Diversity and functions of bacteria associated with springtails

Valeria Agamennnone
This research was financed by grant F07.003.05 from the BE-Basic Foundation (Biotechnology based Ecologically Balanced Sustainable Industrial Consortium).

Diversity and functions of bacteria associated with springtails

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op donderdag 4 oktober 2018 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Valeria Agamennone
geboren te San Giovanni Rotondo, Italië
promotoren: prof.dr. N.M. van Straalen
            prof.dr. A. Brouwer

copromotoren: dr.ir. T.F.M. Roelofs
                dr. W.F.M. Röling†
# Table of contents

Chapter 1 – Introduction ........................................................................... 7

Chapter 2 – The microbiome of *Folsomia candida*: an assessment of bacterial diversity in a *Wolbachia*-containing animal .............................................. 17

Chapter 3 – Antimicrobial activity and carbohydrate metabolism in the bacterial metagenome of the soil-living invertebrate *Folsomia candida* .............................................................. 39

Chapter 4 – Antimicrobial activity in culturable gut microbial communities of springtails .......................................................................................................................... 65

Chapter 5 – Genome sequence and antimicrobial properties of *Bacillus toyonensis* VU-DES13, isolated from the *Folsomia candida* gut, and its role in the colonization resistance to entomopathogenic fungi ........................................ 85

Chapter 6 – Discussion ............................................................................. 117

Bibliography .......................................................................................... 129

Summary .............................................................................................. 149

Samenvatting ....................................................................................... 151

Acknowledgements .............................................................................. 155

Curriculum vitae .................................................................................. 157

List of publications ................................................................................ 158

Annex – Genome draft of *Bacillus toyonensis* VU-DES13 .................... 159
Chapter 1 – Introduction

1.1. A microbial world

Microorganisms have been the foundation for the evolution of all other life forms on Earth. Unicellular, prokaryotic microorganisms were the first living organisms to populate the planet, approximately 3.8 billion years ago. Eukaryotic microorganisms appeared around 2 billion years ago, and only around 1 billion years ago more complex, multicellular life forms began to emerge. The specific functions and metabolic capabilities evolved in the first microorganisms during millions of years shaped the environmental characteristics of the planet, thereby influencing the evolution of other life forms. For example, oxygenic photosynthesis by Cyanobacteria led to the accumulation of oxygen in the atmosphere, and aerobic metabolism allowed more complex organisms to evolve (Stamati, Mudera and Cheema 2011). Today, microbial metabolism supports the existence of all life on Earth by contributing to complex biogeochemical processes on a global scale (Gilbert and Neufeld 2014).

In terms of biodiversity and variety of environments colonized, microorganisms are the most successful life forms on the planet. Recent estimates of global biodiversity, taking into account the many unique symbionts of insects, amount to 2 billion species, of which Bacteria form three quarters (Larsen 2017). Bacteria and Archaea contain unicellular prokaryotic organisms, and are therefore entirely microbial. The domain Eukarya, to which plants, animals and fungi belong, also includes many microorganisms. Furthermore, microbes are found in almost every habitat on Earth, and they possess specific adaptations that allow them to occupy the most diverse environmental niches. More complex, multicellular organisms, such as plants and animals, can themselves constitute a habitat for microbes.

1.2. Microbial symbioses in hexapods

The intestinal tract of animals is a typical example of an environment that can be colonized by microbes, offering opportunities for symbioses to arise. The word symbiosis (from the Greek “life together”) is used to describe different kinds of interactions between organisms. While some microorganisms are parasitic and can cause diseases, others are mutualistic, and they can contribute to the fitness of their host by mediating ecologically important traits (Feldhaar 2011). Specific and evolutionarily ancient symbioses can deeply influence the evolution of the
interacting organisms. For example, in obligate symbioses, genome reduction is often observed in the symbiont (Boscaro et al. 2017), while hosts become dependent, to various degrees, upon services provided by the symbionts (Fisher et al. 2017). These and other observations indicate that microorganisms constitute important aspects of the biology of their hosts, and that studying microbe-host interactions is important to achieve a good understanding of the natural world (McFall-Ngai et al. 2013).

After microorganisms, hexapods and in particular insects are the most diverse and widespread group of organisms on Earth (Stork et al. 2015). With more than 1 million species described, insects represent approximately 66% of the total number of described animal species (Zhang et al. 2011), and estimates suggest 5.5 million species are yet to be discovered (Stork 2018). This incredible biodiversity offers many good opportunities to study microbial symbioses. The small size of insects, their short generation time, the ability of most species to metamorphose and to fly, and coevolution with plants are often recognized as factors that have contributed to the success of these animals. Additionally, microbial symbioses have been recognized as important drivers in the diversification of insects because they can confer different benefits to the host (Klepzig et al. 2009).

Nutritional symbioses are good examples of hexapod-microbe associations providing benefits for the host. Some microbes allow their hosts to thrive on nutritionally poor or highly specific diets by directly synthesizing additional nutrients, or by facilitating the digestion of recalcitrant compounds. For example, aphids feeding on phloem sap with a low content of essential amino acids can obtain the nutrients missing from their diet from a symbiotic bacterium (Akman Gunduz and Douglas 2009). Another good example is constituted by termites, wood-feeding insects that harbor specific communities of bacteria and protists in their hindgut. These microbes hydrolyze cellulose and hemicellulose, making it possible for their host to obtain nutrients from these plant compounds (Warnecke et al. 2007). Aside from nutritional symbionts, there are also defensive ones, that protect their host from pathogens or parasites (Ballinger and Perlman 2017), and there is a variety of other roles that microbes can play in the biology of their hexapod host, for example in relation to detoxification, development or communication (Engel and Moran 2013; van den Bosch and Welte 2017). Figure 1 summarizes the known functions of gut bacteria in insects.
Gut bacteria in insects are involved in a variety of processes: they contribute to digestion, by supplementing nutrients and breaking down dietary components; to immunity, by protecting the host from pathogens through colonization resistance; and to communication, by producing specific molecules (figure in Engel and Moran, 2013).

The above examples indicate that microbes may contribute in different ways to the fitness and adaptations of their host. Studying hexapod-associated microbial communities can provide important insights in these animals' ecology, as well as lead to important practical applications. For examples, symbiotic microbial communities can be used as targets in agricultural pest management strategies, by eliminating microorganisms required for the survival and reproduction of the host (Arora and Douglas 2017). Furthermore, the biosynthetic potential of microbial communities associated with hexapods can be explored and exploited for a variety of biotechnological applications.

### 1.3. Drugs from bugs

Nature is a source of many products with potential commercial value. In particular, microorganisms are important targets for product discovery, both because of their biodiversity and corresponding biosynthetic potential, and because of the possibility to make use of their short generation time to establish efficient production processes (Abdel-Rahman, Tashiro and Sonomoto 2013). Many secondary metabolites produced by microbes are used in the pharmaceutical and medical sector. A good example is constituted by antibiotics. The first antibiotic to be discovered, penicillin, is produced by *Penicillium* fungi, and, even with the advances of synthetic chemistry, many other antibacterial and antifungal compounds are still isolated from microorganisms or are obtained through semisynthetic modifications of natural products. Another class of microbial products with important applications is represented by antitumor drugs (Chang *et al.* 2011).
Microbial symbioses may offer good opportunities for drug discovery. In the past, interesting compounds have been isolated from marine organisms such as bryozoans and sponges, and more recent studies have revealed the microbial origin of these products (Sudek et al. 2007; Kennedy et al. 2008; Yung et al. 2011). Also, among the insects, there are examples of biologically active compounds produced by their associated microbes. Some of these microbial products may bring benefits to the host by functioning as defense mechanisms against pathogens or predators. For example, a Pseudomonas symbiont of the beetle Paederus fuscipes produces pederin, a polyketide with antitumor activity that protects the beetle against its natural predators (Piel, Höfer and Hui 2004). Similarly, in fungus-growing ants a bacterial symbiont of the genus Streptomyces produces antibiotics that are active against a parasitic fungus, preventing it from infecting the garden (Currie et al. 1999).

In the search for new molecules, the likelihood of finding compounds with activities of interest can be increased by “mining” the appropriate environment. For example, the energy industry shows interest in lignocellulose degradation, as this allows the conversion of plant biomass to simple sugars, which can subsequently be fermented to obtain ethanol as well as other bulk chemicals. Enzymes responsible for this catalytic activity may be found in the digestive systems of organisms feeding on wood, and, indeed, the microbial community of termites contains many functions related to the degradation of plant materials, such as cellulose and xylan hydrolysis (Warnecke et al. 2007). Similarly, other specific environments can be screened for possible new catalytic or biosynthetic activities using exploratory techniques such as sequence-based and functional metagenomics.

The traditional view on evolution of antibiotic biosynthesis results from the evolutionary arms race taking place between microorganisms in the environment. Recently, this view has been challenged by observations that, in low concentrations, antibiotics also modulate important functions in bacterial populations, such as gene expression, motility and the formation of biofilms. This indicates that antibiotics may have evolved as signaling molecules in bacteria (Townsley and Shank 2017). Either way, it is clear that the production of antibiotics is a social phenomenon that occurs as a result of interactions between microorganisms (Abrudan et al. 2015). Therefore, environments rich in microbes and in microbial interactions, may constitute good sources of antibiotics. With up to 10 billion bacterial cells per gram, the soil is clearly a microbe-dominated environment, and soil microorganisms are indeed the source of most antibiotics in use today. Soil-born animals that feed on dead organic matter, such as collembolans and earthworms, also ingest high amounts of microorganisms as part of their diets. The digestive tract of these soil invertebrates constitutes another microbe-rich environment (Brito-Vega and Espinosa-Victoria 2009; Pawar et al. 2012) that may be explored for drug discovery.
1.4. High-throughput sequencing and metagenomic analyses

Traditional microbiology relied on culturing, microscopy and biochemical tests to study microorganisms. Later on, nucleic acid-based assays such as sequencing were introduced, and these techniques advanced incredibly after the development of the polymerase chain reaction (PCR), which allowed to obtain high levels of DNA. Sequence-based identification quickly became the standard for classifying microorganisms, as opposed to the previous phenotypic and physiological identification. In bacteria and fungi, sequence-based identification is mostly based on ribosomal RNA genes. These genes are ubiquitous, and they present both highly conserved regions, allowing the design of primers to target a wide variety of species, and variable ones, that allow to differentiate among them. Sequencing information is stored in public databases, which have grown exponentially in size in the last decade with the introduction and development of high-throughput sequencing methods.

High-throughput sequencing is a powerful tool, because of the incredible amount of information that it can generate. Furthermore, it can be successfully applied to study uncultured microbial communities. This term refers to the fact that most microbes in the environment are not accessible through culturing (Rappé and Giovannoni 2003). This constitutes a limit in microbial ecology, however, information on unculturable microbes can be obtained by studying the total genetic material isolated from the microbial communities. This approach, called metagenomics, combined with the depth of sequencing resulting from high-throughput methods, has provided unprecedented insight in this hidden microbial diversity. Metagenomic approaches allow us to study microbes in their natural environment, thus leading to a better understanding of the complexity of microbial communities with regard to both their composition and functions.

Microorganisms are important members of soil ecosystems, which harbor a large portion of Earth’s biodiversity. Soil microbial communities are responsible for many key functions that ultimately maintain and support life on Earth. These life support functions (LSF) or ecosystem services define the health and quality of soils, and they include the decomposition of organic matter and mineralization, nutrient cycling, the maintenance of soil structure, and the establishment of symbioses with other microorganisms and with plants and animals. Metagenomics offers the opportunity to study soil microbial communities with increasing depth and resolution, providing an insight into their roles and their impact on soil quality (Nesme et al. 2016; Schloter et al. 2018).

When applied to host-associated microbes, metagenomics can shed a light on the characteristics of these symbiotic systems, and possibly on the mechanisms underlying the interactions between the host and its microbes. For example,
metagenomics has been used to study the nutritional roles of gut microbes in wood-feeding termites (Warnecke et al. 2007; He et al. 2013). The digestive tract of these insects harbors complex microbial communities comprising bacteria, archaea, fungi and protists, that are responsible for the breakdown of the ingested wood material through the production of an array of glycoside hydrolases, cellulases and hemicellulases (Brune and Dietrich 2015). Similarly, metagenomic methods have been applied to study the complex microbial interactions in the fungal agricultural systems of herbivorous insects, providing an insight into their plant biomass degradation capacities (Suen et al. 2010; Aylward et al. 2014). In honey bees, metagenomics was applied to identify the candidate pathogens possibly responsible for the colony collapse disorder that threatens bee populations worldwide (Cox-Foster et al. 2007).

Aside from contributing to the field of microbial ecology, metagenomics is also a valid approach in the search for new molecules with potential biotechnological applications. Metagenomic libraries (collections of environmental DNA) can be screened for functions of interest using both sequence- and function-driven methods (Iqbal, Feng and Brady 2012). Sequence-based metagenome mining identifies potentially interesting genes based on similarity with known database sequences. This type of metagenomics was traditionally conducted using microarrays with a large number of nucleotide probes designed to bind DNA sequences from environmental samples (e.g. the GeoChip, Azarbad et al. 2015). Due to the rise of next generation sequencing technology this approach is now mostly replaced by direct shotgun sequencing of environmental DNA. Functional metagenomics, instead, does not require sequence information, and is based on phenotypic detection of the desired metabolic activity, host complementation or gene expression (Simon and Daniel 2011). While sequence-based strategies can evaluate the overall functional potential of microbial communities and identify potentially interesting genes, targeted studies are necessary to confirm the functions of individual candidate genes and also to discover enzymatic activities that cannot be predicted based on DNA sequence alone (Cheng et al. 2017). Furthermore, confirming predicted functions is important for a meaningful interpretation of the data in an ecological perspective.

### 1.5. *Folsomia candida*

#### 1.5.1. A model springtail

*Folsomia candida* is a small soil invertebrate belonging to the class Collembola (Hexapoda). These animals, commonly known as springtails, were the first hexapods to evolve more than 400 million years ago, and, together with Protura and Diplura, they constitute a basal lineage of Hexapoda and the sister group of insects (Misof et al. 2014). Specific features distinguish springtails from insects, such as
entognathous mouth parts, the ancestral absence of wings, the absence of Malpighian tubules and the lack of metamorphosis. Springtails are also characterized by the presence of a furcula, an appendage functioning as a jumping organ, and a colophore, or ventral tube, playing a role in water intake and osmoregulation. The over 8000 described species of springtails are especially abundant in moist environments and soils, where they feed on fungal hyphae, decaying plant material and microorganisms. In turn, springtails constitute a food source for different predators, such as spiders, coleopterans and mites, and they can host a variety of parasitic protozoans, bacterial and fungal pathogens. By disintegrating plant material and macro- and megafauna droppings, springtails contribute to important soil processes such as litter decomposition and the formation of soil microstructures (Rusek 1998).

*F. candida* is an euedaphic springtail widespread in soils throughout the world. Because of its sensitivity to soil pollution and the ease of culturing in the lab, *F. candida* has been used as a biological indicator for soil quality and as a model organism in ecotoxicology for decades (Fountain and Hopkin 2005). Aside from traditional ecotoxicity tests, based on monitoring the effects of stress exposure on survival and reproduction, *F. candida* has also been used in expression-based studies, that have provided insights in the molecular basis of stress response in this springtail (Timmermans *et al.* 2009). The study of gene expression in *F. candida* also provided the first indication of the presence of antibiotic-related genes in the genome of this animal (Nota *et al.* 2008), later confirmed by specific studies (Roelofs *et al.* 2013; Suring *et al.* 2016). Further research showed that beta-lactam biosynthesis genes are present in several other springtail species (Suring *et al.* 2017), likely originating from a horizontal gene transfer (HGT) event close to the origin of the clade. Genes with bactericidal functions have been previously identified as transferred genes in eukaryotic genomes (Husnik and McCutcheon 2018).

### 1.5.2. Horizontal gene transfer

HGT is the lateral movement of genetic material between organisms, as opposed to the vertical transmission of DNA from parent to progeny characteristic of reproduction. Lateral transfer of genes is widespread among prokaryotes, where it can allow for rapid adaptation to changing environmental conditions. However, over the years, increasing evidence has also accumulated for the occurrence of this phenomenon across different life domains, with many cases of gene transfer being described between bacteria and animals (Dunning Hotopp 2011). Horizontal gene transfer constitutes a source of innovation and an important factor in the evolution of animals (Moran and Jarvik 2010; Boschetti *et al.* 2012): by acquiring novel functions related to nutrition, detoxification and pathogen defense, recipients of
novel genes may become able to adapt to previously unexploited environmental niches (Husnik and McCutcheon 2018).

Many years ago, conjugative gene transfer was observed between microbial species in the gut of _F. candida_ (Hoffmann _et al._ 1998). More recently, many genes of foreign origin have been found in the genome of this springtail (Faddeeva-Vakhrusheva _et al._ 2017) suggesting that the animal itself may be able to acquire genes through HGT. Most of the foreign genes in _F. candida_ are of bacterial or fungal origin (Faddeeva-Vakhrusheva _et al._ 2017), and were probably obtained from microorganisms living in association or in close proximity with the animal (Figure 2). Furthermore, many foreign genes in _F. candida_ have functions related to carbohydrate metabolism and plant biomass degradation, indicating that their acquisition may have been beneficial for the springtail by allowing access to important food sources in the soil.

![Figure 2](image_url) Left: pie chart shows the origin of foreign genes in _Folsomia candida_. Right: bars represent the numbers of foreign gene sequences in _Folsomia_ annotated with CAZy (Carbohydrate-Active Enzymes) domains. GH: glycoside hydrolases; GT: glycosyltransferases; AA: auxiliary activity module; CE: carbohydrate esterases; SLH: S-layer homology module; PL: polysaccharide lyases; CBM: carbohydrate binding module. Modified from Faddeeva-Vakhrusheva _et al._, 2017.

### 1.5.3. The microbiome of _F. candida_

The microbiota of _F. candida_ was first studied using culture-based approaches and 16S rRNA sequencing (Thimm _et al._ 1998; Czarnetzki and Tebbe 2004a). These early studies identified a few bacterial and one fungal species, including common soil species and isolates related to nitrogen fixing bacteria and plant pathogenic ones (Thimm _et al._ 1998; Czarnetzki and Tebbe 2004a). Despite changes in the abundance of microbial cells caused by molting, the gut of _F. candida_ was shown to harbor a specific microbial community, which may in part be explained by specific food preferences by the springtail (Thimm _et al._ 1998). Furthermore, with a high density of bacterial cells and high rates of gene transfer among bacteria, the gut of _F. candida_ was suggested to be a hot spot for HGT (Hoffmann _et al._ 1998).
F. candida is a soil invertebrate and its diet includes fungal hyphae and fragmented plant biomass. These factors may select for a symbiotic microbial community that can assist the springtail with the digestion of recalcitrant compounds. Carbohydrate-degrading functions may therefore be enriched in the gut of F. candida, and, as previously shown, they may also be transferred to the springtail through HGT (Faddeeva-Vakhrusheva et al. 2017). Furthermore, the high density of bacterial cells inside the gut of the springtail (Hoffmann et al. 1998) and the changeable nature of this microhabitat due to molting (Thimm et al. 1998) may create optimal conditions for a microbial warfare, which in turn may lead to antimicrobial production. Interestingly, high rates of HGT may be the result of the presence of antibiotics in low concentrations in the environment (Jutkina et al. 2016), supporting the possible presence of antimicrobial compounds in the gut of F. candida. These observations suggest that the microbial community of F. candida may constitute a target for the discovery of interesting functions, specifically related to carbohydrate metabolism and antibiotic production.

Metagenomics and high-throughput sequencing, by providing in-depth information on the functions of microbial communities, constitute appropriate tools to investigate these functions. Furthermore, they enabled us to better characterize the symbiosis between F. candida and its microbes. Sequencing and assembling the genome and the transcriptome of F. candida were important stepping stones that allowed for a deeper understanding of the ecology and the evolutionary history of this species; similarly, studying the springtail’s metagenome would provide important insights into the ecology of its interacting microbes.

1.6. Scope, approach and outline of the thesis

In this thesis we studied the microbial community associated with the springtail F. candida. This springtail is well characterized as a model organism in ecotoxicology and ecogenomics, and its genome and transcriptome have recently been assembled, however our knowledge about its associated microbes is rather limited. Symbiotic microorganisms are an integrated part of the biology of animals, and can determine important aspects of their ecology and evolution. For example, microorganisms can be involved in functions such as nutrition and immunity, and they can directly provide animals with new metabolic functions through HGT. Furthermore, studying host-associated microbes may lead to the discovery of compounds with potentially important biotechnological and medical applications.

The main questions that were addressed in this research were:

1) What is the structure of the microbial community of F. candida?
2) What potential functions are present in this community?
3) Can we identify specific antimicrobial compounds or resistances?
Chapter 1

4) How does the microbiome contribute to the ecology or evolution of its hosts?

To address these questions, we used a combination of sequence-based (amplicon and shotgun) and function-based approaches to study the microbial community of *F. candida*, adding to the previous knowledge about this animal’s microbiota.

In Chapter 2 we describe the composition of the microbial community of the springtail *F. candida*. The scope of this chapter was to assess the bacterial diversity associated with this animal. To do this, we compared two populations of *F. candida* by applying an amplicon-based high-throughput method to obtain a complete overview of the bacterial community. The most abundant bacterial taxa were shared, but we found important differences in the abundances of the bacterial endosymbiont *Wolbachia pipientis* and of rare species between the two *F. candida* populations.

In Chapter 3 we provide a detailed overview of the potential functions present in the gut microbiota of *F. candida*. We sequenced the metagenome of the springtail’s gut, and we used a bioinformatics approach to highlight some of the features of this metagenome. We found a number of antibiotic-biosynthesis gene clusters, antibiotic resistance and carbohydrate metabolism functions, both probably relevant for the ecology of the springtail.

In Chapter 4 we describe the inhibitory activity of bacteria isolated from the gut of springtails. The scope of this chapter was to investigate the antimicrobial potential of springtails’ gut microflora. To study this, we performed a functional study where we tested the ability of springtails’ isolates to inhibit common fungal pathogens. We observed a widespread capacity for fungal inhibition in all the springtail species tested, suggesting a high potential for drug discovery.

In Chapter 5 we describe a *Bacillus toyonensis* isolate obtained from the gut of *F. candida*. The scope of this chapter was to study the basis of the inhibitory activity against pathogens that was shown by the isolate. We characterized the metabolic profile of the isolate, we tested its antibiotic resistance, we studied its genome and we performed various agar assays to study its antimicrobial activity. The strain showed high resistance to penicillin, substantiated by the presence of resistance genes. Genome analysis also revealed the presence of secondary metabolite clusters. Evidence of antimicrobial production was found both in live bacteria and in their extract, pointing to the possibility of extracting antimicrobials from the strain.

In the final Chapter 6, I provide a synthesis and a discussion of the results from the previous chapters in the light of the research questions of the thesis.
Chapter 2 – The microbiome of *Folsomia candida*: an assessment of bacterial diversity in a *Wolbachia*-containing animal

Agamennone V, Jakupović D, Weedon JT, Suring WJ, van Straalen NM, Roelofs D, Röling WFM. *FEMS Microbiology Ecology* 2015; 91: fiv128
Chapter 2

2.1. Abstract

The springtail *Folsomia candida* is an important model organism for soil ecology, ecotoxicology and ecogenomics. The decomposer activities of soil invertebrates like *Folsomia* depend on their relationship with microbial communities including gut symbionts. In this paper, we apply high-throughput sequencing to provide a detailed characterization of the bacterial community associated with parthenogenetic *F. candida*. First, we evaluated a method to suppress the amplification of DNA from the endosymbiont *Wolbachia*, to prevent it from interfering with the identification of less abundant Operational Taxonomic Units (OTUs). The suppression treatment applied was effective against *Wolbachia* and did not interfere with the detection of the most abundant OTUs (59 OTUs, contributing over 87% of the reads). However, this method did affect the inferred community composition. Significant differences were subsequently observed in the composition of bacterial communities associated with two different strains of *F. candida*. A total of 832 OTUs were found, of which 45% were only present in one strain and 17% only in the other. Among the 20 most abundant OTUs 16 were shared between strains. Denaturing Gradient Gel Electrophoresis (DGGE) and clone libraries, although unable to capture the full diversity of the bacterial community, provided results that supported the NGS data.

2.2. Introduction

Microorganisms living in symbiosis with animals are intimately connected with virtually every aspect of their hosts’ lives, from early development to long-term adaptation and evolution (Zilber-Rosenberg and Rosenberg 2008; McFall-Ngai et al. 2013). Microbial symbionts are fundamental for the well-being of many insects through their contribution to nutrition (Warnecke et al. 2007; Engel, Martinson and Moran 2012), pathogen defense (Behar, Yuval and Jurkevitch 2008; Teixeira, Ferreira and Ashburner 2008) and other important functions (Moran and Jarvik 2010). Such intimate functional interactions support the idea that microorganisms and their hosts can form a single unit in the context of evolution (Mcfall-Ngai 2008; Zilber-Rosenberg and Rosenberg 2008), and suggest that a deep understanding of the biology of an organism cannot exclude the study of its associated microbes.

Microarthropods belonging to the group of springtails (Collembola; Hexapoda) are found in high-humidity areas around the world, and are very abundant in soils and litter, where they feed on detritus and microorganisms (Thimm et al. 1998). The springtail *Folsomia candida* has been extensively used as a model organism in soil ecology, ecotoxicology and ecogenomics (Fountain and Hopkin 2005; Nota et al. 2010). Recently, antibiotic biosynthesis genes were discovered in the genome of *F. candida* (Roelofs et al. 2013). Antibiotic biosynthesis by animals themselves has never been observed before, therefore we hypothesize that these genes originated
from a symbiotic microorganism through a lateral gene transfer (LGT) event. Studying the microbial community of *F. candida* can help us identify the potential source of these genes as well as increase our understanding of the relationships between this springtail and its associated microbes.

On the basis of the study of clone libraries, it was postulated that *F. candida* harbors a specific bacterial community that differs from that of other springtails because of selective factors such as food preferences and the presence of specific gut metabolites (Czarnetzki and Tebbe 2004a). The gut of *F. candida* has been described as a selective habitat for microbes (Thimm et al. 1998) and as a hot spot for lateral gene transfer (Hoffmann et al. 1998). Most populations of *F. candida* are infected with the bacterial endosymbiont *Wolbachia pipiens*. The endosymbionts are mainly associated with brain and ovary tissues and are assumed to impose a system of parthenogenetic reproduction in the springtail (Czarnetzki and Tebbe 2004b).

Czarnetzki & Tebbe (2004a) showed that the high abundance of *Wolbachia* can interfere with the identification of other, less abundant, microbial species in studies relying on PCR-amplified 16S rRNA genes. Therefore, after exclusively detecting *Wolbachia* in libraries obtained with primer set F27-R1492, Czarnetzki & Tebbe (2004a) used an alternative reverse primer (R1525) to avoid amplification of *Wolbachia* DNA and identify other bacterial sequences. Gene libraries constructed with primer R1525 did not contain sequences of *Wolbachia*, but rather those of various other phylogenetic groups, many of which were identified as soil bacteria. However, in addition to selecting against *Wolbachia*, primer R1525 may also lead to a failure to detect other bacteria. Furthermore, Czarnetzki & Tebbe (2004a) characterized 3 clone libraries, containing 50, 93 and 95 clones. Comparing communities on the basis of clone libraries with relative small sizes limits the sensitivity of analyses, as often only a small part of the actual diversity is recovered (Hughes et al. 2016).

Nowadays, Next Generation Sequencing (NGS) offers the opportunity to determine the compositions of bacterial communities in great detail. This has proven to be an effective tool to identify some rare groups that were not easily detected with older methods (Kautz et al. 2013). In this paper we present a NGS-centered study of the bacterial community associated with *F. candida*. We first tested the method suggested by Czarnetzki & Tebbe (2004a) for its efficacy to avoid amplification of undesired *Wolbachia* DNA and evaluated the potential introduction of biases resulting from the use of this method. A reliable and minimally biased method to filter out *Wolbachia* is needed to characterize the rest of the microbiome, including rare species whose detection might be hindered by the presence of dominant groups. Rare species constitute the largest portion of microbial diversity in the environment, and their functional potential and genomic diversity are widely unexplored (Sogin et al. 2006).
In our study we compared two strains of *F. candida*, one cultured in the laboratory for a long time (over ten years), one recently sampled from the field, to assess the degree of intraspecies variability in the composition of bacterial communities. Finally, we compared the NGS data on the bacterial community of *F. candida* to DGGE and clone libraries data, to evaluate whether the three methods would give comparable results. Answering these questions will provide the basis for a systematic study of the microbiome of *F. candida* for the purpose of better understanding the relationship between this springtail and its microbiome.

### 2.3. Materials and methods

**Test organism**

The study was conducted on the springtail *Folsomia candida* Willem, 1902 (Hexapoda: Collembola, Isotomidae). Two strains of *F. candida* were used: a laboratory population derived from the “Berlin strain” (henceforth referred to as “Berlin”), a well-established breeding stock maintained for more than 20 years in climate rooms at the VU University Amsterdam; and a field-derived population established from approximately 30 animals collected at a sampling site in Zaandam (The Netherlands) (henceforth “Zaandam”) in November 2012. The field-derived Zaandam population was kept in the laboratory for 4 months, and individuals that hatched from eggs that were laid in the lab were used for the analysis, in order to reduce the influence of the original environment on their microbiome composition and therefore focus on population-specific differences in community composition. All springtails reproduced by parthenogenesis and were bred in PVC rings of 25 cm diameter, on a plaster of Paris-charcoal substrate, and were fed dry baker’s yeast (Dr. Oetker, Bielefeld, Germany). The breeding conditions in climate rooms were stable at 20°C temperature, 75% humidity and a 12 hour light/dark cycle.

**Experimental design**

To test the strategy suggested by Czarnetzki & Tebbe (2004a) to suppress amplification of *Wolbachia* DNA, we used springtails from the Berlin strain. Four replicate samples, each consisting of a pool of ten adult individuals, were taken from this population. Each replicate was subjected to normal amplification (referred to as: control) and to amplification with the primer pair used by Czarnetzki & Tebbe (2004a) (referred to as: suppression treatment). This resulted in a total of eight amplified DNA samples consisting of four pairs of control vs. suppression treatment. To assess differences in microbial composition at the population-level, the Berlin population was compared to the Zaandam one. Four biological replicates were taken for each of the two populations, each replicate consisting of a pool of ten springtails, and the suppression treatment was applied to all samples. This second experiment
The microbiome of *Folsomia candida* resulted in two sets of four non-paired samples of amplified DNA. An overview of the experimental set-up is shown in Figure S1 (Supplementary Information).

**DNA isolation**

Samples were taken by placing the springtails (ten individuals combined per extraction) in Eppendorf tubes and snap-freezing them in liquid nitrogen prior to DNA extraction. Total DNA was isolated using the Wizard® SV Genomic DNA Purification System (Promega, Madison, WI, USA) with modifications. Frozen springtails were ground in 100 µl Phosphate Buffer Saline (PBS) using a plastic pestle in the presence of liquid nitrogen, to ensure disruption of bacterial cell walls. 100 µl Nuclei Lysis Solution (NLS) was added, along with 1 µl of Proteinase K. The samples were then incubated for 15 min at 65°C. Following this, 170 µl DNA Lysis Buffer was added, and the samples were centrifuged for 10 min at 14000 rpm. The supernatant was then transferred to a DNA-binding minicolumn and washed three times with 500 µl Wizard® SV Wash Solution, by centrifuging each time at 14 000 rpm for 1 min. Finally, the DNA was eluted in 40 µl Nuclease-Free Water and its concentration measured on a NanoDrop ND-1000 spectrophotometer (Wilmington, DE, USA).

**Amplification of 16S rRNA genes**

Bacteria specific primers were used to amplify 16S rRNA gene fragments (Table S1, Supplementary Information). The hypervariable V3 region was targeted using primer pair 357F-518R, and the fragment obtained was used for DGGE analysis and high-throughput sequencing; a larger fragment was amplified with primers 357F-1392R and used for clone libraries and subsequent sequencing (Figure S2, Supplementary Information).

Czarnetzki & Tebbe (2004a) obtained 100% *Wolbachia* sequences when using primers 27F and 1492R to amplify 16S rRNA genes in *F. candida*. They subsequently used primers 8F and 1525R and successfully obtained sequences from bacteria other than *Wolbachia*. Following this finding, in this study we applied a nested PCR approach: primers 8F and 1525R were used for a first round of amplification, and a second PCR with primer pairs 357F-518R or 357F-1392R allowed us to obtain fragments of appropriate size for subsequent analysis (respectively, DGGE/high-throughput sequencing or cloning). To amplify fragments for DGGE profiling, primer 357F with a GC-clamp was used. For the preparation of samples for Illumina sequencing, forward and reverse primers with barcodes and Illumina adapters were used (IDT, Leuven, Belgium) (Bartram *et al.* 2011).

