Microbiome and aging: A study of microbial evolution and community structure across model organisms

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Francisco Daniel Davila Aleman

aus Mexiko-Stadt

Köln 2022

Gutachter: Dr. Dario Riccardo Valenzano

- Para mis padres, hermano y esposa -

Thesis Structure

The present thesis is structured in two main parts:

1. I explored the evolution of the gut microbiome in response to host aging.
2. I characterized the gut microbiota of the aging model, the turquoise killifish *Nothobranchius furzeri*, using short- and long-read metagenomics and culturomics approaches.

All scheme figures were generated using Biorender (https://biorender.com/)

Kurzzusammenfassung

Abschnitt I: Entwicklung kommensaler Darmmikroben während der Alterung des Wirts

Der Darm von Tieren enthält eine große Anzahl ansässiger Mikroorganismen (Darmmikrobiota), die einen erheblichen Einfluss auf die Physiologie des Wirts haben. Die Zusammensetzung der Darmmikrobiota ändert sich während der gesamten Lebensspanne des Wirts dramatisch, da die Darmmikroben mit zahlreichen Veränderungen in ihrer ökologischen Nische konfrontiert sind. Altersbedingte Veränderungen im Umfeld des Wirts üben einen starken Selektionsdruck auf die mikrobielle Gemeinschaft aus. Ob und wie sich die Darmmikroben während der Lebensspanne des Wirts entwickeln und wie sich dies auf den altersbedingten Phänotyp des Wirts auswirkt, ist eine offene Frage. Insbesondere können neue genetische Varianten im Mikrobiom dem Wirt neue adaptive Eigenschaften verleihen oder einen virulenten Verlauf nehmen, der die Homöostase zwischen Wirt und Mikrobiota beeinträchtigt und die Lebensspanne des Wirts begrenzt. Frühere Studien haben gezeigt, dass sich das Darmmikrobiom schnell anpassen und in kurzer Zeit weiterentwickeln kann. Über die evolutionäre Dynamik dieser komplexen mikrobiellen Gemeinschaft während der Alterung des Wirts innerhalb desselben Individuums ist jedoch weniger bekannt. Mein Forschungsprojekt zielt darauf ab, zu untersuchen, wie sich die Darmmikrobiota während der Lebensspanne des Wirts entwickelt und ob evolutionäre Prozesse in den mikrobiellen Gemeinschaften des Darms die mikrobielle Fitness beeinflussen.

Ich nutze drei verschiedene metagenomische Sequenzierungsstrategien, um mehr als 200 verschiedene metagenomisch assemblierte Genome von drei verschiedenen individuellen

Wirten wieder zusammenzusetzen. Anhand dieses Genomkatalogs identifizierte und verfolgte ich Einzelnukleotidvarianten, die Signaturen der Selektion während des Wirtslebens aufweisen. Insgesamt stellte ich fest, dass die meisten neu entdeckten Varianten von einer reinigenden Selektion betroffen sind. Ich fand jedoch heraus, dass genetische Varianten in äußeren Membranproteinen, kohlenhydrataktivierenden Enzymen und Transposasegenen Signaturen positiver Selektion ohne offensichtliche positive Auswirkungen auf die mikrobielle Fitness aufweisen. Die in dieser Arbeit erzielten Ergebnisse werden dazu beitragen, die Auswirkungen des Alterns des Wirts auf die Evolution des Mikrobioms zu verstehen und herauszufinden, ob die Evolution zur Entstehung krankheitsfördernder Bakterienstämme bei älteren Menschen beiträgt.

Abschnitt II: Charakterisierung der mikrobiellen Darmgemeinschaft des alternden Modells *Nothobranchius furzeri* mit Hilfe von Short- und Long-Read-Metagenomik

Der Afrikanische Türkis-Killerfisch (*Nothobranchius furzeri*) ist ein aufstrebendes Modell für die Altersforschung, da er von Natur aus eine kurze Lebensspanne hat, unter Laborbedingungen leicht zu halten ist und typische Alterungsmerkmale von Wirbeltieren aufweist. Jüngste Arbeiten an dieser Art haben gezeigt, dass die Übertragung der Darmmikrobiota von jungen auf mittelalte Killifische die Lebensspanne der Empfänger verlängert und den altersbedingten Verhaltensabbau verzögert. Im Vergleich zu anderen Wirbeltieren (Zebrabärblinge, Mäuse, Menschen) besitzt der Afrikanische Türkis Killerfisch eine komplexe mikrobielle Vielfalt, die während des natürlichen Alterungsprozesses deutlich abnimmt, während gleichzeitig die relative Häufigkeit potenziell pathogener Bakterien bei älteren Individuen zunimmt. Während sich die Charakterisierung der mit Killifischen assoziierten Mikrobiota auf die 16S-Amplikon-Sequenzierung konzentriert hat, fehlt eine Darstellung des intestinalen Metagenoms dieser Art.

Um diese Lücke zu schließen, kombiniere ich die Short-Read- mit der Long-Read-Sequen zierungstechnologie, um das Metagenom des Stuhls von in Gefangenschaft lebenden türkisen Killifischen zu analysieren. Ich fand heraus, dass Bakterien den größten Teil der Stuhlmikrobiota ausmachen (73%), wobei Proteobakterien der häufigste Stamm sind (93%), gefolgt von Actinobakterien (4%), Firmicutes (2%) und Bacteroidetes (2%). Viren

machen 0,04% des gesamten Stuhlmikrobioms aus, gefolgt von Archaea (0,02%) und Fungi (0,02%). Bemerkenswert ist, dass 27% des gesamten Stuhl-Metagenoms keiner Referenzdatenbank (RefSeq) zugeordnet werden können, was potenziell neue, mit dem Wirt assoziierte Arten darstellt.

Um einen Katalog von Bakteriengenomen zu erstellen, die mit dem Darm von Killifischen assoziiert sind, habe ich die Genome von 8 der häufigsten Bakterienarten durch Kultivierung, Isolierung, Sequenzierung und Hybridassemblierung mit Illumina- und Long-Read-Nanopore-Sequenzierungstechnologie assembliert. Die Analyse der vollständig assemblierten bakteriellen Genome des Darms ergab das Vorhandensein von mobilen Elementen, Prophagen, antibiotikaresistenten Genen und anderen genetischen Elementen, die möglicherweise zu bakteriellen Anpassungen im Wirt beitragen könnten. Insgesamt stellt die Metagenomik beim Türkisen Killifisch eine wichtige neue Ressource für diese Art dar, die dazu beitragen wird, diese Art als natürliches kurzlebiges Modell zur Untersuchung der Interaktionen zwischen Wirt und Mikrobiom während des Alterns weiterzuentwickeln.

Abstract

Section I : Evolution of commensal gut microbes during host aging

The animal intestine contains a large number of resident microorganisms (gut microbiota), which has a significant impact on host physiology. The gut microbiota composition changes dramatically throughout the host lifespan as the gut microbes face multiple changes in their ecological niche. Age-associated changes in the host environment exert strong selective pressures on the microbial community, and whether and how gut microbes evolve during host life span and how this affects host age-associated phenotype is an open question. Specifically, novel genetic variants in the microbiome can confer new adaptive traits to the host or follow a virulent trajectory that affects host-microbiota homeostasis and delimit host lifespan. Previous studies demonstrated that the gut microbiome could rapidly adapt and evolve in a short timescale. However, less is known about the evolutionary dynamics of this complex microbial community during host aging within the same individual host. My research project aims to survey how the gut microbiota evolves during the host´s life span and whether evolutionary processes in the gut microbial communities affect microbial fitness.

I leverage three different metagenomic sequencing strategies to reassemble more than 200 different metagenomic-assembled genomes from three different individual hosts. Using this genome catalog, I identified and tracked single-nucleotide variants that show signatures of selection during host life. Overall, I found that most new variants detected are impacted by purifying selection. However, I found that genetic variants in outer membrane proteins, carbohydrate-activate enzymes, and transposase genes show signatures of positive selection with no obvious beneficial effect on microbial fitness. The results

obtained in this thesis will help understand the impact of host aging in determining microbiome evolution and whether evolution acts for the emergence of disease-promoting bacterial bacteria strains in the elderly.

Section II: Characterization of the gut microbial community of the aging model *Nothobranchius furzeri* using short and long read metagenomics

The African turquoise killifish (*Nothobranchius furzeri*) is an emerging model for aging research, thanks to its naturally short lifespan, easy husbandry under laboratory conditions, and the display of typical vertebrates hallmarks of aging. Recent work in this species has shown that gut microbiota transfer from young to middle-age killifish increases the recipients' life span and delays age-related behavioral decline. The African turquoise killifish possesses a complex microbial diversity compared to other vertebrates (zebrafish, mouse, humans), which significantly decline during natural aging, with concurrent increase in the relative abundance of potentially pathogenic bacteria in older individuals. While the characterization of killifish-associated microbiota has focused on 16S amplicon sequencing, an account of the intestinal metagenome in this species is lacking.

To fill this gap, I combine short-read with long-read sequencing technology to analyze the stool metagenome in captive turquoise killifish. I found that bacteria represent the majority of the stool microbiota (73%), with Proteobacteria as the most abundant Phylum (93%), followed by Actinobacteria (4%), Firmicutes (2%), and Bacteroidetes (2%). Viruses represent 0.04% of the total stool microbiome, followed by Archaea (0.02%) and Fungi (0.02%). Notably, 27% of the total stool metagenome does not map to any reference database (RefSeq) representing potentially novel host-associated species.

Additionally, to generate a catalog of bacterial genomes associated with killifish gut, I assembled the genome of 8 of the most abundant bacterial species by culturing, isolating, sequencing, and hybrid assembly strategy, involving Illumina and long-read Nanopore sequencing technology. The analysis of the fully-assembled intestinal bacterial genomes revealed the presence of mobile elements, prophages, antibiotic-resistant genes, and other genetic elements that could potentially contribute to bacterial adaptations in the host.

Together, metagenomics in the turquoise killifish represents a critical new resource for this species, which will help further develop this species as a naturally short-lived model to study host-microbiome interactions during aging.

Acknowledgments

First of all, I would like to thank my supervisor, Dr. Dario Valenzano, for giving me the opportunity to be part of his wonderful research group and for the guidance during the whole project. To Dr. Prof. Marcel Bucher and Dr. Prof. Michael Lässig for their willingness to examine this work and join my defense committee. To Dr. Angela Hancock for their support, ideas, discussion, and valuable advice during our TAC meetings. The whole Valenzano group for their helpful advice, discussion, and excellent time that we spent together in the lab. This work would not be possible without the valuable collaboration with Dr. Linda Partridge Lab; I will always be grateful. To Dr. Karen Guillemin Lab for sharing all the bacterial strains, support, and knowledge concerning the genetic modification of gut microbes. I want to thank the Bioinformatic core facility for their help and guidance during the whole project. Special thanks to all my true friends from the CGA, particularly Gabriel, Stephanie, and Paulo, who always made me feel at home. To the CGA coordinators, Dr. Daniela Morick, Jenny Ostermann, and Dr. Julia Zielinski, for all the support and for always helping me find my way in Germany. Finally, to my mom, dad, brother, and wife. You teach me to fight for my dreams and always keep moving forward.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Gut microbiota role in host health and aging

The gut microbiota is a diverse and dynamic community of microorganisms composed of bacteria, archaea, fungi, unicellular eukaryotes, and viruses (161). This microbial community is required for the development and maturation of the host's intestinal epithelium and immune system, facilitates host metabolism, produces essential vitamins, influences organ morphogenesis, tissue regeneration, bone homeostasis, and behavior, regulates the gut endocrine function and neurological signaling, modifies drug action, and provides a physical barrier that protects against foreign pathogens (95, 161, 162, 163). For example, non-digestible dietary fibers by intestinal epithelial cells in humans, such as xyloglucans, fructooligosaccharides, and oligosaccharides, are degraded by bacteria species of Bacteroides, Lactobacillus, and Bifidobacterium, providing energy to the host intestinal epithelium and promoting gut homeostasis (163).

On the other hand, the gut microbiota also benefits from host colonization, as the intestine represents a nutrient-rich environment (162). However, the intestinal environment is a dynamic habitat that undergoes physiological changes influenced by multiple host factors, such as lifestyle, diet, medication, hygiene, aging, and genetics, which strongly modulate the structure and composition of the microbiota (164) (Figure 1). Thus, the microbiota community composition and the collection of all intestinal microbial genes (the gut microbiome) can rapidly change as a response to intestinal environmental modifications,

resulting in modified microbial function and influencing host physiology and metabolism (162, 164, 166).

The microbiome community has to be controlled and restricted to ensure the benefits of the mutualistic association with the host. Consequently, a disruption of the balance in the composition and function of the gut microbiota leads to detrimental effects and disturbance of the host-microbiota homeostasis (165). The interruption of the host-microbiota homeostasis, defined as dysbiosis, is associated with the onset of gastroenterological disorders, neurological and cardiovascular illness, asthma, and metabolic disorders including obesity, type 2 diabetes, and cardio-metabolic diseases (95, 164, 167, 168). For example, dysbiosis is associated with intestinal inflammation and reduced integrity of the gut barrier, facilitating the translocation of bacterial components (Lipopolysaccharides, short-chain fatty acids), causing inflammatory responses, and promoting the development of cardiovascular diseases (168, 169).

One of the main contributing factors to microbial dysbiosis is age-related physiological and immune changes (171). Intestinal physiology changes with age, and because of the tight relationship between microbiota and the host environment, these physiological changes contribute to the outgrowth of specific microbial species and the reduction in the frequency of other microbes (172). For example, in mice, mucin production (highly glycosylated proteins which protect intestinal epithelial cells (173)) decreases during aging (172, 174), reducing the nutrients required by Akkermansiaceae, Clostridiaceaea, Bifidobacteriaceae families, changing their abundance in the community (174). Importantly, reduced levels of the mucin-metabolizing Akkermansia muciniphila have been observed in humans with inflammatory bowel disease and metabolic disorders (175).

While the gut microbiota composition suffers rapid changes in early life and remains relatively stable throughout adulthood, aging induces significant changes in microbiota structure, a decline in microbial diversity, and an increase in frailty (8, 177, 178, 179). This decline in microbial diversity and changes in the microbial composition is related to a decrease in the immune response observed in the elderly (immunosenescence) (179).

Immunosenescence is characterized by reducing the quality and quantity of the immune response and the effectiveness of the cellular response against infections (179, 180). It is

hypothesized that the progressive decline of the immune fitness during host aging directly affects the regulation of the gut microbiome by the immune system, impairing the community balance (dysbiosis) and resulting in severe consequences to host health and frailty (7 ,179, 181).

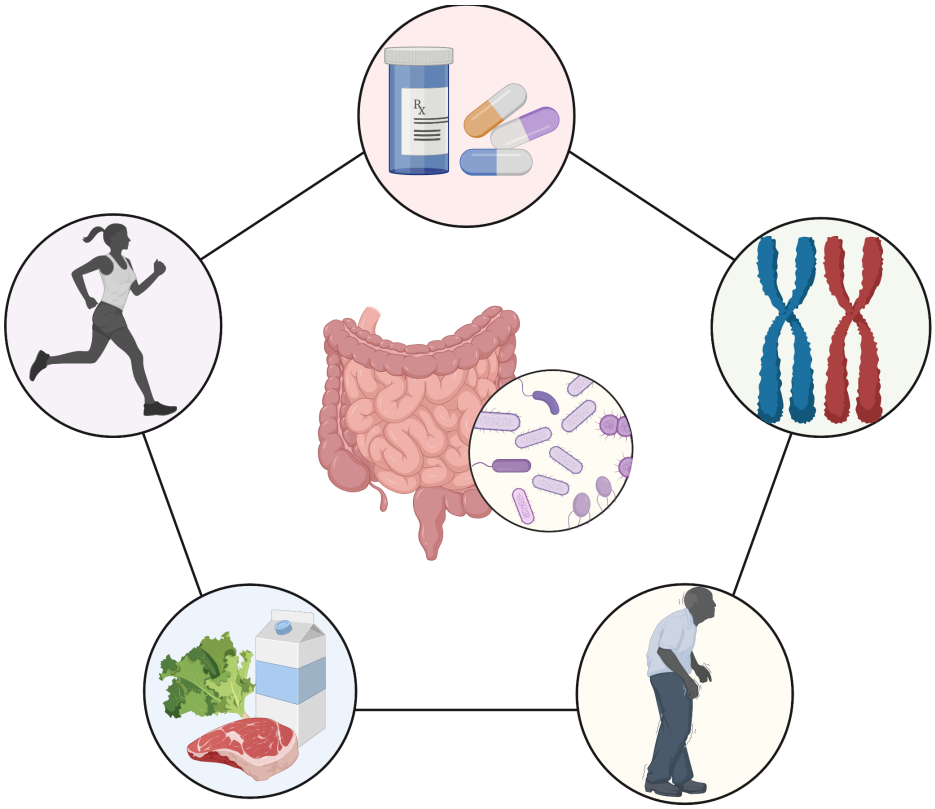


Figure 1.1: Factors that influence gut microbiota composition. The composition of the gut microbiota is influenced by host lifestyle, diet, medication, aging, hygiene and host genetics. Microbiota response to these factors by genetic, transcriptional and metabolics changes (Adapted from Sommer et al. 2013). Created with BioRender.com.

Increasing evidence using model organisms demonstrated the role of the gut microbiome to modulate aging and host lifespan. Hartsough et al. showed in the model organism *Caenorhabditis elegans* that controlling metabolite production of bacteria residing in the host gut can significantly extend the host lifespan (182). Guo et al. described that preventing gut dysbiosis in the aging intestine of *Drosophila melanogaster* promotes tissue homeostasis and expands host lifespan (183). Moreover, Smith et al. observed the extension lifespan and a delay in age-related behavior by transplanting of young microbiota community to a middle-age turquoise killifish (8). Similarly, Bárcena et al. describe that mouse models of accelerating aging phenotype (progeria) are characterized by intestinal

dysbiosis and showed that fecal microbiota transplantation (FMT) from wild-type mice increases host lifespan. Interestingly, progeria mice model administered with *Akkermansia muciniphila* enhanced health span and conducted lifespan extension, highlighting how FMT could be adopted as a therapeutic intervention to improve healthspan and lifespan (184).

Finally, previous studies in human centenarians (aged > 100 years) demonstrated the close relationship between the gut microbiome and longevity (185, 186, 187). Remarkably, Sato et al. identified centenarian-specific gut microbiota capable of producing secondary bile acids, including lithocholic acid and isoallolithocholic acid. Interestingly, isoallolithocholic acid shows antimicrobial effects against Gram-positive microorganisms. The authors suggest that secondary bile acids may play a role in intestinal homeostasis by selective colonization mechanisms and regulate the host immune response (188).

Although there is a starting understanding of the factors that modulate the gut microbiota and the taxonomic and functional responses that the microbial community undergoes during host aging, there is a lack of knowledge on the evolutionary changes of the gut microbes in response to aging (7). Microbiome genetic variants can confer new adaptive host traits or enable age-related dysbiosis by enabling bacteria to escape host immune surveillance directly limiting host lifespan (80, 179). Therefore, one of the main objectives of my thesis was to explore the evolution of the microbial commensal in the gut during host aging.

1.2 Genetic variation in the gut microbiome

Next-generation sequencing technologies applied to studying the gut microbiome reveal that most microbial species harbor a large amount of genetic variation between and within-hosts over time (199). Genetic variation is the result of single-nucleotide variants (SNVs) (103), gene flow (159), short insertions and deletions (indels) (152), structural variants (SVs) (1), and gene copy number variants (CNVs) (200) in gut microbes (Figure 2).

Genetic diversity continuously increases in the gut microbiome owing to errors in DNA

replication, DNA damage, error in the DNA repair process, and recombination. It is estimated that the human gut microbiome can generate from 10^9 to 10^{12} novel mutations every day (4). The accumulation of novel mutations and increasing the genetic diversity within species is subjected to natural selection and genetic drift (201).

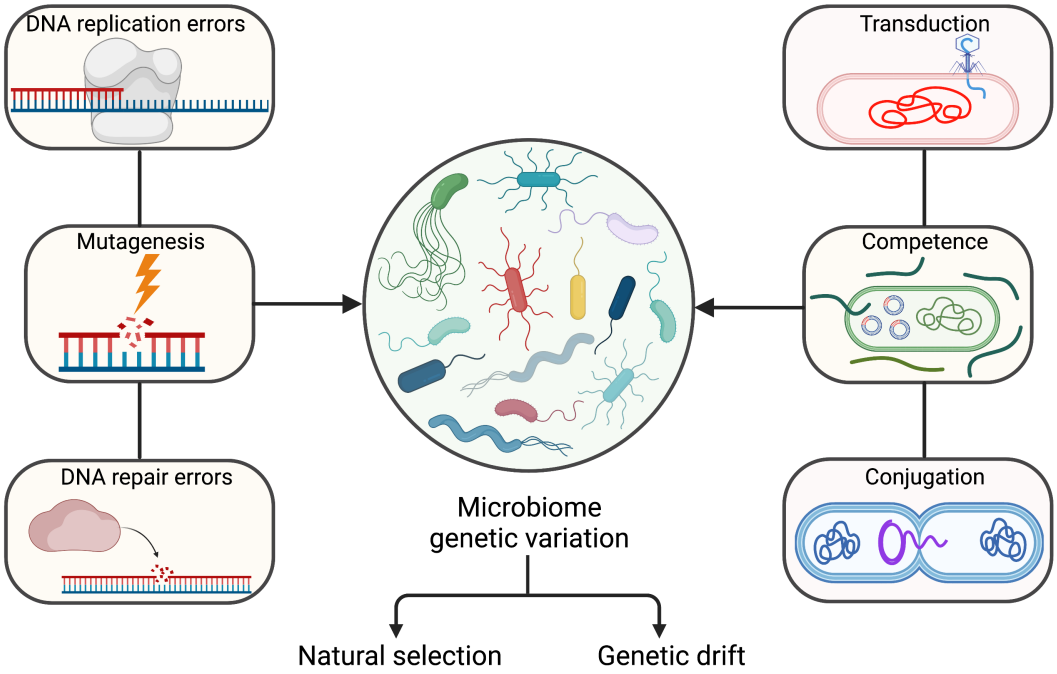


Figure 1.2: Factors affecting microbiome genetic variation. Factors affecting microbiome genetic variation. Mutations (DNA replication errors, mutagenesis, DNA repair errors) and horizontal gene transfer events (transduction, competence, conjugation) create genetic variation in the microbial population which then are subject to natural selection and genetic drift to main in the population (Rossum et al. 2020). Created with BioRender.com.

Notably, it is reported that accessory genes (genes not present in all strains of a bacterial species) accumulate more mutations than core genes (genes present in all strains of a bacterial species) (203). Accessory genes constitute a "genetic reservoir" that may have a critical role in bacterial adaptation (202). Remarkably, it is known that mutations within one bacterial lineage can be transmitted vertically to daughter cells or horizontally to close or distant lineages through recombination events (3, 202).

Recombination in prokaryotes, or horizontal gene transfer, is the process of transferring DNA to a distant bacterial recipient that allows the gain or loss of genetic information

(152). Recombination includes the transfer of genes, complete metabolic pathways, plasmids, and transposable elements (203). Bacteria can uptake new DNA material via conjugation (direct contact with another bacterial cell), transduction (transfer dependent of bacteriophages), competence (direct uptake of exogenous DNA) and membrane vesicles (201). Incorporating novel genetic material can generate novel phenotypes that improve the adaptation of the bacterial host to dynamic environments such as the intestinal host (152). A clear example of a recombination event and bacterial-host behavior changes came from the commensal bacterium *Escherichia coli*. A horizontal genetic exchange encoding Shiga toxin 2 and antibiotic-resistance factors generated a diarrhea outbreak in Germany in May 2011 (205).

De novo mutations and recombination events can facilitate host adaptation. However, these events could also generate detrimental effects on the bacterial population. Most de novo mutations have no or negative impact on the individual fitness, and incorporating novel genetic material can bring incompatibility with the genome and detrimental effects if the mobile gene has no function (209). Therefore, natural selection and genetic drift are the major determinants of the accumulation or elimination of within-species variation introduced by mutations or recombination (105, 201).

Genetic drift eliminates genetic variation randomly in the population, with the potential to spread neutral or deleterious mutations in the population, while natural selection maintains or removes novel genetic variants based on beneficial or detrimental fitness effects (201, 204, 205). Natural selection can modulate the microbial community and genetic diversity depending on the environmental context. Purifying selection purges deleterious variants in the population, while positively selected variants increase their frequency in the population (152, 201).

When a novel advantageous mutation emerges in a given gene and becomes beneficial for the microbial populations, it can rapidly increase in frequency and eventually fix in the population, resulting in a selective sweep (206). Selective sweeps can reduce intra-species genetic diversity, produce genetic dissimilarities between lineages, and eventually generate bacterial speciation (201). In contracts, multiple adaptive alleles can emerge and spread simultaneously, coexisting within the microbial population. This phenomenon

is known as soft sweeps, and it is described as the first step of commensal gut bacteria adaptation when encountering novel environmental conditions (75).

Two powerful strategies to detect genetic variants under positive, neutral, or purifying selection are focusing on the ratio of non-synonymous to synonymous fixations (dN/dS) and the ratio of non-synonymous to synonymous polymorphism (pN/pS). The dN/dS ratio is commonly used to characterize selective pressures between individual species, while pN/pS is focused on identifying selective constraints within-species level (103, 152). Low ratio values (dN/dS < 1 , pN/pS < 1) are indicative of purifying selection because non-synonymous mutations are restricted and eliminated in the population. Values of dN/dS and pN/pS >1 indicate that functional change is positively selected.

Therefore, same frequency of non-synonymous and synonymous variants are indicative of neutrality (dN/dS = 1, pN/pS = 1) (152, 199). Genome-wide estimation of dN/dS and pN/pS ratios in the human gut microbiome shows that purifying selection is the main constraint force that impacts the microbial community, indicating that novel genetic variants are generally detrimental in the intestinal environment (3,4,5).

1.3 Microbiome diversity affects host health and disease

While we are starting to understand how the gut microbial community and microbial diversity change during host life, little is known about how the microbiota adapt to environmental perturbations at the genetic level (5,7). Moreover, it is still unknown whether the bacterial evolution influences community composition and affects host-microbiota homeostasis, host health, and lifespan determination (7).

Understanding how microbial communities adapt to changing environmental conditions in the intestine could shed light on the mechanism of how host health and lifespan are determined by microbiota genomics (7, 80). For example, previous reports showed that genetic diversity in the gut microbiome could confer new traits to the microbial community and the host (80). Specifically, Hehemman et al. observed that the transfer of carbohydrate-active enzyme genes from marine Bacteroidetes to the gut microbe *Bac-*

teroides plebeius in Japanese individuals confer the ability to digest complex polysaccharides present in traditional nutrient sources (190). Kent et al. observed the acquisition of antimicrobial resistance genes in the gut microbiome during clinical antibiotic treatments, increasing the load of pathogenic bacteria in the community (191). Finally, Zimmermann et al. described a substantial difference in drug-metabolizing activities of the human gut bacteria in different individuals based on bacteria genetic composition and microbiome variability (192).

There is evidence that genetic diversity in the gut microbiota can significantly alter gut bacteria behavior and affect host health and lifespan. Sokurenko et al. observed that genetic variation in the FimH gene (an adhesin responsible for bacterial binding to host cells) can dramatically change *E. coli* behavior from commensal to virulent phenotype (193). Zou et al. conducted a high-throughput screening in *Caenorhabditis elegans* using an *E. coli* library and identified a single nucleotide polymorphism (SNP) in the gene encoding blc (an outer membrane lipoprotein involved in the storage or transport of lipid for maintenance of the membrane under stressful conditions) (194) that promotes epithelial barrier disruption, intestinal inflammation, and immune activation (195). Remarkably, Zou et al., report that gut bacteria with blc gene variant are enriched in patients with inflammatory bowel disease (195).

Moreover, Zhu et al. explored the evolutionary changes in intestinal species and identified a large number of SNPs in the genome of *Bacteroides vulgatus* in patients with Graves disease, implicating a connection between *B. vulgatus* diversity with the onset of the disease (196). Finally, Zeevi et al. analyzed metagenomic samples of the human gut microbiome and identified the high presence of microbial structural variants (SVs) in the human gut microbiome. The sight of SVs associated with CRISPR genes and antimicrobial metabolites suggests the potential contribution of these SVs in microbial adaptation in response to their environment. However, specific SVs were also associated with host risk factors. For example, an SV in *Anaerostipes hadrus* may produce butyrate, a short-chain fatty acid strongly associated with protection from multiple human diseases (197) and lower metabolic risk (1).

Nevertheless, Han et al. reported one of the most relevant direct associations between

bacterial genetic variation and host physiology. By designing a high-throughput screening of an *E. coli* collection of 3,983 non-essential single-knockout mutants, Han et al. discovered 29 microbial genetic factors that regulate *C. elegans* health and lifespan. These longevity-promoting bacterial mutants belong to different functional categories, including transcription and translation, metabolism and respiration, membrane and transport, and protease and chaperones (2).

In addition, the authors discovered that two bacterial mutants in genes that regulate the biosynthesis of colanic acid (lon and hns) overproduce colanic acid in bacterial culture medium, extend *C. elegans* lifespan and improve tissue maintenance by modulating mitochondrial unfolded protein response signaling pathway (2). Together, these results reveal the significance of microbial genetic variations in modulating and activating longevity regulatory mechanisms in the host, opening the opportunity to develop genetically modified pro-longevity probiotic therapies (2).

1.4 Within-host adaptive evolution of the microbiota

The gut microbiota is a complex and dynamic ecosystem that responds to environmental (e.g., diet, medication, species invasion) and within-host perturbation (e.g., mucin production, transcriptional changes, intestinal structures) (3, 4, 5, 199). In these harsh conditions, microbes must adapt to the transient environmental and within-host perturbations to persist in the community (4). It has been reported that the gene expression profiles of individual bacteria taxa and composition of the microbiota change within days or weeks under nutrient alteration. These ecological changes involved altering microbial species abundance and increasing the expression of enzymes capable of degrading carbohydrates present in the diet and outer membrane proteins efficient for nutrient transportation (207). While ecological responses have been extensively described, less is known about the adaptive mutation that arises within the populations that reside in the host gut critical for long-term persistence in the host (3,4,5).

The application of metagenomic sequencing of samples and whole-genome sequencing in bacterial isolates has enabled comprehensive pictures of the molecular processes involved

in within-host bacterial evolution (208).

Previous evidence of within-host evolution in pathogen bacteria demonstrated that different bacteria species accumulate within-host point mutations in a short period of time (months or years) (208). For example, longitudinal studies in the pathogen *Helicobacter pylori* reported the accumulation of 30 mutations per year per genome (209). A high number of point mutation accumulation, known as hypermutation, is a common adaptation mechanism for pathogen and commensal bacteria to new ecological niches (209, 210). These novel mutations are in control of natural selection, which favors the evolution of beneficial variants. In fact, new variants in gene encoding outer membrane proteins have shown signatures of adaptive evolution. The evolution of outer membrane protein in bacterial pathogen and commensal have been observed and could suggest that these proteins have evolved to evade the host immune system or phage predation (208, 4).

On the other hand, the acquisition of novel genetic material through HGT, such as plasmids, transposons, or phages, has played a vital role in pathogen adaptation to changing selective pressures within-host, been the acquisition of antibiotic resistance the most common transferred genetic pathway in the human gut microbiome (191, 211).

While the evolution of pathogenic bacteria and their effects on public health have been widely studied, the evolutionary dynamics and adaptation mechanism of commensal gut bacteria have been recently described (191).

Commensal gut microbes coevolved within their host for tens of thousands of years, suggesting that selection or purging novel variants in microbial populations would not be possible in short timescales (e.g., host lifespan) (152, 208). However, commensal microbes could adapt rapidly to novel environmental stress due to their large population size and short generation time, resulting in a large number of daily de novo mutations in the whole microbiome ($2x10^9$ - $6x10^{12}$) (4, 152). In fact, recent metagenomic studies demonstrated that microbial populations could evolve on short time scales within the host (within weeks, months) (3, 4, 5, 6, 75, 207, 210).

Two different perspectives have been conducted to study the evolutionary dynamics of the microbial commensals; 1) Evolution of a single commensal bacteria species in a gnoto-

biotic host and 2) Evolutionary dynamics of multiple bacterial species in the human gut microbiome.

Early studies in mice described the accumulation of beneficial mutations in the bacteria species E. coli during host colonization (75). Barroso-Batista et al. describe that the first adaptive steps to the mice's gut environment are observed within the first 24 days after colonization. Moreover, independent evolving populations showed parallel adaptive paths in genes involved in galactitol and sorbitol metabolism, transmembrane transporters, and DNA transcription factors. These adaptive mechanisms may respond to the host metabolism and microbe-microbe competition (75). Interestingly, the authors also describe the presence and dynamics of multiple beneficial mutations in the population that change their frequency over time. The observed dynamics were consistent with soft sweeps, meaning that beneficial mutations with the same phenotype compete between them without reaching fixation (clonal interference) (75).

Using the same model strategy, Ramiro et al. follow the evolutionary adaptation steps over long-term colonization. Here, the authors described the development of a hypermutation phenotype caused by mutations in DNA polymerase III proofreading and catalytic subunits that cause the accumulation of a large number of deleterious mutations. However, beneficial mutations involved carbon metabolism and stress responses pathways were also detected in independent populations. Remarkably, Ramiro et al. propose that partial sweeps of beneficial mutations combined with hitchhiking (allele frequency changes not by selection but because of the proximity to a selected variant) of weak-effect deleterious mutations maintained within-species genetic diversity in the microbiome (210).

However, these studies describe the evolutionary dynamics of commensal bacteria within a single gut microbial community. The gut microbiome represents a complex and diverse community with microbial interactions that may impose novel selective pressures on bacterial species, driving different adaptation pathways (4). Therefore, several studies focused on understanding the evolutionary dynamics and adaptive evolution in complex microbiome environments like the human gut microbiome using multiple sequencing strategies (3,4,5).

Garud et al. implemented metagenomic samples from the human gut microbiome to

quantify the evolutionary dynamics within- and between hosts. The authors introduced a methodology that confidently assigns major alleles to the dominant lineage in the population. This methodology leverages the quantification of the evolutionary dynamics that operate within and between hosts (5). Importantly, this approach detects within-host adaptation on 6-month timescales where de novo mutations and recombination sweeps are involved in the adaptation process. However, genome-wide estimation of dN/dS suggests that purifying selection governs the microbiome evolution over long-timescales (5).

Cheng et al. leverage the same approach and monitor the evolutionary changes that the human gut microbiome exerts during the early days of life. Interestingly, evolutionary modifications and ecological replacements increased in the infant's gut compared to rates in adults. These results align with the particular dynamic ecological successions that occur during early day life before the microbial community becomes more stable with host maturation. Notably, Cheng et al. identify genes that undergo parallel adaptive evolution belonging to the susC/susD family, which are outer membrane proteins implicated in processing complex glycans and degradation of breast-milk oligosaccharides (3).

On the other hand, Zhao et al. combine culture-based population genomics and metagenomics to investigate the within-host adaptive evolutionary process in the prevalent commensal bacteria *Bacteroides fragilis*. This analysis reveals that *Bacteroides fragilis* show multiple de novo mutations with signals of within-host adaptation in 16 genes. Interestingly, six of these genes under adaptive selection are members of the susC/susD family outer membrane protein, providing solid evidence for these genes to be involved in commensal gut colonization and long-term persistence within the host (4).

In contrast to detecting SNVs and linked adaptive signatures on novel genetic variants, Yaffe et al. implemented high-throughput chromosome conformation capture (Hi-C) to capture the role of horizontal gene transfer for bacterial evolution in the gut environment. Surprisingly, the authors found that bacterial strains can persist within the host over ten years and evolve in a combination of adaptive selection of novel genetic variants and dynamically exchanged accessory genes (159).

Finally, Roodgar et al. implemented linked-read metagenomic sequencing to track longitudinal trajectories of single nucleotide variants of a single individual under a period of

health, disease, antibiotic treatment, and recovery. This approach showed that antimicrobial therapy rapidly shifts genetic diversity in the microbial population but no change in species abundance, demonstrating the role of microbial genetic diversity in the ecological resilience of the microbial population to encounter environmental perturbations (58).

Together, these previous works describe the molecular mechanism of gut microbial adaptation to encounter rapid shifts in environmental conditions. However, how the gut microbiome evolves during host aging is still unexplored. As previously described, the intestinal environment undergoes significant compositional changes during host aging that can lead to microbial evolution (7). In recent work using aged mice as a model, Barreto et al. described that the pace and targets of adaptation in commensal bacteria are different from young individuals, selecting genetic variants associated with stress responses and virulence factors (76). However, a comprehensive overview of within-host adaptive evolution during aging is still lacking, and that is the main approach that I propose to address in the present work.

1.5 Characterization of the gut microbial community of the aging model *Nothobranchius furzeri*

The turquoise killifish *Nothobranchius furzeri* is a naturally short-lived vertebrate with an exceptionally short life span of 4 to 6 months which recently emerged as a novel experimental model for aging research (212, 213). *Nothobranchius furzeri* is used as a model system to understand the functional relationship between gut microbiota and host lifespan determination (8, 214). The composition of the *Nothobranchius furzeri* microbiota has a complex taxonomic diversity that undergoes dramatic compositional and functional changes during host aging (8). Moreover, an experimental study using this aging model exposed the gut microbiota's role in improving health and extending the host's lifespan (8). Although the killifish's gut microbiota is being analyzed using 16S rRNA gene amplicons, relatively little is known about the functional characterization of bacteria taxa present in the turquoise killifish microbiota.

Next-generation sequencing and long-read metagenomics allow a detailed taxonomic and functional characterization of microbial species present in a complex environmental sample (62). Developed bioinformatics pipelines can recover complete closed bacterial genomes that allow creating a reference genome database to identify critical microbial genes and pathways relevant to host colonization and life span determination (8, 218).

However, to fully understand the role of microbial gene function and microbial isolates in determining the host lifespan, it is required to create a microbial isolate collection. Therefore, implementing high-throughput culture conditions (culturomics) to isolate and identify new genes and genetic pathways present in the host's microbiome is required (13). Implementing culturomics in the turquoise killifish gut will leverage the power to modify bacterial isolates genetically, re-colonize the host with a specific genetic variant, and monitor the effects of novel genetic variants in host physiology and aging (127).

Chapter 2

Thesis goal

Section I : Evolution of commensal gut microbes during host aging

Recent studies in the human microbiome described the evolutionary changes encountered during infant maturation [3] and adulthood in humans [4,5] and the rapid adaptation of bacterial species to the aged gut environment in mice [6]. However, how the gut microbiome evolves across host aging within the same individual is still unclear. Undercover microbial evolutionary changes during host aging are crucial. The appearance of microbial genetic variations could modulate host health and longevity [1,2] or contribute to the onset of the imbalance of the microbial community homeostasis, increase the individual frailty, and reduce host life expectancy [7]. The primary goal of this thesis is to characterize the genetic changes that appear in the mouse microbiome during host aging, identify the genes that are the target of selection, and the effects of the novel genetic changes on bacterial fitness and host physiology.

Section II: Characterization of the gut microbial community of the aging model *Nothobranchius furzeri* using short and long read metagenomics

The Turquoise killifish - Nothobranchius furzeri - is a short-lived emerging experimental model organism to study the link between host aging and gut microbiota [8] and test candidate interventions for lifespan modulation [9]. The gut microbiota of the turquoise killifish is a taxonomic diverse microbial community comparable to mice and humans. Detailed characterization and cultivation of the gut microbial community will significantly expand the knowledge of the genetic diversity and microbial repertoire in this vertebrate model system. Therefore, the second aim of this thesis is to isolate and characterize the gut microbes from the turquoise killifish implementing a high throughput culturing approach in combination with short- and long-read metagenomics.

Chapter 3

Materials and Methods

Samples

The stool samples used in the present work are part of a large mice longitudinal study. I focused the analysis on stool samples collected from 3 female hybrid mice (SG0476 / SG0696 / SG0503) housing under pathogen-free conditions (generated by mating C57BL/6N males with C3H/HeOuj females) at 12 (SG0476-1 / SG0696-1 / SG0503-1), 20 (SG0476-5 / SG0696-5 / SG0503-5), and 26 (SG0476-8 / SG0696-8 / SG0503-8) months old. All mice were fed with commercial rodent chow ad libitum.Stool samples were immediately frozen in liquid nitrogen and stored at -80ºC before High Molecular DNA extraction. Mouse SG0476 died at the age of 988 days (33 months old), mouse SG0696 at the age of 962 days (32 months old), and mouse 0503 at the age of 894 days (30 months old).

For the characterization of the turquoise killifish, six male turquoise killifish *Nothobranchius furzeri* (GRZ-AD strain) were raised and single-housing under standard laboratory husbandry protocol for nine weeks until sample collection [10]. Fish were housed under 12 h of day/night cycle at water temperature set to 28ºC. The fish diet consisted of brine shrimp nauplii during the first week of life and changed to blood worms twice a day during weekdays and once a day during weekends [8, 10].

Before sample collection, fish were transferred to a clean 2.8 L tank containing autoclaved tank water and monitored until stool was collected (20 mg). Stool samples were

immediately transferred to 900 µl pre-reduced 1x PBS buffer and placed in low oxygen conditions (<10 ppm) for bacterial cultivation [12] or transferred to 200 µl of DNA/RNA shield (Zymo Research) for High Molecular DNA extraction [11].

Culturomics

Culturomics is a high-throughput culturing strategy implemented to isolate and identify uncharacterized bacteria species from highly diverse microbial environments [12, 13]. Our culturomics study concerned six stool samples and 25 different culturing conditions (Table 7). I selected those culturing conditions for growing fastidious aerobic and anaerobic bacteria. Bacterial identification was performed by 16S rRNA sequencing. Anaerobic conditions were created using the BD GasPack EZ container system and hypoxia conditions under a Coy NeoLab anaerobic chamber.

Diluted stool samples in reduced 1x PBS (25 µl) were inoculated in the different agar plate media and incubated under distinct culturing conditions. Visualizing inspection of the bacterial colony morphology was performed three days after inoculation. Uniquely distinguished colonies were isolated in a new agar plate and cultured for three days. Bacterial identification was performed by 16S rRNA gene sequencing using the primers Bakt-341F (5'-CCTACGGGNGGCWGCAG-3') and Bakt-1061R (5'-CRRCACGAGCTGACGAC-3') [14,15]. Uniquely bacterial species were inoculated into a culture flask for three days, followed by High Molecular Weight DNA extraction for Nanopore sequencing and Shotgun metagenomic sequencing.

High Molecular Weight (HMW) DNA extraction of stool samples

The extraction of high molecular weight DNA allowed the performance of long-read sequencing to increase the quality and contiguity of complete bacterial genomes [16]. To extract High Molecular Weight (HMW) DNA, I utilized the previously reported three peaks fecal High Molecular Weight DNA extraction method [11] using 30 mg of the stool sample. HMW DNA was eluted in 50 µl of low EDTA-TE buffer (1x) (Quality Biological) and stored at 4ºC. HMW DNA concentration, purity, and fragment size distribution were performed using Qubit fluorometer (Thermo Fisher Scientific), Nanodrop (Thermo Fisher Scientific), and TapeStation 2200 (Agilent Technologies).

The eluted HMW DNA from the mice longitudinal samples were implemented for long-read metagenomics (Nanopore Technology), linked-read metagenomic (Tell-Seq Technology), and short-read shotgun metagenomic sequencing (Illumina / Nextera). DNA samples from the turquoise killifish microbiota isolates were sequenced using short- (Illumina / Tn5) and long-read (Nanopore) sequencing.

High Molecular Weight DNA extraction from bacterial cultures

Bacterial isolates were inoculated in 50 mL of LB media at 28ºC overnight. HMW DNA was extracted from bacterial culture using MagAttract HMW DNA Kit (Qiagen) following the manufacturer's instructions.

Tn5 Library preparation of isolated bacteria from the killifish gut

Tagmentation-based library construction was followed using an in-house generated Tn5 transposase [17, 18] to create high-quality sequencing libraries from the isolated bacteria species. For the assembly of the transposome in solution, 10 µl of preannealed Tn5MEDS-A (100 µM) and 10 µl of preannealedTn5MEDS-B (100 µM) oligonucleotides were mixed with 20 µl of 2× Tn5 dialysis buffer (100 HEPES-KOH at pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 2 mM DTT, 0.2% Triton X-100, 20% glycerol) and 20 µl of Tn5 enzyme (3.4 µg/µl) and incubated for 60 min at room temperature. Tagmentation was performed in 2µl of 5x TAPS Buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH 8.5) at 25°C), 6 µl of Nuclease-free water, 1 µl of pre-charged Tn5 and 1 µl of MNW DNA. Reactions were incubated at 55ºC for 5 minutes, stopped with 5µl of 0.2% SDS, and incubated at 70ºC for 10 minutes. Clean-up was performed using 15 µl of SeraSure beads at Room temperature and washed with 150 µl of Ethanol 80% twice. Samples were eluted in 6 µl of nuclease-free water.

PCR amplification was performed using 2x KAPA HiFi HotStart ReadyMix PCR Kit using primer i5 (N5xx) and i7 (N7xx) with unique barcodes for sample demultiplexing. Library amplification underwent 19 rounds of PCR cycles described as follow: 3 minutes of initial activation (72 ºC), 30 seconds for initial DNA denaturation (95ºC), 19 cycles of 10 seconds denaturation (95ºC), 30 seconds annealing (55ºC), 2:40 minutes extension (72ºC), and a final extension of 5 minutes (72ºC). The amplified library underwent a final

clean-up step using 0.5 vol of AMPureXP beads (Beckmann), eluted in 20 µl of water and stored at -20ºC before sequencing. Libraries were sequenced with 2 x 150 bp paired-end using the Illumina HiSeq4000 platform.

Illumina library DNA prep of the longitudinal mouse samples

HMW DNA samples from the longitudinal mouse study were prepared using the Illumina DNA Prep Kit following the manufacture's instructions. The DNA input of every sample was 150 ng to avoid the quantification and normalization steps during library pooling. Individual libraries and library pool were quantified using Qubit. The index 1 (i7) and index 2 (i5) adapters were included using the Nextera^{TM} DNA CD Indexes Kit (Table S2). The pooled library was sequenced using the Illumina HiSeqX platform 2x150 bp paired-end.

Nanopore long-reads sequencing

Extracted HMW DNA from the longitudinal stool samples was sequenced using the Ligation Sequencing Kit (SQK-LSK109) and three MinION flow cells (R10.3) according to the manufacturer's protocols. The samples were pooled by mouse (1.5 µg DNA), and one pool host sample per flow cell was sequenced. The sequencing run was scheduled for 48 hours and allowed to run until fewer than ten pores remained functional. The sequencing run yielded raw sequencing data of 8.4 Gb per sample.

HMW DNA extracted from fish stool samples was pooled (400ng total DNA input) and sequenced using the Ligation Sequencing Kit (SQK-RAD004) and one GridION flow cell (R9.4.1) according to the manufacturer's protocols. The final raw sequencing yield of the sequencing run was 3.4 Gb. The long-read sequencing of the isolated bacteria cultures was sequenced using the Rapid Barcoded Kit (SQK-RBK004) and one MinION flow cell (R9.4.1) according to the manufacturer's protocols. Each bacterial genome was barcoded (RB0X) for pooling the final libraries and demultiplexing using bioinformatic tools (Porechop v.0.2.4, https://github.com/rrwick/porechop).

Transposase enzyme-linked long-read sequencing (Tell-Seq^{TM})

Transposase enzyme-linked long-read sequencing (Universal Sequencing) is a next- gen-

eration sequencing technology that generates barcoded linked reads from complex metagenomic samples [19]. Tell-Seq technology enables high-accuracy short-read sequencing to generate over 100 kb of long-range sequencing information [19,20]. The Tell-Seq WGS Library Prep Kit was implemented to generate the barcoded linked reads libraries from the longitudinal mouse samples. Sample indexes were implemented for pooling and sequenced following the manufacturer's protocol (Table S3). Tell-Seq library was sequenced following the manufacturer's specific sequencing requirements on the Illumina NextSeq High Output platform with 2x150 paired-end reads.

Genetic modification of commensal gut bacteria from the turquoise killifish using fluorescent tagging vectors

The genetic modification of the isolated gut bacteria species *Pseudomonas lundensis* was performed using the well-established Tn7 chromosomal integration system [21]. The Tn7 system is an efficient genetic manipulation tool for diverse bacterial isolates and was recently implemented for the genetic transformation of proteobacterial isolates native to zebrafish's intestinal microbiota [22]. The protocol S1 [22] was followed for engineering chromosomal insertions using the Tn7 system in *Pseudomonas lundensis*. Briefly, helper (carrying pTNS2 Tn7 transposition tool - Addgene Plasmid No.64968) and donor (carrying pUC18-mini-Tn7T-Gm-eyfp fluorescent tagging vector - Addgene Plasmid No.65031) *E. coli* SM10 strains and target strain (*Pseudomonas lundensis*) were grown overnight in 5 ml of LB+ampicillin media at 37ºC 180 rpm (OD600 > 1) for donor and helper strains and LB media at 30ºC for target strain. Subculture donor and helper strain 1:50 into 5 ml of fresh LB+ampicillin media at 37ºC with shaking and target strain into 5 ml of fresh LB media at 30ºC with shaking until bacterial cultures reach OD600 of 0.4 - 0.6.

The subcultured donor, helper, and target strains were mixed 1:1:1 (500µl each), centrifuged at 7,000xg for 2 minutes, and washed with 1 ml of wash solution (0.7% NaCl sterile solution). The bacterial pellet was resuspended in 25 µl of washing solution. The mating reaction was incubated in a TSA plate onto the surface of a sterile 25mm-wide 0.45µm filter disc (Millipore) at 30ºC overnight. The filter disc containing the mating reaction was transferred to a 50 ml conical tube containing 1 ml of wash solution, and bacteria were

removed from the filter by pipetting. Bacterial mix was plated on TSA+gentamicin and incubated overnight at 30ºC. Plate screening and colony isolation of putative modified clones were performed using a fluorescent stereomicroscope. Putative modified clones were streaked on TSA plates and incubated overnight at 30ºC.

Target strain colonies that exhibit stable fluorescence were confirmed with colony PCR (FW glmS 5'-TGGAGAGGGTGAAGGTGATG-3' / Rv GmR-Tn7R 5'- CACAGCATAA CTGGACTGATTTC-3') and culture in 5ml of LB+gentamicin media with shaking at 30ºC. Confirmed target strains were prepared for glycerol stock and stored at -80ºC.

Preparation of Calcium competent and heat shock transformation of E. coli cells

Generation of home-made competent cells of the *E. coli* strain SM10 and heat shock transformation were performed following the standard and highly efficient calcium chloride protocol [23] with a slight modification in the CACl2 solution (0.1M CaCl2, 15% glycerol).

Nanopore reads data processing

Generated fast5 files from Nanopore sequencing run were based-called using Guppy v3.6.1 (Oxford Nanopore Technologies) using the pipeline guppy_on_slum (https://github.com/colindaven/guppy_on_slurm) retaining only sequences with a minimum Q-score of 7 and specifying the flow cell and library kit parameters. Output DNA sequence reads were saved as fastq files.

DNA sequencing data from the gut bacterial isolates were demultiplexed and adapters trimmed using Porechop v 0.2.4 (https://github.com/rrwick/porechop) under the parameters –barcode_threshold 75 –barcode_diff 5 –discard_middle. Adapters from DNA metagenomic sequencing reads from stool samples (mice and fish) were trimmed using the –discard_middle parameter.

Read statistics from all samples, including the number of reads, read lengths, and quality scores were assessed using NanoPlot v1.0.0 (https://github.com/wdecoster/NanoPlot) using the parameters –loglength –format pdf –N50.

Short-read metagenomic data processing

Illumina fastq sequencing data generated from shotgun metagenomics of the mice and fish stools samples and isolated gut bacteria genomes were pre-processed using KneadData v0.10.0 (https://github.com/biobakery/kneaddata) to remove host reads, perform quality filtering, and quality check analysis. Mouse (C57BL/6NL), turquoise killifish (GRZ), and bloodworm larvae (Chironomus spp.) genomes were used for host de-contamination in the KneadData pipeline.

Transposase Enzyme Linked Long-read sequencing (Tell-Seq^{TM}) data processing

Tell-Seq^{TM} raw fastq files were processed using the Tell-Read pipeline (Universal Sequencing), part of the Tellysis software explicitly developed for Data Analysis of paired-end barcoded linked reads generated by Tell-Seq^{TM} technology. Tell-Read generates linked-read fastq data and a quality control report of each longitudinal mouse sample. Tell-Read pipeline ran the analysis without a reference genome parameter (-g "NONE"), and QC reports were summarized in HTML files.

Metagenomic assembly of fish stool samples using long-reads

Metagenomic assemblies were constructed with metaflye assembler v2.8.3 (https://github.com/fenderglass/Flye) [26] with an estimated metagenome size of 250 Megabases and 5x polishing iterations to correct structural errors. The novo assembly graphs were visualized using Bandage v0.8.1 (https://rrwick.github.io/Bandage/) [27].

Hybrid assembly of Bacterial isolates from the killifish gut

A hybrid assembly was performed using both Illumina and Nanopore pre-processed data for each bacterial isolate using the hybrid assembly pipeline Unicycler v0.4.9 (https://github.com/rrwick/Unicycler) [24] with a final polishing of the assembly with Illumina reads using Pilon v1.23 (https://git hub.com/broadinstitute/pilon/wiki) [25].

Hybrid assembly of metagenomic longitudinal mouse samples

Hybrid metagenomic assembly of the mouse longitudinal study was performed using both Illumina and Nanopore pre-processed read data following the OPERA-MS v0.9.0 assembly workflow (https://github.com/CSB5/OPERA-MS)[28] using default parameter and OPERA-MS genome database for genome clustering. Using the Illumina reads, hybrid assemblies were polished using Pilon v1.23 (https://github.com/broadinstitute/pilon/wiki) [25]. Assembly statistics and taxonomic information were reported for each genome.

Tell-link assembly

De Novo Assembly of metagenomes using linked-read data from Tell-Read was implemented using the new de novo assembler TuringAssembler [19], which uses linked-read sequences to perform a more efficient local assembly and scaffolding to generate high-quality metagenomes. TuringAssembler is part of the Tell-Link pipeline included in the Tellysis software (Universal Sequencing). Tell-Link assembly pipeline was performed under the parameters -d metagenomics -k 100 -lc 75 -j 40. Assembly reports were generated using Quast v5.0.0.

Short-read data assembly

De Novo metagenome assembly using quality-controlled short-reads data were performed using SPAdes v3.11.1 [30] using the pipeline for metagenomic data metaspades.py (https://github.com/ablab/spadesmeta). Assembled metagenomes were quality controlled for a minimum length of 2Kb and further polished using Pilon v1.23 (https://github.com/broadinstitute/pilon/wiki)[25].

Binning, bin-refinement, and reassemblies of assemblies using MetaWrap

For metagenomic binning of the assembled metagenomes, the metagenomic wrapper suite MetaWRAP v1.3.2 (https://github.com/bxlab/metaWRAP) was implemented. MetaWRAP extracts high-quality draft genomes (bins) from metagenomic data by combining three different binning methods: MetaBAT v2.12.1 [32], Concoct v1.0.0 [33], and Maxbin2 v2.26 [34]. All binning methods use tetranucleotide and abundance frequency information for extracting draft genomes. Binning was filtered for a minimum contig length of 2Kb. High-quality bins with >=50% completeness and <=10% contamination were

retained using the bin_refinement module. Retained bins were reassembled using the reassemble_bins module of MetaWRAP to improve bin quality. High-quality bins statistics were plotted using a custom R script.

Bin dereplication

Bin dereplication eliminates redundancy in a genome dataset based on sequence similarity [35]. Dereplication at 95% ANI (Average nucleotide identity) is equivalent to removing redundancy at the species level [36]. High-quality bins were dereplicated at 95% ANI using dRep v2.6.2 [37]. Primary clustering dendrogram and clustering scores were plotted.

Taxonomic classification using DTDB-TK

Taxonomic classification of each high-quality bin was performed with the standardized taxonomic pipeline GTDB-TK v1.5.1 (https://ecogenomics.github.io/GTDBTk/index.html) [38] under the prokaryote genome database R202 [39]. Classify_wf workflow (identify, align and classify) was used with default parameters. The taxonomic classification of each high-quality bin was contained in summary.tsv output file.

