0042302~ structural constituent of cuticle	56	7.35204E-49	<i>col-169,col-168,col-170,col-167,col-107,col-175,dpy-5,col-174,col-120,col-97,dpy-13,col-60,col-104,col-130,col-63,col-125,col-58,col-157,col-12,col-10,col-166,col-14,col-13,col-117,col-109,col-71,col-73,col-113,bli-6,col-65,bli-2,bli-1,lon-3,cut-2,col-152,col-77,col-76,col-38,col-39,col-74,col-33,col-147,rol-1,col-182,rol-8,sqt-3,sqt-2,sqt-1,col-89,col-91,ram-2,col-144,col-48,col-145,dpy-4,col-138</i>
6	GO:0005581~ collagen trimer	39	2.18413E-20	<i>Y69H2.14,col-106,col-133,col-80,col-98,col-62,emb-9,col-124,col-129,col-156,col-103,col-154,col-101,col-155,col-149,col-81,col-160,col-178,col-159,col-181,col-180,col-179,col-184,col-20,col-19,col-7,col-162,col-8,col-161,col-139,col-140,col-92,col-142,col-143,col-88,col-49,col-146,col-93,col-137</i>
6	GO:0042302~	38	3.67345E-16	<i>Y69H2.14,col-106,col-133,col-80,col-98,col-62,col-124,col-129,col-156,col-103,col-154,col-101,col-155,col-149,col-81,col-160,col-178,col-159,col-181,col-180,col-179,col-184,col-20,col-</i>

	structural constituent of cuticle			<i>19,col-7,col-162,col-8,col-161,col-139,col-140,col-92,col-142,col-143,col-88,col-49,col-146,col-93,col-137</i>
6	GO:0008152~ metabolic process	71	2.4078E-06	<i>acs-5,H25K10.1,pgp-2,ZK228.4,sur-5,tba-4,F08A8.4,B0395.3,F08A8.2,F08A8.3,suca-1,fbk-5,B0286.3,F26C11.1,lips-17,gst-26,cpt-5,C45E5.1,T05C3.6,gst-27,gst-28,ugt-37,ugt-30,F49D11.3,pes-9,pmp-2,cbl-1,Y39E4A.3,ugt-60,ugt-47,pmp-5,F45E1.4,pmt-2,ugt-53,pho-1,fbk-3,D1022.3,W05B2.7,Y105C5B.15,ugt-63,K11G9.3,unc-54,T24C12.3,ugt-46,F11C1.5,alh-9,ugt-21,F08F3.4,F18E2.1,ilys-5,aman-3,pmp-1,F54C1.1,fah-1,ech-7,poml-2,tkl-1,acdh-2,C44C1.5,C23H4.7,C53A3.2,aagr-4,Y62E10A.13,acs-1,dgat-2,catp-3,sucl-1,aman-1,T09B4.8,alh-13,ugt-22</i>
6	GO:0005604~ basement membrane	8	0.000598574	<i>him-4,epi-1,ost-1,emb-9,unc-52,nid-1,lam-1,lam-2</i>
6	GO:0005865~striated muscle thin filament	6	0.00820371	<i>lev-11,act-4,unc-87,act-1,atn-1,act-3</i>

**Table S2: Enriched WormExp gene sets for differentially expressed genes upon treatments**

Cluster	Category	Gene set <sup>a</sup>	Counts <sup>b</sup>	FDR
C1	Microbes	down by <i>B. thuringiensis</i> at 6h (BT247, 1:10)	143	2.35E-119
	TF Targets	Low-complexity elt-2 targets	178	4.57E-93
C2	Microbes	UP by PA14, 24h	191	3.85E-212
	Microbes	UP <i>S. marcescens</i> 24h, RNASeq	290	2.49E-202
	Microbes	UP by <i>B. thuringiensis</i> (BT247), 24h	226	3.64E-179
	Microbes	UP by PA14 (Miller)	117	2.65E-145
	Microbes	UP <i>S. marcescens</i> 24h, TillingArray	256	1.06E-141
	DAF/Insulin/food	UP by <i>daf-16</i> mutant under <i>daf-2;rsks-1</i> mutant	187	3.96E-136
	Microbes	UP by PA14, 12h	118	4.88E-135
	Microbes	UP <i>P. luminescens</i> 24h, RNASeq	254	4.96E-126
	TF Targets	Low-complexity elt-2 targets	220	9.73E-110
	C3	Microbes	UP by <i>B. thuringiensis</i> at 6h (BT247, 1:2)	217
Microbes		UP on <i>S. marcescens</i>	227	1.39E-169
Tissue		Positive Enriched Intestinal genes	248	1.68E-125
Microbes		UP by PA14, 24h	140	1.88E-121
Microbes		UP by Bt toxin, Cry5B	103	2.27E-90
C4	Microbes	UP by <i>S. aureus</i>	213	3.14E-186
	Microbes	UP on <i>S. marcescens</i>	221	3.33E-156
	DAF/Insulin/food	UP by starved	232	2.91E-141
	Microbes	UP by <i>B. thuringiensis</i> at 12h (BT247, 1:2)	186	2.13E-127
	Microbes	UP on <i>X. nematophila</i>	139	5.81E-96
C5	DAF/Insulin/food	Fasting-Induced	184	4.65E-91
	Microbes	down by <i>S. enterica</i> SL2048 under <i>fer-1</i> mutant	274	4.14E-300
	Mutants	UP by <i>daf-12</i> mutant (Delaney)	283	1.07E-296
	DAF/Insulin/food	UP by <i>daf-16</i> mutant (Delaney)	284	6.10E-267
	Mutants	down by <i>lin-45</i> mutant (Burton)	264	2.15E-240
	DAF/Insulin/food	UP <i>daf-2</i> ( <i>e1370</i> ) vs. wildtype	356	2.82E-232
	Mutants	down by <i>pmk-1</i> ( <i>km25</i> ) (Bond)	228	2.00E-231
C6	Microbes	down by <i>B. thuringiensis</i> at 12h (BT247, 1:2)	362	0
	Microbes	Down infected by <i>X. Nematophila</i>	200	7.49E-195
	DAF/Insulin/food	UP by <i>daf-16</i> mutant under <i>daf-2;rsks-1</i> mutant	235	4.26E-180
	Mutants	down by <i>aak-2</i> mutant	196	8.09E-169
	Mutants	down by <i>crtc-1</i> and <i>aak-2</i> mutant	198	3.68E-159
	DAF/Insulin/food	down by Fasting	131	1.65E-153

<sup>a</sup>The enriched gene sets are ordered according to the cluster they are classified under

<sup>b</sup>Counts shows the number of genes present in the current data set and that from the database



**Table S3: A selection of enriched GO terms, gene counts, statistical significance, and the list of all the genes in the respective enrichment (on the significant transcriptome and proteome combined dataset).**

Cluster	Term	Count	FDR	All genes
1	GO:0055114~oxidation-reduction process	16	5.05E-08	<i>K09H11.1, dao-3, Y38F1A.6, ddo-2, alh-12, F32D8.12, F53C11.3, ACOX-1, R04B5.5, alh-10, dhs-2, CYP-13A12, CTL-2, fat-6, hpd-1, F58F9.7</i>
	GO:0008152~metabolic process	17	5.42E-05	<i>K09H11.1, pho-11, alh-12, tre-3, Y25C1A.13, FAAH-3, aagr-1, acox-1, icl-1, alh-10, W01A11.1, C02D5.4, ugt-23, F58A6.1, F13H6.3, acs-22, F58F9.7</i>
2	GO:0045087~innate immune response	18	3.37E-14	<i>C17H12.8, ZK6.11, gst-13, F55G11.2, hpo-6, ZK1055.7, irg-3, K11H12.4, F19C7.1, lys-1, F54E2.1, C29F3.7, pgp-3, dod-17, C49C3.9, M60.2, dod-19, clec-41</i>
	GO:0050829~defense response to Gram-negative bacterium	6	4.56E-04	<i>C17H12.8, lys-1, pgp-3, F55G11.4, C49C3.9, M60.2</i>
	GO:0005882~intermediate filament	3	0.008727925	<i>ifp-1, ifd-2, ifb-2</i>
5	GO:0031672~A band	3	0.003184548	<i>unc-89, unc-22, myo-3</i>
	GO:0005198~structural molecule activity	4	0.011489773	<i>deb-1, ifd-1, ifc-1, let-2</i>
	GO:0003779~actin binding	4	0.020750636	<i>vab-10, tni-1, deb-1, myo-3</i>
6	GO:0003824~catalytic activity	19	4.92E-05	<i>sucl-1, sur-5, gst-42, spl-1, acs-1, F38B6.4, cth-2, sptl-3, T09B4.8, T05C3.6, R102.4, UCR-2.2, AAGR-4, F08F3.4, bas-1, spds-1, aman-1, cbl-1, R12C12.1</i>
	GO:0005604~basement membrane	5	4.10E-04	<i>lam-2, nid-1, lam-1, ost-1, unc-52</i>
	GO:0005581~collagen trimer	10	4.10E-04	<i>col-156, col-124, col-103, col-140, col-80, col-19, col-93, col-81, col-160, col-98</i>
	GO:0042302~structural constituent of cuticle	10	0.014424327	<i>col-156, col-124, col-103, col-140, col-80, col-19, col-93, col-81, col-160, col-98</i>
	GO:0006629~lipid metabolic process	11	0.016412323	<i>elo-5, sur-5, sams-1, Y54G2A.45, spl-1, pept-1, let-767, fat-7, sptl-3, F28H7.3, elo-2</i>

**Table S4: Gene network analysis used to prioritize genes in cluster 2 combined dataset using WormNet v2\*, a probabilistic functional gene network model.**

Gene name	Rank	Score	Connected genes
C29F3.7	1	3.25	<i>ugt-48 C31H5.6 ifp-1 ckb-4 pcp-2 fbxa-72 F54E2.1 F55G11.2 F55G11.4 dod-17 K11H12.4 M60.2 gst-13 ZC416.6 ugt-16 pgp-3 dod-19 ZK6.11</i>
M60.2	2	3.04	<i>C29F3.7 C31H5.6 ifp-1 ckb-4 pcp-2 fbxa-72 K11H12.4 crt-1 ZC416.6 pgp-3</i>
ZC416.6	3	3.01	<i>ugt-48 C29F3.7 C31H5.6 ifp-1 ifb-2 ckb-4 pcp-2 K11H12.4 M60.2 pgp-3</i>
<i>ifp-1</i>	4	3.01	<i>ugt-48 C29F3.7 C31H5.6 ifb-2 ckb-4 pcp-2 fbxa-72 F40F4.6 M60.2 ZC416.6 pgp-3</i>
<i>dod-19</i>	5	2.84	<i>clec-41 C29F3.7 C49C3.9 F19C7.1 fbxa-72 clec-65 F54E2.1 F55G11.4 bcat-1 K06G5.1 lys-1</i>
F19C7.1	6	2.82	<i>lec-9 F55G11.2 F56C9.7 bcat-1 K06G5.1 T21H3.1 lys-1 dod-19 ZK6.11</i>
<i>lys-1</i>	7	2.74	<i>clec-41 C17H12.8 F19C7.1 clec-65 F37C4.5 F55G11.2 F56C9.7 bcat-1 K06G5.1 T21H3.1 dod-19 ZK6.11</i>
<i>pcp-2</i>	8	2.46	<i>C29F3.7 ifp-1 ckb-4 F40F4.6 M60.2 ZC416.6</i>
ZK6.11	9	2.46	<i>C29F3.7 C49C3.9 F19C7.1 fbxa-72 F54E2.1 F55G11.2 F55G11.4 T21H3.1 lys-1</i>
<i>pgp-3</i>	10	2.25	<i>ugt-48 C29F3.7 ifp-1 M60.2 abt-4 ZC416.6 ugt-16</i>
<i>ckb-4</i>	11	2.18	<i>C29F3.7 C31H5.6 ifp-1 pcp-2 fbxa-72 M60.2 ZC416.6</i>
<i>ugt-16</i>	12	2.01	<i>ugt-48 C29F3.7 pgp-3</i>
K06G5.1	13	1.9	<i>C17H12.8 C49C3.9 F19C7.1 clec-65 F37C4.5 F54E2.1 F56C9.7 bcat-1 T21H3.1 lys-1 dod-19</i>
<i>ugt-48</i>	14	1.84	<i>C29F3.7 ifp-1 ZC416.6 ugt-16 pgp-3</i>
C31H5.6	15	1.82	<i>C29F3.7 ifp-1 ifb-2 ckb-4 fbxa-72 M60.2 ZC416.6</i>
<i>clec-65</i>	16	1.77	<i>lec-9 C17H12.8 C49C3.9 F37C4.5 F55G11.4 bcat-1 K06G5.1 T21H3.1 lys-1 dod-19</i>
C17H12.8	17	1.75	<i>clec-65 F55G11.2 K06G5.1 lys-1</i>
F54E2.1	18	1.74	<i>clec-41 C29F3.7 C49C3.9 fbxa-72 bcat-1 K06G5.1 T21H3.1 dod-19 ZK6.11</i>
T21H3.1	19	1.7	<i>F19C7.1 clec-65 F37C4.5 F54E2.1 F56C9.7 bcat-1 K06G5.1 lys-1 ZK6.11</i>
<i>fbxa-72</i>	20	1.67	<i>C29F3.7 C31H5.6 ifp-1 ckb-4 F54E2.1 M60.2 dod-19 ZK6.11</i>
<i>bcat-1</i>	21	1.58	<i>clec-41 cdd-1 F19C7.1 clec-65 F37C4.5 F54E2.1 K06G5.1 T21H3.1 lys-1 dod-19</i>
F55G11.2	22	1.52	<i>C17H12.8 C29F3.7 F19C7.1 pgp-7 lys-1 ZK6.11</i>
<i>clec-41</i>	23	1.44	<i>C49C3.9 F54E2.1 bcat-1 lys-1 dod-19</i>
F37C4.5	24	1.44	<i>C49C3.4 clec-65 bcat-1 K06G5.1 T21H3.1 lys-1</i>

C49C3.9	25	1.42	<i>clec-41 clec-65 F54E2.1 K06G5.1 dod-19 ZK6.11</i>
F40F4.6	26	1.38	<i>ifp-1 C49C3.4 pcp-2 abt-4</i>
F56C9.7	27	1.36	<i>F19C7.1 K06G5.1 T21H3.1 lys-1</i>
<i>ifb-2</i>	28	1.31	<i>lec-9 C31H5.6 ifp-1 ZC416.6</i>
<i>dod-17</i>	29	1.26	<i>C29F3.7 pgp-7</i>
<i>pgp-7</i>	30	1.25	<i>F55G11.2 dod-17</i>
F55G11.4	31	1.23	<i>C29F3.7 clec-65 dod-19 ZK6.11</i>
<i>lec-9</i>	32	1.21	<i>ifb-2 F19C7.1 clec-65</i>
C49C3.4	33	1.21	<i>F37C4.5 F40F4.6 abt-4</i>
K11H12.4	34	1.14	<i>C29F3.7 M60.2 ZC416.6</i>
<i>abt-4</i>	35	1.12	<i>C49C3.4 F40F4.6 pgp-3</i>
<i>cdd-1</i>	36	0.92	<i>bcat-1</i>
<i>crt-1</i>	37	0.88	<i>M60.2</i>
<i>gst-13</i>	38	0.78	<i>C29F3.7</i>

\*WormNet v2 was used to query the genes in the cluster 2 of the combined transcriptome and proteome dataset, with the 51 genes that were significantly differentially expressed in response to at least of one the treatment conditions compared to the worm grown on only OP50. 38 of the 51 genes were connected within the network and are ordered based on WormNet score, which is based on the number of connections that gene has and the strength of the evidence for those connections (score). C= number of connected genes to the cluster 2 combined dataset.

## REFERENCES

- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169.
- Baugh, L. R., DeModena, J., & Sternberg, P. W. (2009). RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science*, *324*(5923), 92–94.
- Berg, M., Stenuit, B., Ho, J., Wang, A., Parke, C., Knight, M., Alvarez-Cohen, L., & Shapira, M. (2016a). Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *The ISME Journal*, *10*(8), 1998–2009.
- Berg, M., Zhou, X. Y., & Shapira, M. (2016b). Host-specific functional significance of *Caenorhabditis* gut commensals. *Frontiers in Microbiology*, *7*(1622).
- Bischof, L. J., Kao, C.-Y., Los, F. C. O., Gonzalez, M. R., Shen, Z., Briggs, S. P., van der Goot, F. G., & Aroian, R. V. (2008). Activation of the Unfolded Protein Response is required for defenses against bacterial pore-forming toxin *in vivo*. *PLOS Pathogens*, *4*(10), e1000176.
- Boehnisch, C., Wong, D., Habig, M., Isermann, K., Michiels, N. K., Roeder, T., May, R. C., & Schulenburg, H. (2011). Protist-type lysozymes of the nematode *Caenorhabditis elegans* contribute to resistance against pathogenic *Bacillus thuringiensis*. *PLoS ONE*, *6*(9), e24619.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
- Bosch, T. C. G., & McFall-Ngai, M. J. (2011). Metaorganisms as the new frontier. *Zoology*, *114*(4), 185–190.
- Bulcha, J. T., Giese, G. E., Ali, M. Z., Lee, Y. U., Walker, M. D., Holdorf, A. D., Yilmaz, L. S., Brewster, R. C., & Walhout, A. J. M. (2019). A persistence detector for metabolic network rewiring in an animal. *Cell Reports*, *26*(2), 460–468.
- Cabreiro, F., & Gems, D. (2013). Worms need microbes too: Microbiota, health and aging in *Caenorhabditis elegans*. *EMBO Molecular Medicine*, *5*(9), 1–11.
- Caragata, E. P., Rancès, E., Hedges, L. M., Gofton, A. W., Johnson, K. N., O’Neill, S. L., & McGraw, E. A. (2013). Dietary cholesterol modulates pathogen blocking by wolbachia. *PLoS Pathogens*, *9*(6), e1003459.
- Cassidy, L., Petersen, C., Treitz, C., Dierking, K., Schulenburg, H., Leippe, M., & Tholey, A. (2018). The *Caenorhabditis elegans* proteome response to naturally associated microbiome members of the genus *Ochrobactrum*. *Proteomics*, *18*(8), 1700426.
- de Hoon, M. J. L., Imoto, S., Nolan, J., & Miyano, S. (2004). Open source clustering software. *Bioinformatics*, *20*(9), 1453–1454.
- Deng, X., Hiatt, J. B., Nguyen, D. K., Ercan, S., Sturgill, D., Hillier, L. W., Schlesinger, F., Davis, C. A., Reinke, V. J., Gingeras, T. R., Shendure, J., Waterston, R. H., Oliver, B., Lieb, J. D., & Disteche, C. M. (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nature Genetics*, *43*(12), 1179–1185.
- Dirksen, P., Assié, A., Zimmermann, J., Zhang, F., Tietje, A.-M., Marsh, S. A., Félix, M.-A., Shapira, M., Kaleta, C., Schulenburg, H., & Samuel, B. S. (2020). CeMbio - The *Caenorhabditis elegans* microbiome resource. *G3: Genes, Genomes, Genetics*, *10*(8).
- Dirksen, P., Marsh, S. A., Braker, I., Heitland, N., Wagner, S., Nakad, R., Mader, S., Petersen, C., Kowallik, V., Rosenstiel, P., Félix, M.-A., & Schulenburg, H. (2016). The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biology*, *14*(38), 1–16.

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21.
- Douglas, A. E. (2019). Simple animal models for microbiome research reviews. *Nature Reviews Microbiology*, 17, 764–775.
- Evans, E. A., Chen, W. C., & Tan, M. W. (2008). The DAF-2 insulin-like signaling pathway independently regulates aging and immunity in *C. elegans*. *Aging Cell*, 7(6), 879–893.
- Ezcurra, M. (2018). Dissecting cause and effect in host-microbiome interactions using the combined worm-bug model system. *Biogerontology*, 19, 1–12.
- Fischbach, M. A. (2018). Microbiome: Focus on Causation and Mechanism. *Cell*, 174(4), 785–790.
- Gallotta, I., Sandhu, A., Peters, M., Haslbeck, M., Jung, R., Agilkaya, S., Blersch, J. L., Rödelsperger C., Röseler W., Huang C., Sommer R. J., David, D. C. (2020). Extracellular proteostasis prevents aggregation during pathogenic attack. *Nature*, 584(7821), 410–414.
- Garsin, D. A., Villanueva, J. M., Begun, J., Kim, D. H., Sifri, C. D., Calderwood, S. B., Ruvkun, G., & Ausubel, F. M. (2003). Long-lived *C. elegans daf-2* mutants are resistant to bacterial pathogens. *Science*, 300(5627), 1921.
- Geisler, F., Coch, R. A., Richardson, C., Goldberg, M., Denecke, B., Bossinger, O., & Leube, R. E. (2019). The intestinal intermediate filament network responds to and protects against microbial insults and toxins. *Development*, 146(2).
- Geisler, F., & Leube, R. E. (2016). Epithelial intermediate filaments: Guardians against microbial infection? *Cells*, 5(3).
- Goya, M. E. et al. (2020) Probiotic *Bacillus subtilis* protects against  $\alpha$ -synuclein aggregation in *C. elegans*, *Cell Reports*, 30(2), 367–380.e7.
- Haller, D. (Ed.) (2018). *The Gut Microbiome in Health and Disease*. Cham, Switzerland: Springer International Publishing.
- Hamilton, P. T., Peng, F., Boulanger, M. J., & Perlman, S. J. (2016). A ribosome-inactivating protein in a *Drosophila* defensive symbiont. *Proceedings of the National Academy of Sciences of the United States of America*, 113(2), 350–355.
- Hasin, Y., Seldin, M., & Lusi, A. (2017). Multi-omics approaches to disease. *Genome Biology*, 18(1), 83.
- Hasshoff, M., Böhnisch, C., Tonn, D., Hasert, B., & Schulenburg, H. (2007). The role of *Caenorhabditis elegans* insulin-like signaling in the behavioral avoidance of pathogenic *Bacillus thuringiensis*. *The FASEB Journal*, 21(8), 1801–1812.
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57.
- Huffman, D. L., Abrami, L., Sasik, R., Corbeil, J., van der Goot, F. G., & Aroian, R. V. (2004). Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins. *Proceedings of the National Academy of Sciences*, 101(30), 10995–11000.
- Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., & Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols*, 14(1), 68–85.
- Iatsenko, I., Sinha, A., Rödelsperger, C., & Sommer, R. J. (2013). New role for DCR-1/Dicer in *Caenorhabditis elegans* innate immunity against the highly virulent bacterium *Bacillus thuringiensis* DB27. *Infection and Immunity*, 81(10), 3942–3957.
- Ideo, H., Fukushima, K., Gengyo-Ando, K., Mitani, S., Dejima, K., Nomura, K., & Yamashita, K. (2009). A *Caenorhabditis elegans* glycolipid-binding galectin functions in host