The initial PCR reactions were set up as follows: 1 µl DNA template, 1 µl of each primer (10 pmol), 11 µl Nuclease-Free Water (Promega) and 11 µl GoTaq Colorless Master Mix (Promega), for a final volume of 25 µl. The PCR reactions with Illumina primers were set up as follows: 2 µl template (concentration: 2.5 ng µl⁻¹), 1.25 µl of
each primer, 8 µl Nuclease-Free Water (Promega) and 12.5 µl GoTaq Colorless Master Mix (Promega), for a final volume of 25 µl. To minimize contamination, all PCR mixes were prepared in a UV sterilized cabinet. The following PCR programs were used: (1) For the primer pair 357F/GC357F-518R: initial denaturation for 5 min at 94°C, followed by 35 amplification cycles (30 s at 94°C, 30 s at 55°C and 1 min at 72°C) and a final extension step of 8 min at 72°C. (2) For the primer pairs 8F-1525R and 357F-1392R: initial denaturation for 5 min at 94°C, followed by 35 amplification cycles (30 s at 94°C, 30 s at 55°C and 1 min 30 s at 72°C) and a final extension step of 8 min at 72°C. (3) For the Illumina PCR: initial denaturation for 2 min at 95°C, followed by 25 amplification cycles (30 s at 95°C, 30 s at 50°C and 45 s at 72°C) and a final extension step of 5 min at 72°C. Amplification products were verified by electrophoresis on 1.5% (wt/vol) agarose gels and by staining with ethidium bromide.

**DGGE analysis**

DGGE was performed with a Bio-Rad DCode system (Bio-Rad, Hercules, CA, USA). PCR products were loaded on an 8% (wt vol⁻¹) polyacrylamide gel (37.5:1 acrylamide-bisacrylamide) with a 30 to 55% linear denaturing gradient. Electrophoresis was performed in 1× TAE buffer (40 mM Tris-acetate, 1 mM Na-EDTA, pH 8.0) at a constant voltage of 200 V and a temperature of 60°C for 4 h. The gel was stained by incubation in 1 X TAE buffer with ethidium bromide. Bands were visualized with UV light and images were recorded using the Digi Doc Photo Documentation System (Bio-Rad). Dominant bands were excised from the gel using a scalpel and dissolved in 1 X TE buffer at 4°C overnight. The DNA contained in the TE buffer was re-amplified with primers 357F-518R. PCR products were submitted to Macrogen (Amstelveen, The Netherlands) for purification and sequencing.

**Clone libraries**

Eight clone libraries, one for each of the four samples from both the Zaandam and the Berlin populations (see “Experimental Design”), were constructed using the fragments obtained after nested PCR and subsequent amplification with primers 357F-1392R. PCR products were purified with the Wizard SV Gel and PCR Clean-Up System (Promega). DNA fragments were inserted into plasmid pGEM-T and transformed into *Escherichia coli* XL-1 Blue following the T4 DNA Ligase Blue/White Cloning protocol (Promega). For each of the eight transformation reactions, twelve positive transformants were screened by PCR with primers SP6 and T7 to verify the insertion of the 16S rRNA gene fragment. Correctly sized amplicons were then subjected to restriction fragment length polymorphism (RFLP) analysis, by incubating with 10 U of restriction enzyme RsaI at 37°C for 3 h. The digested fragments were visualized on a 1.5% (wt/vol) agarose gel and, based on differences of RFLP patterns, six to eight clones for each transformation were chosen for sequencing. Transformants were cultured at 37°C overnight, after which the plasmid
DNA was isolated with the Wizard Plus SV Minipreps DNA Purification System (Promega). The purified plasmid DNA was sent to Eurofins (Amsterdam, The Netherlands) for sequencing. Raw sequence reads were cleaned using the ContigExpress Module in Vector NTI Software (Life Technologies, Bleiswijk, The Netherlands). BLAST searches were performed using the Sequence Match tool on Ribosomal Database Project (RDP) (Cole et al. 2014). Further phylogenetic analysis was conducted with MEGA (Tamura et al. 2013): sequences were aligned using the ClustalW algorithm, and a maximum likelihood analysis was performed to obtain a phylogenetic tree (Hall 2013).

Illumina sequencing and bioinformatic analysis

After normalization of samples and amplification with Illumina primers (for primers see Table S1 - details on amplification are included in the “Amplification of 16S rRNA genes” section above), barcoded PCR products were cleaned up with the Agencourt AMPure XP system (Beckman Coulter, Brea, CA, USA), and combined in equimolar ratios, after assessing the quality and quantity of the products with a Bioanalyzer System (Agilent Technologies, Santa Clara, CA, USA). Paired-end sequencing was performed on a Illumina MiSeq Sequencing System (Micro Array Facility, VUmc, Amsterdam). Data analysis was performed using QIIME, version 1.8.0 (Caporaso et al. 2010b). Paired end reads were merged and primer sequences were removed. The reads were filtered using Phred quality scores, using a minimum quality score of 20. In QIIME, usearch_qf (usearch quality filter) was used to filter sequences, remove chimeras and cluster sequences in Operational taxonomic units (OTUs) at 97% sequence similarity (Edgar 2010). Singletons (OTUs represented by only one sequence) were discarded. The most abundant sequence in each cluster was chosen to represent that OTU. Representative sequences were aligned using PyNAST (Caporaso et al. 2010a) and assigned taxonomic status using the RDP Classifier 2.2 (Wang et al. 2007). Both steps were performed using the Greengenes 13_8 reference database (DeSantis et al. 2006). The alignment was filtered using the greengenes lanemark. Phylogenetic analysis was performed using FastTree (Price, Dehal and Arkin 2010) and the online tool iTOL was used for tree visualization (Letunic and Bork 2007). The OTU map generated in QIIME using the result of usearch_qf was used for further diversity analyses.

Statistical analyses

DGGE images were analyzed with GelCompar II software package (Applied Maths, Sint-Martens-Latem, Belgium). Similarities between profiles were calculated using Pearson product-moment correlation coefficient. Cluster analysis on the resulting similarity matrix was performed using the unweighted pair group method with arithmetic means (UPGMA).
In QIIME, nonparametric ANOSIM was used to test whether differences in community composition (occurrences of OTUs and their abundances) between groups (Suppression vs. Control and Zaandam vs. Berlin) were significant. A paired t-test (for the paired Suppression vs. Control samples) and two-sample t-test (for Zaandam vs. Berlin samples) were used to test the significance of differences in OTU abundances and richness between groups. A Mantel test was applied to compare dissimilarity matrices obtained from DGGE and Illumina analysis. Alpha diversity was calculated using the Chao1 estimator, after rarefying to the lowest read number between samples. Rarefaction plots were generated for each sample separately and for the samples pooled per population. Differences between samples in community composition were assessed using weighted Unifrac metrics to calculate similarity matrices based on the phylogenetic distances among reads (Lozupone and Knight 2005). Similarities between samples were visualized using principal coordinate analysis (PCoA) plots.

**Accession numbers**

Sequencing data are available in GenBank’s Sequence Read Archive under accession number SRP055488.

### 2.4. Results and Discussion

**Wolbachia suppression**

One of the main objectives of this study was to test PCR approaches that would avoid amplification of abundant Wolbachia DNA, thereby allowing the identification of the less abundant bacterial species in the Folsomia microbiome. To this end, we conducted Illumina amplicon sequencing on pairs of DNA samples subject to either normal PCR protocols, or a pre-amplification using primers designed to suppress amplification of Wolbachia-type sequences. Our intention was to maximize the reproducibility of the method and the coverage of the targeted bacterial communities to better characterize the impact of the different PCR approaches, and later of Folsomia population, on microbial community composition. For this reason we chose the strategy of pooling ten springtails per biological replicate to reduce variation between replicates, with the obvious drawback that we could not use these datasets to assess individual differences in microbial composition of F. candida.

Across all samples, 1 007 238 non-singleton reads were counted, and 719 OTUs were identified. 661 OTUs were found in the Control treatment (average per replicate (M)=315, with standard deviation (SD)=124), 373 in the Suppression treatment (M=224, SD=23). Seven of these OTUs were identified as Wolbachia. One of them accounted for 97.4% of all Wolbachia reads, while the remaining OTUs had low abundances, but were not singletons. Previous research on ants and lepidopterans suggests that multiple infections are possible, with individual insects harboring more
than one *Wolbachia* strain (Russell et al. 2012). Therefore, the multiple *Wolbachia* OTUs found in our study may well represent real variation, suggesting that *Folsomia* can be infected by multiple *Wolbachia* strains, although we cannot exclude the possibility that at least some of these OTUs originated from sequencing errors. Because 73.7% of the reads in the Control corresponded to *Wolbachia*, the suppression treatment was successful in lowering the abundance of *Wolbachia* by more than a factor of ten, to on average 6.9% of the reads.

Principal coordinate analysis (PCoA) on the basis of pair-wise weighted Unifrac distances revealed that the distance between samples of the Suppression and Control groups was reduced when *Wolbachia* sequences were not included in the analysis (compare Figure 1 B to Figure 1 A), but also suggested that the revealed community composition (here referred to as the community members detected and their relative abundances) between the two treatments remained different (Figure 1 B). ANOSIM confirmed that the separation between Suppression and Control groups was significant, whether *Wolbachia* reads were taken into account or not (p < 0.05, 10 000 permutations). This suggests that the effect of the suppression treatment on PCR amplification was not limited to *Wolbachia*, but it extended to other bacterial sequences.

![Figure 1](image.png)

**Figure 1.** Weighted UniFrac 2D PCoA plots illustrating the distance between samples of the springtail *Folsomia candida* subjected to different bacterial 16S rRNA gene amplification procedures. *Wolbachia* reads were either included (A) or excluded (B) in the calculation of distances between samples. Blue dots correspond to samples resulting from the *Wolbachia* suppression treatment; red squares are the samples resulting from normal amplification. The percentages in brackets on the axis indicate the percent of variation explained by that axis.

Figure 2 illustrates the identities and relative abundances of the OTUs that were found in *F. candida* when using either normal amplification or suppression, and that were considered abundant if they accounted for more than 0.2% of total number of
reads. Wolbachia OTUs are not included in the figure. The relative abundances of seven OTUs, corresponding to two Bacillus sp. (Firmicutes, Bacillaceae), two Pseudomonas sp. (Gammaproteobacteria, Pseudomonadaceae), Escherichia sp. (Gammaproteobacteria, Enterobacteriaceae), Olivibacter sp. (Bacteroidetes, Sphingobacteriaceae), and a member of the Actinomycetales (Actinobacteria), were significantly influenced by the suppression treatment (p<0.05). With the exception of Bacillus sp., all these OTUs showed lower relative abundances when suppression was applied compared to normal amplification.

**Figure 2.** Column chart showing the bacterial OTUs identified in samples of the springtail Folsomia candida when using a Wolbachia-suppression step in bacterial 16S rRNA gene amplification (left) and with normal amplification (right). Family names are indicated, unless otherwise stated. Wolbachia reads are excluded from the dataset. The category “other” includes OTUs that account for less than 0.2% of the total number of reads.

Of the 17 highly abundant OTUs shown in Figure 2, 16 were present in all samples, both from the Suppression and the Control group. Only one OTU, identified as Pseudomonas, was absent in two samples of the Suppression group. Therefore, the suppression treatment did not interfere with the identification of the most abundant OTUs.

Among the 719 OTUs detected, 98 OTUs were present in all four Suppression samples, and 95 OTUs in all four Controls (Figure 3). Fifty-nine OTUs were recovered from all eight samples (Figure 3). Most of the reads in both the Suppression (98.2% of total reads) and the Control group (87.5%) belonged to these 59 shared OTUs.
The microbiome of *Folsomia candida*

**Figure 3.** Venn diagrams showing the number of bacterial OTUs shared within and between groups of samples. The groups correspond to different approaches in the 16S rRNA gene amplification of DNA isolated from the springtail *Folsomia candida*: a conventional control and a PCR using primers that suppress the amplification of the 16S rRNA genes of the endosymbiont *Wolbachia*. A: number of shared OTUs between samples for each of the two groups, Suppression and Control. B: number of shared OTUs between the Suppression and the Control groups. *Wolbachia* OTUs were not included in the OTU count.

Only three OTUs were differentially represented between the Suppression and the Control groups, meaning that they were present in all four samples of one of these two groups, and absent in all four samples from the other group. The relative abundances of these OTUs were low: one OTU, assigned to the family Pseudomonadaceae (phylum Gammaproteobacteria), was present in all Control samples (average abundance across samples: 0.026%) and absent from all Suppression samples; in contrast, a member of group SJA4 in the phylum TM6, and an OTU assigned to the family Paenibacillaceae (phylum Firmicutes), were always present when suppression was performed (respective average abundances: 0.003% and 0.001%), and never in the control conditions.

Species richness seemed to be lower in the suppression treatment (Table 1), however, a paired t-test revealed that the average difference in both the observed and estimated number of OTUs between the two treatments (M=90, SD=131, and M=95, SD=111, respectively) was not significant (t (3)=−1.4, two-tailed p=0.3, and t(3)=−1.7, two-tailed p=0.2, respectively), showing that the *Wolbachia* treatment did not significantly affect the number of retrieved OTUs.

Overall, our observations showed that neither the presence of a highly abundant endosymbiont nor the use of a suppression treatment significantly affected the retrieval of bacterial species in our samples. However, the use of the suppression treatment did significantly affect the relative abundances of the identified OTUs, and the resulting community composition. Using suppressive methods to target highly abundant species can still be useful when these bacteria are suspected to interfere with the identification of rare groups. Furthermore, suppression steps can lead to a higher diversity of sequences per sequencing effort, reducing the amount of reads to
discard and therefore allowing a more efficient use of sequencing data. In this study, suppression of the highly abundant endosymbiont *Wolbachia* was achieved through a nested PCR approach. An alternative method involves specific inhibition of PCR amplification through the use of non-extendable primers (Yu *et al.* 1997). We were not successful in applying this method in our study. Dissection is also frequently used in studies of insect-associated gut microbiota (Engel, Martinson and Moran 2012), and in our case it could allow the isolation and study of *Wolbachia*-free tissues. For *F. candida*, however, isolation of the gut or dissection of the ovaria is not practical due to the size of the animal (0.5-2 mm). One could also take advantage of the intracellular location of *Wolbachia* to separate it from the other bacteria. Filters and/or differential centrifugation can be used to separate *Wolbachia*-containing eukaryotic cells from prokaryotic cells, allowing the study of the *Wolbachia*-free prokaryotic fraction only.

**Comparison of bacterial diversity in *F. candida* populations**

To compare population-level bacterial diversity in *F. candida*, we further compared the bacterial composition of two populations of *F. candida*. The two populations were both represented by four biological replicates, each representing ten springtails.

Figure 4 shows a phylogenetic tree summarizing the bacterial diversity found in the two populations of *F. candida*, and the OTUs resulting from Illumina sequencing that are present in the tree are listed in Table S2 (Supplementary Information). A total of 832 OTUs were found across the four samples from the Berlin population and the four samples from the Zaandam population. 689 OTUs were found in the Zaandam population (M=345, SD=63), and 457 in the Berlin population (M=292, SD=15). Seven OTUs were identified as *Wolbachia*, and one of these OTUs accounted for 97.0% of all *Wolbachia* reads in both *Folsomia* strains. The *Wolbachia* suppression approach was used for all samples. *Wolbachia* reads constituted 3.8% of the reads from the Berlin population, but 32.8% of the reads from the Zaandam population, despite the use of the suppression treatment. The difference in relative abundances observed between the two populations of *F. candida* can either be due to higher abundance of *Wolbachia* in the Zaandam population or to lower numbers of other bacteria compared to the Berlin population. The former hypothesis is supported by studies showing that densities of endosymbiotic *Wolbachia* are highly variable among populations of insects. Hoffmann *et al.* (2014) observed significant differences in *Wolbachia* densities among populations of the citrus psyllid *Diaphorina citri*. Similarly, Unckless *et al.* (2009) estimated a 20 000 fold variation of *Wolbachia* densities among natural populations of *Drosophila innubila*. A study by Ahantarig *et al.* (2008) found even higher variations of relative density (bacteria-to-host ratio) of *Wolbachia* in field-collected populations of the mosquito *Aedes*
The microbiome of *Folsomia candida* albopictus, in contrast to lower density variations of the endosymbiont in a previously characterized laboratory strain.

**Figure 4.** Maximum likelihood phylogeny of the bacterial OTUs resulting from Illumina sequencing and cloning of two populations of the springtail *Folsomia candida*. The inner circle and the branch color indicate the bacterial phyla to which the OTUs belong, the middle circle indicates the method used to obtain the sequences, and the outer circle indicates the population (Zaandam or Berlin) where the OTUs were found. The tree was rooted on *Deinococcus* and *Thermus* species, like in a phylogenetic tree obtained in a previous study (Wu et al. 2009). For legibility, some tips represent combinations of a small number of tips, which were given the same taxonomic assignment.

PCoA revealed a clear separation of samples between the two populations, confirmed by ANOSIM (p<0.05, 10 000 permutations), also when excluding *Wolbachia* reads from the calculation of beta-diversity. The first axis (PC1) already separates the two populations completely (Figure S3, Supplementary Information). Since *Wolbachia* sequences were not included in this analysis, this shows that other OTUs contribute significantly to the differences between the communities. The second axis separates the biological replicates within each population, reflecting within-population differences. The plots also suggest that the variation is higher in the Zaandam population, compared to the Berlin population.
Figure 5 shows that the most abundant OTUs and their relative abundances varied between the two populations of *F. candida*. In the Zaandam population, 11 OTUs accounted each for more than 1% of non-*Wolbachia* reads. The most abundant OTUs were identified as (in order of abundance): *Paracoccus* sp. (Alphaproteobacteria, Rhodobacteraceae, 30.3% of reads), a member of the Actinomycetales (9.9%), *E. coli* (6.0%), *Thermomonas* (Gammaproteobacteria, Xanthomonadaceae, 5.1%), and *Bacillus cereus* (4.0%). In the Berlin population, 12 OTUs accounted each for more than 1% of non-*Wolbachia* reads. The most abundant OTUs in this population were *Staphylococcus sciuri* (Firmicutes, Staphylococcaceae, 27.2% of reads), *Paracoccus* sp. (15.4%), *B. cereus* (12.1%), *Ochrobactrum* (Alphaproteobacteria, Brucellaceae, 5.9%) and a member of the Flavobacteriaceae (Bacteroidetes, 5.4%).

![Column chart showing the bacterial OTUs identified in two populations of the springtail *Folsomia candida*, Zaandam (left) and Berlin (right). Family names are indicated, unless otherwise stated. *Wolbachia*-suppression was used on all samples and the *Wolbachia* reads retrieved were excluded from the dataset. The category “other” includes OTUs that account for less than 0.5% of the total number of reads.](image)

Sixteen of the 20 high-abundance OTUs shown in Figure 5 were present in all Berlin and Zaandam samples, indicating that most dominant bacterial taxa were shared between populations. Overall, 81 OTUs were shared between all samples of both populations of *F. candida*, constituting 71.9% of the total number of reads in the Zaandam population and 90.3% in the Berlin one.

In total, 45.1% of OTUs were present only in the Zaandam population, and they accounted for 13.7% of the total reads from that population; instead, 17.2% were found only in the Berlin population, where they accounted for 1.9% of the reads.
The microbiome of *Folsomia candida*

When the search was restricted to OTUs that were present in all samples of one population and absent from all samples of the other, 41 OTUs were differentially represented between two populations. However, these OTUs were not major components of the microbial communities: 15 of these OTUs were present in the Zaandam population with a total relative abundance of 1.0%, and the remaining 26 were characteristic of the Berlin population, for a total abundance of 0.4%.

Table 1 shows that the Zaandam population has a higher variation in species diversity than the Berlin population. A t-test revealed that the average difference in both the observed and the estimated species number between the two populations (M=53, SD=66 and M=67, SD=91, respectively) was not significant (t(3)=1.65, two tailed p=0.2 and t(3)=1.58, two tailed p=0.2, respectively), showing that the species richness is not significantly different between the two populations.

<table>
<thead>
<tr>
<th></th>
<th>Observed diversity</th>
<th>Estimated diversity: Chao1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zaandam</td>
<td>338 ± 63</td>
<td>386 ± 84</td>
</tr>
<tr>
<td>Berlin</td>
<td>285 ± 15</td>
<td>319 ± 10</td>
</tr>
<tr>
<td>Suppression</td>
<td>218 ± 23</td>
<td>297 ± 29</td>
</tr>
<tr>
<td>Control</td>
<td>308 ± 124</td>
<td>391 ± 106</td>
</tr>
</tbody>
</table>

Diversity is expressed as number of OTUs and calculated based on bacterial 16S rRNA genes isolated from *F. candida*. Indicated values are average ± standard deviation (n=4). Means do not differ statistically (p>0.05). *Wolbachia* OTUs were not included in the count.

Despite sharing most of the dominant OTUs, the Berlin and the Zaandam populations were characterized by different bacterial profiles. The community composition of different *Folsomia* populations might reflect strain- or environmental-specific effects. Multiple studies have shown that environmental factors can have a strong effect on host-associated bacterial communities. Staubach *et al.* (2013) studied laboratory-reared and wild populations of two *Drosophila* species, and concluded that environmental factors, specifically food-substrate, have the biggest effect on bacterial communities associated with the flies. Although they could find a host species effect, it was subtle and only detectable in the wild flies. Zouache *et al.* (2009) observed that the microbial community composition of long-term laboratory colonies of the parasitoid wasp *Asobara tabida* is different from that of both natural populations and lines recently established from field-collected insects. The authors concluded that long-term laboratory rearing affects the composition of the bacterial community. Broderick *et al.* (2004) found similar T-RFLP bacterial profiles in gypsy moths from lab and field populations. However, field populations were originated from eggs collected in field sites and subsequently
surface sterilized in the lab, and the authors concluded that diet, rather than egg source, had a significant effect on the bacteria in the gypsy moth. Higher microbial diversity in field populations compared to lab populations has been observed in other insect species, such as *Drosophila* (Chandler *et al.* 2011; Staubach *et al.* 2013) and the ant *Cephalotes varians* (Kautz *et al.* 2013).

In order to focus on population effects, in our study we tried to reduce the impact of the environment of origin on the microbial community composition of the Zaandam population by acclimating it to lab conditions for four months (three to four generations). It may be expected that environmental effects on the microbiota would disappear over this time. However, we cannot exclude a still lasting impact of the original environment in the Zaandam population, and therefore we cannot unequivocally attribute the differences in microbial community composition that we observed to population-effects only.

In addition to Illumina sequencing, we applied DGGE, a conventional microbial community fingerprinting technique, to analyze and compare the Zaandam and the Berlin populations. Our aim was to assess the comparability of the information that these two techniques provided on the bacterial community of *F. candida*.

Bacterial profiles for the samples of the Berlin and Zaandam populations of *F. candida* were obtained through amplification using the suppression treatment for *Wolbachia* DNA and subsequent DGGE analysis of the 16S V3 region. UPGMA-based cluster analysis of the DGGE profiles revealed two clusters at high levels of similarity, corresponding to the two populations (Figure S4, Supplementary Information).

The banding pattern was more complex and heterogeneous in the Zaandam profiles, suggesting a higher bacterial diversity in this population (Figure S4). Furthermore, the differences among samples are higher in the Zaandam population: the overall average similarity between replicates of the Zaandam population was 75%. Within the Berlin population, the overall average similarity between replicates was higher, 90%. A Mantel test indicated a positive correlation between the DGGE and the Illumina dissimilarity matrices, meaning that the outputs of the two methods are in accordance (p<0.05). The dominant bands in all profiles of Zaandam samples corresponded to *Wolbachia, Bacillus* and *Paracoccus*. In the profiles of the Berlin samples, bands corresponding to *Staphylococcus, Bacillus* and *Paracoccus* were dominant. These groups were also among those dominant in the NGS dataset.

We were not able to assign an identity to all the bands in the DGGE profiles, and multiple bands corresponded to the same species, probably as a result of the intraspecies heterogeneity of 16S rRNA genes (Dahllöf, Baillie and Kjelleberg 2000). To confirm the NGS data on the identity and abundances of the dominant species, we prepared clone libraries from the Zaandam and Berlin samples using 1000 bp
The microbiome of *Folsomia candida*

fragments (V3/V8 region) and sequenced 52 clones. We found 19 different clones, 16 of which corresponded to OTUs identified through Illumina sequencing, including seven that were among the abundant OTUs as observed by Illumina sequencing: *Wolbachia*, *Paracoccus*, *Ochrobactrum*, *Delftia*, *Pseudomonas*, *Bacillus* and *Staphylococcus* (Figure 5, contributing 66% of reads). The remaining three (*Haloferula* sp., *Chitinofaga* sp. and a member of the family Chitinofagaceae) did not have corresponding OTUs in the Illumina dataset, although this contained unclassified members of the family Verrucomicrobiaceae and of the order Sphingobacteriales, to which the genus *Haloferula* and the family Chitinofagaceae respectively belong to. Some of the cloning sequences are included in the phylogenetic tree in Figure 4.

Results from DGGE and clone library analysis confirmed the reliability of NGS technology in characterizing our bacterial community. The most abundant bacterial species were represented in the clone libraries, and DGGE confirmed the diversity trend revealed by the Illumina data. Furthermore, Illumina sequencing detected hundreds of OTUs that were overlooked with DGGE and cloning in this and in previous studies. In their survey of bacterial diversity in collembolans based on 16S rRNA PCR and cloning, Czarnetzki & Tebbe (2004a) studied two populations of *F. candida*, both originating from lab cultures. The gene library of one population produced exclusively sequences related to *Rickettsiella*. In the other population, the authors were able to identify 8 different clones, attributed to *Bacillus weihenstephaniensis*, *Stenotrophomonas maltophilia*, *Mesorhizobium* sp., *Paracoccus* sp., *Bacillus* sp., *Paenibacillus* sp., and members of the Bacteroidetes and of the Planctomycetes groups. Thimm *et al.* (1998) found 11 types of agar-cultured bacterial colonies from a lab population of *F. candida*, and identified three of them as *Erwinia amylovora*, *Staphylococcus capitis* and *Pantoea agglomerans*. With the exception of *Pantoea*, all the 8 genera found in these two studies were represented in our NGS dataset, constituting 57.6% of non-*Wolbachia* reads, but in addition we found 269 more genera. This study confirms that high-throughput sequencing qualifies as an adequate tool to identify rare species in host-associated bacterial communities (Kautz *et al.* 2013).

**Acknowledgements**

We thank the Microarray Facility, VU Medical Center, Amsterdam, The Netherlands, for performing Illumina sequencing.
Chapter 2

Supplementary Information

**Figure S1**: Experimental layout

**Figure S2**: 16S rRNA gene fragment

**Figure S3**: PCoA plot of the distance between populations

**Figure S4**: Clustering of DGGE profiles

**Table S1**: Primer sequences

**Table S2**: List of OTUs from Illumina sequencing (available at the online version of this article)
The microbiome of *Folsomia candida*

**Figure S1.** Layout of the experiments reported in this study. For each experiment, the following are indicated: which population of *F. candida* was used, the treatment applied to the samples (normal amplification or amplification to suppress *Wolbachia*), and the methods used.
**Figure S2.** Overview of the 16S rRNA gene fragment. Red arrows indicate the primers used in this study (see Table 1 for more details). The blue arrow indicates the hypervariable region V3. The brackets indicate the fragments obtained through PCR amplification with different primers. The orange bracket indicates the fragment obtained after the first amplification round of a nested PCR, using primers 8F and 1525R to avoid amplification of *Wolbachia* DNA. The second amplification round was performed with primers 357F and 518R (green bracket, fragment was used for DGGE analysis and high-throughput sequencing), or with primers 357F and 1392R (red bracket, fragment was cloned and sequenced).
The microbiome of *Folsomia candida*

**Figure S3.** Weighted UniFrac 2D PCoA plots illustrating the distance between samples from the Zaandam (blue dots) and Berlin (red squares) populations. *Wolbachia* reads were not used to build the plot.

**Figure S4.** Clustering of bacterial 16S DGGE profiles in two populations of *F. candida*. The dendrogram was obtained through UPGMA-based cluster analysis. Cophenetic correlation coefficients are indicated at the root and on the nodes of the dendrogram. Arrows indicate the taxonomic identity of corresponding bands.
Table 2

Table S1. Primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>8F</td>
<td>AGA GTT TGA TYM TGG CTC AG</td>
<td>Lane et al., 1991</td>
</tr>
<tr>
<td>357F*</td>
<td>CCT ACG GGA GGC AGC AG</td>
<td>Muyzer et al., 1993</td>
</tr>
<tr>
<td>518R*</td>
<td>ATT ACC GCG GCT GCT GG</td>
<td>Muyzer et al., 1993</td>
</tr>
<tr>
<td>1392R</td>
<td>ACG GCC GGT GTG TAC A</td>
<td>Lane et al., 1991</td>
</tr>
<tr>
<td>1512R</td>
<td>ACG GYT ACC TTG TTA CGA CTT</td>
<td>Lane et al., 1991</td>
</tr>
<tr>
<td>1525R*</td>
<td>AAG GAG GTG WTC CAR CC</td>
<td>Lane et al., 1991</td>
</tr>
<tr>
<td>GC clamp</td>
<td>CGC CCG CCG CGC CCC GCG CCC GCG CCC GCG CCC CCG CCC C</td>
<td>Sheffield et al., 1989</td>
</tr>
</tbody>
</table>

* these primers, with appropriate barcodes and Illumina adapters, were also used for Illumina sequencing.

** primer R1525 is known to select against Wolbachia sequences
Chapter 3 – Antimicrobial activity and carbohydrate metabolism in the bacterial metagenome of the soil-living invertebrate *Folsomia candida*

Agamennone V, LeNgoc G, van Straalen NM, Brouwer A, Roelofs D - submitted
3.1. Abstract

The microbiome associated with an animal’s gut and other organs is considered an integral part of its ecological functions and adaptive capacity. To better understand how microbial communities influence activities and capacities of the host, we need more information on the functions that are encoded in a microbiome. Until now, the information about soil invertebrate microbiomes is mostly based on taxonomic characterization, achieved through culturing and amplicon sequencing. Using shotgun sequencing and various bioinformatics approaches we explored functions in the bacterial metagenome associated with the soil invertebrate *Folsomia candida*, an established model organism in soil ecology with a fully sequenced, high-quality genome assembly. Our metagenome analysis revealed a remarkable diversity of genes associated with antimicrobial activity and carbohydrate metabolism. The microbiome also contains several homologs to *F. candida* genes that were previously identified as candidates for horizontal gene transfer (HGT). We suggest that the carbohydrate- and antimicrobial-related functions encoded by *Folsomia’s* metagenome play a role in the digestion of recalcitrant soil-born polysaccharides and the defense against pathogens, thereby significantly contributing to the adaptation of these animals to life in the soil. Furthermore, the transfer of genes from the microbiome may constitute an important source of new functions for the springtail.

3.2. Introduction

Microorganisms inhabit every type of environment, and many of them live in association with eukaryotic hosts. These microbes they can influence their host’s ecology and evolution by contributing to a variety of processes such as digestion, immunity, and protection from pathogens (Engel and Moran 2013). Hexapods are good models to study host-associated microorganisms: they constitute the most diverse and abundant group of eukaryotic organisms on earth, and in many cases the establishment of specific microbial symbioses may have provided the key for their evolutionary success. Some hexapods depend on microbial symbionts for nutritional or defensive purposes (Kroiss et al. 2010; Douglas 2016), suggesting that a good understanding of their biology should include the study of their associated microbes. This has been described as a “new imperative for the life sciences” (McFall-Ngai et al. 2013).

The majority of microorganisms is not accessible through traditional culturing techniques (Rappé and Giovannoni 2003), and metagenomic sequencing is an appropriate tool to study microbial diversity in different ecosystems (Streit and Schmitz 2004). Metagenomics of insect-associated microbial communities has provided important insights in the interactions between microorganisms and their hosts, including the discovery of metabolites that may find specific biotechnological
applications. For example, metagenomics of a termite’s gut microbiota has elucidated the mechanisms underlying wood degradation in this environment, while also identifying bacterial enzymes with interesting hydrolytic functions (Warnecke et al. 2007). Other studies have found that microbial symbionts of insects are important sources of novel antimicrobials (Wang et al. 2015).

The springtail *Folsomia candida* Willem 1902 (Hexapoda: Collembola) is a small invertebrate living in soil environments, where it feeds on fungal hyphae, decaying organic material and microorganisms. This springtail species has been used as a test organism in ecotoxicology and in ecogenomics for many years (Fountain and Hopkin 2005) and recently its genome and transcriptome have been sequenced (Faddeeva-Vakhрусheva et al. 2017). It is estimated that approximately 2.8% of the genes in the genome of *F. candida* are of foreign origin, having been acquired from bacteria and fungi through HGT (Faddeeva-Vakhрусheva et al. 2017). Many of these genes are involved in carbohydrate metabolism, specifically in cell wall degradation; these functions may aid the animal in extracting nutrients from polysaccharides resulting from the degradation of plant and fungal biomass in the soil. In addition, several foreign genes are involved in antibiotic biosynthesis (Roelofs et al. 2013; Suring et al. 2017). These genes are strongly induced by stress exposure (Nota et al. 2008; Suring et al. 2016), and, while their function still needs to be investigated, it is hypothesized that they may be involved in regulatory mechanisms determining the composition of gut microbial communities (Thimm et al. 1998), or in defense mechanisms protecting these animals from pathogens. Microbial pathogens are abundant in soil environments, and *F. candida* has been shown to be non-susceptible to such pathogens (Broza, Pereira and Stimac 2001; Dromph and Vestergaard 2002). Recently, we have shown that bacteria isolated from springtails are able to suppress a wide array of pathogens, indicating a potential for antimicrobial production (Agamennone et al. 2018). This suggests that gut bacteria may be an important factor interacting with *F. candida*, possibly providing physiological traits advantageous to thrive in a microbe-dominated environment such as the soil.