Taxonomic classification of fish stool metagenomics

Using Kraken2 v2.1.2 [40], taxonomic classification was performed using the PlusPF genome collection, including archaea, bacteria, viral, plasmid, human, UniVec_Core, protozoa, and fungi genomes (https://benlangmead.github.io/aws-indexes/k2). For Visualization of the Kraken2 output, created taxonomic reports were transformed into Metaphlan-compatible format using the script kraken2mpa.py from the KrakenTools scripts suite (https://github.com/jenniferlu717/KrakenTools). export2graphlan.py conversion software (https://github.com/SegataLab/export2graphlan) and graphlan.py visualization tools (https ://huttenhower.sph.harvard.edu/graphlan/)[41] were used to generate a circular representation of the taxonomic and phylogenetic tree of the most abundant species present in the turquoise killifish microbiome.

Functional profiling of the stool metagenome of the fish

High-quality MAGs reconstructed from short-read metagenomics were annotated using

the rapid prokaryotic genome annotation tool Prokka v1.14.6 (https://github.com/tseemann/prokka) [42] with the –metagenome option to improve gene prediction. Genes annotated in Prokka were assigned to the Kyoto Encyclopedia of Genes and Genomes Orthologs (KEGG Orthologs) using the online tool GhostKOALA (https://www.kegg.jp/ghostkoala/) [43].

Within-species genetic diversity analysis of the mice longitudinal study

Gene prediction of the high-quality dereplicated bacterial genomes was performed using Prodigal v2.6.3 [44] (-p meta), and all predicted gene files were merged into a single fasta file. All genomes files were merged into a single fasta file, and a Bowtie2 index was created using the bowtie2-build command [45]. Pre-processed metagenomic reads were mapped to the merged metagenome index using Bowtie2 with the –very-sensitive parameter and converting the mapping output files into bam format. All the metagenomic reads were mapped against its corresponding host reference metagenomic index. A scaffold-to-bin file per host was created using the parse_stb.py script of the dRep v2.6.2 software.

Instrain v1.5.4 (https://instrain.readthedocs.io/en/latest/index.html) [36] was used to analyze within-sample genetic diversity. inStrain profile command was run on the mapping output files using the scaffold-to-bin and concatenated gene files. inStrain profile was run within the most abundant bacterial species in the metagenomic samples (minimum mean genome coverage of 20x). inStrain parameter aimed to avoid mismapping of reads were using a minimum coverage to call a variant of 6x (-c 6), minimum SNP frequency to confirm an SNV of 5% (-f 0.05), a minimum percent identity of read pairs to consensus of 99% (-l 0.99), a minimum mapQ score of 20 (–min_mapq 20), skipping the mm level analysis (–skip_mm_profiling). All non-paired reads were filtered out.

inStrain classified SNVs as non-synonymous, synonymous, and intergenic based on gene file. Bacterial replication rate metric (iRep), allele frequency of each SNV, nucleotide diversity (π), percentage of genome coverage by read data (breadth), and selection estimators dN/dS and pN/pS were computed per genome by inStrain.

Identification of single-colonized bacteria species

Identifying evolutionary changes using metagenomic data represents a challenging approach in populations with multiple colonizing strains [5]. The distribution of SNV allele frequencies provides information about the lineage structure of the bacterial population. A single-colonized population is expected to generate a major allele frequency distribution dominated by a single peak at a higher frequency (>80%) [46]. In contrast, the distribution of major allele frequencies with one or more peaks at intermediate frequencies suggests a mixture of multiple distinct lineages [5,46].

To exclude mapping error cases, sites with coverage (C) 0.3 lower than the mean genome coverage (G) (0.3C < G) or three times higher (3C > G) were excluded. SNV allele frequencies were plotted and examined using a custom R script [5].

Major Allele frequency distributions dominated by a single peak at higher frequency (>80%) and a low number of SNV at lower frequency were selected as single-colonized populations.

To track evolutionary dynamics of single-colonized populations over host aging, the same bacterial population needs to remain in the host for at least two-time points. To corroborate that the same population remains in the host over time, popANI value [36], which is a novel calculation of Average Nucleotide Identity (ANI) considering major and minor alleles, was calculated using the inStrain compare command.

inStrain compare command was used to compare young, adult, and old pre-processed metagenomic samples using –database_mode, a minimum coverage to call a variant of 6x (-c 6) and minimum SNP frequency to confirm an SNV of 5% (-f 0.05). popANI value >0.99999 represents a highly stringent definition for identical strains for metagenomic analysis [36].

Bacterial populations that passed the popANI threshold and showed a major allele frequency distribution dominated by a single peak of over 80% with a low number of SNVs at lower frequencies were selected for population genetic analysis.

Detection of SNV changes in single-colonized bacteria species

SNV changes were defined as SNV were allele frequency change from <=0.2 to >=0.8

between two-time points [3]. SNV tracking during host aging and identification of SNV changes were computed using a custom R script. Genes that acquired SNV changes were identified using blastx, and dN/dS and pN/pS values were calculated as possible signatures of selection [3].

Chapter 4

Results

Section I: Detecting microbial evolution in the gut microbiome during host aging

To fully characterize and identify the evolutionary mechanism of the gut microbiome during host aging, I selected longitudinal stool samples of three mice that are part of a mice cohort involved in a longitudinal collaborative study conducted at the Max Planck Institute for Biology of Ageing in Cologne. In collaboration with the Partridge's Laboratory, I selected individual mouse samples based on the criteria that samples were collected at the same host's age and the time distance between samples of at least six months.

The mice ID are SG0476, SG0696, and SG0503 for host 1, host2, and host3, respectively. The age of the host at the time that the sample was collected is 12 months, 20 months, and 26 months old. I considered the sample collected at 12 months of age as the young microbiome sample, 20 months old as the adult microbiome sample, and 26 months old as the old microbiome sample [48].I weight each stool sample using a cold metal surface (-20ºC) and sterile tweezers (Table 4.1).

High Molecular Weight DNA extraction

Based on the sample's weight, I performed a test experiment to optimize the input sample

weight to get a high quantity and quality of HMW-DNA of each sample. I followed the well-established Three peaks fecal DNA extraction method for long-read sequencing [11] using different input weights of mouse stool (Table 4.2).

Sample ID	Age of the host	Sample weight
SG0476-1	12 months	33mg
SG0476-5	20 months	65mg
SG0476-8	26 months	50mg
SG0696-1	12 months	31mg
SG0696-5	20 months	25mg
SG0696-8	26 months	29mg
SG0503-1	12 months	22mg
SG0503-5	20 months	80mg
SG0503-8	26 months	29mg

Table 4.1: Sample ID and weight of each stool sample

I quantified the extracted HMW-DNA concentration using the Qubit dsDNA BR assay kit (Thermo Fisher - Q32850) and the Qubit Fluorometer using the manufacturer guidelines. To assess the purity of the extracted HMW-DNA, I measured the absorbance of the sample at 230 nm, 260 nm, and 280 nm using the Nanodrop 2000c spectrophotometer. Nucleic acids, organic compounds, and protein have maximal absorbance at 260 nm, 230 nm, and 280 nm, respectively. A 260/280 ratio of 1.8 and 260/230 ratio range of 2.0 - 2.2 is considered a high-quality DNA sample. A low 260/280 and 260/230 ratio may result from the presence of contaminants, proteins, or unwanted organic compounds [48]. I used a Genomic DNA ScreenTape (Agilent Technologies) assay to observe the sample size distribution and integrity of each HMW-DNA test extraction (Figure 4.1). DNA Integrity is measured using the algorithm DIN (DNA Integrity Number - Agilent Technologies), which provides a numerical assessment of HMW-DNA integrity scaled from 1 to 10. A high value of DIN indicates intact HMW-DNA, while a low value of DIN suggests a highly degraded DNA sample [49].

Sample input of 24 mg and 35 mg showed the highest values of DNA concentration (46.6 ng/µl and 100 ng/µl, respectively), optimal ratio value for DNA purity (260/280 ratio of

Sample weight	Concentration (ng/µl)	260/280	260/230	DNA Integrity Number
11mg	5.3	2.04	0.91	7.4
24mg	46.6	1.94	2.02	7.1
35mg	100	1.89	1.95	6.2
40mg	80	1.88	1.78	4.0

Table 4.2: Quality, quantity and integrity assay of test experiment to optimize HMW-DNA extraction from stool samples.

1.94 and 1.89; 260/230 ratio of 2.02 and 1.95, respectively), and an integrity value of 7.1 and 6.2, respectively. Sample input of 11 mg showed a low concentration of HMW-DNA and a low 260/230 ratio indicating contamination for inorganic compounds. Finally, sample input of 40 mg showed a high DNA concentration and optimal purity ratios with the drawback of high degraded DNA (Table 4.2). Based on these results, I selected 24 mg and 35 mg as sample input of mouse stool for HMW-DNA extraction.

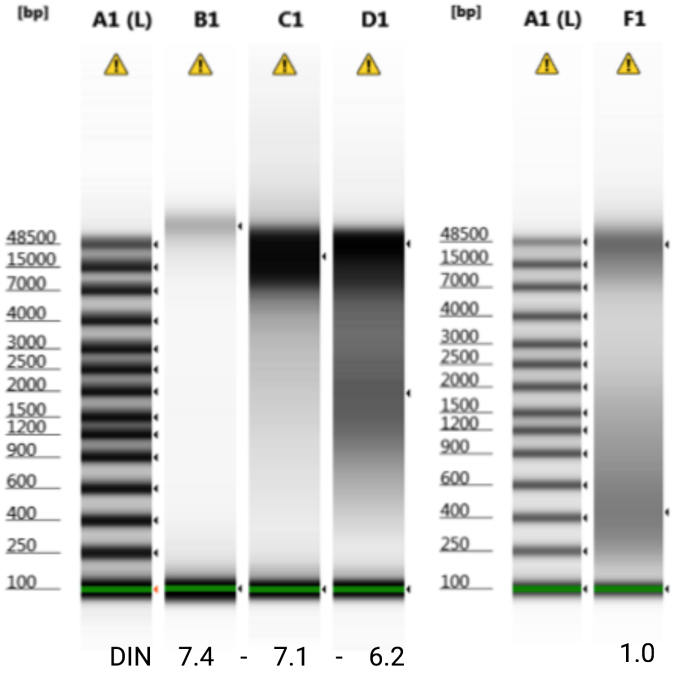


Figure 4.1: HMW-DNA extraction test with different stool sample weight. A1 - Genomic DNA Ladder; B1 - 11 mg; C1 - 24 mg; D1 - 35mg; F1 - 40 mg

I resized the DNA stools samples to a range of 24 to 35 mg using a frozen scalpel, sterile tweezers, and metal surface (-20ºC). The final nine stool samples ranged from 22 to 37 mg (Table 3). I followed the same HMW-DNA extraction protocol for the nine selected

longitudinal samples and quantified DNA concentration, purity, and integrity using the Qubit Spectrophotometer, Nanodrop 2000c (Table 4.3).

Sample ID	Weight	DNA concentration	260/280	260/230	DNA Integrity Number
SG0476-1	33mg	97.8 ng/µl	1.85	1.3	7.1
SG0476-5	37mg	149 ng/µl	1.86	1.5	6.4
SG0476-8	33mg	90.3 ng/µl	1.85	1.8	7.5
SG0696-1	31mg	161.6 ng/µl	1.86	1.85	6.4
SG0696-5	25mg	40.1 ng/µl	1.85	1.62	7.7
SG0696-8	29mg	92 ng/µl	1.84	1.61	7.3
SG0503-1	22mg	37.7 ng/µl	1.9	1.74	6.6
SG0503-5	33mg	65.4 ng/µl	1.89	2.18	5.6
SG0503-8	29mg	61.7 ng/µl	1.9	2.05	6.6

Table 4.3: DNA concentration, purity, and integrity of the selected longitudinal stool samples

All extracted HMW-DNA samples showed an optimal DNA concentration, 260/280 ratio, 260/230 ratio, and DNA Integrity number for downstream sequencing application for Nanopore and Illumina platforms (Figure 4.2).

High-coverage metagenomic sequencing of selected longitudinal samples using short-, long-, and linked-read sequencing.

To create the host's reference metagenomes and detect signatures of microbial evolution driven by age-related physiological changes of the host's environment, I performed high-coverage metagenomic sequencing of the selected longitudinal stool samples using three different sequencing strategies. Short-read shotgun metagenomics is characterized by generating billions of short sequencing reads (<300 bp), which provide accurate detection of within-species genetic diversity and bacterial evolution (4,5,36).

However, metagenome assembly using short sequencing reads generates fragmented, inaccurate, and incomplete genomes due to the complexity of the microbial community and the presence of repeat elements (16, 50, 51, 52). Long-read metagenomics sequencing, characterized for the reads with an average 20-30 Kb length, can directly improve

metagenomic assemblies by resolving repeat regions, increasing assembly contiguity (16, 28, 53, 56, 57), allowing the study of mobile and repeat elements in microbial adaptation and evolution (28, 55).

However, long-read sequencing technologies generate reads with higher sequencing error than short-read platforms. To overcome this issue, novel assembly strategies that combine short, high-fidelity reads to correct the sequencing errors in long-read sequencing data (54) generate complete and accurate microbial genomes of complex natural communities (16,23, 28).

Finally, linked-read is a sequencing technology characterized by isolating a large DNA fragment, fragmented in shorter DNA pieces, and labeled with a specific DNA barcode (58,59). Previous studies demonstrated that Linked-read metagenomics improves metagenome assembly and taxonomic classification in human gut communities (60) and could be implemented on haplotype phasing, population genetic approaches, and resolve complex structural variation in gut microbial communities (58, 59, 61).

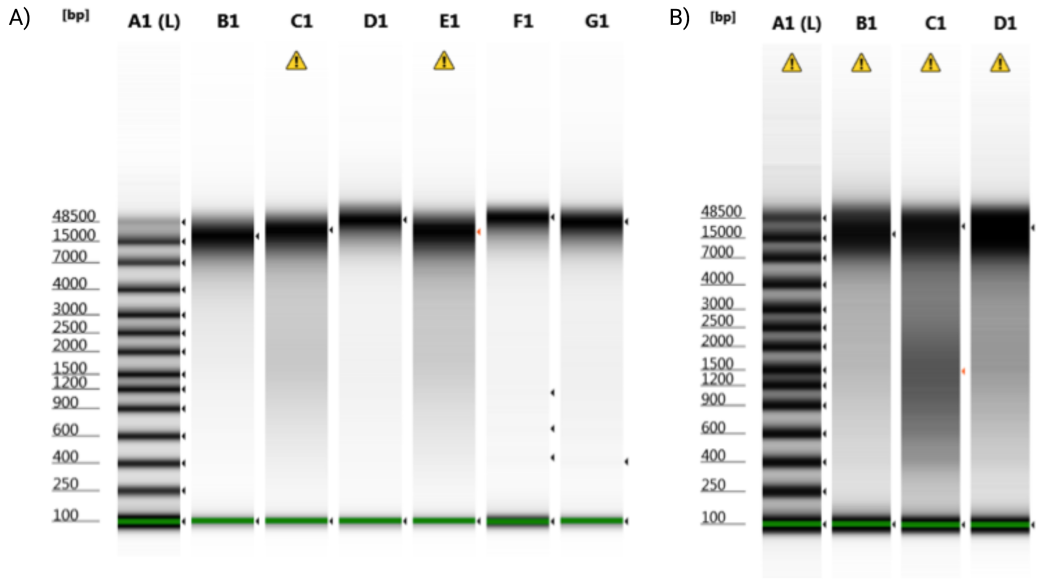


Figure 4.2: HMW-DNA extraction of longitudinal stool samples A) Samples from mouse B1) SG0476 Young; C1) SG0476 Adult; D1) SG0476 Old; E1) SG0696 Young; F1) SG0696 Adult; G1) SG0696 Old; L) Genomic DNA Ladder. B) HMW-DNA extraction from mouse B1) SG0503 Young; C1) SG0503 Adult; D1) SG0503 Old; L) Genomic DNA Ladder.

By combining short-, long-, and linked read metagenomics sequencing, I aim to reconstruct host-specific metagenome-assembled genomes (MAGs) and identify genetic changes on the gut microbial population produced by age-related physiological changes of the host's environment during aging. The analysis of the genetic modifications could shed light on the possible effects of the genetic modifications on microbial adaptation (Figure 4.3).

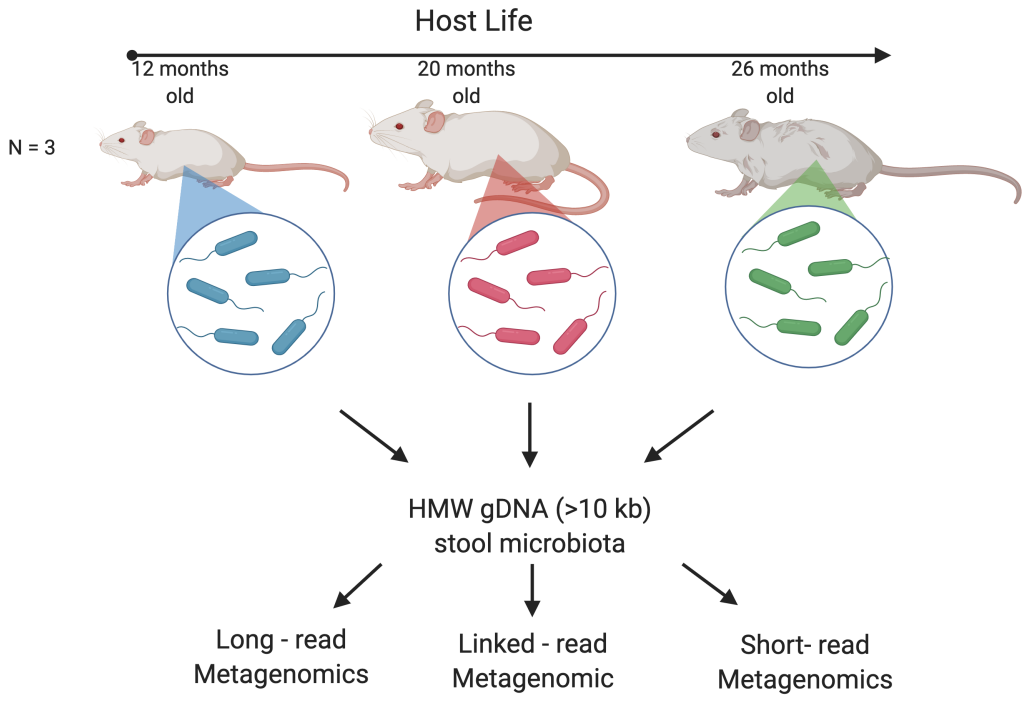


Figure 4.3: Experimental strategy to fine-scale characterize the genetic changes on the gut microbial population during host aging. Created with BioRender.com.

Short-read metagenomic sequencing preprocessing

I performed high-coverage short-read sequencing of the nine selected longitudinal samples to recover host-specific genomes and identify DNA sequence variations that occur in a single nucleotide position in the genome (SNV). Short-read libraries were generated using Illumina DNA prep kit (Methods) and sequenced in three lanes of a HiSeqX Illumina platform to generate 740 million of 2x150 raw reads. I removed low-quality reads, sequencing adapter trimming, and host reads decontamination of the total raw sequencing reads using the KneadData tool (Table 4.4). After short-read metagenomic preprocessing, I plotted the quality control reports using FastQC, and no quality problems were observed for downstream analysis (64) (Figure 4.4).

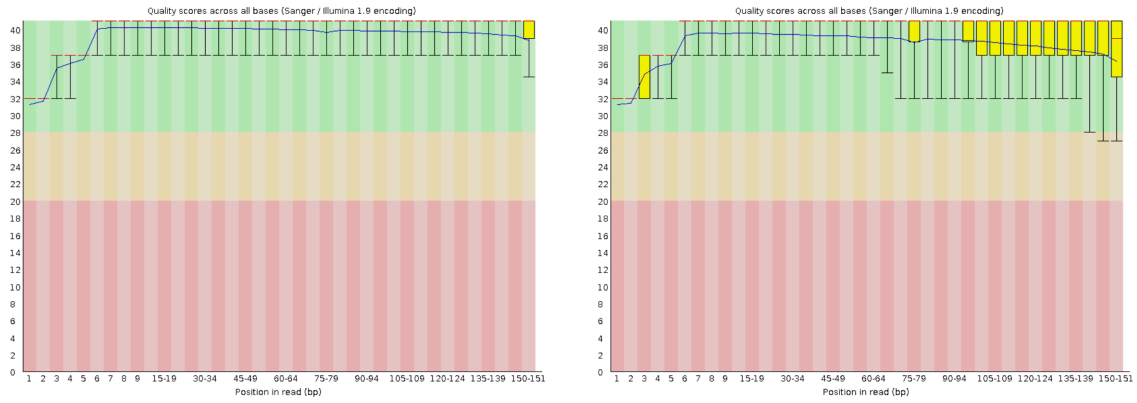


Figure 4.4: Quality control plots of all of the selected longitudinal samples after quality filtering and host removal of host reads.

Long-read metagenomic sequencing and data preprocessing

Nanopore sequencing can generate very long reads (> 2 Mb), allowing significant improvement in de novo genome assembly, resolution of complex genomic structures (DNA repetitive elements and structural variants), and reconstruction of high-contiguity metagenome-assembled genomes (MAGs) (107). Therefore, I performed a metagenomic sequencing approach using the GridION sequencing platform from Oxford Nanopore Technologies. Metagenomic samples from the same individual were pooled and sequenced in individual flow cells. After libraries were basecalled, and fastq files were generated using Guppy pipeline (Methods), nanopore reads with quality score < 7 and library adapters were removed from the dataset (Table 4.5). I used the NanoPlot tool to visualize the overall quality and length of the generated long-reads and observe the correct performance of the filtering tools for downstream assembly of de novo genome assembly (Figure 4.5).

Linked-read metagenomic sequencing and data preprocessing

Linked-read sequencing leverages long-range information in the form of molecular barcodes with a low error rate of short-read sequencing technology (106, 108). I implemented linked-read sequencing technology to identify the different haplotypes or lineages (phasing) present in each microbial population and track their longitudinal trajectories during host life. I used the TELL-Seq technology to partition long DNA fragments and tag the 3' end with an 18 nucleotides barcode (19, 20). Raw fastq reads served as input files for the Tellysis software, the manufacturer's developed bioinformatic tool for processing

Sample ID	Total read	Read after filtering	Final reads	Throughput (Gb)
SG0476-1	104,985,209	75,429,273	73,368,281	10.79
SG0476-5	112,364,851	82,570,056	80,825,350	11.84
SG0476-8	128,240,383	93,367,988	92,031,094	13.54
SG0696-1	107,613,584	77,437,992	76,405,368	11.22
SG0696-5	99,002,947	71,250,268	64,688,898	9.48
SG0696-8	98,914,708	72,023,221	66,413,422	9.77
SG0503-1	72,947,158	52,852,769	49,088,772	7.19
SG0503-5	85,569,094	62,163,567	54,400,806	7.92
SG0503-8	122,191,923	88,554,261	83,865,240	12.35

Table 4.4: Total number of reads per sample after quality trimming

Sample ID	Input DNA	Number of reads	Throughput(Mb)	Read length N50
SG0476-1-5-8	1.1 µg	4,539,306	7,808.7	3,070
SG0696-1-5-8	1.2 µg	4,368,134	10,466.6	4,138
SG0503-1-5-8	0.9 µg	1,324,686	7,570.3	11,055

Table 4.5: Total number of long-reads after quality filtering and adapter trimming of each host metagenome

Tell-Seq sequencing data. Tellysis comprises three main pipelines: Tell-Read, Tell-Sort, and Tell-Link. Tell-Read generates linked-read FASTQ data and quality control reports (Supplementary Figure 1) for downstream assembly (Tell-Link) and phasing (Tell-Sort) pipelines (Table 4.6).

I examined the quality control reports of each metagenomic sample, and no low-quality sequences were observed. I conclude that the generated FASTQ linked-read files were optimal for the downstream barcode-aware *de novo* assembly pipeline. However, the sample from host 2 at a young age contains fewer reads than the rest of the metagenomic samples. This reduced number of reads could impact the complete reconstruction of the metagenome assembly for host 2.

Reconstruction of MAGs using short-read metagenomic sequencing data

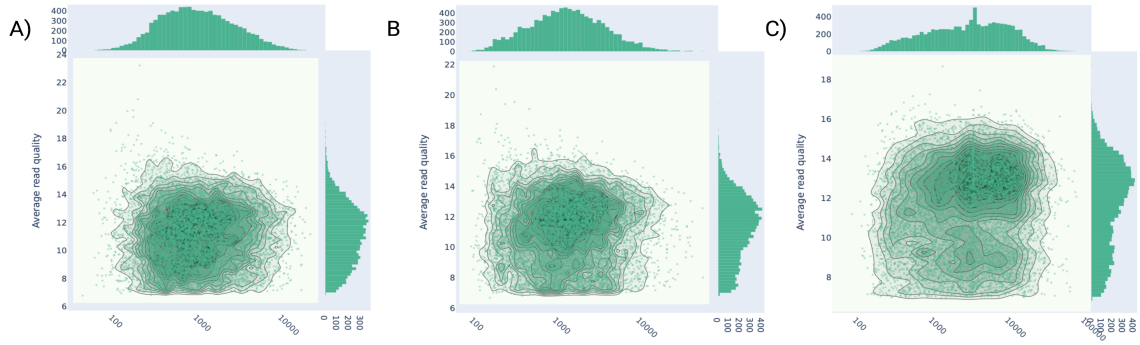


Figure 4.5: Comparison plot of log transformed read length with average Phred quality score of each metagenome dataset.A) Host 1 - SG0476; B) Host 2 - SG0696; and C) Host 3 - SG0503

Sample ID	Raw reads	Corrected reads	Sample index
SG0476-1	63,075,388	59,079,238	T501
SG0476-5	62,446,548	58,513,402	T505
SG0476-8	59,439,456	55,748,969	T508
SG0696-1	10,243,204	9,584,697	T502
SG0696-5	49,532,827	46,532,302	T504
SG0696-8	42,492,520	39,822,926	T506

Table 4.6: Number of reads generated after quality filtering of each metagenomic sample.

The reconstruction of host-specific metagenome-assembled genomes (MAGs) allows the recovery and analysis of genomes for organisms absent in metagenome reference databases and reduces read mismapping produced by intraspecies genetic divergence (62). The major advantage of host-specific reconstruction of MAGs is that it allows the accurate detection of SNV and intraspecies genetic diversity analysis (36,63). Therefore, the subsequent step was metagenome assembly, which merged overlapping reads into longer contiguous DNA fragments (62).

I performed a host-specific assembly using only metagenomic samples at the host's age of 12 months (young sample). Our experimental approach is a read-based mapping approach that assumes the reconstructed MAGs at a young age as metagenomic baseline (ancestral metagenome). Mapping metagenomic samples from 20 and 26 months old to their corresponding host-specific reference metagenome allow the detection of SNV, pro-

file intra-population genetic diversity, and determine essential population genetic metrics in the gut microbiome (36,65).

I utilized the metagenomic pipeline of the SPAdes toolkit called metaSPAdes for the metagenomic assembly of the short-read sequencing (30). I selected the metaSPAdes pipeline because of its optimal assembly performance of large and complex microbial communities, reduced computational resources, and creation of high-quality consensus genomes (30,62). MetaSPAdes assembly generates a large number of small contigs for each reference metagenome (Figure 4.6).

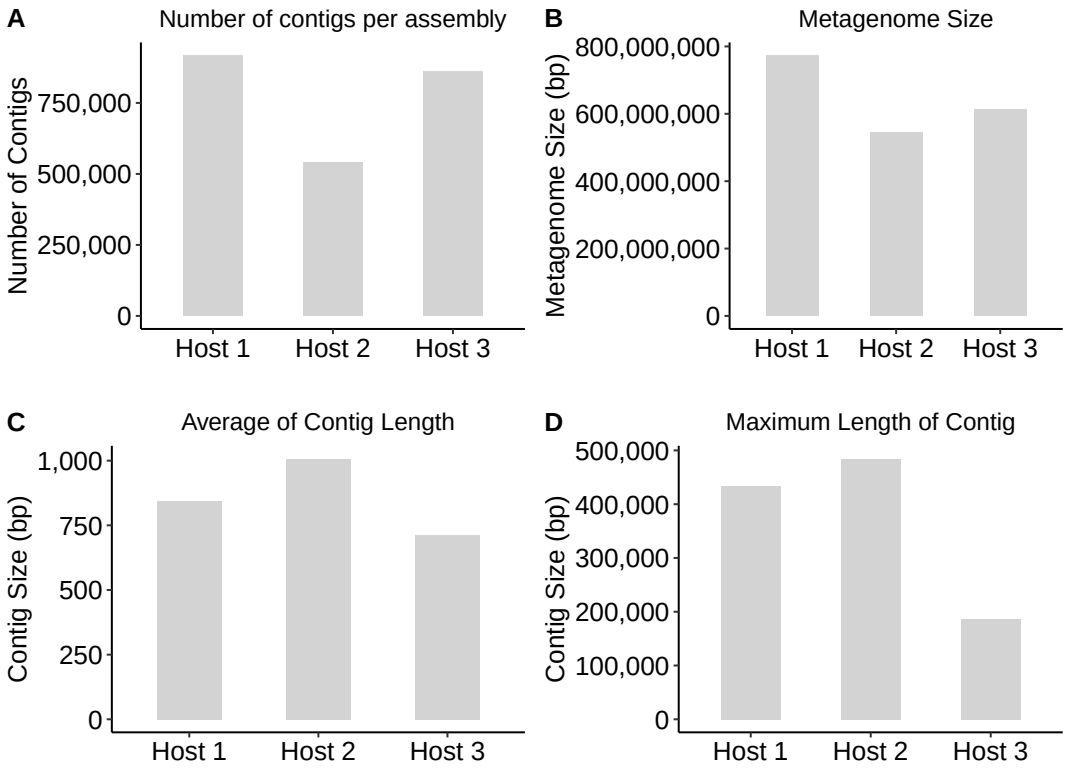


Figure 4.6: Statistics of the metagenome assembly of the three different host using short-read sequencing data after removing small contigs (<2kb)

Metagenomic assembly indicates that host 1 (SG0476) shows the higher number of contigs and largest metagenome size respective to host 2 (SG0696) and host 3 (SG0503). However, metagenomic data from host 2 shows the largest contig assembled (484,357 bp) and the highest average of contig length. These results correlate with the number of total reads after quality filtering and host-read decontamination, where host 2 contains the higher number of sequencing reads (Table 4.4). Previous studies indicated that a higher

number of short-read increases assembly contiguity, reaching a maximum contig size of 100 kbps using this sequencing technology (28).

The generated assemblies are composed of highly fragmented genomes. It is difficult to predict which contigs are coming from the same genome and determine the number of genomes present in the metagenomic sample (62). To solve this issue, metagenomic binning is essential for recovering high-quality MAGs (63).

Metagenomic binning consists of clustering contigs with similar sequence properties (e.g., codon usage, k-mer nucleotide frequencies, GC content) and read-abundance profiles (31, 62, 67, 68, 69). However, binning tools performed poorly on contigs smaller than 2 kb (73). Therefore, contigs smaller than 2kb were removed from the binning step.

To take advantage of multiple binning approaches and extract high-quality MAGs, I used the bin assignment, bin refinement, and bin reassembly modules of the toolkit metaWRAP to recover high-quality metagenome-assembled genomes (MAGs). Based on the definition of the Minimum Information about a Metagenome-Assembled Genome (MIMAG) (72), recovered MAGs with genome completeness >=50% and contamination level <=10% using CheckM metrics (71) are used for downstream analysis (70). Completeness and contamination scores are estimated by the presence and the copy number of single-copy marker genes in the evaluated MAGs (71).

metaWRAP recovered 95, 75, and 74 high-quality MAGs from hosts 1, 2, 3, respectively, after bin assignment and bin refinement modules (Supplementary Table 1, Supplementary Figure 1). To improve the quality of the MAGs after the refinement module, I used the reassembly module from the metaWRAP toolkit. The refinement module extracts the reads that belong to each MAGs and assembles them separately (31). metaWRAP reassembly implements SPAdes assembler for individual reassembly of each bin, improving the quality of the high-quality MAGs (Figure 4.7, Supplementary Table 2, Supplementary Figure 2).

After bin reassembly, some genomes have a quality and contiguity improvement (Figure 4.7). There is a general improvement in N50 contiguity in almost all the MAGs reconstructed from host 3 and the middle-ranked genomes from host 1 and host 2 (Sup-

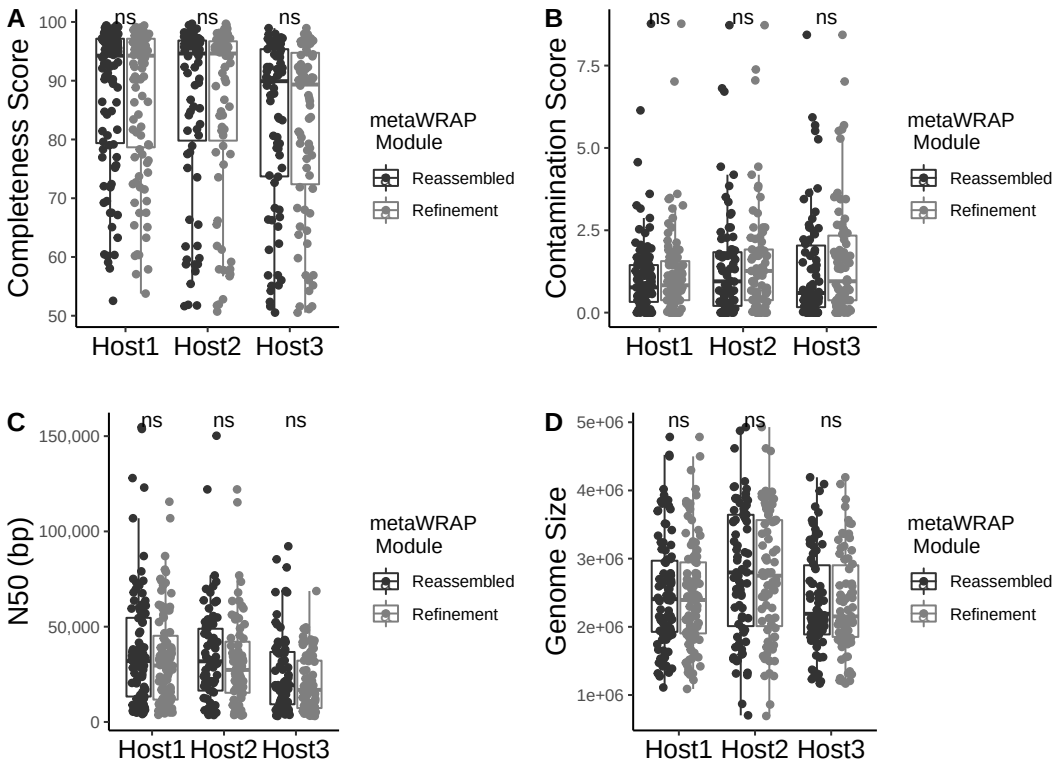


Figure 4.7: Statistics of metagenome binning modules of metaWRAP toolkit. Comparison of A) Completeness Score; B) Contamination Score; C) N50; D)Genome Size between Refinement and Reassembled module.

plemental Figure 2). However, bin reassembly slightly improves the completeness score in all hosts and contamination score in host 1 and host2. Notably, the reassembly module reduced the contamination score in almost all the reconstructed MAGs of the host 3 (Supplemental Figure 2). The Wilcoxon test showed non-significant differences between mean comparisons in all MAGs metrics (Figure 4.7).

Metagenomic assembly could generate highly similar MAGs across the sample data set, constraining downstream analysis and leading to incorrect population genomic inferences by erroneous recruiting reads (mismapped reads) during mapping sequencing reads (36). Genome dereplication is applied to remove genome redundancy to the MAGs collection. Dereplicaion cluster highly similar genomes and establish the highest quality genome representative for each group (62). The similarity between two genomes is estimated through a genome comparison at the nucleotide level (Average Nucleotide Identity - ANI)(62). For bacterial genomes from complex microbial environments (e.g., gut, soil, ocean), a

dereplication species threshold of 95% is suggested to ensure that the data set contains no redundant genomes (74).

I used the dRep dereplication tool (37) to remove genome redundancy on each reconstructed MAGs data set, setting the species threshold to 95% ANI. After performing the dereplication pipeline, the ANI value of all of the MAGs was lower than the 95% ANI species threshold. Therefore, I conserved the same number of MAGs per individual, 95 MAGs for host 1, 75 MAGs for host2, and 74 MAGs for host3 (Supplemental figure 3). The obtained reference metagenome collection is the final list of MAGs for analyzing intraspecies genomic diversity.

It is essential to identify which bacterial species are included in each reference metagenome collection. Within-species comparison between individuals could elucidate the adaptive strategies to encounter novel environmental changes (75, 76). I used the Genome Taxonomic Database Toolkit (GTDB-TK) (38, 39) to accurately taxonomic classify the reconstructed MAGs of each reference metagenome collection. GTDB-TK identifies and concatenates 120 ubiquitous single-copy proteins of each MAGs and places them into a protein reference tree, providing an objective and accurate taxonomic classification of bacterial and archaeal genomes (39).

GTDB-TK identified all the MAGs from the three hosts as bacterial genomes. Several MAGs are classified as bacterial species previously reported as commensal residents of mouse gut microbiota (70, 77, 78, 79) and were present in the three mouse individuals. *Muribaculum intestinale*, a strictly anaerobic bacteria species recently described as commensal species in the murine microbiota, was classified in the three mouse hosts (bin.49.strict, bin.42.strict, bin.55.orig for host 1, host2, and host3, respectively) as high-quality draft genome (Completeness score >96% and contamination score < 1.5%) with similar genome size (2.5 Mb) and N50 contiguity scores (36 kb) (Figure 4.8). Identifying bacterial species in different hosts might elucidate genes under selective forces and show host-specific molecular adaptations.

Reconstruction of MAGs using a hybrid assembly approach

Hybrid assembly combines the strengths of both short-read and long-read sequencing

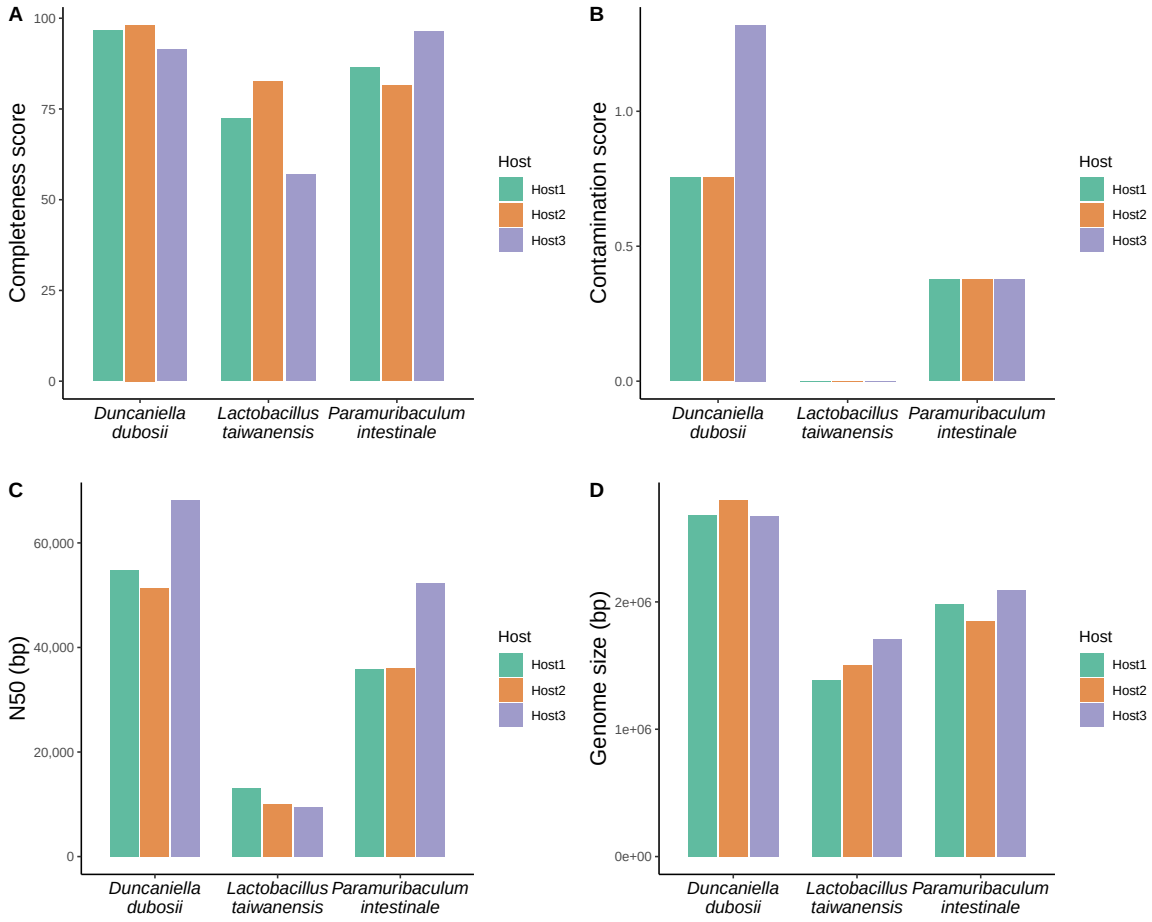


Figure 4.8: Comparison of three host-specific reconstructed genomes from shotgun metagenomic. Reconstructed genomes from three different bacterial species show similar assembly metrics. Each genome represents a host-specific version of the bacterial species. Comparative genomics approach can identify targets of selection in independent hosts (80).

approaches generating high accuracy and near-complete genomes from complex metagenomic samples. OPERA-MS, a novel hybrid assembler, first creates contigs using short-read data and then uses the long-read information to bridge the gap between contigs (28, 107). I used the hybrid assembler OPERA-MS to reconstruct high-quality metagenome-assembled genomes from each host by combining short-read metagenomic samples with their corresponding pooled long-read dataset. Conversely, OPERA-MS assembly performance was lower than short-read only metagenome assembly (Table 4.7).

Hybrid assembly reconstructed 91, 75, and 88 High-quality bins from hosts 1,2, and 3, respectively. MAGs reconstructed from OPERA-MS show a lower completeness score and

Host	Metagenome Size(Mb)	Average contig length	Large contig
SG0476	720.4	1,013	110,178
SG0696	525.4	1,147	94,369
SG0503	625.09	1,421	2,211,450

Table 4.7: Performance of the OPERA-MS hybrid metagenome assembly

N50 contig length than short-read metaSPADES assembly. Notably, hybrid-assembled bins possess a higher contamination score than metaSPADES assembly (Figure 4.9). I hypothesized that pooling the HMW-DNA input sample could originate from the low performance of the hybrid assembly. Pooling metagenomic samples with similar microbial compositions could increase the assembly performance of shallow abundant bacterial species by increasing the coverage of such low abundant species. However, pooling samples could also increase strain diversity and generate an opposite effect (114, 115).

Interestingly, hybrid assembly from Host 3 performs better than metaSPADES assembly in terms of contiguity (N50 contig length) and the number of reconstructed bins, with no significant differences in contamination and completeness score. This improvement could result from this particular sample's sequencing performance (N50 Read length > 11 Kb and mean read quality >11.5). Overall, hybrid assembly significantly increases the contiguity and number of quality bins in one of the three metagenomes with no sacrifice of completeness and contamination scores. However, a reduction in bin quality is observed in the other two metagenomes. Therefore, I continued the population genomic analysis using the short-read metagenomic dataset.

Reconstruction of MAGs using linked-read metagenomic sequencing data

I used the Linked-read FASTQ data and the Tell-Link pipeline to perform a barcode-aware de novo assembly of young metagenomic samples of host 1 (SG0476) and host 2 (SG0696). I examined the statistics of both generated assemblies and observed a reduced performance on genome assembly (Table 4.8). Even though the assembly output shows a reduced performance of barcode-aware assembly, I assessed the generation of drat genomes by binning metagenome contigs using the binning refinement tool metaWRAP. I evaluated the resulting bins by determining the completeness and contamination scores

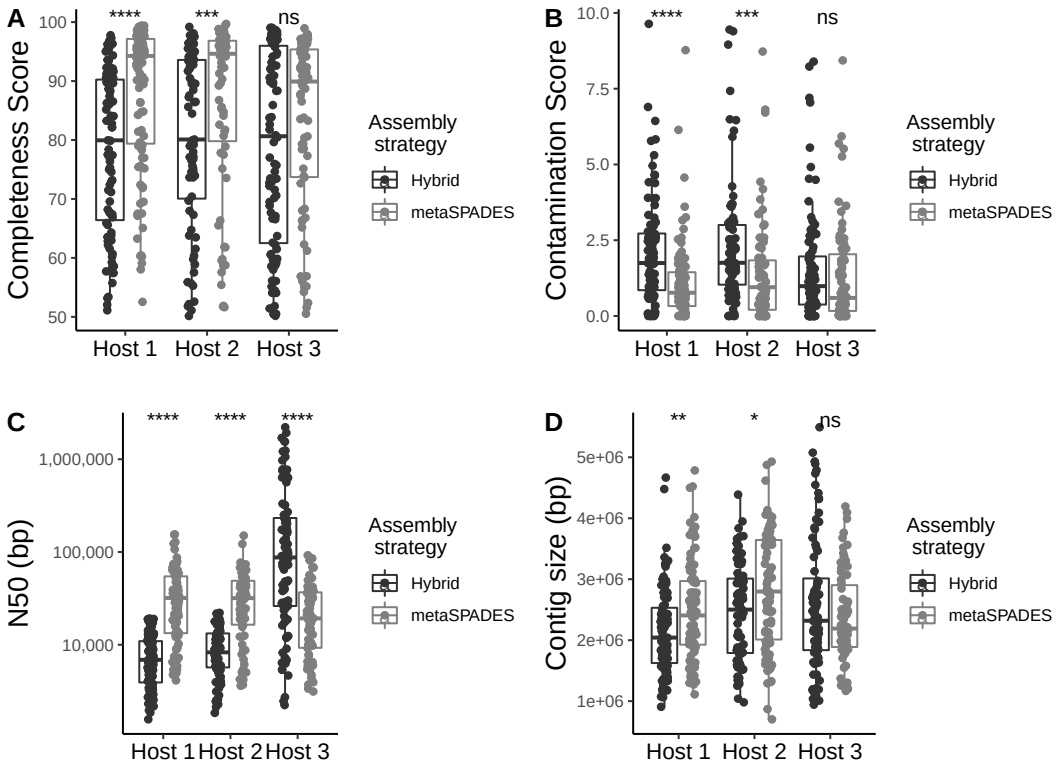


Figure 4.9: Comparison of quality and contiguity scores of hybrid assembly and short-read assembly. Hybrid assembly generates bins with lower completeness and high contamination scores than assemblies generated with short-read only. Except for the mouse 3, short-read assembly produces longer N50 score and larger contig size distribution than hybrid assembly.

and classified them as high-quality bins if the completeness score is >50% and contamination <10%. I conclude that barcode-aware *de novo* assembly shows a limited number of high-quality MAGs compared with short- and hybrid metagenome assembly, limiting the resolution for detecting novel mutations and the number of species to analyze in the gut microbiome over host life (Figure 4.10).

Host	Metagenome Size(Mb)	Number of contigs	Contig length N50	Large contig
SG0476	94.8	48,017	35,720	1,493,941
SG0696	20.9	9,443	3,343	512,574

Table 4.8: Assembly stats of barcode-aware *de novo* metagenome assembly

Microdiversity analysis of reconstructed microbial genomes using short-read metagenomic data

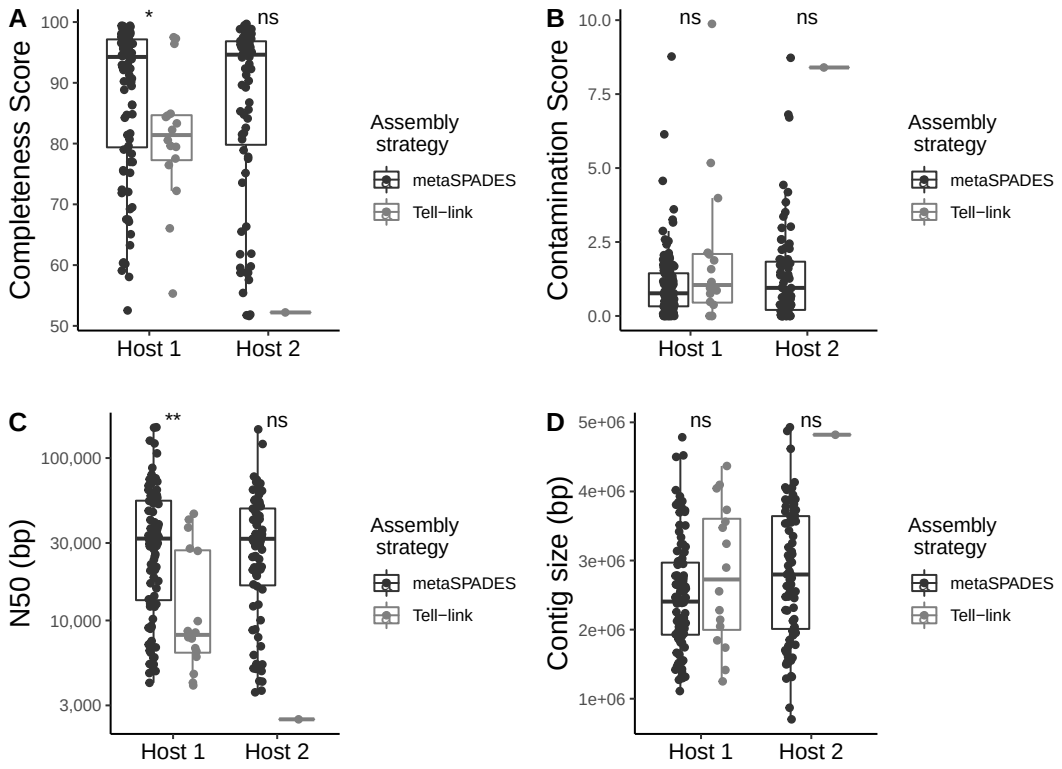


Figure 4.10: Comparison of high-quality bins generated by short-read and linked-read assemblies. Linked-read assembly generates a smaller number of high-quality bins than short-reads meta assembly. Moreover, linked-read bins show a reduced completeness score and N50 contigs length in comparison with metaSPADES assembly.

By combining longitudinal sampling and comparative population analysis of the same bacterial species in independent hosts, I can identify recurrent targets of selection (parallel evolution) and examine whether these targets of selection change over time (80). Therefore, the next step was to identify the targets of selection in each host and determine whether these targets are recurrent in each host at a specific age (age-specific selection) or are only present in one host (host-specific selection).

To identify the genes in a microbial community under selection, I simultaneously analyzed all the reconstructed microbial genomes and recalled the genetic variants and their changes in the frequency of these variants over time. An increase in the frequency of a genetic variant over time can indicate a positive effect on the bacterial fitness to specific environmental conditions. On the other hand, a low frequency of a genetic variant can reveal a deleterious effect on microbial fitness and be subject to purifying selection (80).

I used the inStrain microbial metagenomic tool (36) to identify SNV in all the reconstructed genomes and examine SNV frequency changes over time. inStrain identifies SNVs on a reference-based approach, meaning that sequencing reads are mapped to reference genomes to identify SNVs in all microbial species (36,81). Importantly, inStrain can perform intraspecies comparison in different metagenomic samples to track genomic dynamics of the same species over time.

To detect SNV and calculate the genetic diversity within a microbial species, InStrain requires three steps: 1) Mapping sequencing reads to reference genome database; 2) Profiling genetic diversity in each reference genome, and 3) microdiversity-aware genomic comparison between metagenomic samples (Figure 4.11).

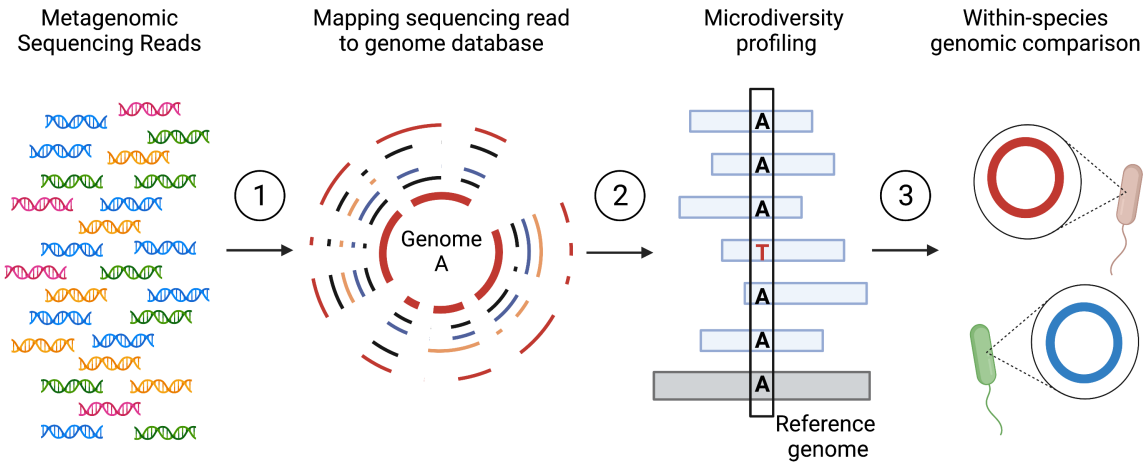


Figure 4.11: Overview of inStrain steps to measure within species diversity from metagenomic sequencing. Sequencing samples are mapped to a generated *de novo* genome database to generate mapping files (1). InStrain calculates population microdiversity using the mapping files (2) and performs population-aware comparisons between metagenomic samples (3). Created with BioRender.com.

The first step in the inStrain pipeline is to map each metagenomic sample to its corresponding reference metagenome database. Metagenomic samples SG0476-1 (Host 1 - Young), SG0476-5 (Host 1 - Adult), and SG0476-8 (Host 1 - Old) are mapped to Reference metagenome database Host 1. Metagenomic samples SG0696 - X (Host 2) are mapped to the reference metagenome database Host 2 and SG0503 - X samples to the reference metagenome database Host 3, respectively. One drawback concerning mapping metagenomic sequencing data from a complex community to a reference database

is that sequencing reads can be assigned to the genome where they are not originated. This mapping error is called read mis-mapping. Read mis-mapping originated when a sequencing read mapped equality to multiple genomes, leading to incorrect population diversity metrics (36).

To reduce read mis-mapping, I implemented a competitive mapping approach using the Bowtie2 mapping tool. Competitive mapping involves the mapping of the sequencing reads to multiple genomes simultaneously. By including more reference genomes when performing mapping, Bowtie2 maps reads to a single location. If a read maps at two different sites, Bowtie2 will place the read at the site with the higher identity with the reference. Overall, competitive mapping performs mapping more accurately (36). The first step to perform competitive mapping, I concatenated all reconstructed genomes sequences from each host in a single host-specific fasta file and mapped the corresponding metagenomic samples using Bowtie2 –very-sensitive parameter. I converted the output mapping file to .bam format using samtools (Methods). I generated three different mapping files per host reference genome database (mapping.hostX.young.bam, mapping.hostX.adult.bam, and mapping.hostX.old.bam).

InStrain can perform genome and gene-level microdiversity profiles. The genome-level profile provides the overall diversity metrics of each bacterial special, providing information about genome coverage, nucleotide diversity (π), the total number of SNV, and the total number of base substitutions (SNS). In contrast, the gene-level profile provides diversity metrics in each gene present in the reference metagenome database and gives insights about genes under selective pressures (dN/dS and pN/pS). inStrain requires a list file indicating which scaffolds are assigned to each genome (scaffold-to-bind file) and a list of the genes present in all the reference metagenome to generate the gene-level microdiversity profile.

I used the parse_stb.py script from the dRep pipeline to generate a scaffold-to-bind file for each reference metagenome database (host1.stb, host2.stb, and host3.stb, respectively) and prodigal for predicting protein-coding genes. I could obtain the largest number of scaffolds for Host 1 between the three reference databases. However, Host2 contains the highest number of protein-coding genes (Figure 4.12).

