UNIVERSIDAD DE COSTA RICA

SISTEMA DE ESTUDIOS DE POSGRADO

ESTRUCTURA, COMPOSICIÓN Y POTENCIAL GENÉTICO PARA DEGRADAR CELULOSA DE LA MICROBIOTA INTESTINAL DEL ESCARABAJO *VETURIUS* SP. (COLEOPTERA: PASSALIDAE)

THE GUT MICROBIOME OF THE NEOTROPICAL BEETLE *VETURIUS* SP. (COLEOPTERA: PASSALIDAE): STRUCTURE, COMPOSITION AND GENETIC POTENTIAL FOR CELLULOSE DEGRADATION

Tesis sometida a la consideración de la Comisión del Programa de Posgrado en Microbiología, Parasitología, Química Clínica e Inmunología para optar al grado y título de Maestría Académica en Microbiología

JUAN GABRIEL VARGAS ASENSIO

Cuidad Universitaria Rodrigo Facio, Costa Rica

2019

Agradecimientos

Director de tesis: Adrian Pinto por su apoyo, mentoria y amistad durante el desarrollo de esta investigación. Por su motivación y contagiosa pasión por la investigación que fue vital para completar esta tesis.

Comité de tesis: Giselle Tamayo y Cesar Rodríguez por sus aportes para completar este trabajo.

Garret Suen por el entrenamiento en bioinformática que recibí en su laboratorio en UW Madison y discusión sobre los resultados

Melissa Duhaime y Greg Dick por el entrenamiento en bioinformática que recibí en sus laboratorios en Universidad de Michigan.

La sección de Bioprospección del INBio por sus contribuciones logísticas, trabajo de campo y de laboratoria.

Carol Angulo y mi familia por su apoyo durante estos años.

Acknowledgements

Thesis advisor Adrian Pinto for his support, mentorship and friendship during this project. Your motivation and passion for research was vital to complete this thesis.

Thesis committee Giselle Tamayo and Cesar Rodríguez for their contribution to complete this project.

Garret Suen for bioinformatics training at UW Madison and insightful discussion about the results.

Melissa Duhaime and Greg Dick for bioinformatics training at University of Michigan.

Bioprospecting department at INBio for their contribution in planning, field work and laboratory work for the project.

Carol Angulo and my family for their support over these years.

"Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Microbiología, Parasitología, Química Clínica e Inmunología de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Microbiología"

PhD. Max Chavarría Vargas

Representante del Decano Sistema de Estudios de Posgrado

PhD. Adrian Pinto Tomás

Profesor Guía

PhD. **César Rodríguez Sánchez**

Lector

PhD. Giselle Tamayo Castillo

Lectora

PhD. Luis Alberto Acuña Amador

Representante Programa de Posgrado en Microbiología, Parasitología, Química Clínica e Inmunología

Juan Gabriel Vargas Asensio

Sustentante

Table of contents

Resumen

El ciclo del carbono mantiene en equilibrio la concentración de dióxido de carbono y otros gases en la atmosfera que influyen directamente en el clima del planeta y en la acidez del océano. Los organismos herbívoros cumplen un papel importante en el ciclo del carbono. Sin embargo, los microorganismos que habitan sus intestinos son los verdaderos responsables de la degradación de celulosa y la consecuente producción de gases con efecto invernadero, como CO_2 y metano. En este trabajo investigamos el microbioma intestinal del escarabajo Pasálido *Veturius* sp. en el parque nacional de Braulio Carrillo de Costa Rica. *Veturius* sp. se alimenta únicamente de madera en descomposición y además presenta comportamiento subsocial, características que sugieren que el Pasálido adquiere y comparte simbiontes microbianos que son fundamentales para una eficiente producción de biomasa y energía. Para evaluar esta hipótesis se realizó un análisis de los microbiomas del intestino de adultos y larvas, así como del material que rodea las galerías en las que los Pasálidos viven en los troncos. Se utilizó secuenciación de amplicones del gen de ARNr 16S para determinar la estructura y composición de las comunidades microbianas, luego utilizamos estos resultados para seleccionar un grupo de individuos para realizar un análisis metagenómico de las funciones presentes en estos microorganismos y reconstruir sus genomas. Los resultados obtenidos sugieren que las comunidades microbianas asociadas con los intestinos de larvas y adultos y con el material de la galería son distintas entre sí, compartiendo menos del 3% de las especies presentes (total de OTUs en el sistema= 11 712). Entre los tres tipos de muestra estudiados, las galerías presentaron la mayor diversidad (3178 especies) y riqueza (chao=6714) entre los microbiomas. Firmicutes y Euryarchaeota fueron los filos dominantes en los intestinos de los adultos y las larvas. Entre las familias mas abundantes de Firmicutes se encontraron los Clostridiaceae, Lachnospiraceae y Ruminococcaceae, todos degradadores de celulosa. Un total de 766 genomas parciales de bacterias y arqueas fueron reconstruidos utilizando las secuencias de los metagenomas de adultos, larvas y sustrato asociado a las galerías; 101 de los cuales fueron clasificados como genomas completos (MAGs por sus siglas en inglés). Todos estos MAGs cuentan con un gran espectro de genes envueltos en degradación de celulosa y vías de metabolismo de carbono. Nueve genomas parciales reconstruidos del intestino de las larvas mostraron genes utilizados como marcadores moleculares de metanogénesis, lo cual sugiere que en este lugar es donde ocurren los pasos finales para completar la degradación de celulosa. Finalmente, se compararon los metagenomas del Pasalido con los de otros sistemas que degradan celulosa y encontramos evidencia de evolución convergente de las funciones metabólicas encontradas en las larvas del Pasálido con el rumen de las vacas. Los resultados obtenidos sugieren que *Veturius sp*., así como otros escarabajos similares son hasta ahora no reconocidos pero importantes contribuidores en el ciclo del carbono y la producción biótica de metano, importante gas de invernadero; y que el microbioma intestinal del Pasálido es importante para la ecología y fisiología del escarabajo.

Abstract

The global carbon cycle is responsible for maintaining the carbon dioxide and methane concentration in the atmosphere, influencing the weather and ocean acidification. Herbivores play an important role in the carbon cycle. However, the microorganisms that inhabit their guts are the ones responsible for the cellulose breakdown and consequent release of greenhouse gases such as CO_2 and methane. Here we studied the gut microbiome of the Passalid beetle *Veturius* sp., from Braulio Carrillo National Park, Costa Rica. *Veturius* sp. only feed on decay wood and presents a subsocial behavior that may lead to the acquisition and sharing of microbial symbionts for efficient biomass and energy production. Family groups from different logs in the forest were sampled and the gut microbiome of larvae and adults as well as the woody gallery material (substrate) in which they resided was analyzed. The structure and composition of the communities was determined using amplicon sequencing of 16S rRNA genes. These results were then used in order to select the samples to perform metagenomic sequencing for further functional analysis and genome reconstruction. The results showed that adult, larvae and gallery material harbor significantly different communities, sharing less than 3% of total OTUs (total OTUs in the system= 11 712), with the gallery woody substrate having the higher diversity (3178 observed species) and richness (chao=6714). Firmicutes and Euryarchaeota were the dominant phyla in adults and larvae gut. The most abundant families of Firmicutes included Clostridiaceae, Lachnospiraceae and Ruminococcaceae, all known for its cellulose degradation capacity. A total of 766 partial genomes were reconstructed using the metagenomic sequences from adult, larvae and substrate; 101 of were classified as metagenome assembled genomes (MAG). Larvae and adults are enriched in

microorganism with genomes having a myriad of glycosyl hydrolases, and other functions related to carbon metabolism; furthermore, methanogenesis markers were found in the larval partial genomes, suggesting that those performed the final steps of cellulose decomposition. Finally, the *Veturius* sp. metagenomes were compared with datasets from other cellulolytic systems, the results showed evidence of convergent evolution of functions between the larvae and cow rumen, suggesting that *Veturius* sp. and other beetle larvae are unexplored, yet important contributors to the carbon cycle and the biotic production of the greenhouse gas methane. The Passalid microbiome is important to the ecology and physiology of these beetles.

List of Tables

List of Figures

List of abbreviations

IMG: Integrated Microbial Genomes

JGI: Joint Genome Institute

MAG: Metagenome Assembled Genome

TRFLP: terminal restriction fragment polymorphisms

INBio: Instituto Nacional de Biodiversidad

OTU: Operational Taxonomic Unit

COG: Clusters of Orthologous Genes

CAZy: Carbohydrate Active Enzymes database

Pfam: Protein families

ORF: open reading frame

ANI: Average Nucleotide Identity

PCoA: Principal Coordinate Analysis

ANOSIM: Analysis of Similarity

GH: glycosyl hydrolase

GT: glycosyl transferase

CE: carbohydrate esterase

PL: Polysaccharide lyase

CBM: carbohydrate binding module

AA: auxiliary activity

TCA: tricarboxylic acid cycle

UNIVERSIDAD DE COSTA RICA

SEP Sistema de Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Juan Gabriel Vargas A, con cédula de identidad 1-1208-0579, en mi condición de autor del TFG titulado Estructura, composición y potencial genético para degradar celulosa de la microbiota intestinal del escarabajo Vetovia sp. (Coleoptera: Passalidae)

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. **SI** [ ] **NO** *[X]

***En caso de la negativa favor indicar el tiempo de restricción:** 1 **año (s).**

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

INFORMACIÓN DEL ESTUDIANTE:

Nombre Completo: Juan Gabriel Vargas Arcenio

Número de Carné: A25268 Número de cédula: 1-1208-0579

Correo Electrónico: jgabriel1206@gmail.com

Fecha: 1-18-20 .Número de teléfono: (+1) 312-478-6526

Nombre del Director (a) de Tesis o Tutor (a): Adrien Pinto Tomás

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

1 Introduction

Insects are one of the most abundant and diverse groups of animals in the planet. They have evolved different strategies to colonize every environment in the planet, distinct types of behavior, and stablish symbiotic relationships with microorganisms[1]. Furthermore, because of this great ecological versatility, insects are able to feed on almost any available substrate in the environment. Among insects, the order Coleoptera is the most diverse group containing over 380 000 described species distributed in 178 families[2,3]. Beetles account for 45% of insects and about 38% of all described animals[2]. They have inhabited our planet for about 285 million years (Permian), and can be found in any habitat presenting both diurnal and nocturnal lifestyles. In the forest, most beetles inhabit the treetops, but they can also be found under rocks, associated with trees, in the foliage, and colonizing fallen logs in decomposition[4–6]. They have a broad diet that includes plants and roots, some eat seeds and grains, some are predators of insects or other small organisms, others are fungivores, and many feed on decaying organic matter[5,6]. The life cycle of the Coleoptera can range from a few weeks to two or three years, it starts with the eggs from which larvae emerge (which can be mobile or sessile). Beetle larvae have two to five larval instars, after which they pass into the state of pupae, in this stage the individuals do not require food and are immobile. The final stage is the adult, in which they reach sexual maturity and therefore is the reproductive phase. It has been estimated that Costa Rica harbors over 35 000 species of beetles belonging to 110 different families[5].

The beetle from the Passalidae family are a homogeneous and cosmopolitan group that preferentially inhabits tropical environments[5,7–9]. The Passalid beetles in Costa Rica are distributed

in about 50 species, ranging in size between 15mm to 50mm. They have black elytra and prominent jaws[10]. The Passalid diet consists of decomposing wood, the adults colonize decaying tree logs in the forest by chewing inside and creating chambers that they inhabit for the rest of their life. Particularly, the beetles from Passalidae show subsocial behavior, they live in familiar groups in which one can found individuals from three simultaneous generations living inside the tree logs[8,10,11]. Furthermore, the adults take care of their larvae and prepare specific chambers that they later use to lay the eggs. Inside the chambers the adults and larvae feed on the substrate material (chewed wood mixed with feces) that covers the totally of the chambers; it has been proposed that the substrate is rich on partially digestive molecules that can easily absorbed in the beetle gut. Also, the substrate represents an ideal vehicle for the horizontal transmission of microbes between adults and larvae[12,13].

The microbiota of the insect gut plays a major role in their physiology, development and environmental adaptability. Close interactions between insects and microbes are abundant and widespread; moreover, it has been estimated that the majority of insects in the planet have developed symbiotic relationships with a microbial partner[14,15]. Several experiments imposing nutrient depleted diets on insects have demonstrated that the microbial communities are able to supply the insect host with the missing nutrients required for its growth[1,16]. These findings are not surprising given the large amount of evidence of this type of interactions between microbes and animal hosts, including humans.

Recent advances in sequencing technologies have allowed researchers to study the microbial diversity of the planet, their role in the biogeochemical cycles, and the interactions of these microbes with the different components of the ecosystem, including animals. Herbivores have evolved different yet efficient strategies to host microbial communities that assist them on the decomposition of plant material, playing an important role in the global carbon cycle. Invertebrate herbivores as termites and beetles are abundant and widespread in tropical forests worldwide proving to be efficient agents of degradation of recalcitrant plant material and significantly contributing to the carbon cycle[17–19]. The ability of insects to breakdown cellulose lays on the establishment of a microbial community in their guts responsible for the production of secreted enzymes for the hydrolysis of cellulose molecules as well as its many degradation byproducts[17,20–22]. They acquire their microbiota mainly from the materials on which they feed; many of these microorganisms have adapted to the digestive tract environment and over time, developed mutually beneficial relationships with their host which provides the host with the ability to utilize nutrients that otherwise will be unavailable for them.[18,23,24] The microorganisms in insect guts and in the rumen of some mammals are able to completely decompose plant material producing intermediate metabolic products that can be uptake by the host and fermented by the microbial community[25]. The final steps of the remineralization of organic matter produce the greenhouse gasses CO_2 and CH_4. Methane, which is primarily produced in nature by anaerobic archaea and by the alphaproteobacteria *Rhodopseudomonas palustris*[26], is considered the most potent greenhouse gas, as it has 34 times the warming potential of CO_2 by mass[26]. It has been estimated that these archaea produce one Gigaton of methane each year in anoxic environments, representing about 2% of the CO_2 fixed annually via photosynthesis[27].

The carbon cycle is fundamental for the life on Earth[28,29]. It regulates the temperature in the surface working as the thermostat of the planet by controlling the equilibrium of CO_2 concentrations in the atmosphere and the ocean[29]. Microorganism actively participate in the carbon cycle, especially in the fast carbon cycle. The photosynthetic primary producers fix inorganic carbon from the atmosphere and transform it in organic molecules that are later respire by heterotrophic organisms[28]. Herbivory in the rain forest largely contribute to the fast carbon cycle by the breakdown of carbon fixed by plants in the form of cellulose to CO_2 and methane[30]. However, human activity since the industrial revolution have alter the carbon thermostat in the planet increasing the levels of CO_2 in the atmosphere causing global warming and ocean acidification[28,29,31]. In recent years, international efforts have been proposed to keep CO_2 emissions under 450 ppm by the year 2100 aiming to decrease the effects of global warming. The use of bioenergy and other renewable and clean energies have been getting attention as potential strategies to reduce CO_2 in the atmosphere[32,33]. Currently, bioenergy produced from plant biomass provides about 10% of the global supply with 50EJ/year, bioethanol and biodiesel represent only 3EJ/year[34]. Despite the efforts in recent years to optimize the enzymatic production of bioenergy, the efficiencies obtained for this process are still limiting the wider industrial production. The molecular heterogeneity and recalcitrance of cellulose and similar compounds are the major challenges that have to be overcome in order to achieve increased efficiencies in biorefineries[34,35].

The production of clean energies, depend on the discovery of new enzymes in nature capable of hydrolyze lignocellulose in plant biomass. Glycosyl hydrolases (GH) encoded in the genomes of microorganisms are important for the degradation of these complex polysaccharides[36]. The

conversion of cellulose to glucose require the action of a combination of these enzymes including endo-1-4-beta-glucanases, exo-1-4-beta-glucanases, beta-glucosidases and carbohydrate binding modules; the latter attached with different affinities to catalytic domains facilitating the action of the hydrolases[34,37]. Many glycosyl hydrolases have been discovered in microorganism and classified in different groups according to their activity, amino acid sequence, and other intrinsic properties. The family GH5 is one of the largest including enzymes with different substrate specificities including cellulases, mannanases, xylanases and xyloglucanases which hydrolyze the larger cellulose molecules[36,38]. Because of the vast diversity of sequences in GH5, this was divided in several subfamilies of CAZy enzymes and then included in the CAZy GH Clan A together with other 19 families of GHs that share similar catalytic properties[38]. The GH10 family, mainly containing xylanases, GH3 mainly glucosidases and GH43 which includes beta-xylosidases, beta-galactatosidase and other endo-glucanase activities; are just some of the GH families that include phylogenetically diverse enzymes that are involved in cellulose degradation[39–41]. Moreover, a new mechanism has been recently discovered in GH61s and CBM33s which are capable of catalyze the oxidative cleavage of cellulose[34]. Glycosyl hydrolases act as individual enzymes or as multienzymatic complexes called cellulosomes[34,42]. Similar to GHs, cellulosomes are phylogenetically diverse and have been isolated in multiple environments. They present sequence and affinity heterogeneity but one common feature is the high efficiency of their hydrolytic activity[43]. Therefore, cellulosomes have been studied because of their biotechnological potential. Systems involving insects and gut symbionts have shown great potential for the discovery of new CAZy enzymes, though the exploration of these have only recently started[34,44,45].

There are several examples of host-bacteria symbiotic associations enabling the organisms to efficiently degrade carbon to produce energy and biomass. One of the first studied systems is the beetle *Melolontha melolontha* (Coleoptera:Scarabaeidae), where terminal restriction fragment length polymorphisms (t-RFLP) were exploited to study the microbial diversity associated with this beetle and revealed that the midgut and the hindgut harbor different microbial communities[46]. However, their most important finding was that the hindgut microbiota was specialized in carbohydrate fermentation and methanogenesis. They found that Clostridiales, Bacillales, Lactobacillales and Actinobacteria were the most important bacterial groups, and suggested that those were responsible for cellulose degradation. The only methanogenic archaea found in this study was *Methanobrevibacter* sp.[46]. Another study of beetle-associated microbiota, this time in *Pachnoda ephippiata* (Coleoptera:Scarabaeidae), showed a similar structure of the community that were consistent with previous findings, but proposing that there were differences between the midgut and the hindgut due to the high alkalinity of the midgut. The hindgut instead consisted in an anaerobic environment ideal for fermentation[47,48]. Similar findings were later described for the bark beetle *Dendroctonus frontalis* (Curculionidae), unlike the previous studies, the authors found a high abundance of Proteobacteria in the gut of this beetle[49].

More recently, a metagenomic study of the hindgut of *Anaplophora glabripennis* (Cerambycidae), an important agricultural pest, showed that Firmicutes and Actinobacteria were the dominant groups. Their metagenomes were rich in laccases, xylanases and other glycosyl hydrolases necessary for cellulose degradation[21]. A second study of the microbial community of the wood-boring beetle *A. glabripennis* demonstrated that these beetles acquire specific bacterial groups

from the environment, as well as from vertical transmission; these bacteria were involved in nutrient acquisition, nitrogen metabolism and other important processes.[22]

The microbial communities of insects and other animals that feed on plant material have also been studied. These studies include the higher termites *Nasutitermes* and *Amitermes*, which host Spirochaetes and Fibrobacteres in their paunch (P3 segment), these microbes, along with other nitrogen fixing bacteria, encode hydrolytic enzymes involved in the degradation of lignocellulose[17,19,50,51]. Other environmental and economically important symbiotic cellulose degradation system is the mammal rumen, especially cow rumen. This particular environment harbors a diverse and dynamic community consisting in Fibrobacteres, Firmicutes as *Ruminococcus* and others, and Bacteroidetes that contribute to the efficient degradation of cellulose and influence the physiology of the host [52,53]. Many other mammal rumens have been sequenced and studied for their capability of cellulose degradation including goat[54], sheep[55,56], the tammar wallaby[57], and even the baleen whale[58]. All of them were consistent in demonstrating the primarily roles of the microbes that produce the glycosyl hydrolases and other enzymes for cellulose degradation.

In this study we use amplicon sequencing of the molecular barcode gene 16S rRNA to carefully explore the microbial diversity of adult, larval guts and the substrate (explained above) of Passalid beetles in Costa Rica. I then used shotgun metagenomics aiming to explore functional capabilities of the diverse taxa in the *Veturius* sp. (Coleoptera: Passalidae) gut, specifically mining for homologs of known genes involved in cellulose degradation and other vital metabolic processes

such as nitrogen fixation. Furthermore, I reconstructed the genomes of the most important members of the community to determine their role in the carbon transformations, the ecology and physiology of the beetle, and overall functional capacity and novelty of these organisms.

2 Scientific question

Do Passalid beetles from Costa Rica harbor gut microbial communities enriched in bacteria and archaea that aid them in the complete degradation of cellulose all the way to methane as well as other metabolic processes involved in the carbon cycle, contributing to their physiology and ecology?

3 Justification

Advances in biotechnology and molecular ecology, specifically those involving the so called "next generation sequencing" technologies, have allowed the microbiological exploration of almost every existing environment on the planet. Over the last 15 years, researchers have uncovered the microbial diversity in the biosphere and its complexity, however, we don't fully understand its functional capabilities and dynamics, as well as the interactions between microbial communities and the environment.

The dramatic declines of fossils fuel reservoirs actively affect the economies of developing countries and have produced military conflicts around the world. Moreover, the excessive use of fossil fuels has caused severe damage to the planet. Therefore, it is crucial to find alternative and cleaner sources of energy. In recent years, researchers have turned their attention to natural systems looking for organisms able to use enzymes that can potentially facilitate various industrial processes in fields as medicine, food, energy and environmental conservation.

Environmental systems that depend on cellulose degradation are of great biogeochemical and biotechnological interest. These consist on microbes that breakdown cellulose and other

recalcitrant derivates converting them into smaller molecules that then are available to the host and other members of the community to feed; and in the process they largely contribute to the carbon cycle in the planet and the production of greenhouse gases. From a biotechnological point of view, the microbes that are involved in cellulose degradation are important since their genomes encode for the specific enzymes that carry the process. Furthermore, these enzymes have been subjected to evolutionary processes that have selected for efficiency and affinity for thousands of years.

Costa Rica is known worldwide because of its environmental consciousness and large efforts towards conservation of natural resources. More recently, the country has committed itself to achieve carbon neutrality by the year 2021 (Decreto N°37926[59]) by promoting the use of clean energies. The purpose of this project is to contribute to these efforts by exploring the Passalid beetle gut in search for new enzymes, specially glycosyl hydrolases, with potential use in bioenergy production and biotechnology. These beetles feed only on decomposing wood and also show subsocial behavior which is infrequent in beetles, characteristics that suggest that the microbes in their guts harbor important enzymes for cellulose degradation. Therefore, I intended to study the microbial communities that inhabit the gut of adults and larvae of five *Veturius* sp. family groups, as well as the microbial communities of their surrounding environment, seeking to understand: i) the taxonomical composition of the system by using high-throughput sequencing of the 16S rRNA gene to survey the members of these communities and to make comparisons between the samples in order to find the potential host-microbe interactions. ii) how some microorganisms play important roles in the communities by reconstructing their genomes in order to explore in detail their metabolic capabilities and mine for the specific enzymes that breakdown

cellulose and drive other important processes. iii) how is the Passalid beetle cellulose degradation compared to other model systems like termites and rumen, and how this influence their ecology. In order to do this, I first performed a preliminary analysis of 69 amplicon sequencing samples of the 16S rRNA gene in order to determine the influence, if any, of our molecular biology methods in the diversity of the system. Then, I selected 15 samples of adults, larvae and substrate corresponding to 5 family groups of *Veturius* sp. consistently extracted with one method to carry out the in deep analysis of the community composition. Giving the abundance and novelty of archaea found on the previous analysis, we decided to sequence the gut contents of five larvae in full lanes of Illumina HiSeq 2000, and to pool two adults and two substrate samples to be sequenced in one lane each. These seven metagenomes were used for the read-based analysis and for the metagenomic comparisons done in IMG/M platform. For consistency, we co-assembled and binned the adult and substrate pool metagenomes with the corresponding two larvae in order reconstructed genomes of the taxonomical groups in the system.

4 Hypothesis

The Passalid beetle *Veturius* sp. from Costa Rica harbor gut microbial communities enriched in bacteria and archaea that completely breakdown cellulose producing methane and CO_2; and aid them in other biogeochemically important functions contributing to the host physiology and ecology.

5 Research aims

5.1 Main aim

To determine the composition of the gut microbial communities of five family groups of *Veturius* sp. (Passalidae) beetles from Costa Rica and the substrate present in their habitat, and explore their potential for cellulose degradation.

5. 2 Specific aims

I. To optimize the DNA isolation methods to the study the taxonomical and functional distributions in the Passalid beetle *Veturius* sp. microbial communities.

II. To determine the structure and composition of the microbial communities in the guts of adults and larvae of the Passalid beetle *Veturius* sp. as well as in the substrate material present in the galleries that they inhabit.

III. To identify enzymes and metabolic pathways potentially involved in cellulose degradation in the *Veturius* sp. system.

IV. To reconstruct the genomes of the microorganisms likely involved in cellulose degradation and other processes important for the ecology and physiology of the beetles.

V. To compare the Passalid beetle microbial communities with those of other cellulolytic systems to identify similarities and unique features.

6. Methods

6.1 Sample collection and processing

Beetle specimens were collected in collaboration with INBio's Bioprospection research staff during two field trips to the Braulio Carrillo National Park, Quebrada González sector, Heredia Province (10° 9′ 36″ N, 83° 58′ 28″ W), between August 2011 and March 2012 (permit: R-CM-INBio-094-202-OT). The interior of decomposing logs was sampled (logs sampled showed a decomposition level that allows colonization by beetles), to collected larvae and adults from the Passalid beetle genus *Veturius* sp. Only logs that presented at least three adult and three larvae specimens were collected, all the individuals found were collected. All insects were transported to the laboratory at room temperature inside vented polyethylene containers and surrounded by the woody material in which they were collected (hereafter referred as substrate). All the individuals were alive at the time of the DNA extraction in the lab. The specimens were placed at -20°C in a sterile petri dish for 10-20 minutes prior the dissection. Additionally, we sterilely collected 1 –5 g of the substrate from where the beetles were found in 50mL Falcon tubes for metagenomic DNA extraction. Once in the lab, all beetles were dissected in sterile conditions, using scissors and forceps to extract the digestive tract. Beetle taxonomic identification was carried out in lab by experts at the Costa Rican National Biodiversity Institute (INBio) and voucher specimens were deposited at INBio's insect collection.

6.2 DNA extraction and library preparation

The DNA extractions of 15 samples (5 adult guts, 5 larvae guts and 5 substrates) belonging to five family groups of *Veturius sp*. were performed by INBio's research staff using three commercial DNA extraction kits: MP, MoBio PowerSoil kit and Epicentre including some modifications to the manufacturer's protocols. These modifications included different incubation times in buffers and the use of previously heated elution buffers to obtain improved extraction yields. Each sample was fractioned to in six equal parts (by weight) before performing the DNA extraction. All adult and larvae gut samples were processed using the three DNA extraction kits with and without sonication, none of the substrate samples were sonicated because these lack of animal tissues. The additional sonication steps were done with the intention to test whether this facilitates the extraction of gDNA from microorganisms that were strongly attached to the gut tissue. The detection of these taxa was performed by pairwise comparison of the 16S rRNA sequences obtained for the sonicated/non-sonicated samples. Samples including the gut contents and gut epithelium (called epithelium samples) were not sonicated either in order to detect if tightly attached taxa that could be absent or underrepresented in the gut content samples, similarly the differences were determined using the 16S rRNA data. DNA samples were shipped with dry ice to the Joint Genome Institute (JGI) in Walnut Creek California for 16S rRNA amplicon sequencing (454 Titanium FLX) of the V4-V5 region using the EMP primers, and shotgun metagenomic sequencing (Illumina HiSeq 2000). See Table 1 for a summary of samples, treatments and DNA extraction methods.

After performing a preliminary analysis of the microbial community composition of all the samples, comparing the diversity and richness obtained for the 16S rRNA amplicons of the 69

samples, we selected seven samples extracted using the MoBio extraction kit (more consistent in terms of yields and diversity) for metagenomic sequencing (Illumina runs) in collaboration with JGI. These samples included the gut microbiome of five individual larvae (larvae from logs 1-5), one pooled sample from two adult guts (logs 3 and 4) and one pooled sample from the gallery material associated with these insects (from logs 3 and 4 as well).

6.3 16S rRNA amplicon and metagenome analysis

All 16S rRNA sequence libraries were quality filtered and processed using Qiime[41] as previously described. Briefly, sequences that showed Phred quality scores lower than 33, as well as the ones that contained homopolymers and mismatches in the barcode and adapter sequences were filtered out of the analysis. In order to include all samples in the analysis, these were randomly rarified to 46 500 sequences (number of reads in the samples with lowest yield) per sample for alpha and beta diversity calculations. OTUs were defined at 97% sequence similarity to create the OTU table using the greengenes database v13.8. Unifrac distances were calculated and used for ordination analysis of beta diversity.

The metagenome sequences were quality filtered (quality score > 30) using cutadapt[60] v1.9, and the remaining reads were assembled in megahit[61] v1.0.3 using kmer sizes from 31 – 91 nucleotides with a kmer step of 30, the minimum contig length was set to 1000 base pairs. We used Bowtie2[62] to map the reads back to the assembly using alignment preset options local, mode very-fast, the mapping files were sorted and indexed with SAMtools[63] v1.8. The unsupervised binning of the contigs was performed using Metabat2[64] using differential coverage and tetranucleotide frequencies. Finally, Anvio[65] v3 was used to construct the metagenomic database in which we

stored the contigs, the coverage profiles and the functional annotations. Anvio was also used for the manual curation of the metagenomic bins, each bin was visualized individually and evaluated based on its: differential coverage, sequence composition, as well as the completion (% of single copy core genes in bin) and contamination (% of redundant single copy core genes in bin) in terms of the number of single copy core genes found in each bin. The manual curation involved the careful examination of each bin coverage, separating the contigs with similar tetranucleotide frequencies but with coverages inconsistent with the rest of the bin. If necessary, the gene content and order of dubious contigs was examined in order to detect chimeric contigs, or to confirm the quality of the contigs. After manual curation, the reconstructed metagenomic bins that presented completion >70%, contamination less than 7% and were larger than 1 Megabase were classified as metagenome assembled genomes (MAGs). We called open reading frames (ORFs) in all metagenomic bins and MAGs using Prodigal[66], and performed taxonomical and functional annotation by sequence similarity against the ones in publicly available databases including: RefSeq, COG, CAZY and others with Diamond[67] using the more-sensitive setting (evalue < 0.00001). Taxonomical classification of MAGs was attempted using average nucleotide identities (ANI) with FastANI[25], and kmer-based sequence classifiers Kraken[26] and Centrifuge[27]. The final taxonomical approximation of the MAGs was done using sequence similarity of single copy core genes (Campbell gene set[68]) compared with the KBase bacteria and archaea genomes reference database. The taxonomical information of the top hit obtained for each gene was tabulated to determine the highest taxonomical level that was common for the majority of the genes. The metabolic model reconstruction of MAGs was performed in the KBase platform[69].

The comparative metagenomics analysis of the Passalid metagenomes versus other cellulose and insect-associated metagenomes (Supplementary Table 1) was performed using the open access platform Integrated Microbial Genomes and Metagenomes (IMG/M) from the JGI; by using IMG/M we ensure that all metagenomes were quality filtered and processed using the same pipeline. IMG/M uses state- of-the-art algorithms for the processing of metagenomic data including quality filtering with BBMap tools, assembly programs as Newbler and SPADES, gene calls with Prodigal and annotations by sequence similarity using BLAST and HMMER for Pfam and other protein profile searches[70]. Moreover, IMG/M contains functional and taxonomic annotations for reads and assembled genes that are constantly being updated. The normalization of the data was done by total reads and by gene content for unassembled and assemble data, respectively[71,72].

The R packages RColorBrewer (https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf) and Viridis (https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html) were implemented to define colorblind and print friendly color schemes for the figures.

6.4 Phylogenetic analysis

The phylogenetic analysis were performed in MEGA7[73]. Multiple sequence alignments were done in MUSCLE implemented in MEGA using the default settings. Reference sequences were retrieved from GenBank. The evolutionary model was selected based on Akaike values calculated

in MEGA7 selection model tool. Maximum likelihood was used to construct the phylogeny. The trees were edited in MEGA7.

6.5 Statistical analysis

All the anova tests used to evaluate the significance found in the comparisons made in this study were performed in R Studio Version 1.1.383. The ANOSIM analysis was implemented in Qiime, the R statistic varies from 0 to 1, values closer to 1 indicates higher separation between the groups, and values close to 0 suggest no difference between groups. Furthermore, the difference obtained for sample type is statistically significant ($p < 0.05$). In contrast, we obtained a R value close to 0 for family group (adult and larvae collected from the sample log along with their corresponding substrate). The Metastats analysis to determine the number of OTUs that were differentially abundant in between two sample types was performed in MOTHUR. The metagenomic comparative analysis was performed using Spearman correlations clustering implemented in IMG/M. The metagenomic gene hits data was normalized to the corresponding metagenome total gene count.

7 Results

7.1 Gut dissections and influence of DNA isolation on the microbial diversity

7.1.1 The gut of adults and larvae present anatomical and physiological differences

After successfully collecting five familiar groups of Passalid beetles from the sampling site, the plastic containers were inspected in order to discard any dead individuals or other organisms that were not part of the sampling effort. Then, the specimens were dissected in order to obtain the guts preventing the contamination of the sample with the environment or other tissues. The first observation we recorded is that the gut of the adults and larvae are anatomically different (Figure 1B, 1D). The adult gut was similar to the one in other insects, this was compartmentalized and the contents consisted of a mixture of wood materials that were wet and dark suggesting that these were subject to chemical and enzymatical transformations. On the other hand, the larval gut consists of a straight tube with no evidence of compartmentalization, at least to the naked eye and after stereoscopic inspection. The contents of the larval gut also differed from the adult's as these were dryer and less mixed, more similar to the substrate material, in which it is possible to recognize wood fibers. These differences are important for further analysis and data interpretation.

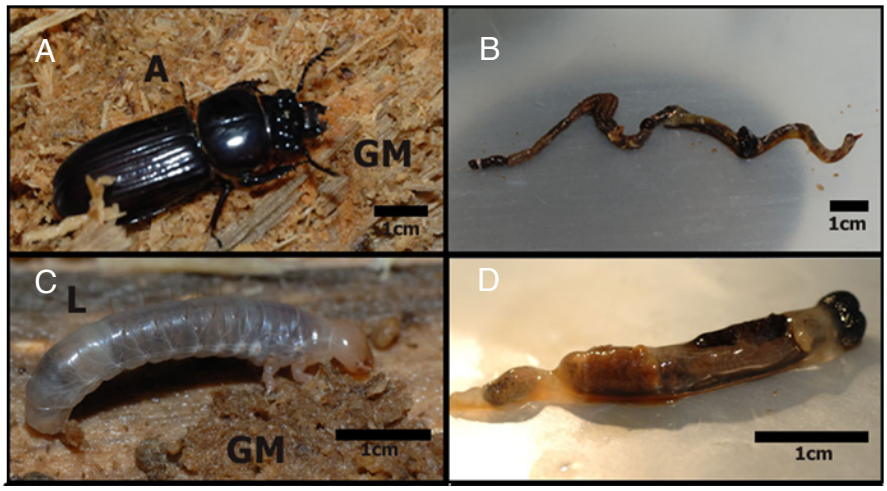


Figure 1. *Veturius* sp. (Coleoptera: Passalidae) individuals, guts and gut contents. A- *Veturius* sp. adult and the gallery material (substrate). B- Passalid adult dissected gut. C- Passalid larvae and the gallery material (substrate). D- Passalid larvae dissected gut. Black letters code for: A= Adult in the trunk gallery surrounded by GM= gallery material, L= larvae.

In total we processed 69 samples for 16S rRNA gene amplicon sequencing and shotgun metagenomics (Table 1). The samples were distributed as follows: 29 adult, 29 larvae and 11 substrate samples. From the total, 44 of the samples corresponded to gut content, and 14 samples included gut contents plus gut epithelium. Finally, 22 of the 69 samples received the sonication treatment.

Table 1. Samples for DNA isolation and 16S rRNA amplicon sequencing.

Sample*	Sample Type	Stage	Treatment	Extraction Method
Fam1_BA	Gut content	Adult	No sonication	MP
Fam1_BL	Gut content	Larvae	No sonication	MP
Fam1_BSA	Gut content	Adult	Sonication	MP
Fam1_BSL	Gut content	Larvae	Sonication	MP
Fam1_BSU	Gut content	Substrate	No sonication	MP
Fam1_MA	Gut content	Adult	No sonication	Mobio
Fam1_ML	Gut content	Larvae	No sonication	Mobio
Fam1_MSA	Gut content	Adult	Sonication	Mobio
Fam1_MSL	Gut content	Larvae	Sonication	Mobio
Fam1_MSU	Gut content	Substrate	No sonication	Mobio
Fam2_BA	Gut content	Adult	No sonication	MP
Fam2_BL	Gut content	Larvae	No sonication	MP
Fam2_BSA	Gut content	Adult	Sonication	MP
Fam2_BSL	Gut content	Larvae	Sonication	MP
Fam2_BSU	Gut content	Substrate	No sonication	MP
Fam2_MA	Gut content	Adult	No sonication	Mobio
Fam2_ML	Gut content	Larvae	No sonication	Mobio
Fam2_MSA	Gut content	Adult	Sonication	Mobio
Fam2_MSL	Gut content	Larvae	Sonication	Mobio
Fam2_MSU	Gut content	Substrate	No sonication	Mobio
Fam3_BA	Gut content	Adult	No sonication	MP
Fam3_BA	Epithelium	Adult	No sonication	MP

Fam3_BepA	Epithelium	Larvae	No sonication	MP
Fam3_BL	Gut content	Larvae	No sonication	MP
Fam3_BSA	Gut content	Adult	Sonication	MP
Fam3_BSL	Gut content	Larvae	Sonication	MP
Fam3_BSU	Gut content	Substrate	No sonication	MP
Fam3_MA	Gut content	Adult	No sonication	Mobio
Fam3_MepA	Epithelium	Adult	No sonication	Mobio
Fam3_MepA	Epithelium	Larvae	No sonication	Mobio
Fam3_ML	Gut content	Larvae	No sonication	Mobio
Fam3_MSA	Gut content	Adult	Sonication	Mobio
Fam3_MSL	Gut content	Larvae	Sonication	Mobio
Fam3_MSU	Gut content	Substrate	No sonication	Mobio
Fam4_BA	Gut content	Adult	No sonication	MP
Fam4_BepA	Epithelium	Adult	No sonication	MP
Fam4_BepL	Epithelium	Larvae	No sonication	MP
Fam4_BL	Gut content	Larvae	No sonication	MP
Fam4_BSA	Gut content	Adult	Sonication	MP
Fam4_BSL	Gut content	Larvae	Sonication	MP
Fam4_BSU	Gut content	Substrate	No sonication	MP
Fam4_MA	Gut content	Adult	No sonication	Mobio
Fam4_MepA	Epithelium	Adult	No sonication	Mobio
Fam4_MepL	Epithelium	Larvae	No sonication	Mobio
Fam4_ML	Gut content	Larvae	No sonication	Mobio
Fam4_MSA	Gut content	Adult	Sonication	Mobio

Fam4_MSL	Gut content	Larvae	Sonication	Mobio
Fam4_MSU	Gut content	Substrate	No sonication	Mobio
Fam5_BA	Gut content	Adult	No sonication	MP
Fam5_BepA	Epithelium	Adult	No sonication	MP
Fam5_BepL	Epithelium	Larvae	No sonication	MP
Fam5_BL	Gut content	Larvae	No sonication	MP
Fam5_BSA	Gut content	Adult	Sonication	MP
Fam5_BSL	Gut content	Larvae	Sonication	MP
Fam5_BSU	Gut content	Substrate	No sonication	MP
Fam5_EA	Gut content	Adult	No sonication	Epicentre
Fam5_EepA	Epithelium	Adult	No sonication	Epicentre
Fam5_EepL	Epithelium	Larvae	No sonication	Epicentre
Fam5_EL	Gut content	Larvae	No sonication	Epicentre
Fam5_ESA	Gut content	Adult	Sonication	Epicentre
Fam5_ESL	Gut content	Larvae	Sonication	Epicentre
Fam5_ESU	Gut content	Substrate	No sonication	Epicentre
Fam5_MA	Gut content	Adult	No sonication	Mobio
Fam5_MepA	Epithelium	Adult	No sonication	Mobio
Fam5_MepL	Epithelium	Larvae	No sonication	Mobio
Fam5_ML	Gut content	Larvae	No sonication	Mobio
Fam5_MSA	Gut content	Adult	Sonication	Mobio
Fam5_MSL	Gut content	Larvae	Sonication	Mobio
Fam5_MSU	Gut content	Substrate	No sonication	Mobio

*The number at the beginning of the sample name represents the beetle family group. The letter after the number is for the method as in column 4: M= Mobio, B= MP, E= Epicentre. After method: A= Adult, L= larvae, SU= substrate, epA= epithelium adult, epL= epithelium larvae, SA= sonicated adult, SL= sonicated larvae.

7.1.2 Effect of methods on the diversity of 16S rRNA gene amplicons

Three runs of 454 sequencing worth of data from 69 metagenomic DNA samples (Table 1) were processed in order to determine any biases due to the methods or the sampling procedures. In order to be able to compare the 69 samples, and determine the influence of the DNA extraction method, the sonication treatment, and the presence of gut tissue (epithelium) in the microbial diversity; the samples were rarefied to 5600 sequences per sample. Then, diversity was evaluated using several alpha diversity metrics which calculate richness (chao1), richness and evenness (Shannon), or the sum of phylogenetic distances (PD whole tree). The results of the analysis are summarized in Table 2.

Table 2. Alpha diversity of the 69 samples of 16S rRNA amplicons from adults, larvae and substrate of the *Veturius* sp.

Sample*	chao1	PD whole tree	Shannon	Extraction Method
Fam1_BA	6942.58147	110.60405	8.196718684	MP
Fam1_BL	5317.602941	96.33563	7.693401658	MP
Fam1_BSA	5845.305785	108.26982	8.197369362	MP
Fam1_BSL	4513.257511	81.35535	7.581689168	MP
Fam1_BSU	10071.21812	146.03182	10.83567555	MP
Fam1_MA	11860.00714	201.49126	9.838713453	Mobio
Fam1_ML	6130.487252	107.51615	8.649605054	Mobio
Fam1_MSA	6597.715953	128.61134	9.050413274	Mobio
Fam1_MSL	6468.926366	118.85203	9.035112493	Mobio
Fam1_MSU	8572.207607	127.29079	10.66957529	Mobio
Fam2_BA	6732.472585	113.36913	8.4694718	MP
Fam2_BL	3892.151899	72.8708	7.107822532	MP
Fam2_BSA	5267.219595	94.92654	7.648265796	MP
Fam2_BSL	3923.141221	78.46411	7.578412522	MP
Fam2_BSU	7886.784884	111.56631	10.49600311	MP
Fam2_MA	5345.952778	103.96622	8.739499319	Mobio
Fam2_ML	5052.514851	90.81476	8.427981824	Mobio
Fam2_MSA	6500.016393	119.67019	8.482940299	Mobio
Fam2_MSL	6775.200441	119.61558	8.874203626	Mobio
Fam2_MSU	7896.357877	102.24554	10.43546708	Mobio

Fam3_BA	5959.131894	108.78714	8.739192729	MP
Fam3_BA	634.9782609	14.9262	2.030833283	MP
Fam3_BepA	1873.878505	40.22981	3.883629677	MP
Fam3_BL	16558.19203	240.59072	8.921270477	MP
Fam3_BSA	5488.84573	101.18137	7.993706307	MP
Fam3_BSL	10093.18579	168.87937	8.059502417	MP
Fam3_BSU	5449.307159	73.69483	9.902841423	MP
Fam3_MA	6930.610338	125.45082	9.031743059	Mobio
Fam3_MepA	1640.04	30.11073	2.914706628	Mobio
Fam3_MepA	2213.783333	54.64485	5.554540806	Mobio
Fam3_ML	6847.960265	115.92201	8.731634176	Mobio
Fam3_MSA	5311.94012	96.88501	8.670947195	Mobio
Fam3_MSL	6858.62954	110.90446	8.743254759	Mobio
Fam3_MSU	9874.310345	136.36452	10.39611817	Mobio
Fam4_BA	5173.315789	95.71513	6.404241302	MP
Fam4_BepA	3126.604938	70.43244	5.437377767	MP
Fam4_BepL	2511	54.41522	5.606142574	MP
Fam4_BL	5328.854701	98.36846	6.955393578	MP
Fam4_BSA	3267.421053	70.7106	6.511341229	MP
Fam4_BSL	4960.690625	91.47625	7.077217359	MP
Fam4_BSU	6773.696356	99.10992	9.796913211	MP
Fam4_MA	6092	112.89472	7.200994434	Mobio
Fam4_MepA	3163.755556	63.38128	5.54966757	Mobio

Fam4_MepL	3054.654639	59.22491	6.796098721	Mobio
Fam4_ML	7515.4	128.46517	9.06802721	Mobio
Fam4_MSA	5737.935714	113.84106	7.218707653	Mobio
Fam4_MSL	7085.911565	116.29369	8.705472019	Mobio
Fam4_MSU	4860.90099	81.9166	9.509991505	Mobio
Fam5_BA	5501.703072	92.62821	7.289944649	MP
Fam5_BepA	2234.179487	49.64845	3.31633203	MP
Fam5_BepL	3857.595331	75.29856	5.711206892	MP
Fam5_BL	6551.242236	108.81746	8.623448361	MP
Fam5_BSA	7240.629371	114.16677	7.225756317	MP
Fam5_BSL	6454.378453	111.25496	8.565411997	MP
Fam5_BSU	8558.906526	101.75848	10.02140707	MP
Fam5_EA	5963.5	98.54444	6.981134298	Epicentre
Fam5_EepA	1337.690141	26.98344	2.243813716	Epicentre
Fam5_EepL	3503.384236	68.59777	5.569195931	Epicentre
Fam5_EL	8190.873171	136.60473	9.436930599	Epicentre
Fam5_ESA	7059.246154	103.39204	7.601668905	Epicentre
Fam5_ESL	32811.49072	435.3632	10.80010548	Epicentre
Fam5_ESU	6979.301471	83.95233	9.904690052	Epicentre
Fam5_MA	8601.918972	129.16966	7.704518241	Mobio
Fam5_MepA	1977.075188	45.32512	2.989647676	Mobio
Fam5_MepL	4554.937759	81.18563	7.230727698	Mobio
Fam5_ML	8969.196078	153.23885	9.071661036	Mobio

Fam5_MSA	33	3.70337	3.424122123	Mobio
Fam5_MSL	11197.38046	188.97264	9.849371502	Mobio
Fam5_MSU	4260.38247	60.75013	8.958851233	Mobio

*Fam= beetle family group. The letter after the number is for the method as in column 4: M= Mobio, B= MP, E= Epicentre. After method: A= Adult, L= larvae, SU= substrate, epA= epithelium adult, epL= epithelium larvae, SA= sonicated adult, SL= sonicated larvae.

