A simple path to complexity:

horizontal gene transfer in microbial communities

Dissertation

in fulfilment of the requirements for the degree of **Dr. rer. nat.**

of the Faculty of Mathematics and Natural Sciences

at Kiel University

Submitted by Pauline Buffard

Plön, 2023

I, **Pauline Buffard**, declare that apart from my supervisor's guidance, the content and design of this thesis are my own work and only use the sources listed. This thesis has neither as a whole nor partially been submitted to another examining body and none of it has been published or submitted for publication. The thesis has been prepared according to the Rules of Good Scientific Practice of the German Research Foundation. No academic degree has ever been withdrawn.

Contents

Abstract

In nature, microbial communities are inherent components of all ecosystems, actively shaping and being shaped by complex eco-evolutionary feedback processes. Importantly, these communities have a substantial impact on the health and disease of eukaryotic organisms. Genetic variation is the ultimate driver of phenotypes and ecosystem functioning, and Horizontal Gene Transfer (HGT) plays a crucial role in generating this variability. Nonetheless, a consequent knowledge gap remains in understanding the dynamic role of HGT in biologically meaningful microbial communities, and detecting HGT in complex microbial communities is challenging. Recent work has shown that Mobile Genetic Elements (MGEs) impact the movement of ecologically relevant genes, influencing community function. Building upon this, my thesis investigates HGT dynamics and its influence on community function in ecologically relevant microbial communities. For this, I experimentally evolved the nematode *Caenorhabditis elegans* with compost-derived microbial communities cultured on a single carbon source of cellulose paper. Regular transfers of communities, incorporating pooled MGEs, enabled the tracking of HGT.

In **Chapter 3**, I demonstrate the effectiveness of a bioinformatic pipeline to identify and trace HGT events facilitated by MGEs across diverse free-living communities, using metagenomic data from cellulose paper-grown microbial communities. **Chapter 4** explores the impact of HGT on the functioning of host-associated communities. Periodic nematode counting in repeated evolution experiments revealed fitness changes due to HGT, with both detrimental and beneficial effects observed. Microbial community composition was unaffected by HGT. Approaches were then sought to follow HGT dynamics in the nematode gut. Metagenomics proved challenging to use in this context. An in-house barcoded library of *Pseudomonas fluorescens* SBW25 was not suitable to track single lineages due to the strain's inability to establish a long-term association with *C. elegans*, despite observed beneficial interactions. The gut was dominated by *Ochrobactrum* and *Pseudochrobactrum*, diverse genera with a high propensity for gene gain. Individual genomes from these genera were successfully tracked in real-time, and multiple horizontally transferred sequences carrying ecologically relevant genes were detected.

Overall, this thesis demonstrates the utility of simplified experimental designs in comprehending the intricate eco-evolutionary dynamics of microbial communities, and paves the way for a deeper understanding of the significance of HGT in shaping community function.

Abstrakt

Translated by Manuela Spagnuolo and Christopher Böhmker

In der Natur sind mikrobielle Gemeinschaften ein fester Bestandteil aller Ökosysteme, die sowohl durch komplexe öko-evolutionäre Rückkopplungsprozesse geformt werden, diese aber auch selbst aktiv gestalten können. Wichtig ist, dass diese Gemeinschaften einen erheblichen Einfluss auf Gesundheit und Krankheit eukaryontischer Organismen haben. Genetische Variation ist die oberste Triebkraft für Phänotypen und das Funktionieren von Ökosystemen und horizontaler Gentransfer (HGT) spielt eine entscheidende Rolle diese Variabilität zu generieren. Dennoch besteht nach wie vor eine Wissenslücke, um die dynamische Rolle von HGT in biologisch bedeutsamen mikrobiellen Gemeinschaften zu verstehen und der Nachweis von HGT in komplexen mikrobiellen Gemeinschaften ist kompliziert. Neuste Arbeiten haben gezeigt, dass Mobile Genetische Elemente (MGEs) die Bewegung ökologisch relevanter Gene und damit die Funktion von Gemeinschaften beeinflussen. Darauf aufbauend untersucht meine Dissertation die Dynamik von HGT und deren Einfluss auf die Gemeinschaftsfunktionen in ökologisch relevanten mikrobiellen Gemeinschaften. Zu diesem Zweck habe ich den Fadenwurm *Caenorhabditis elegans* experimentell mit aus Kompost gewonnenen mikrobiellen Gemeinschaften, die auf Zellulosepapier als einziger Kohlenstoffquelle kultiviert wurden, evolviert. Regelmäßige Transfers von Gemeinschaften, die gepoolte MGEs enthielten, ermöglichten die Verfolgung von HGT.

In **Kapitel 3** zeige ich die Effektivität einer bioinformatischen Pipeline zur Identifizierung und Nachverfolgung von HGT-Ereignissen, die durch MGEs in verschiedenen freilebenden Gemeinschaften begünstigt werden, unter Verwendung metagenomischer Daten mikrobieller Gemeinschaften, die auf Zellulosepapier angezüchtet wurden. **Kapitel 4** untersucht die Auswirkungen von HGT auf die Funktion von wirts-assoziierten Gemeinschaften. Periodische Zählungen von Nematoden in wiederholten Evolutionsexperimenten ergaben Änderungen der Fitness aufgrund von HGT, wobei sowohl schädliche als auch positive Auswirkungen beobachtet wurden. Die Zusammensetzung mikrobieller Gemeinschaften wurde nicht durch HGT beeinflusst. Anschließend wurde nach Ansätzen gesucht, um die HGT-Dynamik im Fadenwurmdarm zu verfolgen. In diesem Zusammenhang erwies sich die Verwendung von Metagenomik als schwierig. Eine hauseigene barcodierte Bibliothek von *Pseudomonas fluorescens* SBW25 war nicht geeignet, um einzelne Linien zu verfolgen, da der Stamm keine langfristige Verbindung mit *C. elegans* aufbauen konnte, obwohl positive Interaktionen beobachtet wurden. Der

Darm wurde von *Ochrobactrum* und *Pseudochrobactrum*, diversen Gattungen mit einer hohen Neigung zum Gengewinn, dominiert. Individuelle Genome dieser Gattungen wurden erfolgreich in Echtzeit verfolgt und es wurden mehrere horizontal übertragene Sequenzen mit ökologisch relevanten Genen gefunden.

Insgesamt zeigt diese Doktorarbeit den Nutzen eines vereinfachten Versuchsaufbaus für das Verständnis komplizierter öko-evolutionären Dynamiken mikrobieller Gemeinschaften und ebnet den Weg für ein tieferes Verständnis der Bedeutung von HGT bei der Gestaltung von Gemeinschaftsfunktionen.

Acknowledgments

I first wish to express my gratitude to my supervisor **Prof. Dr. Paul Rainey** for providing me with the opportunity to work on this fascinating and expansive project. I am particularly appreciative of his trust in me, allowing me the freedom to explore and take ownership of the project. I am deeply appreciative of his guidance and mentorship, which have been precious in navigating the complexities of experimental design, scientific writing, and addressing challenging scientific queries. I am immensely appreciative of **Dr. Bram van Dijk** for his continuous presence and guidance throughout this journey. His expertise in metagenomics and bioinformatics, as well as his scientific discussions, have been invaluable. I value his relentless moral support and encouragement. Bram's contributions have significantly shaped my understanding and growth, and I am thankful for his mentorship.

I acknowledge the **Collaborative Research Center 1182** for its generous funding, which has underpinned this research work.

I am grateful to **Prof. Dr. Hinrich Schulenburg** for extending an invitation to join his laboratory and providing me with the opportunity for initial training. I would also like to express my thanks to the other members of Prof. Dr. Hinrich Schulenburg's team, particularly to **Dr. Agnes Piecyk**, **Lena Bluhm**, and **Lena Peters**, for their invaluable support and guidance, specifically regarding the study of *C. elegans*.

I am thankful to the members of my thesis advisory committee, **Prof. Dr. Hinrich Schulenburg**, **Prof. Dr. Eva Stuckenbrock**, and **Dr. Javier Lopez-Garrido** for their valuable insights and constructive feedback.

I would also like to extend my heartfelt thanks to **all the members of the MPB group**, as well as from the **ECB group**, both past and present, for their collaboration and contributions. In particular, I would like to thank **Dr. David Rogers** for his continuous availability, positive contributions to the collaborative ambiance within the department, and his consistent appreciation of my work and the work of others. I am also sincerely grateful to **Dr. Elisa Brambilla** for her proactive support, commitment, and problem-solving skills in the lab. I would like to express my warm gratitude to **Anja Baade** for her kindness, the laughter we shared, and her invaluable assistance in all aspects of the lab work, particularly for her significant support with the core experiment of my thesis. I am sincerely appreciative to all the

1   Introduction

"Complex is not sum of many simples, it is simply one complex."

Deepak Kripal, Sense of the Quiet

1.1   A World Ruled by Microorganisms

Microorganisms have evolved to adapt to almost all possible environmental niches[1]. They are a key component in a wide range of biogeochemical processes on Earth[2]. Furthermore, their ability to establish relationships with host organisms makes them very relevant in health and disease[3].

1.2   Embracing the Complex Eco-Evolutionary Dynamics of Microbial Communities

In the early stages of microbiology, the ability to characterize microbes relied solely on the capacity to grow cultures in the laboratory. This prevented access to the incredible phenotypic and genomic microbial diversity we nowadays hold[4]. In fact, much of this technical limitation has been overcome by extensively sequencing and profiling microbial communities[5]. On the down side, it is now easy to get lost in the complexity and abundance of these data. To disentangle the stunning complexity of microbial communities, it is necessary to put these systems back into their ecological and evolutionary context[6]. Microorganisms can neither be separated from their ability to interact with the biodiversity they are part of, nor from the structured and constantly changing environment where they establish niches. Very importantly, microorganisms are deeply intertwined within the evolutionary framework from which they are derived[7]. Although simplification through, for example, two-species experiments can give some insights into the mechanistic interactions between species, such simplification may hamper the ability to understand processes that are inherent to the complexity itself[8,9]. Environmental samples grown under well-controlled laboratory conditions are referred to as mesocosms. They provide a good trade-off between overly complex natural environments and over-simplified experiments. This intermediate level of environmental and genotypic complexity allows testing ecological theories on populations and communities[10–12].

1.3 The Crucial Role of HGT in the Ecology and Evolution of Microbial Communities

1.3.1 HGT, the necessary source of variation for microorganisms

Bacteria divide via binary fission to give rise to cells that in the absence of mutation, are genetically identical. However, short generation times and large population sizes generate a wealth of mutational diversity. This allows bacteria to rapidly adapt to their environment[13]. Nonetheless, the mutation-driven variability in bacteria is not sufficient to explain the extent of bacterial evolution[14]. In fact, a process similar to meiotic recombination in eukaryotes allows the acquisition of new genes and may compensate for the accumulation of slightly deleterious mutations[15]. Horizontal gene transfer (HGT), the ability of bacteria to acquire genes from sources other than their parents, the latter being called vertical gene transfer, satisfies this condition[16]. Typically, essential housekeeping genes show a pattern of vertical inheritance, whereas genes associated with more specialized functions tend to undergo horizontal transfer[17–19]. HGT plays a crucial role in shaping the course of microbial evolution and occurs at high rates within microbial communities[20–23]. Moreover, emerging evidence strongly indicates that HGT exerts a far greater influence than point mutations in driving the adaptive dynamics of bacterial populations[24,25]. Therefore, HGT certainly generates the variability necessary to fuel bacterial evolution.

1.3.2 MGEs, with their ability to self-sustain, are crucial vehicles for HGT

Mobile Genetic Elements (MGEs) serve as primary vehicles for HGT. For instance, plasmids and bacteriophages can transport fragments of bacterial genomes through conjugation and transduction processes, respectively[26,27]. Recent studies have begun to unravel the diversity of MGEs and the mechanisms underlying their transfer[26]. MGEs act as independent Darwinian entities capable of self-replicating, and self-integrating into bacterial genomes[28]. This means that their evolutionary trajectory may deviate from that of the host. Nonetheless, the persistence of MGEs is often attributed to their ability to transfer host-beneficial genes, as carrying cargo genes may also promote their own survival[29–31].

1.3.3 A direct link between HGT and community function

One central question to answer in microbial ecology and evolution focuses on the understanding of ecosystem functioning and its link with biodiversity[32]. While multiple levels of selection affect microbial communities through complex eco-evolutionary feedback processes, genes are the primary entity that determines the functioning of ecosystems[33,34]. It has been shown that HGT maintains species diversity

and helps against environmental disturbance, thereby promoting ecosystem stability[35,36]. Recently, Quistad *et al.*[37] further strengthened this notion by establishing a direct link between HGT and the functioning of complex microbial communities.

1.4   HGT in Host-Associated Microbial Communities

1.4.1   Host-microbe interactions

Biotic interactions within ecosystems extend beyond interactions among microorganisms and also include the interplay with higher organisms. The theory of endosymbiosis, which posits that eukaryotes emerged through the engulfment of bacteria that subsequently evolved into mitochondria and chloroplasts, highlights the essential role of these interactions in evolution[38]. Host-microbe interactions exist along a parasitism-mutualism spectrum. At one extreme of this spectrum, mutualism characterizes a reciprocal relationship where both the host and microbial symbiont rely on each other for their survival. In contrast, parasitism describes a scenario in which the microbial parasite benefits at the expense of the host, without providing any reciprocal benefits[39]. Whether positively or negatively, the presence of microorganisms greatly impacts the health of eukaryotic organisms[40–42].

1.4.2   HGT within the gut microbiome

1.4.2.1   *Influence of microorganisms in the gut*

The gut serves as a vast habitat for microbial communities within a host. Microorganisms contribute to various functions related to host health, including immune protection and digestion, and alteration of the gut microbiome can lead to various host diseases[40].

1.4.2.2   *The Gut, a hotspot for HGT*

The gut has been repeatedly shown to be a hotspot for HGT[24,43,44]. In fact, the large microbial density and diversity in this environment, and the presence of multiple niches, enhance interactions between community members[45]. The extreme conditions encountered by microorganisms are both signals for HGT and selection pressures for the maintenance of horizontally transferred ecologically relevant genes within the gut[44].

1.4.2.3   The effect of HGT on host-associated microbial communities

While the effect of HGT on the interaction of a single bacterial strain with a host has been considered, this may not fully capture the implications of HGT in a natural and complex environment[46]. In the past decade, the rise of metagenomics has enabled researchers to start disentangling the extensive diversity of MGEs *in vivo*[26,47–49]. Interestingly, certain plasmids have been found to be enriched with genes of ecological significance for bacterial life in the rumen of bovines[50]. Moreover, studies have observed specific HGT events among bacteria residing in the human gut[51,52]. Although research in this field is still in its early stages, these recent advancements indicate the crucial role of HGT in the ecological dynamics of microbial communities within hosts. However, the precise impact of HGT on the interaction between these communities and their host remains relatively understudied.

1.4.3   *C. elegans*, a good model for microbiome research

Caenorhabditis elegans, an easily maintainable model organism in the laboratory, has been extensively utilized to investigate various health-related processes[53]. As a result, there is now a comprehensive understanding of the nematode's biology and life-history traits. *C. elegans* is a bacterivorous nematode, and it has predominantly been studied in the context of single-strain bacterial nutrition. Nonetheless, in recent years, it has repeatedly been shown to be associated with diverse microorganisms in nature[54,55]. Consequently, there has been a growing emphasis on using *C. elegans* as a host model organism for microbiome research. *C. elegans* is specifically useful to test the effects of the microbiome on host fitness due to its ability to generate genetically homogeneous populations. This property allows for clonal reproduction, therefore, eliminating the response of host genotypic variation[56]. Moreover, the fitness of *C. elegans* has been found to be related to its microbial components, notably to its gut colonizers[57,58].

1.5   Detecting HGT in Microbial Communities

1.5.1   Limitations of current HGT detection methods

Detecting HGT poses a challenge, as distinguishing it from differential gene loss can be convoluted. To address this difficulty, two primary approaches are commonly employed: sequence composition analysis and phylogenetic tree comparisons. However, HGT detection can be hindered by small sample sizes and the rapid adjustment of sequences following HGT events. Moreover, these approaches often focus on

past instances of HGT, limiting our ability to directly observe the impact of HGT on microbial adaptation to new environments[59].

1.5.2 Leveraging the power of metagenomics to detect HGT

In recent years, the advent of metagenomics has provided unprecedented opportunities for detecting HGT that may go undetected by traditional methods[60]. Metagenomic approaches utilize read-mapping techniques that are capable of detecting more recent HGT events. By analysing structural variants (SVs) in donor-acceptor pairs, read-mapping can provide evidence of HGT. Additionally, split mapping and read pair information can enhance the detection of SVs. However, it is important to note that prior knowledge about the donor and acceptor genomes is required with this method for accurate analysis. Furthermore, determining the underlying cause and timing of SVs necessitates further investigation. It is worth mentioning that SVs can be attributed to various mechanisms beyond HGT, including chromosomal rearrangements, deletions, duplications, inversions, or translocations[61]. Consequently, relying on information derived from SVs presents similar limitations, as addressed above.

1.5.2.1 Detecting HGT in real-time

Real-time identification and tracking of HGT events are crucial for comprehending the eco-evolutionary dynamics of microbial communities. Recently, Quistad *et al.*[37] have proposed an innovative experimental evolution approach for detecting HGT in microbial communities. At the beginning of the experiment, microbial communities were introduced into mesocosms consisting of nitrogen-limited media and cellulose paper as the sole carbon source. After two weeks, they were divided into two separate communities named "horizontal" and "vertical". These communities underwent bi-weekly transfers. To monitor the progression of horizontally acquired sequences, DNA was selectively filtered and combined from all "horizontal" communities during each transfer to introduce MGEs from other communities. Conversely, the "vertical" communities were not exposed to this filtrate. Genomic DNA was extracted from the communities at each transfer, and metagenomic analyses were conducted to trace the dynamics of horizontally acquired genes, thereby identifying sequences unique to the "horizontal" community compared to their ancestor. In contrast to conventional approaches, this design thus renders possible read mapping across samples rather than within individual samples. Moreover, it allows the detection of the movement of DNA between communities without the need for prior knowledge of the community composition or specific donor and acceptor genomes. More recently, this approach has been

streamlined into a bioinformatic pipeline called Xenoseq[62]. Therefore, with this experimental design, and the metagenomic data delivered by it, the complex role of HGT in evolving microbial communities can begin to be disentangled.
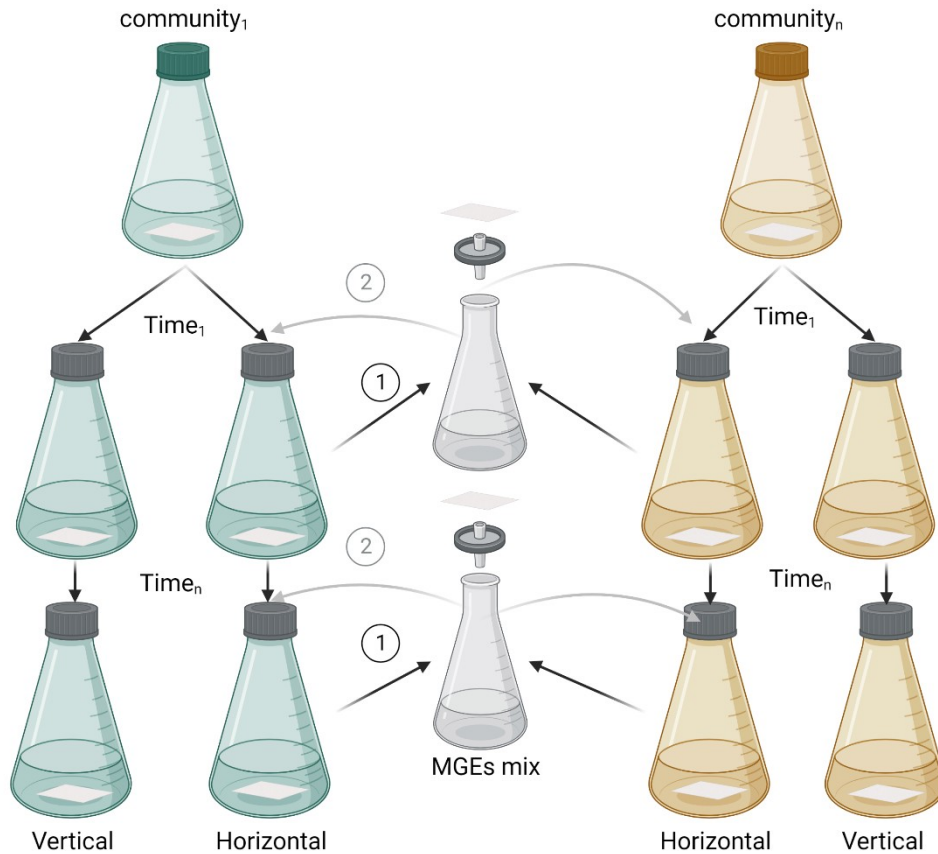
Figure 1.1. A simple experimental design to manipulate MGEs in microbial communities (adapted from Quistad *et al.*[37]). Multiple communities (from 1 to n) are grown in parallel in a minimum environment consisting of nitrogen-limiting medium and cellulose paper as a sole carbon source. At initial time (Time1), communities are split, and subsequently transferred on a bi-weekly basis. At each time of transfer, communities termed "horizontal" are additionally treated with a pooled filtrate of all "horizontal" communities (MGEs mix).

1.6   Summary and Aims

Microbial communities mould ecosystems, and in turn, are influenced by them, resulting in intricate eco-evolutionary dynamics. Previous studies have often dissected ecosystems to concentrate on individual components, potentially overlooking the intricate nature of their interactions. This approach may result in conclusions that lack applicability to real-world scenarios. Genetic variation is the primary entity that ultimately shapes ecosystem functioning. It can arise through two main mechanisms in

microorganisms: mutations and HGT. Previous research has provided valuable insights into the extensive nature of HGT and its remarkable adaptive potential, both in the context of free-living and host-associated microorganisms. In fact, HGT has emerged as a significant driving force behind microbial adaptation, at times surpassing the impact of mutations. This growing body of evidence emphasizes the importance of recognizing HGT as a major contributor to microbial evolution. However, despite the rapid progress in newer and more sensitive sequencing technologies, the technical complexities associated with real-time detection of HGT in complex systems pose one of the most formidable challenges in microbial ecology and evolution, namely achieving a comprehensive understanding of the role of HGT within environmentally-relevant communities.

In this thesis, I exploit an approach originally developed and demonstrated by Quistad *et al*.[37] to follow the dynamics of HGT and its impact on community function, while maintaining the integrity of the entire ecosystem. Specifically, I aim to demonstrate the utility of a newly developed pipeline, called Xenoseq, for tracking HGT across complex microbial communities and to explore the link between the presence of MGEs and community function in relation to the well-studied host nematode *C. elegans*.

1.6.1   Objective 1 – Tracing HGT in complex microbial communities using the Xenoseq pipeline

The first objective involves the use of the Xenoseq bioinformatics pipeline, which was developed in-house by van Dijk *et al*.[62], to accurately detect horizontally transferred genetic material across complex microbial communities. To achieve this, I employ the mesocosm-based approach previously established by Quistad *et al*.[37] that allows the movement of MGEs between communities as a framework for metagenomic sequencing, which serves as the input for the Xenoseq pipeline. Additional approaches are applied to the sequences identified as horizontally transferred by Xenoseq with the aim to gain further knowledge about the nature and function of these sequences.

1.6.2   Objective 2 – Investigating the impact of HGT on the functioning of gut microbiome communities in *Caenorhabditis elegans*

Secondly, the mesocosm-based approach established by Quistad *et al*.[37] is also applied to investigate the potential impact of HGT on the functioning of microbial communities in their interaction with the host nematode *C. elegans*. Different strategies are then investigated to track the real-time acquisition of HGT within the nematode gut. Finally, single dominant lineages of the *C. elegans* gut microbiome are tracked to follow the acquisition of MGE-driven HGT events.

2 Material and Methods

2.1 General Methods

2.1.1 Data availability

The raw data from this thesis, along with supplementary material that could not be included in the manuscript, can be accessed through the following link:

https://doi.org/10.17617/3.NJO672

The provided dataset includes raw sequencing data from whole-genome sequencing of the main and pilot experiments, 16S metabarcoding sequencing of gut-associated communities, metagenomics sequencing of cellulose paper and gut-associated communities, and nanopore sequencing of individual genotypes. Additionally, the raw data encompass all the assays conducted with *P. fluorescens* SBW25. Furthermore, the link provides access to supplementary tables from **Chapters 3 and 4**, as well as the scripts and files from **Chapter 2** necessary for reproducing the analysis.

2.1.2 Nematode and bacterial strains

species	strain	genotype	source
Caenorhabditis elegans	N2	wild-type	Caenorhabditis Genetics Center
Escherichia coli	OP50	wild-type	Caenorhabditis Genetics Center
Pseudomonas fluorescens	SBW25	wild-type	Two overnight cultures removed from PBR340[63]
Pseudomonas fluorescens	SBW25	attTn7::Tn7-PntpII-mScarlet-I-Gm	Integrated Tn7 construct from plasmid pMRE-Tn7-145* into SBW25, itself from two overnight cultures removed from PBR341, a second aliquot of PBR340
Pseudomonas lurida	Myb11	wild-type	Natural isolate from Northern Germany[55]

* https://www.addgene.org/118561/

2.1.3 Nematode maintenance

2.1.3.1 *Nematode transfer*

The N2 *C. elegans* strain was grown and maintained at 20°C on solid (2%) nematode growth media (NGM) inoculated with the *E. coli* OP50 strain[64]. The OP50 lawns were prepared by inoculating liquid LB media with a frozen stock of OP50, followed by an overnight incubation at 37°C. 700 µL of the overnight culture were then spread on the NGM plate and left to dry before introducing the transfer of nematodes.

2.1.3.2 *Synchronisation of nematode populations*

Nematodes were cultured on NGM plates seeded with OP50, following the above-mentioned procedure. To obtain synchronized nematode populations, when plates were crowded with nematode eggs, they were washed with 5 mL of M9 buffer (1X M9 salts; 0.1 mM $CaCl_2$; 2 mM $MgSO_4$). The washed eggs were then transferred to a sterile tube and mixed with 1 mL bleaching solution (0.5 mL 12% NaClO and 0.5 mL 5M NaOH). The tube was vortexed for 5-7 min until no more nematodes were visible. Next, the eggs were centrifuged for 1 minute at 3500 rpm and washed three times in 5 mL of M9 buffer. The eggs were then suspended in 2-5 mL of M9 buffer and incubated overnight at 20°C with gentle shaking until they hatch.

2.1.4 Compost microbial communities sampling and acclimation

2.1.4.1 *Sampling of compost communities*

Compost matter was sampled from three independent compost heaps located in the German county of Plön. 10 g samples from each compost heap were resuspended in 30 mL M9 buffer and vortexed with 20-30 glass beads for 2 min at maximum speed. The samples were prepared for storage by adding 500 µL of glycerol saline solution (70% glycerol + 8.5 g/L NaCl) to 1.2 mL aliquots of the resuspended compost samples and frozen at -80°C. The communities were named after the owners of the compost heaps they came from: Paul Rainey (P), Ines Schultz (I), and Norma Rivera (N).

2.1.4.2 *Acclimation of the communities to the laboratory environment*

Communities were acclimated prior to the commencement of the evolution experiments. For this, a frozen aliquot was initially thawed, and 100 µL of the sample was pipetted onto a fresh mesocosm plate. In this study, a mesocosm plate consists of a sterile 4cm2 square of cellulose paper placed in the centre of a 1.5% solid agar M9 buffer plate. After incubating the community mesocosms for two weeks at 20°C, the cellulose paper was removed from the M9 agar plates and placed in a tube with 20 mL M9 buffer. The tube was vortexed for 2 minutes at maximum speed and centrifuged for 10 minutes at 3500 rpm. The supernatant was discarded, and the tubes were refilled with 2 mL of M9 buffer. Next, 100 µL of the cellulose slurry solution was transferred onto the cellulose paper from a fresh mesocosm. This process was repeated for each community. After another two weeks of incubation, the cellulose paper was

processed in the same manner, and multiple frozen aliquots were prepared in 70% glycerol saline for each community.

2.1.5   The evolution experiments

2.1.5.1   *Reference genomes inoculation for the main evolution experiment*

In the main evolution experiment, before introducing microbial communities, eight known genotypes that were isolated from the pilot experiment were added to the mesocosms (**Table 4.S3**). To prepare for this, the isolates were individually cultured in 5 mL of Tryptic Soy Broth (TSB) at a $1/4^{th}$ dilution and incubated overnight at 30°C. Then, 1 mL of each culture was combined and centrifuged for 10 minutes at 5000 rpm. After removing the supernatant, the cells were washed three times in 8 mL of M9 buffer and resuspended in about 100 µL. The entire cell suspension was placed on a fresh mesocosm plate.

2.1.5.2   *Microbial community inoculation*

After completing the washing and acclimatizing processes described in **Chapter 2** sections **2.1.4.1** and **2.1.4.2**), the microbial community inoculum was prepared. From each frozen cellulose slurry, 300 µL was collected and thawed. The samples were then washed three times in 250 µL of M9 buffer and centrifuged for 1 minute at 13,000 rpm after each wash. Next, 100 µL of the washed slurry from each community was pipetted onto a fresh mesocosm plate. The three plates were then dried and incubated at 20°C. To serve as a negative control for microbe contamination by nematodes, one plate was incubated without microbial communities.

2.1.5.3   *First time of transfer*

2.1.5.3.1   Transfer of microbial communities

Following the inoculation of the microbial communities onto the mesocosms, the mesocosms were incubated for a period of two weeks. After the incubation period, each community was transferred to a 15-mL tube containing 2 mL of M9 buffer. The tube was vortexed for 2 minutes at maximum speed to ensure proper mixing. Subsequently, 100 µL of the resulting community paper slurry was transferred onto two separate solid fresh mesocosm plates.

2.1.5.3.2   Filtrates preparation and inoculation

Following the transfer of microbial communities, the remaining 1.8 mL of the paper slurry was centrifuged for 5 minutes at 5000 rpm. The supernatant was then filtered using a 0.2 μm pore size filter and a 1-mL syringe. 105 μL of each filtered community were combined to create the MGEs mix. For each pair of divided community plates, the "horizontal" treatment plate received a 100 μL aliquot from the MGEs mix, while the "vertical" treatment plate received a 100 μL aliquot filtered from its original community without inter-community mixing.

2.1.5.3.3   Storage of microbial communities and filtrates

After filtrate inoculation, any leftover filtrates were stored at -20°C. To store microbial communities, 4.8 mL of M9 buffer was added to the cellulose pellets. After thorough mixing, 1.2 mL of the mixture was dispensed into three cryo-tubes pre-filled with 500 μL of 70% glycerol saline and stored at -80°C. The remaining 1.2 mL was centrifuged for 5 minutes at 5000 rpm, and the pellet was stored at 4°C for a few hours before DNA extraction, which was performed on the same day.

2.1.5.3.4   Nematode Inoculation

Approximately 100 nematodes were pipetted onto the cellulose paper of each plate after filtrate inoculation. The number of nematodes was determined by counting the nematodes in five replicate 5 μL drops using a microscope. The developmental stage of the nematodes that were inoculated depended on the particular experiment being conducted. For the pilot experiment, sterile L1 larvae of nematodes were introduced to the mesocosms following their synchronisation using the procedure described in **Chapter 2** section **2.1.3.2**. In the main evolution experiment, nematodes were introduced after being fed a mixture of the eight genotypes isolated from the pilot experiments, which were prepared and mixed as described in **Chapter 2** section **2.1.5.1**. A lawn was created on peptone-free nematode growth media (PFM) plates using the entire pellet of the mixture, and sterile L1 nematodes were placed in the centre and allowed to feed for three days.

2.1.5.4   Subsequent times of transfer

The community mesocosms were incubated at 20°C and underwent transfers every two weeks after the initial division. In the pilot experiment, there were three transfers over a six-week period, while in the

main experiment, which was conducted as two replicates on two separate days, there were seven transfers over a 14-week period.

2.1.5.4.1 Nematode transfer

To separate the nematodes from the cellulose microbial community, 10 mL of M9 buffer was added to the plates and gently shaken for 5 minutes at 88 rotations per minute. The liquid containing the nematodes was then pipetted out and transferred into a 15-mL falcon tube. For each community, the number of live worms was determined as previously explained.

2.1.5.4.2 Microbial communities transfer

The microbial communities grown on the cellulose paper were transferred vertically, as described previously, and treated with either pulled (horizontal) or non-pulled (vertical) filtrates. The "horizontal" treatment involved adding the MGEs mix from the three filtrated communities named "horizontal" at the initial time of division, while the "vertical" treatment involved adding the filtrates from the individually filtrated communities named "vertical".

2.1.5.4.3 Storage of microbial communities, filtrates, and nematodes

The microbial communities and filtrates were stored as stated above. An average of 100 nematodes were transferred onto the cellulose paper of each corresponding plate, and nine aliquots of nematodes were stored in a final concentration of about 15% glycerol saline.

2.2 Methods Chapter 3 - Cellulose microbiomes

2.2.1 DNA extraction and sequencing

At the initial time of division and every time of transfer, DNA extraction was performed on about $1/4^{th}$ of the cellulose paper pellet using the DNeasy Powersoil kit from Qiagen. Subsequently, DNA underwent in-house library preparation using a self-made Tn5 enzyme[65]. The prepared libraries were then sequenced on an Illumina NextSeq platform, generating 150 bp paired-end short reads.

2.2.2   Xenoseq

2.2.2.1   Data preparation for Xenoseq

The raw reads were first copied to the working directory with simplified names, and unzipped (**Script 2.S1**). Afterwards, the names of the raw reads were replaced with the corresponding sample names (**Script 2.S2** and **File 2.S1**). The samples are named as follows:

R#1_T#2_#3#4/Control

#1: the replicate experiment (1 or 2)

#2: the time point (from 0 to 7)

#3: the community (P, I, or N)

#4: the filtrate treatment (H or V)

For samples without any community, #3 and #4 are replaced with "Control" in the sample name.

2.2.2.2   Running Xenoseq

The renamed raw reads were organized into separate folders based on the experiment they belonged to. For Xenoseq to detect novel sequences, it requires a reference library containing all samples from a given community where no new DNA is introduced. The references used in this study consisted of concatenated raw reads from the initial time point (T0) and all subsequent vertical time points (T1 to T7) for each specific community. Xenoseq was used to query samples manipulated with MGE cocktails (horizontal regime) for the emergence of novel sequences, hereafter referred to as unique. To establish the sequence origin, Xenoseq was run with the linking option through a BLAST search with a minimum alignment length of 500 base pairs and a minimum percentage identity of 98%. The sequences that were linked to an origin are hereafter referred to as xenotypic. As a negative control, the run included querying samples that have only received MGEs from their own community (vertical regime), using concatenated raw reads from the initial time point and subsequent horizontal time points for each community.

The analysis of sequences identified in the "horizontal" communities also involved the tracing option, mapping reads from longitudinal samples to obtain sequence abundance estimates across communities and in time (**Script 2.S3**). For each experiment, a metadata file was utilised, containing the relevant query and reference sample names (**Files 2.S2 and 2.S3**).

2.2.2.3 Xenoseq data analysis

2.2.2.3.1 Tracing of xenotypic sequences

The processed data output was utilized for visualizing the trajectories of the xenotypic sequences identified in each of the three "horizontal" communities (**Script 2.S4**). For each sample, the longest contig with a 100% BLAST match to its corresponding reference was manually selected for each community, based on the information provided in the output text and tabular files.

2.2.2.3.2 GC content analyses

For the GC analyses, only xenotypic sequences that were shorter than their corresponding sequences of destination or origin were considered. The destination sequences of the xenotypic sequences were subjected to a BLAST against a custom database consisting of the assembly of the community in which they were identified, and the GC content of the alignments was determined. The aligned portions of the sequences of origin with the xenotypic sequences were directly extracted from the BLAST output generated by the pipeline (.tbl files), and the GC content was calculated based on these alignments. For the analysis of GC composition differences between xenotypic sequences and their respective sequences of destination and origin, only the xenotypic sequences that had defined sequences of both destination and origin were included in the analysis. The GC content analysis can be replicated using **Script 2.S5** along with the instructions provided in **File 2.S4**.

2.2.2.4 MGEs identification among xenotypic sequences

To predict plasmids, PlasClass was employed (**Script 2.S6**), while VirSorter2 version 2.2.3 (**Script 2.S7**) was used for virus prediction[66,67].

2.2.2.5 Functional annotations of Genes encoded on xenotypic sequences

Xenotypic sequences were annotated using Prokka version 1.14.6 (**Script 2.S8**). The functional annotation of genes was assigned based on the KEGG Orthology (KO) database (https://www.genome.jp/kegg/pathway.html). The top six categories from the pathway maps were used to assign the three annotation levels to each gene.

2.3 Methods Chapter 4 - Nematode Gut Microbiomes

2.3.1 *C. elegans* and SBW25 community: exploration of the use of metagenomics

2.3.1.1 *Nematode feeding and microbiome extraction*

Nematodes were cultured in the presence of either *P. fluorescens* SBW25 (hereafter SBW25) or the washed and acclimatized microbial community P. After either a three-day (for nematodes fed on SBW25) or a two-week (for nematodes fed on the microbial community P) incubation period, the nematodes were harvested using 5 mL of M9 buffer, and pelleted by centrifugation at 5000 rpm for two minutes. The supernatant was carefully removed, and the nematode pellet was treated with 1 mL of 0.5% NaClO. The tube was incubated for 2 minutes at room temperature, followed by three times in 1 mL M9 buffer with centrifugation steps of 1 minute at 5000 rpm. Nematodes counts were performed as described in **Chapter 2** section **2.1.5.3.4**. Aliquots of either 100 (for nematodes fed on SBW25) or 1000 (for nematodes fed on the microbial community P) were prepared. To lyse the nematodes, a tissue lyser (TissueLyser II, Qiagen) was used for 5 minutes at a speed of 25s-1.

2.3.1.2 *DNA extraction procedures*

The impact of freeze-thaw treatments, microbiome compositions, and DNA extraction kits was investigated in two separate trials.

2.3.1.2.1 First trial

In the first trial, two different treatments were applied after lysis:

-   LFW (lysing-freezing-washing) treatment: Nematodes were subjected to a freeze-thaw cycle by freezing them at -80°C for 5 minutes, followed by thawing at 42°C for 2 minutes. This freeze-thaw process was repeated three times. Afterward, three washes with 1 mL of M9 buffer were performed.
-   L (lysis) treatment: Nematodes underwent lysis without any post-lysis treatment.

For DNA extraction, two different kits were used: the DNeasy Ultraclean Microbial kit from Qiagen and the Nucleospin Blood & Tissue kit from Macherey and Nagel.

2.3.1.2.2   Second trial

In the second trial, two treatments were applied after the lysis step. The cycles of freeze-thaw were repeated either once (N1) or 10 times (N10). The specific details of each treatment are as follows:

- LFW treatment: Nematodes were subjected to a freeze-thaw cycle by freezing them in liquid nitrogen for 10-15 seconds, followed by thawing at 42°C for 1.30 minutes. This freeze-thaw process was repeated three times. Afterward, three washes with 500 uL of PBS buffer.

- LF treatment: Nematodes were subjected to a freeze-thaw cycle by freezing them in liquid nitrogen for 10-15 seconds, followed by thawing at 42°C for 1.30 minutes. This freeze-thaw process was repeated three times. No subsequent washes were performed prior to DNA extraction.

DNA extraction was performed with the Nucleospin Blood & Tissue kit from Macherey and Nagel.

2.3.1.3   DNA sequencing

DNA underwent in-house library preparation using a self-made Tn5 enzyme[65]. The prepared libraries were then sequenced on an Illumina MiSeq platform, generating 150 bp paired-end short reads.

2.3.1.4   Data analysis

In order to obtain an initial understanding of the taxonomic distribution of the DNA, the raw reads underwent annotation using the tool diamond (**Script 2.S9**). The annotated reads were then analyzed using R (**Script 2.S10**).

2.3.2   *C. elegans* and *P. fluorescens* SBW25: exploration of the use of a barcoded library

2.3.2.1   Bacterial growth conditions

For all subsequent assays, each bacterium was cultured from a freezer stock in liquid LB medium for 18 hours at 28°C with shaking. A sample of each culture was then subjected to centrifugation for 2 minutes at 6000 rpm. The resulting pellet was washed twice with Ringer's solution, with each wash involving centrifugation for 2 minutes at 6000 rpm. Afterward, different quantities of this washed culture were used to inoculate PFM plates.