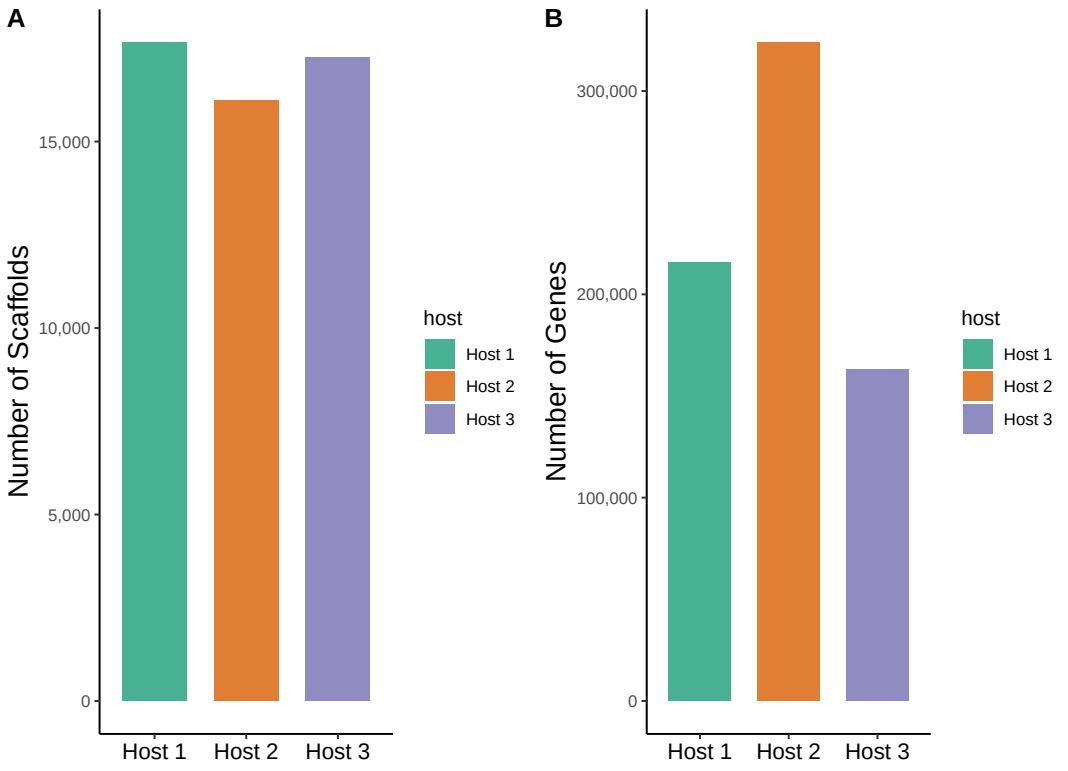


Figure 4.12: Total number of scaffolds (A) and protein-coding genes (B) in each reference metagenome database. Protein-coding genes were predicted using Prodigal with default parameters.

After generating all input files, I performed inStrain profiles of each longitudinal metagenomic sequencing using its corresponding reference metagenome database. I set the minimum ANI value between reads and reference genome that they map to 99% (–min_read_ani) and MapQ score >20 (remove reads that map in two different positions) to reduce mis-mapping events. Therefore, all reads with a minimum ANI value below 99% and MapQ score <20 were discarded for the inStrain profile. I applied these stringent read-filtering thresholds to minimize sequencing or read mapping errors (46). To calculate more precise population diversity estimates in each sample, I only focused on the most abundant species with a minimum genome coverage of 20x (5).

After filtering for removing read mis-mapping and minimum genome coverage thresholds, young metagenomic samples in all individuals show the highest number of genomes and scaffolds with a minimum coverage of 20x, followed by old and adult metagenomic samples (Figure 4.13).

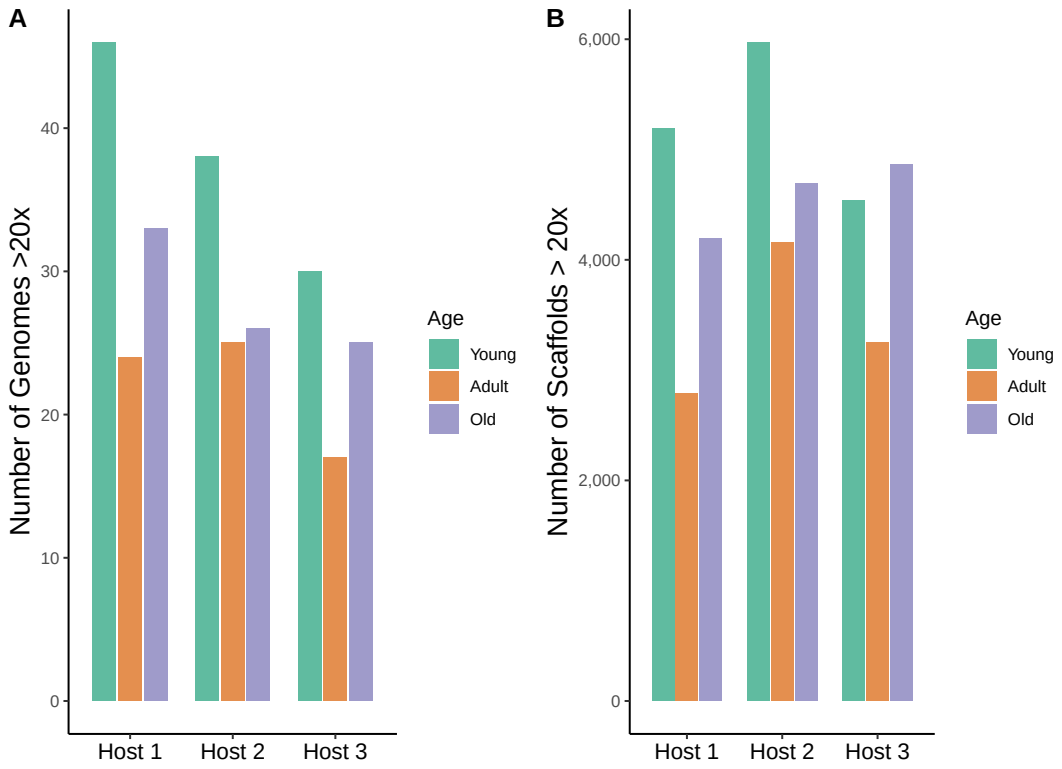


Figure 4.13: Total number of Genomes (A) and Scaffolds (B) with a coverage >20X in each metagenome database. I focus on genomes with a minimum coverage of 20x to minimize sequencing and read mapping error for reducing calculation errors in population diversity estimates.

inStrain generated multiple population diversity metrics in all the genomes and scaffolds with a coverage >20x. To identify signatures of selection, I focused on diversity metrics of nucleotide diversity (π), single nucleotide variants (SNVs), single nucleotide substitutions (SNSs), dN/dS, pN/pS, and iRep. Nucleotide diversity is measured using the method from Nei and Li (82). Nucleotide diversity is calculated in all sites in the genomes using the mapped reads. The equation for calculating nucleotide diversity at a specific site is defined:

$$\pi = 1 - [(Frequency of A)^2 + (Frequency of C)^2 + (Frequency of T)^2 + (Frequency of G)^2]$$

Single Nucleotide Variant is defined as a single nucleotide change present in a fraction of a population(36). On the other hand, Single Nucleotide Substitutions are SNV that reach fixation in a microbial population. Using SNV and SNS changes over time, inStrain calculates if those genetic mutations are under evolutionary pressure using the ratio dN/dS

and pN/pS.

dN/dS ratio measures evolutionary pressures on protein-coding regions and quantifies the mode and the strength of selection (84). This measure quantifies selection based on the rate of substitution at synonymous sites (dS) to the rate of substitution at sites (dN). Mutations at non-synonymous sites change the amino acid composition of the protein. Therefore, non-synonymous sites are exposed to selection pressures (85). A value of dN/dS equal to one is indicative of neutrality. Values greater than one indicate that natural selection promotes changes in the protein sequence, while values less than one are expected when selection suppresses protein changes (84, 85). While dN/dS ratio focuses on nucleotide substitutions, pN/pS ratio estimates the rate of non-synonymous polymorphism or variants (pN) to the rate of synonymous polymorphism (pS). pN/pS ratio > 1 can be indicative of balancing selection or purifying selection, while purifying selection is suggested when pN/pS < 1 (86).

inStrain quantified the total number of Single Nucleotide Variants (SNVs) and Single Nucleotide Substitutions (SNSs) per genome based on their corresponding reference database (Figure 16). The number of SNVs ranges from 10^2 to 10^4 (Figure 4.14 - A), and SNSs range from 0 to 10^3 (Figure 4.14 - B). SNVs and SNSs could have originated from evolutionary changes (*de novo* mutations) or by ecological changes (fluctuations of colonized lineages within the same bacterial population (58) coinfection/replacement of a divergent bacterial strain from the environment (5, 46)).

To distinguish if the origin of SNVs and SNSs are from evolutionary or ecological effects, I analyzed the distribution of the SNVs frequencies of each bacterial population. A single evolving population (single-colonized population), assuming that synonymous mutations are neutral over time, is expected to generate an SNVs frequency distribution dominated by a single peak at high frequency. In contrast, bacterial populations with a mixture of distinct lineages are expected to generate a distribution of allele frequencies with one or more peaks at intermediate allele frequencies (5, 46). Identifying *de novo* mutations in single-colonized populations could imply an adaptive response to novel environmental perturbations (evolutionary changes) (5). Therefore, I focused on identifying single-colonized populations with no signal of coinfection over the host's lifespan with at

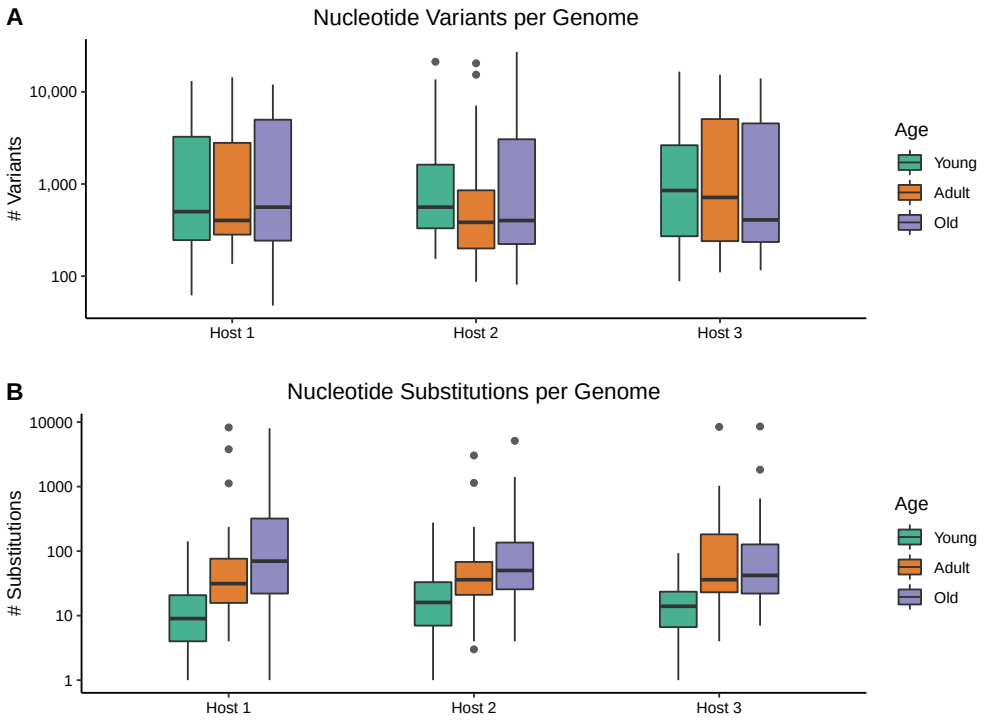


Figure 4.14: Nucleotide variation per genome in all reference metagenome. A) Number of Single Nucleotide Variants (SNVs); B) Single Nucleotide Substitutions (SNSs) per each genome from the three reference metagenome dataset over time.

least two time points.

Detected SNVs were filtered before plotting allele frequency distribution per each genome. I excluded SNV sites coverage with very low or high values (c) compared to the median genome coverage (C). SNV sites with low coverage (c<C) arise from mapping errors or reads donated from less abundant species. On the other hand, SNV sites with high coverage (c>C) arise from multi-copy genes or repetitive regions (5). Therefore, I excluded SNV sites if c <0.3C or c>3C (5).

After site coverage filtering, I examined the SNVs frequency distribution in every recovered genome to select single-colonized populations with no signal of coinfection over the host's lifespan in each metagenome database (Figure 4.15 - Supplementary Figure 4). Examining the SNV distribution of all the genomes in our datasets revealed that the three metagenomic datasets show single peak distribution populations (Supplementary Figure 4). Interestingly, I identified most single-colonized populations at the young age of the host (12 months old). Host 1 shows eight single-colonized microbial populations at a

young age (Supplementary Figure 4 - Bin 4, 9, 15, 40, 45, 74, 81, and 96), while only one shows single-colonized distribution at adult (Bin 90) and old (Bin 47) age. Host 2 and Host 3 show the same pattern. Host 2 only shows five single-colonized microbial populations at a young age (Bin 1, 6, 10, 16, and 17). In comparison, host 3 shows five single-colonized populations at a young age (Bin 6, 33, 44, 60, and 71), one population at adult (Bin 12), and two at old age (Bin 20 and 74).

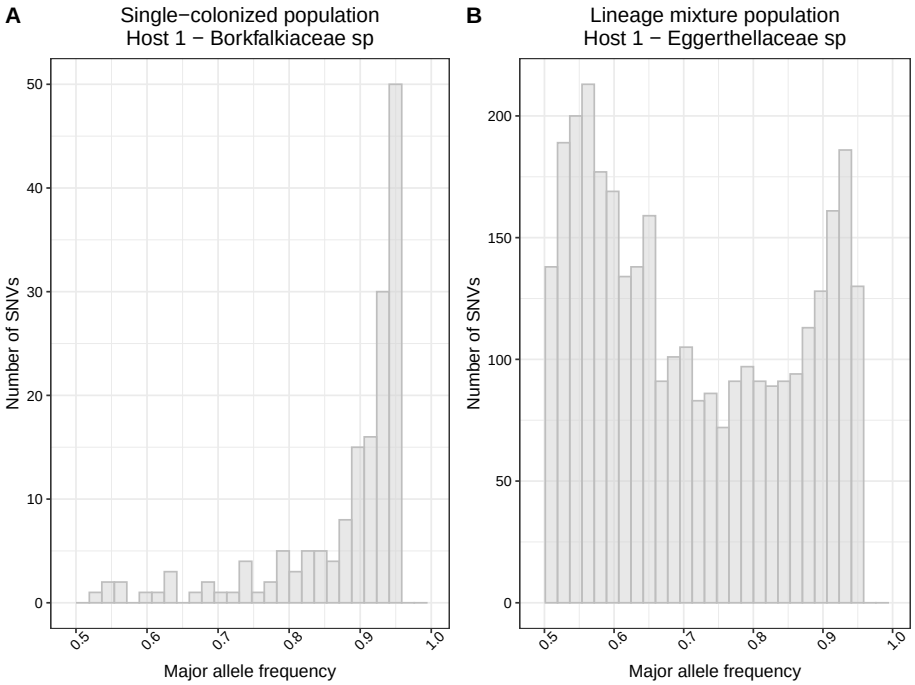


Figure 4.15: SNV frequency distribution. A) Single-colonized population shows a single peak at higher frequencies; B) Lineage mixed population shows one peak at intermediate frequencies.

The Lachnospiraceae, Muribaculaceae, and Bacteroidalesfamilies comprehend the single-colonized microbial populations in our dataset. Lachnospiraceae and Muribaculaceae families belong to the core microbial community in mice gut and are among the main producers of short-chain fatty acids (90) and complex carbohydrates degradation (91). Interestingly, the Muribaculaceaea family was characterized as Gram-negative, non-motile, not sporulated bacteria (91). Recent evolutionary studies in the human microbiome associate the lack of sporulation with single intraspecific competition and host adaptation (92, 93), partially explaining single-colonized SNV distribution in the Muribaculaceae family members.

However, the most critical question to address using the SNV frequency distribution was determining which microbial populations are characterized by single-colonized populations and show no signs of coinfection or lineage replacement from environmental strains over the time course of host lifespan. I examined microbial populations where SNV frequency distribution from at least two longitudinal metagenomic samples show a dominated peak at high frequency (>=80%). Host 1 shows two microbial populations with an SNV frequency distribution dominated by a peak at high frequency at young and old age (Bin 42 - UBA7182 sp009774145 and Bin 86 -Alistipes sp002362235).

Host 2 shows three microbial populations suggestive of single-colonized populations. Bin 52 - *Muribaculum sp.* and Bin 54 - *14-2 sp.* show simple lineage structure at young and adult age, while Bin 70 - *14-2 sp.* shows single population structures over the three longitudinal samples. In comparison, host 3 presents three possible single-colonized microbial populations. Bin 30 - *Duncaniella freteri* displays a single structure at young and old age samples, while Bin 57 - *Alistipes sp002428825* and Bin 59 - *Alistipes sp002358415* shows single populations in all the metagenomic samples (Figure 4.16).

Interestingly, I observed microbial populations that showed a change in structure over time. In host 2, bin 41 - *UBA7182 sp.*, shows an SNV frequency distribution characteristic of a single-colonized population at a young age. However, the frequency distribution of SNV at old age shows multiple peaks at intermediate frequencies, suggesting a fluctuation in lineage frequencies or coinfection (3,5). In contrast, I observed multiple strains of bin 3 - *UBA7182 sp.* from host 1 at adult age, which changed again to one dominated peak at high frequency at old age (Supplementary Figure 4). This finding could suggest that one bacterial lineage became the dominant strain in the population during host aging. However, while changes in strain numbers for each bin could reflect evolutionary and adaptation events, they could also reflect stochastic oscillations.

However, a possible scenario in longitudinal single-colonized populations is that a new bacterial lineage coinfecs the host's environment between sample timepoints and replaces the existing lineage completely. To uncover if the microbial population remains stable throughout host life, I compared the eight selected lineage genomes at different ages using the microdiversity-aware ANI calculation popANI (36). inStrain performs a

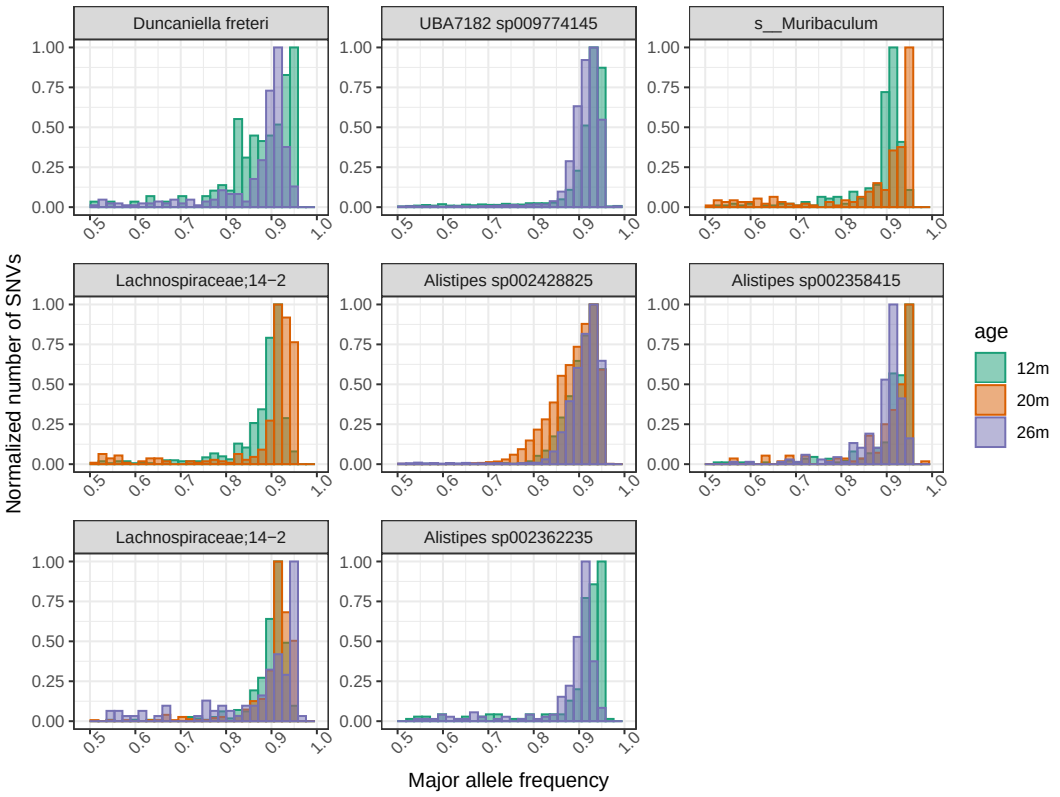


Figure 4.16: SNV frequency distribution of selected microbial populations. Eight microbial populations show a single-colonized model over host lifespan selected based on high genome coverage (>=20x) with threshold frequency of 0.8 and low rate of intermediate frequency polymorphism.

microdiversity-aware comparison metric that considers major and minor alleles. The first step in calculating popANI value is identifying all positions in the genome with a minimum coverage of 20x. These bases are compared between samples to determine if there is a popANI substitution at that position. (Supplementary Table 5 - compared_bases_count). A popANI substitution is called if the two compared populations do not share any alleles at that position. popANI is calculated using the following formula:

$$popANI = (comparedbasescount - popANIsubstitutions)/comparedbasescount$$

For metagenomic analysis, identical strain is defined by a 99.999% popANI value (36). This stringent popANI value represents less than ten years of divergence time between samples based on a previously reported accumulation rate of 0.9 SNS per genome per year in the human microbiome (4). In addition to the popANI value, the required fraction of the genomes with a minimum coverage of 20x between both samples is >=50%.

inStrain profile comparison between different longitudinal samples reveals that all microbial populations possess a popANI value >99.999% with a genome comparison >=50% suggestive for identical strains (Table 4.9).

Host	Species	Sample1	Sample2	% compared	popANI
1	*UBA7182 sp009774145*	Young	Old	97.1%	99.999%
1	*Alistipes sp002362235*	Young	Old	65.1%	100%
2	*Muribaculum sp.*	Young	Adult	68.3%	100%
2	*Lachnospiraceae 14-2 sp.*	Young	Adult	50.8%	100%
2	*Lachnospiraceae 14-2 sp.*	Young	Old	52.6%	100%
2	*Lachnospiraceae 14-2 sp.*	Adult	Old	85.1%	99.999%
3	*Duncaniella freteri*	Young	Old	70.1%	99.999%
3	*Alistipes sp002428825*	Young	Adult	98.3%	100%
3	*Alistipes sp002428825*	Adult	Old	98.4%	100%
3	*Alistipes sp002428825*	Young	Old	99%	100%
3	*Alistipes sp002358415*	Young	Adult	93.1%	99.999%
3	*Alistipes sp002358415*	Adult	Old	68.8%	100%
3	*Alistipes sp002358415*	Young	Old	68.1%	99.999%

Table 4.9: popANI of each of the microbial population show unique dominant strain.

After confirming that selected microbial populations represent the same lineage throughout host aging, I examined whether the population's genetic variation (Nucleotide diversity, π) changes over time. I hypothesized that under a model of positive or purifying selection, nucleotide diversity of host-associated microbial populations should decrease during host aging (7). A decrease in microbial diversity that characterizes host aging is a strictly ecological phenomenon, increases host frailty (8, 95), and could be followed by increases in the population size of the most abundant bacterial species, creating the opportunity for microbial evolution in response to novel environmental perturbations (7) (Figure 4.17). Therefore, the emergence of beneficial variants could spread over the microbial population, reducing genetic variation in the microbial community. This reduction in nucleotide diversity might alter the community structure affecting host physiology. A recent study supports this hypothesis by showing that nucleotide diversity of the microbiome is associated with host fitness effects and adaptability in stressful environments

(94).

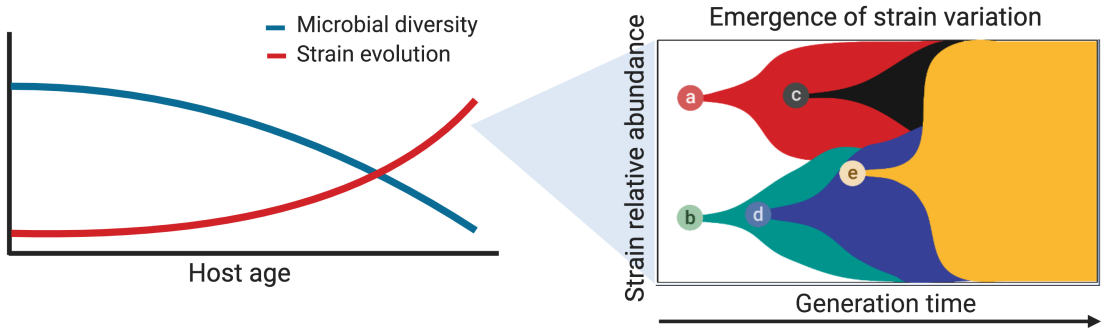


Figure 4.17: Decrease in taxonomic diversity during aging could lead microbiome evolution. Age-related decrease in taxonomic diversity in microbial community could leads the increase in population size of persistent microbial species, increasing the opportunity for evolving novel potentially pathogenic microbial strains (adapted from Davila-Aleman & Valenzano, 2019). Created with BioRender.com.

I observed that the nucleotide diversity of protein-coding genes in most microbial populations changes over time (Figure 4.18). Consistent with our hypothesis, the nucleotide diversity of *Muribaculum sp., Lachnospiraceae 14-2*, and *Alistipes sp002358415* significantly decrease during aging (p <= 0.0001). Interestingly, *Alistipes sp00248825* increases nucleotide diversity at adult age but drops at old age. *UBA7182 sp009774145* shows an increase in nucleotide diversity at old age, and *Duncaniella freteri* and *Alistipes sp002362235* show no significant difference over time.

Adaptive evolution can be detected by tracking the drastic increase in the frequency of SNV over time or by recurrent mutations in genes under selection (parallel evolution) (4, 5, 75, 96). These changes might be host-specific or age-specific adaptations and may contribute to bacterial resilience to age-dependent environment changes and community shifts over time (4, 5). To track the increase in the frequency of SNV during host aging, I identified SNV that shifts allele frequency from 0.2<= to >=0.8 in each genome between two-time points (SNV changes) (3,5).

I quantified SNV changes between sampled time points in all the single colonized microbial populations (Figure 21). Interestingly, most microbial populations experience no SNV changes over host life timescales (63%). However, three microbial species *(Duncaniella freteri, Muribaculum sp., and Alistipes sp002358415)* undergo <=15 SNV

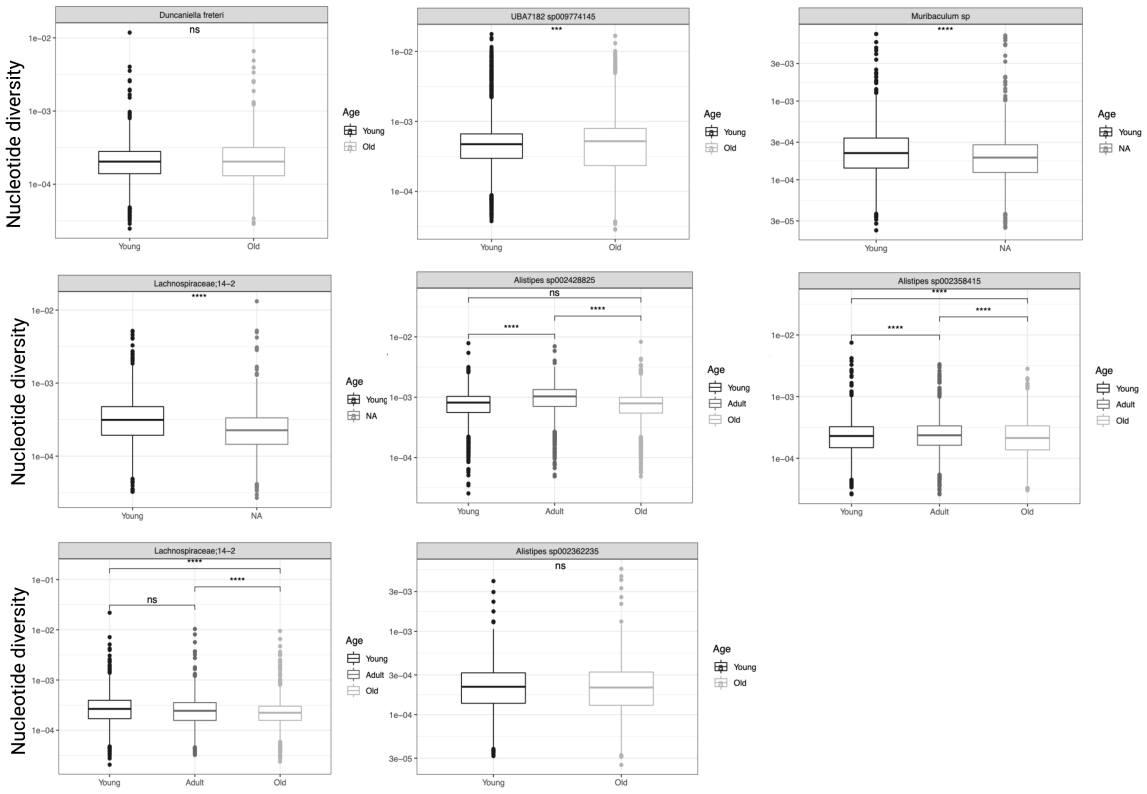


Figure 4.18: Nucleotide diversity of protein-coding genes of the selected single-colonized microbial populations over time. Changes in nucleotide diversity is observed in the majority of microbial populations over time. Reduction of π is observed in *Lachnospiraceae 14-2, Muribaculum* and *Alistipes sp002358415*. ns: p >0.05 ; *** : p <=0.001, **** : p <=0.0001

changes over a time scale of 14 months or less. The number of SNV changes in these populations is consistent with evolutionary changes of the population within a host in the human gut (<= 20 SNV changes), and no populations harbor >=10^4 SNV differences which is consistent with strain replacements (3,5). I next qualified the number of SNV that change allele frequency from <=0.5 to >=0.5 (Consensus SNV) and SNV that do not increase allele frequencies to <0.5 over host times scale (Variant SNV). I observed that all microbial populations undergo changes in consensus alleles and variant SNV (Figure 4.19 - blue and gray lines, respectively). Consensus SNV describes minor alleles that became dominant in the population (major allele) at old age. On the other hand, variant SNV describes minor alleles that remain at low frequencies. Consensus and variant SNV could be signals of balancing selection, beneficial mutations in the process of reach fixation in the population, or detrimental mutations getting purified from the microbial

community (97). However, further investigation is needed to explore the nature of these intermediate-frequency SNV.

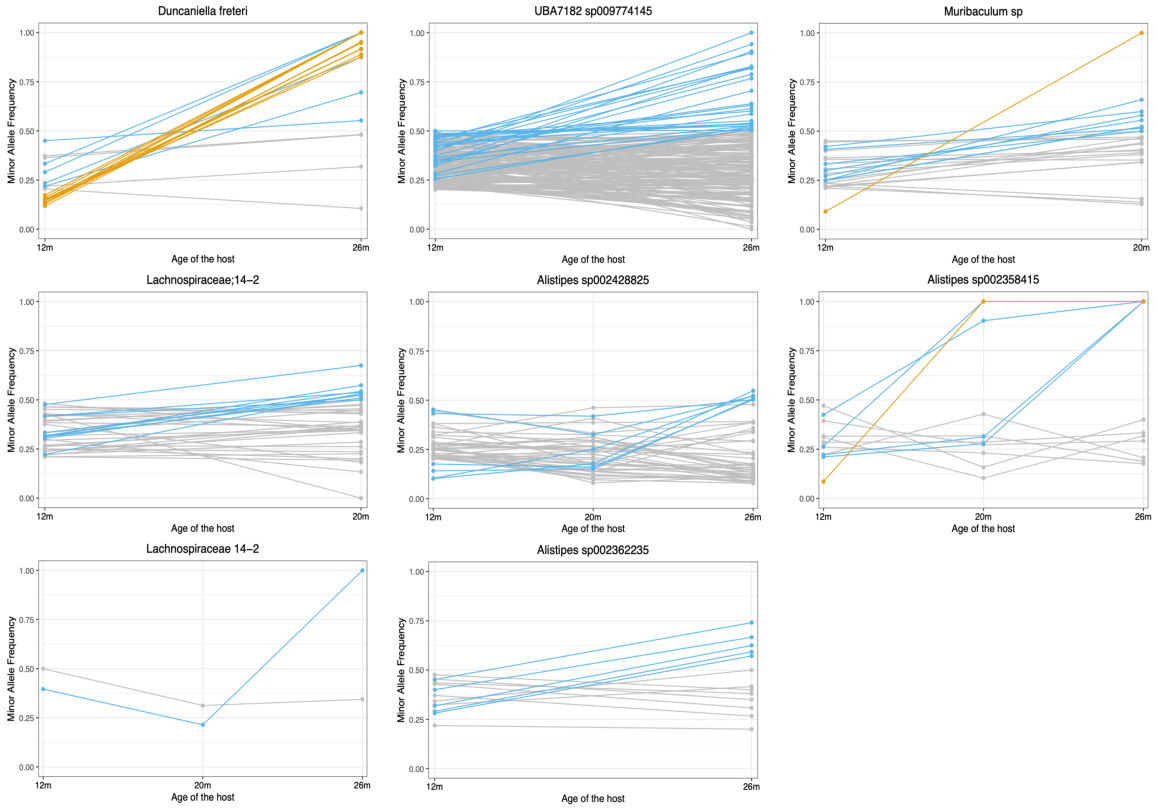


Figure 4.19: SNV changes in all the microbial populations over time. Most of the samples experience no SNV changes over host lifespan but three populations undergo a small number of SNV changes (Yellow lines) (<= 15 SNV changes) consistent with expected evolutionary modifications (3,5). However, all microbial populations experienced changes in consensus SNVs (blue lines).

To assess which genes undergo SNV changes, Consensus SNV, and Variant SNV changes, I aligned the DNA sequence of each gene against the NCBI database using BLASTX (98). Interestingly, *Alistipes sp002358415* and *Muribaculum sp.* show SNV changes in SusD family nutrient uptake outer membrane protein (Supplementary Table 6). Notably, *Alistipes sp 002358415, Duncaniella freteri*, and *Lachnosporaceae 14-2* also show consensus SNV changes in SusC family outer membrane proteins (SusC, SusF/SusE, and TonB dependent receptor) (Supplementary Table 7). SusC/SusD outer-membrane polysaccharide importers have been previously identified as gene families undergoing parallel evolution in the human gut microbiome (3,4) and showed a critical role for IgA-mediated colonization in mice (99). Another gene class that undergoes SNV changes over the host time scale

is the Carbohydrate Active Enzyme (CAZy) glycoside Hydrolase Family 43 *(Duncaniella freteri)*. In particular, the glycoside hydrolase family 43 is required to degrade complex glycans and is among the most abundant CAZymes in the human gut microbiome (100).

Consensus SNV changes are represented by distinct gene classes, including transcriptional factors, ABC transporters, and different transposase families (Supplementary Table 7). ATP-binding cassette (ABC) transporters are responsible for importing glycosaminoglycans (GAGs), the primary component of extracellular matrix in animals, into bacterial cells. Notably, a recent study demonstrates that commensal microbes from the human microbiome targets GAGs for adhesion to intestinal cells, and the presence of GAG-degrading pathways might be involved in their prevalence in the gut microbial community (101). On the other hand, transposase families can significantly contribute to the adaptation and evolution of bacterial genomes to overcome changing environmental conditions. For example, transposase genes flanking resistance genes are associated with the spread of antimicrobial resistance genes in the human microbiome (102).

To investigate whether SNV changes, consensus SNV, and variant SNV are adaptive events, I examined the dN/dS and pN/pS ratio computed by the inStrain pipeline. dN/dS ratio was calculated between the number of observed non-synonymous SNS divided by non-synonymous sites (dN) and the number of observed synonymous SNS divided by non-synonymous sites (dS). Analogous to dN/dS, pN/pS ratio is computed between non-synonymous (pN) and synonymous polymorphism (pS). dN/dS and pN/pS ratio are only calculated on genes that contain non-synonymous and synonymous substitutions/variants (Supplementary Table 8). dN/dS ratio was not possible to compute in most genes because of the absence of synonymous substitutions (synonymous SNS) in the gene sequence. However, pN/pS ratio was calculated in almost all genes (Figure 4.20).

I observed that the majority of the polymorphic genes showed pN/pS ratios <1 (Figure 4.20). Low pN/pS ratio values denote purifying selection of non-synonymous polymorphisms in the population (86). Previous work observed that genetic diversity in species of the human gut microbiota is strongly constrained by purifying selection (3, 5, 103). Remarkably, transposase genes in *UBA7182 sp 009774145* and *Muribaculum sp* show a pN/pS ratio > 1, indicating potential adaptive polymorphisms. Positive selection of

transposase genes could be associated with adaptation mechanisms related to the bacterial response to new environments (97, 104). Transposons can affect genes by disrupting coding sequences or upregulating neighboring genes by introducing promoter sequences, responding to physiological stress, and resulting in beneficial fitness effects for the bacterial host (105, 106).

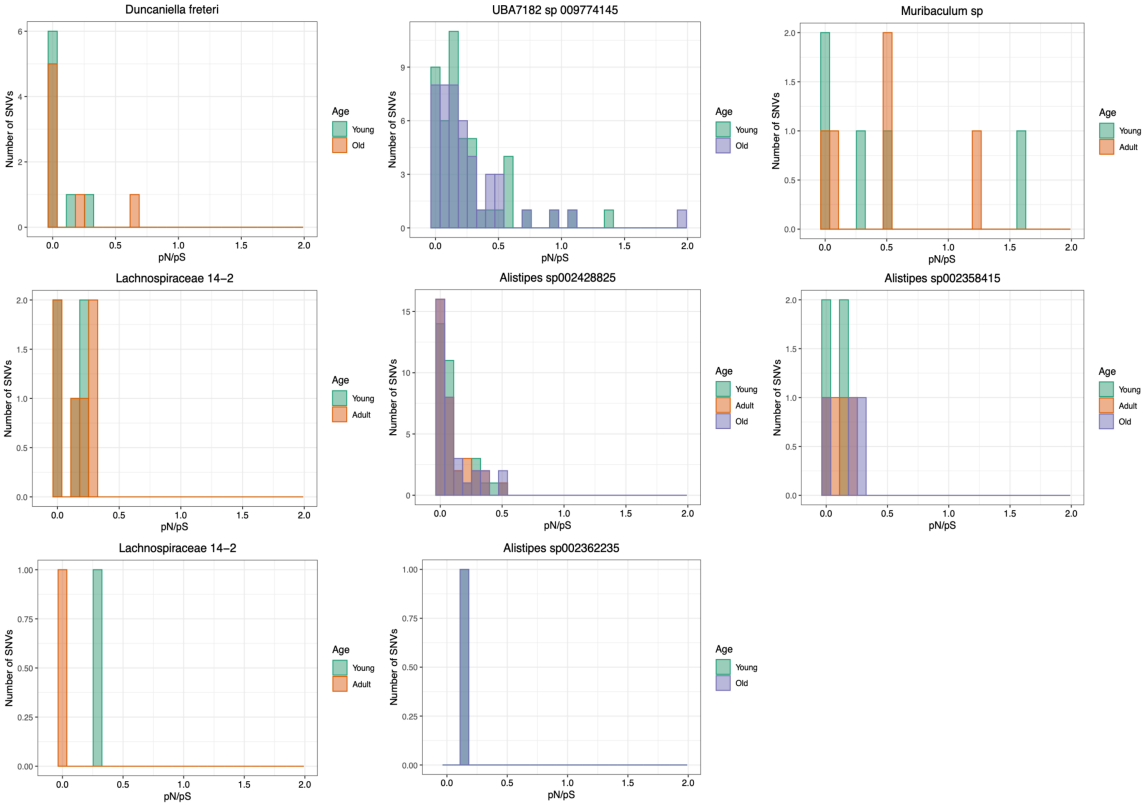


Figure 4.20: pN/pS ratio for SNV changes, consensus SNV and variant SNV changes in all single colonized microbial populations. As previously described by Garud et al., 2019, genetic diversity in gut microbial communities is strongly constrained by purifying selection on amino acid replacement (pN/pS <1). Interestingly, transposase genes in *Muribaculum sp.* and *BA7182 sp009774145* shows pN/pS > 1.

Finally, to examine whether SNVs and SNSs under natural selection affect bacterial fitness, I used a bioinformatic method to determine the bacterial fitness in metagenomic samples denominated index of replication (iRep) (83). Bacterial fitness is commonly evaluated as the potential of reproduction (Bacterial growth rate) (89). Bacterial replication proceeds bidirectionally with a single origin of replication (87). Therefore, copies of the genome accumulate during bacterial growth at the origin of replication (88). iRep index determines bacterial replication measuring the rate of sequencing coverage from

the origin to the terminus of replication (83). In a bacterial population in which all the cells replicate a single copy of their chromosome, iRep values will be equal to two. An index value of 1.5 indicates that half of the bacterial population is replicating (83). I hypothesized that bacterial species with signatures of positive or purifying selection in protein-coding genes would increase or maintain their iRep index over time.

I found that the replication rate decreased in *Muribaculum sp, Lachnospiraceae 14-2,* and *Alistipes sp002358415*, and remain stable in *UBA7182 sp009774145, Alistipes sp002428825,* and *Alistipes 002358415*. However, *Duncaniella freteri* population shows an increase in growth rate (Figure 4.21). A possible explanation of a decreased replication rate could involve the accumulation of weak-effect deleterious mutations linked to beneficial mutations (hitchhiking). Increasing or maintaining the replication rate could implicate the presence of beneficial mutations (210).

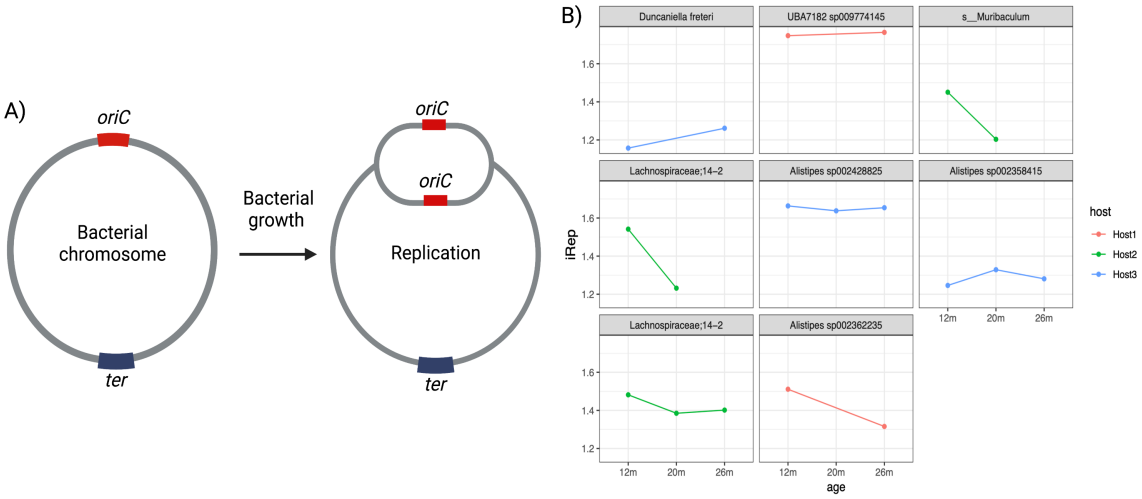


Figure 4.21: Replication rate (iRep) of the selected microbial population over host aging. A) Replication rate calculation relies on directional bacterial replication. Bacterial chromosomes have a single replication origin (oriC) from which DNA replication starts, proceeds in both directions and terminates at the ter region. B) iRep values of the single-colonized microbial population over host aging. Created with BioRender.com.

Section II: Characterization of the gut microbial community of the aging model *Nothobranchius furzeri* using short and long read metagenomics

I followed a well-established three peaks fecal DNA extraction protocol to extract HMW DNA from complex microbial samples (11). I pooled six stool samples of adult males (9 weeks old) to obtain a sample input of 30 mg. To measure the purity of the extracted HMW DNA from the pooled stool sample, I measure the absorbance of the sample at the wavelength of 230 nm, 260 nm, and 280 nm using the Nanodrop 2000c spectrophotometer. The extracted HMW-DNA was measured using the Qubit dsDNA BR assay kit (Table 4.10).

Sample	Quantification	260/280	260/230	DNA Integrity Number
Fish Stool	19.3 ng/µl	1.88	1.97	9.0

Table 4.10: Quality of HMW-DNA extracted from fish stool metagenome

Extracted HMW-DNA from a fresh stool sample showed an optimal fragment size distribution with a higher DNA Integrity Number, calculated using a Genomic DNA ScreenTape from Agilent Technologies (Figure 4.22). HMW-DNA extracted from fresh stool samples shows higher integrity than DNA extracted from mice's frozen-store stool samples. It is reported that freezing/thaw cycles promote DNA degradation, and larger DNA fragments are the most sensitive to these cycles (109). Therefore, I conclude that HMW-DNA extraction from fresh stool samples generates optimal DNA fragment size for downstream Nanopore library preparation and long-read sequencing.

High-throughput culture approach to isolate bacteria species from the killifish gut microbiota.

Culturomics consists of rapidly identifying bacteria by multiple culture conditions (110). Culturomics implements matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry and/or 16S rRNA sequencing to identify bacterial species (111). I performed the first culturomics approach to isolate, identify and create a collection of

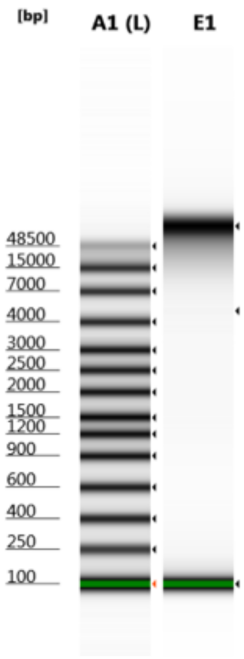


Figure 4.22: DNA fragment size distribution of HMW DNA extraction from fresh stool sample. Fresh fish stool DNA fragment distribution using a Genomic DNA Screen Tape. Extracted DNA from fresh stool material possess a high integrity and fragment size higher than >50 kb.

bacteria species in the turquoise killifish gut microbiota using 16S rRNA sequencing. Culturomics of the killifish gut consists of 25 different culturing conditions (Table 4.11), including two temperature conditions, three oxygen level conditions, and five nutrient-rich growth media. After three days of culturing, I examined phenotypically distinct bacteria colonies and counted the total number of colonies in each culturing plate (Figure 4.23). I observed the major number of distinct bacteria colonies under culturing conditions of normoxia at 28 ℃ in all media. Previous studies have demonstrated that facultative anaerobic, aerobic, and obligate anaerobic are the prevalent colonizers in the teleost fish gastrointestinal tract (112, 113), consistent with the number of distinct colonies observed in the culturomic approach under normoxia and hypoxia conditions. However, I observed a lower number in distinct colonies under anaerobic conditions. For anaerobic cultures, stool samples were first resuspended in a pre-reduced 1x PBS buffer and plated under hypoxia conditions before anaerobic culturing. Therefore, culturing protocol implemented in the culturomic approach could explain the lower number in anaerobic conditions.

Next, I isolated the phenotypically distinct bacterial colonies from each plate and ampli-

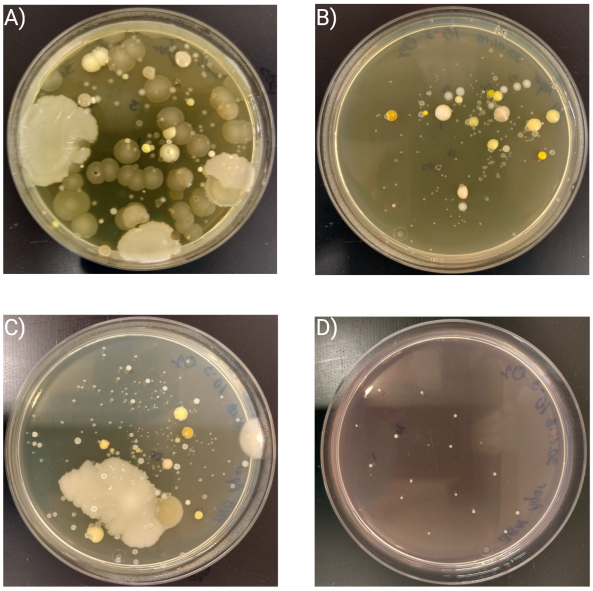


Figure 4.23: Culturing microbial species of the turquoise killifish gut in different media conditions at 28ºC in Normoxia conditions. A) Brain Heart Infusion; B) Lysogeny broth; C) Thioglycollate medium; D) Bryant and Burkey medium.

fied the hypervariable regions V3, V4, V5, and V6 from the 16S rRNA genes by colony PCR. Amplicons were sequenced using Sanger sequencing, and DNA sequences were aligned to the NCBI database for taxonomic classification (Figure 4.24). Bacterial isolates are members of the Halomonas, Bacillus, Shewanella, Vibrio, Carnobacterium, and Enterococcus genera (Table 4.12). Previous studies describe that microbial members of the phylum Proteobacteria, Firmicutes, and Bacteroidetes are the dominant intestinal microbiota in most teleost fish (112).

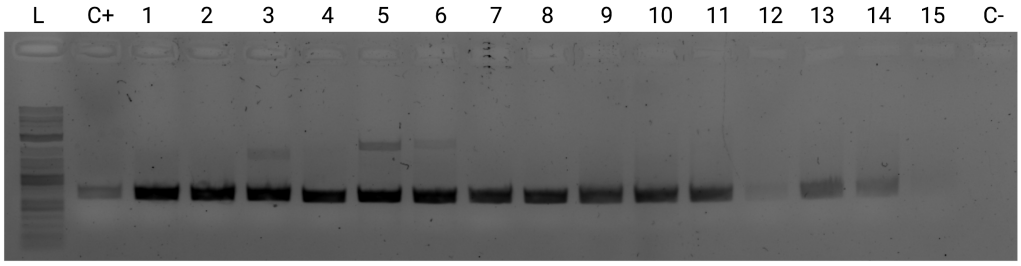


Figure 4.24: Colony PCR of unique colonies isolated from the turquoise killifish intestine. Fifteen different bacterial colonies were isolated and hypervariable regions of the 16S rRNA were amplified using the universal bacteria primers Bakt-341F and Bakt-1061R. L = Quick-Load 1 Kb Plus DNa Ladder; C+ = Positive control (Extracted bacterial DNA from *E. coli*); C- = Negative control (Water sample).

Next, to fully identify the bacterial species of the isolated colonies, I grew the bacterial isolates on single culture tubes with yeast extract, casitone, and fatty acid (YCFA) medium, a broad-range bacteriological medium used to culture bacteria from the human gut microbiota (12). I extracted HMW-DNA from culture media using the MagAttract HMW Kit and obtained optimal concentration and quality DNA for long-read sequencing (Table 4.13).

Nanopore sequencing and de novo genome assembly of killifish gut metagenome.

Long-read sequencing of the killifish gut metagenome was conducted using the ligation sequencing kit from Oxford Nanopore technologies and one R9.4.1 MinION flow cell. The whole raw sequencing of the killifish metagenome yielded 3.4 GB and generated a total number of 131,520 long-reads after 48 h of the sequencing run. I basecalled and filtered out reads with a quality score <7 using Guppy v3.6.1 (Figure 4.25). Next, Nanopore adapters were trimmed using Porechop v.0.2.4 and obtained a total number of reads of 73,805 and a final throughput of 976,472,653 bp.

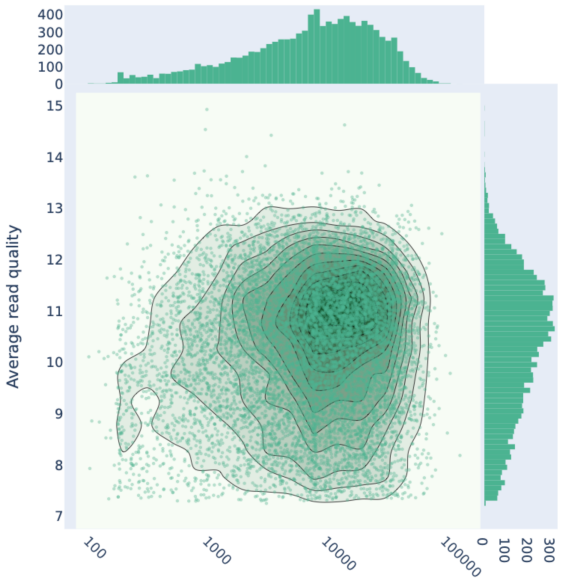


Figure 4.25: Size distribution of the nanopore sequencing reads of the killifish gut metagenome. Raw reads were trimmed and quality filtered before metagenome assembly. HMW-DNA extraction from fresh stool samples improves the extraction of long DNA fragments. Read length N50 was 24 Kb.

To get a comprehensive overview of the microbial community diversity present in the kil-

lifish metagenome, I used the long-read taxonomic classifier BugSeq, an accurate metagenomic classifier specific for Nanopore reads (130). BugSeq classified 78% of the long-read metagenomic. In comparison, 22% of the total reads were not assigned to any reference from the RefSeq database. For the classified reads, 90% of the metagenomic data belong to Proteobacteria phylum, 5% are members of the Bacteroidetes phylum, 3% Firmicutes, and 1 % Actinobacteria (Figure 4.26). This result is consistent with the previous characterization of the turquoise killifish gut raised under laboratory conditions (8). The Proteobacteria community comprises 14% of Alphaproteobacteria, 33% Betaproteobacteria, and 52% of Gammaproteobacteria. Notably, the five most abundant bacterial genera in the turquoise killifish gut comprehend Aeromonas, Shewanella, Pseudomonas, Undibacterium, and Ralstonia. Interestingly, BugSeq identified the resistance markers cphA2, ampS, blaOXA-252, and blaMOX-4 from Aeromonas and Klebsiella genera.

For *de novo* assembly of the fish metagenome and asset, whether long-read metagenomic sequencing can reconstruct complete close bacterial or viral genomes from the killifish microbiota, I employed the long-read metagenome assembler metaFlye (26). metaFlye is a state-of-the-art long-read assembler specialized in metagenome assembly capable of reconstructing complete or nearly complete bacterial genomes within single contigs of complex microbial communities (26).

MetaFlye successfully assembles large contigs and reconstructs complete circular contigs with the size of bacterial and viral genomes from the turquoise killifish gut (Figure 4.27). Remarkably, metaFlye reconstructed a 5.05 Mb circular contig, likely representing a complete bacterial genome, and a dozen of 30 - 60 Kb circular genomes, likely representing complete viral genomes or plasmids.

To taxonomic classify the reconstructed bacterial size circular bin, I used the GTDB-TK toolkit to assign taxonomy to the 5.05 Mb contig. GTDB-Tk classified the circular bin in the Proteobacteria phylum, Burkholderiaceae family, and *Undibacterium* genus. No species was assigned by GTDB-tk but identified the closest species, *Undibacterium squillarum*, with a 77.54% ANI value. The *Undibacterium* genus is characterized by Gram-staining-negative, oxidase-positive, rod-shaped, motile by a single polar flagellum and chemoheterotrophic cells (116, 117).

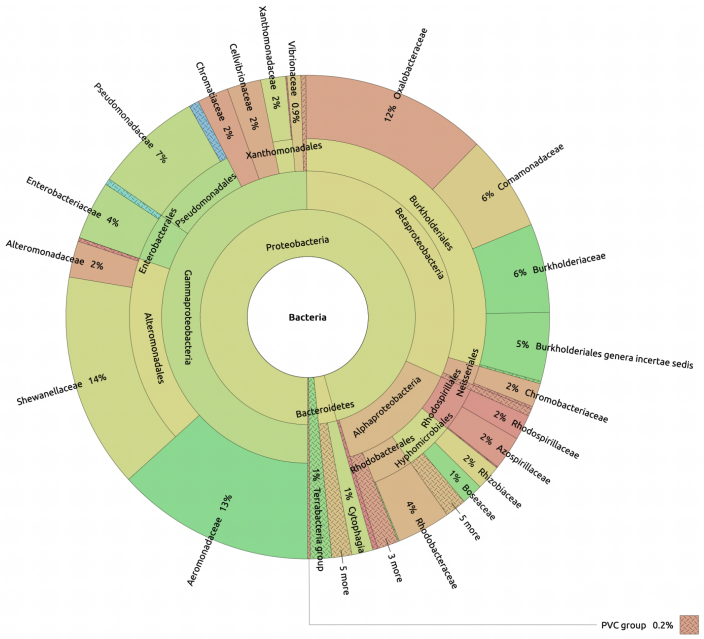


Figure 4.26: Taxonomic classification of the long-read Nanopore metagenome using BugSeq tool. Proteobacteria is the most prevalent bacteria phylum present in the killifish metagenome dominated by the Aeromonadaceae, Shewanellceaea, Oxalobacteraceae, and Pseudomonadaceae family.