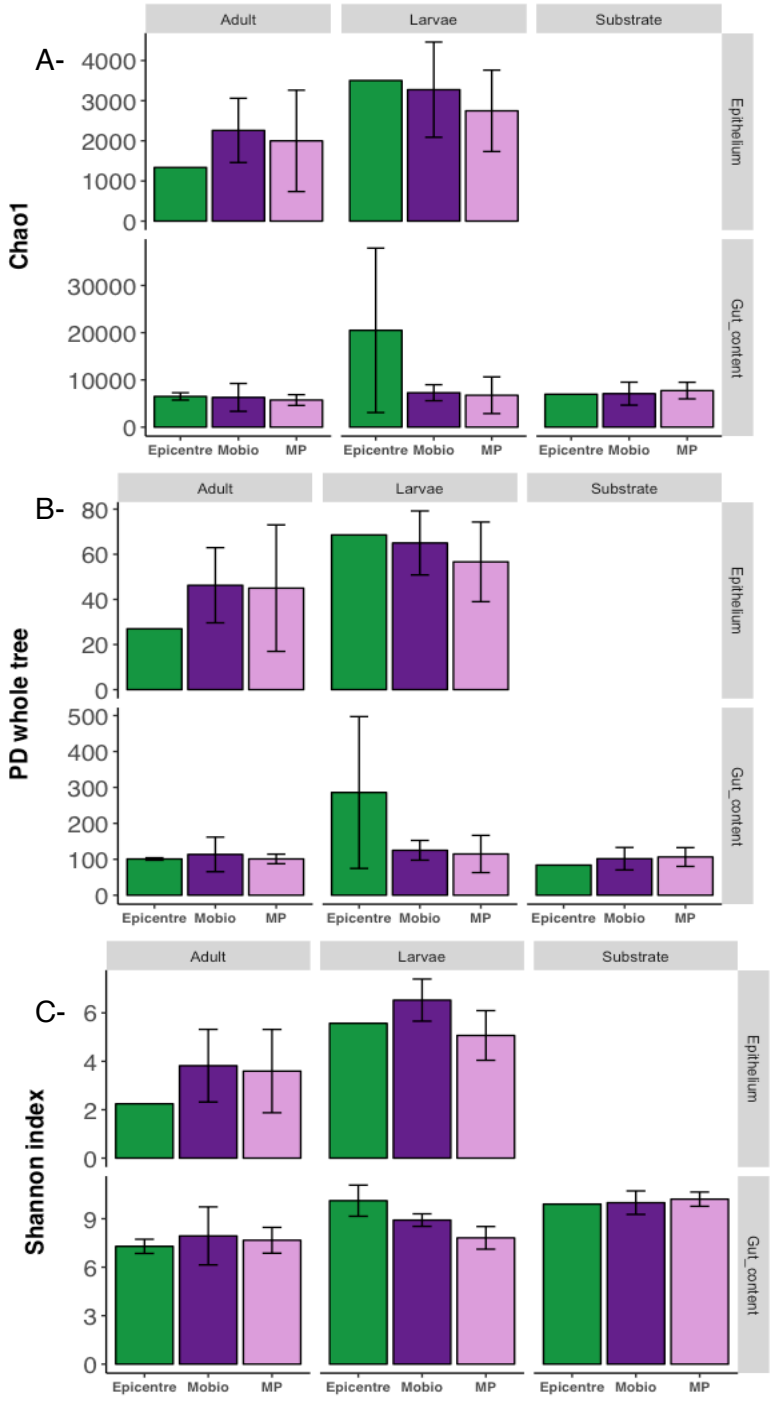


Figure 2. Effect of the DNA isolation method on the alpha diversity of the sample. A- Chao1 richness index. B- PD whole tree or "Faith's diversity index". C- Shannon diversity index.

This analysis showed that there are not significant differences ($p > 0.05$) related to the DNA isolation method (Figure 2), however, the samples extracted using the Mobio extraction kit presented more consistent results in terms of total number of reads and alpha diversity. Therefore, in order to dereplicate the dataset, we selected the 15 samples of adult, larvae and substrate of each family group extracted using the Mobio kit for further diversity analysis. These samples showed comparable yields allowing us to rarified this subset to 10 900 sequences per sample. The taxonomical classification of the reads was not considered at the time of the selection of these samples. The epithelium samples were not considered for further analysis as these showed a high abundance of Eukaryotic reads, most of which were classified as Coleoptera. Therefore, the epithelium samples presented the lowest yields and diversity of microbial reads.

7.2 The larvae and adults of *Veturius* sp. harbor specialized microbial communities

7.2.1 Structure and composition of the *Veturius* sp. gut microbial community

In order to study the structure and composition of the microbial communities in the *Veturius* sp. gut and their substrate, we analyzed the alpha and beta diversity using several metrics and statistical tests. The values obtained for the Shannon alpha diversity index, which considers richness and evenness, showed that the substrate is slightly more diverse than the adult and the larvae gut. The Inverse Simpson index, which is more weighted on the abundant taxa than Shannon, also showed a similar trend (Figure 3). There was not an obvious difference between the adult and larvae diversity when using these metrics.

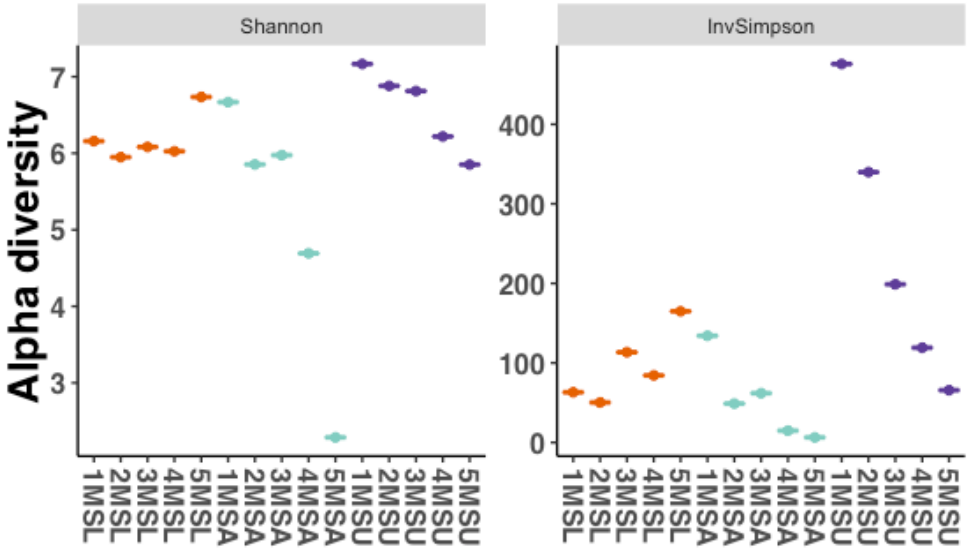


Figure 3. Shannon and Inverse Simpson (InvSimpson) alpha diversity in the Passalid beetle samples. Larvae= orange, Adult= cyan, Substrate= purple.

Next, I evaluated the richness of the communities in the different sample type (Figure 4), and found that the substrate showed the larger number of observed species (OTU criteria= 97%) compared to the adult and the larvae gut ($p < 0.05$). However, the difference observed in the richness estimator Chao1 was not significant ($p < 0.05$) between any of the sample types. Moreover, the Faith's Phylogenetic Diversity "PD whole tree", which estimates phylogenetical richness of the sample by adding the distances of the branch of the phylogenetic that is produced with the OTUs of each sample, also did not show differences between the sample types. These data showed that the only difference between the communities in the substrate, adult and larvae is in the total observed OTUs, the other richenss metrics were not significantly distinct.

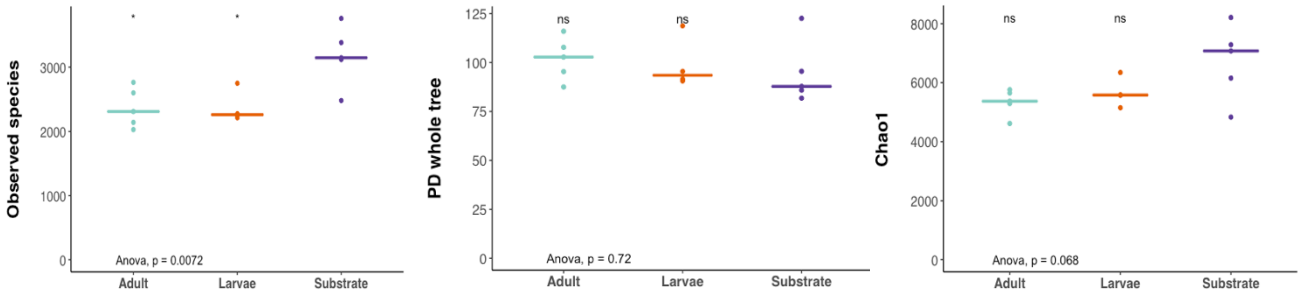


Figure 4. Microbial richness of the communities and the substrate of *Veturius* sp. Observed species calculated using an OTU criteria of 97% similarity. Ns= non-significant. *= significant compared to the substrate. Larvae= orange, Adult= cyan, Substrate= purple.

Next, I compared the beta diversity of the samples, using principal component analysis (PCoA) based on unifrac distances (calculated phylogenetic distances between the OTUs found in the system), and obtained three discrete clusters of samples, one group including the five adults, one for the five larvae, and one containing the five substrate samples; when weighted unifrac distances were considered the two principal components explained 79.7% of the variation (Figure 5A). The principal component analysis of unweighted unifrac (which does not consider the abundance of the taxa), produced the same clustering of the samples, but the two principal components explained 40.1% of the variation (Figure 5B). These results indicate that the microbial communities of the substrate, the larvae gut and the adult gut are composed of phylogenetically distinct microbial taxa.

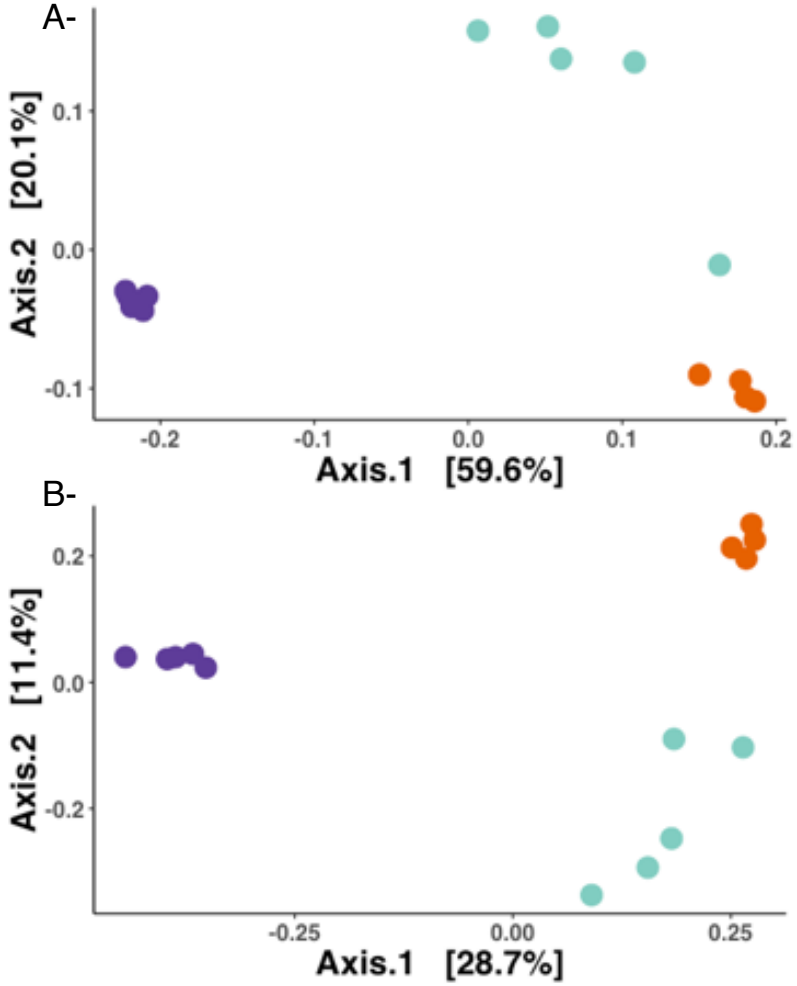


Figure 5. PCoA of five *Veturius* sp. families based on 16S rRNA sequences based on weighted (A) and unweighted (B) Unifrac. Larvae= orange, Adult= cyan, Substrate= purple.

An analysis of similarities ANOSIM was performed to statistically test the robustness of the clustering observed in Figure 5. The results are summarized in Table 3.

Table 3. Analysis of similarities of the *Veturius sp*. gut and substrate microbial communities.

Variable	R statistic	P value*
Sample type**	0.9633	0.001
Family group	-0.3393	0.994
Extraction method	0.1755	0.133

*999 permutations per test.

**Sample type= adult gut, larvae gut, substrate

These results support the clustering observed in Figure 5 as the R value obtained for sample type is 0.9633.

Being aware of the differences in the adult gut, larvae gut and substrate; I explored the uniqueness of the OTUs found in each sample type. To do this we calculated the number of unique and shared OTUs between sample type. Finally, we applied the statistical test Metastats to determine the number of OTUs that were differentially abundant in between two sample types (Table 4). From the total 11 712 OTUs in the analysis, the substrate was the sample type with the most unique OTUs with 5230, doubling the number of unique OTUs found in the adult gut. The larvae showed the lowest unique OTUs with 1981. The larvae, adult and substrate share only 360 OTUs (3.07%) (Supplementary Figure 1). The pairwise comparison between adult and larvae to the substrate showed that the digestive tracts of adults and larvae shared a low number of OTUs with the substrate, with the larvae being the lowest with 79. The adult gut and the larvae gut shared the most OTUs with 1109. The Metastats analysis also showed that adult and larvae gut had the lowest

value of differential OTUs with 628; the adult and the substrate had 1198 differential OTUs, and larvae and substrate had 1421 of these OTUs (Table 4).

In summary, the analysis showed that the microbial communities in the guts of the adult, the larvae, and the substrate of *Veturius* sp., are different in terms of community composition and structure.

Table 4. Distribution of OTUs in the *Veturius* sp. gut and substrate microbial communities.

	Adult	Larvae	Substrate	Shared	Different OTUs ($p<0.05$) *
Unique OTUs	2526	1981	5230		
Adult - Larvae				1109	628
Adult -Substrate				427	1198
Larvae - Substrate				79	1421
All				360	

*Different OTUs calculated with Metastats.

Total OTUs= 11712 (97%)

7.2.2 Phylogenetic distribution of the microbial communities in *Veturius* sp.

The length and quality of our amplicon sequences, and the universality of the primer set used to sequence our samples, allows the use of the reads to assess the distribution of the sequences at the kingdom level. The communities in the Passalid beetle were dominated by bacteria, however, archaeal sequences were abundant in our dataset, especially in the larvae samples (average= 18% ±3), followed by the adults (average= 3.5% ± 3.5) (Figure 6). Interestingly, archaeal sequences were barely detected in the substrate samples (0.08% ± 0.1), indicating an enrichment in the larval

gut environment. A 15.28% of the sequences passed the quality filtering but didn't obtain a taxonomic classification when compared to the sequences available in the greengenes database (Fig 6). These sequences probably belong to novel taxa with no current known close relative in the tree of life. Even though we consider these unclassified sequences interesting, we did not perform any further classification efforts, however, they were not removed from downstream analyses and were summarized under the category "unclassified" in our phylogenetic distributions.

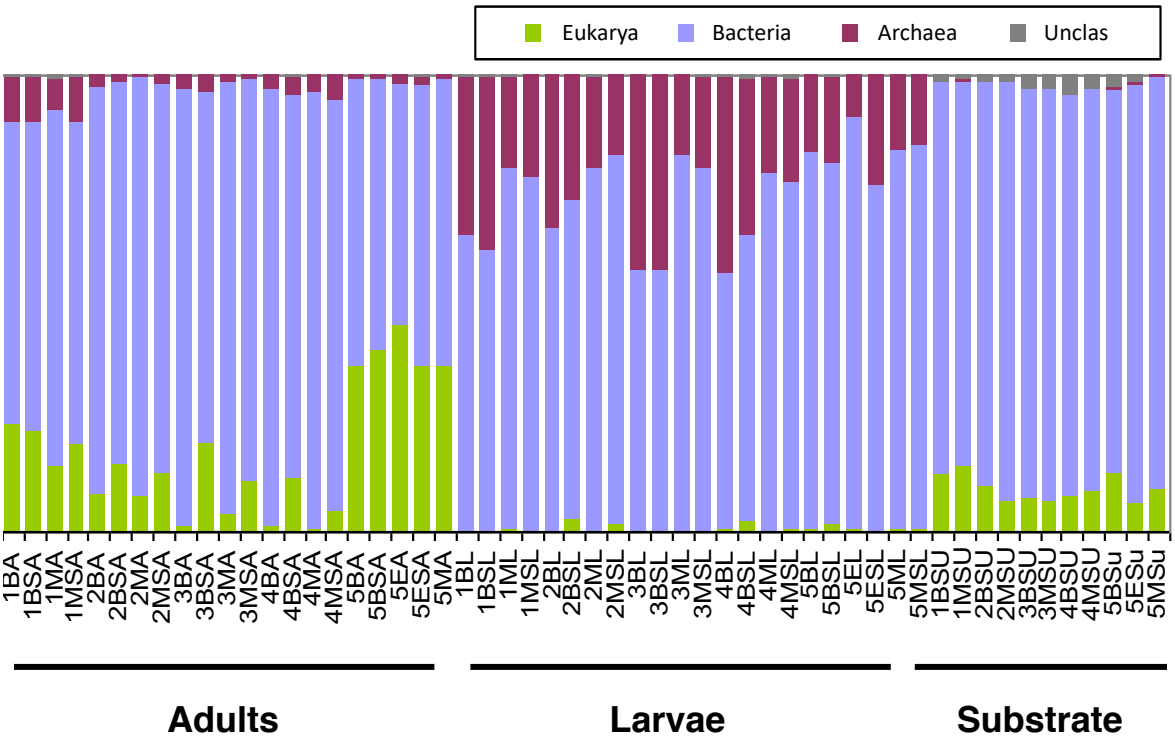


Figure 6. Phylogenetic distribution at the kingdom level of *Veturius sp*. microbial communities. Sample codes as in Table 1. Epithelium samples were not considered in this analysis.

The OTUs in the microbial communities were taxonomically classified to evaluate their distribution at the phylum level, and facilitate the identification of high order differences between

the microbial communities. Similar to other environmental microbiomes, the substrate samples are dominated by members of the phylum Proteobacteria (48.0% ± 4.7), followed by Bacteroidetes (9.6% ± 3.5), Acidobacteria (7.8% ± 4.5) and Verrucomicrobia (6.2 ± 2.8). On average 14.9% of sequences in the substrate samples were not assigned to any known taxa (Figure 7). The larvae microbial communities were dominated by Firmicutes (46.9% ± 6.3), followed by the archaeal phylum Euryarchaeota (18.7% ± 3.2) and the Proteobacteria (8.6% ± 2.8). In the larval communities, 6.74% of the sequences could not been assigned to any known taxonomic group. Finally, in the adult microbial communities the most predominant groups were the Firmicutes (23.6% ± 5.0) and the Tenericutes (23.9% ± 11.9), followed by the Proteobacteria (9.2% ± 4.9). The adults showed the most unassigned sequences with 24.2% ± 12.0, suggesting that these samples harbor the largest diversity of novel taxa.

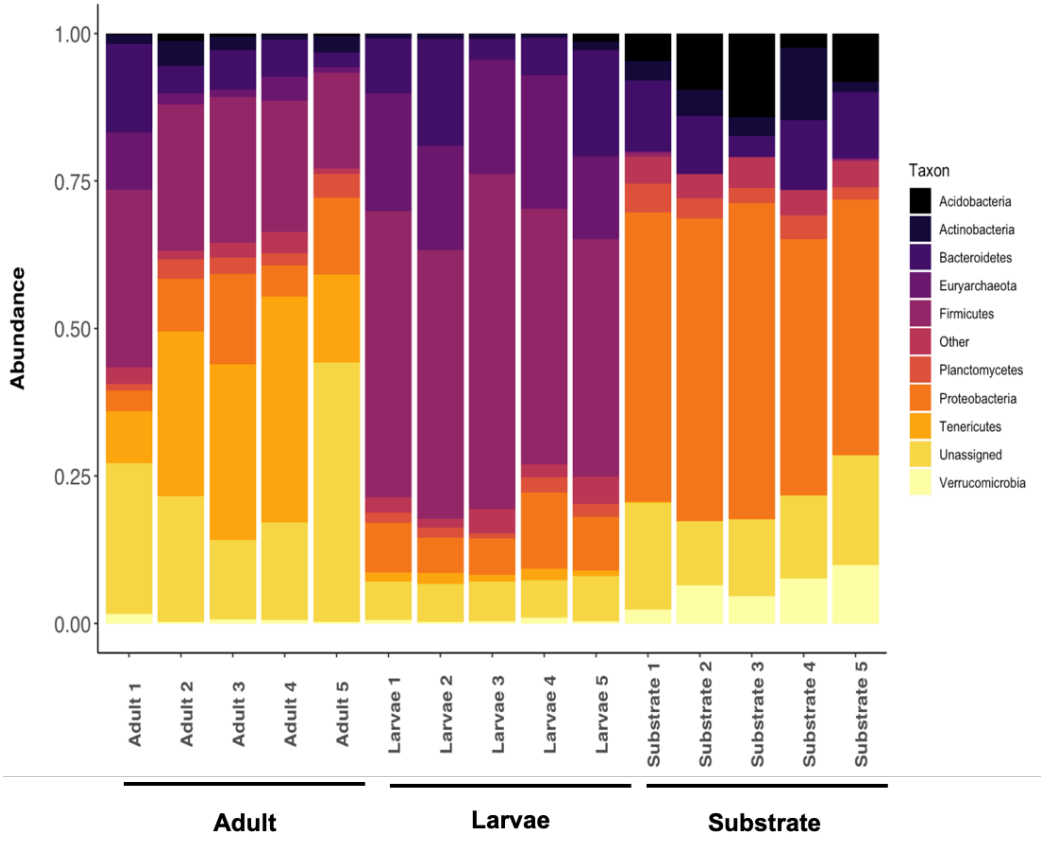


Figure 7. Phylogenetic distribution at the phylum level of *Veturius* sp. microbial communities. Sample codes as in Table 1.

The phylum Firmicutes, which includes members of families as Clostridiaceae, Lachnospiraceae and Ruminococcaceae, is the most abundant in the *Veturius* sp., especially in the larvae where it constitutes nearly half of the bacteria found in their gut. Besides the already mentioned presence of archaeal sequences, most predominantly Euryarchaeota, another unexpected group found to be abundant in the system is the Tenericutes. The physiological involvement of these groups in the physiology and ecology of the Passalids will be discussed later. The phyla Firmicutes (23.0% in adults, 46.9% in larvae, 0.3% in substrate), and Euryarchaeota (3.6% in adults, 18.7% in larvae, 0.1% in substrate) were abundant in the adult and larvae guts compared to the substrate. The

Tenericutes were abundant in the adults (24.0%) but presented low abundance in larvae (1.5%), and substrate (0.1%). The heatmap in Figure 8 shows the abundance of sequences from these three phyla at the family level. Sequences belonging to the several families showed in the heatmap are mostly absent (<1%) in the substrate and highly present in the larvae and adults. The clustering at the bottom of the heatmap also shows that the diversity in the sequences from the three phyla is enough to group the three different communities in discrete clusters (Bray Curtis similarity). These results agree with the hypothesis that the larvae and adults Passalids harbor specialized microbial communities.

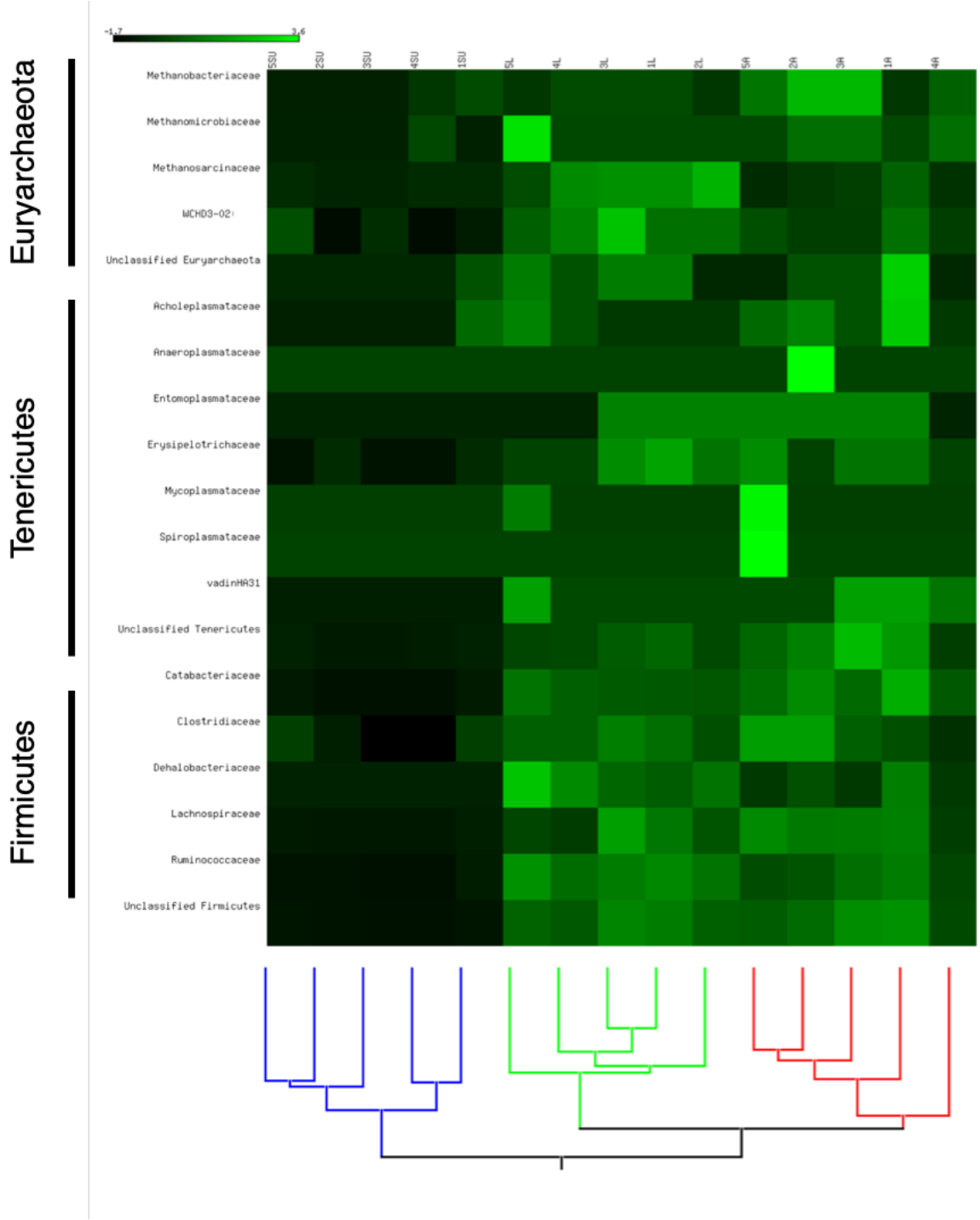


Figure 8. Relative abundance of families from the phyla Firmicutes, Tenericutes and Euryarchaeota in the *Veturius* sp. gut and substrate microbial communities. Cluster colors: Blue=

Substrate, Green= Larvae, Red= Adult. The sample clustering shown at the bottom was based on Bray Curtis distances.

A high taxonomical diversity was found in the dataset. The most dominant phylum in the adult and larvae communities was the Firmicutes. The tree in Supplementary figure 2 shows that the majority and the most abundant Firmicutes OTUs in the dataset were Clostridia (triangles). This figure also shows the lower abundance and diversity of Firmicutes in the substrate. To improve the resolution of the results in Supplementary figure 2 (too busy given the large amount of Firmicutes OTUs in our samples), I then filtered the OTUs to keep the ones that have a minimum of five reads in at least 25% of the samples (Figure 9). After filtering, all remaining OTUs were classified in the order Clostridia; most in the families Ruminococcaceae, and Lachnospiraceae as the second family in abundance. Isolates from these families are recognized cellulose degraders and have biotechnological relevance. The most abundant OTUs were commonly detected in both larvae and adults, and very sparsely in the substrate. Only two Ruminococcaceae OTUs that passed the filtered were found in the substrate sample, similar to seven OTUs of Lachnospiraceae that were also detected in the substrate.

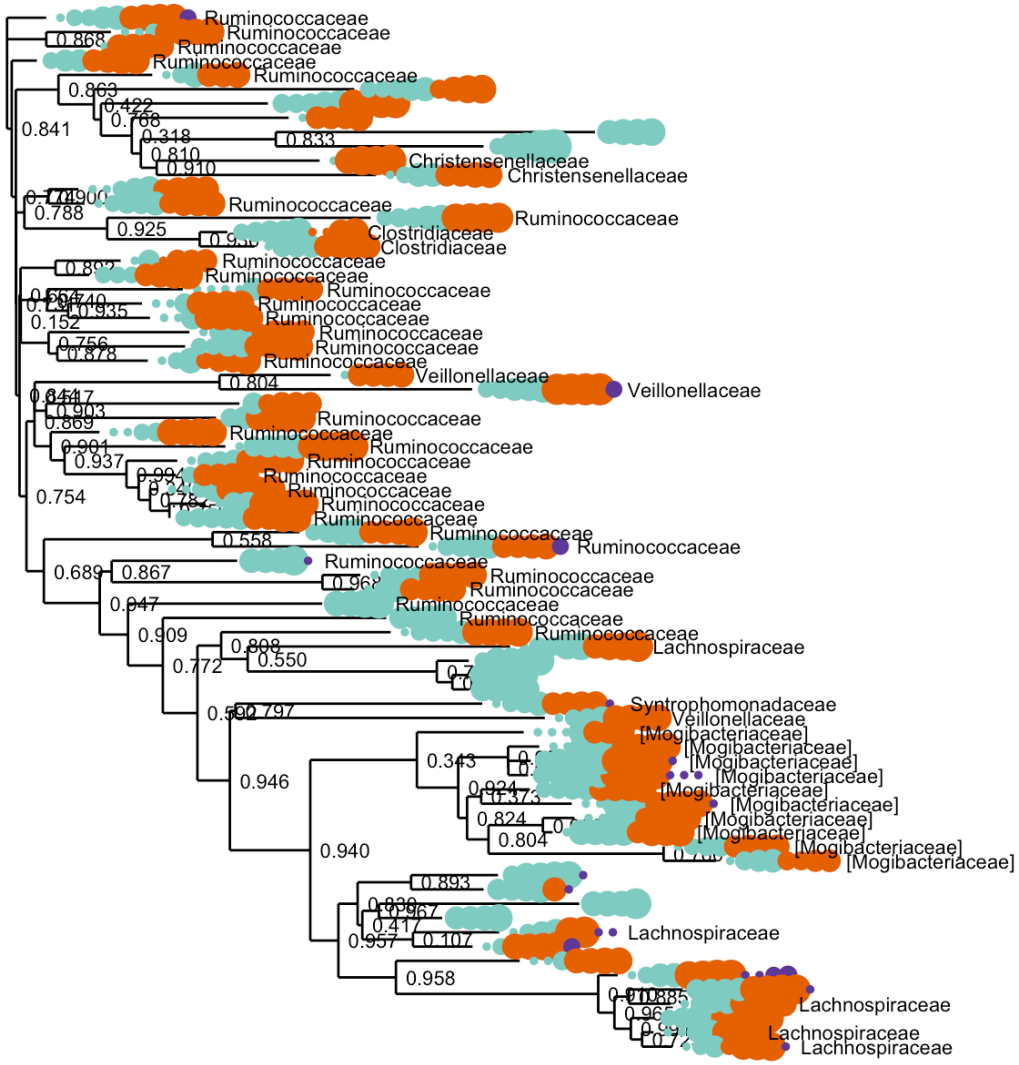


Figure 9. Neighbor joining tree of filtered Firmicutes OTUs in the *Veturius sp*. gut (Larvae= orange, Adult= cyan,) and substrate (purple). Only OTUs with at least five reads in >25% of the samples are shown. The size of the circle correlates with the abundance of the OTU.

OTUs from the order Bacilli were found in both adult and larvae guts, however they showed a higher diversity and abundance in the adult gut. Similar to the results presented in Figure 9, the

order Bacilli was underrepresented and sparse in the substrate samples. None of the Bacilli OTUs passed the data filtering threshold described above (Supplementary Figure 3).

The Euryarchaeota were also abundant, especially in the larvae (18.73%), and to a lesser extent in the adult samples. Supplementary Figure 4 shows the large diversity of sequences belonging to this archaeal phylum. The distribution of Euryarchaeotal sequences in the three sample types resemble the one observed for the Firmicutes, being which were scattered and underrepresented in the substrate. This figure also shows the higher abundance of archaea in the larval gut. The archaeal class of methanogens Methanomicrobia showed the most diversity and abundance, this class includes the orders Methanosarcinales and Methanomicrobiales which are strictly carbon reducing methanogens. Next, I applied the same filtering to the Euryarchaeota OTUs (Figure 10) and obtained seven OTUs, including six similar to known methanogenic archaea. Four OTUs were classified as Methanimicroccocus, family Methanosarcinales. The other OTU was classified as vadinCA11, which corresponds to a single uncultured lineage in the archaea clade that has been previously detected in rumen and anaerobic digestors using molecular methods. It is noteworthy that, even at very low abundance, four of the seven OTUs were detected in substrate samples.

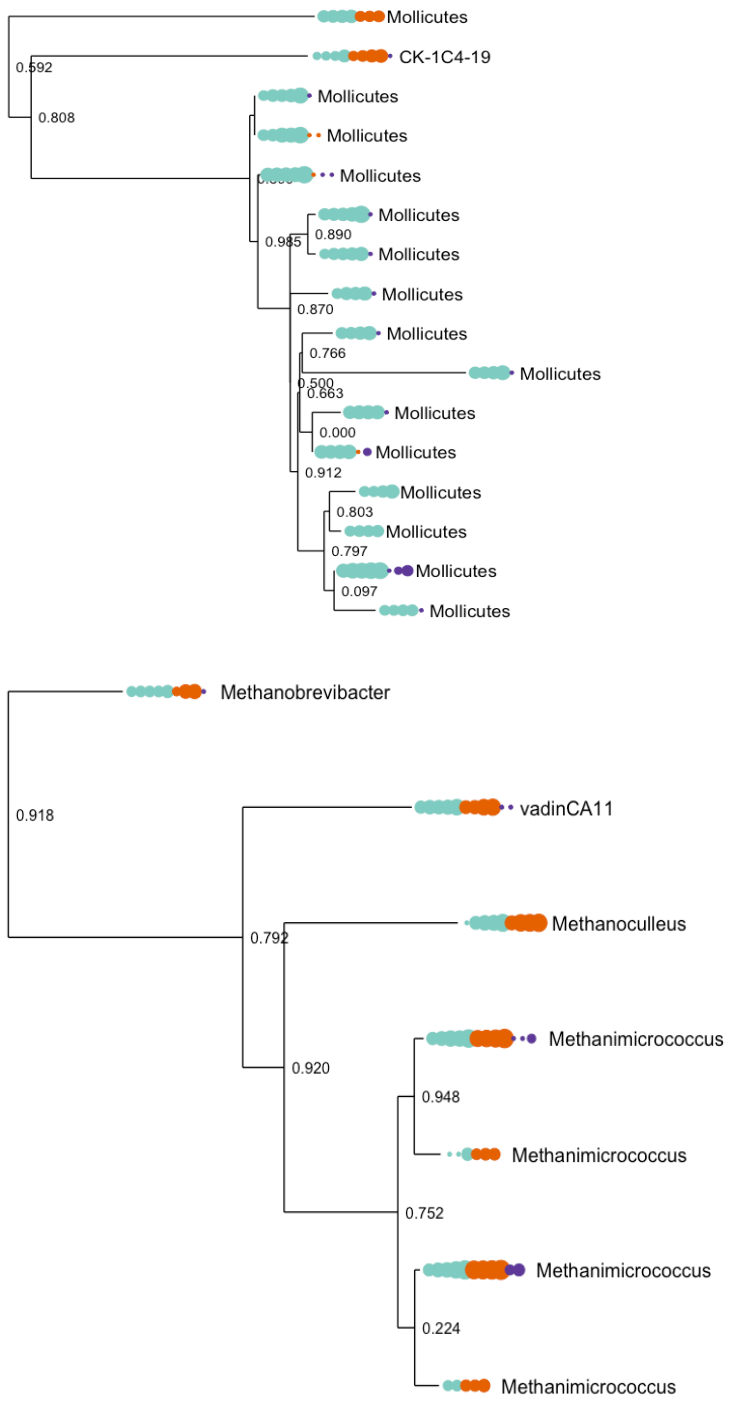


Figure 10. Neighbor joining trees of parsed of Tenericutes and Euryarchaeota OTUs in the *Veturius* sp. gut and substrate (Larvae= orange, Adult= cyan,) and substrate (purple). Only OTUs with at

least five reads in >25% of the samples are shown. The size of the circle correlates with the abundance of the OTU.

Finally, Figure 10 and the Supplementary Figure 5 shows the diversity of Tenericutes (known intracellular parasites with small genomes) and the high abundance of OTUs of this phylum in the adult samples. A total of 16 OTUs (15 of which were classified as Mollicutes and the other in the group CK-1C4-19) presented at least 5 reads in 25% of the samples. These Mollicutes OTUs were poorly detected in the larvae and substrate samples.

In summary, the community survey performed showed that the communities in the larval gut, adult gut and the substrate are different; and that the major driver of this differences is the sample type over family group and other important environmental variables. Our results were consistent over five different family groups.

7.3 The microbiomes of *Veturius* sp. include organisms with the genetic capacity of cellulose breakdown and other carbon transformations

7.3.1 Metagenomic sequencing of the *Veturius* sp. gut and substrate

The results obtained in the community composition analysis showed that larvae and adult guts are diverse and differentially enriched in microorganism from the phyla Firmicutes and Euryarchaeota, the latter being especially abundant in the larvae gut. These results, combined with the expectation that the larvae is more efficient degrading cellulose, motivated us to invest most of our efforts in deeply sequencing the larvae samples. Therefore, the five larvae samples were

sequenced in individual Illumina HiSeq 2000 lanes. The remaining two lines in the run were used to sequence a pooled DNA samples of the adults from families 3 and 4, and a pool of DNA from the substrate also from families 3 and 4. The results obtained for the metagenomic sequencing and the assembly and annotations performed in the JGI platforms are summarized in Table 5.

Table 5. Summary of the metagenomes of Passalid beetle and substrate.

	Larvae 1	**Larvae 2**	**Larvae 3**	**Larvae 4**	**Larvae 5**	**Adult**	**Substrate**
Metagenome size (Kb)	511 654	52 750	804 003	402 414	77 567	142 922	170 323
Gene count	1 063 212	125 251	1 802 558	832 561	206 349	425 036	433 068
CDS count *	1 051 900	123 727	1 786 624	823 392	204 189	424 225	428 463
CDS % *	98.94	98.78	99.12	98.9	98.95	99.81	98.94
Function Prediction %*	43.04	50.56	39.37	53.29	36.14	57.5	43.57
No function prediction %*	55.9	48.22	59.75	45.61	62.81	94.06	55.37
Orthologs %	0	0	0	0	0	0	0
COG %*	47.87	46.25	42.2	49.15	43.4	6.02	49.15
KOG %	0	0	0	0	0	0	0
Pfam %*	48.13	44.64	42.48	48.72	39.73	6.25	46.56
TIGRfam %*	10	0	0	0	0	0	7
Enzyme %*	21.21	19.83	17.75	20.27	18.84	2.67	23.29
KEGG % *	21.91	21.02	18.61	21.41	19.63	3.14	24.98

Not KEGG %*	77.02	77.76	80.5	77.49	79.32	96.67	73.96
MetaCyc %*	14.09	13.4	11.96	13.68	12.5	1.29	16.39

* Assembled data

The data showed a wide range of metagenomic reads obtained from each sample. The most data were obtained for the Larvae 3 followed by Larvae 1 and Larvae 4, with almost 500Mb. The adult and substrate pools produced 143 Mb and 170 Mb, respectively. The poorest samples in terms of data produced were Larvae 2 and Larvae 5. On average we obtained more than 140 million genes in the seven metagenomes from which more than 98% were protein coding sequences. The percentages of pseudogenes and dubious sequence calls are negligible in the dataset. One interesting, though difficult to overcome, feature of the data is that only 42.3% of the genes have predicted functions. Compared to other functional databases as TIGRfam and KEGG, the COG and Pfam annotations produced the highest percentage of hits (ca. 49%). The Larvae 5 metagenome presented the lowest number of gene annotations.

7.3.2 Genome reconstruction of dominant taxa

The 16S rRNA amplicon sequence analysis of the microbial communities related to *Veturius* sp, revealed that they were diverse and enriched in bacteria and archaea. In order to explore the genetic potential of these organism and further understand their possible ecological roles, I attempted to reconstruct the genomes of the most abundant taxonomical groups using the metagenomic reads. The co-assembly of the metagenomes of the Larvae 3, Larvae 4 and the pool metagenomes of the adults 3+4 and substrate 3+4 produced 1 161 239 contigs larger than 1kb containing a total 2 .734 713 841 bp. The largest contig was 412 893bp long (average contig size= 2 355bp), and the N50

was 2 600bp. All the metagenomic reads were mapped back to the co-assembly to generate coverage maps for each contig (Table 6). The contigs were then binned and manually curated to create comprehensive genome bins. The final collection consisted in 766 bins accounting for 1 022 999 923 nucleotides, which represent 37.41% of all nucleotides stored in the contigs database, and 72.61% of nucleotides stored in the profile database. From the total 766 bins, 391 were larger than 1Mb, from which 101 were >70.00% completion and <7.00% contamination, these bins were then classified, and further refer to those as MAGs (Table 7). The summary of the entire collection of bins is in Supplementary table 2.

Table 6. Number of metagenomic reads mapped to the co-assembly.

Metagenomic sample	Mapped reads
Adult	96 407 495
Larvae 3	262 766 215
Larvae 4	161 378 257
Substrate	138 542 298

The clustering in Figure 11 shows the distribution of the bins and MAGs in the four co-assembled metagenomes based on the coverage of the reads mapped to the bins. Most of the bins were mapped back to the larvae metagenomes, however, only a percentage is present in both larvae, the majority of the bins were found in one or the other larvae. Finally, the substrate has a distinct group of bins that are not represented in the beetles. The clustering on top of the metagenomes was calculated

using the differential coverage of bins that were found across the samples (Supplementary Table 2). The clustering shows the substrate as an outgroup presenting a different set of bins than the beetle guts. As expected, the adult and larvae are grouped in the other clade with the larvae next to each other. These results are consistent with our community composition analysis based on 16S rRNA sequences.