2.3.2.2 Lawn leaving assay

OP50 and SBW25 lawns were prepared on solid PFM 9-cm plates. Around the edge of each plate, 80 µL of the washed OP50 sample was spread, while at the center, 30 µL of either OP50 or SBW25 samples were applied. As a control, plates were inoculated with 30 µL of Ringer's buffer only. Each of these three conditions was performed in triplicate. The lawns were left to dry, and approximately 20 synchronized stage 4 (L4) larvae nematodes, previously grown on OP50, were transferred onto the central lawns using multiple 2-5 µL aliquots of Ringer's solution. After inoculation, the nematodes were counted on each plate and incubated at 20°C. The worms located outside the central lawn were manually counted after 24 and 48 hours. The toxicity of the bacteria was calculated over time for each replicate and condition using the leaving index:

Leaving index = Number of worms outside the central lawn / Total number of worms

2.3.2.3 Reproductive rate assay

In replicate assay 1, 50 µL OP50 and SBW25 lawns were prepared on the first day of the experiment. These lawns were prepared on 3-cm PFM plates. Each condition was performed in 10 replicates. In replicate assays 2 and 3, 50 µL OP50 and SBW25 lawns were freshly prepared every day. In replicate assay 2, the lawns were also prepared on 3-cm PFM plates, and the condition was performed in 10 replicates. In replicate assay 3, the lawns were prepared on 6-well PFM plates, and the condition was performed in 12 replicates. In each assay, single L4 *C. elegans* N2 nematodes were inoculated onto the bacterial lawns. Every day for six days, a single nematode was transferred to a fresh PFM plate, and the offspring were counted every second day from the old plate.

2.3.2.4 Colonization assay

60 µL OP50, Myb11, and SBW25 lawns were prepared, with each condition performed in six replicates. Synchronized L1 nematodes were placed on the lawns at the start of the experiment and incubated for three days at 20°C. To harvest the nematodes, 600 µL of 0.025% PBS-Triton X-100 (PBS-T) with 10 mM Tetramisole was used. The nematodes were pipetted onto a removable 20 µm pore size filter held by a filter holder. Subsequently, the nematodes were washed with 5 mL of PBS-T. To ensure surface sterilization, 100 µL of the flow-through was plated on LB agar medium as a control. The filter was then removed from the holder, and the nematodes were manually counted under a microscope. Next, the

filter was placed in a tube filled with 1 mL of PBS-T and 1 mm sterile beads. The nematodes were lysed using a tissue lyser as previously described. Various dilutions of the lysate in PBS-T were plated on LB agar medium and incubated at 28°C. The number of colonies was counted to calculate the number of colony-forming units (CFUs) per nematode.

2.3.2.5 Persistence assay

2.3.2.5.1 Bleach assay

Preliminary tests were conducted to evaluate the impact of NaClO concentration on the fluorescently tagged SBW25 strain with mScarlet (hereafter SBW25-mScarlet), which was used for subsequent assays. In these tests, 100 µL of an overnight culture of SBW25-mScarlet was passed through a filter held by a holder. 5 mL of diluted NaClO (0, 0.05%, and 0.5%) were pipetted through the filter. Following that, 5 mL of PBS was used to wash the filter. The filter was then removed from the holder and placed into 1 mL of water. Dilutions of the solution were plated on LB plates and incubated at 28°C. The number of colonies was subsequently counted to calculate the number of CFUs per nematode. This assay was performed on three separate days, and for each assay, six technical replicates were conducted.

2.3.2.5.2 Inside *C. elegans* gut

120 µL lawns were prepared from 5 times concentrated overnight cultures of SBW25-mScarlet. Approximately 30-50 synchronized L1 nematodes were inoculated on PFM-SBW25 6-well plates and incubated for three days at 20°C. After three days, the nematodes were washed off the plates using 600 µL of PBS-Triton X-100 (PBS-T) solution and pipetted onto a filter with a 20 µm pore size. The nematodes were then washed with 5 mL of PBS-T, and the filter was placed in a tube containing 1 mL of PBS-T. The nematodes were centrifuged at 6000 rpm for 30 seconds, and approximately 950 µL of the supernatant was discarded. The remaining approximate 50 µL, containing the nematodes, was pipetted onto a fresh PFM 6-well plate spotted with 120 µL OP50 lawns. To prevent overcrowding of nematodes on the plates in subsequent generations, the nematodes were transferred to fresh PFM-OP50 6-well plates every day using filters in a similar manner. This helped to eliminate some of the smaller worms generated after the switch to OP50 as the food source.

Every day, the presence of SBW25-mScarlet within the nematode gut was assessed in two ways. First, the bacterial fluorescence in the gut was visualized using microscopy. A few nematodes were pipetted onto a microscope glass slide using approximately 10-20 µL of 10 mM Tetramisole solution

(dissolved in PBS-T) to immobilize the worms. A cover slip was added, and the nematodes were observed using brightfield or the Scarlet (excitation filter: 538-562 nm; emission: 570-640 nm) fluorescent channel of the AxioZoom.V16 microscope. Secondly, the bacterial load present within the nematode gut was quantified by plating dilutions on LB + Gm agar medium. This assessment was performed in six independent populations grown on PFM-OP50 (or PFM-SBW25 for time point 0) 6-well plates. The procedure was similar to the colonization assay but with the use of 0.5% NaClO for surface sterilization.

2.3.2.5.3   Outside *C. elegans* gut

After incubation, each plate was washed following a procedure similar to the previously described assay, but only 900 µL of the solution was removed from the tube after centrifugation. The tubes were vortexed, and from the remaining 100 µL, 2-5 µL was used to transfer the nematodes to fresh OP50 lawns. To quantify the bacterial load, 20 µL of the remaining solution was pipetted for dilutions and plated on LB + Gm agar medium, being careful to avoid pipetting any nematodes. This procedure was repeated every few days. Nematodes were sampled and visualized using microscopy as described above.

2.3.3   *C. elegans* and Community: Exploration of the Use of Single Genome Sequencing

2.3.3.1   Nematode surface-sterilization

During the evolution experiments, at times of transfer, the remaining nematodes were collected and subjected to a series of treatments to remove microbes from their environment and surface. They were first centrifuged at 5000 rpm for 2 minutes, and the supernatant was removed. Then, 1 mL of 10 mM Tetramisole was added, and the mixture was incubated at room temperature for 10 minutes after gentle mixing. After another centrifugation step, the supernatant was removed, and 1 mL of 0.5% NaClO was added. The tube was incubated for 1 minute at room temperature, followed by another centrifugation step for 1 minute at 5000 rpm. Nematodes were washed three times in 1 mL of M9 buffer with centrifugation steps of 1 minute at 5000 rpm. Aliquots of about 1000 nematodes, calculated from counting at the time of transfer, were prepared. The aliquots were then centrifuged at 13,000 rpm for 1 minute and left in 20 µL of buffer. Most aliquots were frozen at -80°C for backup after the addition of 3 µL of glycerol saline.

2.3.3.2   Microbiome extraction

After each transfer, two fresh aliquots per community were collected and processed for microbiome extraction within a few hours. Alternatively, pulled frozen aliquots of surface-sterilized nematodes were used for 16S metagenomics analysis. For whole-genome sequencing (WGS), either frozen surface-sterilized pulled aliquots or fresh frozen surface-sterilized aliquots from pulled frozen live nematodes were used. In all cases, 980 µL of PBS-Triton X-100 (PBS-T) buffer was added to each aliquot tube, along with 1 mm beads. The surface-sterilized nematodes were lysed for 5 minutes at 25s-1 using a tissue lyser (TissueLyser II, Qiagen).

2.3.3.2.1   16S microbiome meta barcoding

2.3.3.2.1.1   DNA extraction

One of the tubes containing the processed nematodes was used for DNA extraction using the NuceloSpin Blood&Tissue kit and DNA was stored at -20°C prior to sequencing.

2.3.3.2.1.2   Sequencing and analysis

The hypervariable V1V2 region of the 16S rRNA gene was amplified using primers 27F and 338R and the library was prepared using the Miseq Reagent Kit v2 (MS-102-2003). Sequencing was performed in-house using the Illumina MiSeq platform to generate 250-bp paired-end reads. Data were prepared and analyzed using a combination of R tools (**Scripts 2.S11 and 2.S12**).

2.3.3.2.2   Whole-genome microbiome isolate sequencing

2.3.3.2.2.1   DNA extraction

The other tube was diluted 10 and 100 times and each diluted and non-diluted sample was plated onto TSA 1/4th solid agar plates and incubated at 30°C for 24h to 48h until the colony have grown. For each plate, 20 colonies were re-streaked on 1/4th solid agar TSA plates. This last step was repeated one more time. About 20 colonies from each derived community were stored at -80°C after resuspension in 100 µL M9 buffer and the addition of 200 µL 70% glycerol saline. To store and extract DNA from colonies, stored colonies were thawed and re-grown in 5 mL LB for 24 hours at 30C under shaking. DNA was extracted using the NucleopSpin Blood&Tissue kit.

2.3.3.2.2.2 Sequencing and data preparation

2.3.3.2.2.2.1 Short-read sequencing

Libraries were prepared either in-house using the self-made Tn5 enzyme (pilot experiment) or by Eurofins (main experiment), and 150 bp paired-end short-read genomic DNA was sequenced using the Illumina Nextseq platform, either in-house (pilot experiment) or by Eurofins (main experiment)[65]. The raw reads were first copied to the working directory, and their names were simplified (**Scripts 2.S13** and **2.S14**). Adapters were removed and quality trimming was performed using fastp (**Script 2.S15**). Quality was summarized using multiqc, for each of the pilot and main experiment isolates, respectively (**File 2.S5** and **2.S6**). Short reads were assembled using metaSpades (**Script 2.S16**). Assembly quality was assessed using CheckM and genomes having a contamination threshold below 5% or a completion threshold below 98% were discarded (**Script 2.S17**).

2.3.3.2.2.2.2 Long-read sequencing

In addition to Illumina sequencing, the genomic DNA of eight selected isolated genotypes from the pilot experiment was also barcoded using the Rapid Barcoding Kit SQK-004 from Oxford Nanopore Technologies, and long-read sequencing was performed using the MinION platform. Barcodes 1 to 8 were used for the eight reference isolates, respectively for references 1, 1', 2, 3, 4, 5, 6, and 7. Reads from each sample were concatenated, and files were manually renamed with the corresponding sample names. Adapters were removed using porechop and quality filtering was done using fastp (**Script 2.S18**). Quality was summarized using multiqc (**File 2.S7**). Long reads and short reads were assembled using Unicylcer hybrid assembly tool (**Script 2.S19**).

2.3.3.2.2.3 Data analysis

2.3.3.2.2.3.1 Functional annotations and phylogeny building

Short-read assemblies were annotated using Prokka version 1.14.6 (**Script 2.S8**). *gyrA* nucleotide sequences were extracted using R (**Script 2.S20**). *gyrA* sequences from the long-read sequenced isolates were extracted in the same way. *gyrA* sequences were then aligned using Clustal Omega[68,69]. A neighbor-joining phylogeny (Tamura-Nei distance method) was inferred from the alignment for each experiment. Phylogenies were built using Geneious, and visualized using iTOL.

The .fna output files from the Prokka analysis were categorized into groups based on their classification as either belonging to *Ochrobactrum*, *Pseudochrobactrum*, or a specific clade within these genera, determined by the *gyrA* phylogeny. Pangenome analysis was conducted for each group using the pagoo R tools on the assemblies that were previously annotated using Prokka version 1.14.6 (**Script 2.S21**).

2.3.3.2.2.3.3   *Detecting MGEs within end-point isolates in the main evolution experiment*

References were grouped with the end-point isolates that shared the same *gyrA* sequence as them, for each replicate of the main experiment. Within each group, the filtered and trimmed short reads were mapped onto the assembly of the reference, and SNP calling was performed using Bcftools version 1.9 (**Scripts 2.S22 and 2.S23**). If the alignment resulted in less than 20 SNPs, the reads that did not map to the reference were assembled (**Script 2.S24**). Contigs above 10,000 bp were retained for further analysis. The contigs were first annotated using Prokka version 1.14.6 (**Script 2.S8**). The presence of MGEs was determined by using online tools such as Phaster (for prophages) and ICE finder (for ICE)[70,71]. Plasmids were predicted using PlasClass, and viruses were identified using VirSorter2 version 2.2.3 (**Scripts 2.S6 and 2.S7**)[66,67].

3 Exploring the Dynamics of Horizontal Gene Transfer in Free-Living Microbial Communities

3.1   Introduction

Microorganisms reside in immensely complex communities. Recent advances in metagenomics have enabled microbiologists to start unraveling the complexity of these microbial communities[5,37,72]. Nonetheless, there is still a lack of knowledge when it comes to understanding microbial communities as dynamic systems[9]. HGT is a mechanism of utmost significance in the microbial world as it provides the genetic variation necessary to fuel microbial evolution[20–23,73]. Therefore, HGT can substantially act on ecosystem functioning[33,34]. Despite this crucial influence on eco-evolutionary processes, HGT has thus far poorly been studied in the context of natural or semi-natural communities. In fact, detecting HGT in these complex systems is intricate, and current detection methods hinder real-time detection, limiting our ability to explore the processes that are inherent to these dynamic systems[59,61]. The experimental approach introduced by Quistad *et al*.[37] addresses this challenge by implementing a filtering and pooling strategy for MGEs obtained from different evolving communities. By utilizing metagenomics techniques to identify unique sequences specific to each community, this approach enables the tracking of horizontally acquired sequences and their dynamics across different communities. To streamline the analysis process, a bioinformatic pipeline named Xenoseq has been developed[62]. In addition to its efficiency in analysis, Xenoseq was specifically designed to overcome one shortcoming from the analysis performed by Quistad *et al*.[37], namely to differentiate between the factors contributing to the uniqueness of sequences, providing direct evidence for their origin.

The primary aim of this chapter is to employ the Xenoseq bioinformatic pipeline to detect HGT in conjunction with a short-term evolution experiment during which communities are repeatedly supplemented with MGEs obtained from other communities (**Figure 3.1**). Subsequently, complementary analyses are presented to gain insights into the nature and function of the horizontally transferred sequences identified by Xenoseq.

3.2   The Experimental Design

The entire thesis revolves around an experimental design that has been adapted from the aforementioned work of Quistad *et al.*[37]. Hereafter, I describe the shared characteristics of this design, which apply to all versions of the experiments conducted in this study.

The three initial communities, named P, I, and N, were derived from different garden compost and have been established in a laboratory environment consisting of a solid nitrogen-limited media and a single carbon source of cellulose paper. At the start of the experiment, communities were inoculated on the mesocosms and, after two weeks, were split into two communities, respectively named "horizontal" and "vertical". At this point, nematodes were introduced to the mesocosms. Communities were transferred bi-weekly together with nematodes. To follow the dynamics of horizontally acquired sequences, the DNA was filtered and pooled from all three "horizontal" communities at times of transfer to enhance HGT. In contrast, the "vertical" communities were only given a filtered sample of their own (**Figure 3.1**).

In the course of this thesis, two versions of the experiment will be analysed, hereafter referred to as "pilot" and "main" evolution experiments. In this section, I delineate the disparities between the pilot and main evolution experiments and explain their specific objectives and applications, associated with references to the corresponding results within the thesis for further clarification.

The pilot evolution experiment was conducted for a total of six weeks, i.e. three transfers. The aims of this experiment were, first, to assess whether HGT impacts nematode fitness (see **Chapter 4** section **4.2**), and then to acquire knowledge about the members of the nematode gut microbiome (see **Chapter 4** section **4.5**). This experiment distinguishes itself in that communities were inoculated on fresh sterile mesocosms and the introduced nematodes were stage 1 (L1) larvae that had not previously come into contact with any microorganisms[74] (**Figure 3.1**).

The main evolution experiment was performed for 14 weeks, i.e. seven transfers. This experiment was implemented to follow the movement of HGT between the free-living communities cultured on cellulose paper using the Xenoseq pipeline (**Chapter 3**). Moreover, similarly to the pilot experiment, it aimed to gain a better understanding of the effect of HGT on community function (see **Chapter 4** section **4.2**). Finally, this experiment was specifically designed to track HGT acquisition within the nematode gut (see **Chapter 4** section **4.6**) based on the exploration of the microbiome performed in the pilot experiment (see **Chapter 4** section **4.5**). To achieve this goal, nematodes were fed with a

mixture of eight known bacterial genotypes isolated during the course of the pilot experiment and whose genomes had been thoroughly sequenced. Because nematodes acquire their microbiomes mostly horizontally through feeding, colonization of these eight genotypes by subsequent generations of nematodes was necessary to ensure their persistence within the nematode gut until the end of the experiment. Therefore, the mixture was also displayed on the mesocosms before introducing the communities. To account for the non-independence of horizontal samples caused by the mixing of the filtrates, the experiment was not expanded by adding more communities to increase representativeness. Instead, the entire experiment was replicated twice on separate days (**Figure 3.1**).
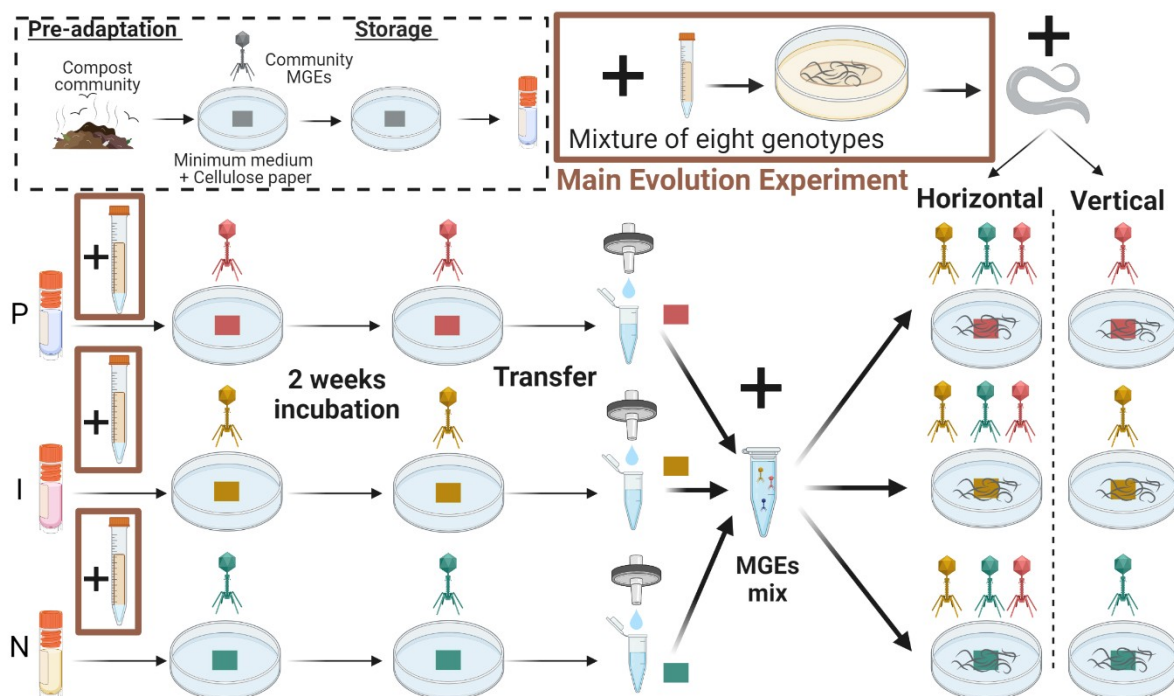
Figure 3.1. Design of pilot and main evolution experiments. Pre-adaptation (top-left dotted line rectangle): Initial communities (colours: P, I, and N are respectively highlighted in red, yellow, and green) come from garden compost and have been established in a laboratory environment consisting of a solid nitrogen-limiting medium and a single carbon source. Evolution Experiments: Two variations of the evolution experiment were conducted: the pilot experiment spanned six weeks (three transfers), while the main evolution experiment lasted 14 weeks (seven transfers). For the main experiment, a mixture of eight genotypes isolated during the pilot experiment was introduced to fresh mesocosms before introducing the communities. Communities were added to the mesocosms at the start of the experiments. After two weeks, each community was split in two and inoculated with either sterile larvae nematodes (pilot) or adult nematodes previously fed on the eight aforementioned genotypes (main). Communities were transferred bi-weekly together with the nematodes and DNA was filtered and pooled from either all communities (horizontal regime) or their original community (vertical regime) at times of transfer.

3.3   Xenoseq, a Straightforward Bioinformatic Pipeline for Detecting HGT between

Communities

Xenoseq is a recently developed computational pipeline that aims to identify sequences of non-native origin derived from metagenomics data, referred to as xenotypic (from *xenos* meaning foreign and *typic* meaning type). In combination with the "main" version of the evolution experiment described earlier, Xenoseq was implemented to directly identify these sequences (**Figures 3.1** and **3.2**). To achieve this, before the initial splitting of the communities and at each subsequent time of transfer, DNA was extracted and sequenced from the communities grown on cellulose paper, i.e. exempted from the nematodes, that received either a pooled (horizontal regime) from other communities or a non-pooled (vertical regime) filtrate.

Xenoseq takes raw short reads obtained from metagenomics sequencing as input and performs all preparation steps including filtering and trimming. The pipeline then proceeds by identifying unique sequences in a particular "horizontal" community through comparison with a reference community. For this, reads from the query "horizontal" community are mapped onto its equivalent vertical reference assembly, and reads that do not map, are assembled into unique sequences. Two potential explanations exist for identifying sequences as unique. One possibility is that these sequences were introduced from an external source and were not originally present in the mesocosm. Alternatively, these sequences may have been initially present but below the detection threshold and subsequently increased in frequency. To differentiate between uniqueness resulting from demographic changes and that arising from the presence of horizontally transferred sequences originating from other communities, the next step involves determining the foreign origin of these unique sequences. This is accomplished by conducting a BLAST search against the cross-assemblies of the two other communities. Finally, the resulting xenotypic sequences are tracked across time and all communities by mapping the reads from these communities back onto the assembled xenotypic sequences (**Figure 3.2**).
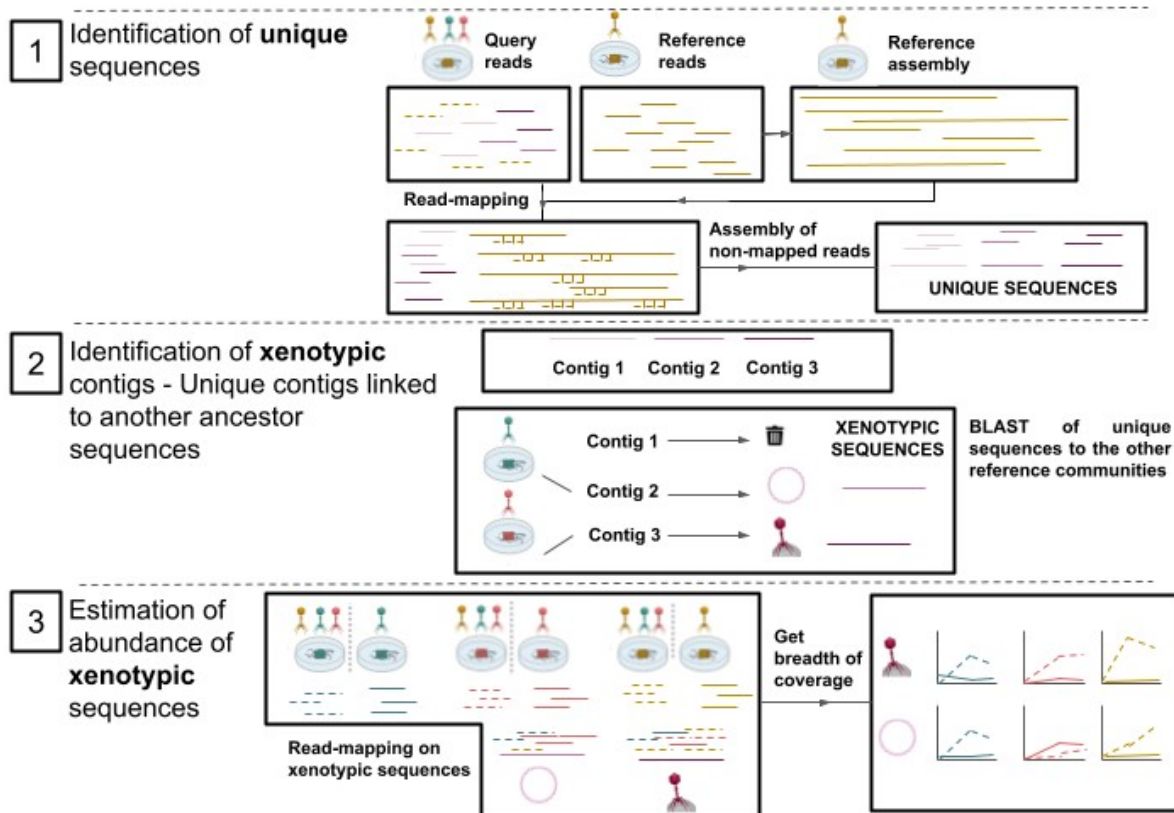
Figure 3.2. Xenoseq, a bioinformatic pipeline to detect and trace MGEs between complex communities (adapted from van Dijk *et al.*[62]). (1) All reads from each given ancestral community and its descendants from the vertical regime are cross-assembled to build the reference. Reads that do not map to this reference are assembled to constitute the unique sequences. (2) Unique sequences are blasted against the other reference assemblies allowing the unique sequences to be linked to their community of origin; these unique sequences are called xenotypic (*xeno-*, foreign; *typic-*, type). (3) Abundance of xenotypic sequences within each community and across time is traced. The reads from each community are mapped against xenotypic sequences and the likelihood of the presence of the sequence is estimated using the breadth of coverage (percentage of xenotypic contig bases covered by at least one read).

3.4   Assessing the Efficacy of the Xenoseq Pipeline for Detecting Unique and Xenotypic Sequences in Microbial Communities

3.4.1   Effect of conservative pipeline conditions on the proportion of xenotypic to unique sequences

The pipeline's efficacy was first evaluated by computing the number of sequences classified as unique and xenotypic. To minimize the possibility of confusing highly dynamic within-sample sequences with foreign sequences, the pipeline used a cross-assembly of all vertical time points as the reference for each community, instead of simply using the ancestor communities as references. Besides, a

conservative BLAST with a high minimum percentage identity, i.e. 98%, was performed to identify the origin of unique sequences. The minimum alignment length selected for the BLAST was 500 bp.

Unique and xenotypic sequences were detected in both replicate experiments across all "horizontal" communities. In both experiments, the number of xenotypic sequences represented only approximately 9% (400/9570) and 2% (406/26230) of the numerous unique sequences. The ratio of xenotypic sequences compared to unique sequences was similarly low when considering each community separately. In the first experiment, the ratios were 0.5% (7/1300) for community P, 4% (41/1075) for community I, and 5% (352/7195) for community N. In the second experiment, the ratios were 1% (37/3790) for community P, 0.8% (137/17302) for community I, and 4% (232/5408) for community N (**Figure 3.3 and Table 3.S1**). This shows that most of the sequences identified as unique were not newly introduced but likely present below the threshold for metagenomic detection in the references. This is consistent with the conservative conditions that were deliberately imposed on the pipeline. Therefore, although less numerous, the xenotypic sequences are likely to be true foreign sequences.

3.4.2   Assessing the effectiveness of the pipeline through Xenoseq analysis of "vertical" samples

To verify the efficacy of the pipeline, Xenoseq was run to detect unique and xenotypic sequences in the "vertical" samples, using a cross-assembly of all "horizontal" communities at all time points as a reference. As the "vertical" communities were not exposed to pooled filtrates from other communities, no foreign sequences should be detected using this control approach. The results showed that the number of unique and xenotypic sequences detected in the vertical samples was substantially lower compared to those found in the horizontal samples. In fact, the maximum number of unique and xenotypic sequences detected in the vertical samples was two and one, respectively. Moreover, in most samples, no xenotypic sequences were detected at all, thereby confirming the effectiveness of the pipeline (**Figure 3.S1**).

3.4.3   Partitioning of xenotypic sequences among communities

In both experiments, the partitioning of xenotypic sequences among the different communities was similar. Most of the xenotypic sequences were found to have transferred in community N, i.e. 88% (352/400) and 57% (232/406) in both experiments respectively. While the number of xenotypic sequences in the I community varied between the two experiments, i.e. and 10% (41/400) and 34%

(137/406), the P community always exhibited a lower number of xenotypic sequences compared to the two other communities, i.e. 2% (7/400) and 9% (37/406) (**Figure 3.3**).

3.4.4   Dynamic of unique and xenotypic sequences over time

The timing of the appearance of both unique and xenotypic sequences was strikingly similar in both replicate experiments. To start with, the detection of these sequences began at the same time, i.e., eight weeks (time four of transfer) after the start of the experiment. In all communities of both experiments, a peak in abundance was observed between times four and six of transfer, followed by a decline in most cases. In the first experiment, the highest number of unique and xenotypic sequences was observed at time five, with 2520 and 76 sequences, respectively. In the second experiment, this community did not show a clear peak in the number of unique sequences, but a peak in the number of xenotypic sequences was observed at time six, with 116 sequences. For the I community of the second experiment, the highest number of unique and xenotypic sequences, 6283 and 57, respectively, was detected at times four and five. Despite the limited number of sequences detected in community P, small but noticeable peaks of xenotypic sequences were observed, particularly at time four of transfer during the second experiment, with a total of 18 sequences detected (**Figure 3.3**).
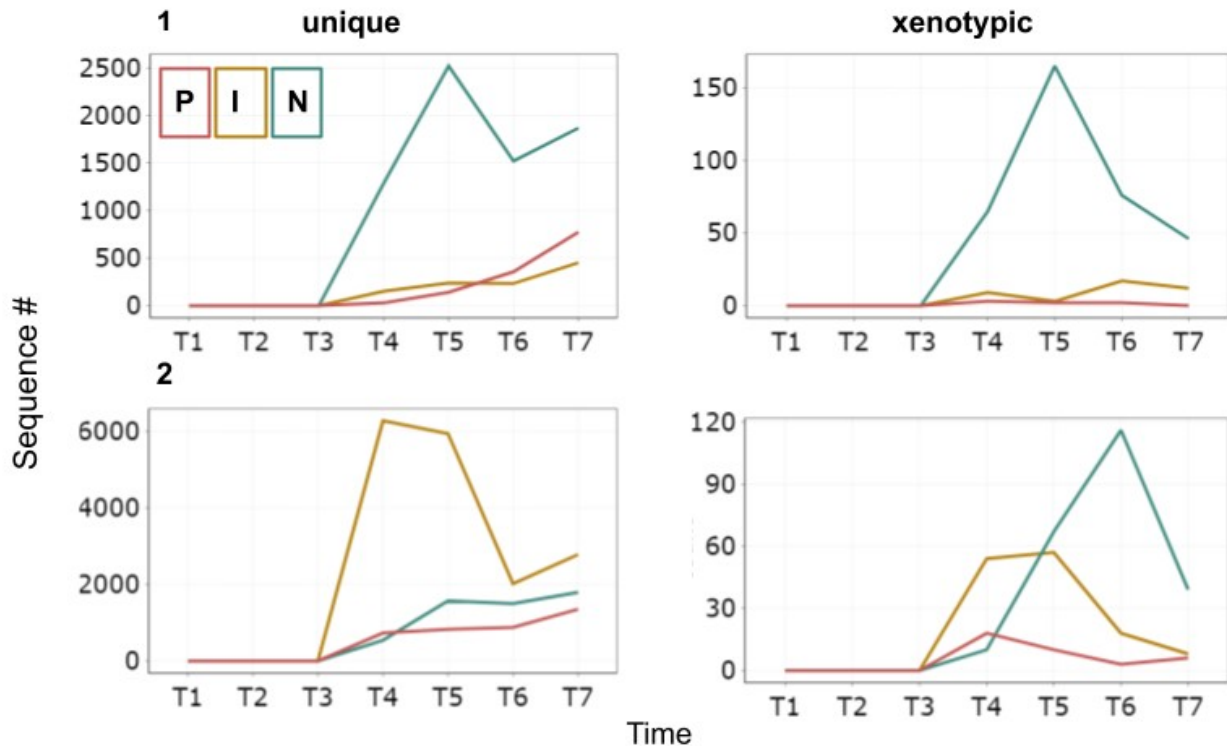
Figure 3.3. Conservative use of the Xenoseq pipeline allows the identification of foreign sequences within microbial communities. Number of unique (left) and xenotypic (right) sequences identified within the three "horizontal" communities (colours: P, I, and N are shown in red, yellow, and green respectively) across time (T1-T7) in both replicate experiments (top: 1; bottom: 2).

3.5 Short but Strong: Evidence for the Presence of Foreign Sequences in Microbial Communities through a Conservative Use of the Xenoseq Pipeline

In each community and both experiments, certain unique sequences were identified as xenotypic sequences, thereby present in one of the two other communities. Here, the quality of this evidence is scrutinized. Unique sequences were blasted against communities to identify their origin using the aforementioned options. In both replicate experiments and all three communities, most of the xenotypic sequences blasted to their sequence of origin in their entirety, as shown by the linear relationship between xenotypic sequence length and the length of the alignment. However, xenotypic sequence lengths were always found to be rather short, almost all being shorter than 1000 bp. Moreover, despite their short sizes, there were many xenotypic sequences longer than their sequence of origin (**Figure 3.4A**). Most of the alignments of the xenotypic sequences to their original community had a percentage identity between 99 and 100%. In fact, during experiment one at time five and experiment

two at time six, there were twice as many sequences identified in community N with 99-100% identity to their reference compared to those with less than 99% identity. Similar trends were as well clearly visible at times four and five in community I in the second experiment. A noticeable fraction of xenotypic sequences even had a perfect match to their sequence of origin. Although this happened for all communities, this was notably visible in community N in both experiments due to the high number of xenotypic sequences detected in this community (**Figure 3.4B**). Therefore, despite the short lengths of the xenotypic sequences identified in both experiments, these results provide strong evidence for their foreign origin.

A
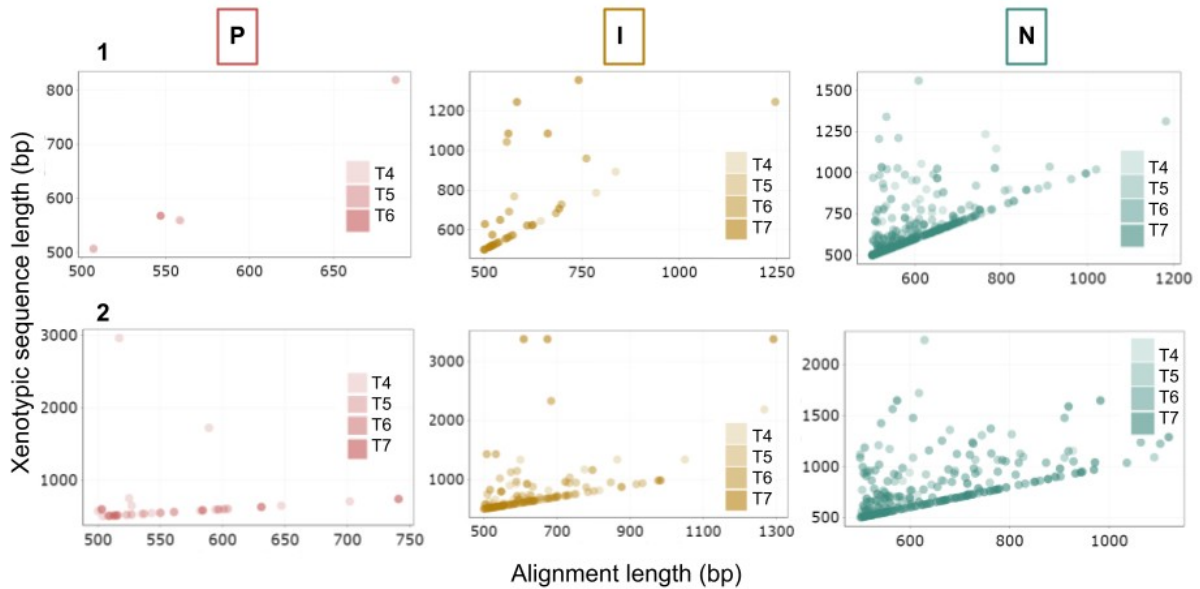
B
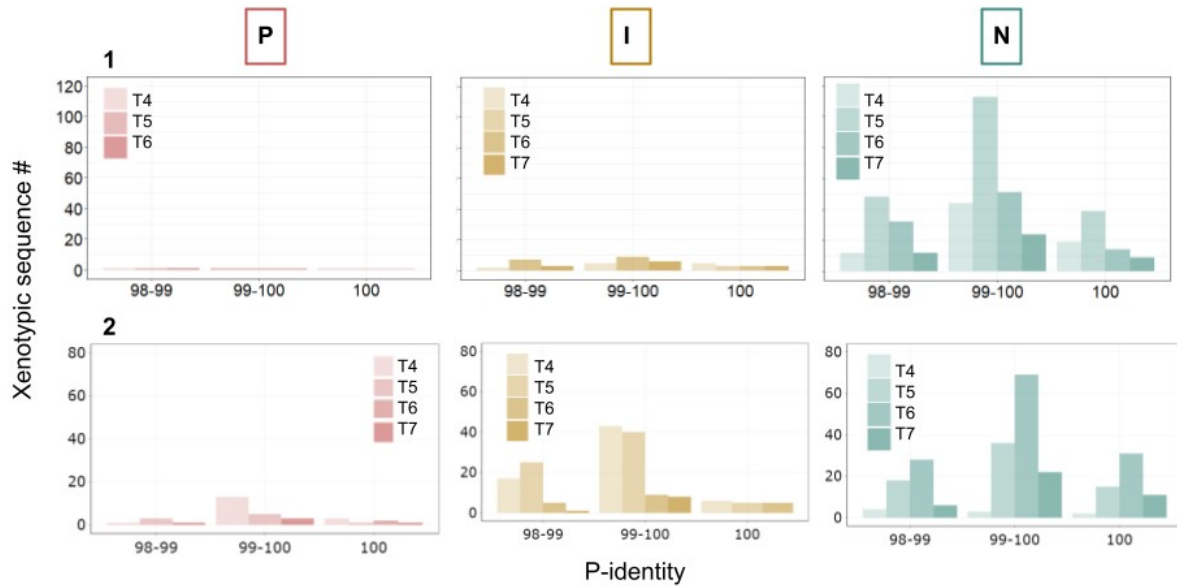
Figure 3.4. The xenotypic sequences are well aligned to their communities of origin. A BLAST of the unique sequences present within each horizontal community was performed to identify xenotypic sequences (see **Figure 3.2**, step 2). For xenotypic sequences present in the three communities (colours: P, I, and N are shown in red, yellow, and green, respectively), across time (color shades: T4-T7 with lighter shades representing earlier time points) in both replicate experiments (top:1 and bottom: 2): (A) Length of xenotypic sequences respective to the length of their alignment to their sequence of origin. (B) The number of xenotypic sequences for given ranges of percentage identity of their alignment to their sequence of origin.

3.6   Tracing the movement of xenotypic sequences between communities

3.6.1   Tracing all xenotypic sequences

As an additional verification for the movement of foreign sequences between communities, the identified sequences were traced within all communities by mapping community reads to them (**Figure 3.2**). Specifically, the breadth of coverage, defined as the percentage of base pairs of a given xenotypic sequence covered by at least one read, was plotted across time and in each community as an indicator of the confidence of the presence of the sequences.

For vertical communities, the expectation would be to have a low breadth of coverage. Indeed, the breadth of nearly all xenotypic sequences was very low (below 50%) in their equivalent vertical community. There was a single exception of a sequence that reached 100% breadth of coverage within the vertical N community in the second replicate experiment. As expected, the xenotypic sequences originally identified in a given community at a specific time point were all traced with the highest confidence in that community at that time point. The trajectories of xenotypic sequences in time in the

different communities seemed generally distinct, although the high number of sequences visualized made it difficult to distinguish every sequence trace (**Figure 3.S2**).