Undibacterium bacterial strains have been isolated from freshwater samples, freshwater shrimp culture ponds, soil, permafrost soil, freshwater fish species *(Danio rerio)* (118), and intestinal tract of freshwater fish species (119). I examined the quality of the assembled *Undibacterium sp.* genome using the CheckM toolkit and classified it as a high-quality draft genome (completeness score of 91.43% and 1.46% as contamination score). Genome annotation of *Undibacterium sp.* identified 6,656 coding sequences, three CRISPR arrays with 37, 114, and 37 spacers, respectively, 12 rRNAs and 63 tRNA.

To identify the reconstruction of viral genomes or plasmids, I annotated all the circular contigs using the Basic Local Alignment Search Tool integrated into the bacterial bioinformatics Resource center PATRIC (120). PATRIC Annotation service identified two distinct contings (with a length of 39.5 Kb and 33.9 Kb contig size) with genes encoding for phage proteins, phage replication proteins, and phage structural proteins. The 39.5 Kb contig was annotated with 17 structural and DNA excision proteins belonging to the Bacteriophage Mu. The second contig annotated 49 structural proteins (capsid and tail proteins), DNA replication machinery, integrase, host lysis, and phage transcriptional

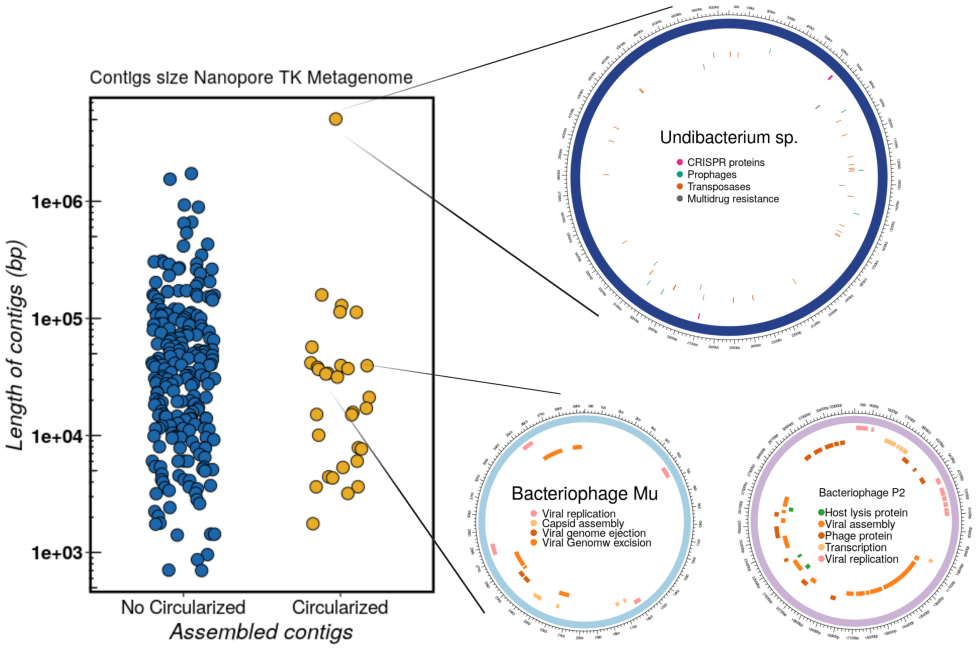


Figure 4.27: Size distribution of the contigs assembled by metaFlye. Contigs are divided in two groups: Circularized and no circularized contigs. No Circularized contigs shows a maximum contig size of 2 Mb. On the other hand, circularized assembled contigs show a circularized bacterial genome contig of 5Mb and viral genomes corresponding to bacteriophages.

regulator from Bacteriophage P2 (Supplementary Table 9 - Figure 29). Therefore, I conclude that long-read sequencing of the fish stool metagenome can provide high-quality bacterial and viral reference genomes, leveraging gene content and metabolic pathway analysis in this complex microbial community.

Short- and long-read sequencing and hybrid de novo assembly on cultured isolated from the killifish gut.

To evaluate the quality of the long-read sequencing from the bacterial isolates, I pooled the extracted HMW-DNA samples from each bacterial isolate and performed Nanopore sequencing using a flow cell R9.4.1 and the GridION sequencer (Figure 4.28). Nanopore sequencing generates 10.06 Gb with a total number of 1,582,012 long reads. I visualized the distribution of the long-reads using NanoPlot and observed that the average read size of the sequencing is 6.3 Kb with an average Phred score of 9.1 (Figure 4.29). Demultiplexing of the sequencing data was performed using Porechop by sorting sequencing

reads in separate directories for each bacterial isolate in the sample. Short-read sequencing generates 1.3 - 2.2 million reads per bacterial isolate. However, only five out of ten bacterial isolates successfully generate short-reads sequencing data (Table 4.14).

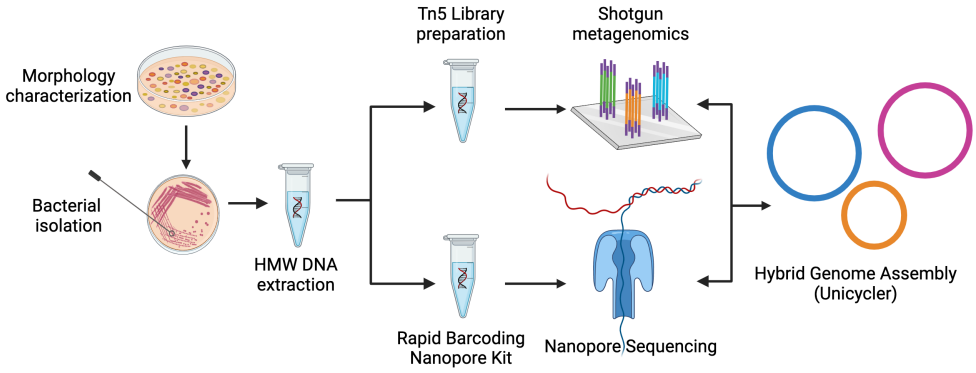


Figure 4.28: Hybrid Metagenomic approach for reconstruction of bacterial genomes. Culturomics detects distinctive bacterial colonies and HMW-DNA from distinctive isolates were extracted. HMW-DNA served as DNA input for Tn5 Illumina and Nanopore library preparation. Sequencing dataset was demultiplexed and hybrid genome assembly was performed using the Unicycler pipeline. Created with BioRender.com.

For bacterial isolates with only long-read sequencing data and hybrid assembly, I used the Unicycler assembler. Unicycler can perform hybrid assembly using Illumina and Nanopore read data and long-read only assembly. Hybrid assembly first generates a short-read assembly graph and uses long-read to build bridges between the assembly gaps and resolve repeat genomic regions, resulting in complete genome assemblies. Long-read only assembly undergoes multiple rounds of genome polishing to improve sequence accuracy (24). Unicycler generates complete circular genomes in all hybrid assemblies (Table 4.14). Notably, hybrid assemblies from bacterial isolates BC07 and BC08 reconstructed a circular bacterial genome and circular plasmids. On the other hand, the long-read only assembly could not reconstruct a complete circular and generate highly fragmented draft genomes.

To examine the quality of the generated draft genomes, I used CheckM to assess completeness and contamination scores in each assembly. Overall, all reconstructed genomes show completeness and contamination scores characteristics for high-quality draft genomes (>50% completeness and <10% contamination). However, hybrid assemblies show higher

completeness scores and genome contiguity (N50) and lower contamination scores compared with long-read only assemblies (Figure 4.30 - Supplementary Table 10). Interestingly, BC01, BC04, BC05, and BC06 show a high score of contamination (>10%), and the incorrect assignment of reads during the demultiplexing step could generate genome mis assemblies and contamination.

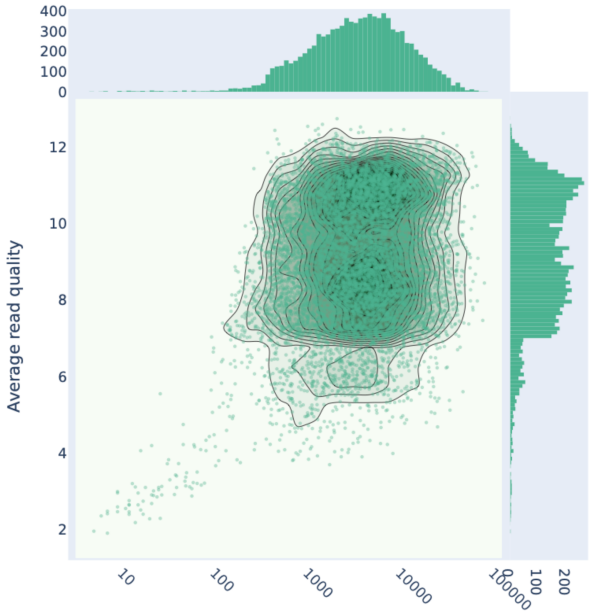


Figure 4.29: Long-read size distribution of Nanopore sequencing of HMW-DNA samples isolated from bacterial isolates. Read size distribution comprehens from 10^2 to 10^5. Average Phred score of read quality is 9.1.

Next, I assessed the taxonomic classification of each bacterial isolate using the GTDB-TK pipeline. Genomic classification places the bacterial genomes in Proteobacteria, Firmicutes, and Actinobacteria phylum (Supplementary Table 11). Bacterial isolates belong to Aeromonas, Enterococcus, Carnobacterium, Pseudomonas, Vibrio, Bacillus, Shewanella, Vagococcus, and Leucobacter genera (Table 4.15). Recently, opportunistic pathogenic genera, including Pseudomonas, Aeromonas, and Vibrio, have been identified as part of teleost's healthy gut microbiota community (112). Under healthy microbial community conditions, these pathogenic bacteria do not threaten the host. However, under stress conditions, these opportunistic pathogens can invade the body through wounds in the mucosa (112, 113).

On the other hand, *Bacillus* and *Shewanella* genera have been linked to beneficial effects within the host. Both genera have demonstrated positive effects in intestinal mucosal

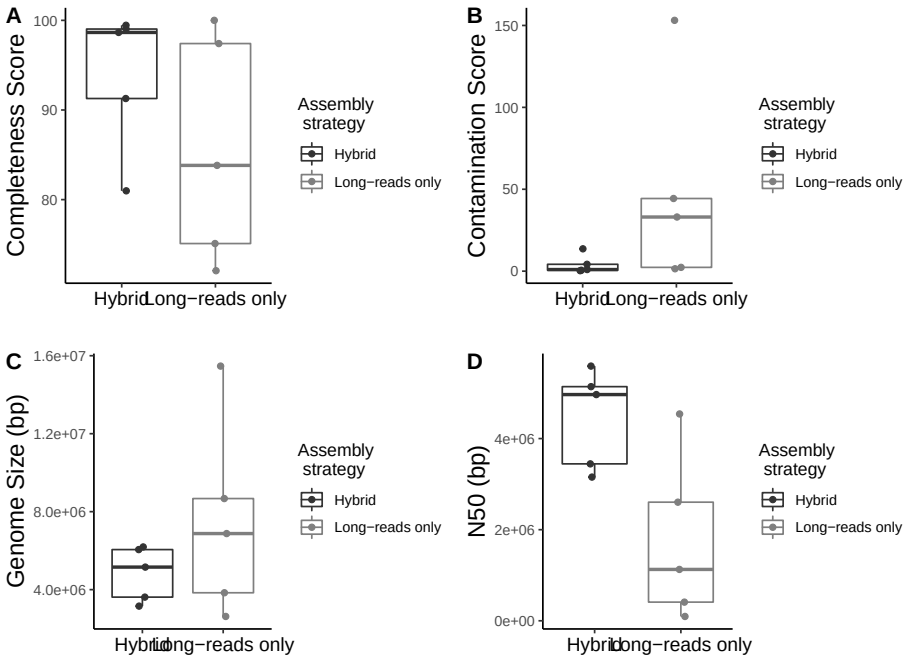


Figure 4.30: Quality and contiguity comparison between hybrid and long-reads only assembly of the isolated bacteria for the turquoise killifish gut. Draft genomes assembled using Illumina and Nanopore reads show higher completeness score and N50 contiguity and lower contamination score compared with genomes assembled using long-reads only.

cells and immune response stimulation when administered as probiotics in teleost fish (122). Significantly, the *Shewanella* genus serves as probiotics in farmed marine fish, protecting against skin infections by upregulating the gene expression of immune-related genes (123). Therefore, to better understand the potential role of these microbial members in the community, it is crucial to perform gene identification and genome annotation.

I used the PATRIC annotation pipeline to predict protein-coding sequences, tRNA, rRNA, and CRISPR locus and perform functional annotation of the predicted proteins (Figure 4.31). Notably, the PATRIC pipeline identifies prophage sequences, antimicrobial resistance genes, transposase elements, virulence factors, and host-modulators in the complete circularized genomes (Figure 4.32). The identification and annotation of these coding sequences are crucial to characterize the bacterial adaptation strategies in the intestine and influence the composition of the microbial community. For example, prophages (bacteriophages genomes integrated into the host genome) are considered beneficial for

the bacterial host (124) by bringing new traits that improve host colonization, inhibit other prophage infections, affect bacterial pathogenicity, and favor host stress resistance (124,125). On the other hand, transposable elements can affect gene regulation and function by disrupting the gene coding sequence or introducing promoter sequences (106). Therefore, this effort to annotate and characterize protein-coding sequences and mobile elements in novel gut bacteria species helps to expand our knowledge of these bacterial species' genetic repertoire, adaptation strategies, and the emergence of genetic variants associated with microbial function and host health (1).

Temperature	Oxygen level	Medium	Distinct colonies	Total No. colonies
28°C	Anaerobic	BHI	-	-
28°C	Normoxia	BHI	7	>100
28°C	Hypoxia	BHI	5	>100
37°C	Normoxia	BHI	3	>100
37°C	Anaerobic	BHI	2	>100
28°C	Anaerobic	Th	1	>100
28°C	Normoxia	Th	8	>100
28°C	Hypoxia	Th	6	>100
37°C	Normoxia	Th	2	>100
37°C	Anaerobic	Th	2	>100
28°C	Anaerobic	LB	1	>100
28°C	Normoxia	LB	7	>100
28°C	Hypoxia	LB	4	>100
37°C	Normoxia	LB	2	>100
37°C	Anaerobic	LB	2	>100
28°C	Anaerobic	Sch	1	>100
28°C	Normoxia	Sch	6	>100
28°C	Hypoxia	Sch	5	>100
37°C	Normoxia	Sch	2	>100
37°C	Anaerobic	Sch	2	>100
28°C	Anaerobic	BBM	-	-
28°C	Normoxia	BBM	2	18
28°C	Hypoxia	BBM	1	23
37°C	Normoxia	BBM	1	>100
37°C	Anaerobic	BBM	2	>100

Table 4.11: Number of phenotypically distinct bacterial colonies and the total number of bacteria per plate per condition. BHI = Brain heart infusion, Th = Thioglycollate medium; LB = Lysogeny broth; Schaedler = Schaedler medium; BBM = Bryant and Burkey medium.

Bacterial isolate	Length of the amplicon sequenced	Genera	Phylum
1	717bp	Halomona	Proteobacteria
2	718bp	Enterococcus	Firmicutes
3	722bp	Vagococcus	Firmicutes
4	722bp	Bacillus	Firmicutes
5	717bp	Shewanella	Proteobacteria
6	718bp	Halomona	Proteobacteria
7	716bp	Vibrio	Proteobacteria
8	719bp	Shewanella	Proteobacteria
9	718bp	Halomona	Proteobacteria
10	697bp	Pseudomonas	Proteobacteria
11	698bp	Carnobacterium	Firmicutes
12	711bp	Leucobacter	Actinobacteria
13	706bp	Aeromonas	Proteobacteria
14	699bp	Halomona	Proteobacteria
15	710bp	Vibrio	Proteobacteria

Table 4.12: Genera of the isolated bacterial species from the turquoise killifish gut

Bacterial	DNA concentration	260/280	260/230
Shewanella	80 ng/µl	1.68	2.02
Vagococcus	240 ng/µl	1.85	2.14
Vibrio	644 ng/µl	1.78	1.62
Carnobacterium	1427 ng/µl	1.81	1.61
Halomonas	1200 ng/µl	1.82	1.85
Leucobacter	10 ng/µl	1.58	1.36
Bacillus	860 ng/µl	1.66	1.9
Enterococcus	402 ng/µl	1.82	1.94
Pseudomonas	240 ng/µl	1.65	1.78
Aeromonas	324 ng/µl	1.84	1.79

Table 4.13: Concentration and quality of extracted HMW-DNA from isolated bacteria

Bacterial isolate	Long-reads	Short-reads(M)	genome size	Largest contig	Contigs	Circularized?
BC01	51,519	-	8.6 Mb	4.54 Mb	103	N
BC02	62,137	-	2.6 Mb	2.603 Mb	3	N
BC03	51,991	1.36	3.88 Mb	3.49 Mb	3	Y
BC04	21,175	-	6.87 Mb	0.3 Mb	93	N
BC05	120,474	2.13	6.18 Mb	4.965 Mb	21	Y
BC06	11,156	-	15.46 Mb	2.85 Mb	96	Y
BC07	143,137	1.42	6.08 Mb	5.59 Mb	7	Y
BC08	210,480	2.26	5.13 Mb	5.13 Mb	2	Y
BC09	67,248	1.51	3.15 Mb	3.15 Mb	1	Y
BC10	35,165	-	3.83 Mb	0.66 Mb	52	N

Table 4.14: Short- and long-read sequencing output and *de novo* assembly of the ten bacterial isolates from the turquoise killifish intestine.

Bacterial isolate	Taxonomic classification by GTDB-TK
BC01	*Aeromonas veronii*
BC02	*Enterococcus aquimarinus*
BC03	*Carnobacterium maltaromaticum*
BC04	*Pseudomonas lundensis*
BC05	*Pseudomonas lundensis*
BC06	*Vibrio metschnikovii*
BC07	*Bacillus thuringiensis*
BC08	*Shewanella baltica*
BC09	*Vagococcus salmoninarum*
BC010	*Leucobacter luti*

Table 4.15: Taxonomic classification of the ten bacterial isolates from the turquoise killifish gut.

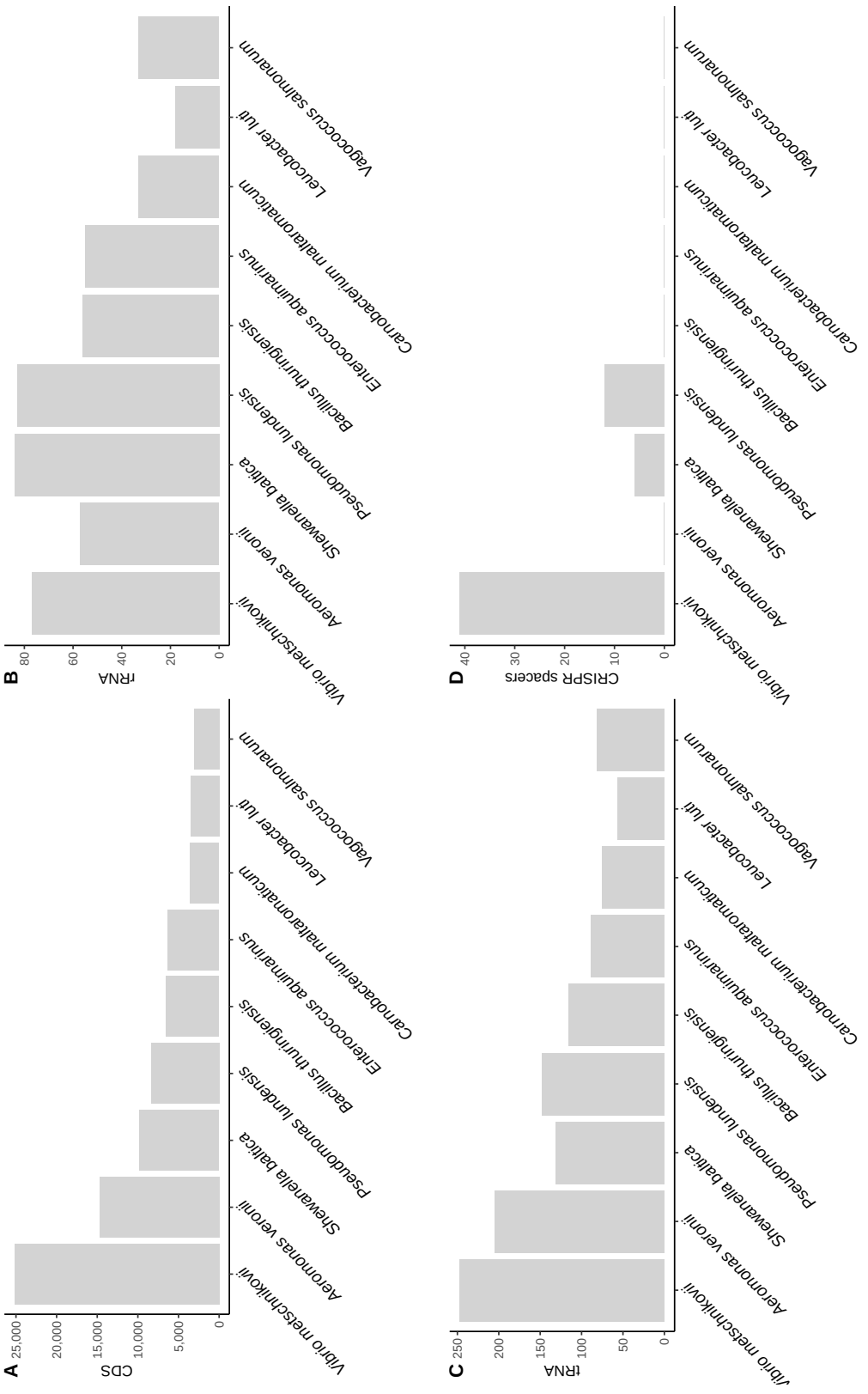


Figure 4.31: Number of detected coding sequences. tRNAs, tRNAs, and CRISPR spacers in each bacterial isolate of the killifish gut.

Genetic modification of gut microbial isolates using a bacterial conjugation protocol

Characterizing the biology between host cells and commensal bacteria, inter-species microbial interaction, and the spatial distribution of bacterial communities along the gut axes is critical to understanding the function and effect of the gut microbiota in host development, immune activation, and health and disease states (126). Therefore, precise genetic manipulation tools are required for studying the function and effect of a particular commensal member on the microbial community and host physiology (127). Stable chromosomal insertions of fluorescent proteins allow detailed microscopy and flow cytometry analysis to study cellular interactions within host tissue or other gut commensals (22). I implemented an established conjugation method to reliably deliver DNA into bacterial cells followed by a single chromosomal integration by homologous recombination using the Tn7 transposon-based chromosomal insertion system (22, 128). Tn7 transposition method allows the site-specific integration of genes in a neutral evolved site in the bacterial chromosome without compromising the fitness of the host cell and removing the necessity of continued antibiotic selection, allowing studies in complex host-microbial communities such as the intestine.

I followed the detailed tagging protocol reported by Wiles et al. using the pTNS2 helper plasmid and the pUC18T-mini-Tn7T-Gm-eyfp as a delivery plasmid. Helper plasmid contains the encoding site-specific transposition pathway TnsABCD. In contrast, the delivery plasmid contains the Tn7L and Tn7R end integration sites, the gene encoding the enhanced yellow fluorescent protein (eyfp), and the gentamicin (GmR) resistance selection marker (21, 22, 128) (Figure 4.33). I used *E. coli SM10* strains as cellular containers due to their specialized function for mobilization and plasmid transfer to target bacterial cells. Dr. Guillemin's lab from the Institute of Molecular Biology at the University of Oregon donated the *E. coli SM10* strain.

The implementation of eyfp as an integration marker helps to target cells that are effectively recombined using the Tn7 system. I implemented the conjugation and Tn7 recombination system into the *Pseudomonas lundensis* isolate from the turquoise killifish gut due to their high prevalence in the gut microbiota community (8) and the previously re-

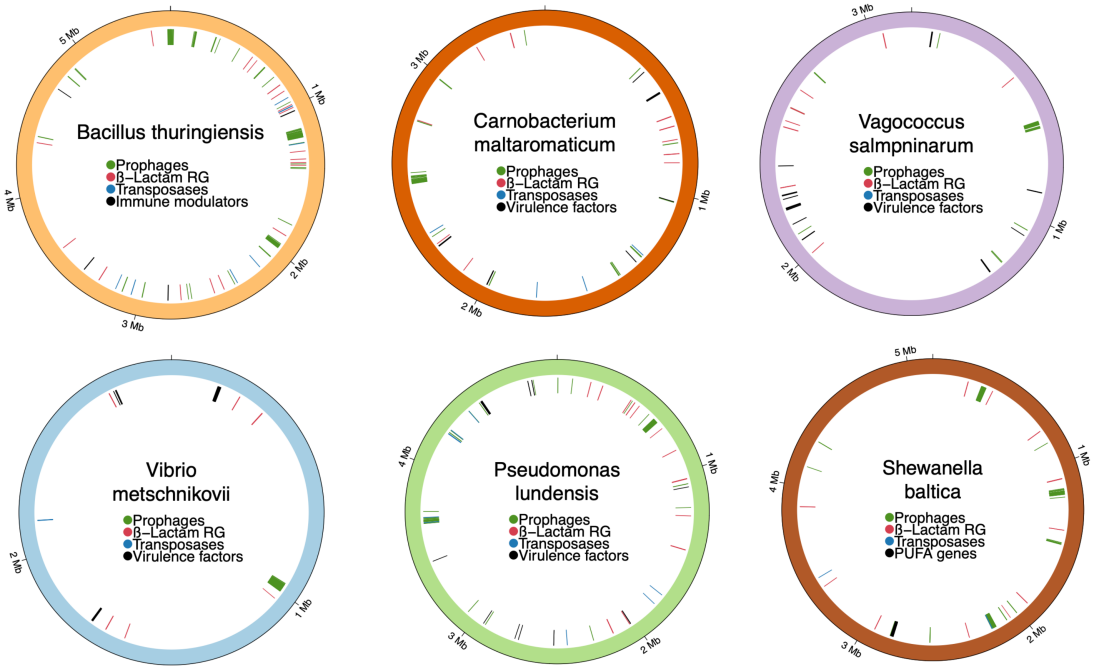


Figure 4.32: Circular plots of close reconstructed genomes from hybrid and long-read only assembly. In all plots, the other rings represent the complete, closed and circularized genome of the given bacteria species. The inner ring represents annotated genes with a potential host-specific adaptation role.

ported effective genomic integration of the *Pseudomonas* genus isolated from zebrafish intestinal microbiota (22). During the mating or conjugation reaction, I combined two *E. coli SM10* carrying the helper and delivery plasmid and incubated the mating reaction overnight to increase the number of modified *Pseudomonas lundensis* cells. After mating, the selection of genetically modified cells relies on transferring the cellular mixture to a gentamicin media. However, since *E. coli SM10* that carries the pUC18T-mini-Tn7T-Gm-eyfp plasmid could grow under this selective media, I implemented a double selective media condition. I screen for the resistance of *Pseudomonas lundensis* and *E. coli SM10*, carrying the delivery plasmid to a broad spectrum of antibiotics (Table 4.16). I identified that *E. coli SM10* carrying the delivery plasmid is sensitive to Spectinomycin and Trimetropin antibiotics, while *Pseudomonas lundensis* presents a resistant phenotype to these antibiotics.

Therefore, I implemented a double antibiotic selection method that reliably selects for targeting bacterial cells with the presence of the delivery plasmid. I screened multiple colonies using a fluorescent stereomicroscope, and eyfp-positive colonies were isolated

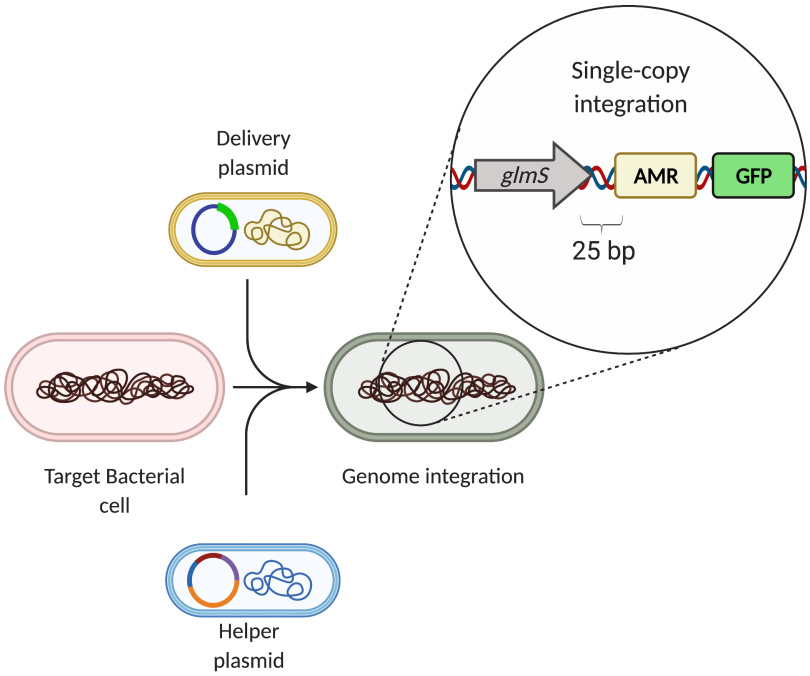


Figure 4.33: Overview of the Tn7 integration system and conjugation protocol. Helper and delivery plasmids are transferred into the recipient cell before conjugation with the target bacteria cells. Insertion of the fluorescent protein and the antibiotic resistance selection marker occurs 25 nucleotides downstream of glmS gene (Adapted from Choi & Schweizer, 2006). Created with BioRender.com.

to a new bacterial plate. However, one problem can arise with this methodology. Fluorescent *Pseudomonas* cells can only contain the delivery plasmid and present the same phenotype as those cells with successful chromosome integration of the genetic markers. Therefore, I purified the plasmid-based markers phenotype by liquid culturing and plating all fluorescent and antibiotic resistance isolates to a non-selective media. Under these conditions, isolates with no chromosome integration lost fluorescent positive phenotypes. Only target cells with the successful integration of the marker genes conserve the fluorescent and antibiotic resistance markers.

Next, I performed genotyping by colony PCR of six fluorescent-positive bacterial isolates to confirm the integration of the fluorescent and antimicrobial marker downstream of the chromosomal gene gmlS. The *Pseudomonas lundensis* genotype implements the first PCR for amplifying the glmS gene to confirm the presence of the glmS gene in the bacterial isolate and that the primer sequences are specific for this gene (Fw gmlS *Pseudomonas* + Rv gmlS *Pseudomonas*). The second PCR amplifies the start of the glmS gene + 100

Antibiotic	*Escherichia coli*	*Pseudomonas lundensis*
Spectinomycin	N	Y
Gentamicin	Y	N
Trimetropin	x	Y
Neomicin	Y	N

Table 4.16: Broad spectrum antibiotic screening in E. coli SM10 and Pseudomonas lundensis isolate.

bp of the intergenic region at the end of the glmS gene (Fw gmlS *Pseudomonas* + Rv Intergenic *Pseudomonas*). The second PCR confirms that no external DNA sequence integrates downstream the glmS gene. Finally, a third genotype PCR amplifies the glmS gene and the Tn7R integration region (Fw glmS *Pseudomonas* + Rv GnR Tn7) (Figure 4.34 A).

I observed that all the tested isolates showed the genotype pattern indicative for fluorescent and antimicrobial chromosomal integration downstream of the gmlS gene (Figure 4.34 B - D). All positive colonies were cultured on 5 mL of LB media with Gentamicin, and 1.5 mL glycerol stock of each isolate was stored at -80°C. The generation of this novel *Pseudomonas lundensis* strain isolated demonstrates the feasibility of the generation of knock-in strain in novel and uncharacterized bacterial species of complex microbial communities. A particular interest will be developing strategies that implement fluorescent commensal bacteria to study microbial behavior in the host intestine, the distribution of different gut commensals across the gut axes, and tracking evolutionary changes of commensal using lineage tracking technologies (129).

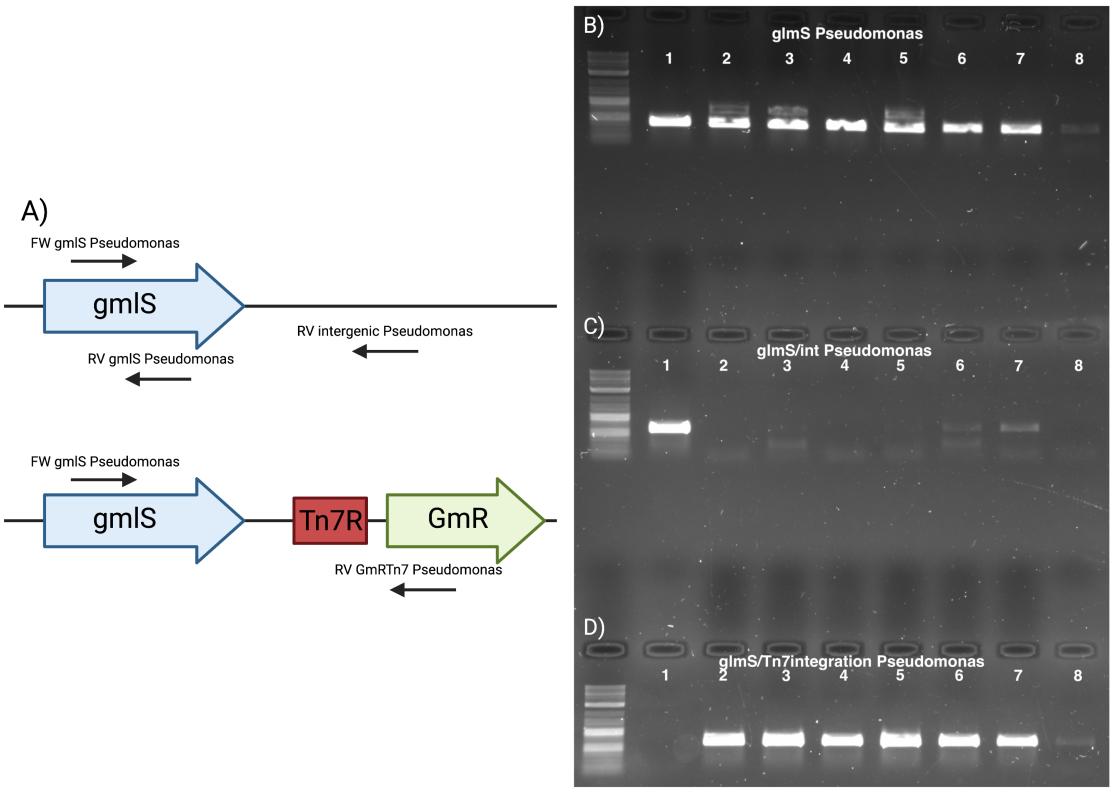


Figure 4.34: Genotyping for confirmation of marker integration in six different bacterial isolates of *Pseudomonas lundensis*. A) Graphical representation of *Pseudomonas genotyping* assay; B) Genotyping PCR for detecting the presence of the gmlS gene; C) PCR for intergenic region of glmS gene is only present in isolates without integration by the TN7 system; D) PCR for detection of integration of fluorescent and antibiotic markers downstream the gmlS gene. 1 - *Pseudomonas lundensis* wt, 2 - 7 *Pseudomonas* isolates with fluorescent positive and GmR phenotype, 8 - Negative control. Created with BioRender.com.

Chapter 5

Discussion

This thesis examined two distinct approaches to studying the gut microbiome in two different model organisms. First, I asked whether the targets of selection in the gut microbiome during host aging are different from short-term evolutionary changes. Second, I characterized the stool microbiota community of the novel aging model, the turquoise killifish Nothobranchius furzeri, implementing long-read metagenomics technologies, creating the first bacterial collection and reference genomes of these novel bacterial strains.

Section I: Detecting evolution of the commensal bacteria in the timescale of host life using longitudinal hybrid metagenomic sequencing

Short-read metagenome assembly strategy creates a comprehensive reference genome dataset.

The creation of a comprehensive, high-quality host-specific reference metagenome catalog represents a valuable resource for accurate characterization of the taxonomic and functional repertoire of the gut microbial community (70, 131) and a crucial tool for determining the evolutionary changes that occur in the community over time using reference-based metagenomic read mapping strategies (36, 66). Using three different metagenomic sequencing technologies, I examined the quality of the reconstructed metagenome catalog by each metagenomic strategy (Figure 4.5 - 4.6). I found that host-specific reference

genomes assembled using short-read sequencing data generate more contiguous and reliable bacterial genomes than those created using barcoded short-read libraries (linked-read sequencing) and hybrid assemblies (long-read + short-read sequencing data).

These results contradict previous reports that evaluated the impact of different metagenome assemblies strategies of complex microbial environments, supporting hybrid and barcoded short-read assemblies as reliable strategies for generating accurate, high-quality metagenomic assemblies. Moreover, hybrid and barcoded short-read assemblies outperform short-reads strategies enabling more contiguous assemblies (>200 Kb N50) and generating complete close bacterial, plasmid, or phage sequences (28, 60, 132).

Two factors can explain the low performance of hybrid assemblies for generating high-quality draft genomes. One factor may be the low throughput generated during long-read sequencing compared to the short-read metagenomic sequencing. Prior observations suggest that assembly of low coverage species is more likely to overlap reads and incorporate them into chimeric contigs in hybrid assemblers (132). The second influencing factor relies on the sequencing of pooling metagenomic samples to increase the coverage of low abundant species for improving draft genome assembly. However, pooling metagenomic samples could increase community and strain diversity, reducing the assembly performance of complex microbial strains (114, 115).

On the other hand, low performance by barcoded short-read assemblies could rely on utilizing the recently created TurinAssembler and the barcode deconvolution problem inherent in read-clod sequencing. TurinAssembler is a recently developed microbial genome assembler specialized for Tell-Seq data (19). However, even though TuringAssembler successfully assembles microbial isolates genomes, it is not optimized for assembling complex microbial communities like the mouse microbiota. Therefore, implementing metagenomic-aware barcoded short-read assemblers (Athena, cloudSPADES (60, 133)) is critical for leveraging the longitudinal linked-read sequencing data generated in the present work. A second limitation for the linked-read assembly was previously described as the barcode deconvolution problem (134, 135). The barcode deconvolution problem occurs when two or more genomic segments originating from different hosts are tagged with the same 3' barcode and are assigned within the same genomic subgroup. However,

multiple computational tools have been developed to tackle the barcode deconvolution problem (134, 135, 136). Therefore, implementing a robust barcode-aware metagenome assembler combined with a read deconvolution tool will leverage the long-range linked information and short-read sequencing accuracy to generate more contiguous and accurate reference metagenome catalogs from the generated data.

Host-specific reference genomes catalogs allow the detection of individual genomic variants and evolutionary changes across host life

The three generated host-specific reference catalogs in the present study allowed the detection and tracking of novel genetic variants in the microbial community throughout the host life using state-of-the-art metagenomic variant caller inStrain (36). inStrain profiles intra-population genetic diversity by mapping the short-read metagenomic samples to the reference genome catalog. The microdiversity analysis reveals that single nucleotide substitutions (SNS) in the whole metagenome increase over time (Figure 4.9). The number of SNS can increase from a combination of ecological dynamics (strain replacement) and evolutionary modification (adaptive mutations)(3,5). Microbiome ecological changes refer to those changes in species or strains composition over time. Although it is crucial to thoroughly understand the ecological dynamics of the gut microbiota over the host life and their effect on host physiology, I focused on determining the evolutionary changes at the nucleotide level that may be relevant for gut bacterial adaptation over host aging.

To confidently distinguish genetic variant changes due to evolutionary processes from fluctuation in strain frequency, I leverage the quasi-phasing approach developed by Garud et al. (5, 46). Quasi-phasing established that a single evolving population is expected to generate a distribution of allele frequencies dominated by a single peak at high frequency (>80%) with low rates of intermediate-frequency polymorphism (5,46) with a minimum genome coverage of 20x (Figure 4.10). Therefore, I selected eight different microbial populations that showed single-colonized allele frequency behavior across host life based on these criteria.

However, single-colonized allele frequency distribution can also be generated via complete strain replacement by a more-adapted bacterial lineage and becoming the dominant strain in the population (4,5). To confirm that the microbial populations are not affected by

co-infection and replaced by novel environmental bacterial strains, I leveraged the novel popANI metrics calculated by the inStrain pipeline (36). Overall, popANI performs a community-aware comparison between two metagenomic samples. popANI compares genetic variants between the two samples and computes the similarity percentage. To determine that the two microbial populations are identical, I follow inStrain stringent popANI threshold of >= 99.999%. popANI analysis indicates that bacteria populations can stably colonize the host over the host lifespan. This finding is in line with findings reported by Yaffe et al., which demonstrated that microbial strains could persist over 10 years in the human gut (159) (Supplementary Table 4).

Nucleotide diversity of coding protein genes decreases during host aging.

In a previous publication (7), I hypothesized that the decrease in the microbial taxonomic diversity during aging leads to larger population size for a few commensal species, increasing the opportunity for the evolution of more adapted microbial strains that sweep and dominate the microbial population. As a result, the microbial population reduces genetic heterogeneity, experiencing a gene-wide selective sweep (137). Next, I tested this hypothesis and examined whether single-colonized microbial populations experience a reduced intra-species microdiversity over time. My results revealed that six of the eight selected single-colonized microbial populations experienced a significant change in microdiversity profile in protein-coding sequences (14). Moreover, *Muribaculum sp, Lachnospiraceae 14-2*, and *Alistipes sp002358415* significantly decrease coding sequences' genetic diversity, representing supporting data to my proposed hypothesis. Overall, my findings revealed a reduction in the genomic diversity of the protein-coding sequences over host aging (Figure 4.12). These findings might indicate the presence of a single evolved dominant allele that confers bacterial adaptation to harsh environmental conditions (137).

SusD outer-membrane protein shows a rapid increase in SNV frequency during host aging

Adaptive evolution can be detected by identifying SNV that rapidly increase in frequency in the population (4). Therefore, I tracked SNV frequencies to identify genetic variants that increased in frequency in the populations during host aging. Interestingly, I identified

that SNVs that rapidly increase in frequency in the population (<=0.2 to >=0.8, SNV change) (3,5) are located in the genes involved in the starch utilization system family and processing complex glycans (susD) and glycoside hydrolase family protein (4). SusC and SusD family genes have been previously reported as top-ranking genes that undergo parallel adaptive evolution in gut commensal from the human microbiome and infant microbiome (3,4) (Figure 4.19 - Supplementary Figure 6).

SusC/SusD encodes outer-membrane polysaccharide importers necessary to convert extracellular polysaccharides into cellular monosaccharides (138), and variations in these genes may be an adaptation to host-derived glycans (139). Moreover, it is hypothesized that mutations on these cell surface proteins may reflect an adaptive strategy for IgA-mediated colonization to the host (4), evasion of host immune processes (140), or phage predation (4).

On the other hand, the OmpA family is characterized by heat-modifiable, surface-exposed proteins present in high-copy numbers in the outer membrane of mainly Gram-negative bacteria. OmpA proteins have a substantial role in bacterial adhesion, invasion, and evasion of host immune defenses in pathogenic bacteria (146).

Finally, Glycoside hydrolase family 43 proteins are among the most abundant carbohydrate-active enzymes present in the human microbiome, with an essential role in digesting complex polysaccharides (100), and adaptations on these genes might be related to the degradation of specific host-derived glycans. Overall, my findings indicate that adaptive mutations on the susD outer-protein family and enzymes involved in the degradation of complex glycans are in line with previous human microbiome evolution studies and suggest that these genes are essential for long-term host adaptation in the mammalian gut.

SNVs in transposase genes show positive selection signatures

To investigate whether age-dependent onset of SNVs reflect adaptive events, I examined pN/pS values computed by inStrain of each gene with variants at intermediate frequencies. pN/pS is the ratio between the number of non-synonymous mutations per non-synonymous sites (pN) and the number of synonymous mutations per synonymous sites (pS). pN/pS denotes signatures of selection in polymorphisms segregating within a pop-

ulation. Purifying selection is suggested when pN/pS ratio <1, and positive selection is inferred when pN/pS value is greater than 1 (86). My results found that most SNVs have pN/pS values <0.5, reflecting that purifying selection acts on these polymorphisms and purges deleterious variants from the population (Figure 4.20). Notably, SNV present in the SusD family and glycoside hydrolase family proteins do not have a pN/pS value since calculating pN/pS of a single non-synonymous SNV becomes trivial. However, I found SNVs with pN/pS values >1, indicating that they are potentially adaptive (Figure 4.20 - *Muribaculum sp*. and *UBA7182* genomes).

Interestingly, these genetic variations are located in transposase genes (Supplementary Table 14). Transposases have been suggested to play an essential role in bacterial genome plasticity and adaptation in response to environmental perturbations (142). Transposases can influence bacterial adaptation by altering the expression of host genes and facilitating genomic rearrangements and gene duplications (142, 143). Moreover, a previous report in the cyanobacteria *Crocosphaera watsonii WH8501* found positive selection signatures in transposase genes (141). Mess and Doeleman suggest that positive selection signatures in transposase genes may be linked to controlling the transposition rate within the host by adjusting the number of gene variants in transposase genes (141). Therefore, my work proposes that transposase genes could play a crucial role in long-term adaptation to the mammalian gut.

Finally, I examined whether SNVs with adaptive signatures affect bacterial fitness. The most common model for evaluating bacterial fitness measures reproductive potential throughout bacterial growth rate (89). The calculation of the bacterial growth rate in metagenomic samples relies on the sequencing coverage from the origin and the replication terminus (83). My results indicate that the bacterial growth rate in most single-colonized microbial populations decreases (Figure 4.21 - *Muribaculum sp.*) or remains stable (*Alistipes 002358415*) during host aging. A possible explanation for the decrease in growth rate can be connected to interspecies interactions or new associations with novel colonized bacteria strains (144).

In contrast, microbial populations with a stable growth rate may align with the Red Queen hypothesis, which states that species require ongoing adaptation to persist in the commu-

nity (145). Interestingly, *Duncaniella freteri* population showed an increase in bacterial growth rate over host aging. One possible explanation for this result could be explained by the presence of a genetic variant in a helix-turn-helix domain-containing protein. Helix-turn-helix is a central structural motif capable of binding DNA. Proteins containing helix-turn-helix motifs are involved primarily in DNA replication and repair and transcription regulation (147). Therefore, genetic variation in helix-turn-helix motifs might influence the DNA-binding properties resulting in repression or activation of DNA replication or transcription (147, 148).

Characterization of the gut microbial community of the aging model *Nothobranchius furzeri* using short and long read metagenomics.

Long-read metagenomic sequencing assembles complete close bacterial and viral genomes from the turquoise killifish gut microbiome.

De novo generation of microbial genomes from the gut microbiota using short-read metagenomics allows the taxonomic and functional characterization of novel species present in complex microbial communities, expanding our understanding of the microbial world within us (62, 131). However, short-read metagenomics assembly yields incomplete, highly fragmented, and chimeric draft genomes and often fails to recover plasmid or viral genomes due to their size limitation in resolving highly repetitive genomes (62, 149). Therefore, applying long-read sequencing technologies to complex microbial communities can dramatically improve the quality of metagenomes assembly and generate single-contig complete bacterial genomes (16). Using pooled high molecular weight DNA samples, I sequenced the gut microbiome of the novel aging model organism, the turquoise killifish using Nanopore sequencing technologies. My results revealed that performing high molecular weight DNA extraction from fresh-recovered stool samples generates >48 Kb DNA fragments for the gut microbiome (Figure 24), which generates a median read length >13 Kb and reads >110 Kb (Figure 4.25).

Taxonomic classification of the microbial community using long-read and highly accurate metagenomic classifier (130) identified Proteobacteria as the main prevalent phylum

of the killifish microbiome, followed by Bacteroidetes phylum (Figure 4.26). These results are in line with the previous characterization of the turquoise killifish gut using 16S rRNA gene sequencing old wild populations and laboratory fish (8). Moreover, the most dominant bacterial families, Aeromonadaceae, Shewanellaceae, Oxalobacteraceae, and Pseudomonadaceae, constitute 46% of the total metagenomic reads. However, genome assembly is required to create host-associated genomes collections to understand better the microbial diversity present in the killifish gut and elucidate their ecological roles and interactions of these gut commensals with the host environment (63).

The metaFlye assembler reconstructed 31 contiguous circular assemblies, including a high-quality complete bacterial genome (>5 Mb) and two phage genomes (>30 Kb)(Figure 4.27). Classification of the circular genomes placed the bacteria genome in Undibacterium genera with no specific species identification. The two viral genomes show phages specific-proteins attributed to Bacteriophage Mu and P2 (Supplementary Table 9). Overall, my results demonstrate the potential of long-read metagenomics to reconstruct high-quality and complete genomes from gut commensal. Remarkably, the functional characterization of novel microbial species opens the investigation of microbial function, microbial evolution, and microbiome-editing interventions(150, 151) to elucidate the role of the microbiome in regulating host lifespan and disease (8).

Cultivation and isolation of the killifish microbiota identify host-associated bacteria species and generate reference genomes catalog.

The generation of a genome catalog of commensal bacteria from the turquoise killifish would afford to get insights into the ecology of host-associated microbes, mining for genes of interest (Antimicrobials genes or CRISPR-Cas) and metabolic pathways, and create reference genomes for comparative and evolutionary studies (66, 131, 151). Therefore, I implemented the first high-throughput culturing approach coupled with 16S sequencing on commensal bacteria from the killifish gut to create high-quality bacterial reference genomes. Next, I isolated ten phenotypically distinct bacteria colonies and sequenced their genomes using long-read and/or short read sequencing technologies (Figure 4.28). My results indicate that *de novo* genome assembly using both short- and long-reads (hybrid assembly) outperforms long-reads-only assembly (Figure 4.30). Overall, hybrid

assembled genomes show higher completeness and genome contiguity than assemblies using only long-read data. Remarkably, hybrid assembly generates complete, closed bacterial genomes in all bacterial species. For example, *Bacillus thuringiensis* hybrid assembly successfully reconstructs six circular contigs that are likely to represent bacterial plasmids. Finally, annotation of the closed bacterial genomes identified genes responsible for antimicrobial resistance, virulence factors, CRISPR arrays, hots-immune modulators, and transposase genes that have been described as potential factors for host-microbial adaptation (124, 125, 143).

Tn7 transposon-mediated system successfully integrates genetic markers in gut commensal chromosomes

Chromosomal modification of the gut commensal bacteria will examine microbial genomes involved in host-microbe and microbe-microbe interactions (150). Therefore, I implemented the Tn7 transposase-mediated integration system to knock-in genetic markers for the first time in gut commensal bacteria species from the turquoise killifish gut. My results indicate that the Tn7 system efficiently integrates a fluorescent protein reporter and antimicrobial resistance marker downstream of the ubiquitous glmS gene in *Pseudomonas lundensis* (Figure 4.34). Therefore, this integration system represents a helpful tool for making genetic knock-in across a broad range of bacterial commensal species (22). For example, the utilization of fluorescent commensal bacteria can be a valuable tool for live imaging studies, investigating microbial behavior and niche localization in vivo (22), or analyzing intestinal immune cells' effectiveness on the bacteria growth rate.

Limitations

In the present work, I describe that the gut microbiota harbors wide genetic variation within a host and observed that some microbial species decrease this genetic variation during host life. Genetic variation is associated with the adaptive evolutionary responses in stressful environments (153). I focused on identifying single-nucleotide variants in the gut metagenome that support bacterial adaptation signals to the host environment. However, the genetic variation can be originated from single-nucleotide variants (SNVs), short insertions and deletions (indels), and structural variants (duplications, deletions, insertions, and inversion), and these evolutionary changes are more difficult to detect using

short-read metagenomics sequencing (1, 152). Therefore, utilizing the host-specific long-read metagenomic sequencing data generated in this work will provide a higher resolution to uncover larger genomic regions associated with microbial adaptations and host health (80).

A second limitation of this study relies on the assumption that the lineage structure of the gut microbiome is simple with a single dominant strain such that major alleles can be assigned to the dominant strain (5). The quasi-phasing approach allows the confident assignment of alleles to the dominant strain and elucidates evolutionary changes in gut microbial population over host life. However, the major challenge of this approach is the absence of strain resolution without knowing the absolute lineage frequency and the exclusion of samples that do not show single lineage domination (5). As an example, more than 95% of the reconstructed genomes were excluded from the microdiversity and evolutionary changes analysis due to their complex population structure. One possible solution to disentangle genetic changes within-population lineages is to estimate haplotypes of each species in the population and track their allele frequencies over time. Recently, different computational algorithms can identify and reconstruct bacterial and viral haplotypes within microbial communities from short-read and linked-read metagenomic sequencing data (154, 155). Haplotype reconstruction could be a powerful tool for disentangling within species dynamics and shed light on evolutionary adaptations to host colonization (5).

On the other hand, technical limitations of the characterization of the turquoise killifish's gut microbiota are inherent in the culturomics approach. The main weakness of culturomics is the great amount of time and workload in cultivating and isolating all the bacterial strains (156). One solution involves utilizing optimized culturomics conditions to capture the higher number of bacterial isolates reducing the culturomics workflow (157). For example, Diakite el at. established a list of 16 optimal conditions that allow the recovery of 98% of the total number of previously isolated species using 58 culturomics conditions, reducing the number of conditions by more than a half (157).

The second technical limitation in our culturomics approach relies on medium composition. My culturomics approach implemented nutrient-rich media to cultivate nutrient-rich

fastidious aerobic and anaerobic microorganisms. However, some bacteria taxa could find these growth conditions nutrient-limited, limiting their growth and leading to microbial analysis biases (158). The implementation of enriched non-selective media with fatty acids (YCFAC) or rumen fluids supplementations have been demonstrated to capture the >300 bacterial species in one single condition (158).

Chapter 6

Conclusion and future perspectives

The study of the microbiome evolution in the context of host aging represents a practical approach to understanding how gut microbes evolve and adapt under age-related environmental perturbation and how these genetic modifications can affect host physiology and health (80, 156). Recent analyses leverage metagenomic data of the human microbiome and demonstrated that microbial populations could evolve on short time scales (3,4,5).

However, there is still limited understanding of how within-host evolutionary processes occur throughout host aging. To address this question, I leveraged a unique longitudinal stool collection from three different mouse individuals collected during mice lifespan. This collection represents a unique opportunity to understand whether evolutionary changes occur over longer timescales and which genes or molecular pathways are the targets of selection.

My research showed evidence that specific microbial populations can colonize the host gastrointestinal tract and remain present throughout host life. Next, I explored the genetic changes that these microbial communities experience during host life and found that some microbial populations exert a decrease in the genetic diversity in coding-protein genes. This result supports our hypothesis, which suggests that gut commensal microbes evolve to age-related environmental perturbation (e.g., changes in intestinal architecture and cell composition, transcriptional shifts), and beneficial genetic changes increase in frequency in the population, reducing genetic diversity.

Following, I asked which genes present an increase in the frequency of a genetic variant and found that outer-membrane proteins (SusD) and polysaccharide degradation enzymes possess non-synonymous variants that get fixed in the population at host old age. SusD outer-membrane family protein is involved in Ig-A mediated host colonization, and novel genetic variants in SusD might be responsible for evasion of the host's immune system or phage predation that allows the strain maintenance in the microbial community (3,4,99). Then, I asked whether these genetic changes possess signatures of selection and found that purifying selection impacts the majority of these variants in the microbiome.

However, transposase genes show clear signatures of adaptation. A previous study in the human microbiome showed that transposase genes exert high pN/pS ratios (103), confirming that transposase plays a crucial role in bacterial adaptation in the host's intestine over aging.

Finally, I asked whether genetic variants emerging during aging increase bacterial fitness and found three different scenarios where the bacterial growth rate increases, decreases, or remains stable. The results of my thesis showed that microbial evolution takes place over long-time scales and genetic variants in outer-membrane genes, polysaccharide pathways, and transposase genes play an essential role in microbial adaptation during the constantly changing aged mammalian gut.

Future perspectives of my research work should focus on other evolutionary changes that I excluded from my analysis. Recombination, insertion/deletion (indels), structural variants (SVs), duplications, and inversions affect genetic variation within microbial populations (152). Implementation of long-read sequencing, linked reads sequencing, and Chromosome conformation capture techniques (Hi-C) will substantially improve the resolution of metagenome-assembled genomes and reveal the role of extensive genomic modification in adaptive bacterial evolution in the mammalian gut (58, 159, 160). Finally, it is necessary to understand the adaptation mechanisms produced by the evolutionary adaptations of the microbiome and the effect on the microbial community composition and host health.