- defense against bacterial infection. *The Journal of Biological Chemistry*, 284(39), 26493–26501.
- Johnke, J., Dirksen, P., & Schulenburg, H. (2020). Community assembly of the native *C. elegans* microbiome is influenced by time, substrate and individual bacterial taxa. *Environmental Microbiology*, 22(4), 1265–1279. <https://doi.org/10.1111/1462-2920.14932>
- Johnstone, I. L. (2000). Cuticle collagen genes expression in *Caenorhabditis elegans*. *Trends in Genetics*, 16(1), 21–27.
- Johnstone, I. L., & Barry, J. D. (1996). Temporal reiteration of a precise gene expression pattern during nematode development. *The EMBO Journal*, 15(14), 3633–3639.
- Kao, C.-Y., Los, F. C. O., Huffman, D. L., Wachi, S., Kloft, N., Husmann, M., Karabrahimi, V., Schwartz, J.-L., Bellier, A., Ha, C., Sagong, Y., Fan, H., Ghosh, P., Hsieh, M., Hsu, C.-S., Chen, L., & Aroian, R. V. (2011). Global functional analyses of cellular responses to pore-forming toxins. *PLOS Pathogens*, 7(3), e1001314.
- Kennedy, E. A., King, K. Y., & Baldridge, M. T. (2018). Mouse microbiota models: Comparing germ-free mice and antibiotics treatment as tools for modifying gut bacteria. *Frontiers in Physiology*, 9(1534).
- Khoruts, A. (2018). Targeting the microbiome: From probiotics to fecal microbiota transplantation. *Genome Medicine*, 10(80).
- Kim, D. H., Feinbaum, R., Alloing, G., Emerson, F. E., Garsin, D. A., Inoue, H., Tanaka-Hino, M., Hisamoto, N., Matsumoto, K., Tan, M.-W., & Ausubel, F. M. (2002). A conserved p38 map kinase pathway in *Caenorhabditis elegans* innate immunity. *Science*, 297(5581), 623–626.
- Kim, Dennis H., & Ewbank, J. J. (2018). Signaling in the innate immune response, WormBook, ed. The *C. elegans* Research Community, WormBook.
- Kissoyan, K. A. B., Drechsler, M., Stange, E.-L., Zimmermann, J., Kaleta, C., Bode, H. B., & Dierking, K. (2019). Natural *C. elegans* microbiota protects against infection via production of a cyclic lipopeptide of the viscosin group. *Current Biology*, 29(6), 1030–1037.
- Krediet, C. J., Ritchie, K. B., Alagely, A., & Teplitski, M. (2013). Members of native coral microbiota inhibit glycosidases and thwart colonization of coral mucus by an opportunistic pathogen. *The ISME Journal*, 7(5), 980–990.
- Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A. G., & Marcotte, E. M. (2010). Predicting genetic modifier loci using functional gene networks. *Genome Research*, 20(8), 1143–1153.
- Lemon, K. P., Armitage, G. C., Relman, D. A., & Fischbach, M. A. (2012). Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med*, 4(137), 137–142.
- MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J., & Walhout, A. J. M. (2013). Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell*, 153(1), 240–252.
- Mallo, G. V., Kurz, C. L., Couillault, C., Pujol, N., Granjeaud, S., Kohara, Y., & Ewbank, J. J. (2002). Inducible antibacterial defense system in *C. elegans*. *Current Biology*, 12(14), 1209–1214.
- Manzoni, C., Kia, D. A., Vandrovcova, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, 19(2), 286–302.
- McGhee, J. D., Fukushige, T., Krause, M. W., Minnema, S. E., Goszczynski, B., Gaudet, J., Kohara, Y., Bossinger, O., Zhao, Y., Khattra, J., Hirst, M., Jones, S. J. M., Marra, M. A., Ruzanov, P., Warner, A., Zapf, R., Moerman, D. G., & Kalb, J. M. (2009). ELT-2 is the

- predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult. *Developmental Biology*, 327(2), 551–565.
- Mesbahi, H., Pho, K. B., Tench, A. J., Guerrero, V. L. L., & MacNeil, L. T. (2020). Cuticle collagen expression is regulated in response to environmental stimuli by the GATA transcription factor ELT-3 in *Caenorhabditis elegans*. *Genetics*, 215(2), 483–495.
- Mohri-Shiomi, A., & Garsin, D. A. (2008). Insulin signaling and the heat shock response modulate protein homeostasis in the *Caenorhabditis elegans* intestine during infection. *Journal of Biological Chemistry*, 283(1), 194–201.
- Montalvo-Katz, S., Huang, H., Appel, M. D., Berg, M., & Shapira, M. (2013). Association with soil bacteria enhances p38-dependent infection resistance in *Caenorhabditis elegans*. *Infection and Immunity*, 81(2), 514–520.
- Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., & Pettersson, S. (2012). Host-gut microbiota metabolic interactions. *Science*, 336(6086), 1262–1267.
- Pees, B., Kloock, A., Nakad, R., Barbosa, C., & Dierking, K. (2017). Enhanced behavioral immune defenses in a *C. elegans* C-type lectin-like domain gene mutant. *Developmental and Comparative Immunology*, 74, 237–242.
- Pees, B., Yang, W., Kloock, A., Petersen, C., Peters, L., Fan, L., Friedrichsen, M., Butze, S., Zárate-Potes, A., Schulenburg, H., Dierking, K. (2021). Effector and regulator: Diverse functions of *C. elegans* C-type lectin-like domain proteins. *PLoS Pathogens*, in press.
- Peters L. (2020). *The effect of the C. elegans microbiota on intestinal epithelium structure*. [Master thesis, Christian Albrechts Universität zu Kiel].
- Radeke, L. J., & Herman, M. A. (2020). Identification and characterization of differentially expressed genes in *Caenorhabditis elegans* in response to pathogenic and non-pathogenic *Stenotrophomonas maltophilia*. *BMC Microbiology*, 20(1), 170.
- Reinke, V., Smith, H. E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S. J. M., Davis, E. B., Scherer, S., Ward, S., & Kim, S. K. (2000). A global profile of germline gene expression in *C. elegans*. *Molecular Cell*, 6(3), 605–616.
- Ripert, G., Racedo, S. M., Elie, A.-M., Jacquot, C., Bressollier, P., & Urdaci, M. C. (2016). Secreted compounds of the probiotic *Bacillus clausii* strain o/c inhibit the cytotoxic effects induced by *Clostridium difficile* and *Bacillus cereus* toxins. *Antimicrobial Agents and Chemotherapy*, 60(6), 3445–3454.
- Rolhion, N., & Chassaing, B. (2016). When pathogenic bacteria meet the intestinal microbiota. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1707).
- Saldanha, A. J. (2004). Java Treeview-extensible visualization of microarray data. *Bioinformatics*, 20(17), 3246–3248.
- Sampson, T. R., & Mazmanian, S. K. (2015). Control of brain development, function, and behavior by the microbiome. *Cell Host and Microbe*, 17(5), 565–576.
- Samuel, B. S., Rowedder, H., Braendle, C., Félix, M. A., & Ruvkun, G. (2016). *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), E3941–E3949.
- Sassone-Corsi, M., & Raffatellu, M. (2015). No vacancy: how beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *The Journal of Immunology*, 194(9), 4081–4087.
- Schulenburg, H., & Müller, S. (2004). Natural variation in the response of *Caenorhabditis elegans* towards *Bacillus thuringiensis*. *Parasitology*, 128(4), 433–443.
- Schwarz, E. M., Kato, M., & Sternberg, P. W. (2012). Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 109(40), 16246–16251.

- Sellegounder, D., Liu, Y., Wibisono, P., Chen, C. H., Leap, D., & Sun, J. (2019). Neuronal GPCR NPR-8 regulates *C. elegans* defense against pathogen infection. *Science Advances*, 5(11).
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504.
- Shapira, M., Hamlin, B. J., Rong, J., Chen, K., Ronen, M., & Tan, M.-W. (2006). A conserved role for a GATA transcription factor in regulating epithelial innate immune responses. *Proceedings of the National Academy of Sciences*, 103(38), 14086–14091.
- Shi, J., Yang, W., Chen, M., Du, Y., Zhang, J., & Wang, K. (2011). AMD, an automated motif discovery tool using stepwise refinement of gapped consensus. *PLoS ONE*, 6(9), E24576.
- Singh, R. N., & Sulston, J. E. (1978). Some observations on moulting in *Caenorhabditis elegans*. *Nematologica*, 24(1), 63–71.
- Sison-Mangus, M. P., Mushegian, A. A., & Ebert, D. (2015). Water fleas require microbiota for survival, growth and reproduction. *The ISME Journal*, 9(1), 59–67.
- Stiernagle, T. (2006). Maintenance of *C. elegans*. *WormBook: The Online Review of C. elegans Biology*.
- Taffoni, C., & Pujol, N. (2015). Mechanisms of innate immunity in *C. elegans* epidermis. *Tissue Barriers*, 3(4).
- Treitz, C., Cassidy, L., Höckendorf, A., Leippe, M., & Tholey, A. (2015). Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematicidal *Bacillus thuringiensis*. *Journal of Proteomics*, 113, 337–350.
- Troemel, E. R., Chu, S. W., Reinke, V., Lee, S. S., Ausubel, F. M., & Kim, D. H. (2006). p38 MAPK regulates expression of immune response genes and contributes to longevity in *C. elegans*. *PLOS Genetics*, 2(11), e183.
- Visvikis, O., Ihuegbu, N., Labed, S. A., Luhachack, L. G., Alves, A.-M. F., Wollenberg, A. C., Stuart, L. M., Stormo, G. D., & Irazoqui, J. E. (2014). Innate host defense requires TFEB-mediated transcription of cytoprotective and antimicrobial genes. *Immunity*, 40(6), 896–909.
- Vogt, S. L., & Finlay, B. B. (2017). Gut microbiota-mediated protection against diarrheal infections. *Journal of travel medicine*, 24(1), S39–S43.
- Watson, E., Macneil, L. T., Ritter, A. D., Yilmaz, L. S., Rosebrock, A. P., Caudy, A. A., & Walhout, A. J. M. (2014). Interspecies systems biology uncovers metabolites affecting *C. elegans* gene expression and life history traits. *Cell*, 156(4), 759–770.
- Yang, W., Dierking, K., Esser, D., Tholey, A., Leippe, M., Rosenstiel, P., & Schulenburg, H. (2015). Overlapping and unique signatures in the proteomic and transcriptomic responses of the nematode *Caenorhabditis elegans* toward pathogenic *Bacillus thuringiensis*. *Developmental & Comparative Immunology*, 51(1), 1–9.
- Yang, W., Rosenstiel, P. C., & Schulenburg, H. (2016a). ABSSeq: a new RNA-Seq analysis method based on modelling absolute expression differences. *BMC genomics*, 17(541).
- Yang, W., Dierking, K., & Schulenburg, H. (2016b). WormExp: a web-based application for a *Caenorhabditis elegans* -specific gene expression enrichment analysis. *Bioinformatics*, 32(6), 943–945.
- Yang, W., Petersen, C., Pees, B., Zimmermann, J., Waschina, S., Dirksen, P., Rosenstiel, P., Tholey, A., Leippe, M., Dierking, K., Kaleta, C., & Schulenburg, H. (2019). The inducible response of the nematode *Caenorhabditis elegans* to members of its natural microbiota across development and adult life. *Frontiers in Microbiology*, 10, 1793.



- Zárate-Potes, A., Yang, W., Pees, B., Schalkowski, R., Segler, P., Andresen, B., Haase, D., Nakad, R., Rosenstiel, P., Tetreau, G., Colletier, J.-P., Schulenburg, H., & Dierking, K. (2020). The *C. elegans* GATA transcription factor *elt-2* mediates distinct transcriptional responses and opposite infection outcomes towards different *Bacillus thuringiensis* strains. *PLOS Pathogens*, *16*(9), e1008826.
- Zhang, F., Berg, M., Dierking, K., Félix, M.-A., Shapira, M., Samuel, B. S., & Schulenburg, H. (2017). *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, *8*(485).
- Zimmermann, J., Obeng, N., Yang, W., Pees, B., Petersen, C., Waschina, S., Kissoyan, K. A., Aidley, J., Hoepfner, M. P., Bunk, B., Spröer, C., Leippe, M., Dierking, K., Kaleta, C., & Schulenburg, H. (2019). The functional repertoire contained within the native microbiota of the model nematode *Caenorhabditis elegans*. *The ISME Journal*, *14*, 26–38.



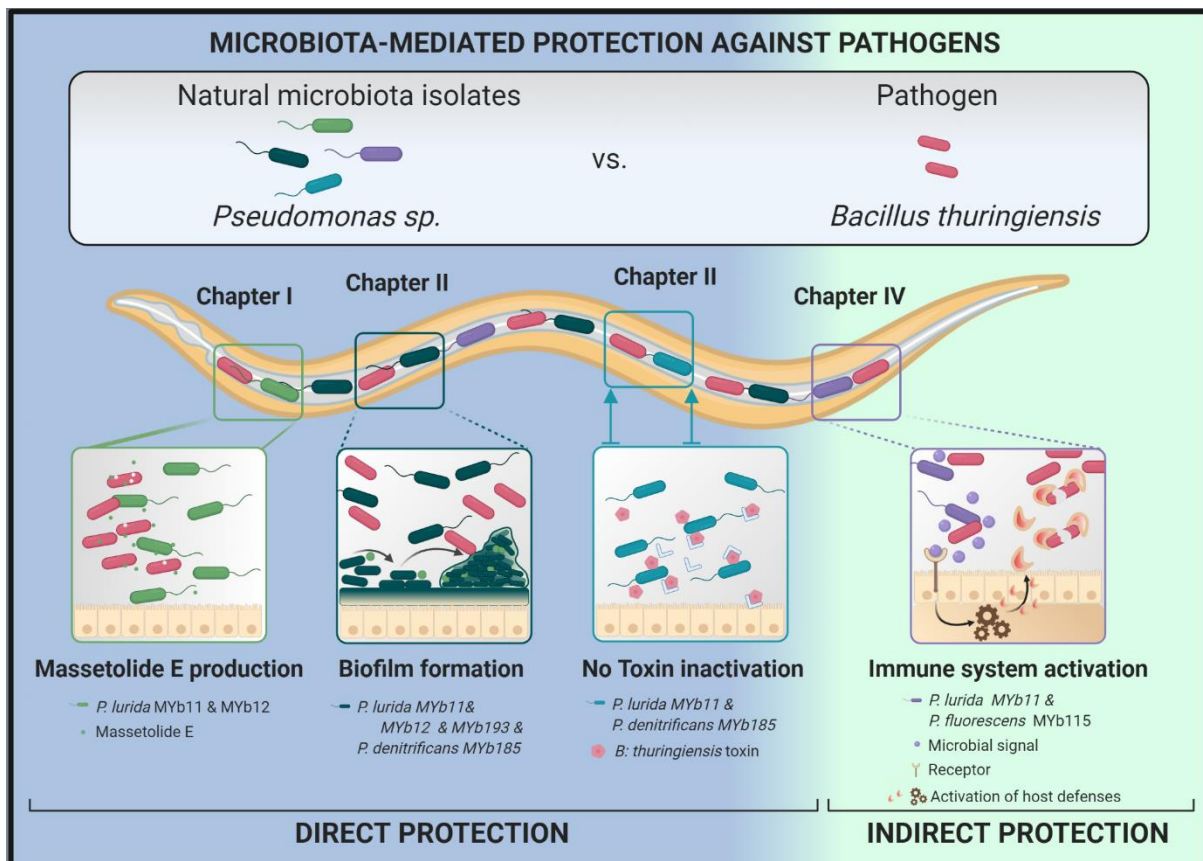
## GENERAL DISCUSSION

Throughout evolution, metazoans and microorganisms have formed complex associations (Sampson and Mazmanian, 2015). In these associations, the microbiota, affect the host's physiology (Douglas, 2019; Mcfall-Ngai et al., 2013). One of these associations' beneficial effects on the host is microbiota-mediated protection against pathogen infection (Masson & Lemaitre, 2017). Microbiota's inherent complexity and that of the host-microbiota-pathogen interactions make understanding microbiota-mediated protection mechanisms more complicated (Ezcurra, 2018; Koch & McFall-Ngai, 2018; Zheng et al., 2020). To untangle these layers of complexity, I used the *C. elegans*-microbiota-pathogen tripartite model system to study microbiota-mediated protection. In this thesis project, I explored the role of *C. elegans* natural microbiota in protecting the host against pathogen infection. Subsequently, I investigated the underlying mechanisms of microbiota-mediated protection on the bacterial and the host side.

### Summary of the main findings of this thesis

First, in **Chapter I**, I identified six natural *C. elegans* microbiota isolates of the genus *Pseudomonas* that protect the worm from pathogen infection with *Bacillus thuringiensis*. Therefore, I mainly focused on these *Pseudomonas* isolates throughout the rest of this work. The protective microbiota isolates were *P. lurida* MYb11, MYb12, and MYb193, *P. fluorescens* MYb115, *P. denitrificans* MYb185, and *P. alkylphenolia* MYb187 (**Figure 1**). Next, I showed that microbiota-mediated protection in *C. elegans* can have distinct protection mechanisms: direct (i.e., host-independent) or indirect (i.e., host-dependent) protection mechanisms (see **Introduction** for more on direct and indirect protection mechanisms). I showed that direct microbiota-mediated protection in *C. elegans* can operate via the cyclic lipopeptide massetolide E production (**Chapter I**). Further, in **Chapter II**, I showed biofilm production of protective microbiota *in vitro*, potentially another protection mechanism, if confirmed *in vivo* (**Figure 1**). In **Chapter II**, I also demonstrated that microbiota isolates cannot inactivate the pathogen toxins; instead, they render the worms more susceptible to toxin exposure, indicating their dual nature in *C. elegans*, where the protective efficacy of the microbiota isolates is contextual and depends on pathogen exposure conditions. This context-dependent nature of MYb11 is also shown *in vitro* and *in silico* growth experiments, in

monoculture and co-culture with another microbiota isolate (i.e., *Ochrobactrum sp.* MYb71) in **Chapter III**, where the nutrient context (i.e., the carbon source in medium) is the central modulator in switching bacterial growth from being competitive to parasitic. Moreover, indirect microbiota-mediated protection may operate via the activation of host defenses, via priming the host's innate immune system and preparing it for a subsequent pathogen attack, as suggested by microbiota-induced differential abundance patterns at both the transcript and protein levels in **Chapter IV (Figure 1)**.



**Figure 1: Summary of findings on microbiota-mediated protection mechanisms in *C. elegans* against the pathogen *B. thuringiensis*.** Protection against *B. thuringiensis* is mediated by the *C. elegans* natural microbiota *Pseudomonas* isolates, *P. lurida* MYb11, *P. lurida* MYb12, *P. fluorescens* MYb115, *P. denitrificans* MYb185, *P. alkylphenolia* MYb187, and *P. lurida* MYb193. Microbiota-mediated protection can occur via direct or indirect protection mechanisms. Direct microbiota-mediated protection was shown to occur via massetolide E, an antimicrobial lipopeptide of the viscosin group, production by *P. lurida* MYb11 and *P. lurida* MYb12 (**Chapter I**). Direct microbiota-mediated protection is also suggested to occur via biofilm production by *P. lurida* MYb11, *P. lurida* MYb12, *P. denitrificans* MYb185, and *P. lurida* MYb193 (**Chapter II**). Direct microbiota-mediated protection is not achieved via toxin inactivation; instead, the protective microbiota isolates *P. lurida* MYb11, and *P. denitrificans* MYb185 are rendered more pathogenic to the worm in the presence of purified pathogen toxins (**Chapter II**). Indirect microbiota-mediated protection is induced by *P. lurida* MYb11 and *P. fluorescens* MYb115, via the up-regulation of gene expression, at the transcript and protein level of genes involved in the worm innate immune response; in addition to priming the host defense even in the absence of a

pathogen (**Chapter IV**). Solid lines: indicate that empirical evidence is available, arrowheads: indicate that the reverse of the expected phenotype was observed (i.e., harmful effects instead of protective one), dashed lines: indicate that further empirical evidence is needed for verification. This figure was created with Biorender.com.

In the following sections, I will discuss the insights gained from this thesis in the light of the current literature focusing on three concepts: 1) the context-dependent dual nature of microbiota in the section "Microbiota: a friend or a foe?", 2) microbiota as an extension of the *C. elegans* immune system, and 3) the transition of *C. elegans* from a "living test tube" to a metaorganism. I will then discuss the limitations of this work and propose future research perspectives based on this project and an outlook on microbiota research.

## **Microbiota: a friend or a foe?**

Our view on microbiota-mediated protection more often than not tends to be host-centered, likely since we are hosts ourselves. However, hosts and microorganisms coexist in nature, where each organism has its own priorities and is subject to intense selective pressures. While all organisms, hosts and microorganisms alike try to survive, each has a set of strategies to do so. The microorganisms have to interact and often compete for space, resources, and ultimately survival. These microbial interactions might benefit the host in some contexts and be detrimental to the host in others, which could be manifested in the dual nature of microbiota, as shown in this thesis (**Chapters I & II**).

The dual nature of microbiota raises the question of whether the microbiota is to be considered a friend or foe? It is easy to label a microbiota isolate as "friend" and a pathogen as "foe"; however, symbiotic associations appear to be more complicated than this (Dale & Moran, 2006; Douglas, 2008; Kiers & West, 2015). While features of microbes such as microbe-associated molecular patterns (MAMPs) (e.g., lipopolysaccharides) are traditionally referred to as virulence factors, and features such as the production of antimicrobial peptides are traditionally referred to as defense strategies; these features are shared by microbiota and pathogens alike, only used in different contexts (Koch & McFall-Ngai, 2018). Thus, the microbiota has properties that allow it to be alternatively or even simultaneously protective or harmful, such as colonizing the host or producing antibiotics.

There is further evidence that supports the dual nature of microbiota isolates reported in this thesis. For example, Broderick *et al.* have shown how the gypsy moth (*Lymantria dispar L.*) survives exposure to *B. thuringiensis* toxins in the absence of its natural microbiota. However,

when the moth's natural microbiota is present, the Bt toxins, which create pores and make the intestinal epithelium permeable, allow the microbiota to pass through these pores into the insect hemolymph. Thus, the presence of Bt toxins induces the otherwise benign gut microbiota to exert pathogenic effects via invading the bloodstream, causing sepsis and killing the animal (Broderick et al., 2006; Broderick et al., 2009). The microbiota dual effect could also be rooted in host genetics, inducing changes at the level of the host molecular signaling pathways. The Shapira lab has shown that in mutant *C. elegans* strains with disrupted Transforming Growth Factor (TGF) $\beta$ /Bone Morphogenetic Protein (BMP) signaling (i.e., immune-compromised worms), the otherwise beneficial *Enterobacter* microbiota member turns into a pathogenic one (Berg et al., 2019). Similarly, Martin Blaser has reported on the human gut microbiota member, *Helicobacter pylori*, which can be both harmful (i.e., causing ulcer and stomach cancer), and protective (i.e., against esophageal cancer) (Blaser, 2015; Blaser, 2010). Thus, microbiota appears to be more complicated than simply labeled a "friend". Microorganisms appear to swiftly slide across the continuum of commensalism, between mutualism and parasitism, according to the environmental conditions. As Martin Blaser points out in his book, *Missing Microbes*, this phenomenon was first coined amphibiosis by the microbial ecologist Theodore Rosebury in the 1960s (Rosebury, 1962); now, we commonly refer to an amphibiont as pathobiont instead (Blaser, 2015). A pathobiont was initially defined as "a symbiont that is able to promote pathology only when specific genetic or environmental conditions are altered in the host" (Chow & Mazmanian, 2010; Mazmanian et al., 2008). This definition of pathobiont implies the knowledge of normality, i.e., an unaltered state (Jochum & Stecher, 2020). However, in microbiota research, the term pathobiont, although commonly used, remains vague since the assumed normality (i.e., the healthy microbiota) remains undefined in most hosts (Aguirre De Cárcer, 2018; Almeida et al., 2019). Regardless of the terms used, amphibiosis (as initially coined) is common in nature (Iebba et al., 2016), in which the constant natural selection pressure creates many such nuanced interactions (Blaser, 2015). Thereby, there appears to be no inherently "friend" or "foe" classification of microbiota, rather a more fluid and contextual one, depending on environmental and genotypic factors.