3.6.2   Tracing best-candidate xenotypic sequences

Because of the difficulty to visualize the presence of individual sequences, especially in communities with numerous sequences, a specific focus was made on specific sequences of interests. To illustrate the quality of the xenotypic sequences identified, the longest sequence from each community that showed 100% identity when aligned to its original sequence was plotted, thus representing the best-candidate xenotypic sequence from each community (**Figure 3.5**).

Results confirmed that these sequences were not found in their respective reference community. The minimum breadth of coverage for these xenotypic sequences was rather high, i.e. 50%, in their respective communities of origin. In the first experiment, the xenotypic sequences from both I and P communities were even found with 100% confidence in their respective communities of origin. As previously observed, the visible trace indicated the reliable presence of the xenotypic sequences in the horizontal communities where they were identified. Furthermore, the analysis also revealed evidence for the transfer of xenotypic sequences to communities that were neither the community of origin nor the destination identified by the pipeline. For instance, in the second experiment, a sequence that was originally detected in community P and believed to have originated from community I was also found to be present in horizontal community N at times one and two, with a breadth of coverage of about 50% in this community. Intriguingly, most of these xenotypic sequences were not found in the horizontal equivalent of the vertical community of origin. The xenotypic sequence detected in the I community was the sole sequence that could also be identified in the corresponding horizontal community of its vertical community of origin, here the P community (**Figure 3.5**). This phenomenon could be a result of stochasticity in the amplification of these sequences. Nonetheless, despite their short size, this analysis overall strongly suggests that xenotypic sequences do, in fact, move horizontally between communities.
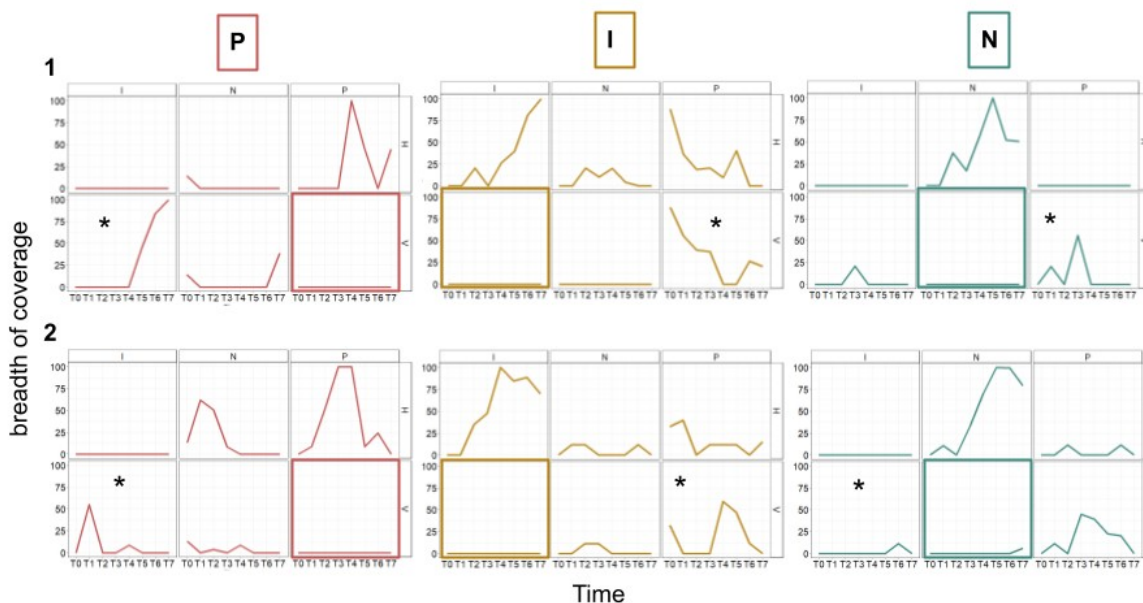
Figure 3.5. The best-candidate xenotypic sequences move across communities. Reads from the three communities (colours: P, I, and N are shown in red, yellow, and green respectively) were mapped onto the xenotypic contigs of each community (column panels: P, I, N, and row panels: H, V) (see **Figure 3.2**, step 3). The breadth of coverage was plotted across time. Colored squares represent the reference communities for the considered xenotypic sequence and the * is the community of origin identified by Xenoseq with the BLAST (see **Figure 3.2**, step 2).

3.7 Complementary Approaches for Gaining Insights into the Nature and Function of Xenotypic Sequences

The preceding analysis showed that the foreign sequences identified using the Xenoseq pipeline are very likely to have moved between communities. Nonetheless, these sequences were generally found to be rather short in length. Short sequences are less likely to be MGEs, as they only contain a few genes at best. Therefore, the xenotypic sequences identified in these experiments may have moved in another way. Distinct complementary approaches were therefore used in the attempt to gain insights into the nature and function of these horizontally transferred sequences.

3.7.1 A compositional-based approach to decipher the nature of xenotypic sequences

Compositional-based approaches have been long used by researchers to detect HGT events between genomes[75,76]. It is indeed well known that sequence composition, namely codon usage and base content, varies considerably between microbial species[77]. This entails that a horizontally transferred sequence

should exhibit greater content similarity to its genome of origin than to its destination genome. Xenoseq allows access to both the genomic contexts of the origin and destination of the xenotypic sequences. While the sequence of origin was determined by blasting unique contigs to community assemblies, the destination sequence can be found within the assembly of the sample where the xenotypic sequence was identified. Here, a GC content analysis was chosen to compare the composition of xenotypic sequences with both the sequence they originate from and the one they were transferred to.

To compare the GC content of the xenotypic sequences with their origin and destination sequences, only the xenotypic sequences shorter than their respective origin or destination sequences were retained. Overall, 45% and 65% of sequences were retained for the comparison with the sequences of origin, and 47% and 58% for the comparison with the sequences of destination in the two experiments, respectively (**Table 3.S2**).

3.7.1.1 *Comparison of GC content of xenotypic sequences with their origin and destination sequences*

First, the GC content of xenotypic sequences was compared to that of their sequence of origin. To do this, the absolute difference in GC content between each xenotypic sequence and its corresponding origin sequence was calculated (Δ_{origin}, **Figure 3.6a**) and results were represented as a distribution (**Figure 3.7A**). For all communities in both experiments, the Δ_{origin} distribution showed a significant deviation from zero mean, indicating that, on average, the GC content of xenotypic sequences differs from their origin sequences. Moreover, while the majority of the considered sequences showed a small Δ_{origin} value (about 1-2%), several sequences had a remarkably high Δ_{origin}, with a maximum difference of up to to 18% observed in the N community of the second experiment (**Figure 3.7A**).

The GC content of the xenotypic sequences were then compared to that of their sequence of destination, i.e. the genomic context in which the sequence appears to have inserted. Similarly to the previous comparison, the absolute difference in GC content between each xenotypic sequence and its corresponding destination sequence was calculated ($\Delta_{destination}$, **Figure 3.6b**) and results were represented as a distribution (**Figure 3.7B**). This analysis did not include any xenotypic sequences in community P of the first experiment as none were smaller than their destination sequence. The $\Delta_{destination}$ distribution displayed a statistically significant deviation from a mean of zero for all considered communities in both experiments, suggesting that the GC content of xenotypic sequences on average differed from their destination sequences. The $\Delta_{destination}$ distributions were similar to the Δ_{origin} distributions as they both

centred around 1-2% GC difference and contained sequences with greatly different GC content from their genomic context (**Figure 3.7B**).

Altogether, the results of this investigation revealed that the xenotypic sequences, not only expectedly differ in GC content from their destination sequence, but also from their origin (**Figure 3.7A** and **B**). This observation implies that the original genomic context of these xenotypic sequences may not be their true origin, but rather the result of a recent location change, which indicates a high frequency of movement. Therefore, these findings support the notion that these xenotypic sequences are indeed part of MGEs. However, it should be noted that these results do not disprove the possibility that the origin and destination sequences have identical genomic contexts. Indeed, in such a scenario, the opposite conclusion would be drawn, namely that the xenotypic sequences have not experienced any movement. The subsequent analysis will aim to evaluate the likelihood of both possibilities.

3.7.1.2 *Comparing the degrees of differences in GC content of xenotypic sequences from their sequence of origin and destination*

To determine whether the genomic context of the origin and destination of given xenotypic sequences are different, the absolute difference between Δ_{origin} and $\Delta_{destination}$ ($\Delta(\Delta_{origin}, \Delta_{destination})$, **Figure 3.6c**) was calculated and plotted as a distribution (**Figure 3.7C**). For this analysis, only xenotypic sequences shorter in length than both their sequence of destination and their sequence of origin were considered. For all communities in both experiments, the mean of $\Delta(\Delta_{origin}, \Delta_{destination})$ was significantly different from zero (**Figure 3.7C**). Therefore, in addition to exhibiting a distinct GC profile, the analysis of $\Delta(\Delta_{origin}, \Delta_{destination})$ provides evidence that the flanking regions surrounding horizontally transferred sequences differ before and after the transmission event. This observation strongly suggests that the xenotypic sequences have likely integrated into new genomes or have been transferred between distinct replicons, such as chromosomes and plasmids.
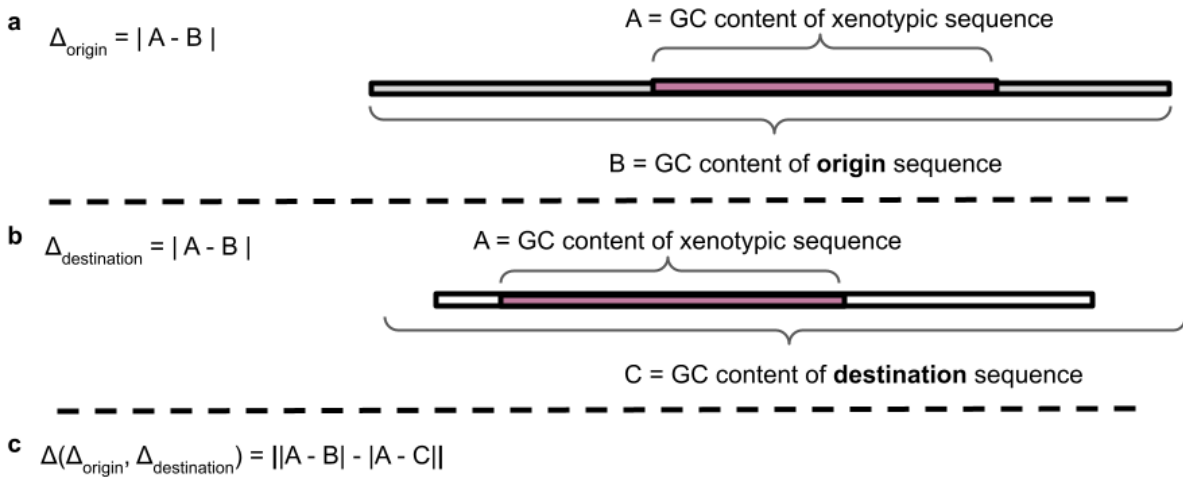
Figure 3.6. Analysis of GC content of xenotypic sequences and their sequences of origin and destination (a) Δ_{origin}: Difference in GC content between a given xenotypic sequence (pink) and their sequence of origin (grey). (b) $\Delta_{destination}$: Difference in GC content between a given xenotypic sequence (pink) and their sequence of origin (white). (c) $\Delta(\Delta_{origin}, \Delta_{destination}$: Difference in GC content between Δ_{origin} and $\Delta_{destination}$ for a given xenotypic sequence.
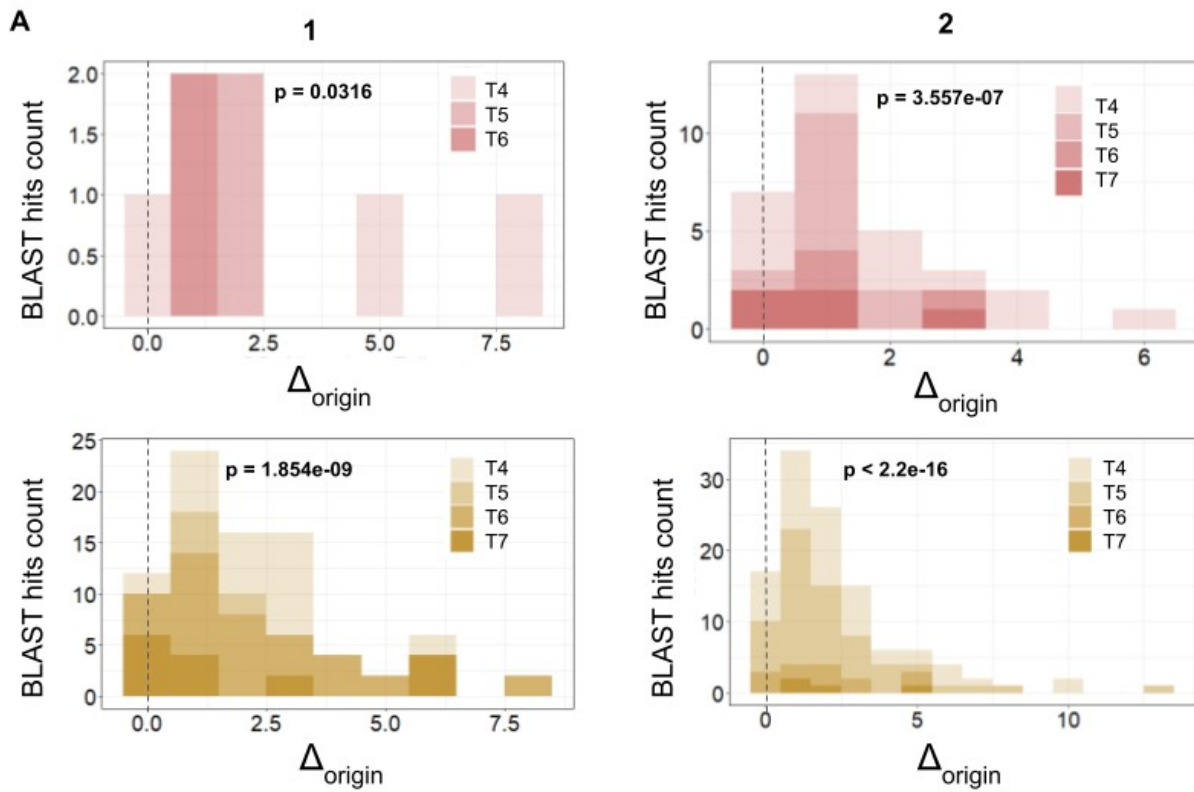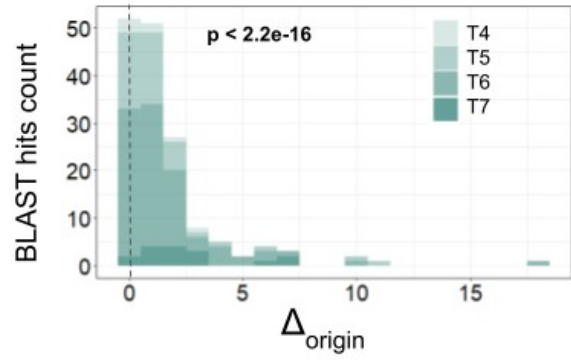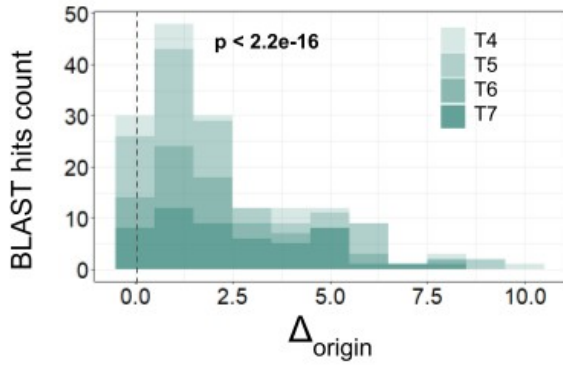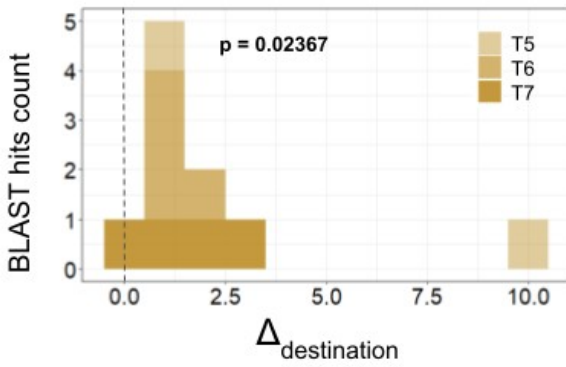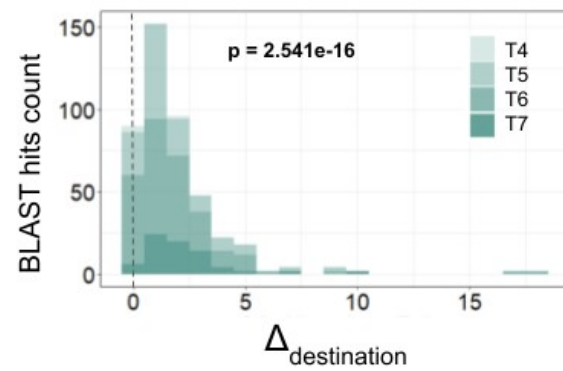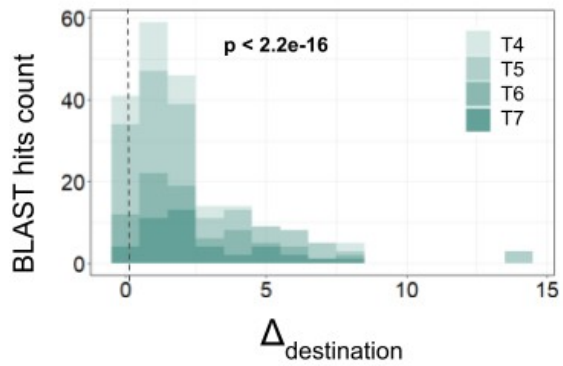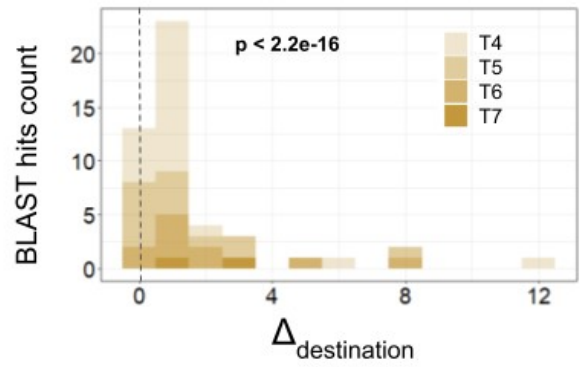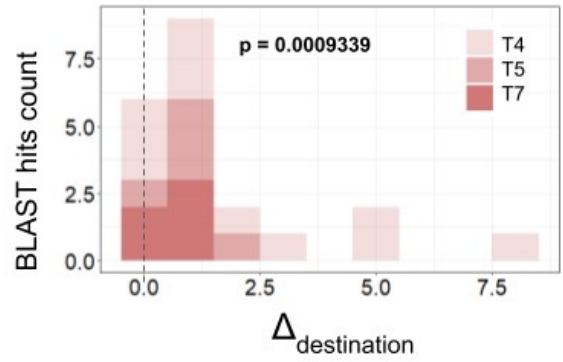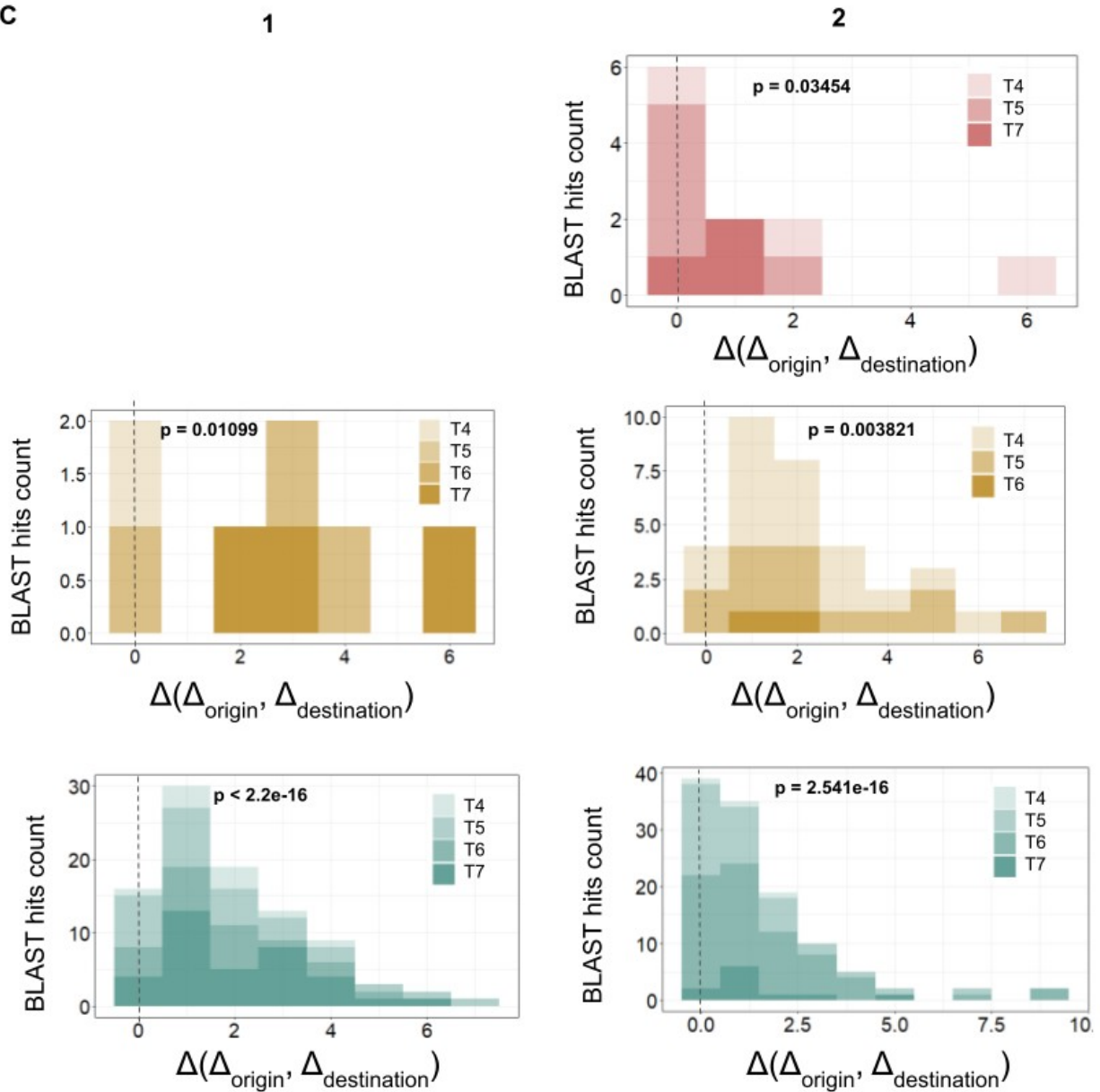
Figure 3.7. The xenotypic sequences differ in their GC content from both their origin and destination sequences. For both experiments (left column:1 and right column:2) and the three communities (colours: P, I, and N are shown in red, yellow, and green respectively) and time points (color shades: T4-T7 with lighter shades representing earlier time points) where xenotypic sequences were identified: Distribution of differences in GC content between xenotypic sequences and their sequences of origin (Δ_{origin}) (A) or destination (Δ_{origin}) (B), or differences in GC content between Δ_{origin} and Δ_{origin} ($\Delta(\Delta_{origin}, \Delta_{destination})$). One-tailed Student t-tests indicate statistical significance ($p < 0.05$).

3.7.2   A search for MGEs among xenotypic sequences

MGEs serve as the primary carriers of HGT, predominantly in the form of plasmids and bacteriophages [26].

In order for the Xenoseq pipeline to detect MGEs, they must be present extracellularly in the donor

community, allowing for subsequent amplification either within or outside recipient cells. Plasmids, commonly transferred through conjugation and relying on direct cell-to-cell contact, are not anticipated to be easily detectable by Xenoseq[78]. Conversely, bacteriophages, viruses that specifically infect bacteria and have the ability to package and transduce genomic DNA between cells in their free form, are anticipated to be abundant in these communities, as demonstrated by previous research[37,79]. To identify the specific MGEs involved in the transfer of the xenotypic sequences, prediction analyses were conducted.

3.7.2.1   *Searching for plasmids among xenotypic sequences*

Plasmids were sought within the xenotypic sequences by using the plasmid classifier called PlasClass, a tool trained on plasmid databases. PlasClass is a performant tool for short sequences, as it was tested with good results on output contigs from metagenomics data[67]. It is therefore a relevant and likely reliable tool to identify plasmids among the overall short xenotypic sequences found in this study. A sequence was classified as a plasmid when the probability of belonging to a plasmid was above 0.5. Otherwise, it was considered chromosomal.

This analysis predicted plasmid sequences among xenotypic sequences detected in all three communities. Strikingly, a large proportion of the xenotypic sequences detected in both experiments were predicted to be plasmids in both experiments, sometimes outpacing the chromosomal fraction. For the first experiment, the percentage of sequences classified as plasmids in communities P, I, and N were approximately 43% (4/7), 63% (26/41), and 41% (146/352), respectively. In the second experiment, around 46% (17/37), 32% (44/137), and 67% (156/232) of the sequences were annotated as plasmids for these respective communities. Strikingly, at times five and six in the second experiment, the ratio of xenotypic sequences predicted to be plasmids in community N was more than twice as high as those categorized as chromosomal, with 2.5X (48/19) and 2.5X (83/33), respectively. In the first experiment, the largest proportion of sequences classified as plasmids were found at the same time points in this community, representing 41% (60/146) and 26% (39/146) of all sequences annotated as plasmids across all time points. In contrast, in the I community of the first experiment, slightly more sequences classified as plasmids were annotated at times six and seven, with around 36% (10/28) of sequences predicted as plasmids at each of these time points, respectively. In the second experiment, more plasmids were predicted at times four and five in this community, constituting approximately 39% (17/44) and 34% (15/44) of the plasmid sequences, respectively. Although the number of sequences in community P was

small, plasmid sequences were predicted to be most abundant at time four in the second experiment, accounting for around 65% (11/17) of the plasmid sequences. These findings suggest that plasmids may be an important part of xenotypic sequences and that their presence may be highly dynamic (**Figure 3.8A**). This result is surprising, given that plasmids are typically thought to require cell-to-cell contact to transfer horizontally and are thus not expected to be found in the MGEs filtrates, as opposed to free MGEs.

3.7.2.2 Searching for viruses among xenotypic sequences

The identification of prophage sequences is reliable for input sequences with a minimum length of 20 000 base pairs, and none of the xenotypic sequences identified reached that length. Instead, the virus identification tool VirSorter2 was used to get an overview of the viral content of the xenotypic sequences[66]. This tool provides accurate predictions about the identity of sequences as viral and categorizes them as either double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) phage sequences.

A relatively small number of xenotypic sequences identified in the I and N communities were classified as viral. This may reflect the lower performance of this tool on short sequences[66]. In community I, a maximum of zero to four dsDNA sequences and one to two ssDNA sequences were predicted in the two experiments, respectively. Community N had even fewer viral sequences with only one dsDNA phage and one ssDNA phage in the first experiment, and zero dsDNA phages and one ssDNA phage in the second experiment. These results showed that both ssDNA and dsDNA phages were likely present, but dsDNA phages were probably slightly more abundant overall (**Figure 3.8B**). Consistent with this outcome, while both phage types are thought to participate in HGT, dsDNA phages are known to be more common vectors. One crucial element that contributes to this trend is the larger size of dsDNA phage genomes, which provides greater flexibility for the integration and spread of potentially beneficial novel genes[80].

3.7.2.3 Searching for small MGEs among xenotypic sequences

To verify the existence of smaller, potentially unknown, MGEs among the xenotypic sequences, genes were annotated using Prokka[69] (**Table 3.S3**). Notably, among the 244 and 350 predicted Open Reading Frames (ORFs), two genes were functionally annotated as putative insertion sequences (ISs) from IS110 and IS3 families. Both genes originated from xenotypic sequences found in the N community of the first

replicate experiment. ISs are the smallest MGEs, and are very dynamic within and across genomes, therefore likely to have moved horizontally during this experiment[81]. The subsequent results provide further insights into the gene content of the annotated xenotypic sequences.
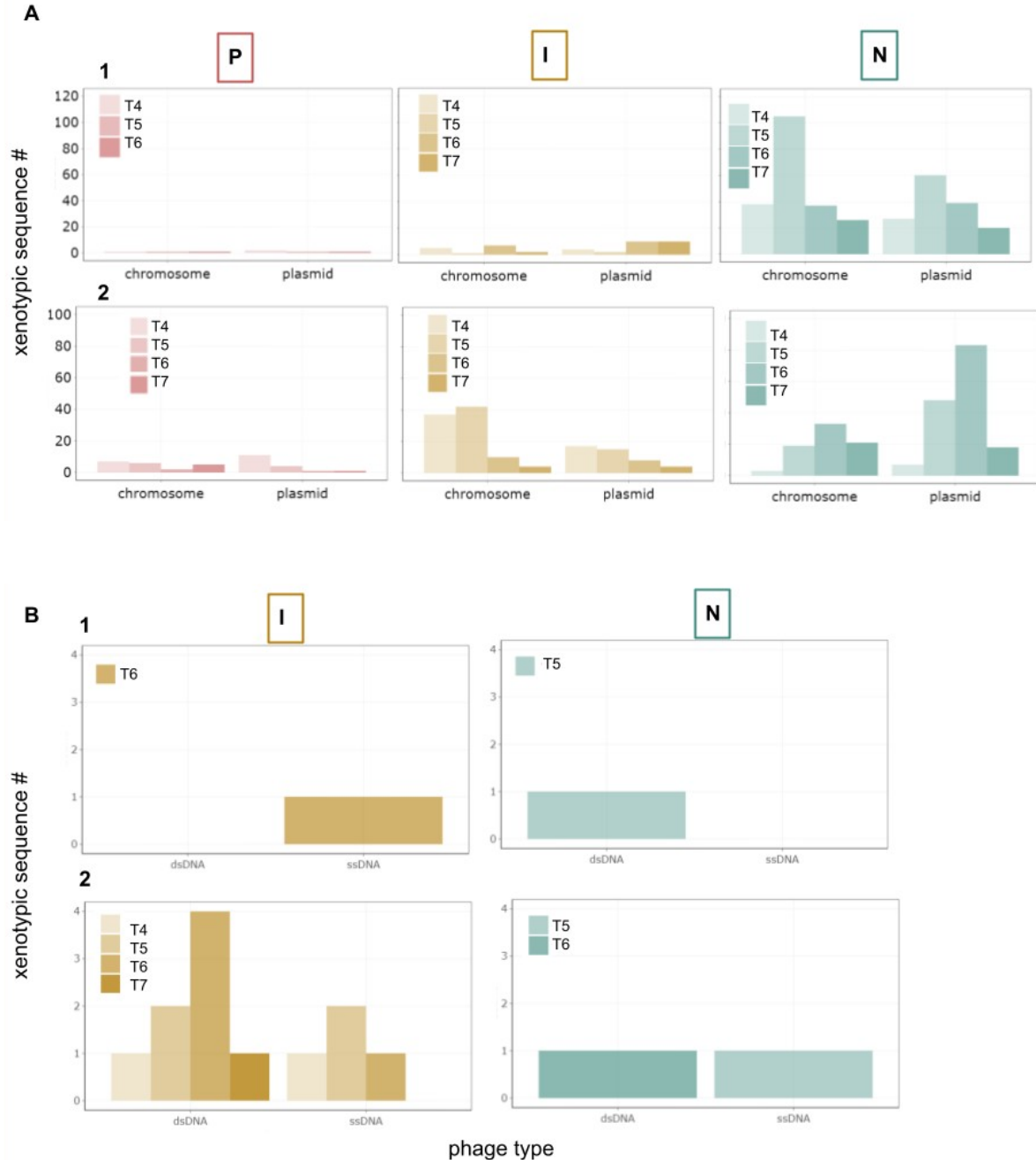
Figure 3.8. Small pieces of MGEs are found within xenotypic sequences. For each community (colours: P, I, and N are shown in red, yellow, and green respectively) and each time point (T4-T7) where xenotypic sequences were identified: (A) Number of xenotypic sequences classified as plasmid or chromosome using PlasClass. (B) Number of xenotypic sequences classified as viral dsDNA or ssDNA phages using VirSorter2.

3.7.3 Functional analysis of horizontally transferred genes comprised within xenotypic sequences

For a long time, it has been understood that certain types of genes are more commonly transferred horizontally. These are often called "operational genes" and tend to be more susceptible to HGT than "informational genes"[82]. Researchers have suggested that the likelihood of a gene being horizontally transferred is linked to its connectivity, in other words, the number of components it interacts with[83]. Since then, studies have supported this idea, as well as specifically shown that most functional categories follow this pattern[84–86]. Given the previous evidence of xenotypic sequence movement, this study aimed to analyze the functional categories of horizontally transferred genes identified in the xenotypic sequences to determine if this bias holds in the present investigation.

To test the aforementioned hypothesis, an investigation of the gene content of all xenotypic sequences was conducted. To begin with, the xenotypic sequences were sorted based on the community in which they were discovered. For both experiments, all the identified sequences, consisting of 400 (P: 7, I: 41, N: 352) and 406 (P: 37, I: 137, N: 232), were included in the analysis. Next, the genes identified were subjected to annotation using Prokka[69]. As a result, 36 out of 244 genes (15%) and 48 out of 350 predicted ORFs (14%) were successfully annotated with a potential function for both experiments, respectively. All remaining ORFs were classified as unnamed "CDS". Once duplicate gene names were eliminated, 32 and 39 distinct genes were found in the respective experiments and included in the analysis (**Table 3.S3**). Each gene was then functionally annotated using the KEGG Orthology (KO) database, specifically with the three annotation levels of the top six categories from pathway maps. Findings indicate that about 56% (18/32) and 49% (19/39) of the genes were attributed to at least one of these six top categories for both experiments respectively, meaning that the majority of the genes remained unannotated. All genes were associated with xenotypic sequences from either community N (94%, 17/18, and 74% 14/19) or I (6%, 1/18, and 26%, 5/19), and none were attributed to community P (**Figure 3.9 and Table 3.S4**).

3.7.3.1 First annotation level

The majority of functionally annotated genes in both experiments were related to metabolism and environmental information processing categories. In the first experiment, 36% (8/22) and 37% (10/22) of the annotated genes were associated with these categories, respectively. In the second experiment, 37% (10/27) of the annotated genes were related to both metabolism and environmental information

processing categories. These genes were present in both the N and I communities, except for the metabolism genes in the first experiment, which were only found in the N community. Additionally, only one or two genes associated with these categories were identified in xenotypic sequences detected in the community I. This is likely due to the limited number of annotated genes in this community. The category of cellular processes included a small percentage of genes, specifically 5% (1/22) and 1% (4/27) in the first and second experiments, respectively. In both experiments, genetic information processing and human diseases categories had even fewer genes, accounting for only 5% (1/22) and 7% (2/27), and 9% (2/22) and 4% (1/27) of the annotated genes, respectively. The only category absent from this gene dataset was the organismal system one, corroborating the absence of contamination by genes from eukaryotic organisms (**Figure 3.9A**). Therefore, strikingly consistent with the literature, the majority of genes found within xenotypic sequences belonged to "operational" categories such as metabolism, and environmental information processing, while the "informational" category of genetic information processing contained the fewest number of genes[82].

Previous studies have shown that metabolism genes are often found in higher proportions within accessory genomes[87,88]. Additionally, environmental information processing genes have been identified as the second most abundant function within accessory genomes, following metabolism[87]. Enrichment refers to a higher prevalence of these genes in the accessory genome compared to the remainder of the genomes. In an attempt to confirm whether the xenotypic sequences are truly enriched in these genes, the proportion of genes belonging to each category was also examined in a given entire community, in which the xenotypic sequences were detected. In fact, the observed results may have been biased if the KEGG database includes a larger number of genes in these categories. Community I at time four was selected as an example. The most enriched category was environmental information processing and metabolism, which is further divided into multiple subcategories in the figure. This indicates a potential bias in the KEGG database towards these categories. However, the analysis indicated that only a very small fraction of the ORFs were functionally annotated, specifically 6.5% (**Figure 3.S3**). This is even lower than the aforementioned percentages of annotated ORFs for the xenotypic sequences. Thus, due to the limited number of functionally annotated ORFs, it is not possible to draw a definitive conclusion about the enrichment of xenotypic sequences in these categories. Despite these findings, the prevalence of metabolic and environmental processing genes in xenotypic sequences is still noteworthy.

To gain a deeper understanding of the gene composition of the two main categories, namely metabolism and environmental information processing, an analysis of the second and third levels of classification was conducted for each of these categories.

Based on data collected from both experiments, the most prevalent metabolic functions found were amino acid (70%, 7/10, and 30%, 3/10) and carbohydrate metabolism (0%, 0/10, and 30%, 3/10). Xenobiotic degradation (0%, 0/10, and 20%, 2/10) and nucleotide metabolism (0%, 0/10, and 20%, 2/10) were present to a lesser extent (**Figure 3.9B**). The different types of amino acid metabolism observed in the xenotypic sequences, such as arginine and proline, lysine, tyrosine, phenylalanine, and tryptophan, indicate that these sequences may have a high potential to facilitate niche adaptation. Similarly, the category of carbohydrate metabolism exhibited a wide range of functions, such as fructose and mannose metabolism, pentose and glucuronate metabolism, and starch and sucrose metabolism (**Figure 3.9C**).