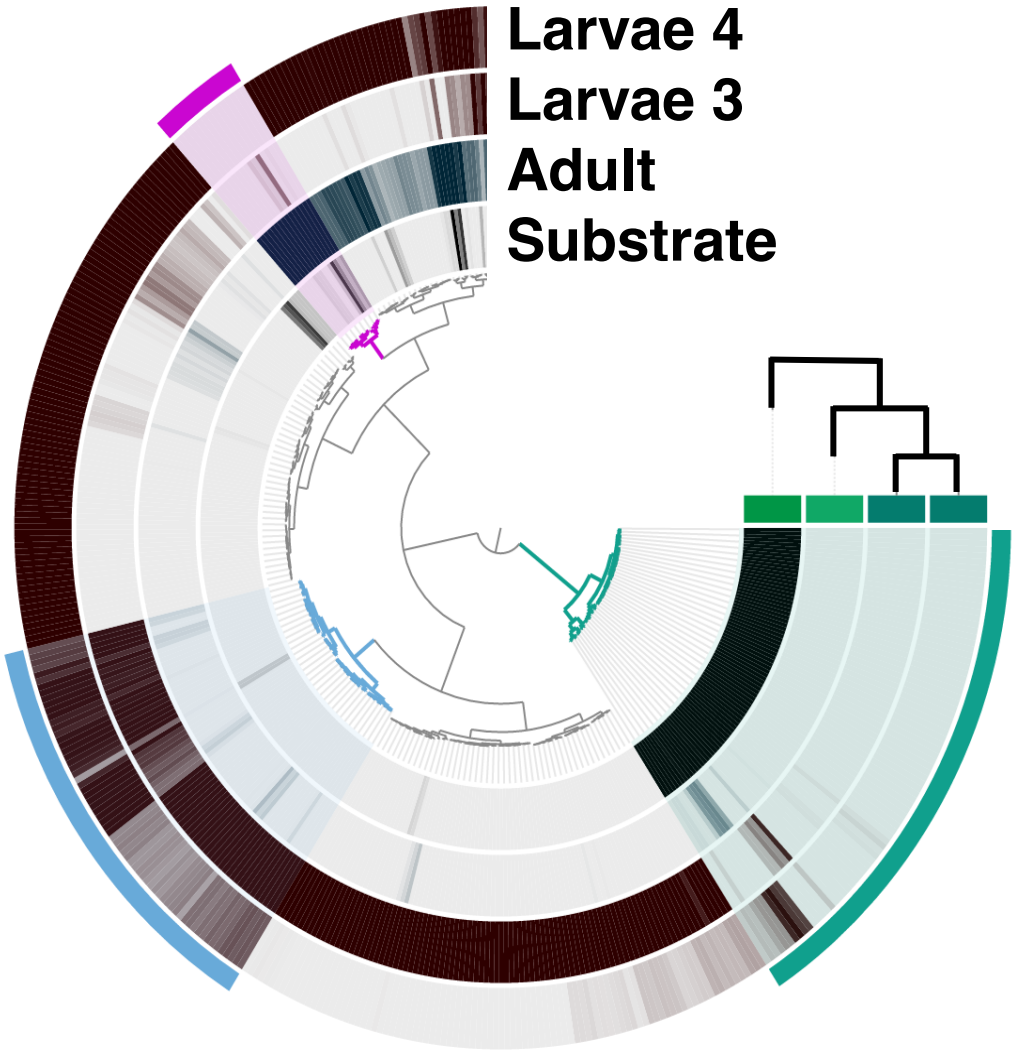


Figure 11. Metagenomic community analysis based on 766 bins. Pink branch: metagenomic bins dominant in adult gut. Green branch: metagenomic bins enriched in substrate Blue branch: metagenomic bins found in both larvae samples. Non-colored branch: metagenomic bins dominant in only one larvae sample. The intensity of black of each bin was calculated using Maximum normalized ratios being grey= absent, and black= 100% of bin present in the sample.

Table 7. Summary of 101 MAGs reconstructed from the *Veturius* sp. gut and substrate metagenomes.

Bin	Length (Mb)	contigs	N50	%GC	%completion	%redundancy
Bin 78	4.059491	125	56286	43.04	99.28	0.72
Bin 315	1.643355	38	59693	27.53	99.28	0.72
Bin 340	1.785043	152	17065	51.08	99.28	2.88
Bin 552	2.254344	143	23041	33.34	99.28	4.32
Bin 178	4.475214	99	71984	36.18	97.84	1.44
Bin 438	1.078999	23	77003	38.56	97.12	0.00
Bin 57	4.487127	61	102697	46.02	97.12	0.72
Bin 322	2.224147	222	13019	34.83	97.12	0.72
Bin 191	2.639116	114	33214	52.88	97.12	2.16
Bin 470	2.477736	159	20932	50.18	97.12	2.16
Bin 11	4.077	96	75577	53.18	97.12	3.60
Bin 127	2.762037	90	55910	50.58	97.12	3.60
Bin 505	2.24753	178	17326	41.76	97.12	3.60
Bin 74	1.828126	45	67879	38.90	96.40	2.88

Bin 330_1	1.501955	174	11303	25.93	96.40	6.47
Bin 167	2.969866	28	176920	60.29	95.68	1.44
Bin 485	1.824596	187	13368	32.12	95.68	3.60
Bin 124_1	2.792969	247	17612	35.41	95.68	6.47
Bin 41	2.140966	150	21377	58.30	94.96	0.72
Bin 445	4.87312	112	79576	41.36	94.96	3.60
Bin 397_2	4.254739	312	20569	60.41	94.96	3.60
Bin 405	1.727392	110	25836	27.27	94.96	5.76
Bin 317_1	1.478132	129	15668	41.40	94.96	7.19
Bin 184	2.122613	212	13166	53.06	94.24	1.44
Bin 195	2.071625	131	25436	41.30	94.24	1.44
Bin 270	1.728146	114	22373	52.55	94.24	1.44
Bin 553	1.111015	44	36367	39.27	93.53	0.00
Bin 163	2.294155	88	41857	55.01	92.81	2.16
Bin 123_1	1.928882	160	18352	37.79	92.81	2.16
Bin 316	1.264829	76	21964	29.40	92.81	2.88
Bin 526	2.034256	139	24169	46.54	92.81	6.47
Bin 407	5.446711	295	25761	60.09	92.09	0.00
Bin 503	2.145093	153	18626	54.07	92.09	0.72
Bin 284	3.116704	179	24358	57.57	92.09	1.44
Bin 461	2.90979	163	26579	68.61	92.09	1.44
Bin 361	1.817726	157	14407	56.50	92.09	1.44
Bin 174	3.168773	265	15063	49.08	92.09	2.88
Bin 260	2.163051	60	66296	42.95	91.98	6.79

Bin 379	1.284114	74	24513	45.65	91.37	0.00
Bin 266	3.625865	129	40612	38.88	91.37	0.72
Bin 4	2.107329	99	36416	53.04	91.37	2.16
Bin 211	3.547183	151	39950	64.07	90.65	3.60
Bin 511	2.231954	178	15218	51.16	89.93	2.16
Bin 363	4.425913	391	14809	57.77	89.21	2.16
Bin 398	1.667669	218	8935	54.92	89.21	3.60
Bin 433	1.641575	162	12858	50.03	88.49	0.72
Bin 106	1.713734	173	11847	53.76	88.49	1.44
Bin 64_1	1.442541	93	20121	28.34	88.49	3.60
Bin 399	2.504482	95	56972	34.57	88.49	5.04
Bin 230	2.176001	279	9680	37.48	87.77	0.72
Bin 440	1.745623	175	13785	46.15	87.77	0.72
Bin 442	1.342311	61	35512	25.41	87.05	2.88
Bin 518_1	1.669738	192	11340	37.38	87.05	4.32
Bin 72	2.02646	266	8858	48.10	86.33	3.60
Bin 89	3.267142	65	90990	49.90	85.61	2.16
Bin 113	1.33199	156	10480	39.65	85.61	2.16
Bin 58	1.353409	196	7932	49.22	84.89	0.72
Bin 225_1	3.670433	192	27043	60.04	84.89	3.60
Bin 201	3.020484	168	26512	52.71	83.45	0.00
Bin 14	1.573518	158	12697	49.53	83.45	2.16
Bin 161	2.601191	272	13020	59.77	82.01	2.16
Bin 85	1.307863	75	31481	27.15	81.29	0.00

Bin 446	1.362137	236	6395	53.30	80.58	0.00
Bin 481	2.634059	170	22640	51.76	80.58	1.44
Bin 250	3.029708	284	14034	66.90	80.58	2.16
Bin 16	2.897163	381	9874	58.11	80.58	2.88
Bin 155	2.586332	230	14613	37.93	80.58	2.88
Bin 272	2.814037	148	24555	40.70	79.86	1.44
Bin 208	2.241333	131	25527	51.32	78.42	1.44
Bin 232_1	1.210797	114	13147	30.71	78.42	1.44
Bin 86	2.251322	115	26149	55.34	78.42	2.16
Bin 172	1.152459	125	10543	40.44	78.42	2.16
Bin 549	1.920274	206	10868	49.22	78.42	2.88
Bin 143	2.059154	370	6147	37.92	78.42	5.76
Bin 96	2.067989	67	54941	41.82	77.70	0.72
Bin 372	4.351562	403	14336	53.03	76.98	0.72
Bin 502	3.970695	170	38348	62.04	76.98	2.88
Bin 536	1.582449	368	4308	40.22	76.26	2.88
Bin 288	2.455874	215	16120	54.01	75.54	3.60
Bin 223	2.83604	342	10232	54.32	75.54	5.04
Bin 133	1.616156	103	20840	54.85	74.82	0.72
Bin 206	2.648174	335	10027	39.04	73.38	2.16
Bin 193	2.17889	175	16450	51.94	73.38	2.16
Bin 217	2.841096	220	18244	51.70	73.38	2.88
Bin 522	1.455003	116	17665	33.58	72.66	0.00
Bin 202	2.084695	381	6004	39.54	72.66	0.72

Bin 23	1.350307	75	29196	27.68	72.66	0.72
Bin 519	3.346523	450	9284	43.05	72.66	1.44
Bin 406_1	1.281203	184	8045	27.86	72.66	2.88
Bin 301	1.750208	136	25956	38.19	72.66	4.32
Bin 248	1.570354	193	10065	53.25	71.94	2.16
Bin 508_1	3.793885	380	12401	61.88	71.94	2.88
Bin 17	1.546744	279	6096	35.60	71.94	3.60
Bin 308	1.494432	216	8487	31.50	71.22	2.88
Bin 139_2	3.240482	442	9080	47.81	71.22	4.32
Bin 83	2.531595	254	14000	42.63	71.22	4.32
Bin 357	1.191913	150	9727	52.93	70.50	0.00
Bin 134	1.21814	195	7104	49.94	70.50	0.72
Bin 24	1.67139	148	15144	42.48	70.50	1.44
Bin 441	1.621338	290	6028	54.30	70.50	1.44
Bin 207_1	2.497164	213	15131	48.04	70.50	3.60

The histogram in Figure 12 shows the length distribution of the 766 bins in the collection. The average bin size obtained was 1.3Mb. Bin length, % completeness and %redundancy were proxies used to defined the MAGs and to estimate the quality of the rest of the metagenomic bins. The dominant members of the community (with some exceptions due to methodological limitations) present the best metagenomic bins because these are more deeply covered. In our dataset, 391 bins were larger than 1Mb, 235 were between 0.3-1Mb, and only 141 were smaller than 300 000bp. Furthermore, the average number of ORFs obtained was 1 290 per bin. The other 290 genomic

bins that did not pass the cutoff for MAG but were larger than 1Mb were no considered for further analysis; these bins included many genes and genomic context that are valuable to explore.

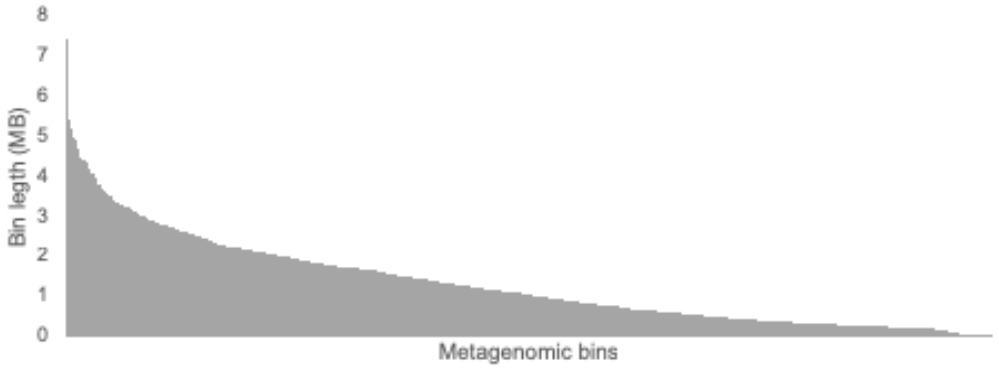


Figure 12. Histogram of the lengths in Mb of the reconstructed bins from *Veturius* sp. gut and substrate metagenomes.

Another important proxy to consider in a collection of bins is the GC content. One can hypothesize that in a given environmental sample that shows large microbial diversity, the expected average %GC of the community will be around 50%. The average %GC of our collection was 46.52%, presenting a slight bias towards the low GC content.

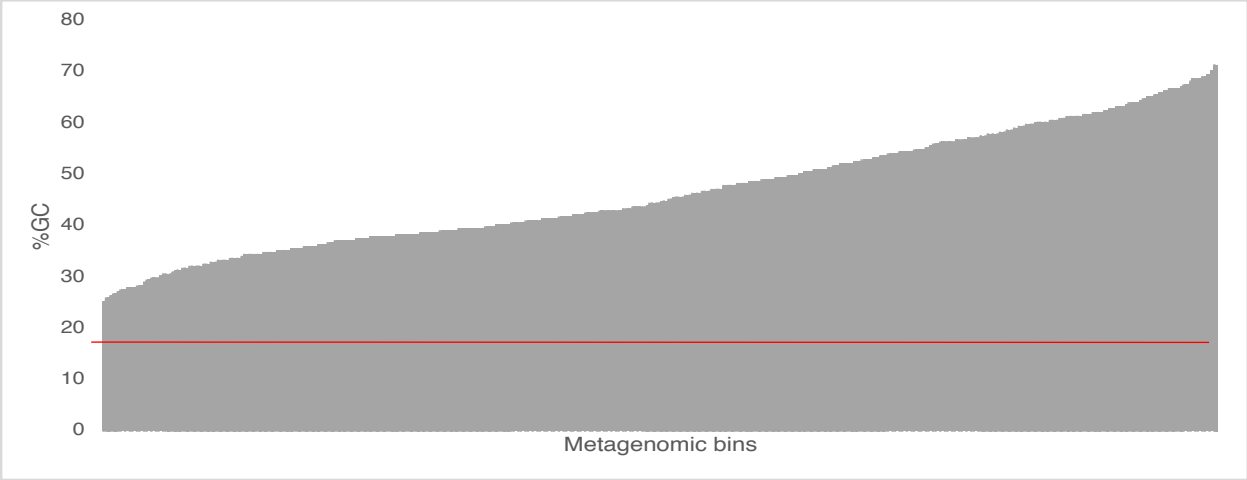


Figure 13 %GC of the reconstructed bins from *Veturius* sp. gut and substrate metagenomes. Red dashed line represents the average %GC of the collection= 46.52.

The collection of metagenomic bins reconstructed here are large, have low redundancy and contain thousands of genes. The further analysis is focused on the 101 MAGs and in the archaeal bins. These comprise the dominant microorganism in the system but also represent novel taxonomical groups from which little is known about their genome content and functional capacities. Only one of the bins that we classified as archaeal met our definition of MAG in terms of redundancy or completion, the reasons why this happened are discussed later. However, these archaeal bins are large and high-quality bins that can be analyzed.

From the 101 MAGs, 78 were present in the larvae samples, 15 in the substrate and 9 in the adult. It is important to consider that all the reads from the four metagenomic samples were included for the co-assembly, therefore it is possible for any MAG to contain contigs that have good coverage in all samples if this microorganism was present and abundant in all four samples. The distribution

of the MAGs in the four metagenomes as well as its approximated taxonomy is summarized in Table 8. The anvio pipeline includes the taxonomic classification of bins by comparing the set of single copy core genes found in a given bin to the corresponding homologs in the MG-RAST reference database of genomes. This method was able to classify only 21 of our 766 bins. This suggests that our bins belong to novel organisms that are not represented in the current curated databases.

Several strategies were used in order to assign taxonomy to the 101 MAGs. First, kmer base algorithms like FastANI were used to compare the MAGs to the genomes in the refseq reference genome database. This method failed to classified all the MAGs with the exceptions of two: Bin 399 classified as *Lactococcus lactis* (ANI>97%) and Bin 299 as *Limnobacter* sp. (ANI>97%). The rest of the MAGs presented ANIs below than 90% when compared to the GenBank database, suggesting again the novelty of these MAGs. Similar results were obtained with Kraken and Centrifuge. Then, a custom strategy was designed to rank the MAGs to an approximate taxonomy by comparing the set of single copy core genes from each MAG to the genes in the KBase curated database of bacterial and archaeal genomes. The taxonomic information of the top hit to the database was extracted for each gene in the MAG, and tabulated to manually determined the highest taxonomical rank that was consistent in the majority of the genes in the MAG (Table 8).

Table 8. Normalized ratio of the MAGs in the metagenomic samples and their approximate taxonomic classification.

MAG	Adult*	Larvae 3*	Larvae 4*	Substrate*	Approximated taxonomy**
Bin 407	0.00	0.00	0.00	1.00	Actinobacteria (27/47)
Bin 398	0.00	1.00	0.00	0.00	*Alistipes* sp. (66/82)
Bin 143	1.00	0.00	0.00	0.00	Bacillales (37/63)
Bin 553	0.27	0.00	1.00	0.00	Bacillales (22/41)
Bin 106	0.00	0.03	1.00	0.00	Bacteria (47/47)
Bin 139_2	0.00	0.00	1.00	0.00	Bacteria (18/18)
Bin 16	0.00	0.00	1.00	0.00	Bacteria (63/63)
Bin 167	0.00	0.00	0.00	1.00	Bacteria (70/70)
Bin 17	0.00	0.11	1.00	0.00	Bacteria (27/29)
Bin 172	0.00	1.00	0.00	0.00	Bacteria (29/30)
Bin 207_1	0.00	1.00	0.00	0.00	Bacteria (39/39)
Bin 217	0.00	0.00	1.00	0.00	Bacteria (48/48)
Bin 248	0.00	1.00	0.00	0.00	Bacteria (39/39)
Bin 260	1.00	0.00	0.00	0.00	Archaea (4/4) ***
Bin 284	0.08	0.06	1.00	0.00	Bacteria (41/41)
Bin 379	0.00	0.00	0.01	1.00	Bacteria (34/36)
Bin 433	0.00	0.00	1.00	0.00	Bacteria (49/49)
Bin 446	0.00	1.00	0.00	0.00	Bacteria (49/49)
Bin 504	0.00	1.00	0.00	0.00	Bacteria (49/49)
Bin 536	1.00	0.00	0.06	0.00	Bacteria (29/29)
Bin 549	0.00	0.00	1.00	0.00	Bacteria (46/46)

Bin 74	0.00	0.00	1.00	0.00	Bacteria (48/48)
Bin 78	0.00	0.00	0.00	1.00	Bacteria (80/81)
Bin 85	0.00	0.00	1.00	0.00	Bacteria (45/46)
Bin 89	0.00	0.00	1.00	0.00	Bacteria (59/59)
Bin 83	0.00	1.00	0.01	0.00	Bacteroidetes (34/46)
Bin 57	0.00	0.00	0.00	1.00	Bacteroidetes (34/51)
Bin 441	0.00	0.00	1.00	0.00	Betaproteobacteria (56/66)
Bin 211	0.01	0.00	0.03	1.00	*Chthoniobacter flavus* Ellin428 (32/87)
Bin 397_2	0.00	0.00	0.00	1.00	*Chthoniobacter flavus* Ellin428 (43/88)
Bin 184	0.00	0.00	1.00	0.00	Clostridia (29/68)
Bin 552	0.00	1.00	0.00	0.00	Clostridia (61/88)
Bin 123_1	0.00	1.00	0.00	0.00	Clostridiales (31/53)
Bin 124_1	0.00	1.00	0.08	0.00	Clostridiales (43/65)
Bin 127	0.00	0.00	1.00	0.00	Clostridiales (23/48)
Bin 134	0.01	1.00	0.00	0.00	Clostridiales (25/39)
Bin 161	0.00	1.00	0.05	0.00	Clostridiales (23/70)
Bin 272	0.00	1.00	0.04	0.00	Clostridiales (37/54)
Bin 322	0.00	1.00	0.00	0.00	Clostridiales (43/63)
Bin 330_1	0.00	1.00	0.00	0.00	Clostridiales (69/95)
Bin 340	0.00	1.00	0.00	0.00	Clostridiales (40/72)
Bin 361	0.00	1.00	0.00	0.00	Clostridiales (26/63)
Bin 405	0.00	0.00	1.00	0.00	Clostridiales (61/68)

Bin 41	0.00	1.00	0.27	0.00	Clostridiales (43/76)
Bin 440	0.00	1.00	0.01	0.00	Clostridiales (31/52)
Bin 442	0.00	1.00	0.00	0.00	Clostridiales (64/86)
Bin 470	0.00	0.00	1.00	0.00	Clostridiales (29/53)
Bin 481	0.00	1.00	0.00	0.00	Clostridiales (30/58)
Bin 485	0.00	0.02	1.00	0.00	Clostridiales (40/79)
Bin 505	0.00	1.00	0.00	0.00	Clostridiales (29/51)
Bin 511	0.00	1.00	0.15	0.00	Clostridiales (29/42)
Bin 518_1	0.00	1.00	0.00	0.00	Clostridiales (44/58)
Bin 72	0.00	0.00	1.00	0.00	Clostridiales (27/51)
Bin 96	0.00	1.00	0.01	0.00	Clostridiales (28/55)
Bin 406_1	0.00	0.00	1.00	0.00	Clostridium sp. (41/56)
Bin 64_1	0.00	1.00	0.00	0.00	*Clostrium* sp. (51/82)
Bin 461	0.00	0.01	0.00	1.00	*Conexibacter* sp. (49/49)
Bin 178	0.00	0.00	0.00	1.00	Cytophagales (41/80)
Bin 266	0.00	0.00	0.00	1.00	Cytophagales (30/68)
Bin 163	0.00	0.00	1.00	0.00	Deltaproteobacteria (35/57)
Bin 288	0.00	0.00	1.00	0.00	*Desulfovibrio* sp. (27/68)
Bin 202	0.20	0.00	1.00	0.00	Enterobacteriaceae (79/92)
Bin 372	0.03	0.00	1.00	0.00	Enterobacteriaceae (107/108)
Bin 232_1	0.00	0.00	1.00	0.00	Erysipelotrichaceae (22/71)
Bin 315	0.03	0.00	1.00	0.00	Erysipelotrichaceae (72/102)
Bin 14	0.00	1.00	0.10	0.00	Firmicutes (38/61)
Bin 208	0.00	1.00	0.10	0.00	Firmicutes (32/47)

Bin 223	0.00	1.00	0.01	0.00	Firmicutes (36/55)
Bin 23	0.00	0.00	1.00	0.00	Firmicutes (38/51)
Bin 230	0.00	1.00	0.00	0.00	Firmicutes (32/38)
Bin 270	0.00	0.00	1.00	0.00	Firmicutes (24/49)
Bin 308	0.00	0.03	1.00	0.00	Firmicutes (41/53)
Bin 316	0.00	1.00	0.00	0.00	Firmicutes (41/62)
Bin 4	0.02	0.00	1.00	0.00	Firmicutes (33/61)
Bin 522	0.00	1.00	0.02	0.00	Firmicutes (40/48)
Bin 526	0.00	1.00	0.07	0.00	Firmicutes (46/60)
Bin 113	0.00	1.00	0.00	0.00	Gammaproteobacteria (13/41)
Bin 438	0.00	0.00	1.00	0.00	Gammaproteobacteria (22/48)
Bin 193	0.00	1.00	0.00	0.00	Lachnospiraceae (20/51)
Bin 301	0.00	1.00	0.19	0.00	Lachnospiraceae (33/55)
Bin 155	1.00	0.00	0.06	0.04	*Lactococcus lactis* (82/101)
Bin 399	0.03	0.00	1.00	0.00	*Lactococcus lactis subsp. Lactis* (126/129)
Bin 201	0.12	0.01	0.23	1.00	*Limnobacter* sp. (105/110)
Bin 445	0.00	0.00	0.00	1.00	*Mucilaginibacter* sp. (43/90)
Bin 225_1	0.00	0.00	0.00	1.00	Rhizobiales (61/99)
Bin 133	0.00	0.00	1.00	0.00	Ruminococcaceae (25/53)
Bin 174	0.00	1.00	0.05	0.00	Ruminococcaceae (35/70)
Bin 195	0.00	1.00	0.01	0.00	Ruminococcaceae (20/48)
Bin 206	0.00	0.00	1.00	0.00	Ruminococcaceae (17/42)
Bin 24	0.00	1.00	0.00	0.00	Ruminococcaceae (44/52)

Bin 357	0.00	1.00	0.61	0.00	Ruminococcaceae (33/49)
Bin 503	0.00	1.00	0.03	0.00	Ruminococcaceae (28/79)
Bin 519	0.00	1.00	0.02	0.00	Ruminococcaceae (17/40)
Bin 58	0.00	0.03	1.00	0.00	Ruminococcaceae (26/59)
Bin 86	0.00	0.00	1.00	0.00	Ruminococcaceae (33/61)
Bin 508_1	0.00	0.00	0.00	1.00	*Solimonas* sp. (36/81)
Bin 250	0.00	0.00	0.00	1.00	*Sphingobium* sp. (59/111)
Bin 363	0.00	0.00	0.00	1.00	*Terriglobus* sp. (80/98)
Bin 11	0.00	0.01	0.00	1.00	Verrucomicrobia (30/66)
Bin 191	0.00	0.00	0.00	1.00	Verrucomicrobia (41/64)
Bin 502	0.00	0.00	0.00	1.00	Verrucomicrobia (25/62)

*Maximum normalized ratio= number of reads recruited to a contig divided by the maximum number of reads recruited to that contig in any sample.

**Approximated taxonomy was determined based on sequence similarity of single copy core genes (Campbell set of genes[68]) to the KBase genome database.

***The archaeal MAG260 was confirmed by phylogenetic analysis of single copy core genes and methanogenesis markers (below).

The phylogenetical distribution of the MAGs agrees with the dominant phyla found in the 16S rRNA analysis (Figure 14). The Firmicutes was the most represent phylum with 57 MAGs, including 27 MAGs classified in the order Clostridiales, 10 Ruminococcaceae, two Lachnospiraceae; four MAGs were classified in the class Bacilli, including two *Lactococcus,* and the rest with best classification of Firmicutes. The only Firmicutes MAGs in the adults were 1 *Lactococcus* and one Bacillales. The second group in importance were the MAGs with no known

phylum classification with 18 MAGs, followed by the Proteobacteria with 11. The Proteobacteria were distributed across the main classes with two Alpha-, two Beta-, two Delta- and five Gammaproteobacteria; only two MAGs were classified as Actinobacteria. The five Verrucomicrobia MAGs reconstructed were exclusively present in the substrate, similar to the two Actinobacteria and the only one Acidobacteria seen. The one archaeal MAG was only present in an adult sample. six MAGs were classified as Bacteroidetes, four of which were present in the Substrate and two in the larvae.

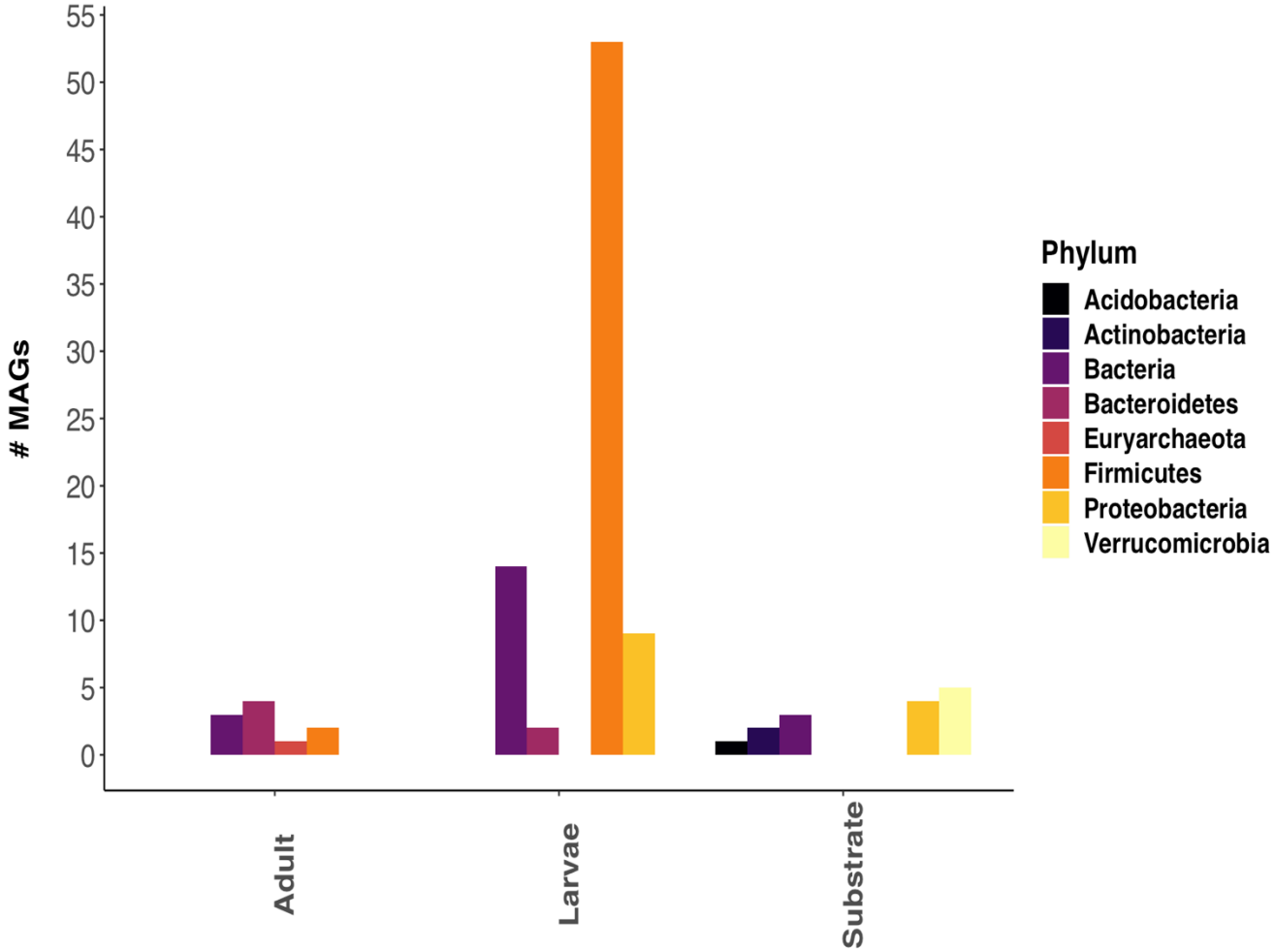


Figure 14. Taxonomical distribution of MAGs in the three metagenome sample types. Bacteria= MAGs with no classification at the phylum level.

7..3.3 Cellulose degradation in the *Veturius* sp. gut

The next step was to explore the MAGs in order to understand their potential role in the system. Passalid beetles feed only on wood material consisting mainly of lignocellulose. Thus, the first function that were searched in our MAGs were glycosyl hydrolases and other functions involved in the digestion of wood material. To do this, the ORFs predicted in the 101 MAGs were compared to the sequences in the CAZY database for functional classification. We parsed the data to keep

the top hit of every function with a cutoff e-value = 0.00001, then the genes with hits were grouped based on CAZy clans and family classification.

Data on Figure 15 show the distribution of glycosyl hydrolases (GH) clans (CAZy classification system for glycosyl hydrolases families that share a fold and catalytic machinery), glycosyl transferases (GT), carbohydrate binding modules (CBM) and carbohydrate esterases (CE) in the MAGs (these were normalized by the total number of ORFs in each MAG). Additionally, the MAGs were clustered based on the presence of these functions (Figure 16). As predicted, the MAGs are enriched in cellulose degradation and other carbon transformation functions; furthermore, I found a great diversity of genes related to this process (Supplementary Table 4). The GH-A clan presented the highest abundance among the glycosyl hydrolases, these include several families of β-glucosidases and β-galactosidases, including GH2 (1.24% of hits), GH5 (0.63% of hits), GH9 (0.12% of hits), GH 10 (0.3% of hits), GH28 (0.62% of hits) and GH30 (0.11% of hits) families of cellulases with endoglucanase (cellulase), exoglucanase and endoxylanases, between other activities. Clans GH-B and GH-H were also highly abundant in the MAGs, the former includes GH7 which cleave β-1,4 glycosidic bonds in cellulose and also has xylanase activity, and GH16 that consists of a large variety of hydrolases that are active on plant polysaccharides. The GH-H clan comprises GH13, which shows activity on polysaccharides such as amylose, and the amylopectin and the transglucosylases GH70 and GH77. The clan GH-K showed intermediate abundance in the MAGs, this contains the family GH18 of chitinases. GH43 was abundant within the MAGs (GH-F) representing 1.17% of the hits.

All the MAGs also showed significant presence of GTs, CEs and CBMs (Supplementary Table 4). GTs are expected to be abundant as they form the glycosidic bond, taking part in the biosynthesis

of oligo and polysaccharides; these molecules are metabolically important for biomass and energy production. CEs were consistently abundant among the MAGs, especially in the Firmicutes. CEs are interesting as they have been shown to catalyze the hydrolysis of ester bonds in hemicellulose and pectin, two of the more recalcitrant molecules in the plant cell wall. CBMs were also abundant across all MAGs, however, they showed higher abundances in terms of genes and sequences in the larval MAGs (Figure 15 and Supplementary Table 4). Finally, AAs were significantly less abundant in our MAGs suggesting that lignin breakdown is being performed by other microorganisms that we failed to assemble and bin, or that some of the GHs in our MAGs have this activity (Supplementary Table 4).

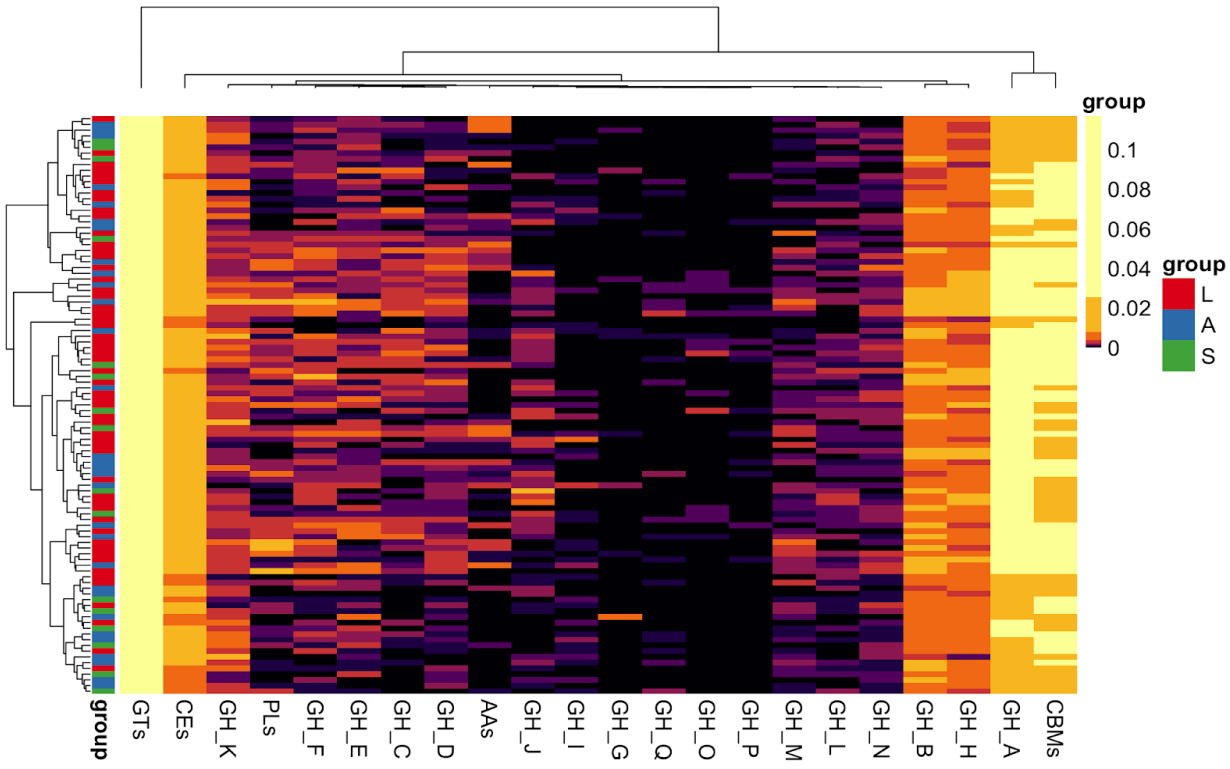


Figure 15. Heatmap of the glycosyl hydrolase clan and other CAZy functions related to cellulose breakdown on reconstructed MAGs from the *Veturius* sp. gut and substrate metagenomes. GH=

glycosyl hydrolase, GT= glycosyl transferases, CE= carbohydrate esterases, AAs= auxiliary activities, PL= polysaccharide lyases, CBMs= carbohydrate binding modules. The origin of each MAG is showed in groups: Red (L)= Larvae, Blue (A)= Adult, and Green (S)= Substrate. Data was normalized by dividing the number of genes belonging to a GH clan or CAZy function to the MAG gene count.

Figure 16 shows the hierarchical clustering of the MAGs based on the presence of carbohydrate active enzymes, and therefore their genetically capability to use similar strategies for cellulose breakdown based on gene content. Looking closer at the clustering, 4 semi-discrete groups of MAGs were distributed in two big branches (Figure 16). The first branch includes Cluster 1, which contains the majority of MAGs and is highly dominated by Firmicutes. Cluster 1 also includes one Proteobacterial MAG, two Bacteroidetes MAGs, four Unclassified bacterial MAGs and the single archaeal MAG detected. Cluster 2a consists of seven MAGs, four of these are Unclassified bacteria. The second branch contains Cluster 2b and Cluster3, the former shows the most diversity of MAGs in terms of taxonomic affiliation including unclassified bacteria, Proteobacteria, Actinobacteria and Bacteroidetes MAGs. Finally, Cluster 3 includes the Verrucomicrobial MAGs. They were exclusively present in the substrate, suggesting that these bacteria use its own particular strategy for cellulose degradation.

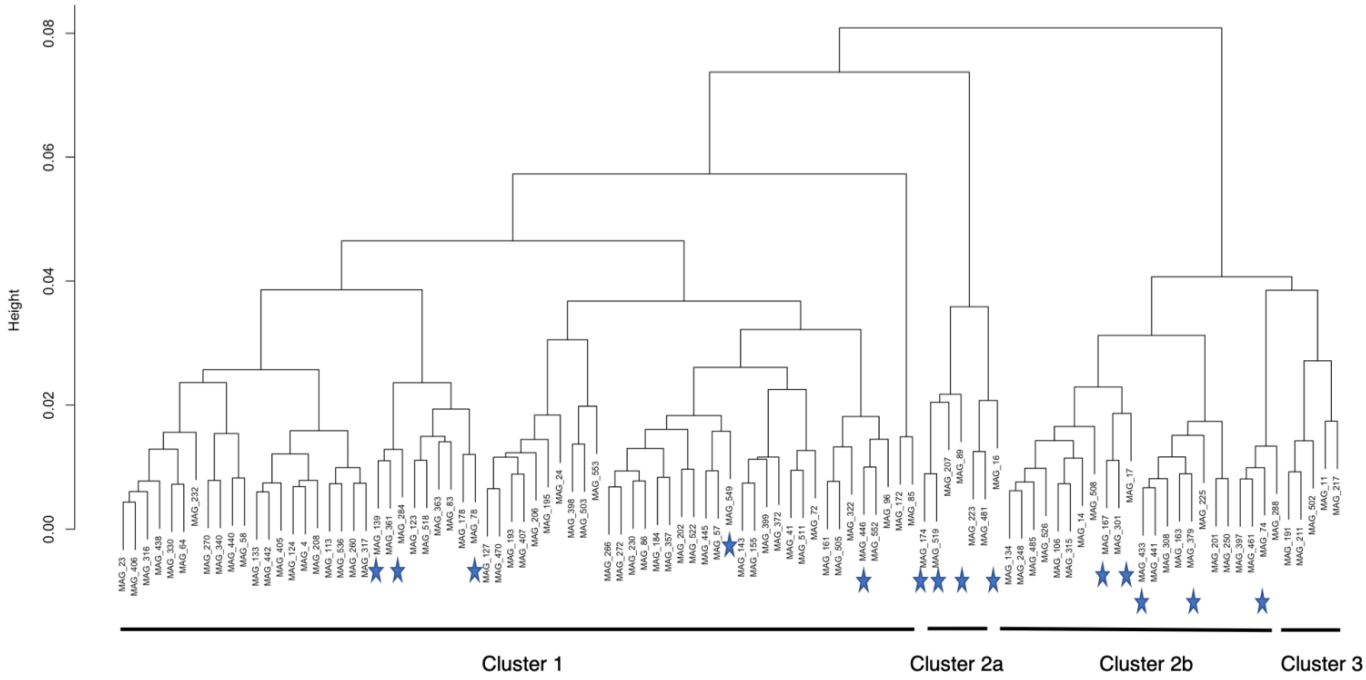


Figure 16. Hierarchical clustering of the abundances of carbohydrate active enzymes in the MAGs reconstructed from the Passalid metagenomes. The clustering was calculated from Figure 15. Blue stars= Unclassified MAGs.

I then searched in the IMG/M annotation for carbohydrate active genes assembled from the seven metagenomes in order to consider genes in other microorganisms that did not make it to bins and did not passed our definition of MAG. For this calculation, I normalized the number of hits obtained for each function to the total number of genes in the respective metagenome. The hierarchical clustering on the top and right of Figure 17 was calculated based on the relative gene abundance.

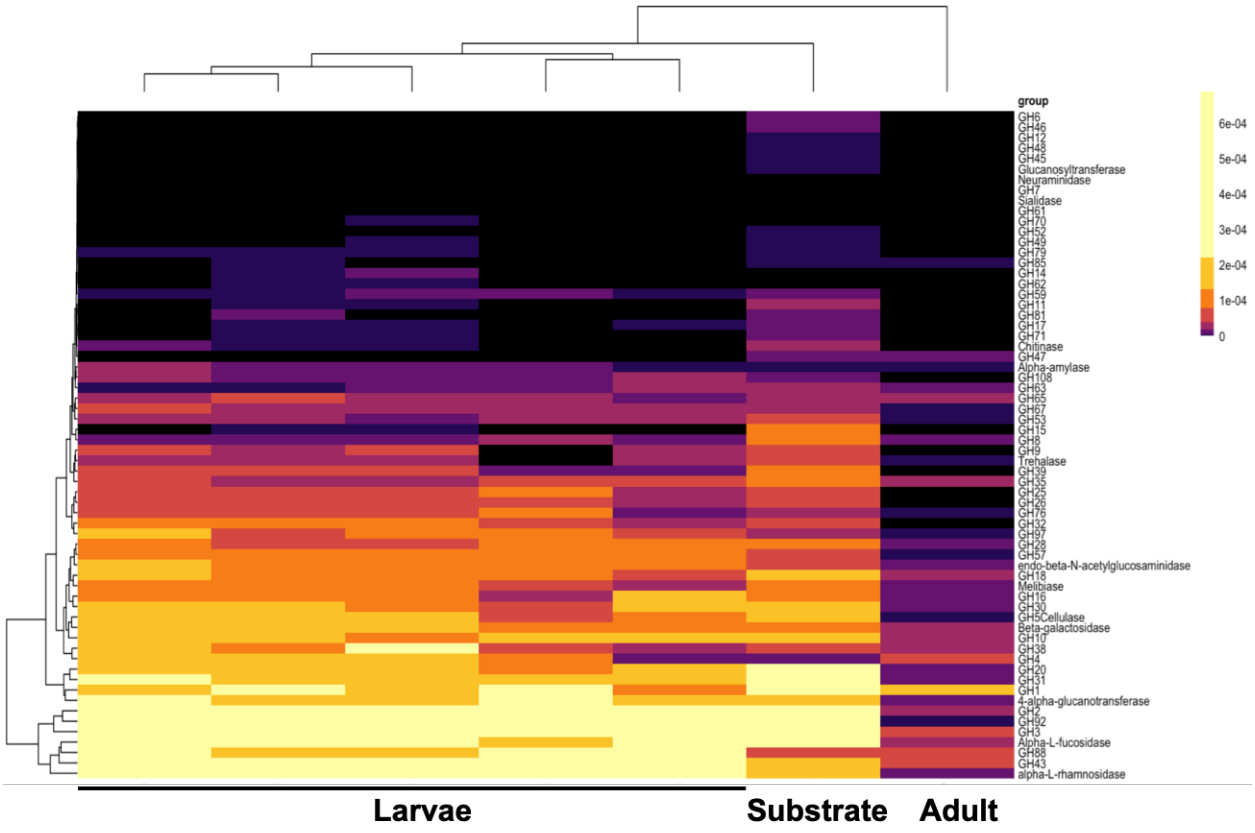


Figure 17. Heatmap of the functions related to cellulose breakdown on the complete metagenomes of *Veturius* sp. gut (larvae and adult) and substrate. GH= glycosyl hydrolase. Data was normalized to the total gene count of each metagenome.

The adult metagenome presented the lowest abundance and diversity of these cellulolytic functions. On the contrary, the substrate showed the highest diversity of genes. Both larval metagenomes and the substrate showed a high abundance of glycosyl hydrolases (Supplementary Table 5). Interestingly, this is the first comparison in which the adult metagenome clustered separated from the larvae given the lowest abundance of glycosyl hydrolases. The cellulases GH2 (β-mannosidase), GH3 (β-glucosidase) and GH92 (α-mannosidase) were the most abundant in the

larvae and substrate. These showed low abundance in the adult. GH43 (α-L-arabinofuranosidases and β-D-xylosidase) and GH88 (glucuronyl hydrolase) showed higher abundance in the larvae metagenomes (Supplementary Table 5). Besides being one of the largest families of glycosyl hydrolases, GH5 were not in the group with the highest abundance in our metagenomes, suggesting that the system relies on other cellulases for plant material degradation. Chitinases were only detected in the substrate metagenomes. The differences between the MAGs and the entire community analysis will be discussed later.