In this paper, we provide the first functional description of the gut bacterial community of a springtail based on a whole-metagenome sequencing approach. We observed a number of genes involved in carbohydrate metabolism, antibiotic resistance and secondary metabolite biosynthesis. These functions may constitute beneficial traits for an animal living in the soil environment, and may also represent good targets for drug discovery and for the development of biotechnological applications. We also performed a comparative analysis between genes of the gut microbiome and foreign genes in *F. candida*, to identify functions possibly assimilated by the host through HGT.
3.3. Materials and methods

Test organism

*Folsomia candida* were cultured in plastic boxes with a bottom of plaster of Paris and charcoal. Cultures were kept in climate rooms at 20°C temperature, 75% humidity and a 12 hour light-dark cycle. The springtails were fed dry baker’s yeast (Dr. Oetker, Bielefeld, Germany), and they were starved for 2 days prior to DNA isolation.

Sample preparation and DNA isolation

DNA was isolated from four different source samples. Two samples (Fc1 and Fc3) consisted of guts dissected from *F. candida* individuals; one sample (Fc4) consisted of whole springtails; one sample (Fc2) consisted of a mixture of whole animals and dissected guts. Dissected guts were rinsed in sterile PBS and whole springtails were rinsed three times in sterile water before processing. After the washing steps, DNA was directly isolated from two of the samples (Fc3 and Fc4) while additional steps were applied to prepare samples Fc1 and Fc2. For these two samples, we separated bacterial cells from *F. candida*’s cells by using the method described by Engel *et al.* (2012), with modifications. The samples were crushed in PBS in a 1.5 ml microcentrifuge tube, using a plastic pestle. The samples were then gently vortexed, to encourage separation of cells, before being passed through 20 µm and 8 µm filters in succession. The filtered samples were centrifuged at 10 000 g for 30 min to harvest cells, and the pellet was resuspended in 200 µl TE buffer. For sample Fc2, an additional step with a density gradient was applied. An 80% Percoll solution in 0.15 mol l⁻¹ NaCl was prepared. 1 ml of this solution was placed in a 2 ml microcentrifuge tube and spun at 20 000 g for 20 min to create a gradient. The 200 µl of TE buffer containing the cells was gently placed on top of the gradient, and the tube was centrifuged at 400 g for 20 min. Bacterial cells were then visible as a band and were collected using a pipette. The cells were centrifuged at max speed for 5 min and washed with TE buffer to remove residual Percoll solution. DNA was extracted from all samples using the PowerSoil DNA Isolation Kit (MOBIO Laboratories Inc., Carlsbad, CA, USA) and quantified using a Qubit 2.0 (Invitrogen, Carlsbad, CA, USA).

Library preparation and sequencing

Metagenomic libraries for the four samples were prepared using the TruSeq Nano DNA Library Preparation Kit (Illumina Inc., San Diego, CA, USA) with the following modifications. First, genomic DNA (250 ng) was sheared in a Covaris S2 (Covaris Inc., Woburn, MA, USA) with the following settings: duty cycle 10%, intensity 5.0, bursts per second 200, duration 300 s, mode frequency sweeping, power 23W, temperature 5.5°C to 6°C. Fragmented DNA was cleaned using Agencourt AMPure XP beads (Beckman Coulter Inc., Brea, CA, USA) to remove short fragments. After
end repair, cleaning was performed again to select the appropriate library size (180 bp). Then, 3’ end adenylation and adapter ligation were performed, and the ligated fragments were subjected to two rounds of clean-up. PCR was used to enrich the ligated DNA fragments. The PCR program started with 3 min at 98°C, followed by eight amplification cycles (20 s at 98°C, 15 s at 60°C and 30 s at 72°C) and a final extension step of 5 min at 72°C. The amplified library was cleaned and its quality was assessed with a Bioanalyzer on a DNA 7500 chip (Agilent Technologies, Santa Clara, CA, USA). Finally, libraries were equimolarly combined and the concentration of the final pool was checked using a High Sensitivity DNA chip. 10 pmol of barcoded DNA was sequenced on an Illumina HiSeq 2500 using 125 base, paired end run mode.

**Data analysis**

Raw reads of the four samples obtained from the sequencer were trimmed using Trimmomatic version 0.36 (Bolger, Lohse and Usadel 2014) to remove adapters and low quality reads, with the following options: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:36. Metaphlan2 was used to characterize the taxonomic profile of the metagenome (Truong et al. 2015). Bowtie2 (Langmead and Salzberg 2012) was used to create reference genomes for *Folsomia candida* (BioProject accession: PRJNA299291) (Faddeeva-Vakhrusheva et al. 2017), *Wolbachia pipientis* (BioProject accession: PRJNA300838) (Faddeeva-Vakhrusheva et al. 2017), *Saccharomyces cerevisiae* (Assembly accession: ASM105121v1) and *Homo sapiens* (Assembly accession: GRCh38.p7), and to align and identify reads originating from these organisms in the metagenome. SAMtools was used to remove aligned reads from the metagenome. This program was also used to merge all the four sequencing samples together for comprehensive bioinformatic analysis (Li et al. 2009). Only paired ends were extracted with Bedtools (Quinlan and Hall 2010). FastQC (Andrews 2010) was used to check the quality of the reads at different processing stages. Assembly was done using SPAdes version 3.9.0 with the (--meta) setting for metagenomic and k-mer values 21, 41, 65, 75, 87, 91, 95. This range of K-mer was found to give the best assembly result (Bankevich et al. 2012). The quality of contigs was checked with Quast 4.2 (Gurevich et al. 2013). Prodigal (version 2.6.3) was used for genes prediction with the option -m-p meta for predicting metagenomic genes with no gaps (Hyatt et al. 2010). Taxonomic assignment was done using Metaphlan2. The predicted proteins were uploaded to GhostKOALA webservice for KEGG assignment (Kanehisa, Sato and Morishima 2016). For functional annotation, blastp was performed against the Swiss-Prot, refseq and NR databases, with a threshold e-value of 1e-6. InterProScan5 was used with the addition of panther database to identify protein domains using HMM model (Quevillon et al. 2005). Blast2GO was used to integrate the blastp and interproscan results for further improving functional
Chapter 3

annotation (Götz et al. 2008). HMMER version 3.0 was used with CAZy database (version 6) using HMM model to identify carbohydrate-active genes (Lombard et al. 2014). These genes were subjected to filtering using an e-value threshold of 1e-5 for alignments over 80aa, and a threshold of 1e-3 for shorter alignments. The CARD database was used to identify resistance genes (Jia et al. 2017). All the amino acid sequences of anti-resistance proteins were merged and subjected to blastp with a threshold e-value of 1e-6. All the sequences with more than 60% identity with their top blast hit were collected. Descriptions of the ARO terms was obtained from the online database (https://card.mcmaster.ca/). The KEGG, Pfam and NR databases were used to confirm the accuracy of the functional annotations obtained with CAZY and CARD. Secondary metabolite biosynthetic gene clusters were identified for contigs larger than 3 000 bp using the antiSMASH2 program (Weber et al. 2015). To identify homologies and orthologies between the genome of F. candida and the metagenome, a reciprocal blast was performed. The metagenomic protein sequences were blasted against the host proteins, and vice versa. Sequences that were top hits of each other were extracted using a homemade script, and those matching F. candida's foreign genes were identified (Faddeeva-Vakhrusheva et al. 2017).

Data deposition

The raw sequencing data was deposited in NCBI’s Sequence Read Archive (SRA) under accession number SRP149127. The Whole Genome Shotgun (WGS) project was deposited at DDBJ/ENA/GenBank under accession number QIRE00000000. The version described in this paper is version QIRE01000000.

3.4. Results

Sequencing results, assembly and annotation

Table 1 summarizes the sequencing results by indicating, for each sample, the preparation method used and the number of raw and filtered reads obtained. Approximately 90% of the reads passed the trimming step. Most of these reads (more than 97%) originated from the host Folsomia candida, and were removed during the next filtering step along with reads from Wolbachia pipiens, Saccharomyces cerevisiae and human DNA. The proportion of reads of prokaryotic origin was slightly higher in dissected gut samples compared to whole springtail samples (compare sample Fc3 to Fc4), and it was much higher in samples treated with the cell-separation method compared to untreated samples (compare sample Fc2 to Fc4, and Fc1 to Fc3). When combining dissection and cell-separation, the proportion of prokaryotic reads increased by a factor 5 (compare sample Fc1 to Fc4). The lowest proportion of Wolbachia was observed in the FC3 sample (untreated dissected guts).
Table 1. Preparation method and number of raw and filtered reads obtained for each sample.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Sample type</th>
<th>Sample preparation method</th>
<th>Raw reads</th>
<th>Reads after trimming</th>
<th>Reads after bowtie</th>
<th>Filtered reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fc1</td>
<td>Dissected guts (~1 000)</td>
<td>Filter and DNA isolation</td>
<td>138 555 106</td>
<td>121 428 759 (87.6%)</td>
<td>3 605 008 (2.6%)</td>
<td></td>
</tr>
<tr>
<td>Fc2</td>
<td>Whole springtails (300) and dissected guts (~400)</td>
<td>Filter + Percoll and DNA isolation</td>
<td>133 586 006</td>
<td>116 187 374 (87%)</td>
<td>1 811 553 (1.36%)</td>
<td>5 806 361 (1.23%)</td>
</tr>
<tr>
<td>Fc3</td>
<td>Dissected guts (250)</td>
<td>Direct DNA isolation</td>
<td>103 864 717</td>
<td>93 503 412 (90%)</td>
<td>535 052 (0.52%)</td>
<td></td>
</tr>
<tr>
<td>Fc4</td>
<td>Whole springtails (60)</td>
<td>Direct DNA isolation</td>
<td>94 686 416</td>
<td>84 746 773 (89.5%)</td>
<td>372 193 (0.39%)</td>
<td></td>
</tr>
</tbody>
</table>

For each sample, the number of raw reads and the numbers of reads surviving each processing step is indicated. The percentages in bracket indicate the numbers of reads after each step relative to the number of raw reads.

A total of 5 806 361 high quality paired reads was used for assembly, which resulted in 107 138 contigs with a total length of 69 Mb (Table 2). Prodigal predicted 147 851 protein-coding sequences (CDSs), and 133 594 of these were annotated in Swiss-Prot (Supplementary File 1). 132 657 genes (99%) were of bacterial origin, 665 genes were annotated as Eukaryota, 209 as viruses, 33 as Archaea, 30 as vectors or uncultured microorganisms and 14 257 were unassigned. Supplementary Figure 1 shows the length distribution of the contigs. The 20 longest contigs (more than 100 000 bp each) were assigned either to either *Pseudomonas* or *Microbacterium*.

Table 2. Results of assembly and annotation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of contigs</td>
<td>107 138</td>
</tr>
<tr>
<td>Largest contig (bp)</td>
<td>1 306 495</td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>69 108 988</td>
</tr>
<tr>
<td>N50 (bp)</td>
<td>2 514</td>
</tr>
<tr>
<td>N75 (bp)</td>
<td>853</td>
</tr>
<tr>
<td>L50</td>
<td>1 835</td>
</tr>
<tr>
<td>L75</td>
<td>10 181</td>
</tr>
<tr>
<td>GC%</td>
<td>60.2%</td>
</tr>
<tr>
<td>Gene count</td>
<td>147 851</td>
</tr>
<tr>
<td>Genes with function prediction</td>
<td>133 500</td>
</tr>
</tbody>
</table>
**Taxonomic classification**

The dominant bacterial taxa in the metagenome of *F. candida* were Proteobacteria (50% of the reads), Actinobacteria (32%), Bacteroidetes (12%) and Firmicutes (6%) (Figure 1). These phyla constituted 99.5% of all the reads. 35 additional phyla were found in the remaining 0.5% of reads. 826 bacterial genera (excluding singletons) were identified. 23 of these genera covered 83% of the reads. The most abundant genus was *Microbacterium* (Actinobacteria, 13.1% of the reads), followed by *Paraburkholderia* (Betaproteobacteria, 7.2%), *Pseudomonas* (Gammaproteobacteria, 6.3%), *Staphylococcus* (Firmicutes, 5.6%), *Sphingopixis* (Alphaproteobacteria, 5.5%), *Stenotrophomonas* (Gammaproteobacteria, 5.4%), *Pseudoxanthomonas* (Gammaproteobacteria, 5.4%), *Gordonia* (Actinobacteria, 4.1%), *Burkholderia* (Betaproteobacteria, 3.4%) and 14 other genera each with a relative abundance higher than 1%. The overview of the identified taxonomic groups at the phylum, class and genus level is given in Supplementary Figure 2.

![Phylogenetic distribution of the bacterial community in the metagenome of *F. candida*. The size of the circles is proportionate to the abundance of the taxa. The phylogeny was built based using Metaphlan on high quality raw reads.](image-url)

**Figure 1.** Phylogenetic distribution of the bacterial community in the metagenome of *F. candida*. The size of the circles is proportionate to the abundance of the taxa. The phylogeny was built based using Metaphlan on high quality raw reads.
Overall functional analysis

Comparison of the genes with the KEGG database recovered a number of functions. The most abundant functional categories were associated with membrane transport, signal transduction, carbohydrate and amino-acid metabolism, and the genetic information processes replication and repair and translation (Figure 2A).

Mapping of the functions on the phylogenetic tree shows that most predicted genes within any functional category are assigned to few bacterial species, namely the Proteobacteria *Acinetobacter johnsonii*, *A. luofii*, *Pseudomonas stutzeri*, *Paraburkholderia phytofirmans*, *Azoarcus toluclasticus*, *Sphingopixis alaskensis*, the Actinobacteria *Gordonia araii*, *Cutibacterium acnes* and three *Propionibacterium* species, and the Firmicutes *Staphylococcus equorum* (Figure 2B). The next sections present the functions related to carbohydrate metabolism, secondary metabolite production and antibiotic resistance identified in *F. candida*’s microbiome.
Figure 2. Functional annotation. A: detailed representation of the functional classes belonging to different categories. B: functions mapped on the phylogenetic tree. The bars indicate the abundance of genes encoding proteins for each functional category.

**Carbohydrate metabolism**

Carbohydrate metabolism was investigated by comparing predicted genes in *F. candida*’s microbiome with the carbohydrate-active enzymes (CAZY) database. 2004 genes were predicted to code for enzymes involved in carbohydrate metabolism. 1988 (99.2%) of these genes were of bacterial origin and they mostly originated from Proteobacteria (43%) and Actinobacteria (36%). The complete list of CAZymes is presented in Supplementary File 2, and an overview of the pathways involved in sucrose and starch metabolism is given in Figure 3.
The metagenome of *Folsomia candida*

Figure 3. Diagram of the pathways involved in starch and sucrose metabolism. Pink boxes indicate the genes identified in the microbiome.

The carbohydrate-related genes were assigned to five CAZy classes and three modules (Figure 4). 664 genes were identified as glycosyltransferases (GT, 33.1% of the total), 598 as glycoside hydrolases (GH, 30%), 420 as carbohydrate esterases (CE, 21%) and 206 as carbohydrate-binding modules (CBM, 10.1%). The GT, GH and CE CAZymes classes were overrepresented in the metagenome compared to the genome of *F. candida* (data not shown). Instead, enzymes with a carbohydrate-binding module (CBM) were more abundant in the genome of the host. 23 of the genes encoding carbohydrate-active enzymes had a best reciprocal blast hit against foreign genes in the genome of *F. candida*.

Figure 4. Column chart indicating the distribution of CAZy domains among the bacterial phyla retrieved in the metagenome. CBM: carbohydrate-binding module; CE: carbohydrate esterase; GH: glycoside hydrolase; GT: glycosyltransferase; AA: auxiliary activity; PL: polysaccharide lyase.
Secondary metabolites

We screened the gut microbiome for the presence of secondary metabolite biosynthesis pathways related to antimicrobial activity. In total, 166 pathways were identified, 96 of which are putatively involved in the production of an unknown type of secondary metabolite (Table 3). 32 pathways are related to saccharide or fatty acid containing metabolites, while one cluster showed similarity to metabolites with both a saccharide and fatty acid component. Thirteen clusters are represented by non-ribosomal protein synthases (NRPS), which encode multi-domain and multifunctional enzymes involved in the biosynthesis of a large class of biologically active natural products. Another group of ribosomally-synthesized antimicrobial peptides, bacteriocins, are represented by four biosynthetic clusters. We also identified known antibiotics classes among the antismash clusters, namely rifamycin, spectinomycin, chalcomycin, and the antifungal antibiotic bacillomycin.

<table>
<thead>
<tr>
<th>Cluster type</th>
<th>Number of contigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arylpolyene</td>
<td>1</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>4</td>
</tr>
<tr>
<td>Cf_fatty_acid</td>
<td>14</td>
</tr>
<tr>
<td>Cf_fatty_acid -Cf_saccharide</td>
<td>1</td>
</tr>
<tr>
<td>Cf_putative</td>
<td>96</td>
</tr>
<tr>
<td>Cf_saccharide</td>
<td>18</td>
</tr>
<tr>
<td>Cf_saccharide-Cf_fatty_acid</td>
<td>1</td>
</tr>
<tr>
<td>Ectoine</td>
<td>1</td>
</tr>
<tr>
<td>Hserlactone</td>
<td>3</td>
</tr>
<tr>
<td>Nrps</td>
<td>13</td>
</tr>
<tr>
<td>Nrps-Arylpolyene</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
</tr>
<tr>
<td>Siderophore</td>
<td>2</td>
</tr>
<tr>
<td>T1pks</td>
<td>1</td>
</tr>
<tr>
<td>T1pks-Nrps</td>
<td>1</td>
</tr>
<tr>
<td>T3pks</td>
<td>1</td>
</tr>
<tr>
<td>T3pks-Cf_saccharide</td>
<td>1</td>
</tr>
<tr>
<td>Terpene</td>
<td>5</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>166</strong></td>
</tr>
</tbody>
</table>

For each type of secondary metabolite cluster, the number of contigs in which the cluster was detected is indicated. The complete output of the antiSMASH analysis is given in Supplementary File 3.
**Antibiotic resistance**

Predicted genes were mapped against the CARD database to determine the occurrence of antibiotic resistance genes (ARGs) in the gut microbiome of *F. candida* (Jia *et al.* 2017). The analysis recovered 811 genes, corresponding to 209 unique terms in the CARD database. Figure 5 provides an overview of the identified antibiotic resistance mechanisms and of the drug classes to which resistance is conferred. The complete list of genes with accession and classification in CARD is provided in Supplementary File 4. Most antibiotic resistance mechanisms retrieved involved antibiotic target alteration (52%), followed by efflux processes (33%) and antibiotic target replacement (8%). The most abundant class of antibiotics associated with resistance was that of fluoroquinolones (16%), followed by aminocoumarins (10%), peptide antibiotics, lipopeptide antibiotics and tetracyclines (9% each), macrolides, beta-lactams and rifamycin (5% each). Several classes of ARGs involved in resistance to clinically relevant antibiotics, such as β-lactams and tetracycline, were identified (Figures 5 and 6).

![Figure 5](image-url). Overview of the drug mechanisms (left) and classes (right) associated with antibiotic resistance recovered in the metagenome of *F. candida*. The data was obtained by mapping predicted genes against the CARD database and by extracting the “resistance mechanism” and “drug class” categories from the results.
**Figure 6.** Genes from the *F. candida*’s metagenome predicted to be involved in β-lactam resistance are represented as colored items in KEGG’s β-lactam resistance pathway.

**Host-microbiome interaction and horizontal gene transfer**

A reciprocal blast was performed between the *F. candida* genome and the metagenome, to identify homologies between the springtails’ genome and metagenome. The list of best reciprocal blast hits was then compared with the list of 809 foreign genes in the genome of *F. candida*, to identify genes that may have undergone HGT from the gut microbiome into the host genome.

1 204 genes in the microbiome had a best reciprocal blast hits in the host genome. Most of these genes are involved in basic metabolic functions that are highly conserved across most life forms, such as transcription, translation, fatty acid metabolism, chaperone activity, amino acid biosynthesis, nucleic acid biosynthesis and ATP biosynthesis. Of these 1 204 genes, 113 had a best reciprocal blast hit against one of the 809 foreign genes in *F. candida* (Figure 7). The complete list of these 113 genes is given in Supplementary File 5. Taxonomic and functional annotation suggests that *Pseudomonas*, *Microbacterium* and *Gordonia* may be the potential donors of 26, 12 and 9 genes respectively, accounting for almost 50% of them (Supplementary File 5). 23 of the 113 genes are CAZymes. Furthermore, we
identified a non-ribosomal peptide synthase potentially involved in bacteriocin synthesis, a polyketide synthase and several enzymes associated with detoxification, such as monooxygenases ABC transporters, glutathione-S-transferases, and copper oxidase. Most of the 71 remaining annotated genes are related to basic metabolic processes.

**Figure 7.** Venn diagram showing overlap (best-reciprocal blast hits) between proteins from *F. candida*’s genome (light-blue) and proteins from its gut microbiome (green). The red circle contains the horizontally transferred genes, and the number in red indicates the overlap with the gut microbiome.

### 3.5. Discussion

**Sequencing results, assembly and annotation**

In this study, we applied both dissection and a cell separation method to enrich the bacterial component of springtail samples, with the aim of increasing the proportion of bacterial reads after sequencing. The cell separation method was developed by Engel *et al.* (2012) to enrich the bacterial fraction of a sample, and it was more effective than dissection when applied to *F. candida*. Although dissection normally helps to effectively target the microbial component (Gontang *et al.* 2017), this may be more complicated in microarthropods such as springtails because of their small size. A combination of dissection and cell separation method proved to be most effective in increasing the proportion of prokaryotic reads. Still, more than 97% of the reads in any sample belonged to the host *Folsomia candida*, which is probably not surprising: recovery of genetic material from symbiotic microorganisms can be problematic in microhabitats such as insect guts, due to the much higher abundance of host DNA (Paula *et al.* 2016).

Cell separation was also expected to reduce the amount of *Wolbachia* DNA in the samples. *Wolbachia* is an endosymbiont that dominates the bacterial population in *F. candida* (Agamennone *et al.* 2015). Because of the intracellular location of *Wolbachia* (gut epithelium, ovaries and brain), a method that separates the eukaryotic cells from the prokaryotic ones without lysing them should be effective in
reducing the amount of host and Wolbachia DNA in the same step. However, in this study dissection was more effective than cell separation in reducing the occurrence of the endosymbiont in the samples, probably because it allowed to discard organs containing high amounts of Wolbachia.

The number of contig and total length after assembly are comparable to other soil invertebrate-associated metagenomes (Suen et al. 2010; Cheng et al. 2013; He et al. 2013). Although this was not attempted here, it may be possible to recover the genome of one or more species using the data collected in this study (Sangwan, Xia and Gilbert 2016).

**Taxonomic classification**

With 826 bacterial genera identified, the level of diversity in *F. candida* approaches that described in the hindgut of termites, wood-feeding insects that have one of the most complex microbiota of any animal group (Bourguignon et al. 2018). Other soil invertebrates are characterized by comparable or even higher levels of microbial diversity. For example, Pass et al (2015), studied the microbiome of the earthworm *Lumbricus rubellus* and found no less than 9 120 host-specific OTUs. This very diverse community was dominated by Proteobacteria and Actinobacteria, totalling 80% of all OTUs, very similar to the situation in *F. candida*. High diversity was also observed in the gut of two cockroach species, with approximately 1 000 OTUs (Berlanga et al. 2016), whereas slightly lower counts were detected in the ant *Cephalotes varians* (445 OTUs), in the compost worm *Eisenia fetida* (338 OTUs) and in the isopod *Armadillidium vulgare* (153 OTUs) (Kautz et al. 2013; Dittmer et al. 2016; Liu et al. 2018).

The bacterial community in *F. candida* was dominated by species belonging to the phylum Proteobacteria, and within this group members of the Gammaproteobacteria were particularly abundant (21% of the reads). Proteobacteria, a large taxon of functionally diverse bacteria, dominate the microbiome of terrestrial insects and other soil invertebrates such as earthworms, nematodes and isopods (Yun et al. 2014; Pass et al. 2015; Berg et al. 2016; Bouchon, Zimmer and Dittmer 2016; Esposti and Romero 2017). *Pseudomonas*, one of the most abundant bacteria detected in *F. candida*, is commonly found in the microbiome of soil arthropods like termites, ants and beetles and in their environment (Aylward et al. 2014; Esposti and Romero 2017), and was also identified in isopods and nematodes (Dittmer et al. 2016; Liu et al. 2018). *Pseudomonas*, together with *Rickettsia* and *Chryseobacterium*, was also the most abundant OTUs in the microbiome of the springtail *Orchesella cineta* (Bahrndorff et al. 2018). Another abundant bacterium in *F. candida* was *Paraburkholderia*. This genus includes many soil species, a few of which are used as plant probiotics thanks to their growth-promoting and possibly defensive properties (Rahman et al. 2018). Other members of the Proteobacteria identified in *F. candida*’s microbiota were
The metagenome of *Folsomia candida*

*Sphingopixis, Stenotrophomonas, Pseudoxanthomonas, Burkholderia,* all of which were detected in soil invertebrates (worms, cockroaches, termites, ants and beetles) (Esposti and Romero 2017). The most abundant bacterium in *F. candida* was *Microbacterium*. Members of the Microbacteriaceae have been previously identified in different species of beetles (Kelley and Dobler 2011; Scully *et al.* 2013), and Actinobacteria in general (although in low amounts) have been found in cockroaches (Gontang *et al.* 2017) and in a few species of insects (ants, beetles and termites) characterized by nutritional symbioses with fungi (Kautz *et al.* 2013; Aylward *et al.* 2014). Actinobacteria are also one of the dominant bacterial groups in other soil invertebrates such as earthworms (Pass *et al.* 2015; Ma *et al.* 2017; Liu *et al.* 2018).

The observed bacterial diversity in *F. candida* is comparable to that previously detected by 16S high-throughput sequencing in the same lab-reared population of springtails (Agamennone *et al.* 2015). However, the taxonomic distribution between the two studies is very different. Based on 16S sequencing, *Pseudomonas* was the most abundant bacterial genus with 42% of the reads (Agamennone *et al.* 2015). Nine other dominant OTUs were identified, including *Bacillus* (19% of the reads), a member of the Actinomycetales (9%), *Escherichia sp.* (4%), and *Ochrobactrum sp.* (3%). *Microbacterium* accounted only for 0.3% of the read, and *Paraburkholderia* was not identified. This discrepancy can be explained by the difference in sequencing methods applied. High-throughput amplicon sequencing is subjected to PCR bias, with differences in the amplification efficiency of DNA from different bacterial species; in shotgun metagenomic sequencing, on the other hand, biases can be caused by using all genes, rather than only markers genes, for taxonomic assignment, possibly leading to misidentifications (Tessler *et al.* 2017).

**Overall functional analysis**

The majority of reads in *F. candida*’s metagenome originated from pathways involved in membrane transport, carbohydrate and amino acid metabolism, replication, translation and repair. The abundance of genes involved in carbohydrate and amino acid metabolism may suggest a nutritional role of the microbiota. Carbohydrate-related functions are often enriched in the gut microbiome of soil invertebrates like beetles, nematodes and isopods (Cheng *et al.* 2013; Scully *et al.* 2013; Brune and Dietrich 2015; Bouchon, Zimmer and Dittmer 2016; Smith *et al.* 2017), some of which rely on symbiotic microbes for the breakdown of long polymers such as lignin, cellulose and other plant-derived products (Cheng *et al.* 2013; Brune and Dietrich 2015). *F. candida* is an euedaphic springtail species whose diet includes decaying plant material, and it could benefit from the presence of such catalytic functions in its microbiome. Recently, the microbiota of another springtail species, the epiedaphic *Orchesella cincta*, was studied, and some of the main functions predicted based on the microbial community structure were related to the breakdown of dietary components and of plant secondary metabolites (Bahrndorff
et al. 2018). Amino acid-related functions may also be beneficial for the host. Some intracellular endosymbionts biosynthesize essential amino acids that are lacking in their diet of their host (Douglas 2016), and gut bacteria may exert similar functions (Leitão-Gonçalves et al. 2017). A contribution to the host’s nutrition may also explain the abundance of functions related to membrane transport in *F. candida*. Transport allows host-symbiont exchanges and therefore it constitutes one of the most important functions in the maintenance of the symbiosis with bacteria providing nutrients (Charles et al. 2011).

In accordance with the taxonomic assignment, most genes in the above discussed categories were predicted to belong to Proteobacteria and Actinobacteria species. Many genes were annotated to *Acinetobacter johnsonii*, a member of the Gammaproteobacteria that has been described as an opportunistic pathogens for animals as well as a possible reservoir of antibiotic resistance genes (Montaña et al. 2016; Tian et al. 2016). *Acinetobacter* was also a dominant genus in the microbiome of the earthworm *Eisenia fetida* (Liu et al. 2018), and it was identified in other soil invertebrates such as the *Longitarsus* beetle and the isopod *Armadillidium vulgare* (Kelley and Dobler 2011; Dittmer et al. 2016). Many functions were also assigned to the genus *Propionibacterium*. This group of Actinobacteria it includes species with good probiotic potential due to their capacity to modulate microbiota, gut metabolic activity and the immune system (Cousin et al. 2010). Interestingly, the immunomodulatory and anti-inflammatory properties of *Propionibacterium* have been observed not only in human and mouse models (Cousin et al. 2010), but also in soil invertebrates (Kwon, Lee and Lim 2016). An abundance of genes was taxonomically assigned to a few other groups, among which *Gordonia*, a genus of Actinomycetes including many symbionts of terrestrial invertebrates (Sowani, Kulkarni and Zinjarde 2018), and *Pseudomonas*, commonly found in soils and in soil invertebrates (Esposti and Romero 2017).

**Carbohydrate metabolism**

Carbohydrate-degrading enzymes are commonly found in the digestive tract of these organisms, for example in the bovine rumen (Jose et al. 2017), in the gut of wood-feeding insects such as termites and woodwasps (Warnecke et al. 2007; Adams et al. 2011) and in the microbial community of fungus gardens associated with leaf-cutter ants (Aylward et al. 2012). These enzymes are often of microbial origin, suggesting that herbivorous animals can exploit the catalytic activities of microbial symbionts to access nutrients stored in plant biomass (Suen et al. 2010). In termites, the symbiotic relationship with a complex community of bacteria, archaea and protists in the gut enables the digestion of lignocellulose, conferring these insects a unique ecological position in tropical and subtropical ecosystems (Brune and Dietrich 2015). Whether similar relationships between Collembola and their microbiome exist is unknown at the moment, but microbial functions related to carbohydrate
metabolism are likely to significantly contribute to the ecological role of springtails as members of the soil decomposer community.

Warnecke et al. found 700 glycoside hydrolase (GH) catalytic domains found corresponding to 45 CAZY families (Warnecke et al. 2007). In the microbiome of F. candida, we identified a comparable number of genes encoding for enzymes with a capacity to break down long chain carbohydrates such as starch, lignin and cellulose. In nature, these enzymes may aid F. candida in extracting nutrients from the plant biomass that constitutes part of its diet, as was suggested for the springtail O. cincta (Bahrndorff et al. 2018).

A large number of glycoside hydrolases was also observed among F. candida foreign genes (Faddeeva-Vakhrusheva et al. 2017). Interestingly, some of the foreign genes that were also best reciprocal hits between the genome and the metagenome of F. candida were identified as CAZymes (Supplementary File 2). HGT of cellulose-degrading enzymes has been previously observed in plant-feeding insects (Pauchet and Heckel 2013) and may be an important mechanism providing soil invertebrates with advantageous traits for living in the soil (Eyun et al. 2014).

Secondary metabolites

The microbiome of F. candida contained several pathways responsible for the biosynthesis of secondary metabolites. This is a class of compounds that are often involved in competition and interaction between species and that may find applications in the biotechnological and medical sector. The main contributors to the identified pathways seem to be Gordonia, Pseudomonas fluorescens, Bacillus and Streptomyces.

A few of the identified pathways were represented by NRPSs, a class of enzymes responsible for the biosynthesis of natural products with a broad range of biological activities and pharmaceutical properties. Cluster 10 and 28 show resemblance with an NRPS producing pyoverdines, siderophores well known for their high affinity for Fe³⁺ under low iron availability (Schalk and Guillon 2013). Another NRPS involved in the biosynthesis of the siderophore nocobactin was identified in antismash cluster 95. Three clusters show homology to NRPSs involved in antibacterial and antifungal activity. Cluster 31 shows substantial similarity (47%) to an NRPS producing orfamide, a compound of bacterial origin with antifungal properties and with good potential as biocontrol agent against fungal pathogens (Ma et al. 2016). Cluster 130 represents an NRPS involved in microsclerodermin biosynthesis, an antifungal compound produced by a marine sponge (Zhang et al. 2012). A recent study also showed that this compound has properties of pharmaceutical relevance, as it can inhibit NFκB transcription in a human pancreatic cell line leading to apoptosis (Guzmán et al. 2015). Finally, the NRPS identified in cluster 48 showed similarity to the NRPS involved in biosynthesis of the antibiotic caryoyencin, a compound
originally isolated from a plant pathogen. Very recently it has been shown that this compound is produced by a symbiont of a herbivorous beetle, protecting its eggs against detrimental microbes (Flórez et al. 2017).