A considerable proportion of the genes allocated to the environmental signalling processing category were involved in transport, i.e. 40% (3/8) and 70% (7/10), in both experiments respectively (**Figure 3.9B**). All of these genes encode for ABC transporters (**Figure 3.9C**). ABC transporters are a broad family of ATP-binding transmembrane proteins at the interface between the cell and the environment. They can have various functions, including nutrient uptake, secretion of waste products, energy generation, and cell signalling[89]. All of these functions are considered competitive traits for environmental adaptations and are commonly believed to be primarily horizontally transferred[18]. It is noteworthy that among the identified genes, two were also found to be quorum-sensing genes, with one of them, named *oppF*, also implicated in antibiotic resistance (**Table 3.S4**). Indeed, these two functions are related to competitive interactions, and it is well-established that genes responsible for antibiotic resistance are frequently disseminated through horizontal transfer by MGEs[90]. Additionally, *oppF* has been shown the be involved in microbial competence[91]. Several signal transduction genes were also found in the environmental signaling processing category, and all were attributed to two-component systems (**Figure 3.9B and C**). Bacteria rely primarily on two-component signal transduction systems to detect and react to environmental signals[92]. Novel histidine kinases, often together with their associated response regulators, can be acquired through HGT by certain species. This provides the bacteria with novel functions and consequently widens their ability to adapt to various niches[93].
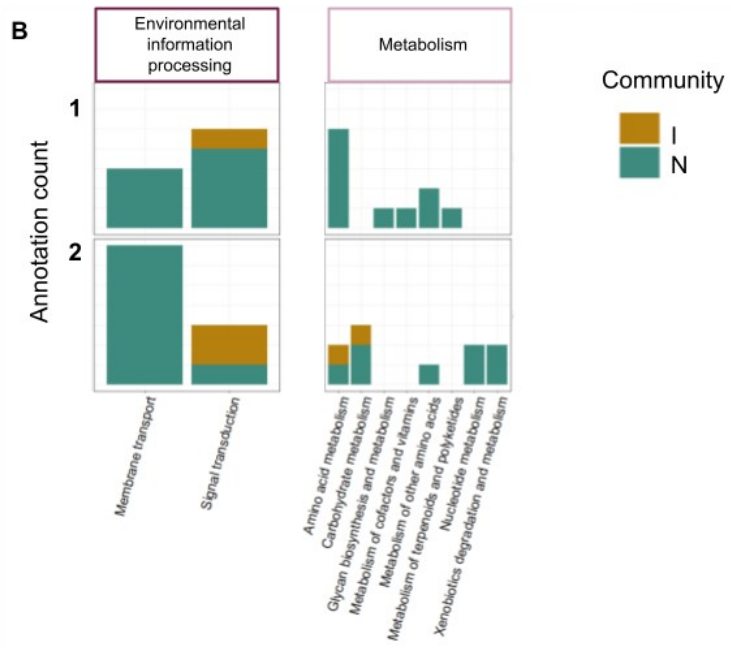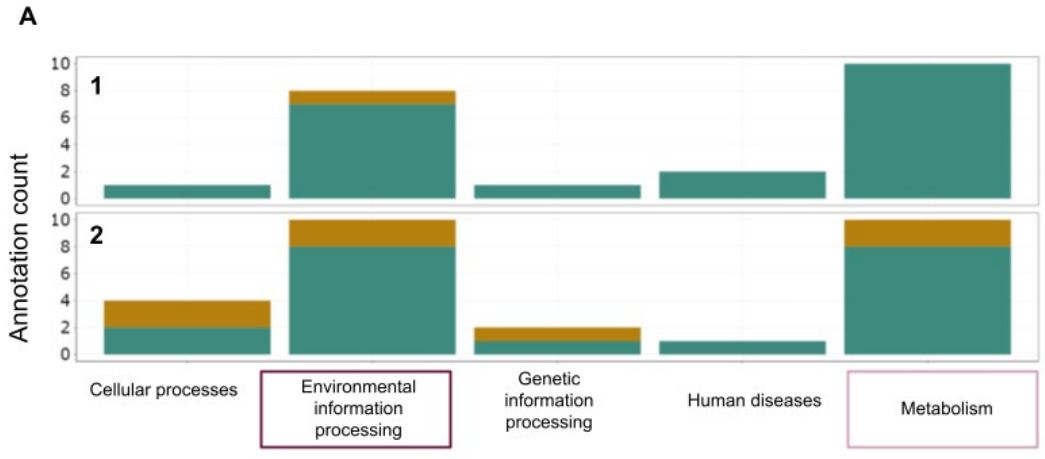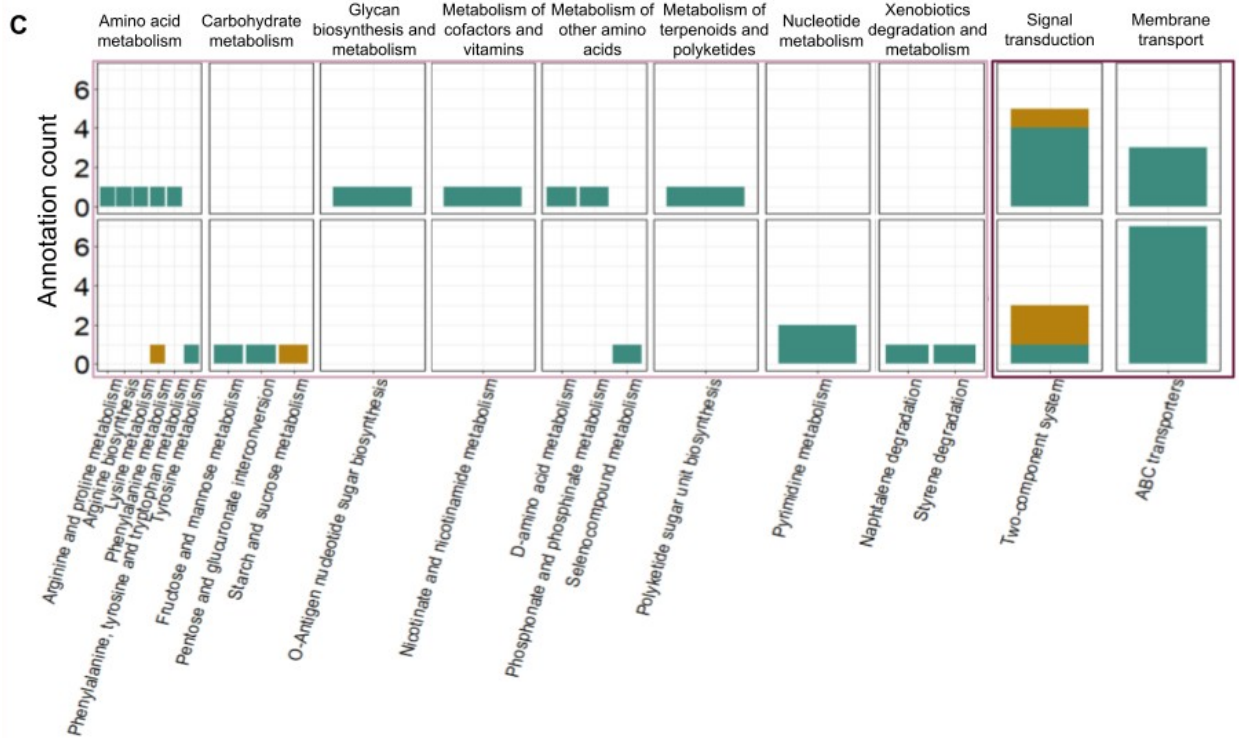
Figure 3.9. Xenotypic sequences encompass protein-coding genes involved in metabolism and environmental processing signaling. Xenotypic sequences were first annotated using Prokka. For each community in which the xenotypic sequences were identified, each gene with a provided name, i.e. all that were not categorized as "hypothetical", was associated with a KEGG orthology (KO), and the three corresponding annotation levels of the top six categories from pathway maps. Bar plots of the number of genes found in xenotypic sequences identified in the two replicate experiments (experiment replicates 1 and 2 are shown in top and bottom panels, respectively) and the different horizontal communities (colours: I, and N are shown in yellow, and green respectively) in each category, for (A) First annotation level. (B) Second annotation level. (C) Third annotation level for environmental information processing (light pink) and metabolism (dark pink).

3.8   Discussion

3.8.1   Robustness of the Xenoseq pipeline in detecting foreign sequences

In this chapter, the efficiency of the Xenoseq pipeline to detect horizontally acquired sequences was tested for the first time in the frame of a short-term evolution experiment (**Figures 3.1** and **3.2**). The pipeline's stringent criteria allowed for the detection of authentic foreign sequences, distinguishing them from sequences that remained undetected in the reference samples because they fell below the detection threshold (**Figure 3.3**). The quality of alignment to their communities of origin provides good evidence for the foreign origin of the xenotypic sequences identified in both experiments, despite their short lengths (**Figure 3.4**). The movement of foreign sequences between communities was successfully

verified by tracing the identified sequences in all communities through the alignment of community reads to them (**Figures 3.S1** and **3.5**).

3.8.2 Consistent amplification of foreign sequences in microbial communities

The findings of this study revealed a remarkable level of predictability in the occurrence of undetectable sequences and the acquisition of foreign sequences by each community (**Figure 3.3**). This is particularly surprising considering the high complexity of the microbial communities involved. One would expect a high degree of stochasticity in the movement of foreign sequences, making these findings particularly intriguing. The early emergence of these sequences repeatedly observed in this study aligns with the findings of Quistad *et al.*[37], who identified unique sequences just two weeks after separating the communities. In this study, it was nonetheless not confirmed which of these sequences originated from another community. However, it was observed that these sequences were shared among multiple horizontal communities and contained a significant number of phage-like elements[37].

Evolutionary processes are primarily driven by stochastic factors, such as random mutations and genetic drift. However, deterministic effects can also play a role, primarily stemming from either natural selection or mutational biases, or both[94]. As Darwinian entities, MGEs are subject to these processes as well[26]. For example, a recent study found that a phage-plasmid repeatedly infected its host after the same number of generations in multiple independent biological replicates, highlighting the consistency of the timing of this phenomenon. The reason for this was attributed to frequent mutations that resulted in loss-of-function of the phage repressor, which is responsible for regulating prophage induction[95].

Ecological factors could also account for the observed deterministic increase in sequences. Predator-prey dynamics is characterized by oscillating patterns in the abundance of predators and prey [96]. MGEs such as bacteriophages can be regarded as predators of bacteria in this study. As a result, bursts in communities often lead to the emergence of MGEs[97]. In line with this, a sharp increase in the abundance of xenotypic sequences was observed in nearly all communities, followed by a subsequent decline. However, the two-week interval between transfers provides ample time for various events to occur that could influence sequence dynamics. Therefore, more precise measurements could provide a more in-depth understanding of the behaviour of these sequences over time. Alternatively, longer evolution experiments could reveal longer cycles in the abundance of these sequences.

3.8.3 Mechanisms of transfer of xenotypic sequences

MGEs are thought to be key drivers of HGT. The analysis of the disparities in GC content between the xenotypic sequences and their source and target sequences provided preliminary evidence for the mobile nature of the xenotypic sequences (**Figure 3.7**). The utilization of Xenoseq in this study thereby highlighted its ability to provide valuable insights into the direction of gene flow and the genetic similarity necessary for successful gene transfer. Additionally, a considerable number of the xenotypic sequences harboured predicted MGEs, with sequences annotated as plasmids being by far the most prevalent (**Figure 3.8**). However, plasmids are commonly believed to require cell-to-cell contact for transfer, making it seemingly impossible for these MGEs to be transferred through community filtration. Therefore, it is plausible that alternative mechanisms of transfer are utilized by plasmids. For example, nanobacteria have been shown to pass through selection filters and proliferate within microbial communities, and could potentially carry plasmids with them[62]. Additionally, although much is still unknown about this process, outer membrane vesicles (OMVs) are a recently discovered means of HGT that has been shown to transport DNA, including plasmids, between bacteria[98]. The recent discovery of phage-plasmids, with plasmid elements packaged inside phage capsids, grants an alternative explanation for transmission mechanisms without direct cell-to-cell contact[99].

Not all xenotypic sequences exhibited characteristics typical of the most common MGEs, and it is possible that some were mistakenly attributed to MGEs due to databases biases or their small sizes. On the one hand, the results of this experiment may imply the movement of previously undescribed MGEs, which are not part of existing databases. It could also suggest that MGEs-independent mechanisms of HGT may play a crucial role in the dissemination of genes across communities. This could happen for example through the uptake of environmental DNA that was released from lysing bacterial cells, which can lead to a higher frequency of HGT than previously assumed[100]. Even without MGEs, theory has shown that DNA uptake has a significant impact on microbial genomes, microbial diversity, and community function[101–103]. Finally, there could be other mechanisms that are currently unknown and might be contributing to the observed HGT. Therefore, Xenoseq holds the potential to illuminate novel MGE-driven mechanisms of HGT.

3.8.4 Rethinking the prevalence of HGT: movement of small DNA sequences

Findings indicate the movement of much smaller pieces of DNA than previously analysed. Possibly, the short sequences identified in the study could be the result of incompletely assembled sequences, and

the multiple detected sequences may be less numerous than reported. However, the observed seemingly stochastic trajectories of the xenotypic sequences suggest that these elements may be separated individuals and not simply the result of incomplete assembly (**Figure 3.S1**). The horizontal transmission of such short sequences could be easily mistaken for a few point mutations, leading to the possibility that the prevalence of HGT in nature may be even more underestimated than currently thought. This implies that certain evolutionary consequences, which were previously attributed to point mutations, may have actually resulted from HGT.

3.8.5   Operational genes in xenotypic sequences: implications for community function

The predominance of "operational" genes, such as those involved in metabolism and environmental information processing, rather than "informational" genes related to genetic information processing, was consistent with the literature on the functional bias of horizontally transferred genes[82] (**Figure 3.9**). Metabolic and environmental processing information genes were the most abundant among xenotypic sequences. Because these categories were also significantly more present when considering the entire community, it is difficult to draw conclusions about the enrichment of the xenotypic sequences in these genes compared to the whole community due to the low number of functionally annotated genes. Regardless, the presence and diversity of functional genes belonging to these categories within the xenotypic sequences suggest that they may play an important role in ecosystem functioning. Consistent with this, previous research demonstrated the role of HGT in shaping microbial metabolic dependencies and the adaptiveness of metabolic genes[104–106]. Goyal *et al*.[105] have thereby suggested that accessory genomes may carry adaptive metabolic traits, highlighting the potential contribution of the diverse metabolic genes identified in our study to the functioning of the community[105]. The next chapter aims to provide further evidence for the role of HGT in community function.

3.8.6   Conclusion

Our understanding of the eco-evolutionary dynamics that unfold within natural communities remains limited. HGT is a pivotal process that profoundly influences microbial ecology and evolution. This study serves as a proof-of-concept, demonstrating the feasibility of real-time detection of HGT in complex microbial communities through a straightforward experimental design, metagenomics, and a simple bioinformatic pipeline.

4 Investigating the Impact of Horizontal Gene Transfer on Microbial Community Function in the Gut of *Caenorhabditis elegans*

4.1 Introduction

From the results discussed in **Chapter 4**, it is apparent that, despite significant technical challenges, the proposed experimental design coupled with a simple bioinformatic pipeline is sufficient to detect HGT events in complex free-living microbial communities. Whether or not HGT influences the role of microorganisms in ecosystem functioning remains to be tested.

HGT is a highly relevant process affecting the evolution of microorganisms and microorganisms are known to impose a great toll on the eukaryotic organisms they interact with [107–109]. Due to the various niches, immune system pressures, and challenging conditions present in this environment, the eukaryotic gut is believed to serve as a hotspot for HGT [43,44]. Various studies have confirmed the significant role of HGT in the gut, with particular emphasis on its importance in the evolution of microorganisms [24,51,51,52,110,111]. However, despite its apparent role in health and disease, similar to free-living communities, real-time detection of HGT in ecologically relevant gut communities lags far behind.

This chapter aims to understand how HGT affects ecosystem functioning in the context of host-associated microbial communities. To do so, the pilot and main versions of the aforementioned experimental design are utilized to study the impact of MGE acquisition on the fitness of the host nematode *C. elegans* (**Chapter 4** section **4.2**, and **Figure 1.1**). Additionally, this chapter explores different methods for detecting HGT in the nematode gut microbiome (**Chapter 4** sections **4.3**, **4.4**, and **4.5**). Finally, the studied approach investigated in **Chapter 4** section **4.5**, is employed to evaluate the ability of crucial members of the microbiome to acquire horizontally transferred elements in an evolutionary experiment (**Chapter 4** section **4.6**, and **Figure 1.1**).

4.2 Effect of MGEs Acquisition on the Functioning of Host-Associated Microbial Communities

4.2.1 Assessing microbial community function through nematode fitness

The bacterivore nematode *C. elegans* presents a promising opportunity to contribute to the understanding of the role of HGT within the gut microbiome. In fact, certain microorganisms are known

to remain intact and persist within *C. elegans* gut while affecting the nematode life history traits[54,55,112]. Building on previous research, it will thereafter be assumed that the fitness of the nematode is closely tied to the colonization potential of its gut microbiome components, and will thus be contemplated as community function[113,114].

4.2.2 Tracking changes in nematode fitness during microbial community evolution experiments

To evaluate the impact of HGT on nematode fitness, microbial communities were evolved in the presence of *C. elegans*, and MGEs were filtered and pooled from either all parallel communities (horizontal regime) to enhance HGT or only from the native community (vertical regime) at times of transfer (**Figure 1.1**). Hence, any observed fluctuations in nematode fitness are hypothesized to be related to horizontally-acquired changes by its microbiome components. Changes in community function upon the addition of MGEs from other communities were investigated by measuring nematode population size as a proxy for fitness during the evolution experiments. To do so, nematodes were counted every two weeks at times of transfer. For this analysis, results from both the previously described pilot and main evolution experiments were considered (**Figure 1.1**).

4.2.2.1 Nematode fitness in the pilot evolution experiment

In the pilot evolution experiment, a sharp and significant decrease in nematode population size was observed for all three communities at times of third transfer, i.e. after six weeks, when communities were fuelled with pooled MGEs compared to when they only received MGEs originating from their community. The extent of the decrease varied among the different communities. Specifically, the nematodes in the P and N communities experienced the smallest decrease, with their population dropping from an average of 7000 to 2548, and 3760 to 1504, respectively. In contrast, those in the I community faced the greatest decrease, from 10152 to 6600. Additionally, a drop in nematode count from 7000 to 2548 was seen in community P under horizontal treatment between transfers two and three, which was significant and comparable to the reduction observed between the horizontal and vertical treatments of this community at transfer three. Noticeably, the nematode population size in the I community was significantly larger, approximately twice as large, compared to the P and N communities (**Figure 4.1A**).

Across both replicates of the main evolution experiment, it was consistently observed that nematode fitness was considerably higher in the horizontal treatment than in the vertical treatment when nematodes were grown in the I community. In both replicate experiments, this increase was extensive, i.e. reaching more than two-fold change at transfer four, and persistent, starting from transfer two and lasting until at least transfer five. For the first replicate, the difference between the horizontal and vertical treatments was statistically significant for transfers four and five, while for the second replicate, this significance was observed for transfers two and three. This effect was also seen at transfer seven in the second experiment. In contrast to the nematodes grown in the I community, the ones in the two other communities did not exhibit any obvious or consistent pattern in the disparity of fitness between the horizontal and vertical treatments. In fact, at most transfer times, no differences were observed. In community P, there was a small decrease from 7560 to 4140 in nematode number at the first transfer in the first experiment, but an increase from 900 to 6000 at the third transfer in the second experiment. In community N, there was both a slight increase from 2400 to 7020 and a decrease from 5676 to 3724 observed in the first experiment, specifically at transfers two and five, respectively. A slight increase from 3072 to 5184 was observed at time six in the second experiment. The results strongly suggest that the pooled filtrates had a positive impact on nematode fitness and that this effect was exclusive to the I community (**Figure 4.1A**).

While nematode fitness was higher in the horizontal treatment of community I at various time points, the nematode population size did not increase over time in these communities. Instead, the difference between the treatments arose from a drop in nematode fitness in the vertical community. This initially emerged between transfers two and three, where the number of nematodes in the I vertical community sharply dropped from 23360 to 8480, and from 10360 to 8640, respectively in both replicate experiments. These observations imply that nematode communities may decline in the absence of the pooled filtrates, but this decline can potentially be reversed by the treatment (**Figure 4.1A**).

Significant and large differences were observed in nematode population sizes, with community I showing larger populations compared to communities P and N. Additionally, a slight but significant difference was observed between communities N and P in the first replicate experiment (**Figure 4.1A**).

4.2.3 Investigating nematode fitness following microbiome transplants

To search for direct evidence that the observed changes in nematode fitness are caused by variations within the microbiome, microbiome transplants were performed at the end of the main evolution experiments. Briefly, surface-sterilized nematodes harvested from mesocosms were ground and the extract was displayed on M9 buffer agar plates. A calculated average of 100 freshly hatched sterile nematode larvae were inoculated on the bacterial lawn and let to feed for six days. This was performed in triplicate (**Figure 4.1B**).

The number of nematodes grown on the microbiomes extracted from the first replicate experiment noticeably decreased by approximately ten-fold for microbiomes derived from horizontal community P, compared to vertical community P, with an average count dropping from 1133 to 66. Similarly, for microbiomes originating from horizontal community N, the number of nematodes decreased by approximately two-fold in comparison to microbiomes originating from vertical community N, from 1013 to 546. Despite the lack of statistical significance, the observed differences were substantial, and even striking for community P. In contrast, no discernible discrepancy was observed in the nematode count grown on the microbiomes of the P and N communities from the second replicate experiment when supplemented with pooled filtrates. In both experiments, no effect of the pooled MGEs was observed when nematodes were fed with microbiomes originating from the I community. These findings provide direct evidence of the pooled filtrates having negative impacts on nematode fitness, which is likely specific to P and N communities (**Figure 4.1B**). Interestingly, these effects are different from the trends observed during the course of these experiments. In fact, previous results showed no striking effect of the pooled filtrates on communities P and N. However, at the time when microbiomes were extracted for transplant, only microbiomes extracted from the I community in the second replicate experiment were expected to show differences between treatments (**Figure 4.1A**). Nonetheless, the transplant experiments provide more fine-tuned and direct evidence for the effect of fuelling communities with MGEs on nematode fitness.
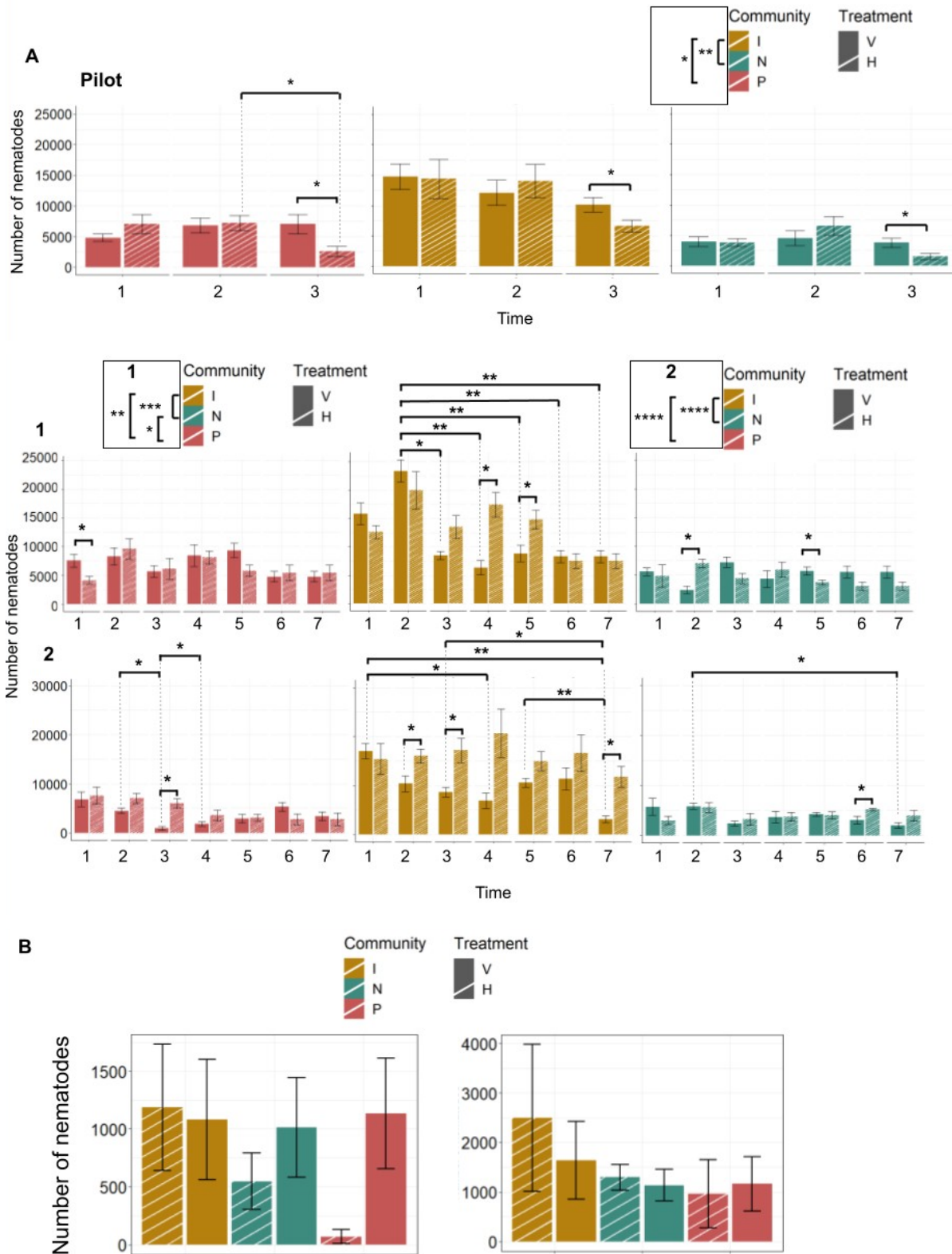
Figure 4.1. Fuelling communities with MGEs correlates with changes in nematode fitness. (A) Nematodes were harvested at each time point. The number of live nematodes was derived by counting worms in 5 replicate 5 µL drops after harvesting them from mesocosms using M9 buffer. Number of nematodes in each community (colours:

P, I, and N are shown in red, yellow, and green respectively) and for each treatment (plain and striped bars: vertical (V) and horizontal (H), respectively) across times of transfer in the pilot experiment (Pilot) and both replicates of the main experiment (1 and 2). Error bars represent the standard error of the mean for the five replicate counted drops. One-tailed Student t-tests corrected for multiple comparisons indicate statistical significance between samples (on the graphs) or between communities (above the graphs) (*: p.adj < 0.05; **: p.adj < 0.01; *** < 0.001; **** < 0.0001) (B) Nematodes were harvested at the last time point (T7) and the microbiome was extracted after surface sterilization. Extracted microbiomes were plated as triplicates on M9 agar plates and gnotobiotic nematodes were inoculated on the plates. The number of worms was calculated as previously explained after six days of growth on the microbiomes. Number of nematodes in each community (colours: P, I, and N are shown in red, yellow, and green respectively) and for each treatment (plain and striped bars: vertical (V) and horizontal (H), respectively). Error bars represent the standard error of the mean for the three replicates.

4.2.4   Taxonomic structure of host-associated microbial communities

C. elegans is impacted by MGEs and MGEs can affect communities in various ways. Darwinian entities of their own, the ability of these MGEs to survive in their host can on occasion conflict with their host's[26]. On the other end, MGEs drive the acquisition of horizontally transferred genes that might provide adaptive fitness benefits to certain types over others[115]. In both cases, as a consequence, MGEs are expected to lead to variations in community composition. The effect of MGEs on community composition was therefore investigated. To decipher whether MGEs might affect gut microbiome composition, 16S metabarcoding sequencing was performed on DNA extracted from the gut of surface-sterilized nematodes during the pilot experiment.

Within-sample (alpha) diversity was compared between samples that received pooled filtrates (horizontal, H) and those that did not (vertical, V). Results showed no significant difference in the Shannon alpha diversity index between the two treatments (Wilcox test: W-value = 37, p-value = 0.7962; **Figure 4.2A**). Between-samples (beta) diversity was then evaluated. Bray-Curtis dissimilarities between microbiome samples were visualized using non-metric multidimensional scaling (NMDS). The ordination graph showed clear and statistically significant differences in composition between the different communities (PERMANOVA test: F-value = 6.5502, p-value = 0.001). However, there was no significant difference when communities were treated with MGEs from other communities (PERMANOVA test: F-value = 0.3589, p-value = 0.924). Therefore, microbiome compositions were not significantly changed when fuelled with novel MGEs (**Figure 4.2B**). Nonetheless, microbial community function has been repeatedly shown to be determined by gene content, and not composition[116–119]. Thus, even without disturbance of the community composition, MGE-driven HGT could explain the observed difference in nematode fitness between the two treatments. A deeper look into genomic differences within the gut is,

therefore, necessary to decipher how HGT might affect the functioning of the *C. elegans* gut microbiome.
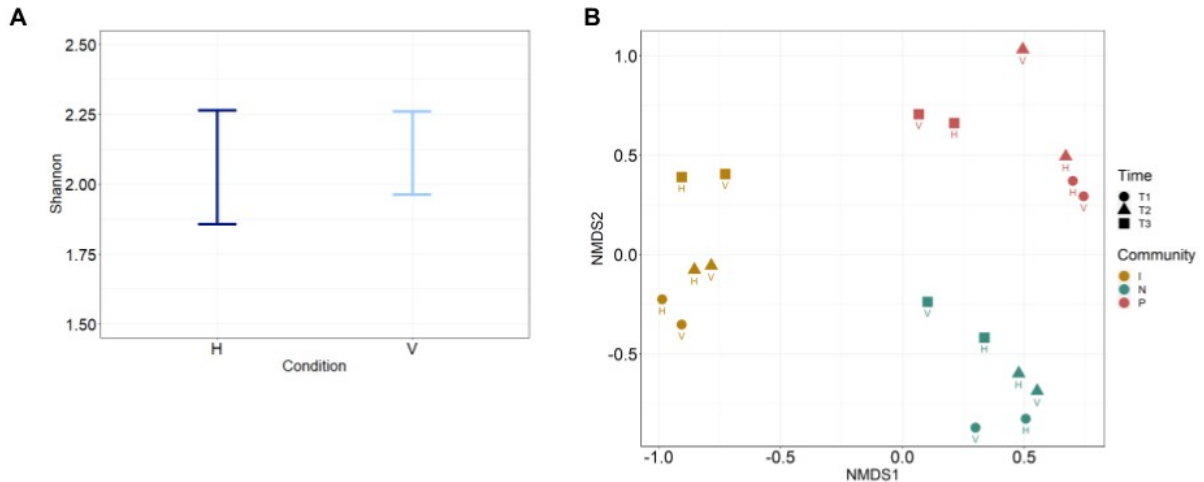
Figure 4.2. MGEs do not impact community composition. DNA was extracted from surface-sterilized and lysed surface-sterilized nematodes and 16S metabarcoding and sequencing were performed. (A) 16S Shannon index as a measure of alpha diversity. (B) 16S NMDS ordination plot of Bray-Curtis dissimilarities between microbiomes as a measure of beta diversity. H and V labels on the plot respectively describe the Horizontal and Vertical origins of the community.

4.2.5   Exploring the relationship between gene content and community function in the nematode gut microbiome

The previous findings indicate that the introduction of MGEs from other communities significantly impacts nematode fitness, while community composition remains unchanged. This suggests that HGT-mediated changes in gene content, rather than community composition, are key in shaping nematode fitness. However, the specific connection between horizontally transferred genes and their influence on community function requires further investigation. Ensuring accurate detection of HGT within the context of host-microbe interactions requires the development of reliable approaches. Subsequent sections of **Chapter 4** (**4.3**, **4.4**, and **4.5**) will delve into the exploration of various approaches to assess their feasibility in detecting HGT within the nematode gut microbiome. The most promising one will be employed to uncover specific HGT events in real-time within the nematode gut (**Chapter 4** section **4.6**).

4.3   Exploring the Possibility of Using the Xenoseq Pipeline to Detect HGT within *C. elegans* Gut

4.3.1   Introduction

Chapter 3 demonstrated the successful application of the Xenoseq pipeline to detect HGT in complex microbial communities. As a result, the feasibility of performing metagenomic sequencing on the gut microbiomes of *C. elegans* was explored. To assess the viability of this approach, microbiomes were extracted from surface-sterilized nematodes that were fed either a single laboratory bacterial strain, specifically, *Pseudomonas fluorescens* SBW25, or one of the compost-derived microbial communities.

4.3.2   Challenges in DNA extraction from nematodes for microbiome sequencing

4.3.2.1   *Gut microbiome sequencing: microbial DNA signal overwhelmed by nematode DNA signal in gut microbiome sequencing*

The NucleoSpin Blood and Tissue kit (NS) from Machery & Nagel was previously used to sequence DNA from nematodes, and the DNeasy UltraClean Microbial kit (DN) from Qiagen. Despite the high number of sampled nematodes, extractions resulted in extremely low DNA yields, i.e. in the order of magnitude of 10^{-1} to 10^{0}, independently of the microbiome composition (**Table 4.S1**). After sequencing the genomic DNA of triplicate samples from each condition, the analysis revealed that the average percentage of nematode DNA in samples prepared using the NS kit was 93.68% for nematodes grown on SBW25 and 88.55% for nematodes grown in the community. For samples prepared using the DN kit, the percentage of nematode DNA was 86.95% and 86.26% for nematodes grown on SBW25 and the community, respectively. This indicates that regardless of the strategy used, the majority of the sequenced DNA was attributed to the Nematoda clade, which covered most of the bacterial signal (**Figure 4.3A**). This can be a significant issue when trying to *de novo* assemble sequences.

4.3.2.2   *Attempt to increase microbial DNA yield*

In order to mitigate the issue of overwhelming nematode DNA signal in sequencing the nematode gut microbiome, a strategy was employed to decrease the amount of nematode DNA prior to sequencing. Nematode tissues were lysed, and different cycles and types of freezing and thawing were applied to selectively break open eukaryotic cells while preserving microbial cells. The supernatant, potentially enriched with eukaryotic DNA, was discarded, and subsequent washing steps were performed. This

approach seeks to reduce nematode DNA contamination and enhance the detection of microbial DNA in gut microbiome sequencing. Initially, a comparison was made between DNA extracted from nematodes that had simply been lysed following surface sterilization and samples that were additionally frozen at - 80 and thawed once before the DNA extraction. This did not show any noteworthy improvement in the bias of nematode DNA detection. In fact, the nematode DNA signal solely decreased from 93.68% to 84.96% and from 86.95% to 89.86% when nematodes were previously fed on a single known strain, for the NS and the DN kits, respectively. When samples came from nematodes fed on the community, the proportion of nematode DNA was even more similar to that of the control treatment, changing from 88.55 to 87.32% and 86.22 to 86.62%, for both kits, respectively (**Figure 4.3A**). Therefore, using nematodes fed on the single bacterial strain only, the freezing temperature was decreased using liquid nitrogen, and the number of cycles of freezing and thawing increased to 10, with or without washing the sample afterward. Neither the number of cycles nor the washing of the samples visibly altered the amount of nematode DNA sequenced (**Figure 4.3B**).

A

Microbiome	Kit	Treatment
SBW25	NS	LFW
SBW25	NS	L
SBW25	DN	LFW
SBW25	DN	L
Community	NS	LFW
Community	NS	L
Community	DN	LFW
Community	DN	L

B

Treatment		phylum
LFW N1	6.61 % ... 89.77 %	Actinobacteria
LFW N10	.46 % ... 92.41 %	Chordata
LF N1	8.44 % ... 87.44 %	Firmicutes
LF N10	.23 % ... 92.32 %	Nematoda
		Proteobacteria

Hits (%)
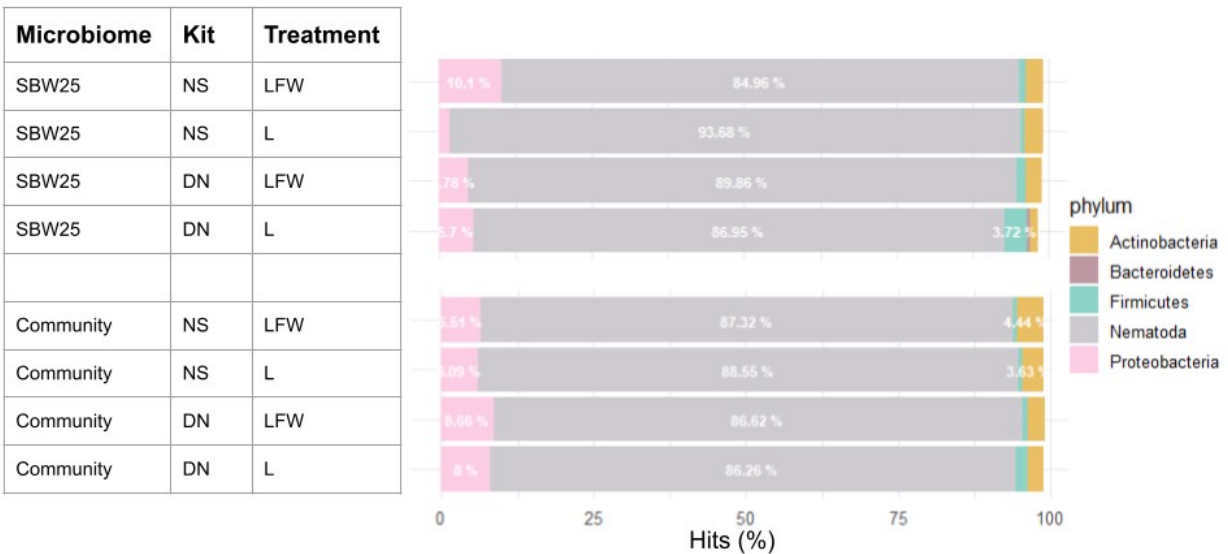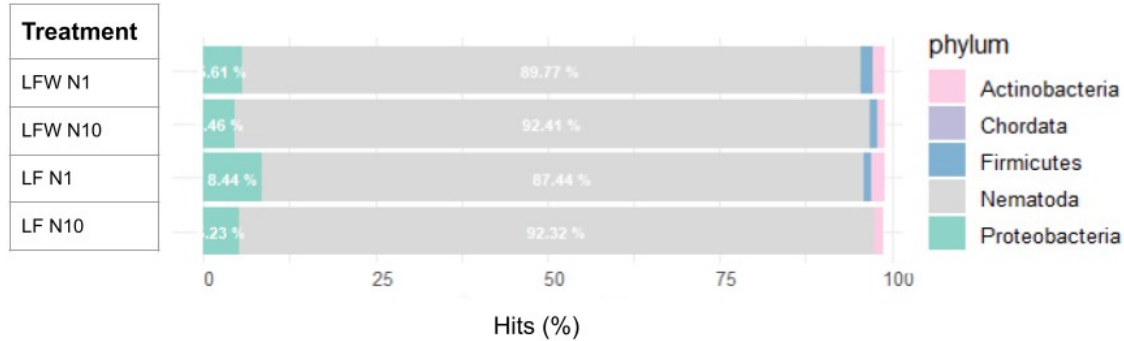
Figure 4.3. DNA signal from nematodes overwhelms microbial DNA signal from the gut microbiome. Percentage of reads attributed to a given phylum using diamond. (A) Nematodes were grown on either a rich-media plate seeded with SBW25 for three days or on the cellulose mesocosms inoculated with community P and incubated for two weeks. Samples were either simply lysed (L) or subjected to a lysis-freezing-thawing (LFW) cycle with freezing at -80C. DNA was extracted using either the DNeasy UltraClean Microbial kit (DN) or the Nucleospin Blood & Tissue kit (NS). (B) Nematodes were grown on the cellulose mesocosms inoculated with community P and incubated for two weeks. Samples were either simply lysed (L) or subjected to one (N1) or 10 (N10) lysis-freezing-thawing (LFW) cycle(s) with freezing in liquid nitrogen.

4.3.3   Conclusion

Overall, based on the findings from the conducted experiments, it can be concluded that achieving a higher yield of bacterial DNA is a complex and challenging task. As a result, it was determined that an alternative approach would be more reliable in detecting HGT within the nematode gut microbiome.

4.4   Assessing the Possibility of Tracing HGT in the Nematode Gut with a Barcoded Library of *P. fluorescens* SBW25

4.4.1   Introduction

In recent years, field experiments have shown that wild *C. elegans* nematodes possess a core gut microbiome characterized by the presence of the same bacterial genera irrespective of the host habitat[120]. *Pseudomonas* is one of these core genera. A Tn*7* barcoded library of the genetically well-known *P. fluorescens* SBW25 strain has been previously engineered in-house[121]. Here my goal is to investigate the suitability of this barcoded library to study *in vivo* evolution of single lineages of SBW25 within the gut of nematodes. The ultimate goal is to trace potential HGT events occurring within the individual lineages of this focal strain. The following section presents the results regarding the toxicity of

SBW25 to nematodes, the feeding and reproductive capabilities of *C. elegans* nematodes on SBW25, and the colonization and persistence of the bacteria within the nematode gut.

4.4.2   Exploration of the type of interaction between *P. fluorescens* SBW25 and *C. elegans*

First, the toxicity of SBW25 to nematodes was assessed in comparison to the commonly used *E. coli* OP50 *C. elegans* food source. Briefly, either SBW25 or OP50 bacterial lawns were spotted on the centre of 6-cm diameter PFM plates prepared from overnight cultures washed in Ringer's solution. Bacterial-free Ringer's solution lawns were also made as a control for the nutritive role of the buffer itself. An additional outer ring of OP50 lawn was spread on each plate to provide an alternative food source. Each condition was performed in triplicates. A known number of stage 4 larvae (L4) nematodes were inoculated on the central lawn. All plates were incubated at 20°C and nematodes located outside of the central lawns were counted after 24 and 48 hours. The leaving index was calculated by dividing the number of nematodes outside the lawn by the total number of nematodes on the plate. This assay was replicated three times.

The leaving index of nematodes from the Ringer's lawn after 24 hours showed significant variability across replicate assays, with values of 86%, 18%, and 75%, respectively. Notably, within the second assay, the standard error was very high, indicating further fluctuation across technical replicates. Nematodes displayed on the Ringer's lawn hence expectedly tended to leave the centre of the plate, supposedly to reach to outer *E. coli* ring. In contrast, none of the nematodes incubated with either SBW25 or OP50 left the lawn within the first 24 hours in any of the three replicate assays. This strongly suggests that SBW25 is a non-toxic bacterium for *C. elegans* (**Figure 4.4**).