From the bacterial perspective, microbes' human-centric classification as friend or foe is irrelevant in the broader context of environmental challenges that threaten microbial survival. While we, the human host, may perceive microbial responses and try to classify them as either protective or pathogenic, from the microbes' perspective, they are simply the opportunities and

challenges faced when living on host tissues and among many other microorganisms (Wiles and Guillemin, 2020). Moreover, to be able to distinguish a microbiota friend from foe successfully, we first need to decipher casual relationships and have a detailed mechanistic-level understanding of the total functional capacity of microbiota (e.g., in all possible microbial interactions, hosts, and environmental conditions) (Jochum & Stecher, 2020). We have yet to accumulate such knowledge. Thereby, this idea prompts a shift in the conceptual framework around host-microbe relationships and microbe classification in general (Koch & McFall-Ngai, 2018). One suggestion would be to refrain from such classifications to avoid confusion and instead adapt the concept of pathogenic potential to microbiota, quantifying the microbiota's virulence capacity considering the changes in the microbiota, host and environmental conditions (Casadevall, 2017; Jochum & Stecher, 2020).

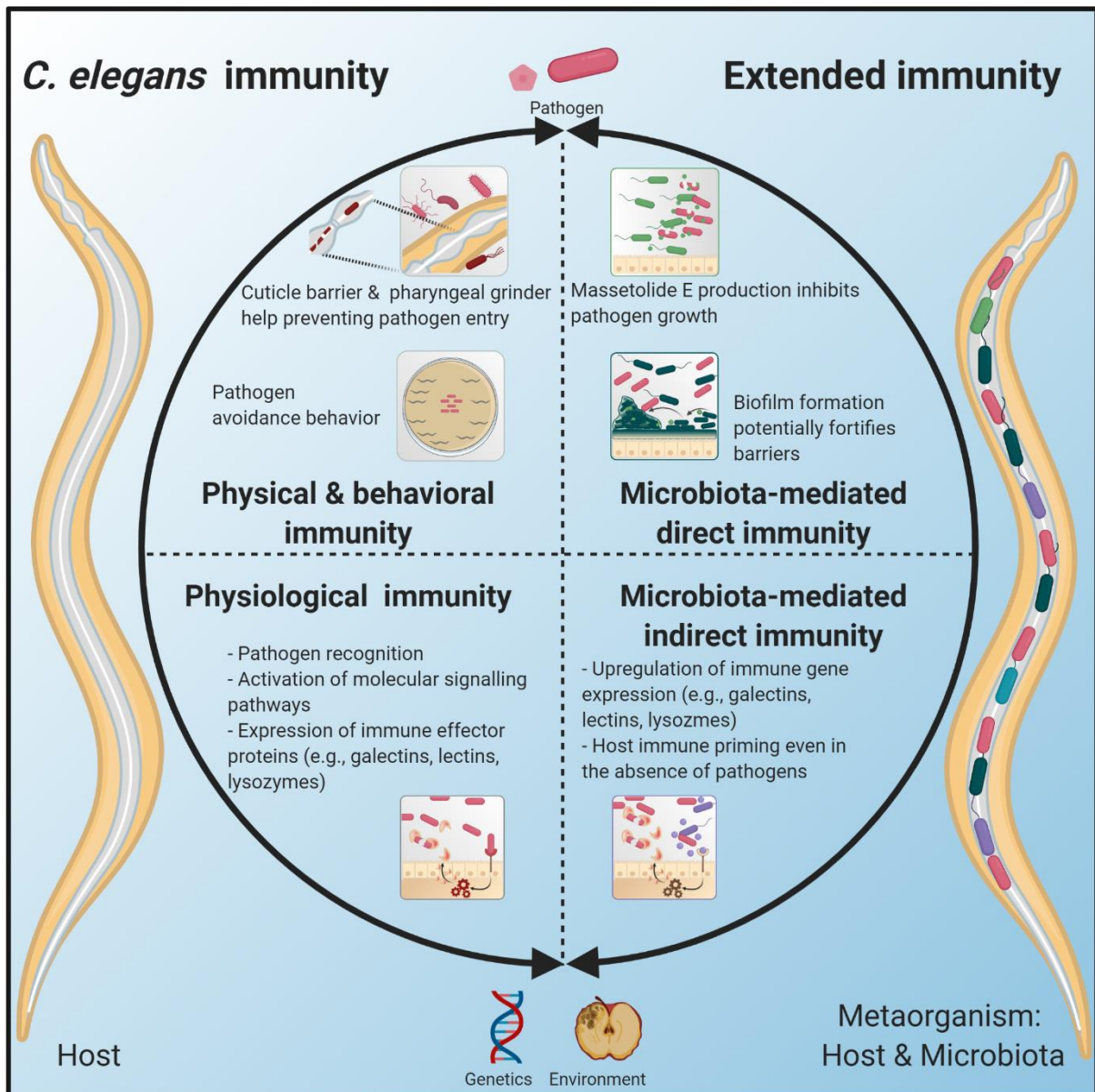
### **Microbiota can act as an extension of the *C. elegans* immune system**

In its classical sense, the immune system distinguishes "self" from "non-self" to ensure the protection of the host against invading pathogens and intruders (i.e., toxins, allergens). Based on this discrimination of self and non-self, and upon identifying the non-self, the immune system, both innate and adaptive, uses the host defensive arsenal to protect the host (Chaplin, 2010). More recently, many researchers have come to view the host as a metaorganism, an entity composed of the host and its microbiota (Bosch & McFall-Ngai, 2011). Others even consider that microbiota changes the human self's perspective by challenging all three classical biological explanations of the individual self; the adaptive immune system, the brain, and the genome (Rees et al., 2018). Thereby, the classical definition of self in describing immune systems, both innate and adaptive, should be adjusted accordingly to include host-microbiota as well. Microbiota can be protective to the host, deploying its own defensive arsenal, which benefits the host, and extends host defensive capacities. Microbiota-mediated protection presented in this thesis and the line of microbiota-mediated protection evidence in plants (Vannier et al., 2019) and animals (see **Introduction**) support the notion that microbiota is an extension of the host immune system with the function of protecting the host, given the contextual settings as discussed earlier.

The protective repertoire of the *C. elegans* innate immune system consists of the following three complementary modes: 1) Physical barriers, such as; the worm cuticle, a barrier between the worm and its environment; and the pharyngeal grinder, that crushes bacteria in preparation

for their mechanical and chemical digestion in the intestine. 2) Behavioral responses, such as pathogen avoidance that decrease the chances of worm-pathogen physical encounter. 3) Physiological innate immune responses, which consist of recognizing a pathogen, the activation of molecular signaling pathways, and the expression of immune effector molecules (Dierking et al., 2016; Schulenburg & Müller, 2004) (**Figure 2**). My findings on microbiota-mediated protection extend the *C. elegans* immune response repertoire to include the following non-exclusive protection mechanisms: microbiota-mediated direct protection via the production of massetolide E, which inhibits the pathogen *B. thuringiensis* (**Chapter I**); microbiota-mediated biofilm formation that could potentially fortify the host defensive barriers against pathogen infection (**Chapter II**); and microbiota-mediated indirect protection mechanism including up-regulation of genes involved in the host immune response, metabolism and structural components (**Chapter IV**). The interplay of the microbiota and the host immune system results in signaling and cascading pathways, which could eventually lead to the production of effector proteins with antimicrobial capacities such as lectins and lysozymes. The crosstalk of host- and microbiota-mediated protection mechanisms extends the worm's defensive repertoire and likely increases survival chances of the metaorganism in its natural habitat under the myriad of fluctuating environmental conditions and genetic variables (**Figure 2**).





**Figure 2: Microbiota-mediated extended immunity against pathogen infection in the *C. elegans* metaorganism.** The *C. elegans* immune repertoire consists of physical immunity, which includes the cuticle and grinder as a physical barrier and disruptor against the pathogen, respectively. It could also include the physical distancing from pathogen via the nematode avoidance behavior. *C. elegans* also has physiological immunity, where the worm's innate response recognizes pathogens and activates a series of pathways resulting in the expression and production of immune effector proteins. *C. elegans* microbiota offers an additional immune repertoire to the worm, thereby extending the worm immune response. Microbiota-mediated immunity can be via direct microbiota-mediated protection such as the production of antimicrobial compounds (e.g., massetolide E) or biofilm production as a fortified barrier against pathogens. Microbiota-mediated protection can also be indirect via the up-regulation of host immune gene expression. Thus, the crosstalk of microbiota with the worm immune system and the activation of signaling pathways lead to immune effector protein production. This figure was created with Biorender.com.

## The transition of *C. elegans* from a "living test tube" into a metaorganism

In its natural environment, *C. elegans* is associated and interacts with a myriad of microorganisms. The nature of microorganism-nematode interactions ranges from microbes serving as food, pathogens, or microbiota to the worms (see **Introduction**). *C. elegans*, being a bacterivore, ingests bacteria as food. The subsequent acquisition of nutrients then influences worm metabolism (MacNeil et al., 2013). The worm feeds and acquires nutrients from bacteria, deploying an armamentarium of mechanical and chemical facilitators to digest bacteria. This armamentarium includes the nematode pharynx grinder, immune and digestive enzymes, peristaltic movements, and the weakly acidic intestinal lumen pH (~4) (Allman et al., 2009; Bender et al., 2013; Dimov & Maduro, 2019). However, worms also get infected with pathogenic bacteria and get colonized with protective microbiota, as shown in this thesis, indicating that bacteria can escape worm digestion, at least to a certain extent, which could then influence worm immunity. Therefore, it is likely that microorganisms have more than one function for the worm (e.g., metabolism and immunity). Such similarities in microorganism-induced responses by the digestive and immune systems could be due to their common evolutionary origin (Broderick, 2016). Nevertheless, this phenomenon seems to apply also to the protective microbiota, which can thus be partly digested by the worm, producing metabolites (e.g., vitamins, amino acids) needed by the worm as shown in **Chapter III** most likely affecting its metabolism, while some escape to colonize the gut and manifest their protective effects on the nematode, affecting its immunity as shown in **Chapters I, II, and IV**.

Although *C. elegans* has been extensively used in biomedical research, this research has completely neglected the worm's natural microbiota instead of treating *C. elegans* as a sort of a "living test tube", implying a rather microbiota-free model organism. While the nematode thrives on a variety of bacteria in its natural habitat, in the laboratory, it is traditionally fed on a single *E. coli* strain: OP50. It is only recently that we have started to understand the ecology of *C. elegans* (Petersen et al., 2015) and to appreciate *C. elegans* as a metaorganism in its own right, with a defined, species-rich, and non-random microbiota (Berg et al., 2016a; Dirksen et al., 2016; Samuel et al., 2016; Zhang et al., 2017).

Following the pioneering work on the worm microbiota, particularly by Dirksen *et al.*, and the isolation of *C. elegans* natural microbiota, and as initial steps in untangling host-microbiota complexity, my colleagues and I began characterizing the *C. elegans* microbiota isolates. In **Chapter I**, I selected a list of 13 microbiota isolates, which are representative of the most abundant genera in the *C. elegans* microbiota (e.g., *Pseudomonas*, *Ochrobactrum*, and *Acinetobacter*), and can persist in the worm gut (Dirksen *et al.*, 2016; Sieber *et al.*, 2018; Zhang *et al.*, 2017). Then, I screened for their potential effects on host fitness, population growth, and host protection capacity against pathogen infection. In **Chapter III**, as a collaborative effort with my colleagues, we characterized the functions of a set of 77 *C. elegans* microbiota isolates using integrative approaches (i.e., whole-genome sequencing, metabolic modeling, and experimentation). We generated a resource that may help understand the nematode's biology in a more ecologically meaningful context. However, we are still only beginning to understand *C. elegans* as a metaorganism. Moreover, our knowledge of the physiological significance of the worm natural microbiota, the extent to which it can influence worm life history characteristics, and the underlying mechanisms thereof is in its infancy.

## Limitations and perspectives

In this thesis project, individual microbiota isolates and the Bristol laboratory strain of *C. elegans* (N2) were used, which could be considered as limitations. Individual microbiota isolates rather than microbiota as a community was used to explore microbiota-mediated protection in *C. elegans*. Microbiota, particularly gut microbiota, is usually described as a dynamic community of a host's resident microbes (Sommer & Bäckhed, 2013). The gut microbiota community hosts a consortium of isolates that likely interact among themselves and with the host, individually or collectively, cooperatively or competitively. Thereby, conclusions made by studying single microbiota isolates, protective of the host, may (Dahan *et al.*, 2020) or may not (Lenhart & White, 2017) hold when the protective microbiota isolates are put together with a microbiota community in the host. My decision to use a binary association between the worm and its microbiota isolate was taken to allow a clear focus on their interactions during a pathogen exposure. The use of such a strategy based on binary associations has proven valuable to study both naturally occurring binary associations and zoom-in on the host responses to specific partners in a community (Koch & McFall-Ngai, 2018). This strategy also offers the opportunity to explore the contributions of individual

microbiota isolates to the community's function on the one hand while allowing better control of the participating partners on the other hand. Thereby, such simplistic models and model systems provide valuable insights that can be extended to the study of more complex interactions (Douglas, 2019). This strategy was in line with our aims to focus on dissecting the underlying mechanisms of microbiota-mediated protection. Therefore, it was beyond the aims and scope of this study to include a community of microbes; however, this can be an exciting avenue of investigation in the future, particularly with the availability of the CeMBio community, composed of 12 natural *C. elegans* microbiota isolates, of which MYb11 is a member (Dirksen et al., 2020).

Another limitation to consider in this thesis is that it may be insufficient to focus only on the reference Bristol laboratory strain of *C. elegans* (N2) in exploring microbiota-mediated protection. Indeed the *C. elegans* N2 strain has been the most commonly used strain to be grown and maintained in laboratory conditions (e.g., constant temperature, unlimited food, monoxenic culture, and two-dimensional living space) and the absence of the nematode's native microbiota for decades (Félix & Braendle, 2010; Sterken et al., 2015). Having been transported from their natural habitat to the artificial laboratory conditions, N2 worms underwent strong and artificial selective pressures that have affected their physiology, phenotype, and genotype (Duveau & Félix, 2012; Sterken et al., 2015). These changes most likely blur our understanding of the nematode biology on which basis we build the conclusions to understand better the evolutionarily conserved pathways in higher organisms (Schulenburg & Müller, 2004). However, years of experimentation on N2 have also yielded vast amounts of data, a sizeable experimental toolkit, and knowledge of N2 biology. This wealth of information is highly experimentally useful, especially when using omics approaches where a reference database is indispensable for the analysis. Platforms containing massive databases on the worm genome, transcriptome, and proteome (e.g., WormBase and UniProt), for instance, are readily available for N2. Thus, N2 use also allows for deeper insights into microbe-mediated protection. However, to compensate for this bias towards N2, I tested microbiota-mediated protection on the wild *C. elegans* strain MY316.

Microbiota-mediated protection was also observed in the MY316 strain, from which many of the natural microbiota isolates used in this thesis were originally isolated (**Chapter I**). Nevertheless, hundreds of wild *C. elegans* isolates have been isolated. Thus, to screen across the entire species would be more insightful in studying the variation in *C. elegans* response to

microbiota and a more laborious approach. Thankfully, the Andersen lab has facilitated this process by establishing a platform, *C. elegans* Natural Diversity Resource (CeNDR) (Cook et al., 2017). CeNDR platform makes genome-wide association (GWA) analyses much more feasible in identifying natural variation to microbiota-mediated protection. As a first step in understanding natural variation in microbiota-mediated protection, initial data on microbiota-mediated protection on wild *C. elegans* strains of the CeNDR divergent set (a group of 12 different wild strains selected for their genetic distance from each other) show a variation in microbiota-mediated protective phenotype in four of the tested strains (Rackow, 2019, Bachelor thesis, unpublished). However, further experiments and screening of hundreds of more *C. elegans* wild strains are needed to properly understand this area. In order for this to be practically feasible, a high-throughput approach and readout are needed. Interestingly, it has been shown that wild *C. elegans* can harbor distinct microbiota communities which could be due to changes in environment (Berg et al., 2016a; Zhang et al., 2017) or host genetics (Berg et al., 2016b; Dirksen et al., 2016; Zhang et al., 2017). Thus, it would be an interesting aspect to explore in the future the *C. elegans* natural variation in responsiveness to microbiota-mediated protection. The natural variation could be explored with the CeNDR platform and a GWA approach. Together with the incorporation of higher throughput measures (e.g., worm sorter or worm-tracker to automatically sort or track worms rather than manually do so), a less laborious readout (e.g., assessing worm locomotion instead of survival), and the combination of wild worm strains (CeNDR) and natural microbiota community (CeMbio) would provide a more natural setting to investigate microbiota-mediated protection in the nematode.

This thesis project's insights further highlight exciting aspects that could be considered in subsequent studies, such as *C. elegans* natural microbiota-mediated protection against human pathogens. In **Chapter I**, we observed that *C. elegans* natural microbiota isolates protect the worm against *P. aeruginosa* PA14 *in vivo*, while other *C. elegans* natural microbiota isolates inhibit *S. aureus* growth *in vitro*. Both pathogens are clinically relevant to humans (Balasubramanian et al., 2017; Horcajada et al., 2019). Therefore, studying these pathogens in the worm-microbiota context may reveal new insights and potentially provide inferences applicable in higher organisms, as was shown earlier with both pathogens in the *C. elegans* without its microbiota (Sifri et al., 2003; Tan et al., 1999).

Furthermore, the wealth of omics data generated in this thesis merits future experiments aiming for functional validation and complementing the multi-omics dataset. It is worth complementing the multi-omics dataset with functional assays using *C. elegans* mutants lacking some of the candidate genes involved in microbiota-mediated protection to understand the underlying mechanistic level pathways and the players involved in more detail. Moreover, *C. elegans* transcriptome level up-regulation of gene expression (**Chapter IV**) does not necessarily signify that the expressed gene is translated into a protein. Therefore, I complemented the transcriptome analysis with that of the proteome. However, even if there is an overlap of gene expression at transcript and protein levels, it still is possible that the translated protein is not functional. Post-translational modifications (e.g., proteolysis, glycosylation, phosphorylation) could change protein function (Hasin et al., 2017; Manzoni et al., 2018). Due to such post-translational modifications, it is also not completely surprising to have discordance between the abundance in transcript level gene expression and protein abundance (Hasin et al., 2017). Thus, metabolome acquisition would offer another valuable dimension on which metabolic processes are activated in the host at a given time and space (Manzoni et al., 2018; Wang et al., 2019). The information on which metabolites are present, and their specific levels would better reflect the microbiota-affected metabolic functions. On top of that, a single-cell multi-omics approach (i.e., the isolation and analysis of individual cells rather than the whole worm) would be an exciting addition to have a higher resolution of the microbiota-induced changes in the worm and to identify cellular or tissue-specific functions. Such an approach would also shed light on the possible scenario of functional worm intestine compartmentalization. Intestinal compartmentalization of microbiota is already established in many other animals, including humans, but has not been shown in worms (The Human Microbiome Project Consortium et al, 2012). Moreover, a single-cell multi-omics approach would immensely benefit from the now available fluorescently labeled protective microbiota isolates (e.g., *P. lurida* MYb11 and *P. fluorescens* MYb115) in tandem with using the Fluorescence-Activated Cell Sorting (FACS) facility. Therefore, complementing our multi-omics dataset with functional analyses, metabolomics, and single-cell transcriptomics can add to our knowledge of microbiota-mediated protection mechanisms in *C. elegans*.

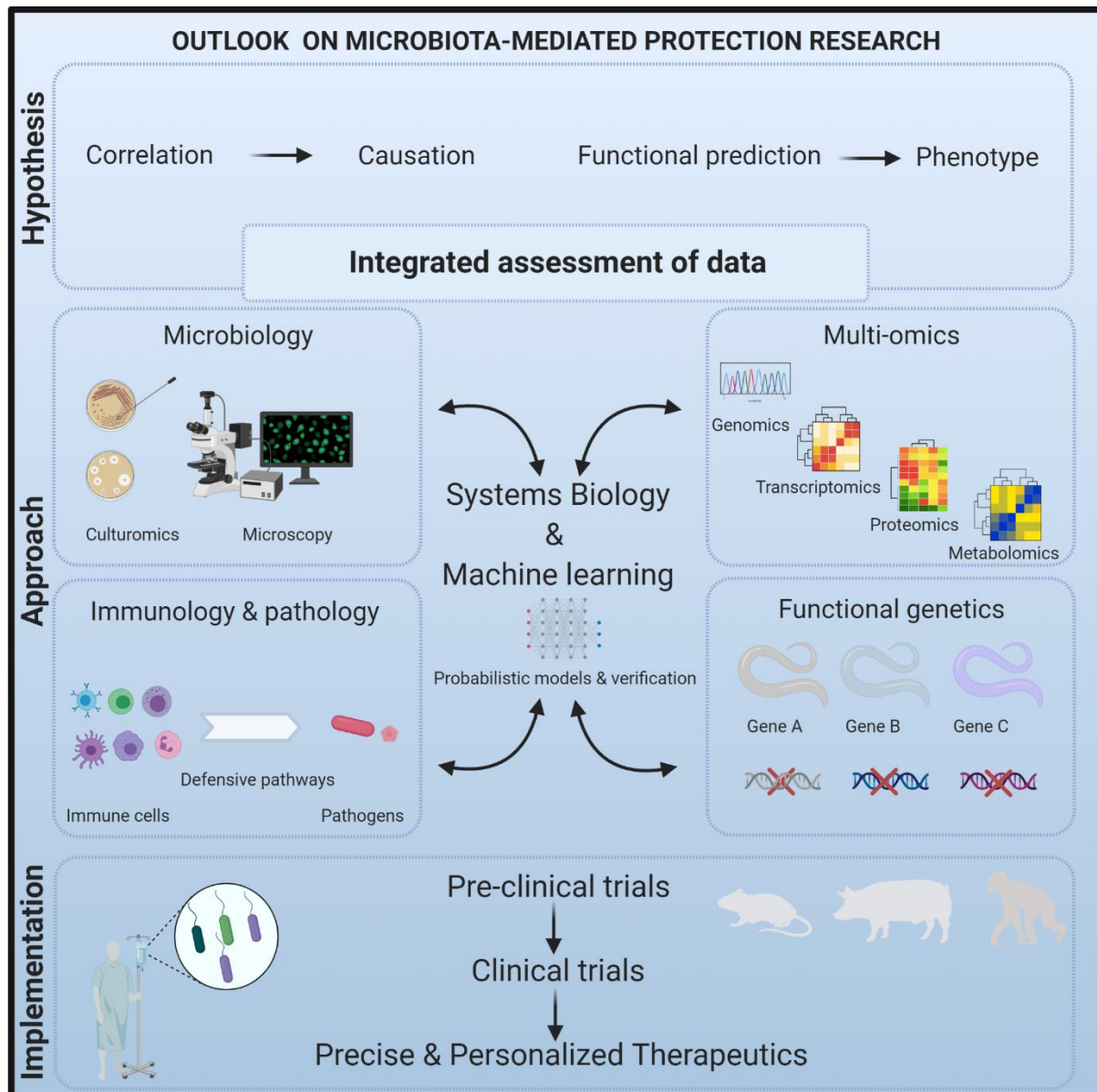
## Outlook on microbiota-mediated protection research

Microbiota affects host health or disease states. This effect can be exploited via microbiota's targeted use to leverage the desired outcome. Microbiota affects host physiology (Douglas, 2019; Mcfall-Ngai et al., 2013) via, for instance, the production of antimicrobial compounds, such as the lipopeptide massetolide E production by MYb11 (**Chapter I**). Microbiota also affects host physiology via the production of essential vitamins, including Vitamin B12 (**Chapter III**), which promote the host's growth (MacNeil et al., 2013). These concepts, defined as probiotic i.e., “live microorganisms which, when administered in adequate amounts, confer a health benefit on the host”, and prebiotic, i.e., the ingestion of nutrients that favor the growth of protective bacteria, are currently used to promote health benefits in humans (Hill et al., 2014; World Health Organization and Food and Agriculture Organization of the United Nations, 2011)(see **Introduction**). Simultaneously, the microbiota is also used to promote disease, such as the use of insect microbiota as a biocontrol strategy (Dennison et al., 2014). The World Mosquito Program (formerly known as the Eliminate Dengue program), for instance, infects the mosquito *Aedes aegypti* with the insect symbiont *Wolbachia pipientis* *wMel* strain. *Wolbachia* reduces the transmission of human viruses such as dengue, zika and chikungunya via inhibiting the replication of the virus in the mosquito, interrupting the transmission cycle of the virus, thus controlling the viral transmission in affected communities (Flores & O'Neill, 2018; O'Neill et al., 2018).