We also identified a number of bacteriocins, a class of compounds with potential as natural food preservative (Gálvez et al. 2007). Many bacteriocins are biosynthesized by lactic acid bacteria, and in Folsomia’s gut microbiome these clusters are homologous to Pseudomonas fluorescens and Gordonia effusa.

Several other interesting biosynthesis clusters with functions related to medical applications were found, such as lymphostin, a known immunosuppressant isolated from Streptomyces (Aotani, Nagata and Yoshida 1997), and chartreusin, that exerts strong chemotherapeutic activity against various tumor cell lines (Xu et al. 2005). We also identified a mangotoxin biosynthesis cluster. Mangotoxin causes apical necrosis of plant tissue, which may aid in food processing and digestion by the host (Arrebola et al. 2003). Biosynthesis of the volatile compound homoserine lactone (hserlactone) may be related to communication between fungi and bacteria (Shiner, Rumbaugh and Williams 2005), while ectoine may serve as osmolyte conferring resistance to salt, dessication and temperature stress (Mosier et al. 2013).

**Antibiotic resistance**

The distribution of antibiotic resistance genes (ARGs) in microbiomes sampled across environments and organisms is still not well understood. A large-scale metagenomics study conducted by Nesme et al. (2014) indicated that soils harbor most classes of ARGs. In the gut microbiome of F. candida, we identified over 200 unique terms associated with antibiotic resistance distributed over more than 800 genes, more than twice the amount detected in human microbiomes and almost eight times the amount detected in the giant African snail Achatina (Fitzpatrick and Walsh 2016). This might be explained by the intimate association between the springtail and the soil ecosystem.

The presence of ARGs does not indicate whether Folsomia’s gut is a source or a sink of antibiotic resistance. Instead, it is more informative to investigate ecological connectivity with antibiotic biosynthesizing bacteria. Such connectivity may be very important in shaping a balanced gut microbiome community structure supporting host health. In this context, it is noteworthy that we identified a substantial amount of β-lactamases, probably resulting from the selective pressure caused by β-lactam production by the host itself (Suring et al. 2017). For example, Bacillus toyonensis, a member of F. candida’s microbiota, is highly resistant to β-lactams (Janssens et al. 2017). Furthermore, interactions between bacterial communities with antibiotic biosynthesis capacity and communities showing resistance to such antibiotics can also be expected. Observations from this and other studies indicate a potential for Pseudomonas, Streptomyces and Gordonia strains isolated from F. candida to
The metagenome of *Folsomia candida*

synthesize antibiotics (see section above, Supplementary File 3 and Agamennone et al. 2018), while *Streptomyces, Enterococcus* and *Staphylococcus* are abundant among ARG-containing bacterial strains in *Folsomia’s* gut (Supplementary File 4). This supports the notion that antibiotics regulate the homeostasis of microbial communities, and may even be beneficial for commensal bacteria in environments such as the animal gut (Linares et al. 2006). Finally, Engel & Moran (2013) suggested that this balance may be important in facilitating colonization resistance against parasites and bacteria pathogenic to the host. The data provided in this study will be highly relevant in formulating concrete hypotheses to investigate the ecological connectivity of antibiotic-biosynthetic and ARG-containing bacteria in gut microbiomes.

**Host-microbiome interaction and horizontal gene transfer**

A previous study had identified 809 foreign genes in *F. candida’s* genome (Faddeeva-Vakhrusheva et al. 2017). We found 113 best reciprocal blast hits to these genes within the gut microbiome of the springtail, possibly indicating HGT from the gut microbiome to the host genome. The foreign genes without a best reciprocal blast hit within the gut microbiome may have been transferred from other microbial sources, for example the over 30% of foreign genes that had a top blast hit with fungal donors (Faddeeva-Vakhrusheva et al. 2017). Alternatively, other genes may have been transferred to the host genome early in the evolution of *F candida*. In that case, the accumulation of mutations over time would lead to low similarity with members of the microbiome, preventing the identification of the possible source of these genes. A number of foreign genes with best reciprocal blast hit with genes in the microbiome were CAZymes, involved in the degradation of polymers such as cell wall components. Gene transfer of carbohydrate-active enzymes may increase the capacity of *F. candida* and other soil invertebrates to extract nutrients from their diet (Faddeeva-Vakhrusheva et al. 2016), thereby contributing to their adaptation to life in the soil.

Horizontal gene transfer from prokaryotes to eukaryotic host genomes has become a highly controversial topic. There are claims that gene transfer only occurs between hosts and mitochondria, plastids and endosymbionts, and that other HGT cases are the result of differential loss of ancestral genes, that originated prior to the last eukaryotic common ancestor (Martin 2017). However, this hypothesis overestimates gene contents of ancestral genomes, and is therefore unlikely (Leger et al. 2018). The foreign genes in *Folsomia’s* genome are most likely acquired via horizontal gene transfer (Faddeeva-Vakhrusheva et al. 2017). Here, we suggest that part of these HGT events could have taken place by interaction with the gut microbiota. In the gut environment host and microorganisms maintain an intimate physical association with many opportunities for interaction, thus increasing chances for gene transfer to occur (Huang 2013). Two recent studies provide evidence for bacterial DNA transfer
into somatic human cells (Schroder et al. 2011; Riley et al. 2013) through bacterial type IV secretion system (T4SS). This system is known to mediate interbacterial conjugative DNA transfer and transkingdom protein transfer into eukaryotic host cells during bacterial pathogenesis. Schroder et al. showed that T4SS-dependent DNA transfer into host cells may occur naturally during human infection with Bartonella (Schroder et al. 2011). Furthermore, Ridley et al. identified a Pseudomonas strain as a donor of foreign DNA detected in human stomach carcinomas (Riley et al. 2013). It is still unclear why functions that can be provided by the microbiome would be incorporated and maintained in F. candida’s genome. In the case of foreign genes involved in lignocellulose breakdown, we speculate that such functions, when controlled by the host, could provide fitness advantage in terms of energy balance and nutrient acquisition. Similarly, transferred genes involved in detoxification may protect the host for natural toxins that are quite common in the soil. These and other hypotheses should be tested by conducting gene knockdown and other experiments.

We have provided an insight in the functional metagenome of a Collembolan species, F. candida. More than 99% of the biodiversity is attributed to four phyla, that are also representative for soil microbial ecosystems. This confirms the interaction of F. candida with its surrounding soil ecosystem. A broad spectrum of gene functions was identified through several annotation approaches, most notably related to carbohydrate metabolism, antibiotic resistance and secondary metabolite production. These functions were presented and discussed in the context of their ecological relevance and in the light of potential biotechnological applications. Finally, we presented data suggesting that the gut microbiome may have been a source of genes acquired by the host through HGT. These genes may have conferred a fitness advantage to the springtail, probably during adaptive evolution in the soil ecosystem.
The metagenome of *Folsomia candida*

**Supplementary Information**

**Supplementary Figure 1**: Contig length distribution

**Supplementary Figure 2**: Identified taxonomic groups

**Supplementary File 1**: List of all the predicted genes, including taxonomies and functional annotations*

**Supplementary File 2**: Complete list of the genes involved in carbohydrate metabolism*

**Supplementary File 3**: Complete list of antiSMASH results*

**Supplementary File 4**: Complete list of the predicted antibiotic resistance genes*

**Supplementary File 5**: Complete list of best reciprocal hits between *Folsomia candida*’s genome and metagenome that are also HGT*

* available online at the VU University Library: www.ub.vu.nl
Supplementary Figure 1. Contig length distribution.
Supplementary Figure 2. Identified taxonomic groups at the phylum, class and genus level.
Chapter 4 – Antimicrobial activity in culturable gut microbial communities of springtails

Agamennone V, Roelofs D, van Straalen NM, Janssens TKS. *Journal of Applied Microbiology* 2018; doi: 10.1111/jam.13899
4.1. Abstract

Aims
The rise of antibiotic resistance pushes the pharmaceutical industry to continually search for substances with new structures and novel mechanisms of action. Many environmental niches are still to be explored as sources of antimicrobials. In this paper we assess the antimicrobial potential of gut microbes of springtails, soil invertebrates which live in a microbe-dominated environment and are known to be tolerant to entomopathogenic microorganisms.

Methods and Results
Bacteria isolated from the guts of five springtail species were tested for inhibitory activity against different microbial pathogens. We identified 46 unique isolates belonging to 17 genera and 15 families. Thirty-five of these isolates (76%) showed inhibitory activity, and 18 inhibited both bacterial and fungal pathogens. One isolate was active against all the pathogens tested.

Conclusions
We demonstrated a range of antimicrobial activities in bacteria isolated from the guts of springtails, indicative of complex interactions within the gut community, possibly relating to nutrition or defense against pathogens.

Significance and impact of the study
Our results suggest that a large proportion of cultivatable microbes associated with Collembola have a potential for antimicrobial production. We propose that soil invertebrates and their associated microbes are interesting targets for drug discovery.

4.2. Introduction

Various biologically active compounds obtained from plants, animals and microbes have been used as therapeutic agents in traditional and modern medicine (Zhang, Sun and Wang 2013; Kong and Tan 2015). Despite an increased focus on synthetic and semi-synthetic compounds in the past decades (Patridge et al. 2016), natural products and their derivatives are still important leads in drug development (Li and Vederas 2009), and several sources remain to be explored (Taylor 2013; Pidot et al. 2014). Most natural products are of microbial origin (Peláez 2006) and many are derived from actinomycetes (Actinobacteria, Actinomycetales), a group of prolific antibiotic producers (Genilloud et al. 2011). These bacteria can be isolated from a variety of terrestrial and marine environments (Gontang et al. 2010; Guo et al. 2015) and are also found in association with animal and plant hosts (Qin et al. 2011).

Invertebrates’ guts constitute a habitat for distinctive communities of microorganisms (König 2006) and offer opportunities for the evolution of specific
Antimicrobial activity in gut bacteria of springtails

and complex symbioses (Noda et al. 2009). Gut microbes can contribute to the health and fitness of their host by participating in important functions such as nutrition, regulation of metabolism, development, immunity and defense against pathogens (Engel and Moran 2013). Importantly, resident microbial communities may contribute to defense against pathogens by preventing foreign microbes from establishing in the host, a phenomenon called colonization resistance (He et al. 2014). Different mechanisms can drive colonization resistance: symbiotic bacteria can contribute indirectly to pathogen defense by modulating host immunity mechanisms (Dong, Manfredini and Dimopoulos 2009). Alternatively they can interfere directly with colonization by external microorganisms through resource competition (Maltby et al. 2013) or through interference competition, by producing antimicrobials targeting the invading microorganisms (Rea et al. 2010).

Host-associated microbes, whether symbiotic or entomopathogenic, have significant potential as sources of secondary metabolites of biotechnological interest (Bode 2009; Piel 2009). For example, a Pseudomonas symbiont of the beetle Paederus fuscipes produces pederin-type polyketides, which are used as chemical weapons by the host. These compounds were also shown to contain antitumor properties (Piel et al. 2005). In this context, microbial communities associated with Collembola (Hexapoda) deserve special attention.

Collembolans, commonly known as springtails, are microarthropods that are found in soil ecosystems throughout the world. Recently, it was discovered that beta-lactam biosynthesis genes are widespread among collembolan families, while they have not been found in other animals (Roelofs et al. 2013; Suring et al. 2017). These genes probably originated by horizontal gene transfer from microorganisms living in close proximity with the springtails (Faddeeva-Vakhrusheva et al. 2016). While the function of these genes is still to be elucidated, it is suggested that they are involved in mechanisms regulating the composition of the gut microbial communities (Thimm et al. 1998) and possibly protecting the springtails from invasion by pathogenic microorganisms. Previous studies have shown that Beauveria bassiana and other entomopathogenic fungi are generally of low virulence to springtails (Broza, Pereira and Stimac 2001; Dromph and Vestergaard 2002), but the mechanisms explaining this resistance have not yet been described. Symbiotic bacteria may contribute to host defense by producing antimicrobial compounds that inhibit pathogen growth. Therefore, evaluating the inhibitory activity of bacteria isolated from collembolans, resistant to entomopathogenic microbes, has ecological relevance, because it could suggest a possible mechanism to explain the non-susceptibility of springtails to these pathogens.

Here, we screened microbial communities associated with soil invertebrate guts for inhibitory activity against pathogens. First, we isolated cultivable bacteria from the guts of five springtail species belonging to three different families within the order
Entomobryomorpha: *Folsomia candida* and *F. fimetaria* (Isotomidae), *Orchesella cincta* and *Sinella curviseta* (Entomobryidae), and *Tomocerus minor* (Tomoceridae). To enrich the bacterial pool for potential antibiotic-producers, we used culture media targeting actinomycetes. The colonies were then genotyped and dereplicated using repetitive element sequence-based PCR (rep-PCR) and random amplified polymorphic DNA (RAPD) PCR and identified with 16S rRNA gene sequencing. Finally, biological activity of the isolates was assessed by testing their ability to inhibit bacterial, fungal and oomycete pathogens.

### 4.3. Materials and methods

*Test organisms*

The animals were kept in plastic boxes with a plaster of Paris bottom and reared under constant conditions in climate rooms with 20°C air temperature, 75% humidity and a 12:12 light dark regime. *F. candida, F. fimetaria* and *S. curviseta* were fed dried baker’s yeast (Dr. Oetker, Bielefeld, Germany); *T. minor* and *O. cincta* were fed algae growing on twigs of pine trees. *F. candida* (“Berlin strain”, Vrije Universiteit Amsterdam) originated from a population that had been cultured in our laboratory for more than 10 years. The other springtail strains were sampled from different locations in the Netherlands and reared in the laboratory for several generations before conducting the experiments.

*Isolation and growth of bacteria from springtails*

Guts were dissected from adult springtails (ten individuals for *F. candida, F. fimetaria* and *S. curviseta*, five for *T. minor* and *O. cincta*) using sterile forceps and tweezers. Dry ice was used as a source of carbon dioxide to anesthetize the animals. Dissected guts were placed in sterile Phosphate Buffer Saline (PBS) and crushed using a sterile plastic pestle. Ten-fold dilutions of the original extract were prepared, until dilution factor $10^{-5}$. 100 µl of the undiluted extract and of the four dilutions was spread on Actinomycete Isolation Agar (M490) medium (per liter: sodium propionate 4 g; sodium caseinate 2 g; dipotassium phosphate 0.5 g; L-asparagine 0.1 g; magnesium sulphate 0.1 g; ferrous sulphate 0.001 g). The procedure was repeated twice, obtaining ten plates per species. 100 µl of sterile PBS was used as a negative control. Plates were incubated at 30°C and observed daily for bacterial growth. Colonies of different morphologies were selected, transferred individually to Tryptic Soy Agar (TSA) plates (Tryptic Soy Broth, agar 1.5%; Sigma-Aldrich, St. Louis, MO, USA) to obtain pure cultures and kept at 4°C until further characterization and screening steps.
Genotyping of isolates

Rep-PCR, RAPD-PCR and 16S rRNA gene sequencing were used to characterize and dereplicate the pure cultures obtained from the animals’ guts. Genomic DNA was isolated from bacterial colonies using the PowerSoil DNA Isolation Kit (MOBIO Laboratories Inc., Carlsbad, CA, USA). First, rep-PCR was used to identify unique strains, to avoid redundant screening. The strains with unique rep-PCR profiles and the ones that did not produce bands were identified using 16S rRNA gene sequencing. We then used RAPD-PCR to 1) characterize the isolates that were not amplified by the rep primers, 2) obtain clear banding patterns for the isolates that did not have very clear rep-PCR profiles, and 3) try to differentiate between isolates identified as the same species based on 16S rRNA gene sequencing, but obtained from different springtail species.

Repetitive element sequence-based PCR (rep-PCR) and random amplified polymorphic DNA (RAPD) PCR

The first screening, to identify unique strains, was performed using primers REP1R-I (5’ - III ICG ICG ICA TCI GGC - 3’) and REP2-I (5’ - ICG ICT TAT CIG GCC TAC - 3’) (Versalovic, Koeth and Lupski 1991). The second screening, to differentiate isolates that were identified as the same genus by 16S rRNA gene sequencing and to characterize the isolates that did not have a rep profile or had unclear banding pattern, was performed using the RAPD primer (GTG)₅ (5’- GTG GTG GTG GTG GTG GTG - 3’) (Wiid et al. 1994). PCR reactions were set up as follows: 1 µl genomic DNA (approximately 10 ng), 5 µl of each primer (5 µmol l⁻¹), 1 µl dNTPs (10 mmol l⁻¹ each), 5 µl DreamTaq Green Buffer (10X) (Thermo Fisher Scientific, Waltham, MA, USA), 27.5 µl high-purity water (Purelab Flex, ELGA LabWater, Veolia Water Technologies), 5 µl DMSO, 0.5 µl DreamTaq DNA Polymerase (5 U µl⁻¹). For primer pair REP1R-I – REP2-I, the PCR program was the following: initial denaturation for 3 min at 95°C, 35 amplification cycles (30 s at 95°C, 30 s at 40°C, 4 min at 72°C), and final extension step of 10 min at 72°C. For RAPD primer (GTG)₅ the PCR program was the following: initial denaturation for 5 min at 94°C, 30 amplification cycles (30 s at 95°C, 1 min at 45°C, 5 min at 65°C), and final extension step of 16 min at 65°C. The PCR products were loaded on a 0.8% (wt vol⁻¹) agarose gel containing 5% ethidium bromide solution and electrophoresed for 90 min at a constant voltage of 120 V in 1X TAE (40 mmol l⁻¹ Tris, 20 mmol l⁻¹ acetic acid, 1 mmol l⁻¹ EDTA, pH 8.3). Rep-PCR profiles were visualized with UV light and images were captured using a Molecular Imager Gel Doc XR System (Bio-Rad, Hercules, CA, USA). The profiles were analysed using the GelCompar II software (Applied Maths, Sint-Martens-Latem, Belgium). Similarities between the banding patterns were calculated using the Pearson correlation coefficient. To visualize the similarities, a dendrogram was built using the unweighted pair group method with arithmetic
means (UPGMA) algorithm. Clustering of RAPD-PCR banding patterns is shown in Figure S1 (Supporting Information).

16S rRNA gene sequencing and phylogenetic analysis

16S rRNA gene fragments were amplified using the primers 27F (5’ - AGA GTT TGA TCM TGG CTC AG - 3’) (Lane 1991) and 1492R (5’ - CGG TTA CCT TGT TAC GAC TT - 3’) (Turner et al. 1999). The PCR reactions were set up as follows: 1 µl genomic DNA (approximately 10 ng), 2.5 µl of each primer (5 µmol l⁻¹), 1 µl dNTPs (10 mmol l⁻¹ each), 10 µl Phusion Buffer (5X) (New England Biolabs, Ipswich, MA, USA), 31 µl ELGA water, 1.5 µl DMSO, 0.5 µl Phusion DNA polymerase (2 U µl⁻¹). The PCR program was the following: initial denaturation for 30 s at 98ºC, 35 amplification cycles (10 s at 98ºC, 10 s at 53ºC, 45 s at 72ºC), and final extension step of 10 min at 72ºC. The PCR products were verified by electrophoresis on 1.5% (wt vol⁻¹) agarose gels and sent to Macrogen (Amsterdam, The Netherlands) for purification and sequencing. Raw sequence reads were trimmed and assembled in CLC Genomics Workbench software, version 5.1 (CLC Bio-Qiagen, Aarhus, Denmark). The BLAST Sequence Analysis Tool (Madden 2002), the Sequence Match Tool of the Ribosomal Database Project (RDP) (Cole et al. 2014) and the EzBioCloud database (Yoon et al. 2017) were used to assign taxonomic identity to the 16S rRNA gene sequences from the isolates. The EzBioCloud database was also used to obtain accurate similarity values between the isolates and type strains. 16S rRNA gene sequences of the unique isolates were submitted to GenBank, where they are available under accession numbers MF801315-MF801360.

Screening of isolates against pathogens

The unique strains were screened for inhibitory activity against nine pathogens: the bacteria Staphylococcus aureus 533R4 (DSMZ 20231) (Firmicutes), Escherichia coli WA321 (DSMZ 4509) (Gammaproteobacteria) and Micrococcus luteus ATCC49732 (Actinobacteria); the fungi Rhizoctonia solani AG2.2 IIIB (Basidiomycota), Candida albicans BSMY 212 (DSMZ 10697), Fusarium oxysporum and Beauveria bassiana (ARSEF 2597) (Ascomycota); and the oomycetes Pythium ultimum P17 and Saprolegnia diclina (1152F4) (Heterokontophyta). The choice of these microorganisms was based on their relevance as pathogens for humans (E. coli, S. aureus, M. luteus and C. albicans), plants (F. oxysporum, R. solani and P. ultimum), fish (S. diclina) or arthropods (B. bassiana).

S. aureus, E. coli, C. albicans, and M. luteus were grown on LB agar (Lennox LB Broth Base, agar 1.5%; Thermo Fisher Scientific, Waltham, MA, USA); S. diclina and R. solani were grown on Potato Dextrose Agar (PDA) (Potato Dextrose Broth, agar 1.5%; Sigma-Aldrich, St. Louis, MO, USA); P. ultimum and F. oxysporum were grown on 1/5 PDA; and B. bassiana was grown on Sabouraud agar (SDA) (Jaronski
Antimicrobial activity in gut bacteria of springtails

and Jackson 2012) with 1% yeast extract as a supplement. *S. aureus*, *E. coli* and *C. albicans* were grown at 37°C, the remaining organisms were grown at 28°C. To test the isolates against the bacteria and the fungus *C. albicans*, we used the agar overlay method. The agar plug method was used to test antimicrobial activity against the remaining fungi and against the oomycetes. For the agar overlay method, the isolates were grown overnight in Tryptic Soy Broth (TSB; Sigma-Aldrich, St. Louis, MO, USA), at 30°C with shaking at 400 rpm. Subsequently, 5 µl of liquid culture of each isolate was transferred to solid medium. Square petri dishes (120x120 mm), were used for the screening and 16 isolates were tested on each plate. *Pseudomonas protegens* Pf5, a broad range biocontrol strain, was used as a positive control on each plate (Ramette et al. 2011). The plates were incubated overnight at 28°C. On the same day, liquid cultures of the bacterial pathogens and of *C. albicans* were prepared by inoculating colonies in 4 ml of LB broth base (Lennox LB Broth Base, Thermo Fisher Scientific, Waltham, MA, USA) and growing them overnight at 30°C with shaking at 400 rpm. The following day, growth of the strains of interest on the agar plates was observed, and the pathogens grown during the night were prepared for the overlay. 500 µl of liquid culture of the pathogen was inoculated in 10 ml top agar (LB Broth Base, 0.7% agar), mixed well by vortexing and poured over the plate to completely cover the colonies of the isolates. The plates were then incubated overnight at the appropriate temperature for the pathogens and the next day the presence and diameter of inhibition zones were recorded. For the agar plug method, agar plugs covered in mycelium of the pathogens were placed on the plates between the isolates’ colonies, after the isolates of interest had grown on the plates. The plates were sealed with parafilm and incubated at the appropriate temperature for the pathogens until the growth allowed the observation of clear inhibition zones (between three and 19 days). For both the agar overlay and the agar plug method, the screening was conducted on three different media: actinomycete isolation agar (M490), Potato Dextrose Agar (PDA) and 1/10 Tryptic Soy Agar (1/10TSA) (Tryptic Soy Broth, NaCl 0.5%, KH₂PO₄ 0.1%, agar 1.5%; Sigma-Aldrich, St. Louis, MO, USA).

4.4. Results

Isolation and growth of bacteria

After dissecting the guts from the springtails *F. candida*, *F. fimetaria*, *T. minor*, *S. curviseta* and *O. cincta*, and plating them on M490 medium, colony growth was observed between day 3 and day 7. No growth was observed on negative control plates. Based on morphological differences, a total of 78 colonies were picked and transferred to TSA plates to obtain pure cultures.
Rep-PCR, RAPD-PCR and 16S rRNA gene sequencing

Genomic DNA was isolated from the pure cultures and subjected to rep-PCR to prevent further analysis of identical strains (dereplication). Isolates that produced unique molecular profiles or that did not amplify during the rep-PCR screening step, were then identified using 16S rRNA gene sequencing. Finally, RAPD-PCR was used to obtain profiles for the isolates that did not have a rep-PCR profile and possibly to differentiate between isolates with the same 16S-based identity, but obtained from different animal species. Following these molecular identification and dereplication procedures, all unique isolates were further investigated to assess their inhibitory activity against pathogens.

In total, 46 unique isolates were selected. The number of unique isolates that were obtained from each collembolan species are summarized in Table 1. Most isolates were obtained from S. curviseta (18) and F. candida (11), while only four were isolated from O. cincta guts. Morphological characteristics and rep- and RAPD-PCR profiles of all unique isolates are available in Table S1 (Supporting Information). Rep-PCR profiles were not obtained for 17 of the 46 unique isolates (37%), corresponding to the Firmicutes (Staphyloccoccus Fc5, Fc17c and Ff5 and Bacillus Sc23) some Actinobacteria (all Streptomyces: Fc1, Fc7, Fc12, Ff1, Ff4b Tm3, Tm6b, Sc7a and Sc8, Glutamicibacter Sc3 and Cellulosimicrobium Tm1) and some Proteobacteria (Halomonas Oc4 and Sc22) isolates (Table S1).

Based on their 16S rRNA gene sequence, the isolates were assigned a taxonomic identity. The top BLAST and EzBioCloud hits of the 16S rRNA gene sequences obtained from each unique isolate are represented in Table 1. Based on the comparison with type strains in the EzBioCloud database, 17 different genera were identified across the five springtail species: the Proteobacteria Ochrobactrum (Alphaproteobacteria, Rhizobiales), Delftia (Betaproteobacteria, Burkholderiales), Stenotrophomonas, Pseudoxanthomonas (Gammaproteobacteria, Xanthomonadales), Pseudomonas, Acinetobacter (Gammaproteobacteria, Pseudomonadales), Serratia, Pantoea (Gammaproteobacteria, Enterobacteriales) and Halomonas (Gammaproteobacteria, Oceanospirillales); the Firmicutes Staphylococcus and Bacillus (Bacilli, Bacillales); the Actinobacteria Cellulosimicrobium (Micrococcales), Streptomyces, Glutamicibacter, Gordonia, Nocardioides and Microbacterium (Actinomycetales). No significant differences in the occurrence of the different genera was observed between the five animal species (Chi-squared test, data not shown).

Fifteen isolates had 100% identity with their top EzBioCloud hit, 25 isolates had higher than 99% identity, and six isolates (Streptomyces Fc1, Microbacterium Fc2, Halomonas Oc4 and Sc22, and Pantoea Oc5 and Tm9b) had less than 98.65% similarity with reference sequences in EzBioCloud.
To discriminate between isolates with the same top BLAST hit, a RAPD-PCR was performed using (GTG)$_5$ primer. A profile was obtained for most of the isolates, except for Fc1 and Ff1. Isolates with similar morphologies and similar rep-PCR profile from the first screening step also showed similar RAPD-PCR profiles (see Table S1 – Supporting Information: Pantoea Oc5 and Tm9b; Stenotrophomonas Ff7b and Sc15; Ochrobactrum Fc6 and Ff3). In some cases, the first rep-PCR screening step was sufficient to differentiate strains (for example, compare profiles of Microbacterium Fc2, Fc13 and Fc16a). However, when a REP profile was not available, RAPD-PCR allowed us to differentiate between strains with similar morphology but different genotype (compare Streptomyces Sc7a and Sc8) or to identify additional replication among strains, which was also confirmed by the results of 16S rRNA gene sequencing. Isolates with the same EzBioCloud top hit, also had similar RAPD-PCR profiles (compare Staphylococcus Fc5 and Ff5); and isolates assigned by EzBioCloud to the same genus but to different species, had different RAPD-PCR profiles (compare Microbacterium Fc2, Fc13 and Fc16a, or Staphylococcus Fc5 and Fc17c). Overall, the greatest redundancy of isolates was found among the F. candida and S. curviseta isolates.

**Antimicrobial screening**

The antimicrobial potential of the bacterial isolates was assessed by testing their ability to inhibit growth of a variety of pathogenic microorganisms. Figure 1 shows an example of inhibition of pathogens’ growth by strain Streptomyces Fc1 isolated from F. candida. The results of the inhibition assays are summarized in Table 1 and presented in more detail in Table S1 (Supporting Information). Forty-six springtail isolates were tested against nine pathogens, on three culture media, resulting in 1242 specific inhibition tests. In 222 tests (18%) inhibition was observed. As expected, the positive control Pseudomonas protegens Pf5 inhibited growth of all pathogens on all media. A surprisingly high number of 35 out of 46 (76%) springtail isolates inhibited at least one pathogen; 18 of these (50%) were active against both fungal and bacterial pathogens (Streptomyces Fc1, Fc7, Fc12, Ff1, Ff4b, Tm3, Tm6b, Sc7a and Sc8, Stenotrophomonas Fc3a and Ff7a, Pseudomonas Fc8, Oc2, Tm10a and Tm10b, Acinetobacter Sc11, Serratia Sc14, Nocardiooides Sc24); 11 strains were active only against bacteria (Staphylococcus Fc5 and Ff5, Ochrobactrum Fc6 and Ff3, Microbacterium Fc16a, Pantoea Oc5 and Tm9b, Cellulosimicrobium Tm1, Glutamicibacter Sc3, Acinetobacter Sc17, Pseudomonas Sc19); while six were active only against fungi and/or oomycetes (Staphylococcus Fc17c, Stenotrophomonas Ff7b and Oc1, Halomonas Oc4, Ochrobactrum Tm9a and Sc9). One isolate, Streptomyces Sc8, inhibited the growth of all pathogens tested, like the positive control.
Figure 1. Pictures showing the strain *Streptomyces* sp. Fc1 (isolated from *Folsomia candida*), growing in the left upper corner, inhibiting the growth of different pathogens on a variety of media. a: *Fusarium oxysporum* on 1/10 TSA; b: *Beauveria bassiana* on M490; c: *Micrococcus luteus* on PDA; d: *Saprolegnia diclina* on 1/10 TSA.

All isolates from *F. fimetaria*, *O. cincta* and *T. minor* showed some kind of inhibitory activity (Table 1), while only 50% (nine out of 18) of *S. curviseta* isolates and 82% (nine out of 11) of *F. candida* isolates showed activity. Moreover, all of the *Streptomyces*, *Pseudomonas*, *Acinetobacter*, *Staphylococcus* and *Pantoea* isolates were active against some of the pathogens, regardless of the animal species from which they were derived. Finally, when considering all isolates from a springtail species, they were collectively active against all of the pathogens tested, with the exception of isolates from *O. cincta* which could not inhibit *E.coli* growth.

We observed differences in the inhibitory phenotypes between replicates of the same strains (Table S1). At the same time, patterns of inhibition were very similar for some isolates from different springtails suspected (based on 16S rRNA gene sequencing, rep- and RAPD-PCR profiles and morphologies) to belong to the same species, for example *Streptomyces* Fc7 and Ff4b, and *Ochrobactrum* Tm9a and Sc9. No significant differences in the average number of inhibitions per isolate could be observed between the five different hosts (Kruskal-Wallis test, Dunn’s multiple pairwise comparison, data not shown).
Table 1. Total number of cultured isolates, top BLAST hits of the unique isolates and number of pathogens inhibited by each unique isolate for each springtail species.