After 48 hours, the response of nematodes to SBW25 varied greatly both within and between assays. In the first assays, all nematodes left the SBW25 lawn, but the same occurred for nematodes on Ringer's and OP50 lawns. This phenomenon could potentially be attributed to the nematodes depleting the food resources at the centre of the plate and subsequently moving toward the periphery. Besides, the nematodes sampled for this experiment came out of a densely packed plate. It is likely that the nematodes inoculated at the start of the experiment were already lacking food, inducing an accelerated eating behaviour and leading them to reach out for the outer ring OP50 lawn faster. In the second assay, nematodes did not leave the SBW25 lawn, while nematodes on Ringer's lawn escaped. In the third experiment, nearly all nematodes were found outside the OP50 and Ringer's lawns, whereas the large variation in their number on SBW25 lawns prevented any definitive conclusions about their behaviour. The feeding behaviour of nematodes could also have varied between the different assays due to the

variable number of nematodes inoculated at the start of each assay, i.e. ranging from about 10 to 50 worms **(Figure 4.4).** Nonetheless, these results do suggest that nematodes, not only tolerate SBW25, but might even consume it as a food source.
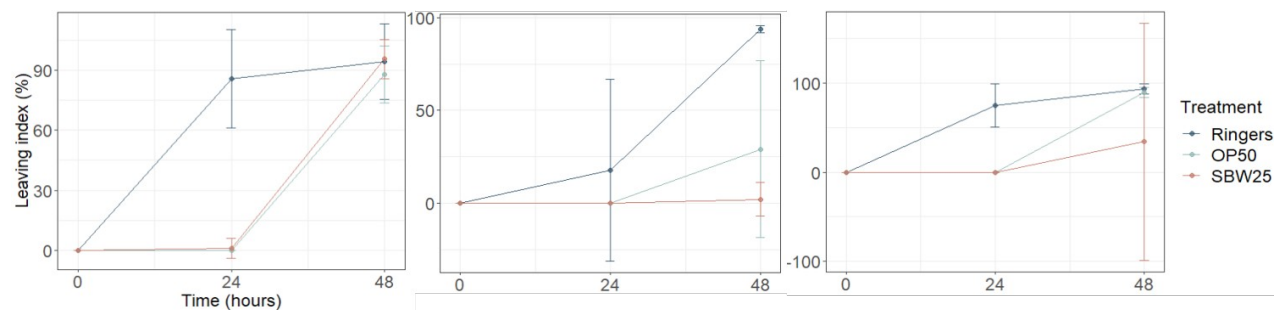
Figure 4.4. ***P. fluorescens* SBW25 is not toxic *to C. elegans*.** L4 nematodes were incubated for 48 hours at 20°C on different lawns (colours: control Ringer's solution, *E. coli* OP50, and *P. fluorescens* SBW25 lawns are shown in blue, green, and red, respectively). The number of worms having escaped the central lawn was calculated over time. Each condition was performed in triplicates. The leaving index (# worms outside the central lawn / total # of worms) averaged for triplicates is represented here over time with 95% confidence interval error bars. Each panel represents an independent replicate assay.

4.4.3   Assessment of the reproductive behaviour of *C. elegans* on *P. fluorescens* SBW25

To investigate whether *C. elegans* can feed on SBW25, the reproductive capacity of the nematodes when grown with this bacterial genotype was examined. The experiment involved inoculating single self-fertilizing L4 nematodes onto SBW25 or OP50 lawns. The experiment was conducted with 10 or 12 replicates for each treatment. Every day, the individual worms were transferred to a fresh lawn, and the number of offspring laid on the plate was counted every second day.

Nematodes produced an average of 225 and 269 offspring across the three replicate experiments when grown on SBW25 and OP50, respectively. This difference was statistically significant, with nematodes producing approximately 19% fewer offspring on SBW25 than on OP50 (Two-way ANOVA, F-value = 11.825, df = 1, p-value = 0.00153). This difference was expected since the *C. elegans* nematode laboratory strain used in these assays is highly adapted to its common food source, OP50. However, despite the slight decrease in offspring, these results strongly indicate that *C. elegans* is capable of feeding on and ingesting SBW25.

The reproductive patterns of the nematodes were similar when they were grown on both SBW25 and OP50 bacterial strains. The maximum number of offspring was observed between 120 and 150 for both strains. However, there were slight differences in the timing of reproduction. Nematodes on SBW25 exhibited a higher number of offspring during the first two to three days but reached sterility

slightly earlier, around four to five days. In contrast, nematodes on OP50 remained fertile for a longer period, approximately five to six days. These findings demonstrate that the duration of fertility differed between the two bacterial strains (**Figure 4.5B**).
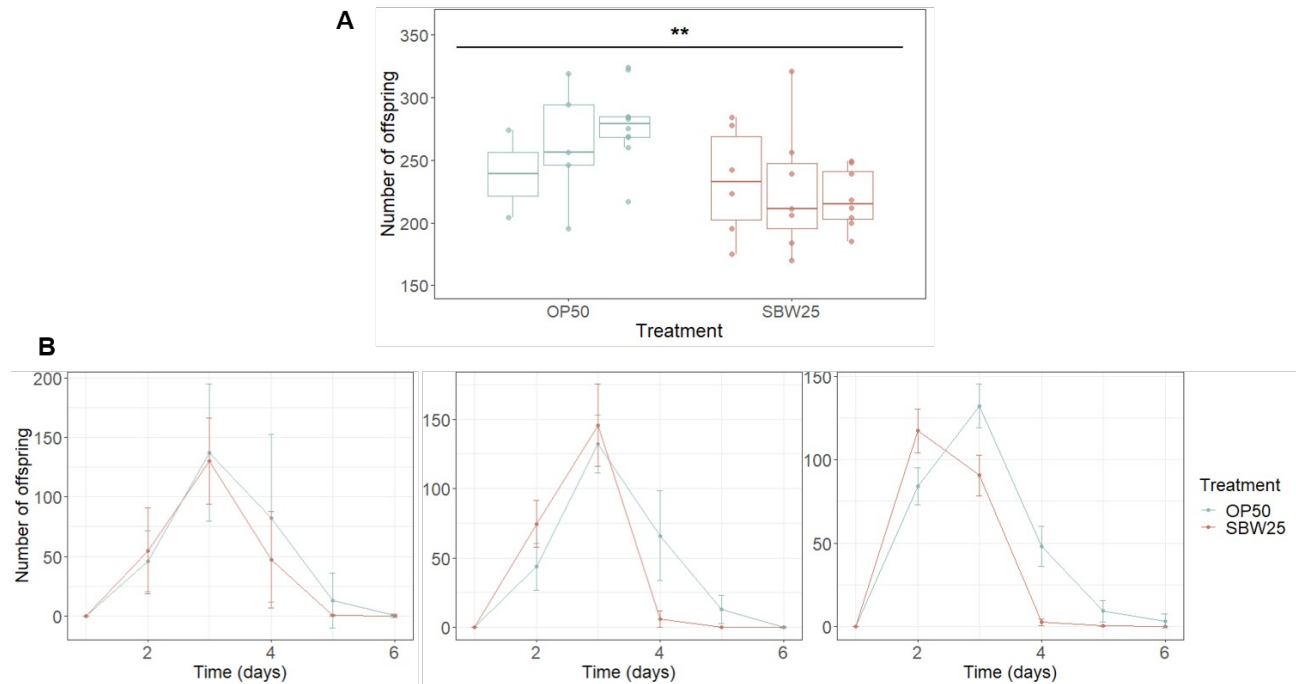
Figure 4.5. *C. elegans* can reproduce on *P. fluorescens* SBW25. Single L4 nematodes were inoculated on either OP50 or SBW25 bacterial lawns (colours: *E. coli* OP50, and *P. fluorescens* SBW25 lawns are shown in green and red, respectively) and incubated for 6 days at 20°C. Every day, single nematodes were transferred to fresh lawns while the offspring were let to hatch and counted. Each condition was performed in 10 to 12 replicates. (A) Boxplot of the number of offspring laid by a single worm over the six-day course experiment for three independent replicate assays (panels from left to right). Statistical analysis performed using two-way ANOVA showed a significant difference between the number of offspring laid on OP50 compared to SBW25 (**: p-value < 0.01). (B) The average number of offspring laid by a single worm every day. Each panel represents an independent replicate assay. Error bars represent 95% confidence intervals.

4.4.4 Investigating the colonization and persistence of *P. fluorescens* SBW25 in the gut of *C. elegans*

4.4.4.1 *A novel method for surface-sterilizing nematodes: improved accuracy and reliability in nematode counting*

For the following colonization and persistence assays, a novel way of surface-sterilizing nematodes was developed. This method is based on the use of filters fitting into autoclavable holders. The procedure renders possible a precise count of nematodes by removal of the filter from the holder and consecutive microscopy visualization. Additionally, contrary to "in-tube" sterilization, which is based on multiple

pipetting of supernatants, nematodes cannot be wrongfully discarded during the procedure, thereby increasing the reliability of the nematode count.

4.4.4.2   *The colonization ability of* C. elegans *gut by* P. fluorescens *SBW25*

As previously demonstrated, nematodes can use SBW25 as a food source. However, in order to evolve SBW25 within the nematode gut, at least a portion of the bacterial population should be able to survive within this environment. Therefore, the ability of SBW25 to colonize the gut of *C. elegans* was tested. *Pseudomonas lurida* Myb11 has previously been shown to colonize the nematode gut and was thus used as a positive control for gut colonization[122]. OP50 was also included for comparison. After three days of growth on the bacteria, the lawns were washed with Tetramisole solution to prevent subsequent bleach intake by the nematodes. Nematodes were then retained on a filter. Soft bleaching was performed to remove surface-attached bacteria, followed by thorough washing. Nematodes were then counted under a microscope after removing the filter. Next, they were lysed to release the gut bacteria, and multiple dilutions were plated on LB agar for colony counting.

OP50 was barely detectable in the gut of *C. elegans*, while Myb11 exhibited significantly higher colonization compared to OP50. SBW25 also consistently colonized the nematode gut more effectively than OP50. The bacterial loads of Myb11 and SBW25 in the gut were approximately three orders of magnitude higher than that of OP50. The mean number of CFUs for Myb11 and SBW25 were 1188 and 1217, respectively, compared to only 32 for OP50. This difference between the two strains and OP50 was statistically significant (**Figure 4.6**). These findings collectively indicate that SBW25 is not only ingested by the nematodes but can also survive and thrive within their gut.
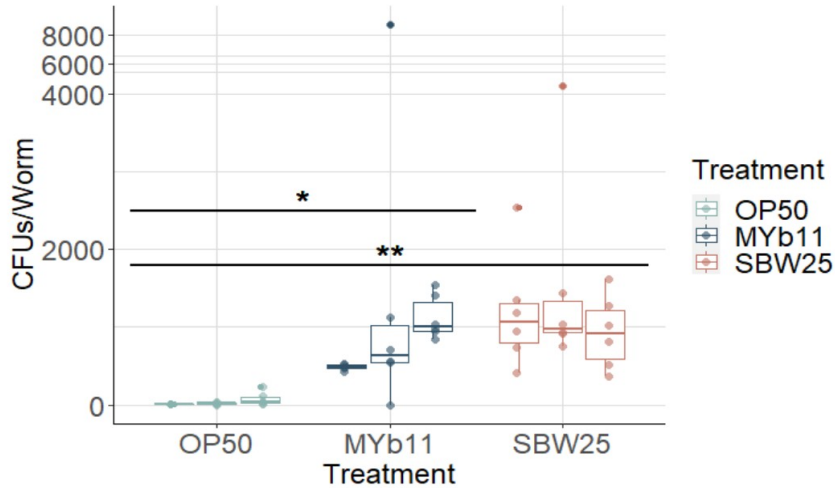
Figure 4.6. *P. fluorescens* SBW25 can colonize the gut of *C. elegans* N2. L1 nematodes were grown for three days at 20°C on different bacterial lawns (colours: *E. coli* OP50, *P. lurida* Myb11, and *P. fluorescens* SBW25 lawns are shown in green, blue, and red, respectively). Nematodes were collected on a filter, surface sterilized, counted, and lysed for bacterial plating and counting. Boxplots of the number of CFUs per worm are shown in the figure for the three treatments. Statistical analysis performed using Two-way ANOVA followed by the Tukey test showed a significant difference in the bacterial load between OP50 and SBW25 (**: p-value < 0.01) and between OP50 and Myb11 treatments (*: p-value < 0.05).

4.4.4.3   Persistence of P. fluorescens *SBW25 in the* C. elegans *gut*

4.4.4.3.1   Within the nematode gut

To establish SBW25 as an effective model system for studying bacterial evolution in the nematode gut, it is crucial to investigate its capacity not only to colonize but also to persist within the gut over time. To address this question, nematodes were initially cultured on PFM medium supplemented with SBW25-mScarlet, a strain that expresses the mScarlet fluorescent marker and is resistant to gentamicin (Gm), for a duration of three days. Subsequently, the nematodes were washed and transferred to fresh OP50 lawns, with daily transfers to ensure continuous exposure to new bacteria. At each time point, the nematodes were harvested, surface-sterilized, and lysed. Two distinct assays were performed to assess SBW25 persistence. Firstly, the SBW25-mScarlet bacterial load was quantified by plating dilutions of the nematode lysate on LB medium supplemented with Gm. Concurrently, a subset of nematodes was imaged to confirm the presence of the fluorescent signal emitted by SBW25-mScarlet.

The results revealed that the initial bacterial load of SBW25 in the nematode gut was relatively low, with approximately 200 CFUs on the first day (**Figure 4.7A**). Noticeably, this finding contrasts with the higher bacterial load observed in the previous colonization assay (**Figure 4.6**). The likely cause for this discrepancy was identified as the difference in surface-sterilization protocols. In this experiment, a

higher concentration of NaClO (0.5%) was used compared to the colonization assays (0.05%). At this increased concentration, it is expected that all bacteria attached to the nematode surface would be eliminated, providing an explanation for the lower initial bacterial load observed, and ensuring that solely gut bacteria are considered in this analysis (**Figure 4.S1**). The lower initial bacterial load of SBW25 in the nematode gut does not contradict the previous findings of its colonization ability, as a significantly higher load was then observed compared to the negative control strain OP50 (**Figure 4.7**).

Although there was no significant decrease in bacterial load between the initial time point and subsequent time points, it appeared that the bacterial load declined to nearly zero after either two days or three days in all replicate assays. Additionally, the absence of the mScarlet signal in the gut two days after the switch in food source indicated the lack of SBW25 persistence, contrasting with the clear signal observed at the start of the experiment. These findings suggest that SBW25 is unable to establish long-term colonization in the nematode gut **(Figure 4.7A)**.

4.4.4.3.2   Outside the nematode gut

The microbiome of *C. elegans* is believed to be mainly inherited during ex-utero development, specifically through feeding, rather than being directly transmitted from parent to offspring[55]. As a result, newly hatched larvae rely on acquiring bacteria from their environment to establish their microbiome. Given this mode of horizontal transmission, it was hypothesized whether the presence of nematodes could facilitate the persistence of SBW25 in a minimal medium. To investigate this hypothesis, SBW25 lawns were incubated with or without nematodes for three days on a PFM plate. Subsequently, the plates were washed, and a subset of the wash was plated onto PFM-OP50 and transferred every three or four days. At each time point, the number of bacteria outside the gut was quantified by counting the number of CFUs.

During the 14-day experiment, a significant decline in the bacterial load of SBW25-mScarlet was observed early on, after three days, in the presence of nematodes. The bacterial load decreased approximately threefold, from 3e07 to below 1e07, and this lower level was maintained throughout the entire experiment. Furthermore, it is worth noting that there was a small but significant difference in bacterial load between the absence and presence of nematodes at multiple time points, specifically on days seven, 10, and 14. Although there was no significant drop in the bacterial load in plates without nematodes at the early time point, a small significant decrease was observed between times three and 14, as well as between times seven and 14 (**Figure 4.7B**).

Noticeably, there was an initial disparity in the bacterial load between the two conditions. Additionally, it was anticipated that both conditions would experience a decrease in bacterial load due to dilution effects. In order to evaluate whether the presence of nematodes counteracts the dilution effect and to mitigate any potential bias from the initial disparity, the rate of decrease between the initial and final time points was compared, offering a more robust and unbiased measure. Results showed that the SBW25-mScarlet load significantly decreased approximately three times faster when the bacteria were grown in the presence of *C. elegans* (**Figure 4.7B**). These findings collectively suggest that nematodes do not effectively maintain SBW25 on solid minimal media.

A

B
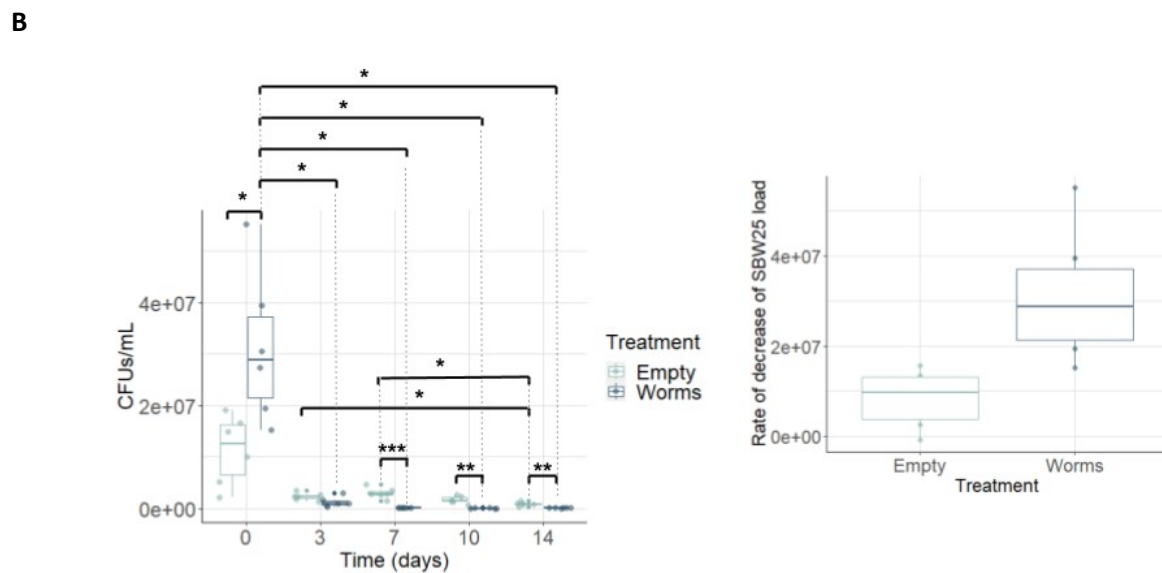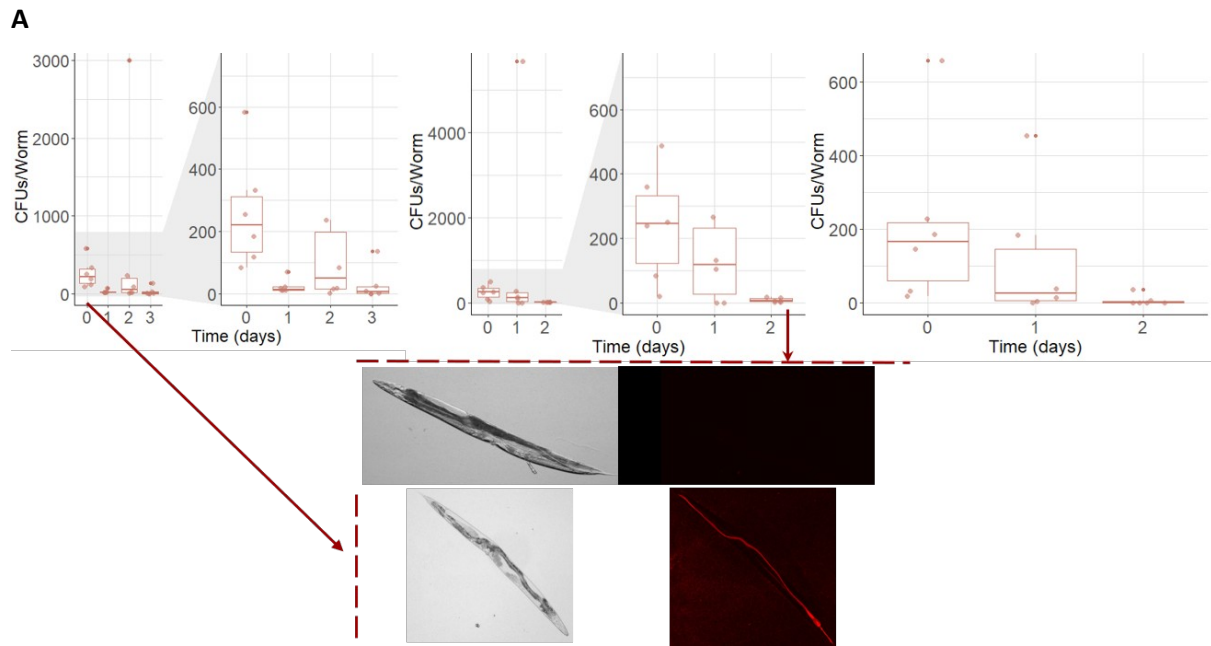
Figure 4.7. *P. fluorescens* SBW25 does not persist in the gut of *C. elegans* and the presence of the nematodes does not allow the bacteria to be maintained outside the gut. (A) Nematodes were harvested after being grown for three days on SBW25-mScarlet and every day after the food switch to OP50. Number of CFUs of SBW25 per worm before (Time 0) and after food switch (Times 1, 2, 3) for three independent replicate assays. Two representative examples of nematode pictures taken in brightfield and fluorescent (Scarlet) channels from day 0 and 2 post-switch are shown with respective exposure times. (B) SBW25-mScarlet was grown in the presence (Worms; dark blue) or absence (Empty; light blue) of nematodes. Number of CFUs on the plate (left plot) and rate of decrease of SBW25 load (right plot) between time points 0 and 14. One-tailed Student t-tests corrected for multiple comparisons indicate statistical significance between samples (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

4.4.5   Conclusion

To summarize, the study demonstrates that *P. fluorescens* SBW25 is safe for *C. elegans* nematodes to consume as a food source and does not exhibit toxicity (**Figures 4.4** and **4.5**). However, SBW25 does not establish long-term colonization in the nematode gut and shows limited persistence on solid minimal media in the presence of nematodes (**Figures 4.6** and **4.7**). These findings indicate that SBW25 is highly unlikely to survive in an evolution experiment within the nematode gut community. Consequently, SBW25 is not suitable as a barcoded library for studying *in vivo* evolution and HGT acquisition in the *C. elegans* gut. The next phase of this study is therefore dedicated to identifying persistent colonizers within the nematode gut when nematodes are grown in community mesocosms.

4.5   Uncovering Candidate Genomes Prone to HGT through a Short-Term Evolution Experiment

4.5.1   Introduction

Previous attempts to detect HGT within the nematode gut microbiome, whether through metagenomics or the use of an in-house barcoded library of a well-known bacterial genome, proved thus far unsuccessful (**Chapter 4** sections **4.4** and **4.5**). Challenges arose from both technical limitations and inherent biological complexities. As a result, a shift in strategy was warranted. Rather than examining the entire community or population, a decision was made to search for individual genomes of interest that could be tracked and traced during a pilot evolution experiment, increasing the likelihood of capturing HGT events (**Figure 1.1**). By adopting this targeted approach, the aim is to overcome the limitations encountered thus far and shed light on the elusive mechanisms of gene transfer within the nematode gut microbiome.

4.5.2 Prevalence of *Ochrobactrum* and *Pseudochrobactrum* genera in the nematode gut

To be traceable across time, and to have a noticeable effect on nematode fitness, genomes of interest should be both dominant and persistent within the nematode gut. In the search for possible dominant and persistent types within the gut, the genus composition of the different samples was inferred from 16S metabarcoding sequencing (**Figure 4.8A**).

For all samples, independently of the communities, times of transfer, and whether or not fuelled with MGEs, two genera were found to represent most of the gut microbiome. These genera, called *Ochrobactrum* and *Pseudochrobactrum*, were strikingly predominant, with either of them having a relative abundance above 50% in all communities over the course of the entire experiment, and independently of whether or not they received a pooled filtrate (**Figure 4.8A**). Consistent with these findings, *Ochrobactrum* was previously documented to be a major and persistent part of the nematode gut microbiome [55].

4.5.3 Diversity of *Ochrobactrum* and *Pseudochrobactrum* genera in the nematode gut microbiome

Bacteria were isolated from the nematode gut at times of transfer in the hope to recover genomes from these two prevalent genera. A total of 348 genomes were isolated from all samples at different time points and DNA was extracted for short-read sequencing. Sequenced genomes were *de novo* assembled. 273 genomes were retained for analysis after a quality check (**Table 4.S2**). The assemblies were then annotated using Prokka and *gyrA* nucleotide sequences were extracted and aligned[69]. A neighbor-joining phylogeny was inferred from the alignment (**Figure 4.8B**).

The *gyrA* alignment contained 280 sequences after the elimination of partial sequences (below 1500 base pairs). The number exceeded the total number of genomes due to the presence of duplicated *gyrA* genes within certain genomes. Unique *gyrA* nucleotide sequences were blasted against the nucleotide database for taxonomic annotations. Strikingly, most of the *gyrA* sequences (207/280, 74%) were found to belong to either *Ochrobactrum* (152/280, 54%) or *Pseudochrobactrum* (55/280, 20%). The diversity of *gyrA* sequences within both genera was extensive and identical sequences were repeatedly found in all communities and times of transfer independently of MGEs treatment (**Figure 4.8B**). Therefore, besides their striking predominance and persistence within the nematode gut, these genera, and notably *Ochrobactrum*, encompass a wide variety of cultivable types.
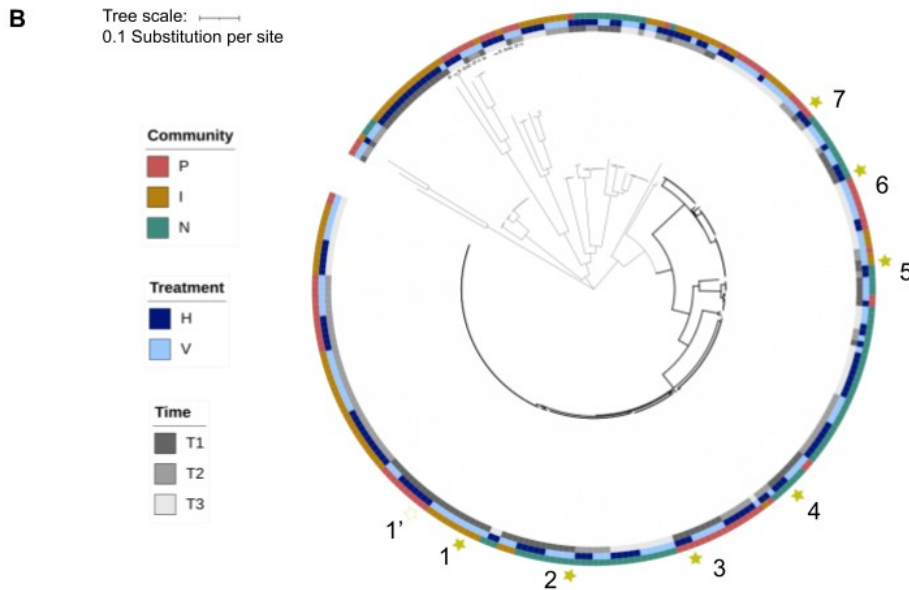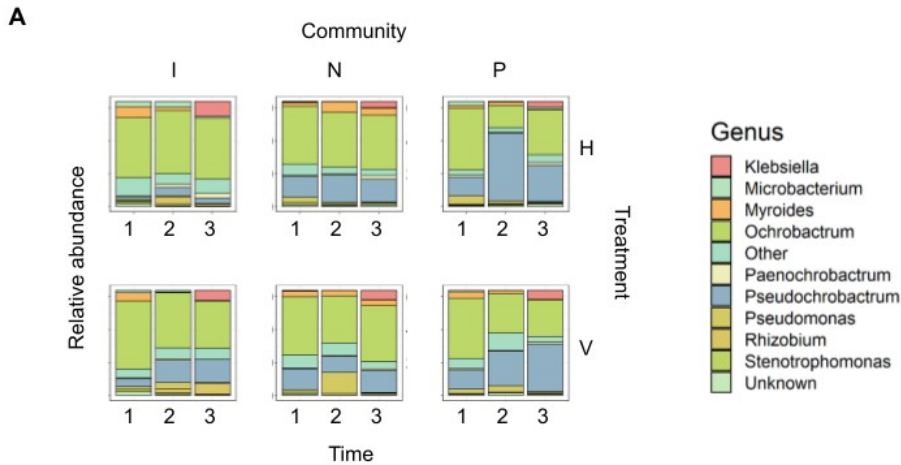
Figure 4.8. *Ochrobactrum* and *Pseudochrobactrum* are the most abundant genera found within the gut microbiome of *C. elegans*. (A) 16S relative abundances of the 10 most abundant genera within the gut microbiome derived from 16S metabarcoding sequencing using phyloseq. (B) *gyrA* neighbor-joining phylogeny (Tamura-Nei distance method) of isolated genomes reconstructed from multiple sequence alignment using Clustal omega. In bold *Ochrobactrum* and *Pseudochrobactrum* genera. Yellow stars indicated by arrows are the strains chosen for long-read sequencing. The clades corresponding to the eight isolated reference genomes used for the main experiment are highlighted as stars and numbered 1, 1', 2, 3, 4, 5, 6, and 7 (plain and empty stars for reference 1 and 1' refer to two references that belong to the same clade). Isolates labelled with identical letters (a to g) correspond to the same isolates with non-identical gyrA sequences.

4.5.4   The capacity of *Ochrobactrum* and *Pseudochrobactrum* populations for gene gain

In the pilot evolution experiment, *Ochrobactrum* and *Pseudochrobactrum* appeared to constitute the majority of the microbiome population within the gut in all samples. Both in the pilot and the main

experiment, MGEs had a clear impact on nematode fitness. At this point, whether or not this observed effect is ecologically or evolutionary relevant for the gut microbial components remains to be tested. Analyses were therefore performed to find out if the main genera present within the gut microbiome have a noteworthy ability for gene gain.

4.5.4.1   Deciphering the flexibility of the dominant gut genera through pangenome analysis

The pangenome is the entire set of genes present within a given population of genomes. It is constituted of the accessory genome and the core genome. While the core genome refers to the set of genes that reside in all or nearly all genomes (here, in 95% of the genomes), the accessory genome is a set of genes that belong to certain genomes only (here, in less than 95% of the genomes). The accessory pangenome can be divided into two parts; while the cloud is restricted to the genes present in a single genome, the shell comprises the genes that are found in a small number of genomes (here, higher than 1 genome and less than 95% of the genomes). The size ratio between accessory and core genomes reflects genome flexibility for a given population. It is therefore a suitable way to decipher whether the dominant genera identified in *C. elegans* gut might have a favourable ability to gain novel genes via HGT.

4.5.4.2   Accounting for genome size variations between Ochrobactrum and Pseudochrobactrum

The greater the differences between genomes within a studied population, the more open the pangenome is. However, if genomes are too different in size, and consequently in gene content, this comparison might not be entirely relevant. To avoid bias in the analysis due to variations in genome sizes, the distribution of the number of ORFs of isolates from both genera was plotted. This revealed two separate ORFs distributions for both genera (**Figure 4.S2**). On account of these results, two separate pangenome analyses were carried out on the two sub-population of genomes. This analysis also revealed the presence of three *Pseudochrobactrum* genomes outside the genus ORFs distribution. Those were therefore not considered for further analysis.

4.5.4.3   Analysis of the Ochrobactrum *pangenome*

Ochrobactrum is a likely candidate that has changed its gene content due to MGE acquisition (**Figure 4.8A**). To confirm the possibility for genes to move between *Ochrobactrum* strains, the pangenome of the genus was reconstructed using the 165 genomes identified as *Ochrobactrum* based on their *gyrA* sequence. Pangenome and core genome curves shown in **Figure 4.9A** are representations of the rate at

which genes are incorporated in the population as genomes are progressively considered. For the *Ochrobactrum* population, a continuous increase in the number of newly added genes was observed. Notably, genomes separated in two different populations, likely due to differences in the distribution of numbers of clusters of orthologous genes (**Figure 4.9A**). In fact, the distribution of the number of ORFs showed a large bimodal distribution for *Ochrobactrum* (**Figure 4.S2**). Coincidentally with this increase in new genes in the pangenome, a sharp decrease in the rate of newly added core genes was detected early on. Besides, the number of genes newly incorporated in the core genome reached zero before half of the genomes were considered (**Figure 4.9A**). 97% of the genes were attributed to the accessory pangenome, with 84.6% corresponding to the shell and 12.4% to the cloud. The core genome itself only represented 3.03%, of all genes (**Figure 4.9B**). These results strongly suggest an open pangenome for *Ochrobactrum*. This signifies that genes are likely to be gained by these genomes. These conclusions are consistent with a precedent pangenome study on *Ochrobactrum* clinical and environmental isolates. Indeed, the authors also showed an open pangenome for *Ochrobactrum* and highlighted the presence of MGEs hallmarks, such as transposases, integrases, and recombinases[123].

The principal component analysis (PCA) plots did not show any clustering of the pangenomes depending on the community, treatment, or time points (**Figure 4.9C**). In agreement with previous results, it thus seems that fuelling communities with MGEs does not influence the composition of *Ochrobactrum* gut populations. However, it would be necessary to isolate a higher number of genomes to confirm this hypothesis. Nonetheless, the lack of population structure variations between samples at the genus level, while an impact on nematode fitness is observed, reinforces the idea of particular HGT occurrences affecting community function (**Figures 4.1** and **4.9**).
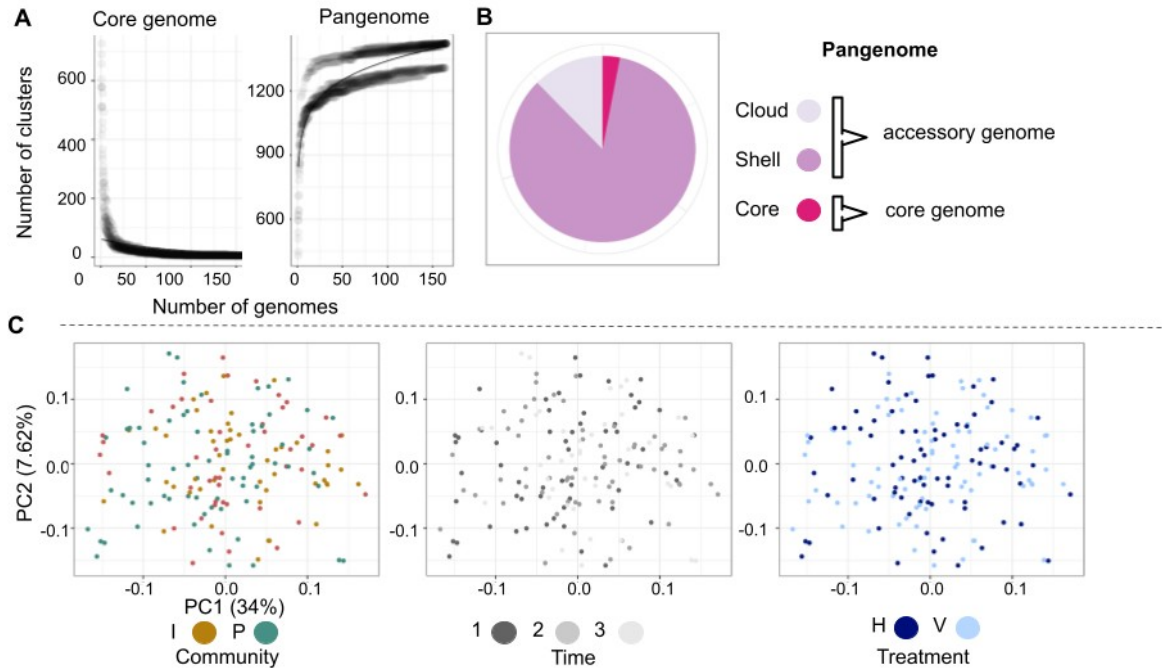
Figure 4.9. The *Ochrobactrum* population appears to have an open pangenome. (A) Core genome and pangenome curves showing the number of clusters of orthologous genes subsequently discovered as more genomes are considered (a total of 165 *Ochrobactrum* genomes). (B) Diagram showing the number of genes present within the core and accessory genomes. (C) PCA plots of the pangenome with respecting grouping by community, time points, and treatment (plots from left to right).

4.5.4.4 Analysis of the Pseudochrobactrum *Pangenome*

Pseudochrobactrum was the second most common genus found within the nematode gut in this study (**Figure 4.8A**). The same pangenome analysis was therefore implemented on the 41 retained *Pseudochrobactrum* genomes (**Figure 4.10**). The small number of core genes, together with the steady increase of newly added genes in the pangenome indicated that the *Pseudochrobactrum* population also harbored an open pangenome (**Figure 4.10A**). Also similar to *Ochrobactrum*, this population had a narrow core genome (4.34% of the genes) and a large accessory genome, i.e. 95.65% corresponding to 5.05% of genes in the cloud and 90.6% in the shell (**Figure 4.10B**). Although no obvious differences in the clustering of the genomes based on a community, time or treatment could be observed, this may be a result of the small number of the considered genomes for the analysis (**Figure 4.10C**).

Figure 4.10. The *Pseudochrobactrum* population appears to have an open pangenome. (A) Core genome and pangenome curves showing the number of clusters of orthologous genes subsequently discovered as more genomes are considered (a total of 41 *Pseudochrobactrum* genomes). (B) Diagram showing the number of genes present within the core and accessory genomes. (C) PCA plots of the pangenome with respecting grouping by community, time points, and treatment (plots from left to right).

4.5.4.5   *Intra-clade* Ochrobactrum *and* Pseudochrobactrum *pangenomes*

To assess the extent of genomic diversity within closely related strains of *Pseudochrobactrum* and *Ochrobactrum* genera, two new pangenome analyses were then performed on genomes with identical *gyrA* sequences respectively from the two genera (**Figure 4.11**). Remarkably and even when the number of genomes in the population was very small, both clades displayed a vast accessory genome, respectively 95.68 and 94.31% for *Ochrobactrum* and *Pseudochrobactrum* clades (**Figure 4.11B and D**). Genes in the *Ochrobactrum* clade kept being incrementally added, though at a slow pace, after considering the totality of genomes. In contrast, no further genes were observed in the *Pseudochrobactrum* population after considering as few as six genomes among 15. Therefore, this particular *Ochrobactrum* clade displayed a more open pangenome than the studied *Pseudochrobactrum* clade (**Figure 4.11A and C**).
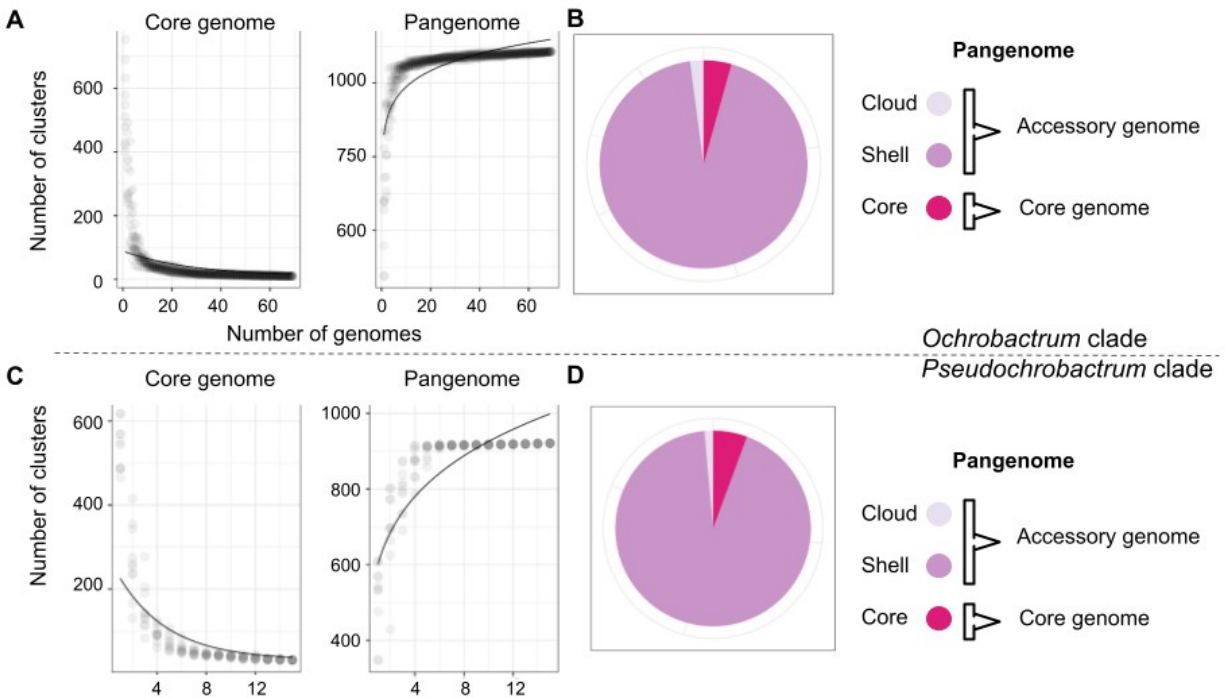
Figure 4.11. Two clades of *Ochrobactrum* and *Pseudochrobactrum* strains respectively display an open and a closed pangenome. Respectively for two clades, each characterized by a given *gyrA*, and taxonomically annotated as *Ochrobactrum* (total of 69 genomes) and *Pseudochrobactrum* (total of 15 genomes): (A) and (C) Core genome and pangenome curves showing the number of clusters of orthologous genes subsequently discovered as more genomes are considered, (B) and (D) Diagrams showing the number of genes present within the core and accessory genomes.