In summary, I was able to produce a set of high-quality MAGs and metagenomic bins comprising a considerable phylogenetic diversity. Moreover, I found evidence that the genomes I reconstructed could be capable of cellulose degradation since they encode a myriad of enzymes with activities involved in this process. By examining the distribution of the functions in the MAGs, I found that these genomes may use different strategies to degrade recalcitrant plant molecules. In terms of the entire community, the larvae and substrate were both rich in carbohydrate active enzymes, however, they showed different patterns in their distribution. The adult has the lowest abundance and diversity of glycosyl hydrolases.

7.3.4 Methanogenesis in the *Veturius* sp. gut

The phylogenetic analysis based on 16S rRNA sequences showed a large abundance of archaea, especially in the larvae samples. Hence, I searched our metagenomic bins for archaeal bins that can give us a better hint into their role in the community. To do this, I looked at the single copy core genes and other phylogenetic markers traditionally used for archaea, such as genes involved

in methane metabolism. The selection of these markers was guided by the taxonomical classification of the archaea in the 16S rRNA data, as it is more probable to assemble and bin genomes of the abundant microorganisms in the system. By this procedure, I was able to classify 9 metagenomic bins as archaea along with the previously identified archaeal MAG260.

Table 9. Summary of archaeal bins reconstructed from the Passalid gut and substrate metagenomes.

	Length (Mb)	contigs	N50	%GC	%completion	%redundancy
MAG 260	2.163051	60	66296	42.95	91.98	6.79
Bin 348	1.027575	122	9507	50.57	54.94	2.47
Bin 517	0.573118	98	6534	48.59	24.07	1.23
Bin 146	0.505845	99	5287	47.83	16.05	1.85
Bin 101	0.447055	83	5636	49.20	16.67	1.23
Bin 5	0.428931	88	4698	48.02	16.05	4.32
Bin 435	0.328622	56	6783	48.80	7.41	0.00
Bin 299	0.306028	48	6730	48.37	17.90	2.47
Bin 87	0.239842	25	13173	42.23	6.79	0.00
Bin 458	0.228178	37	6988	43.52	7.41	0.62

The reconstructed archaeal bins ranged in size from 2.16Mb to 0.23Mb and the average %GC was 47. The % redundancy of the bins < 7%; and with the exception of the MAG, the %completion ranged from 55% - 6.8% based on the presence of single copy core genes (Supplementary Table

6). Interestingly, MAG260 was only detected in the adult metagenome, all the other archaeal bins were detected in the larvae. None of these organisms was detected in the substrate metagenome.

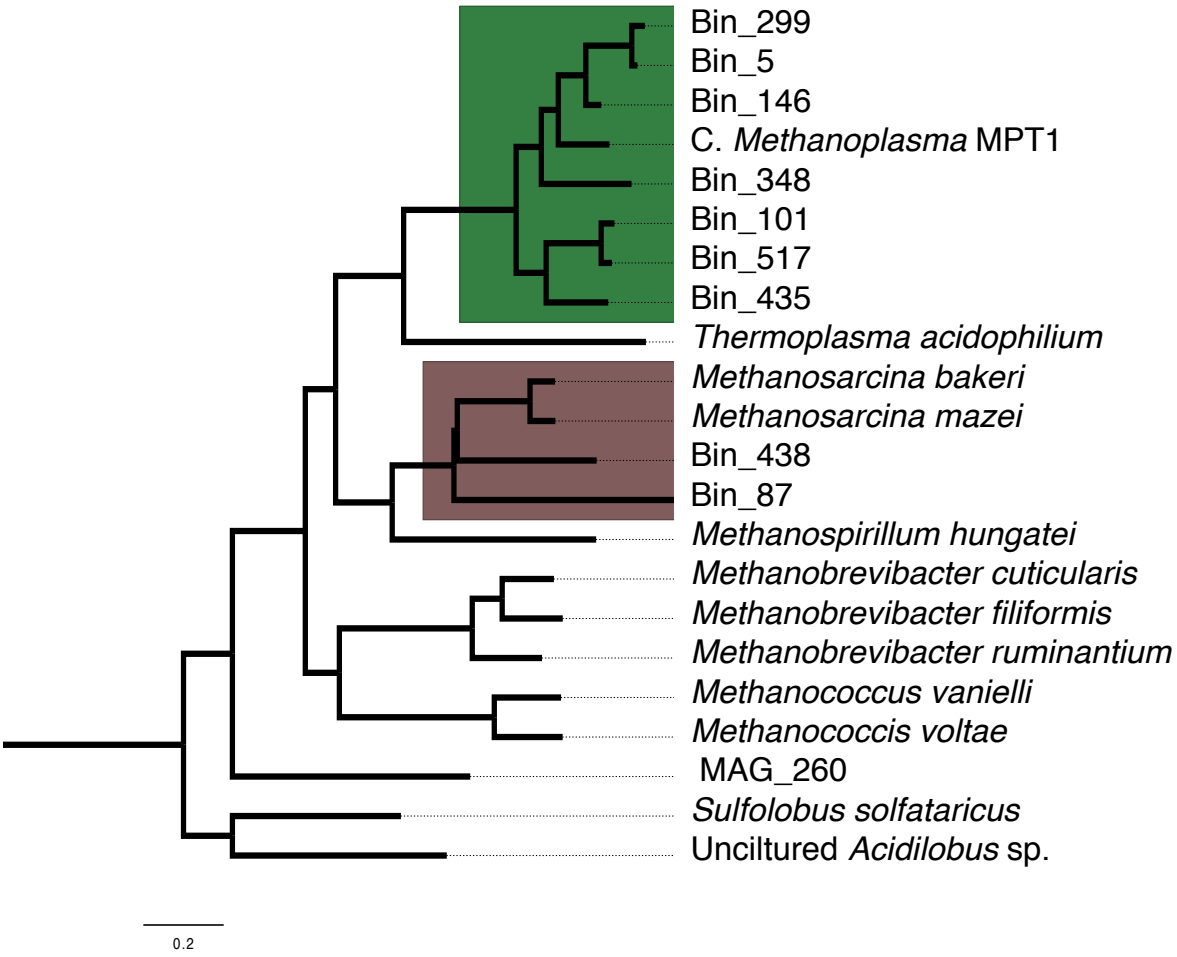


Figure 18. Phylogenomics of archaeal bins. Green: putative Candidatus *Methanoplasma* clade. Brown: putative *Methanosarcina* clade. Phylogenomics performed based on 13 concatenated single copy core genes. The reference sequences included in this tree were retrieved from full genomes in GenBank.

In order to have a better taxonomic classification of the bins, I concatenated the single copy core genes that were shared between the bins to perform a phylogenomics analysis. I included the closest homologs of these genes from known archaea as reference. The majority of the bins cluster with Candidatus *Methanoplasma termitum*, this strain was isolated from an enriched culture of termite gut contents and its genome was recently sequenced (Figure 18). Ca. *Methanoplasma termitum* has been classified in a new order of methanogenic bacteria, the Methanomassiliicoccales. It has been proposed that these archaea use a new metabolic strategy for energy metabolism (Figure 19): a Fpo-like coenzyme related to F_{420} to reduce hydrogen. Therefore, I searched for these genes in the bins included in this cluster and was able to find homologs for the mentioned hydrogen reducing coenzyme in Bin348 only, which encodes all the subunits of Coenzyme F_{420}-reducing hydrogenase. These results suggest that the Methanomassiliicoccales use similar strategies for energy production.

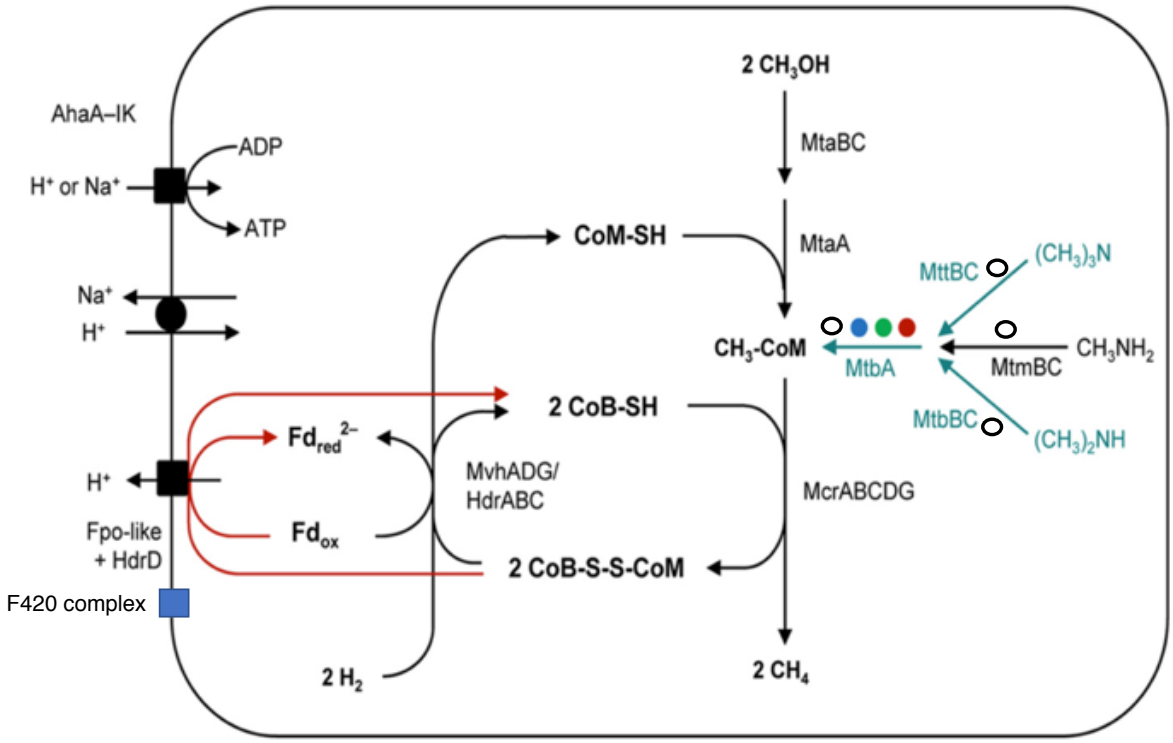


Figure 19. Proposed energy metabolism for members of the order Methanomassilicoccales. Black arrows= reactions encoded in all genomes, Red arrows= proposed reaction of HdrD coupled to the Fpo-like complex. Blue-green arrows= enzymes not present in "Ca. *Methanoplasma termitum*" but are present in other members of the order. Empty black circles= genes not found in our bins. Blue square= F_{420} complex found in bin 348. Adapted from Lang et al[74].

Next, I searched these bins for methyl-coenzyme M reductase (*mcr*) genes, key enzymes for methanogenesis. Full length sequences were recovered for *mcrC* (subunit C in the complex) in four of the bins, all of which are part of the Methanomassiliicoccales clade in the phylogenomics analysis. The *mcrC* in Bin101, Bin517 and Bin146 clustered with their homolog in Ca. *Methanoplasma termitum* (Figure 20). Interestingly, the *mcrC* of Bin348 was positioned in a

different clade outside this group. This result, in addition of the evidence of the presence of the Coenzyme F_{420}-reducing hydrogenase genes in the same bin, indicates that Bin348 does not belong to the Methanomassiliicoccales clade, and that it was probably grouped within this clade because of lack of phylogenetic resolution (Further discussion later).

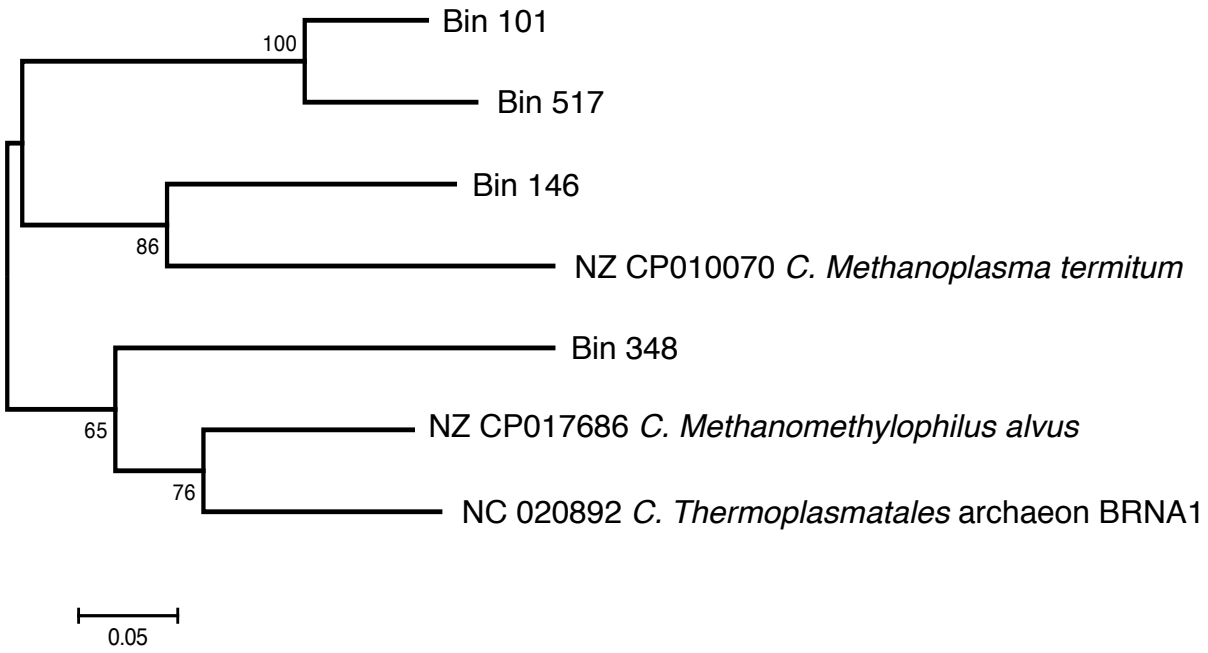


Figure 20. Maximum likelihood tree of *mcrC* full length genes from reconstructed archaeal bins. Bootstrap= 100 000 generations. The reference sequences included in this tree were retrieved from full genomes in GenBank.

No homologs of *mcr* were found in Bin458 and Bin87, these were classified as putative *Methanosarcina* based on phylogenomics. Given that all known *Methanosarcina* are methanogens, the absence of homologs of *mcr* in these bins can be explained by their size. These are the two shortest archaeal bins with 0.23Mb and 0.24Mb, and % completion of 7.4 and 6.8 respectively.

Homologs of *mcr* were also absent in MAG260. Contrary to Bin458 and Bin87, MAG260 is 2.16Mb in size, has a % completion of 92 and encodes for 2 257 ORFs. Therefore, the absence of this key gene suggests that this organism is not capable of performing methanogenesis. The clustering of MG260 in the phylogenomics tree is also interesting, it is positioned in a clade separated from the other Euryarchaeota indicates that it might belong to a new class in this phylum. However, a much comprehensive phylogenetic analysis needs to be done to confirm this. In order to obtain more information about the possible role of MAG260 in the adult gut, a metabolic model of was constructed employing the 2 257 ORF predicted to be coded by this MAG. No gap filling methods were used in the model as we are confident of the completion of the MAG, and preferred to avoid possible biases in our interpretation of the model. The lack of genes in methane metabolism pathways was confirmed with the model (Supplementary figure 6). Moreover, the model shows that the MAG260 is capable of fixing nitrogen as it encodes the nitrogenase complex, nitrogen fixation has been previously found in methanogenic archaea. Incomplete TCA cycle and Pentose phosphate pathways were found, but no evidence of oxidative phosphorylation genes and cytochromes. The MAGs also lacks of RuBisCo based carbon fixation genes suggesting a fermentative metabolism.

The collection of 766 metagenomic bins includes high quality partial genomes of microorganism from two domains in the tree of life, and eight different phyla. The average genome size was >1Mb and an average of more than 1000 genes per bin (about half of them lacking annotations in the current database), therefore this collection constitutes a valuable resource to the study of microbial ecology and for bioprospecting purposes.

7.4 Comparative metagenomics of *Veturius* gut and substrate

7.4.1 The *Veturius* sp. gut has convergently evolved to resemble rumen

In order to compare the *Veturius* sp. system and other cellulolytic metagenomes, as well as the metagenomes from other beetles, I aimed to perform a global comparison of the proteins found in the systems. Then, to compare the functional capabilities of these systems, I utilized the COG (Clusters of Orthologous Genes) annotations to all the predicted ORFs of a set of metagenomes available in the IMG database along with the *Veturius* metagenomes. The gene hits to these functions were collected in all the metagenomes, then the data was normalized to the total gene count of the respective metagenome. The resulting matrix was used to generate a hierarchical cluster by calculating the correlation coefficient based on the similarity of the functional characterization of the genomes (Figure 21). The cluster showed two major branches, one branch included the adult metagenome by itself, suggesting that it is functionally different from the other metagenomes. The second branch also shows two major groups, the first includes the leaf-cutter ant fungus garden and refuse dump metagenomes, the metagenomes associated with the fungus garden of bark beetles: the mountain pine beetle *Dendroctonus frontalis* and *Dendroctonus*

ponderosae, and the Passalid substrate (gray branch in Figure 21). Again, clustering of this groups reflects functional similarities between the microbes in these metagenomes.

The most interesting result was obtained for the larvae, they clustered in the same branch as the termite hindgut and the rumen from different mammals. The five larvae metagenomes are included in one branch having the rumen from Bovines from Vernon Texas, the goat rumen and the Musk oxen rumen as its closest neighbors. The next group of metagenomes in proximity to the Passalid larvae are the metagenomes of the hindgut of *Cubitermes* termites. The second half of the cyan branch includes the *Nasutitermes* termites, the sheep and other bovine rumen. This result indicates that the microbes of the Passalid larvae encode for similar proteins that the ones encoded in the rumen and in the hindgut of termites. Furthermore, our analysis implies the existence of evolutionary processes selecting for microbes carrying these functions.

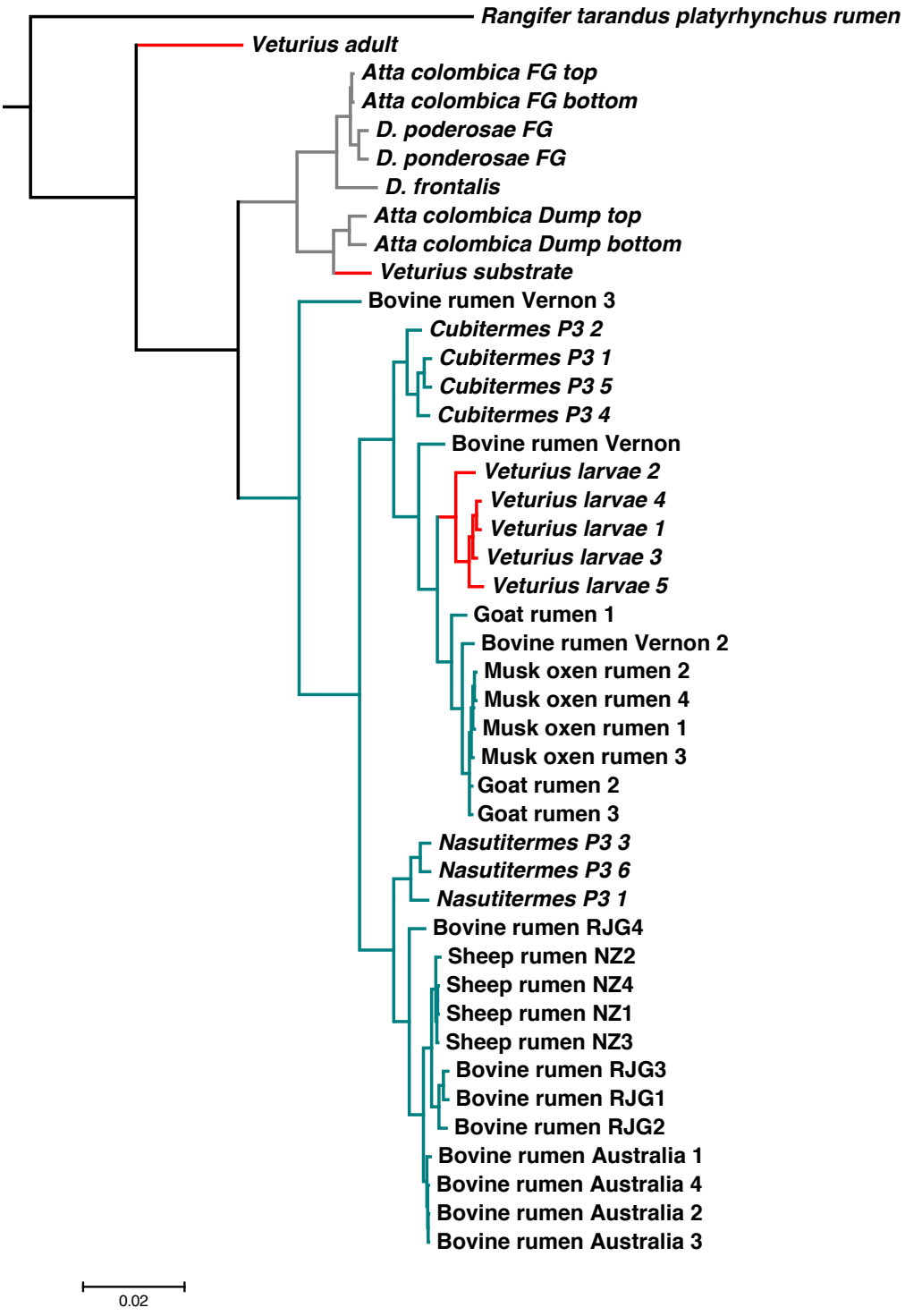

Rangifer tarandus platyrhynchus rumen
Veturius adult
Atta colombica FG top
Atta colombica FG bottom
D. poderosae FG
D. ponderosae FG
D. frontalis
Atta colombica Dump top
Atta colombica Dump bottom
Veturius substrate
Bovine rumen Vernon 3
Cubitermes P3 2
Cubitermes P3 1
Cubitermes P3 5
Cubitermes P3 4
Bovine rumen Vernon
Veturius larvae 2
Veturius larvae 4
Veturius larvae 1
Veturius larvae 3
Veturius larvae 5
Goat rumen 1
Bovine rumen Vernon 2
Musk oxen rumen 2
Musk oxen rumen 4
Musk oxen rumen 1
Musk oxen rumen 3
Goat rumen 2
Goat rumen 3
Nasutitermes P3 3
Nasutitermes P3 6
Nasutitermes P3 1
Bovine rumen RJG4
Sheep rumen NZ2
Sheep rumen NZ4
Sheep rumen NZ1
Sheep rumen NZ3
Bovine rumen RJG3
Bovine rumen RJG1
Bovine rumen RJG2
Bovine rumen Australia 1
Bovine rumen Australia 4
Bovine rumen Australia 2
Bovine rumen Australia 3
0.02

Figure 21. Functional clustering of cellulolytic and other insect-associated metagenomes. Spearman correlation clustering based on COG functions. Branch containing the Passalid beetle metagenomes are shown in red.

A similar analysis but this time using the taxonomic classification of the ORFs in the metagenomes was performed to test whether this clustering matches the one obtained for the functional comparison. I used the same approach, calculating the correlation coefficient, but in this case of the taxonomy assigned to the genes in the metagenomes at the class level. The taxonomical clustering is mostly consistent with the functional clustering, as it shows that the taxa present in the rumen, termite and Passalid larvae metagenomes share taxa (Figure 22). The Passalid substrate clustered again with the bark beetles and the leaf cutter ants (gray branch). The position of the adult metagenome differed from that seen in the functional analysis, as this clustered in the cyan branch, more similar to the Passalid larvae, the termites and the rumen.

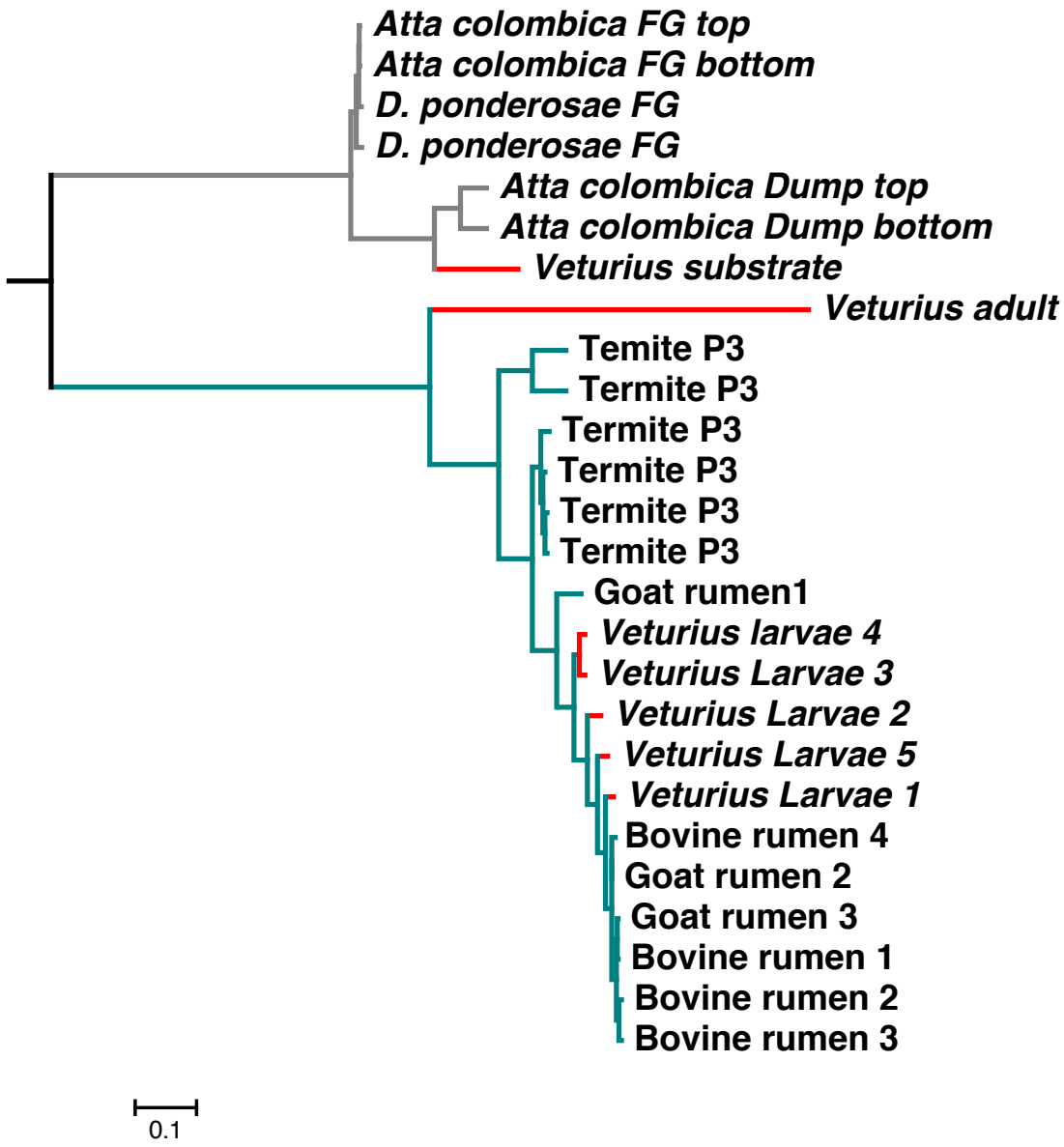


Figure 22. Taxonomical clustering of cellulolytic and other insect-associated metagenomes. Spearman correlation clustering based on taxonomy at the class level. Branches containing the Passalid beetle metagenomes are shown in red.

7.4.2 Cellulose degradation and methanogenesis in *Veturius* sp. and other cellulolytic systems

The results from the previous sections showed that Passalid larvae, rumen and the termite hindgut have similar genes, and therefore, they might be using similar strategies to generate energy and biomass from plant material. Therefore, I expect that cellulose degradation and methanogenesis are the most important catabolic processes in these systems. I previously showed that the Passalid larvae microbial communities are dominated by Firmicutes bacteria and methanogenic archaea. The Firmicutes are known to encode cellulases that help termites and rumen to digest cellulose, and methanogenic archaea are also active and abundant in the termite hindgut and the rumen. Thus, next I compared the distribution of genes of carbohydrate active enzymes and methanogenesis functions in insect and rumen metagenomes.

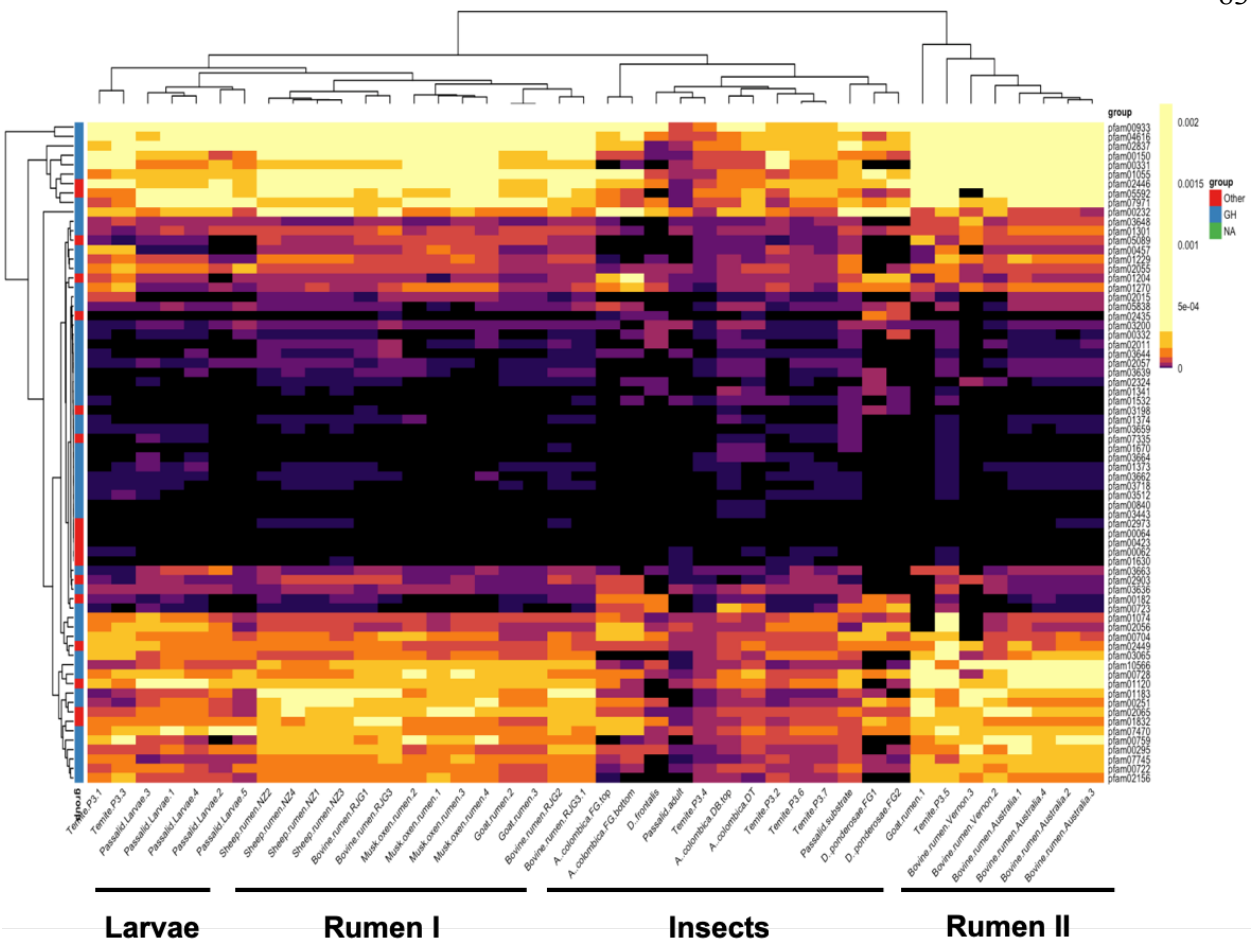


Figure 23. Heatmap of the functions related to cellulose breakdown on cellulolytic and other insect-associated metagenomes. Blue= glycosyl hydrolase, Red= other carbohydrate active enzymes. Data was normalized to the total gene count of each metagenome.

The heatmap in Figure 23 shows that the Passalid larvae were grouped again with the termites and the rumen in the first cluster (left to right). The Passalid adult and substrate grouped in a second cluster with the bark beetles and the ants; the heatmap shows that the major cause of the separation of this second cluster is the difference in the abundance of the functions in the top of the figure. These include the Pfams for the most important cellulases and GHs: GH2, GH3, GH5, GH10 and

GH43. There is a third cluster on the far right that similarly to the second cluster, shows a significantly high abundance of GH2, GH3, GH5, GH10 and GH43; however, these metagenomes also showed higher abundance of GH97, GH20 and fucosidase in comparison to the other two clusters.

7.4.3 Methanogenesis in insect metagenomes

Encouraged by the results obtained in the previous sections where I found evidence of the presence of methanogenic archaea in the Passalid adults and larvae, I compared the abundance of the functions related to methanogenesis in the Passalid metagenomes to other insects capable of producing methane.

The heatmap in Figure 24 shows that the Passalid larvae and the termites have similar abundances and distribution of functions related to methanogenesis (left cluster). The adult showed significantly less abundance, however, it was grouped in an intermediate abundance group along with other termites. This result is consistent with the 16S rRNA data that showed the presence of methanogenic archaea in the Passalid adult but in a lower abundance compared to the larvae. Finally, the Passalid substrate clustered with the ants and the bark beetles on the extreme right cluster, this branch shows the lowest abundance and lowest number of genes involved in these pathways.

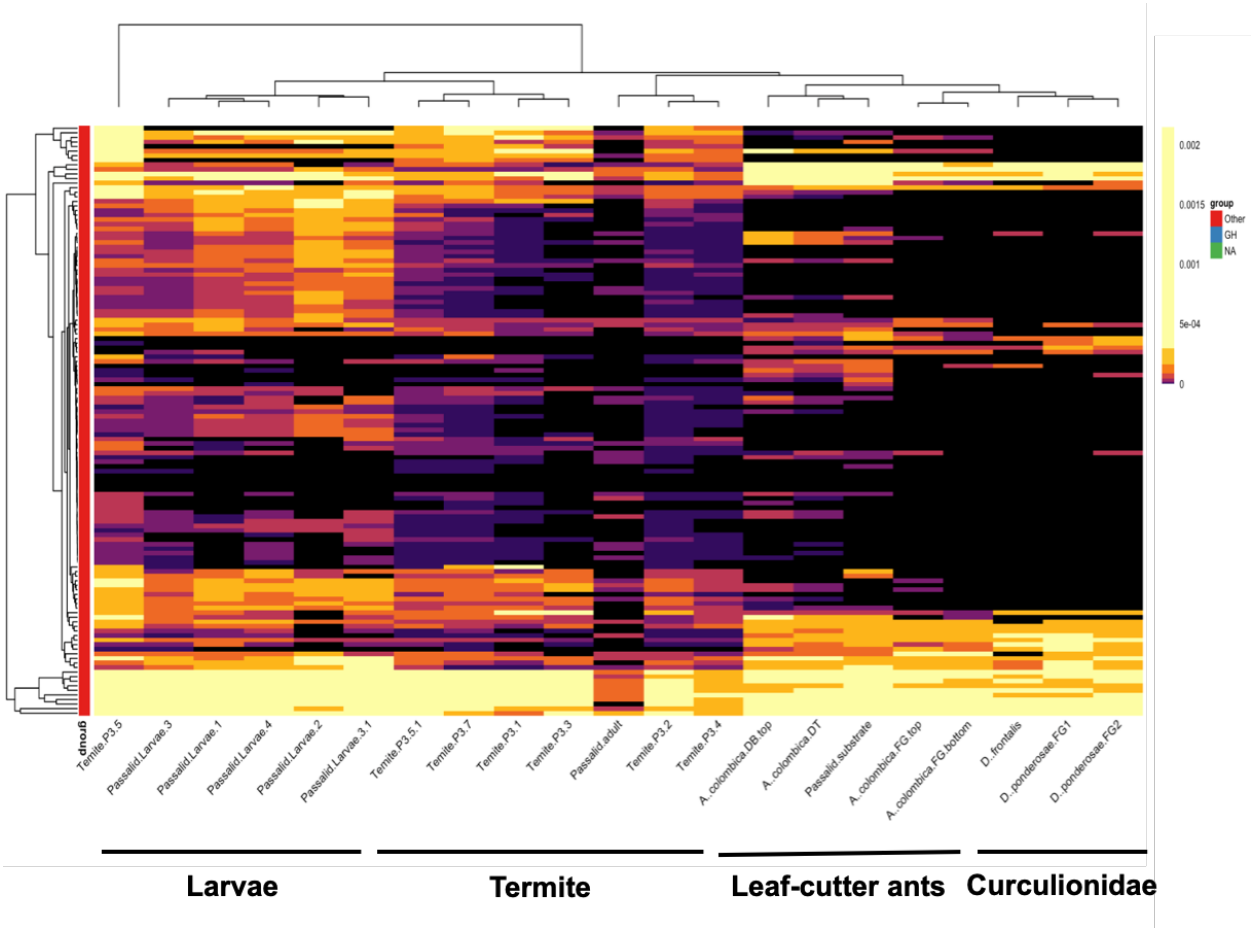


Figure 24. Heatmap of the functions related to methanogenesis and methane metabolism on insect-associated metagenomes including *Veturius* sp. Data was normalized to the total gene count of each metagenome.

Together, these results suggest the existence of convergent evolution between the *Veturius* sp. larvae guts and the rumen microbial communities. The similarities between the two systems lay both on the phylogenetical affiliation of the members of the community as well as in their genetic potential.

7.4.4 Nitrogen fixation in the *Veturius* sp. gut and substrate metagenomes

Insects often establish symbiotic relationships with nitrogen fixing microorganisms to supply their nitrogen requirements. The genomes of potential nitrogen fixers were reconstructed from the metagenomic sequences of *Veturius* sp. A total of 92 open reading frames of the subunits of the nitrogenase (*nif*) were found in 32 of the bins in the collection. The nitrogenase was found in 11 MAGs (10.9% of reconstructed MAGs), one of which was the archaeal MAG260. The phylogeny in Figure 25 shows a clade in the tree with no known similar sequences when the public databases were searched. Nitrogenase subunits were also found in bacterial and archaeal bins (shown in red in Figure 25)

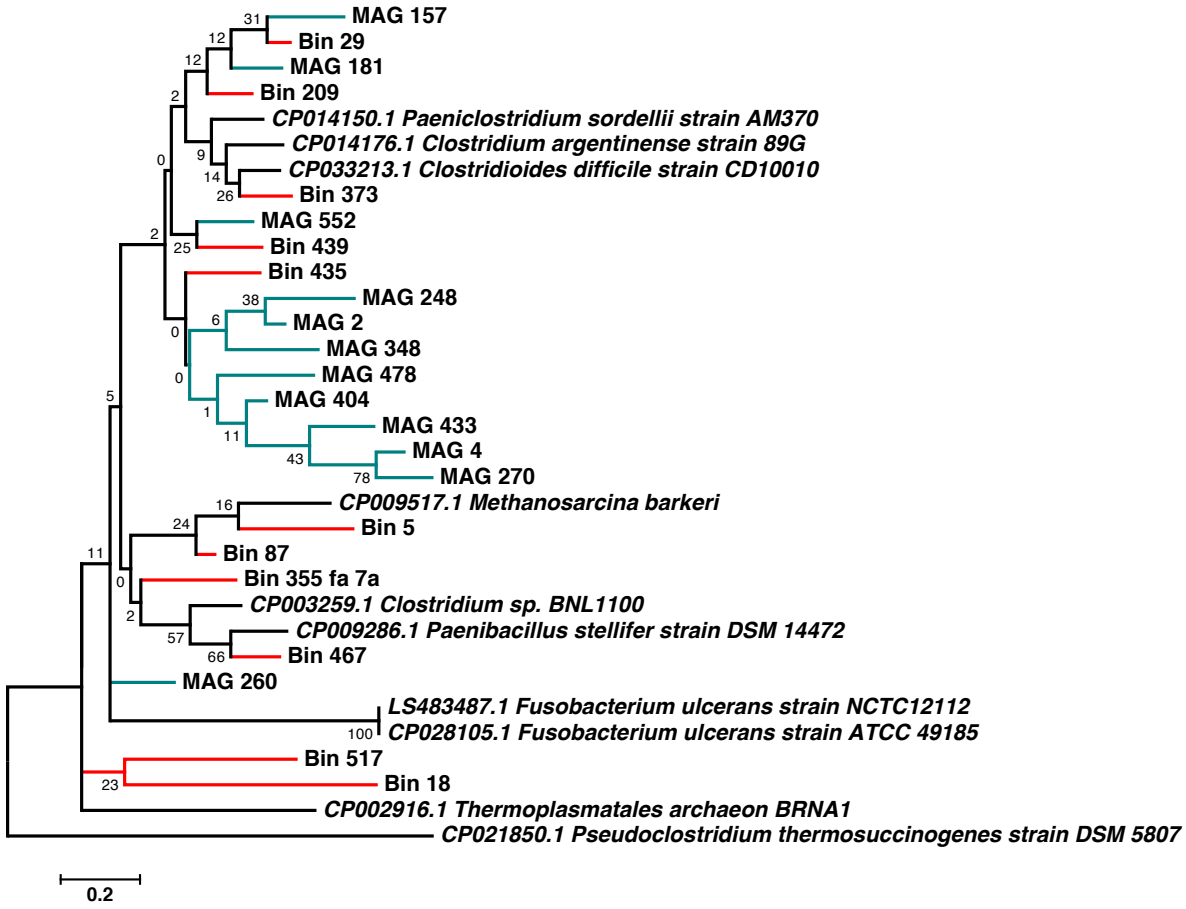


Figure 25. Maximum likelihood tree of *nifH* full length genes from reconstructed in *Veturius* sp. MAGs and bins. Bootstrap= 100 000 generations. The reference sequences included in this tree were retrieved from full genomes in GenBank. Branches containing MAGs are shown in cyan, branches containing bins are shown in red.

In order to explore the genetic potential for nitrogen fixation in the seven *Veturius* sp. metagenomes, I compared the abundance of the nitrogenase subunits in the adult, larvae and substrate metagenomes, with the metagenomes of other insects. The three nitrogenase subunits *nifD*, *nifH* and *nifK* were found in similar abundances in all the metagenomes included in the

analysis. The gene *anfG* was the one that showed the lower abundance in all metagenomes (Figure 26). The relative abundances of these subunits are summarized in Supplementary Table 7. There were not significant differences in the average relative abundances of the metagenomes per insect type.

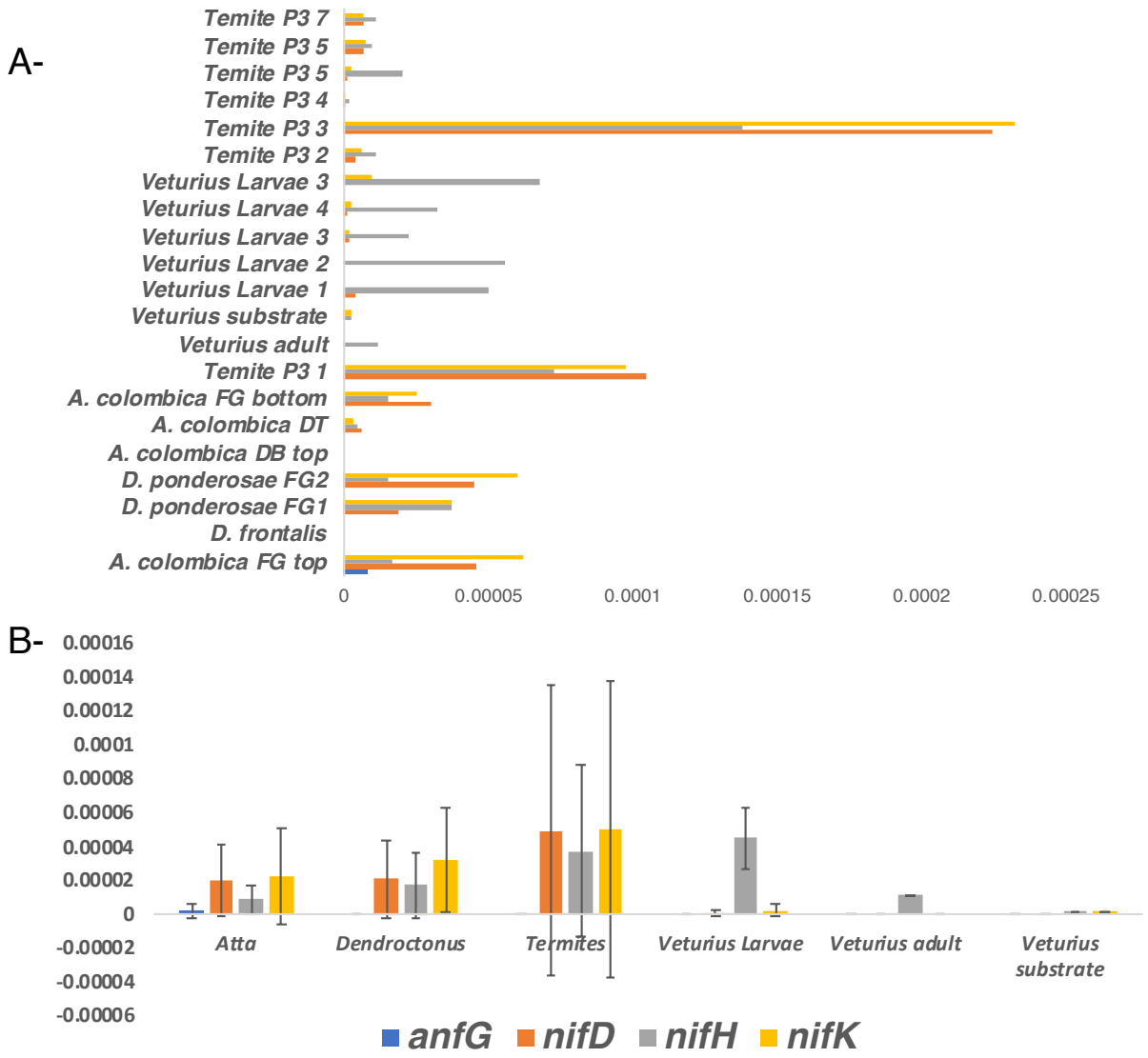


Figure 26. Relative abundance of genes for the nitrogenase complex subunits in the *Veturius* sp. and other insect metagenomes (A). Average per insect for replicates of metagenomes related to the

same type of insect but not biological replicated (B). Passalid adults and substrate only have one replicate. Data was normalized to the total gene count of each metagenome.