<table>
<thead>
<tr>
<th>Springtail</th>
<th>Total number of isolates growing on M490 medium</th>
<th>Unique isolates based on rep-PCR profiles, RAPD-PCR profiles and 16S rRNA gene sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Isolate</td>
<td>Top BLAST Hit</td>
</tr>
<tr>
<td>Folsomia candida</td>
<td>FC1</td>
<td>Streptomyces sp. TTHQ-DM-1</td>
</tr>
<tr>
<td></td>
<td>FC2</td>
<td>Microbacterium sp. SIB_Cu_R3</td>
</tr>
<tr>
<td></td>
<td>FC3a</td>
<td>Stenotrophomonas sp. MACLI3A</td>
</tr>
<tr>
<td></td>
<td>FC5</td>
<td>Staphylococcus sp. S07 ①</td>
</tr>
<tr>
<td></td>
<td>FC6</td>
<td>Ochrobactrum sp. MYb103 ②</td>
</tr>
<tr>
<td></td>
<td>FC7</td>
<td>Staphylococcus sp. strain BV316 ③</td>
</tr>
<tr>
<td></td>
<td>FC8</td>
<td>Pseudomonas sp. strain B16</td>
</tr>
<tr>
<td></td>
<td>FC12</td>
<td>Streptomyces sp. strain KIB-H0945</td>
</tr>
<tr>
<td></td>
<td>FC13</td>
<td>Microbacterium sp. T6220-7-3b strain: K13218-3-2b</td>
</tr>
<tr>
<td></td>
<td>FC16a</td>
<td>Microbacterium natoriense strain NW41</td>
</tr>
<tr>
<td></td>
<td>FC17c</td>
<td>Staphylococcus sciuri strain F9A92685</td>
</tr>
<tr>
<td>Folsomia fimetaria</td>
<td>FF1</td>
<td>Streptomyces sp. strain K1</td>
</tr>
<tr>
<td></td>
<td>FF3</td>
<td>Ochrobactrum sp. JCM 28827 strain: T7426-9-3b</td>
</tr>
<tr>
<td></td>
<td>FF4b</td>
<td>Streptomyces sp. strain BV316 ③</td>
</tr>
<tr>
<td></td>
<td>FF5</td>
<td>Staphylococcus sp. S07 ①</td>
</tr>
<tr>
<td></td>
<td>FF7a</td>
<td>Uncultured Stenotrophomonas sp. clone OTUM16</td>
</tr>
<tr>
<td></td>
<td>FF7b</td>
<td>Stenotrophomonas terrae strain R-32768④</td>
</tr>
<tr>
<td>Orchesella cincta</td>
<td>Oc1</td>
<td>Stenotrophomonas sp. LY-2</td>
</tr>
<tr>
<td></td>
<td>Oc2</td>
<td>Pseudomonas putida strain B-18</td>
</tr>
<tr>
<td></td>
<td>Oc4</td>
<td>Halomonas qiaohouensis strain YIM QH88</td>
</tr>
<tr>
<td></td>
<td>Oc5</td>
<td>Uncultured gamma proteobacterium clone P9 ⑤</td>
</tr>
</tbody>
</table>
Table 1 (continued)

| Tomocerus minor | 14 | Sc1  | Pseudoxanthomonas sp. strain Oil-1 | 100% | MF405126 | Pseudoxanthomonas indica | 99.64% | N.I. |
|                |    | Sc3  | Arthrobacter sp. RCB237            | 100% | KT260449 | Glutamicibacter soli     | 99.56% | 2   |
|                |    | Sc7a | Streptomyces sp. strain BV316      | 100% | MF511802 | Streptomyces pratensis   | 100%   | 7   |
|                |    | Sc7b | Gordonia sp. COL11                | 100% | JN695023 | Gordonia malaquae        | 99.93% | N.I.|
|                |    | Sc8  | Streptomyces sp. H16 LR-2016      | 100% | LC230101 | Streptomyces luridiscabiei | 100% | 9   |
|                |    | Sc9  | Ochroactrum sp. strain Ktm-7      | 100% | MF405118 | Ochroactrum anthropi     | 100%   | 2   |
|                |    | Sc10 | Microbacterium shaanxiense        | 100% | KJ735510 | Microbacterium shaanxiense | 100% | N.I.|
|                |    | Sc11 | Acinetobacter sp. strain LSN1-10  | 99%  | KY054572 | Acinetobacter cauvalinii | 99.21% | 4   |
|                |    | Sc14 | Serratia marcescens strain FF6    | 100% | KR778806 | Serratia marcescens      | 99.93% | 7   |
|                |    | Sc15 | Stenotrophomonas terrae strain R-32768 | 99% | NR_042569 | Stenotrophomonas terrae | 99.49% | N.I.|
|                |    | Sc16 | Delfia sp. strain As-37            | 100% | MF353935 | Delfia lacustris         | 100%   | N.I.|
|                |    | Sc17 | Acinetobacter sp. RS206            | 100% | EU912468 | Acinetobacter vivianii   | 99.35% | 1   |
|                |    | Sc19 | Pseudomonas orovenensis strain E8  | 100% | KY938127 | Pseudomonas donghuensis  | 99.85% | 3   |
|                |    | Sc20b| Glutamicibacter arilaitensis strain MLS-4-4 | 100% | KT997452 | Glutamicibacter arilaitensis | 99.18% | N.I.|
|                |    | Sc21b| Ochroactrum sp. MYb103            | 100% | KX079833 | Ochroactrum pituitusom    | 99.85% | N.I.|
|                |    | Sc22 | Halomonas sp. ADMK29              | 98%  | KU850985 | Halomonas stenophila     | 97.93% | N.I.|
|                |    | Sc23 | Bacillus sp. Ant-1b              | 100% | HF678912 | Bacillus weihenstephanensis | 100% | N.I.|
|                |    | Sc24 | Nocardoides sp. strain Bt-B       | 100% | MF405108 | Nocardoides daejeonensis | 100%   | 3   |

The numbers in circles in the “top BLAST hit” and in the “Top EzBioCloud hit” columns indicate isolates with the same top BLAST hit. E-values for the BLAST hits in the table were all lower than 1E-70. N.I. = no inhibition.
Antimicrobial activity in gut bacteria of springtails

Discussion

In this study, we show that bacteria isolated as cultivatable isolates from the guts of springtails are active against a broad range of microbial pathogens that are relevant for human health, agriculture and fish culturing. We identified 46 unique isolates, 35 of which (76%) showed inhibition of at least one of the pathogens tested. In principle, these isolates could serve as biocontrol agents or be further characterized for applications in the pharmaceutical sector.

Isolation of potential antimicrobial producers

With the exception of Staphylococcus and Bacillus isolates (Firmicutes), all bacteria with antimicrobial activity belong to the phyla Actinobacteria and Proteobacteria. 18 of the 46 isolates (40%) are Actinobacteria. This suggests that M490 medium can be effectively used to target this microbial group, while still allowing growth of other bacteria. Previous microbial community studies, using both culture-based and cultivation-independent approaches on eight springtail species, did not identify Streptomyces strains in collembolan gut (Thimm et al. 1998; Czarnetzki and Tebbe 2004a; Agamennnone et al. 2015). In culture-based studies, this may have been due to specific culturing conditions, for example the use of media selecting against Streptomyces, or to the slow growth of Actinobacteria compared to other bacteria. In culture independent analyses, such as electron microscopy or in situ hybridization, underrepresentation of Streptomyces may be due to less optimal binding of primers, or to very low abundances of these bacteria. In contrast to Streptomyces, in this study we found common genera like Stenotrophomonas, Ochrobactrum and Pseudomonas that had been previously identified in collembolan gut (Thimm et al. 1998; Czarnetzki and Tebbe 2004a).

Many of the isolates did not amplify with the rep-PCR primers, which may have been due to lack of annealing of the primers with the template or to suboptimal amplification conditions. However, RAPD-PCR profiles were obtained for most isolates and were useful in discriminating between different bacterial species with the same top BLAST hit.

Activity of isolates against pathogenic fungi and bacteria

Of the 46 bacterial isolates tested, 76% showed inhibitory activity. However, pathogen inhibition was observed in only 18% of the inhibition tests performed with different combinations of isolate, pathogen and growth medium. This suggests that interactions between microorganisms are specific, and that inhibition and competition may be modulated by environmental factors (Hoek et al. 2016).

In this study, we confirm the broad inhibitory activity of Streptomyces, a genus of Gram-positive bacteria well known for the production of secondary metabolites including antibiotics (de Lima Procópio et al. 2012). Three Pseudomonas isolates
also showed broad inhibitory activity, inhibiting almost all pathogens. Furthermore, all *Acinetobacter*, *Staphylococcus* and *Pantoea* isolates showed inhibitory activity. These bacteria have previously been shown to inhibit pathogens. *Pseudomonas* species are known for their antagonistic activity against plant and fish pathogens (Berg et al. 2005; Liu et al. 2015b). *Acinetobacter* strains were previously shown to inhibit *Ralstonia solanacearum*, a plant pathogen causing bacterial wilt in tomato (Xue et al. 2009). The growth-inhibiting activities exerted by *Serratia marcescens* have also been observed before, and this species is known to synthesize the red antimicrobial compound prodigiosin (Suryawanshi et al. 2016).

Most likely, the observed inhibitions by the gut isolates are due to the production of antimicrobials (Makras and De Vuyst 2006). Microorganisms can use such compounds to compete with other species or with individuals of the same species (Hibbing et al. 2010). Competition can be particularly intense in environments such as the rhizosphere and the soil ecosystem in general. Here, complex interactions between plant roots, pathogenic and beneficial microorganisms take place, driving a complex warfare based on the production of chemical toxins and on the evolution of strategies to resist them (Raaijmakers et al. 2009).

Springtails can be particularly active in the rhizosphere of plants, where they are attracted by high microbial activity and biomass and they may establish symbiosis with microorganisms (Endlweber, Ruess and Scheu 2009). In the rhizosphere, collembolans feed on a diversity of fungi, including arbuscular mycorrhizal (AM), saprotrophic and pathogenic ones (Broza, Pereira and Stimac 2001; Jonas et al. 2007). For example, the root pathogens *F. culmorum* and *R. solani* constitute high quality food sources for *F. candida* and *F. fimetaria* (Larsen et al. 2008). *F. fimetaria* not only feeds on *R. solani*, but is also effective in reducing infection by this fungus in soil under experimental conditions (Lootsma and Scholte 1997). Moreover, *S. curviseta* was observed to graze on *F. oxysporum* (Nakamura, Matsuzaki and Itakura 1992), while *F. candida* is known to feed on *F. solani* and is even reproductively active on it (Bastian et al. 2010). These observations suggest that springtails may be beneficial for plant health, by using plant pathogens as food source. At the same time, microorganisms associated with springtails may respond with antimicrobial production to the presence of pathogens in their environment. For example, we observed that *Streptomyces* and *Pseudomonas* isolated from springtails were active against *R. solani*, *P. ultimum* and *F. oxysporum*, plant pathogens that mainly infect plant roots.

Antimicrobial compounds produced by springtails’ isolates could be used as biocontrol agents. In the past, biocontrol agents have been obtained from bacteria isolated from the rhizosphere based on their antagonistic activity against pathogens (Trotel-Aziz et al. 2008). Compounds with antibacterial and antifungal activity may also find applications in the pharmaceutical and medical sector. Many of the
antibacterial and antifungal products in use today are natural products. The majority of known antibiotics, such as streptomycin, tetracycline and chloramphenicol, are produced by *Streptomyces*, a genus of soil-living bacteria that has been a source of antimicrobial compounds for decades and continues to be of interest for drug discovery nowadays (Antoraz et al. 2015). The most famous antibiotic and the first to be described, penicillin, is produced by a fungus. And filamentous fungi, especially the genera *Penicillium* and *Aspergillus*, also provide promising candidate for the development of antifungal therapies (Garrigues et al. 2017).

Using methods that aimed to enrich for antimicrobial production, we found higher levels of inhibitory activity compared to those previously found in soil bacteria. For example, one study found that 20 out of 62 soil strains (32%) had moderate to high activity against *B. subtilis*, *S. aureus*, *E. coli* and *C. albicans* (Singh et al. 2009), while another found that only 10 strains out of 160 isolated from soil showed antifungal activity against plant pathogens (Petatán-Sagahón et al. 2011). Other studies have found high levels of antimicrobial activity in bacteria associated with both marine (Zheng et al. 2000; Yung et al. 2011) and terrestrial invertebrates (Fredenhagen et al. 1987; Piel et al. 2005), suggesting that microbial communities in invertebrate guts may constitute important targets for the discovery of new therapeutic agents. Whether host-associated microbial communities are enriched for antimicrobial functions compared to the free-living ones in water and soil environments is an interesting question that could be addressed by future comparative studies.

**Ecological relevance of animal-microbe associations**

One might argue that soil invertebrates, living in a microbe-dominated environment, would benefit from antimicrobial activity in their gut, however, care must be taken in assuming benefits or even coevolution when there is in fact no evidence (Moran and Sloan 2015). Antimicrobial biosynthesis is often triggered by interactions and signaling between microorganisms (Tyc et al. 2014) or by the presence of specific elicitors in the environment (Rigali et al. 2008). In some cases, antimicrobial production may result in indirect benefits for the host by conferring protection against predators and parasites. For example, fungus-growing ants harbour *Streptomyces* bacteria that produce antibiotics specifically targeting the parasitic fungus *Escovopsis*. As a result, the ants and their gardens are protected from fungal infection (Haeder et al. 2009). Another example is provided by pederin, a polyketide produced in the beetle *Paederus fuscipes* by its symbiont *Pseudomonas aeruginosa*. Pederin functions as a chemical defense and protects the larvae of the beetle from predatory spiders (Kellner and Dettner 1996; Piel et al. 2005). Similarly, in larvae of the marine bryozoan *Bugula neritina* bryostatins, polyketide metabolites produced by the symbiont *Endobugula sertula*, confer protection from predators (Lopanik, Lindquist and Targett 2004). Also among vertebrates there are examples of host-
microbe associations where antimicrobials produced by symbiotic bacteria contribute to the host fitness (Soler et al. 2008).

The gut of springtails has been described as a selective habitat for microorganisms (Thimm et al. 1998). It is possible that antimicrobials affect microbial community dynamics in this habitat, by regulating the microbial density or by inhibiting pathogenic microorganisms. It has been shown that entomopathogenic fungi (EPF), used as microbial control agents against soil pests (Erler and Ates 2015), are of low virulence to springtails (Broza, Pereira and Stimac 2001; Dromph and Vestergaard 2002) as compared to target hosts (Zimmermann 2007). For example, collembolans are highly resistant against *B. bassiana*, a fungus parasitic to a wide range of arthropods such as ants and beetles (Zimmermann 2007). *B. bassiana* has been found on the surface and in the guts of collembolans (Greif and Currah 2007; Anslan, Bahram and Tedersoo 2016). In this study, we isolated bacteria inhibiting *B. bassiana* from all springtail species. However, whether antimicrobials are produced in the guts of springtails, and whether they are active against pathogenic microorganisms in that environment, still needs to be investigated. Susceptibility tests with germ-free animals may provide clues about possible roles of microbes and microbial products in pathogen defense.

**Future directions**

This study has identified bacterial isolates that inhabit the gut of springtails and that are able to inhibit pathogenic microorganisms. Six of these isolates share less than 98.65% similarity to type strains in EzBioCloud. Although 16S rRNA gene sequence similarity suggests that these isolates may constitute novel species (Kim et al. 2014), this possibility should be confirmed by more extensive sequencing and phenotypic characterization. Of these six isolates, four were active against pathogens. This suggests that soil invertebrate guts are underexplored environmental niches that have a potential for the discovery of new microbial species and, in turn, new antibiotics (Taylor 2013).

Following these findings, a good strategy to identify molecules with interesting properties would be to extend the functional screens like the one performed in this study. For example, pathogenic strains known to activate specific secondary metabolism pathways could be included in the inhibition assays. Also, focus could be placed on the isolates that showed broad inhibitory activity against pathogens. Promising strains from this study, for example, would be *Streptomyces* Sc8, Fc12, Ff1 and Tm6b, and *Pseudomonas* Fc8 and Tm10a, all showing a broad spectrum of pathogen inhibition. These strains could be further characterized to identify the molecular basis of inhibition. Such studies are underway in our group. For instance, from a previous survey of *F. candida*’s gut, we isolated a *Bacillus toyonensis* (VU-DES13, Janssens et al. 2017) showing a broad spectrum of pathogen inhibition. We subsequently harvested from this isolate an antimicrobial fraction inducing cell wall
stress activity (as revealed by a reporter assay) and inhibiting the growth of the fungus C. albicans (Agamennone et al. unpublished).

In addition to functional studies, metagenomic profiling of complex microbial communities can provide clues for potentially useful catalytic functions. Recently, we sequenced the functional metagenome of F. candida’s gut, and we are currently exploring the annotations which may be associated with antibiotic biosynthesis. Cloning and functional screening of these genes can integrate metagenomic information and potentially lead to antimicrobial compound discovery.

To our knowledge, this is the first study to investigate antimicrobial potential in springtails’ gut microbial communities. We showed that bacteria isolated from the gut of springtails inhibit a variety of pathogens, a possible sign of antimicrobial production. Although the gut microbial communities of the springtails studied here show a clear potential for drug discovery, they also likely constitute the tip of the iceberg in this regard. Antimicrobial compounds have been previously isolated from microbes associated with different invertebrate hosts, suggesting that host-associated microbial communities are rich sources of antimicrobials in general. This is likely the result of natural interactions between resident and pathogenic microorganisms, leading to antimicrobial warfare within the habitat constituted by the host. Further investigations should aim to explore the potential of gut microbiota as a source of novel antimicrobials, identify targets of interest for the pharmaceutical industry and for agricultural applications, and understand the ecological relevance of antimicrobial production for springtails and other soil invertebrates.

Acknowledgements

This work was supported by a grant from the BE-Basic Foundation (Biotechnology-based Ecologically Balanced Sustainable Industrial Consortium, Delft, The Netherlands) (project number F07.003.05). We thank USDA-ARS for providing the P. protegens Pf-5, Claudio A. Valero-Jiménez at Wageningen University for providing strains of B. bassiana, and Victor Carrion and Olaf Tyc at The Netherlands Institute of Ecology for providing the other target strains for the antimicrobial screening. We also thank MicroLife Solutions for providing the equipment and resources to perform the experiments.
Supplementary Information

**Figure S1**: Clustering of RAPD-PCR banding patterns of bacterial isolates obtained from springtails

**Table S1**: Morphological and molecular characterization and pattern of pathogen inhibition of the unique bacterial strains isolated from springtails’ guts (available at the online version of this paper)
Figure S1. Clustering of RAPD-PCR banding patterns of bacterial isolates obtained from springtails. The profiles were obtained through RAPD-PCR with (GTG)$_5$ primer. The dendrogram was obtained through UPGMA-based cluster analysis. Cophenetic correlation coefficients are indicated at the root of the nodes on the dendrogram. The cluster contains the RAPD profiles of 60 isolates that had previously been selected based on their rep-PCR profiles. The names of the isolates in the cluster correspond to the top EzBioCloud hits. The cluster also contains markers (indicated with the letter M), a blank, and the RAPD-PCR patterns of four other isolates used as additional references (a Bacillus subtilis, and three strains that had been previously isolated from F. candida: B. toyonensis VU-DES13, Pseudomonas D1 and Pseudomonas C11).
Chapter 5 – Genome sequence and antimicrobial properties of *Bacillus toyonensis* VU-DES13, isolated from the *Folsomia candida* gut, and its role in the colonization resistance to entomopathogenic fungi

Agamennone V, van Straalen J, de Boer TE, Hensbergen P, Zaagman N, Braster M, Roelofs D, van Straalen NM, Janssens TKS

The genome of *Bacillus toyonensis* VU-DES13 is published in *Genome Announcements* (see annex at the end of this book)
5.1. Abstract

The issue of antibiotic resistance urges the search for new bioactive compounds with novel mechanisms of action. Natural products derived from bacteria and fungi have been widely explored and applied in the fields of medicine for treatment of infectious diseases and there is a need to explore new environments and new sources of antimicrobials. Springtails harbor complex microbial communities in their guts which have been proven to show inhibitory activity against pathogens. The objective of this study was to characterize a dominant bacterial strains isolated from the gut of the springtail *Folsomia candida* that showed a high potential for antimicrobial production.

Using aerobic cultivating conditions we isolated a *Bacillus* strain from the gut of the springtail *F. candida*, named *Bacillus* VU-DES13. We characterized the strain by using the API system and the MIC assay to test for antibiotic resistance. We then used a combination of agar overlay assay and agar disk diffusion assay to test the inhibitory activity of the *Bacillus* strain and of its extract against a variety of pathogens, and we applied reporter assays to investigate the mode of action. We then used HPLC to analyze and to fractionate the extract of bacterial cultures, and we performed additional assays with the obtained fractions. Finally, we sequenced the genome of the strain and we screened it for the presence of antibiotic resistance genes and secondary metabolite gene clusters.

The isolate was identified as *Bacillus toyonensis*, and the results of the API system highlighted some differences in metabolic profile with the type species. The isolate was not susceptible to penicillin, even at the highest concentrations tested. Genome analysis revealed an enrichment of resistance genes for β-lactams antibiotics compared to the type isolate and identified secondary metabolite clusters involved in the production of siderophores, bacteriocins and NRPSs. The isolate and its extract inhibited the growth of a variety of pathogenic microorganisms, and some of the fractions analyzed with HPLC also showed activity against these pathogens.

In conclusions, *Bacillus* VU-DES13 was isolated from the gut of the springtail *F. candida* and evidence for antimicrobial activity was obtained from both the living bacteria and the bacterial extracts. This, coupled with the high resistance to penicillin showed by the MIC assay and substantiated by the presence of resistance genes, points to the potential of *Bacillus* VU-DES13 to provide a new source of antimicrobial compounds.

5.2. Introduction

Symbiotic microorganisms can have different functions in their arthropod host. They can play an important role in nutrition, by providing essential nutrients and helping with digestion (Akman Gunduz and Douglas 2009). They can also contribute to
Bacillus toyonensis VU-DES13

protection from pathogens through a variety of mechanisms, for example by competitive exclusion and the production of defense compounds (Kaltenpoth and Engl 2014). Various antibiotics have been isolated from insect symbionts (Fredenhagen et al. 1987; Flórez et al. 2017). These and other molecules with antimicrobial activity have potential applications in the food and in the pharmaceutical industry.

Recently, we showed that a number of bacteria isolated from the gut of the springtail Folsomia candida show antagonistic potential against pathogenic microorganisms (Agamennone et al. 2018) and we hypothesized that this activity results from the production of antimicrobial compounds. Here, we argue that specific bioassays can be successfully used to guide the discovery of new bioactive compounds, and we apply this approach to a Bacillus species isolated from the gut of F. candida. This animal is an established model in ecotoxicology and ecogenomics, and it is resistant to and able to feed on entomopathogens (Broza, Pereira and Stimac 2001). Its genome contains secondary metabolite biosynthetic gene clusters, such as for β-lactam compounds (Roelofs et al. 2013), which are specifically induced in the gut epithelium upon adverse conditions (Nota et al. 2008). In the past, amplicon-sequencing studies revealed a prominent presence of Bacillus sp. in the microbial community of this springtail (Czarnetzki and Tebbe 2004a; Agamennone et al. 2015).

Bacteria of the Bacillus group and their products have a variety of applications. They are widely used as biopesticides and biofertilizers in agricultural practice (Pérez-García, Romero and de Vicente 2011), as probiotics for both animals and humans (Lodemann et al. 2008; Cutting 2011), and they constitute a source of antibacterial and antifungal compounds (Gebhardt et al. 2002; Raaijmakers et al. 2010). Furthermore, they are even evaluated as biocontrol agent to reduce healthcare-associated infections (Vandini et al. 2014). Such approach relies on the colonization of non-pathogenic probiotic Bacillus strains on hard surfaces of hospitals, counteracting proliferation of pathogenic strains.

B. cereus (sensu lato) is a group of Gram-positive, spore-forming, facultative anaerobic bacteria that display a wide ecological diversity and variety of lifestyles. Members of this group are commonly found in soils, where they can live as saprophytes (Vilain et al. 2006), but they can also establish symbioses with plants and soil insects, and act as pathogens in different animal hosts (Swiecicka 2008; Ceuppens, Boon and Uyttendaele 2013). It has been suggested that all members of the B. cereus group can go through a life cycle during which they establish a symbiotic relationship with an appropriate invertebrate host (Jensen et al. 2003). Filamentous spore-forming bacteria belonging to the B. cereus cluster were found in the guts of soil-dwelling arthropods, and described as a possible symbiotic life cycle (Margulis et al. 1998). Bacillus spores present in the soil are ingested by arthropods
and can germinate in their guts, a moist and nutrient-rich environment. There, they grow in their filamentous form and attach to the intestinal epithelium. Cells are then released from the distal end of the filaments and return to the soil with the animal’s feces (Margulis et al. 1998).

In this study we present the *B. toyonensis* strain VU-DES13 isolated from the gut of the soil-dwelling springtail *Folsomia candida* (Collembola: Hexapoda). We provide the initial description of this isolate, focusing on antibiotic resistance and antimicrobial production. The potential for the production of inhibitory substances was assessed using growth inhibition assays targeted against a variety of pathogenic microorganisms. Reporter assays were then applied to elucidate mechanisms of action. We also describe the genome of this strain, with an emphasis on the presence of secondary metabolite biosynthetic gene clusters that could represent a genetic basis for antibiotic activity. Finally, we discuss the ecological relevance of our findings in the context of colonization resistance of the springtail *F. candida* from entomopathogenic fungi.

### 5.3. Materials and methods

#### 5.3.1. Isolation procedure

*Folsomia candida* was reared in plastic boxes with a bottom of plaster of Paris. The animals were kept in climate rooms in stable conditions (20°C temperature, 75% humidity and a 12:12 light dark regime) and they were fed dried baker's yeast (Dr. Oetker, Bielefeld, Germany). Guts were dissected from ten adult springtails and crushed in sterile Phosphate Buffer Saline (PBS) using a plastic pestle. Ten-fold dilutions of the extract were prepared and 100 µl of each dilution was spread on nutrient agar (15.0 g peptone, 3.0 g yeast extract, 6.0 g sodium chloride, 1.0 g D(+) glucose, 15.0 g agar). Plates were incubated at 30°C and bacterial growth was checked daily. Colonies with different morphologies were transferred to fresh plates and kept at 4°C until characterization (see further details about 16S rRNA gene sequencing).

#### 5.3.2. Description of the isolate

**Phenotypic characterization with the API system**

Two API identification kits, the API 20E (first 12 tests) and the API 50CH (bioMérieux, Marcy-l’Étoile, France), were used to determine the physiological profile of *Bacillus* VU-DES13. The method is based on detection of metabolism of substrates by the microorganisms, revealed by a color change due to the precipitation or enzymatic conversion. Alternatively fermentation is detected by a drop in pH. The biochemical panel of the API 20E test measures the metabolism of various carbohydrates and amino-acids and detects the presence of specific enzymes,
such as gelatinase and β-galactosidase. The API 50CH test provides a detailed carbohydrate fermentation profile.

Bacteria were grown overnight on nutrient agar and harvested in 2 ml normal saline (NaCl 0.85%, w/v). Two different suspensions were then prepared, one in normal saline for the API 20E test, and one in API 50 CHB/E medium for the API 50CH test. The wells of the strips were inoculated with 120 µl of microbial dilution. The test was conducted in duplicate. The strips were incubated at 34°C for 48 h and checked at 24h and 48h. The last 8 tests of the API20E strips were not used, since the pH indicator is not suitable for measuring the fermentation of carbohydrates by Bacillus (Logan and Berkeley 1981).

**Antibiotic resistance**

We tested the susceptibility of Bacillus VU-DES13 to penicillin by using the broth microdilution method described by Wiegand *et al.* to determine the minimal inhibitory concentration (MIC) (Wiegand, Hilpert and Hancock 2008). Bacteria were grown overnight in Nutrient Broth (Sigma-Aldrich, St. Louis, MO, USA). The optical density of liquid bacterial cultures at 600 nm was measured on a spectrophotometer (SmartSpec Plus Spectrophotometer, Bio-Rad, Hercules, CA, USA). The cultures were then diluted to OD_{600} = 0.1, corresponding approximately to 5x10^6 cfu/ml (Biesta-Peters *et al.* 2010) to perform the test. We tested a range of ten two-fold increasing concentrations of penicillin G sodium salt (Sigma-Aldrich), from 800 µg/ml to 1.5 µg/ml. *Micrococcus luteus*, a gram-positive bacterium highly sensitive to β-lactam antibiotics, was used as a positive control. The test was performed in a 96-well microtiter plate. For each row of the plate, ten wells were used to test the different concentrations of antibiotic, one well was used to test bacterial growth and one well was used to control for contamination. To minimize evaporation during the incubation, the test was conducted in a volume of 200 µl per well. The test wells were inoculated with 100µl antibiotic dilution and 100 µl bacterial culture; the growth control with 100 µl bacterial culture and 100 µl NB; and the sterility control wells with 100 µl NB. The plate was incubated overnight at 30°C in a microplate incubator (THERM0star, BMG Labtech GmbH, Ortenberg, Germany) and optical density was measured the next day on a microplate reader (Spark 10M, Tecan, Männedorf, Switzerland).
5.3.3. Antimicrobial activity

Pathogen inhibition by isolate

The strains isolated from the springtail guts were initially screened for inhibitory activity against nine pathogens: the bacteria *Staphylococcus aureus*, *Escherichia coli* (Proteobacteria) and *Bacillus subtilis* (Firmicutes), the fungi *Rhizoctonia solani* (Basidiomycota), *Fusarium solani* and *F. oxysporum* (Ascomycota), and the oomycetes *Pythium ultimum*, *Saprolegnia diclina* and *Phytophthora capsici* (Oomycota). The strains and growing conditions for each of these microorganisms are specified in Table 1.

Table 1. Pathogens used in the inhibition assays with *Bacillus* VU-DES13 isolate and extract

<table>
<thead>
<tr>
<th>Pathogen</th>
<th>Growth medium</th>
<th>Growth temperature</th>
<th>Seeding medium</th>
<th>Positive control (30 μg)</th>
<th>Seeding</th>
<th>Incubation time</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Staphylococcus aureus</em> 533R4 (DSMZ 20231) (strain ACT12600)</td>
<td>LB</td>
<td>37°C</td>
<td>MH-agar</td>
<td>kanamycin</td>
<td>0.5 McFarland</td>
<td>24h</td>
</tr>
<tr>
<td><em>Escherichia coli</em> WA321 (DSMZ 4509) (Dam mutant of <em>E.coli</em> K12 AB1157)</td>
<td>LB</td>
<td>37°C</td>
<td>MH-agar</td>
<td>kanamycin</td>
<td>0.5 McFarland</td>
<td>24h</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em> strain 168</td>
<td>LB</td>
<td>37°C</td>
<td>MH-agar</td>
<td>kanamycin</td>
<td>0.5 McFarland</td>
<td>24h</td>
</tr>
<tr>
<td><em>Pseudomonas syringae</em> DC3000</td>
<td>LB low salt</td>
<td>28°C</td>
<td>MH-agar</td>
<td>kanamycin</td>
<td>0.5 McFarland</td>
<td>24h-48h</td>
</tr>
<tr>
<td><em>Micrococcus luteus</em> (strain ATCC49732)</td>
<td>LB</td>
<td>30°C</td>
<td>MH-agar</td>
<td>kanamycin</td>
<td>0.5 McFarland</td>
<td>24h</td>
</tr>
<tr>
<td><em>Saprolegnia diclina</em> 1152F4</td>
<td>PDA</td>
<td>25°C</td>
<td>NA</td>
<td>chlorothalonil</td>
<td>NA</td>
<td>3-7 days</td>
</tr>
<tr>
<td><em>Pythium ultimum</em> P17</td>
<td>1/5 PDA</td>
<td>28°C</td>
<td>NA</td>
<td>kanamycin</td>
<td>NA</td>
<td>3-7 days</td>
</tr>
<tr>
<td><em>Phytophthora capsici</em></td>
<td>PDA</td>
<td>28°C</td>
<td>NA</td>
<td>kanamycin</td>
<td>NA</td>
<td>3-7 days</td>
</tr>
<tr>
<td><em>Fusarium oxysporum</em></td>
<td>1/5 PDA</td>
<td>28°C</td>
<td>NA</td>
<td>chlorothalonil</td>
<td>NA</td>
<td>3-7 days</td>
</tr>
<tr>
<td><em>Fusarium solani</em></td>
<td>1/5 PDA</td>
<td>28°C</td>
<td>NA</td>
<td>chlorothalonil</td>
<td>NA</td>
<td>3-7 days</td>
</tr>
<tr>
<td><em>Rhizoctonia solani</em> AG2.2 IIB</td>
<td>PDA</td>
<td>28°C</td>
<td>NA</td>
<td>pyraclastrobin</td>
<td>NA</td>
<td>3-7 days</td>
</tr>
<tr>
<td><em>Candida albicans</em> BSMY 212 (DSMZ 10697)</td>
<td>LB</td>
<td>37°C</td>
<td>MH-agar</td>
<td>ketoconazole</td>
<td>1.0 McFarland</td>
<td>24h</td>
</tr>
</tbody>
</table>
To test the isolates against the bacteria, we used the agar overlay method. The agar plug method was used to test antimicrobial activity against fungi and oomycetes. For the agar overlay method, the isolates were grown overnight in Triptic Soy Broth (TSB) (Sigma-Aldrich, St. Louis, MO, USA), at 30°C with shaking at 400 rpm. Subsequently, 5 µl of liquid culture of each isolate was transferred to solid medium. *Pseudomonas protegens* Pf5, a broad range biocontrol model strain, was used as a positive control on each plate. The plates were incubated overnight at 28°C. On the same day, liquid cultures of the bacterial pathogens were prepared by inoculating colonies in 4 ml of LB broth base (Lennox L Broth Base, Thermo Fisher Scientific) and growing them overnight at 30°C with shaking at 400 rpm. The following day, growth of the strains of interest on the agar plates was observed, and the pathogens grown during the night were prepared for the overlay. 100 µl of liquid culture of the pathogen was inoculated in 2 ml top agar (Supplementary Information), mixed well by vortexing and poured over the plate to completely cover the colonies of the isolates. The plates were then incubated overnight at the appropriate temperature for the pathogens (Table 1) and the next day the presence of inhibition zones was recorded. For the agar plug method, agar plugs covered in mycelium of the pathogens were placed on the plates between the isolates’ colonies, after the isolates of interest had grown on the plates. The plates were sealed with parafilm and incubated at the appropriate temperature for the pathogens (Table 1) until the growth allowed the observation of clear inhibition zones. For both the agar overlay and the agar plug method, the screening was conducted on three different media: 1/10 Tryptic Soy Agar (1/10TSA), Potato Dextrose Agar (PDA) and Nutrient Agar (NA) (Supplementary Information).