Thus, microbiota affecting host health covers a broad range of applications, from agriculture to public health and the clinic. However, microbiota applications are still used with hesitation, primarily due to the missing causal links between the microbiota and the observed short- and long-term effects. Thus, microbiota-mediated effects and their broad scope of application highlight the necessity to better understand the intricacies of microbiota-induced effects, including the context-, strain-, and dose-dependent specific conditions that induce the desired outcome in the host. Many open questions remain to be answered using microbiota therapeutics, including concerns on the host side (e.g., genetic, lifestyle or geographic variations and their effect on host responsiveness to microbiota), in order to ensure an optimal efficacy of the probiotic and prebiotic effects of the microbiota. Here, the insights we gain from the worm can help extend our understanding of how to approach microbiota research in the future and aim to leverage its protective benefits in the clinic.

In the broader picture of microbiota research, improving microbiota implementation in the clinic could be achieved via some changes in the investigation process (**Figure 3**). First, investigators should focus more on studying causal relationships rather than correlations between the microbiota and their effects on the host. Next, aiming to use the available functional predictions to identify phenotypes would be beneficial. Finally, to shift towards integrating the assessed data from various disciplines for a more rapid transfer of knowledge from the bench-side to the bed-side (Berg et al., 2020; Zheng et al., 2020). Thus, aiming to augment our understanding of microbiota-mediated protection by leaping from reductionism to the global-integrative and more holistic approach (Berg et al., 2020). A holistic approach can be achieved by incorporating inter-disciplinary tools and platforms (Manzoni et al., 2018); including and not limited to: a) microbiological tools (e.g., culturomics, microscopy), b) omics approaches (e.g., transcriptomics, proteomics, metabolomics), c) immunological and pathological tools (e.g., immunological assays, model organisms), and d) functional genomics (e.g., *C. elegans* mutant knockout or knockdown strains). Findings from inter-disciplinary experimentation can then feed into systems biology and machine learning processes to develop probabilistic models (Hasin et al., 2017). Probabilistic models can then be used to inform experimentation and vice versa. Finally, an experimentally verified microbiota with the relevant metadata obtained from such a pipeline can enter rigorous pre-clinical and clinical trials. The chances are that such an approach speeds-up and yields the translation of microbiota-mediated protection into the clinic and supports precision medicine with personalized therapy to ensure better diagnosis, prognosis, and therapeutic options (Manzoni et al., 2018; Q. Wang et al., 2019) (**Figure 3**).





**Figure 3: Outlook on microbiota-mediated protection research.** Schematic representation of the future perspective on microbiota research. Future researchers should shift their hypothesis formulation to focus on causation and phenotype. At the same time, data from various fields of expertise should be assessed in an integrated manner. Thereby, the focus should be on the inter-disciplinary integration of data to produce probabilistic models, taking advantage of systems biology and machine learning advances. Such an integrated assessment of data facilitates knowledge transfer and translation from the bench-side to the bed-side. This pipeline would more likely improve the current microbiota-mediated therapy toward a more precise and personalized one. This figure was created with Biorender.com.

## Concluding remarks

Overall, this thesis project provides new insights into the *C. elegans* natural microbiota. It underscores the *C. elegans* natural microbiota functional significance, particularly in protecting

*C. elegans* from pathogen infection. Also, it reveals a more complex picture of the host-microbiota-pathogen interactions in *C. elegans*, highlighting the context-dependent protective nature of microbiota. Moreover, the findings of this thesis provide insights at the mechanistic level for microbiota-mediated protection, viewed from both the bacterial and the host perspectives. Finally, this work puts forth a repertoire of genomic, transcriptomic, and proteomic datasets on host-microbiota interactions in *C. elegans*, providing a framework for further experimentation. Thereby, this thesis highlights the strength of *C. elegans* as a model for 1) the study of host-microbiota-pathogen tripartite interactions and 2) the identification of bacterial or immune-protective mechanisms, which is crucial for the development of microbiota-based therapeutic strategies.

## REFERENCES

- Aguirre De Cárcer, D. (2018). The human gut pan-microbiome presents a compositional core formed by discrete phylogenetic units. *Scientific Reports*, 8, 14069.
- Allman, E., Johnson, D. and Nehrke, K. (2009). Loss of the apical V-ATPase  $\alpha$ -subunit VHA-6 prevents acidification of the intestinal lumen during a rhythmic behavior in *C. elegans*. *American Journal of Physiology - Cell Physiology*, 297(5), C1071.
- Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D., & Finn, R. D. (2019). A new genomic blueprint of the human gut microbiota. *Nature*, 568(7753), 499–504.
- Balasubramanian, D., Harper, L., Shopsin, B., & Torres, V. J. (2017). *Staphylococcus aureus* pathogenesis in diverse host environments. *Pathogens and Disease*, 75(1).
- Bender, A., Woydziak, Z.R., Fu, L., Branden, M., Zhou, Z., Ackley, B.D., and Peterson, B.R. (2013). Novel acid-activated fluorophores reveal a dynamic wave of protons in the intestine of *Caenorhabditis elegans*. *ACS Chemical Biology*, 8, 636–642.
- Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M. C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., ... Schlöter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, 8(1).
- Berg, M., Monnin, D., Cho, J., Nelson, L., Crits-Christoph, A., & Shapira, M. (2019). TGF $\beta$ /BMP immune signaling affects abundance and function of *C. elegans* gut commensals. *Nature Communications*, 10(1), 604.
- Berg, M., Stenuit, B., Ho, J., Wang, A., Parke, C., Knight, M., Alvarez-Cohen, L., & Shapira, M. (2016a). Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *The ISME Journal*, 10(8), 1998–2009.
- Berg, M., Zhou, X. Y., & Shapira, M. (2016b). Host-specific functional significance of *Caenorhabditis* gut commensals. *Frontiers in Microbiology*, 7(1622).
- Blaser, M. J. (2010). *Helicobacter pylori* and esophageal disease: Wake-up call? *Gastroenterology*, 139(6), 1819–1822.
- Blaser, M. J. (2015). Missing microbes: How the overuse of antibiotics is fueling our modern plagues. New York: Picador.
- Bosch, T. C. G., & McFall-Ngai, M. J. (2011). Metaorganisms as the new frontier. *Zoology*, 114(9), 185–190.
- Broderick, N. A., Raffa, K. F., & Handelsman, J. (2006). Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. *Proceedings of the National Academy of Sciences*, 103(41), 15196–15199.
- Broderick, N. A. (2016). Friend, foe or food? Recognition and the role of antimicrobial peptides in gut immunity and drosophila-microbe interactions. *Philosophical Transactions of the Royal Society*, 371(1695).
- Broderick, N. A., Robinson, C. J., McMahon, M. D., Holt, J., Handelsman, J., & Raffa, K. F. (2009). Contributions of gut bacteria to *Bacillus thuringiensis*-induced mortality vary across a range of Lepidoptera. *BMC Biology*, 7(1), 1–9.
- Casadevall, A. (2017). The pathogenic potential of a microbe. *MSphere*, 2(1).
- Chaplin, D. D. (2010). Overview of the immune response. *Journal of Allergy and Clinical Immunology*, 125(2), S3.
- Chow, J., & Mazmanian, S. K. (2010). A pathobiont of the microbiota balances host colonization and intestinal inflammation. *Cell Host and Microbe*, 7(4), 265–276.
- Cook, D. E., Zdraljevic, S., Roberts, J. P., & Andersen, E. C. (2017). CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Research*, 45.

- Dahan, D., Preston, G. M., Sealey, J., & King, K. C. (2020). Impacts of a novel defensive symbiosis on the nematode host microbiome. *BMC Microbiology*, *20*(1), 159.
- Dale, C., & Moran, N. A. (2006). Molecular interactions between bacterial symbionts and their hosts. *Cell*, *126*(3), 453–465.
- Dennison, N. J., Jupatanakul, N., & Dimopoulos, G. (2014). The mosquito microbiota influences vector competence for human pathogens. *Current Opinion in Insect Science*, *3*, 6–13.
- Dierking, K., Yang, W., & Schulenburg, H. (2016). Antimicrobial effectors in the nematode *Caenorhabditis elegans*: an outgroup to the Arthropoda. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1695), 20150299.
- Dimov, I. and Maduro, M. F. (2019). The *C. elegans* intestine: organogenesis, digestion, and physiology. *Cell and Tissue Research*, *377*, 383–396.
- Dirksen, P., Assié, A., Zimmermann, J., Zhang, F., Tietje, A.-M., Marsh, S. A., Félix, M.-A., Shapira, M., Kaleta, C., Schulenburg, H., & Samuel, B. S. (2020). CeMbio - The *Caenorhabditis elegans* microbiome resource. *G3: Genes, Genomes, Genetics*, *10*(8).
- Dirksen, P., Marsh, S. A., Braker, I., Heitland, N., Wagner, S., Nakad, R., Mader, S., Petersen, C., Kowallik, V., Rosenstiel, P., Félix, M.-A., & Schulenburg, H. (2016). The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biology*, *14*(38).
- Douglas, A. E. (2008). Conflict, cheats and the persistence of symbioses. *New Phytologist*, *177*(4), 849–858.
- Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, *17*, 764–775.
- Duveau, F., & Félix, M. A. (2012). Role of pleiotropy in the evolution of a cryptic developmental variation in *Caenorhabditis elegans*. *PLOS Biology*, *10*(1), e1001230.
- Ezcurra, M. (2018). Dissecting cause and effect in host-microbiome interactions using the combined worm-bug model system. *Biogerontology*, *19*, 1–12.
- Félix, M.-A., & Braendle, C. (2010). The natural history of *Caenorhabditis elegans*. *Current Biology*, *20*(22), 965–969.
- Flores, H. A., & O'Neill, S. L. (2018). Controlling vector-borne diseases by releasing modified mosquitoes. *Nature Reviews Microbiology*, *16*(8), 508–518.
- Haller, D. (Ed.) (2018). *The Gut Microbiome in Health and Disease*. Cham, Switzerland: Springer International Publishing.
- Hasin, Y., Seldin, M., & Lusi, A. (2017). Multi-omics approaches to disease. *Genome Biology*, *18*(1), 83.
- Hill, C., Guarner, F., Reid, G., Gibson, G. R., Merenstein, D. J., Pot, B., Morelli, L., Canani, R. B., Flint, H. J., Salminen, S., Calder, P. C., & Sanders, M. E. (2014). Expert consensus document: The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nature Reviews Gastroenterology and Hepatology*, *11*(8), 506–514.
- Iebba, V., Totino, V., Gagliardi, A., Santangelo, F., Cacciotti, F., Trancassini, M., Mancini, C., Cicerone, C., Corazziari, E., Pantanella, F., & Schippa, S. (2016). Eubiosis and dysbiosis: the two sides of the microbiota. *New Microbiologica*, *39*, 1–12.
- Jochum, L., & Stecher, B. (2020). Label or Concept – What Is a Pathobiont? *Trends in Microbiology*, *28*(10), 789–792.
- Kiers, E. T., & West, S. A. (2015). Evolving new organisms via symbiosis: When, how do symbiotic partnerships become new, integrated organisms? *Science*, *348*(6233), 392–394.

- Koch, E. J., & McFall-Ngai, M. (2018). Model systems for the study of how symbiotic associations between animals and extracellular bacterial partners are established and maintained. *Drug Discovery Today: Disease Models*, 28, 3–12.
- Lenhart, P. A., & White, J. A. (2017). A defensive endosymbiont fails to protect aphids against the parasitoid community present in the field. *Ecological Entomology*, 42(5), 680–684.
- MacNeil, L. T., Watson, E., Arda, H. E., Zhu, L. J., & Walhout, A. J. M. (2013). Diet-induced developmental acceleration independent of TOR and Insulin in *C. elegans*. *Cell*, 153(1), 240–252.
- Manzoni, C., Kia, D. A., Vandrovцова, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, 19(2), 286–302.
- Masson, F., & Lemaitre, B. (2017). Protection from within. *ELife*, 6, e24111.
- McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lo, T., Douglas, A. E., Düblier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Neelson, K., Pierce, N. E., ... Kremer, N. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, 110(9), 3229–3236.
- O'Neill, S. L., Ryan, P. A., Turley, A. P., Wilson, G., Retzki, K., Iturbe-Ormaetxe, I., Dong, Y., Kenny, N., Paton, C. J., Ritchie, S. A., Brown-Kenyon, J., Stanford, D., Wittmeier, N., Anders, K. L., & Simmons, C. P. (2018). Scaled deployment of *Wolbachia* to protect the community from *Aedes* transmitted arboviruses. *Gates Open Research*, 2, 36.
- Petersen, C., Dirksen, P., & Schulenburg, H. (2015). Why we need more ecology for genetic models such as *C. elegans*. *Trends in Genetics*, 31(3), 120–127.
- Rackow, B. (2019). Protective effects of microbiota-isolates on the *C. elegans* Natural Diversity Resource (CeNDR) strains against *Bacillus thuringiensis* pathogenic infection. [Bachelor thesis, Christian Albrechts Universität zu Kiel].
- Rees, T., Bosch, T., & Douglas, A. E. (2018). How the microbiome challenges our concept of self. *PLoS Biology*, 16(2), e2005358.
- Rosebury T. (1962). *Microorganisms Indigenous to Man*. New York: McGraw Hill Book Co.
- Sampson, T. R., & Mazmanian, S. K. (2015). Control of brain development, function, and behavior by the microbiome. *Cell Host and Microbe*, 17(5), 565–576.
- Samuel, B. S., Rowedder, H., Braendle, C., Félix, M. A., & Ruvkun, G. (2016). *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proceedings of the National Academy of Sciences*, 113(27), E3941–E3949.
- Schulenburg, H., & Müller, S. (2004). Natural variation in the response of *Caenorhabditis elegans* towards *Bacillus thuringiensis*. *Parasitology*, 128(4), 433–443.
- Sieber, M., Pita, L., Weiland-Bräuer, N., Dirksen, P., Wang, J., Mortzfeld, B., Franzenburg, S., Schmitz, R. A., Baines, J. F., Fraune, S., Hentschel, U., Schulenburg, H., Bosch, T. C. G., & Traulsen, A. (2018). Neutrality in the metaorganism. *PLOS Biology*, 17(6), e3000298.
- Sifri, C. D., Begun, J., Ausubel, F. M., & Calderwood, S. B. (2003). *Caenorhabditis elegans* as a model host for *Staphylococcus aureus* pathogenesis. *Infection and Immunity*, 71(4), 2208–2217.
- Sommer, F. and Bäckhed, F. (2013) The gut microbiota — masters of host development and physiology. *Nature Reviews Microbiology*, 11(4), 227–238.
- Sterken, M. G., Snoek, L. B., Kammenga, J. E., & Andersen, E. C. (2015). The laboratory domestication of *Caenorhabditis elegans*. *Trends in Genetics*, 31(5), 224–231.

- Tan, M. W., Mahajan-Miklos, S., & Ausubel, F. M. (1999). Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *PNAS*, *96*(2), 715–720.
- The Human Microbiome Project Consortium, Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*, 207–214.
- Vannier, N., Agler, M., & Hacquard, S. (2019). Microbiota-mediated disease resistance in plants. *PLoS Pathogens*, *15*(6), e1007740.
- Wang, Q., Wang, K., Wu, W., Giannoulatou, E., Ho, J. W. K., & Li, L. (2019). Host and microbiome multi-omics integration: applications and methodologies. *Biophysical Reviews*, *11*(1), 55–65.
- Wiles, T. J., & Guillemin, K. (2020). Patterns of Partnership: Surveillance and Mimicry in Host-Microbiota Mutualisms. *Current Opinion in Microbiology*, *54*, 87-94.
- World Health Organization and Food and Agriculture Organization of the United Nations. (2011). Food and Agriculture Organization of the United Nations. World Health Organization. Evaluation of certain contaminants in food. *WHO Technical Report Series*, *959*, 55–64.
- Zhang, F., Berg, M., Dierking, K., Félix, M.-A., Shapira, M., Samuel, B. S., & Schulenburg, H. (2017). *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, *8*(485).
- Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Research*, *30*(6), 492–506.

## LIST OF EQUIPMENT

EQUIPMENT	BRAND
Dissecting Scope	Leica S6e
Stereomicroscope	M80 Leica
CLSM	Zeiss Axio Observer
Leica Fluorescence Dissecting Scope	Leica M205fa
Chemidoc™ Touch Imaging System	Biorad
PCR Cycler	SensoQuest Labcycler
Shaker	Gfl 3006
Laboratory Dish Washer	MIELE G 7883 PROFESSIONAL
Fine Balance	Pcb Kern
PH Meter	Hi 221
Incubator 37°C	Binder
Thermostatschrank 20°C	Lovibond
Rumed Incubator	Rumed
Minus 80°C Freezer	Scala
Chemidoc™ Touch Imaging System	Biorad
Tissue Shredder	Geno / Grinder
Clean bench Heraeus	Heraeus
Spectrophotometer	Jenway 6300
Microcentrifuge 5415	Eppendorf
Mini-Centrifuge	Ika Mini G S000
Power source	300V VWR





## **ANNEX**

### **Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms**

Philipp Rausch, Malte Rühlemann, Britt M. Hermes, Shauni Doms, Tal Dagan, Katja Dierking, Hanna Domin, Sebastian Fraune, Jakob von Frieling, Ute Hentschel, Femke-Anouska Heinsen, Marc Höppner, Martin Thomas Jahn, Cornelia Jaspers, Kohar Annie B. Kissoyan, Daniela Langfeldt, Ateequr Rehman, Thorsten B. H. Reusch, Thomas Roeder, Ruth A. Schmitz, Hinrich Schulenburg, Ryszard Soluch, Felix Sommer, Eva Stukenbrock, Nancy Weiland-Bräuer, Philip Rosenstiel, Andre Franke, Thomas Bosch, John F. Baines.

**Published manuscript in *Microbiome* (2019), 7(133)**



## RESEARCH

## Open Access

# Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms



Philipp Rausch<sup>1,2,3\*†</sup>, Malte Rühlemann<sup>4†</sup>, Britt M. Hermes<sup>1,2,5</sup>, Shauni Doms<sup>1,2</sup>, Tal Dagan<sup>6</sup>, Katja Dierking<sup>7</sup>, Hanna Domin<sup>8</sup>, Sebastian Fraune<sup>8</sup>, Jakob von Frieling<sup>9</sup>, Ute Hentschel<sup>10,11</sup>, Femke-Anouska Heinsen<sup>4</sup>, Marc Höppner<sup>4</sup>, Martin T. Jahn<sup>10</sup>, Cornelia Jaspers<sup>11,12</sup>, Kohar Annie B. Kissoyan<sup>7</sup>, Daniela Langfeldt<sup>6</sup>, Ateequr Rehman<sup>4</sup>, Thorsten B. H. Reusch<sup>11,12</sup>, Thomas Roeder<sup>9</sup>, Ruth A. Schmitz<sup>6</sup>, Hinrich Schulenburg<sup>7</sup>, Ryszard Soluch<sup>6</sup>, Felix Sommer<sup>4</sup>, Eva Stukenbrock<sup>13,14</sup>, Nancy Weiland-Bräuer<sup>6</sup>, Philip Rosenstiel<sup>4</sup>, Andre Franke<sup>4</sup>, Thomas Bosch<sup>8</sup> and John F. Baines<sup>1,2\*</sup>

## Abstract

**Background:** The interplay between hosts and their associated microbiome is now recognized as a fundamental basis of the ecology, evolution, and development of both players. These interdependencies inspired a new view of multicellular organisms as “metaorganisms.” The goal of the Collaborative Research Center “Origin and Function of Metaorganisms” is to understand why and how microbial communities form long-term associations with hosts from diverse taxonomic groups, ranging from sponges to humans in addition to plants.

**Methods:** In order to optimize the choice of analysis procedures, which may differ according to the host organism and question at hand, we systematically compared the two main technical approaches for profiling microbial communities, 16S rRNA gene amplicon and metagenomic shotgun sequencing across our panel of ten host taxa. This includes two commonly used 16S rRNA gene regions and two amplification procedures, thus totaling five different microbial profiles per host sample.

**Conclusion:** While 16S rRNA gene-based analyses are subject to much skepticism, we demonstrate that many aspects of bacterial community characterization are consistent across methods. The resulting insight facilitates the selection of appropriate methods across a wide range of host taxa. Overall, we recommend single- over multi-step amplification procedures, and although exceptions and trade-offs exist, the V3 V4 over the V1 V2 region of the 16S rRNA gene. Finally, by contrasting taxonomic and functional profiles and performing phylogenetic analysis, we provide important and novel insight into broad evolutionary patterns among metaorganisms, whereby the transition of animals from an aquatic to a terrestrial habitat marks a major event in the evolution of host-associated microbial composition.

**Keywords:** Animal microbiome, Evolution, Phylosymbiosis, Holobiont, Metaorganism

\* Correspondence: philipp.rausch@bio.ku.dk; baines@evolbio.mpg.de

†Philipp Rausch and Malte Rühlemann contributed equally to this work.

<sup>1</sup>Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

## Background

Dynamic host-microbe interactions have shaped the evolution of life. Virtually all plants and animals are colonized by an interdependent complex of microorganisms, and there is growing recognition that the biological processes of hosts and their associated microbial communities function in tandem, often as biological partners comprising a collective entity known as the metaorganism [1]. For instance, symbiotic bacteria contribute to host health and development in critical ways, ranging from nutrient metabolism to regulating whole life cycles [2] and in turn benefit from habitats and resources the host provides. Moreover, it is well established that perturbations of the microbiome likely play an important role in many host disease states [3]. However, researchers have yet to elucidate the mechanisms driving these interactions, as the exact molecular and cellular processes are only poorly understood.