The lack of significant differences between the abundance of genes related to nitrogen fixation in the insect metagenomes suggest that *Veturius* sp. microbial communities also include organisms capable to perform this process.

8 Discussion

The results obtained in this study suggest that the *Veturius* sp. adult and larvae gut, as well as their substrate's microbial communities, have a vast phylogenetical diversity, and that their communities are distinct to each other. These microbial communities are enriched in microorganisms with the genetic potential to play important roles for the community itself and for the host, such as cellulose degradation, nitrogen fixation and methanogenesis. Furthermore, we showed that the microbial communities of *Veturius* sp. are abundant and diverse in terms of the genes involved in these processes. Furthermore, our findings suggest that the adult and larvae gut communities cooperatively degrade cellulose, this characteristic might be one major driver for the subsocial behavior of these beetles. The functional capacity and the phylogenetic diversity in the larvae microbial communities resembles the one of the mammal rumen, suggesting that these animals have convergent evolutionary trajectories toward an effective digestion of plant material in their guts. As part of our analysis we created a collection of partial genomes of relevant members of the Passalid beetle microbial communities comprising a large phylogenetic and functional diversity. This collection represents a valuable resource for further microbial ecology studies of Costa Rica biodiversity. Finally, I combined the findings in the study to propose a model for the ecology of the Passalid beetle and their microbial communities.

8.1 Influence of the methodological strategy in the microbial diversity

Advances in molecular biology and deep sequencing have allowed researchers to explore the microbial diversity of many ecosystems in the planet. However, the field of microbial ecology is still in the early stages; hence, the validation of methods is required to obtain meaningful and

consistent conclusions. There are several variables that could have influenced the final results of the sequencing runs in terms of both sequence yield and detected microbial diversity. Some of these variables included the anatomical differences in the adult and larval gut, the differential physical conditions of the gut contents and the substrate samples, the presence of chemical inhibitors in the sample; and some others that are inherent of molecular biology protocols as the use of different buffer solutions and reagents, the method used for cell lysis and the persistence of nucleases among others. Because of the lack of a gold standard methods for DNA extraction in ecological studies, we tested different protocols for metagenomic DNA isolation in the lab and tested their impact in the diversity obtained for the community. By comparing the results of several alpha diversity metrics we demonstrated that the extraction kit used does not significantly influence the microbial community diversity and composition. It is vital to test the influence of the methodology in the experimental results as there are several examples in which DNA extraction protocols or other molecular biology techniques significantly impacted the diversity obtained of the microbial communities. In fact, in a study of human fecal microbiome samples that used two of the same DNA isolation kits that we used in our experiment, the authors found that the method introduced significant differences in the abundances of Firmicutes including Ruminococcaceae, Clostridiaceae and Lachnospiraceae, as well as other members of Enterobacteriaceae[75]. We did not detect this effect in our data. Gram positive bacteria represent a challenge for molecular studies based on DNA isolation; they have strong cell walls and some of them produce spores that are hard to lyse. Different approaches can have been used aiming to obtain good DNA from communities that include these bacteria, the most commonly used involve physical methods such as vortexing, bead beating and sonication, and the use of chemicals to dissolve the components of

the cell wall[76–78]. The physical methods, especially the ones involving bead beating, have proven to produce consistent results in terms of diversity and yields when applied to different sample types such as mock communities[79] , activated sludge[80], or other more difficult samples to extract such as plant woodchips[81]. In order to test the possibility of missing interesting taxonomical groups that could be tightly attached to gut epithelia, as these can be involved in intricate relationships with the host, I compared the alpha diversity of samples that were processed using the same protocols of DNA isolation but including only gut contents versus gut contents plus epithelia. No significant differences were found in diversity between the samples or total sequencing yields; however, the epithelium samples presented a large amount of eukaryotic sequences that had to be filtered out causing a significant reduction of the microbial sequence yields. Given the universality of the primers used in this study we compared our sequences to the ones in available databases and confirmed that more than 90% of the eukaryotic sequences were similar to beetle rRNA sequences. It has been previously shown that the primer set used in this study (as well as other common primers used for amplicon sequencing of 16S rRNA regions) can be reliably used to study the diversity of eukaryotes and archaea[82]. Therefore, the archaeal sequences were kept for the analysis and only the eukaryotic sequences, as they were highly dominated by the own beetle, were filtered out.

8.2 The larvae and adults of *Veturius* sp. harbor specialized microbial communities

One of the most interesting results obtained was the striking difference found between the adult, the larvae and the substrate microbial communities, and the consistency of the composition of this communities showed in the beta diversity comparison. The sample type was the most important

driver of the variation to the microbial community composition, overpowering the contribution of beetle family and other variables like geographical location, tree species, and other environmental factors. The beta diversity clustering observed in the ordination analysis was supported by the differences in the phylogenetic groups that dominated each sample type. The Passalid substrate samples were largely dominated by Proteobacteria, the adults by Firmicutes and Tenericutes, and the larvae by Firmicutes and Euryarchaeota. Contrary to the findings, one could hypothesize that the communities in the adult and the larvae are similar, and that the subsocial behavior of these beetles is due the need to share the microbes that encode the genes for cellulose degradation. Moreover, the initial expectation was that the microbial community in the substrate was also similar to the beetle gut, as it constitutes the ideal vehicle for the horizontal transmission of microbes between the larvae and the adult. This was the case for the wood feeding beetle *Anoplophora glabripennis*. In a more modest clone study of the microbial diversity related to this beetle, the authors found more similar microbial communities between different life stages of the beetle and the wood material, and concluded that the microbes in the larvae gut where either vertically transmitted or, environmentally acquired from the bark they feed on[22]. Also, a metagenomic study of the Mountain pine beetle *Dendroctonus ponderosae* showed that the beetle fungus garden and the Lodgepole pine where they were sampled have similar microbial communities dominated by Proteobacteria[83]. Likewise, the results of another 16S rRNA clone library in the bark beetle *Dendroctonus valens* LeConte also showed similar microbial communities in the larvae and adult, again dominated by Proteobacteria[84]. Finally, a study of the bacterial communities associated to another Curculionidae, the pine weevil beetle *Hylobius abietis,* found that these were very similar in individuals across Europe and dominated by Proteobacteria

followed by Firmicutes[85]. Contrary to those previous findings in other beetle species, these results suggest that the communities in *Veturius* sp. are different, only sharing about 3% of the more than 11 000 OTUs. Moreover, there are high order differences in the dominant taxa even at the phylum level. I hypothesize that adult and larvae have distinct ecological roles and therefore, have evolved to harbor specialized microbial taxa that facilitates the metabolic processes necessary for their ecology.

The Passalid adult and larval guts are enriched in microorganisms capable of cellulose degradation all the way to methane and carbon dioxide. Their communities are abundant in Clostridiales including members of Clostridiaceae, Ruminococcaceae and Lachnospiraceae, as well as some other Bacilli. The larvae are also enriched in methanogenic archaea comprising members of several classes of Euryarchaeota. These results are consistent with the ones found in other cellulolytic microbial communities such as the cow rumen and the termite hindgut. Several studies of mammal rumen have identified the Firmicutes as the taxa carrying the largest and more diverse repertoires of carbohydrate active enzymes[20,52,86,87]. In a study of the hindgut of two different species of termites, the authors found that the taxa carrying the cellulolytic capacity differ between the species; in *Amitermes wheeleri* the most abundant phyla where the Firmicutes and Spirochaetes whereas in *Nasutitermes corniger* the community was dominated by the Spirochaetes and Fibrobacteres[17,19]. Both Spirochaetes and Fibrobacteres showed significant abundance in the *Veturius* sp. metagenomes. Similar to the Passalid beetles, cow rumen is also enriched in methanogenic archaea, most of these are classified as *Methanobrevibacter* based on the results of a recent metagenomic study[88]. All together, these results suggest the *Veturius* sp. gut as a phylogenetically and functionally diverse cellulolytic system with characteristics similar to the cow rumen, and to

lesser degree the termite hindgut. The Passalid gut is a source of novel microorganisms capable of perform important metabolic processes that has not been explored yet.

8.3 The microbiomes of *Veturius* sp. include organisms with the genetic capacity of cellulose breakdown and methanogenesis

The analysis of the metagenomes of *Veturius* sp. showed that their associated microbial communities harbor diverse microorganisms capable to perform the complete degradation of cellulose all the way to methane (data not shown). In order to understand more about the biology of these microorganisms were mined the genes that populate the pathways for the degradation of these compounds with the advantage of being able to consider the genomic context of the genes in the genomes reconstructed from the metagenomic reads. In total 766 partial genomes were reconstructed from microorganisms that inhabit the *Veturius* sp. gut and its substrate. These genomes are distributed among seven bacterial phyla and the archaeal phyla Euryarchaeota. From the 766, 101 met the criteria for a MAG, and therefore the analysis was concentrated in those near complete, high quality genomes. Because of the implicit limitations of metagenomics, the taxa with the highest probability of being successfully binned are the ones showing abundant and evenly distributed populations. Hence, the binned genomes often represent the important organisms in the system[64,88,89]. The majority of the MAGs reconstructed from the Passalid metagenomes correspond to Firmicutes, and they contain a large number of glycosyl hydrolases and other genes coding for carbohydrate active enzymes. This result adds evidence to the key role that the Firmicutes have in the degradation of cellulose in the Passalid gut. Other researchers have attempted to assemble genomes from cellulolytic metagenomes, specifically from the cow rumen.

In one of the first studies, the authors were able to assemble 15 genomes of uncultured bacteria from the phyla Firmicutes, Bacteroidetes and Spirochaetes[52]. More recently, others have taken advantage of the methodological improvements and the computational efficiency of novel bioinformatic algorithms to assemble larger number of genomes. In a 2018 study the authors used a combination of HiSeq Illumina sequencing and Hi-C Illumina sequencing[90] (alternative library preparation method that uses crosslinked strategies to capture DNA molecules that are physically adjacent), to reconstruct draft genomes of 898 microorganism assembled from 43 cow rumen metagenomes. The majority of their MAGs were classified in the Firmicutes and Bacteroidetes phyla, followed by genomes of Actinobacteria, Proteobacteria and Euryarchaeota[88]. Their collection resembles ours in terms of taxonomical affiliation of the genomes and in their overall functional capacity. Both datasets showed thousands of novel sequences that encode for carbohydrate metabolism (estimation based on the percentage of sequences with good matches to the public databases), however there is a significant difference in terms of the phylogenetical affiliation of the bacteria that carry the enzymes. While in cow rumen MAGs the majority of the glycosyl hydrolases are detected in Bacteroidetes (Prevotellaceae), in the Passalid beetle these are significantly more abundant in the Firmicutes, specifically in the Clostridiales. The approximated taxonomy of the archaea found in both collections is also distinct, 25 of the 28 archaeal genomes in the cow rumen dataset were similar to known ruminant *Methanobrevibacter*, whereas our partial genomes are more similar to the Methanomassiliicoccales. This group is of particular interest as it has been shown to encode for novel coenzymes that allowed them to perform a novel type of energy metabolism, this archaea has been detected also in the rumen of Chinese goats[54].

There is evidence of the presence of other archaea in the 16S rRNA data comprising the five major classes of Euryarchaeota, including an abundant OTU similar to a unique clade that contains vadinCA11, and WCHD302, an archaea obtained from an aquifer contaminated with organic solvents[91]. Unfortunately, no genomes were yet assembled for these other classes. Together the cow rumen and the Passalid beetle MAGs substantially improved our knowledge of microbes specialized in degrading cellulose, producing a new set of more than 1000 genomes of microorganisms from cellulolytic environments. Most of these organisms cannot be grown in the laboratory under current culturing methods. Both datasets are a valuable resource for bioprospecting purposes, as will be the study of the ecology of other taxa such as Verrucomicrobia and Acidobacteria.

8.4. The *Veturius* sp. microbiomes contain microorganisms capable of preforming nitrogen fixation

The microbiomes of *Veturius* sp. contain homologs for nitrogenase subunits similar in abundance and diversity to what has been described in termites[17,19], a well-known case of symbiotic gut microbes that fix molecular nitrogen from the atmosphere[17,92]. Termites, as well as other wood feeding insects in the forest, live on diets that are dramatically depleted in nitrogen; with carbon to nitrogen ratios up to two orders of magnitude higher in wood material than in insect biomass[51,93]. Nitrogen is a limiting factor in the ecosystem as it is abundant in proteins and other biomolecules. Therefore, the establishment of symbiotic relationships with nitrogen fixing microorganism are essential for the survival and ecological success of wood feeding insects. Similar to termites, which depend on Spirochaetes and other bacteria to supplement their nitrogen requirement, leaf-cutter

ants species rely on Proteobacteria to provide them with this valuable resource[24,94]. Nitrogen fixing bacteria have been isolated also from bark beetles *Dendroctonus terebrans* (Coleoptera:Curculionidae)[95]. In-situ acetylene reduction has been demonstrated in larvae of the stag beetle *Dorcus rectus* (Coleoptera:Luconidae), however, the microorganisms responsible were not identified[96]. Moreover, nitrogenase activity was also demonstrated in *Odontotaenius disjunctus* (Coleopteroa:Passalidae), in this case, the major contributor to nitrogen fixation was a member of the Bacteroidetes, *Paludibacter propionicigenes*, according to *nif* sequence analysis[97]. The microbiome of *Veturius* sp., also a wood feeding beetle, is likely to have symbiotic nitrogen fixers; the nitrogenase sequences found in this study include bacteria and archaeal homologs. One of these homologs was included in the archaeal MAG260, binned from the adult metagenome of *Veturius* sp. To our knowledge, this is the first report of nitrogen fixing archaea in insect gut microbiomes. The in-situ activity, as well as the isolation of this archaea in culture, is yet to be experimentally proven.

8.5. Model of cooperative cellulose degradation in *Veturius* sp.

Wood feeding animals rely on microbes to degrade recalcitrant cellulose and lignin molecules into more easily digestible compounds for the host. Therefore, the microbial communities in the gut of Passalid beetles, and other animals, heavily influence their ecology in terms of habitat selection, diet and behavior. Based on our findings, we propose a model to explain how the enrichment of specific microbial taxa involved in carbon metabolism in the gut of the Passalid beetles contributed in the evolution of the characteristic subsocial behavior of this Coleopteran family (Figure 27). In the model we propose that the larvae are more intensively contributing to the cellulose breakdown

in the system. The beetle larvae are very voracious and rapidly growing[10,12], thus, the larvae have a high energy demand to produce biomass. Additionally, the larval gut occupies the majority of its body, consisting of a tube that lacks anatomical compartmentalization. It is because of this features that we believe that the gut contents transit rapidly thought the gut preventing for a complete digestion of the plant material before its ejection in the feces. Another piece of empirical support is the important level of integrity of the gut contents on the direct observations of the material inside the larval gut at the moment of extraction. Therefore, these feces of the larvae are rich in smaller sugars product of cellulose breakdown that accumulates in the substrate surrounding the galleries that *Veturius* sp. inhabits inside decay tree logs. Both larvae and adults then feed on the more easily digestible sugars in the substrate. The larvae eventually completely transform cellulose to CO_2 and CH_4 by the metabolic action of Firmicutes and methanogenic archaea. The adult metagenomes, on the other hand, were significantly depleted in cellulases like GH5, GH10 and others involved in the first steps of cellulose breakdown; this is why we believe they mostly consume the small sugars released by the larvae. Thus, the subsocial behavior of the Passalid can be explained by the dynamics of their relationship where the larvae assist with the digestion of the adult, and the adult provides the larvae with protection and a habitat.

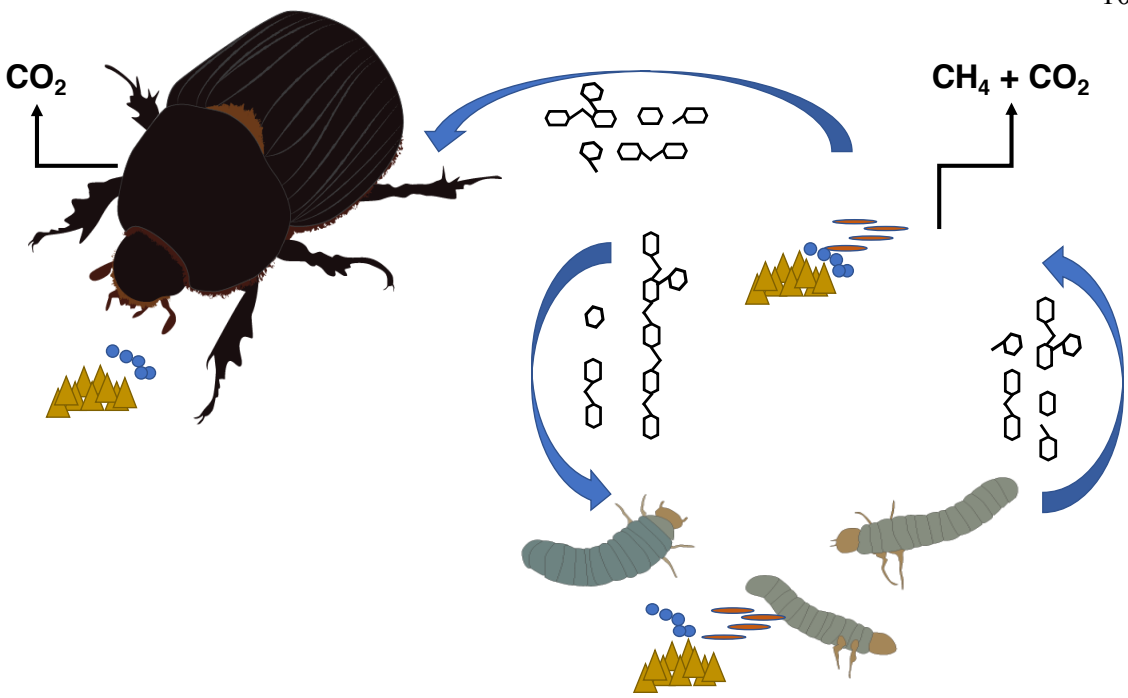


Figure 27. Proposed model of microbe-assisted cellulose degradation in the Passalid beetle *Veturius* sp.

It is important to consider the limitations of our models as it was purely created based on a sample set of metagenomes and 16S rRNA data for a particular time in five families of Passalid beetles. The validation of the model requires future experiments involving gene expression measurements and controlled laboratory experiments.

This study aims to contribute with our knowledge of the biological diversity of Costa Rica, and to obtain a better understanding the contribution of microorganisms that inhabit the guts of beetles in the forest in biogeochemical processes with global relevance. This work also contributes with a dataset of partial genomes that can be further used to better understand the ecology and natural history of these and other beetles in the tropics. The genome collection and the assemblies generated in this project are a source of thousands of genes with potential biotechnological interest

worth to further explore. These enzymes can be potentially use in the generation of clean energy, cooperating to the country efforts towards carbon neutrality and conservation.

9 Conclusions

- The DNA isolation methods tested did not affect the yield and the diversity of the sequences. However, the MoBio protocol produced higher yield and more consistent results.
- The gut of the *Veturius* sp. adult, larvae and their substrate harbor different microbial communities enriched in specific taxonomic groups. The main driver of the functional and taxonomical variation is the sample type itself.
- Adult and larval gut microbial communities are dominated by Firmicutes that encode for thousands of carbohydrate active enzymes.
- The larval microbial communities are enriched in methanogenic archaea.
- The *Veturius* sp. harbor bacteria and archaea capable to perform methanogenesis and nitrogen fixation. These processes have not been previously documented in this insect.
- I found evidence of functional convergence evolution between the larvae and the cow rumen. Both systems selected for microbes with similar enzymatic capabilities for cellulose degradation.
- The *Veturius sp*. gut microbial communities influence the ecology and physiology of these beetles.

10 References

1. Shi, W., Syrenne, R., Sun, J.-Z. & Yuan, J. S. Molecular approaches to study the insect gut symbiotic microbiota at the 'omics' age. *Insect Sci*. **17,** 199–219 (2010).
2. Zhi-qiang Zhang. Animal biodiversity : An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* 1–237 (2011).
3. Smith, D. M. & Marcot, J. D. The fossil record and macroevolutionary history of the beetles. *Proc. R. Soc. B* **282,** (2015).
4. Hunt, T. *et al*. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318,** 1913–6 (2007).
5. Solís, A. *Escarabajos de Costa Rica*. (INBio: National Institute of Biodiversty, 2002).
6. Grimaldi, D. & Engel, M. S. *Evolution of the Insects*. (Cambridge Univ. Press, 2005).
7. Reyes-castillo, A. P., Schuster, J. C., The, S., Bulletin, C. & Mar, N. The Coleopterists Society Notes on Some Mesoamerican Passalidae (Coleoptera): Petrejoides and Pseudacanthus. **37,** 49–54 (2014).
8. Reyes-Castillo, P. Coleoptera Passalidae de Mexico. *PRIBES Hacia un Proy. CYTED para el Inventar. y Estimación la Divers. Entomológica en Iberoamérica* (2000).
9. Jiménez-Ferbans, L., Reyes-Castillo, P., Schuster, J. C. & Beza-Beza, C. Revista Mexicana de Biodiversidad The passalid beetles (Coleoptera : Passalidae) from Costa Rica , with the description of two new species of Passalus. *Rev. Mex. Biodivers*. **88,** 608–615 (2017).
10. Solís, A. *Escarabajos de Costa Rica Beetles*. (National Institution of Biodiversity, 1999).
11. Castillo, M. L. & Reyes-Castillo, P. Copulation in Natura of Passalid Beetles (Coleoptera:

Passalidae). *Coleopt. Bull.* **43,** 162–164 (1989).

12. Reyes-Castillo, P. & Schuster, J. C. Notes on Some Mesoamerican Passalidae (Coleoptera): Petrejoides and Pseudacanthus. *Coleopt. Bull.* **37,** 49–54 (1983).
13. Mason, W. & Odum, E. The effect of Coprophagy on retention and bioelimination of radionucloetids by detritus-feeding animals. *Radioecolgy* 721 – 724 (1969).
14. Mccutcheon, J. P. & Moran, N. A. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. (2007).
15. Moran, N. a. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl. Acad. Sci. U. S. A.* **104 Suppl,** 8627–33 (2007).
16. Tamas, I. *et al.* 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science* **296,** 2376–9 (2002).
17. Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450,** 560–5 (2007).
18. Suh, S. O., Marshall, C. J., McHugh, J. V. & Blackwell, M. Wood ingestion by passalid beetles in the presence of xylose-fermenting gut yeasts. *Mol. Ecol.* **12,** 3137–3145 (2003).
19. He, S. *et al.* Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. *PLoS One* **8,** (2013).
20. Liu, N. *et al.* Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *ISME J.* (2018). doi:10.1038/s41396-018-0255-1
21. Scully, E. D. *et al.* Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. *PLoS One* **8,** e73827 (2013).

22. Geib, S. M. *et al*. Microbial community profiling to investigate transmission of bacteria between life stages of the wood-boring beetle, Anoplophora glabripennis. *Microb. Ecol.* **58,** 199–211 (2009).

23. Andert, J., Marten, A., Brandl, R. & Brune, A. Inter- and intraspecific comparison of the bacterial assemblages in the hindgut of humivorous scarab beetle larvae (Pachnoda spp.). *FEMS Microbiol. Ecol.* **74,** 439–49 (2010).

24. Nardi, J. B. *et al*. Communities of microbes that inhabit the changing hindgut landscape of a subsocial beetle. *Arthropod Struct. Dev.* **35,** 57–68 (2006).

25. Tremaroli, V. & Backhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* 1–8 (2012). doi:10.1038/nature11552

26. Zheng, Y. *et al*. A pathway for biological methane production using bacterial iron-only nitrogenase. *Nat. Microbiol.* **3,** 281–286 (2018).

27. Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: Ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6,** 579–591 (2008).

28. Worden, A. Z. *et al*. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science (80-.).* **347,** (2015).

29. Archer, D. Carbon cycle: Checking the thermostat. *Nat. Geosci.* **1,** 289–290 (2008).

30. Metcalfe, D. *et al*. Herbivory makes major contributions to ecosystem carbon and nutrient cycling in tropical forests. *Ecol. Lett.* **17,** (2014).

31. Chau, C. K., Leung, T. M. & Ng, W. Y. A review on Life Cycle Assessment , Life Cycle Energy Assessment and Life Cycle Carbon Emissions Assessment on buildings. *Appl.*

Energy **143,** 395–413 (2015).

32. Kemper, J. Biomass and carbon dioxide capture and storage: A review. *Int. J. Greenh. Gas Control* **40,** 401–430 (2015).

33. Cherubini, F., Peters, G. P., Berntsen, T., Strømman, A. H. & Hertwich, E. CO2 emissions from biomass combustion for bioenergy: atmospheric decay and contribution to global warming. *GCB Bioenergy* **3,** 413–426 (2011).

34. Horn, S. J., Vaaje-kolstad, G., Westereng, B. & Eijsink, V. G. H. Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* **5,** (2012).

35. Klein-Marcuschamer, D., Oleskowicz-popiel, P., Simmons, B. A. & Blanch, H. W. The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels. *Biotechnol. Bioeng.* **109,** 1083–1087 (2012).

36. Park, B., Karpinets, T. V, Syed, M. H., Leuze, M. R. & Uberbacher, E. C. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20,** 1574–1584 (2010).

37. Carrard, G., Koivula, A., Soderlund, H. & Beguin, P. Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *PNAS* **97,** (2000).

38. Aspeborg, H., Coutinho, P. M., Wang, Y., Iii, H. B. & Henrissat, B. Evolution , substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12,** (2012).

39. Bayer, E. A., Chanzyt, H., Lamed, R. & Shoham, Y. Cellulose , cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8,** 548–557 (1998).

40. Mewis, K., Lenfant, N. & Lombard, V. Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies : a Motivation for Detailed Enzyme Characterization. *Appl. Environ. Microbiol.* **82,** 1686–1692 (2016).

41. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Curr. Biol.* **3,** (1995).

42. Bayer, E. A. Cellulosomes and designer cellulosomes : why toy with Nature ? *Environ. Microbiol. Rep.* **9,** 14–15 (2017).

43. Artzi, L., Bayer, E. A. & Moraïs, S. Cellulosomes : bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **15,** 83–95 (2016).

44. Shi, W., Syrenne, R., Sun, J. & Yuan, J. S. Molecular approaches to study the insect gut symbiotic microbiota at the ' omics ' age. *Insect S* **17,** 199–219 (2010).

45. Suen, G. *et al*. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6,** (2010).

46. Egert, M. *et al*. Structure and Topology of Microbial Communities in the Major Gut Compartments of Melolontha melolontha Larvae (Coleoptera : Scarabaeidae) Structure and Topology of Microbial Communities in the Major Gut Compartments of Melolontha melolontha Larvae (Col. (2005). doi:10.1128/AEM.71.8.4556

47. Egert, M., Wagner, B., Lemke, T., Brune, A. & Friedrich, M. W. Microbial Community Structure in Midgut and Hindgut of the Humus-Feeding Larva of Pachnoda ephippiata (Coleoptera : Scarabaeidae). **69,** 6659–6668 (2003).

48. Lemke, T., Stingl, U., Egert, M., Friedrich, M. W. & Brune, A. Physicochemical Conditions and Microbial Activities in the Highly Alkaline Gut of the Humus-Feeding

Larva of Pachnoda ephippiata (Coleoptera : Scarabaeidae). **69,** 6650–6658 (2003).

49. Vasanthakumar, A. *et al*. Characterization of Gut-Associated Bacteria in Larvae and Adults of the Southern Pine Beetle, <I>Dendroctonus frontalis</I> Zimmermann. *Environ. Entomol.* **35,** 1710–1717 (2006).

50. Burnum, K. E. *et al*. Proteome insights into the symbiotic relationship between a captive colony of Nasutitermes corniger and its hindgut microbiome. *ISME J.* **5,** 161–164 (2011).

51. Breznak, J. a & Brune, A. Role of microorganisms in the digestion of lignocellulose by termites. *Ann. Rev. Entomol.* **39,** 453–487 (1994).

52. Hess, M. *et al*. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331,** 463–7 (2011).

53. Jewell, K. A., McCormick, C. A., Odt, C. L., Weimer, P. J. & Suen, G. Ruminal bacterial community composition in dairy cows is dynamic over the course of two lactations and correlates with feed efficiency. *Appl. Environ. Microbiol.* **81,** 4697–4710 (2015).

54. Jin, W., Cheng, Y. & Zhu, W. The community structure of Methanomassiliicoccales in the rumen of Chinese goats and its response to a high- grain diet. *J. Anim. Sci. Biotechnol.* **8,** 1–10 (2017).

55. Kamke, J. *et al*. Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea - enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* **4,** 1–16 (2016).

56. Shi, W. *et al*. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* 1517–1525 (2014). doi:10.1101/gr.168245.113.1

57. Pope, P. B. *et al*. Adaptation to herbivory by the Tammar wallaby includes bacterial and

glycoside hydrolase pro fi les different from other herbivores. (2010). doi:10.1073/pnas.1005297107

58. Sanders, J. *et al*. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat. Commun*. 8 (2015).

59. Presidencia, R. de C. R. La Gaceta. Lunes 11 de Noviembre 2013. *217* 1–100 (2013).

60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads,. *EMBnet J*. **17,** 10–12 (2011).

61. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2015).

62. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

63. Li, H. *et al*. The Sequence Alignment / Map format and SAMtools. **25,** 2078–2079 (2009).

64. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3,** e1165 (2015).

65. Eren, A. M. *et al*. Anvi ' o : an advanced analysis and visualization platform for ' omics data. 1–29 (2015). doi:10.7717/peerj.1319

66. Hyatt, D. *et al*. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11,** (2010).

67. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12,** 59–60 (2014).

68. Campbell, J. H., Donoghue, P. O., Campbell, A. G., Schwientek, P. & Sczyrba, A. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl. Acad. Sci.* **110,** 5540–5545 (2013).

69. Arkin, A. P., Cottingham, R. W., Henry, C. S. & et al. KBase : The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36,** (2018).

70. Chen, I. A. *et al.* IMG / M v . 5 . 0 : an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 1–12 (2018). doi:10.1093/nar/gky901

71. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5,** 1–11 (2014).

72. Markowitz, V. M. *et al.* IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* **40,** D123-9 (2012).

73. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).

74. Lang, K., Schuldes, J., Klingl, A., Poehlein, A. & Daniel, R. New Mode of Energy Metabolism in the Seventh Order of Methanogens as Revealed by Comparative Genome Analysis of " Candidatus Methanoplasma termitum ". **81,** 1338–1352 (2015).

75. Kennedy, N. A. *et al.* The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing. *PLoS One* **9,** 1–9 (2014).

76. Zaveri, P., Patel, R., Patel, M., Sarodia, D. & Munshi, N. S. Modification of extraction method for community DNA isolation from salt affected compact wasteland soil samples.

MethodsX **4,** 63–67 (2017).

77. Loreille, O. M., Diegoli, T. M., Irwin, J. A., Coble, M. D. & Parsons, T. J. High efficiency DNA extraction from bone by total demineralization. *Forensic Sci. Int.* **1,** 191–195 (2007).

78. Miller, D. N., Bryant, J. E., Madsen, E. L. & Al, M. E. T. Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples. *Appl. Environ. Microbiol.* **65,** 4715–4724 (1999).

79. Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *PLoS One* **7,** (2012).

80. Albertsen, M., Karst, S. M., Ziegler, A. S. & Kirkegaard, R. H. Back to Basics – The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. *PLoS One* 1–15 (2015). doi:10.1371/journal.pone.0132783

81. Nnadozie, C. F., Lin, J. & Govinden, R. Optimisation of protocol for e ff ective detachment and selective recovery of the representative bacteria for extraction of metagenomic DNA from Eucalyptus spp . woodchips. *J. Microbiol. Methods* **148,** 155–160 (2018).

82. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters : assessing small subunit rRNA primers for marine microbiomes with mock communities , time series and global field samples. *Environ. Microbiol.* **18,** 1403–1414 (2016).

83. Adams, A. S. *et al*. Mountain Pine Beetles Colonizing Historical and Naïve Host Trees Are Associated with a Bacterial Community Highly Enriched in Genes Contributing to Terpene Metabolism. *Appl. Environ. Microbiol.* **79,** 3468–3475 (2013).

84. Morales-jiménez, J., Zúñiga, G., Villa-Tanaca, L. & Hernandez-Rodriguez, C. Bacterial Community and Nitrogen Fixation in the Red Turpentine Beetle , Dendroctonus valens LeConte (Coleoptera : Curculionidae : Scolytinae). *Invertebr. Microbiol.* **58,** 879–891 (2009).

85. Berasategui, A., Axelsson, K., Nordlander, R. A. N., Schmidt, A. & Kaltenpoth, M. The gut microbiota of the pine weevil is similar across Europe and resembles that of other conifer-feeding beetles. *Mol. Ecol.* **49,** 4014–4031 (2016).

86. Seshadri, R. *et al.* Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* **36,** 359–367 (2018).

87. Beguin, P. & Aubert, J.-P. The biological degradation of cellulose. *FEMS Microbiol. Rev.* **13,** 25–58 (1994).

88. Stewart, R. D. *et al.* Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **9,** 1–11 (2018).

89. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11,** 1144–1146 (2014).

90. Belton, J.-M. *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58,** 268–276 (2012).

91. Wright, G. *et al.* Molecular Diversity of Rumen Methanogens from Sheep in Western Australia. *Appl. Environ. Microbiol.* **70,** 1263–1270 (2004).

92. Breznak, J. A., Brill, W., Mertins, J. & Coppel, H. Nitrogen Fixation in Termites. *Nature* 577–580 (1973).

93. Nardi, J. B., Mackie, R. I. & Dawson, J. O. Could microbial symbionts of arthropod guts

contribute significantly to nitrogen fixation in terrestrial ecosystems ? *J. Insect Physiol.* **48,** 751–763 (2002).

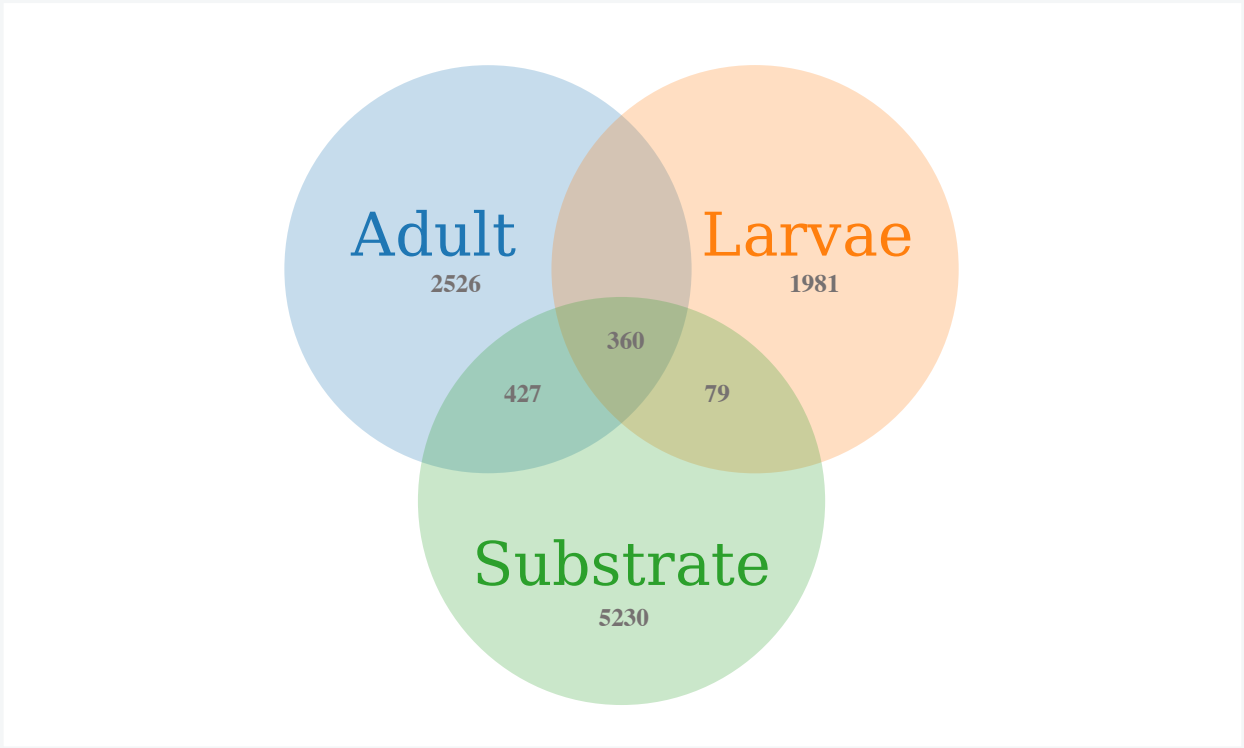
94. Pinto-tomás, A. A. *et al*. Symbiotic Nitrogen Fixation in the Fungus Gardens of Leaf-Cutter Ants. *Science (80-.).* **326,** 1120–1124 (2009).

95. Bridges, J. R. Nitrogen-Fixing Bacteria Associated with Bark Beetles. *Microb. Ecol.* **137,** 131–137 (1981).

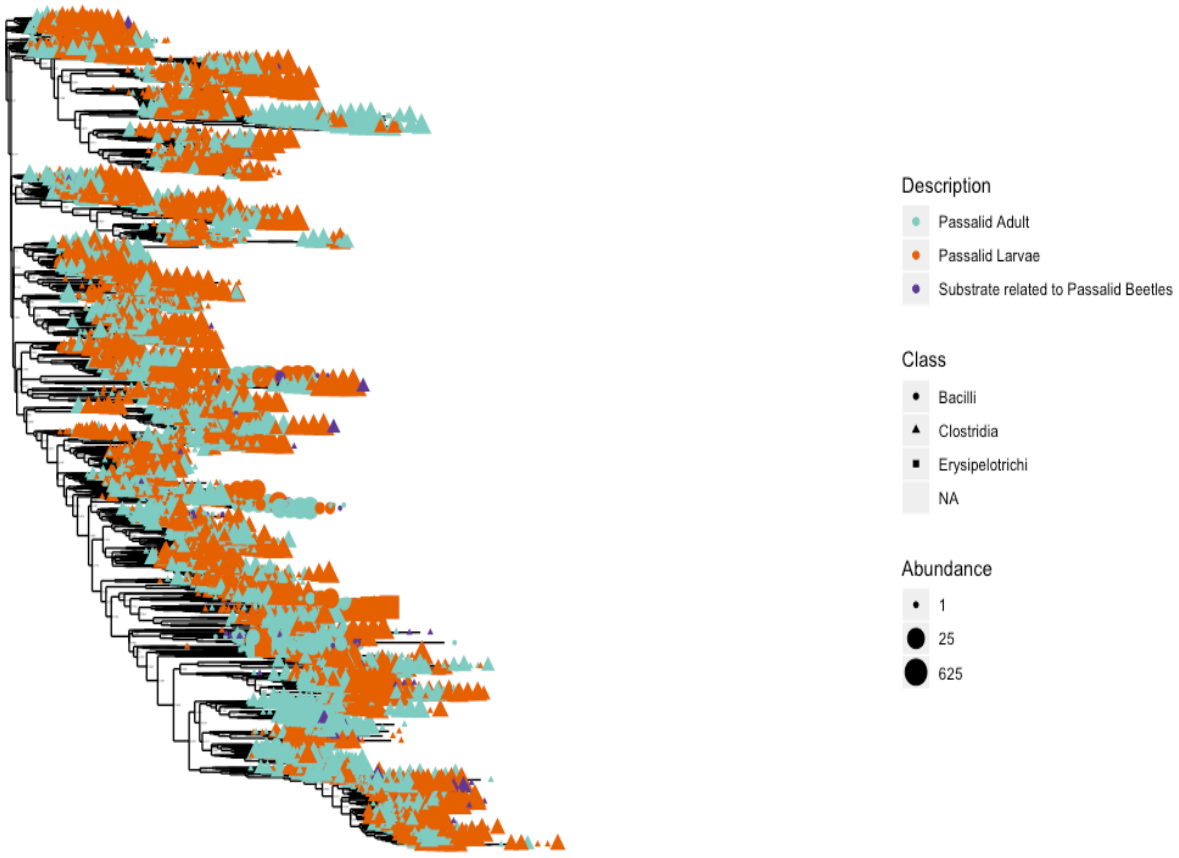
96. Kuranouchi, T. *et al*. Nitrogen fixation in the stag beetle , Dorcus (Macrodorcus) rectus (Motschulsky) (Col ., Lucanidae). *J. Appl. Entomol.* **130,** 471–472 (2006).

97. Ceja-Navarro, J. A. *et al*. Compartmentalized microbial composition , oxygen gradients and nitrogen fixation in the gut of Odontotaenius disjunctus. *ISME* **8,** 6–18 (2014).

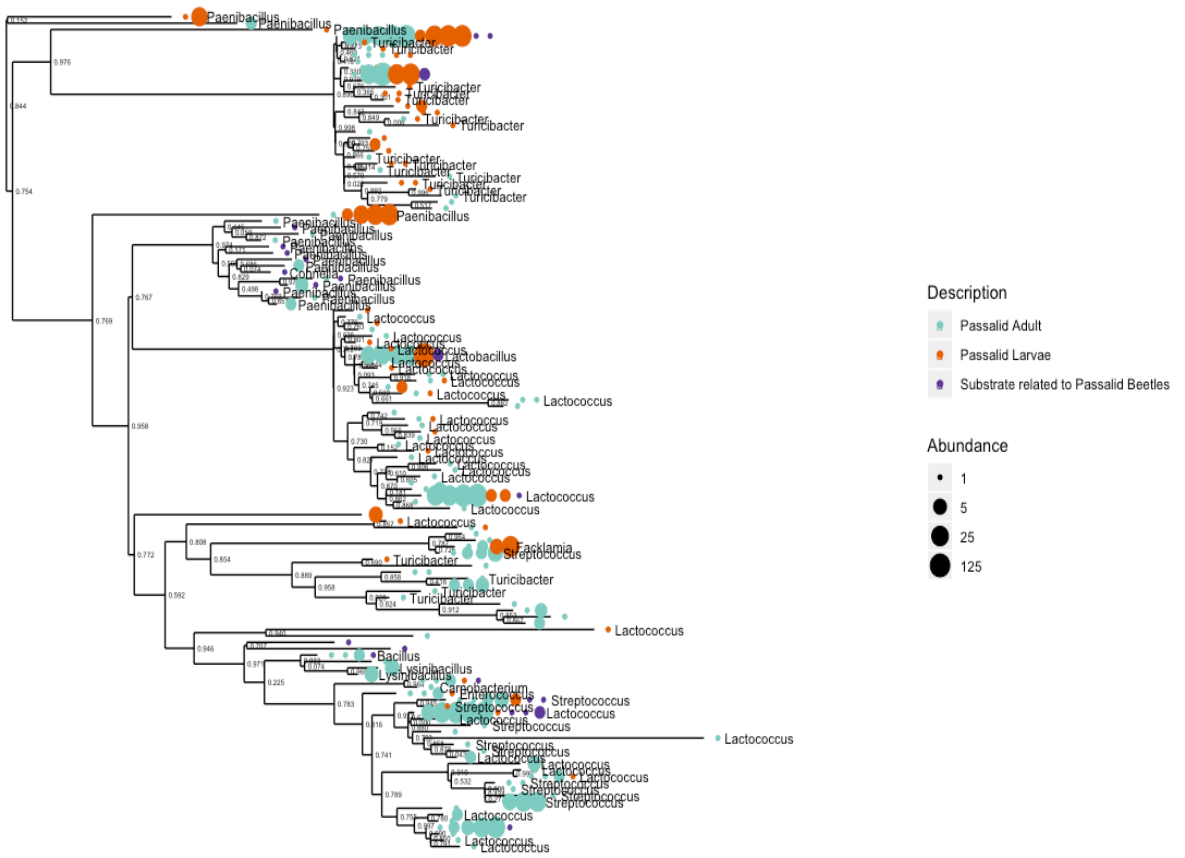
11 Supplementary information



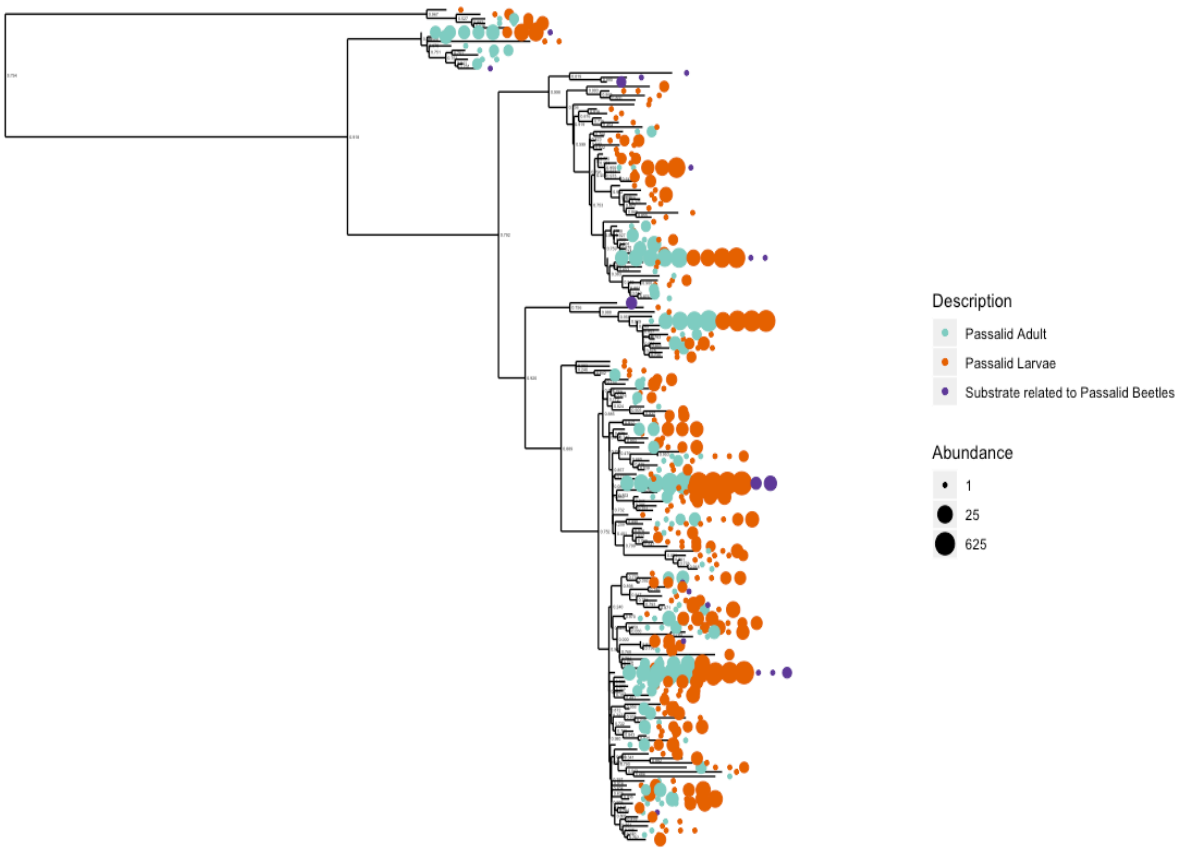
Supplementary figure 1. Venn diagram of OTUs detected in the *Veturius* sp. gut and related substrate.