**Pathogen inhibition by culture extracts made by SPE**

*Solid Phase Extraction (SPE)*

*Bacillus* VU-DES13 was grown overnight in 10 ml Nutrient Broth (NB) at 30°C with 225 RPM shaking. 2 ml of culture was inoculated in 250 ml Potato Dextrose Broth (PDB) or NB and incubated for three days with shaking at 150-200 rpm. The culture was then centrifuged for 20 minutes at 3000 rpm and the supernatant was extracted by solid phase extraction (SPE). In order to capture and purify a broad range of secondary metabolites with unknown chemistry the ABN Evolute® Express columns for acidic, basic and neutral analytes were used. To that end, 6 ml column volume with 200 mg of the above mentioned sorbent mass were used to extract *Bacillus* VU-DES13 cultures grown in PDB and NB medium. Columns with 500 mg sorbent mass were used in bioassay-guided screening. Before applying the samples, the columns were conditioned with 6 ml methanol and equilibrated with 6 ml water. Afterwards, the columns were washed with 6 ml of water and finally eluted with 6 ml of methanol to collect the analytes. The extracts obtained with SPE from liquid *Bacillus* VU-DES13 cultures were used for i) agar diffusion assays; ii) HPLC analysis; iii) reporter
assays; and iv) HPLC fractionation for bio-assay guided screening. For both the HPLC analysis and the reporter assays, 3 ml of the extract resulting from SPE was used. These volumes were completely evaporated in autosampler point vials using two nitrogen evaporators (Dionex ASE 500 and Pierce Reacti-Therm III & Reacti-Vap III). The sample to be used for HPLC was dissolved in 1 ml methanol, and the one used for the reporter assays was dissolved in 50 μl DMSO.

*Inhibition from SPE extracts*

Agar diffusion assays were performed to investigate the antimicrobial properties of extracts of *Bacillus* VU-DES13 grown in NB and PDB for three days. The assays were performed against the bacteria *E. coli, M. luteus, Pseudomonas syringae, B. subtilis* and *S. aureus* and the yeast species *Candida albicans* (Table 1). The assays were performed in triplo with slight modification of the EUCAST guidelines (EUCAST Disk diffusion Manual version 5.0, January 2015). Liquid cultures of the pathogens in LB were diluted to McFarland standard 0.5, with the exception of cultures of *C. albicans* that were diluted to McFarland standard 1.0, and the dilutions were used to evenly seed plates of MH agar with a sterile cotton swab. The plates were rimmed and left to dry for a few minutes, then wells were made in the agar using a sterile cork borer. 50 μl of *Bacillus* VU-DES13 extract, medium only extract and 30 μg of ketoconazole or kanamycine as a positive control for *C. albicans* or bacteria respectively were applied to the wells. The plates were then incubated at the appropriate temperature for the pathogen (Table S1). After one or more days, inhibition zones were recorded.

*Reporter assay: mode of action*

Bacterial luciferase reporter assays were constructed as described in Tyc *et al.* (submitted) and were used to investigate the mode of action of *Bacillus* VU-DES13 extract at sub-inhibitory levels. The extract was exposed to *E. coli*-based luciferase reporter bacteria in 96 well plates. The reporter bacteria carries the Gr-luciferase reporter plasmid pCS26Pac in which the autonomous lux operon of *Photorhabdus luminescens* is transcriptionally regulated by the mechanism of choice. Derivative plasmid pBAD33 (Guzman *et al.* 1995; Bjarnason, Southward and Surette 2003) was used in the ΔampD (CGSC 8388) background, given the kanamycin resistance used in the deletion strategy (Baba *et al.* 2006). These genetically modified bacteria generate a quantifiable signal of luminescence in the presence of specific environmental conditions (Van der Meer and Belkin 2010). Table 2 summarizes the bacterial reporters used in this study and the mode of action they can detect.
### Table 2. *E. coli* reporters used in this study.

<table>
<thead>
<tr>
<th>Reporter</th>
<th>Indicative for</th>
<th>Plasmid</th>
<th>Donor</th>
<th>Transcription factor</th>
<th>Promoter</th>
<th>Host</th>
<th>Reference (transcriptional mechanism)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPHZlux-1</td>
<td>redox-cycling compounds</td>
<td>pCS26Pac</td>
<td><em>Pseudomonas aeruginosa</em> PA14</td>
<td>SoxR</td>
<td>PA35160</td>
<td>DH5α</td>
<td>(Dietrich et al. 2008)</td>
</tr>
<tr>
<td>PSOSlux-2</td>
<td>DNA damage (SOS)</td>
<td>pCS26Pac</td>
<td><em>E. coli</em> CDA plasmid</td>
<td>NA</td>
<td>colD</td>
<td>MG1655</td>
<td>(Tecon et al. 2010)</td>
</tr>
<tr>
<td>pBLAx-2</td>
<td>β-lactams and cell wall stress</td>
<td>pBAD33</td>
<td><em>Pseudomonas protegens</em> Pf-5</td>
<td>AmpR</td>
<td>ampC</td>
<td>ΔampD (CGSC 8388)</td>
<td>(Jacobs, Frere and Normark 1997) (Valtonen, Kurittu and Karp 2002)</td>
</tr>
<tr>
<td>pBPlux</td>
<td>general toxicity</td>
<td>pCS26Pac</td>
<td>NA</td>
<td>NA</td>
<td>artificial</td>
<td>DH5α</td>
<td>Braatsch et al. (2008)</td>
</tr>
<tr>
<td>pBPlux-2</td>
<td>general toxicity</td>
<td>pBAD33</td>
<td>NA</td>
<td>NA</td>
<td>artificial</td>
<td>ΔampD (CGSC 8388)</td>
<td>Braatsch et al. (2008)</td>
</tr>
</tbody>
</table>

The bacterial luciferase reporter assays were performed in 96-well plates. 1.5 μl of *Bacillus* VU-DES13 extract or reference substance was pipetted into the wells, followed by 148.5 μl of bacterial reporter solution (5% bacterial reporter in MOPS minimal medium + glucose). Subsequently, the plates were measured on luminescence after incubation with shaking at 400 rpm (temperature and duration depending on the type of reporter) with a luminometer (Berthold Centro XS3 LB 960). Relative Light Unit (RLU) data were analyzed using GraphPad Prism version 5.03 and compared to the RLU data of a calibration curve of known reference compounds and concentrations including a DMSO blank. Values were converted to a fraction of the DMSO blank, constituting the induction factor, and a percentage of the maximum value in the calibration curve. Values lower than the detection limit were excluded. The limit of quantification was calculated by adding the mean DMSO blank value to the standard deviation multiplied by ten. Toxic concentrations in the calibration curve were excluded in order to obtain a reliable goodness of fit (R²≈1.0).

### 5.3.4. HPLC analysis

**HPLC analysis of SPE extract**

The total eluate volume of 6 ml resulting from the SPE of *Bacillus* VU-DES13 grown in PDB was used to perform HPLC fractionation (Agilent Technologies 1260 Infinity Binary LC System). The extract was evaporated to dryness and dissolved in 1 ml methanol. 10 μl of extract was injected for UV diode array detector (DAD) HPLC fractionation along with the extract of a PDB blank. This extracts was separated on a Phenomenex Kinetex 2.6u Biphenyl 100Å 150 x 4.6 mm column and eluted with a 3% to 30% gradient of acetonitrile in water. During the complete time of elution a
steady flowrate of 1.5 ml/min was used at a maximum pressure of 600 bar and the absorbance in the 200 to 900 nm spectral range was recorded by DAD.

**Bio-assay guided fractionation of SPE extract**

Half of the 6 ml SPE extract from a three day culture in PDB was dissolved in 1 ml of methanol. From this, 100 μl aliquots were injected in a similar HPLC setup as described above, but adapted to make use of a semi-prep HPLC column of the same chemistry, i.e. Kinetex 5u Biphenyl 250 x 10 mm column. The adjusted flow speed conditions were 5ml/min and 100 ul injection (of consecutive samples). Fractions of the extract were collected during the time of elution in windows of one minute using a Teledyne ISCO Foxy 200 fraction collector for further investigation of their antimicrobial properties and mode of action. The absorbance in the 200 to 900 nm spectral range was recorded by DAD. The fractions were dried under vacuum (Genevac Rocket evaporator) and subsequently dissolved in 100 μl DMSO. They were then used for reporter assays (conducted at 1% exposure volume of ten-fold diluted fractions) and agar well diffusion assays (conducted with 50 ul of undiluted fractions).

5.3.5. Genome analysis

**De novo genome shotgun sequencing, assembly and annotation**

Genomic DNA was extracted from a pelleted 5 ml overnight nutrient broth culture with the Macherey-Nagel Nucleospin Soil kit. A genomic library was prepared by enzymatic shearing of the gDNA with the Ion Xpress Plus Fragment Library kit (ThermoFisher) in combination with size selection on a E-Gel ® SizeSelect TM 2% Agarose Gel (ThermoFisher). The sequencing template was prepared with 10 pM of the library Ion One Touch 2 (ThermoFisher). All intermediate steps were monitored on a 2100 Bioanalyzer (Agilent). Finally, a 400 bp run was executed on an Ion Torrent PGM, making use of the Ion PGM™ Hi-Q™ Sequencing Kit (ThermoFisher). After adapter removal, the sequences were assembled using Spades (Bankevich et al. 2012), which performs autocorrection, Phred score selection and trimming. The assembly was optimized using MeGAMerge (Scholz, Lo and Chain 2014). The genome was annotated automatically using the Prokka pipeline (Seemann 2014), extended with the analysis for carbohydrate active enzymes, CAZY (Lombard et al. 2014) and for secondary metabolism biosynthetic gene clusters, anti-SMASH (v3.0.5) (Weber et al. 2015). Additionally, BAGEL3 (van Heel et al. 2013) was performed in order to confirm the bacteriocin clusters. The tRNA genes were predicted using the package ARAGORN (Laslett and Canback 2004).

The assembled genome was screened using blastn for similarities with the plasmids and associated virulence genes from *B. anthracis*: plasmid pXO1 with associated
genes cya, lef pagA and repX, and plasmid pXO2 with capA, capB, capC, capD, capE and repS. The assembly was also checked for similarities to virulence plasmids from B. thuringiensis pAW63 and pBT9727 (Van der Auwera, Andrup and Mahillon 2005). Delta-endotoxines, related to the Bt-toxin, were searched by checking the Pfam family output from the Prokka pipeline for the following families: PF03945, PF00555, PF09131, PF03944, PF05431, PF01338, PF12495 and PF03538 (Finn et al. 2016).

Antibiotic resistance genes were detected by applying hidden Markov models (HMMs) scanning of the predicted amino acid sequences against the curated Resfams v1.2 database of protein families and ontologies associated with verified antibiotic resistances. The core database was trained with CARD (Comprehensive Antibiotic Resistance Database), LacED (Lactamase Engineering Database) and Jacoby’s and Bush’s collection, while the full database was supplemented with verified accessory functions from protein families that contribute to the resistance (Gibson, Forsberg and Dantas 2015). In the case of multiple hits, the hit with the lowest E-value was reported. For the detection of plasmids, the coverage of mapped reads on the respective contigs, as mapped by the bwa mem algorithm (Li 2013), and as implemented in Unipro UGENE (Golosova et al. 2014) was calculated. The PlasmidFinder 1.3 tool (Carattoli et al. 2014) was deployed in order to find replication elements of Gram positive plasmids.

**Phylogenetic analysis**

In order to characterize the isolate of this study, we initially amplified the 16S rRNA gene with the primer pair 27F (5’ - AGA GTT TGA TCM TGG CTC AG - 3’, (Lane 1991)) and 1492R (5’ - CGG TTA CCT TGT TAC GAC TT - 3’, (Turner et al. 1999)) and conducted Sanger sequencing (Macrogen, Amsterdam). The PCR reactions were set up as follows: 1µl genomic DNA, 2.5 µl of each primer (5 µM), 1 µl dNTPs (10mM each), 10 µl Phusion Buffer (5X) (New England Biolabs, Ipswich, MA, USA), 31 µl ELGA water, 1.5 µl DMSO, 0.5 µl Phusion DNA polymerase (2 U/µl), in order to conduct an amplification with the following program: initial denaturation for 30 s at 98°C, 35 amplification cycles (10 s at 98°C, 10 s at 53°C, 45 s at 72°C), and a final extension step of 10 min at 72°C. After visual verification by electrophoresis on 1.5% (wt/vol) agarose gels stained with ethidium bromide the PCR fragments were sent to Macrogen (Amsterdam, the Netherlands) for purification and Sanger sequencing. The raw chromatograms and base-called reads were trimmed and subsequently assembled in CLC Genomics Workbench software, version 5.1 (CLC Bio-Qiagen, Aarhus, Denmark). By means of the BLAST Sequence Analysis Tool (Madden 2002) and the Sequence Match Tool of the Ribosomal Database Project (RDP) (Cole et al. 2014) the sequences were assigned to the closest taxonomic identity.
Two methods of whole genome phylogeny were conducted. First, the novel genome was compared to other *Bacillus* genomes using feature frequency profiling (FFP), a method based on counting the number of features of a particular length that occur in a genome (Sims *et al.* 2009; Wang and Ash 2015). The frequency of features was determined for the novel strain and for 30 additional strains representative of the phylogenetic clusters of the *B. cereus* clade (Zwick *et al.* 2012; Liu *et al.* 2016), taking *B. subtilis* strain 168 as an outgroup. Prior to the calculation of the distance matrix, the optimal feature length was determined in the newly sequenced genome. One thousand replicated neighbor-joining trees were constructed, based on a Jensen-Shannon divergence distance matrix calculated from the normalized feature frequency profiles. Subsequently, a consensus tree was made with PHYLIP (Felsenstein 2002) and plotted with FigTree v1.4.1 (Rambaut 2014). The second approach used for whole genome phylogenetic positioning was the Genome BLAST Distance Phylogeny, which allows to perform digital DNA:DNA hybridization (dDDH) (Meier-Kolthoff *et al.* 2013) to obtain whole genome sequence similarity scores. This approach was implemented using a webtool (available on http://ggdc.dsmz.de/) and according to the recommendations by the authors. The method uses a species threshold level of 70% dDDH, and is robust for the analysis of incomplete genome sequences.

**Comparative genomics**

The genome rearrangements, synteny and gene content were compared with the type strain *B. toyonensis* BCT-7112T (Jiménez *et al.* 2013a), by making use of the progressiveMauve alignment algorithm in Mauve (Darling, Mau and Perna 2010). Differential genes compared to the two other *B. toyonensis* genomes were detected by performing a comparison with Panseq (Laing *et al.* 2010). The respective GO terms associated with these protein coding genes were compared and summarized with the redundancy reducing and semantic analysis offered by Revigo (Supek *et al.* 2011).
5.4. Results

5.4.1. Isolation procedure

Growth of colonies from *Folsomia candida*’s dissected guts plated on agar was observed from day 2 of culturing. Colonies of different morphologies were picked and the 16S region was sequenced to confirm their identity. Bacteria belonging to *Pseudomonas*, *Bacillus*, *Ochrobactrum* were identified. During a preliminary screening, a *Bacillus* strain showed high inhibitory activity against a wide range of microbial pathogens, and was therefore selected to conduct further analyses.

5.4.2. Description of the isolate

The genus *Bacillus* comprises Gram-positive, aerobic endospore-forming rods. *Bacillus* VU-DES13 cells have rod-shape, observe singularly, in pairs or in filaments, and produce endospores. They also grew in anaerobic conditions. Colonies appeared flat, milky-white and slightly granulated (Figure 1).

![Image of Bacillus toyonensis VU-DES13](image)

**Figure 1.** Micrographs of *Bacillus toyonensis* VU-DES13. The picture on the right shows sporulating cells.

**Phenotypic characterization with the API system**

We used two API systems, API 20E and API 50CH, to characterize the phenotype of *Bacillus* VU-DES13. The API system is an efficient method to discriminate *Bacillus* based on their physiological profile (Logan and Berkeley 1984). Table 3 shows the phenotypic characters of *Bacillus* VU-DES13, in comparison with *B. toyonensis* BCT-7112, *B. cereus* CECT-148, *B. thuringiensis* CECT-197 and *B. anthracis* (as reported by Jiménez et al. 2013b). The results indicate that *Bacillus* VU-DES13 has a physiological profile distinct from closely related *Bacillus* species, including the previously described *B. toyonensis* BCT-7112⁷. Relative to the other *B. cereus* clade species in the table, our isolate was not able to ferment saccharose, methyl-αD-glucopyranoside and D-turanose in the API 50CH test. The API 20E test showed that *Bacillus* VU-DES13 cannot utilize citrate as substrate nor produce acetoin. Furthermore, *Bacillus* VU-DES13 does not show L-arginine dehydrolase activity when compared to the type strain *Bacillus* BCT-7112⁷.
Table 3. Phenotypic characteristics of *Bacillus* VU-DES13, compared to closely related strains in the *Bacillus* genus, as revealed by selected tests from the API 20E and API 50CH systems.

<table>
<thead>
<tr>
<th></th>
<th><em>Bacillus toyonensis</em> VU-DES13</th>
<th><em>Bacillus toyonensis</em> BCT-7112&lt;sup&gt;+&lt;/sup&gt;</th>
<th><em>Bacillus cereus</em> (CECT 148)</th>
<th><em>Bacillus thuringiensis</em> (CECT 197)</th>
<th><em>Bacillus anthracis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>API 20E</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaerobic growth</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>L-Arginine dihydrolase</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Citrate utilization</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Acetoin production</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gelatinase</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td><strong>API 50CH</strong> (carbohydrate fermentation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>±</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Ribose</td>
<td>±</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>+</td>
</tr>
<tr>
<td>D-Mannose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Methyl-α-D-Glucopyranoside</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Amygdalin</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arbutin</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>Salicin</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>D-Cellobiose</td>
<td>±</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>D-Saccharose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-Trehalose</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Starch</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Glycogen</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-Turanose</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

± indicates a weak positive result. A question mark indicates that different strains gave different results. Data obtained from Jiménez et al. 2013b.

**Antibiotic resistance**

Figure 2 shows the results of the broth microdilution method to determine the minimum inhibitory concentration (MIC) of penicillin. The figure suggests that *Bacillus* VU-DES13 is not susceptible to penicillin, even at the highest concentration tested. In fact, bacterial cultures of *Bacillus* VU-DES13 did not show a significant difference in OD<sub>600</sub> growth between the control conditions (M=0.84, SD=0.02) and 800 µg/ml of antibiotic (M=0.81, SD=0.02). In contrast, *Micrococcus luteus* growth was inhibited by penicillin already at the lowest concentration. Indeed, the OD<sub>600</sub> bacterial cultures of *M. luteus* grown in control conditions (M=0.23, SD=0.02)
significantly differed from bacterial cultures grown in any of the concentrations of penicillin.

![Graph showing cell density of Bacillus VU-DES13 and Micrococcus luteus cultures grown overnight at 30°C at increasing concentrations of penicillin. The cell density was inferred from optical density measurements at 600 nm (assuming that OD_{600} = 0.1 corresponds to 5x10^6 cfu/ml). The starting OD_{600} for both bacterial cultures was 0.1. The sterility control contained only broth.]

**Figure 2.** Cell density of *Bacillus* VU-DES13 and *M. luteus* cultures grown overnight at 30°C at increasing concentrations of penicillin. The cell density was inferred from optical density measurements at 600 nm (assuming that OD_{600} = 0.1 corresponds to 5x10^6 cfu/ml). The starting OD_{600} for both bacterial cultures was 0.1. The sterility control contained only broth.

### 5.4.3. Antimicrobial activity

**Pathogen inhibition by isolate**

We tested the ability of *Bacillus* VU-DES13 to inhibit the growth of several microbial pathogens. The results of the growth inhibition assay are shown in Table 4. *Bacillus* VU-DES13 did not inhibit bacterial pathogens, but it inhibited both oomycetes and fungi, with the strongest effect observed against fungi under poor media conditions (1/10 TSA and PDA plates).

**Table 4. Growth inhibition of target pathogens by Bacillus VU-DES13**

<table>
<thead>
<tr>
<th></th>
<th><em>S. aureus</em></th>
<th><em>E. coli</em></th>
<th><em>B. subtilis</em></th>
<th><em>P. ultimum</em></th>
<th><em>P. capsici</em></th>
<th><em>S. diclina</em></th>
<th><em>F. oxysporum</em></th>
<th><em>F. solani</em></th>
<th><em>R solani</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1/10 TSA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>++</td>
<td>++</td>
<td>+/++</td>
<td>++</td>
</tr>
<tr>
<td>PDA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-/+++</td>
</tr>
<tr>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
</tr>
</tbody>
</table>

x = no growth of pathogen; - = no inhibition; + = growth reduction; ++ = inhibition (size of inhibition < 5mm), +++ = inhibition (size of inhibition > 5mm)
Chapter 5

**Pathogen inhibition by SPE extracts**

We also exposed *Bacillus* VU-DES13 PDB and NB extracts to six non-filamentous pathogens: the bacteria *Escherichia coli*, *M. luteus*, *Pseudomonas syringae*, *B. subtilis* and *Staphylococcus aureus* and yeast species *Candida albicans*. The PDB extract inhibited growth of *C. albicans* and *M. luteus*, and the NB extract inhibited the growth of all tested pathogens, except *E. coli* (Figure 3). While extracts from growth in NB extract was effective against a wider variety of pathogens, the inhibitory effect from PDB extract was stronger and more specific.

![Figure 3](image)

**Figure 3.** MH agar diffusion assays with extract of *Bacillus* VU-DES13 grown in NB medium. The extract inhibited the growth of 1) *B. subtilis*; 2) *M. luteus*; 3) *P. syringae*; and 4) *S. aureus*.

**Reporter assay: mode of action**

The strongest inhibitory effects of *Bacillus* VU-DES13 was observed with the extract grown for 3 days in PDB. Therefore, we investigated the mechanism of antimicrobial activity within this extract. We used the bacterial reporters pPHZlux-1 (DH5α), pSOSlux-2 (MG1655) and pBLAlux-2 (ΔampD) to detect redox cycling activity, DNA damage (SOS response) and cell wall stress respectively, since these are common mode of actions of antimicrobials. Furthermore, the general toxicity of the samples was monitored by using the extract to expose the reporters pBPlux (DH5α), pBPlux (MG1655) and pBPlux-2 (ΔampD). Results from the reporter assays for cell wall stress and general toxicity are shown in Figure 4.
Figure 4. A: Graphs of fold induction of the pBLAlux-2 (A) and the pBPlux-2 (C) bacterial luciferase reporters exposed to dilutions of Bacillus VU-DES PDB extract. The calibration curve for induction of the pBLAlux-2 Δ ampD by the reference compound (penicillin G) is given in B.

The reporter assays revealed that Bacillus VU-DES13 PDB extract induces cell wall stress, but not redox cycling or DNA damage. The signal indicating cell wall stress increased with further dilutions of Bacillus VU-DES13 PDB extract (Figure 4A), and was highest for dilution factor 48, where reporter activity increased threefold relative to the DMSO control. Subsequently, increasing dilutions produced weaker signals, which were not explainable by general toxicity (Figure 4C). The general stress
reporter induction also did not indicate toxicity of less diluted extracts (Figure 4C). The signal from the 48 times diluted extract was 27% of the maximum induction by penicillin G, used as reference compound, in the calibration curve, which in turn was about 12 times compared to baseline (compare figure 4A to 4B).

5.4.4. HPLC analysis and bioassay-guided fractionation

Gradient HPLC was used to separate the PDB extracts of Bacillus VU-DES13. Absorption peaks were detected at various wavelengths (Figure 5A and 5B). Differential absorption peaks between the bacterial extract and the sterile medium are evident at 250 nm between minutes 4.5 and 19.5.

In order to effectively separate the Bacillus VU-DES13 extracts and collect the fractions, a Kinetex 5u Biphenyl 250 x 10 mm semi-prep column was used with an adjusted flow rate. This changed the peak profile compared to the analytical column (compare blue line in Figure 6A to blue line in Figure 5). Fractions were collected at one minute intervals and were used to perform agar well diffusion assays to test inhibition of C. albicans (Figure 6B) and to study cell wall stress-induced bioactivity and general toxicity by using the pBLAlux-2 ($\Delta$ampD) and pBPlux-2 ($\Delta$ampD) reporter assays (Figure 6C). For the reporter assays, ten-fold dilutions of each of the fractions were used.

Fourteen out of the 26 fractions inhibited the growth of C. albicans with different intensities, with fraction 21 showing the strongest inhibitory effect (Figure 6A and 6B). This fraction also clearly induced cell wall stress in the E. coli reporter assays (figure 6C).
Figure 6. A: UV-VIS DAD chromatogram at 250 nm absorbance of the *Bacillus* VU-DES13 PDB extract separated on the Kinetex 5u Biphenyl 250 x 10 mm HPLC column (blue line); and inhibition of *C. albicans* by the 26 one-minute fractions represented by the x-axis (orange line). Secondary y-axis indicated the size of inhibition zones. B: effect of the *Bacillus* VU-DES13 fractions 15 to 21, of the negative control DMSO and of 30 µg the positive control ketoconazole (KC) on *C. albicans* in an agar well diffusion assay. C: Induction (expressed as ratio of the DMSO blank on the y-axis) of the *pBLAlux*-2Δ*ampD* and the *pBPux*-2Δ*ampD* reporters by the *Bacillus* VU-DES13 fractions, indicated on the x-axis.
5.4.5. Genome analysis

**Genome assembly and annotation**

After assembly of the raw reads with SPADES and post-assembly treatment by MeGAMerge, a draft genome was available for *Bacillus* VU-DES13 (Table 6). The total genome size was estimated to be 5.45 Mbp, corresponding to 40 contigs, with an N50 of more than 25 kbp and a GC content of 35%. A detailed description of assembly and initial annotation is published (Janssens *et al.* 2017) and available as an Appendix to this chapter. The final genome sequence was submitted to Genbank under project accession number MWMG00000000 (Bioproject PRJNA374700, biosample SAMN06330328). The average coverage was estimated at 119X and no contigs were overrepresented. 5 512 open reading frames (ORFs) were predicted, in addition to 53 tRNA genes and 6 rRNA clusters. Due to the repetitive nature of the latter, and the auto correction option in the SPADES assembler, artefacts were created in the assembly sequence at the respective rRNA clusters. According to the distribution of the coverage of the mapped reads on the contigs, no evidence for the presence of a plasmid could be given. The PlasmidFinder tool failed to indicate any plasmid elements in our dataset.

**Table 6**: Summary of the assembly and annotation statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of contigs</td>
<td>40</td>
</tr>
<tr>
<td>Largest contig (bp)</td>
<td>470,007</td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>5,452,021</td>
</tr>
<tr>
<td>GC (%)</td>
<td>35.06</td>
</tr>
<tr>
<td>N50 (bp)</td>
<td>254,013</td>
</tr>
<tr>
<td>L50</td>
<td>9</td>
</tr>
<tr>
<td>Average coverage</td>
<td>119 ± 49</td>
</tr>
<tr>
<td>Number of ambiguous bases per 100 kbp (bp)</td>
<td>0.81</td>
</tr>
<tr>
<td>Predicted ORFs</td>
<td>5,512</td>
</tr>
<tr>
<td>Hypothetical proteins</td>
<td>1,650</td>
</tr>
<tr>
<td>tRNA</td>
<td>53</td>
</tr>
<tr>
<td>rRNA clusters</td>
<td>6</td>
</tr>
</tbody>
</table>

**Virulence-related genes**

The well-known human pathogen, *B. anthracis* confers its virulence by the presence of the virulence plasmids pXO1 and pXO2, essential for pathogenicity. Related plasmids pAW65 and pBT9727 have also been detected in *B. thuringiensis*. Sequences of these plasmids were used to query the *Bacillus* VU-DES13 genome
using blastn. None of these plasmids showed similarity to sequences in the Bacillus VU-DES13 genome. Some relatively short blastn hits pointed to similarities with the backbone sequences, mobile elements or encoding non-virulence genes. This was supported by the tblastx analysis of the anthrax-related virulence factors, which pointed to a short stretch of homology to the capA capsular biosynthesis gene, but failed to indicate that anthrax-related virulence genes were present in the assembly. This was also confirmed by Pfam family annotations related to various classes of Bt-toxin in the Prokka annotation output. Finally, the B. cereus toxin cereulide encoded and synthesized by a specific NRPS was not detected in the antiSMASH analysis (see further), although the annotation pointed to four hemolytic enterotoxin genes.

**Antibiotic resistance**

The predicted amino acid sequences of the assembled genome were analyzed for the occurrence of antibiotic resistance genes, by comparing them to the 166 HMMs of the full Resfam database. The results are summarized in Table 7.

<table>
<thead>
<tr>
<th>Antibiotic mechanism</th>
<th>Bacillus VU-DES-13</th>
<th>Bacillus toyonensis BCT-7112(^\text{r}) (Khatri et al. 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC Transporter</td>
<td>124</td>
<td>123</td>
</tr>
<tr>
<td>Acetyltransferase</td>
<td>124</td>
<td>113</td>
</tr>
<tr>
<td>Aminotransferase</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Beta-lactamase</td>
<td>54</td>
<td>41</td>
</tr>
<tr>
<td>Chloramphenicol Resistance</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>D-ala Ligase</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Gene Modulating Resistance</td>
<td>43</td>
<td>72</td>
</tr>
<tr>
<td>Glycopeptide Resistance</td>
<td>12</td>
<td>106</td>
</tr>
<tr>
<td>Methyltransferase</td>
<td>36</td>
<td>41</td>
</tr>
<tr>
<td>MFS Transporter</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Other Efflux</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>Quinolone Resistance</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>RND Antibiotic Efflux</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>rRNA Methyltransferase</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Stress Response</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Target Redundancy/Overexpression</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>Tetracycline MFS Efflux</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Tetracycline Ribosomal Protection</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>576</strong></td>
<td><strong>693</strong></td>
</tr>
</tbody>
</table>

Resfam analysis indicates that the Bacillus VU-DES13 isolate is resistant to β-lactams, tetracyclin, chloramphenicol and various glycopeptide antibiotics. The total number of antibiotic resistance-related genes is slightly lower in Bacillus VU-DES13
compared to the type isolate BCT-7112\textsuperscript{T}, and it is shifted towards the resistance to β-lactam compounds (54 in \textit{Bacillus} VU-DES13 compared to 41 in BCT-7112\textsuperscript{T}). Furthermore, the type isolate BCT-7112\textsuperscript{T} has almost ten times more resistance genes against glycopeptide antibiotics, and contains additional tetracycline and quinolone resistance genes.

\textit{Secondary metabolism}

AntiSmash 3.0 identified 10 secondary metabolite-related clusters (Table 8). Five gene clusters were involved in the production of bacteriocins (of which three were confirmed by the BAGEL3 package), including one lantipeptide and one lasso peptide. Furthermore two NRPS were identified, of which one cluster encodes the machinery for the biosynthesis of a siderophore, and one cluster involved in the production of a terpene (Table 8). Four clusters showed similarity hits with the MIBiG (minimum information on biosynthetic gene clusters) database (Supplementary Information). One of these clusters showed 83% similarity to a petrobactin (siderophore) and another showed 83% similarity to a thuricin (lantipeptide) biosynthetic gene cluster.

\textbf{Table 8}: Results of the antiSMASH analysis of the \textit{B. toyonensis} VU-DES-13 assembly, including information on the nature of the biosynthetic gene cluster (BGC) and the presence in the conspecific strains \textit{B. toyonensis} BCT-7112\textsuperscript{T} and BAG6O-1. A detailed table is provided in the Supplementary Information.

<table>
<thead>
<tr>
<th>Type</th>
<th>Most similar known cluster</th>
<th>Detected by BAGEL3</th>
<th>present in BCT-7112\textsuperscript{T}</th>
<th>present in BAG6O-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terpene</td>
<td>Molybdenum cofactor biosynthetic gene cluster (11% of genes show similarity)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Lasso peptide</td>
<td>-</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Nrps</td>
<td>Bacillibactin biosynthetic gene cluster (38% of genes show similarity)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>-</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Nrps</td>
<td>-</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Siderophore</td>
<td>Petrobactin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>-</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>-</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Lantipeptide</td>
<td>Thuricin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
**Phylogenetic analysis**

Blastn of the 16S rRNA gene sequence against the nr and RDP databases, indicated that the isolate belonged to the *B. cereus* clade, with 100% similarity to *B. cereus, thuringiensis* and *toyonensis* (data not shown). The 16S rRNA does not offer enough resolution to phylogenetically position between these strains (Ash *et al.* 1991), therefore we applied two whole genome methods.