On the other hand, the characterization of gut microbes inthe turquoise killifish represents a unique opportunity to understand the role of the gut commensal on determin-

ing host lifespan (8). *Nothobranchius furzeri* possesses a complex microbial community whose characterization obtained via metagenomic sequencing will give valuable information about the genes and metabolic pathways involved in determining host lifespan. I performed the first reported long-read metagenomic sequencing on six adult fish samples. I found that the Proteobacteria family and the Aeromonadaceae, Shewanellaceae, Oxalobacteraceae, and Pseudomonadaceae genera dominate the killifish gut microbial community under laboratory conditions. Furthermore, assembly of the gut metagenome could completely reconstruct a bacterial genome belonging to the Undibacterium genus and two bacteriophage genomes from the Myoviridae family.

Further, cultivation, isolation, and hybrid *de novo* genome sequencing of commensal microbes from the killifish gut completely reconstruct commensal genomes creating the first bacterial collection and reference database from the killifish gut microbiota. Finally, to leverage the power of the bacterial isolates, I established the Tn7 transposase-mediated chromosome modification on the commensal *Pseudomonas lundensis*, opening the opportunity to study microbe-host and microbe-microbe interactions in vivo.

Further perspective relies on establishing optimal culturomic conditions to recover more bacterial isolates to reduce the workload inherent in the culturomics approach. Furthermore, the establishment of reported genome editing strategies in vivo, using the CRISPR-CAS transposase system (151), will uncover microbial gene function in the context of the gut microbial community and host aging.

Chapter 7

References

1. Zeevi, D. et al. Structural variation in the gut microbiome associates with host health. Nat. 2019 5687750 568, 43–48 (2019).

2. Han, B. et al. Microbial Genetic Composition Tunes Host Longevity. Cell 169, 1249-1262.e13 (2017).

3. Chen, D. W. & Garud, N. R. Rapid evolution and strain turnover in the infant gut microbiome. bioRxiv 2021.09.26.461856 (2021). doi:10.1101/2021.09.26.461856

4. Zhao, S. et al. Adaptive Evolution within Gut Microbiomes of Healthy People. Cell Host Microbe 25, 656-667.e8 (2019).

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

6. Barreto, H. C., Sousa, A. & Gordo, I. The Landscape of Adaptive Evolution of a Gut Commensal Bacteria in Aging Mice. SSRN Electron. J. 1–8 (2019). doi:10.2139/ssrn.3455477

7. Aleman, F. D. D. & Valenzano, D. R. Microbiome evolution during host aging. PLOS Pathog. 15, e1007727 (2019).

8. Smith, P. et al. Regulation of life span by the gut microbiota in the short-lived african turquoise killifish. Elife 6, (2017).

9. Valenzano, D. R., Aboobaker, A., Seluanov, A. & Gorbunova, V. Non-canonical aging model systems and why we need them. EMBO J. 36, 959–963 (2017).

10. Han, B. et al. Microbial Genetic Composition Tunes Host Longevity. Cell 169, 1249-1262.e13 (2017).

11. Quick, J., Nicholls, S. M. & Loman, N. J. The 'Three Peaks' faecal DNA extraction method for long-read sequencing. Available at: https://www.protocols.io/view/the-39-three-peaks-39-faecal-dna-extraction-method-7rshm6e. (Accessed: 10th February 2022)
12. Browne, H. P. et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533, 543–546 (2016).
13. Lagier, J. C. et al. Culture of previously uncultured members of the human gut microbiota by culturomics. Nat. Microbiol. 1, 1–8 (2016).
14. Ong, S. H. et al. Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences. PLoS One 8, e60811 (2013).
15. Herlemann, D. P. R. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5, 1571–1579 (2011).
16. Moss, E. L., Maghini, D. G. & Bhatt, A. S. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat. Biotechnol. 38, 701–707 (2020).
17. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24, 2033–2040 (2014).
18. Hennig, B. P. et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. G3 Genes, Genomes, Genet. 8, 79–89 (2018).
19. Chen, Z. et al. Ultra-low input single tube linked-read library method enables short-read NGS systems to generate highly accurate and economical long-range sequencing information for de novo genome assembly and haplotype phasing. Genome Res. (2020). doi:10.1101/852947
20. Chen, T. Simple and Scalable Genome Analysis with Transposase Enzyme Linked Long-Read Sequencing (TELL-Seq): From Haplotype Phasing to De Novo Assembly in a Tube. J. Biomol. Tech. 30, S37 (2019).
21. Choi, K. H. et al. A Tn7-based broad-range bacterial cloning and expression system. Nat. Methods 2, 443–448 (2005).
22. Wall, E. S. et al. Modernized Tools for Streamlined Genetic Manipulation and Comparative Study of Wild and Diverse Proteobacterial Lineages. MBio 9, 1–19 (2018).
23. Anh-Hue T. Tu. Transformation of Escherichia coli Made Competent by Calcium Chloride Protocol. ASM. 1 (2008).
24. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacte-

rial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, (2017).

25. Walker, B. J. et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963 (2014).

26. Kolmogorov, M. et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat. Methods 17, 1103–1110 (2020).

27. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31, 3350–3352 (2015).

28. Bertrand, D. et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat. Biotechnol. 37, 937–944 (2019).

29. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075 (2013).

30. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 27, 824–834 (2017).

31. Uritskiy, G. V., Diruggiero, J. & Taylor, J. MetaWRAP - A flexible pipeline for genome-resolved metagenomic data analysis 08 Information and Computing Sciences 0803 Computer Software 08 Information and Computing Sciences 0806 Information Systems. Microbiome 6, 1–13 (2018).

32. Kang, D. D. et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019, e7359 (2019).

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

34. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607 (2016).

35. Evans, J. T. & Denef, V. J. To Dereplicate or Not To Dereplicate? mSphere 5, (2020).

36. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat. Biotechnol. 39, 727–736 (2021).

37. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 11, 2864–2868 (2017).

38. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36, 1925–1927 (2020).

39. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996 (2018).

40. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol. 20, 1–13 (2019).

41. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ 2015, (2015).

42. Seemann, T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).

43. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J. Mol. Biol. 428, 726–731 (2016).

44. Hyatt, D. et al. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 1–11 (2010).

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

46. Levade, I. et al. A Combination of Metagenomic and Cultivation Approaches Reveals Hypermutator Phenotypes within Vibrio cholerae-Infected Patients. mSystems 6, (2021).

47. Toth, L. A. Identifying and Implementing Endpoints for Geriatric Mice. Comp. Med. 68, 439–451 (2018).

48. Koetsier, G. & Cantor, E. A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers. New Engl. Biolabs 1, 1–8 (2019).

49. Agilent Technologies Inc. Agilent D1000 ScreenTape System Quick Guide. Agil. Technol. Inc. 3, 1–4 (2015).

50. Bharti, R. & Grimm, D. G. Current challenges and best-practice protocols for microbiome analysis. Briefings in Bioinformatics 22, 178–193 (2021).

51. Li, Y. et al. Recovery of human gut microbiota genomes with third-generation sequencing. Cell Death Dis. 12, 1–11 (2021).

52. Sczyrba, A. et al. Critical Assessment of Metagenome Interpretation - A benchmark

of metagenomics software. Nat. Methods 14, 1063–1071 (2017).

53. Frank, J. A. et al. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. Sci. Rep. 6, 1–10 (2016).

54. Koren, S. et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotechnol. 30, 693–700 (2012).

55. Derakhshani, H., Bernier, S. P., Marko, V. A. & Surette, M. G. Completion of draft bacterial genomes by long-read sequencing of synthetic genomic pools. BMC Genomics 21, 1–11 (2020).

56. Nicholls, S. M., Quick, J. C., Tang, S. & Loman, N. J. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. Gigascience 8, (2019).

57. Sevim, V. et al. Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies. Sci. Data 6, 1–9 (2019).

58. Roodgar, M. et al. Longitudinal linked-read sequencing reveals ecological and evolutionary responses of a human gut microbiome during antibiotic treatment. Genome Res. 31, 1433–1446 (2021).

59. Zhang, L. et al. A comprehensive investigation of metagenome assembly by linked-read sequencing. Microbiome 8, 1–11 (2020).

60. Bishara, A. et al. High-quality genome sequences of uncultured microbes by assembly of read clouds. Nat. Biotechnol. 36, 1067–1080 (2018).

61. Spies, N. et al. Genome-wide reconstruction of complex structural variants using read clouds. Nat. Methods 14, 915–920 (2017).

62. Saheb Kashaf, S., Almeida, A., Segre, J. A. & Finn, R. D. Recovering prokaryotic genomes from host-associated, short-read shotgun metagenomic sequencing data. Nature Protocols 16, 2520–2541 (2021).

63. Chen, L. X., Anantharaman, K., Shaiber, A., Murat Eren, A. & Banfield, J. F. Accurate and complete genomes from metagenomes. Genome Research 30, 315–333 (2020).

64. Simon Andrews. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Soil 5, 47–81 (2020). 65. Andreu-Sánchez, S. et al. A Benchmark of Genetic Variant Calling Pipelines Using Metagenomic Short-Read Sequencing. Front. Genet. 12, (2021).

66. Nayfach, S., Rodriguez-Mueller, B., Garud, N. & Pollard, K. S. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and

biogeography. Genome Res. 26, 1612–1625 (2016).

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

68. Kang, D. D. et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019, e7359 (2019).

69. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607 (2016).

70. Lesker, T. R. et al. An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome. Cell Rep. 30, 2909-2922.e6 (2020).

71. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015).

72. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology 35, 725–731 (2017).

73. Evans, J. T. & Denef, V. J. To Dereplicate or Not To Dereplicate? mSphere 5, (2020).

74. Olm, M. R. et al. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. mSystems 5, (2020).

75. Barroso-Batista, J. et al. The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. PLoS Genet. 10, (2014).

76. Barreto, H. C., Sousa, A. & Gordo, I. The Landscape of Adaptive Evolution of a Gut Commensal Bacteria in Aging Mice. SSRN Electron. J. 1–8 (2019). doi:10.2139/ssrn.3455477

77. Liu, C. et al. The Mouse Gut Microbial Biobank expands the coverage of cultured bacteria. Nat. Commun. 11, 1–12 (2020).

78. Yang, J., Park, J., Park, S., Baek, I. & Chun, J. Introducing murine microbiome database (MMDB): A curated database with taxonomic profiling of the healthy mouse gastrointestinal microbiome. Microorganisms 7, (2019).

79. Lagkouvardos, I. et al. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nat. Microbiol. 2016 110 1, 1–15 (2016).

80. Shoemaker, W. R., Chen, D. & Garud, N. R. Comparative Population Genetics in the

Human Gut Microbiome. Genome Biol. Evol. 14, (2022).

81. Hildebrand, F. Ultra-resolution Metagenomics: When Enough Is Not Enough. mSystems 6, (2021).

82. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76, 5269–5273 (1979).

83. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. Nat. Biotechnol. 34, 1256–1263 (2016).

84. Mugal, C. F., Wolf, J. B. W. & Kaj, I. Why time matters: Codon evolution and the temporal dynamics of dN/dS. Mol. Biol. Evol. 31, 212–231 (2014).

85. Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. PLoS Genet. 4, e1000304 (2008).

86. Fuller, Z. L. et al. Genome-wide analysis of signatures of selection in populations of African honey bees (Apis mellifera) using new web-based tools. BMC Genomics 16, 1–18 (2015).

87. McKenna, W. G. & Masters, M. Biochemical evidence for the bidirectional replication of DNA in Escherichia coli. Nature 240, 536–539 (1972).

88. Korem, T. et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science (80-.). 349, 1101–1106 (2015).

89. Pope, C. F., McHugh, T. D. & Gillespie, S. H. Methods to determine fitness in bacteria. Methods Mol. Biol. 642, 113–121 (2010).

90. Nava, G. M., Friedrichsen, H. J. & Stappenbeck, T. S. Spatial organization of intestinal microbiota in the mouse ascending colon. ISME J. 5, 627–638 (2011).

91. Lagkouvardos, I. et al. Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. Microbiome 7, 1–15 (2019).

92. Browne, H. P. et al. Host adaptation in gut Firmicutes is associated with sporulation loss and altered transmission cycle. Genome Biol. 22, 1–20 (2021).

93. Hildebrand, F. et al. Dispersal strategies shape persistence and evolution of human gut bacteria. Cell Host Microbe 29, 1167-1176.e9 (2021).

94. Oersted, M., Yashiro, E., Hoffmann, A. A. & Kristensen, T. N. Population bottlenecks constrain microbiome diversity and host genetic variation impeding fitness. bioRxiv 2021.07.04.450854 (2021).

doi:10.1101/2021.07.04.450854 95. Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease. N. Engl. J. Med. 375, 2369–2379 (2016).

96. Lieberman, T. D. et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. 43, 1275–1280 (2011).

97. N'Guessan, A., Brito, I. L., Serohijos, A. W. R. & Shapiro, B. J. Mobile Gene Sequence Evolution within Individual Human Gut Microbiomes Is Better Explained by Gene-Specific Than Host-Specific Selective Pressures. Genome Biol. Evol. 13, (2021).

98. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402 (1997).

99. Lee, S. M. et al. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature 501, 426–429 (2013).

100. Kaoutari, A. El, Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat. Rev. Microbiol. 11, 497–504 (2013). 101. Kawai, K., Kamochi, R., Oiki, S., Murata, K. & Hashimoto, W. Probiotics in human gut microbiota can degrade host glycosaminoglycans. Sci. Rep. 8, 1–13 (2018). 102. Broaders, E., Gahan, C. G. M. & Marchesi, J. R. Mobile genetic elements of the human gastrointestinal tract: Potential for spread of antibiotic resistance genes. Gut Microbes 4, 271–280 (2013).

103. Schloissnig, S. et al. Genomic variation landscape of the human gut microbiome. Nature 493, 45–50 (2013).

104. Mes, T. H. M. & Doeleman, M. Positive selection on transposase genes of insertion sequences in the Crocosphaera watsonii genome. J. Bacteriol. 188, 7176–7185 (2006).

105. Glansdorff, N., Charlier, D. & Zafarullah, M. Activation of gene expression by IS2 and IS3. Cold Spring Harb. Symp. Quant. Biol. 45 Pt 1, 153–156 (1981).

106. Ciuffreda, L., Rodríguez-Pérez, H. & Flores, C. Nanopore sequencing and its application to the study of microbial communities. Comput. Struct. Biotechnol. J. 19, 1497–1511 (2021).

107. Zlitni, S. et al. Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. Genome Med. 12, 1–17 (2020).

108. Kang, J. B. et al. Intestinal microbiota domination under extreme selective pressures characterized by metagenomic read cloud sequencing and assembly. BMC Bioinformatics 20, 1–13 (2019).

109. Shao, W., Khin, S. & Kopp, W. C. Characterization of Effect of Repeated Freeze and Thaw Cycles on Stability of Genomic DNA Using Pulsed Field Gel Electrophoresis. https://home.liebertpub.com/bio 10, 4–11 (2012).

110. Lagier, J. C. et al. Microbial culturomics: paradigm shift in the human gut microbiome study. Clin. Microbiol. Infect. 18, 1185–1193 (2012).

111. Lagier, J. C. et al. Culturing the human microbiota and culturomics. Nat. Rev. Microbiol. 2018 169 16, 540–550 (2018).

112. Yu, Y. Y., Ding, L. G., Huang, Z. Y., Xu, H. Y. & Xu, Z. Commensal bacteria-immunity crosstalk shapes mucosal homeostasis in teleost fish. Rev. Aquac. 13, 2322–2343 (2021).

113. Wang, A. R., Ran, C., Ringø, E. & Zhou, Z. G. Progress in fish gastrointestinal microbiota research. Rev. Aquac. 10, 626–640 (2018).

114. Sczyrba, A. et al. Critical Assessment of Metagenome Interpretation - A benchmark of metagenomics software. Nat. Methods 14, 1063–1071 (2017).

115. Gerner, S. M., Rattei, T. & Graf, A. B. Assessment of urban microbiome assemblies with the help of targeted in silico gold standards. Biol. Direct 13, 1–21 (2018).

116. Sheu, S. Y., Lin, Y. S., Chen, J. C., Kwon, S. W. & Chen, W. M. Undibacterium squillarum sp. nov., isolated from a freshwater shrimp culture pond. Int. J. Syst. Evol. Microbiol. 64, 3459–3466 (2014).

117. Kämpfer, P. et al. Undibacterium pigrum gen. nov., sp. nov., isolated from drinking water. Int. J. Syst. Evol. Microbiol. 57, 1510–1515 (2007).

118. Kämpfer, P. et al. Undibacterium danionis sp. Nov. isolated from a zebrafish (Danio rerio). Int. J. Syst. Evol. Microbiol. 66, 3625–3631 (2016).

119. Lee, S. Y. et al. Undibacterium piscinae sp. Nov., isolated from Korean shiner intestine. Int. J. Syst. Evol. Microbiol. 69, (2019).

120. Davis, J. J. et al. The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. Nucleic Acids Res. 48, D606–D612 (2020).

121. Li, T. et al. Alterations of the gut microbiome of largemouth bronze gudgeon (Coreius guichenoti) suffering from furunculosis. Sci. Reports 2016 61 6, 1–9 (2016).

122. Egerton, S., Culloty, S., Whooley, J., Stanton, C. & Ross, R. P. The gut microbiota of marine fish. Front. Microbiol. 9, 873 (2018).

123. Chen, Z., Ceballos-Francisco, D., Guardiola, F. A. & Esteban, M. Á. Dietary admin-

istration of the probiotic Shewanella putrefaciens to experimentally wounded gilthead seabream (Sparus aurata L.) facilitates the skin wound healing. Sci. Reports 2020 101 10, 1–13 (2020).

124. Cornuault, J. K. et al. The enemy from within: a prophage of Roseburia intestinalis systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. ISME J. 14, 771–787 (2020).

125. Bondy-Denomy, J. & Davidson, A. R. When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. Journal of Microbiology 52, 235–242 (2014).

126. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. 2015 141 14, 20–32 (2015).

127. Rubin, B. E. et al. Species- and site-specific genome editing in complex bacterial communities. Nat. Microbiol. 7, 34–47 (2022).

128. Choi, K. H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. Nat. Protoc. 2006 11 1, 153–161 (2006).

129. Bakshi, S. et al. Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. Nat. Microbiol. 2021 66 6, 783–791 (2021).

130. Fan, J., Huang, S. & Chorlton, S. D. BugSeq: a highly accurate cloud platform for long-read metagenomic analyses. BMC Bioinformatics 22, 1–12 (2021).

132. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat. Biotechnol. 39, 105–114 (2021).

132. Brown, C. L. et al. Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. Sci. Rep. 11, 1–12 (2021).

133. Tolstoganov, I., Bankevich, A., Chen, Z. & Pevzner, P. A. CloudSPAdes: Assembly of synthetic long reads using de Bruijn graphs. in Bioinformatics 35, i61–i70 (Oxford Academic, 2019).

134. Danko, D. C., Meleshko, D., Bezdan, D., Mason, C. & Hajirasouliha, I. Minerva: An alignment- and reference-free approach to deconvolve Linked-Reads for metagenomics. Genome Res. 29, 116–124 (2019).

135. Mak, L. et al. Ariadne: Barcoded Linked-Read Deconvolution Using de Bruijn Graphs. bioRxiv 2021.05.09.443255 (2021). doi:10.1101/2021.05.09.443255

136. Shajii, A., Numanagić, I., Whelan, C. & Berger, B. Statistical Binning for Barcoded Reads Improves Downstream Analyses. Cell Syst. 7, 219-226.e5 (2018).

137. Bendall, M. L. et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 10, 1589–1601 (2016).

138. Chijiiwa, R. et al. Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. Microbiome 8, 1–14 (2020).

139. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J. Biol. Chem. 284, 24673–24677 (2009).

140. Carrow, H. C., Batachari, L. E. & Chu, H. Strain diversity in the microbiome: Lessons from Bacteroides fragilis. PLoS Pathog. 16, e1009056 (2020).

141. Mes, T. H. M. & Doeleman, M. Positive selection on transposase genes of insertion sequences in the Crocosphaera watsonii genome. J. Bacteriol. 188, 7176–7185 (2006).

142. Vigil-Stenman, T., Ininbergs, K., Bergman, B. & Ekman, M. High abundance and expression of transposases in bacteria from the Baltic Sea. ISME J. 11, 2611–2623 (2017).

143. Casacuberta, E. & González, J. The impact of transposable elements in environmental adaptation. Molecular Ecology 22, 1503–1517 (2013).

144. Emiola, A. & Oh, J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. Nat. Commun. 9, 1–8 (2018).

145. Ferreiro, A., Crook, N., Gasparrini, A. J. & Dantas, G. Multiscale Evolutionary Dynamics of Host-Associated Microbiomes. Cell 172, 1216–1227 (2018).

146. Confer, A. W. & Ayalew, S. The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. Veterinary Microbiology 163, 207–222 (2013).

147. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. FEMS Microbiology Reviews 29, 231–262 (2005).

148. Hoskisson, P. A. & Rigali, S. Chapter 1 Variation in Form and Function. The Helix-Turn-Helix Regulators of the GntR Superfamily. Advances in Applied Microbiology 69, 1–22 (2009).

149. Maguire, F. et al. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands. Microb. Genomics 6, 1–12 (2020).

150. Lam, K. N. et al. Phage-delivered CRISPR-Cas9 for strain-specific depletion and genomic deletions in the gut microbiome. Cell Rep. 37, (2021).

151. Rubin, B. E. et al. Species- and site-specific genome editing in complex bacterial communities. Nat. Microbiol. 7, 34–47 (2022).

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

152. Garud, N. R. & Pollard, K. S. Population Genetics in the Human Microbiome. Trends Genet. 36, 53–67 (2020).

153. Ørsted, M., Hoffmann, A. A., Sverrisdóttir, E., Nielsen, K. L. & Kristensen, T. N. Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. PLoS Genet. 15, (2019).

154. Luo, C. et al. ConStrains identifies microbial strains in metagenomic datasets. Nat. Biotechnol. 2015 3310 33, 1045–1052 (2015).

155. Cao, C. et al. Reconstruction of Microbial Haplotypes by Integration of Statistical and Physical Linkage in Scaffolding. Mol. Biol. Evol. 38, 2660–2672 (2021).

156. Naud, S. et al. Proof of Concept of Culturomics Use of Time of Care. Front. Cell. Infect. Microbiol. 10, 714 (2020).

157. Diakite, A. et al. Optimization and standardization of the culturomics technique for human microbiome exploration. Sci. Rep. 10, 9674 (2020).

158. Pereira, A. C. & Cunha, M. V. An effective culturomics approach to study the gut microbiota of mammals. Res. Microbiol. 171, 290–300 (2020).

159. Yaffe, E. & Relman, D. A. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. Nat. Microbiol. 5, 343–353 (2020).

161. Demaere, M. Z. & Darling, A. E. Bin3C: Exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes. Genome Biol. 20, 1–16 (2019).

161. Esser, D. et al. Functions of the Microbiota for the Physiology of Animal Metaorganisms. J. Innate Immun. 11, 393–404 (2019).

162. Sommer, F. & Bäckhed, F. The gut microbiota — masters of host development and physiology. Nat. Rev. Microbiol. 2013 114 11, 227–238 (2013).

163. Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The Human Microbiota in Health and Disease. Engineering 3, 71–82 (2017).

164. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. Nature Reviews Microbiology 19, 55–71 (2021).

165. Schultz, M., Burton, J. P. & Chanyi, R. M. Chapter 11 - Use of Bacillus in Human Intestinal Probiotic Applications A2 - Floch, Martin H. Microbiota Gastrointest. Pathophysiol. 119–123 (2017).

1. Wu, Z. A. & Wang, H. X. A Systematic Review of the Interaction Between Gut Microbiota and Host Health from a Symbiotic Perspective. SN Compr. Clin. Med. 1, 224–235 (2019).

1. Safadi, J. M., Quinton, A. M. G., Lennox, B. R., Burnet, P. W. J. & Minichino, A. Gut dysbiosis in severe mental illness and chronic fatigue: a novel trans-diagnostic construct? A systematic review and meta-analysis. Molecular Psychiatry 1–13 (2021). doi:10.1038/s41380-021-01032-1

168. Novakovic, M. et al. Role of gut microbiota in cardiovascular diseases. World Journal of Cardiology 12, 110–122 (2020).

169. Scheithauer, T. P. M. et al. Gut Microbiota as a Trigger for Metabolic Inflammation in Obesity and Type 2 Diabetes. Frontiers in Immunology 11, 2546 (2020).

170. Yang, Y. et al. Dysbiosis of human gut microbiome in young-onset colorectal cancer. Nat. Commun. 2021 121 12, 1–13 (2021).

171. Kim, S. & Jazwinski, S. M. The Gut Microbiota and Healthy Aging: A Mini-Review. Gerontology 64, 513–520 (2018).

172. DeJong, E. N., Surette, M. G. & Bowdish, D. M. E. The Gut Microbiota and Unhealthy Aging: Disentangling Cause from Consequence. Cell Host and Microbe 28, 180–189 (2020).

173. Pelaseyed, T. & Hansson, G. C. Membrane mucins of the intestine at a glance. J. Cell Sci. 133, (2020).

174. Elderman, M. et al. The effect of age on the intestinal mucus thickness, microbiota composition and immunity in relation to sex in mice. PLoS One 12, (2017).

175. Derrien, M., Belzer, C. & de Vos, W. M. Akkermansia muciniphila and its role in regulating host functions. Microbial Pathogenesis 106, 171–181 (2017).

176. Soenen, S., Rayner, C. K., Jones, K. L. & Horowitz, M. The ageing gastrointestinal tract. Curr. Opin. Clin. Nutr. Metab. Care 19, 12–18 (2016).

177. O'Toole, P. W. & Jeffery, I. B. Gut microbiota and aging. Science (80-.). 350,

1214–1215 (2015).

178. Wilmanski, T. et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. Nat. Metab. 3, 274–286 (2021).

179. Bosco, N. & Noti, M. The aging gut microbiome and its impact on host immunity. Genes Immun. 2021 225 22, 289–303 (2021).

180. Montecino-Rodriguez, E., Berent-Maoz, B. & Dorshkind, K. Causes, consequences, and reversal of immune system aging. Journal of Clinical Investigation 123, 958–965 (2013).

181. Ragonnaud, E. & Biragyn, A. Gut microbiota as the key controllers of "healthy" aging of elderly people. Immun. Ageing 2021 181 18, 1–11 (2021). 182. Hartsough, L. A. et al. Optogenetic control of gut bacterial metabolism to promote longevity. Elife 9, 1–16 (2020).

183. Guo, L., Karpac, J., Tran, S. L. & Jasper, H. PGRP-SC2 promotes gut immune homeostasis to limit commensal dysbiosis and extend lifespan. Cell 156, 109 (2014).

184. Bárcena, C. et al. Healthspan and lifespan extension by fecal microbiota transplantation into progeroid mice. Nat. Med. 2019 258 25, 1234–1242 (2019).

185. Odamaki, T. et al. Age-related changes in gut microbiota composition from newborn to centenarian: A cross-sectional study. BMC Microbiol. 16, 1–12 (2016).

186. Sato, Y. et al. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. Nature 599, 458–464 (2021).

187. Biagi, E. et al. Gut Microbiota and Extreme Longevity. Curr. Biol. 26, 1480–1485 (2016).

188. Sato, Y. et al. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. Nature 599, 458–464 (2021).

189. Jackson, M. et al. Signatures of early frailty in the gut microbiota. Genome Med. 8, 1–11 (2016).

190. Hehemann, J. H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nat. 2010 4647290 464, 908–912 (2010).

191. Kent, A. G., Vill, A. C., Shi, Q., Satlin, M. J. & Brito, I. L. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. Nat. Commun. 11, 1–9 (2020).

192. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L.

Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature 570, 462–467 (2019).
193. Sokurenko, E. V. et al. Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. Proc. Natl. Acad. Sci. U. S. A. 95, 8922–8926 (1998).
194. Campanacci, V. et al. The crystal structure of the Escherichia coli lipocalin Blc suggests a possible role in phospholipid binding. FEBS Lett. 562, 183–188 (2004).
195. Zou, D. et al. A SNP of bacterial blc disturbs gut lysophospholipid homeostasis and induces inflammation through epithelial barrier disruption. EBioMedicine 52, (2020).
196. Zhu, Q. et al. Compositional and genetic alterations in Graves' disease gut microbiome reveal specific diagnostic biomarkers. ISME J. 2021 1511 15, 3399–3411 (2021).
197. Clark, R. L. et al. Design of synthetic human gut microbiome assembly and butyrate production. Nat. Commun. 2021 121 12, 1–16 (2021).
198. Ma, C. et al. Probiotic consumption influences universal adaptive mutations in indigenous human and mouse gut microbiota. Commun. Biol. 2021 41 4, 1–12 (2021).
199. Zhang, J. & Knight, R. Genomic mutations within the host microbiome: Adaptive evolution or purifying selection. Engineering (2022). doi:10.1016/J.ENG.2021.11.018
200. Greenblum, S., Carr, R. & Borenstein, E. Extensive strain-level copy-number variation across human gut microbiome species. Cell 160, 583–594 (2015).
201. Van Rossum, T., Ferretti, P., Maistrenko, O. M. & Bork, P. Diversity within species: interpreting strains in microbiomes. Nat. Rev. Microbiol. 2020 189 18, 491–506 (2020).
202. Segerman, B. The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. Front. Cell. Infect. Microbiol. 2, 116 (2012).
203. Cooper, V. S., Vohr, S. H., Wrocklage, S. C. & Hatcher, P. J. Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. PLoS Comput. Biol. 6, 1000732 (2010).
204. Redfield, R. J. Do bacteria have sex? Nat. Rev. Genet. 2, 634–639 (2001).
205. Rasko, D. A. et al. Origins of the E. coli Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany . N. Engl. J. Med. 365, 709–717 (2011).
206. Vogan, A. A. & Higgs, P. G. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. Biol. Direct 6, 1–14 (2011). 207. Vogan, A. A. & Higgs, P. G. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. Biol. Direct 6, 1–14 (2011).
208. Ghalayini, M. et al. Evolution of a Dominant Natural Isolate of Escherichia coli

in the Human Gut over the Course of a Year Suggests a Neutral Evolution with Reduced Effective Population Size. Appl. Environ. Microbiol. 84, (2018).

209. Gordo, I., Demengeot, J. & Xavier, K. Escherichia coli adaptation to the gut environment: a constant fight for survival. http://dx.doi.org/10.2217/fmb.14.86 9, 1235–1238 (2014).

210. Yilmaz, B. et al. Long-term evolution and short-term adaptation of microbiota strains and sub-strains in mice. Cell Host Microbe 29, 650-663.e9 (2021).

211. Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W. & Wilson, D. J. Within-host evolution of bacterial pathogens. Nat. Rev. Microbiol. 2016 143 14, 150–162 (2016).

212. Kennemann, L. et al. Helicobacter pylori genome evolution during human infection. Proc. Natl. Acad. Sci. U. S. A. 108, 5033–5038 (2011).

213. Ramiro, R. S., Durão, P., Bank, C. & Gordo, I. Low mutational load and high mutation rate variation in gut commensal bacteria. PLOS Biol. 18, e3000617 (2020).

214. Rau, M. H., Marvig, R. L., Ehrlich, G. D., Molin, S. & Jelsbak, L. Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. Environ. Microbiol. 14, 2200–2211 (2012).

215. Reichard, M. & Polačik, M. Nothobranchius furzeri, an 'instant' fish from an ephemeral habitat. Elife 8, (2019).

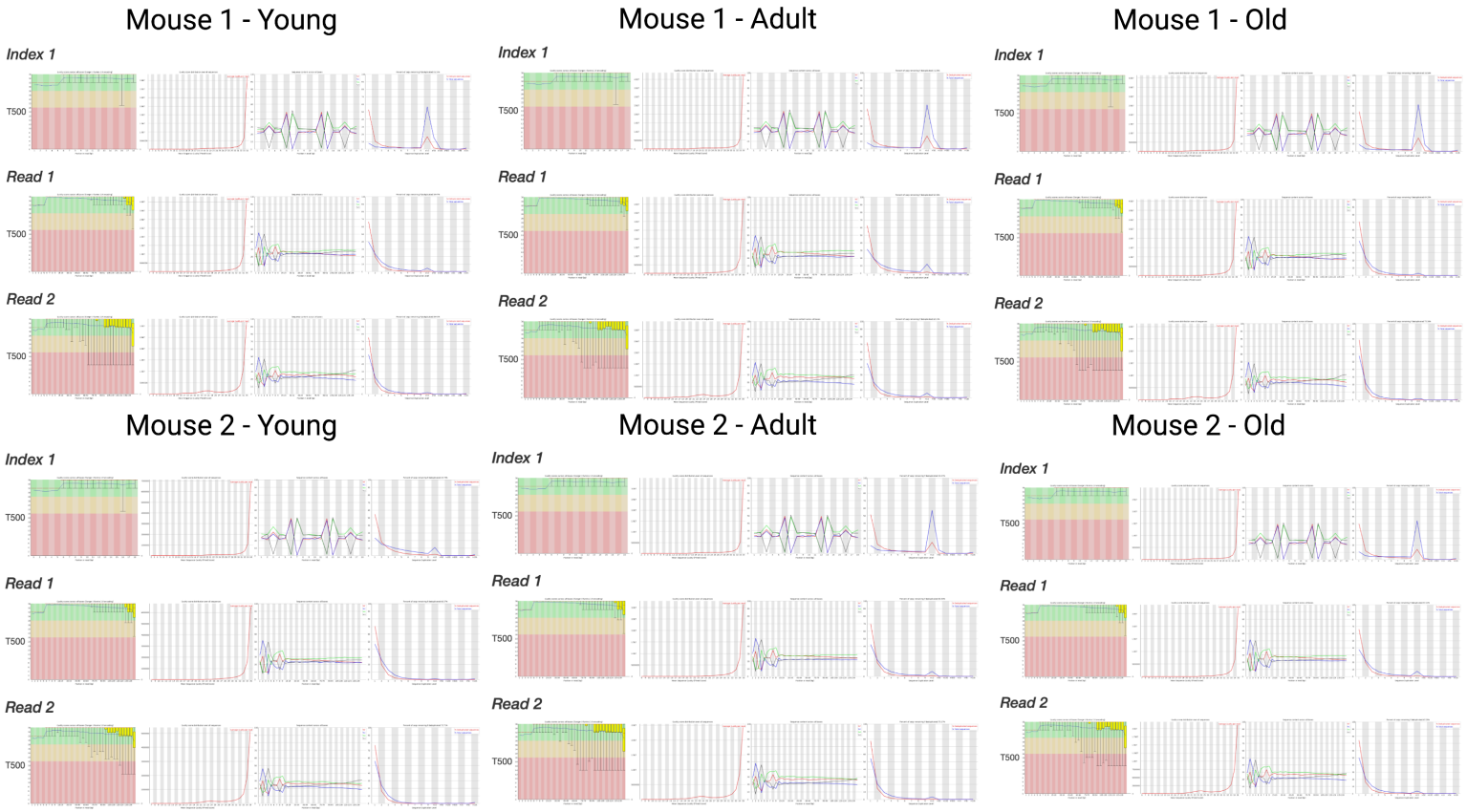
216. Kim, Y., Nam, H. G. & Valenzano, D. R. The short-lived African turquoise killifish: an emerging experimental model for ageing. Dis. Model. Mech. 9, 115–129 (2016).

217. Ragonnaud, E. & Biragyn, A. Gut microbiota as the key controllers of "healthy" aging of elderly people. Immun. Ageing 2021 181 18, 1–11 (2021).

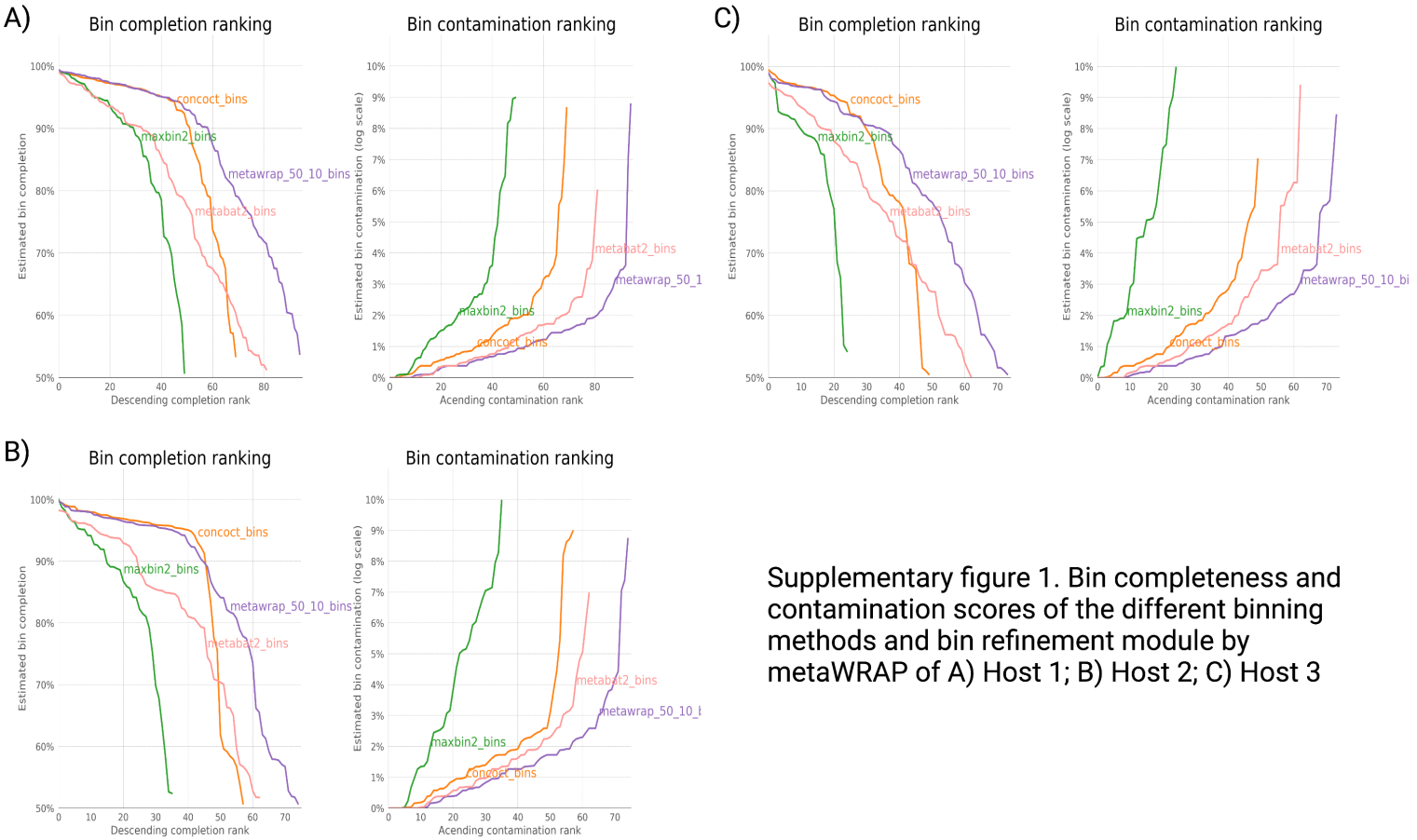
218. Bradley, P. H., Nayfach, S. & Pollard, K. S. Phylogeny-corrected identification of microbial gene families relevant to human gut colonization. PLOS Comput. Biol. 14, e1006242 (2018).

Appendix A

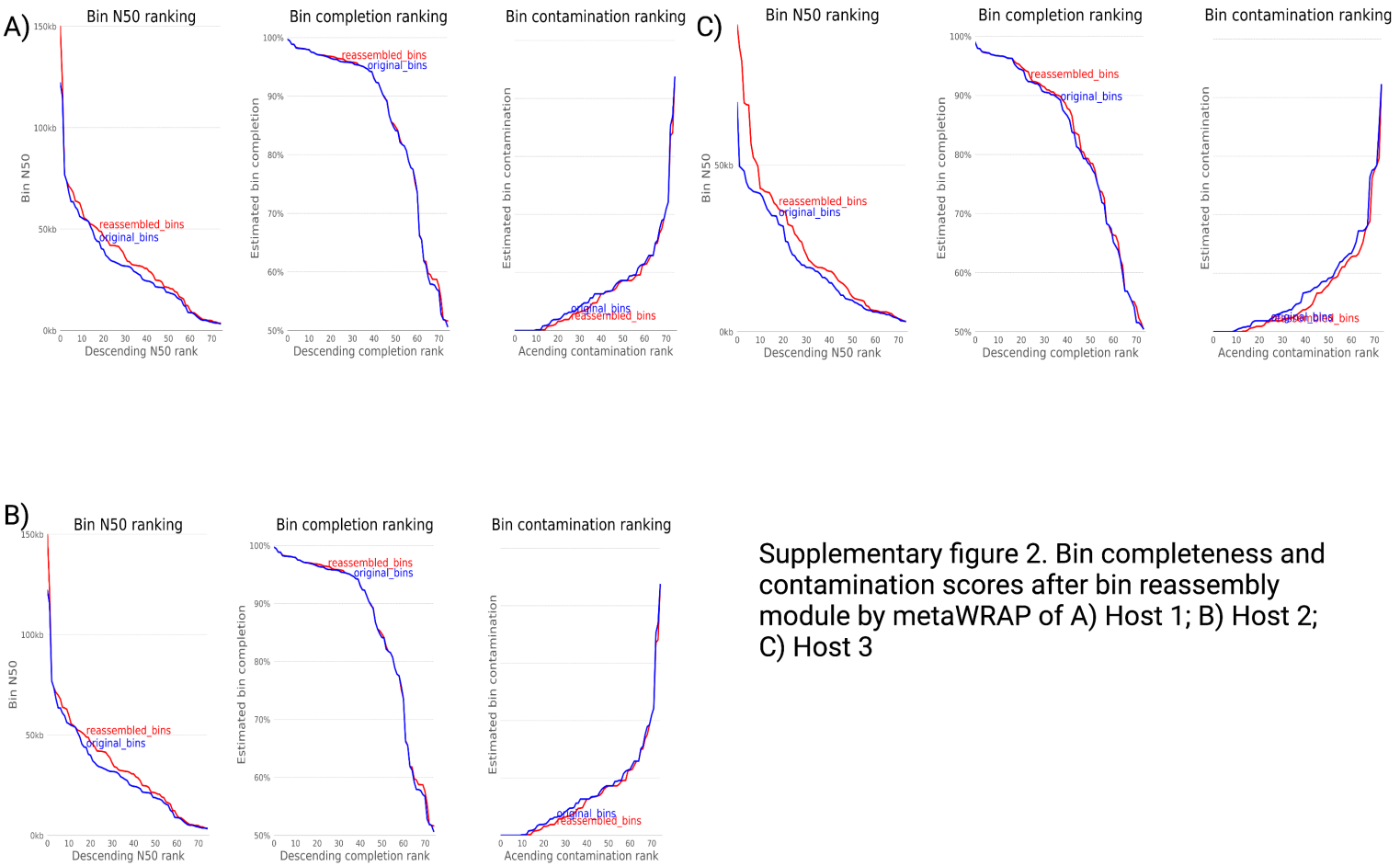
Supplementary Figures



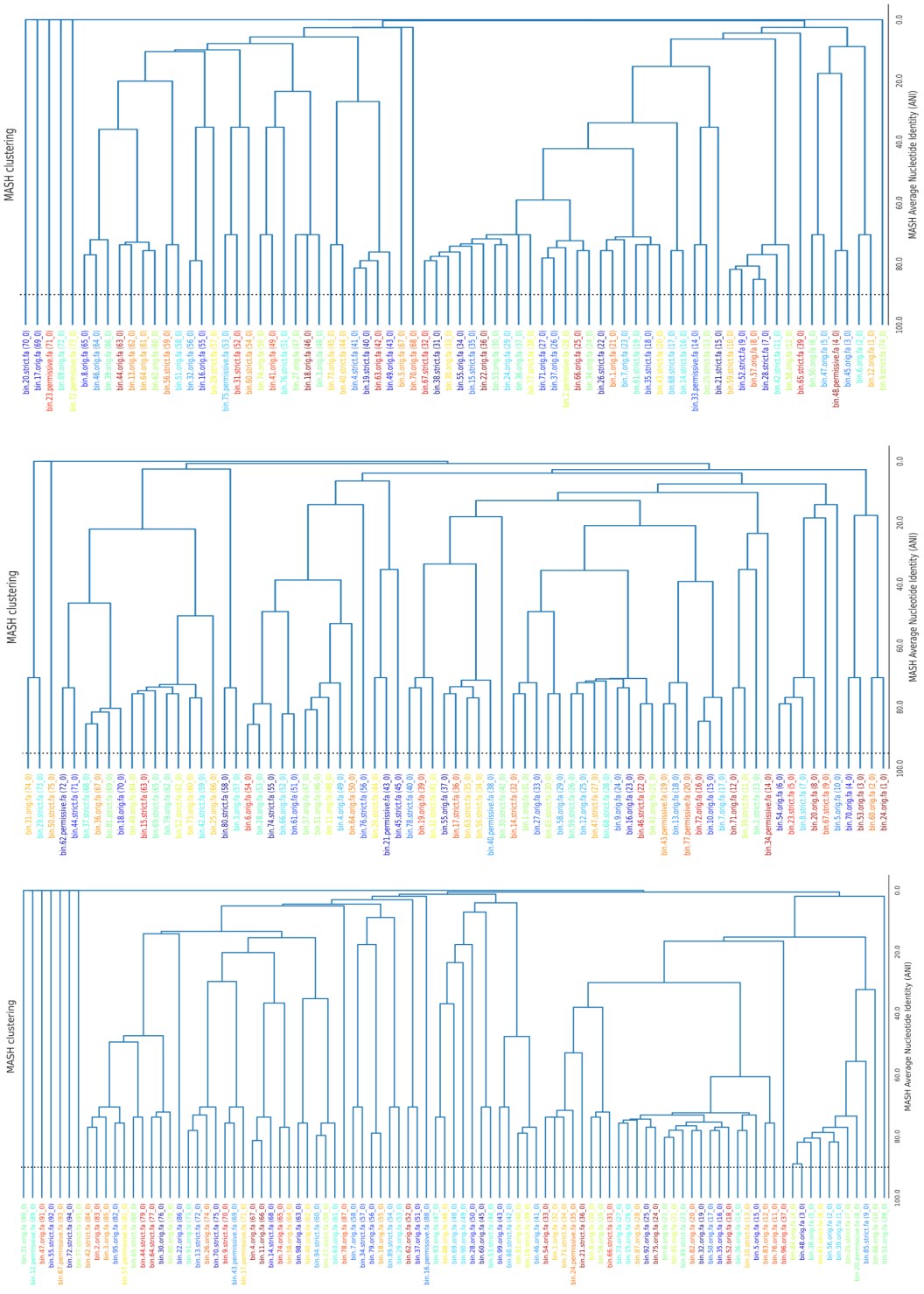
Supplementary Figure 1. Quality control reports for each metagenomic sample sequenced using the Tell-Seq technology.

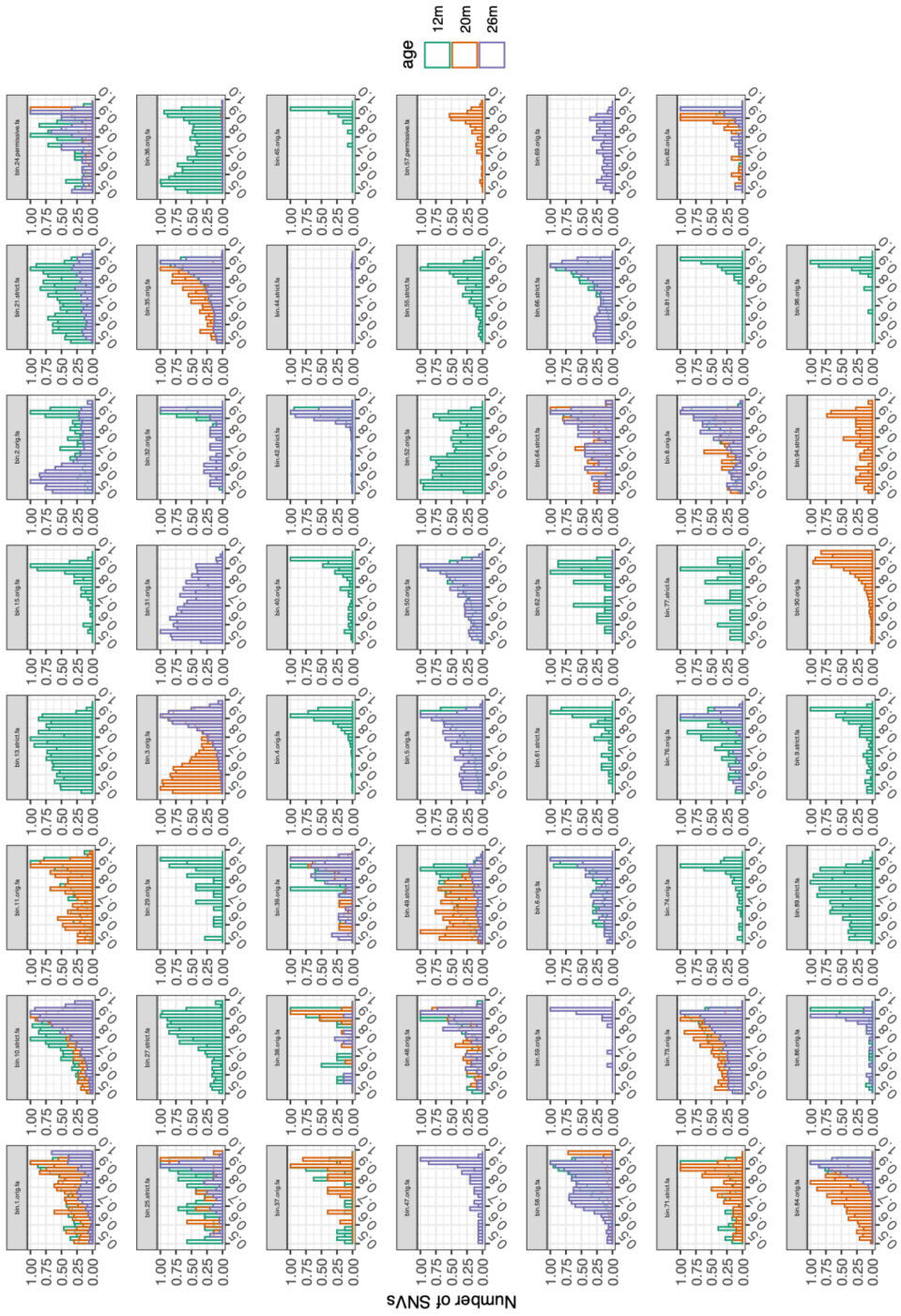


Supplementary figure 1. Bin completeness and contamination scores of the different binning methods and bin refinement module by metaWRAP of A) Host 1; B) Host 2; C) Host 3

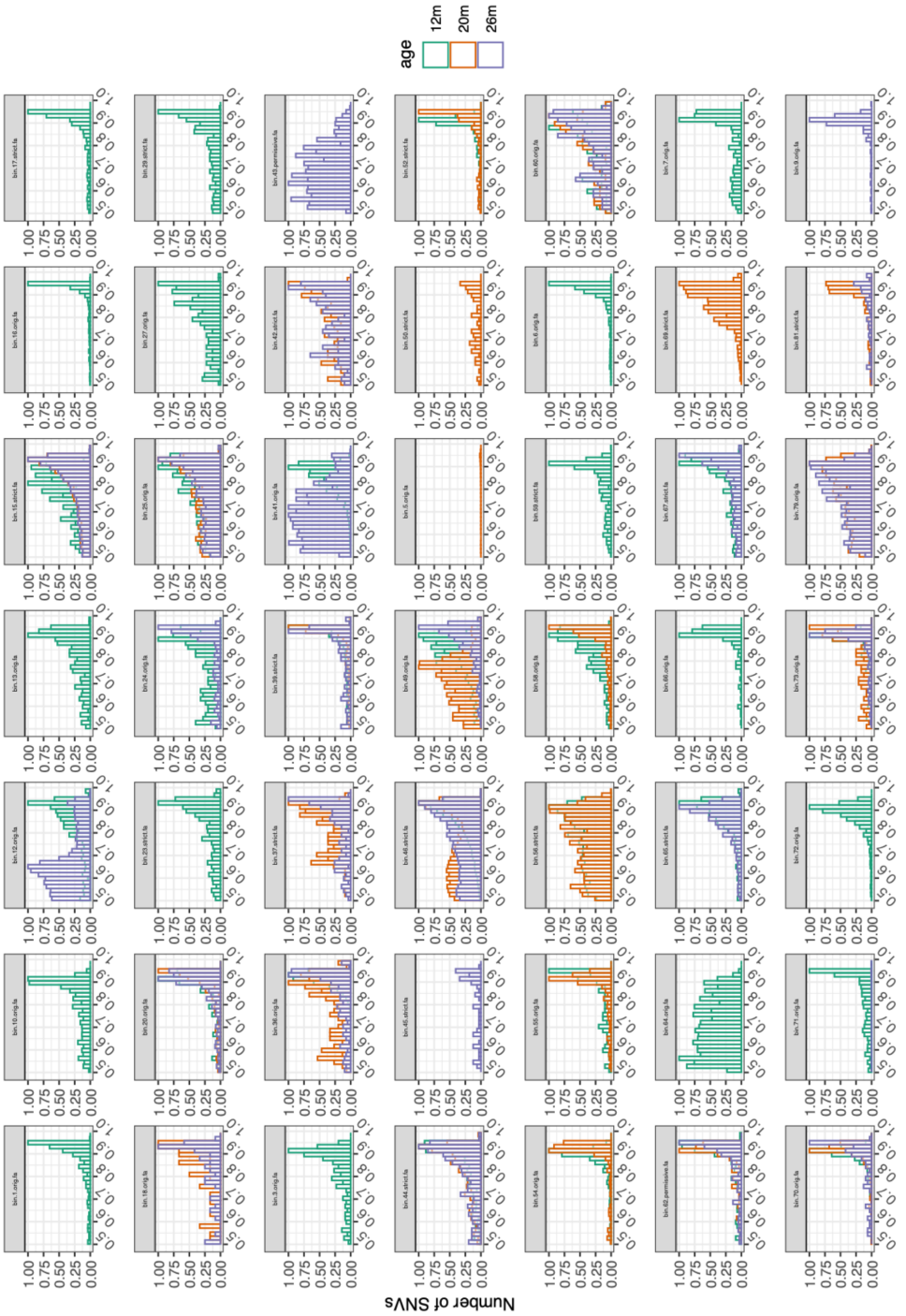


Supplementary figure 2. Bin completeness and contamination scores after bin reassembly module by metaWRAP of A) Host 1; B) Host 2; C) Host 3

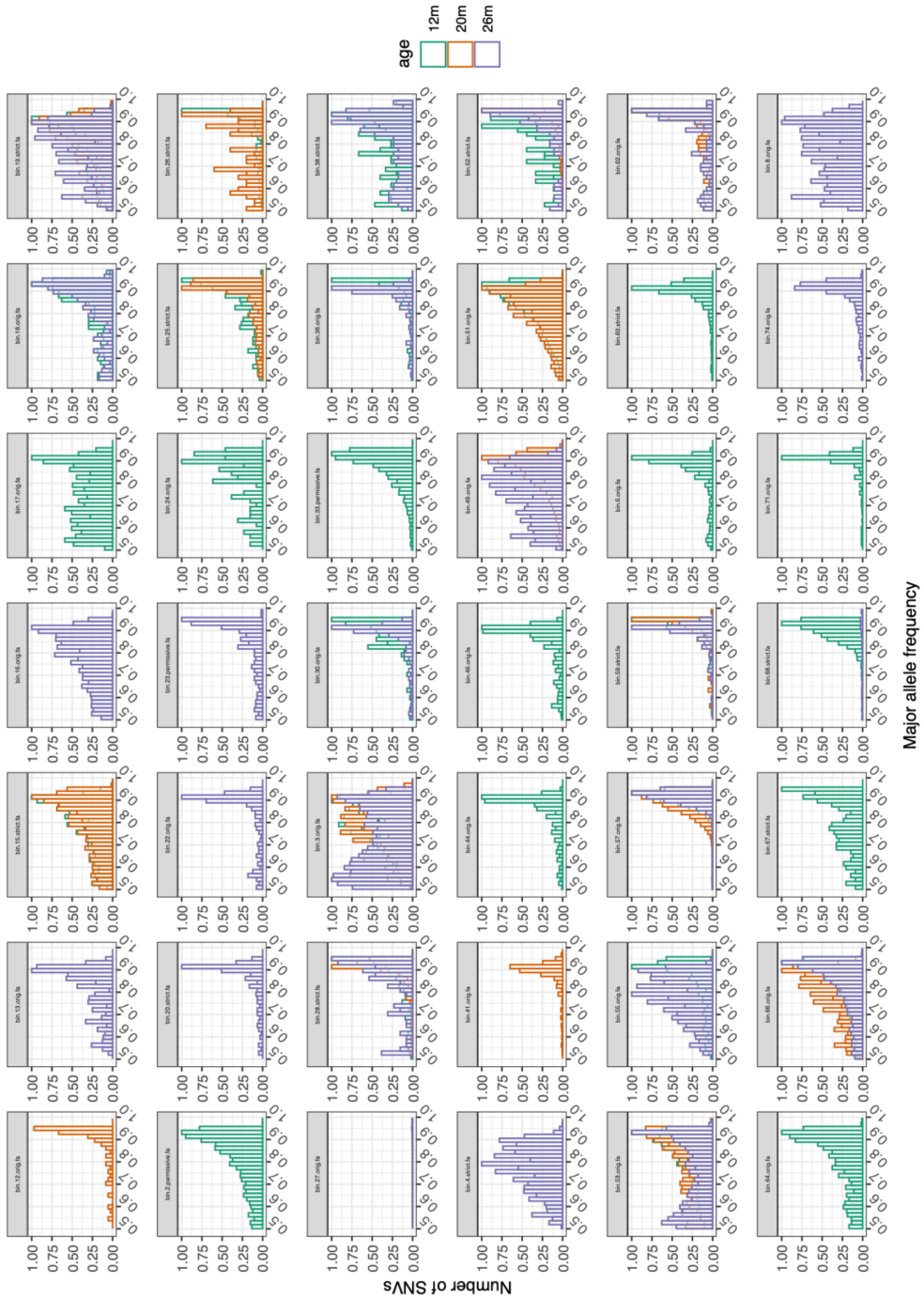

MASH clustering
MASH Average Nucleotide Identity (ANI)
0.0
20.0
40.0
60.0
80.0
100.0



Supplementary Figure 4. SNV frequency distribution of each reconstructed genome from Host 1 during host aging



Supplementary Figure 4. SNV frequency distribution of each reconstructed genome from Host 2 during host aging



Supplementary Figure 4. SNV frequency distribution of each reconstructed genome from Host 3 during host aging.