4.6   Investigating HGT Acquisition in *Ochrobactrum* and *Pseudochrobactrum* Genomes during a Short-Term Evolution Experiment

The remarkable persistence, diversity, and abundance of *Ochrobactrum* and *Pseudochrobactrum* genomes within the nematode gut, coupled with their open pangenome, strongly suggest their involvement in HGT in this particular context (**Chapter 4** section **4.5**). Driven by these compelling findings, the decision was made to proceed with the proposed single-genome tracking methodology. This approach holds the tremendous potential to follow the eco-evolutionary dynamics of individual lineages within the nematode gut. Specifically, it offers a direct means to investigate the hypothesis of HGT occurrence in real-time within these specific genera during a short-term evolution experiment.

4.6.1   Tracing of selected genotypes in the nematode gut

The hypothesis that HGT can occur within those genera in the time frame of a short-term evolution experiment was therefore tested in the main version of the previously described evolution experiment (**Figure 1.1**). Briefly, *C. elegans* was initially fed with a mixture of eight reference strains before being introduced to the community mesocosms. To increase the possibility for these strains to persist in the communities over time, the mixture was also added to mesocosms before the inoculation of microbial communities and the nematode introduction to the mesocosms. These eight reference strains were chosen from the pilot experiment as representatives of the seven *Ochrobactrum* (references 1, 1', 2, 3, and 4) and *Pseudochrobactrum* (references 5, 6, and 7) largest clades, i.e. containing more than 5 isolates, identified in the *gyrA* phylogeny of the pilot experiment (**Figure 4.8B**). They were sequenced using long-read sequencing with Oxford Nanopore Technologies and complemented with short reads for high-quality assemblies (**Table 4.S3**). Numerous single colonies were isolated at the end of the evolution experiment to maximize the opportunity to detect HGT events, in real-time. Illumina short-read sequencing was performed on end-point isolated genomes. Short reads were assembled and high-

quality genomes were retained for further analysis (**Table 4.S4**). Respectively for both replicate experiments, 103 and 92 genomes were retained for further analysis (**Table 4.S4**).

4.6.2 Abundance and diversity of *Ochrobactrum* and *Pseudchrobactrum* strains in isolated genomes

To verify the presence of *Ochrobactrum* and *Pseudochrobactrum* strains in the isolated population of genomes, an overview of the abundance and diversity of these genera within all isolated genomes and in relation to the eight reference genomes was first depicted. For this, a phylogenetic tree was reconstructed from the alignment of *gyrA* sequences extracted from both annotated genomes of end-point and reference isolates.

Respectively for both experiments, 120 and 123 *gyrA* sequences were retrieved and aligned. This number accounts for duplicate *gyrA* sequences within single isolates. Once again and in both experiments, the predominance of *Ochrobactrum* (53/120, 44% and 50/92, 54%) and notable presence of *Pseudochrobactrum* (12/120, 10% and 8/92, 9%) in the isolated population of genomes was striking (in total 65/120, 54% and 50/92, 54%). Also consistent with results from the pilot experiment, the diversity of *gyrA* sequences found in those isolates was extensive, indicating substantial strain diversity (**Figure 4.S3 and Table 4.S4**).

4.6.3 Identification of closely related *Ochrobactrum* and *Pseudochrobactrum* strains to reference genomes

To identify which strains are most closely related to the reference genomes, a specific focus was then made on *gyrA* sequences identical to those of the reference genomes. In both experiments, at least one and often multiple isolates with *gyrA* sequences identical to a reference were recovered, except for references 3 and 6 in the second experiment. The number of closely related strains differed depending on the reference, with references 1 and 1' having the highest numbers of close relatives (24 in each experiment). Reference 2 had 13 and 17 closely related isolates in the two experiments, respectively. The remaining references had between one and six closely related isolates (**Figure 4.12**).

Genetic distances were measured between the strains of a given *gyrA* sequence and their respective references. Briefly, reads from each genome were mapped to their respective reference assembly, and SNP calling was performed. Isolates that were less than 40 SNPs apart from their respective reference were considered close relatives to their respective references. Although references 1 and 1' had identical *gyrA* sequences, two separate sets of strains were closely related to each of these

76

references (**Table 4.S5**). The number of SNPs found when comparing isolated genomes to reference 1 was three to four. However, when aligning isolated genomes to reference 1', the number of SNPs ranged from 14 to 19. Genomes related to reference 2 were found to be about 30-40 SNPs apart from their reference in both experiments and may therefore not originate from the same lineage. Interestingly, in several instances, very closely related genomes could be recovered at the end of the experiment, with a number of SNPs below ten, and a few with one or no SNPs (**Figure 4.12**). These results collectively show that lineages can easily be followed in the nematode gut through single-genome isolation. For downstream analysis, genomes with a low number of SNPs, i.e. less than 20, will be assumed to be direct descendants from the initially sequenced isolates, and MGE acquisitions occurring during the scope of our experiment will be investigated.

A
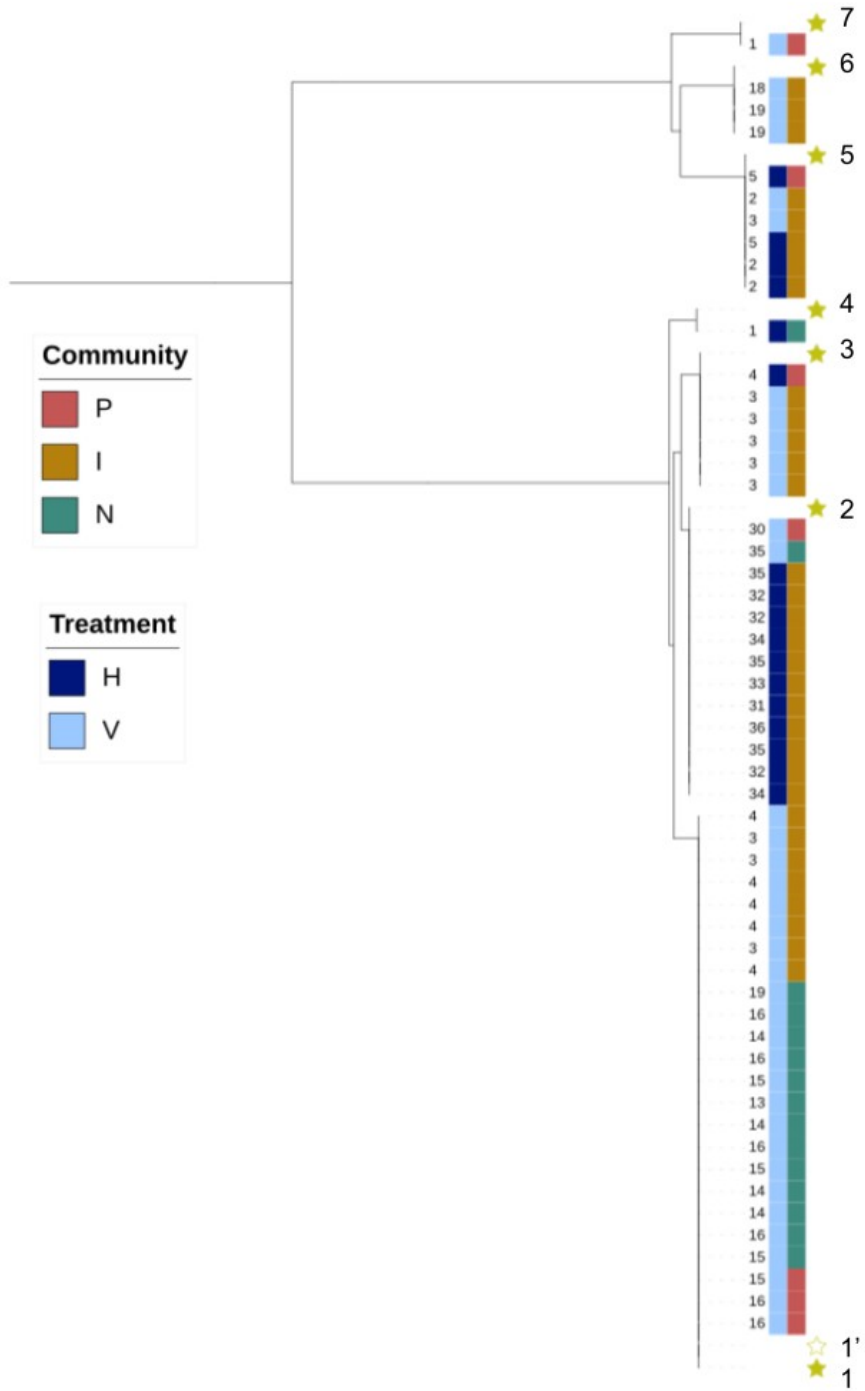
B
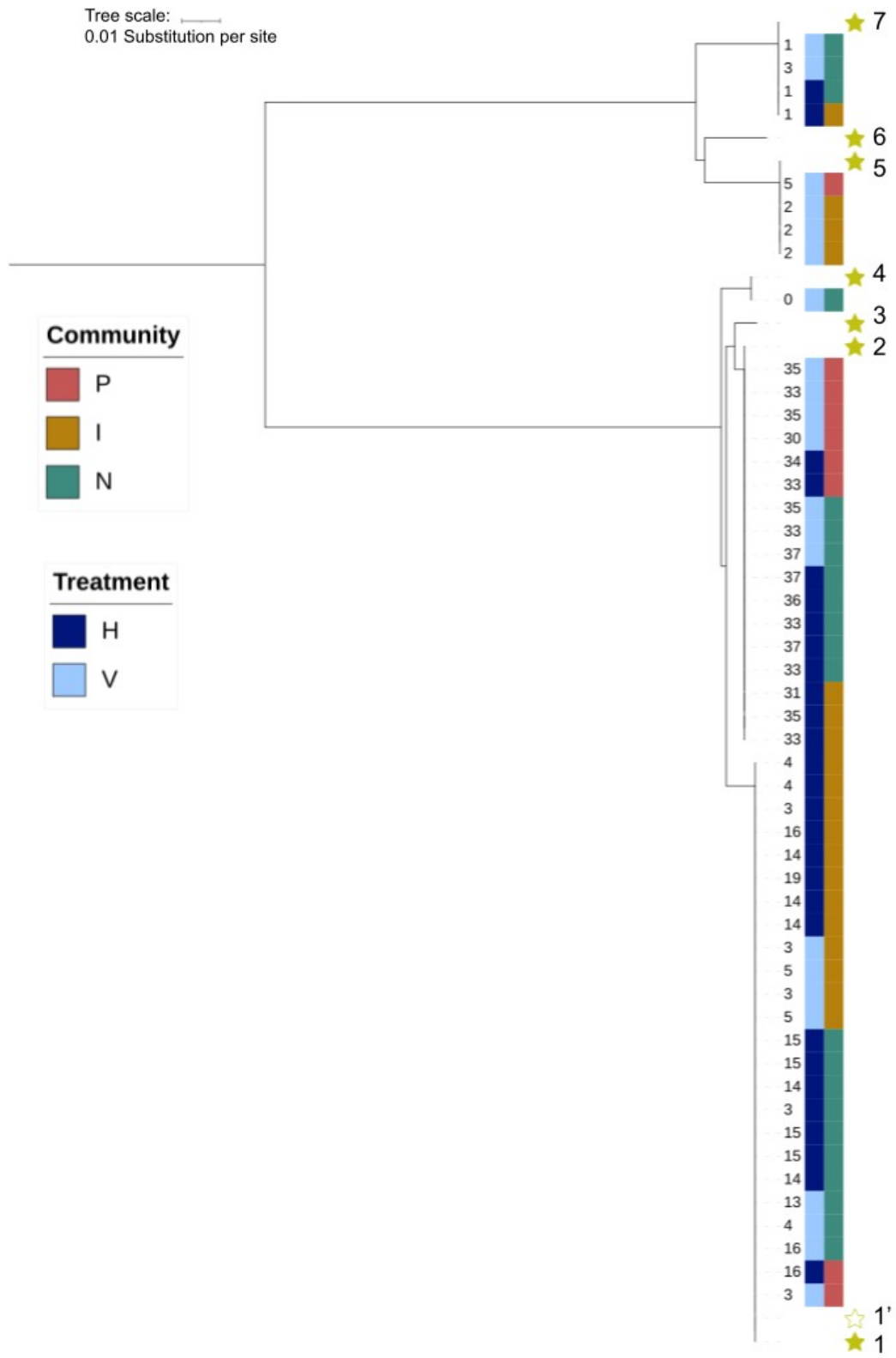
Figure 4.12. *Ochrobactrum* and *Pseudochrobactrum* end-point isolated genotypes are closely related to their respective references. For replicate experiments 1 (A) and 2 (B) respectively: *gyrA* neighbor-joining phylogeny (Tamura-Nei distance method) reconstructed from multiple nucleotide sequence alignment using Clustal omega of *gyrA* sequences from isolated genomes with a *gyrA* sequence identical to one of the reference genomes. Yellow stars indicate the reference isolates. For each clade defined by a given *gyrA* sequence, isolate query reads were mapped to their assembled reference contigs, and SNPs were called. The numbered labels represent the number of SNPs identified for each query isolate when mapped to their reference assembly. The bottom clade contains two sets of genomes, each of them mapping to a given reference (for both trees, empty star: strains with 3-4 SNPs, plain star: strains with 13-19 SNPs).

4.6.4 Investigating real-time HGT within *Ochrobactrum* and *Pseudochrobactrum* genomes

Intending to detect HGT events within *Ochrobactrum* and *Pseudochrobactrum* genomes isolated at the end of the experiment, for every given aforementioned *gyrA* clade related to a given reference, the reads of each genome were mapped to their respective reference assembly, and the unmapped reads were extracted and assembled into contigs. These contigs are therefore uniquely found in these genomes in comparison to their reference. Assembled contigs that had a size above 10 000 bp were retained as candidate transferred sequences for detailed analysis (**Table 4.S5**).

4.6.4.1 *Four candidate horizontally transferred sequences in the first replicate experiment*

In the first replicate experiment, four sequences of interest were identified. Each of these four sequences was only found in a given community under a given treatment. Therefore, it was not possible to suggest anything about their movement between communities. However, and importantly, all presented representative features of one or multiple MGEs. A description of the content of these sequences is detailed here.

4.6.4.1.1 First candidate

The first candidate was a 46,342 bp-long sequence acquired by strains isolated from community P of vertical origin (**Figure 4.13Aa**). Interestingly, while isolated from only 3 isolates, one of those differed from the other by 1 SNP. This means that at least two different lineages acquired this sequence. Evidence for a 13920 bp-long prophage insertion within this sequence was highlighted using the online prophage annotation tool Phaster[70]. Indeed, this region harbors the site-specific recombination sites *attL* and *attR* that are generated upon the integration of temperate phages into bacterial genomes. This is known to occur following recombination between the *attP* site of temperate phages and the *attB* bacterial site[124]. Besides, the prophage sequence harbors a tyrosine recombinase (YR). It is indeed well

known that YRs are used by MGEs, such as Repetitive Extragenic Palindromic Inverted (REPIN) sequences or bacteriophages, to integrate into bacterial genomes[125–127]. In addition, the sequence was confidently attributed to both a dsDNA phage using VirSorter2 and a plasmid using PlasClass[66,67]. Unfortunately, of the multiple ORFs found in this contig, nearly none had a functional annotation, making it challenging to associate a putative effect of the presence of this MGE on the host bacteria.

4.6.4.1.2 Second candidate

The second assembled contig was over 102356-bp long (**Figure 4.13Ab**). This sequence was found in eight isolates representative of two different lineages that were all recovered from the community I under vertical treatment. Similarly to the first contig, this sequence had both characteristics of a phage and a plasmid. It included a recombinase, known to be involved in homologous recombination. Moreover, one ORF was also annotated as a topoisomerase, also known to play roles in recombination[128]. Among the functionally annotated ORFs was a flagellin-encoded gene (fliC), which shares homologies to type three secretion systems[129]. It is noteworthy that this gene, in addition to its role in motility, has been shown to be involved in the translocation of virulence factors[130].

4.6.4.1.3 Third candidate

The third candidate contig was found in a single isolate from community P under horizontal treatment (**Figure 4.13Ac**). Both prophage and ICE regions were detected respectively using Phaster and ICEfinder online tools[70,71]. Besides, this contig contains multiple ISs, that are small transposable elements and therefore MGEs of their own[131]. Remarkably, the gene *oppF*, which was previously identified among xenotypic sequences, was also found to be present in this contig, providing additional evidence for the transfer of functionally important genes[91,132]. Interestingly, this contig shares similar genes with the second candidate. For that reason, the relatedness of those two candidates was investigated by performing a Multiple Alignment using Fast Fourier Transform (MAFT) (**Figure 4.14**). Results show an overall high percentage of identity and a lot of gene synteny between the two sequences upstream of the MGEs. This confirms the high relatedness of the studied genomes. However, the quality of the alignment sharply decreases from the first IS and spanning part though not all of the regions identified as prophage and ICE. These observations strengthen the evidence for the recent acquisition of these MGEs within this genome.

4.6.4.1.4　Fourth candidate

The fourth and last contig found to have been recently acquired was located in community I under vertical treatment within three isolates representative of two different genotypes (**Figure 4.13Ad**). The sequence was neither annotated as phage nor as plasmids or ICEs. However, two recombinases were annotated as such, including again a YR. Interestingly, this contig carried a lot of annotated ORFs. Notably, the cytochrome bd-I ubiquinol oxidase subunit 1 (CydA) and 2 (CydB) is thought to be involved in the aerobic respiratory chain under low oxygen level, which are relevant environmental conditions within the gut[133]. In the intestine, oxygen exists along a gradient and can therefore reach extremely low values[134]. Similar morphologies are shared between the mammalian and nematode gut[135]. Therefore, the horizontal transmission of CydA and CydB via this MGE could be giving a beneficial advantage to this *Ochrobactrum* genome in the gut. These genes were not found in the reference genome, providing further evidence for actual MGE-driven gene gain. Moreover, this complex is also thought to be involved in nitrogen fixation, also providing another potential beneficial trait to the host bacteria[136]. In fact, in this experiment, *C. elegans* nematodes were grown under nitrogen-limited conditions. Besides, a nitrogen fixation protein, fixK, was also found to be carried by this recently acquired contig, and also not found in the reference genome.

4.6.4.2　*Two candidate horizontally transferred sequences in the second replicate experiment*

In the second replicate experiment, two contigs were of interest. In this experiment, candidates were also found in a single community under a given treatment (**Figure 4.13B**). The first candidate contig was inserted in two different sets of strains, each mapping well to a given reference. This suggests that this contig might have been acquired by different and thus more distantly related strains. This sequence was identical to the second candidate of the first experiment. While found in the community I under V treatment in the latter experiment, here the contig was isolated from the I community under H treatment (**Figure 4.13Ba**). It is thus possible that this MGE moved from the V to the H community. However, a PCR of this sequence would be necessary to verify that this MGE was not originally present in the H community. The second candidate was a 41978-bp long sequence with once again, hallmarks of dsDNA phage and plasmid. Very few ORFs were annotated on this sequence (**Figure 4.13Bb**).
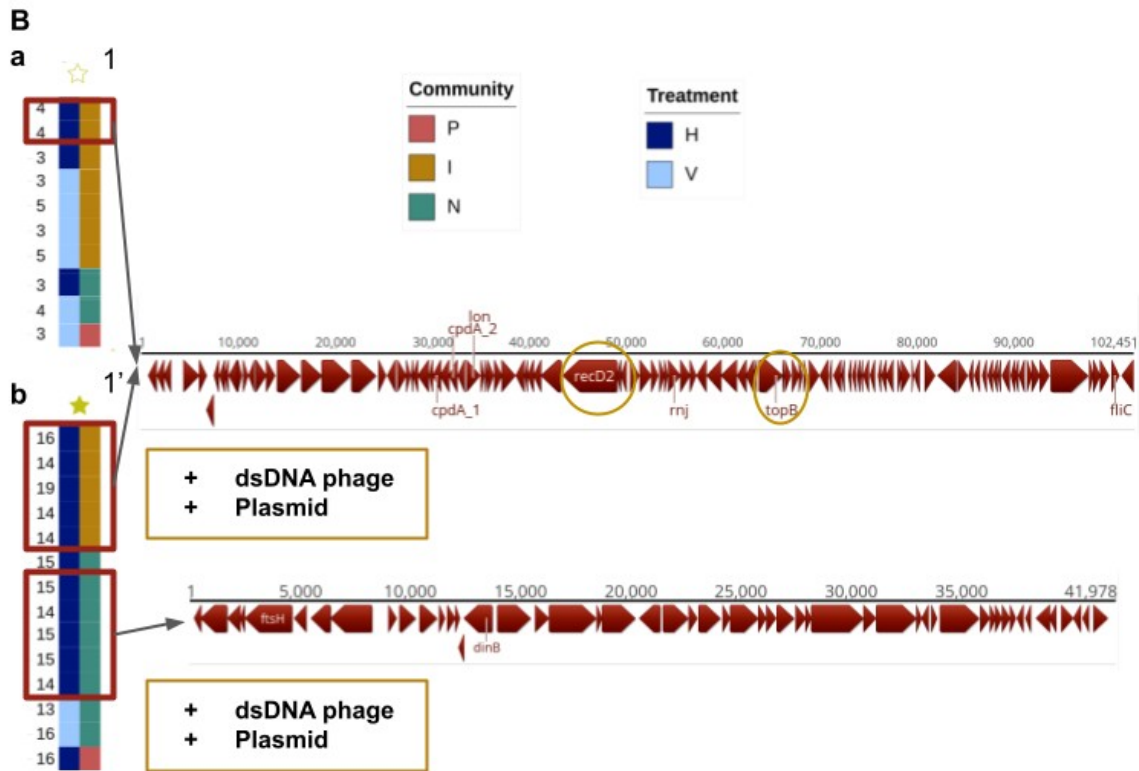
Figure 4.13. Candidate horizontally transferred sequences within *Ochrobactrum* and *Pseudochrobactrum* genera. (A) and (B) Sequences respectively found in each replicate experiment. Short reads of end-point isolated genomes were mapped on to their most closely related reference assembly, SNPs were called and unmapped reads assembled. Longest contigs (> 10 000 bp) are presented here (a-d and a-b respectively for both replicate experiments). Left: *gyrA* clade for each of the associated numbered reference with corresponding number of SNPs (see Figure 4.8). Red squares represent the isolates that acquired this sequence. Right: Sequence annotated for ORFs using Prokka (red) and for MGEs (yellow): viruses (VirSorter2), prophages (Phaster), ICE (ICEfinder), plasmids (PlasClass, score > 0.5). Putative hallmarks of MGEs are surrounded in yellow.
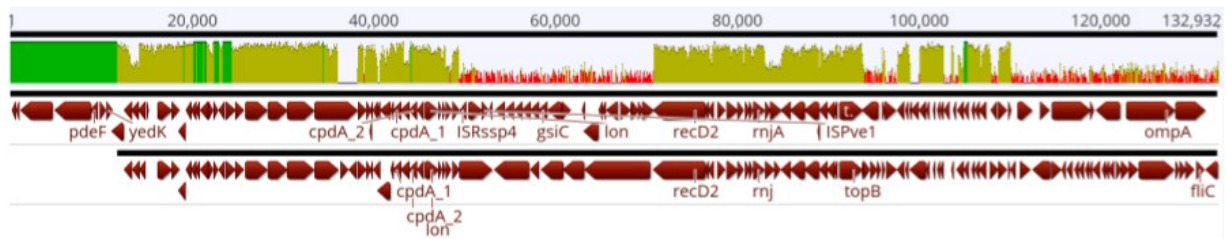
Figure 4.14. Two horizontally acquired sequences within *Ochrobactrum* genomes mostly differ from the presence of certain MGEs. Pairwise alignment of the sequences of second and third of interest in the first replicate experiment using Multiple Alignment using Fast Fourier Transform (MAFT).

85

4.7   Discussion

4.7.1   Positive and negative effects of HGT within the nematode gut microbiome

Both in the pilot and the subsequently performed evolution experiments, the addition of MGEs to communities changed the fate of the nematode populations. While in the pilot, this treatment reduced nematode population size after six weeks in all biological samples, the two subsequent evolution experiments showed an increase in nematode fitness in one community, namely I, while no evident and consistent trends were seen in the two other communities, namely P and N (**Figure 4.1A**). However, when microbiomes extracted from these communities were fed to naïve gnotobiotic nematodes, direct evidence for a drastic loss of fitness for nematodes fed on the horizontal microbiome was highlighted in the latter two communities (**Figure 4.1B**). Moreover, differences were observed between communities in terms of beta diversity, reinforcing the idea that gene movement is likely to differentially affect communities. As *C. elegans* fitness is correlated with its microbiome composition, the differential effect of gene movement on communities is likely to be a reflection of the differences observed in nematode fitness[57,58].

Although it appears that manipulating compost communities with MGEs can have both positive and negative effects on nematode fitness, the transplant experiments suggest the effects on the gut microbiome are mostly either neutral or deleterious. It is well known that virulence genes are often encoded on all types of MGEs[137–139]. Besides, the *C. elegans* microbiome is solely horizontally transmitted, meaning that nematodes have to get colonized by environmental microorganisms rather than receive them from their parents. In fact, horizontal transmission of microorganisms has repeatedly been predicted and shown to be associated with virulence[39,109,140]. This is intuitive because the fate of the microbial community becomes sufficiently unlinked to that of its host. The same intuition holds for MGEs, which horizontally transfer between microbial cells. In this light, the dissemination of MGEs can have both positive and negative effects on the nematodes. Firstly, by transmitting virulence factors to the microbial population, the microbes are benefited to the detriment of the nematodes. However, because MGEs are horizontally transmitted, they may instead have negative effects on the microbes and may be able to protect against such exploitations. Multiple levels of selection are thus simultaneously at play here.

Complicating things even further, the direction of microbial fitness change induced by horizontally transferred genes can also vary in time. For example, virulence factors may at first benefit the bacteria, but if it kills all the host nematodes, then it will eventually be outcompeted by other non-

pathogenic types[138]. In the current study, nematode fitness changed over time, and this may have partly arisen as a consequence of such trade-offs between microbial virulence and transmission. This gives a preview of the complexity of interactions that may exist in nature. Further study was thus pursued to gain insights into the causes of changes in nematode fitness

4.7.2 Causes for changes in nematode fitness upon MGEs addition

4.7.2.1 Microbiome composition

Gut microbiome composition was unaffected by gene movement for any given community. A possible explanation for the absence of differences observed resides in the high diversity of the communities. In fact, high alpha diversity is known to positively influence community resistance to disturbances[141]. In the current study, initial communities grown on cellulose paper were incredibly diverse, and this diversity was maintained during the experiment (**Figure 4.S4**). In the nematode gut, alpha diversity remained rather high, despite the restricted environmental conditions and the consideration of a taxonomic level higher than species, i.e. genus (**Figure 4.2**). Nevertheless, microbial community function has been repeatedly shown to be determined by gene content, and not composition[117–119]. Therefore, the investigation was narrowed to the gene content of microbiome components to identify potential causes for changes in nematode fitness.

4.7.2.2 The gene content of the nematode gut microbiome

4.7.2.2.1 Focus on genera of interest within the gut microbiome

4.7.2.2.1.1 *Ochrobactrum* and *Pseudochrobactrum*: promising candidates as vectors of HGT in *C. elegans* gut?

The pilot experiment allowed identifying two dominating and persisting genera within the gut microbiome of *C. elegan*s: *Ochrobactrum* and *Pseudochrobactrum*. The multitude of isolates obtained from the evolution experiments provides a substantial and diverse library, opening up opportunities for diving into further population genetic investigations. These genera were not highly prevalent in the community extracted from the cellulose paper (**Figure 4.S5**). This suggests an added benefit of *Ochrobactrum* and *Pseudochrobactrum* to colonize *C. elegans* in the given environment of the experiment and therefore a potentially relevant role in *C. elegans* fitness. Consistent with this, researchers have found that the assembly of the gut microbiome of nematodes occurs through a

deterministic process, where the community composition of the gut significantly differs from that of their substrate[55,142].

Ochrobactrum has been repeatedly found to induce a neutral or positive response in the host nematode both in the frame of simplified single or two-strain experiments and more complex reconstituted microbiome communities[112,143–145]. Moreover, while interspecies competition is thought to be of great importance in community assembly, the results presented here also contrast with previous findings showing *Ochrobactrum's* low competitive ability within the nematode gut in a two-species interaction framework[142,146]. However, these experiments are mostly based on a single focal strain of *Ochrobactrum* and were carried out in simple rich laboratory environments. These conclusions may thus not be fully transposable to this study, where nematodes grow on complex communities that involve an extensive diversity of *Ochrobactrum* strains. Besides, *Ochrobactrum*, a close relative of Brucella, is a major opportunistic pathogen[147]. It is also thought to interact with *C. elegans* metabolism, notably with its immune system[144,148]. In certain contexts, *Ochrobactrum* could therefore very well be driving the evolution of virulence through HGT.

4.7.2.2.1.2 Successful tracking of candidate reference strains within the nematode gut

The enrichment of both the nematode gut and the microbial communities grown on the paper at the start of the experiment was sufficient to pick up close relatives of the reference strains at the end of the experiment. Genomes less than 20 SNPs apart from the introduced genotypes were assumed to be direct descendants of those. While there is no direct evidence of this, the high error sequencing rates, notably with nanopore, are likely to account for most of these few differences. Interestingly, the number of SNPs in comparison to one of the reference strains (reference five) was found to vary depending on whether the strain was sequenced alone (1 SNP, **Table 4.S5**) or together with the other seven references on a single flow cell (18-19 SNPs, **Figure 4.12,** and **Table 4.S5**). One possible explanation for this observation would be a difference in the quality of the assembly of this genome following multiplexed or not multiplexed sequencing. However, both assemblies were of very good quality as reflected by their low contamination and high completeness (**Table 4.S3**). It is also possible that the indexes were misattributed in the pool of samples, a phenomenon known as index hopping, consequently incorporating sequences from another strain (**Figure 4.S6)**. The high error rate in long-read sequencing and the close relatedness of the strains used in this study might have amplified this issue. This suggests that, in this experiment, there is more confidence in downstream analysis when running strains separately. However, for reasons of time, the decision was made to sequence all eight reference

genomes on the same chip. Despite this choice, owing to these observations, one can learn that the number of found SNPs might in reality be lower. Therefore, for investigating recent MGE acquisition, and potentially occurring during the experiment, genomes that were less than 20 SNPs apart from their reference were reasonably retained. Conducting an evolution experiment using fluorescently tagged inoculated strains would provide the opportunity to confirm this assumption.

This experiment thus provided rather compelling evidence that it is possible to easily recover lineages introduced within the nematode gut after as long as 14 weeks. This finding demonstrates the feasibility of tracking the eco-evolutionary dynamics of individual genomes within a complex microbial community. It highlights the potential to study the intricate interactions and evolutionary changes occurring within the gut microbiome over an extended period.

4.7.2.2.1.3 Direct evidence for horizontally acquired sequences within *Ochrobactrum* and *Pseudochrobactrum* genomes

Not only were the initially inoculated genomes successfully recovered at the end of the experiment, but multiple sequences were horizontally acquired within these same genotypes. This was observed in nearly all of the eight lineages. Furthermore, the results suggest that this phenomenon may have taken place in slightly distinct genotypic backgrounds, as evidenced by the presence of a few detected SNPs in the various backgrounds where given HGT events were observed (**Figure 4.12**). These findings suggest two important aspects: first, these MGEs were inserted multiple times during the course of the experiment, suggestive of a high transfer rate; and second, in addition to the MGEs, certain mutations also emerged. Investigating the phenotypic effects of these mutations and their contribution to community function could be an interesting avenue for future research. These observations once again highlight the possibility of tracking the *in vivo* evolution of microorganisms within the context of complex microbial communities.

Interestingly, many of the identified horizontally transferred sequences contained mixed features of MGEs. Notably, several inserted sequences showed features of both phages and plasmids (**Figure 4.12**). Therefore, these sequences may contain a phage-plasmid, a type of MGE associated with temperate bacteriophages able to act as plasmids by replicating and segregating at the time of division [149–151]. Phage-plasmids have recently come to attention due to their extensive diversity [152]. One sequence harboured features of prophage, ICE, and ISs altogether. MGEs are thought to be able to facilitate each other's insertions into genomes [153]. This may be an example of nested MGEs, where initially present recombinases recruited ISs, which then helped prophage and ICE recombination. Though the order of

appearance of these MGEs is unknown, ISs have been shown to enhance site-specific recombination of other MGEs[153]. Multiple genes involved in the activity of the MGEs themselves were highlighted, implying that these elements might be active. Specifically, YRs, known markers for MGEs detection, were found on several MGEs[154]. Although only a fraction of the ORFs were functionally annotated, ecologically relevant genes involved in oxygen and nitrogen metabolisms, as well as in environmental information processing, were found to be carried by some of these MGEs. This investigation hence confirms the hypothesis that these genomes are subject to HGT and points toward a possible role of MGE-driven HGT in microbial community function, and consequently nematode fitness.

To gain direct evidence for the role of MGEs acquisition by *Ochrobactrum* genomes on nematode fitness, a simplified experimental design could involve applying pooled filtrates from communities grown in the presence of nematodes to a naive nematode population exclusively fed with single *Ochrobactrum* strains. Besides, the transparency of the nematode body allows for visualisation of these bacteria to determine their sub-localisation within the host and gain a better understanding of their interaction within the host.

4.7.2.2.2   Possibilities and challenges of other approaches to tracking HGT in the gut microbiome

Although a lot was gained from the analysis of individual genomes within the nematode gut, single genome isolation is insufficient to draw a direct link between the MGEs discovered in this study within specific genomes and the previously observed changes in nematode fitness. A PCR of these sequences would confirm the movement of these sequences between communities. Nonetheless, many potentially HGT events may be missed using this approach. It should also be noted that the addition of excess *Ochrobactrum* strains in the communities at the start of the main evolution experiment may have given a competitive advantage to *Ochrobactrum*, independently of HGT, in comparison to other types in the community.

4.7.2.2.2.1   Metagenomics and the Xenoseq pipeline

Ideally, a comprehensive analysis of the horizontally transferred genes within the gut microbiome would involve metagenomics, enabling the application of the Xenoseq pipeline to identify newly acquired genes. However, this study has revealed that there are additional challenges to the utilization of metagenomics in host-associated communities.

Preliminary attempts were made to sequence nematode metagenomes by analysing nematodes that were fed either a single known laboratory strain or complex communities. However, the majority of

the sequenced DNA originated from the nematodes themselves, overshadowing the bacterial signal (**Figure 4.3**). While extracting microbiomes from nematodes is relatively straightforward, the main difficulty lies in effectively separating the nematode DNA from the microbial DNA. To address this, whole-genome amplification (WGA) strategies, such as multiple displacement amplification, could be employed to minimize biases in detecting eukaryotic DNA[155]. Nevertheless, limitations of these methods exist, including the potential introduction of biases due to the formation of chimeras, which could impede the detection of HGT.

Another challenge encountered was the consistently low DNA concentrations obtained when extracting DNA from nematodes. These necessitated multiple rounds of PCR amplification to obtain a sufficiently strong signal for sequencing the samples. However, increasing the number of PCR cycles can lead to nonspecific amplification, introducing bias into the sequencing results. Initial attempts to enhance the ratio of bacterial DNA to nematode DNA did not yield favourable outcomes (**Figures 4.3** and **4.S1**). Deep sequencing of the nematodes together with their microbiomes can be a potential solution to compensate for the low DNA concentration.

4.7.2.2.2.2   A barcoded library of a genome of interest

Library barcoding of single bacterial strains is a valuable tool for tracing the evolutionary fate of individual lineages within a given species[156]. Given the existence of a library for the widely studied *Pseudomonas fluorescens* SBW25 laboratory strain, and considering the prevalence of this genus in the natural microbiome of *C. elegans*, selecting this strain was a logical decision to enhance the likelihood of establishing a strong association with the nematode. However, despite the benefits of SBW25 as a food source for nematodes, long-term interaction between the bacteria and the nematodes was not established, making it impractical to utilize the barcoded library in the context of a more complex environment (**Figures 4.4**, **4.5**, **4.6** and **4.7**). Nevertheless, this study successfully established several assays to investigate the *in vivo* colonization of gut bacteria over time, both qualitatively and quantitatively, which can greatly contribute to the field of host-microbe interactions using *C. elegans* as a model system. Furthermore, future experiments could involve the development of a custom-designed barcoded library for the *Ochrobactrum* and *Pseudochrobactrum* genomes, known to be prevailing members of the *C. elegans* gut and here demonstrated to participate in HGT. Such an experiment, beyond the scope of simple fluorescent labelling, could directly assess the influence of HGT acquisition on bacterial fitness and establish potential connections to the fate of the nematode population.

5 Outlook

5.1 Project Background

Microbial communities shape ecosystems, consisting of populations whose fate is influenced by the phenotypes that arise from genes[1]. However, it is important to recognize that evolution does not always follow a linear path. Instead, complex eco-evolutionary feedback loops occur, involving interactions between different levels of selection, from genes to cells, to collectives of cells[33,34]. Hence, it is the collective interplay of these interactions that ultimately shapes ecosystem functioning. Advancements in sequencing technologies have significantly improved the understanding of the composition and functional characteristics of microbial communities in diverse ecosystems[145,157–162]. Nevertheless, it is crucial to recognize that these studies often represent mere snapshots and may not fully capture the dynamic intrinsic to microbial communities in nature[9]. Acknowledging this knowledge gap, recent studies have hinted at the fact that certain processes cannot be fully understood or explained when isolated from the intricate complexity and evolutionary framework to which they are inherently linked. For example, Garud *et al.*[163] revealed the inadequacy of population genetics models in comprehending the evolutionary dynamics of gut-associated microbial communities[163]. HGT is a mechanism that generates substantial genetic diversity and, as a result, is expected to have a significant impact on ecosystem functioning[15,73]. Not only is HGT ubiquitous in nature, but it also yields profound evolutionary consequences[22,24,164,165]. Recently, Quistad *et al.*[37] provided direct evidence of the impact of the dynamic process of HGT on the functioning of free-living microbial communities.

Microbiomes have been widely acknowledged for their significant impact on the health and well-being of host organisms[3,166]. Notably, resident gut microorganisms contribute to vital metabolic reactions, including the process of digestion. Disruptions in the environment can unsettle the symbiotic relationship between eukaryotic hosts and their associated microbes, potentially leading to dysbiosis and the development of diseases[167,168]. However, the precise mechanisms underlying these interactions remain poorly understood. It has been observed that rates of HGT can vary within the gut depending on the type of environment in which humans reside, indicating that HGT may be a significant contributing factor[169]. In the gut environment, where multiple contributing factors are present, HGT is anticipated to have a substantial influence on microbial dynamics[44]. This was demonstrated in a study highlighting that HGT can exceed the impact of individual mutations within this ecosystem[24].

In their study, Quistad *et al.*[37] proposed a novel approach for identifying HGT events in microbial communities. This method involves the detection of unique sequences within communities that received pooled MGEs from other communities (**Figure 1.1**). However, considering the dynamic nature of microbial communities, it is conceivable that certain of these unique sequences were present initially but went undetected. In this thesis, I first employed a streamlined metagenomic pipeline recently designed by van Dijk *et al.*[62] to determine the causes of appearance of unique sequences detected in microbial communities, and further analysed the nature of these foreign sequences (**Chapter 3**). Building upon the findings presented by Quistad *et al.*[37] regarding the impact of HGT on ecosystem functioning, as well as the compelling evidence supporting the role of HGT in shaping microbial interactions within the gut, I conducted an investigation to explore the influence of HGT dynamics on the functioning of microbial communities within the gut of the widely studied *C. elegans* nematode (**Chapter 4**).