An integrated view on the metaorganism encompasses a cross-disciplinary approach that addresses how and why microbial communities form long-term associations with their hosts. Despite widespread agreement that the interdependencies of microbes and their hosts warrant study, there remains considerable incongruity between researchers regarding the best methodologies to study host-microbe interactions. The development of standardized protocols for characterizing and analyzing host-associated microbiomes across the tree of life is thus crucial to understand the evolution and function of metaorganisms without the issues of technical inconsistencies or data quality.

The rapidly growing interest in microbiome research has been bolstered by the ability to profile diverse microbial communities using next-generation sequencing (NGS). This culture-free, high-throughput technology enables identification and comparison of entire microbial communities, so-called metagenomics [4]. Metagenomics typically encompasses two particular sequencing strategies: amplicon sequencing, most often of the 16S rRNA gene as a phylogenetic marker; or shotgun sequencing, which captures the complete breadth of DNA within a sample [4].

The use of the 16S ribosomal RNA gene as a phylogenetic marker has proven to be an efficient and cost-effective strategy for microbiome analysis and even allows for the imputation of functional content based on taxon abundances [5]. However, PCR-based phylogenetic marker protocols are vulnerable to biases through sample preparation and sequencing errors. The choice of which hypervariable regions of the 16S rRNA gene are targeted for sequencing seems to be among the biggest factors underlying technical differences in microbiome composition [6–8]. Furthermore, 16S rRNA gene amplicon sequencing is typically limited to taxonomic

classification at the genus level depending on the database and classifiers used [9], and provides only limited functional information [5]. These well-recognized limitations of amplicon-based microbial community analyses have raised concerns about the accuracy and reproducibility of 16S rRNA phylogenetic marker studies and have led to an increased interest in developing more reliable methods for amplicon library preparation and sequencing [8, 10].

Shotgun metagenomics, on the other hand, offers the advantage of species- and strain-level classification of bacteria. Additionally, it allows researchers to examine the functional relationships between hosts and bacteria by determining the functional content of samples directly [9, 11], and enables the exploration of yet unknown microbial life that would otherwise remain unclassifiable [12]. However, the relatively high costs of shotgun metagenomics and more demanding bioinformatic requirements have precluded its use for microbiome analysis on a wide scale [4, 9].

In this study, we set out to systematically compare experimental and analytical aspects of the two main technical approaches for microbial communities profiling, 16S rRNA gene amplicon and shotgun sequencing, across a diverse array of host species studied in the Collaborative Research Center 1182, “Origin and Function of Metaorganisms.” The ten host species range from basal aquatic metazoans [*Aplysina aerophoba* (sponge) and *Mnemiopsis leidyi* (comb jelly)]; to marine and limnic cnidarians (*Aurelia aurita*, *Nematostella vectensis*, *Hydra vulgaris*), standard vertebrate (*Mus musculus*), and invertebrate model organisms (*Drosophila melanogaster*, *Caenorhabditis elegans*); to *Homo sapiens*; and in addition to wheat (*Triticum aestivum*) and a standardized mock community. This setup provides a breadth of samples in terms of taxonomic composition and diversity. Conducting standardized data generation procedures on these diverse samples on the one hand provides a unique and powerful opportunity to systematically compare alternative methods, which display considerable heterogeneity in performance. On the other hand, this information enables researchers working on these or similar host species to choose the experimental (e.g., hypervariable region) or analytical pipelines that best suit their needs, which will be a valuable resource to the greater community of host-microbe researchers. Finally, we identified a number of interesting, broad-scale patterns contrasting the aquatic and terrestrial environment of metaorganisms, which also reflect their evolutionary trajectories.

## Results

Our panel of hosts includes ten species, for which five biological replicates each were included (see

Additional file 1: Figure S1). The majority of hosts are metazoans, including the “golden sponge” (*Aplysina aerophoba*), moon jellyfish (*Aurelia aurita*), comb jellyfish (*Mnemiopsis leidyi*), starlet sea anemone (*Nematostella vectensis*), fresh-water polyp *Hydra vulgaris*, roundworm (*Ceanorhabditis elegans*), fruit fly (*Drosophila melanogaster*), mouse (*Mus musculus*), human (*Homo sapiens*), and the inclusion of wheat (*Triticum aestivum*), which can serve as an outgroup to the metazoan taxa. *Drosophila melanogaster* was additionally sampled using two different methods targeting feces and intestinal tissue. Nucleic acid extraction procedures were conducted according to the needs of the individual host species (see the “Methods” section and Additional file 1), after which all DNA templates were subjected to a standard panel of sequencing procedures. For 16S rRNA gene amplicon sequencing, we used primers flanking two commonly used variable regions, the V1 V2 and V3 V4 regions. Further, for each region, we compared a single-step fusion-primer PCR to a two-step procedure designed to improve the accuracy of amplicon-based studies [8]. Finally, all samples were also subjected to shotgun sequencing, such that five different sequence profiles were generated for each sample. While a single classification pipeline was employed for all four 16S rRNA gene amplicon sequence profiles, community composition based on shotgun data was evaluated using MEGAN [13], due to the advantage of simultaneously performing taxonomical and functional classification of shotgun reads and an overall good performance (for additional description, see Additional file 1).

#### Performance of data processing and quality control

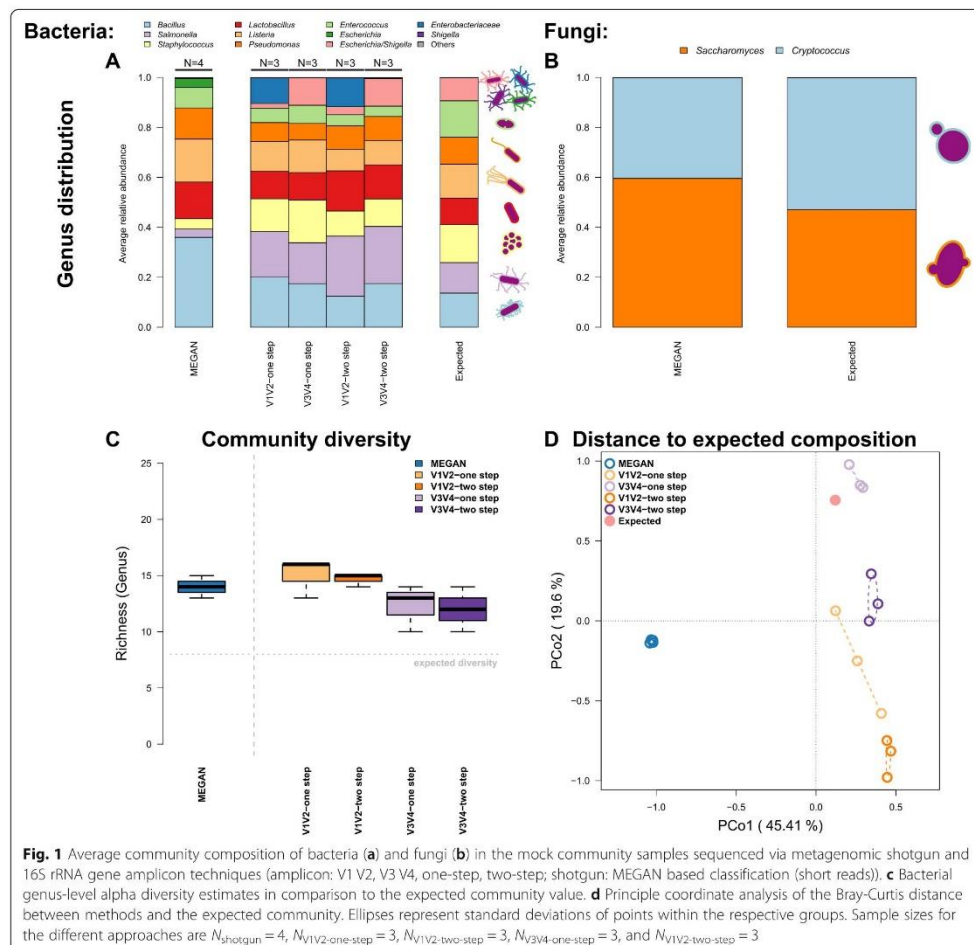
All data generated from amplicons were subject to the same stringent quality control pipeline including read-trimming, merging of forward and reverse reads, quality filtering based on sequence quality and estimated errors, and chimera removal (see the “Methods” section). The one-step V1 V2 amplicon data showed the highest rate of read-survival ( $62.13 \pm 23.90\%$ ; mean  $\pm$  sd) followed by the corresponding two-step method ( $49.85 \pm 23.90\%$ ; mean  $\pm$  sd), in large part due to the greater coverage of this comparatively shorter amplicon (~312 bp). In contrast,  $42.02 \pm 16.41\%$  and  $36.88 \pm 23.89\%$  of the total reads were included in downstream analysis for the one-step and two-step V3 V4 data, respectively. The longer V3 V4 amplicon (~470 bp) was more affected by drops in quality at the end of the reads, which decreases the overlap of forward and reverse reads and thus increases the chances of sequencing errors (Additional file 1: Figure S2; for final sample sizes, see Additional file 2: Table S1). Overall, aside from chimera removal, each quality control step resulted in a comparatively greater loss of V3 V4 compared to V1 V2 data. On the other hand, the

V3 V4 one-step method yields the lowest number of chimeras, suggesting a lower rate of chimera formation and/or detection in this approach (variable region  $F_{1,214} = 3.8881$ ,  $P = 0.0499$ ; PCR protocol  $F_{1,214} = 8.1751$ ,  $P = 0.0047$ ; variable region  $\times$  PCR protocol  $F_{1,214} = 6.4733$ ,  $P = 0.0117$ ; linear mixed model with organism as random factor). Among all host taxa, we observe the highest proportion of retained reads in the V1 V2 one-step method and the lowest in the V3 V4 two-step method (Additional file 1: Figure S2B; variable region  $F_{1,215} = 74.9989$ ,  $P < 0.0001$ ; PCR protocol  $F_{1,215} = 21.0743$ ,  $P < 0.0001$ ; linear mixed model with organism as random factor). After quality filtering and the identification of bacterial reads, an average of 0.46 Gb of shotgun reads per sample was achieved (range 0.03–2.1 Gb) (Additional file 1: Figure S3A; for final sample sizes, see Additional file 2: Table S1). To provide an initial assessment and comparison between the amplicon and shotgun-based techniques, we plotted the discovered classifiable taxa and functions for the entire pooled dataset. Although the methods differ distinctly, each method shows a plateau in the number of discovered entities (see Additional file 1: Figures S3C, S3D).

#### Mock community

The analysis of standardized mock communities is an important measure to ensure general quality standards in microbial community analysis. In this study, we employed a commercially available mixture of eight bacterial and two yeast species. Comparison among the amplification procedures (one- and two-step PCR), 16S rRNA gene regions (V1 V2, V3 V4), and shotgun data reveals varying degrees of similarity to the expected microbial community composition (Fig. 1). One discrepancy is apparent due to the misclassification of *Escherichia/Shigella*, whose close relationship makes delineation at the genus level difficult based on the V1 V2 region and is subsequently classified to *Enterobacteriaceae* (Fig. 1a, Additional file 1: Figure S4). Classification of this bacterial group also differs based on the shotgun analysis employed, due to different naming and taxonomic standards of the respective databases (*Escherichia*, *Shigella*, and *Enterobacteriaceae* refer to the *Escherichia/Shigella* cluster) [14]. However, overall, the amplicon-based profiles show the closest matches to the expected community. The V3 V4 one-step method shows the lowest degree of deviation between observed and expected abundances of the focus taxa (Table 1; Additional file 1: Figure S4). In addition, the relative abundances of fungi in the mock community were relatively well predicted by MEGAN (see Fig. 1).

Next, we evaluated alpha and beta diversity across the different technical and analytical methods. Interestingly, most methods overestimate taxon richness but



underestimate complexity (as measured by the Shannon index) of the mock community, which could reflect biases arising from grouping taxon abundances based on slightly differing taxonomies (Fig. 1c, Additional file 1: Figures S4, S5A and Additional file 2: Table S2). Overall, the amplicon methods appear to more accurately reflect alpha diversity, although significant differences are present with regard to the amplified region (species richness: variable region  $F_{1,10} = 6.3657$ ,  $P = 0.0302$ ; Shannon H: method  $F_{1,9} = 3.330$ ,  $P = 0.1014$ , variable region  $F_{1,9} = 6.110$ ,  $P = 0.0354$ ; ANOVA best model). With regard to beta diversity, the largest distance to the expected composition is observed for the shotgun-based data, while the amplicon-based techniques, in particular V3 V4,

show the lowest distance (Fig. 1d, Additional file 1: Figure S5B). Pairwise tests show almost no differences between the amplicon-based techniques, while the shotgun-based data significantly differs from all amplicon profiles (Additional file 2: Table S3). Thus, in conclusion, shotgun-based analysis yields a higher degree of error compared to the amplicon-based approaches for the simple mock community used in our study.

#### Taxonomic diversity within and between hosts

To evaluate the performance of our panel of metagenomic methods over the range of complex host-associated communities in our consortium, we next employed a series of alpha and beta diversity analyses to

**Table 1** Differences between expected and observed genus abundances in the mock communities ( $N_{\text{shotgun}} = 4$ ,  $N_{\text{amplicon}} = 3$ ) via a one-sample *t* test (two-sided) of relative abundances (*P* values are adjusted via Hommel procedure)

Members mock community	Shotgun	Amplicon			
	MEGAN	V1 V2 one-step	V3 V4 one-step	V1 V2 two-step	V3 V4 two-step
<i>Staphylococcus</i>	0.00002	0.52446	0.09200	0.03994	0.21564
<i>Listeria</i>	0.00395	0.34964	0.53267	0.03003	0.00545
<i>Bacillus</i>	0.00006	0.21420	0.02818	0.29671	0.30589
<i>Pseudomonas</i>	0.13668	0.36721	0.05776	0.38147	0.59037
<i>Escherichia/Shigella</i> <sup>a</sup>	NA	0.00462	0.45612	0.00237	0.59037
<i>Shigella</i> <sup>a</sup>	$4.6372 \times 10^{-10}$	NA	NA	NA	NA
<i>Escherichia</i> <sup>a</sup>	0.00001	NA	NA	NA	NA
<i>Enterobacteriaceae</i> <sup>a</sup>	NA	0.87898	0.00004	0.19274	0.00055
<i>Salmonella</i>	$3.8092 \times 10^{-5}$	0.34964	0.05838	0.09712	0.08851
<i>Lactobacillus</i>	0.00297	0.87898	0.53267	0.38147	0.59037
<i>Enterococcus</i>	0.00012	0.04816	0.03746	0.01159	0.00954

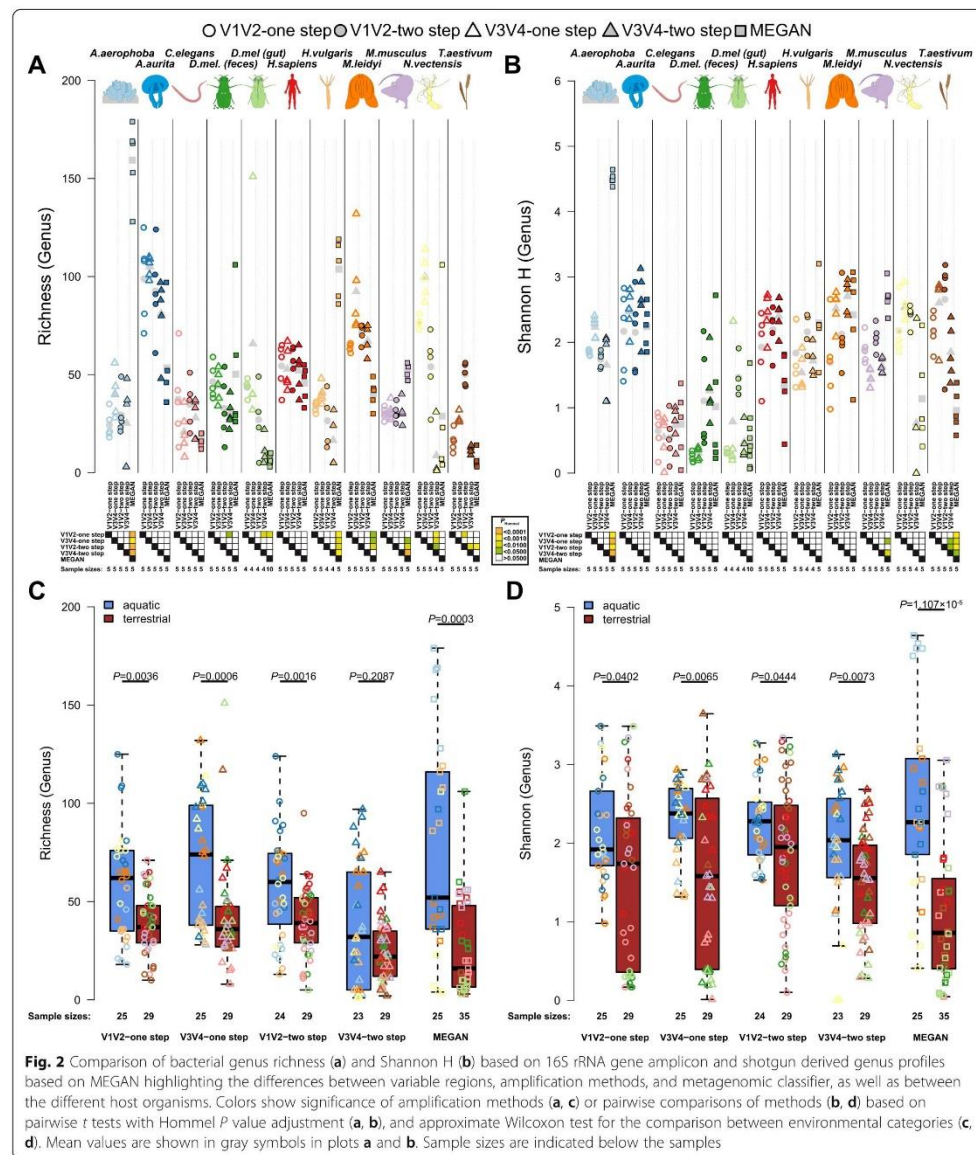
<sup>a</sup>*Escherichia/Shigella* relatives counted as equivalent

these samples, which also provides an opportunity to infer broad patterns across animal taxa based on a standardized methodology. Measures of alpha diversity display overall consistent values with respect to host species, although many significant differences between methods are present, which are mostly host-specific (Fig. 2a, b). However, several host taxa display high levels of consistency across methods including *A. aurita*, *C. elegans*, *D. melanogaster*, and *H. sapiens*, which show almost no significant differences between methods. Discrepancies and individual recommendations for each host species are discussed in Additional file 1: Figures S6–S16 and Additional file 2: Table S4. An intriguing observation is the tendency of aquatic hosts to display higher alpha diversity values than those of terrestrial hosts, which is supported by average differences between aquatic and terrestrial hosts and by relative consistent comparisons among single host species as well (Fig. 2c, d; Additional file 2: Table S5).

In order to investigate broad patterns of bacterial community similarity according to metagenomic procedure and host species, we performed beta diversity analyses including all host samples and each of their five different methodological profiles. This analysis reveals an overall strong signal of host species, irrespective of the method used to generate community profiles (Table 2; Fig. 3). Pairwise comparisons between hosts are significant in all cases except for samples derived from the V3 V4 two-step protocol, which did not consistently reach significance after correction for multiple testing (Additional file 2: Table S6). Further, complementary to the observations made for alpha diversity, we also find strong signals of community differentiation between the aquatic and terrestrial hosts (Table 2; Fig. 3b, d). The separation between these environments appears to be stronger based on amplicon data, whereas the separation between hosts is

stronger based on shotgun-derived data (Table 2). To further evaluate the variability among biological replicates, we evaluated intra-group distances according to host species, which reveals organisms with generally higher community variability (i.e., *C. elegans*, *A. aurita*, *H. sapiens*, *H. vulgaris*, *T. aestivum*, and *M. leidyii*) than other host organisms in our study (*N. vectensis*, *M. musculus*, *D. melanogaster*, and *A. aerophoba*; Additional file 1: Figure S17A and C). Interestingly, intra-group distances also significantly differ between the aquatic and terrestrial environments, whereby aquatic organisms tend to display less variable communities than terrestrial ones (Additional file 1: Figure S17B and D). Thus, this suggests higher sample sizes may be necessary for experimental analysis of the higher variability/terrestrial taxa. The low performance of *T. aestivum* in subsequent analyses possibly originates from its commercial origin and low bacterial biomass relative to host material.

To identify individual drivers behind patterns of beta diversity, we performed indicator species analysis [15] at the genus level with respect to method, host species, and environment. Based on the amplicon data, we identified 56 of 313 indicators to display consistent associations across all four amplicon techniques, such as *Bacteroides*, *Barnesiella*, *Clostridium IV*, and *Faecalibacterium* in *H. sapiens* and *Helicobacter* and *Mucispirillum* in *M. musculus*, whereas other associations were limited to, e.g., only one variable region (Additional file 2: Tables S7 and S8). However, the overall pattern of host associations is largely consistent across methods (Additional file 1: Figure S18). We also identified numerous indicator genera for aquatic and terrestrial hosts (Additional file 2: Tables S9 and S10). Indicator analyses based on shotgun data reveals a smaller and less diverse set of host-specific indicators, which however show many congruencies with the amplicon-based data.