Appendix B

Supplementary Tables

Supplementary Table 1

Statistics of recovered high-quality MAGs by metaWRAP bin assigment and bin refinement modules						
Host 3						
bin	completenes	contaminati	GC	lineage	N50	size
bin.17	98.95	0	0.402	Lactobacillales	19280	1946310
bin.18	97.98	1.83	0.352	Clostridiales	68709	4093372
bin.26	97.89	0.566	0.503	Bacteroidales	48945	2455208
bin.3	97.37	3.062	0.468	Clostridiales	18917	3519338
bin.31	97.31	1.006	0.555	Clostridiales	14663	2437912
bin.47	97.23	0.754	0.492	Bacteroidales	35934	2010931
bin.1	97.17	0.754	0.513	Bacteroidales	18181	2384148
bin.36	96.9	0.476	0.488	Bacteroidetes	41732	2424486
bin.53	96.83	0.377	0.471	Bacteroidales	42725	2848092
bin.6	96.72	0	0.385	Lactobacillales	49642	1776407
bin.24	96.7	0.377	0.461	Bacteroidales	42101	3762673
bin.34	96.64	0.377	0.505	Bacteroidales	26909	2490427
bin.56	96.63	1.7	0.408	Lachnospiraceae	20044	3157578
bin.66	96.51	0.022	0.539	Bacteroidales	31931	2503497
bin.55	96.3	0.188	0.51	Bacteroidales	36867	2596212
bin.44	96.28	0.158	0.464	Clostridiales	32268	2632571
bin.68	96.27	0.377	0.552	Bacteroidales	34713	2059098
bin.43	95.32	0.377	0.536	Bacteroidales	14714	2125252
bin.14	94.83	0	0.526	Bacteroidales	20005	2365610
bin.46	94.49	2.586	0.508	Lachnospiraceae	19325	4193204
bin.60	94.37	2.849	0.474	Lachnospiraceae	22777	3868745
bin.27	94.15	1.34	0.481	Lachnospiraceae	18056	3183055
bin.25	92.8	0.706	0.466	Bacteroidales	34875	3364803

bin.32	92.33	0.69	0.508	Clostridiales	9783	3121770
bin.71	92.32	0	0.534	Bacteroidales	16311	2633102
bin.22	92.18	0.125	0.475	Bacteroidales	38643	2091212
bin.40	91.99	2.684	0.57	Clostridiales	16269	2033422
bin.16	91.98	1.323	0.464	Clostridiales	13695	3364973
bin.51	91.61	1.898	0.501	Clostridiales	21553	3374735
bin.23	90.76	0.161	0.486	Clostridia	10966	1562723
bin.57	90.55	5.689	0.602	Bacteroidetes	41917	1770622
bin.38	90.49	1.509	0.488	Bacteroidales	43248	2606976
bin.15	90.44	0.105	0.559	Bacteroidales	41458	1966238
bin.35	90.11	1.363	0.515	Bacteroidales	10989	2084524
bin.62	90.1	1.525	0.5	Lachnospiraceae	40490	3276383
bin.21	89.82	0.625	0.445	Proteobacteria	7097	2149715
bin.42	89.52	1.378	0.558	Bacteroidetes	48197	1519006
bin.64	89.17	2.083	0.516	Lachnospiraceae	23154	3564100
bin.7	87.53	0.188	0.569	Bacteroidales	17360	2075839
bin.29	87.02	0.671	0.595	Clostridiales	12096	2155444
bin.8	86.54	1.594	0.513	Lachnospiraceae	9987	3303725
bin.65	85.8	3.508	0.652	Actinobacteria	6671	2226771
bin.13	83.8	2.38	0.479	Lachnospiraceae	19191	2921087
bin.50	83.55	2.666	0.25	Bacteria	5647	1303130
bin.19	81.33	5.564	0.609	Clostridiales	8133	3002726
bin.2	81.03	1.724	0.484	Bacteria	34795	2084746
bin.61	80.34	0.943	0.46	Bacteroidales	5572	1855560
bin.30	79.31	1.724	0.488	Bacteria	31649	2471349

bin.77	79.27	2.412	0.527	Bacteroidales	5129	1803138
bin.58	78.54	1.442	0.582	Bacteroidetes	6486	1850782
bin.9	78.17	0.6	0.318	Clostridiales	5138	2122298
bin.70	77.31	1.842	0.462	Lachnospiraceae	9607	2990024
bin.33	76.72	3.448	0.425	Bacteria	15709	3511699
bin.37	75.15	0	0.529	Bacteria	20981	1915709
bin.20	73.87	0.061	0.467	Clostridiales	6414	1246122
bin.67	71.88	0.377	0.533	Bacteroidales	44849	1812194
bin.4	71.63	2.516	0.589	Clostridiales	8594	2411806
bin.45	68.33	0.328	0.323	Clostridiales	5979	1200408
bin.74	68.03	0.447	0.451	Clostridiales	5926	1982126
bin.48	67.43	5.285	0.282	Clostridiales	4747	1837616
bin.49	65.17	5.517	0.626	Bacteria	5440	2079370
bin.52	64.48	0.961	0.642	Bacteroidetes	27164	1207004
bin.59	63.79	3.448	0.665	Bacteria	12750	1745634
bin.39	62.27	0	0.514	Lachnospiraceae	24649	2247921
bin.75	59.8	0.373	0.463	Clostridiales	4221	1906285
bin.12	56.89	0	0.334	Bacteria	9463	1704771
bin.5	56.87	3.636	0.47	Bacteria	4360	1167942
bin.41	56.05	2.215	0.49	Clostridiales	5980	2005584
bin.28	55.17	0.862	0.621	Bacteria	8854	1213262
bin.69	54.23	8.434	0.527	Deltaproteobacte	3132	1556324
bin.63	51.59	0	0.647	Bacteria	7879	1346186
bin.76	51.57	7.017	0.475	Bacteria	3259	2285230
bin.72	51.13	0	0.432	Lachnospiraceae	3347	1316651

bin.73	50.53	3.446	0.562	Lachnospiraceae	3941	3072046
HOST 1						
bin	completenes	contaminati	GC	lineage	N50	size
bin.6	99.4	0.503	0.523	Bacteroidales	58854	2474392
bin.37	99.13	2.586	0.311	Bacteria	33412	2152298
bin.92	99.03	0.075	0.459	Bacteroidales	40213	2253611
bin.85	99.03	1.442	0.597	Bacteroidetes	37119	2454249
bin.55	98.95	0.571	0.401	Lactobacillales	27504	1993927
bin.25	98.91	0.096	0.555	Bacteroidetes	78681	1835699
bin.27	98.77	0.862	0.459	Lachnospiraceae	58680	3374616
bin.77	98.66	0.111	0.322	Bacteria	115582	1920454
bin.52	98.65	0.377	0.503	Bacteroidales	35881	2555729
bin.8	98.51	1.445	0.465	Bacteroidales	59517	4018815
bin.2	98.49	1.679	0.436	Lachnospiraceae	61441	3200294
bin.82	98.3	0.83	0.532	Bacteroidales	25561	2739467
bin.71	98.17	0.377	0.462	Bacteroidales	56227	3805920
bin.50	98.1	0.377	0.512	Bacteroidales	36866	2445970
bin.32	98.06	0.754	0.49	Bacteroidales	36323	2425631
bin.3	98.02	0.632	0.428	Clostridiales	21346	2951892
bin.62	97.97	1.689	0.349	Clostridiales	106907	3807194
bin.91	97.66	0.69	0.507	Clostridiales	15785	3236094
bin.86	97.63	0.641	0.575	Bacteroidetes	74096	2017502
bin.5	97.54	0	0.486	Bacteroidales	38075	2165589
bin.90	97.31	0	0.554	Clostridiales	28593	2460616

bin.1	97.23	0.377	0.471	Bacteroidales	58546	2782804
bin.73	97.2	0	0.556	Bacteroidales	75080	2108598
bin.75	97.19	0.22	0.52	Bacteroidales	34780	2435852
bin.59	97.11	0.096	0.451	Bacteroidales	33677	4499876
bin.11	97.01	3.16	0.499	Lachnospiraceae	60142	4784343
bin.47	96.97	0.084	0.613	Deltaproteobacte	31554	2651868
bin.35	96.66	1.908	0.539	Bacteroidales	29588	2548811
bin.45	96.56	0.754	0.488	Bacteroidales	54821	2683452
bin.40	96.43	1.532	0.457	Lachnospiraceae	32929	3930209
bin.81	96.37	0.026	0.484	Clostridia	87062	1947378
bin.64	96.36	1.125	0.408	Lachnospiraceae	43881	3252816
bin.42	96.2	3.256	0.488	Clostridiales	60389	4296927
bin.96	96.1	0.754	0.493	Bacteroidales	73041	2039615
bin.24	95.95	0.476	0.49	Bacteroidetes	47101	2304959
bin.54	95.89	0.377	0.524	Bacteroidales	41593	2101556
bin.69	95.63	1.006	0.577	Clostridiales	23684	2620716
bin.4	95.56	2.873	0.456	Lachnospiraceae	33355	3773938
bin.76	95.31	0.566	0.561	Bacteroidales	29905	2393797
bin.18	95.16	0.099	0.496	Clostridia	71635	1553536
bin.13	95.11	1.582	0.462	Clostridiales	16296	3565720
bin.49	94.99	0	0.511	Bacteroidales	23805	2606383
bin.89	94.97	1.174	0.447	Clostridiales	25197	2697718
bin.84	94.92	1.442	0.595	Bacteroidetes	61027	1884839
bin.26	94.59	1.898	0.52	Clostridiales	16955	2779719
bin.30	94.4	0.531	0.43	Lachnospiraceae	32653	2632394

bin.15	94.36	0.377	0.535	Bacteroidales	17501	2136753
bin.58	94.26	0.766	0.506	Lachnospiraceae	29620	3725586
bin.74	93.9	1.436	0.497	Lachnospiraceae	52444	3418729
bin.63	93.14	1.724	0.484	Lachnospiraceae	33816	3028963
bin.60	92.95	0.671	0.441	Clostridiales	12606	2406081
bin.28	92.9	1.342	0.567	Clostridiales	27416	2237219
bin.95	92.35	0.316	0.461	Clostridiales	22181	2994123
bin.93	92.11	0	0.596	Clostridiales	20279	2158308
bin.9	90.71	0.574	0.455	Lachnospiraceae	39027	3126149
bin.94	90.71	1.221	0.484	Lachnospiraceae	16400	3057216
bin.48	90.24	1.442	0.61	Bacteroidetes	34739	1810155
bin.38	90.24	3.605	0.661	Bacteroidetes	46580	1886446
bin.68	90.15	0.104	0.466	Clostridiales	69014	1516749
bin.19	88.85	1.209	0.524	Clostridia	18169	1529711
bin.20	87.4	1.883	0.58	Bacteroidetes	11227	2054159
bin.36	86.41	0.377	0.553	Bacteroidales	35791	1981770
bin.88	86.32	1.006	0.641	Clostridiales	12191	2124177
bin.22	84.25	1.949	0.506	Lachnospiraceae	9108	3842172
bin.34	83.41	1.22	0.47	Lachnospiraceae	17734	2690015
bin.70	82.17	3.481	0.515	Clostridiales	11252	3029590
bin.46	81.67	2.419	0.544	Clostridia	9019	1298172
bin.57	81.21	2.709	0.466	Clostridiales	9559	2296673
bin.29	80.69	2.125	0.275	Clostridiales	29852	2001281
bin.83	79.16	1.509	0.549	Bacteroidales	13498	1812869
bin.79	79.01	0.94	0.475	Clostridia	39117	1428043

bin.39	78.32	0.24	0.642	Bacteroidetes	67571	1420276
bin.66	77.46	0.943	0.424	Bacteroides	24629	3626183
bin.10	76.98	0	0.537	Bacteroidales	79963	1617642
bin.53	76.43	0.364	0.389	Lactobacillales	6746	1315507
bin.17	76.02	1.545	0.458	Lachnospiraceae	7144	2089840
bin.65	74.37	1.072	0.467	Lachnospiraceae	9192	2383007
bin.16	73.72	0.142	0.4	Clostridiales	4974	1872578
bin.44	72.71	2.186	0.458	Clostridiales	6046	2658867
bin.67	72.41	0	0.342	Bacteria	9283	1372279
bin.51	71.86	0.94	0.685	Actinobacteria	5368	1665037
bin.61	71.5	3.448	0.611	Bacteria	43168	1906036
bin.98	69.51	1.628	0.48	Lachnospiraceae	5392	2939059
bin.31	69.22	0.34	0.511	Clostridia	4948	1472315
bin.78	67.53	1.169	0.451	Lachnospiraceae	6337	1743649
bin.7	67.5	2.011	0.449	Lachnospiraceae	5877	2593795
bin.43	65.39	8.771	0.512	Bacteria	5453	2786641
bin.80	65.08	0	0.463	Bacteria	5928	1920759
bin.56	63.27	0	0.629	Bacteria	9259	1650627
bin.99	60.44	0.832	0.51	Clostridia	4880	1328570
bin.87	60.34	1.724	0.517	Bacteria	7107	1900714
bin.21	60.19	0.377	0.53	Bacteroidales	28213	1477141
bin.14	57.89	7.017	0.522	Bacteria	6908	3339043
bin.12	57.07	0.806	0.61	Actinobacteria	3719	1219567
bin.72	53.77	0.671	0.472	Clostridiales	4493	1088661

HOST 2						
bin	completenes	contaminati	GC	lineage	N50	size
bin.34	99.7	0.403	0.612	Actinobacteria	40424	2082899
bin.63	99.46	0.949	0.458	Clostridiales	24310	3506964
bin.1	98.85	2.445	0.489	Lachnospiraceae	32293	3640206
bin.16	98.83	1.265	0.495	Clostridiales	53843	3709762
bin.24	98.22	2.586	0.312	Bacteria	34195	2126577
bin.73	98.16	0.754	0.486	Bacteroidales	51276	2797940
bin.44	98.14	0.185	0.465	Bacteroidales	43823	3992871
bin.53	98.08	0.632	0.364	Clostridiales	28648	2974023
bin.8	98.07	0.824	0.495	Lachnospiraceae	17552	3286876
bin.60	97.97	1.351	0.349	Clostridiales	122048	3815489
bin.52	97.92	0	0.497	Bacteroidales	63521	2692333
bin.58	97.5	1.724	0.491	Lachnospiraceae	67956	3888161
bin.55	97.46	0.632	0.441	Clostridiales	63483	3551181
bin.25	97.23	0.377	0.473	Bacteroidales	45474	2749441
bin.19	97.07	0.158	0.526	Clostridiales	24713	3645798
bin.13	97.01	2.586	0.496	Lachnospiraceae	55590	4928967
bin.9	96.97	2.237	0.454	Clostridiales	31916	3536371
bin.42	96.87	1.509	0.49	Bacteroidales	25296	2431516
bin.67	96.79	1.614	0.443	Lachnospiraceae	27896	3911318
bin.41	96.63	0.949	0.427	Clostridiales	29220	3092913
bin.64	96.47	0	0.555	Clostridiales	21221	2474903
bin.47	96.33	1.878	0.447	Lachnospiraceae	21323	3365571
bin.12	96.29	2.298	0.39	Lachnospiraceae	44044	4617425
bin.39	96.24	3.787	0.461	Lachnospiraceae	33372	3795876

bin.20	95.99	1.919	0.34	Clostridiales	16228	3567545
bin.70	95.87	2.274	0.41	Lachnospiraceae	33904	3191732
bin.29	95.85	0	0.521	Deltaproteobacte	60915	2607547
bin.54	95.8	0.574	0.46	Lachnospiraceae	27310	3130208
bin.65	95.78	1.34	0.492	Lachnospiraceae	56258	4578749
bin.5	95.75	1.886	0.412	Bacteria	8723	1981678
bin.27	95.68	1.724	0.449	Lachnospiraceae	39939	3877444
bin.15	95.42	0.88	0.537	Bacteroidales	23297	2501869
bin.56	95.35	1.265	0.457	Clostridiales	18087	3981121
bin.66	95.3	0.671	0.585	Clostridiales	59578	2636626
bin.3	95.16	0.026	0.491	Clostridia	76903	1780053
bin.59	95.06	0.966	0.529	Lachnospiraceae	34877	3906988
bin.31	94.82	0.032	0.614	Deltaproteobacte	21416	2537085
bin.7	94.63	1.125	0.48	Lachnospiraceae	49077	3727699
bin.23	94.25	1.915	0.478	Lachnospiraceae	54431	3393297
bin.21	94.18	0.223	0.431	Clostridiales	31723	1731291
bin.43	93.02	3.102	0.459	Lachnospiraceae	18909	3563242
bin.36	92.35	0	0.599	Bacteroidetes	73595	1701131
bin.61	92.28	0	0.575	Clostridiales	30701	2655057
bin.51	91.3	0.167	0.639	Clostridiales	18797	2317017
bin.10	90.33	4.187	0.531	Lachnospiraceae	31827	3951351
bin.74	89.78	7.382	0.564	Clostridiales	18336	3016108
bin.46	89.08	1.265	0.502	Clostridiales	24428	3550909
bin.2	86.76	1.36	0.496	Clostridia	32900	2133156
bin.18	85.59	3.846	0.645	Bacteroidetes	31481	1497047

bin.11	84.77	0.671	0.587	Clostridiales	12524	2799391
bin.72	84.13	3.019	0.567	Lachnospiraceae	23884	3783788
bin.68	84.05	1.265	0.478	Clostridiales	11637	3003624
bin.50	82.13	0.387	0.341	Lactobacillus	8106	1496455
bin.57	81.74	1.677	0.632	Clostridiales	21631	2276749
bin.49	81.5	0.377	0.558	Bacteroidales	36142	1851138
bin.14	80.7	0	0.473	Bacteria	37154	3773315
bin.17	78.9	4.43	0.51	Clostridiales	20869	2947954
bin.6	77.82	0	0.613	Clostridiales	15538	1953481
bin.4	77.51	1.375	0.577	Clostridiales	5365	1668722
bin.30	75.73	2.148	0.456	Lachnospiraceae	6149	2441252
bin.71	73.56	1.724	0.502	Lachnospiraceae	54833	2552853
bin.62	66.18	0.48	0.562	Bacteroidetes	115269	1278404
bin.37	65.51	2.586	0.614	Bacteria	15081	1529062
bin.79	61.88	1.724	0.489	Bacteria	5070	1604227
bin.32	61.26	1.265	0.506	Clostridiales	7045	2261409
bin.80	59.17	8.727	0.443	Clostridia	3366	1570205
bin.81	57.92	7.051	0.661	Bacteroidetes	4369	1510583
bin.78	57.9	0	0.512	Clostridiales	4214	2797792
bin.26	57.74	0.335	0.55	Clostridiales	3846	860750
bin.69	57.05	0	0.542	Bacteria	9010	1279655
bin.76	56.72	0.026	0.512	Clostridia	4811	1317407
bin.45	52.8	1.123	0.294	Bacteria	3708	692391
bin.40	51.91	3.508	0.559	Bacteria	5071	2747653
bin.28	51.75	0	0.62	Bacteria	8818	2039031

bin.77	50.71	0.574	0.484	Lachnospiraceae	3657	1850741

Supplementary Table 2

Improved MAGs statistic after reassembled module of metaWRAP toolkit (Short-read assembly)						
HOST 1						
bin	completenes	contaminati	GC	lineage	N50	size
bin.10.strict	76.98	0	0.537	Bacteroidales	154747	1632377
bin.11.orig	97.01	3.16	0.499	Lachnospiraceae	60142	4784343
bin.12.permissive	59.09	0.073	0.611	Actinobacteria	4131	1273382
bin.13.strict	95.38	1.265	0.461	Clostridiales	31044	3696093
bin.14.strict	58.05	6.14	0.522	Bacteria	7471	3514109
bin.15.orig	94.36	0.377	0.535	Bacteroidales	17501	2136753
bin.16.permissive	75.42	0	0.401	Clostridiales	6827	1931815
bin.17.permissive	76.96	0.821	0.457	Lachnospiraceae	10207	2170035
bin.18.orig	95.16	0.099	0.496	Clostridia	71635	1553536
bin.19.strict	88.85	1.075	0.524	Clostridia	28398	1551096
bin.1.orig	97.23	0.377	0.471	Bacteroidales	58546	2782804
bin.20.permissive	89.48	1.121	0.579	Bacteroidetes	18435	2094506
bin.21.strict	60.19	0	0.529	Bacteroidales	37703	1516329
bin.22.orig	84.25	1.949	0.506	Lachnospiraceae	9108	3842172
bin.24.permissive	95.95	0.476	0.489	Bacteroidetes	56945	2386424
bin.25.strict	99.4	0.016	0.554	Bacteroidetes	153770	1866339
bin.26.orig	94.59	1.898	0.52	Clostridiales	16955	2779719
bin.27.strict	99.35	0.862	0.458	Lachnospiraceae	123009	3507707
bin.28.orig	92.9	1.342	0.567	Clostridiales	27416	2237219
bin.29.orig	80.69	2.125	0.275	Clostridiales	29852	2001281
bin.2.orig	98.49	1.679	0.436	Lachnospiraceae	61441	3200294
bin.30.orig	94.4	0.531	0.43	Lachnospiraceae	32653	2632394
bin.31.orig	69.22	0.34	0.511	Clostridia	4948	1472315

bin.32.orig	98.06	0.754	0.49	Bacteroidales	36323	2425631
bin.34.strict	84.83	1.44	0.47	Lachnospiraceae	24891	2754281
bin.35.orig	96.66	1.908	0.539	Bacteroidales	29588	2548811
bin.36.orig	86.41	0.377	0.553	Bacteroidales	35791	1981770
bin.37.orig	99.13	2.586	0.311	Bacteria	33412	2152298
bin.38.orig	90.24	3.605	0.661	Bacteroidetes	46580	1886446
bin.39.orig	78.32	0.24	0.642	Bacteroidetes	67571	1420276
bin.3.orig	98.02	0.632	0.428	Clostridiales	21346	2951892
bin.40.orig	96.43	1.532	0.457	Lachnospiraceae	32929	3930209
bin.42.strict	96.83	3.256	0.486	Clostridiales	67987	4521859
bin.43.permissive	67.14	8.771	0.512	Bacteria	6518	2987587
bin.44.strict	72.04	1.342	0.458	Clostridiales	7243	2699920
bin.45.orig	96.56	0.754	0.488	Bacteroidales	54821	2683452
bin.46.orig	81.67	2.419	0.544	Clostridia	9019	1298172
bin.47.orig	96.97	0.084	0.613	Deltaproteobacte	31554	2651868
bin.48.orig	90.24	1.442	0.61	Bacteroidetes	34739	1810155
bin.49.strict	96.88	0.251	0.51	Bacteroidales	47357	2655293
bin.4.orig	95.56	2.873	0.456	Lachnospiraceae	33355	3773938
bin.50.orig	98.1	0.377	0.512	Bacteroidales	36866	2445970
bin.51.orig	71.86	0.94	0.685	Actinobacteria	5368	1665037
bin.52.orig	98.65	0.377	0.503	Bacteroidales	35881	2555729
bin.53.strict	75.61	0	0.389	Lactobacillales	8885	1317543
bin.54.orig	95.89	0.377	0.524	Bacteroidales	41593	2101556
bin.55.strict	99.21	0.523	0.4	Lactobacillales	31924	2014931
bin.56.orig	63.27	0	0.629	Bacteria	9259	1650627

bin.57.permissive	81.42	2.056	0.466	Clostridiales	12664	2377846
bin.58.orig	94.26	0.766	0.506	Lachnospiraceae	29620	3725586
bin.59.orig	97.11	0.096	0.451	Bacteroidales	33677	4499876
bin.5.orig	97.54	0	0.486	Bacteroidales	38075	2165589
bin.60.orig	92.95	0.671	0.441	Clostridiales	12606	2406081
bin.61.strict	92.58	4.567	0.61	Bacteroidetes	63622	1941124
bin.62.orig	97.97	1.689	0.349	Clostridiales	106907	3807194
bin.63.orig	93.14	1.724	0.484	Lachnospiraceae	33816	3028963
bin.64.strict	96.5	1.125	0.409	Lachnospiraceae	63718	3336296
bin.65.strict	75.19	1.072	0.466	Lachnospiraceae	12142	2462025
bin.66.strict	79.6	1.213	0.424	Bacteroides	33711	3704909
bin.67.permissive	72.41	0	0.342	Bacteria	13173	1386583
bin.68.strict	90.15	0.104	0.465	Clostridiales	79038	1534140
bin.69.orig	95.63	1.006	0.577	Clostridiales	23684	2620716
bin.6.orig	99.4	0.503	0.523	Bacteroidales	58854	2474392
bin.70.strict	84.72	2.531	0.515	Clostridiales	14138	3121210
bin.71.strict	98.17	0.377	0.462	Bacteroidales	69763	3858799
bin.72.strict	52.54	0.335	0.473	Clostridiales	4745	1110873
bin.73.orig	97.2	0	0.556	Bacteroidales	75080	2108598
bin.74.orig	93.9	1.436	0.497	Lachnospiraceae	52444	3418729
bin.75.orig	97.19	0.22	0.52	Bacteroidales	34780	2435852
bin.76.orig	95.31	0.566	0.561	Bacteroidales	29905	2393797
bin.77.strict	98.66	0.111	0.322	Bacteria	127986	1943622
bin.78.orig	67.53	1.169	0.451	Lachnospiraceae	6337	1743649
bin.79.orig	79.01	0.94	0.475	Clostridia	39117	1428043

bin.7.orig	67.5	2.011	0.449	Lachnospiraceae	5877	2593795
bin.80.orig	65.08	0	0.463	Bacteria	5928	1920759
bin.81.orig	96.37	0.026	0.484	Clostridia	87062	1947378
bin.82.orig	98.3	0.83	0.532	Bacteroidales	25561	2739467
bin.83.orig	79.16	1.509	0.549	Bacteroidales	13498	1812869
bin.84.orig	94.92	1.442	0.595	Bacteroidetes	61027	1884839
bin.85.strict	99.03	0.48	0.597	Bacteroidetes	54475	2463447
bin.86.orig	97.63	0.641	0.575	Bacteroidetes	74096	2017502
bin.87.orig	60.34	1.724	0.517	Bacteria	7107	1900714
bin.88.orig	86.32	1.006	0.641	Clostridiales	12191	2124177
bin.89.strict	94.97	1.174	0.447	Clostridiales	28759	2784066
bin.8.orig	98.51	1.445	0.465	Bacteroidales	59517	4018815
bin.90.orig	97.31	0	0.554	Clostridiales	28593	2460616
bin.91.orig	97.66	0.69	0.507	Clostridiales	15785	3236094
bin.92.orig	99.03	0.075	0.459	Bacteroidales	40213	2253611
bin.93.orig	92.11	0	0.596	Clostridiales	20279	2158308
bin.94.strict	91.29	1.245	0.483	Lachnospiraceae	20481	3137776
bin.95.orig	92.35	0.316	0.461	Clostridiales	22181	2994123
bin.96.orig	96.1	0.754	0.493	Bacteroidales	73041	2039615
bin.98.orig	69.51	1.628	0.48	Lachnospiraceae	5392	2939059
bin.99.orig	60.44	0.832	0.51	Clostridia	4880	1328570
bin.9.strict	90.71	0.574	0.455	Lachnospiraceae	52131	3238417

HOST 2

bin	completenes	contaminati	GC	lineage	N50	size

bin.10.orig	90.33	4.187	0.531	Lachnospiraceae	31827	3951351
bin.11.orig	84.77	0.671	0.587	Clostridiales	12524	2799391
bin.12.orig	96.29	2.298	0.39	Lachnospiraceae	44044	4617425
bin.13.orig	97.01	2.586	0.496	Lachnospiraceae	55590	4928967
bin.14.strict	80.7	0	0.473	Bacteria	50415	3896100
bin.15.strict	96.74	0.943	0.537	Bacteroidales	41971	2543496
bin.16.orig	98.83	1.265	0.495	Clostridiales	53843	3709762
bin.17.strict	78.9	4.43	0.509	Clostridiales	24597	3024349
bin.18.orig	85.59	3.846	0.645	Bacteroidetes	31481	1497047
bin.19.orig	97.07	0.158	0.526	Clostridiales	24713	3645798
bin.1.orig	98.85	2.445	0.489	Lachnospiraceae	32293	3640206
bin.20.orig	95.99	1.919	0.34	Clostridiales	16228	3567545
bin.21.permissive	94.18	0.223	0.431	Clostridiales	38663	1765445
bin.23.strict	94.25	1.915	0.477	Lachnospiraceae	69827	3546007
bin.24.orig	98.22	2.586	0.312	Bacteria	34195	2126577
bin.25.orig	97.23	0.377	0.473	Bacteroidales	45474	2749441
bin.26.strict	58.75	0.335	0.549	Clostridiales	4233	869542
bin.27.orig	95.68	1.724	0.449	Lachnospiraceae	39939	3877444
bin.28.orig	51.75	0	0.62	Bacteria	8818	2039031
bin.29.strict	95.85	0	0.52	Deltaproteobacte	63836	2631747
bin.2.strict	86.76	0.382	0.495	Clostridia	42046	2152096
bin.30.strict	75.15	1.784	0.456	Lachnospiraceae	6934	2478610
bin.31.orig	94.82	0.032	0.614	Deltaproteobacte	21416	2537085
bin.32.strict	61.79	0.316	0.505	Clostridiales	7882	2282916
bin.34.permissive	99.7	0	0.612	Actinobacteria	46370	2127221

bin.36.orig	92.35	0	0.599	Bacteroidetes	73595	1701131
bin.37.strict	65.51	1.724	0.613	Bacteria	20469	1551145
bin.39.strict	96.81	3.361	0.46	Lachnospiraceae	48777	3883412
bin.3.orig	95.16	0.026	0.491	Clostridia	76903	1780053
bin.40.permissive	55.42	3.508	0.557	Bacteria	6131	2925187
bin.41.orig	96.63	0.949	0.427	Clostridiales	29220	3092913
bin.42.strict	96.87	1.32	0.49	Bacteroidales	32166	2457561
bin.43.permissive	93.1	2.978	0.457	Lachnospiraceae	24912	3730681
bin.44.strict	98.14	0.185	0.465	Bacteroidales	52048	4058500
bin.45.strict	51.63	0	0.296	Bacteria	3699	701803
bin.46.strict	89.24	1.265	0.5	Clostridiales	41930	3855259
bin.47.strict	96.41	1.399	0.447	Lachnospiraceae	32638	3432195
bin.49.orig	81.5	0.377	0.558	Bacteroidales	36142	1851138
bin.4.orig	77.51	1.375	0.577	Clostridiales	5365	1668722
bin.50.strict	82.6	0	0.341	Lactobacillus	10036	1504843
bin.51.orig	91.3	0.167	0.639	Clostridiales	18797	2317017
bin.52.strict	97.92	0	0.497	Bacteroidales	71182	2720290
bin.53.orig	98.08	0.632	0.364	Clostridiales	28648	2974023
bin.54.orig	95.8	0.574	0.46	Lachnospiraceae	27310	3130208
bin.55.orig	97.46	0.632	0.441	Clostridiales	63483	3551181
bin.56.strict	97.04	1.476	0.456	Clostridiales	41665	4133192
bin.57.orig	81.74	1.677	0.632	Clostridiales	21631	2276749
bin.58.orig	97.5	1.724	0.491	Lachnospiraceae	67956	3888161
bin.59.strict	95.06	0.966	0.527	Lachnospiraceae	52496	4022326
bin.5.orig	95.75	1.886	0.412	Bacteria	8723	1981678

bin.60.orig	97.97	1.351	0.349	Clostridiales	122048	3815489
bin.61.orig	92.28	0	0.575	Clostridiales	30701	2655057
bin.62.permissive	66.34	0.48	0.561	Bacteroidetes	150217	1315659
bin.63.strict	99.46	0.126	0.458	Clostridiales	30640	3597089
bin.64.orig	96.47	0	0.555	Clostridiales	21221	2474903
bin.65.strict	96.36	1.34	0.489	Lachnospiraceae	62845	4874996
bin.66.orig	95.3	0.671	0.585	Clostridiales	59578	2636626
bin.67.strict	96.79	1.614	0.442	Lachnospiraceae	41429	4052014
bin.68.strict	85.31	0.843	0.477	Clostridiales	16672	3074626
bin.69.strict	58.77	0	0.541	Bacteria	12225	1291558
bin.6.orig	77.82	0	0.613	Clostridiales	15538	1953481
bin.70.orig	95.87	2.274	0.41	Lachnospiraceae	33904	3191732
bin.71.orig	73.56	1.724	0.502	Lachnospiraceae	54833	2552853
bin.72.orig	84.13	3.019	0.567	Lachnospiraceae	23884	3783788
bin.73.orig	98.16	0.754	0.486	Bacteroidales	51276	2797940
bin.74.strict	89.64	6.711	0.562	Clostridiales	19170	3253029
bin.76.strict	57.56	0.026	0.511	Clostridia	5395	1326228
bin.77.permissive	51.86	0.574	0.484	Lachnospiraceae	4205	1917933
bin.78.strict	59.79	0.316	0.512	Clostridiales	4881	2844537
bin.79.orig	61.88	1.724	0.489	Bacteria	5070	1604227
bin.7.orig	94.63	1.125	0.48	Lachnospiraceae	49077	3727699
bin.80.strict	59.55	8.727	0.442	Clostridia	3611	1593799
bin.81.strict	58.72	6.81	0.66	Bacteroidetes	5033	1526372
bin.8.strict	98.31	0.726	0.494	Lachnospiraceae	20624	3352325
bin.9.orig	96.97	2.237	0.454	Clostridiales	31916	3536371

HOST 3						
bin	completenes	contaminati	GC	lineage	N50	size
bin.12.orig	56.89	0	0.334	Bacteria	9463	1704771
bin.13.orig	83.8	2.38	0.479	Lachnospiraceae	19191	2921087
bin.14.strict	94.83	0	0.525	Bacteroidales	38002	2414335
bin.15.strict	90.58	0.022	0.558	Bacteroidales	81112	1995562
bin.16.orig	91.98	1.323	0.464	Clostridiales	13695	3364973
bin.17.orig	98.95	0	0.402	Lactobacillales	19280	1946310
bin.18.orig	97.98	1.83	0.352	Clostridiales	68709	4093372
bin.19.strict	83.28	5.928	0.606	Clostridiales	12877	3217492
bin.1.orig	97.17	0.754	0.513	Bacteroidales	18181	2384148
bin.20.strict	73.87	0.061	0.467	Clostridiales	7603	1256848
bin.21.strict	89.82	0.625	0.445	Proteobacteria	9203	2170466
bin.22.orig	92.18	0.125	0.475	Bacteroidales	38643	2091212
bin.23.permissive	90.96	0	0.485	Clostridia	14391	1580129
bin.24.orig	96.7	0.377	0.461	Bacteroidales	42101	3762673
bin.25.strict	92.95	0.334	0.465	Bacteroidales	51190	3429045
bin.26.strict	97.9	0.451	0.502	Bacteroidales	85358	2496173
bin.27.orig	94.15	1.34	0.481	Lachnospiraceae	18056	3183055
bin.28.strict	55.17	0	0.619	Bacteria	14852	1263389
bin.29.strict	88.72	0.223	0.593	Clostridiales	18380	2211124
bin.2.permissive	95.28	0.566	0.483	Bacteroidales	56459	2125409
bin.30.orig	79.31	1.724	0.488	Bacteria	31649	2471349
bin.31.strict	97.14	0	0.554	Clostridiales	19655	2473807
bin.32.orig	92.33	0.69	0.508	Clostridiales	9783	3121770

bin.33.permissive	78.44	2.586	0.424	Bacteria	28608	3678629
bin.34.orig	96.64	0.377	0.505	Bacteroidales	26909	2490427
bin.35.strict	90.89	0.943	0.515	Bacteroidales	15296	2101670
bin.36.orig	96.9	0.476	0.488	Bacteroidetes	41732	2424486
bin.37.orig	75.15	0	0.529	Bacteria	20981	1915709
bin.38.strict	91.44	1.32	0.487	Bacteroidales	68149	2673349
bin.39.orig	62.27	0	0.514	Lachnospiraceae	24649	2247921
bin.3.orig	97.37	3.062	0.468	Clostridiales	18917	3519338
bin.40.strict	94	2.223	0.57	Clostridiales	27460	2092609
bin.41.orig	56.05	2.215	0.49	Clostridiales	5980	2005584
bin.42.strict	90.02	0	0.557	Bacteroidetes	92196	1549514
bin.43.strict	95.4	0.377	0.534	Bacteroidales	20246	2169639
bin.44.orig	96.28	0.158	0.464	Clostridiales	32268	2632571
bin.45.orig	68.33	0.328	0.323	Clostridiales	5979	1200408
bin.46.orig	94.49	2.586	0.508	Lachnospiraceae	19325	4193204
bin.47.orig	97.23	0.754	0.492	Bacteroidales	35934	2010931
bin.48.permissive	66.26	3.771	0.284	Clostridiales	6388	1904985
bin.49.orig	65.17	5.517	0.626	Bacteria	5440	2079370
bin.4.strict	73.66	2.852	0.587	Clostridiales	11450	2506497
bin.50.orig	83.55	2.666	0.25	Bacteria	5647	1303130
bin.51.orig	91.61	1.898	0.501	Clostridiales	21553	3374735
bin.52.strict	66.81	0	0.64	Bacteroidetes	43030	1232650
bin.53.orig	96.83	0.377	0.471	Bacteroidales	42725	2848092
bin.55.orig	96.3	0.188	0.51	Bacteroidales	36867	2596212
bin.56.strict	96.63	1.125	0.408	Lachnospiraceae	36323	3211107

bin.57.orig	90.55	5.689	0.602	Bacteroidetes	41917	1770622
bin.58.orig	78.54	1.442	0.582	Bacteroidetes	6486	1850782
bin.59.strict	66.37	0.862	0.663	Bacteria	29380	1797291
bin.5.orig	56.87	3.636	0.47	Bacteria	4360	1167942
bin.60.strict	95.73	2.562	0.473	Lachnospiraceae	36149	3993039
bin.61.strict	80.62	0.188	0.46	Bacteroidales	6589	1882964
bin.62.orig	90.1	1.525	0.5	Lachnospiraceae	40490	3276383
bin.63.orig	51.59	0	0.647	Bacteria	7879	1346186
bin.64.orig	89.17	2.083	0.516	Lachnospiraceae	23154	3564100
bin.65.strict	87.81	0.806	0.651	Actinobacteria	10533	2245479
bin.66.orig	96.51	0.022	0.539	Bacteroidales	31931	2503497
bin.67.strict	72.64	0.388	0.532	Bacteroidales	68016	1844859
bin.68.strict	96.27	0.377	0.551	Bacteroidales	52329	2093210
bin.69.orig	54.23	8.434	0.527	Deltaproteobacte	3132	1556324
bin.6.orig	96.72	0	0.385	Lactobacillales	49642	1776407
bin.70.orig	77.31	1.842	0.462	Lachnospiraceae	9607	2990024
bin.71.orig	92.32	0	0.534	Bacteroidales	16311	2633102
bin.72.permissive	52.36	0.136	0.432	Lachnospiraceae	3726	1371025
bin.73.orig	50.53	3.446	0.562	Lachnospiraceae	3941	3072046
bin.74.orig	68.03	0.447	0.451	Clostridiales	5926	1982126
bin.75.permissive	61.22	0.479	0.463	Clostridiales	4740	2002630
bin.76.strict	55.08	5.263	0.474	Bacteria	3368	2349511
bin.77.strict	80.39	2.475	0.526	Bacteroidales	5990	1843706
bin.7.orig	87.53	0.188	0.569	Bacteroidales	17360	2075839
bin.8.orig	86.54	1.594	0.513	Lachnospiraceae	9987	3303725

bin.9.strict	79.34	0.488	0.319	Clostridiales	6279	2160281

Supplementary Table 3

GTDB-TK classification af the MAGs reconstructed from short-read metagenomics

HOST 1

user_genome	classification	fastani_r	fastani_r	fastani_taxonomy	fastani_a	fastani_a	closest_placement_reference
bin.1.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002362	GCF_00383	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002362	98.93	0.92	GCF_003833075.1
bin.10.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA3263;s__UBA3263 sp001689	GCF_01195	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA3263;s__UBA3263 sp001689	99.29	0.68	GCF_011959405.1
bin.11.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774375.1
bin.12.permiss	d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Eggerthellaceae;g__D16-63;s__	N/A	N/A	N/A	N/A	N/A	GCF_003612475.1
bin.13.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__UMGS1370 sp009	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__UMGS1370 sp009	99.06	0.86	GCA_009774615.1
bin.14.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	GCA_002490995.1
bin.15.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.16.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Monoglobales_A;f__UBA1381;g__CAG-41;s__CAG-41 sp010206265	GCF_01020	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Monoglobales_A;f__UBA1381;g__CAG-41;s__CAG-41 sp010206265	97.78	0.91	GCF_010206265.1
bin.17.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__1XD42-69 sp009	GCF_00991	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__1XD42-69 sp009	99.61	0.96	GCF_009911505.1
bin.18.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UBA11940;s__	N/A	N/A	N/A	N/A	N/A	GCA_003979335.1
bin.19.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UMGS1004;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.2.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__1XD42-69 sp003	GCF_00361	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__1XD42-69 sp003	99.21	0.92	GCF_003612565.1
bin.20.permiss	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	GCA_900546065.1
bin.21.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	GCA_00236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	99.16	0.8	GCA_002361215.1
bin.22.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Schaedlerella;s__Schaedler	GCF_00388	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Schaedlerella;s__Schaedler	97.52	0.89	GCF_003885045.1
bin.24.permiss	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA932;g__RC9;s__RC9 sp002298075	GCA_00229	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA932;g__RC9;s__RC9 sp002298075	98.45	0.85	GCA_002298075.1
bin.25.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp003	GCA_00397	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp003	99.81	0.89	GCA_003979135.1
bin.26.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.27.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp009774455	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp009774455	99.19	0.79	GCA_009774455.1
bin.28.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Marseille-P3106;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.29.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__TANB77;f__CAG-508;g__HGM13634;s__	N/A	N/A	N/A	N/A	N/A	GCF_900770385.1
bin.3.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.30.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911505.1
bin.31.orig	d__Bacteria;p__Firmicutes_B;c__Dehalobacteriia;o__UBA4068;f__UBA5755;g__;s__	N/A	N/A	N/A	N/A	N/A	GCA_002398485.1
bin.32.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00476	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.96	0.97	GCF_004766125.1
bin.34.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__ASTD01;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.35.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	97.86	0.88	GCA_009775275.1
bin.36.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Paramuribaculum;s__Paramuri	GCF_00302	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Paramuribaculum;s__Paramuri	98.72	0.85	GCF_003024925.1
bin.37.orig	d__Bacteria;p__Deferribacterota;c__Deferribacteres;o__Deferribacterales;f__Mucispirillaceae;g__Mucispirillum	GCF_00048	95	d__Bacteria;p__Deferribacterota;c__Deferribacteres;o__Deferribacterales;f__Mucispirillaceae;g__Mucispirillum	99.97	0.99	GCF_000487995.1
bin.38.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	GCF_90236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	99.93	0.98	GCF_902362705.1
bin.39.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp009	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp009	98.78	0.8	GCA_009774895.1
bin.4.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__	N/A	N/A	N/A	N/A	N/A	GCF_000403215.1
bin.40.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_J;s__Eubacteri	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_J;s__Eubacteri	99.12	0.8	GCA_009774535.1
bin.42.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__UBA7182 sp009	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__UBA7182 sp009	98.43	0.82	GCA_009774145.1
bin.43.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.44.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.45.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.77	0.98	GCF_004803915.1
bin.46.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UMGS1004;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.47.orig	d__Bacteria;p__Desulfobacterota;c__Desulfovibrionia;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Bilophil	N/A	N/A	N/A	N/A	N/A	N/A
bin.48.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	GCA_002428825.1
bin.49.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	GCF_00220	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	98.96	0.96	GCF_002201515.1
bin.5.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	98.77	0.92	GCA_009775535.1
bin.50.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp001689	GCA_00168	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp001689	99.05	0.9	GCA_001689485.1
bin.51.orig	d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Eggerthellaceae;g__D16-63;s__D16-63	GCF_00361	95	d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Eggerthellaceae;g__D16-63;s__D16-63	98.02	0.9	GCF_003612475.1
bin.52.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp013316	GCA_01331	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp013316	99.25	0.87	GCA_013316675.1
bin.53.strict	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Limosilactobacillus;s__Limosi	GCF_00307	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Limosilactobacillus;s__Limosi	95.7	0.92	GCF_013694365.1
bin.54.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	GCA_00236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	98.59	0.89	GCA_002361155.1
bin.55.strict	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Ligilactobacillus;s__Ligilact	GCF_00159	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Ligilactobacillus;s__Ligilact	96.95	0.84	GCF_001591685.1
bin.56.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp900	GCA_90055	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp900	98.42	0.78	GCA_900552955.1
bin.57.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__CAG-56;s__CAG-56 sp004793	GCF_00479	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__CAG-56;s__CAG-56 sp004793	99.02	0.93	GCF_004793585.1
bin.58.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__UBA3282 sp009	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__UBA3282 sp009	98.23	0.69	GCA_009774655.1
bin.59.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Tannerellaceae;g__Parabacteroides;s__Parabact	GCF_00001	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Tannerellaceae;g__Parabacteroides;s__Parabact	97.53	0.83	GCF_000012845.1
bin.6.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00302	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.83	0.97	GCF_003024805.1
bin.60.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Ruminiclostridium_E;s__	N/A	N/A	N/A	N/A	N/A	GCA_003508605.1
bin.61.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.62.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__TF01-11;s__TF01-11 sp009	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__TF01-11;s__TF01-11 sp009	99.49	0.85	GCA_009774235.1
bin.63.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911185.1
bin.64.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911185.1
bin.65.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7109;s__	N/A	N/A	N/A	N/A	N/A	GCA_002492335.1
bin.66.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	GCF_00168	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	99.72	0.89	GCF_001688725.2
bin.67.permiss	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__Lactobacillu	GCF_00143	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__Lactobacillu	98.3	0.89	GCF_001436695.1
bin.68.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__CAG-917;g__CAG-475;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.69.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	GCA_900543705.1
bin.7.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__COE1 sp000403335	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__COE1 sp000403335	99.07	0.95	GCF_000403335.2
bin.70.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3402;s__	N/A	N/A	N/A	N/A	N/A	GCA_002358555.1
bin.71.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002493	GCF_00376	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002493	99.83	0.99	GCF_003762615.1
bin.72.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__CAG-272;g__UMGS1815;s__	N/A	N/A	N/A	N/A	N/A	GCF_011957885.1
bin.73.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp002490	GCA_00249	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp002490	99.25	0.94	GCA_002490635.1
bin.74.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__	N/A	N/A	N/A	N/A	N/A	GCF_003611805.1
bin.75.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.76.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp002491	GCF_00410	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp002491	99.24	0.92	GCF_004102785.1
bin.77.strict	d__Bacteria;p__Firmicutes;c__Bacilli;o__Acholeplasmatales;f__Anaeroplasmataceae;g__UMGS268;s__	N/A	N/A	N/A	N/A	N/A	GCA_900540705.1
bin.78.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__MD308;s__MD308 sp010206225	GCA_01020	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__MD308;s__MD308 sp010206225	99.15	0.88	GCA_010206225.2
bin.79.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UBA11940;s__	N/A	N/A	N/A	N/A	N/A	GCA_003979335.1
bin.8.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	GCF_00479	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	99.96	0.92	GCF_004793475.1
bin.80.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__1XD42-69 sp011	GCF_01195	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__1XD42-69 sp011	98.91	0.87	GCF_011959925.1
bin.81.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A
bin.82.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00376	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	96.63	0.94	GCF_003762875.1
bin.83.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	98.19	0.94	GCA_009775225.1
bin.84.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	GCA_00242	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	98.66	0.91	GCA_002428825.1
bin.85.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes_A;s__	N/A	N/A	N/A	N/A	N/A	GCA_900549685.1
bin.86.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002362235	GCA_00236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	98.53	0.93	GCA_002362235.1
bin.87.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp003979	GCA_00397	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp003979	99.34	0.7	GCA_003979155.1

bin.88.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UBA9475;s__	N/A	N/A	N/A		N/A	N/A	GCA_004560375.1	
bin.89.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_F;s__	N/A	N/A	N/A		N/A	N/A	N/A	
bin.9.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Kineothrix;s__Kineothrix s	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Kineothrix;s__Kineothrix s	99.01	0.91	GCF_000403275.1		
bin.90.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Anaerotruncus;s__Anaerotr	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Anaerotruncus;s__Anaerotr	99.01	0.97	GCF_000403395.2		
bin.91.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__	N/A	N/A	N/A		N/A	N/A	GCA_009774615.1	
bin.92.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	99.82	0.88	GCF_004801635.1		
bin.93.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Acutalibacteraceae;g__Acutalibacter;s__	N/A	N/A	N/A		N/A	N/A	N/A	
bin.94.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp000403315	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp000403315	99	0.95	GCF_000403315.2		
bin.95.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7160;s__UBA7160 sp00361	GCF_00361	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7160;s__UBA7160 sp00361	99.41	0.96	GCF_003612585.1		
bin.96.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp003979	GCA_00397	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp003979	99.89	0.98	GCA_003979075.1		
bin.98.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_J;s__	N/A	N/A	N/A		N/A	N/A	N/A	
bin.99.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__CAG-552;g__;s__	N/A	N/A	N/A		N/A	N/A	N/A	

HOST 2

user_genome	classification	fastani_r	fastani_r	fastani_taxonomy	fastani_a	fastani_a	closest_placement_reference	
bin.1.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Kineothrix;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.10.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	GCF_003612485.1	
bin.11.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UMGS1872;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.12.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__COE1 sp009774395	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__COE1 sp009774395	98.93	0.8	GCA_009774395.1	
bin.13.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774375.1	
bin.14.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__CAG-95;s__CAG-95 sp0099174	GCF_00991	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__CAG-95;s__CAG-95 sp0099174	98.91	0.91	GCF_009917455.1	
bin.15.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	97.78	0.89	N/A	
bin.16.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	GCF_000875945.1	
bin.17.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3402;s__	N/A	N/A	N/A	N/A	N/A	GCA_002358555.1	
bin.18.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp009	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp009	98.5	0.81	GCA_009774895.1	
bin.19.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Clostridium_Q;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911305.1	
bin.2.strict	d__Bacteria;p__Firmicutes_B;c__Dehalobacteriia;o__UBA4068;f__UBA5755;g__;s__	N/A	N/A	N/A	N/A	N/A	GCA_002398485.1	
bin.20.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.21.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__UBA3700;g__RACS-045;s__	N/A	N/A	N/A	N/A	N/A	GCA_013316135.1	
bin.23.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp000403315	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp000403315	98.81	0.94	GCF_000403315.2	
bin.24.orig	d__Bacteria;p__Deferribacterota;c__Deferribacteres;o__Deferribacterales;f__Mucispirillaceae;g__Mucispirillum	GCF_00048	95	d__Bacteria;p__Deferribacterota;c__Deferribacteres;o__Deferribacterales;f__Mucispirillaceae;g__Mucispirillum	99.97	1	GCF_000487995.1	
bin.25.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002362	GCF_00383	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002362	99	0.92	GCF_003833075.1	
bin.26.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.27.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Kineothrix;s__Kineothrix s	GCF_01195	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Kineothrix;s__Kineothrix s	98.52	0.73	GCF_011958945.1	
bin.28.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Lawsonibacter;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911595.1	
bin.29.strict	d__Bacteria;p__Desulfobacterota;c__Desulfovibrionia;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Desulfov	GCA_00977	95	d__Bacteria;p__Desulfobacterota;c__Desulfovibrionia;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Desulfov	99.88	0.98	GCA_009773975.1	
bin.3.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UBA11940;s__	N/A	N/A	N/A	N/A	N/A	GCA_003979335.1	
bin.30.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__ASTD01;s__	N/A	N/A	N/A	N/A	N/A	GCF_011960625.1	
bin.31.orig	d__Bacteria;p__Desulfobacterota;c__Desulfovibrionia;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Bilophil	N/A	N/A	N/A	N/A	N/A	N/A	
bin.32.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA9502;s__	N/A	N/A	N/A	N/A	N/A	GCF_003669055.1	
bin.34.permiss	d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Eggerthellaceae;g__D16-63;s__	N/A	N/A	N/A	N/A	N/A	GCF_003612475.1	
bin.36.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	GCA_00242	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	98.6	0.94	GCA_002428825.1	
bin.37.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	GCA_002428825.1	
bin.39.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_J;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774535.1	
bin.4.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Marseille-P3106;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.40.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA9502;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.41.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.42.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella freteri	GCF_00476	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.97	0.97	GCF_004766125.1	
bin.43.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__	N/A	N/A	N/A	N/A	N/A	GCF_000403215.1	
bin.44.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	GCF_00479	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	99.8	0.94	GCF_004793475.1	
bin.45.strict	d__Bacteria;p__Firmicutes;c__Bacilli;o__RF39;f__UBA660;g__UMGS1217;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.46.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__UBA7182 sp00977	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__UBA7182 sp00977	98.43	0.86	GCA_009774145.1	
bin.47.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__COE1 sp000403335	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__COE1 sp000403335	98.98	0.93	GCF_000403335.2	
bin.49.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Paramuribaculum;s__Paramuri	GCF_00302	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Paramuribaculum;s__Paramuri	98.77	0.85	GCF_003024925.1	
bin.5.orig	d__Bacteria;p__Firmicutes;c__Bacilli;o__Erysipelotrichales;f__Erysipelotrichaceae;g__C-19;s__	N/A	N/A	N/A	N/A	N/A	GCF_009917405.1	
bin.50.strict	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__Lactobacillu	GCF_00143	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__Lactobacillu	98.24	0.92	GCF_001436695.1	
bin.51.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UBA9475;s__	N/A	N/A	N/A	N/A	N/A	GCA_004560375.1	
bin.52.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum sp002492595	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	96.15	0.92	GCF_004803695.1	
bin.53.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.54.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp009774455	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp009774455	99.1	0.74	GCA_009774455.1	
bin.55.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA9502;s__UBA9502 sp00991	GCF_00991	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA9502;s__UBA9502 sp00991	96.37	0.87	GCF_009911065.1	
bin.56.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__UMGS1370 sp009	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__UMGS1370 sp009	99.02	0.82	GCA_009774615.1	
bin.57.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UMGS1872;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.58.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__	N/A	N/A	N/A	N/A	N/A	GCF_003611805.1	
bin.59.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Roseburia;s__	N/A	N/A	N/A	N/A	N/A	GCA_900548205.1	
bin.6.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Lawsonibacter;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.60.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__TF01-11;s__TF01-11 sp00977	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__TF01-11;s__TF01-11 sp00977	99	0.8	GCA_009774235.1	
bin.61.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Acutalibacteraceae;g__Acutalibacter;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.62.permiss	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	GCA_003979135.1	
bin.63.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.64.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Anaerotruncus;s__Anaerotr	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Anaerotruncus;s__Anaerotr	99.14	0.96	GCF_000403395.2	
bin.65.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__UBA3282 sp00977	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__UBA3282 sp00977	97.91	0.8	GCA_009774655.1	
bin.66.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Acutalibacteraceae;g__Acutalibacter;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.67.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_J;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.68.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Anaerosaccharophilus;s__	N/A	N/A	N/A	N/A	N/A	GCF_009917555.1	
bin.69.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA3263;s__UBA3263 sp001689	GCF_01195	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA3263;s__UBA3263 sp001689	99.05	0.68	GCF_011959405.1	
bin.7.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	GCF_900248245.1	
bin.70.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911185.1	
bin.71.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.72.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	GCF_003612485.1	
bin.73.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.82	0.93	GCF_004803915.1	
bin.74.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Lawsonibacter;s__	N/A	N/A	N/A	N/A	N/A	GCF_009917615.1	
bin.76.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__CAG-552;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.77.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD8-76;s__	N/A	N/A	N/A	N/A	N/A	GCF_003611955.1	
bin.78.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3402;s__	N/A	N/A	N/A	N/A	N/A	GCA_002358555.1	

bin.79.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977		95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	98.12	0.83	GCA_009775535.1	
bin.8.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__	N/A	N/A		N/A	N/A	N/A	N/A	
bin.80.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UBA11940;s__	N/A	N/A		N/A	N/A	N/A	GCA_003503945.1	
bin.81.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	GCF_90236		95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	99.15	0.95	GCF_902362705.1	
bin.9.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A		N/A	N/A	N/A	N/A	