5.2   Review of the Chapters

5.2.1   The Xenoseq metagenomics pipeline traces HGT between microbial communities precisely

In **Chapter 3** of my thesis, I integrated the mesocosm-based approach previously designed by Quistad *et al.*[37] with the Xenoseq pipeline developed by van Dijk *et al.*[62] to detect HGT in evolving free-living microbial communities (**Figures 3.1** and **3.2**). Through this analysis, numerous sequences were identified as unique in all tested communities following the introduction of pooled filtrates from other communities. Despite the relatively short time scale of the evolutionary experiment, and the stringent conditions implemented in the pipeline, several of these unique sequences were determined to originate from foreign communities. Remarkably, a certain degree of consistency was observed in the dynamics of these sequences across replicate experiments (**Figure 3.3**). The migration of foreign sequences was actively tracked within all communities, further substantiating their movement, and variations in the dynamics of these sequences were noted (**Figure 3.5**). The evidence supporting the transfer of xenotypic sequences from other communities was reinforced by high-quality BLAST alignments to their respective community of origin, as well as preliminary compositional analysis of the GC content of the source and destination sequences (**Figures 3.4**, **3.6**, and **3.7**). However, the specific mechanism underlying the transfer of these relatively short sequences remains uncertain, although there were indications suggesting the involvement of MGEs (**Figures 3.4** and **3.8**). Consistent with existing literature, the identified sequences encompassed functional gene categories commonly

associated with HGT, indicating their potential role in microbial adaptation (**Figure 3.9**). Subsequently, I embarked on further investigations using the same experimental design to delve into the impact of HGT on community function, with a specific emphasis on host-associated microbial communities.

5.2.2 HGT affects ecosystem functioning in host-associated microbial communities

In **Chapter 4**, the research emphasis first transitioned to investigating the impact of HGT on the functioning of microbial communities associated with the model organism *C. elegans*. The study encompassed both pilot and main evolution experiments, aiming to examine the impact of introducing MGEs from other communities on the host's fitness. The findings demonstrated a noticeable effect on host fitness. Intriguingly though, the direction of fitness change varied strongly across experiments, communities, and over time (**Figure 4.1**). In contrast, the microbial composition within the gut of *C. elegans* remained unchanged following the introduction of MGEs from other communities (**Figure 4.2**). These results provided compelling evidence in support of the hypothesis that HGT, rather than taxonomic changes, plays a significant role in shaping community function[116,117,119,157]. Consequently, the subsequent investigation focused on developing approaches to validate the occurrence of gene transfer within the gut microbiome of *C. elegans* during the duration of the experiment.

5.2.3 Exploration of approaches to follow HGT within host-associated microbial communities

5.2.3.1 *Metagenomics poses challenges in tracking HGT in host-associated microbial communities*

Initially, metagenomics appeared to hold promise for identifying HGT within host-associated communities in the nematode gut, by leveraging the application of the Xenoseq pipeline associated with the experimental approach from Quistad *et al.*[37] (**Figure 3.1**). However, **Chapter 4** then presented specific challenges associated with studying these communities, including limitations in DNA yields during extraction and the predominance of extracted DNA originating from the eukaryotic genome of *C. elegans* (**Figure 4.3**). Consequently, the application of metagenomics in this context proved to be less straightforward compared to studying free-living communities. Therefore, a decision was made to pursue an alternative approach to address the initial question.

5.2.3.2 P. fluorescens *SBW25 is not suited for HGT detection in the nematode gut using a barcoded library*

The focus subsequently shifted to population-level analysis, with the intention to utilize the well-characterized *P. fluorescens* SBW25 strain as a barcoded library within complex microbial communities to track HGT acquisition (**Chapter 4**). For this, the interaction between this bacterial strain and the nematode *C. elegans* was investigated. Interestingly, the presence of the bacteria did not lead to the escape of the nematodes, and the nematodes were even able to obtain the necessary nutrients for reproduction from the bacteria (**Figures 4.4** and **4.5**). However, this particular bacterial strain did not establish a long-term association with the host, rendering the use of the barcoded library unfeasible (**Figures 4.6** and **4.7**).

5.2.3.3 *Dominant gut genera* Ochrobactrum *and* Pseudochrobactrum *likely participate in HGT*

The research of **Chapter 4** direction then pivoted towards testing the possibility of tracking specific lineages within the gut microbiome, leading to the exploration of potential candidates for HGT acquisition. The gut microbiome composition of *C. elegans* in the presence of microbial communities was analysed using 16S metabarcoding. This revealed the presence of two highly diverse and dominant genera: *Ochrobactrum* and *Pseudochrobactrum* (**Figure 4.8**). This discovery prompted a detailed investigation into their potential for HGT. Numerous genomes from these genera were isolated and cultivated, revealing a significant diversity (**Figure 4.8**). Preliminary pangenome analyses provided insights into the evolutionary history of HGT in *Ochrobactrum* and *Pseudochrobactrum* (**Figures 4.9**, **4.10** and **4.11**). As a result, the approach of single-genome tracking of these particular genera was chosen to investigate HGT acquisition in real-time within the nematode gut.

5.2.3.4 *HGT occurred in the nematode gut within the time frame of a short-term evolution experiment*

Based on the compelling evidence of the potential impact of these genera on host health and their propensity for HGT a pilot experiment was conducted to isolate specific genomes for real-time investigation of HGT acquisition. Eight genomes were carefully sequenced using a combination of short-read and long-read sequencing methods to enhance genome completion. These genomes were introduced at the beginning of a novel evolution experiment, and clones obtained at the end of the experiment were subsequently sequenced. Notably, the read-mapping analysis of the isolated genomes revealed a substantial number of highly similar clones to their respective references, indicating that they

likely originated from the same lineage (**Figure 4.12**). This allowed the disentangling of pre-existing SVs in the compost communities from the process of HGT mediated by the experimental manipulation with MGE cocktails. The investigation into HGT acquisition focused on assembling the non-mapping reads of these genomes, leading to the identification of several candidate sequences exhibiting characteristics commonly associated with MGEs (**Figure 4.13**). Additionally, the discovery of two highly related but not identical sequences provided further evidence for recent acquisition (**Figure 4.14**). Notably, these sequences also contained some ecologically relevant genes (**Figure 4.13**). These findings strongly suggest recent gene acquisition by the main genera present within the nematode gut, occurring simultaneously with changes in community function, and thereby support the notion of HGT influencing the functioning of these host-associated microbial communities.

5.3   Future Directions

5.3.1   Assessing the effect of different types of MGEs and HGT mechanisms on community function

The metagenomic analysis of HGT in free-living communities and the study of genomes within the *C. elegans* gut revealed interesting similarities in the types of MGEs detected. Plasmid-like elements were predicted to be predominant in the free-living communities. The specific mechanism underlying the cell contact-independent transfer of these elements, if it indeed exists, remains unknown. (**Figure 3.7**). A newly identified category known as phage-plasmids, which are widespread and diverse across phylogenetic groups, reflecting their longstanding existence, presents an intriguing possibility[152]. Only a few viral DNA sequences were detected in the communities, which could be underpinned by biological reasons, but is potentially a technical artefact of MGE prediction software, as predicting small viral sequences is a known challenge. Nonetheless, the hypothesis that phage-plasmids can prevail in microbial communities gains support from the fact that using the same analytical tools, four out of the five distinct identified candidate transferred sequences in *Ochrobactrum* and *Pseudochrobactrum* genomes displayed features characteristic of both dsDNA phages and plasmids.

Various types of MGEs have distinct impacts on community function due to several factors. Each MGE imposes different costs on its host bacteria. Bacteriophages can have varying impacts on the survival of their hosts, depending on whether they integrate into the genome or remain lytic[170]. A phage-plasmid has been demonstrated to coexist in both its lysogenic and lytic forms in a marine population[95]. In contrast, plasmids impose subtler costs associated with their transfer and maintenance, rarely leading to cell death[171]. As a result, a community that experiences a sudden influx of bacteriophages is likely to

undergo faster compositional changes compared to a community exclusively exposed to plasmids. Because plasmids are generally non-lethal despite being costly, the compositional changes in plasmid-exposed communities may primarily depend on the fitness effects of the carried traits. Phage-plasmids may experience a combination of these costs. In addition, MGEs exhibit varying host ranges. Conjugative plasmids tend to be limited to closely related hosts, while bacteriophages can sometimes invade a wide range of bacterial genomes[172,173]. While the level of specificity of phage-plasmids to their bacterial host remains uncertain, the widespread occurrence and diversity of these MGEs imply that they may exhibit varying degrees of host specificity[99].

In addition to the impact of costs and host range, different MGEs facilitate the transfer of diverse gene categories[139]. Plasmids are widely recognized for their involvement in the dissemination of antibiotic resistance genes, whereas bacteriophages generally do not play a significant role in this process[139,174]. Instead, bacteriophages commonly carry toxin- and virulence genes. These differences in gene transfer have diverse implications for communities. A host-associated microbial community predominantly shaped by bacteriophages would actively propagate pathogenic traits, causing host disease, whereas a community primarily influenced by plasmids would foster bacterial competition, with potential effects on the host that could either be advantageous or detrimental. It is plausible to assume that phage-plasmids would harbour both types of traits. Thus far, they have been shown to harbour antibiotic resistance genes, a trait related to their plasmid nature[175]. While this study identified traits associated with each type of MGE on distinct horizontally transferred sequences in the nematode gut, it is possible that this separation was due to limited number of functionally annotated genes (**Figure 4.13**).

In order to achieve a more comprehensive understanding of the specific impacts of different MGEs on community dynamics and functioning, additional experimental investigations are necessary. Within a similar experimental framework, these investigations could involve sequencing the pooled filtrates to identify the transferred MGEs, or treating communities with selected amplified sequences to evaluate their individual and combined effects on community functioning. Such approaches would provide valuable insights into the role of MGEs in the overall functioning of microbial communities.

5.3.2   Assessing the adaptive nature of HGT in microbial communities and its implications for community function

Another intriguing question that extends beyond the scope of this study, but can be addressed with minor adaptation, is the relative importance of transfer rates versus fitness benefits in the occurrence of successful HGT events and the subsequent impact on ecosystem functioning[176]. In cases where high

transfer rates by selfish elements take precedence over fitness benefits for the host cells, slightly detrimental genes that are not lethal may become amplified within communities. This scenario can lead to misleading conclusions regarding the adaptiveness of certain traits that may not actually provide any advantage in the wider context of the community. Consequently, the influence on ecosystem functioning may vary depending on the underlying mechanism behind successful HGT events.

While Xenoseq has proven effective in identifying the amplification of horizontally transferred genes within recipient communities, when combined with the appropriate experimental design, it does not offer a means to differentiate between the two above-mentioned causes for such amplification. These events were observed to be numerous and occur within a relatively short time frame, with specific spikes occurring at particular times (**Figure 3.2**). The question of whether the mechanisms underlying the spread of HGT are also dynamic presents an intriguing inquiry. The existence of recently integrated MGEs into multiple genomic backgrounds of *Ochrobactrum*, along with the recurring presence of genes associated with the movement of these sequences, strongly suggests a high transfer rate for these genetic elements. Previous studies have demonstrated that high rates of HGT can exceed the impact of selection in a population of *E. coli*. However, these findings also indicated that the presence of beneficial genes on the transferred sequences was essential for driving the recipient lineages toward fixation[177]. On the other hand, researchers have shown that even with low transfer rates, horizontally transferred ecologically relevant genes have the potential to flourish within a population when subjected to the appropriate selection pressure[25]. Shifting back from a population-level perspective to a community-level approach, to further investigate if and when fitness benefits are associated with the dynamics of HGT, one potential approach is to employ a barcoded library of a focal strain within complex communities. In fact, one might expect that under high rates of transfer, multiple lineages would persist, while the presence of highly adaptive traits with low rates of transfer would favour a single type within a population. By monitoring individual lineages at the population level and combining it with the sequencing and competition of isolated genomes, it becomes possible to assess the prevalence of specific variants and gain insights into the frequency at which fitness benefits impact successful HGT dynamics. Integrated with the approach utilized in this study, it would thereby allow to establish direct connections between the underlying causes of HGT and community-level processes, given the use of an appropriate measure of community function.

5.3.3   Simple approaches to studying complexity

Microbial communities in nature are intricate, and the underlying processes that drive their dynamics are often poorly understood due to technical challenges associated with their complexity. However, this study demonstrates that a simple experimental design can provide insights into these processes and their influence on ecosystem functioning. I advocate for the further adoption of such designs that manipulate genetic diversity in microbial ecology and evolution.

A crucial aspect of this approach involves selecting a complex carbon source, such as cellulose, to maintain the diversity of the community while adapting it to the laboratory environment. Polymers such as cellulose, are degraded extracellularly, allowing the resulting breakdown products to serve as a public resource for the community. The degradation of polymers by primary and secondary degraders has been associated with temporal and spatial dynamics within the community, as was previously demonstrated[178]. Consequently, these complex dynamics have the potential to influence community function.

Similar approaches to the one utilized in this study have demonstrated success in other research endeavours, such as the work conducted by Datta *et al*.[178]. In their study, they explored the assembly and interactions of microbial communities on chitin biopolymers, which closely resemble the conditions observed on marine particles. Such systems could be readily utilized to study the dynamics of HGT and its impact on community function[178]. Moreover, the experimental strategy from Datta *et al*.[178] allows for the high-throughput comparison of a multitude of communities. It would provide valuable insights to compare, for instance, the dynamics of communities inhabiting distinct marine environments, such as communities thriving in the sea sub-surface versus those residing in the deep sea. In fact, using different levels of variations between communities could also help understand the barriers to gene flow. The possibilities are vast, spanning from communities thriving in the open ocean to those inhabiting a solitary leaf. Additionally, this design enables the introduction of environmentally relevant selective pressures, such as antibiotics, limited nutrient resources, or the presence of a predator, to investigate how communities respond to environmental changes.

By employing this specific design across multiple ecosystems, it would be possible to develop new theories that elucidate the processes inherent to the complexity and their implications for the ecology and evolution of microbial communities in nature. The insights gained from these investigations would largely help our understanding of microbial community dynamics and their ecological significance.

Unravelling the intricate dynamics that shape complex communities of microorganisms stands as one of the greatest challenges in the field of microbial ecology and evolution. In order to address this question, it is crucial to employ innovative technologies judiciously. Ecosystems need to be carefully simplified for experimental manipulation while ensuring that the intricate interactions resulting from diverse levels of selection among different entities within the system are not artificially compromised. Overall, this thesis contributes to a deeper understanding of the role of HGT in complex communities and thereby paves the way to the broader comprehension of the global dynamics underlying microbial community complexity.

6   Supplementary material

6.1   Figures
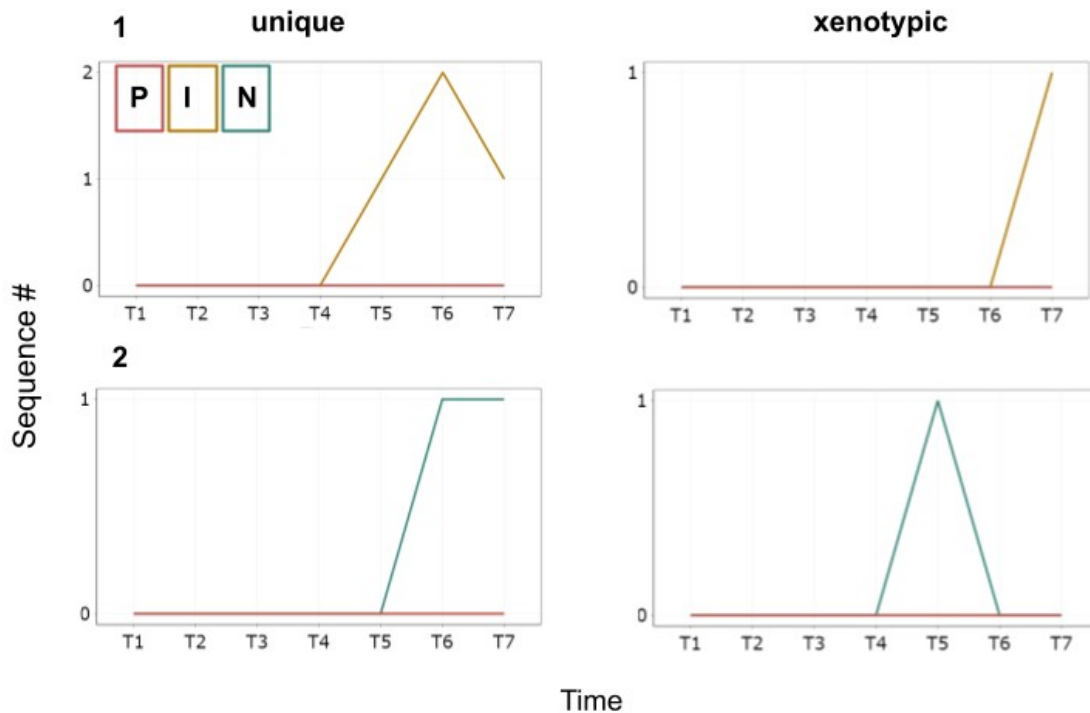
Figure 3.S1. A conservative use of the xenoseq pipeline identified very small numbers of unique and xenotypic sequences within vertical communities. Number of unique (left) and xenotypic (right) sequences identified within the three vertical communities (colours: P, I, and N are shown in red, yellow, and green respectively) across time (T1-T7) in both replicate experiments (top: 1; bottom: 2).

breadth of coverage

1 P

T4 (n = 3) **T5 (n = 2)**

T6 (n = 2)

Time

1 I

T4 (n = 9) **T5 (n = 3)**

T6 (n = 17) **T7 (n = 12)**

breadth of coverage

Time

1 N

T4 (n = 65) T5 (n = 165)

T6 (n = 76) T7 (n = 46)

breadth of coverage

Time

2 P

T4 (n = 18) T5 (n = 10)

breadth of coverage

T6 (n = 3) T7 (n = 6)
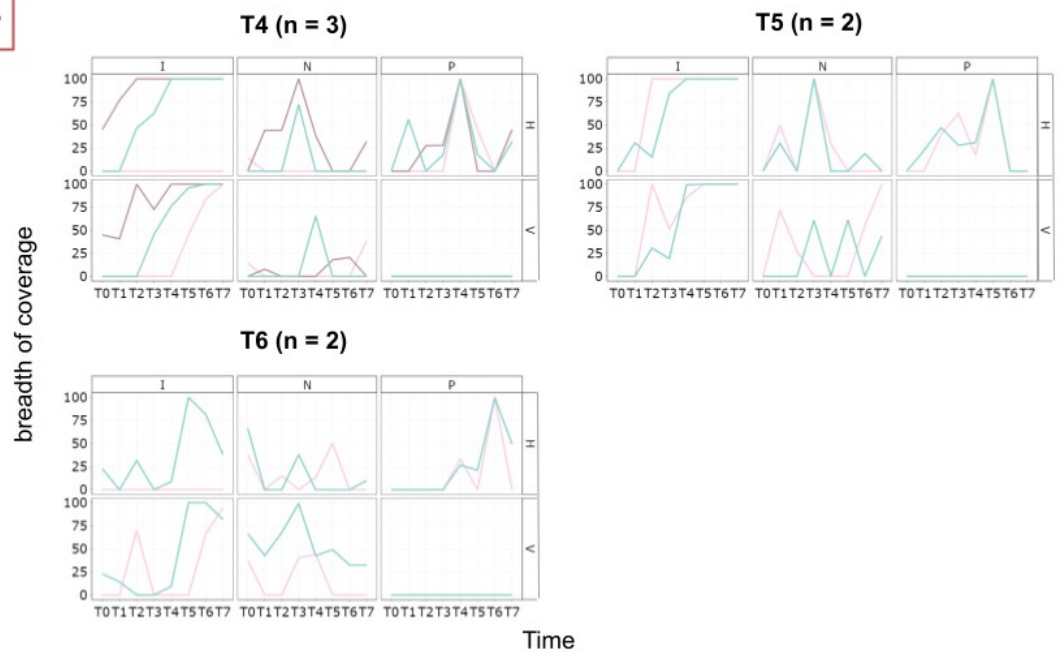
breadth of coverage
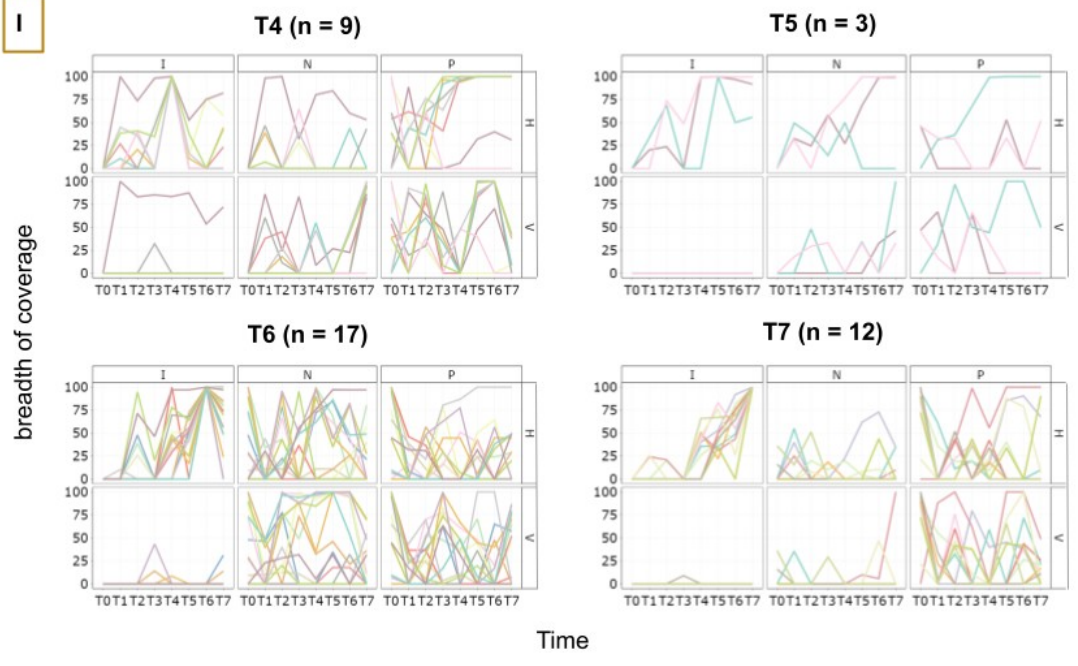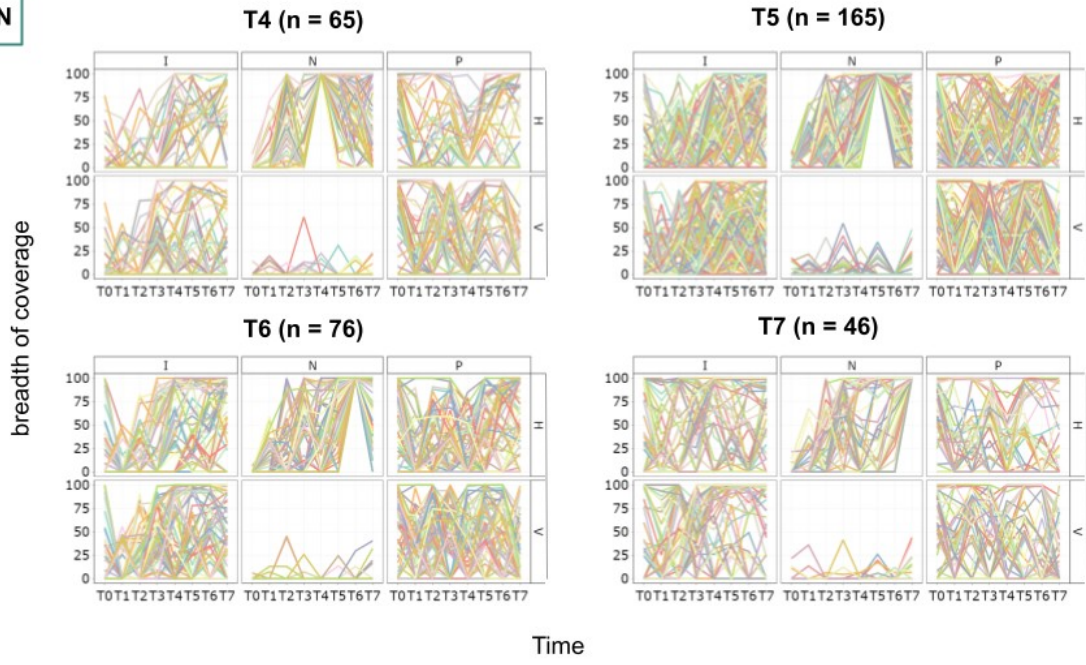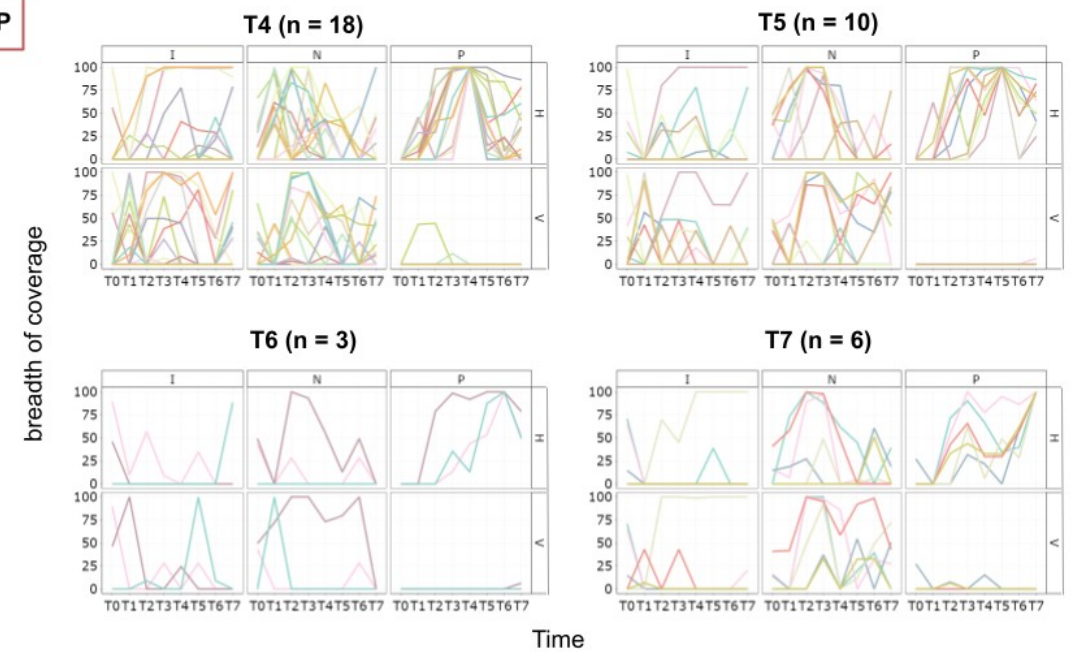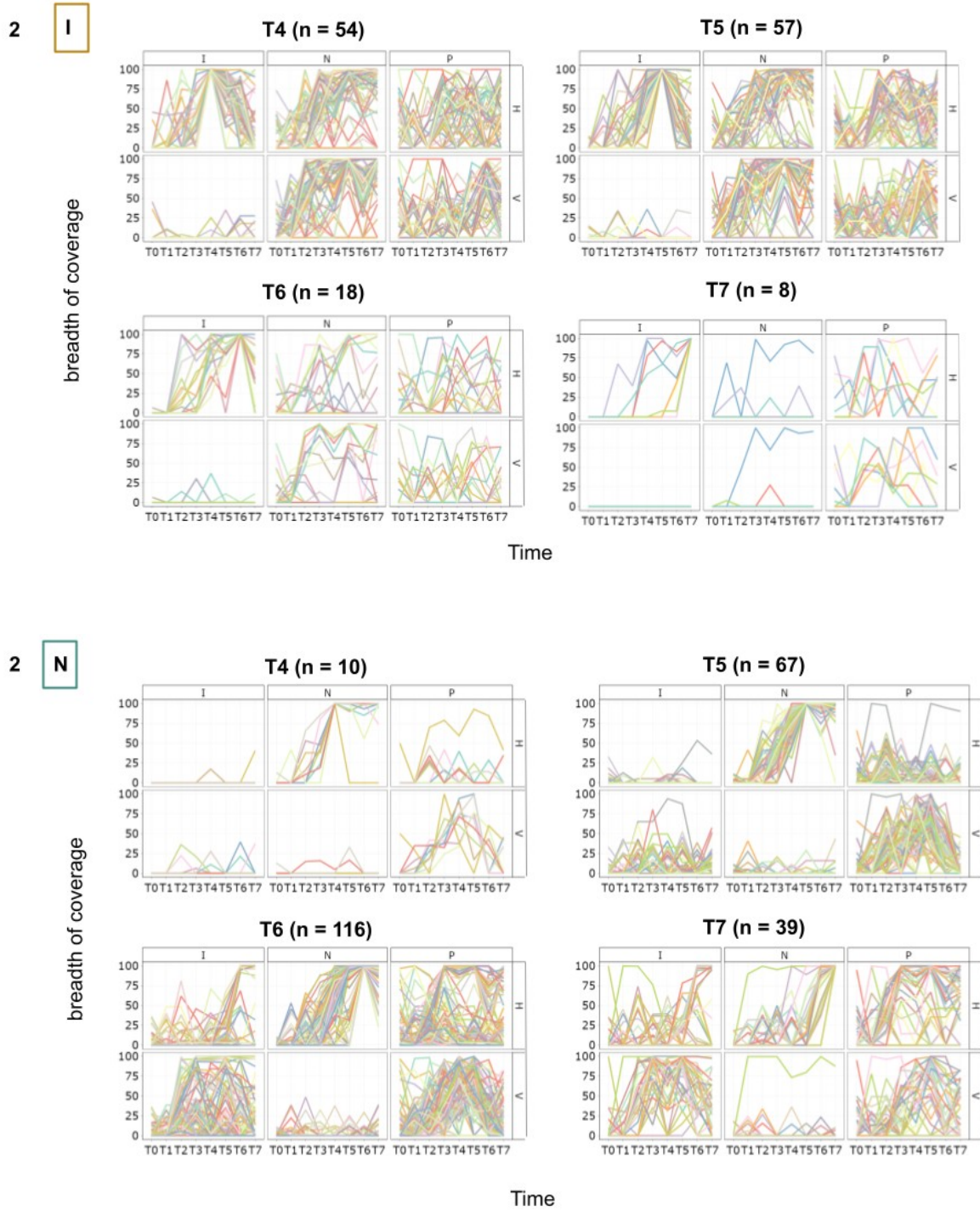
Time

Figure 3.S2. Tracing of all xenotypic sequences across time and communities. Reads from the three communities were mapped onto the xenotypic sequences of each sample. For both experiment replicates (1 and 2), the breadth of coverage of each set (n) of xenotypic sequences (each color represents a different sequence) identified in a given community (P, I, N) at a given time point (T4-T7), across all time points (T0-T7) and communities (P, I, N: column panels and H, V: row panels).

T4 I community

3225 ORFs
6.5% annotated
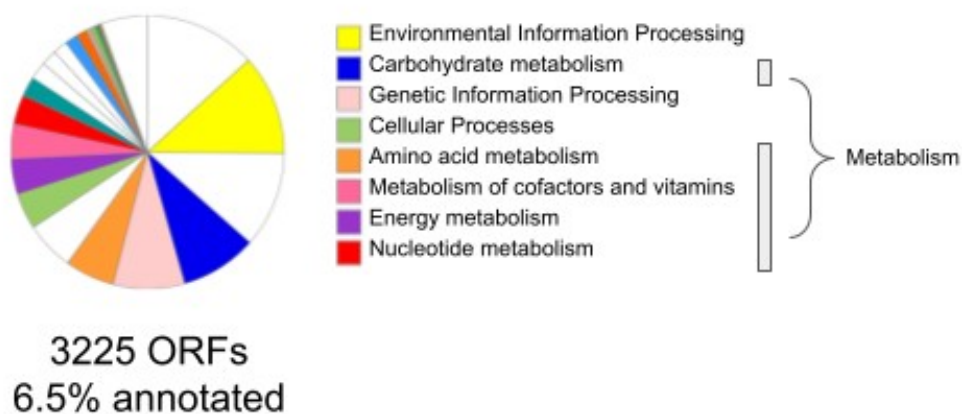
Figure 3.S3. Environmental information processing and metabolism categories may be prevalent in communities. The assembly from community I at time four was annotated using Prokka. The identified ORFs were functionally annotated using the online tool BlastKOALA (KEGG Orthology And Links Annotation). The colours represent the number of functionally annotated ORFs associated to the pathways of the KEGG database.
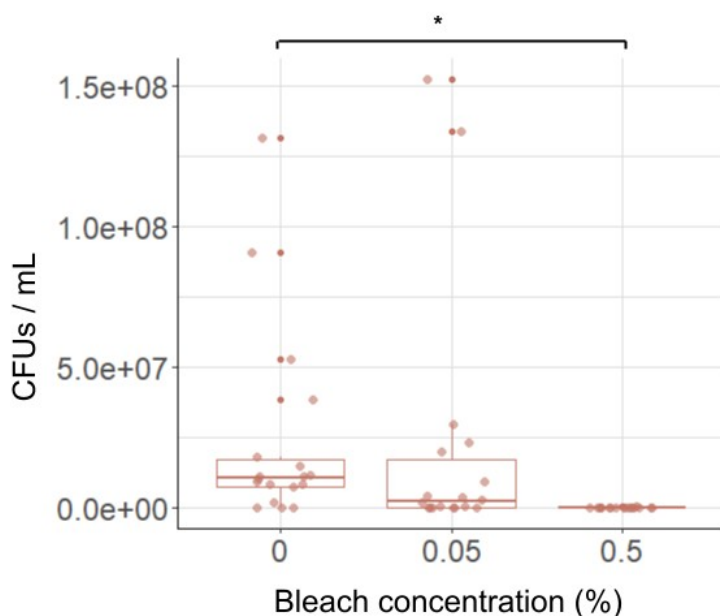
Figure 4.S1. Survival of SBW25 in the presence of different concentrations of bleach. The SBW25-mScarlet strain was cultured overnight, and the culture was then filtered and washed using a filter in a holder. Different concentrations of NaClO diluted in water (0%, 0.05%, 0.5%) were used for the washing step. The filter was collected in tubes containing 1 mL of water, and the contents were vortexed. Dilutions of the solution were then plated and incubated overnight. The boxplot represents the concentration of SBW25 following treatment (CFUs / mL) for the six technical replicates and the three replicate experiments performed. One-tailed