### Functional diversity within and between hosts

To examine the diversity (gene richness) of metagenomic functions across host species, we evaluated EggNOG annotations (evolutionary genealogy of genes: Non-supervised Orthologous Groups [16]) to obtain a

general functional spectrum (assembly-based and MEGAN), in addition to annotations derived from a database dedicated to functions interacting with carbohydrates (CAZY—Carbohydrate-Active enZymes) [17]. Overall, the individual host communities differ



**Table 2** Taxonomic distance-based PERMANOVA results for differences in community composition (genus level) between host species and host environments based on shared abundance (Bray-Curtis) and shared presence (Jaccard), based on whole genome shotgun and different amplicon strategies (*P* values are adjusted via Hommel's procedure)

Distance	Factor	Data	Classifier	DF	F	P	<i>P</i> <sub>Hommel</sub>	<i>R</i> <sup>2</sup>	adj. <i>R</i> <sup>2</sup>			
Bray-Curtis	Organism	Shotgun	MEGAN	10,49	6.3517	0.0001	0.0001	0.5645	0.4756			
			Amplicon	V1 V2 one-step	10,43	7.1026	0.0001	0.0001	0.6229	0.5352		
				V1 V2 two-step	10,42	4.2297	0.0001	0.0001	0.5018	0.3831		
				V3 V4 one-step	10,43	7.8964	0.0001	0.0001	0.6474	0.5654		
	V3 V4 two-step	10,41		3.7917	0.0001	0.0001	0.4805	0.3538				
	Environment	Shotgun	MEGAN	1,58	5.8958	0.0001	0.0004	0.0923	0.0766			
			Amplicon	V1 V2 one-step	1,52	6.1588	0.0001	0.0001	0.1059	0.0887		
		V1 V2 two-step		1,51	4.6185	0.0001	0.0001	0.0830	0.0651			
		V3 V4 one-step		1,52	5.4975	0.0001	0.0001	0.0956	0.0782			
		V3 V4 two-step		1,50	3.3349	0.0001	0.0001	0.0625	0.0438			
		Jaccard		Organism	Shotgun	MEGAN	10,49	4.7458	0.0001	0.0001	0.4920	0.3883
						Amplicon	V1 V2 one-step	10,43	3.6867	0.0001	0.0001	0.4616
V1 V2 two-step			10,42				2.9760	0.0001	0.0001	0.4147	0.2754	
V3 V4 one-step	10,43		4.0248				0.0001	0.0001	0.4835	0.3633		
V3 V4 two-step	10,41	2.9343	0.0001	0.0001	0.4171		0.2750					
Environment	Shotgun	MEGAN	1,58	4.3872	0.0001	0.0004	0.0703	0.0543				
		Amplicon	V1 V2 one-step	1,52	3.8714	0.0001	0.0001	0.0693	0.0514			
	V1 V2 two-step		1,51	3.6541	0.0001	0.0001	0.0669	0.0486				
	V3 V4 one-step		1,52	4.3213	0.0001	0.0001	0.0767	0.0590				
	V3 V4 two-step		1,50	3.6646	0.0001	0.0001	0.0683	0.0497				

drastically in gene richness (EggNOG genes (MEGAN)  $\chi^2 = 52.202$ ,  $P < 2.10 \times 10^{-16}$ ; EggNOG genes (assembly)  $\chi^2 = 49.986$ ,  $P < 2.10 \times 10^{-16}$ ; CAZY  $\chi^2 = 48.815$ ,  $P < 2.10 \times 10^{-16}$ ; approximate Kruskal-Wallis test). Although the values also differ considerably between methods, overall, the functional repertoires are most diverse in the vertebrate hosts, while only *H. vulgaris* and *A. aerophoba* as aquatic hosts carry comparably diverse functional repertoires (Fig. 4a, : Figure S19). Interestingly, in contrast to taxonomic diversity, we observe no difference in functional diversity between aquatic and terrestrial hosts.

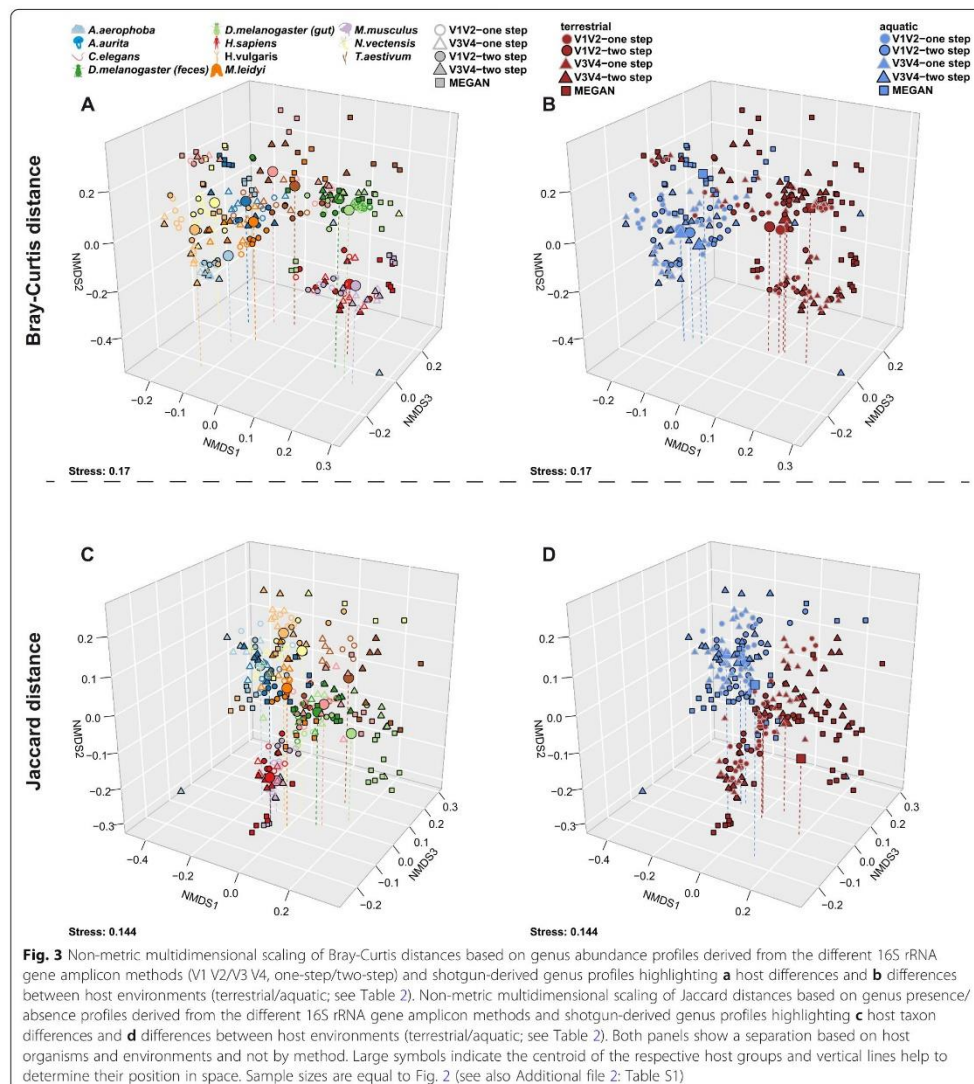
Next we examined community differences (beta diversity) at the functional level, which are overall more pronounced (average adj.  $R^2 = 0.5084$ ; Fig. 4) than those based on taxonomic (genus level) classification (shotgun adj.  $R^2 = 0.4756$ , amplicon average adj.  $R^2 = 0.4594$ ; see Tables 2 and 3; Figs. 3 and 4, Additional file 1: Figure S20). On the functional level, aquatic and terrestrial hosts are considerably less distinct than observed at the taxonomic level (taxonomic shotgun data  $R^2 = 0.0766$ , taxonomic amplicon average adj.  $R^2 = 0.0690$ , functional shotgun data  $R^2 = 0.0441$ ; see Tables 2 and 3; Fig. 4, Additional file 1: Figure S20). Variability of the functional repertoires was lowest in *A. aerophoba*, *D. melanogaster* feces, and *M. musculus* gut contents, while *H.*

*vulgaris*, *C. elegans*, and *D. melanogaster* gut samples displayed the highest intra-group distances, which translates to a higher amount of functional heterogeneity between replicates (Additional file 1: Figure S21). This reflects in large part the patterns we observed in taxonomic variability of those host-associated communities (Additional file 1: Figure S17).

#### Indicator functions

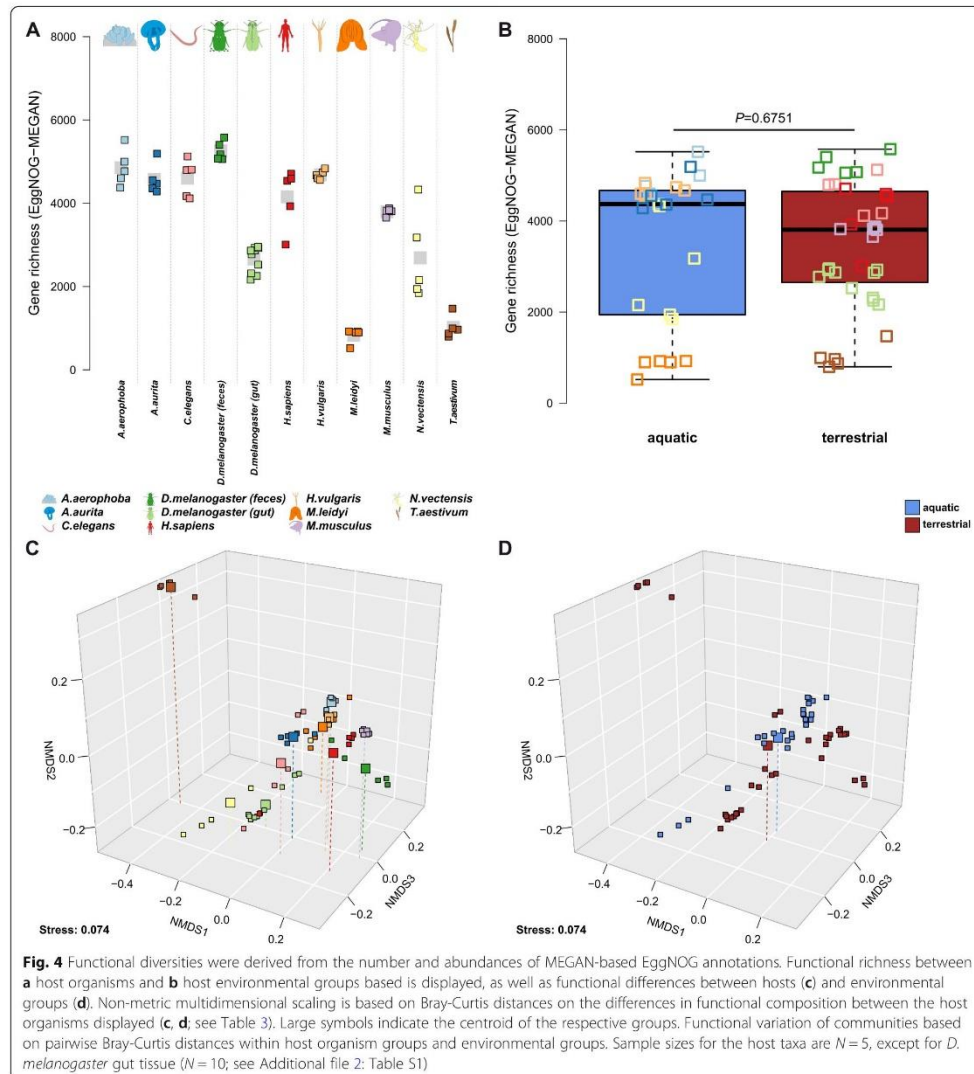
To identify specific functions that are characteristic of individual hosts, we applied indicator analysis to genomic functions. General functions in EggNOG reveal several interesting patterns, including CRISPR-related genes in *A. aerophoba*, *H. sapiens*, and *H. vulgaris*, suggesting a particular importance of viruses in these communities. Further, most species show characteristic genes mainly involved in energy production and conversion, amino acid transport and metabolism, replication, recombination, and repair, as well as cell wall/membrane/envelope biogenesis (Additional file 2: Tables S11–S13).

Analysis of carbohydrate-metabolizing functions based on CAZY [17] (Carbohydrate-Active enZymes) reveals the highest number of characteristic glycoside hydrolases (GH) in *H. sapiens* and *M. musculus*, whereas polysaccharide lyases (PLs) for non-hydrolytic cleavage of



glycosidic bonds are present in *A. aerophoba* and *H. sapiens* (Additional file 2: Table S14). Interestingly, parts of the cellulosome are only associated to *A. aerophoba*, while the freshwater polyp *H. vulgaris* carries characteristic auxiliary CAZymes involved specifically in lignin and chitin digestion, which may reflect adaptations of the host microbial communities to their diets (e.g., *Artemia nauplii*).

**Performance of metagenome imputation from 16S rRNA gene amplicon data using PICRUSt across metaorganisms**  
 Researchers often desire to obtain the insight gained from functional metagenomic information despite being limited to 16S rRNA gene data, for which imputation methods such as PICRUSt can be employed [5]. However, due to their dependence on variable region and database coverage [5], these imputations should be

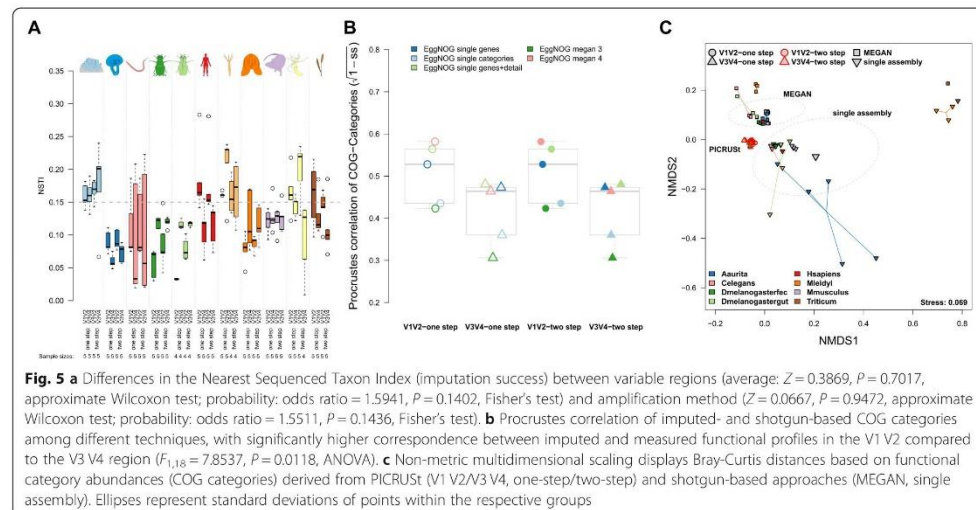


viewed with caution. Given our dataset of both 16S amplicon and shotgun metagenomic sequences, we systematically evaluated the performance of PICRUSt predictions across hosts and amplicon data type (V1 V2/V3 V4, one-step/two-step protocol). Beginning with the mock community, the V1 V2 region displays lower performance for imputing functions compared to V3 V4, as indicated by a higher weighted Nearest Sequenced

Taxon Index (NSTI) ( $t = 17.812$ ,  $P = 1.119 \times 10^{-7}$ ; Additional file 1: Figure S22A). High NSTI values imply low availability of genome representatives for the respective sample, due to either large phylogenetic distance for each OTU to its closest sequenced reference genome or a high frequency of poorly represented OTUs [5]. Comparing the distribution of functional categories based on Clusters of Orthologous Groups (COG) [18] between

**Table 3** Functional distance-based PERMANOVA results for differences in general functional community composition (EggNOG) and carbohydrate-active enzymes (CAZY) between host species and host environments based on shared abundance (Bray-Curtis) and shared presence (Jaccard) of functions ( $P$  values are adjusted via Hommel procedure)

Distance	Factor	Data	DF	F	P	$P_{\text{Hommel}}$	$R^2$	adj. $R^2$		
Bray-Curtis	Organism	CAZY	10,47	7.3323	0.0001	0.0001	0.6094	0.5263		
		EggNOG categories	10,49	5.6088	0.0001	0.0001	0.5337	0.4386		
		EggNOG gene + description	10,49	4.4454	0.0001	0.0001	0.4757	0.3687		
		EggNOG (MEGAN categories)	10,49	12.2594	0.0001	0.0001	0.7144	0.6562		
		EggNOG (MEGAN gene)	10,49	8.2788	0.0001	0.0001	0.6282	0.5523		
	Environment	CAZY	1,56	5.4257	0.0001	0.0007	0.0883	0.0721		
		EggNOG categories	1,58	2.5429	0.0195	0.0195	0.0420	0.0255		
		EggNOG gene + description	1,58	3.0662	0.0001	0.0007	0.0502	0.0338		
		EggNOG (MEGAN categories)	1,58	3.7703	0.0015	0.0030	0.0610	0.0448		
		EggNOG (MEGAN gene)	1,58	3.7271	0.0002	0.0012	0.0604	0.0442		
		Jaccard	Organism	CAZY	10,47	3.9098	0.0001	0.0001	0.4541	0.3380
				EggNOG categories	10,49	3.7179	0.0001	0.0001	0.4314	0.3154
EggNOG gene + description	10,49			2.5275	0.0001	0.0001	0.3403	0.2057		
EggNOG (MEGAN categories)	10,49			7.7781	0.0001	0.0001	0.6135	0.5346		
EggNOG (MEGAN gene)	10,49			5.4989	0.0001	0.0001	0.5288	0.4326		
Environment	CAZY		1,56	2.5866	0.0003	0.0021	0.0442	0.0271		
	EggNOG categories		1,58	1.4180	0.1442	0.1442	0.0239	0.0070		
	EggNOG gene + description		1,58	1.9535	0.0004	0.0024	0.0326	0.0159		
	EggNOG (MEGAN categories)		1,58	3.0425	0.0460	0.0920	0.0498	0.0335		
	EggNOG (MEGAN gene)		1,58	3.1222	0.0001	0.0009	0.0511	0.0347		



the different imputations (no cutoff applied) and the actual shotgun-based repertoires reveals considerable overlap except categories R (general function prediction only) and S (function unknown) (Additional file 1: Figure S22B).

Next we evaluated functional imputations for the different host species and amplification methods. We found no significant difference in average NSTI values or prediction success ( $NSTI < 0.15$ ) between amplification protocols or variable region. However, approximately a third (31.8%) of the samples are lost due to incomplete imputation ( $NSTI > 0.15$ ; Fig. 5a). Notable problematic host taxa are *A. aerophoba* and *H. vulgaris*, for which no sample remained below the NSTI cutoff value. Other host taxa displayed clear differential performance with regard to the variable region used, whereby *H. sapiens*, *N. vectensis*, and *T. aestivum* were successfully predicted based on V3 V4, but not V1 V2. However, when we employ Procrustes tests to compare community functional profiles based on shotgun sequencing (single assembly, MEGAN) and functional imputations at the COG-category level, we find a lower correspondence of the V3 V4-based imputations compared to those based on V1 V2 (Fig. 5b), while the amplification methods displayed no significant difference. A similar pattern is observed when we correlate community differences based on shotgun results and lower level (single functions) COG annotations based on PICRUSt, although the difference is not significant ( $F_{1,18} = 0.6172$ ,  $P = 0.4423$ ; ANOVA).

To investigate the similarities among methods in more detail, we merged shotgun and PICRUSt based annotations at the level of COG categories. Principle coordinate analysis reveals only small differences between imputations with regard to amplification method or variable region (Fig. 5c). However, large differences exist between the PICRUSt and shotgun-based functional repertoires, as well as between the shotgun techniques (MEGAN, single assembly). Differences between the shotgun techniques were significant but smaller than their distance to the imputed functional spectra (Fig. 5c; Additional file 2: Table S15), a pattern also found in the relative abundances of functional categories (Additional file 1: Figure S23).

In summary, the PICRUSt-imputed functional repertoires significantly differ from actual shotgun profiles. While variation in imputation success is largely dependent on the composition of the particular host community, V3 V4 appears to more often yield successful imputations. However, when successful, V1 V2-derived imputations display closer similarity to actual functional profiles. Finally, the amplification method (one-step, two-step) appears to have no significant effect on the quality of functional imputations. These data

therefore support the notion that metagenome imputations should be evaluated with care, as they depend on the underlying variable region and sample source.

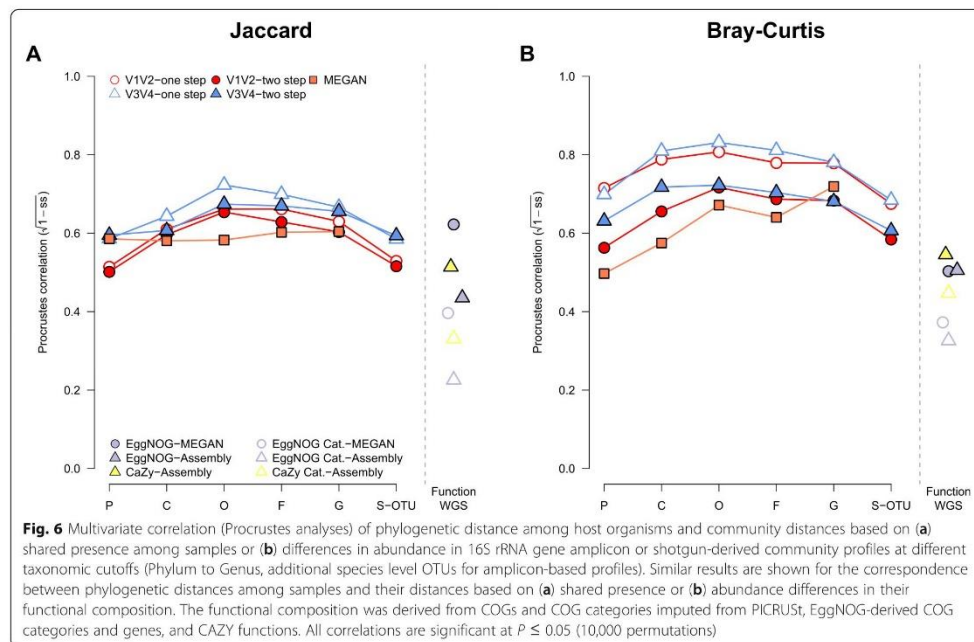
#### Phylogenetic patterns in microbial community composition

The term “phylosymbiosis” refers to the phenomenon where the pattern of similarity among host-associated microbial communities parallels the phylogeny of their hosts [19]. Highly divergent hosts with drastic differences in physiology and life history might be expected to overwhelm the likelihood of observing phylosymbiosis, which can typically be observed within a given host clade [19]. However, the factors driving differences in composition among our panel of hosts may also be expected to vary in terms of the bacterial phylogenetic scale at which they are most readily observed [20]. Thus, we evaluated the degree to which bacterial community relationships (beta diversity) reflect the underlying phylogeny of our hosts at a range of bacterial taxonomic ranks, spanning from the genus to the phylum level.

In order to assess the general overlap between beta diversity and phylogenetic distance of the host species, we performed Procrustes analysis [21]. These analyses reveal that the strongest phylogenetic signal is observed when bacterial taxa are grouped at the order and/or family level, whereby the one-step protocols and the V3 V4 region display greater correlations to phylogenetic distance (Fig. 6). A similar pattern is observed for shotgun-based community profiles (i.e., MEGAN), although its fit increases again at the genus level. Measuring beta diversity based on co-occurrence of bacterial taxa between hosts (Jaccard; Fig. 6a) displays a weaker correspondence to host phylogeny than the abundance-based measure (Bray-Curtis; Fig. 6b).

To assess the fit of individual host taxa, we examined the residuals of the correlation between community composition and phylogenetic distance. This reveals a large variation in correspondence among host taxa, with *M. musculus*, *M. leidy*, *H. sapiens*, and *D. melanogaster* (feces) displaying the highest, while *H. vulgaris*, *C. elegans*, and *A. aerophoba* display the lowest correspondence between their microbiome composition and phylogenetic position (largest residuals; Additional file 1: Figure S24), pointing towards increased environmental influences on these microbial communities. Furthermore, terrestrial hosts display an overall better correspondence between co-occurrences of bacterial genera and host relatedness (V1 V2 one-step:  $Z = 2.9578$ ,  $P = 0.0025$ ), as do measurements based on V3 V4 (one-step:  $Z = 2.7496$ ,  $P = 0.0054$ ; two-step:  $Z = 2.8097$ ,  $P = 0.0046$ ; approximate Wilcoxon test).