Feature frequency profiling (FFP) is based on the frequency of sequence measures across different genomes (Wang and Ash 2015). The optimal feature length, i.e. the length which accounts for the highest number of different features, was estimated to be 11 bp in the newly sequenced genome. The FFP method was applied to genomes representative of every cluster of the *B. cereus* clade and other *B. toyonensis* isolates, resulting in a divergence distance matrix. This matrix was then used to produce 1 000 bootstrapping trees, on which a consensus tree, rooted using *B. subtilis* 168, was based. The second method to perform whole genome phylogeny was based on digital DNA-DNA hybridization (dDDH) measures, which were performed on the same genomes as the FFP method. The FFP method positioned *Bacillus* VU-DES13 between the other *B. toyonensis* strains (BCT-7112\textsuperscript{T} and BAG60-1), and dDDH values higher than 70% on the whole genome alignment supported this result, indicating that *Bacillus* VU-DES13, BCT-7112\textsuperscript{T} and BAG60-1 belong to the same species (Supplementary Figure S1). *Bacillus* VU-DES13 lacks the two plasmids, pBCT8 and pBCT77, that were identified in BCT-7112\textsuperscript{T}.

**Comparative genomics to the type strain BCT-7112\textsuperscript{T}**

Initial analysis by the progressive Mauve aligner, allowed us to confirm syntenic regions between the two genomes of *Bacillus* VU-DES13 and BCT-7112\textsuperscript{T} (data not shown). Subsequent analysis by Panseq revealed differential stretches of DNA between the strains. The length of these stretches amounted to ~672 kbp (159 fragments) for *Bacillus* VU-DES13 and ~260 kbp (144 fragments) for BCT-7112\textsuperscript{T}. These fragments contained 690 and 336 ORFs respectively. By comparing the annotations and gene ontology (GO) terms associated with biological process for these fragments, we interpreted the putative biological differences between both isolates (data not shown). In *Bacillus* VU-DES13, the processes encoded by the differential gene content are related to oxidation-reduction, the transport of metabolites across membranes, and spore germination. Pathogenesis factors and biosynthetic gene clusters were also detected, as well as genes for recombination and integration mechanisms and signal transduction proteins. In BCT-7112\textsuperscript{T}, the most prominent biological process represented in the differential gene content are DNA recombination, phosphorelay transduction systems, metabolism and gluconeogenesis and spore germination.
5.5. Discussion

In this study we have described a putatively important member of the *Folsomia candida* microbiome, *Bacillus toyonensis* VU-DES13. A previous study based on cloning and 16S rRNA gene sequencing isolated 95 clones from the springtail, 83 of which (87%) were identified as *B. weihenstephanensis* (Czarnetzki and Tebbe 2004a). Agamennone et al. (2015) also indicated that members of the *B. cereus* group are among the dominant groups in *F. candida* microbiome, representing 4% of the reads in a natural population of springtails (Agamennone et al. 2015). However, the limited resolution of 16S rRNA in this clade did not allow for a more specific identification within this clade. Just like the life cycle stage-specific association of *B. anthracis* and *B. thuringiensis* with ungulates and insects, the data presented here and in these earlier studies suggest that there might be an association of other *B. cereus* clade members with animals (Margulis et al. 1998; Jensen et al. 2003). *B. cereus* is abundant in soils and has been associated with some cases of food poisoning, because of the rare occurrence of the emetic toxin cereulide (Altayar and Sutherland 2006). To more precisely identify our *Bacillus* VU-DES13 isolate, we used data from Liu et al. (2016). These authors conducted a large scale phylogeny on the complete genomes of 224 *B. cereus* strains. Their results indicated the presence of 30 clusters within this clade. The role of horizontal gene transfer (HGT) in the transfer of virulence genes and plasmids between these strains is significant, yet the presence of these sequences had little correlation with the phylogenetic position within the clade (Liu et al. 2016). By using representative elements from this phylogeny, the genome sequence of *Bacillus* VU-DES13 clustered with the clade comprising *B. toyonensis* BCT-7112T and BAG60-1, two isolates originating from soil ecosystems. Interestingly, BCT-7112T is currently used as a probiotic strain in the commercial formulation TOYOCERIN®, a feed additive (Zwick et al. 2012; Jiménez et al. 2013a) to improve feed efficiency and growth of livestock animals. Another study reported high abundance of this species in the gut of fly larvae (Sánchez-Galván et al. 2017). The authors suggested that *B. toyonensis*, along with other inhabitants of the larval gut, is ingested with the breeding medium, and that it may play a role in the digestion of organic material (although no data was presented to support this hypothesis) (Sánchez-Galván et al. 2017).

Physiological profiling with API tests, and additional phenotyping, is a standardized way to describe new species or isolates of *Bacillus* (Logan and Berkeley 1981, 1984). The diversity of patterns within the genus is significant (Logan and De Vos 2009), revealing differences between *Bacillus* VU-DES13 and the previously described *B. toyonensis* BCT-7112T. This variation may be linked to different ecological niche occupied by the two strains, and is reflected by underlying genetic differences. It was observed that most of the variation between *Bacillus* genomes occurred in genes involved in niche adaptation, reflecting specializations that allow these species to
occupy different habitats (Alcaraz et al. 2010). By comparing the genome content of our isolate with the type strain, we found additional differences in the biosynthetic capacity of secondary metabolites, the presence of β-lactamases and of factors related to virulence and pathogenesis. No known genes related to pathogenesis to human were detected.

We observed a high resistance of the isolate Bacillus VU-DES13 to penicillin. Because of the 16 or 32 μg/ml cutoff in many MIC tests, it is difficult to compare the resistance to 800 μg/ml measured for our isolate (Tyson et al. 2017). From the analysis by the Resfam databases, it became apparent that Bacillus VU-DES13 contains 13 additional β-lactam when compared to the BCT-7112T genome. Antibiotic producers can evolve mechanisms to protect themselves from the compounds they are producing (Hopwood 2007). A study in the US Geological Survey has unraveled very strong correlations between heavy metal concentrations in soil and the resistance of Bacillus isolates to penicillin (Watterson, Nagy and Updegraaf 1984). They observed extreme resistance, with up to 5 mg/ml of penicillin, and the isolates belonging to the B. cereus clade were among the most resistant (Watterson, Nagy and Updegraaf 1984). Bacillus VU-DES13 was isolated from F. candida, that is putatively producing this class of secondary metabolites (Suring et al. 2016). This suggests that the resistance may be directed towards the β-lactam compound produced by the host (Roelofs et al. 2013).

Bacillus VU-DES13 and its extract inhibited the growth of bacterial- and fungal pathogens in agar inhibition assays, especially on poorer media. The different efficacy of inhibition on different media suggests that the activity of Bacillus VU-DES13 is dependent on growth conditions. The mechanism of pathogen inhibition by Bacillus isolate may be different from that underlying inhibition by the extract. Growth inhibition by live microorganisms can be due to a variety of mechanisms leading to competitive exclusion, not only the production of substances such as siderophores, signaling molecules or antimicrobials, but also limited nutrients or space resulting from the competitor’s use of these resources. Inhibition by a bacterial extract is most likely due to the presence of specific compounds, that can be identified and whose mode of action can be elucidated by using bio-assays such as the ones applied in this study. In the future, we are planning to continue this work by analyzing in more detail the most active fraction of the Bacillus VU-DES13 extract, fraction 21, by using MALDI-TOF to link specific peptides to the inhibitory activity. Such peptides could be further investigated for their capacity to inhibit pathogen growth and cause cell wall stress.

The annotation of the genome of Bacillus VU-DES13 with the packages antiSMASH and BAGEL identified 10 biosynthetic gene clusters. Three of them revealed a hit with the MIBiG database and seven were already known from the sequence of the type strain. The ribosomal antimicrobial proteins and the NRPSs were reported in a
previous publication (Zhao and Kuipers 2016). The siderophore biosynthetic cluster 6 identified in our analysis showed similarity to petrobactin, an NRPS-independent siderophore that depends on catecholate condensation (Lee et al. 2007). Cluster 10 exhibited similarity to thuricin, a bacteriocin described from B. thuringiensis (Favret and Yousten 1989). As mentioned earlier, the type strain BCT7112\textsuperscript{T} is used commercially as a probiotic for application in feed for cattle and pigs. A recent study has shed light on a possible role of this probiotic strain in the inhibition of quorum sensing, resulting in reduced biofilm formation by enterotoxic Escherichia coli (González-Ortiz et al. 2016). At this point the nature and identity of the antimicrobial agents in Bacillus VU-DES13 and BCT7112\textsuperscript{T} remains elusive.

From earlier work on F. candida it is known that this species is resistant against bacterial and fungal entomopathogens, and is even able to use them as food source, leading to increased growth and reproduction. This feeding behavior reduces the viability of remaining conidia after passage through the Folsomia gut (Broza, Pereira and Stimac 2001). In light of the high level of penicillin resistance observed here, the presence of biosynthetic gene clusters for putative β-lactam compounds and for other secondary metabolites in the host’s genome (Roelofs et al. 2013; Suring et al. 2016), and the induction of these genes in conditions of stress (Nota et al. 2008), we suggest that these genes affect the growth and metabolism of specific members of the gut bacterial community, possibly modulating colonization resistance of the microbiome against pathogens. Antibiotics are known to increase fitness of the producer in a community by inhibiting sensitive competitors, but they have also been shown to function as signaling molecules (Yim, Huimi Wang and Davies 2007) within and between species. Moreover, inter-kingdom signaling has been shown to have outstanding ecological relevance, for example in plant biocontrol and human gut health (Williams 2007; Fischbach and Segre 2016). Specific elicitors (environmental conditions and chemical triggers) have been described that can explain the highly conditional and specific expression of biosynthetic gene cluster for secondary metabolites (Niu et al. 2016; Urem et al. 2016). These findings imply a highly complex interplay between host, microbiome and environment, taking into account many variables that we have only started to unravel. With the experimental and genomic data from this study we have provided some elements to begin to understand the role of this dominant member of the F. candida’s gut microbiome in the colonization resistance against entomopathogens. The discovery of active fractions that may contain novel antimicrobials constitutes a first step towards the development of new antifungal or antibacterial compounds.
Supplementary Information

Growth media

**Table S1**: Occurrence of antibiotic resistance genes

**Table S2**: antiSMASH analysis of *Bacillus toyonensis* VU-DES13

**Figure S1**: Consensus tree of representative genomes within the *Bacillus* clade
# Growth media

**Tryptic Soy Agar (TSA)**

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>g/l</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptic Soy Broth (TSB) (Sigma-Aldrich)</td>
<td>30.00</td>
</tr>
<tr>
<td>Agar</td>
<td>15.00</td>
</tr>
</tbody>
</table>

**1/10 Tryptic Soy Agar (1/10 TSA)**

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>g/l</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium chloride</td>
<td>5.00</td>
</tr>
<tr>
<td>Potassium dihydrogen phosphate</td>
<td>1.00</td>
</tr>
<tr>
<td>Tryptic Soy Broth (TSB) (Sigma-Aldrich)</td>
<td>3.00</td>
</tr>
<tr>
<td>Agar</td>
<td>15.00</td>
</tr>
</tbody>
</table>

**Potato Dextrose Agar (PDA)**

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>g/l</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potato Dextrose Broth (PDB) (Sigma-Aldrich)</td>
<td>24.00</td>
</tr>
<tr>
<td>Agar</td>
<td>15.00</td>
</tr>
</tbody>
</table>

**Top Agar**

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>g/l</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB Broth Base (Thermo Fisher Scientific)</td>
<td>20.00</td>
</tr>
<tr>
<td>Agar</td>
<td>7.00</td>
</tr>
</tbody>
</table>
Table S1. Output of the Resfam analysis on the occurrence of antibiotic resistance genes.

<table>
<thead>
<tr>
<th>Antibiotic mechanism</th>
<th>Resfam Identifier</th>
<th>Bacillus VU-DES-13</th>
<th>B. toyonensis BCT-7112&lt;sup&gt;T&lt;/sup&gt; (Khatri et al. 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC Transporter</td>
<td>RF0007, RF0008, RF0009, RF0100, RF0143, RF0089, RF0107&lt;br/&gt;RF0011, RF0012, RF0013, RF0014, RF0015, RF0018, RF0031, RF0032, RF0016, RF0045</td>
<td>124</td>
<td>123</td>
</tr>
<tr>
<td>Acetyltransferase</td>
<td>RF0024, RF0025</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Beta-Lactamase</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Class A beta-lactamase</td>
<td>RF0053</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Class B beta-lactamase</td>
<td>RF0037, RF0054, RF0085, RF0086</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>Class C beta-lactamase</td>
<td>RF0055</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Class D beta-lactamase</td>
<td>RF0056</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unclassified beta-lactamase</td>
<td>RF0038, RF0039</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td>Chloramphenicol Resistance</td>
<td>RF0050, RF0051</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>D-ala Ligase</td>
<td>RF0144, RF0060</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Gene Modulating Resistance</td>
<td>RF0153, RF0154, RF0155, RF0156, RF0159, RF0160, RF0157, RF0158&lt;br/&gt;RF0095</td>
<td>12</td>
<td>106</td>
</tr>
<tr>
<td>Glycopeptide Resistance</td>
<td>RF0141, RF0065, RF0102, RF0103, RF0140, RF0109, RF0104&lt;br/&gt;RF0095</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Methyltransferase</td>
<td>RF0048, RF0067</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Other Transporter</td>
<td>RF0077, RF0090</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>Other Efflux</td>
<td>RF0142, RF0169, RF0023</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>Quinolone Resistance</td>
<td>RF0074, RF0113</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>RND Antibiotic Efflux</td>
<td>RF0017, RF0115, RF0164</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>rRNA Methyltransferase</td>
<td>RF0171</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Stress Response</td>
<td>RF0063, RF0165, RF0139, RF0148, RF0146</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>Target Redundancy/Overexpression</td>
<td>RF0134, RF0127</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Tetracycline MFS Efflux</td>
<td>RF0135, RF0133</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Tetracycline Ribosomal Protection</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total antibiotic resistance genes</td>
<td><strong>576</strong></td>
<td><strong>693</strong></td>
<td><strong>693</strong></td>
</tr>
</tbody>
</table>
**Table S2.** Tabular output of the antiSMASH analysis of the *B. toyonensis* VU-DES-13 assembly, giving the nature, the position and the database hits with known BGC, and the presence in the conspecific strains BCT-7112^T^ and BAG6O-1. (MIBiG: minimum information on biosynthetic gene clusters database; BGC: biosynthetic gene cluster)

<table>
<thead>
<tr>
<th>Type</th>
<th>Contig</th>
<th>From</th>
<th>To</th>
<th>kbp</th>
<th>Most similar known cluster</th>
<th>MIBiG BGC-ID</th>
<th>Detected by BAGEL3</th>
<th>present in BCT-7112^T^</th>
<th>present in BAG6O-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terpene</td>
<td>7</td>
<td>78 686</td>
<td>100 548</td>
<td>21.9</td>
<td>Molybdenum cofactor biosynthetic gene cluster (11% of genes show similarity)</td>
<td>BGC0000916 c1</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Lasso peptide</td>
<td>7</td>
<td>208 189</td>
<td>232 096</td>
<td>23.9</td>
<td>Bacillibactin biosynthetic gene cluster (38% of genes show similarity)</td>
<td>n/a</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Nrps</td>
<td>9</td>
<td>64 275</td>
<td>114 000</td>
<td>49.7</td>
<td>Bacillibactin biosynthetic gene cluster (38% of genes show similarity)</td>
<td>BGC0000309 c1</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>10</td>
<td>67 017</td>
<td>77 298</td>
<td>10.3</td>
<td>Petrobactin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>n/a</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Nrps</td>
<td>10</td>
<td>96 053</td>
<td>143 087</td>
<td>47.0</td>
<td>Petrobactin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>n/a</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Siderophore</td>
<td>13</td>
<td>69 587</td>
<td>83 300</td>
<td>13.7</td>
<td>Petrobactin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>BGC0000942 c1</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
<td>1</td>
<td>38 858</td>
<td>38.9</td>
<td>n/a</td>
<td>n/a</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>17</td>
<td>77 886</td>
<td>91 751</td>
<td>13.9</td>
<td>n/a</td>
<td>n/a</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Bacteriocin</td>
<td>22</td>
<td>7 321</td>
<td>17 658</td>
<td>10.3</td>
<td>Thuricin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>n/a</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Lantipeptide</td>
<td>25</td>
<td>10 036</td>
<td>33 123</td>
<td>23.1</td>
<td>Thuricin biosynthetic gene cluster (83% of genes show similarity)</td>
<td>BGC0000626 c1</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
Figure S1. Consensus bootstrap (N=1,000) NJ tree based on the feature frequency profiles of 11 bp among representative genomes from the respective *B. cereus* clades (Zwick et al. 2012; Liu et al. 2015a). The tip labels contain the percent dDDH as calculated on the whole genome alignment of the respective accessions with *B. toyonensis* VU-DES13. *B. anthracis* 3154, NZ_ANFF01000000; *B. mycoides* VD021, NZ_AHES01000000; *B. cereus* 172560W, NZ_CM000717.1; *B. thuringiensis* BAG1O-3, NZ_AHCP01000000; *B. pseudomycoides* Rock 1-4, NZ_CM000743.1; *B. gaemokensis* JCM15801, JOTM00000000; *B. manliponensis* JCM15802, JOTN00000000; *B. cytotoxicus* NVH 391-98, NC_009674.1; *B. toyonensis* BCT-7112T, NC_022781.1; *B. toyonensis* BAG6O-1, GCF_000293525.1; BCG 10, NC_003909.8; BCG 11, NZ_CM000740.1; BCG 12, NC_011658.1; BCG 13, NZ_AHDA01000000; BCG 14, NZ_CM000721.1; BCG 15, NZ_AHFC01000000; BCG 16, NZ_AHDJ01000000; BCG 17, NC_011772.1; BCG 18, NZ_AHFW01000000; BCG 19, NZ_AHCZ01000000; BCG 20, NC_017200.1; BCG 21, NZ_AHEX01000000; BCG 22, NZ_AHDJ01000000; *B. bingmayongensis* FJAT-13831, AKCS01000000; BCG 24, NZ_AHDP01000000; BCG 25, ARSX01000000.; BCG 26, NZ_CM000720.1; BCG 27, NZ_CM000733.1; BCG 28, NZ_CM000718.1; BCG 29, NZ_AHEA01000000; BCG 30, NZ_CM000739.1; *B. subtilis* subsp. *subtilis* ATCC 6051, NC_000964.3
Chapter 6 – Discussion

Microorganisms constitute an integral part of the biology of animals. They can provide nutrients, modulate the immune system and influence the development of their host (McFall-Ngai et al. 2013). By participating in these and other essential functions, interacting microbes can influence the fitness of animals by affecting their reproductive success or their lifespan. For example, the human microbiota is involved in the maintenance of adequate immune function (Lee and Mazmanian 2010), to the point where dysbiosis, an alteration of the healthy microbiota, may lead to the onset or worsening of disease (Levy et al. 2017). Furthermore, the microbiota can co-diversify with its host, with differences in microbial community composition reflecting both phylogenetic and dietary patterns (Davenport et al. 2017). These observations indicate that the study of interacting microbes can provide important insights into the ecology and evolution of their hosts.

Soil-living animals offer several opportunities to study interactions with microbes. The soil environment hosts a high diversity of microorganisms and invertebrates, such as arthropods and earthworms, and interactions between these organisms can evolve into specific symbioses. For example, some animals, such as wood-eating termites and herbivorous ants, depend on complex communities of gut microbes to digest their food and obtain essential nutrients (Brune and Dietrich 2015; Hu et al. 2018), while others rely on symbiotic microorganisms for protection against bacterial and fungal antagonists (Flórez et al. 2017). Symbioses between microbes and invertebrate hosts can serve as models for more complex host-associated microbiomes, and they can become targets for biocontrol strategies in the agricultural sector (Zindel, Gottlieb and Aebi 2011). Nonetheless, relatively few studies have focused on investigating the symbioses between microbes and non-human animals, indicating a potential gap in microbiome research (Stulberg et al. 2016).

*Folsomia candida* (Hexapoda: Collembola) is a springtail that lives in the soil, where it feeds on soil particles, microorganisms and plant material. Because of its crucial role in the soil ecosystem (Rusek 1998) and its sensitivity to soil pollution, this animal has been used for many years as a model organism in ecotoxicology and as a bioindicator of soil quality and health (Fountain and Hopkin 2005; Nogueira Cardoso and Lopes Alves 2012). Traditionally, toxicity tests have been applied to investigate the effects of different stressors on *F. candida*, while more recent
technological advancements, such as transcriptomics, have allowed us to study the molecular mechanisms underlying stress response in this animal (Nota et al. 2010). Recently, the genome and transcriptome of *F. candida* have been assembled, providing important clues on the evolutionary history of this springtail species and of collemboians (Faddeeva et al. 2015; Faddeeva-Vakhrusheva et al. 2017).

In this thesis, we have applied different methods to study another aspect of the biology *F. candida*, namely its microbiology. The microbial community of *F. candida* was studied in the past with culturing and 16S-based sequencing, revealing a rather simple community composed of a few bacterial and fungal groups (Thimm et al. 1998; Czarnetzki and Tebbe 2004a). The springtail is also colonized with the endosymbiont *Wolbachia pipientis*, present in high abundance in its brain and ovaries. Based on these early studies, the gut of *F. candida* was described as a selective, although changeable, microenvironment for microbes and a hotspot for lateral gene transfer (Hoffmann et al. 1998; Thimm et al. 1998). Here, we have further investigated the microbiota associated with this springtail, focusing on its bacterial component and its related functions. First, we used high-throughput sequencing and bioinformatic techniques to study the composition of the microbiota of *F. candida* and its characteristic functions (Chapters 2 and 3). Then, we applied a more functional approach based on growth inhibition assays to evaluate the antimicrobial-producing potential of bacteria isolated from *F. candida* and from other springtail species (Chapter 4). Finally, we studied one bacterial isolate obtained from *F. candida* in detail, focusing on its antibiotic resistance and inhibitory activity (Chapter 5).

Our results show that the bacterial community of *F. candida* is characterized by much higher diversity than previously documented (Thimm et al. 1998; Czarnetzki and Tebbe 2004a). This increased resolution in taxonomic profiling is explained by the tremendous advancements in the methods used to quantify microbiomes compared to earlier studies. Bacterial diversity in *F. candida* is comparable to that observed in other soil invertebrates, with a few dominant taxa which are conserved across populations and many rare groups. We also found dramatic differences in the relative proportion of *Wolbachia pipientis* between different populations of the springtail. Dominant species may mask the presence of rare bacterial groups. To test this, we used a method to suppress the amplification of *Wolbachia*, and we found that the high abundance of this endosymbiont in one population did not affect the identification of other bacteria. In the microbiome of the springtail, we found numerous genes encoding for carbohydrate-active enzymes, some of which showed high similarity to *F. candida*’s foreign genes, which were likely acquired through horizontal gene transfer (HGT). We also discovered pathways related to secondary metabolite biosynthesis and antibiotic resistance, functions that may reflect selective pressures linked to the soil environment in which springtails have evolved. One of
the major findings presented in this work was the *in vivo* inhibition of different microbial pathogens from bacteria isolated from springtails. This observation, combined with the discovery of biosynthetic genes by sequencing, indicates the possibility of antimicrobial production in the springtails' microbiota. In the following sections, these findings are discussed in the context of their ecological relevance and their potential applications in the biotechnology field.

6.1. Factors determining microbial community composition in soil invertebrates

The composition of microbial communities associated with terrestrial animals is fundamentally defined by anaerobic conditions (Esposti and Romero 2017) and influenced by a number of other environmental factors, such as lifestyle and diet. The extent to which these factors affect the structure of gut microbial communities is not the same across all species (Colman, Toolson and Takacs-Vesbach 2012) and depends, at least in part, on the ecology of the host. For example, the digestive tract of soil animals can be temporarily colonized by opportunistic microbes originating from the environment and more stably by microbial symbionts involved in specific dietary adaptations (Kelley and Dobler 2011; Ceja-Navarro *et al.* 2014; Brune and Dietrich 2015; Bouchon, Zimmer and Dittmer 2016). The degree to which the animal host relies on its symbiotic microbes to perform nutrition-related or other functions can determine the presence of key microbial species, forming the core microbiome of that animal group.

Hexapods with a very specialized diet, such as sap- and blood-feeding insects, are generally characterized by less diverse and more defined gut microbial communities (Jing *et al.* 2014) and by the presence of specific nutritional endosymbionts (Engel and Moran 2013). However, other animals with a similarly specialized diet can host highly diverse communities of symbiotic microbes. Termites, for instance, contain one of the most complex microbiota of any animal group (Brune and Dietrich 2015), involving not only bacteria but also archaea and protists. This may partly be explained by the characteristics of the environment that they inhabit. These and other soil-living invertebrates, such as beetles, isopods and earthworms, live in a microbe-dominated environment, and may ingest high loads of microbes along with their food (Yun *et al.* 2014).

*F. candida*’s diet is not very specialized, and it is based on a micro-detritivorous feeding regimen consisting of a variety of food items including fungal hyphae, soil bacteria and decaying plant matter. Without large diverticulae or glands, the gut structure of springtails is very simple compared to the complex and highly compartmentalized digestive tracts observed in other more recently evolved hexapods (Figure 1). Still, like all other hexapods, the gut of *F. candida* is structured in three sections: a foregut and a hindgut, both lined by a cuticle layer, and a midgut,
Chapter 6

representing the primary site of digestion and absorption, secreting a peritrophic matrix. At every molt (every 3-4 days) the entire gut epithelium including the peritrophic matrix are shed. The frequency of the molts suggests that the gut of *F. candida* is a highly changeable environment for microbial colonization, as it was previously described, yet maintaining a certain level of selection (Thimm et al. 1998). The simple structure of the gut, combined with the small size of the animal, suggests that the conditions in this environment are not fully anaerobic. Indeed, multiple bacteria capable of growing in aerobic conditions were isolated from the springtail (Hoffmann et al. 1998; Thimm et al. 1998; Czarnetzki and Tebbe 2004a).

![Diagram of digestive tract](image)

**Figure 1.** Digestive tract of the springtail *Sinella coeca* (left) and generalized gut structure of insects (right). Figures adapted from Hopkin (1997) and from Engel and Moran (2013).

In this thesis, we have compared the bacterial community composition of springtails originating from two *F. candida* populations (one lab culture and one natural population), that were kept in similar lab environmental conditions for a few generations. The same bacterial groups dominated the two *F. candida* populations, although their relative abundances were very different. One of the populations was kept in controlled culturing conditions for several years: this might have affected its microbiome composition by facilitating the acquisition of bacteria from the lab environment or from humans and the subsequent establishment of high abundances of these strains.

The most abundant bacteria identified in *F. candida* were *Paracoccus* and *Ochrobactrum*, commonly isolated from soil and also from other arthropods (Esposti and Romero 2017), as well as *Pseudomonas, Bacillus* and *Staphylococcus*, species that are more broadly associated with various animal hosts, including soil invertebrates such as isopods and earthworms (Bouchon, Zimmer and Dittmer 2016; Liu et al. 2018). We also found bacterial families that constitute core microbiota members in nematodes, namely the Xanthomonadaceae, Pseudomonadaceae, Bacillaceae, Alcaligenaceae and Enterobacteriaceae (Berg et al. 2016).

Previous work showed that *F. candida*’s gut constitutes a selective habitat for microbes (Thimm et al. 1998). Future experiments could attempt to better characterize the microbiota of this animal and to understand what factors determine
its community composition. This could be achieved by studying more *F. candida* populations or by applying specific interventions to identify the drivers of stability and changes in the springtail’s microbiota.

### 6.2. Carbohydrate-active genes

We found a number of genes predicted to encode for carbohydrate-active enzymes (CAZymes) in *F. candida*’s microbiome. These enzymes are involved in the metabolism and transport of carbohydrates in different digestive systems, from the rumen of goats and cows, to the hindgut of wood-feeding termites and to the fungus gardens of leaf-cutter ants (Aylward *et al.* 2014; Brune and Dietrich 2015; Jose *et al.* 2017; Do *et al.* 2018). The main purpose of this enzyme system is to unlock high energy compounds, and plant-eating animals often rely on the carbohydrate-degrading activity of their symbiotic microbes to extract energy from their food. For example, many animals lack their own cellulases and need bacterial enzymes to degrade cellulose (Tanimura *et al.* 2013). *F. candida* is not specifically a herbivore, yet plant material at different stages of decay constitutes an important component of its diet. The soil environment is rich in plant-derived polymers such as lignin and cellulose. The ability to degrade these products is an important trait for insects feeding on decaying plant and fungal material, possibly representing a preadaptation to the evolution of herbivory (Calderón-Cortés *et al.* 2012).

Symbioses with microbes possessing specific catalytic functions may be very beneficial for *F. candida* by allowing access to an important nutrient source. Recently, the microbiota of another springtail was studied (Bahrndorff *et al.* 2018). *Orchesella cincta* is an epedaphic springtail species whose diet includes fungi and algae. A prediction from the taxonomic composition of the community, based on 16S rRNA gene sequencing of its microbiota, revealed an abundance of genes related to the diet of the animal and to soil processes, such as nitrogen and chitin metabolism (Bahrndorff *et al.* 2018). The authors suggested that *O. cincta* may depend on nutritional symbioses with microorganisms, and that these symbionts, by modulating the nutrients and energy available to the host, may constitute fundamental determinants of its physiology and behavior.

Carbohydrate-degrading genes were also identified in the genomes of *F. candida* and *O. cincta*, and many of them were classified as foreign, meaning that they were incorporated in the genomes of the springtails following an HGT event (Faddeeva-Vakhrusheva *et al.* 2016, 2017). Carbohydrate metabolism was one of the largest functional categories in the set of horizontally transferred genes in *O. cincta*, including genes involved in the degradation of cell-wall components of plants, fungi and bacteria (Faddeeva-Vakhrusheva *et al.* 2016). Similarly, in *F. candida* the set of foreign genes was enriched for carbohydrate-related functions, specifically cell-wall degradation, compared to the native genes (Faddeeva-Vakhrusheva *et al.* 2017).
Chapter 6

Interestingly, some of the CAZymes identified in the microbiome of *F. candida* had high homology to the springtail’s foreign genes (Chapter 3), suggesting that carbohydrate-related functions may have been transferred from the gut microbiota to the host. Natural selection can determine whether genes acquired by HGT are maintained and fixed in a recipient genome (Pauchet and Heckel 2013). It is not clear why carbohydrate-related genes would be maintained in *F. candida*’s genome, given that the encoded functions also reside in the gut microbiome. One possibility is that, after HGT, the host acquires transcriptional control of functions that improve fitness, resulting in a selective pressure favoring the maintenance of such genes in its genome.

The presence of carbohydrate-related activities in the gut microbiome is ecologically relevant for a soil-living animal, and also interesting for the bio-based industry. Enzymes capable of breaking down plant biomass can be used in the biofuel industry to extract energy from different sources, for example during the fermentation of sugars to produce ethanol (Raghavendra, Nayaka and Gupta 2016).

### 6.3. Antimicrobial functions in the microbiota of *F. candida*

In this thesis, we used both sequence- and function-based approaches to study potential antimicrobial activity in the microbiota of *F. candida*. Both inhibitory assays and metagenome mining indicated a potential for antimicrobial production in the microbial community of this springtail. In Chapter 3 we identified several pathways involved in the biosynthesis of secondary metabolites, including antibiotics and bacteriocins. In Chapter 4 we observed high inhibitory activity of bacteria isolated from springtails against microbial pathogens, which may indicate antimicrobial production. Although the experiments conducted in this thesis are not conclusive in this regard, we can speculate about the function of antimicrobial production in *F. candida*, and the possible ecological relevance for the host.

Interactions between microorganisms in the environment can lead to competition, which is often mediated by the production of secondary metabolites. This class of compounds includes antimicrobials and other inhibitory substances that have both communication and defensive purposes (Townsley and Shank 2017). Microbial warfare is common in environments such as the rhizosphere and the soil in general, characterized by high densities of microbes and complex interactions between them (Raaijmakers *et al.* 2009). Animals’ guts also constitute a habitat for many beneficial and pathogenic microorganisms, creating ideal conditions for antagonistic mechanisms such as colonization resistance, that may be mediated by antimicrobial production (Kim, Covington and Pamer 2017). These mechanisms (competition and antimicrobial production) can also drive the selection of resistant microbial strains. For example, we observed exceptionally high resistance to penicillin in a *Bacillus*
strain isolated from \textit{F. candida}, which may result from the presence of antibiotic biosynthesis genes both in the microbiome and in the genome of the animal.

In some cases, the production of antimicrobials by microbial symbionts can be beneficial for the animal host. For example, the larvae of the beetle \textit{Paederus} are protected by predatory spiders thanks to a polyketide produced by a bacterial symbiont (Piel \textit{et al.} 2005). Similarly, the eggs of herbivorous beetles are protected from detrimental microbes by a cocktail of antibiotics produced by symbiotic \textit{Burkholderia}, therefore increasing the fitness of the host (Flórez \textit{et al.} 2017).