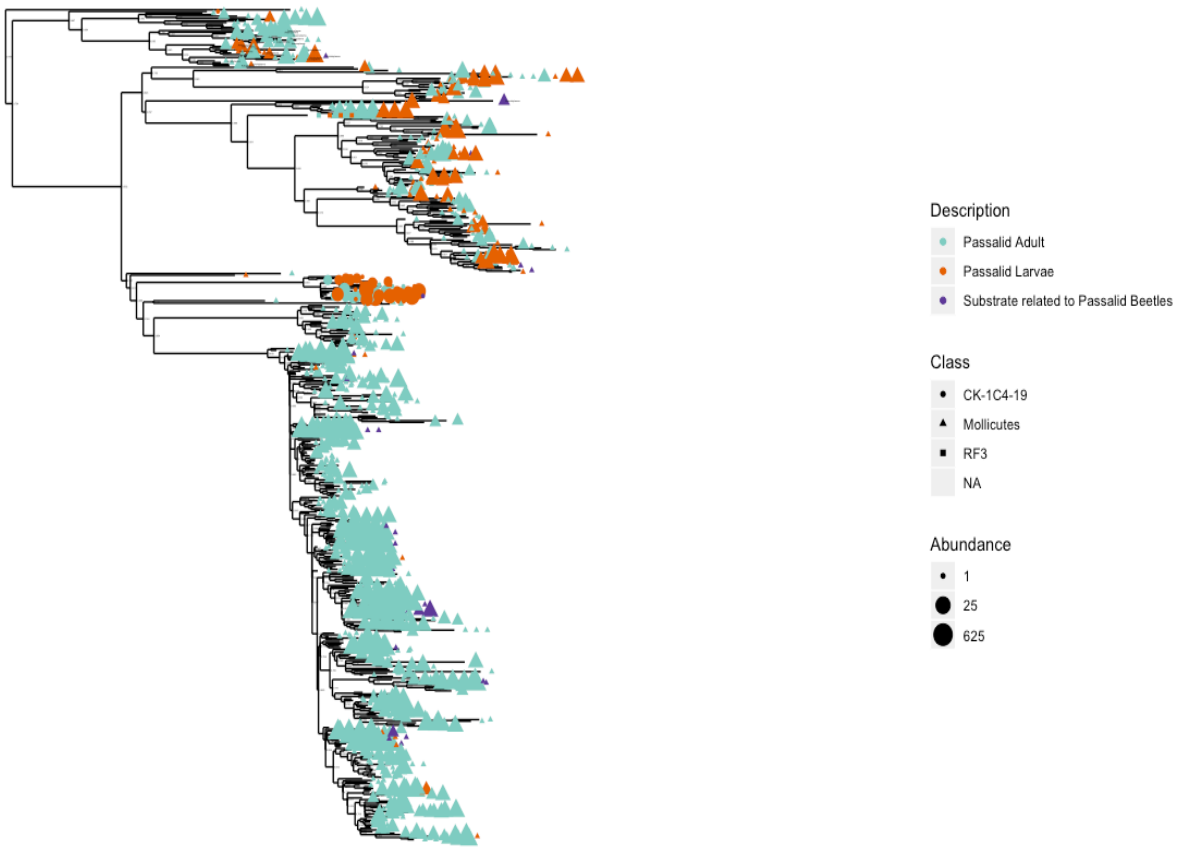
Supplementary figure 2. Neighbor joining tree of Firmicutes in the Passalid beetle gut and substrate. Shape are assigned for OTUs at the class level. The size of the shape correlates with the abundance of sequences in the OUT. Larvae= orange, Adult= cyan, Substrate= purple.



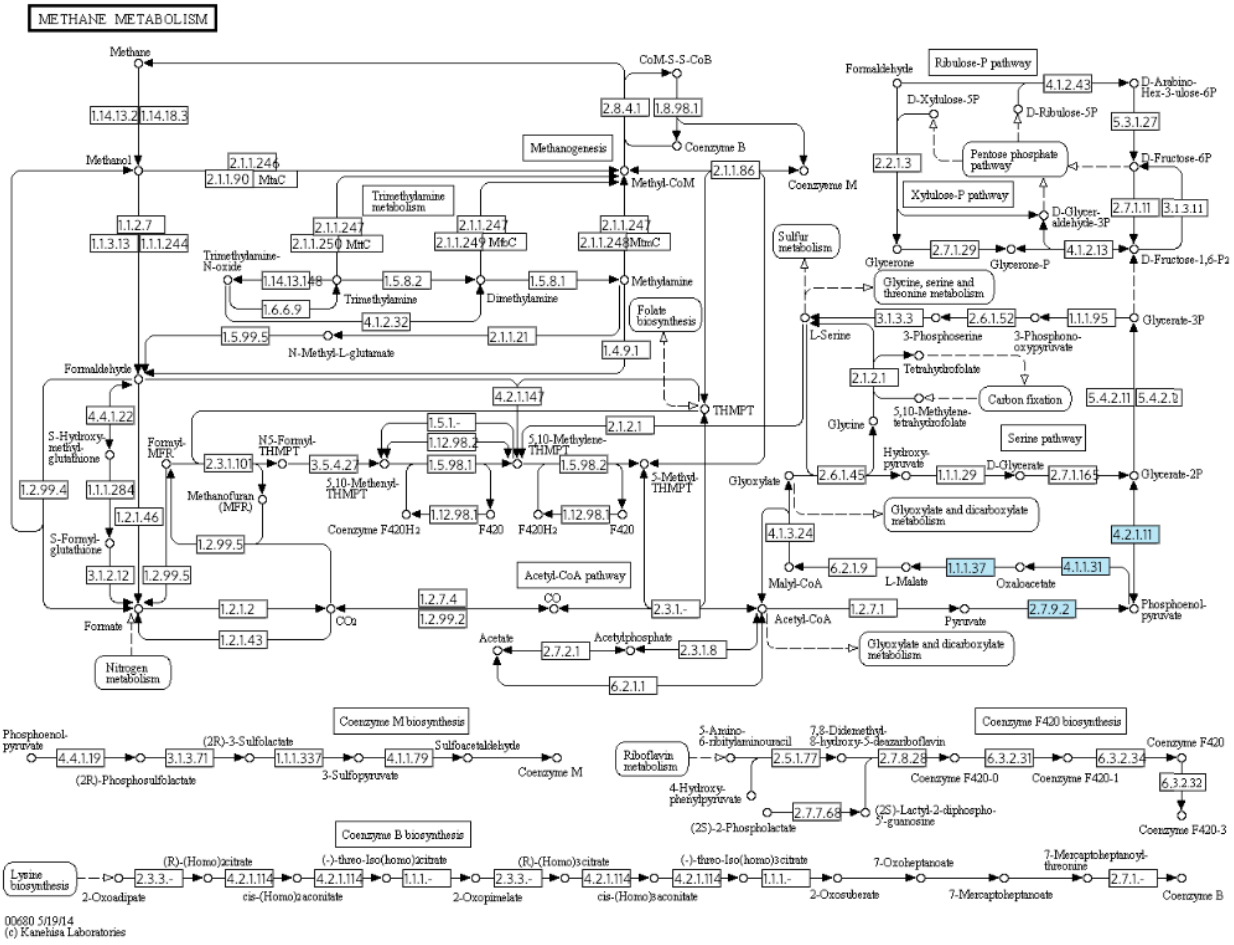
Supplementary figure 3. Neighbor joining tree of Bacilli in the Passalid beetle gut and substrate. Shape are assigned for OTUs at the class level. The size of the shape correlates with the abundance of sequences in the OUT. Larvae= orange, Adult= cyan, Substrate= purple.



Supplementary figure 4. Neighbor joining tree of Euryarchaeota in the Passalid beetle gut and substrate. Shape are assigned for OTUs at the class level. The size of the shape correlates with the abundance of sequences in the OUT. Larvae= orange, Adult= cyan, Substrate= purple.



Supplementary figure 5. Neighbor joining tree of Tenericutes in the Passalid beetle gut and substrate. Shape are assigned for OTUs at the class level. The size of the shape correlates with the abundance of sequences in the OUT. Larvae= orange, Adult= cyan, Substrate= purple.



Supplementary figure 6. Methane metabolism KEGG pathway reconstruction of arqueal MAG260.

Light blue= gene present in the genome.

Supplementary Table 1.

Taxon_oid	Status	Genome Name / Sample Name	Sequencing Center	IMG Genome ID	NCBI Bioproject Accession	Sequencing Method	Genome Size *
3300000064	Draft	Bovine rumen microbial communities (NEW) from Vernon, TX, USA	Texas A and M University	3300000064		Illumina	37196857
2199352026	Draft	Bovine rumen microbial communities from Vernon, TX, USA, Sample 669	Texas A and M University	2199352026		Illumina GAIIx	26083354
2075704000	Draft	Bovine rumen microbial communities from Vernon, TX, USA, Sample 511	Texas A and M University	2075704000		Illumina GAIIx	615647036
2140918032	Draft	Bovine rumen microbial communities from Vernon, TX, USA, Sample 669	Texas A and M University	2140918032		Illumina GAIIx	97713687
3300003523	Permanent Draft	Camel rumen microbial communities from Jandagh-Isfahan, Iran - Sample 1	Beijing Genomics Institute (BGI)	3300003523		Illumina HiSeq 2000	1259982046
3300021254	Draft	Sheep rumen microbial communities from New Zealand - Tag 1494 SPADES assembly	DOE Joint Genome Institute (JGI)	3300021254			2300814014
3300021256	Draft	Sheep rumen microbial communities from New Zealand - Tag 1283 SPADES assembly	DOE Joint Genome Institute (JGI)	3300021256			2245364709
3300021399	Draft	Sheep rumen microbial communities from New Zealand - Tag 1265 SPADES assembly	DOE Joint Genome Institute (JGI)	3300021399			2322534639
3300000079	Draft	Sheep rumen microbial communities from New Zealand - High methane	DOE Joint Genome Institute (JGI)	3300000079	PRJNA202380		160232843
3300021431	Draft	Sheep rumen microbial communities from New Zealand - Tag 1435 SPADES assembly	DOE Joint Genome Institute (JGI)	3300021431			2473673371
3300028327	Permanent Draft	Nasutitermes corniger P3 segment microbial communities from Max Planck Institute, Germany - Nc150P3 (SPAdes)	DOE Joint Genome Institute (JGI)	3300028327	PRJNA366361	Illumina HiSeq 2000, Illumina HiSeq 2500	1301161126
3300002450	Permanent Draft	Termite gut P3 segment microbial communities from Max Planck Institute, Germany - Co191 P3	DOE Joint Genome Institute (JGI)	3300002450	PRJNA405701	Illumina HiSeq 2000, Illumina HiSeq 2500	1316108521
3300002469	Permanent Draft	Termite gut P3 segment microbial communities from Max Planck Institute, Germany - Th196	DOE Joint Genome Institute (JGI)	3300002469	PRJNA405704	Illumina HiSeq 2000, Illumina HiSeq 2500	1212366489
3300006045	Permanent Draft	Termite gut P3 segment microbial communities from Max Planck Institute, Germany - Nt197 P3	DOE Joint Genome Institute (JGI)	3300006045	PRJNA366256	Illumina HiSeq 2000, Illumina HiSeq 2500	3193530320

3300002505	Permanent Draft	Termite gut P3 segment microbial communities from Max Planck Institute, Germany - Nt197 P3	DOE Joint Genome Institute (JGI)	3300002505	PRJNA366256	Illumina HiSeq 2000, Illumina HiSeq 2500	885584058
3300002449	Permanent Draft	Termite gut P3 segment microbial communities from Max Planck Institute, Germany - Mp193 P3	DOE Joint Genome Institute (JGI)	3300002449	PRJNA405703	Illumina HiSeq 2000, Illumina HiSeq 2500	712348561
3300006226	Permanent Draft	Termite gut P3 segment microbial communities from Max Planck Institute, Germany - Th196	DOE Joint Genome Institute (JGI)	3300006226	PRJNA405704	Illumina HiSeq 2000, Illumina HiSeq 2500	3211497060
2044078006	Permanent Draft	Dendroctonus frontalis bacterial communities from Mississippi, USA	DOE Joint Genome Institute (JGI)	2044078006		454 GS FLX Titanium	41255331
2029527007	Permanent Draft	Dendroctonus ponderosae fungus gallery microbial communities from Grand Prairie, Alberta - Hybrid pine	DOE Joint Genome Institute (JGI)	2029527007		454 GS FLX Titanium	28121448
2032320008	Permanent Draft	Dendroctonus ponderosae fungus gallery microbial communities from Grand Prairie, Alberta - MPB hybrid beetle	DOE Joint Genome Institute (JGI)	2032320008		454 GS FLX Titanium	29279222
2038011000	Permanent Draft	Fungus garden microbial communities from Atta colombica in Panama - from dump top	454 Life Sciences, DOE Joint Genome Institute (JGI)	2038011000		454 GS FLX Titanium	465469652
2040502000	Permanent Draft	Fungus garden microbial communities from Atta colombica in Panama - dump bottom	454 Life Sciences, DOE Joint Genome Institute (JGI)	2040502000		454 GS FLX Titanium	482217958
2029527006	Permanent Draft	Fungus garden microbial communities from Atta colombica in Panama -fungus garden bottom	454 Life Sciences, DOE Joint Genome Institute (JGI)	2029527006		454 GS FLX Titanium	83167033
3300001451	Permanent Draft	Goat rumen bacterial communities from Langston, Oklahoma, USA - velvet Assemble	Cornell University	3300001451		Illumina	1986802579
3300001425	Permanent Draft	Goat rumen bacterial communities from Langston, Oklahoma, USA - velvet	Cornell University	3300001425		Illumina	1986802579
3300001376	Permanent Draft	Goat rumen bacterial communities from Langston, Oklahoma, USA	Cornell University	3300001376		Illumina	64563911
2029527005	Permanent Draft	Atta colombica fungus garden Top	DOE Joint Genome Institute (JGI)	2029527005	PRJNA337000	454	100904834
3300000114	Draft	Passalidae beetle gut microbial communities from Costa Rica - Larvae (3ML+3BL)	DOE Joint Genome	3300000114	PRJNA337866	Illumina HiSeq 2000	804003313

			Institute (JGI)				
3300000838	Draft	Passalidae beetle gut microbial communities from Costa Rica - Larvae (5ML+5BL)	DOE Joint Genome Institute (JGI)	3300000838	PRJNA337868	Illumina HiSeq 2000	77567223
2225789003	Draft	Passalidae beetle gut microbial communities from Costa Rica - Larvae (2ML+2BL)	DOE Joint Genome Institute (JGI)	2225789003	PRJNA337865	Illumina HiSeq 2000	52750108
3300000062	Draft	Passalidae beetle gut microbial communities from Costa Rica - Larvae (1ML+1BSL)	DOE Joint Genome Institute (JGI)	3300000062	PRJNA337864	Illumina HiSeq 2000	511653920
3300000036	Draft	Passalidae beetle gut microbial communities from Costa Rica - Gallery material (4MSU+4BSU+3MSU+3BSU)	DOE Joint Genome Institute (JGI)	3300000036	PRJNA337863	Illumina HiSeq 2000	170323120
3300000836	Draft	Passalidae beetle gut microbial communities from Costa Rica - Adult (4MA+4BA+4MSA)	DOE Joint Genome Institute (JGI)	3300000836	PRJNA336847	Illumina HiSeq 2000	142922252
2225789004	Draft	Passalidae beetle gut microbial communities from Costa Rica - Larvae (4BL+4ML+4MSL)	DOE Joint Genome Institute (JGI)	2225789004	PRJNA337867	Illumina HiSeq 2000	402413667
3300005702	Permanent Draft	Rangifer tarandus platyrhynchus rumen microbial communities from Svalbard, Norway - Sample 549	Macrogen	3300005702		454 GS FLX Titanium	1150694
3300028797	Draft	Bovine rumen microbial communities from tropical cattle in Woodstock, Queensland, Australia - Gonzalo_04	DOE Joint Genome Institute (JGI)	3300028797		Illumina HiSeq 2500-1TB, Illumina NovaSeq	2488214290
3300028048	Draft	Bovine rumen microbial communities from Lethbridge, Alberta, Canada - RJG_02	DOE Joint Genome Institute (JGI)	3300028048	PRJNA502607	Illumina HiSeq 2500-1TB	1571803452
3300028886	Draft	Bovine rumen microbial communities from Lethbridge, Alberta, Canada - RJG_04	DOE Joint Genome Institute (JGI)	3300028886		Illumina HiSeq 2500-1TB, Illumina NovaSeq	2155694303
3300026549	Draft	Bovine rumen microbial communities from Lethbridge, Alberta, Canada - RJG_01	DOE Joint Genome Institute (JGI)	3300026549	PRJNA469173	Illumina HiSeq 2500-1TB	1135002002
3300028914	Draft	Bovine rumen microbial communities from tropical cattle in Woodstock, Queensland, Australia - Gonzalo_03	DOE Joint Genome Institute (JGI)	3300028914		Illumina NovaSeq, Illumina HiSeq 2500-1TB	2126034691
3300028888	Draft	Sheep rumen microbial communities from Palmerston North, Manawatu-Wanganui, New Zealand - 1728 DNA GHGlow gp2	DOE Joint Genome Institute (JGI)	3300028888		Illumina HiSeq 2500-1TB, Illumina NovaSeq	3739977822
3300028833	Draft	Sheep rumen microbial communities from Palmerston North, Manawatu-Wanganui, New Zealand - 1742 DNA GHGhigh gp2	DOE Joint Genome Institute (JGI)	3300028833		Illumina HiSeq 2500-1TB, Illumina NovaSeq	3705415280
3300028805	Draft	Sheep rumen microbial communities from Palmerston North, Manawatu-Wanganui,	DOE Joint Genome	3300028805		Illumina HiSeq 2500-1TB,	3185923025

		New Zealand - 1766 DNA GHGlow gp2	Institute (JGI)			Illumina NovaSeq	
3300028832	Draft	Bovine rumen microbial communities from tropical cattle in Woodstock, Queensland, Australia - Gonzalo_01	DOE Joint Genome Institute (JGI)	3300028832		Illumina HiSeq 2500-1TB	3318270658
3300026522	Draft	Bovine rumen microbial communities from Lethbridge, Alberta, Canada - RJG_03	DOE Joint Genome Institute (JGI)	3300026522	PRJNA469174	Illumina HiSeq 2500-1TB	539890075
3300028887	Draft	Bovine rumen microbial communities from tropical cattle in Woodstock, Queensland, Australia - Gonzalo_02	DOE Joint Genome Institute (JGI)	3300028887		Illumina HiSeq 2500-1TB, Illumina NovaSeq	3065012227
3300028591	Draft	Sheep rumen microbial communities from Palmerston North, Manawatu-Wanganui, New Zealand - 1770 DNA GHGhigh gp2	DOE Joint Genome Institute (JGI)	3300028591	PRJNA502469	Illumina HiSeq 2500-1TB	3866467842
2088090023	Draft	Sample MXFH	National Research Council of Canada	2088090023			86883302
2088090024	Permanent Draft	Sample MXSH	National Research Council of Canada	2088090024			90948594
2088090022	Permanent Draft	Sample MXST	National Research Council of Canada	2088090022			86700198
2088090021	Permanent Draft	Sample MXFT-1	National Research Council of Canada	2088090021			75338853
3300014826	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Forg_1366	University of Missouri	3300014826	PRJNA214227		337153485
3300014043	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Forg_1248	University of Missouri	3300014043	PRJNA214227		294134892
3300011989	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Con_7429	University of Missouri	3300011989	PRJNA214227		249485710
3300014047	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Forg_1003	University of Missouri	3300014047	PRJNA214227		312807947
3300014057	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Forg_1208	University of Missouri	3300014057	PRJNA214227		647724957
3300013830	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Con_1239	University of Missouri	3300013830	PRJNA214227		181596033
3300012016	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Forg_1397	University of Missouri	3300012016	PRJNA214227		557664762
3300011977	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Con_1111	University of Missouri	3300011977	PRJNA214227		192833035
3300011979	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Con_1396	University of Missouri	3300011979	PRJNA214227		189845674

3300012007	Draft	Sheep rumen microbial communities from Wyoming, USA - O_aries_Forg_1009	University of Missouri	3300012007	PRJNA214227		421047043
3300000210	Draft	Sheep rumen microbial communities from New Zealand - Low methane emitting sheep	DOE Joint Genome Institute (JGI)	3300000210	PRJNA202380	Illumina HiSeq 2000	212657803
3300001477	Permanent Draft	Sheep rumen microbial communities from New Zealand - Rank02_low	DOE Joint Genome Institute (JGI)	3300001477		Illumina HiSeq 2000	220428912
3300001484	Permanent Draft	Sheep rumen microbial communities from New Zealand - Rank12_low	DOE Joint Genome Institute (JGI)	3300001484		Illumina HiSeq 2000	456493960

Supplementary table 2. Summary of 766 bins reconstructed from the Passalid gut and substrate metagenomes.

Bin	Length (Mb)	contigs	N50	%GC	%completion	%redundancy
Bin 78	4.059491	125	56286	43.04	99.28	0.72
Bin 315	1.643355	38	59693	27.53	99.28	0.72
Bin 340	1.785043	152	17065	51.08	99.28	2.88
Bin 552	2.254344	143	23041	33.34	99.28	4.32
Bin 178	4.475214	99	71984	36.18	97.84	1.44
Bin 438	1.078999	23	77003	38.56	97.12	0.00
Bin 57	4.487127	61	102697	46.02	97.12	0.72
Bin 322	2.224147	222	13019	34.83	97.12	0.72
Bin 191	2.639116	114	33214	52.88	97.12	2.16
Bin 470	2.477736	159	20932	50.18	97.12	2.16
Bin 11	4.077	96	75577	53.18	97.12	3.60
Bin 127	2.762037	90	55910	50.58	97.12	3.60
Bin 505	2.24753	178	17326	41.76	97.12	3.60
Bin 74	1.828126	45	67879	38.90	96.40	2.88
Bin 330_1	1.501955	174	11303	25.93	96.40	6.47
Bin 167	2.969866	28	176920	60.29	95.68	1.44
Bin 485	1.824596	187	13368	32.12	95.68	3.60
Bin 124_1	2.792969	247	17612	35.41	95.68	6.47
Bin 41	2.140966	150	21377	58.30	94.96	0.72
Bin 445	4.87312	112	79576	41.36	94.96	3.60
Bin 397_2	4.254739	312	20569	60.41	94.96	3.60
Bin 405	1.727392	110	25836	27.27	94.96	5.76
Bin 317_1	1.478132	129	15668	41.40	94.96	7.19
Bin 184	2.122613	212	13166	53.06	94.24	1.44
Bin 195	2.071625	131	25436	41.30	94.24	1.44
Bin 270	1.728146	114	22373	52.55	94.24	1.44
Bin 553	1.111015	44	36367	39.27	93.53	0.00
Bin 163	2.294155	88	41857	55.01	92.81	2.16
Bin 123_1	1.928882	160	18352	37.79	92.81	2.16
Bin 316	1.264829	76	21964	29.40	92.81	2.88
Bin 526	2.034256	139	24169	46.54	92.81	6.47

Bin 407	5.446711	295	25761	60.09	92.09	0.00
Bin 503	2.145093	153	18626	54.07	92.09	0.72
Bin 284	3.116704	179	24358	57.57	92.09	1.44
Bin 461	2.90979	163	26579	68.61	92.09	1.44
Bin 361	1.817726	157	14407	56.50	92.09	1.44
Bin 174	3.168773	265	15063	49.08	92.09	2.88
Bin 260	2.163051	60	66296	42.95	91.98	6.79
Bin 379	1.284114	74	24513	45.65	91.37	0.00
Bin 266	3.625865	129	40612	38.88	91.37	0.72
Bin 4	2.107329	99	36416	53.04	91.37	2.16
Bin 211	3.547183	151	39950	64.07	90.65	3.60
Bin 511	2.231954	178	15218	51.16	89.93	2.16
Bin 363	4.425913	391	14809	57.77	89.21	2.16
Bin 398	1.667669	218	8935	54.92	89.21	3.60
Bin 433	1.641575	162	12858	50.03	88.49	0.72
Bin 106	1.713734	173	11847	53.76	88.49	1.44
Bin 64_1	1.442541	93	20121	28.34	88.49	3.60
Bin 399	2.504482	95	56972	34.57	88.49	5.04
Bin 230	2.176001	279	9680	37.48	87.77	0.72
Bin 440	1.745623	175	13785	46.15	87.77	0.72
Bin 442	1.342311	61	35512	25.41	87.05	2.88
Bin 518_1	1.669738	192	11340	37.38	87.05	4.32
Bin 72	2.02646	266	8858	48.10	86.33	3.60
Bin 89	3.267142	65	90990	49.90	85.61	2.16
Bin 113	1.33199	156	10480	39.65	85.61	2.16
Bin 58	1.353409	196	7932	49.22	84.89	0.72
Bin 225_1	3.670433	192	27043	60.04	84.89	3.60
Bin 201	3.020484	168	26512	52.71	83.45	0.00
Bin 14	1.573518	158	12697	49.53	83.45	2.16
Bin 161	2.601191	272	13020	59.77	82.01	2.16
Bin 85	1.307863	75	31481	27.15	81.29	0.00
Bin 504	0.669606	20	60436	42.94	81.29	1.44
Bin 446	1.362137	236	6395	53.30	80.58	0.00
Bin 481	2.634059	170	22640	51.76	80.58	1.44
Bin 250	3.029708	284	14034	66.90	80.58	2.16

Bin 16	2.897163	381	9874	58.11	80.58	2.88
Bin 155	2.586332	230	14613	37.93	80.58	2.88
Bin 272	2.814037	148	24555	40.70	79.86	1.44
Bin 62	0.661441	39	23487	38.68	79.14	0.00
Bin 208	2.241333	131	25527	51.32	78.42	1.44
Bin 232_1	1.210797	114	13147	30.71	78.42	1.44
Bin 86	2.251322	115	26149	55.34	78.42	2.16
Bin 172	1.152459	125	10543	40.44	78.42	2.16
Bin 549	1.920274	206	10868	49.22	78.42	2.88
Bin 143	2.059154	370	6147	37.92	78.42	5.76
Bin 96	2.067989	67	54941	41.82	77.70	0.72
Bin 372	4.351562	403	14336	53.03	76.98	0.72
Bin 502	3.970695	170	38348	62.04	76.98	2.88
Bin 352	0.768602	17	82075	42.94	76.26	0.00
Bin 536	1.582449	368	4308	40.22	76.26	2.88
Bin 288	2.455874	215	16120	54.01	75.54	3.60
Bin 223	2.83604	342	10232	54.32	75.54	5.04
Bin 133	1.616156	103	20840	54.85	74.82	0.72
Bin 206	2.648174	335	10027	39.04	73.38	2.16
Bin 193	2.17889	175	16450	51.94	73.38	2.16
Bin 217	2.841096	220	18244	51.70	73.38	2.88
Bin 522	1.455003	116	17665	33.58	72.66	0.00
Bin 202	2.084695	381	6004	39.54	72.66	0.72
Bin 23	1.350307	75	29196	27.68	72.66	0.72
Bin 519	3.346523	450	9284	43.05	72.66	1.44
Bin 406_1	1.281203	184	8045	27.86	72.66	2.88
Bin 301	1.750208	136	25956	38.19	72.66	4.32
Bin 248	1.570354	193	10065	53.25	71.94	2.16
Bin 508_1	3.793885	380	12401	61.88	71.94	2.88
Bin 17	1.546744	279	6096	35.60	71.94	3.60
Bin 308	1.494432	216	8487	31.50	71.22	2.88
Bin 139_2	3.240482	442	9080	47.81	71.22	4.32
Bin 83	2.531595	254	14000	42.63	71.22	4.32
Bin 357	1.191913	150	9727	52.93	70.50	0.00
Bin 134	1.21814	195	7104	49.94	70.50	0.72

Bin 24	1.67139	148	15144	42.48	70.50	1.44
Bin 441	1.621338	290	6028	54.30	70.50	1.44
Bin 207_1	2.497164	213	15131	48.04	70.50	3.60
Bin 82	2.006976	156	18796	48.48	69.78	0.00
Bin 452	1.872703	277	7959	41.01	69.78	0.72
Bin 43	2.059199	178	15276	47.25	69.78	2.88
Bin 157	2.483054	301	10138	42.07	69.06	2.88
Bin 293	1.457157	234	7187	57.67	69.06	3.60
Bin 20	2.466846	172	21343	54.48	69.06	4.32
Bin 213	1.589973	281	6378	56.04	68.35	0.00
Bin 545	1.84384	117	21858	37.04	68.35	0.72
Bin 205	0.994471	164	7068	35.97	68.35	0.72
Bin 371	1.795609	133	18333	49.38	68.35	2.16
Bin 391	2.317005	272	10044	50.94	68.35	6.47
Bin 478	1.243558	129	11474	51.14	67.63	0.00
Bin 135	2.61259	208	19389	52.64	67.63	3.60
Bin 140	2.340999	131	29767	47.95	66.91	1.44
Bin 510	0.899961	104	10357	45.72	66.91	2.16
Bin 413	3.71797	439	10433	49.39	66.91	2.88
Bin 344	4.355285	440	15921	59.97	66.91	5.76
Bin 67_3	1.744328	260	7988	34.38	66.19	2.16
Bin 294	2.649591	349	9164	40.20	66.19	3.60
Bin 457	2.81087	334	10898	40.21	65.47	2.16
Bin 138_1	2.791617	357	9865	32.21	65.47	5.04
Bin 121	1.901052	258	8850	51.40	64.75	2.16
Bin 253	2.39101	355	8040	51.75	64.75	5.04
Bin 235_2	3.249813	441	8787	43.33	64.75	5.76
Bin 480	1.092801	181	6467	36.04	64.03	1.44
Bin 153	1.455869	95	22177	48.44	64.03	4.32
Bin 245_2	2.848732	494	6561	35.87	64.03	6.47
Bin 460	0.962737	198	5071	34.51	63.31	4.32
Bin 240	2.247855	347	7522	52.21	62.59	2.16
Bin 181	3.087274	266	16815	41.05	62.59	4.32
Bin 137	1.186618	142	10731	26.45	62.59	4.32
Bin 383	2.207743	322	8208	39.40	62.59	5.04

Bin 472	2.985561	404	9337	40.74	62.59	5.76
Bin 171	2.995673	359	10974	60.16	61.15	0.72
Bin 378	2.017208	214	12150	48.13	61.15	0.72
Bin 144	1.966208	229	11050	57.81	61.15	1.44
Bin 198	1.039959	241	4386	43.23	61.15	1.44
Bin 499	1.750481	228	9090	46.67	61.15	4.32
Bin 346	2.553383	421	6784	70.36	60.43	0.00
Bin 91	0.52394	37	19929	48.90	60.43	0.00
Bin 426	3.233951	107	48576	62.98	60.43	2.16
Bin 258	2.434398	413	6823	38.68	60.43	3.60
Bin 408_2	3.345295	598	5988	67.11	60.43	5.04
Bin 281	3.288507	256	17989	69.19	60.43	7.19
Bin 241	4.964589	448	14675	65.53	59.71	2.16
Bin 366	4.057355	343	16094	41.12	59.71	7.19
Bin 151	1.038332	227	4844	39.82	58.99	0.72
Bin 469	1.977611	228	11423	41.61	58.99	1.44
Bin 466_3	2.387109	203	15119	64.94	58.99	4.32
Bin 541	1.413049	164	10738	27.83	58.99	4.32
Bin 544	1.212752	253	5021	41.71	58.27	0.72
Bin 537	1.809605	319	6235	52.89	58.27	3.60
Bin 404	2.098512	285	9179	43.26	56.83	0.00
Bin 2	1.640554	281	6455	49.77	56.83	0.72
Bin 495_1	2.130047	246	11197	49.97	56.83	5.76
Bin 233	0.870713	47	30051	25.48	56.12	0.00
Bin 475	0.615721	86	9195	41.55	56.12	0.00
Bin 165	3.018961	280	14761	69.04	56.12	2.88
Bin 26	1.721404	261	7716	30.65	56.12	3.60
Bin 434	0.831645	173	5118	61.49	55.40	0.00
Bin 314	1.804157	197	12117	48.34	55.40	3.60
Bin 150	1.482521	184	10217	27.89	55.40	3.60
Bin 348	1.027575	122	9507	50.57	54.94	2.47
Bin 52	0.587023	100	6431	34.79	54.68	0.00
Bin 162	1.886773	335	6056	56.73	54.68	0.72
Bin 105	0.605977	126	5107	42.89	54.68	0.72
Bin 276	2.275768	338	8341	52.30	53.96	0.72

Bin 419	1.167591	256	4510	43.78	53.96	2.16
Bin 403	2.180322	396	6199	34.45	53.96	5.04
Bin 354	2.11223	300	8185	56.95	53.24	0.72
Bin 136	1.185556	251	4949	31.41	53.24	5.04
Bin 142	2.249594	287	9650	55.89	52.52	1.44
Bin 525	1.800569	304	6502	43.04	52.52	2.16
Bin 210_1	1.798268	326	6265	63.99	52.52	5.76
Bin 219_1	1.125378	273	4195	66.70	52.52	7.19
Bin 194	1.976427	240	10528	41.65	51.80	0.72
Bin 104	2.00188	255	10221	47.12	51.80	2.16
Bin 368_2	3.110501	383	10216	59.20	51.80	2.88
Bin 410_1	2.765708	327	10753	59.60	51.80	2.88
Bin 395	0.986023	208	4925	43.33	51.80	4.32
Bin 389	0.519342	107	5123	37.75	51.08	0.00
Bin 33	3.129955	381	10352	56.27	51.08	2.16
Bin 59	1.142205	160	8828	37.26	51.08	2.88
Bin 280	2.815397	561	5423	43.20	51.08	7.19
Bin 336	1.294482	229	6181	44.62	50.36	1.44
Bin 158	2.569999	378	8254	40.59	50.36	5.76
Bin 423	2.25367	531	4199	69.27	49.64	0.00
Bin 239	0.607658	81	8714	26.56	49.64	0.00
Bin 159	1.778446	241	9254	48.16	49.64	1.44
Bin 249	2.761357	462	6815	33.59	49.64	5.76
Bin 387	2.015641	314	6989	60.17	48.92	0.72
Bin 261	1.580199	372	4165	61.71	48.92	3.60
Bin 246	0.362533	6	63937	39.59	48.20	0.00
Bin 339	1.548233	237	7058	52.99	48.20	0.72
Bin 228	1.519536	322	5076	27.99	48.20	6.47
Bin 302	1.22176	310	3927	62.11	47.48	0.00
Bin 283	1.189047	183	7303	44.73	47.48	0.00
Bin 7	1.009723	195	5649	48.69	47.48	0.00
Bin 319	1.980545	380	5630	38.29	47.48	5.76
Bin 219_4	4.665721	457	16573	67.36	47.48	7.19
Bin 509	1.11292	214	5681	53.57	46.76	0.72
Bin 338_5	2.94434	343	10756	61.86	46.76	1.44

Bin 273	1.375063	283	5086	37.07	46.76	1.44
Bin 192	0.607282	125	5027	39.69	46.76	2.16
Bin 422	2.889441	639	4774	63.45	46.76	3.60
Bin 291_2	1.904538	410	4874	40.41	46.76	5.04
Bin 231_4	3.207577	208	24141	58.76	46.76	6.47
Bin 444_1	1.95751	350	5981	50.36	46.76	7.91
Bin 189_3	2.423909	293	10050	65.54	46.04	0.72
Bin 298_2	1.824113	268	7858	39.08	46.04	5.04
Bin 512_1	2.153194	371	6390	56.62	46.04	7.19
Bin 92_2	1.657494	365	4604	32.91	46.04	9.35
Bin 429	1.316987	243	6050	40.96	45.32	0.00
Bin 353	0.493503	62	9193	42.74	45.32	0.00
Bin 520_1	1.183599	240	5363	42.34	45.32	1.44
Bin 298_1	1.030341	208	5023	39.92	45.32	1.44
Bin 216	1.702459	395	4368	36.28	45.32	3.60
Bin 190	1.833348	301	6967	46.93	45.32	5.76
Bin 282	4.162415	767	5772	57.87	45.32	7.91
Bin 154_1	2.199717	503	4440	59.13	45.32	8.63
Bin 18	1.142318	215	5764	53.64	44.60	0.00
Bin 455	2.024928	405	5254	55.88	44.60	2.88
Bin 337	4.983503	943	5669	64.23	43.88	0.00
Bin 267	1.8331	280	7345	57.68	43.88	0.00
Bin 179	0.575875	105	5831	34.51	43.88	0.00
Bin 147	1.23268	236	5508	35.34	43.88	1.44
Bin 323	2.676289	369	8824	54.91	43.88	2.16
Bin 297	1.80484	270	7448	30.12	43.88	2.16
Bin 117_2	2.517584	388	7338	44.14	43.88	7.91
Bin 374	0.631403	127	5373	40.49	43.17	0.72
Bin 222	3.347851	709	5066	44.60	43.17	2.88
Bin 484	1.900898	387	5264	49.04	43.17	5.04
Bin 51	2.012832	478	4122	34.44	43.17	6.47
Bin 345_4	0.792659	106	10152	28.36	43.17	10.07
Bin 53	1.570902	374	4124	38.30	42.45	0.00
Bin 271	2.071502	330	7010	65.07	42.45	0.72
Bin 516_1	2.615587	341	11064	37.93	42.45	6.47

Bin 237_2	2.810284	159	25547	67.76	41.73	0.00
Bin 13	1.300732	319	4082	40.47	41.73	0.00
Bin 269	0.673108	137	5141	62.37	41.73	0.72
Bin 231_3	2.745338	180	20757	61.38	41.73	1.44
Bin 148	1.651691	245	7888	30.67	41.73	1.44
Bin 464_2	1.120027	111	14047	69.00	41.73	1.44
Bin 236	2.546217	408	7323	66.33	41.73	2.16
Bin 531_3	1.108292	181	7035	43.84	41.73	2.88
Bin 54_2	1.201301	97	16897	41.60	41.73	5.04
Bin 345_6	1.088092	132	10032	26.74	41.73	6.47
Bin 300	0.95049	83	13575	56.98	41.01	0.00
Bin 94_2	0.863348	205	4252	29.91	41.01	2.16
Bin 454_2	0.793239	85	13838	66.15	41.01	5.76
Bin 237_1	0.246192	29	11359	67.03	41.01	18.71
Bin 333	1.308351	282	4861	41.95	40.29	0.72
Bin 189_1	2.921293	346	10778	66.02	40.29	6.47
Bin 345_1	1.520648	409	3597	28.53	40.29	10.07
Bin 49	1.158838	265	4563	38.33	39.57	0.00
Bin 535_2	3.398883	799	4335	68.49	39.57	7.19
Bin 141	1.087828	254	4196	60.62	38.85	0.00
Bin 232_2	0.294306	38	8623	30.33	38.85	1.44
Bin 289	0.957255	227	4341	37.38	38.85	5.04
Bin 10	0.309427	51	6424	53.94	38.13	0.00
Bin 436_2_1	1.442341	208	9354	34.92	38.13	2.16
Bin 345_5	1.06668	158	8100	27.67	38.13	6.47
Bin 160_1	2.103783	438	4953	37.98	38.13	7.91
Bin 529	0.344268	73	5296	37.29	37.41	0.72
Bin 450	1.402358	288	5234	42.09	37.41	2.16
Bin 495_2	1.106086	206	5677	50.46	37.41	5.76
Bin 345_7	1.035427	201	5357	29.16	37.41	5.76
Bin 238	0.935811	222	4275	28.14	37.41	6.47
Bin 244_3	1.713006	49	55619	41.33	37.41	9.35
Bin 463	0.904236	191	5070	44.75	36.69	0.00
Bin 432	0.399328	34	17330	57.05	36.69	0.00
Bin 464_1	4.148689	350	15380	69.38	36.69	2.16

Bin 338_1	1.450802	264	5875	63.49	36.69	4.32
Bin 306_1_1	2.241718	177	17899	40.09	36.69	5.04
Bin 268_2	2.358883	366	7403	39.54	36.69	8.63
Bin 338_7	0.212555	55	3748	63.04	36.69	8.63
Bin 138_3	1.795755	331	5833	32.30	35.97	0.00
Bin 214	0.856568	111	8939	46.35	35.97	0.00
Bin 183	0.800637	112	8035	53.82	35.97	0.00
Bin 331	1.973955	380	5692	44.33	35.97	0.72
Bin 375	0.895839	215	4171	45.58	35.97	1.44
Bin 483_2	2.64926	486	6329	40.49	35.97	7.91
Bin 390	0.267549	69	4060	28.14	35.80	1.85
Bin 370	1.625371	378	4369	60.64	35.25	0.00
Bin 527	0.305572	36	9958	49.51	35.25	0.00
Bin 27	1.289252	284	4847	52.56	35.25	2.16
Bin 326_1	0.764726	149	5706	59.54	35.25	2.16
Bin 530	0.61537	161	3831	36.08	35.25	2.16
Bin 306_2	0.830075	154	6512	37.92	35.25	7.91
Bin 335	1.164291	205	6242	41.66	34.53	0.72
Bin 32	0.502789	135	3668	56.91	33.81	0.00
Bin 305	2.619543	470	6156	50.88	33.81	1.44
Bin 343	1.358573	277	5075	38.41	33.81	2.16
Bin 220	1.659387	425	3899	38.76	33.81	2.88
Bin 453_1	1.577589	191	10796	39.63	33.81	4.32
Bin 341_3	1.529197	117	16576	58.86	33.81	7.91
Bin 341_1	1.32548	86	21165	57.31	33.09	0.00
Bin 338_8_1	1.125345	132	10728	62.21	33.09	0.00
Bin 540	0.689554	182	3756	38.44	33.09	0.00
Bin 414	0.754326	157	4936	46.42	33.09	0.72
Bin 93	1.686223	305	6233	56.12	33.09	1.44
Bin 54_1	1.463233	126	19252	44.89	33.09	2.16
Bin 268_5	0.896879	166	5838	39.93	33.09	5.04
Bin 55_4	1.69764	232	8935	49.57	33.09	7.91
Bin 295	0.393116	86	4615	42.72	32.37	0.00
Bin 329	2.093323	296	8931	60.47	32.37	2.16
Bin 320_1	0.845149	183	4746	31.68	32.37	7.19