Next, given the peak of correspondence between bacterial community composition and host phylogeny



observed at the order and/or family level, we set out to identify individual community members whose abundances best correlate to host phylogenetic distance using Moran's  $I$  eigenvector method [22]. This reveals 41 bacterial families and 36 orders with significant phylogenetic signal based on one or more amplicon data set, whereby 16 families and 18 orders display repeated associations across methods (e.g., *Clostridia*, *Bacteroidales*, *Desulfovibrionales*; Additional file 2: Table S16; Additional file 1: Figures S25 and S26). Analyzing communities based on shotgun data on the other hand identifies 75 bacterial families and 19 orders associated with phylogenetic distances, whereby 17 and 20 display repeated associations, respectively (Additional file 2: Table S16; Additional file 1: Figure S27). The combined results of these analyses identify several families and orders with strong and consistent phylogenetic associations, in particular for the vertebrate hosts (e.g., *Bacteroidaceae/Bacteroidales*, *Bifidobacteriaceae/Bifidobacteriales*, *Desulfovibrionaceae/Desulfovibrionales*, *Ruminococcaceae/Clostridiales*; see Additional file 2: Table S16). Other individual examples include bacteria related to *Helicobacteraceae/Campylobacterales* in *A. aurita*, which are observed in other marine cnidarians and may be involved in sulfur oxidation [23]. *Alcanivoracaceae*, an alkane-degrading bacterial group, is strongly associated to the

coastal cnidarian *N. vectensis*. This association might originate from adaptation to a polluted coastal environment [24]. *Acidobacteria Gp6* and *Gp9* specifically occur in *A. aerophoba* and are commonly associated to the core microbial community of sponges [25].

#### Phylogenetic patterns in functional community composition

In order to contrast the patterns observed at the taxonomic level to those based on function, we used Procrustes correlation to measure the overlap between phylogenetic distance and community distance based on the panel of functional categories in our analyses. Interestingly, the two functional categories displaying the greatest correspondence to host phylogeny are the CAZY and single EggNOG-based functions (Fig. 6). The remainder of patterns between phylogeny and bacterial functional spectra differed among the host species and functional categories (Additional file 1: Figure S28), and *T. aestivum* and *D. melanogaster* (feces) display the lowest correspondence, while *C. elegans*, *M. musculus*, and *H. sapiens* display the best correspondence (smallest residuals; Additional file 1: Figure S24) between their functional repertoire and phylogenetic position. As observed for the taxonomic analyses, terrestrial hosts again display a slightly better correlation than aquatic hosts (smaller residuals), in particular for the co-abundance of

EggNOG categories ( $Z = 2.2116$ ,  $P = 0.0267$ ), CAZY ( $Z = 2.0393$ ,  $P = 0.0414$ ), and the co-occurrence of EggNOG categories ( $Z = 2.7377$ ,  $P = 0.0061$ ) and genes ( $Z = 3.3062$ ,  $P = 0.0007$ ; approximate Wilcoxon test) among hosts.

Finally, to reveal individual functions correlating to host phylogeny, we used the aforementioned Moran's  $I$  eigenvector analyses with additional indicator analyses to narrow the potential clade associations. Interestingly, most functions that correlate to a specific host taxon/clade (1–3 host taxa) are mainly restricted to vertebrate hosts or in combination with a vertebrate host (Additional file 2: Tables S17–S20). This pattern is repeated across all functional annotations used in this study. Examples include fucosyltransferases, fucosidases, and polysaccharide-binding proteins, as well as different lyases for hyaluronate, xanthan, and chondroitin that stem from CAZY (see Additional file 1: Figure S28; Additional file 2: Table S17). These functions are related to glycan and mucin degradation and interaction, which mediate many intimate host-bacterial interactions and are also observed in subsequent analyses based on general functional databases (EggNOG; Additional file 2: Tables S18–S20). Many other phylogenetically correlated functions appear to be driven by the vertebrate hosts as well, which likely reflects the high functional diversity within this group (Fig. 4 and Additional file 1: Figure S21). Only *LPXC* and *LPXK* (EggNOG), genes involved in the biosynthesis of the outer membrane, are exclusively associated to the non-vertebrate hosts (*LPXC*, UDP-3-O-acyl-*N*-acetylglucosamine deacetylase; *LPXK*, Tetraacyldisaccharide 4'-kinase, as is an oxidative damage repair function (MSRA reductase) associated to *H. vulgaris* (Additional file 2: Table S19; Additional file 1: Figure S28). Finally, antibiotic resistance genes and virulence factors also show frequent phylogenetic and host-specific signals (Additional file 2: Tables S18 and S19; Additional file 1: Figure S28).

## Discussion

Despite the great number of metagenomic studies published to date, which range in their focus on technical, analytical, or biological aspects, our study represents a unique contribution given its breadth of different host samples analyzed with a panel of standardized methods. In particular, the trade-offs between 16S rRNA gene amplicon and shotgun sequencing concerning amplification bias, functional information, and both monetary and computational costs warrant careful consideration when designing research projects. While 16S rRNA gene amplicon-based analyses are subject to considerable skepticism and criticism, we demonstrate that in many aspects similar, if not superior characterization of bacterial communities is achieved by these methods. We also show, however, that important insight can be gained

through the combination of taxonomic and functional profiling, and that imputation-based functional profiles significantly differ from actual profiles. Our findings thus provide a guide for selecting an appropriate methodology for metagenomic analyses across a variety of metaorganisms. Finally, these data provide novel insight into the broad-scale evolution of host-associated bacterial communities, which can be viewed as particularly reliable given the repeatability of observations (e.g., differences between aquatic and terrestrial hosts, indicator taxa) across methods.

Given the concerns regarding the accuracy of 16S rRNA gene amplicon sequencing, other studies such as that of Gohl et al. [8] performed systematic comparisons of different library preparation methods and found superior results for a two-step amplification procedure. This method offers the additional advantage that one panel of adapter/barcode sequences can be combined with any number of different primers. Our first analyses were based on a standard mock community including Gram-positive and Gram-negative bacteria from the Bacilli and Gamma-Proteobacteria (eight species), as well as two fungi, which did not support an improvement of performance based on the two-step protocol. However, a number of changes were made to the Gohl et al. [8] protocol to adapt it to our lab procedures (e.g., larger reaction volumes, polymerase, variable region, heterogeneity spacers) that may contribute to these discrepancies, in addition to our different and diverse set of samples and other factors with potential influence on the performance of amplicon sequencing [6–8, 26–28]. The complexity of the mock community, i.e., the number of taxa, distribution, and phylogenetic breadth, may also have an influence on the discovery of clear trends in amplification biases or detection limits for certain taxonomic groups [29]. Thus, the even and phylogenetically shallow mock community in our study may be less suited than the staggered and diverse mixtures used in other studies [8] but still provides valuable information on repeatability, primer biases, and accuracy [29]. Nonetheless, when applied to our range of complex host-associated communities, we also found that significant differences in most parameters were due to the variable region rather than amplification method, and in many cases, biological signals were either improved or limited to the one-step protocol. Thus, in combination with the less complex laboratory procedures associated with the one-step protocol, we would generally recommend this procedure over two-step protocols.

Additional sources of variation influencing the outcome of our 16S rRNA gene amplicon-based community profiling are nucleic acid extraction procedures and the bioinformatic pipelines we employed. For the former, extraction procedures differed between host species due to

specific optimizations required for individual host species. Thus, certain differences in taxonomic and functional composition may be influenced by the specific protocols employed, as observed elsewhere [30]. Differences in the latter range from trimming and merging to clustering and classification, which are stringent and incorporate more reliable de novo clustering algorithms [31] as well as different classification databases [32]. Heterogeneity among the different amplicon approaches is however smaller than the differences between the amplicon and shotgun methods, as observed in other benchmarking studies [27]. Differences between shotgun approaches have been investigated in detail and also yield varying performances among classifiers, but in general, find a comparatively high performance of MEGAN-based approaches [9, 33, 34], which we also confirm in our study.

Given the limited number of studies that have compared imputed- and shotgun-derived functional repertoires [5, 35], our study also provides important additional insights. As imputation by definition is data-dependent, the differential performance and prediction among hosts in our study may in large part be explained by the amount of bacteria isolated, sequenced, and deposited (16S rRNA or genome) from these hosts or their respective environments. This seems to be most critical for the aquatic hosts. Furthermore, we observe a clear effect of variable region on the prediction performance, which is most obvious based on the mock community. The PICRUSt algorithm was developed and tested using primers targeting V3 V4 16S rRNA, and thus optimization of the imputation algorithm might be biased towards this target over the V1 V2 variable region. Although these performance differences, in particular the bias towards model organisms compared to less characterized communities (e.g., hypersaline microbial mats), were previously shown [5], our study provides additional, experimentally validated guidelines for a number of novel host taxa.

Interestingly, the strongest correspondence between bacterial community similarity and host genetic distance was detected at the bacterial order level for most of the employed methods. This may on the one hand reflect the deep phylogenetic relationships between our host taxa, such that turnover of bacterial taxa erodes phylosymbiosis over time [19, 20]. On the other hand, some of the more striking observations made among our host taxa are the differences between aquatic and terrestrial hosts, both at the level of alpha and beta diversity. Based on a molecular clock for the 16S rRNA gene of roughly 1% divergence per 50 million years [36], bacterial order level divergence corresponds well with the timing of animal terrestrialization (425–500 MYA) [37, 38]. Although evolutionary rates can widely vary among bacteria species [39], other studies of individual gut

microbial lineages such as the *Enterococci* indicate that animal terrestrialization was indeed a likely driver of diversification [40]. Specifically, the changing availability of carbohydrates in the host gut can be seen as a main driver of this diversification, which is consistent with the association of CAZY-based functional repertoires correlating to phylogenetic distance in our data set [19, 41].

In contrast to the patterns observed based on 16S rRNA gene amplicon-based profiles, the differentiation of bacterial communities according to host habitat was less pronounced based on functional genomic repertoires. This raises the possibility that the colonization of land by ancient animals required the acquisition of new, land-adapted bacterial lineages to perform some of the same ancestral functions. The overall observation of increased beta diversity among terrestrial compared to aquatic hosts (Additional file 1: Figure S19) could in part reflect differential acquisition among host lineages after colonizing land, although dispersal in the aquatic environment may on the other hand act as a greater homogenizing factor among aquatic hosts. The stronger correspondence between bacterial community and host phylogenetic distance among terrestrial hosts is also generally consistent with this hypothesis. However, the higher alpha diversity and the slightly lower correspondence with the phylogenetic patterns in aquatic hosts may also indicate a higher influence of environmental bacteria or a lack of physiological control over bacterial communities.

Bacterial taxa and functions involved in carbohydrate utilization were among the most notable associations to individual hosts, groups of hosts, and/or host phylogenetic relationships. Taxa such as *Bacteroidales*, *Ruminococcaceae/Ruminococcales*, and *Clostridia* associated to humans and/or mice include members known for a mucosal lifestyle, and these hosts also display the most diverse and abundant repertoire of carbohydrate-active enzymes (particularly glycosylhydrolases) in their microbiome. Other examples include sialidases, esterases, and fucosyltransferases, as well as different extracellular structures that appear to be specific to aquatic hosts, indicating differences in mucus and glycan composition according to this host environment. Glycan structures provide a direct link between the microbial community and the host via attachment, nutrition, and communication [42, 43], and the composition of mucin and glycan structures themselves show strong evolutionary patterns and are distinct among taxonomic groups [41]. Thus, a high diversity of glycan structures within and between hosts may determine the specific sets of carbohydrate-facilitating enzymes of the respective microbial communities.

In addition to the bacterial carbohydrate hydrolases that digest surrounding host and dietary carbohydrates,



we also identified a number of glycosyltransferases associated with capsular polysaccharide synthesis (Additional file 2: Tables S19 and S20). This type of glycosylation is an important facilitator for host association and survival [44] and plays a crucial role in infections [45] in mutualists and pathogens alike [44, 46]. Thus, capsular and excreted glycan structures are important for the successful colonization and persistence in different environments [47, 48] and host organisms [44, 48].

### Conclusions

In summary, the systematic comparison of five different metagenomic sequencing methods applied to ten different holobiont yielded a number of novel technical and biological insights. Although important exceptions will exist, we demonstrate that broad-scale biological patterns are largely consistent across these varying methods. As many aspects of differential performance in our study are host-specific (more detailed description of individual hosts can be found in Additional file 1), future development and benchmarking analyses would also benefit from including a range of different host/environmental samples.

### Methods

#### DNA extraction and 16S rRNA gene amplicon sequencing

Protocols for each host type are described in Additional file 1: Figures S18–S28. Each library (16S rRNA gene amplicon, shotgun) included at least one mock community sample based on the ZymoBIOMICS™ Microbial Community DNA Standard (Lot: ZRC187324, ZRC187325) consisting of eight bacterial species (*Pseudomonas aeruginosa* (10.4%), *Escherichia coli* (9.0%), *Salmonella enterica* (11.8%), *Lactobacillus fermentum* (10.3%), *Enterococcus faecalis* (14.1%), *Staphylococcus aureus* (14.6%), *Listeria monocytogenes* (13.2%), *Bacillus subtilis* (13.2%)) and two fungi (*Saccharomyces cerevisiae* (1.6%), *Cryptococcus neoformans* (1.8%)).

The 16S rRNA gene was amplified using uniquely bar-coded primers flanking the V1 and V2 hypervariable regions (27F–338R) and V3 V4 hypervariable regions (515F–806R) with fused MiSeq adapters and heterogeneity spacers in a 25- $\mu$ l PCR [28]. For the traditional one-step PCR protocol, we used 4  $\mu$ l of each forward and reverse primer (0.28  $\mu$ M), 0.5  $\mu$ l dNTPs (200  $\mu$ M each), 0.25  $\mu$ l Phusion Hot Start II High-Fidelity DNA Polymerase (0.5 U), 5  $\mu$ l of HF buffer (Thermo Fisher Scientific, Inc., Waltham, MA, USA), and 1  $\mu$ l of undiluted DNA. PCRs were conducted with the following cycling conditions (98 °C, 30 s; 30  $\times$  [98 °C, 9 s; 55 °C, 60 s; 72 °C, 90 s]; 72 °C, 10 min; 10 °C, infinity) and checked on a 1.5% agarose gel. Using a modified version of the recently published two-step PCR protocol by Gohl et al.

2016, we employed for the first round of amplification fusion primers consisting of the 16S rRNA gene primers (V1 V2, V3 V4) and a part of the Illumina Nextera adapter with the following cycling conditions in a 25- $\mu$ l PCR reaction (98 °C, 30 s; 25  $\times$  [98 °C, 10 s; 55 °C, 30 s; 72 °C, 60 s]; 72 °C, 10 min; 10 °C, infinity) [8]. Following the PCR, the product was diluted 1:10 and 5  $\mu$ l were used in an additional reaction of 10  $\mu$ l (98 °C, 30 s; 10  $\times$  [98 °C, 9 s; 55 °C, 30 s; 72 °C, 60 s]; 72 °C, 10 min; 10 °C, infinity) utilizing the Nextera adapter overhangs to ligate the Illumina adapter sequence and individual MID tags to the amplicons, following the manufacturer's instructions. The PCR protocol we used was 1  $\mu$ l of each forward and reverse primer (5  $\mu$ M), 0.3  $\mu$ l dNTPs (10  $\mu$ M), 0.2  $\mu$ l Phusion Hot Start II High-Fidelity DNA Polymerase (2 U/ $\mu$ l), 2  $\mu$ l of 5  $\times$  HF buffer (Thermo Fisher Scientific, Inc., Waltham, MA, USA), and 5  $\mu$ l of the diluted PCR product. The concentration of the amplicons was estimated using a Gel Doc™ XR+ System coupled with Image Lab™ Software (BioRad, Hercules, CA USA) with 3  $\mu$ l of O'GeneRuler™ 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, MA, USA) as the internal standard for band intensity measurement. The samples of individual gels were pooled into approximately equimolar sub-pools as indicated by band intensity and measured with the Qubit dsDNA br Assay Kit (Life Technologies GmbH, Darmstadt, Germany). Sub-pools were mixed in an equimolar fashion and stored at –20 °C until sequencing.

Library preparation for shotgun sequencing was performed using the NexteraXT kit (Illumina) for fragmentation and multiplexing of input DNA following the manufacturer's instructions. Amplicon sequencing was performed on the Illumina MiSeq platform with v3 chemistry (2  $\times$  300 cycle kit), while shotgun sequencing was performed on an Illumina NextSeq 500 platform via 2  $\times$  150 bp Mid Output Kit at the IKMB Sequencing Center (CAU Kiel, Germany).

#### Amplicon analysis

The respective V1 V2 and V3 V4 PCR primer sequences were removed from the sequencing data using cutadapt (v.1.8.3) [49]. Sequence data in FastQ format was quality trimmed using sickle (v.1.33) in paired-end mode with default settings and removing sequences dropping below 100 bp after trimming [50]. Forward and reverse read were merged into a single amplicon read using VSEARCH allowing fragments with a length of 280–350 bp for V1 V2 and 350–500 bp for V3 V4 amplicons [51]. Sequence data was quality controlled using fastq\_quality\_filter (FastX Toolkit) retaining sequences with no more than 5% of per-base quality values below 30 and subsequently with VSEARCH discarding sequences with more than one expected error [51, 52]. Reference-guided chimera removal was performed using the gold.f

reference in VSEARCH (v2.4.3). The UCLUST algorithm was used for a fast classification of the sequence data in order to remove sequences not assigned to the domains Bacteria or Archaea and exclude amplicon fragments from Chloroplasts [53]. Notably, only a total of 15 sequences were assigned to the domain Archaea, all found in two samples of human feces, accounting for less than 0.1% of the clean reads in these samples. The entire cleaned sequence data was concatenated into a single file and dereplicated and processed with VSEARCH for OTU picking using the UCLUST algorithm [54] using a 97% similarity threshold. OTUs were again checked for chimeric sequences, now using the de novo implementation of the UCHIME algorithm in VSEARCH [51, 54, 55]. All clean sequence data of the samples were mapped back to the cleaned OTU sequences using VSEARCH. OTU sequences and clean sequences mapping to the OTUs were taxonomically annotated using the RDP classifier algorithm with the RDP training set 14 [56, 57]. Sequence data were normalized by selecting 10,000 random sequences per sample. Taxon-by-sample abundance tables were created for all taxonomic levels from Phylum to Genus, as well as for OTUs.

#### PICRUSt functional imputations

Species-level OTUs (97% similarity threshold) were further classified using the GreenGenes (August 2013) database [58] via RDP classifier as implemented in mothur (v1.39.5) and merged with the abundances into a biome file which was uploaded to the Galaxy PICRUSt v1.1.1 pipeline (<http://galaxy.morganlangille.com/>) to derive functional imputations (COG predictions) [5]. To achieve accurate functional predictions, samples with NSTI  $\leq 0.15$  (weighted Nearest Sequenced Taxon Index) were pruned from the data set, as recommended by the developers.

#### Shotgun sequencing

Raw demultiplexed sequences were trimmed via Trimmomatic (v0.36) for low-quality regions with a minimum length of 50 bp as well as for adaptor and remaining MID sequences [59]. After trimming reads were mapped to host-specific genome databases and  $\Phi X$  with additional retention databases containing all fully sequenced bacterial and metagenomic genomes (5 September 2015) via DeconSeq (v0.4.3) [60]. Single and paired sequences were repaired using the BBTtools (v37.28) repair function [61]. Combined sequences were searched against the non-redundant NCBI database (28 July 2017) via DIAMOND [62] with ( $E$  value cutoff 0.001, v0.8.28) and MEGAN [13] classifying hits by functions (EggNOG—October 2016) and taxa (May 2017) (v6.6.1). For assemblies of single samples, we used metaSPADES [63] (v3.9.1) using paired reads in addition to unpaired reads

left from the previous steps. PROKKA (v1.12) was used for gene calling and initial genome annotation [64] using the metagenome option with additional identifying rRNA and snRNA via barnap, ARAGORN [65], and Infernal [66]. ORFs were further annotated via EggNOG annotation via HMMER models implemented in the EggNOG-mapper (v0.12.7) [16, 67], CAZY database via dbCAN (v5, July 24, 2016), and HMMER3 [17, 68]. Gene abundances were derived from mapping the all reads back to the predicted ORF via bowtie2 (v2.2.6) [69] and calculated TPM (transcripts per kilobase million) via SamTools (v1.5) [70].

18S rRNA genes were obtained from NCBI GeneBank and aligned via ClustalW (v1.4) [71] for host tree construction, which includes *A. aerophoba* (gi:51095211, AY5917991), *M. leidy* (gi:14517703, AF2937001), *H. vulgaris* (gi:761889987, JN5940542), *A. aurita* (gi:14700050, AY0392081), *N. vectensis* (gi:13897746, AF2543821), *T. aestivum* (gi:15982656, AY0490401), *M. musculus* (gi:374088232, NR\_0032783), *H. sapiens* (gi:36162, X032051), *D. melanogaster* (gi:939630477, NR\_1335591), and *C. elegans* (gi:30525807, AY2681171). Phylogenetic distance was calculated via DNADIST (v3.5c) [72] and a maximum likelihood tree was constructed via FastTree v2.1 CAT+G model [73]. Accuracy was improved via increased minimum evolution rounds for initial tree search [ $-\text{spr } 4$ ], more exhaustive tree search [ $-\text{mlacc } 2$ ], and a slow initial tree search [ $-\text{slownni}$ ].

#### Statistical analysis

Statistical analyses were carried out via R (v3.4.3) [74]. Alpha diversity indices (richness, Shannon-Weaver index) and beta diversity metrics based on the shared presence (Jaccard distance) or abundance (Bray-Curtis distance) of taxa were calculated in the *vegan* package [75] and ordinated via Principal Coordinate Analysis (PCoA, avoiding negative eigenvalues), or via non-metric multidimensional scaling (NMDS) using a maximum of 10,000 random starts to obtain a minimally stressed configuration in three dimensions. Clusters were fit via an iterative process (10,000 permutations) and tested for separation by direct gradient analysis via distance-based redundancy analyses and permutative ANOVA (10,000 permutations) [76, 77]. Univariate analyses were carried out with approximate Wilcoxon/Kruskal tests as implemented in *coin* [78] (10,000 permutations). Procrustes tests were used to relate pairwise community distances based on either different data sources such as functional repertoires or taxonomic composition, as well as phylogenetic distances [21, 79]. Moran's  $I$  eigenvector technique was employed to correlate bacterial community members and their functions to phylogenetic divergence, as implemented in *ape* (10,000 permutations) [22, 80].

Indicator species analysis, employing the generalized indicator value (*IndVal<sub>g</sub>*), was used to assess the predictive value of a taxon for each respective host phenotype/category as implemented in *indicspecies* [15]. Linear mixed models, as implemented in *nlme* were used to compare the influence of amplification method or variable region without the influence of the organism of origin [81]. We employed the Hommel and Benjamini-Yekutieli adjustment of *P* values when advised [82, 83].

### Additional files

**Additional file 1:** Supplementary Materials. (PDF 6900 kb)

**Additional file 2:** Supplementary Tables. (ZIP 1765 kb)

### Acknowledgements

We thank Katja Cloppenburg-Schmidt, Melanie Vollstedt, and Dr. Sven Künzel for the excellent assistance and help during the development of the project and their constant drive to improve its quality.

### Authors' contributions

PRa, PRo, AF, TB, and JFB conceived and designed research. PRa and MR performed data analyses. PRa, MR, BH, SD, and JFB interpreted the results and wrote the manuscript. PRa, MR, TD, KD, HD, SD, SF, JF, UHH, FAH, BH, MH, MJ, CJ, KABK, DL, AR, TBHR, TR, RAS, HS, RS, FS, ES, NWB, PRo, AF, TB, and JFB generated and interpreted host-specific data and gave intellectual input. All authors read and approved the final manuscript.

### Funding

This work was funded by the DFG Collaborative Research Centre (CRC) 1182 "Origin and Function of Metaorganisms" subproject Z3 and the Max-Planck-Society.