HOST 3

user_genome	classification	fastani_r	fastani_r	fastani_taxonomy	fastani_a	fastani_a	closest_placement_reference	
bin.1.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp001689	GCA_00168	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp001689	99.12	0.91	GCA_001689485.1	
bin.12.orig	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__Lactobacillu	GCF_00143	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__Lactobacillu	97.27	0.81	GCF_001436695.1	
bin.13.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__VSOB01;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774625.1	
bin.14.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	GCA_00236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	99.19	0.88	GCA_002361215.1	
bin.15.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp002490	GCA_00249	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp002490	99.27	0.95	GCA_002490635.1	
bin.16.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__UMGS1370 sp009	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__UMGS1370 sp009	99.15	0.89	GCA_009774615.1	
bin.17.orig	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Ligilactobacillus;s__Ligilact	GCF_00159	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Ligilactobacillus;s__Ligilact	96.99	0.86	GCF_001591685.1	
bin.18.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__TF01-11;s__TF01-11 sp00977	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__TF01-11;s__TF01-11 sp00977	98.86	0.78	GCA_009774235.1	
bin.19.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UMGS1872;s__UMGS1872 sp0	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UMGS1872;s__UMGS1872 sp0	97.81	0.66	GCA_009774055.1	
bin.2.permissi	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	98.72	0.91	GCA_009775535.1	
bin.20.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__CAG-917;g__CAG-475;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.21.strict	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Turicimonas;s	GCF_00222	95	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Turicimonas;s	99.93	1	GCF_002221595.1	
bin.22.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__	N/A	N/A	N/A	N/A	N/A	GCF_900768685.1	
bin.23.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.24.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002493	GCF_00376	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002493	99.84	0.98	GCF_003762615.1	
bin.25.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	GCF_00479	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	99.94	0.97	GCF_004793475.1	
bin.26.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	96.26	0.95	GCF_004803695.1	
bin.27.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp000403315	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__14-2 sp000403315	98.95	0.95	GCF_000403315.2	
bin.28.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	GCA_002428825.1	
bin.29.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Acutalibacteraceae;g__Acutalibacter;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.3.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.30.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella freteri	GCF_00476	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.87	0.96	GCF_004766125.1	
bin.31.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Anaerotruncus;s__Anaerotr	GCF_00040	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Ruminococcaceae;g__Anaerotruncus;s__Anaerotr	99.05	0.97	GCF_000403395.2	
bin.32.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774615.1	
bin.33.permiss	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	GCF_00168	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides	99.43	0.92	GCF_001688725.2	
bin.34.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp013316	GCA_01331	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp013316	99.19	0.84	GCA_013316675.1	
bin.35.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp003979	GCA_00397	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp003979	99.67	0.82	GCA_003979155.1	
bin.36.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA932;g__RC9;s__RC9 sp002298075	GCA_00229	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__UBA932;g__RC9;s__RC9 sp002298075	98.5	0.87	GCA_002298075.1	
bin.37.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA3263;s__UBA3263 sp001689	GCF_01195	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA3263;s__UBA3263 sp001689	99.19	0.73	GCF_011959405.1	
bin.38.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.74	0.96	GCF_004803915.1	
bin.39.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.4.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UMGS1872;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.40.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Marseille-P3106;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.41.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774135.1	
bin.42.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp003	GCA_00397	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp003	99.85	0.75	GCA_003979135.1	
bin.43.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.44.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__CAG-56;s__CAG-56 sp0047935	GCF_00479	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__CAG-56;s__CAG-56 sp0047935	99.05	0.94	GCF_004793585.1	
bin.45.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__TANB77;f__CAG-508;g__HGM13634;s__	N/A	N/A	N/A	N/A	N/A	GCF_900770385.1	
bin.46.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__COE1;s__	N/A	N/A	N/A	N/A	N/A	GCA_009774375.1	
bin.47.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp003979	GCA_00397	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp003979	99.92	0.95	GCA_003979075.1	
bin.48.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__TANB77;f__CAG-508;g__HGM13634;s__	N/A	N/A	N/A	N/A	N/A	GCF_900770385.1	
bin.49.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Dysosmobacter;s__Dysosmo	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__Dysosmobacter;s__Dysosmo	97.08	0.75	GCA_009774015.1	
bin.5.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia_A;o__Christensenellales;f__Borkfalkiaceae;g__UBA11940;s__	N/A	N/A	N/A	N/A	N/A	GCA_003979335.1	
bin.50.orig	d__Bacteria;p__Firmicutes;c__Bacilli;o__RF39;f__UBA660;g__UBA3789;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.51.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__UBA7182 sp00977	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA7182;s__UBA7182 sp00977	98.8	0.91	GCA_009774145.1	
bin.52.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp009	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp009	98.69	0.73	GCA_009774895.1	
bin.53.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002362	GCF_00383	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002362	98.88	0.9	GCF_003833075.1	
bin.55.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	GCF_00220	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	98.86	0.96	GCF_002201515.1	
bin.56.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__14-2;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911185.1	
bin.57.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002428825	GCA_00242	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	98.56	0.94	GCA_002428825.1	
bin.58.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__	N/A	N/A	N/A	N/A	N/A	GCA_900546065.1	
bin.59.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002358415	GCF_90236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes;s__Alistipes sp002	99.89	0.98	GCF_902362705.1	
bin.6.orig	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Limosilactobacillus;s__Limosi	GCF_00307	95	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Limosilactobacillus;s__Limosi	95.4	0.85	GCF_014145615.1	
bin.60.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_J;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.61.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	GCF_00480	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Muribaculum;s__Muribaculum	99.53	0.89	GCF_004801635.1	
bin.62.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__	N/A	N/A	N/A	N/A	N/A	GCF_003611805.1	
bin.63.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Oscillospirales;f__Oscillospiraceae;g__UMGS1872;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.64.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	GCF_011959105.1	
bin.65.strict	d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Eggerthellaceae;g__Adlercreutzia;s_	GCF_00042	95	d__Bacteria;p__Actinobacteriota;c__Coriobacteriia;o__Coriobacteriales;f__Eggerthellaceae;g__Adlercreutzia;s_	97.69	0.93	GCF_000422625.1	
bin.66.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	GCA_00977	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-873;s__CAG-873 sp009775	97.84	0.89	GCA_009775275.1	
bin.67.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00302	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	98.77	0.97	GCF_003024805.1	
bin.68.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Paramuribaculum;s__Paramuri	GCF_00302	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Paramuribaculum;s__Paramuri	98.78	0.84	GCF_003024925.1	
bin.69.orig	d__Bacteria;p__Firmicutes_B;c__Dehalobacteriia;o__UBA4068;f__UBA5755;g__;s__	N/A	N/A	N/A	N/A	N/A	GCA_002398485.1	
bin.7.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp002491	GCF_00410	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__UBA7173;s__UBA7173 sp002491	99.22	0.96	GCF_004102785.1	
bin.70.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__ASTD01;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.71.orig	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	GCF_00376	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__Duncaniella;s__Duncaniella	96.67	0.92	GCF_003762875.1	
bin.72.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__1XD42-69;s__	N/A	N/A	N/A	N/A	N/A	GCF_009911505.1	
bin.73.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Acetatifactor;s__	N/A	N/A	N/A	N/A	N/A	GCF_003612485.1	
bin.74.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__Eubacterium_F;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.75.permiss	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UMGS1370;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.76.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__;s__	N/A	N/A	N/A	N/A	N/A	N/A	
bin.77.strict	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	GCA_00236	95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Bacteroidales;f__Muribaculaceae;g__CAG-485;s__CAG-485 sp002361	98.32	0.73	GCA_002361155.1	
bin.8.orig	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__UBA3282 sp00977	GCA_00977	95	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Lachnospiraceae;g__UBA3282;s__UBA3282 sp00977	98.15	0.72	GCA_009774655.1	
bin.9.strict	d__Bacteria;p__Firmicutes_A;c__Clostridia;o__Lachnospirales;f__Anaerotignaceae;g__ASF356;s__	N/A	N/A	N/A	N/A	N/A	GCF_000364165.1	

Supplementary Table 4

inStrain output

Genome-wide population genomic metrics

Host 1 - Young

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_	clinked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	divergent	reads_unf
bin.1.ori	332.199280	0.99974160	0.0003796	2782804	78	78	336	50.0863800	0.03010910	0.9992277	1	0.00036570	0.99998230	0.99999671	2.1106140	FALSE	1819	107.730620	8.84749880	0.9940076	49	9	9	1418	3474526	3601325	0.9980519	1427	7284541
bin.10.st	47.672696	0.99971020	0.0009754	1632377	31	31	47	11.8275710	0.00927490	0.9978448	1	0.00077260	0.99995450	0.99999381	2.3155170	FALSE	1279	50.880375	9.92164250	0.9985601	74	10	10	3868	294943	314756	0.9915341	3878	655094
bin.11.or	107.169550	0.99892500	0.0002632	4784343	127	127	107	23.2442620	0.01065510	0.9975998	1	0.00024680	0.99997380	0.99999161	4.4387693	FALSE	767	102.01694	0.65932100	0.9945299	125	40	40	533	1937908	2035055	0.9970034	573	4161256
bin.13.st	25.027987	0.99645460	0.0015995	3696093	296	296	25	13.8107200	0.00724180	0.96728160	0.99999990	0.00056790	0.99976420	0.9999804		FALSE	2830	43.079151	0.56358410	0.9923998	843	70	71	13085	348374	406081	0.9895173	13156	874827
bin.15.or	20.123013	0.99620870	0.0002345	2136753	161	161	19	33.1936520	0.02288100	0.98333540	0.9999999	5.94E-05	0.99999520	0.99999801	3.6260095	FALSE	98	72.091836	0.47546290	0.9366634	10	4	4	378	162786	174177	0.9978880	382	362839
bin.2.ori	90.108917	0.99962750	0.0002666	3200294	88	88	90	19.8686540	0.01113700	0.9985763	1	0.00025430	0.99998740	0.99999681	5.5787708	FALSE	309	83.074433	0.88661970	0.9945625	40	10	9	406	1083647	1126101	0.9972736	415	2296158
bin.21.st	38.717609	0.99889930	0.0008585	1516329	76	76	37	26.9372070	0.02198580	0.9947893	1	0.00052680	0.99990380	0.99999461	3.3364111	FALSE	872	42.612385	0.91231020	0.9993654	145	8	11	2735	222049	245074	0.9943579	2746	524485
bin.24.pe	172.012730	0.99881160	0.0002550	2386424	126	125	171	71.9480010	0.04682200	0.9959047	1	0.00024000	0.99996670	0.99998271	2.2917565	FALSE	676	122.33136	0.52077190	0.9742720	79	41	40	343	1544922	1620922	0.9891142	383	3373911
bin.25.st	78.623431	0.99930390	0.0002544	1866339	66	66	77	27.6830490	0.02033570	0.9971259	1	0.00023290	0.99997520	0.99998921	6.6957713	FALSE	161	62.341614	0.8314168	1	46	20	20	197	558173	585334	0.9893357	217	1206779
bin.27.st	135.647050	0.99797330	0.0003433	3507707	206	205	137	50.3347140	0.02703470	0.9915112	1	0.00030880	0.99994390	0.99997231	1.1797431	FALSE	1560	101.21794	0.61070360	0.9759324	195	96	97	1196	1791592	1927077	0.9839463	1293	4027345
bin.29.or	31.222630	0.99813660	0.0001721	2001281	120	120	28	14.5767810	0.01036630	0.9892893	1	0.00017270	0.99999340	0.99999542	2.2439058	FALSE	29	45.758620	0.54077050	0.9714131	13	9	9	159	230361	237583	0.9979412	168	483969
bin.32.or	29.302960	0.99886910	0.0002074	2425631	102	102	29	7.1995332	0.00464220	0.9943705	1	0.00022820	0.99999660	0.99999911	2.2203440	FALSE	36	32.888880	0.4627772	1	8	2	2	225	268105	277315	0.9979639	227	567013
bin.35.or	103.890400	0.99814690	0.0009546	2548811	100	100	105	35.7506130	0.02248150	0.9962453	1	0.00092620	0.99991060	0.99999681	5.5650266	FALSE	6841	84.183014	0.76580750	0.9906949	227	8	9	6257	1003215	1149443	0.9947287	6266	2343136
bin.36.or	115.430770	0.99862790	0.0008738	1981770	65	65	117	25.9436770	0.01848980	0.9971474	1	0.00083380	0.99987140	0.99999691	3.3243067	FALSE	3606	88.750138	0.9739010	1	254	6	6	3445	869046	934986	0.9968734	3451	1898371
bin.37.or	27.305583	0.99841000	0.0002117	2152298	88	88	25	14.9677890	0.01024440	0.9768066	1	0.00017140	0.99999040	0.99999853	3.8957409	FALSE	31	35.129032	0.7618478	1	20	3	3	304	216913	222731	0.9983771	307	450965
bin.38.or	63.620804	0.99490620	0.0002316	1886446	69	69	65	16.3008190	0.01191190	0.9903326	1	0.00021870	0.99998660	0.99999091	3.3297069	FALSE	12	23.916666	1	1	25	17	17	62	463914	486208	0.9965507	79	1001262
bin.39.or	199.404030	0.99872340	0.0002594	1420276	35	35	200	46.4079770	0.03903720	0.9977539	1	0.00024410	0.99998230	0.99999431	6.6320130	FALSE	142	107.48591	0.84257440	0.9869690	25	8	8	131	1088799	1124776	0.9982809	139	2271737
bin.4.ori	32.160832	0.99723020	0.0003752	3773938	169	169	32	10.1092050	0.00522720	0.9902282	1	0.00050110	0.99998920	0.99999751	6.6037022	FALSE	409	52.034229	0.82115660	0.9978739	40	9	9	2383	456038	478378	0.9976592	2392	981570
bin.40.or	44.709832	0.99344510	0.0002591	3930209	163	163	44	16.1588510	0.00818480	0.9882359	1	0.00023510	0.99998400	0.99999661	7.0444945	FALSE	103	68.339805	0.8646600	1	62	13	13	482	660878	694674	0.9974102	495	1429831
bin.42.st	108.288350	0.99692270	0.0005623	4521859	338	336	109	37.4568180	0.01774770	0.9862669	1	0.00050110	0.99992080	0.99996791	7.7473506	FALSE	2696	104.43063	0.67108080	0.9887196	353	143	142	4602	1847277	2033211	0.9840325	4744	4310152
bin.45.or	47.053742	0.99958110	0.0002083	2683452	65	65	47	9.2569504	0.00566460	0.9979343	1	0.00019810	0.99999620	0.99999851	1.1955947	FALSE	7	18	1	1	10	4	4	107	475806	488441	0.9982473	111	988035
bin.48.or	120.948410	0.97739140	0.0002667	1810155	74	74	122	42.3770360	0.03162680	0.9704633	1	0.00024210	0.99993900	0.99995611	6.6162545	FALSE	282	82.482269	0.85441640	0.9978820	107	77	77	239	838178	907384	0.9974139	316	1900094
bin.49.st	26.585568	0.99924860	0.0008517	2655293	108	108	27	9.4190146	0.00580390	0.99144760	0.9999999	0.00041280	0.99991260	0.99998631	4.5529519	FALSE	771	31.857328	0.90765790	0.9965417	230	36	36	4956	266957	295411	0.9937201	4992	614315
bin.5.ori	85.354408	0.99903990	0.0008426	2165589	74	74	86	18.9358780	0.01291170	0.9974865	1	0.00078740	0.99994620	0.99998811	3.1160732	FALSE	3876	81.484004	0.84971100	0.9936408	116	4	6	4685	695461	758039	0.9970431	4691	1533205
bin.50.or	30.709030	0.99867080	0.0007776	2445970	105	105	31	7.6501114	0.00491260	0.9931385	1	0.00046840	0.99995220	0.99999951	3.0092844	FALSE	843	35.533807	0.9244270	1	116	1	1	4437	283917	300970	0.9974241	4438	614281
bin.52.or	27.860940	0.99875180	0.0005595	2555729	114	114	28	7.0397593	0.00442320	0.9930869	1	9.92E-05	0.99992080	0.99999841	2.2421704	FALSE	307	29.081433	0.9368550	1	201	4	4	2402	268829	278936	0.9980296	2406	564007
bin.55.st	25.312471	0.99832500	0.0003159	2014931	106	106	25	12.1477290	0.00860320	0.99135500	0.9999999	0.00092450	0.99998490	0.99999741	6.6375630	FALSE	280	90.182142	0.37391260	0.9573353	30	5	5	876	191778	200742	0.9972840	881	413884
bin.58.or	48.526462	0.99856310	0.0009890	3725586	177	177	49	13.1318020	0.00683590	0.9940251	1	0.00090350	0.99996300	0.99999481	6.6361972	FALSE	4202	52.871489	0.84896790	0.9986060	137	19	19	10080	682559	764655	0.9955675	10099	1568251
bin.6.ori	37.030152	0.99922600	0.0002446	2474392	78	78	37	8.3015623	0.00529410	0.9965361	1	0.00023140	0.99999310	0.99999791	2.2867287	FALSE	57	44.877192	0.76610160	0.9884615	17	5	4	405	346808	357972	0.9981392	409	726535
bin.61.st	31.844184	0.99907010	0.0002439	1941124	85	85	31	11.3723790	0.00819850	0.9922828	1	0.00029700	0.99998750	0.99999421	5.5308079	FALSE	76	54.355263	0.66509010	0.9866898	24	11	13	272	237522	252255	0.9886592	285	530645
bin.62.or	277.841160	0.99869690	0.0002092	3807194	66	66	287	79.8724700	0.04100610	0.9980208	1	0.00019920	0.99998970	0.99999471	7.7964500	FALSE	311	95.517684	0.73353480	0.9910115	39	20	20	142	3922046	4036141	0.9982254	162	8153045
bin.64.st	37.363010	0.99872130	0.0003650	3336296	164	164	36	16.2500190	0.00894060	0.9910715	1	0.00045260	0.99997420	0.99998941	7.7055063	FALSE	436	81.353211	0.63673860	0.9799496	85	35	35	1518	468862	505330	0.9889791	1553	1078310
bin.66.st	27.407338	0.99889950	0.0008147	3704909	186	186	27	10.1429830	0.00529620	0.9891233	1	0.00072740	0.99996370	0.99999421	7.7572796	FALSE	1775	38.197746	0.72620200	0.9938549	133	21	21	7602	379037	417514	0.9942503	7623	866663
bin.71.st	148.215400	0.99924430	0.0002466	3858799	124	124	148	48.1802480	0.02460610	0.9975792	1	0.00022280	0.99997610	0.99999351	5.5192614	FALSE	601	101.54742	0.84520450	0.9965888	92	25	26	391	2151625	2235576	0.9916822	417	4556725
bin.73.or	107.263000	0.99938250	0.0013101	2108598	52	52	108	21.1168240	0.01457820	0.9983088	1	0.00126590	0.99995010	0.99999901	2.2306904	FALSE	7008	104.81192	0.76676750	0.9842303	105	2	2	7550	857672	931169	0.9964130	7552	1887785
bin.74.or	24.939880	0.99898900	0.0002317	3418729	99	99	25	7.1815700	0.00389530	0.99451780	0.9999999	5.48E-05	0.99999470	0.99999941	6.6358935	FALSE	70	51.114285	0.6835800	1	18	2	2	504	321936	335506	0.9971688	506	684992
bin.76.or	61.484796	0.99908130	0.0002511	2393797	127	127	62	14.3970450	0.00935500	0.9962026	1	0.00023580	0.99999580	0.99999831	2.2591170	FALSE	152	75.203947	0.56693010	0.9848920	10	4	4	292	560909	578585	0.9980198	296	1176465
bin.77.st	113.121360	0.99974370	0.0001885	1943622	44	44	108	48.6590340	0.03498180	0.9982558	1	0.00018210	0.99999430	0.99999892	6.6808767	FALSE	70	107.38571	0.85241410	0.9987957	11	2	2	82	812204	835124	0.9943336	84	1684529
bin.8.ori	88.272620	0.99868190	0.0002442	4018815	96	96	87	23.1256260	0.01156330	0.9973290	1	0.00022860	0.99998570	0.99999651	5.5760177	FALSE	561	115.01960	0.60684240	0.9661259	57	14	14	499	1331012	1381683	0.9977430	513	2797615
bin.81.or	48.940924	0.99729120	0.0002314	1947378	36	36	49	11.5006030	0.00825650	0.9953573	1	0.00021680	0.99999170	0.99999791	4.4393126	FALSE	53	58.433962	0.65500310	0.9651614	16	4	4	168	359406	371262	0.9982102	172	748172
bin.82.or	43.445337	0.99884750	0.0002278	2739467	150	150	43	30.0917480	0.01828120	0.9940134	1	0.00022600	0.99998970	0.99999811	4.4237243	FALSE	427	132.24121	0.51017800	0.9938100	28	5	5	256	453118	478999	0.9979548	261	981751
bin.84.or	433.973700	0.98942980	0.0009232	1884839	51	51	443	119.781940	0.08748480	0.9835959	1	0.00090340	0.99996330	0.99999461	6.6330698	FALSE	11741	114.12886	0.99021220	0.9999865	68	10	9	5798	3121647	3495877	0.9964189	5807	7132361
bin.86.or	30.261664	0.99937490	0.0002178	2017502	40	40	30	7.9559825	0.00561240	0.9968713	1	0.00014470	0.99999550	0.99999951	5.5110655	FALSE	16	34	0.8277257	1	9	1	1	243	232833	239772	0.9973177	244	483576
bin.89.st	19.810748	0.99746700	0.0004881	2784066	204	204	19	19.0838100	0.01152200	0.97926880	0.9999999	0.00037080	0.99994350	0.99998751	3.7134444	FALSE	298	48.231543	0.54982330	0.9784970	154	34	33	2079	207485	235301	0.9903577	2112	534286
bin.9.str	38.535585	0.99746660	0.0002473	3238417	212	210	38	18.3019590	0.01023740	0.9874830	1	0.00029250	0.99996990	0.99998681	6.6359890	FALSE	332	79.277108	0.50724050	0.9824612	96	42	42	558	469859	507592	0.9861875	600	1100401
bin.96.or	27.719074	0.99934100	0.0002050	2039615	49	49	27	7.0444792	0.00494440	0.9965861	1	0.00013740	0.99999900	0.99999951	4.4142157	FALSE	11	7	0.8211541	1	2	1	1	197	213003	218531	0.9984685	198	442802

Host1 Adult

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_	clinked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	reads_unf	divergent
bin.1.ori	119.551420	0.99936530	0.0003240	2782804	78	78	121	22.0735570	0.01326940	0.9980383	1	0.00028840	0.99984940	0.99999601	2.2994820	FALSE	564	88.879432	0.87473350	0.9951106	418	11	11	1083	1262337	1311312	0.9979928	2650863	1094
bin.10.st	32.067750	0.99904370	0.0006620	1632377	31	31	32	8.4207080	0.00660330	0.9938935	1	0.00054460	0.99975030	0.99999011	3.3233516	FALSE	403	40.019851	0.85710260	0.9952782	405	16	16	2405	200627	212232	0.9903792	437160	2421
bin.11.or	154.288570	0.99787870	0.0002616	4784343	127	127	151	44.0613500	0.02019770	0.9964670	1	0.00024790	0.99996580	0.99998821	1.5767182	FALSE	1298	106.5	0.75439350	0.9944222	163	56	56	634	2825899	3058245	0.9962008	6330936	690
bin.24.pe	38.647530	0.99648670	0.0002327	2386424	126	122	36	66.7131710	0.04341530	0.9900042	1	0.00031900	0.99997120	0.99998891	1.3363797	FALSE	379	105.78891	0.45905980	0.9580053	68	26	25	345	353892	385130	0.9869996	829308	370
bin.25.st	65.820210	0.99751000	0.0002265	1866339	66	65	66	18.9580430	0.01392640	0.9895356	1	0.00019460	0.99996750	0.99998641	1.6051894	FALSE	43	70.126770	0.7841866	1	60	25	25	136	474867	491734	0.9894184	995818	161
bin.3.ori	51.952571	0.99751580	0.0020919	2951892	208	208	54	16.6288030	0.00974740	0.9896124	1	0.00142900	0.99829900	0.99997771	6.6125105	FALSE	7552	53.709216	0.67186250	0.9992233	4969	65	62	14352	582159	674677	0.9945672	1384418	14414
bin.35.or	20.277748	0.99554410	0.0008914	2548811	100	100	20	11.1135430	0.00698860	0.96876260	0.9999999	3.04E-05	0.99980310	0.99998331	6.6437333	FALSE	420	37.352380	0.78257800	0.9962295	486	41	41	4372	197670	229380	0.9945323	469526	4413
bin.36.or	26.986266	0.99069210	0.0001942	1981770	65	65	27	9.3263914	0.00664680	0.9761531	1	0.00035240	0.99941580	0.99942001	3.3800843	FALSE	83	55.602409	0.6875305	1	1130	1122	1121	211	204994	224024	0.9965460	458199	1332
bin.37.or	85.687064	0.99855310	0.0002068	2152298	88	88	63	68.4005070	0.04681560	0.9948705	1	0.00018830	0.99995510	0.99999348	7.7885691	FALSE	82	46.670731	0.67373510	0.9797547	96	14	14	275	687548	707085	0.9982869	1431385	289
bin.38.or	24.976653	0.99073860	0.0002038	1886446	69	69	25	7.8244681	0.00571770	0.97816260	0.9999999	5.48E-05	0.99997720	0.99998101	3.3934039	FALSE	4	8.75	1	1	42	35	35	164	184813	192227	0.9972455	390544	199
bin.39.or	31.697060	0.99385260	0.0002456	1420276	35	35	32	10.8170300	0.00909900	0.9683540	1	0.00023680	0.99996500	0.9999890		FALSE	60	77.266666	0.5901610	1	48	15	15	219	176289	184243	0.9979346	373096	234
bin.48.or	31.513876	0.97097760	0.0002457	1810155	74	74	32	10.5454580	0.00787020	0.9616358	1	0.00020240	0.99997580	0.99998671	5.5527900	FALSE	15	12.4	0.7852620	1	42	23	23	285	222551	234611	0.9974530	482237	308
bin.49.st	19.246187	0.99503030	0.0010842	2655293	108	105	20	7.7572139	0.00477990	0.96147990	0.9999999	2.80E-05	0.99942020	0.99998111	8.8438438	FALSE	172	22.488372	0.9171388	1	1480	48	48	5234	194919	226001	0.9913539	465218	5282
bin.57.pe	21.300480	0.96747350	0.0002535	2377846	310	281	20	22.2618930	0.01462870	0.94215220	0.9999999	6.48E-05	0.99985220	0.99989511	6.5548653	FALSE	167	100.68263	0.49220330	0.9920868	331	235	237	471	194589	215951	0.9910449	470713	708
bin.58.or	230.374060	0.99844990	0.0002523	3725586	177	177	233	47.8466660	0.02490730	0.9952329	1	0.00023170	0.99991880	0.99994171	5.5100436	FALSE	339	112.97935	0.79880760	0.9891697	301	216	226	245	3280370	3404420	0.9980787	6911574	471
bin.62.or	610.175300	0.84699200	0.0002059	3807194	66	61	672	425.906950	0.21865850	0.8424007	1	0.00018000	0.99740230	0.9974248		FALSE	299	72.882943	0.62919980	0.9826673	8331	8259	8252	311	8712137	9880338	0.9912484	20096293	8563
bin.64.st	127.459160	0.99054790	0.0002500	3336296	164	144	129	53.3024980	0.02932650	0.9866873	1	0.00022810	0.99996080	0.99997901	4.4148746	FALSE	417	99.328537	0.70140230	0.9737473	129	69	69	334	1626710	1722728	0.9852785	3588208	403
bin.71.st	18.471268	0.99720530	0.0001978	3858799	124	124	18	9.8252097	0.00501780	0.97887840	0.9999999	0.00028060	0.99998490	0.99999571	6.6904680	FALSE	168	74.904761	0.69280300	0.9971142	57	16	17	470	271032	286157	0.9909567	593026	487
bin.73.or	25.105301	0.99837750	0.0012323	2108598	52	52	25	7.4226013	0.00512420	0.99183010	0.9999999	0.00056770	0.99971350	0.99999661	2.2892610	FALSE	756	37.293650	0.73901590	0.9861824	599	7	7	6135	202635	221456	0.9960913	450557	6142
bin.8.ori	94.244120	0.99847540	0.0002408	4018815	96	96	98	30.3805590	0.01519100	0.9955917	1	0.00021340	0.99997620	0.9999930		FALSE	426	100.71596	0.61867350	0.9295082	95	28	28	548	1438031	1491387	0.9974580	3018794	576
bin.82.or	20.289285	0.99773240	0.0002005	2739467	150	150	20	10.1370200	0.00615840	0.98616590	0.9999999	7.57E-05	0.99998960	0.99999851	3.3822752	FALSE	197	113.15228	0.51579660	0.9925732	28	4	4	307	213824	224142	0.9979891	457971	311
bin.84.or	22.775606	0.98061160	0.0016009	1884839	51	51	19	24.0645240	0.01757590	0.96306100	0.9999999	0.00018110	0.99989030	0.99999771	6.6199038	FALSE	614	33.114006	0.9276423	1	199	4	4	6715	166867	213535	0.9929097	451547	6719
bin.90.or	99.835007	0.96336030	0.0011224	2460616	151	151	109	48.6104850	0.03118090	0.9344379	1	0.00061340	0.99631970	0.9983503		FALSE	2874	67.399791	0.84839180	0.9959720	8462	3793	3789	6784	948087	1209155	0.9914088	2482128	10573
bin.94.st	133.868900	0.98736220	0.0002714	3137776	276	251	136	41.2941340	0.02351960	0.9792850	1	0.00024480	0.99994300	0.99996811	1.1640920	FALSE	381	88.157480	0.74811310	0.9931810	175	98	103	334	1612452	1693602	0.9899207	3503917	437

Host 1 Old

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_c	linked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	reads_unf	divergent
bin.1.ori	217.657380	0.99873070	0.0002594	2782804	78	78	222	38.2923000	0.02301920	0.9963159	1	0.00023770	0.99979080	0.99984651	1.1779396	FALSE	373	114.150130	0.88865710	0.9982229	580	320	320	561	2289815	2382125	0.9978864	4824363	881
bin.10.st	54.4631220	0.99759300	0.0006689	1632377	31	31	56	15.6528940	0.01227470	0.9751417	1	0.00039420	0.99848720	0.99973801	1.3496709	FALSE	593	51.8971330	0.92771040	0.9958778	2408	417	417	2681	338656	370210	0.9894708	766741	3098
bin.2.ori	138.628040	0.98408450	0.0009793	3200294	88	88	143	44.6243460	0.02501350	0.9797718	1	0.00094250	0.99900970	0.99982071	1.6787628	FALSE	4456	100.943220	0.97400860	0.9997256	3105	562	562	5688	1682356	1855174	0.9951339	3812400	6250
bin.21.st	26.9250280	0.99302720	0.0005603	1516329	76	76	27	20.4157200	0.01666310	0.9737392	1	0.00012060	0.99918250	0.99959021	1.3763467	FALSE	170	55.1411760	0.76286040	0.8078901	1	1207	605	605	1482	155470	176185	0.9937300	379717	2087
bin.24.pe	109.388530	0.98438330	0.0002312	2386424	126	119	108	67.9576310	0.04422510	0.9806714	1	0.00021580	0.99996580	0.99998461	1.3081768	FALSE	561	117.604270	0.51145490	0.9623428	80	36	36	269	990494	1049945	0.9858164	2215366	305
bin.25.st	26.5759560	0.99740660	0.0002219	1866339	66	65	24	16.9253260	0.01243320	0.99060080	0.99999990	0.00050270	0.99996860	0.99998971	1.5840409	FALSE	74	79.216216	0.7407056	1	58	19	19	263	190363	204045	0.9878672	428751	282
bin.3.ori	79.1368710	0.99608480	0.0012697	2951892	208	208	85	32.5142260	0.01905920	0.9748093	1	0.00068970	0.99645110	0.99969871	1.9466489	FALSE	4320	69.414351	0.95861280	0.9996791	10212	867	864	10961	880794	1068370	0.9930540	2191908	11825
bin.31.or	24.1347280	0.99268830	0.0021666	1472315	323	323	25	8.8668181	0.00747320	0.95999760	0.99999990	0.00015570	0.99847740	0.9997827		FALSE	979	30.173646	0.87857320	0.9993245	2152	307	306	6637	134620	159635	0.9938986	335353	6943
bin.32.or	49.1380720	0.99903770	0.0002023	2425631	102	102	49	10.1537510	0.00654700	0.9959359	1	0.00018720	0.99997680	0.99998711	1.8825291	FALSE	33	49	0.52425020	0.9863957	56	3	3	162	452275	465765	0.9983043	947401	165
bin.35.or	84.3409190	0.99626720	0.0009456	2548811	100	100	88	28.5447380	0.01795010	0.9931336	1	0.00082250	0.99977750	0.99998491	1.5359266	FALSE	5443	74.174536	0.75303140	0.9882888	563	38	38	6041	818945	939521	0.9942621	1915222	6079
bin.36.or	148.655320	0.99341340	0.0002068	1981770	65	65	153	37.1145370	0.02645120	0.9897909	1	0.00020030	0.99918120	0.9991863		FALSE	162	105.17010	0.76286040	0.9987869	1606	1596	1593	115	1126236	1211524	0.9986891	2457334	1708
bin.38.or	86.6757760	0.99483260	0.0002078	1886446	69	69	89	20.9296100	0.01529440	0.9897495	1	0.00019320	0.99996410	0.99996781	2.2531496	FALSE	10	27.7	1	1	67	60	60	48	635861	659006	0.9974389	1336966	108
bin.39.or	74.5420720	0.99841290	0.0002528	1420276	35	35	74	19.3870950	0.01630790	0.9962535	1	0.00023350	0.99997800	0.99999291	1.5877573	FALSE	121	102.223140	0.83031430	0.9966594	31	10	10	149	410473	428104	0.9978906	869055	159
bin.42.st	249.248950	0.98676180	0.0006134	4521859	338	292	252	86.7289720	0.04109380	0.9779051	1	0.00057450	0.99990430	0.99995341	1.7646930	FALSE	8603	139.060090	0.87487530	0.9946131	423	206	205	8319	4272487	4691918	0.9786857	9838303	8524
bin.44.st	20.8731560	0.94181160	0.0002024	2699920	481	459	22	16.4589970	0.01020010	0.88634030	0.9999999	5.20E-05	0.99884160	0.9988604		FALSE	97	54.948453	0.58499120	0.9943897	2772	2727	2730	315	213791	254874	0.9918413	553425	3045
bin.47.or	47.4164540	0.99903800	0.0002191	2651868	128	128	47	13.4102520	0.00827490	0.9954850	1	0.00019600	0.99997760	0.99998251	1.7181208	FALSE	73	49.972602	0.8372519	0.99125	59	46	46	184	485906	501834	0.9980926	1014586	230
bin.48.or	104.153310	0.97447780	0.0002376	1810155	74	74	108	30.0437600	0.02242220	0.9689722	1	0.00021240	0.99993890	0.99995611	1.5302323	FALSE	115	45.956521	0.82278430	0.9983507	107	77	77	191	727541	765635	0.9974760	1572377	268
bin.49.st	28.8011750	0.98637130	0.0003024	2655293	108	103	29	10.1418810	0.00624930	0.9697393	1	0.00029080	0.99979060	0.99984851	1.2861881	FALSE	163	45.184049	0.79954460	0.9773315	539	390	390	976	291039	310141	0.9946750	647575	1366
bin.5.ori	33.1336620	0.99689410	0.0009863	2165589	74	74	33	16.9824820	0.01157980	0.9843848	1	0.00037920	0.99939950	0.99990801	1.5689042	FALSE	856	42.755841	0.87086100	0.9951426	1280	196	196	4661	271268	315371	0.9959475	642915	4857
bin.50.or	23.7561690	0.99680610	0.0007069	2445970	105	105	24	6.9827457	0.00448400	0.98213140	0.99999990	0.00012550	0.99952210	0.99988511	0.3166809	FALSE	240	30.425	0.9391734	1	1148	276	276	3535	220844	238605	0.9970663	490918	3811
bin.58.or	142.122040	0.99900920	0.0012438	3725586	177	177	143	41.7968470	0.02175800	0.9972954	1	0.00121220	0.99982500	0.99998351	1.7854726	FALSE	12922	111.703760	0.89407070	0.9989825	650	61	70	12000	2022637	2335808	0.9947252	4766667	12070
bin.59.or	49.3005810	0.99928770	0.0001863	4499876	193	193	49	11.2111680	0.00530780	0.9967561	1	0.00017600	0.99999350	0.99999441	1.5263507	FALSE	16	25.4375	1	1	29	25	25	105	837829	858303	0.9986164	1729258	130
bin.6.ori	37.8300130	0.99912660	0.0002287	2474392	78	78	38	8.1131480	0.00517400	0.9960285	1	0.00019310	0.99997440	0.99999551	1.1943490	FALSE	32	37.46875	0.8438708	1	63	11	11	404	356277	367546	0.9981646	747717	415
bin.62.or	385.723430	0.84780070	0.0002093	3807194	66	60	387	297.873980	0.15292700	0.8401899	1	0.00018430	0.99747520	0.9974912		FALSE	217	74.391705	0.62016100	0.9643730	8076	8025	8020	267	5533919	6279719	0.9913573	12773993	8287
bin.64.st	144.257270	0.99155890	0.0002492	3336296	164	149	146	45.3384720	0.02494480	0.9878239	1	0.00023040	0.99995930	0.99997751	1.3644840	FALSE	485	116.742260	0.57065920	0.9816192	134	74	76	325	1824442	1914350	0.9869254	3961972	401
bin.66.st	45.9394800	0.99163940	0.0009786	3704909	186	186	39	28.5476970	0.01490640	0.9771203	1	0.00148130	0.99961460	0.9999820		FALSE	6902	74.461315	0.64120980	0.9942636	1395	65	65	8714	639214	788074	0.9901121	1694803	8779
bin.69.or	22.2134890	0.99731100	0.0003270	2620716	142	142	22	7.1149100	0.00441860	0.98668530	0.99999990	0.00017870	0.99986270	0.99991101	0.6432646	FALSE	83	31.674698	0.89905010	0.9949164	355	230	230	1062	224950	235091	0.9977300	479835	1292
bin.73.or	102.246830	0.99777050	0.0011763	2108598	52	51	104	21.0950170	0.01456310	0.9958583	1	0.00108790	0.99952040	0.99998351	2.2303360	FALSE	5603	99.826521	0.72757760	0.9766753	1007	22	22	6484	821382	890272	0.9966239	1803688	6506
bin.76.or	25.8899300	0.99782350	0.0002068	2393797	127	127	25	8.6586829	0.00562630	0.98901410	0.9999999	6.64E-05	0.99998850	0.99999451	2.2361363	FALSE	46	39.739130	0.6079499	1	27	13	13	243	237563	245839	0.9979035	503562	256
bin.8.ori	59.0853770	0.96827390	0.0002377	4018815	96	96	57	37.2495570	0.01862560	0.9585783	1	0.00022620	0.99997450	0.9999901		FALSE	953	119.206710	0.52087220	0.9274678	98	38	38	607	897457	956721	0.9966279	1968990	645
bin.82.or	39.5971480	0.99773820	0.0002039	2739467	150	149	39	29.8239010	0.01811850	0.9921017	1	0.00021390	0.99998710	0.99999741	2.2261316	FALSE	337	126.50445	0.53158640	0.9880166	35	7	7	244	414939	437172	0.9980053	898214	251
bin.84.or	79.1155290	0.98471910	0.0008096	1884839	51	51	70	69.2509640	0.05057860	0.9783382	1	0.00078300	0.99996470	0.99999451	1.4609897	FALSE	2450	82.538775	0.9883794	1	65	10	10	4978	573427	662230	0.9952756	1381969	4988
bin.86.or	22.1276460	0.99936650	0.0001965	2017502	40	40	22	5.9825553	0.00422020	0.99574720	0.99999990	0.00015040	0.99999300	0.99999951	3.1560045	FALSE	9	53.222222	0.8909090	1	14	1	1	202	171809	176853	0.9976250	356662	203

Host 2 Young

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_c	linked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	divergent	reads_unf
bin.1.ori	37.4003320	0.99878190	0.0002555	3640206	177	177	36	38.7406170	0.02040440	0.9927366	1	0.00023110	0.99998420	0.99999471	1.5783196	FALSE	344	94.058139	0.62489310	0.9960334	57	19	19	465	522662	559870	0.9967348	484	1172853
bin.10.or	20.3167980	0.99837870	0.0003704	3951351	174	174	20	5.8519788	0.00295690	0.99046140	0.9999999	1.45E-06	0.99997520	0.99999711	1.3754963	FALSE	113	25.274336	0.8619785	1	97	11	11	1621	307815	322988	0.9970624	1632	666104
bin.12.or	298.056840	0.99969870	0.0002884	4617425	145	145	312	87.5123430	0.04085420	0.9987932	1	0.00026540	0.99996350	0.99998851	1.5998453	FALSE	1759	106.488340	0.76273830	0.9938584	168	53	53	871	5217900	5500992	0.9969829	924	11246460
bin.13.or	257.362270	0.99979040	0.0002851	4928967	125	125	257	56.0198250	0.02529690	0.9993093	1	0.00026950	0.99997920	0.99999531	1.5843908	FALSE	1448	114.528310	0.70341130	0.9955336	102	23	25	571	4863283	5144795	0.9966663	596	10529665
bin.15.st	18.7596290	0.99804560	0.0009473	2543496	108	108	18	8.0658094	0.00507900	0.97848070	0.9999999	8.55E-05	0.99990590	0.99999151	1.6616118	FALSE	321	30.277258	0.76636230	0.9946718	234	21	21	4561	182592	209244	0.9936317	4582	428218
bin.16.or	44.6011840	0.99951580	0.0003284	3709762	106	106	44	10.6406510	0.00554030	0.9973071	1	0.00031420	0.99995510	0.99998811	1.3950875	FALSE	78	45.294871	0.8877290	1	18	7	8	434	636177	665782	0.9971452	442	1370854
bin.17.st	40.9448690	0.99817380	0.0002527	3024349	226	226	40	22.5752330	0.01307930	0.9903972	1	0.00025640	0.99997930	0.99999191	1.3929321	FALSE	151	59.860927	0.53804790	0.9480224	62	24	24	408	473428	512190	0.9934493	432	1105456
bin.20.or	25.8841670	0.99725970	0.0002046	3567545	296	296	24	16.0813730	0.00858560	0.94926060	0.99999990	0.00015050	0.99999020	0.99999643	2.2882400	FALSE	128	93.679687	0.69182260	0.9942779	33	12	14	321	347602	372567	0.9973966	335	791770
bin.23.st	151.074140	0.99841820	0.0003469	3546007	241	241	152	49.9710280	0.02671900	0.9933234	1	0.00031370	0.99993070	0.99997501	1.1886047	FALSE	1643	107.586730	0.67246780	0.9856325	244	88	92	953	2048440	2260427	0.9840567	1045	4821565
bin.24.or	28.4992730	0.99826900	0.0002059	2126577	84	84	24	18.5193900	0.01274990	0.9744269	1	0.00018780	0.99998790	0.99999855	2.2737683	FALSE	33	28.424242	0.72699200	0.9918414	25	3	3	227	226283	232567	0.9982656	230	470933
bin.25.or	43.8324880	0.99900230	0.0003957	2749441	81	81	44	9.8459143	0.00595540	0.9960486	1	0.00030500	0.99998210	0.99998851	1.3474510	FALSE	440	45.429545	0.82509110	0.9901597	49	4	4	1619	458076	476552	0.9977500	1623	965769
bin.27.or	58.6889160	0.99934770	0.0002522	3877444	135	135	58	23.0320880	0.01173750	0.9969655	1	0.00022820	0.99998050	0.99999481	1.4406101	FALSE	455	89.971137	0.55898200	0.9962262	43	4	4	454	415655	432495	0.9960381	458	884898
bin.3.ori	23.1621570	0.99903930	0.0002474	1780053	36	36	23	6.8379124	0.00513550	0.99438100	0.9999999	0.00012760	0.99997900	0.99999711	1.5664003	FALSE	27	30.703703	0.8387154	1	37	5	5	322	158191	163240	0.9973039	327	329640
bin.36.or	41.1769080	0.99441780	0.0006341	1701131	36	36	41	11.0932800	0.00852330	0.9897967	1	0.00061880	0.99998990	0.99999581	1.5367069	FALSE	586	46.706484	0.9799475	1	17	7	7	2735	271433	294962	0.9966332	2742	599460
bin.39.st	48.0732010	0.99839490	0.0003117	3883412	241	239	46	35.1640320	0.01795570	0.9926737	1	0.00029630	0.99996860	0.99998671	1.4868174	FALSE	526	74.209125	0.63944500	0.9833601	121	51	49	624	716016	797878	0.9881314	673	1756264
bin.41.or	20.3499800	0.99190310	0.0007884	3092913	142	142	20	6.6134503	0.00377780	0.97846230	0.9999999	6.02E-05	0.99998440	0.99999801	1.5863078	FALSE	461	25.704989	0.91674600	0.9968224	47	6	6	5458	240975	257626	0.9971257	5464	528104
bin.44.st	34.1813770	0.99779640	0.0002450	4058500	148	147	33	26.6914720	0.01329770	0.9917597	1	0.00024330	0.99997510	0.99999151	1.8232134	FALSE	333	80.672672	0.64312920	0.9835095	100	34	33	702	526962	560064	0.9921948	735	1169543
bin.46.st	116.054970	0.99601810	0.0017901	3855259	445	444	118	51.8613320	0.02672320	0.9845916	1	0.00167610	0.99976280	0.99993151	1.6653181	FALSE	29584	112.323410	0.45368780	0.9825339	900	260	274	21173	1716014	2159790	0.9800776	21447	4816962
bin.49.or	22.8349310	0.99811620	0.0005691	1851138	73	73	23	7.2310367	0.00533580	0.99131070	0.9999999	4.30E-05	0.99996940	0.99999941	1.3663913	FALSE	176	38.369318	0.8532398	1	56	1	1	2091	162291	171948	0.9974354	2092	349277
bin.52.st	22.7528130	0.99928830	0.0002090	2720290	80	80	22	10.3906760	0.00631850	0.9935885	0.9999999	0.00018900	0.99999550	0.99999741	1.4505776	FALSE	37	106.945940	0.7529994	1	12	7	8	302	235749	245133	0.9961314	310	502065
bin.54.or	19.5916700	0.99803080	0.0003000	3130208	175	175	19	6.7525430	0.00383810	0.98256980	0.9999999	1.28E-05	0.99999540	0.99999831	1.5416671	FALSE	31	25.064516	0.9373386	1	14	5	5	551	243880	253695	0.9976278	556	518966
bin.55.or	56.6457370	0.99924610	0.0002232	3551181	95	95	57	14.4537340	0.00769050	0.9954054	1	0.00020760	0.99999200	0.99999661	1.5480257	FALSE	58	45	0.66154570	0.9819548	28	12	12	162	763791	795375	0.9963847	174	1622548
bin.56.st	40.9042910	0.99779410	0.0014095	4133192	291	291	42	28.8044470	0.01426900	0.9828812	1	0.00103340	0.99983530	0.9999842		FALSE	6197	51.842343	0.62826370	0.9956211	669	64	64	13617	653181	754712	0.9886828	13681	1630018
bin.58.or	74.9027500	0.99939170	0.0002566	3888161	92	92	75	18.4051310	0.00935610	0.9978879	1	0.00023970	0.99998960	0.99999741	1.4892020	FALSE	219	88.333333	0.76435780	0.9869682	40	10	10	287	1115195	1167396	0.9959412	297	2385478
bin.59.st	24.6771730	0.99853290	0.0002833	4022326	247	246	22	40.1605730	0.02014850	0.98783010	0.9999999	0.00057890	0.99996320	0.99998311	1.5705352	FALSE	530	83.069811	0.67543680	0.9944484	146	67	66	926	381411	462490	0.9839931	992	1114325
bin.6.ori	34.7764270	0.99454870	0.0003825	1953481	161	161	35	9.6567736	0.00696680	0.9838877	1	0.00031950	0.99996300	0.99997601	1.3824848	FALSE	103	49.737864	0.64008320	0.9823939	71	46	46	496	266768	305590	0.9920866	542	671961
bin.60.or	385.557490	0.99887580	0.0002191	3815489	51	51	403	158.856930	0.08143520	0.9985126	1	0.00021120	0.99998870	0.9999952		FALSE	427	112.475400	0.65566880	0.9846306	43	18	19	154	5538183	5769818	0.9976636	173	11691146
bin.62.pe	37.4103750	0.99852620	0.0002777	1315659	58	58	36	36.9141250	0.03232540	0.9929343	1	0.00032520	0.99996780	0.99998081	1.5921482	FALSE	114	41.675438	0.66987100	0.9798057	42	25	25	225	190122	207989	0.9838568	250	444704
bin.64.or	22.8675510	0.99599980	0.0016851	2474903	159	159	23	7.8919977	0.00505370	0.98236130	0.9999999	0.00023110	0.99964750	0.99999341	1.7301524	FALSE	1055	23.642654	0.8766595	1	857	16	16	8520	218000	244689	0.9953076	8536	502685
bin.65.st	436.095720	0.99710870	0.0003857	4874996	421	418	448	167.898900	0.07670860	0.9900762	1	0.00035250	0.99988770	0.99994481	1.5098544	FALSE	4928	107.934050	0.68343650	0.9849172	542	266	277	2491	8109787	8776469	0.9815616	2768	18397316
bin.66.or	28.2444480	0.99808990	0.0002395	2636626	75	75	28	6.9980615	0.00432200	0.9942688	1	0.00022490	0.99999610	0.99999881	2.2225558	FALSE	26	24.307692	0.8713217	1	10	3	3	320	286783	298483	0.9974456	323	611218
bin.67.st	60.9513680	0.99796610	0.0003449	4052014	295	294	60	43.6518340	0.02184500	0.9900987	1	0.00029920	0.99996060	0.99998501	1.3978781	FALSE	818	86.683374	0.61951020	0.9727452	158	60	67	828	944784	1044360	0.9876327	895	2286870
bin.7.ori	28.1964350	0.99877350	0.0002576	3727699	115	115	28	7.9670875	0.00413920	0.9937068	1	0.00013960	0.99998560	0.99999751	1.6855643	FALSE	46	29.130434	0.8430041	1	53	9	9	609	400753	416429	0.9977205	618	850359
bin.70.or	20.0214080	0.99825730	0.0002443	3191732	152	152	20	9.4960664	0.00534080	0.97977770	0.9999999	0.00013680	0.99999680	0.99999871	1.4817236	FALSE	149	94.241610	0.45570410	0.9685311	10	4	4	386	247457	258813	0.9978376	390	531637
bin.71.or	41.9766620	0.99888830	0.0002756	2552853	66	66	42	9.7453550	0.00611510	0.9961541	1	0.00024370	0.99998540	0.99999681	1.3947383	FALSE	194	53.768041	0.53469290	0.9965868	37	8	8	359	410862	429369	0.9958061	367	876753
bin.72.or	26.4990680	0.99801460	0.0006216	3783788	242	242	27	7.5344146	0.00389830	0.98858650	0.9999999	0.00021090	0.99997860	0.99999411	1.4863149	FALSE	571	33.001751	0.9154658	1	80	22	22	4774	387600	418488	0.9962030	4796	866573
bin.73.or	23.7996220	0.99910570	0.0002104	2797940	89	89	24	6.3272786	0.00379470	0.99405880	0.9999999	7.94E-05	0.9999974	1	1.4167311	FALSE	22	15.5	0.6935181	1	7	0	0	297	253401	262013	0.9967473	297	534461