Bin 546	1.177937	225	5571	43.22	31.65	0.00
Bin 21	0.973405	219	4599	55.03	31.65	0.00
Bin 309	0.329361	88	3786	29.99	31.65	0.00
Bin 500	0.325751	11	45706	39.15	31.65	0.00
Bin 539	1.357301	265	5276	35.14	31.65	1.44
Bin 516_2	0.508396	112	4619	37.84	31.65	4.32
Bin 103	0.410162	71	6074	53.32	30.94	0.00
Bin 454_1	2.136338	249	10274	66.95	30.94	0.72
Bin 36	1.455488	276	5571	54.29	30.94	0.72
Bin 25	0.588196	138	4292	38.78	30.94	0.72
Bin 321	2.742851	457	6534	39.55	30.94	2.16
Bin 94_1	0.668049	151	4571	29.62	30.94	7.91
Bin 243_1	0.776067	213	3541	37.91	30.94	12.23
Bin 226	1.18807	281	4304	42.70	30.22	0.72
Bin 436_3	1.018908	168	7086	34.78	30.22	2.88
Bin 327_1	0.91353	174	5551	46.13	30.22	5.76
Bin 489	1.124631	223	5539	46.34	29.50	0.72
Bin 382	1.718075	434	3925	38.39	29.50	1.44
Bin 512_2	1.722926	273	7235	55.89	29.50	2.16
Bin 521	2.278774	508	4523	49.18	29.50	3.60
Bin 38_2	1.946472	408	5046	67.44	29.50	3.60
Bin 410_5	1.681584	200	10183	58.00	28.78	0.00
Bin 48	0.810036	150	5695	46.10	28.78	0.00
Bin 90	0.727662	115	7982	54.35	28.78	0.00
Bin 477	0.586263	157	3787	39.12	28.78	0.00
Bin 170	2.081329	423	5229	65.41	28.78	2.16
Bin 268_4	1.348387	214	7392	39.49	28.78	2.88
Bin 45	0.519365	124	4257	33.05	28.78	4.32
Bin 244_4a	2.206646	50	63723	41.12	28.78	5.04
Bin 373	0.635169	145	4238	31.50	28.06	0.00
Bin 459	0.587138	136	4420	41.51	28.06	0.00
Bin 186	0.527526	91	5997	42.67	28.06	0.00
Bin 126	0.426357	86	5331	31.06	28.06	0.00
Bin 402	0.350648	66	6459	29.75	28.06	0.00
Bin 482	2.156403	513	4245	63.48	28.06	2.16

Bin 219_5	3.514206	576	6764	66.43	28.06	2.88
Bin 415_1	0.802386	205	3959	63.23	28.06	4.32
Bin 244_6	1.704828	47	58367	42.83	28.06	6.47
Bin 259	1.266018	166	8606	61.92	27.34	0.00
Bin 30	1.99575	345	6439	56.40	27.34	1.44
Bin 243_5	1.560174	269	6644	37.88	27.34	2.88
Bin 46	1.413761	314	4578	38.34	27.34	4.32
Bin 231_5	0.444641	47	11612	57.84	27.34	5.76
Bin 466_2	3.599738	359	14815	64.34	27.34	7.19
Bin 465	0.54195	115	4946	51.16	26.62	0.00
Bin 417	0.392956	82	4956	36.18	26.62	0.00
Bin 554	1.045159	239	4534	56.45	26.62	0.72
Bin 355_1a	3.510892	669	5640	36.70	26.62	2.16
Bin 231_1	2.268837	245	14126	60.88	26.62	4.32
Bin 291_4	1.008113	249	4197	41.15	26.62	10.07
Bin 80	0.847114	150	5813	49.57	25.90	0.00
Bin 67_2	0.794438	131	7124	34.94	25.90	0.00
Bin 538	0.478305	124	3923	55.36	25.90	0.00
Bin 471	0.302812	57	5459	36.75	25.90	0.00
Bin 99	0.709238	94	9374	34.59	25.90	0.72
Bin 338_4	2.602922	403	7633	64.03	25.90	1.44
Bin 356	1.700859	386	4399	60.11	25.90	1.44
Bin 120	1.672877	389	4406	49.51	25.90	1.44
Bin 453_3	0.570039	104	6028	39.88	25.90	3.60
Bin 494_3	0.498827	119	4104	49.79	25.90	3.60
Bin 345_8	0.870161	88	14103	25.88	25.90	7.91
Bin 494_2	0.921439	141	7004	48.22	25.18	0.00
Bin 207_3	0.129828	19	7486	47.26	25.18	0.00
Bin 92_1	0.480092	110	4599	32.72	25.18	0.72
Bin 488	0.79992	213	3586	38.42	25.18	2.16
Bin 355_6	1.861236	432	4327	37.60	25.18	5.04
Bin 483_1_2	1.74613	280	7610	39.98	25.18	5.76
Bin 355_5a	3.104736	762	3996	38.06	25.18	10.07
Bin 497	1.250291	285	4608	44.68	24.46	0.00
Bin 304	0.745121	154	4757	38.12	24.46	0.00

Bin 313	0.542184	139	3776	39.33	24.46	0.00
Bin 173	0.452732	40	12813	54.07	24.46	0.00
Bin 180	1.510578	282	5702	56.49	24.46	0.72
Bin 535_1	0.793415	164	5090	68.60	24.46	0.72
Bin 351	1.322053	287	4666	51.21	24.46	1.44
Bin 306_6	1.193113	153	10298	37.66	24.46	1.44
Bin 55_2	0.787844	155	5317	48.92	24.46	1.44
Bin 166	1.010032	210	4838	52.17	24.46	2.88
Bin 154_2	2.450763	694	3426	59.40	24.46	3.60
Bin 338_6_1	4.923525	692	8513	63.41	24.46	5.04
Bin 535_5	1.716091	301	6254	68.98	24.46	9.35
Bin 517	0.573118	98	6534	48.59	24.07	1.23
Bin 439	0.52142	120	4457	35.64	23.74	0.00
Bin 345_3	0.199088	47	4476	29.03	23.74	0.00
Bin 139_5	1.173869	177	8130	48.21	23.74	0.72
Bin 176	0.822506	169	5408	42.04	23.74	0.72
Bin 287	0.762878	96	9420	50.64	23.74	0.72
Bin 38_1	0.775275	130	6780	66.73	23.74	5.76
Bin 547	0.261327	52	5656	30.12	23.02	0.00
Bin 63	2.152398	544	3943	66.65	23.02	0.72
Bin 221	0.852017	212	3992	57.45	22.30	0.00
Bin 139_1	0.503634	128	3936	49.27	22.30	0.00
Bin 256	1.107775	220	5347	54.63	22.30	0.72
Bin 355_4	1.772507	459	3734	37.22	22.30	1.44
Bin 479	1.014953	218	4780	35.26	22.30	1.44
Bin 94_4	0.072392	20	3469	30.55	22.30	5.04
Bin 243_2_1	1.79936	409	4510	39.56	22.30	6.47
Bin 69	1.48836	190	10116	26.70	21.60	3.09
Bin 448	1.265774	328	3714	67.51	21.58	0.00
Bin 528	0.639673	150	4258	40.87	21.58	0.00
Bin 427	0.598063	84	7667	62.14	21.58	0.00
Bin 534	0.917646	245	3712	52.20	21.58	0.72
Bin 243_3	1.750278	506	3354	38.36	21.58	7.19
Bin 296	0.514008	91	6048	43.04	20.99	1.23
Bin 368_1	2.239228	357	7126	58.58	20.86	0.00

Bin 506	1.317288	264	5329	54.41	20.86	0.00
Bin 115	0.706579	202	3426	31.82	20.86	0.00
Bin 255	0.564211	105	5558	33.64	20.86	0.00
Bin 108	1.297701	275	4946	38.48	20.86	2.88
Bin 122_2	3.236888	647	5134	57.04	20.86	6.47
Bin 532	0.673953	182	3646	71.32	20.14	0.00
Bin 189_4	0.108787	22	5772	65.47	20.14	0.00
Bin 551	0.714283	115	7084	35.54	20.14	0.72
Bin 70	0.292121	42	8589	27.20	20.14	0.72
Bin 243_6	1.278023	315	4113	36.42	20.14	1.44
Bin 436_2_2	0.583756	104	6416	34.72	20.14	1.44
Bin 483_1_1	1.925685	317	7185	39.55	20.14	2.88
Bin 494_1	0.778487	145	5784	51.06	20.14	3.60
Bin 117_1	2.592079	596	4410	44.63	20.14	4.32
Bin 247	0.82815	189	4466	56.42	19.42	0.00
Bin 209	0.659978	156	4261	35.65	19.42	0.00
Bin 451	0.269572	34	8358	45.84	19.42	0.00
Bin 245_1	0.912233	189	4998	38.09	19.42	1.44
Bin 327_2	0.710519	128	6038	45.85	19.42	1.44
Bin 444_2	0.297126	58	5298	51.07	19.42	1.44
Bin 44_1a	1.263013	274	4569	61.78	19.42	7.19
Bin 42	0.865354	171	5235	56.59	18.71	0.00
Bin 428	0.851258	201	4306	59.71	18.71	0.00
Bin 39	0.526294	117	4826	54.36	18.71	0.00
Bin 35	0.482543	127	3843	41.69	18.71	0.00
Bin 401	0.43548	127	3359	32.59	18.71	0.00
Bin 6	0.377369	53	8952	49.19	18.71	0.00
Bin 535_4	0.250665	62	4144	68.37	18.71	0.00
Bin 9	1.364892	274	5489	54.37	18.71	0.72
Bin 197	0.780669	143	6177	53.43	18.71	0.72
Bin 15	0.592112	138	4408	35.61	18.71	0.72
Bin 152	0.334789	94	3606	40.52	18.71	0.72
Bin 204	1.675556	391	4346	57.96	18.71	1.44
Bin 535_3	3.793178	1087	3331	68.86	18.71	2.16
Bin 187	1.826771	518	3465	40.35	18.71	2.16

Bin 355_3	1.571998	408	3783	37.02	18.71	2.16
Bin 231_2_1	1.203459	239	5210	59.23	18.71	5.04
Bin 359	1.521754	403	3796	37.89	17.99	0.00
Bin 493	0.799573	110	8424	54.02	17.99	0.00
Bin 324	0.312567	84	3701	45.95	17.99	0.00
Bin 415_2a	0.811836	164	5286	61.33	17.99	1.44
Bin 533	0.408367	95	4236	57.31	17.99	1.44
Bin 252_1	3.481543	656	5377	65.56	17.99	2.16
Bin 338_8_2	0.977293	135	9089	62.46	17.99	2.88
Bin 299	0.306028	48	6730	48.37	17.90	2.47
Bin 50	0.736379	57	20090	31.37	17.27	0.00
Bin 523	0.481321	97	5620	42.08	17.27	0.00
Bin 342	0.396219	101	3876	61.06	17.27	0.00
Bin 412	0.637601	156	4215	47.05	17.27	0.72
Bin 264	0.918084	259	3465	37.13	17.27	1.44
Bin 252_3	2.099096	385	5567	64.68	17.27	5.04
Bin 101	0.447055	83	5636	49.20	16.67	1.23
Bin 44_1b	1.787797	405	4452	61.22	16.55	0.00
Bin 306_5	0.693412	83	11098	39.46	16.55	0.00
Bin 303	0.663173	92	8581	49.22	16.55	0.00
Bin 262	0.638441	107	6602	66.75	16.55	0.00
Bin 55_3	0.573052	120	5032	49.48	16.55	0.00
Bin 212	0.37351	85	4698	52.12	16.55	0.00
Bin 111	0.36133	25	19720	31.41	16.55	0.00
Bin 462	0.334473	44	10457	52.02	16.55	0.00
Bin 60	0.318078	72	4746	61.15	16.55	0.00
Bin 77	0.312606	78	4130	52.16	16.55	0.00
Bin 168	0.816799	180	4628	52.76	16.55	1.44
Bin 318	0.671345	185	3640	35.41	16.55	1.44
Bin 355_5b	2.380086	664	3467	35.02	16.55	2.88
Bin 122_1	4.095473	1013	4098	56.29	16.55	4.32
Bin 327_5	1.117962	190	6667	44.82	16.55	4.32
Bin 345_2	0.283475	68	4066	28.07	16.55	5.04
Bin 146	0.505845	99	5287	47.83	16.05	1.85
Bin 5	0.428931	88	4698	48.02	16.05	4.32

Bin 384	0.202846	20	13105	32.08	15.83	0.00
Bin 436_1	0.174393	37	4662	35.27	15.83	2.16
Bin 292	0.619383	176	3475	37.29	15.11	0.00
Bin 251	0.385463	63	6519	46.59	15.11	0.00
Bin 306_3	0.192275	46	4670	38.92	15.11	0.00
Bin 355_1b	2.057098	460	4532	33.39	15.11	0.72
Bin 160_5	1.481529	248	6593	37.35	15.11	0.72
Bin 244_5	1.487714	58	37937	42.89	15.11	1.44
Bin 265	0.69107	177	3845	53.76	15.11	1.44
Bin 306_1_2	0.600013	51	16169	39.20	15.11	1.44
Bin 131	0.53275	126	4398	36.51	15.11	1.44
Bin 520_3	0.117565	29	4200	41.69	15.11	2.88
Bin 406_2	0.858137	176	5330	28.49	15.11	4.32
Bin 177	0.522657	120	4299	33.31	14.39	0.00
Bin 56	0.26237	64	4075	54.28	14.39	0.00
Bin 431	0.385783	67	7192	34.46	14.39	0.72
Bin 550	0.756802	178	4084	48.79	13.67	0.00
Bin 416	0.386598	77	5477	43.61	13.67	0.00
Bin 367	0.336946	72	4648	38.97	13.67	0.00
Bin 385	0.332604	95	3540	33.37	13.67	0.00
Bin 110	0.279386	85	3243	38.87	13.67	0.00
Bin 231_2_2	1.744429	338	5661	60.69	13.67	1.44
Bin 415_3	1.400561	254	5928	62.80	13.67	2.16
Bin 355_7a	1.96656	395	5319	33.52	13.67	4.32
Bin 410_3	1.252776	200	6922	61.56	12.95	0.00
Bin 476	0.923446	285	3125	63.27	12.95	0.00
Bin 119	0.651023	96	8245	45.14	12.95	0.00
Bin 114	0.496845	100	5281	59.61	12.95	0.00
Bin 420	0.416648	94	4356	60.43	12.95	0.00
Bin 182	0.400774	108	3741	35.75	12.95	0.00
Bin 145	0.383933	69	5403	46.94	12.95	0.00
Bin 380	0.315262	62	6289	50.62	12.95	0.00
Bin 328	0.237786	30	9911	48.84	12.95	0.00
Bin 149	0.63913	106	7003	65.27	12.95	0.72
Bin 277	0.351512	48	9104	46.80	12.95	0.72

Bin 320_3	0.268966	83	3070	33.02	12.95	0.72
Bin 65	0.254142	62	4056	39.33	12.95	0.72
Bin 116	1.000033	263	3718	58.78	12.23	0.00
Bin 79	0.597805	130	4486	40.56	12.23	0.00
Bin 514	0.553229	92	7164	55.37	12.23	0.00
Bin 341_4	0.291207	30	11425	57.36	12.23	0.00
Bin 492	0.2344	59	3965	46.98	12.23	0.00
Bin 95	0.201136	18	13805	54.41	12.23	0.00
Bin 199	0.935981	195	4750	32.19	12.23	0.72
Bin 175	0.634282	158	4049	39.63	12.23	0.72
Bin 332	1.558388	381	4056	60.72	12.23	1.44
Bin 474_3	4.563975	1137	3998	34.57	11.73	3.70
Bin 376	0.364649	71	5496	52.56	11.51	0.00
Bin 311	0.282243	75	3833	38.23	11.51	0.00
Bin 275	0.280004	52	5941	42.91	11.51	0.00
Bin 73	0.243632	38	8108	35.12	11.51	0.00
Bin 410_4	1.493432	152	12138	56.89	11.51	0.72
Bin 466_1	0.657866	183	3361	63.13	11.51	0.72
Bin 274	1.149926	207	6265	32.78	11.51	1.44
Bin 244_4b	1.089262	24	70889	41.77	11.51	2.88
Bin 355_7_2	2.706391	467	6765	35.00	10.79	0.00
Bin 219_3	1.742672	320	6193	66.52	10.79	0.00
Bin 125	0.662989	176	3694	52.31	10.79	0.00
Bin 109	0.644386	146	4535	60.98	10.79	0.00
Bin 286	0.449801	130	3416	61.52	10.79	0.00
Bin 1	0.292921	54	6076	55.55	10.79	0.00
Bin 92_3	0.290795	89	3134	32.59	10.79	0.00
Bin 200	0.246079	61	3937	44.68	10.79	0.00
Bin 355_2	3.315196	540	7197	36.70	10.79	0.72
Bin 29	0.920516	220	4157	43.67	10.79	0.72
Bin 524	0.297805	65	4255	49.68	10.79	0.72
Bin 338_3	2.235426	232	13251	58.84	10.79	1.44
Bin 531_2	0.090418	22	4036	43.84	10.79	4.32
Bin 138_2	1.864057	408	4478	33.26	10.07	0.00
Bin 508_3	1.128988	163	8268	63.95	10.07	0.00

Bin 235_1	1.002869	193	5196	43.17	10.07	0.00
Bin 231_7	0.969501	214	4802	63.00	10.07	0.00
Bin 19	0.48902	126	3891	41.39	10.07	0.00
Bin 397_1	0.282208	66	4520	60.09	10.07	0.00
Bin 263	0.224125	37	6263	52.24	10.07	0.00
Bin 443	0.209312	39	5015	40.70	10.07	0.00
Bin 44_2	1.655041	470	3420	61.27	10.07	0.72
Bin 341_5	0.798086	105	9061	59.23	10.07	0.72
Bin 252_4	2.900953	731	3949	64.45	10.07	1.44
Bin 225_2	0.914264	186	4829	60.34	10.07	1.44
Bin 312	0.667491	176	3869	39.77	10.07	2.16
Bin 244_1	0.751352	79	10208	42.24	9.35	0.00
Bin 138_5	0.445894	113	3690	32.17	9.35	0.00
Bin 362	0.439074	108	4242	51.24	9.35	0.00
Bin 358	0.408897	104	3869	61.34	9.35	0.00
Bin 364	0.31597	63	5603	66.70	9.35	0.00
Bin 97	0.261891	58	4809	61.86	9.35	0.00
Bin 326_2	0.081678	20	4079	59.97	9.35	0.00
Bin 64_3	0.034248	3	19274	28.94	9.35	0.00
Bin 468	0.642007	198	3114	38.31	9.35	1.44
Bin 61	0.548405	135	3782	38.86	9.35	1.44
Bin 34	0.586354	121	4918	36.05	9.26	3.09
Bin 156	0.471643	75	7034	49.71	8.64	3.70
Bin 377	0.568026	144	3809	61.33	8.63	0.00
Bin 388	0.272288	31	10104	40.19	8.63	0.00
Bin 123_3	0.019548	6	2946	38.69	8.63	0.00
Bin 130	0.987229	279	3462	58.09	8.63	0.72
Bin 507	0.402047	123	3150	69.62	8.63	0.72
Bin 88	0.321209	92	3354	34.45	8.63	0.72
Bin 243_4	0.55921	116	5018	37.59	8.63	2.88
Bin 139_4	0.747451	156	4759	48.61	7.91	0.00
Bin 218	0.511453	145	3468	63.61	7.91	0.00
Bin 350	0.473466	87	5825	45.52	7.91	0.00
Bin 447	0.43143	98	4609	58.32	7.91	0.00
Bin 498	0.409109	111	3611	61.61	7.91	0.00

Bin 542	0.390128	69	6420	61.12	7.91	0.00
Bin 98	0.380715	99	3772	40.66	7.91	0.00
Bin 393	0.344627	54	7118	55.63	7.91	0.00
Bin 515	0.329988	56	6867	64.01	7.91	0.00
Bin 487	0.294762	68	4290	59.37	7.91	0.00
Bin 501	0.203099	22	11955	55.12	7.91	0.00
Bin 55_1	0.154883	37	4083	49.85	7.91	0.00
Bin 430	0.347173	87	3752	37.68	7.91	1.44
Bin 435	0.328622	56	6783	48.80	7.41	0.00
Bin 458	0.228178	37	6988	43.52	7.41	0.62
Bin 31	0.217146	44	4736	40.43	7.41	0.62
Bin 394_7_3	2.585756	555	4807	42.82	7.41	2.47
Bin 234	0.396574	38	14402	41.19	7.19	0.00
Bin 486	0.367508	88	3970	47.33	7.19	0.00
Bin 185	0.266131	60	4794	51.22	7.19	0.00
Bin 227	0.260554	56	4768	47.27	7.19	0.00
Bin 37	0.225837	42	5685	42.82	7.19	0.00
Bin 84	0.216443	60	3490	57.41	7.19	0.00
Bin 243_2_2	0.213852	53	4357	37.89	7.19	0.00
Bin 347	0.206375	50	3967	57.23	7.19	0.00
Bin 425	0.202511	36	6483	50.09	7.19	0.00
Bin 124_2	0.175077	45	3886	36.27	7.19	0.00
Bin 520_2	0.058726	16	3379	41.52	7.19	0.00
Bin 87	0.239842	25	13173	42.23	6.79	0.00
Bin 394_5b	7.467061	1512	5204	29.50	6.79	2.47
Bin 235_4	1.882315	269	8443	42.35	6.47	0.00
Bin 464_4	1.360458	198	7880	71.36	6.47	0.00
Bin 188	0.527554	110	4932	35.09	6.47	0.00
Bin 437	0.40302	122	3115	68.73	6.47	0.00
Bin 12	0.385343	107	3592	33.95	6.47	0.00
Bin 531_1	0.275466	67	4018	43.74	6.47	0.00
Bin 424	0.268565	62	4354	45.40	6.47	0.00
Bin 325	0.266479	57	4555	36.40	6.47	0.00
Bin 254	0.225419	26	9758	51.57	6.47	0.00
Bin 118	0.209665	46	4837	39.55	6.47	0.00

Bin 411	0.204636	38	5708	44.42	6.47	0.00
Bin 3	0.542165	127	4322	37.78	6.47	0.72
Bin 464_3	0.006992	2	4013	66.82	6.47	0.72
Bin 310	1.075002	297	3580	71.38	6.17	1.23
Bin 252_2	1.871458	415	4424	65.09	5.76	0.00
Bin 189_2	0.933979	165	6464	65.98	5.76	0.00
Bin 215	0.427969	111	3937	39.17	5.76	0.00
Bin 496	0.386196	114	3218	66.67	5.76	0.00
Bin 8	0.324769	94	3457	63.82	5.76	0.00
Bin 92_4	0.264042	50	6903	33.14	5.76	0.00
Bin 279	0.256797	26	11855	39.06	5.76	0.00
Bin 268_1	0.213557	55	3911	39.64	5.76	0.00
Bin 235_3	1.463036	335	4286	42.81	5.76	0.72
Bin 555	0.201666	45	4340	40.93	5.56	0.00
Bin 394_4a	5.944026	1176	5355	34.50	5.56	1.23
Bin 394_4b	0.739808	207	3465	35.07	5.56	1.85
Bin 338_2	1.441389	350	4030	62.62	5.04	0.00
Bin 285	0.499251	148	3249	64.86	5.04	0.00
Bin 75	0.473584	73	7111	60.49	5.04	0.00
Bin 449	0.372478	105	3636	35.36	5.04	0.00
Bin 360	0.237374	69	3421	35.77	5.04	0.00
Bin 169	0.222696	35	6523	47.39	5.04	0.00
Bin 543	0.208216	53	3944	43.55	5.04	0.00
Bin 408_1	0.185351	60	2897	64.21	5.04	0.00
Bin 418	0.270729	55	5148	56.70	5.04	0.72
Bin 128	0.476316	92	4756	41.37	5.04	1.44
Bin 229	0.604408	104	6657	34.22	5.04	5.04
Bin 394_5a	3.77632	895	4159	33.24	4.94	1.85
Bin 394_2c	5.21728	1175	4508	33.76	4.32	0.00
Bin 396	0.230176	38	6907	43.78	4.32	0.00
Bin 394_7_1	3.299299	657	5213	42.66	4.32	0.00
Bin 410_2	1.144817	150	9891	53.69	4.32	0.00
Bin 338_6_2	0.990055	123	10071	60.16	4.32	0.00
Bin 291_1	0.786122	168	4997	38.22	4.32	0.00
Bin 365	0.260547	50	4988	37.18	4.32	0.00

Bin 76	0.20327	17	22038	52.30	4.32	0.00
Bin 210_2	0.002529	1	2529	61.25	4.32	0.00
Bin 421	0.318425	44	9509	61.97	4.32	1.44
Bin 290	0.200298	44	4539	38.86	3.70	0.00
Bin 66	0.293767	82	3553	37.71	3.70	0.62
Bin 394_6c	2.352774	477	5181	39.18	3.70	1.23
Bin 491	1.3278	166	10211	54.96	3.60	0.00
Bin 392	0.434301	66	8075	42.61	3.60	0.00
Bin 100	0.387857	97	3883	37.93	3.60	0.00
Bin 112	0.272214	71	3958	63.93	3.60	0.00
Bin 369	0.260082	35	8919	50.81	3.60	0.00
Bin 68	0.253393	71	3532	33.26	3.60	0.00
Bin 327_3	0.162771	44	3650	44.30	3.60	0.00
Bin 330_2	0.071024	3	33122	27.59	3.60	0.00
Bin 518_4	0.049198	14	3035	38.01	3.60	0.00
Bin 394_6b	2.09528	374	6096	36.21	3.09	0.00
Bin 278	0.314267	81	3697	40.44	3.09	0.00
Bin 71	0.220112	49	4599	48.58	3.09	0.00
Bin 307	0.452158	58	10448	35.50	2.88	0.00
Bin 231_6	0.445649	127	3330	59.68	2.88	0.00
Bin 409	0.293694	25	15627	46.72	2.88	0.00
Bin 129	0.292763	52	5937	47.21	2.88	0.00
Bin 508_6	0.115962	18	7705	60.49	2.88	0.00
Bin 494_5	0.041041	12	3451	47.84	2.88	0.00
Bin 320_2	0.030854	9	3538	32.27	2.88	3.60
Bin 28	0.625623	110	5943	33.09	2.47	0.00
Bin 306_7	0.237568	34	7322	37.99	2.47	0.00
Bin 334	0.231053	44	5171	48.23	2.47	0.00
Bin 513	0.226176	31	7637	46.78	2.47	0.00
Bin 242	0.210937	26	11427	36.16	2.47	0.00
Bin 210_3	0.157683	49	3054	63.07	2.47	0.00
Bin 394_6d	4.433278	961	4714	38.04	2.47	0.62
Bin 394_4c	2.702061	559	4968	33.86	2.47	0.62
Bin 394_7_2	2.190507	532	4162	39.60	2.16	0.00
Bin 327_4	0.36275	64	5567	45.06	2.16	0.00

Bin 453_2	0.067779	15	4585	39.84	2.16	0.00
Bin 317_3	0.059917	17	3752	38.34	2.16	0.00
Bin 81	0.23445	48	5086	61.42	2.16	0.72
Bin 473	0.458124	61	9174	58.19	2.16	1.44
Bin 394_1d	3.37812	766	4433	38.68	1.85	0.62
Bin 224	0.231035	43	5597	45.74	1.85	1.23
Bin 394_1a	5.202305	1184	4480	36.77	1.44	0.00
Bin 381	0.380165	71	5606	33.26	1.44	0.00
Bin 508_4	0.306031	78	3573	62.04	1.44	0.00
Bin 203	0.25495	44	6332	56.56	1.44	0.00
Bin 47	0.240907	41	6918	43.74	1.44	0.00
Bin 341_6	0.193795	52	3621	57.15	1.44	0.00
Bin 520_4	0.144285	35	3820	41.56	1.44	0.00
Bin 160_4	0.13977	33	4667	37.35	1.44	0.00
Bin 244_2	0.075958	20	3518	42.32	1.44	0.00
Bin 257	0.457542	105	4329	36.25	1.44	0.72
Bin 386	0.244948	23	14618	44.68	1.23	0.00
Bin 394_2a	3.706801	951	3799	30.86	1.23	0.62
Bin 394_2d	1.681698	468	3444	33.70	1.23	0.62
Bin 548	0.274899	41	8571	31.26	1.23	0.62
Bin 474_1	3.559292	814	4455	26.57	0.72	0.00
Bin 394_6a	2.718347	578	4914	37.08	0.72	0.00
Bin 394_1c	2.275293	600	3756	40.90	0.72	0.00
Bin 107	1.165633	110	14666	54.49	0.72	0.00
Bin 268_3	0.437107	120	3764	40.51	0.72	0.00
Bin 22	0.382947	65	6593	45.64	0.72	0.00
Bin 474_4	0.320676	80	4208	31.73	0.72	0.00
Bin 400	0.220504	27	10198	25.99	0.72	0.00
Bin 164	0.206671	58	3540	56.27	0.72	0.00
Bin 306_4	0.12781	30	4535	37.53	0.72	0.00
Bin 67_1	0.121248	27	5179	34.67	0.72	0.00
Bin 494_4	0.037226	7	5865	48.14	0.72	0.00
Bin 326_3	0.01634	3	5418	58.25	0.72	0.00
Bin 394_2e	3.219397	829	3788	34.65	0.72	0.72
Bin 394_2b	3.085194	716	4323	29.87	0.72	0.72

Bin 394_1g	1.63323	348	4950	35.45	0.62	0.00
Bin 349	0.306369	52	6963	57.27	0.62	0.00
Bin 394_1e	1.483219	351	4184	41.01	0.62	0.62
Bin 394_1b	1.382079	329	4215	36.29	0.00	0.00
Bin 394_2f	1.163814	325	3486	38.64	0.00	0.00
Bin 474_2	0.432723	96	4467	30.83	0.00	0.00
Bin 132	0.418477	62	8032	42.42	0.00	0.00
Bin 139_3	0.358712	85	4168	45.80	0.00	0.00
Bin 196	0.283303	45	7673	39.24	0.00	0.00
Bin 394_2h	0.275902	77	3347	32.39	0.00	0.00
Bin 394_1f	0.270395	67	3776	34.00	0.00	0.00
Bin 102	0.268748	42	6587	56.62	0.00	0.00
Bin 40	0.220662	31	8769	37.34	0.00	0.00
Bin 456	0.211624	38	7641	67.71	0.00	0.00
Bin 467	0.206618	37	6546	45.21	0.00	0.00
Bin 490	0.2023	18	19414	62.02	0.00	0.00
Bin 394_4d	0.121754	30	4205	31.79	0.00	0.00
Bin 394_2g	0.090906	26	3186	42.53	0.00	0.00
Bin 341_2	0.054557	9	7539	56.25	0.00	0.00
Bin 464_5	0.037187	12	2908	68.72	0.00	0.00
Bin 518_3	0.032617	10	3251	39.09	0.00	0.00
Bin 160_3	0.02997	7	4921	35.46	0.00	0.00
Bin 207_2	0.029094	9	3260	46.42	0.00	0.00
Bin 291_3	0.027917	8	3513	38.83	0.00	0.00
Bin 123_4	0.021154	3	10946	39.56	0.00	0.00
Bin 512_3	0.019879	6	3246	54.12	0.00	0.00
Bin 368_3	0.018076	6	3140	55.03	0.00	0.00
Bin 320_4	0.015919	5	2839	31.81	0.00	0.00
Bin 232_4	0.006979	2	3865	32.05	0.00	0.00
Bin 64_2	0.004673	1	4673	30.69	0.00	0.00
Bin 508_5	0.003819	1	3819	57.89	0.00	0.00
Bin 518_2	0.003146	1	3146	36.55	0.00	0.00
Bin 317_2	0.002995	1	2995	37.90	0.00	0.00
Bin 232_3	0.002587	1	2587	32.66	0.00	0.00
Bin 531_4	0.002517	1	2517	43.66	0.00	0.00

Supplementary table 3. Maximum Normalized ratio of 766 bins reconstructed from the Passalid gut and substrate metagenomes in the metagenomic samples.

Bin	Adult Normalized ratio	Larvae3 Normalized ratio	Larvae4 Normalized ratio	Substrate Normalized ratio
Bin 78	0.00	0.00	0.00	1.00
Bin 315	0.03	0.00	1.00	0.00
Bin 340	0.00	1.00	0.00	0.00
Bin 552	0.00	1.00	0.00	0.00
Bin 178	0.00	0.00	0.00	1.00
Bin 438	0.00	0.00	1.00	0.00
Bin 57	0.00	0.00	0.00	1.00
Bin 322	0.00	1.00	0.00	0.00
Bin 191	0.00	0.00	0.00	1.00
Bin 470	0.00	0.00	1.00	0.00
Bin 11	0.00	0.01	0.00	1.00
Bin 127	0.00	0.00	1.00	0.00
Bin 505	0.00	1.00	0.00	0.00
Bin 74	0.00	0.00	1.00	0.00
Bin 330_1	0.00	1.00	0.00	0.00
Bin 167	0.00	0.00	0.00	1.00
Bin 485	0.00	0.02	1.00	0.00
Bin 124_1	0.00	1.00	0.08	0.00
Bin 41	0.00	1.00	0.27	0.00
Bin 445	0.00	0.00	0.00	1.00
Bin 397_2	0.00	0.00	0.00	1.00
Bin 405	0.00	0.00	1.00	0.00
Bin 317_1	0.00	1.00	0.00	0.00
Bin 184	0.00	0.00	1.00	0.00
Bin 195	0.00	1.00	0.01	0.00

Bin 270	0.00	0.00	1.00	0.00
Bin 553	0.27	0.00	1.00	0.00
Bin 163	0.00	0.00	1.00	0.00
Bin 123_1	0.00	1.00	0.00	0.00
Bin 316	0.00	1.00	0.00	0.00
Bin 526	0.00	1.00	0.07	0.00
Bin 407	0.00	0.00	0.00	1.00
Bin 503	0.00	1.00	0.03	0.00
Bin 284	0.08	0.06	1.00	0.00
Bin 461	0.00	0.01	0.00	1.00
Bin 361	0.00	1.00	0.00	0.00
Bin 174	0.00	1.00	0.05	0.00
Bin 260	1.00	0.00	0.00	0.00
Bin 379	0.00	0.00	0.01	1.00
Bin 266	0.00	0.00	0.00	1.00
Bin 4	0.02	0.00	1.00	0.00
Bin 211	0.01	0.00	0.03	1.00
Bin 511	0.00	1.00	0.15	0.00
Bin 363	0.00	0.00	0.00	1.00
Bin 398	0.00	1.00	0.00	0.00
Bin 433	0.00	0.00	1.00	0.00
Bin 106	0.00	0.03	1.00	0.00
Bin 64_1	0.00	1.00	0.00	0.00
Bin 399	0.03	0.00	1.00	0.00
Bin 230	0.00	1.00	0.00	0.00
Bin 440	0.00	1.00	0.01	0.00
Bin 442	0.00	1.00	0.00	0.00
Bin 518_1	0.00	1.00	0.00	0.00
Bin 72	0.00	0.00	1.00	0.00

Bin 89	0.00	0.00	1.00	0.00
Bin 113	0.00	1.00	0.00	0.00
Bin 58	0.00	0.03	1.00	0.00
Bin 225_1	0.00	0.00	0.00	1.00
Bin 201	0.12	0.01	0.23	1.00
Bin 14	0.00	1.00	0.10	0.00
Bin 161	0.00	1.00	0.05	0.00
Bin 85	0.00	0.00	1.00	0.00
Bin 504	0.00	1.00	0.00	0.00
Bin 446	0.00	1.00	0.00	0.00
Bin 481	0.00	1.00	0.00	0.00
Bin 250	0.00	0.00	0.00	1.00
Bin 16	0.00	0.00	1.00	0.00
Bin 155	1.00	0.00	0.06	0.04
Bin 272	0.00	1.00	0.04	0.00
Bin 62	0.00	0.00	1.00	0.00
Bin 208	0.00	1.00	0.10	0.00
Bin 232_1	0.00	0.00	1.00	0.00
Bin 86	0.00	0.00	1.00	0.00
Bin 172	0.00	1.00	0.00	0.00
Bin 549	0.00	0.00	1.00	0.00
Bin 143	1.00	0.00	0.00	0.00
Bin 96	0.00	1.00	0.01	0.00
Bin 372	0.03	0.00	1.00	0.00
Bin 502	0.00	0.00	0.00	1.00
Bin 352	0.00	0.00	1.00	0.00
Bin 536	1.00	0.00	0.06	0.00
Bin 288	0.00	0.00	1.00	0.00
Bin 223	0.00	1.00	0.01	0.00

Bin 133	0.00	0.00	1.00	0.00
Bin 206	0.00	0.00	1.00	0.00
Bin 193	0.00	1.00	0.00	0.00
Bin 217	0.00	0.00	1.00	0.00
Bin 522	0.00	1.00	0.02	0.00
Bin 202	0.20	0.00	1.00	0.00
Bin 23	0.00	0.00	1.00	0.00
Bin 519	0.00	1.00	0.02	0.00
Bin 406_1	0.00	0.00	1.00	0.00
Bin 301	0.00	1.00	0.19	0.00
Bin 248	0.00	1.00	0.00	0.00
Bin 508_1	0.00	0.00	0.00	1.00
Bin 17	0.00	0.11	1.00	0.00
Bin 308	0.00	0.03	1.00	0.00
Bin 139_2	0.00	0.00	1.00	0.00
Bin 83	0.00	1.00	0.01	0.00
Bin 357	0.00	1.00	0.61	0.00
Bin 134	0.01	1.00	0.00	0.00
Bin 24	0.00	1.00	0.00	0.00
Bin 441	0.00	0.00	1.00	0.00
Bin 207_1	0.00	1.00	0.00	0.00
Bin 82	0.00	0.00	1.00	0.00
Bin 452	0.01	1.00	0.01	0.00
Bin 43	0.00	1.00	0.01	0.00
Bin 157	0.00	1.00	0.47	0.00
Bin 293	0.00	1.00	0.02	0.00
Bin 20	0.00	0.05	1.00	0.00
Bin 213	0.00	0.00	1.00	0.00
Bin 545	0.00	1.00	0.00	0.00

Bin 205	0.00	0.00	1.00	0.00
Bin 371	0.00	1.00	0.00	0.00
Bin 391	0.00	1.00	0.04	0.00
Bin 478	1.00	0.00	0.00	0.00
Bin 135	0.00	1.00	0.02	0.00
Bin 140	0.00	1.00	0.01	0.00
Bin 510	0.00	0.11	1.00	0.00
Bin 413	0.00	0.00	0.00	1.00
Bin 344	0.00	0.00	0.00	1.00
Bin 67_3	0.00	1.00	0.00	0.00
Bin 294	0.00	1.00	0.00	0.00
Bin 457	0.00	1.00	0.03	0.00
Bin 138_1	0.00	1.00	0.00	0.00
Bin 121	0.00	1.00	0.00	0.00
Bin 253	0.00	0.01	1.00	0.00
Bin 235_2	0.00	1.00	0.00	0.00
Bin 480	0.00	1.00	0.00	0.00
Bin 153	0.00	1.00	0.00	0.00
Bin 245_2	1.00	0.00	0.00	0.00
Bin 460	0.00	1.00	0.00	0.00
Bin 240	0.00	1.00	0.10	0.00
Bin 181	0.00	0.00	1.00	0.00
Bin 137	0.00	1.00	0.10	0.00
Bin 383	0.00	0.00	1.00	0.00
Bin 472	0.00	1.00	0.01	0.00
Bin 171	0.00	0.00	1.00	0.00
Bin 378	0.00	1.00	0.01	0.00
Bin 144	0.00	0.00	1.00	0.00
Bin 198	0.00	1.00	0.01	0.00

Bin 499	0.00	0.00	1.00	0.00
Bin 346	0.00	0.01	0.00	1.00
Bin 91	0.00	0.00	0.00	1.00
Bin 426	0.00	0.00	0.00	1.00
Bin 258	0.04	1.00	0.29	0.00
Bin 408_2	0.00	0.00	0.00	1.00
Bin 281	0.03	0.00	0.08	1.00
Bin 241	0.00	0.00	0.02	1.00
Bin 366	0.00	0.00	0.00	1.00
Bin 151	0.00	0.00	1.00	0.00
Bin 469	0.00	1.00	0.03	0.00
Bin 466_3	0.00	0.00	0.00	1.00
Bin 541	0.00	0.00	1.00	0.00
Bin 544	0.00	1.00	0.10	0.00
Bin 537	0.00	1.00	0.09	0.00
Bin 404	0.00	0.00	1.00	0.00
Bin 2	0.00	1.00	0.33	0.00
Bin 495_1	0.00	1.00	0.00	0.00
Bin 233	0.00	1.00	0.00	0.00
Bin 475	1.00	0.00	0.00	0.00
Bin 165	0.00	0.00	0.00	1.00
Bin 26	0.00	1.00	0.02	0.00
Bin 434	0.00	1.00	0.00	0.00
Bin 314	0.00	0.00	1.00	0.00
Bin 150	0.00	1.00	0.00	0.00
Bin 348	0.00	1.00	0.00	0.00
Bin 52	1.00	0.00	0.02	0.01
Bin 162	0.00	1.00	0.07	0.00
Bin 105	0.00	0.00	1.00	0.00

Bin 276	0.00	1.00	0.28	0.00
Bin 419	0.00	1.00	0.00	0.00
Bin 403	0.00	1.00	0.26	0.00
Bin 354	0.00	0.00	1.00	0.00
Bin 136	0.00	1.00	0.00	0.00
Bin 142	0.05	0.00	1.00	0.00
Bin 525	0.00	1.00	0.13	0.00
Bin 210_1	0.00	0.00	0.00	1.00
Bin 219_1	0.00	0.00	0.00	1.00
Bin 194	0.00	1.00	0.00	0.00
Bin 104	0.00	1.00	0.00	0.00
Bin 368_2	0.00	0.00	0.00	1.00
Bin 410_1	0.00	0.00	0.00	1.00
Bin 395	0.00	1.00	0.00	0.00
Bin 389	1.00	0.00	0.00	0.00
Bin 33	0.00	0.00	0.00	1.00
Bin 59	0.00	0.00	1.00	0.00
Bin 280	0.00	0.00	0.00	1.00
Bin 336	0.00	1.00	0.18	0.00
Bin 158	0.14	0.04	1.00	0.00
Bin 423	0.00	0.00	0.02	1.00
Bin 239	0.00	1.00	0.20	0.00
Bin 159	0.00	1.00	0.01	0.00
Bin 249	1.00	0.00	0.00	0.00
Bin 387	0.00	1.00	0.10	0.00
Bin 261	0.00	0.00	0.00	1.00
Bin 246	0.00	0.00	1.00	0.00
Bin 339	0.00	1.00	0.00	0.00
Bin 228	0.00	0.00	1.00	0.00

Bin 302	0.00	0.00	0.00	1.00
Bin 283	0.00	1.00	0.02	0.00
Bin 7	0.00	1.00	0.00	0.00
Bin 319	0.00	1.00	0.02	0.00
Bin 219_4	0.00	0.00	0.00	1.00
Bin 509	0.00	0.04	1.00	0.00
Bin 338_5	0.00	0.01	0.00	1.00
Bin 273	0.00	1.00	0.03	0.00
Bin 192	0.00	1.00	0.00	0.00
Bin 422	0.00	0.00	0.00	1.00
Bin 291_2	0.00	1.00	0.00	0.00
Bin 231_4	0.00	0.00	0.00	1.00
Bin 444_1	0.00	1.00	0.05	0.00
Bin 189_3	0.00	0.00	0.02	1.00
Bin 298_2	0.00	1.00	0.00	0.00
Bin 512_1	0.00	1.00	0.00	0.00
Bin 92_2	0.00	0.00	1.00	0.00
Bin 429	0.00	1.00	0.04	0.00
Bin 353	0.00	0.01	1.00	0.00
Bin 520_1	0.00	1.00	0.13	0.00
Bin 298_1	0.00	1.00	0.03	0.00
Bin 216	1.00	0.00	0.00	0.00
Bin 190	0.00	1.00	0.07	0.00
Bin 282	0.00	0.00	0.00	1.00
Bin 154_1	0.00	0.00	0.00	1.00
Bin 18	0.00	1.00	0.00	0.00
Bin 455	0.00	0.00	1.00	0.00
Bin 337	0.01	0.09	0.01	1.00
Bin 267	0.00	1.00	0.01	0.00

Bin 179	0.00	1.00	0.44	0.00
Bin 147	0.03	0.00	1.00	0.00
Bin 323	0.00	0.01	1.00	0.00
Bin 297	0.00	1.00	0.00	0.00
Bin 117_2	0.00	0.00	0.00	1.00
Bin 374	0.03	0.00	1.00	0.00
Bin 222	0.00	0.00	0.00	1.00
Bin 484	0.00	1.00	0.01	0.00
Bin 51	0.00	0.00	0.00	1.00
Bin 345_4	0.00	1.00	0.00	0.00
Bin 53	0.00	1.00	0.00	0.00
Bin 271	0.15	0.00	0.17	1.00
Bin 516_1	0.00	1.00	0.02	0.00
Bin 237_2	0.00	0.00	0.00	1.00
Bin 13	0.00	0.00	0.00	1.00
Bin 269	0.00	1.00	0.48	0.00
Bin 231_3	0.00	0.00	0.00	1.00
Bin 148	0.01	0.00	1.00	0.00
Bin 464_2	0.00	0.00	0.00	1.00
Bin 236	0.00	0.00	0.00	1.00
Bin 531_3	0.00	0.00	1.00	0.00
Bin 54_2	0.00	0.00	1.00	0.00
Bin 345_6	0.00	1.00	0.02	0.00
Bin 300	0.00	1.00	0.79	0.00
Bin 94_2	0.00	1.00	0.00	0.00
Bin 454_2	0.00	0.00	0.00	1.00
Bin 237_1	0.00	0.00	0.00	1.00
Bin 333	0.00	0.08	1.00	0.00
Bin 189_1	0.00	0.00	0.00	1.00