### Availability of data and materials

Sequence and meta-data are accessible under the study identifier PRJEB30924 ("https://www.ebi.ac.uk/ena/"). Remaining DNA from non-human samples can be made available upon request. All human samples and information on their corresponding phenotypes have to be obtained from the PopGen Biobank Kiel (Schleswig-Holstein, Germany) through a Material Data Access Form. Information about the Material Data Access Form and how to apply can be found at "https://www.uksh.de/g2n/Information+for+Researchers.html".

### Ethics approval and consent to participate

Human samples  
Study participants were randomly recruited from inhabitants of Schleswig-Holstein (Germany) who were recruited for the PopGen cohort. Five individuals from the PopGen biobank (Schleswig-Holstein, Germany) were randomly selected among the healthy and unmedicated individuals and included in the study without corresponding meta-information. Study participants collected fecal samples at home without conservation buffers in standard fecal tubes (sterile feces container 76 × 20 mm, Sarstedt) and shipped them immediately at room temperature or brought them to the collection center (within 24 h). Samples were stored at -80 °C until processing. Human feces (*N* = 4) were sampled and extracted following the procedures as described in Wang et al. 2016 [84]. A biopsy sample of the sigmoid colon was taken from a healthy control individual without macro- or microscopical inflammation (*N* = 1) and DNA was extracted as described in Rausch et al. [85]. Investigators were blinded to sample identities and written informed consent was obtained from all study participants before the study. All protocols were approved by the Ethics Committee of the Medical Faculty of Kiel and by the data protection officer of the University Hospital Schleswig-Holstein in adherence with the Declaration of Helsinki Principles.  
Animal and plant samples  
Wild-derived, hybrid mice were sacrificed according to the German animal welfare law and Federation of European Laboratory Animal Science

Associations guidelines. Hybrid breeding stocks of wild-derived *M. m. musculus* × *M. m. domesticus* hybrids captured in 2008 are kept at the Max Planck Institute Plön (11th lab generation). The approval for mouse husbandry and experiment was obtained from the local veterinary office "Veterinäramt Kreis Plön" (Permit: 1401-144/PLÖ-00469/). All samplings, including invertebrate and plant samples, were performed in concordance with the German animal welfare law and Federation of European Laboratory Animal Science Associations guidelines. Further details for each host type are provided in Additional file 1.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany. <sup>2</sup>Institute for Experimental Medicine, Kiel University, Kiel, Germany. <sup>3</sup>Department of Biology, Laboratory of Genomics and Molecular Biomedicine, University of Copenhagen, Copenhagen Ø, Denmark. <sup>4</sup>Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. <sup>5</sup>Lübeck Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany. <sup>6</sup>Institute of General Microbiology, Kiel University, Kiel, Germany. <sup>7</sup>Department of Evolutionary Ecology and Genetics, Zoological Institute, Kiel University, Kiel, Germany. <sup>8</sup>Zoological Institute, Kiel University, Kiel, Germany. <sup>9</sup>Molecular Physiology, Zoological Institute, Kiel University, Kiel, Germany. <sup>10</sup>Marine Ecology, Research Unit Marine Symbioses, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany. <sup>11</sup>Kiel University, Kiel, Germany. <sup>12</sup>Marine Ecology, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany. <sup>13</sup>Environmental Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany. <sup>14</sup>Environmental Genomics, Botanical Institute, Kiel University, Kiel, Germany.

Received: 8 February 2019 Accepted: 23 August 2019

Published online: 14 September 2019

### References

- Bosch TCG, McFall-Ngai MJ. Metaorganisms as the new frontier. *Zoology*. 2011;114(4):185–90.
- McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci*. 2013;110(9):3229–36.
- Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis*. 2015;26(1):26191.
- Morgan XC, Huttenhower C. Chapter 12: human microbiome analysis. *PLoS Comput Biol*. 2012;8(12):e1002808.
- Langille MGJ, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*. 2013;31(9):814–21.
- Hiergeist A, Glasner J, Reischl U, Gessner A. Analyses of intestinal microbiota: culture versus sequencing. *ILAR J*. 2015;56(2):228–40.
- Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL, Tringe SG. Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol*. 2015;6:771.
- Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol*. 2016;34(9):942–9.
- Walsh AM, Crispie F, O'Sullivan O, Finnegan L, Claesson MJ, Cotter PD. Species classifier choice is a key consideration when analysing low-complexity food microbiome data. *Microbiome*. 2018;6(1):50.
- Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, et al. The long-term stability of the human gut microbiota. *Science*. 2013;341(6141):1237439.
- Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front Microbiol*. 2016;7(459):459.

12. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*. 2013;499(7459):431–7.
13. Huson D, Auch A, Qi J, Schuster S. MEGAN analysis of metagenomic data. *Genome Res*. 2007;17(3):377–86.
14. Hong Nhung P, Ohkusu K, Mishima N, Noda M, Monir Shah M, Sun X, Hayashi M, Ezaki T. Phylogeny and species identification of the family Enterobacteriaceae based on dnaJ sequences. *Diagn Microbiol Infect Dis*. 2007;58(2):153–61.
15. De Cáceres M, Legendre P, Moretti M. Improving indicator species analysis by combining groups of sites. *Oikos*. 2010;119(10):1674–84.
16. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattai T, Mende DR, Sunagawa S, Kuhn M, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*. 2016;44(1):286–93.
17. Cantarel BL. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res*. 2009;37(Database issue):233–8.
18. Fink C, von Frieling J, Knop M, Roeder T. Drosophila Fecal Sampling. *Bio-protocol* 2017;7:e2547.
19. Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol*. 2016;14(11):e2000225.
20. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun*. 2017;8:14319.
21. Peres-Neto P, Jackson D. How well do multivariate data sets match? The advantages of a Procrustes superimposition approach over the Mantel test. *Oecologia*. 2001;129(2):169–78.
22. Gittleman JL, Kot M. Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst Zool*. 1990;39(3):227–41.
23. Murray AE, Rack FR, Zook R, Williams MJM, Higham ML, Broe M, Kaufmann RS, Daly M. Microbiome composition and diversity of the ice-dwelling sea anemone, *Edwardsiella andrillae*. *Integr Comp Biol*. 2016;56(4):542–55.
24. Schneider S, dos Santos VAPM, Bartels D, Bekel T, Brecht M, Buhmester J, Chernikova TN, Denaro R, Ferrer M, Gertler C, et al. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat Biotechnol*. 2006;24(8):997.
25. Heitschel U, Piel J, Degnan SM, Taylor MW. Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol*. 2012;10(9):641–54.
26. Wu JY, Jiang XT, Jiang YX, Lu SY, Zou F, Zhou HW. Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol*. 2010;10:255.
27. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, Shakya M, Podar M, Quince C, Hall N. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*. 2016;17(1):55.
28. Fadrosch D, Ma B, Gajer P, Sengamalay N, Ott S, Brotman R, Ravel J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. 2014;2(1):6.
29. Highlander S. Mock Community Analysis. In: Nelson EK, editor. *Encyclopedia of Metagenomics*. New York: Springer New York; 2013. p. 1–7.
30. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Herczeg R, Jung F E, et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol*. 2017;35(11):1069–76.
31. Westcott SL, Schloss PD. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*. 2015;3:e1487.
32. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, Angenent LT, Knight R, Ley RE. Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J*. 2012;6(1):94–103.
33. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, et al. Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nat Methods*. 2017;14(11):1063–71.
34. Lindgreen S, Adair KL, Gardner PP. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci Rep*. 2016;6:19233.
35. Xu Z, Malmer D, Langille MGJ, Way SF, Knight R. Which is more important for classifying microbial communities: who's there or what they can do? *ISME J*. 2014;8(12):2357–9.
36. Ochman H, Elwyn S, Moran NA. Calibrating bacterial evolution. *Proc Natl Acad Sci U S A*. 1999;96(22):12638–43.
37. Benton MJ. The origins of modern biodiversity on land. *Philos Trans R Soc B*. 2010;365(1558):3667–79.
38. Rota-Stabelli O, Daley Allison C, Pisani D. Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Curr Biol*. 2013;23(5):392–8.
39. Kuo CH, Ochman H. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct*. 2009;4:35.
40. Lebreton F, Manson AL, Saavedra JT, Straub TJ, Earl AM, Gilmore MS. Tracing the Enterococci from Paleozoic origins to the hospital. *Cell*. 2017;169(5):849–861.e813.
41. Bishop JR, Gagneux P. Evolution of carbohydrate antigens—microbial forces shaping host glycomes? *Glycobiology*. 2007;17(5):23R–34R.
42. Pickard JM, Maurice CF, Kinnebrew MA, Abt MC, Schenten D, Golovkina TV, Bogatyrev SR, Ismagilov RF, Pamer EG, Turnbaugh PJ, et al. Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature*. 2014;514(7524):638–41.
43. Schwartzman JA, Koch E, Heath-Heckman EAC, Zhou L, Kremer N, McFall-Ngai MJ, Ruby EG. The chemistry of negotiation: rhythmic, glycan-driven acidification in a symbiotic conversation. *Proc Natl Acad Sci*. 2015;112(2):566–71.
44. Martens EC, Chiang HC, Gordon JL. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe*. 2008;4(5):447–57.
45. Boulnois GJ, Roberts IS. Genetics of capsular polysaccharide production in bacteria. *Curr Top Microbiol Immunol*. 1990;150:1–18.
46. Mahdavi J, Pirincioğlu N, Oldfield NJ, Carlsohn E, Stooj J, Aslam A, Self T, Cawthraw SA, Petrovska L, Colborne N, et al. A novel O-linked glycan modulates *Campylobacter jejuni* major outer membrane protein-mediated adhesion to human histo-blood group antigens and chicken colonization. *Open Biol*. 2014;4(1):130202.
47. Tounkang S, Premkumar D, Gustavo S, Nathalie B, Yann B, Patricia C, Florence L, Olivier N, Brigitte G, Anne L, et al. Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. *Mol Microbiol*. 2008;70(3):762–74.
48. Roberts IS. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol*. 1996;50(1):285–315.
49. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10.
50. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files 1.33 edn. 2011. <https://github.com/najoshi/sickle>.
51. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
52. Hannon G. FASTX-Toolkit. In. [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/); 2010.
53. Edgar RC. UTX algorithm; 2015.
54. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–1.
55. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27(16):2194–200.
56. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261–7.
57. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, et al. The ribosomal database project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res*. 2003;31(1):442–3.
58. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 2012;6(3):610–8.
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
60. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One*. 2011;6(3):e17288.
61. Bushnell B, Rood J: BBTools bioinformatics tools, including BBMap. In. 37.28 edn. <http://sourceforge.net/projects/BBMap/>; 2017.

62. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
63. Nurk S, Meleshko D, Korobeynikov A, Pevzner P: metaSPAdes: a new versatile de novo metagenomics assembler. *arXiv preprint arXiv:160403071* 2016.
64. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–9.
65. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res*. 2004;32(1):11–6.
66. Kolbe DL, Eddy SR. Fast filtering for RNA homology search. *Bioinformatics*. 2011;27(22):3102–9.
67. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol*. 2017;34(8):2115–22.
68. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012;40(1):445–51.
69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
70. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
71. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22(22):4673–80.
72. Felsenstein J. DNADIST – Program to compute distance matrix from nucleotide sequences. 3.5c edn; 1993.
73. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490.
74. Team RC. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing. 3.3.2 edn; 2016.
75. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens MH, Wagner H: vegan: Community Ecology Package 1. 17-6 edn. 2011 <http://CRAN.R-project.org>.
76. Legendre P, Anderson MJ. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr*. 1999;69(1):1–24.
77. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 2001;26(1):32–46.
78. Hothorn T, Hornik K, Van de Wiel MA, Zeileis A. A Lego system for conditional inference. *Am Stat*. 2006;60(3):257–63.
79. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*. 2010;26(11):1463–4.
80. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20(2):289–90.
81. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RDC. nlme: Linear and Nonlinear Mixed Effects Models. 2011 <http://CRAN.R-project.org>.
82. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*. 1988;75(2):383–6.
83. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29(4):1165–88.
84. Wang J, Thingholm LB, Skieceviciene J, Rausch P, Kummel M, Hov JR, Degenhardt F, Heinsen F-A, Ruhlemann MC, Szymczak S, et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet*. 2016;48(1):1396–406.
85. Rausch P, Rehman A, Künzel S, Häslér R, Ott SJ, Schreiber S, Rosenstiel P, Franke A, Baines JF. Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci*. 2011;108(47):19030–5.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)





# ACKNOWLEDGMENTS

This thesis project would not have been possible without many people I was incredibly lucky to accompany me on this journey. You guided, helped, inspired, and supported me, and words are not enough to express my gratitude to you all.

First and foremost, I would like to thank **Hinrich Schulenburg**, who gave me the incredible opportunity to work with this fantastic group and pursue my Ph.D. Also, he kept on guiding and supporting me throughout my journey from the moment I came to Germany.

I want to thank **Katja Dierking** for: meticulously supervising my work, investing hours in discussions and meetings; thoroughly and relentlessly reviewing my writings, connecting me with scientists to collaborate with, providing me honest and generous advising both on the professional and personal level, and for making me feel at home away from home. So, generally for making my Ph.D. journey manageable in every aspect.

I want to thank my Thesis Committee (TAC) members: **Hinrich Schulenburg**, **Matthias Leippe**, and **John Baines**. Thank you for taking the time to read my progress reports, attend my talks, read my thesis, and provide valuable feedback and constructive criticism to guide my work.

I want to thank the **CRC** for the opportunity to work and collaborate in an inspiring and enthusiastic atmosphere. The opportunities the CRC created in attending conferences, retreats, workshops, and simply meeting up to chat with scientists from all over the world helped me to grow not only professionally but also personally in many ways. I also want to thank the **IMPRS** for the opportunities provided to connect with yet another amazing group of scientists.

A huge thanks to the **Evoecogen group members**, both past and present, for the phenomenal working atmosphere. It was a pleasure to be part of such a supportive and inspiring group. You made this journey super special. I am forever grateful, especially to **Carola Peterson**, **Barbara Pees**, and **Alejandra Zárate-Potes**, who taught me a lot and helped me immensely in setting-up my work in the lab. Carola, who taught me everything I know about *C. elegans* work, with her inspiring brutally-honest teaching style, and no-sugar-coating policy. Barbara, who also introduced me to the *C. elegans* world, with great energy and enthusiasm. A huge thanks go to Alejandra, who taught me everything I know about Bt and the survival assays. She taught with generous kindness and a never-ending supply of patience. You all made the long hours of work in the lab fun and full of positive energy.

Thanks to the microbiome crowd and the hours of discussions invested in this topic. I want to thank **Philipp Dirksen**, who did incredible work, and I was happy to follow in his footsteps. Thanks to **Nancy Obeng** for sharing a lab bench with me, and having interesting scientific exchanges. To **Jack Aidley**, for creating a humorous-critical atmosphere and for teaching us all how to criticize published papers to the brink. To **Jule Johnke** for her valuable input as well. Thanks also to the *C. elegans* immunity crowd for helpful discussions.

## ACKNOWLEDGMENTS

To the students I supervised, **Eva Stange**, **Bente Rackow**, and **Christoph Giez**. It was my pleasure to be working with you in the lab. To the HiWis who made the laborious lab-work humanly possible: **Eva**, **Bente**, **Myriam**, **Lena**, **Bentje**, and **Melinda**. It was a privilege to work with you.

Thanks are also due to my office mates: To **Leif** for lending a listening ear whenever needed and jumping in on technical problem-solving. Thanks to **Ashley** for the daily chats and advice, ranging from science to cats, babies, and life in general. It was a joy having you in the office. Thanks to **Roderich** and **Julia** for the enjoyable company as well.

Also to the rest of the group members: **Aditi**, for your funny stories; **Camillo**, for the incredibly fun atmosphere you created; **Christiana**, for sharing artistic appreciation; **Joao**, for the encouraging talks; **Niels**, for taking care of our office plants; **Sabrina**, for offering a helping hand in the lab when needed; **Sabrina K**, for graciously organizing group events at her place; **Sylvia**, for enduring my broken German (supplemented with my inevitable need to use sign language to fill in the missing words).

Many people took the time out of their busy schedule to read sections of my thesis and generously evaluate and proofread its content scientifically and linguistically. For the meticulous peer-review, thank you **Alejandra**, **Barbara**, **Carola**, **Christian**, **Nancy**, **Jule**, and **Jack**. Thanks to Barbara and Jule also for translating the thesis summary to German.

Last but not least, I want to thank my family in Lebanon. To my parents, my mom, **Shakeh**, and my dad, **Berge**. Your constant efforts made it possible for me even to get the chance to be here...without your tremendous sacrifices, relentless encouragement, perseverance, years of dedication in supporting me, pushing me, and investing in my education, none of this would have been even remotely possible! Words are not enough to express my immense gratitude to you. Thanks to my sister **Maria-Lena** and brother **Dikran**, your constant encouragement from overseas made me endure the obstacles and problems I faced throughout this journey. You were always there to motivate me and push me to do my best.

And to my family in Germany. To my husband, **Gomez**, thanks for making it all easy. Thanks for being my rock and for the unconditional support during the good and the bad. Thanks also for the scientific discussions and the critical evaluation of my ideas from the clinical perspective. To my son, **Karl**, you are my inspiration. I hope this goes on to teach you that you can accomplish anything you set your mind and heart to when you invest enough determination and hard work into it, no matter the odds, even in midst of a pandemic!



# CURRICULUM VITAE

## PERSONAL INFORMATION

Date of birth: Oct. 25, 1987

Place of birth: Beirut, Lebanon

Nationality: Lebanese

Place of residence: Kiel, Germany



## EDUCATION

08/2016-12/2021      Christian-Albrechts-Universität zu Kiel      Kiel, DE  
Collaborative Research Center (CRC) 1182  
International Max Planck Research School for Evolutionary Biology  
PhD studies

(01/2020-01/2021      Maternity leave)

09/2010-07/2013      American University of Beirut      Beirut, LB

Master of Science in Microbiology and Immunology  
Supervisor: Prof. Dr. Alexander Abdelnoor  
Thesis title: “Anti-*Aspergillus fumigatus* IgY antibodies and their protective efficacy in cyclosporine A treated-BALB/c mice”

09/2006-06/2010      Haigazian University      Beirut, LB  
Bachelor of Science in Biology

09/2009-06/2010      Haigazian University      Beirut, LB  
Teaching Diploma

## EXPERIENCE

05/2013-07/2016      American University of Beirut Medical Center      Beirut, LB

Experimental pathology, microbiology and immunology  
Principal Investigator: Prof. Dr. Ghassan Matar  
Research scientist

## Publications in peer-reviewed journals

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., Domin, H., Fraune, S., Von Frieling, J., Hentschel, U., Heinsen, F. A., Höppner, M., Jahn, M. T., Jaspers, C., **Kissoyan, K. A. B.**, Langfeldt, D., Rehman, A., Reusch, T. B. H., Roeder, T., ... Baines, J. F. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, 7(1), 133.

Zimmermann, J., Obeng, N., Yang, W., Pees, B., Petersen, C., Waschina, S., **Kissoyan, K. A.**, Aidley, J., Hoepfner, M. P., Bunk, B., Spröer, C., Leippe, M., Dierking, K., Kaleta, C., & Schulenburg, H. (2019). The functional repertoire contained within the native microbiota of

the model nematode *Caenorhabditis elegans*. *The ISME Journal*, 14, 26–38.

**Kissoyan, Kohar A B**, Drechsler, M., Stange, E.-L., Zimmermann, J., Kaleta, C., Bode, H. B., & Dierking, K. (2019). Natural *C. elegans* Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group. *Current Biology*, 29(6), 1030–1037.

Harakeh, S., Azhar, E., Almasaudi, S., **Kissoyan, K. A. B.**, Fadlallah, S., Tanelian, A., & Matar, G. M. (2019). Effects of a specific nutrient combination on ESBL resistance. *Saudi Journal of Biological Sciences*, 26(7).

Kanj, S. S., Tayyar, R., Shehab, M., El-Hafi, B., Rasheed, S. S., **Kissoyan, K. A. B.**, Kanafani, Z. A., Wakim, R. H., Zahereddine, N. K., Araj, G. F., Dbaibo, G., & Matar, G. M. (2018). Increased blaOXA-23-like prevalence in *Acinetobacter baumannii* at a tertiary care center in Lebanon (2007-2013). *Journal of Infection in Developing Countries*, 12(4).

Kanafani, Z. A., Fadlallah, S. M., Assaf, S., Anouti, K., **Kissoyan, K. A. B.**, Sfeir, J., Nawar, T., Yasmin, M., & Matar, G. M. (2017). Sites of colonization in hospitalized patients with infections caused by extended-spectrum beta-lactamase organisms: A prospective cohort study. *Antimicrobial Resistance and Infection Control*, 6(1).

Haidar-Ahmad, N., **Kissoyan, K. A. B.**, Fadlallah, S. M., El-Hajj, R., Saleh, M., Ghosn, N., & Matar, G. M. (2016). Genotypic and virulence characteristics of *Listeria monocytogenes* recovered from food items in Lebanon. *Journal of Infection in Developing Countries*, 10(7).

**Kissoyan, K. A.**, Araj, G. F., & Matar, G. M. (2016). Prevalence of carbapenem resistance genes and corresponding MIC90 in Enterobacteriaceae at a tertiary care center in Lebanon. *The International Arabic Journal of Antimicrobial Agents*, 6(1).

**Kissoyan, K.A.B.**, Bazzi, W., Hadi, U., & Matar, G. M. (2016). The inhibition of *Pseudomonas aeruginosa* biofilm formation by micafungin and the enhancement of antimicrobial agent effectiveness in BALB/c mice. *Biofouling*, 32(7).

Salloum, N. A., **Kissoyan, K. A. B.**, Fadlallah, S., Cheaito, K., Araj, G. F., Wakim, R., Kanj, S., Kanafani, Z., Dbaibo, G., & Matar, G. M. (2015). Assessment of combination therapy in BALB/c mice injected with carbapenem-resistant Enterobacteriaceae strains. *Frontiers in Microbiology*, 6(999).

Rajeh, M., Sabra, A., **Kissoyan, K. A. B.**, Fadlallah, S., & Matar, G. M. (2015). Molecular characterization of staphylococcal cassette chromosome mec and virulence encoding genes in methicillin-resistant staphylococci at a medical center in Lebanon. *The International Arabic Journal of Antimicrobial Agents*, 5(1), 3.

Moukhaiber, R., Araj, G. F., **Kissoyan, K. A. B.**, Cheaito, K. A., & Matar, G. M. (2015). Prevalence of *Clostridium difficile* toxinotypes in infected patients at a tertiary care center in Lebanon. *Journal of Infection in Developing Countries*, 9(7).

Fadlallah, S. M., Rahal, E. A., Sabra, A., **Kissoyan, K. A. B.**, & Matar, G. M. (2015). Effect of rifampicin and gentamicin on shiga toxin 2 expression level and the SOS response in *Escherichia coli* O104:H4. *Foodborne Pathogens and Disease*, 12(1).

**Kissoyan, K.A.B.**, Al-Akl, N. S., Sleiman, F., & Abdelnoor, A. M. (2013). Anti-*Aspergillus Fumigatus* IgY antibodies and their protective efficacy in cyclosporine a treated-BALB/C mice. *IOSR Journal of Pharmacy*, 13(7).