Student t-tests corrected for multiple comparisons indicate statistical significance between samples treated with the different concentrations (*: p.adj < 0.05).
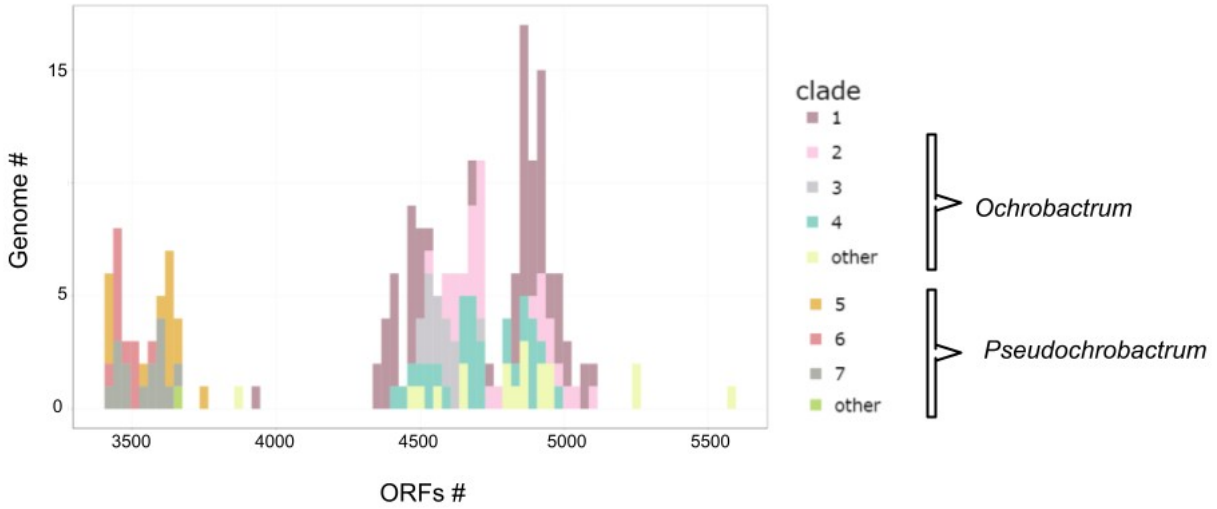
Figure 4.S2. Distribution of the number of ORFs in *Ochrobactrum* and *Pseudochrobactrum* isolated genomes. distribution of the number of ORFs for all isolated *Ochrobactrum* and *Pseudochrobactrum* genera. Colours represent the numbered clades that included at least one of the initially introduced reference genomes.
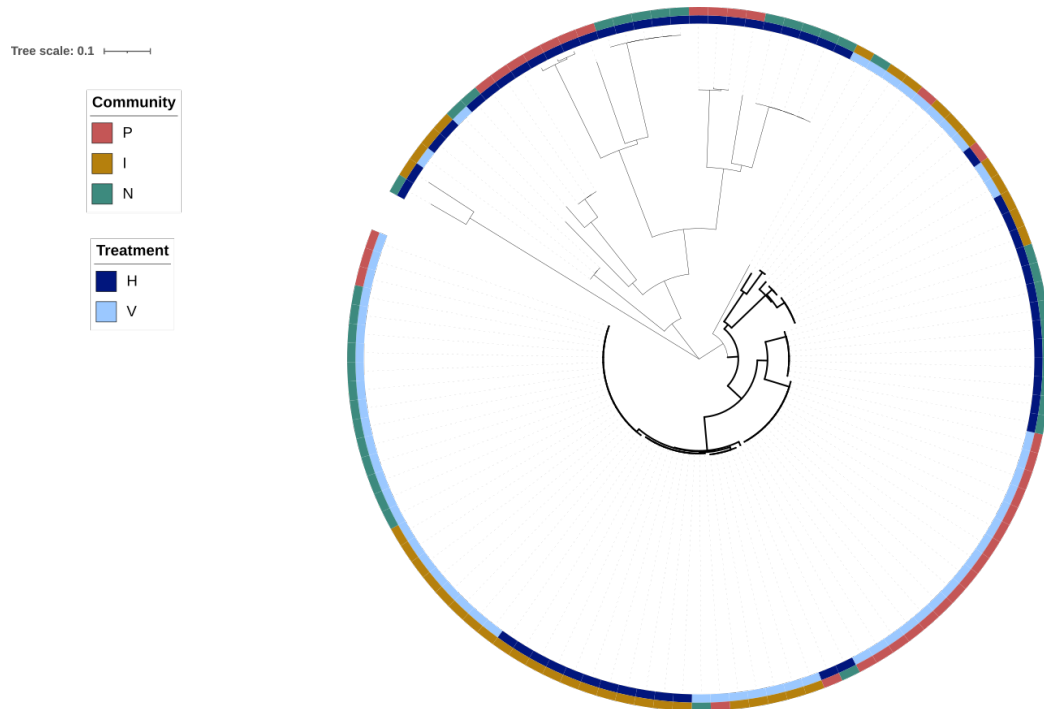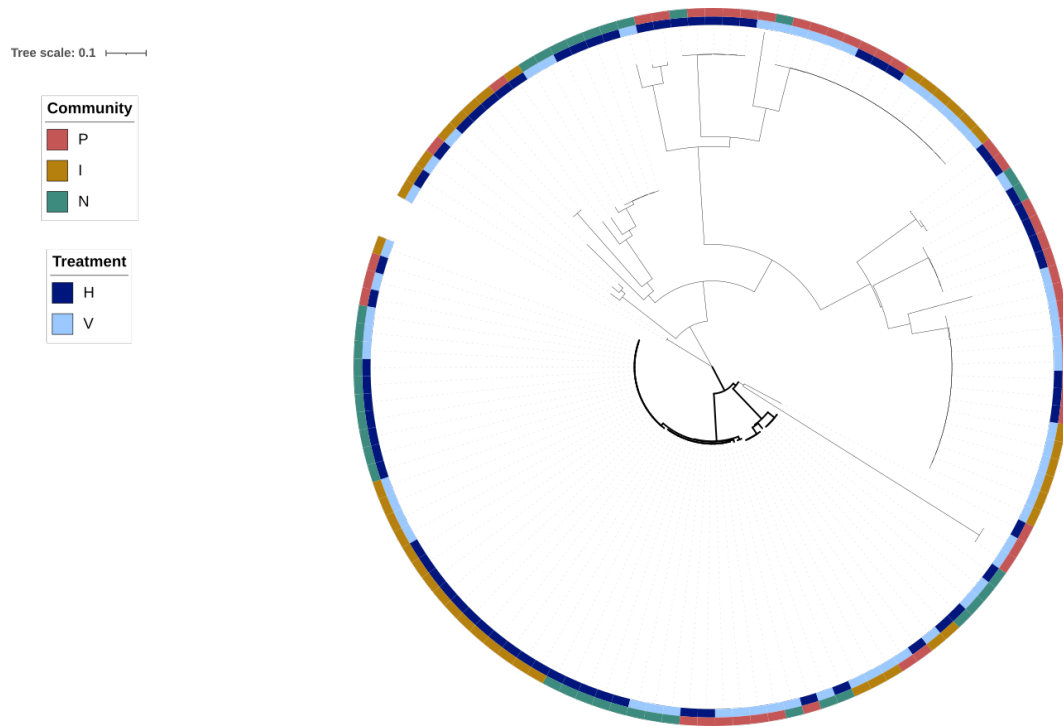
Figure 4.S3. The *gyrA* phylogeny of genomes isolated from *C. elegans* gut at the end of the main evolution experiment. *gyrA* neighbor-joining phylogenies (Tamura-Nei distance method) of isolated genomes reconstructed from multiple sequence alignment using Clustal omega for the two replicates of the main experiment (top and bottom trees, respectively). Outer layer: communities (colours: P, I, and N are shown in red, yellow, and green respectively); inner layer (colours: H and V are shown in dark and light blue respectively): treatment. In bold: *Ochrobactrum* and *Pseudochrobactrum* genera.
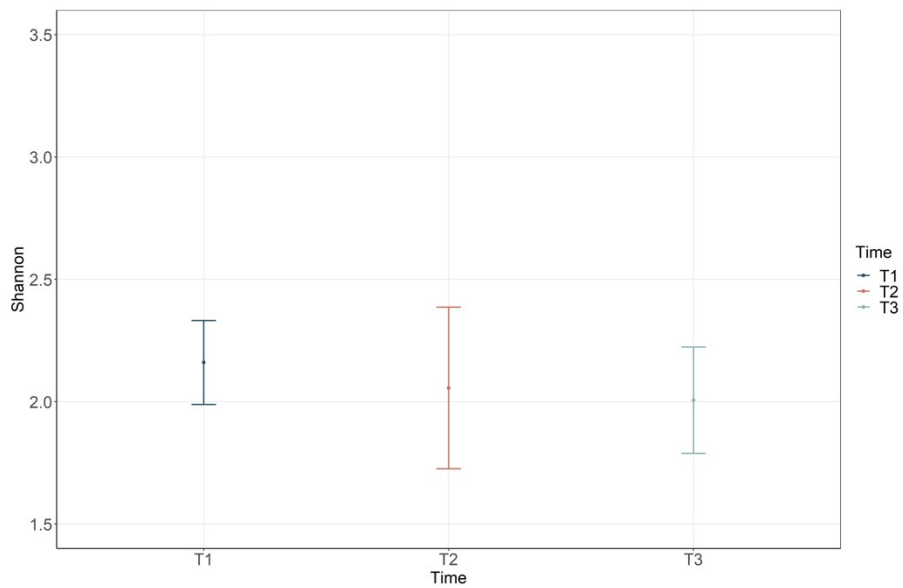
Figure 4.S4. Alpha diversity over time during the pilot evolution experiment. 16S Shannon index as a measure of alpha diversity at the different times of transfer (T1, T2, and T3) during the experiment.
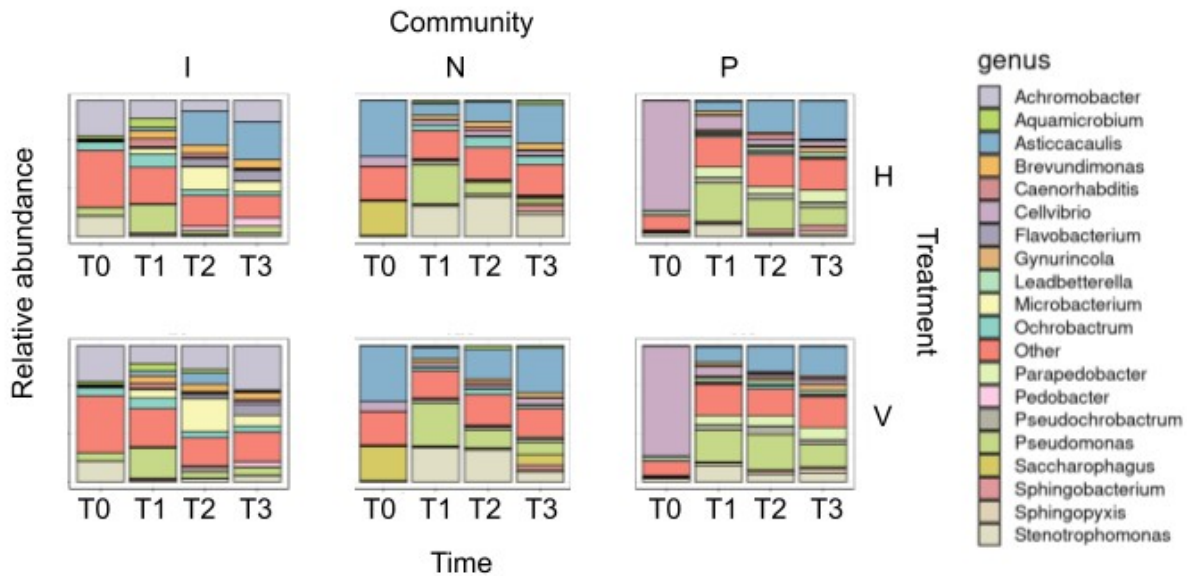
gure 4.S5. *Ochrobactrum* **and** *Pseudochrobactrum* **do not dominate microbial populations outside the gut environment.** Relative abundances of the 20 most abundant genera growing on cellulose were determined using phyloseq tool after taxonomic annotations using Contig Annotation Tool (CAT).
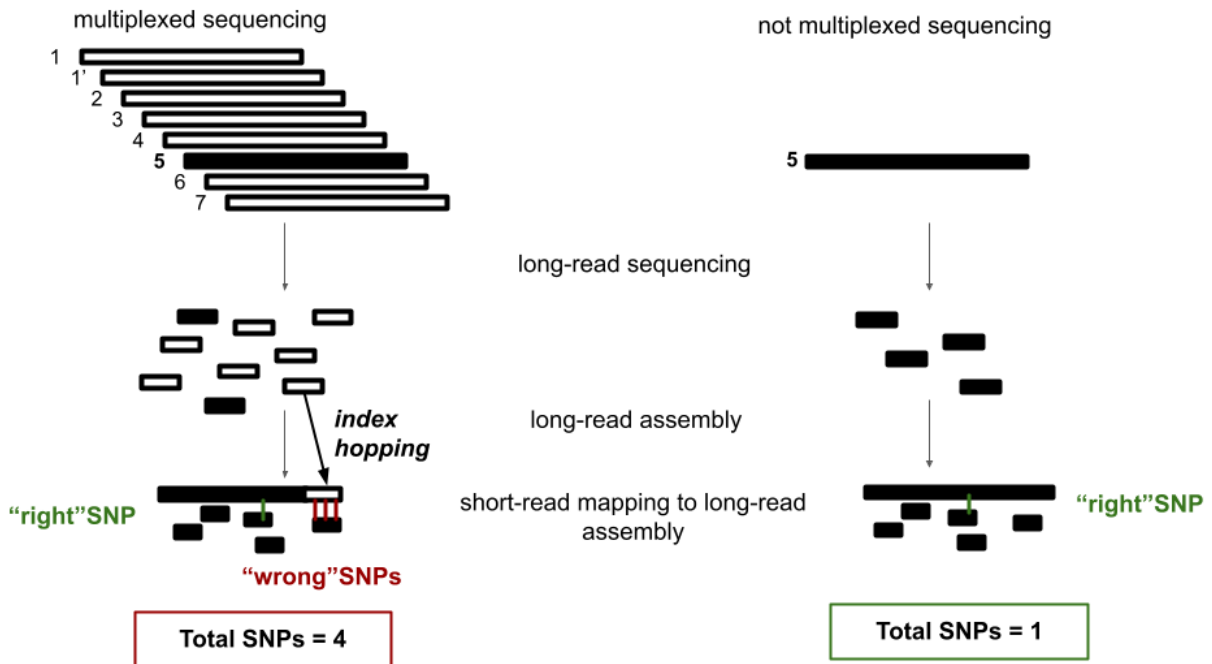
Figure 4.S6. Index hoping in long-read sequencing. When strains are sequenced together on the same nanopore chip (multiplexed sequencing), long reads belonging to two different strains can be mistakenly assembled,

particularly when the strains are closely related. This can lead to misidentifying variations between strains as SNPs, resulting in an overestimation of the number of SNPs compared to individual sequencing where strains are sequenced separately (not multiplexed sequencing).

6.2   Tables

6.2.1   Chapter 3

Table 3.S1. Number of unique and xenotypic sequences. For each experiment (spreadsheets 1 and 2), the sequence Count of a given Type (unique or xenotypic) is given for each Time (T1-T7) and Community (PH, IH, and NH) where the sequences were identified.

Table 3.S2. Number of xenotypic sequences used for the GC content analyses. Summary tables indicate for each experiment replicate (1 or 2) and one (PH, IH, and NH) or all (all) communities, the raw number and percentages of xenotypic sequences that were retained for the comparison with their sequence of origin (column D) and destination (column E), ie those that were shorter than the sequence they were compared with.

Table 3.S3. Summary of genes identified in the xenotypic sequences using Prokka. Each spreadsheet resembles the genes identified in a given community (PH, IH, and NH) and experiment replicates (1 or 2): Name (CDS means that no specific genes were annotated), length (bp), and product description are given in the tables.

Table 3.S4. Summary and functional annotations of genes found within xenotypic sequences. For each replicate experiment (1 and 2) and community (PH, IH and NH), the three levels of annotations of the 6 top levels of pathway maps from the KEGG orthology (KO) database and associated description of the gene.

6.2.2   Tables – Chapter 4

**Table 4.S1. DNA concentrations from *C. elegans* microbiome extractions.** Nematodes were fed for either three days on SBW25 on PFM plates or two weeks on community P on mesocosm plates (Gut bacteria). DNA was extracted using either the NucloSpin Blood and Tissue kit from Macherey&Nagel or the DNeasy Ultraclean Microbial kit from Qiagen (DNA extraction kit) from either 100 or 1000 nematodes (Number of used nematodes) and DNA concentration was measured using the Qubit dsDNA High Sensitivity (HS) dye from ThermoFisher.

Table 4.S2. Summary of Illumina isolate assemblies from the pilot experiment. Summary of illumina short-read assemblies: assessment of completeness and contamination (checkm) as well as the size of the genomes assembled into scaffolds (bbstats.sh) for the 348 isolated genomes. In red are the genomes discarded for further analysis due to low quality (<5%) or poor completion (<98%). The number of ORFs (ORFs #) was extracted using Prokka and *Ochrobactrum* and *Pseudochrobactrum gyrA* clades are numbered based on whether or not they share similar gyrA with one of the references used to track specific genomes during the main evolution experiment. In dark and light blue are the isolates identified as belonging to *Ochrobactrum* and *Pseudochrobactrum* respectively based on a BLAST of their *gyrA* sequence against the NCBI nucleotide database.

Table 4.S3. Summary of hybrid illumina and nanopore isolate reference assemblies from the pilot experiment. Summary of the hybrid illumina short-read and nanopore long-read assemblies for the eight isolates from the pilot experiment used as reference genomes in the main evolution experiment: assessment of completeness and contamination (checkm) as well as the size of the genomes assembled into scaffolds (bbstats.sh) for the eight isolated genomes. Sequencing type indicates whether or not the DNA was sequenced together on the same chip (multiplexed) or not (not multiplexed). In dark and light blue are the isolates identified as belonging to

Ochrobactrum and *Pseudochrobactrum* respectively based on a BLAST of their *gyrA* sequence against the NCBI nucleotide database.

Table 4.S4. Summary of Illumina isolate assemblies from the main evolution experiments. Summary of the illumina short-read assemblies for both replicate experiments (1 and 2 spreadsheets): assessment of completeness and contamination (checkm) as well as the size of the genomes assembled into scaffolds (bbstats.sh) for the 107 (1) and 105 (2) isolated genomes. In red are the genomes discarded for further analysis due to low quality (<5%) or poor completion (<98%). *Ochrobactrum* and *Pseudochrobactrum gyrA* clades are numbered based on whether or not they share similar *gyrA* with one of the references used to track specific genomes during the main evolution experiment. In **dark** and light blue are the isolates identified as belonging to *Ochrobactrum* and *Pseudochrobactrum* respectively based on a BLAST of their *gyrA* sequence against the NCBI nucleotide database.

Table 4.S5. Summary of alignment to their reference of strains isolated at the end of the main experiment. For the two replicate experiments (subsheets 1 and 2), summary of end-point isolates (Query) alignment to their reference (Reference) is shown: Number of SNPs (# SNPs), mapped reads (# mappes reads), and unmapped reads (# unmapped reads), Ratio of unmapped over mappes reads (% unmapped), number of assembled contigs from the unmapped reads (# assembled contigs), the length of the longest contig (maximum contig length), and the candidate contigs retained for analysis (candidate contigs).

7   Bibliography

1.   Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences* **95**, 6578–6583 (1998).
2.   Madsen, E. L. Microorganisms and their roles in fundamental biogeochemical cycles. *Curr Opin Biotechnol* **22**, 456–464 (2011).
3.   Kinross, J. M., Darzi, A. W. & Nicholson, J. K. Gut microbiome-host interactions in health and disease. *Genome Medicine* **3**, 14 (2011).
4.   Větrovský, T. & Baldrian, P. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE* **8**, e57923 (2013).
5.   Garza, D. R. & Dutilh, B. E. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cell Mol Life Sci* **72**, 4287–4308 (2015).
6.   Prosser, J. I. Putting science back into microbial ecology: a question of approach. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190240 (2020).
7.   Prosser, J. I. & Martiny, J. B. H. Conceptual challenges in microbial community ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190241 (2020).
8.   Tecon, R. *et al.* Bridging the Holistic-Reductionist Divide in Microbial Ecology. *mSystems* **4**, (2019).
9.   Rainey, P. B. & Quistad, S. D. Toward a dynamical understanding of microbial communities. 11 (2020).
10.   Benton, T. G., Solan, M., Travis, J. M. J. & Sait, S. M. Microcosm experiments can inform global ecological problems. *Trends in Ecology & Evolution* **22**, 516–521 (2007).
11.   Fraser, L. H. & Keddy, P. The role of experimental microcosms in ecological research. *Trends in Ecology & Evolution* **12**, 478–481 (1997).
12.   Quistad, S. D., Doulcier, G. & Rainey, P. B. *Experimental manipulation of selfish genetic elements links genes to microbial community function*. http://biorxiv.org/lookup/doi/10.1101/608752 (2019) doi:10.1101/608752.
13.   *Prokaryotes and Evolution*. (Springer International Publishing, 2018). doi:10.1007/978-3-319-99784-1.
14.   Muller, H. J. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **1**, 2–9 (1964).
15.   Narra, H. P. & Ochman, H. Of What Use Is Sex to Bacteria? *Current Biology* **16**, R705–R710 (2006).
16.   *Lateral Gene Transfer in Evolution*. (Springer New York, 2013). doi:10.1007/978-1-4614-7780-8.
17.   Hacker, J. & Carniel, E. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* **2**, 376–381 (2001).
18.   Nogueira, T. *et al.* Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Curr Biol* **19**, 1683–1691 (2009).
19.   Rankin, D. J., Rocha, E. P. C. & Brown, S. P. What traits are carried on mobile genetic elements, and why? *Heredity* **106**, 1–10 (2011).
20.   Doolittle, W. F. & Zhaxybayeva, O. On the origin of prokaryotic species. *Genome Res* **19**, 744–756 (2009).
21.   Koonin, E. V. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Res* **5**, F1000 Faculty Rev-1805 (2016).

22. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).

23. Puigbò, P., Lobkovsky, A. E., Kristensen, D. M., Wolf, Y. I. & Koonin, E. V. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC Biology* **12**, 66 (2014).

24. Frazão, N., Sousa, A., Lässig, M. & Gordo, I. Horizontal gene transfer overrides mutation in *Escherichia coli* colonizing the mammalian gut. *Proc Natl Acad Sci USA* **116**, 17906–17915 (2019).

25. Woods, L. C. *et al.* Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *Proceedings of the National Academy of Sciences* **117**, 26868–26875 (2020).

26. Haudiquet, M., de Sousa, J. M., Touchon, M. & Rocha, E. P. C. Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *Philos Trans R Soc Lond B Biol Sci* **377**, 20210234 (2021).

27. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**, 472–482 (2015).

28. Ågren, J. A. & Clark, A. G. Selfish genetic elements. *PLOS Genetics* **14**, e1007700 (2018).

29. Bergstrom, C. T., Lipsitch, M. & Levin, B. R. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**, 1505–1519 (2000).

30. Lili, L. N., Britton, N. F. & Feil, E. J. The persistence of parasitic plasmids. *Genetics* **177**, 399–405 (2007).

31. Slater, F. R., Bailey, M. J., Tett, A. J. & Turner, S. L. Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiology Ecology* **66**, 3–13 (2008).

32. Morris, A., Meyer, K. & Bohannan, B. Linking microbial communities to ecosystem functions: what we can learn from genotype–phenotype mapping in organisms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190244 (2020).

33. Post, D. M. & Palkovacs, E. P. Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play. *Philos Trans R Soc Lond B Biol Sci* **364**, 1629–1640 (2009).

34. Bailey, J. K. *et al.* From genes to ecosystems: an emerging synthesis of eco-evolutionary dynamics. *New Phytologist* **184**, 746–749 (2009).

35. Coyte, K. Z. *et al.* Horizontal gene transfer increases microbiome stability. 2022.02.25.481914 Preprint at https://doi.org/10.1101/2022.02.25.481914 (2022).

36. Fan, Y., Xiao, Y., Momeni, B. & Liu, Y.-Y. Horizontal gene transfer can help maintain the equilibrium of microbial communities. *Journal of Theoretical Biology* **454**, 53–59 (2018).

37. Quistad, S. D., Doulcier, G. & Rainey, P. B. Experimental manipulation of selfish genetic elements links genes to microbial community function. *Philos Trans R Soc Lond B Biol Sci* **375**, (2020).

38. Sagan, L. On the origin of mitosing cells. *Journal of Theoretical Biology* **14**, 225-IN6 (1967).

39. Ewald, P. W. Transmission Modes and Evolution of the Parasitism-Mutualism Continuuma. *Annals of the New York Academy of Sciences* **503**, 295–306 (1987).

40. Hou, K. *et al.* Microbiota in health and diseases. *Sig Transduct Target Ther* **7**, 1–28 (2022).

41. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences* **110**, 3229–3236 (2013).

42. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* **18**, 607–621 (2020).

43. Liu, L. *et al.* The human microbiome: A hot spot of microbial horizontal gene transfer. *Genomics* **100**, 265–270 (2012).

44. Zeng, X. & Lin, J. Factors influencing horizontal gene transfer in the intestine. *Anim. Health Res. Rev.* **18**, 153–159 (2017).

45. Moura De Sousa, J., Lourenço, M. & Gordo, I. Horizontal gene transfer among host-associated microbes. *Cell Host & Microbe* **31**, 513–527 (2023).
46. Portal-Celhay, C., Nehrke, K. & Blaser, M. J. Effect of Caenorhabditis elegans age and genotype on horizontal gene transfer in intestinal bacteria. *FASEB J* **27**, 760–768 (2013).
47. Berg Miller, M. E. *et al.* Phage–bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environmental Microbiology* **14**, 207–227 (2012).
48. Breitbart, M. *et al.* Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *Journal of Bacteriology* **185**, 6220–6223 (2003).
49. Jørgensen, T. S., Xu, Z., Hansen, M. A., Sørensen, S. J. & Hansen, L. H. Hundreds of Circular Novel Plasmids and DNA Elements Identified in a Rat Cecum Metamobilome. *PLOS ONE* **9**, e87924 (2014).
50. Kav, A. B. *et al.* Insights into the bovine rumen plasmidome. *Proc Natl Acad Sci U S A* **109**, 5452–5457 (2012).
51. Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M. & Comstock, L. E. Evidence of extensive DNA transfer between bacteroidales species within the human gut. *mBio* **5**, e01305-01314 (2014).
52. Hehemann, J.-H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
53. Kaletta, T. & Hengartner, M. O. Finding function in novel targets: C. elegans as a model organism. *Nat Rev Drug Discov* **5**, 387–399 (2006).
54. Schulenburg, H. & Felix, M.-A. The Natural Biotic Environment of Caenorhabditis elegans. *Genetics* **206**, 55–86 (2017).
55. Dirksen, P. *et al.* The native microbiome of the nematode Caenorhabditis elegans: gateway to a new host-microbiome model. *BMC Biology* **14**, 38 (2016).
56. Zhang, F. *et al.* Caenorhabditis elegans as a Model for Microbiome Research. *Front Microbiol* **8**, (2017).
57. Samuel, B. S., Rowedder, H., Braendle, C., Félix, M.-A. & Ruvkun, G. Caenorhabditis elegans responses to bacteria from its natural habitats. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3941-3949 (2016).
58. Shapira, M. Host–microbiota interactions in Caenorhabditis elegans and their significance. *Current Opinion in Microbiology* **38**, 142–147 (2017).
59. Sevillya, G., Adato, O. & Snir, S. Detecting horizontal gene transfer: a probabilistic approach. *BMC Genomics* **21**, 106 (2020).
60. Zhang, L. *et al.* Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology* **12**, (2021).
61. Ravenhall, M., Škunca, N., Lassalle, F. & Dessimoz, C. Inferring Horizontal Gene Transfer. *PLoS Comput Biol* **11**, e1004095 (2015).
62. Dijk, B. van *et al.* Identifying and tracking mobile elements in evolving compost communities yields insights into the nanobiome. 2023.02.02.526783 Preprint at https://doi.org/10.1101/2023.02.02.526783 (2023).
63. Rainey, P. B. & Bailey, M. J. Physical and genetic map of the Pseudomonas fluorescens SBW25 chromosome. *Molecular Microbiology* **19**, 521–533 (1996).
64. Stiernagle, T. Maintenance of C. elegans. *WormBook* (2006) doi:10.1895/wormbook.1.101.1.
65. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

66. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

67. Pellow, D., Mizrahi, I. & Shamir, R. PlasClass improves plasmid sequence classification. *PLOS Computational Biology* **16**, e1007781 (2020).

68. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* gkac240 (2022) doi:10.1093/nar/gkac240.

69. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

70. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**, W16-21 (2016).

71. Liu, M. *et al.* ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Research* **47**, D660–D665 (2019).

72. Douglas, G. M. & Langille, M. G. I. Current and Promising Approaches to Identify Horizontal Gene Transfer Events in Metagenomes. *Genome Biol Evol* **11**, 2750–2766 (2019).

73. Takeuchi, N., Kaneko, K. & Koonin, E. V. Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3 (Bethesda)* **4**, 325–339 (2014).

74. Hall, D. H. *WormAtlas*. (Cold Spring Harbor Laboratory Press, 2009).

75. Lawrence, J. G. & Ochman, H. Amelioration of Bacterial Genomes: Rates of Change and Exchange. *J Mol Evol* **44**, 383–397 (1997).

76. Médigue, C., Rouxel, T., Vigier, P., Hénaut, A. & Danchin, A. Evidence for horizontal gene transfer in Escherichia coli speciation. *Journal of Molecular Biology* **222**, 851–856 (1991).

77. Hildebrand, F., Meyer, A. & Eyre-Walker, A. Evidence of Selection upon Genomic GC-Content in Bacteria. *PLOS Genetics* **6**, e1001107 (2010).

78. Norman, A., Hansen, L. H. & Sørensen, S. J. Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci* **364**, 2275–2289 (2009).

79. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brüssow, H. Phage as agents of lateral gene transfer. *Curr Opin Microbiol* **6**, 417–424 (2003).

80. Székely, A. J. & Breitbart, M. Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiology Letters* **363**, fnw027 (2016).

81. Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**, 722–732 (2005).

82. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A* **95**, 6239–6244 (1998).

83. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801–3806 (1999).

84. Cohen, O., Gophna, U. & Pupko, T. The Complexity Hypothesis Revisited: Connectivity Rather Than Function Constitutes a Barrier to Horizontal Gene Transfer. *Molecular Biology and Evolution* **28**, 1481–1489 (2011).

85. Lercher, M. J. & Pál, C. Integration of Horizontally Transferred Genes into Regulatory Interaction Networks Takes Many Million Years. *Molecular Biology and Evolution* **25**, 559–567 (2008).

86. Wellner, A. & Gophna, U. Neutrality of Foreign Complex Subunits in an Experimental Model of Lateral Gene Transfer. *Molecular Biology and Evolution* **25**, 1835–1840 (2008).

87. Goyal, A. Metabolic adaptations underlying genome flexibility in prokaryotes. *PLoS Genet* **14**, e1007763 (2018).

88. McNally, A. *et al.* Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. *PLOS Genetics* **12**, e1006280 (2016).

89. Linton, K. J. Structure and Function of ABC Transporters. *Physiology* **22**, 122–130 (2007).

90. von Wintersdorff, C. J. H. *et al.* Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Frontiers in Microbiology* **7**, (2016).

91. Zafra, O., Lamprecht-Grandío, M., de Figueras, C. G. & González-Pastor, J. E. Extracellular DNA release by undomesticated Bacillus subtilis is regulated by early competence. *PLoS One* **7**, e48716 (2012).

92. Capra, E. J. & Laub, M. T. The Evolution of Two-Component Signal Transduction Systems. *Annu Rev Microbiol* **66**, 325–347 (2012).

93. Alm, E., Huang, K. & Arkin, A. The Evolution of Two-Component Systems in Bacteria Reveals Different Strategies for Niche Adaptation. *PLoS Comput Biol* **2**, e143 (2006).

94. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).

95. Shan, X., Szabo, R. E. & Cordero, O. X. Mutation-induced infections of phage-plasmids. 2022.11.02.514943 Preprint at https://doi.org/10.1101/2022.11.02.514943 (2022).

96. Maslov, S. & Sneppen, K. Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. *Sci Rep* **7**, 39642 (2017).

97. Chevallereau, A., Pons, B. J., van Houte, S. & Westra, E. R. Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* **20**, 49–62 (2022).

98. Domingues, S. & Nielsen, K. M. Membrane vesicles and horizontal gene transfer in prokaryotes. *Current Opinion in Microbiology* **38**, 16–21 (2017).

99. Pfeifer, E., Moura de Sousa, J. A., Touchon, M. & Rocha, E. P. C. Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res* **49**, 2655–2673 (2021).

100. Woegerbauer, M., Bellanger, X. & Merlin, C. Cell-Free DNA: An Underestimated Source of Antibiotic Resistance Gene Dissemination at the Interface Between Human Activities and Downstream Environments in the Context of Wastewater Reuse. *Front Microbiol* **11**, 671 (2020).

101. Dijk, B. van, Bertels, F., Stolk, L., Takeuchi, N. & Rainey, P. B. Transposable elements drive the evolution of genome streamlining. 2021.05.29.446280 Preprint at https://doi.org/10.1101/2021.05.29.446280 (2021).

102. van Dijk, B., Hogeweg, P., Doekes, H. & Takeuchi, N. Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. (2020) doi:10.7554/elife.56801.sa2.

103. van Dijk, B. & Hogeweg, P. In Silico Gene-Level Evolution Explains Microbial Population Diversity through Differential Gene Mobility. *Genome Biology and Evolution* **8**, 176–188 (2016).

104. D'Souza, G. *et al.* Less Is More: Selective Advantages Can Explain the Prevalent Loss of Biosynthetic Genes in Bacteria. *Evolution* **68**, 2559–2570 (2014).

105. Goyal, A. Horizontal gene transfer drives the evolution of dependencies in bacteria. *bioRxiv* 836403 (2019) doi:10.1101/836403.

106. Pande, S. *et al.* Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J* **8**, 953–962 (2014).

107. Herrera, P. *et al.* Molecular causes of an evolutionary shift along the parasitism–mutualism continuum in a bacterial symbiont. *Proceedings of the National Academy of Sciences* **117**, 21658–21666 (2020).

108. Arnold, B. J., Huang, I.-T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 1–13 (2021) doi:10.1038/s41579-021-00650-4.

109. Drew, G. C., Stevens, E. J. & King, K. C. Microbial evolution and transitions along the parasite–mutualist continuum. *Nat Rev Microbiol* **19**, 623–638 (2021).

110. Wang, Y., Baumdicker, F., Schweiger, P., Kuenzel, S. & Staubach, F. Horizontal gene transfer-mediated bacterial strain variation affects host fitness in Drosophila. *BMC Biol* **19**, 187 (2021).

111. Borodovich, T., Shkoporov, A. N., Ross, R. P. & Hill, C. Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. *Gastroenterology Report* **10**, goac012 (2022).

112. Dirksen, P. *et al.* CeMbio - The *Caenorhabditis elegans* Microbiome Resource. *G3 Genes|Genomes|Genetics* **10**, 3025–3039 (2020).

113. Henry, L. P., Bruijning, M., Forsberg, S. K. G. & Ayroles, J. F. The microbiome extends host evolutionary potential. *Nat Commun* **12**, 5141 (2021).

114. Johnson, K. V.-A. & Foster, K. R. Why does the microbiome affect behaviour? *Nat Rev Microbiol* **16**, 647–655 (2018).

115. Hall, J. P. J., Brockhurst, M. A. & Harrison, E. Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160424 (2017).

116. Bittleston, L. S., Gralka, M., Leventhal, G. E., Mizrahi, I. & Cordero, O. X. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nat Commun* **11**, 1440 (2020).

117. Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S. & Thomas, T. Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences* **108**, 14288–14293 (2011).

118. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016).

119. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).

120. Berg, M., Zhou, X. Y. & Shapira, M. Host-Specific Functional Significance of Caenorhabditis Gut Commensals. *Front. Microbiol.* **7**, (2016).

121. Theodosiou, L., Farr, A. D. & Rainey, P. B. Barcoding Populations of Pseudomonas fluorescens SBW25. *J Mol Evol* (2023) doi:10.1007/s00239-023-10103-6.

122. Kissoyan, K. A. B. *et al.* Natural C. elegans Microbiota Protects against Infection via Production of a Cyclic Lipopeptide of the Viscosin Group. *Current Biology* **29**, 1030-1037.e5 (2019).

123. Gohil, K., Rajput, V. & Dharne, M. Pan-genomics of Ochrobactrum species from clinical and environmental origins reveals distinct populations and possible links. *Genomics* **112**, 3003–3012 (2020).

124. Landy, A. DYNAMIC, STRUCTURAL, AND REGULATORY ASPECTS OF λ SITE-SPECIFIC RECOMBINATION. *Annual Review of Biochemistry* **58**, 913–941 (1989).

125. Bertels, F. & Rainey, P. B. Curiosities of REPINs and RAYTs. *Mob Genet Elements* **1**, 262–268 (2011).

126. Das, B., Martínez, E., Midonet, C. & Barre, F.-X. Integrative mobile elements exploiting Xer recombination. *Trends in Microbiology* **21**, 23–30 (2013).

127. Huber, K. E. & Waldor, M. K. Filamentous phage integration requires the host recombinases XerC and XerD. *Nature* **417**, 656–659 (2002).

128. Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* **3**, 430–440 (2002).

129. Aizawa, S.-I. Bacterial flagella and type III secretion systems. *FEMS Microbiology Letters* **202**, 157–164 (2001).

130. He, Y., Xu, T., Fossheim, L. E. & Zhang, X.-H. FliC, a Flagellin Protein, Is Essential for the Growth and Virulence of Fish Pathogen Edwardsiella tarda. *PLOS ONE* **7**, e45070 (2012).

131. Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38**, 865–891 (2014).

132. Ibáñez de Aldecoa, A. L., Zafra, O. & González-Pastor, J. E. Mechanisms and Regulation of Extracellular DNA Release and Its Biological Roles in Microbial Communities. *Front Microbiol* **8**, (2017).

133. Kita, K., Konishi, K. & Anraku, Y. Terminal oxidases of Escherichia coli aerobic respiratory chain. II. Purification and properties of cytochrome b558-d complex from cells grown with limited oxygen and evidence of branched electron-carrying systems. *Journal of Biological Chemistry* **259**, 3375–3381 (1984).

134. Singhal, R. & Shah, Y. M. Oxygen battle in the gut: Hypoxia and hypoxia-inducible factors in metabolic and inflammatory responses in the intestine. *Journal of Biological Chemistry* **295**, 10493–10505 (2020).

135. Jiang, H. & Wang, D. The Microbial Zoo in the C-elegans Intestine: Bacteria, Fungi and Viruses. *Viruses-Basel* **10**, 85 (2018).

136. Kelly, M. J., Poole, R. K., Yates, M. G. & Kennedy, C. Cloning and mutagenesis of genes encoding the cytochrome bd terminal oxidase complex in Azotobacter vinelandii: mutants deficient in the cytochrome d complex are unable to fix nitrogen in air. *J Bacteriol* **172**, 6010–6019 (1990).

137. Bi, L. *et al.* Cross-biome soil viruses as an important reservoir of virulence genes. *Journal of Hazardous Materials* **442**, 130111 (2023).

138. Diard, M. & Hardt, W.-D. Evolution of bacterial virulence. *FEMS Microbiology Reviews* **41**, 679–697 (2017).

139. Takeuchi, N. & Suzuki, H. Prophages and plasmids display opposite trends in the types of accessory genes they carry. 2022.07.21.500938 Preprint at https://doi.org/10.1101/2022.07.21.500938 (2022).

140. van Vliet, S. & Doebeli, M. The role of multilevel selection in host microbiome evolution. *Proceedings of the National Academy of Sciences* **116**, 20591–20597 (2019).

141. Shade, A. *et al.* Fundamentals of microbial community resistance and resilience. *Front Microbiol* **3**, 417 (2012).

142. Berg, M. *et al.* Assembly of the Caenorhabditis elegans gut microbiota from diverse soil microbial environments. *ISME J* **10**, 1998–2009 (2016).

143. Petersen, C., Pees, B., Martínez Christophersen, C. & Leippe, M. Preconditioning With Natural Microbiota Strain Ochrobactrum vermis MYb71 Influences Caenorhabditis elegans Behavior. *Front Cell Infect Microbiol* **11**, 775634 (2021).

144. Zhang, F. *et al.* Natural genetic variation drives microbiome selection in the Caenorhabditis elegans gut. *Current Biology* **31**, 2603-2618.e9 (2021).

145. Zimmermann, J. *et al.* The functional repertoire contained within the native microbiota of the model nematode Caenorhabditis elegans. *ISME J* **14**, 26–38 (2020).

146. Ortiz, A., Vega, N. M., Ratzke, C. & Gore, J. Interspecies bacterial competition regulates community assembly in the C. elegans intestine. *The ISME Journal* 1–15 (2021) doi:10.1038/s41396-021-00910-4.

147. Ryan, M. P. & Pembroke, J. T. The Genus Ochrobactrum as Major Opportunistic Pathogens. *Microorganisms* **8**, 1797 (2020).

148. Yang, W. *et al.* The Inducible Response of the Nematode Caenorhabditis elegans to Members of Its Natural Microbiota Across Development and Adult Life. *Front Microbiol* **10**, 1793 (2019).

149. Gilcrease, E. B. & Casjens, S. R. The genome sequence of Escherichia coli tailed phage D6 and the diversity of Enterobacteriales circular plasmid prophages. *Virology* **515**, 203–214 (2018).

150. Łobocka, M. B. *et al.* Genome of Bacteriophage P1. *J Bacteriol* **186**, 7032–7068 (2004).

151. Utter, B. *et al.* Beyond the Chromosome: The Prevalence of Unique Extra-Chromosomal Bacteriophages with Integrated Virulence Genes in Pathogenic Staphylococcus aureus. *PLOS ONE* **9**, e100502 (2014).

152. Pfeifer, E., Moura de Sousa, J. A., Touchon, M. & Rocha, E. P. C. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Research* **49**, 2655–2673 (2021).

153. Horne, T., Orr, V. T. & Hall, J. P. How do interactions between mobile genetic elements affect horizontal gene transfer? *Current Opinion in Microbiology* **73**, 102282 (2023).

154. Smyshlyaev, G., Bateman, A. & Barabas, O. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol Syst Biol* **17**, e9880 (2021).

155. Wang, X. *et al.* Recent advances and application of whole genome amplification in molecular diagnosis and medicine. *MedComm (2020)* **3**, e116 (2022).

156. Lebonah, D. E. *et al.* DNA Barcoding on Bacteria: A Review. *Advances in Biology* **2014**, e541787 (2014).

157. Louca, S. *et al.* Function and functional redundancy in microbial systems. *Nat Ecol Evol* **2**, 936–943 (2018).

158. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

159. Nayfach, S. *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol* **39**, 499–509 (2021).

160. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, (2015).

161. Lagkouvardos, I. *et al.* The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat Microbiol* **1**, 1–15 (2016).

162. Rühlemann, M. C. *et al.* Comparative metagenomics reveals host-specific functional adaptation of intestinal microbiota across hominids. 2023.03.01.530589 Preprint at https://doi.org/10.1101/2023.03.01.530589 (2023).

163. Garud, N. R., Good, B. H., Hallatschek, O. & Pollard, K. S. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLOS Biology* **17**, e3000102 (2019).

164. McDonald, M. J., Rice, D. P. & Desai, M. M. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233–236 (2016).

165. Goddard, M. R., Godfray, H. C. J. & Burt, A. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**, 636–640 (2005).

166. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* **26**, 10.3402/mehd.v26.26191 (2015).

167. Rosenfeld, C. S. Gut Dysbiosis in Animals Due to Environmental Chemical Exposures. *Frontiers in Cellular and Infection Microbiology* **7**, (2017).

168. Bäumler, A. J. & Sperandio, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).

169. Groussin, M. *et al.* Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell* **184**, 2053-2067.e18 (2021).

170. Makky, S. *et al.* The bacteriophage decides own tracks: When they are with or against the bacteria. *Current Research in Microbial Sciences* **2**, 100050 (2021).

171. Carroll, A. C. & Wong, A. Plasmid persistence: costs, benefits, and the plasmid paradox. *Can. J. Microbiol.* **64**, 293–304 (2018).

172. Cury, J., Oliveira, P. H., de la Cruz, F. & Rocha, E. P. C. Host Range and Genetic Plasticity Explain the Coexistence of Integrative and Extrachromosomal Mobile Genetic Elements. *Molecular Biology and Evolution* **35**, 2230–2239 (2018).

173. Koskella, B. & Meaden, S. Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses* **5**, 806–823 (2013).

174. Enault, F. *et al.* Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J* **11**, 237–247 (2017).

175. Pfeifer, E., Bonnin, R. A. & Rocha, E. P. C. Phage-Plasmids Spread Antibiotic Resistance Genes through Infection and Lysogenic Conversion. *mBio* **13**, e01851-22.

176. Baumdicker, F. & Kupczok, A. Tackling the Pangenome Dilemma Requires the Concerted Analysis of Multiple Population Genetic Processes. *Genome Biology and Evolution* **15**, evad067 (2023).

177. Maddamsetti, R. & Lenski, R. E. Analysis of bacterial genomes from an evolution experiment with horizontal gene transfer shows that recombination can sometimes overwhelm selection. 30.

178. Datta, M. S., Sliwerska, E., Gore, J., Polz, M. F. & Cordero, O. X. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat Commun* **7**, 11965 (2016).