Bin 345_1	0.00	1.00	0.00	0.00
Bin 49	0.00	1.00	0.03	0.00
Bin 535_2	0.00	0.00	0.00	1.00
Bin 141	0.00	1.00	0.00	0.00
Bin 232_2	0.00	0.00	1.00	0.00
Bin 289	0.01	1.00	0.29	0.00
Bin 10	0.00	0.30	1.00	0.00
Bin 436_2_1	0.00	0.00	1.00	0.00
Bin 345_5	0.00	1.00	0.00	0.00
Bin 160_1	0.00	0.06	1.00	0.00
Bin 529	0.00	0.00	1.00	0.00
Bin 450	0.00	0.00	1.00	0.00
Bin 495_2	0.00	1.00	0.00	0.00
Bin 345_7	0.00	1.00	0.00	0.00
Bin 238	0.00	0.00	1.00	0.00
Bin 244_3	0.00	0.00	0.00	1.00
Bin 463	0.00	1.00	0.00	0.00
Bin 432	0.00	1.00	0.87	0.00
Bin 464_1	0.00	0.00	0.00	1.00
Bin 338_1	0.01	0.00	0.02	1.00
Bin 306_1_1	0.00	0.00	1.00	0.00
Bin 268_2	0.00	0.00	1.00	0.00
Bin 338_7	0.02	0.00	0.04	1.00
Bin 138_3	0.00	1.00	0.01	0.00
Bin 214	0.00	0.77	1.00	0.00
Bin 183	0.00	0.27	1.00	0.00
Bin 331	0.00	1.00	0.06	0.00
Bin 375	0.00	1.00	0.00	0.00
Bin 483_2	0.00	1.00	0.16	0.00

Bin 390	0.00	1.00	0.30	0.00
Bin 370	0.00	0.04	0.00	1.00
Bin 527	0.00	0.50	1.00	0.00
Bin 27	0.00	0.00	1.00	0.00
Bin 326_1	0.00	0.00	1.00	0.00
Bin 530	0.01	1.00	0.45	0.00
Bin 306_2	0.01	0.02	1.00	0.00
Bin 335	0.00	0.04	1.00	0.00
Bin 32	0.00	1.00	0.00	0.00
Bin 305	0.00	0.00	1.00	0.00
Bin 343	0.01	1.00	0.18	0.00
Bin 220	0.00	1.00	0.00	0.00
Bin 453_1	0.00	0.00	1.00	0.00
Bin 341_3	0.00	0.00	0.00	1.00
Bin 341_1	0.00	0.00	0.00	1.00
Bin 338_8_1	0.00	0.00	0.00	1.00
Bin 540	0.00	1.00	0.00	0.00
Bin 414	0.00	1.00	0.04	0.00
Bin 93	0.00	0.01	1.00	0.00
Bin 54_1	0.00	0.00	1.00	0.00
Bin 268_5	0.01	0.01	1.00	0.00
Bin 55_4	0.00	1.00	0.00	0.00
Bin 295	0.00	0.00	1.00	0.00
Bin 329	0.03	0.24	0.02	1.00
Bin 320_1	0.00	0.00	1.00	0.00
Bin 546	0.00	0.99	0.25	0.00
Bin 21	0.00	1.00	0.07	0.00
Bin 309	1.00	0.00	0.00	0.00
Bin 500	0.00	0.03	1.00	0.00

Bin 539	0.03	0.00	1.00	0.00
Bin 516_2	0.01	1.00	0.08	0.00
Bin 103	0.00	0.09	1.00	0.00
Bin 454_1	0.00	0.00	0.00	1.00
Bin 36	0.00	1.00	0.18	0.00
Bin 25	1.00	0.00	0.00	0.00
Bin 321	0.01	0.07	1.00	0.00
Bin 94_1	0.00	1.00	0.00	0.00
Bin 243_1	0.00	0.03	1.00	0.00
Bin 226	0.01	0.00	1.00	0.00
Bin 436_3	0.00	0.00	1.00	0.00
Bin 327_1	0.00	0.04	1.00	0.00
Bin 489	0.03	0.06	1.00	0.00
Bin 382	0.00	1.00	0.14	0.00
Bin 512_2	0.00	1.00	0.00	0.00
Bin 521	0.00	0.02	1.00	0.00
Bin 38_2	0.00	0.00	0.00	1.00
Bin 410_5	0.00	0.00	0.00	1.00
Bin 48	0.00	0.00	1.00	0.00
Bin 90	0.00	0.33	1.00	0.00
Bin 477	1.00	0.00	0.00	0.00
Bin 170	0.00	0.00	0.00	1.00
Bin 268_4	0.00	0.00	1.00	0.00
Bin 45	0.00	1.00	0.00	0.00
Bin 244_4a	0.00	0.00	0.00	1.00
Bin 373	0.00	1.00	0.29	0.00
Bin 459	0.01	1.00	0.57	0.00
Bin 186	0.00	0.12	1.00	0.01
Bin 126	0.00	1.00	0.35	0.00

Bin 402	0.00	0.04	1.00	0.00
Bin 482	0.00	0.00	0.00	1.00
Bin 219_5	0.00	0.00	0.00	1.00
Bin 415_1	0.02	0.03	1.00	0.00
Bin 244_6	0.00	0.00	0.00	1.00
Bin 259	0.00	0.31	0.00	1.00
Bin 30	0.00	1.00	0.00	0.00
Bin 243_5	0.00	0.00	1.00	0.00
Bin 46	0.00	1.00	0.47	0.00
Bin 231_5	0.00	0.00	0.00	1.00
Bin 466_2	0.00	0.00	0.00	1.00
Bin 465	0.00	1.00	0.50	0.00
Bin 417	0.00	1.00	0.21	0.00
Bin 554	0.00	1.00	0.17	0.00
Bin 355_1a	0.00	1.00	0.00	0.00
Bin 231_1	0.00	0.00	0.00	1.00
Bin 291_4	0.00	1.00	0.00	0.00
Bin 80	0.05	0.00	1.00	0.00
Bin 67_2	0.00	1.00	0.00	0.00
Bin 538	0.04	0.00	1.00	0.00
Bin 471	0.00	0.00	1.00	0.00
Bin 99	0.00	1.00	0.74	0.00
Bin 338_4	0.00	0.00	0.00	1.00
Bin 356	0.00	0.00	0.00	1.00
Bin 120	0.01	1.00	0.31	0.00
Bin 453_3	0.00	0.03	1.00	0.00
Bin 494_3	0.00	1.00	0.00	0.00
Bin 345_8	0.00	1.00	0.00	0.00
Bin 494_2	0.00	1.00	0.00	0.00

Bin 207_3	0.00	1.00	0.04	0.00
Bin 92_1	0.00	0.04	1.00	0.00
Bin 488	0.01	0.03	1.00	0.00
Bin 355_6	0.02	1.00	0.01	0.00
Bin 483_1_2	0.00	1.00	0.07	0.00
Bin 355_5a	0.00	1.00	0.00	0.00
Bin 497	0.00	1.00	0.08	0.00
Bin 304	0.00	1.00	0.16	0.00
Bin 313	1.00	0.00	0.00	0.00
Bin 173	0.00	0.23	1.00	0.00
Bin 180	0.00	0.00	1.00	0.00
Bin 535_1	0.00	0.00	0.00	1.00
Bin 351	0.00	0.01	1.00	0.00
Bin 306_6	0.00	0.00	1.00	0.00
Bin 55_2	0.01	1.00	0.04	0.00
Bin 166	0.00	1.00	0.12	0.00
Bin 154_2	0.00	0.00	0.00	1.00
Bin 338_6_1	0.00	0.00	0.00	1.00
Bin 535_5	0.00	0.00	0.00	1.00
Bin 517	0.00	1.00	0.00	0.00
Bin 439	1.00	0.00	0.00	0.03
Bin 345_3	0.00	1.00	0.00	0.00
Bin 139_5	0.00	0.00	1.00	0.00
Bin 176	0.00	1.00	0.27	0.00
Bin 287	0.00	0.93	0.99	0.00
Bin 38_1	0.00	0.00	0.00	1.00
Bin 547	0.00	0.03	1.00	0.00
Bin 63	0.00	0.00	0.00	1.00
Bin 221	0.00	1.00	0.02	0.00

Bin 139_1	0.00	0.06	1.00	0.00
Bin 256	0.00	0.01	1.00	0.00
Bin 355_4	0.01	1.00	0.17	0.00
Bin 479	0.00	1.00	0.00	0.00
Bin 94_4	0.00	1.00	0.02	0.00
Bin 243_2_1	0.00	0.00	1.00	0.00
Bin 69	0.00	0.00	0.00	1.00
Bin 448	0.00	0.01	1.00	0.00
Bin 528	0.00	0.01	1.00	0.00
Bin 427	0.00	0.27	0.00	1.00
Bin 534	0.00	0.00	0.00	1.00
Bin 243_3	0.00	0.00	1.00	0.00
Bin 296	0.00	0.14	1.00	0.00
Bin 368_1	0.00	0.00	0.00	1.00
Bin 506	0.00	0.01	1.00	0.00
Bin 115	0.00	0.00	1.00	0.00
Bin 255	0.00	1.00	0.36	0.00
Bin 108	0.01	0.08	1.00	0.00
Bin 122_2	0.00	0.00	0.00	1.00
Bin 532	0.00	0.00	0.08	1.00
Bin 189_4	0.00	0.00	0.00	1.00
Bin 551	0.00	1.00	0.00	0.00
Bin 70	0.00	1.00	0.13	0.00
Bin 243_6	0.00	0.00	1.00	0.00
Bin 436_2_2	0.00	0.01	1.00	0.00
Bin 483_1_1	0.00	1.00	0.01	0.00
Bin 494_1	0.00	1.00	0.00	0.00
Bin 117_1	0.00	0.00	0.00	1.00
Bin 247	0.00	1.00	0.00	0.00

Bin 209	0.00	0.04	1.00	0.00
Bin 451	0.00	1.00	0.00	0.00
Bin 245_1	1.00	0.03	0.00	0.01
Bin 327_2	0.00	0.01	1.00	0.00
Bin 444_2	0.00	1.00	0.18	0.00
Bin 44_1a	0.00	0.00	0.00	1.00
Bin 42	0.00	0.03	1.00	0.00
Bin 428	0.02	0.00	0.10	1.00
Bin 39	0.00	0.00	1.00	0.00
Bin 35	0.05	1.00	0.01	0.00
Bin 401	0.00	1.00	0.01	0.00
Bin 6	0.17	0.00	1.00	0.00
Bin 535_4	0.00	0.00	0.00	1.00
Bin 9	0.00	1.00	0.14	0.00
Bin 197	0.00	1.00	0.02	0.00
Bin 15	1.00	0.00	0.00	0.01
Bin 152	0.00	1.00	0.00	0.00
Bin 204	0.00	0.02	0.00	1.00
Bin 535_3	0.00	0.00	0.00	1.00
Bin 187	0.00	0.00	0.00	1.00
Bin 355_3	0.01	1.00	0.06	0.00
Bin 231_2_1	0.00	0.00	0.00	1.00
Bin 359	0.00	0.00	0.00	1.00
Bin 493	0.00	0.36	1.00	0.00
Bin 324	0.00	1.00	0.00	0.00
Bin 415_2a	0.00	0.01	1.00	0.00
Bin 533	0.00	0.03	1.00	0.00
Bin 252_1	0.00	0.00	0.00	1.00
Bin 338_8_2	0.00	0.00	0.00	1.00

Bin 299	0.00	0.64	0.89	0.00
Bin 50	0.17	0.63	1.00	0.02
Bin 523	0.00	0.03	1.00	0.00
Bin 342	0.00	1.00	0.00	0.00
Bin 412	0.00	1.00	0.01	0.00
Bin 264	0.00	1.00	0.04	0.00
Bin 252_3	0.00	0.00	0.00	1.00
Bin 101	0.05	0.01	1.00	0.00
Bin 44_1b	0.00	0.00	0.00	1.00
Bin 306_5	0.03	0.00	1.00	0.00
Bin 303	0.00	0.99	0.91	0.00
Bin 262	0.00	0.24	1.00	0.00
Bin 55_3	0.17	1.00	0.04	0.00
Bin 212	0.00	1.00	0.14	0.00
Bin 111	0.18	0.68	1.00	0.01
Bin 462	0.00	0.39	1.00	0.00
Bin 60	0.00	1.00	0.38	0.00
Bin 77	0.00	1.00	0.00	0.00
Bin 168	0.23	1.00	0.15	0.00
Bin 318	0.00	1.00	0.17	0.00
Bin 355_5b	0.00	1.00	0.00	0.00
Bin 122_1	0.00	0.00	0.00	1.00
Bin 327_5	0.00	0.00	1.00	0.00
Bin 345_2	0.00	1.00	0.10	0.00
Bin 146	0.00	0.90	0.57	0.00
Bin 5	0.00	0.85	0.64	0.00
Bin 384	0.19	0.58	1.00	0.01
Bin 436_1	0.00	0.09	1.00	0.00
Bin 292	0.00	0.01	1.00	0.00

Bin 251	0.00	0.58	1.00	0.00
Bin 306_3	0.02	0.10	1.00	0.00
Bin 355_1b	0.00	1.00	0.00	0.00
Bin 160_5	0.00	0.02	1.00	0.00
Bin 244_5	0.00	0.00	0.00	1.00
Bin 265	0.00	1.00	0.00	0.00
Bin 306_1_2	0.00	0.00	1.00	0.00
Bin 131	0.00	0.08	1.00	0.00
Bin 520_3	0.00	1.00	0.31	0.00
Bin 406_2	0.00	0.01	1.00	0.00
Bin 177	0.01	0.01	1.00	0.00
Bin 56	0.00	1.00	0.12	0.00
Bin 431	0.00	1.00	0.65	0.00
Bin 550	0.00	1.00	0.23	0.00
Bin 416	0.00	0.08	1.00	0.02
Bin 367	0.01	1.00	0.45	0.00
Bin 385	0.00	1.00	0.00	0.00
Bin 110	0.00	0.00	0.00	1.00
Bin 231_2_2	0.00	0.00	0.00	1.00
Bin 415_3	0.02	0.00	1.00	0.00
Bin 355_7a	0.00	1.00	0.00	0.00
Bin 410_3	0.00	0.00	0.00	1.00
Bin 476	0.00	0.00	0.00	1.00
Bin 119	0.00	0.19	0.98	0.00
Bin 114	0.00	0.98	0.92	0.00
Bin 420	0.00	1.00	0.07	0.00
Bin 182	0.00	1.00	0.04	0.00
Bin 145	0.00	1.00	0.10	0.00
Bin 380	0.00	1.00	0.92	0.00

Bin 328	0.00	0.46	1.00	0.00
Bin 149	0.00	0.27	1.00	0.00
Bin 277	0.00	0.84	1.00	0.00
Bin 320_3	0.00	0.00	1.00	0.00
Bin 65	0.00	0.68	1.00	0.00
Bin 116	0.00	0.00	0.00	1.00
Bin 79	0.00	1.00	0.00	0.00
Bin 514	0.00	0.22	1.00	0.00
Bin 341_4	0.00	0.00	0.00	1.00
Bin 492	0.00	0.00	1.00	0.00
Bin 95	0.00	0.24	1.00	0.00
Bin 199	0.17	0.05	1.00	0.00
Bin 175	0.00	1.00	0.01	0.00
Bin 332	0.01	0.00	0.08	1.00
Bin 474_3	1.00	0.00	0.00	0.00
Bin 376	0.00	1.00	0.50	0.00
Bin 311	0.00	1.00	0.01	0.00
Bin 275	0.00	0.14	1.00	0.00
Bin 73	0.00	0.99	0.84	0.00
Bin 410_4	0.00	0.00	0.00	1.00
Bin 466_1	0.00	0.00	0.00	1.00
Bin 274	0.00	0.01	1.00	0.00
Bin 244_4b	0.00	0.00	0.00	1.00
Bin 355_7_2	0.00	1.00	0.00	0.00
Bin 219_3	0.00	0.00	0.00	1.00
Bin 125	0.00	0.01	1.00	0.00
Bin 109	0.00	0.00	0.00	1.00
Bin 286	0.03	0.00	0.09	1.00
Bin 1	0.00	1.00	0.87	0.00

Bin 92_3	0.00	0.00	1.00	0.00
Bin 200	0.00	1.00	0.00	0.00
Bin 355_2	0.01	1.00	0.01	0.00
Bin 29	0.00	1.00	0.16	0.00
Bin 524	0.00	1.00	0.06	0.00
Bin 338_3	0.00	0.00	0.00	1.00
Bin 531_2	0.01	0.04	1.00	0.00
Bin 138_2	0.00	1.00	0.05	0.00
Bin 508_3	0.00	0.00	0.00	1.00
Bin 235_1	0.00	1.00	0.13	0.00
Bin 231_7	0.00	0.00	0.00	1.00
Bin 19	0.00	1.00	0.07	0.00
Bin 397_1	0.00	0.00	0.00	1.00
Bin 263	0.00	1.00	0.83	0.00
Bin 443	0.00	1.00	0.36	0.00
Bin 44_2	0.00	0.00	0.00	1.00
Bin 341_5	0.00	0.00	0.00	1.00
Bin 252_4	0.00	0.00	0.00	1.00
Bin 225_2	0.00	0.00	0.00	1.00
Bin 312	0.01	0.01	1.00	0.00
Bin 244_1	0.00	0.00	0.00	1.00
Bin 138_5	0.01	1.00	0.01	0.00
Bin 362	0.00	1.00	0.02	0.00
Bin 358	0.00	0.01	1.00	0.00
Bin 364	0.00	0.22	1.00	0.00
Bin 97	0.00	0.34	1.00	0.00
Bin 326_2	0.00	0.06	1.00	0.00
Bin 64_3	0.00	1.00	0.00	0.00
Bin 468	0.00	0.00	0.00	1.00

Bin 61	0.00	1.00	0.28	0.00
Bin 34	0.02	0.00	0.00	0.98
Bin 156	0.00	1.00	0.00	0.00
Bin 377	0.02	0.00	0.13	1.00
Bin 388	0.55	0.40	1.00	0.00
Bin 123_3	0.00	1.00	0.02	0.00
Bin 130	0.00	0.00	0.00	1.00
Bin 507	0.00	0.00	0.03	1.00
Bin 88	0.15	1.00	0.25	0.00
Bin 243_4	0.01	0.00	1.00	0.00
Bin 139_4	0.00	0.01	1.00	0.00
Bin 218	0.00	0.00	0.00	1.00
Bin 350	0.00	1.00	0.10	0.00
Bin 447	0.00	0.20	0.00	1.00
Bin 498	0.00	0.00	0.04	1.00
Bin 542	0.00	0.93	0.99	0.00
Bin 98	0.00	0.00	1.00	0.00
Bin 393	0.00	0.25	1.00	0.00
Bin 515	0.00	0.28	1.00	0.00
Bin 487	0.00	0.99	0.86	0.00
Bin 501	0.00	0.27	1.00	0.00
Bin 55_1	0.07	0.99	0.27	0.00
Bin 430	0.00	0.14	1.00	0.00
Bin 435	0.07	0.25	0.97	0.00
Bin 458	0.00	0.16	1.00	0.00
Bin 31	0.00	1.00	0.54	0.00
Bin 394_7_3	1.00	0.00	0.05	0.00
Bin 234	0.00	1.00	0.01	0.00
Bin 486	0.00	0.00	0.00	1.00

Bin 185	0.00	0.00	1.00	0.00
Bin 227	0.00	1.00	0.00	0.00
Bin 37	0.00	0.17	1.00	0.00
Bin 84	0.00	1.00	0.19	0.00
Bin 243_2_2	0.00	0.00	1.00	0.00
Bin 347	0.00	0.02	1.00	0.00
Bin 425	0.00	0.99	0.93	0.00
Bin 124_2	0.01	1.00	0.23	0.00
Bin 520_2	0.02	1.00	0.18	0.00
Bin 87	0.53	0.21	0.99	0.00
Bin 394_5b	1.00	0.00	0.03	0.00
Bin 235_4	0.00	1.00	0.01	0.00
Bin 464_4	0.00	0.00	0.00	1.00
Bin 188	0.00	0.07	1.00	0.00
Bin 437	0.00	0.00	0.00	1.00
Bin 12	0.01	0.01	1.00	0.00
Bin 531_1	0.00	0.01	1.00	0.00
Bin 424	0.00	1.00	0.19	0.00
Bin 325	0.00	1.00	0.37	0.00
Bin 254	0.00	0.26	1.00	0.00
Bin 118	0.01	0.03	1.00	0.13
Bin 411	0.00	0.00	1.00	0.00
Bin 3	0.01	0.03	1.00	0.00
Bin 464_3	0.00	0.00	0.01	1.00
Bin 310	0.00	0.00	0.00	1.00
Bin 252_2	0.00	0.00	0.00	1.00
Bin 189_2	0.00	0.00	0.00	1.00
Bin 215	0.00	0.00	1.00	0.00
Bin 496	0.00	0.00	0.00	1.00

Bin 8	0.00	0.00	0.00	1.00
Bin 92_4	0.00	0.00	1.00	0.00
Bin 279	0.62	0.31	0.99	0.00
Bin 268_1	0.01	0.05	1.00	0.00
Bin 235_3	0.00	1.00	0.01	0.00
Bin 555	0.00	1.00	0.47	0.00
Bin 394_4a	1.00	0.00	0.03	0.00
Bin 394_4b	1.00	0.00	0.03	0.01
Bin 338_2	0.00	0.00	0.00	1.00
Bin 285	0.00	0.00	0.00	1.00
Bin 75	0.03	0.20	0.02	1.00
Bin 449	0.01	0.01	1.00	0.00
Bin 360	0.09	0.04	1.00	0.00
Bin 169	0.00	0.54	1.00	0.00
Bin 543	0.00	1.00	0.33	0.00
Bin 408_1	0.00	0.00	0.00	1.00
Bin 418	0.00	0.18	0.00	1.00
Bin 128	0.02	0.05	1.00	0.00
Bin 229	0.00	1.00	0.00	0.00
Bin 394_5a	1.00	0.00	0.02	0.00
Bin 394_2c	1.00	0.00	0.05	0.00
Bin 396	0.00	1.00	0.09	0.00
Bin 394_7_1	1.00	0.00	0.04	0.00
Bin 410_2	0.00	0.00	0.00	1.00
Bin 338_6_2	0.00	0.00	0.00	1.00
Bin 291_1	0.00	1.00	0.00	0.00
Bin 365	0.00	0.07	1.00	0.00
Bin 76	0.00	1.00	0.74	0.00
Bin 210_2	0.00	0.00	0.00	1.00

Bin 421	0.05	0.22	0.03	1.00
Bin 290	0.00	0.99	0.08	0.00
Bin 66	0.00	0.01	1.00	0.00
Bin 394_6c	1.00	0.00	0.04	0.00
Bin 491	0.00	0.00	0.00	1.00
Bin 392	0.00	0.01	1.00	0.00
Bin 100	0.01	1.00	0.23	0.00
Bin 112	0.00	0.00	0.00	1.00
Bin 369	0.00	0.00	0.00	1.00
Bin 68	0.00	1.00	0.00	0.00
Bin 327_3	0.00	0.00	1.00	0.00
Bin 330_2	0.00	1.00	0.00	0.00
Bin 518_4	0.00	1.00	0.07	0.00
Bin 394_6b	1.00	0.00	0.04	0.00
Bin 278	0.01	1.00	0.07	0.00
Bin 71	0.00	0.00	1.00	0.00
Bin 307	0.00	0.01	1.00	0.00
Bin 231_6	0.00	0.00	0.00	1.00
Bin 409	0.00	1.00	0.02	0.00
Bin 129	0.00	0.00	0.00	1.00
Bin 508_6	0.00	0.00	0.00	1.00
Bin 494_5	0.00	1.00	0.00	0.00
Bin 320_2	0.00	0.01	1.00	0.00
Bin 28	0.03	0.01	1.00	0.00
Bin 306_7	0.00	0.00	1.00	0.00
Bin 334	1.00	0.00	0.04	0.00
Bin 513	0.00	0.35	1.00	0.00
Bin 242	0.00	1.00	0.09	0.00
Bin 210_3	0.00	0.00	0.00	1.00

Bin 394_6d	1.00	0.00	0.04	0.00
Bin 394_4c	1.00	0.00	0.04	0.00
Bin 394_7_2	1.00	0.00	0.03	0.00
Bin 327_4	0.00	0.00	1.00	0.00
Bin 453_2	0.00	0.02	1.00	0.00
Bin 317_3	0.00	0.99	0.24	0.00
Bin 81	0.04	0.29	0.03	1.00
Bin 473	0.00	0.08	0.00	1.00
Bin 394_1d	1.00	0.00	0.05	0.00
Bin 224	0.00	0.01	1.00	0.00
Bin 394_1a	1.00	0.00	0.06	0.00
Bin 381	0.01	0.02	1.00	0.00
Bin 508_4	0.00	0.00	0.00	1.00
Bin 203	0.00	1.00	0.00	0.00
Bin 47	0.00	0.00	1.00	0.00
Bin 341_6	0.00	0.00	0.00	1.00
Bin 520_4	0.01	1.00	0.28	0.00
Bin 160_4	0.00	0.03	1.00	0.00
Bin 244_2	0.00	0.00	0.00	1.00
Bin 257	0.00	0.00	0.00	1.00
Bin 386	0.00	1.00	0.06	0.00
Bin 394_2a	1.00	0.00	0.06	0.00
Bin 394_2d	1.00	0.00	0.05	0.01
Bin 548	0.00	0.00	0.00	1.00
Bin 474_1	1.00	0.00	0.00	0.00
Bin 394_6a	1.00	0.00	0.04	0.00
Bin 394_1c	1.00	0.00	0.07	0.00
Bin 107	0.00	0.00	0.00	1.00
Bin 268_3	0.00	0.00	1.00	0.00

Bin 22	0.00	0.01	1.00	0.00
Bin 474_4	1.00	0.00	0.00	0.00
Bin 400	0.00	1.00	0.00	0.00
Bin 164	0.00	0.00	1.00	0.00
Bin 306_4	0.00	0.00	1.00	0.00
Bin 67_1	0.00	1.00	0.07	0.00
Bin 494_4	0.00	1.00	0.03	0.00
Bin 326_3	0.00	0.00	1.00	0.00
Bin 394_2e	1.00	0.00	0.06	0.00
Bin 394_2b	1.00	0.00	0.04	0.00
Bin 394_1g	1.00	0.00	0.05	0.00
Bin 349	0.00	0.00	0.00	1.00
Bin 394_1e	1.00	0.00	0.05	0.00
Bin 394_1b	1.00	0.00	0.07	0.00
Bin 394_2f	1.00	0.00	0.07	0.00
Bin 474_2	1.00	0.00	0.00	0.00
Bin 132	0.00	0.01	1.00	0.00
Bin 139_3	0.01	0.00	1.00	0.00
Bin 196	0.00	0.01	1.00	0.00
Bin 394_2h	1.00	0.00	0.05	0.00
Bin 394_1f	1.00	0.00	0.04	0.00
Bin 102	0.01	0.00	0.06	1.00
Bin 40	0.00	0.00	1.00	0.01
Bin 456	0.01	0.01	1.00	0.00
Bin 467	0.00	1.00	0.05	0.00
Bin 490	0.00	0.00	0.00	1.00
Bin 394_4d	1.00	0.00	0.04	0.00
Bin 394_2g	1.00	0.00	0.07	0.00
Bin 341_2	0.00	0.00	0.00	1.00

Bin 464_5	0.00	0.00	0.01	1.00
Bin 518_3	0.00	1.00	0.02	0.00
Bin 160_3	0.09	0.01	1.00	0.00
Bin 207_2	0.00	1.00	0.00	0.00
Bin 291_3	0.03	1.00	0.20	0.00
Bin 123_4	0.00	1.00	0.07	0.00
Bin 512_3	0.00	1.00	0.00	0.00
Bin 368_3	0.00	0.00	0.00	1.00
Bin 320_4	0.00	0.01	1.00	0.00
Bin 232_4	0.00	0.04	1.00	0.00
Bin 64_2	0.00	1.00	0.00	0.00
Bin 508_5	0.00	0.00	0.00	1.00
Bin 518_2	0.00	1.00	0.00	0.00
Bin 317_2	0.02	1.00	0.00	0.00
Bin 232_3	0.00	0.00	1.00	0.00
Bin 531_4	0.00	0.00	1.00	0.00

Supplementary Table 4. *Veturius* sp. genes of the major CAZy enzymes classification groups normalized to total gene count of the MAG.

MAG	GHs	GTs	CBMs	CEs	AAs	PLs
MAG106	0.03607333	0.09047901	0.0277942	0.01301005	0	0.00059137
MAG11	0.06710875	0.09973475	0.03156499	0.02254642	0.00265252	0.00557029
MAG113	0.02760524	0.07522429	0.01863354	0.00483092	0	0
MAG123	0.07440311	0.07773459	0.047196	0.01499167	0.00055525	0.00666297
MAG124	0.03963893	0.07731554	0.02433281	0.00588697	0	0
MAG127	0.08756941	0.07090987	0.02947458	0.01366937	0	0.00170867
MAG133	0.03730846	0.07461692	0.02598268	0.01199201	0	0.00199867
MAG134	0.0407498	0.08394458	0.02852486	0.00814996	0	0
MAG139	0.06689342	0.06386999	0.03590325	0.01284958	0.00037793	0.00340136
MAG14	0.02758621	0.08401254	0.02570533	0.01191223	0	0
MAG143	0.05448065	0.07433809	0.01527495	0.01069246	0.00050916	0.00050916
MAG155	0.06357173	0.07695525	0.01589293	0.01045588	0	0.00041824
MAG16	0.09223847	0.09561305	0.02587177	0.01724784	0.00037495	0.00187477
MAG161	0.0691009	0.07027876	0.02630546	0.01099333	0	0.00235571
MAG163	0.03082852	0.10115607	0.01926782	0.01589595	0.00192678	0
MAG167	0.05087659	0.09659677	0.02990718	0.01512547	0.00309385	0
MAG17	0.04940711	0.085639	0.03623188	0.00988142	0	0.00065876
MAG172	0.02491103	0.04181495	0.02313167	0.00622776	0	0.00088968
MAG174	0.13458012	0.07224056	0.03630363	0.00990099	0.0003667	0.00256692
MAG178	0.05584281	0.07549121	0.03826267	0.0188728	0.00206825	0.00853154
MAG184	0.05428987	0.07561803	0.02278236	0.01357247	0.00048473	0
MAG191	0.05645816	0.10966694	0.03452478	0.01665313	0.0016247	0.0016247
MAG193	0.07286673	0.07142857	0.03403643	0.01486098	0.00047939	0.0033557
MAG195	0.08690987	0.07403433	0.03165236	0.01180258	0.00053648	0.00053648
MAG201	0.03332211	0.08515651	0.02356109	0.01581959	0.00639515	0.00067317

MAG202	0.04580897	0.07651072	0.01900585	0.01461988	0.00146199	0.00097466
MAG206	0.08336463	0.07960946	0.03492302	0.01502065	0	0.0022531
MAG207	0.12779553	0.07348243	0.0397079	0.01277955	0.00091283	0.01232314
MAG208	0.04443383	0.07978978	0.02341137	0.00668896	0.00047778	0
MAG211	0.0516129	0.11520737	0.0328725	0.01996928	0.0015361	0.00675883
MAG217	0.08096367	0.09992101	0.0335703	0.01184834	0	0.00078989
MAG223	0.09594645	0.0792116	0.02491633	0.01375976	0	0.00446263
MAG225	0.04558707	0.09117413	0.01219189	0.02120329	0.0045057	0.00079512
MAG23	0.03483607	0.06215847	0.0239071	0.00751366	0	0.00068306
MAG230	0.05736138	0.07552581	0.02772467	0.01481836	0.00047801	0.00573614
MAG232	0.0300158	0.05924171	0.02290679	0.01105845	0	0
MAG24	0.08698413	0.08253968	0.03746032	0.01650794	0.00063492	0.00253968
MAG248	0.0377943	0.08612144	0.0291202	0.0086741	0.00123916	0.00061958
MAG250	0.04011461	0.08627826	0.01973894	0.01878383	0.00541229	0.00095511
MAG260	0.02791316	0.07487816	0.01595038	0.01019052	0.0013292	0.00044307
MAG266	0.05645876	0.07833284	0.0268992	0.01684895	0.00354715	0.00177357
MAG270	0.04307692	0.07384615	0.03692308	0.01046154	0	0.00123077
MAG272	0.04593373	0.07567771	0.02673193	0.0188253	0.00075301	0.00150602
MAG284	0.06871838	0.06576728	0.04637437	0.01602024	0.00210793	0.00210793
MAG288	0.0306834	0.11715481	0.0204556	0.0153417	0.0009298	0.0004649
MAG301	0.05864752	0.09275883	0.02274087	0.01137044	0.00119689	0
MAG308	0.02822322	0.09108403	0.01667736	0.01539448	0.00064144	0
MAG315	0.03424223	0.08877616	0.03360812	0.01331642	0	0
MAG316	0.03099017	0.06198035	0.02191988	0.00680272	0	0
MAG317	0.03046716	0.07379824	0.01624915	0.01150982	0	0
MAG322	0.06428571	0.06116071	0.02455357	0.009375	0	0.00223214
MAG330	0.03411029	0.05628198	0.02728823	0.00795907	0	0.0005685
MAG340	0.0385289	0.07589025	0.03327496	0.01517805	0	0.00058377
MAG357	0.05622837	0.08217993	0.02422145	0.01211073	0	0

MAG361	0.05170316	0.0663017	0.0377129	0.01216545	0	0.00304136
MAG363	0.06432459	0.07991094	0.03884216	0.0212766	0.00593765	0.00074221
MAG372	0.06277007	0.07596088	0.01546509	0.01751194	0.00204685	0.00068228
MAG379	0.04041204	0.09587956	0.01743265	0.01030111	0	0.00079239
MAG397	0.05087719	0.11278195	0.02631579	0.01654135	0.00175439	0.00150376
MAG398	0.09022082	0.06687697	0.0340694	0.01388013	0	0.00252366
MAG399	0.06356099	0.08132649	0.01223845	0.01381761	0.00039479	0.00039479
MAG4	0.03924491	0.07848982	0.02136115	0.00943865	0	0.00099354
MAG405	0.02834225	0.07272727	0.02620321	0.00748663	0	0.00053476
MAG406	0.03016086	0.06367292	0.02345845	0.00737265	0.00067024	0.00201072
MAG407	0.08317472	0.07213552	0.03102398	0.02055577	0.00247431	0.00190331
MAG41	0.06481013	0.09265823	0.02278481	0.01367089	0.00050633	0.00050633
MAG433	0.03715776	0.09322034	0.01955671	0.01173403	0	0.00065189
MAG438	0.02520161	0.06653226	0.02318548	0.00403226	0	0
MAG440	0.03512623	0.06860593	0.02744237	0.01536773	0.00109769	0.00164654
MAG441	0.03974654	0.0921659	0.02073733	0.01382488	0.00057604	0.00115207
MAG442	0.03275261	0.07317073	0.02926829	0.00905923	0	0.00139373
MAG445	0.06034858	0.08518519	0.02374728	0.01851852	0.00566449	0.00130719
MAG446	0.05684211	0.06877193	0.02245614	0.00912281	0	0.00070175
MAG461	0.03716091	0.10850985	0.02526942	0.01263471	0.00371609	0.00222965
MAG470	0.08040422	0.07601054	0.02899824	0.01362039	0	0.0013181
MAG481	0.10627247	0.07710747	0.02556932	0.0155813	0	0.00239712
MAG485	0.04231228	0.08224076	0.03098927	0.00774732	0	0.0011919
MAG502	0.0734447	0.11175115	0.03197005	0.01612903	0.00201613	0.00345622
MAG503	0.10147213	0.06992639	0.03259727	0.01209253	0	0.00052576
MAG505	0.06852444	0.06989493	0.03197807	0.01096391	0	0.00091366
MAG508	0.03811406	0.09119142	0.03585545	0.01637493	0.0053642	0.00254094
MAG511	0.07778915	0.09160696	0.0174002	0.01023541	0	0.00204708
MAG518	0.0566879	0.07770701	0.04140127	0.01082803	0.00063694	0.00063694

MAG519	0.13147139	0.06982289	0.04087193	0.01089918	0.0003406	0.0023842
MAG522	0.05054645	0.08128415	0.01775956	0.01912568	0.00068306	0
MAG526	0.04177411	0.08767406	0.03094379	0.00928314	0.00051573	0.00103146
MAG536	0.04539474	0.07565789	0.02236842	0.00723684	0	0.00131579
MAG549	0.0472534	0.08387478	0.0218547	0.02008269	0.001772	0
MAG552	0.05078125	0.07080078	0.03027344	0.00927734	0.00048828	0.00097656
MAG553	0.06543967	0.06237219	0.03271984	0.00715746	0	0.00408998
MAG57	0.06980938	0.08140134	0.02782071	0.02395672	0.00489438	0.00360639
MAG58	0.03283369	0.07066381	0.02712348	0.01498929	0	0
MAG64	0.03789731	0.05806846	0.02322738	0.00672372	0	0.00061125
MAG72	0.08145766	0.08145766	0.02197213	0.01393355	0	0.00107181
MAG74	0.04705882	0.1124183	0.02614379	0.00980392	0.00065359	0.00130719
MAG78	0.04814004	0.07795405	0.04595186	0.01695842	0.0035558	0.01531729
MAG83	0.08452594	0.08005367	0.04025045	0.01788909	0.00044723	0.009839
MAG85	0.0251046	0.05299861	0.03138075	0.00488145	0	0
MAG86	0.05185185	0.07314815	0.02361111	0.01435185	0.00092593	0.00138889
MAG89	0.11995899	0.08373206	0.02768284	0.01435407	0.00068353	0.00205058
MAG96	0.05298701	0.06857143	0.0187013	0.01246753	0	0.00571429

Supplementary Table 5. Glycosyl hydrolases in the seven *Veturius* sp. metagenomes. Data normalized to the metagenome bin count.

	Adult	**Substrate**	**Larvae 1**	**Larvae 2**	**Larvae 3**	**Larvae 4**	**Larvae 5**
Chitinase	0	2.3091E-05	6.5838E-06	0	2.7738E-06	2.4022E-06	0
Glycoside hydrolase 97	2.3528E-06	3.9255E-05	0.00013168	7.984E-05	8.4879E-05	7.567E-05	4.8462E-05
Glycosyl hydrolase 1	0.00014117	0.00029326	0.00020222	0.00026347	0.00015811	0.00028106	8.2385E-05
Glycosyl hydrolase 10	3.7644E-05	0.00017318	0.00015895	0.00016766	0.00010152	0.00014173	0.00015992
Glycosyl hydrolase 12	0	4.6182E-06	0	0	0	0	0
Glycosyl hydrolase 14	0	0	9.4055E-07	0	7.7667E-06	4.8045E-06	0
Glycosyl hydrolase 20	1.1764E-05	0.0002794	0.00019375	0.00011178	0.00016809	0.00015855	0.00016477
Glycosyl hydrolase 26	0	5.7728E-05	7.4303E-05	7.1856E-05	6.3243E-05	5.405E-05	3.3923E-05
Glycosyl hydrolase 3	7.0583E-05	0.00061884	0.00054928	0.0004471	0.0004133	0.00041438	0.00052339
Glycosyl hydrolase 30	1.6469E-05	0.00013393	0.00014484	6.3872E-05	0.00011206	0.00014293	0.00017931
Glycosyl hydrolase 4	5.1761E-05	1.6164E-05	0.00017024	0.00011976	0.00018918	0.00016695	9.6923E-06
Glycosyl hydrolase 45	0	4.6182E-06	0	0	0	0	0
Glycosyl hydrolase 46	0	6.9273E-06	0	0	0	0	0
Glycosyl hydrolase 47	1.4117E-05	1.6164E-05	0	0	0	0	0
Glycosyl hydrolase 48	0	4.6182E-06	0	0	0	0	0
Glycosyl hydrolase 49	0	2.3091E-06	9.4055E-07	0	1.6643E-06	0	0
Glycosyl hydrolase 52	0	2.3091E-06	0	0	5.5477E-07	0	0
Glycosyl hydrolase 53	2.3528E-06	6.4655E-05	1.787E-05	3.1936E-05	1.276E-05	2.162E-05	2.4231E-05
Glycosyl hydrolase 57	2.3528E-06	6.4655E-05	0.00010628	0.00010379	7.7667E-05	9.2486E-05	0.00011146
Glycosyl hydrolase 59	0	6.9273E-06	2.8216E-06	7.984E-06	7.7667E-06	1.2011E-06	4.8462E-06
Glycosyl hydrolase 61	0	0	0	0	0	0	0
Glycosyl hydrolase 62	0	0	0	0	4.4381E-06	3.6033E-06	0
Glycosyl hydrolase 63	1.6469E-05	2.7709E-05	4.7027E-06	7.984E-06	7.212E-06	3.6033E-06	1.9385E-05
Glycosyl hydrolase 65	2.1175E-05	2.54E-05	2.6335E-05	3.1936E-05	1.9972E-05	3.9637E-05	9.6923E-06
Glycosyl hydrolase 67	4.7055E-06	2.7709E-05	4.1384E-05	2.3952E-05	3.6615E-05	2.8827E-05	3.3923E-05
Glycosyl hydrolase 7	0	0	0	0	0	0	0
Glycosyl hydrolase 70	0	0	0	0	1.1095E-06	0	0
Glycosyl hydrolase 71	0	1.1546E-05	9.4055E-07	0	1.1095E-06	1.2011E-06	0
Glycosyl hydrolase 76	2.3528E-06	1.8473E-05	6.0195E-05	0.00011178	3.9388E-05	6.0056E-05	9.6923E-06
Glycosyl hydrolase 79	0	2.3091E-06	1.8811E-06	0	1.1095E-06	1.2011E-06	0
Glycosyl hydrolase 81	0	1.6164E-05	0	0	0	6.0056E-06	0
Glycosyl hydrolase 85	4.7055E-06	2.3091E-06	0	0	0	4.8045E-06	0
Glycosyl hydrolase 9	0	6.0037E-05	4.8908E-05	0	4.2717E-05	3.8436E-05	3.3923E-05
Glycosyl hydrolase 92	4.7055E-06	0.00028633	0.00064898	0.00033533	0.00042495	0.00056092	0.00068815

Glycosyl hydrolase 108	0	9.2364E-06	3.1979E-05	1.5968E-05	9.9858E-06	1.3212E-05	2.4231E-05
Glycosyl hydrolase 5	4.7055E-06	0.00015471	0.00017118	5.5888E-05	0.00019195	0.00019698	0.00013085
Glycosyl Hydrolase Family 88	5.1761E-05	5.5419E-05	0.00034236	0.00030339	0.00018862	0.00020659	0.00035862
Glycosyl hydrolases 11	0	1.8473E-05	9.4055E-07	0	1.1095E-06	2.4022E-06	0
Glycosyl hydrolases 15	0	8.7746E-05	9.4055E-07	0	4.4381E-06	1.2011E-06	0
Glycosyl hydrolases 16	1.4117E-05	0.00012469	0.00012509	3.1936E-05	9.0427E-05	8.5279E-05	0.00014054
Glycosyl hydrolases 17	0	9.2364E-06	9.4055E-07	0	1.6643E-06	1.2011E-06	4.8462E-06
Glycosyl hydrolases 18	2.1175E-05	0.00017087	0.00014578	8.7824E-05	0.00011151	8.7681E-05	7.7539E-05
Glycosyl hydrolases 2	2.8233E-05	0.00029557	0.00066214	0.00042315	0.00060525	0.00056452	0.000504
Glycosyl hydrolases 25	0	4.8491E-05	5.9254E-05	9.5808E-05	5.936E-05	6.9665E-05	2.4231E-05
Glycosyl hydrolases 28	7.0583E-06	8.5437E-05	0.00010158	0.00010379	6.9901E-05	7.567E-05	0.00010177
Glycosyl hydrolases 31	1.6469E-05	0.00023784	0.00027182	0.00020758	0.00017253	0.00018978	0.00020354
Glycosyl hydrolases 32	0	4.1564E-05	9.4055E-05	6.3872E-05	8.2106E-05	0.00012371	2.4231E-05
Glycosyl hydrolases 35	1.8822E-05	7.851E-05	5.0789E-05	4.7904E-05	2.9957E-05	3.3631E-05	5.8154E-05
Glycosyl hydrolases 38	2.8233E-05	4.3873E-05	0.00015331	7.1856E-05	0.0002502	0.00012612	3.8769E-05
Glycosyl hydrolases 39	0	9.0055E-05	5.6433E-05	1.5968E-05	5.4367E-05	4.4441E-05	1.4538E-05
Glycosyl hydrolases 43	6.3525E-05	0.00019396	0.0004524	0.00036726	0.00028903	0.000436	0.00046039
Glycosyl hydrolases 6	0	6.9273E-06	0	0	0	0	0
Glycosyl hydrolases 8	7.0583E-06	9.4673E-05	1.0346E-05	2.3952E-05	8.8763E-06	1.2011E-05	9.6923E-06

Supplementary Table 6. Normalized ratio of the archaeal bins in the metagenomic samples.

	Adult *	Larvae3 *	Larvae4 *	Substrate *
MAG 260	1.00	0.00	0.00	0.00
Bin 348	0.00	1.00	0.00	0.00
Bin 517	0.00	1.00	0.00	0.00
Bin 146	0.00	0.90	0.57	0.00
Bin 101	0.05	0.01	1.00	0.00
Bin 5	0.00	0.85	0.64	0.00
Bin 435	0.07	0.25	0.97	0.00
Bin 299	0.00	0.64	0.89	0.00
Bin 87	0.53	0.21	0.99	0.00
Bin 458	0.00	0.16	1.00	0.00

*Maximum normalized ratio= number of reads recruited to a contig divided by the maximum number of reads recruited to that contig in any sample.

Supplementary Table 7. Nitrogenase subunits and *anfG* in the *Veturius* sp. gut and substrate metagenomes. Data normalized to the metagenome gene count.

	Adult	Substrate	Larvae 1	Larvae 2	Larvae 3	Larvae 4	Larvae 3
anfG	0	0	0	0	0	0	0
nifD	0	0	3.7622E-06	0	1.6643E-06	1.2011E-06	0
nifH	1.1764E-05	2.3091E-06	4.9849E-05	5.5888E-05	2.2745E-05	3.243E-05	6.7846E-05
nifK	0	2.3091E-06	0	0	1.6643E-06	2.4022E-06	9.6923E-06