\textit{F. candida} and other species of springtails are resistant to entomopathogenic fungi that are deadly for many other soil arthropods (Broza, Pereira and Stimac 2001; Dromph 2001). We have shown that bacteria isolated from springtails are able to inhibit one of these fungi, \textit{Beauveria bassiana} (Chapter 4). Further experiments could test whether bacteria from springtails can also inhibit the fungus \textit{in vivo}, therefore conferring resistance to the host. Potentially, such strains may also be applied as biological control against entomopathogenic fungi in situations where they can cause economic damage, for instance in entomology farms, where insects are mass-reared for food or feed and infections from pathogens and parasites are one of the biggest concerns (Eilenberg \textit{et al.} 2015).

6.4. HGT as a source of evolutionary innovation

Aside from performing functions that benefit their hosts, microorganisms can also provide them with the ability to carry out those functions directly. This can be achieved through horizontal gene transfer (HGT), the transfer of genes between organisms belonging to different species, and even to different domains. HGT is well described among microorganisms, that use this mechanism to adapt to rapidly changing environmental conditions (Polz, Alm and Hanage 2014). HGT can also occur between prokaryotes and eukaryotes (Pontiroli \textit{et al.} 2009; Crisp \textit{et al.} 2015). Some well-known examples include the biosynthesis of carotenoids in aphids, a capacity acquired with the transfer of genes from an endosymbiont, and by the presence of phenylpropanoid metabolism in plants, which drove important adaptations to life on land (Emiliani \textit{et al.} 2009; Moran and Jarvik 2010). HGT has been described as a mechanism that can contribute to the biochemical diversification of eukaryotic species, and therefore as an important factor in their long-term evolution (Boschetti \textit{et al.} 2012).

The gut of springtails has been described as a hotspot for HGT (Hoffmann \textit{et al.} 1998). We recently confirmed this by detecting over 800 genes of foreign origin in the genome of \textit{F. candida}, most of them originating from bacteria and fungi (Faddeeva-Vakhrusheva \textit{et al.} 2017). It was suggested that gene transfer from symbionts or from environmental microbes may have contributed to the adaptation of the springtail to life in the soil. The fact that most of the foreign genes in \textit{F. candida}
are involved in carbohydrate metabolism seems to support this hypothesis. The soil is an environment rich in plant degradation products, that, as mentioned above, are part of the diet of the springtail. By contributing to the degradation of plant polymers, carbohydrate-active enzymes could allow *F. candida* to access an important nutrient source in the soil. HGT of cellulose-degrading enzymes had been previously observed in other plant-feeding insects (Pauchet and Heckel 2013) confirming that it may be an important mechanism providing soil invertebrates with advantageous traits (Eyun *et al.* 2014). In the gut microbiota of *F. candida*, we identified several genes predicted to encode carbohydrate-active enzymes, and a few of them showed high homology to some of *F. candida*'s foreign genes (Chapter 3). This observation supports the possibility that members of the gut microbial community of the springtail may constitute a source of genes and new functions through HGT.

Another important class of genes that was acquired by *F. candida* through HGT are antibiotic-related ones. An entire pathway of genes involved in the biosynthesis of a β-lactam compound was found in the genome of *F. candida* (Roelofs *et al.* 2013; Suring *et al.* 2016; Faddeeva-Vakhruševa *et al.* 2017). Interestingly, antibiotic-related genes have been found not only in the genome of the springtail *F. candida*, but also in six other collembolan species (Suring *et al.* 2017). In *F. candida*, these genes are markedly upregulated after exposure to different stressors (Nota *et al.* 2009; Chen *et al.* 2015), and they are transcriptionally active (Suring *et al.* 2016). Suring also detected a beta-lactam compound using an ELISA assay (Suring *et al.* 2016). Although the exact function of this compound is still unknown, a possibility is that it may be involved in regulating the microbial community of *F. candida*, or that it may have a defensive role beneficial for the springtail.

Whether the beta-lactam biosynthesis genes in *F. candida* are of bacterial or fungal origin is currently unknown (Roelofs *et al.* 2013; Suring *et al.* 2016). Both bacteria and fungi have been found in the gut of this animal (Thimm *et al.* 1998) and may constitute a source of genes through HGT. Fungi known for the production of medically important antibiotics, such as *Penicillium* species, have been previously detected in springtails (Anslan, Bahram and Tedersoo 2016). Entomopathogenic fungi, to which springtails are resistant, have also been detected in these animals (Greif and Currah 2007). Interactions between these and other microorganisms in the springtails’ gut may result in a microbial warfare characterized by the production of inhibitory compounds and the evolution of strategies to counteract them. It is therefore possible that the gut microbiota of *F. candida* constitutes a reservoir of genes encoding compounds with antimicrobial and inhibitory activities, that can be subject to HGT and become integrated in the genome of the host.
6.5. Applications: the bio-based industry

Microbes are prolific sources of compounds with many different functions, and for decades their biosynthetic potential has been harvested and applied in different sectors. Many common drugs, such as antibiotics and antifungals, were originally isolated from microorganisms, and other compounds of medical importance have been chemically engineered based on structures obtained from natural sources (Kim, Moore and Yoon 2015). Microbial polyketides and non-ribosomal peptides are often the targets of pharmaceutical interest because of their antimicrobial, anticancer and immunosuppressant properties. These compounds are biosynthesized in an assembly-line fashion, with building blocks being successively assembled into more complex structures. This type of biosynthetic system can be easily modified through gene replacements, domain substitutions and module exchanges, and these strategies are frequently applied to obtain novel antibiotics (Nguyen et al. 2006).

Some animals are associated with microorganisms that affect their host’s fitness through the production of antimicrobial compounds. For example, symbiotic *Burkholderia* bacteria produce a cocktail of antibiotics that protect the eggs of their host, a herbivorous beetle, against detrimental microbes (Flórez et al. 2017). Other insect species use antibiotic-producing symbionts to control parasites and protect their fungus gardens (Haeder et al. 2009). In this thesis, we showed that springtails’ gut bacteria are able to inhibit the growth of various microbial pathogens (Chapters 4 and 5) and we suggested that this antagonistic activity may be the result of antimicrobial production. We also identified multiple secondary metabolite pathways in the microbiome of *F. candida* (Chapter 3), which seems to support this hypothesis. A few of these pathways were represented by non-ribosomal peptide synthases (NRPSs) showing high similarity with previously described antibiotic and antifungal compounds, such as orfamide and caryoyncin, two antifungal compounds isolated from rhizosphere-derived bacteria and from an insect’s symbiont (Ma et al. 2016; Flórez et al. 2017). We identified additional pathways involved in the synthesis of medically relevant compounds, such as the immunosuppressant lymphostin and the compounds microsclerodermin and charthreusin, both active against cancer cell lines (Aotani, Nagata and Yoshida 1997; Xu et al. 2005; Guzmán et al. 2015). These microbial products could be harvested and investigated for potential applications in the pharmaceutical sector. Biosynthetic pathways containing novel combinations of genes are the most interesting to explore, as this could lead to the discovery of novel bioactive molecules (Nguyen et al. 2006; Suring et al. 2017).

The industrial sector also benefits from the biosynthetic potential of many fungal and bacterial species. Microbially derived enzymes can be applied in fermentation and catalytic processes (Raghavendra, Nayaka and Gupta 2016). Carbohydrate-degrading enzymes are particularly interesting, as they can be used to convert plant
biomass, such as lignocellulose, to simple sugars (Florencio et al. 2016). These can in turn be fermented to produce ethanol, which can be used as biofuel, and other building blocks for the chemical industry (Chen, van Straalen and Roelofs 2016). Many efforts are spent on searching novel enzymes and on improving their characteristics and effectiveness in order to optimize the conversion of plant biomass into products of higher value (Yang et al. 2011). One of the critical steps in this process is the pre-treatment of plant biomass to remove the lignin and recover the cellulose, which is then accessible to hydrolytic enzymes (Anwar, Gulfraz and Irshad 2014). Microorganisms can also be employed directly in industrial processes. *Saccharomyces cerevisiae* and *Zymomonas mobilis*, for instance, are very effective ethanol producers, whereas *Caldicellulosiruptor bescii* has high rates of cellulose degradation (Liao et al. 2016). These organisms can be used in combination or they can be genetically engineered to achieve plant biomass degradation and sugar fermentation (Figure 2).

**Figure 2.** Components of plant biomass, such as lignocellulose, can be converted to fuels in two subsequent steps (hydrolysis followed by fermentation) or through bioprocessing, which combines the two processes in one step. Bioprocessing can be achieved by genetically engineering microorganisms. Figure adapted from Liao et al. (2016).

Various environments have been explored for the purpose of isolating bioactive molecules. The soil ecosystem is rich in microbes, and it seems to be the biggest reservoir of genes and functions with important practical applications (Zhuang et al. 2013). For example, many medically important strains of microbes, such as *Penicillium, Aspergillus* and *Streptomyces*, are common in soils (Antoraz et al.
2015; Garrigues et al. 2017). We suggest that soil invertebrates also constitute promising environments to explore.

The guts of various wood- and plant-feeding insects, such as termites, ants and weevils, have been widely investigated as sources of carbohydrate-degrading functions (Suen et al. 2010; Franco Cairo et al. 2011; Mohamed et al. 2018). The microbial communities in these microenvironments produce a mixture of enzymes that break down lignocellulose and other polysaccharides, allowing the host to extract energy from plant food (Lima et al. 2014; Ali et al. 2017). These enzymes have great potential for application in the biofuel industry, as discussed above, and some have already been isolated, characterized and genetically engineered to improve their catalytic efficiency (Kumar, Dangi and Shukla 2018). In this context, knowledge of the biology of the host, for example its dietary regimen and the pH and temperature in the gut, can provide valuable information on the specific functions of carbohydrate-active enzymes and on the optimal conditions needed for their activity (Franco Cairo et al. 2011; Zhang et al. 2014). In Chapter 3, we showed that the microbiome of *F. candida* contains several genes predicted to encode for carbohydrate-degrading enzymes, probably linked with the detritivorous lifestyle of the animal. Additional work is needed to isolate these enzymes and to assess their potential applicability in the bio-based industry.

### 6.6. Conclusion and future work

This thesis constitutes an in-depth study of the microbial community associated with the springtail *F. candida*, achieved by using a combination of next-generation sequencing techniques, bioinformatics approaches and function-based studies. We have described the composition of *F. candida*’s microbiota and its functions, focusing on those with ecological relevance for a soil-living animal and with an interest for the bio-based industry.

We discovered many functions related to carbohydrate metabolism, which may confer advantageous functional traits to an animal living in the soil and that are likely subject to HGT in the gut of microarthropods. The presence of antibiotic and secondary metabolite biosynthesis genes in the microbiome and the inhibitory activity of bacteria isolated from the springtail indicate a potential for antimicrobial production. Furthermore, we found antibiotic resistance genes in the microbiome and we observed resistance in bacterial isolates. Antibiotic resistant phenotypes are consistent with the presence of antibiotic-related genes in the genome of the host, which were shown by previous studies to be active and were suggested to have a function in regulating the microbial community.

The work presented in this thesis shines a light on the microbes within *F. candida*, adding a new layer to the information available on this model organism. It also constitutes the basis for future research, whether aimed at a better understanding of
host-microbe interactions, or at the discovery of novel bioactive molecules through microbiome-mining.

Future studies may attempt to move beyond the homology-based analysis and the \textit{ex vivo} assays presented in this thesis, in order to explain the function of antimicrobial production in the gut of \textit{F. candida}. A possible approach might consist of experimentally removing the antimicrobial producers from the community, for example through antibiotic treatment, and observing the effect on the phenotype. A possible drawback of this method is the fact that antibiotics can also affect non-target microorganisms in the microbial community. This can create a dysbiosis that can in turn lead to other systemic effects, constituting additional confounding factors. Despite the potential biases introduced, this kind of experiments could provide clues on the importance of symbiotic microbes in \textit{F. candida}, possibly indicating whether they play defensive roles against soil pathogenic organisms, which may confirm a contribution of microbes to the adaptation of the springtail to life in the soil.

Based on the results of the metagenome mining and the inhibitory assays presented in this thesis, it also seems worth to invest additional efforts in the identification of specific antimicrobial compounds. Screening of culturable bacteria against microbial pathogens is a quick method to select potential antimicrobial producers, and sequencing of the promising strains can confirm the presence of biosynthetic pathways. A combination of methods like that applied in Chapter 5 could be used to identify good targets for drug discovery.


Alcaraz L, Moreno-Hagelsieb G, Eguiarte LE *et al.* Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* 2010;**11**:1–17.


Altayar M, Sutherland AD. *Bacillus cereus* is common in the environment but emetic toxin producing isolates are rare. *J Appl Microbiol* 2006;**100**:7–14.


Bibliography


Ceuppens S, Boon N, Uyttendaele M. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. *FEMS Microbiol Ecol* 2013;84:433–50.
Bibliography


Bibliography


Feldhaar H. Bacterial symbionts as mediators of ecologically important traits of insect hosts. Ecol Entomol 2011;36:533–43.


He X, McLean JS, Guo L et al. The social structure of microbial community involved in colonization resistance. *ISME J* 2014;8:564–74.

Bibliography


Kelley ST, Dobler S. Comparative analysis of microbial diversity in *Longitarsus* flea beetles (Coleoptera: Chrysomelidae). *Genetica* 2011;139:541–50.


Bibliography


Lee YK, Mazmanian SK. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* 2010;**330**:1768–73.


Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;**0**:1–3.


Logan N a, Berkeley RC. Identification of *Bacillus* strains using the API system. *J Gen Microbiol* 1984;130:1871–82.


Bibliography


Bibliography


Raaijmakers JM, de Bruijn I, Nybroe O et al. Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol Rev* 2010;34:1037–62.


Rambaut A. FigTree. 2014.


Bibliography


Thimm T, Hoffmann A, Borkott H *et al.* The gut of the soil microarthropod *Folsomia candida* (Collembola) is a frequently changeable but selective habitat and a vector for microorganisms. *Appl Environ Microbiol* 1998;**64**:2660–9.


Watterson JR, Nagy LA, Updegraaf DM. Penicillin Resistance in Soil Bacteria Is an Index of Soil Metal Content near a Porphyry Copper Deposit and near a Concealed Massive Sulfide Deposit., 1984.


Summary

Diversity and functions of bacteria associated with springtails

Springtails are tiny invertebrates that inhabit moist environments all over the world. In soil ecosystems, they play important ecological roles by contributing to decomposition processes through the fragmentation of organic matter and the control of soil microbial communities. Some springtail species, such as *Folsomia candida*, are used as model organisms in ecotoxicological research and as bioindicators of soil quality. Decades of research have provided a wealth of information on the anatomy, physiology, ecology and behavior of these model species, and more recently genomic sequencing studies have offered new insights into their evolution and phylogenetic relationships to other arthropod groups. Like all other animals, springtails constitute a habitat for microorganisms, which may play important roles for the functioning of their hosts. Therefore, associated microbes constitute another aspect of the biology of these animals, which deserves attention.

In this thesis, we studied the microbial community of the springtail *F. candida*. We explored the composition of this community and its potential functions by using a combination of sequence-based and more targeted functional approaches, as well as bioinformatic tools for data analysis. We specifically focused on the identification and description of potential antimicrobial functions, which are ecologically relevant for *F. candida* and may have a variety of applications.

In Chapter 2 we studied the composition of the bacterial communities in two populations of *F. candida*. By using amplicon-based high-throughput sequencing we identified over 800 bacterial groups. Pronounced differences were observed in the relative abundances of bacteria between the two populations, especially in the case of *Wolbachia*, the endosymbiont that causes parthenogenesis in *F. candida*. When applying a method to target bacteria other than *Wolbachia*, we were still able to successfully identify the most dominant bacterial groups, which were shared between the two populations of springtails. While in line with the results of previous studies on the microbial community of *F. candida*, the findings presented in this chapter deepen our understanding of bacterial diversity in this animal.

In Chapter 3 we studied the metagenome of *F. candida*. By using shotgun sequencing and a variety of bioinformatics approaches, we obtained a detailed picture of the pooled genomic content of the gut microbiota of the springtail. The analysis revealed the presence of multiple genes involved in antimicrobial activity and carbohydrate metabolism. These functions may contribute to the digestion of plant material and
to defense against pathogens in the gut of *F. candida*, possibly indicating a supportive role of the microbes for the fitness of their host. In the metagenome, we also identified several homologs to horizontally transferred genes in *F. candida*, suggesting that specific functions may have been transferred from the gut microbiota to the springtail.

In Chapter 4 we investigated the potential for antimicrobial activity of springtails’ bacteria. We isolated and cultivated gut bacteria from four springtail species, and we tested their capacity to inhibit the growth of a variety of microbial pathogens. Of 46 unique isolates identified, 35 showed inhibitory activity, indicating a high potential for antimicrobial production. The isolates were active against insect and plant pathogens, suggesting that they may play a role in host defense or in the control of environmental microbes. The results of this chapter point to springtails’ bacteria as possible sources of biocontrol methods and as interesting targets for drug discovery.

In Chapter 5 we described one *Bacillus* strain isolated from the gut of *F. candida*. We studied its metabolic profile, antibiotic resistance and inhibitory activity against pathogenic microorganisms. The isolate was identified as a *Bacillus toyonensis*, and displayed extremely high resistance to penicillin. Both the living bacteria and their extract inhibited the growth of a variety of pathogenic microorganisms. In line with these observations, genome analysis of the strain revealed an enrichment of resistance genes for β-lactam antibiotics and the presence of secondary metabolite clusters.

The work presented in this thesis expands the knowledge of springtails’ microbiota and constitutes a foundation for future research on this topic. Building on previous findings, which mainly resulted from culturing and 16S amplicon sequencing, we applied a combination of methods to study the bacterial community of *F. candida*. This community appears to be highly diverse and it contains a number of potential functions with possible useful applications as well as high ecological relevance. The finding of many genes associated with carbohydrate metabolism suggests a contribution of gut microbes to digestive processes. The similarity of these genes to foreign genes in the genome of *F. candida* suggests that bacteria may not only break down the plant material in the gut of the springtail, but also transfer genes involved in this process to the host. This would constitute an important adaptation to life in the soil, an environment rich in plant degradation products. A high potential for antimicrobial production was revealed both by metagenome mining and by functional assays. Antimicrobial substances in *F. candida*’s gut may be involved in a variety of processes, namely pathogen defense, microbial interactions, modulation of microbial community composition and communication between microorganisms and their host. These and other processes could be explored by future research aimed at elucidating the function of antimicrobials in springtails and in their associated microbes.
Samenvatting

Diversiteit en functies van bacteriën bij springstaarten

Springstaarten zijn kleine bodemdieren die in vochtige omgevingen over de hele wereld te vinden zijn. In bodemecosystemen zijn ze van enorm ecologisch belang door hun bijdrage aan decompositieprocessen die tot stand komt door fragmentatie van organisch materiaal en controle over microbiële bodemgemeenschappen. Sommige springstaarten, zoals *Folsomia candida*, worden als modelorganismen en indicatorsoorten in ecotoxicologisch onderzoek gebruikt. Na tientallen jaren onderzoek, is veel informatie aanwezig over de anatomie, fysiologie, ecologie en het gedrag van deze dieren, en recente genomische studies hebben nieuwe inzichten in hun evolutie en fylogenetische relaties met andere groepen geleedpotigen opgeleverd. Net als alle dieren, zijn springstaarten een habitat voor microorganismen, die belangrijke functies voor hun gastheer kunnen uitoefenen. Daarom verdient ook de microbiële gemeenschap, een nog slecht onderzocht aspect van de biologie van deze dieren, aandacht.

In dit proefschrift hebben we de microbiële gemeenschap van de springstaart *F. candida* bestudeerd. Door een combinatie van genomische en bioinformatica-analyses, en functionele testen, hebben we de samenstelling en de potentiële functies van deze gemeenschap onderzocht. We waren specifiek gericht op het identificeren en beschrijven van potentiële antimicrobiële functies die ecologisch relevant zijn voor *F. candida* en verschillende toepassingen zouden kunnen hebben.

In Hoofdstuk 2 hebben we de samenstelling van de bacteriële gemeenschappen in twee populaties van *F. candida* bestudeerd. Door high-throughput amplicon sequencing identificeerden we meer dan 800 bacteriële groepen. Grote verschillen in de relative abundanties van bacteriën tussen de twee populaties werden waargenomen, vooral voor *Wolbachia*, de endosymbiont die bij *F. candida* parthenogenese veroorzaakt. Door het toepassen van een methode om bacteriën andere dan *Wolbachia* op te sporen, konden we de meest abundante bacteriegroepen die in beide populatie springstaarten aanwezig waren, nog steeds identificeren. Onze bevindingen komen overeen met de resultaten van eerdere studies over de microbiële gemeenschap van *F. candida*, terwijl ze ons begrip van bacteriële diversiteit bij dieren verdiepen.

In Hoofdstuk 3 hebben we het metagenoom van *F. candida* bestudeerd. Door shotgun sequencing en bioinformatische analyses, kregen we een gedetailleerd beeld van de genomische samenstelling van de darmmicrobiota van de springstaart. De analyse onthulde de aanwezigheid van meerdere genen die betrokken zijn bij
antimicrobiële activiteit en koolhydraatmetabolisme. Door bijdragen aan de afbraak van plantenmateriaal en aan de bescherming tegen ziekteverwekkers in de darm van *F. candida*, zouden deze microbiële functies de fitness van de gastheer kunnen ondersteunen. In het metagenoom identificeerden we ook verschillende homologen van horizontaal overgedragen genen in *F. candida*: dit suggereert dat sommige functies van de darmmicrobiota naar het genoom van de springstaart zijn overgedragen.

In Hoofdstuk 4 hebben we de potentiële antimicrobiële activiteit van bacteriën uit springstaarten onderzocht. Darmbacteriën van vier soorten springstaarten zijn geïsoleerd en gecultiveerd, en hun capaciteit om de groei van verschillende microbiële pathogenen te remmen werd getest. Van de 46 unieke isolaten die werden geïdentificeerd, hadden er 35 een remmende activiteit, wat wijst op een hoog potentieel voor antimicrobiële productie. Als ze ook *in vivo* actief zouden zijn, zouden ze een rol kunnen spelen bij de afweer van de gastheer of bij de controle van microben in de omgeving. De resultaten van dit hoofdstuk wijzen erop dat de bacteriën van springstaarten mogelijke biocouremethoden zouden kunnen leveren en interessante doelwitten kunnen zijn voor het ontdekken van geneesmiddelen.

In Hoofdstuk 5 hebben we een *Bacillus*-stam beschreven die geïsoleerd werd uit de darm van *F. candida*. We bestudeerden het metabole profiel, antibioticaresistentie en remmende activiteit tegen pathogene micro-organismen. Het isolaat werd geïdentificeerd als een *Bacillus toyonensis* en vertoonde een extreem hoge resistentie tegen penicilline. Zowel de levende bacteriën als hun extract remden de groei van verschillende pathogene micro-organismen. In lijn met deze waarnemingen, onthulde genoomanalyse van de stam een verrijking van resistentiegenen voor β-lactam-antibiotica en de aanwezigheid van secundaire metabolietclusters.

Het werk in dit proefschrift vergroot de kennis van de microbiota van bodemdieren en vormt een basis voor toekomstig onderzoek. Eerdere studies over dit onderwerp waren voornamelijk gebaseerd op kweken en 16S amplicon-sequencing. Hier hebben we een combinatie van methoden toegepast om de bacteriële gemeenschap van *F. candida* te bestuderen. Deze gemeenschap lijkt zeer divers en bevat potentiële functies die van toepassing nut en van grote ecologische relevantie zouden kunnen zijn. De aanwezigheid van genen geassocieerd met koolhydraatmetabolisme suggereert een bijdrage van darmmicroben aan de spijsvertering van de gastheer. De gelijkenis van deze genen met genen die deel van het genoom van *F. candida* zijn suggereert dat bacteriën niet alleen het plantenmateriaal in de darm van de springstaart kunnen afbreken, maar ook genen die bij dit proces betrokken zijn op hun gastheer kunnen overdragen. Dit zou belangrijk zijn als aanpassing aan het leven in de bodem, een omgeving die rijk is aan afbraakproducten van planten. Verder
hebben we een hoog potentieel voor antimicrobiële productie onthuld, zowel door in het metagenoom als door functionele analyses. Antimicrobiële stoffen in de darm van *F. candida* zouden bij verschillende processen betrokken kunnen zijn, namelijk bescherming tegen pathogenen, microbiële interacties, controle van darmmicrobiële gemeenschapssamenstelling en communicatie tussen micro-organismen en hun gastheer. Het is wenselijk voor toekomstig onderzoek om deze en andere processen te verkennen, om de functie van antimicrobiële stoffen in springstaarten en in hun geassocieerde microben op te helderen.
Acknowledgements

One of the most important things that I learned during my PhD is that teamwork is essential in science. As a testament to that, many people deserve to be acknowledged for the work presented in this thesis. Some have provided scientific guidance and practical support, others their friendship, and they have all helped me along the way.

Firstly, thank you to my promotors Nico and Bram and to my supervisors Dick and Wilfred for giving me the opportunity of doing this PhD. It has been a great learning and growing experience. I am grateful for your support, guidance and patience and for believing in me during these years.

Dick, you first welcomed me in your group as a student and you involved me in different projects. I admire your enthusiasm as a scientist, and the positivity with which you supervise your students. It has been at the same time fun and instructive to work with you, and I am happy that we are setting up future collaborations.

Nico, you are a very special teacher: you can bring science to anyone with a combination of wisdom and entertainment. You also care deeply about your students and you are always available for them. It has been a privilege to assist you with your teaching and a lot of fun to see you in action.

Thank you Bram for your collaboration within BE-Basic. Wilfred, thank you for welcoming me in your group and for helping me become familiar with a new field of research. I remember you with gratitude and respect.

I am extremely grateful to the people I collaborated with, especially Wouter, James, Giang, Joeri and Thierry, whose input was precious for the making of this thesis. Wouter, I really enjoyed working with you and Anna: I am glad that we could find ways to collaborate and to help each other during our PhDs. James, thank you for sharing your knowledge of bioinformatics pipelines. Giang and Joeri, thank you for your important contributions to my thesis. Thierry, your ideas and participation have been fundamental in shaping this work.

Thank you to the technicians of the Animal Ecology and Molecular Cell Physiology departments, especially Janine, Riet, and Martin for their assistance and their expert advice. Thank you to Niels, Wendy and Tjalf at MicroLife Solutions for helping me during the months I spent at their labs, and to Paul for performing the sequencing.

Thank you to my students, Dennis, Elianne and Roy, for their help with my experiments, and for challenging me to become better at supervising others.
During the years I have crossed roads with many other PhD students and postdocs in the department, and I have many good memories of the time we spent together. Cécile, André, Tjalf, Ben, Elaine, Daniel G., Elferra, Valentina, Astra, Daniel E., Wouter, Simon, Anna, Mima, Dré, Monica, Oscar, Jeroen, Milou, Mark, Wendy, Guang, Claudia, Masoud, Estefania, Anouk... Thank you all for making the department such a friendly environment to work in.

I also want to express my appreciation to my teachers at the University of Siena for their knowledge and teachings, which I feel have provided me with a good foundation in the understanding of science.

I am very thankful to the opponents for the time and effort they dedicated to reviewing my thesis and for being present at my defense.

Thank you so much Kim and Mima for being my paranymphs and my dear friends.

Thank you Peter for teaching me Dutch: it is very important in my new job, and it allows me to connect in new ways with my ex-colleagues and friends at the VU.

Ingrid, thank you for coaching me and for offering new insights and different perspectives on my PhD and on my future.

Thank you to my roommates of Villa Zebrazicht and to all the friends who shared fun and adventures with me: you all made my years in Amsterdam unforgettable.

Finally, a very special thank you to my mother Rossella, my father Mauro and my sister Chiara and to Cyril for being a wonderful family and for their love and support during these years.

Valeria
Amersfoort, August 2018
Curriculum vitae

I, Valeria Agamennone, was born on the 19th August 1986 in San Giovanni Rotondo, Italy. From 2006 to 2009 I studied Biological Sciences at the University of Siena. I graduated in December 2009 with a thesis on the bioaccumulation of heavy metals in isopods from a dismissed mine site, supervised by prof. Roberto Bargagli. From 2009 to 2011 I conducted my master study in Biodiversity and Nature Conservation at the University of Siena. During that time, I spent 6 months at the Department of Animal Ecology at the Vrije Universiteit Amsterdam, where I became familiar with molecular techniques, ecotoxicological assays, and the use of microarrays. I graduated in December 2011 with a thesis on comparative transcriptomics of two closely related springtail species, supervised by dr.ir. Dick Roelofs and prof. Francesco Frati. In June 2016 I started my PhD at the Animal Ecology Department, under the supervision of dr.ir. Dick Roelofs, dr. Wilfred Röling, prof. Nico van Straalen and prof. Abraham Brouwer. During my PhD I applied a combination of microbiology, molecular and genomics techniques and bioinformatics to study the microbial community of the springtail *Folsomia candida*. This work resulted in a thesis titled “Diversity and functions of bacteria associated with springtails”. Since November 2016 I work as a Junior Scientist at the Microbiology and Systems Biology department at TNO, a research institute for applied science in The Netherlands.
List of publications


Janssens TKS, de Boer TE, **Agamennone V**, Zaagman N, van Straalen NM, Roelofs D. Draft genome sequence of *Bacillus toyonensis* VU-DES13, isolated from *Folsomia candida* (Collembola: Entomobryidae). *Genome Announcements* 2017; 5(19): e00287-17.


Draft Genome Sequence of *Bacillus toyonensis* VU-DES13, Isolated from *Folsomia candida* (Collembola: Entomobryidae)

Thierry K. S. Janssens, Tjalf E. de Boer, Valeria AgamenNONE, Niels Zaagman, Nico M. van Straalen, Dick Roelofs
Department of Ecological Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; MicroLife Solutions, Amsterdam, The Netherlands

**ABSTRACT** We present here the draft genome of *Bacillus toyonensis* VU-DES13, which was isolated from the midgut of the soil-living springtail *Folsomia candida*. Previous research revealed the presence of gene clusters for the biosynthesis of various secondary metabolites, including β-lactam antibiotics, in the host’s genome. The genome data are discussed in the light of the antimicrobial properties against fungi and oomycetes and a high level of β-lactam resistance of the isolate.

*Bacillus toyonensis* strain VU-DES13 was isolated from the midgut of the soil-dwelling springtail *Folsomia candida*, which displays resistance to and can thrive on entomopathogens (1). The host’s genome contains biosynthetic gene clusters for secondary metabolites, such as for β-lactams (2), which are induced in the gut epithelium upon stress (3). Previous amplicon-sequencing studies (4, 5) revealed the prominence of *B. cereus* in the *F. candida* midgut. An association between the propagation cycle of members of the *B. cereus* clade and the internal environment of animals has been suggested (6, 7).

In this study, we provide the draft genome of the isolate *B. toyonensis* VU-DES13, which exhibited antimicrobial properties against fungi and oomycetes and a high level of β-lactam resistance in a MIC assay (>800 μg/mL penicillin G). We hypothesize that this microorganism represents a key player in colonization resistance to entomopathogens selected upon by the host.

High-molecular-weight gDNA was extracted from an overnight culture with the Macherey-Nagel Nucleospin soil kit. The genomic library was made by enzymatic shearing with the Ion Xpress Plus fragment library kit (Thermo Fisher) and size selection on a 2% agarose E-Gel SizeSelect Gel (Thermo Fisher). The template was prepared with 10 pM of the library on an Ion One Touch 2 system (Thermo Fisher). A 400-bp run was executed on an Ion Torrent PGM sequencer, with the Ion PGM Hi-Q sequencing kit (Thermo Fisher). After removal of adapters, the sequences were assembled with SPAdes (8). The assembly was annotated with the Prokka pipeline (9). Biosynthetic gene clusters for secondary metabolites, plasmid-related elements, and open reading frames (ORFs) related to antibiotic resistance were screened by anti-SMASH version 3.0.5 (10), PlasmidFinder version 1.3 (11), and ResFam (12). Feature frequency profiling (13) and genome BLAST distance phylogeny (14) were performed against representative *B. cereus* clade genomes in order to determine the phylogeny of this isolate, which was previously identified by 16S rRNA sequencing as belonging to this clade.

The draft genome consisted of 5.45 Mb (35%GC), in 40 contigs, with 119X average coverage, an N₅₀ of 254 kb, and an L₅₀ of 9. Moreover, 5,512 ORFs, 53 tRNAs, 6 rRNA clusters, and 10 biosynthetic gene clusters for secondary metabolites were predicted.
Annex

Janssens et al.

We were not able to confirm the presence of any plasmid, although seven contigs exhibited BLASTn hits with a number of Bacillus plasmids. Our isolate was positioned within B. toyonensis, of which the type strain BCT-7112, isolated from soil, is used as a feed supplement (15). Compared to BCT-7112T, VU-DES13 exhibited hits for three extra biosynthetic gene clusters by antiSMASH, as well as 13 additional β-lactamases. No virulence genes related to other B. cereus clade members (B. anthracis, B. cereus, and B. thuringiensis) were observed.

Accession number(s). Genome sequence data have been deposited in GenBank under accession number MWMG00000000.

ACKNOWLEDGMENTS

This work was supported by funding from the BE-Basic Foundation (Biotechnology based Ecologically Balanced Sustainable Industrial Consortium, Delft, The Netherlands; project no. FO7.003.05).

REFERENCES