Host 2 adult

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_c	linked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	reads_unf	divergent
bin.15.st	97.8585250	0.99857280	0.0010756	2543496	108	107	100	42.5324050	0.02678280	0.9960833	1	0.00089520	0.99972010	0.99997471	1.5278687	FALSE	6812	75.375073	0.76843300	0.9888597	709	64	64	7068	928246	1083021	0.9929569	2224299	7132

bin.18.or	156.27020	0.9960114	0.0002390	1497047	72	72	157	37.554031	0.0308416	0.9942426	1	0.0002224	0.9999724	0.9999798	1.6394340	FALSE	80	69.5875	0.7377438	0.9895114	41	30	30	87	887792	920909	0.9976308	1873161	117
bin.20.or	41.895943	0.9974251	0.0001967	3567545	296	296	42	14.682785	0.0078389	0.9880766	1	0.0001709	0.9999846	0.9999917	2.3064677	FALSE	136	48.036764	0.6840259	0.9982589	54	29	29	321	546310	569720	0.9977800	1176147	350
bin.25.or	332.48647	0.9995860	0.0003866	2749441	81	81	338	51.074562	0.0308934	0.9989736	1	0.0003714	0.9999107	0.9999938	1.1478203	FALSE	2535	100.60710	0.8030555	0.9848398	245	17	19	1523	3381677	3513435	0.9977795	7110576	1542
bin.36.or	158.45670	0.9963295	0.0002778	1701131	36	36	160	34.348988	0.0263916	0.9935384	1	0.0002621	0.9999538	0.9999846	1.4118867	FALSE	173	98.774566	0.9606631	0.9997107	78	26	26	199	1015802	1083797	0.9963080	2227986	225
bin.37.st	143.42521	0.9906037	0.0002480	1551145	139	139	147	41.196090	0.0333777	0.9872094	1	0.0002323	0.9999386	0.9999582	1.4573873	FALSE	75	33.173333	0.8499884	0.9831256	94	64	64	151	841217	880331	0.9961173	1827254	215
bin.39.st	92.496153	0.9915520	0.0003114	3883412	241	221	91	57.590134	0.0294071	0.9838338	1	0.0002858	0.9999667	0.9999816	1.8617811	FALSE	388	80.363402	0.5632501	0.9739922	127	70	68	574	1328557	1403435	0.9864341	2939955	642
bin.42.st	85.597357	0.9989192	0.0002286	2457561	124	124	84	78.041183	0.0500350	0.9969445	1	0.0002156	0.9999685	0.9999853	1.1398382	FALSE	185	77.875675	0.4914096	0.9488376	77	36	36	244	780757	817741	0.9959275	1734052	280
bin.44.st	124.01323	0.9974589	0.0002536	4058500	148	147	118	68.006723	0.0338811	0.9948870	1	0.0002320	0.9999621	0.9999853		FALSE	1247	107.23576	0.6756557	0.9847799	153	59	58	747	1858845	1992866	0.9901108	4190250	805
bin.46.st	79.060543	0.9783386	0.0022782	3855259	445	423	81	66.753728	0.0343969	0.9509846	1	0.0018678	0.9979363	0.9999356		FALSE	17811	75.415305	0.5716005	0.9890404	7566	236	237	20393	1128129	1518358	0.9752774	3307223	20630
bin.49.or	172.35827	0.9990400	0.0004517	1851138	73	73	176	48.522633	0.0358050	0.9979834	1	0.0004074	0.9996411	0.9999821		FALSE	1233	78.274939	0.9818675	0.9981576	663	33	34	1276	1196143	1266941	0.9973682	2571713	1310
bin.5.ori	33.527216	0.9664890	0.0002000	1981678	284	278	33	17.622485	0.0127018	0.9463959	1	0.0002165	0.9983635	0.9983822	1.7208542	FALSE	102	97.970588	0.5531078	0.9861728	3069	3034	3034	246	243698	270871	0.9941698	563620	3280
bin.50.st	66.368909	0.9963039	0.0002043	1504843	195	194	67	18.575469	0.0153425	0.9899989	1	0.0001900	0.9999637	0.9999724	1.7678706	FALSE	51	63.882352	0.6644340	0.9289542	54	41	41	103	363984	378245	0.9978741	776165	144
bin.52.st	42.099654	0.9993901	0.0002060	2720290	80	80	38	57.298021	0.0348428	0.9971569	1	0.0002218	0.9999852	0.9999948	1.2033818	FALSE	157	75.127388	0.7075717	0.9913733	40	14	16	294	425156	457111	0.9948391	986210	310
bin.54.or	28.981002	0.9979739	0.0002287	3130208	175	175	29	7.5884256	0.0043132	0.9829375	1	0.0002205	0.9999876	0.9999931	1.2317106	FALSE	24	33.458333	0.9180749	1	38	21	21	398	336382	347207	0.9980959	705284	419
bin.55.or	25.933990	0.9942813	0.0002052	3551181	95	95	26	8.7924297	0.0046782	0.9661402	0.9999999	0.0001678	0.9999906	0.9999988	1.7288240	FALSE	24	49.083333	0.7459608	1	32	4	4	384	340014	355835	0.9948651	728813	388
bin.56.st	25.886548	0.9844372	0.0017049	4133192	291	283	27	15.389256	0.0076234	0.9258135	0.9999999	0.0006803	0.9986813	0.9997010		FALSE	3399	40.604883	0.5014325	0.9934522	5046	1144	1142	15334	394789	469580	0.9857812	997417	16476
bin.58.or	35.118534	0.9962342	0.0002351	3888161	92	92	35	11.637093	0.0059156	0.9854710	1	0.0002856	0.9999785	0.9999950	1.6432205	FALSE	138	63.181159	0.6684966	0.9902069	82	19	19	647	506880	534052	0.9948113	1095629	666
bin.60.or	575.45604	0.9988696	0.0002097	3815489	51	51	583	145.29681	0.0744838	0.9985747	1	0.0002025	0.9999887	0.9999944	1.6889643	FALSE	507	112.96252	0.6302930	0.9854155	43	21	22	174	8014353	8264614	0.9982881	16683620	196
bin.62.pe	37.691549	0.9977767	0.0002444	1315659	58	58	28	51.329993	0.0449492	0.9938418	1	0.0008139	0.9999518	0.9999717		FALSE	233	96.729613	0.6978169	0.9928917	63	37	37	279	186154	223838	0.9797348	517506	316
bin.69.st	38.897907	0.9975084	0.0003622	1291558	140	140	39	12.232826	0.0108825	0.9886060	1	0.0002866	0.9997893	0.9999130	1.2136836	FALSE	150	48.313333	0.9347946	0.9852284	269	111	111	854	187752	196950	0.9974239	410680	965
bin.70.or	26.973285	0.9982476	0.0002378	3191732	152	152	27	10.809964	0.0060798	0.9904518	1	0.0003685	0.9999927	0.9999958	1.3849070	FALSE	223	112.51121	0.3447639	0.9746364	23	13	13	508	318422	331709	0.9978639	679275	521
bin.73.or	96.154772	0.9996618	0.0002133	2797940	89	89	96	23.680465	0.0142022	0.9987515	1	0.0002060	0.9999906	0.9999989	1.1359526	FALSE	350	125.45714	0.4289080	0.9955462	26	3	3	182	997118	1038059	0.9953462	2144479	185
bin.79.or	41.775250	0.9942445	0.0010375	1604227	352	352	44	12.973243	0.0104751	0.9746494	1	0.0005767	0.9994870	0.9998874		FALSE	1613	44.541227	0.9004650	0.9953345	802	176	176	3810	248140	279534	0.9957581	582030	3986
bin.81.st	38.094112	0.9920176	0.0002239	1526372	342	340	30	36.327478	0.0300857	0.9695886	1	0.0001107	0.9999236	0.9999405		FALSE	25	33.64	0.7142050	1	113	88	88	200	221205	236851	0.9964356	509226	288

Host2 Old

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_c	linked_SN	SNV_dista	r2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	divergent	reads_unf
bin.12.or	502.83608	0.9946097	0.0012390	4617425	145	145	549	175.13286	0.0817590	0.9933822	1	0.0011904	0.9984841	0.9996912		FALSE	26182	115.17443	0.9631612	0.9975950	6953	1416	1411	11729	8467013	9663824	0.9945123	13140	19769152
bin.15.st	120.58391	0.9985099	0.0011164	2543496	108	107	125	37.407839	0.0235558	0.9960385	1	0.0009709	0.9996786	0.9999727	1.5680796	FALSE	8485	86.300766	0.7598969	0.9874194	814	69	69	7335	1136988	1324464	0.9924196	7404	2697034
bin.18.or	84.169511	0.9964944	0.0002348	1497047	72	72	85	21.418296	0.0175900	0.9932092	1	0.0002254	0.9999771	0.9999852	1.6582089	FALSE	60	73.133333	0.8439974	0.9900829	34	22	22	93	474938	493720	0.9974415	115	1011343
bin.20.or	26.176164	0.9938484	0.0001979	3567545	296	295	25	12.628874	0.0067423	0.9771913	0.9999999	0.0001302	0.9999853	0.9999931	2.4755464	FALSE	126	68.507936	0.5781721	0.9636768	51	24	24	435	340374	362223	0.9973939	459	763242
bin.24.or	38.826195	0.9985441	0.0001937	2126577	84	84	39	11.422944	0.0078642	0.9947243	1	0.0002318	0.9999706	0.9999853	1.8766492	FALSE	48	48.020833	0.7776396	0.9891975	62	31	31	224	299474	307237	0.9984294	255	622104
bin.25.or	99.457167	0.9992984	0.0003744	2749441	81	81	100	17.721856	0.0107193	0.9979530	1	0.0003438	0.9999077	0.9999923	1.1733134	FALSE	1101	92.132606	0.7674757	0.9946327	253	21	21	1556	1007140	1047597	0.9977076	1577	2124063
bin.36.or	135.04022	0.9959203	0.0003328	1701131	36	36	136	29.349980	0.0225507	0.9933579	1	0.0003147	0.9999568	0.9999857	1.4326070	FALSE	183	95.945355	0.9519869	1	73	24	24	264	859702	909181	0.9971080	288	1856813
bin.37.st	55.900425	0.9897424	0.0002391	1551145	139	138	56	17.492428	0.0141726	0.9839847	1	0.0002094	0.9999554	0.9999737	1.4949111	FALSE	48	40.854166	0.8322502	0.9548288	68	40	40	168	325177	344111	0.9953705	208	728596
bin.39.st	260.79842	0.9941136	0.0003450	3883412	241	225	259	158.49123	0.0809301	0.9880991	1	0.0003194	0.9999530	0.9999744	1.5767714	FALSE	618	68.145631	0.6767679	0.9728944	180	98	100	518	3762331	3971959	0.9872728	618	8310241
bin.41.or	28.830314	0.9911866	0.0017203	3092913	142	142	30	11.968953	0.0068371	0.9703628	1	0.0004260	0.9976029	0.9999740	2.0501108	FALSE	1240	36.345967	0.8561459	0.9968731	7194	78	78	10869	326390	384641	0.9942989	10947	789709
bin.42.st	83.160754	0.9984993	0.0002201	2457561	124	122	83	38.086903	0.0244188	0.9961506	1	0.0002097	0.9999709	0.9999861	1.1625562	FALSE	113	77.814159	0.4877211	0.9530086	71	34	33	227	754960	782374	0.9959241	260	1623340
bin.43.pe	23.437326	0.9906700	0.0011477	3730681	377	361	22	29.424721	0.0153904	0.9524097	0.9999999	0.0004518	0.9986780	0.9999600	1.8847391	FALSE	627	59.586921	0.7437071	0.9848968	4697	142	146	8013	321238	395405	0.9865411	8159	930874
bin.44.st	93.163726	0.9969252	0.0002467	4058500	148	147	87	54.709323	0.0272563	0.9940046	1	0.0002204	0.9999613	0.9999898		FALSE	991	102.21089	0.6896512	0.9856147	156	41	42	697	1390667	1485159	0.9899743	739	3110343
bin.45.st	29.721622	0.9832759	0.0001907	701803	195	194	32	12.553788	0.0154199	0.9268341	1	9.41E-05	0.9999108	0.9999308		FALSE	9	25.111111	0.6812667	1	58	45	44	81	75623	79309	0.9971222	125	169312
bin.46.st	103.94828	0.9794968	0.0027032	3855259	445	427	110	68.183577	0.0351337	0.9560877	1	0.0023119	0.9978958	0.9999253		FALSE	37929	96.419731	0.4291236	0.9855225	7756	275	276	27099	1476511	2048773	0.9747119	27375	4478856
bin.49.or	36.913921	0.9953882	0.0002907	1851138	73	73	37	14.754504	0.0108874	0.9755852	1	0.0002246	0.9995498	0.9997923		FALSE	83	99.168674	0.8222399	0.9870300	813	375	375	638	254365	271523	0.9971052	1013	554139
bin.60.or	120.42702	0.9986730	0.0002126	3815489	51	51	123	39.962067	0.0204858	0.9977966	1	0.0002025	0.9999900	0.9999950	2.0332285	FALSE	384	101.17968	0.7425424	0.9927376	38	19	20	196	1675543	1768190	0.9968570	216	3617362
bin.62.pe	40.829288	0.9977859	0.0002514	1315659	58	58	35	39.291188	0.0344069	0.9939536	1	0.0003667	0.9999472	0.9999747		FALSE	216	72.504629	0.6452152	0.9975958	69	33	33	271	200282	222695	0.9814314	304	485676
bin.65.st	31.071781	0.9881780	0.0009071	4874996	421	397	28	38.655130	0.0176605	0.9709735	1	0.0015947	0.9998383	0.9999649	1.8827433	FALSE	4445	74.596625	0.5400699	0.9814795	765	166	164	10748	559398	837195	0.9750320	10912	2099763
bin.67.st	48.213648	0.9920999	0.0003001	4052014	295	282	46	66.348482	0.0332032	0.9795792	1	0.0002637	0.9999450	0.9999788	1.5648260	FALSE	823	86.744835	0.5933225	0.9697756	218	84	94	942	719320	809357	0.9852804	1036	1799893
bin.70.or	59.756517	0.9990315	0.0002366	3191732	152	152	59	22.704302	0.0127694	0.9953956	1	0.0002241	0.9999867	0.9999937	1.4015042	FALSE	232	116.70258	0.4789644	0.9723125	42	20	19	149	701271	725335	0.9982229	168	1482954
bin.71.or	45.467532	0.9545723	0.0002080	2552853	66	65	50	19.711813	0.0123691	0.9243403	1	0.0001771	0.9978077	0.9978281		FALSE	44	62.340909	0.4678536	0.9295660	5173	5125	5123	223	428275	497065	0.9908997	5346	1018897
bin.73.or	26.368996	0.9991093	0.0001982	2797940	89	89	26	9.3093863	0.0055832	0.9929330	0.9999999	0.0002148	0.9999859	0.9999985	1.2132581	FALSE	133	101.94736	0.6336598	0.9939405	39	4	4	371	272230	286648	0.9946688	375	603725
bin.79.or	26.564354	0.9874656	0.0010990	1604227	352	352	28	9.9989331	0.0080735	0.9567124	0.9999999	9.89E-05	0.9994318	0.9998592		FALSE	626	33.707667	0.8745497	0.9941031	872	216	215	3825	157033	179021	0.9953927	4040	372630
bin.81.st	52.385417	0.9929925	0.0002152	1526372	342	341	54	18.387190	0.0152279	0.9794755	1	0.0001995	0.9999143	0.9999270		FALSE	27	38.666666	0.9008328	0.9912955	128	109	109	107	302216	318637	0.9964636	216	672248
bin.9.ori	22.888681	0.9982422	0.0001960	3536371	160	160	23	6.7809670	0.0036223	0.9914491	0.9999999	1.45E-05	0.9999811	0.9999837	1.6045996	FALSE	9	10.666666	0.8897577	1	66	57	57	353	297770	309879	0.9978597	410	634869

Host 3 Young

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_c	linked_SN	SNV_dista	r2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_	reads_unf	reads_mea	reads_unf	divergent
bin.15.st	26.767770	0.9991300	0.0011268	1995562	64	63	26	15.488212	0.0109993	0.9945323	0.9999999	0.0007826	0.9998730	0.9999934	1.2331930	FALSE	943	45.564156	0.7296796	0.9869985	252	13	13	5430	199243	222122	0.9913478	474607	5443
bin.17.or	23.820612	0.9970266	0.0003715	1946310	131	131	24	7.6108464	0.0054924	0.9858162	0.9999999	6.34E-05	0.9999463	0.9999942	1.6442333	FALSE	114	32.684210	0.8543428	1	103	11	11	1011	170145	177337	0.9976978	362048	1022
bin.18.or	207.42623	0.9993250	0.0002397	4093372	107	107	212	58.631887	0.0290557	0.9980026	1	0.0002229	0.9999789	0.9999946	1.6816097	FALSE	1045	121.04689	0.4252837	0.9712716	86	22	22	359	3104081	3204481	0.9979246	6472690	381
bin.19.st	23.850275	0.9909040	0.0005566	3217492	529	526	23	16.371522	0.0092809	0.9547569	0.9999999	0.0002902	0.9998922	0.9999694	1.9586389	FALSE	997	60.619859	0.3927279	0.9744840	331	94	93	2801	290511	351614	0.9867991	803421	2894
bin.2.per	71.730989	0.9983509	0.0009470	2125409	101	99	68	117.05206	0.0806736	0.9940938	1	0.0007973	0.9999361	0.9999914	1.3422699	FALSE	3719	66.703414	0.8660035	0.9944650	135	18	18	5094	563913	627525	0.9930231	1318313	5112
bin.24.or	397.47294	0.9996805	0.0002267	3762673	129	129	399	80.282236	0.0415303	0.9990450	1	0.0002193	0.9999880	0.9999957	1.5019766	FALSE	312	88.733974	0.9396759	0.9977824	45	16	15	165	5532644	5680544	0.9984147	11456380	180
bin.25.st	62.274081	0.9980125	0.0002490	3429045	130	129	62	22.281044	0.0120781	0.9946297	1	0.0002269	0.9999870	0.9999929	1.2703453	FALSE	290	74.782758	0.6804397	0.9804918	44	24	24	397	791135	821847	0.9936398	1676285	421
bin.26.st	50.854921	0.9993526	0.0002210	2496173	88	87	50	20.951133	0.0133078	0.9967165	1	0.0002121	0.9999939	0.9999975	1.1764499	FALSE	82	63.341463	0.8701112	1	15	6	6	165	472022	492265	0.9930208	1024307	171
bin.28.st	42.845310	0.9963922	0.0002636	1263389	146	146	43	13.476197	0.0121304	0.9861127	1	0.0002295	0.9999791	0.9999911	1.5517507	FALSE	72	39.486111	0.8503215	0.9820312	26	11	11	207	205615	216524	0.9953365	455769	218
bin.3.ori	22.831970	0.9953084	0.0010044	3519338	205	205	23	8.0861022	0.0043356	0.9821065	0.9999999	0.0002406	0.9998634	0.9999892	1.8230348	FALSE	1100	30.625454	0.8367117	0.9978068	472	37	36	7777	299076	327627	0.9957517	670532	7813
bin.30.or	56.849597	0.9992926	0.0002201	2471349	119	119	57	11.437001	0.0073104	0.9967179	1	0.0002058	0.9999967	0.9999995	1.1571353	FALSE	41	44.804878	0.8620688	0.9642678	8	1	1	175	521650	537399	0.9980222	1093467	176
bin.33.pe	39.793294	0.9981639	0.0006825	3678629	299	299	41	19.849317	0.0104342	0.9854146	1	0.0007163	0.9999401	0.9999853		FALSE	2467	60.749493	0.7133133	0.9905556	217	53	52	6126	538696	591687	0.9940080	1240824	6178
bin.36.or	46.240564	0.9991086	0.0002327	2424486	92	92	46	12.957556	0.0083534	0.9949828	1	0.0002254	0.9999970	0.9999987	1.2305418	FALSE	58	45.344827	0.9182691	1	7	3	3	203	417964	433793	0.9981361	885393	206
bin.38.st	123.48342	0.9995058	0.0002415	2673349	116	116	121	86.278857	0.0529992	0.9973579	1	0.0002255	0.9999763	0.9999891	1.1541652	FALSE	373	104.82037	0.6382943	0.9719589	63	29	34	232	1224273	1299089	0.9932197	2733282	266
bin.44.or	19.628981	0.9982032	0.0002322	2632571	123	123	19	6.4337333	0.0039839	0.9873765	0.9999999	0.0001018	0.9999873	0.9999965	1.6484529	FALSE	35	35.828571	0.7802644	1	33	9	9	443	191735	198867	0.9979140	405464	452
bin.46.or	19.414443	0.9976263	0.0002403	4193204	310	310	19	6.1152855	0.0030086	0.9843074	0.9999999	5.91E-05	0.9999917	0.9999985	1.4624102	FALSE	58	37.189655	0.5443694	0.9783152	34	6	6	714	303798	314330	0.9980110	643258	720
bin.51.or	85.698968	0.9981945	0.0016459	3374735	224	224	86	26.164391	0.0143381	0.9940718	1	0.0015743	0.9998581	0.9999883	1.7544410	FALSE	15567	87.965696	0.5800322	0.9938405	476	39	41	16550	1077436	1228509	0.9943601	2512295	16591
bin.52.st	86.660065	0.9994840	0.0002637	1232650	59	59	87	22.904124	0.0207292	0.9961132	1	0.0002449	0.9999918	0.9999967	1.5783020	FALSE	100	81.07	0.7090085	1	10	4	4	88	405649	424469	0.9953642	871994	92
bin.53.or	316.34351	0.9995196	0.0003647	2848092	117	117	320	54.705771	0.0325497	0.9987546	1	0.0003477	0.9999792	0.9999933	1.1677056	FALSE	2019	78.800396	0.8675979	0.9860143	59	19	21	1391	3332771	3451973	0.9980164	7003913	1412
bin.55.or	29.830077	0.9988999	0.0003891	2596212	105	105	30	7.8967263	0.0049208	0.9947843	1	0.0002092	0.9999670	0.9999980	1.2178951	FALSE	209	35.574162	0.9736306	1	85	5	5	1469	288277	301091	0.9979945	610495	1474
bin.57.or	103.03235	0.9983124	0.0008238	1770622	86	86	105	27.861499	0.0210407	0.9959539	1	0.0007898	0.9999603	0.9999914	1.6635035	FALSE	2882	103.93858	0.9885735	1	70	15	15	4692	688921	761577	0.9955699	1559904	4707
bin.59.st	33.746889	0.9983324	0.0002371	1797291	137	137	34	10.027235	0.0075371	0.9875618	1	0.0002198	0.9999842	0.9999887	1.2460144	FALSE	29	23.724137	0.7342279	1	28	20	20	256	231348	243748	0.9932285	510617	276
bin.6.ori	20.719087	0.9990779	0.0003998	1776407	59	59	20	5.9356333	0.0044683	0.9927798	0.9999999	0.0001669	0.9999795	0.9999971	1.4306437	FALSE	76	33.618421	0.8025491	0.9979555	36	5	5	1110	134331	140994	0.9977899	285102	1115
bin.60.st	27.189461	0.9976671	0.0003882	3993039	242	241	27	12.610842	0.0063495	0.9852573	1	0.0004867	0.9999722	0.9999895	1.6875750	FALSE	391	52.774936	0.6907312	0.9799635	109	41	40	2186	402395	433835	0.9893952	922010	2226

bin.62.or	34.450609	0.9992732	0.0002519	3276383	124	124	34	9.75682130	0.00541070	0.9950198	1	0.00023650	0.99988800	0.99999631	1.5505172	FALSE	116	55.362068	0.63707730	0.9924879	39	12	12	514	422190	437994	0.9978246	887414	526	
bin.64.or	70.881336	0.9983813	0.0003311	3564100	225	225	71	17.663015	0.00941560	0.9942936	1	0.00029850	0.99995880	0.99998981	1.5264566	FALSE	881	68.489216	0.49970820	0.9836198	146	36	36	1072	945514	994152	0.9949715	2028682	1108	
bin.66.or	93.880149	0.9980846	0.0010020	2503497	99	99	96	24.956484	0.01583550	0.9959017	1	0.00088320	0.99986440	0.99999471	1.4670452	FALSE	6218	74.310228	0.74243290	0.9848367	338	13	13	6458	876560	998612	0.9954765	2022362	6471	
bin.67.st	51.586505	0.9987841	0.0002622	1844859	74	74	51	19.826789	0.01465610	0.9951421	1	0.00023550	0.99998310	0.99999341	1.1975142	FALSE	206	92.679611	0.54203800	0.9913332	31	12	12	323	354652	375235	0.9933878	790337	335	
bin.68.st	58.411961	0.9990693	0.0004445	2093210	73	73	58	26.402263	0.01831270	0.9962330	1	0.00040680	0.99991820	0.99999131	1.1920828	FALSE	674	66.737388	0.88924810	0.9984786	39	18	17	1933	457618	486071	0.9947075	1019497	1950	
bin.71.or	23.878905	0.9975492	0.0002175	2633102	226	226	23	8.625941	0.00536200	0.98720360	0.9999999	3.41E-05	0.99999570	0.99999841	2.1742924	FALSE	20	17.9	0.9113017	1	11	4	4	359	235544	243067	0.9982891	496811	363	

Host 3 Adult

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_	clinked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_reads_unf	reads_mea	reads_unf	divergent	
bin.12.or	30.493618	0.8970970	0.0001905	1704771	218	217	33	16.229417	0.01259200	0.8463078	1	0.00027120	0.99994870	0.9999535		FALSE	7	28.285714	0.70262700	0.9002926	74	67	67	113	187641	195258	0.9966034	403883	180	
bin.15.st	35.811075	0.9974393	0.0010549	1995562	64	61	36	9.8195663	0.00697360	0.9924737	1	0.00066910	0.99967630	0.99998431	2.2914363	FALSE	1339	45.586258	0.75611190	0.9873164	641	31	31	5058	265472	287659	0.9916938	589901	5089	
bin.19.st	19.267119	0.9803001	0.0006733	3217492	529	520	14	36.277912	0.02056570	0.90622940	0.9999999	0.00043310	0.99979380	0.9999375		FALSE	2186	66.060841	0.36195830	0.9577676	601	182	182	3556	235538	406701	0.9767745	1041361	3738	
bin.25.st	29.401572	0.9959994	0.0002575	3429045	130	127	29	12.208496	0.00661800	0.9879590	1	0.00039040	0.99997430	0.99998931	1.4885617	FALSE	156	50.570512	0.71222080	0.9859486	87	36	36	715	371868	387803	0.9917715	793008	751	
bin.26.st	80.210941	0.9973287	0.0002243	2496173	88	84	80	25.458762	0.01617090	0.9935497	1	0.00021230	0.99998750	0.99999831	2.2037146	FALSE	54	43.481481	0.69683600	0.9700776	31	4	4	110	741541	763811	0.9922606	1550485	114	
bin.27.or	42.124404	0.8983702	0.0002524	3183055	235	233	48	23.414916	0.01322210	0.8560615	1	0.00025400	0.99689710	0.9969077		FALSE	45	53.222222	0.74800680	0.9979122	8455	8426	8422	228	498152	657029	0.9913415	1363505	8650	
bin.28.st	27.176548	0.9939353	0.0003068	1263389	146	145	27	9.6518728	0.00868800	0.9808253	1	0.00045530	0.99998220	0.99999031	1.5316771	FALSE	31	40.387096	0.6929273	1	22	12	12	319	130640	138831	0.9943397	293746	331	
bin.3.ori	21.934791	0.9869788	0.0011793	3519338	205	205	23	10.965313	0.00587940	0.90304280	0.9999999	0.00026340	0.99803680	0.99967652	2.3289625	FALSE	494	36.787449	0.7141249	1	6239	1028	1028	8061	287758	354563	0.9916611	738885	9089	
bin.41.or	20.184088	0.9932079	0.0002634	2005584	364	364	21	6.7553440	0.00485910	0.96332340	0.9999999	2.61E-05	0.99988870	0.9998954		FALSE	32	21.46875	0.8151292	1	215	202	201	410	150514	157306	0.9976680	323809	611	
bin.49.or	27.277215	0.9659031	0.0015518	2079370	426	422	28	13.423530	0.00950570	0.9129616	1	0.00105700	0.99931360	0.9997113		FALSE	2311	45.452185	0.63558640	0.9948565	1303	548	547	7546	216592	302102	0.9853715	663378	8093	
bin.51.or	29.674430	0.9912873	0.0017927	3374735	224	224	29	12.792079	0.00701000	0.9728153	1	0.00078420	0.99978280	0.99996551	1.6574900	FALSE	5130	44.558089	0.52678920	0.9940534	713	113	112	15263	372136	456528	0.9917772	970799	15375	
bin.52.st	41.472238	0.9976148	0.0003381	1232650	59	58	41	12.150551	0.01099670	0.9904246	1	0.00033250	0.99998680	0.99999501	1.5298476	FALSE	47	58.680851	0.6600903	1	16	6	6	240	193784	205521	0.9947839	425303	246	
bin.53.or	100.64018	0.9943709	0.0003400	2848092	117	117	102	22.599109	0.01344630	0.9898774	1	0.00031880	0.99992690	0.99999181	1.2213805	FALSE	637	78.832025	0.80409760	0.9964698	206	23	23	1092	1056732	1097840	0.9977898	2228840	1115	
bin.57.or	79.401135	0.9981786	0.0010406	1770622	86	86	80	23.106430	0.01744970	0.9951858	1	0.00103140	0.99996310	0.99999261	1.6377458	FALSE	3223	80.154204	0.9899516	1	65	13	13	5828	531596	601145	0.9951272	1229637	5841	
bin.59.st	46.609675	0.9947226	0.0002585	1797291	137	135	47	12.962427	0.00974340	0.9850669	1	0.00022110	0.99997450	0.99998021	1.3286111	FALSE	42	53.119047	0.8150972	1	45	35	35	173	319103	332836	0.9937661	682462	208	
bin.62.or	60.661153	0.9971050	0.0002932	3276383	124	124	61	15.349855	0.00851240	0.9937888	1	0.00027130	0.99998280	0.99999041	1.5254827	FALSE	128	71.851562	0.68431460	0.9709391	56	31	31	268	744765	775011	0.9970278	1568903	299	
bin.66.or	22.606010	0.9961925	0.0009395	2503497	99	99	23	7.973696	0.00505950	0.97863660	0.9999999	0.00033960	0.99971220	0.99996201	1.6318367	FALSE	555	24.861261	0.81346120	0.9931841	705	93	93	4749	209975	242871	0.9950360	493355	4842	

Host 3 Old

genome	coverage	breadth	nucl_dive	length	true_scaf	detected_	coverage_	coverage_	coverage_	breadth_m	breadth_e	nucl_dive	conANI_re	popANI_re	iRep	iRep_GC_	clinked_SN	SNV_distar	2_mean	d_prime_m	consensus	populatio	SNS_count	SNV_count	filtered_reads_unf	reads_mea	reads_unf	divergent	
bin.13.or	19.475823	0.9943948	0.0002715	2921087	244	244	19	8.793484	0.00518850	0.97019910	0.9999999	6.48E-05	0.99995200	0.99997451	1.6028132	FALSE	84	37.738095	0.7499441	1	136	72	72	708	211628	224213	0.9953354	470289	780	
bin.16.or	31.150011	0.9955443	0.0014448	3364973	334	334	31	10.662664	0.00587120	0.9785189	1	0.00098490	0.99963430	0.99997351	1.7189061	FALSE	3297	39.171671	0.57988090	0.9968116	1204	87	87	11887	389640	434188	0.9954067	909961	11974	
bin.18.or	312.45874	0.9968251	0.0002357	4093372	107	107	321	86.194505	0.04271460	0.9955598	1	0.00021910	0.99997150	0.99999231	1.6039743	FALSE	1237	133.75747	0.44240020	0.9745845	116	31	31	419	4673929	4845388	0.9973173	9833560	450	
bin.19.st	19.756978	0.9819393	0.0007123	3217492	529	521	12	63.124432	0.03578480	0.88498210	0.9999999	0.00051460	0.99971860	0.9999290		FALSE	3294	81.902550	0.31789090	0.9583396	801	202	202	3890	240086	458140	0.9743433	1273249	4092	
bin.20.st	21.195759	0.9920324	0.0002478	1256848	200	200	21	8.981841	0.00814230	0.96674530	0.9999999	8.66E-06	0.99994730	0.99996541	1.4785191	FALSE	32	49.75	0.7497126	1	64	42	42	224	100240	107689	0.9964590	224827	266	
bin.22.or	22.524133	0.9981168	0.0002087	2091212	75	75	22	6.071160	0.00421340	0.99160050	0.9999999	0	0.99996470	0.99998931	1.3036060	FALSE	8	15	0.7550553	1	73	22	22	266	174877	179249	0.9979516	361804	288	
bin.23.pe	36.995978	0.9501933	0.0002926	1580129	166	155	38	14.504732	0.01166200	0.9382759	1	0.00027550	0.99994670	0.9999682		FALSE	151	52.423841	0.59584390	0.9920021	79	47	47	332	220881	233000	0.9958776	477994	379	
bin.27.or	42.188883	0.8995999	0.0002386	3183055	235	234	49	23.358170	0.01319000	0.8563254	1	0.00022990	0.99684700	0.9968591		FALSE	47	72.531914	0.67444560	0.9843554	8594	8561	8559	198	498257	653646	0.9915391	1369696	8757	
bin.28.st	75.595956	0.9955635	0.0002681	1263389	146	145	77	22.676629	0.02041210	0.9889392	1	0.00023470	0.99996790	0.99998711	1.4984490	FALSE	77	67.801948	0.85138260	0.9779613	40	16	16	119	363338	381440	0.9948680	800690	135	
bin.3.ori	49.470478	0.9960054	0.0014443	3519338	205	205	50	17.500428	0.00938340	0.9884628	1	0.00092150	0.99933160	0.99999051	1.9570735	FALSE	5230	48.160803	0.84576620	0.9988065	2325	33	33	11122	647538	749058	0.9937031	1555714	11155	
bin.30.or	22.354989	0.9976830	0.0002015	2471349	119	119	22	6.391953	0.00408570	0.98784100	0.9999999	1.31E-05	0.99998070	0.99999181	1.2613794	FALSE	10	27.3	0.8040827	1	47	20	20	295	205512	212341	0.9977874	434431	315	
bin.36.or	19.883698	0.9973544	0.0002035	2424486	92	92	20	7.262188	0.00468170	0.98780030	0.9999999	7.75E-05	0.99999030	0.99999581	1.2781862	FALSE	18	24.444444	0.9052539	1	23	10	10	283	179239	186263	0.9979953	383535	293	
bin.38.st	83.805784	0.9973613	0.0002223	2673349	116	114	82	34.999476	0.02149940	0.9927697	1	0.00020890	0.99997470	0.99998901	1.1859318	FALSE	315	100.14285	0.55097830	0.9569813	67	29	29	235	831168	874614	0.9932788	1818134	264	
bin.4.str	18.806468	0.9768597	0.0010051	2506497	363	362	18	18.366644	0.01177270	0.88597630	0.9999999	0.00028550	0.99979330	0.9999437		FALSE	743	58.072678	0.53335130	0.9601873	459	125	124	4667	177489	233189	0.9870990	561707	4791	
bin.49.or	19.499674	0.9772431	0.0031769	2079370	426	426	18	12.544427	0.00888320	0.88874220	0.9999999	0.00078700	0.99875430	0.9998712		FALSE	3386	39.859716	0.38174940	0.9844356	2302	238	237	13698	153278	328107	0.9757853	812560	13935	
bin.52.st	53.206372	0.9984245	0.0002660	1232650	59	59	53	15.261321	0.01381210	0.9937906	1	0.00024670	0.99997950	0.99999421	1.5504206	FALSE	48	53.208333	0.6800829	1	25	7	7	116	249098	263376	0.9948030	548145	123	
bin.53.or	139.20877	0.9930557	0.0003309	2848092	117	117	141	31.355909	0.01865660	0.9903314	1	0.00030520	0.99990710	0.99999321	1.2147374	FALSE	679	92.580265	0.83811890	0.9904525	262	19	19	1018	1466586	1520574	0.9979422	3088593	1037	
bin.55.or	19.815305	0.9963238	0.0007153	2596212	105	105	20	6.844748	0.00426530	0.98305600	0.9999999	5.96E-06	0.99986710	0.99993881	1.4577859	FALSE	198	22.505050	0.9256659	1	339	156	156	3654	191633	210529	0.9966589	427752	3810	
bin.57.or	200.94273	0.9986524	0.0008189	1770622	86	86	205	53.557568	0.04044620	0.9971349	1	0.00078290	0.99994560	0.99998921	1.6540669	FALSE	4979	113.21992	0.99068990	0.9999754	96	19	19	4543	1348490	1483871	0.9958128	3029260	4562	
bin.59.st	22.727750	0.9926945	0.0002132	1797291	137	135	22	9.543909	0.00717380	0.97711720	0.9999999	0.00012520	0.99997660	0.99998001	1.2807442	FALSE	23	21.260869	0.5987579	1	41	35	35	205	155630	165894	0.9920904	355698	240	
bin.62.or	99.432166	0.9990361	0.0002589	3276383	124	124	100	23.812801	0.01320570	0.9946309	1	0.00023390	0.99997260	0.99998891	1.6247158	FALSE	164	75.304878	0.70166110	0.9659723	89	36	36	277	1215098	1260663	0.9972785	2559382	313	
bin.66.or	96.689531	0.9979532	0.0010433	2503497	99	99	100	27.004912	0.01713530	0.9957275	1	0.00088480	0.99972880	0.99997671	1.5549328	FALSE	6510	74.533179	0.76286800	0.9873813	676	58	58	6767	903248	1037417	0.9952794	2101361	6825	
bin.68.st	116.95738	0.9906005	0.0002294	2093210	73	72	120	35.741114	0.02479020	0.9852370	1	0.00020740	0.99967170	0.9996819		FALSE	99	74.787878	0.83404870	0.9979423	677	656	655	166	916404	968109	0.9942947	1986859	821	
bin.74.or	28.225597	0.9941991	0.0002793	1982126	372	372	29	10.904087	0.00789460	0.9659148	1	0.00011230	0.99990280	0.9999336		FALSE	34	40.588235	0.6298355	1	186	127	127	408	207850	215678	0.9978404	448474	535	
bin.8.ori	21.819957	0.9778985	0.0021518	3303725	333	333	23	11.678023	0.00649060	0.87978050	0.9999999	0.00055320	0.99806360	0.99936932	1.4050491	FALSE	2225	33.849887	0.67111150	0.9977179	5628	1833	1829	13962	268282	360396	0.9907949	752072	15791	

Supplementary Table 5

inStrain Compare output		popANI comparison genome wide							
Host 1									
genome	name1	name2	coverage_	compared_	consensus	populatio	popANI	conANI	percent_compared
bin.1.ori	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9975583	2764831	321	1	0.9999996	0.9998838	0.9935414
bin.1.ori	mapping.host1.young.index.sorte	mapping.host1.adult	0.9959420	2766904	511	0	1	0.9998153	0.9942863
bin.1.ori	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9968854	2769496	649	15	0.9999945	0.9997656	0.9952177
bin.10.st	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9478779	1529097	2800	169	0.9998894	0.9981688	0.9367303
bin.10.st	mapping.host1.young.index.sorte	mapping.host1.adult	0.9705833	1568761	527	0	1	0.9996640	0.9610286
bin.10.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9725946	1574058	2576	3	0.9999980	0.9983634	0.9642735
bin.11.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.9967324	4750371	254	12	0.9999974	0.9999465	0.9928993
bin.2.ori	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9855231	3120343	4606	567	0.9998182	0.9985238	0.9750176
bin.21.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.8963197	1329111	1339	0	1	0.9989925	0.8765320
bin.24.pe	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9840740	2286085	27	6	0.9999973	0.9999881	0.9581855
bin.24.pe	mapping.host1.young.index.sorte	mapping.host1.adult	0.9825605	2316420	29	3	0.9999987	0.9999874	0.9709000
bin.24.pe	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9974923	2326152	27	3	0.9999987	0.9999883	0.9749791
bin.25.st	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.8089457	1445105	67	4	0.9999972	0.9999536	0.7745275
bin.25.st	mapping.host1.young.index.sorte	mapping.host1.adult	0.9947834	1812317	79	1	0.9999994	0.9999564	0.9713408
bin.25.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.8097526	1455064	68	0	1	0.9999532	0.7798652
bin.3.ori	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9688664	2747919	3316	0	1	0.9987932	0.9309009
bin.32.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9424592	2260884	48	0	1	0.9999787	0.9320807
bin.35.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.6390001	1515380	515	0	1	0.9996601	0.5945438
bin.35.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.6360784	1516204	432	0	1	0.9997150	0.5948671
bin.35.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9929573	2487548	1081	0	1	0.9995654	0.9759640
bin.36.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.8955337	1753917	5	0	1	0.9999971	0.8850255
bin.36.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.8927108	1753917	1078	0	1	0.9993853	0.8850255
bin.36.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9942306	1960726	1943	4	0.9999979	0.9990090	0.9893812
bin.37.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.7413990	1366662	59	0	1	0.9999568	0.6349780
bin.38.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.8202564	1501590	5	0	1	0.9999966	0.7959888
bin.38.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.8229687	1501629	35	26	0.9999826	0.9999766	0.7960095
bin.38.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9944899	1839823	40	35	0.9999809	0.9999782	0.9752852
bin.39.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9438153	1289821	82	8	0.9999937	0.9999364	0.9081481
bin.39.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.9385107	1289904	59	0	1	0.9999542	0.9082065
bin.39.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9941109	1406528	40	6	0.9999957	0.9999715	0.9903201
bin.42.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9954210	4371002	279	21	0.9999951	0.9999361	0.9711490
bin.48.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9470663	1649748	13	0	1	0.9999921	0.9113849
bin.48.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.9451485	1649748	28	1	0.9999993	0.9999830	0.9113849
bin.48.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9963277	1743294	59	1	0.9999994	0.9999661	0.9630633
bin.49.st	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.6300301	1487829	1969	2	0.9999986	0.9986765	0.5606042

bin.49.st	mapping.host1.young.index.sorte	mapping.host1.adult	0.6405901	1456868	1810	2	0.9999986	0.9987576	0.5489383
bin.49.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.8711609	2181905	397	0	1	0.9998180	0.8221274
bin.5.ori	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9552937	2052285	2263	2	0.9999990	0.9988973	0.9476798
bin.50.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.8220260	1911884	1115	1	0.9999994	0.9994168	0.7816465
bin.58.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9952067	3693087	1077	0	1	0.9997083	0.9912768
bin.58.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.9859402	3649375	77	1	0.9999997	0.9999789	0.9795438
bin.58.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9855830	3649139	1007	0	1	0.9997240	0.9794805
bin.6.ori	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9868175	2415240	74	0	1	0.9999693	0.9760943
bin.62.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9965049	3218052	19	0	1	0.9999940	0.8573122
bin.62.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.8939828	3226624	9550	9381	0.9970926	0.9970402	0.8595959
bin.62.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.8921993	3218557	9185	9043	0.9971903	0.9971462	0.8574468
bin.64.st	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9979168	3272381	66	0	1	0.9999798	0.9827893
bin.64.st	mapping.host1.young.index.sorte	mapping.host1.adult	0.9676389	3172290	97	7	0.9999977	0.9999694	0.9527291
bin.64.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9671992	3176879	75	8	0.9999974	0.9999763	0.9541073
bin.66.st	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9026283	2988643	2864	1	0.9999996	0.9990417	0.8066710
bin.73.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.8638243	1804930	1160	0	1	0.9993573	0.8559858
bin.73.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.8627908	1805168	860	0	1	0.9995235	0.8560986
bin.73.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9977570	2092613	1646	0	1	0.9992134	0.9924191
bin.76.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.8486124	1996167	27	8	0.9999959	0.9999864	0.8338915
bin.8.ori	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.9904720	3804414	128	0	1	0.9999663	0.9466506
bin.8.ori	mapping.host1.young.index.sorte	mapping.host1.adult	0.9845780	3863136	88	0	1	0.9999772	0.9612624
bin.8.ori	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9949937	3824353	142	14	0.9999963	0.9999628	0.9516121
bin.82.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.5890105	1490992	17	0	1	0.9999885	0.5442635
bin.82.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.5872835	1492889	12	0	1	0.9999919	0.5449560
bin.82.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9879499	2652478	22	0	1	0.9999917	0.9682460
bin.84.or	mapping.host1.adult.index.sorte	mapping.host1.old.i	0.6619135	1162400	444	0	1	0.9996180	0.6167104
bin.84.or	mapping.host1.young.index.sorte	mapping.host1.adult	0.6547768	1162424	449	0	1	0.9996137	0.6167232
bin.84.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.9885993	1830242	39	0	1	0.9999786	0.9710336
bin.86.or	mapping.host1.young.index.sorte	mapping.host1.old.i	0.6924232	1315087	8	0	1	0.9999939	0.6518392

Host 2

genome	name1	name2	coverage_	compared_	consensus	populatio	popANI	conANI	percent_compared
bin.12.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9905703	4555858	11404	1399	0.9996929	0.9974968	0.9866663
bin.15.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9966754	2516142	913	0	1	0.9996371	0.9894742
bin.15.st	mapping.host2.young.instrain.in	mapping.host2.adult	0.5670701	1291551	543	0	1	0.9995795	0.5079031
bin.15.st	mapping.host2.young.instrain.in	mapping.host2.old.i	0.5665657	1291141	666	2	0.9999984	0.9994841	0.5077419
bin.18.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9958401	1478638	13	0	1	0.9999912	0.9877031

bin.20.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.8115458	2514841	37	0	1	0.9999852	0.7049220	
bin.20.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.7138523	2195593	34	0	1	0.9999845	0.6154352	
bin.20.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.7444498	1957635	15	0	1	0.9999923	0.5487344	
bin.24.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.7148483	1271745	33	1	0.9999992	0.9999740	0.5980244	
bin.25.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9965858	2735259	187	0	1	0.9999316	0.9948418	
bin.25.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.9879018	2711206	306	0	1	0.9998871	0.9860935	
bin.25.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9911231	2711026	306	0	1	0.9998871	0.9860280	
bin.36.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9986321	1684190	19	0	1	0.9999887	0.9900413	
bin.36.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.9860220	1662437	87	0	1	0.9999476	0.9772539	
bin.36.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9868607	1662440	72	0	1	0.9999566	0.9772557	
bin.37.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9870659	1501737	18	0	1	0.9999880	0.9681474	
bin.39.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9946335	3789209	78	0	1	0.9999794	0.9769641	
bin.39.st	mapping.host2.young.instrain.in	mapping.host2.adult	0.9840105	3737910	115	0	1	0.9999692	0.9637377	
bin.39.st	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9814657	3753508	173	4	0.9999989	0.9999539	0.9677593	
bin.41.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.6354910	1655108	5998	8	0.9999951	0.9963760	0.5351291	
bin.42.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9974908	2434693	36	0	1	0.9999852	0.9912090	
bin.44.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9952322	3979414	99	0	1	0.9999751	0.9805134	
bin.44.st	mapping.host2.young.instrain.in	mapping.host2.adult	0.9454162	3780390	98	0	1	0.9999740	0.9314746	
bin.44.st	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9470863	3757578	123	0	1	0.9999672	0.9258538	
bin.46.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9683086	3451694	11284	2	0.9999994	0.9967308	0.8965532	
bin.46.st	mapping.host2.young.instrain.in	mapping.host2.adult	0.9787695	3566180	13494	1	0.9999997	0.9962161	0.9262901	
bin.46.st	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9644358	3498600	12338	13	0.9999962	0.9964734	0.9087367	
bin.49.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9696531	1779231	13	0	1	0.9999926	0.9611552	
bin.49.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.7500300	1360498	755	0	1	0.9994450	0.7349522	
bin.49.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.7577995	1339916	636	24	0.9999820	0.9995253	0.7238336	
bin.52.st	mapping.host2.young.instrain.in	mapping.host2.adult	0.7175435	1858092	16	0	1	0.9999913	0.6830492	
bin.54.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.5574526	1526660	12	0	1	0.9999921	0.5077183	
bin.55.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.8664210	2875001	24	1	0.9999996	0.9999916	0.8095901	
bin.56.st	mapping.host2.young.instrain.in	mapping.host2.adult	0.8582111	3095344	5378	0	1	0.9982625	0.7495449	
bin.58.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.9751395	3693381	85	0	1	0.9999769	0.9499043	
bin.60.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9947982	3788912	35	5	0.9999986	0.9999907	0.9930344	
bin.60.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.9995117	3807100	15	0	1	0.9999960	0.9978013	
bin.60.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9952539	3788911	34	5	0.9999986	0.9999910	0.9930341	
bin.62.pe	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.8949826	1148884	53	0	1	0.9999538	0.8732384	
bin.62.pe	mapping.host2.young.instrain.in	mapping.host2.adult	0.8873040	1133215	30	0	1	0.9999735	0.8613288	
bin.62.pe	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9653941	1244135	44	0	1	0.9999646	0.9456363	
bin.65.st	mapping.host2.young.instrain.in	mapping.host2.old.i	0.8950268	4173711	1229	8	0.9999980	0.9997055	0.8582304	
bin.67.st	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9873389	3869290	257	2	0.9999994	0.9999335	0.9560863	

bin.70.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.8733239	2717462	16	1	0.9999996	0.9999941	0.8514066
bin.70.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.5696722	1551778	8	0	1	0.9999948	0.4861868
bin.70.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.5636491	1680426	14	0	1	0.9999916	0.5264934
bin.71.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.9386370	2270639	5632	5591	0.9975376	0.9975196	0.8894515
bin.73.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.8686583	2389408	49	0	1	0.9999794	0.8539882
bin.73.or	mapping.host2.young.instrain.in	mapping.host2.adult	0.7874358	2141517	23	0	1	0.9999892	0.7653906
bin.73.or	mapping.host2.young.instrain.in	mapping.host2.old.i	0.7278704	1877834	27	0	1	0.9999856	0.6711487
bin.79.or	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9127538	1354621	813	0	1	0.9993998	0.8444073
bin.81.st	mapping.host2.adult.instrain.in	mapping.host2.old.i	0.9046807	1309221	24	0	1	0.9999816	0.8577338

Host 3	popANI comparison								
genome	name1	name2	coverage_	compared_	consensus	populatio	popANI	conANI	percent_compared
bin.15.st	mapping.host3.young.instrain.in	mapping.host3.adult	0.8883539	1741640	809	0	1	0.9995354	0.8734604
bin.18.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9985691	4062117	56	1	0.9999997	0.9999862	0.9923644
bin.25.st	mapping.host3.young.instrain.in	mapping.host3.adult	0.9225360	3097452	63	22	0.9999928	0.9999796	0.9034504
bin.26.st	mapping.host3.young.instrain.in	mapping.host3.adult	0.9964165	2461797	30	0	1	0.9999878	0.9871222
bin.27.or	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.9786710	2721320	5	0	1	0.9999981	0.8633552
bin.28.st	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.8645044	1053278	8	0	1	0.9999924	0.8340708
bin.28.st	mapping.host3.young.instrain.in	mapping.host3.adult	0.8736146	1051559	5	0	1	0.9999952	0.8327095
bin.28.st	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9831798	1211350	12	0	1	0.9999900	0.9592450
bin.3.ori	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.7714689	2483893	2696	0	1	0.9989146	0.7057841
bin.3.ori	mapping.host3.young.instrain.in	mapping.host3.adult	0.7191279	1984769	4694	12	0.9999939	0.9976349	0.5639608
bin.3.ori	mapping.host3.young.instrain.in	mapping.host3.old.i	0.7633652	2457132	3553	0	1	0.9985540	0.6981801
bin.30.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.7170170	1733266	34	1	0.9999994	0.9999803	0.7013440
bin.36.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.5469130	1273971	13	1	0.9999992	0.9999897	0.5254602
bin.38.st	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9970112	2637183	46	2	0.9999992	0.9999825	0.9868585
bin.49.or	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.8026440	1245605	4091	0	1	0.9967156	0.6025073
bin.51.or	mapping.host3.young.instrain.in	mapping.host3.adult	0.9166200	3009488	826	2	0.9999993	0.9997255	0.8917701
bin.52.st	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.9871709	1192404	9	0	1	0.9999924	0.9673500
bin.52.st	mapping.host3.young.instrain.in	mapping.host3.adult	0.9818410	1194006	8	0	1	0.9999932	0.9686496
bin.52.st	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9921345	1206751	17	2	0.9999983	0.9999859	0.9789891
bin.53.or	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.9976705	2804902	238	0	1	0.9999151	0.9848354
bin.53.or	mapping.host3.young.instrain.in	mapping.host3.adult	0.9950200	2806040	281	0	1	0.9998998	0.9852350
bin.53.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9964804	2810221	362	0	1	0.9998711	0.9867030
bin.55.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.6003342	1447912	242	26	0.9999820	0.9998328	0.5577017
bin.57.or	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.9901919	1743077	57	0	1	0.9999672	0.9844433
bin.57.or	mapping.host3.young.instrain.in	mapping.host3.adult	0.9929066	1742157	27	0	1	0.9999845	0.9839237

bin.57.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9954146	1753044	52	0	1	0.9999703	0.9900724	
bin.59.st	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.7183841	1236565	7	0	1	0.9999943	0.6880160	
bin.59.st	mapping.host3.young.instrain.in	mapping.host3.adult	0.9611619	1673859	28	14	0.9999916	0.9999832	0.9313233	
bin.59.st	mapping.host3.young.instrain.in	mapping.host3.old.i	0.7243710	1224206	18	7	0.9999942	0.9999852	0.6811395	
bin.62.or	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.9945583	3212466	32	0	1	0.9999900	0.9804915	
bin.62.or	mapping.host3.young.instrain.in	mapping.host3.adult	0.9681773	3118527	42	5	0.9999983	0.9999865	0.9518200	
bin.62.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9655367	3123292	47	5	0.9999983	0.9999849	0.9532743	
bin.66.or	mapping.host3.adult.instrain.in	mapping.host3.old.i	0.7593160	1848357	753	0	1	0.9995926	0.7383100	
bin.66.or	mapping.host3.young.instrain.in	mapping.host3.adult	0.7590511	1848466	795	0	1	0.9995699	0.7383535	
bin.66.or	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9972990	2478498	1196	0	1	0.9995174	0.9900143	
bin.68.st	mapping.host3.young.instrain.in	mapping.host3.old.i	0.9870559	2034819	877	586	0.9997120	0.9995690	0.9723484	

Supplementary Table 6

SNV changes present in microbial populations	
Duncaniella freteri	
Assembly gene name	Gene name
NODE_1783_length_24024_cov_39.683216_pilon_pilon_pilon_pilon_pilon_9	PL29 family lyase N-terminal domain-containing protein
NODE_2068_length_21847_cov_40.153910_pilon_pilon_pilon_pilon_pilon_5	Hypothetical protein
NODE_21242_length_3379_cov_38.154633_pilon_pilon_pilon_pilon_pilon_1	Insulinase family protein
NODE_4047_length_13678_cov_39.049108_pilon_pilon_pilon_pil	OmpA family protein
NODE_4945_length_11834_cov_39.761185_pilon_pilon_pilon_pilon_pilon_6	Glycoside hydrolase family 43 protein
NODE_508_length_45345_cov_37.865599_pilon_pilon_pilon_pilon_pilon_12	Sodium/solute symporter
NODE_8939_length_7215_cov_40.918017_pilon_pilon_pilon_pilon_pilon_1	Helix-turn-helix domain-containing protein
NODE_971_length_33758_cov_39.519746_pilon_pilon_pilon_pilon_pilon_23	DUF4153 domain-containing protein
NODE_9977_length_6516_cov_39.052933_pilon_pilon_pilon_pilon_pilon_2	Sigma-54-dependent transcriptional regulator
Alistipes sp002358415	
NODE_8_length_54086_cov_16.410598_2	RagB/SusD family nutrient uptake outer membrane protein
Muribaculum sp	
NODE_6_length_105711_cov_11.523108_25	RagB/SusD family nutrient uptake outer membrane protein

Supplementary Table 7

List of genes that posses Consensus SNV	
Muribaculum sp	
Assembly genome name	Gene name
NODE_12_length_80848_cov_12.193621_12	Peptidase domain-containing ABC transporter
NODE_12_length_80848_cov_12.193621_12	Peptidase domain-containing ABC transporter
NODE_21_length_55343_cov_9.899486_50	NAD(P)-dependent oxidoreductase
NODE_5_length_119085_cov_10.114757_104	Hypothetical protein EEL48
NODE_55_length_2544_cov_11.543981_2	GNAT family N-acetyltransferase
NODE_55_length_2544_cov_11.543981_2	GNAT family N-acetyltransferase
NODE_55_length_2544_cov_11.543981_2	GNAT family N-acetyltransferase
Alistipes sp002358415	
Assembly genome name	Gene name
NODE_123_length_543_cov_1.873391_1	glucosamine-6-phosphate deaminase
NODE_123_length_543_cov_1.873391_1	glucosamine-6-phosphate deaminase
NODE_6_length_60732_cov_16.918045_5	DUF4906 domain-containing protein
NODE_8_length_54086_cov_16.410598_3	SusC/RagA family TonB-linked outer membrane protein
Duncaniella freteri	
Assembly genome name	Gene name
NODE_1799_length_23938_cov_38.620358_pilon_pilon_pilon_pilon_pilon_4	SusF/SusE family outer membrane protein
NODE_1996_length_22309_cov_44.583086_pilon_pilon_pilon_pilon_pilon_2	RNA polymerase sigma factor
NODE_893_length_35356_cov_38.321351_pilon_pilon_pilon_pilon_pilon_18	Hypothetical protein
Alistipes sp002428825	
Assembly genome name	Gene name
NODE_2223_length_20893_cov_80.945772_pilon_pilon_pilon_pilon_pilon_14	Methionine adenosyltransferase
NODE_2223_length_20893_cov_80.945772_pilon_pilon_pilon_pilon_pilon_14	Methionine adenosyltransferase

NODE_3553_length_15080_cov_80.518270_pilon_pilon_pilon_pilon_pilon_9	Helix-turn-helix transcriptional regulator
NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_pilon_37	Acetyl-CoA carboxylase biotin carrier protein subunit
NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_pilon_27	glutamate formimidoyltransferase
NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_pilon_67	leucine-rich repeat protein
Lachnospiraceae 14-2	
Assembly genome name	Gene name
NODE_8668_length_8461_cov_15.564240_pilon_pilon_pilon_pilon_pilon_pilon_pilon_11	Sensor histidine kinase
Lachnospiracea 14-2	
Assembly genome name	Gene name
NODE_10935_length_6698_cov_10.942797_pilon_pilon_pilon_pilon_pilon_pilon_pilon_6	IS3 family transposase
NODE_10935_length_6698_cov_10.942797_pilon_pilon_pilon_pilon_pilon_pilon_pilon_6	IS3 family transposase
NODE_1406_length_38287_cov_13.853447_pilon_pilon_pilon_pilon_pilon_pilon_pilon_5	Hypothetical gene
NODE_1406_length_38287_cov_13.853447_pilon_pilon_pilon_pilon_pilon_pilon_pilon_5	Hypothetical gene
NODE_3064_length_22622_cov_18.802987_pilon_pilon_pilon_pilon_pilon_pilon_pilon_13	IS110 family transposase
NODE_3064_length_22622_cov_18.802987_pilon_pilon_pilon_pilon_pilon_pilon_pilon_13	IS110 family transposase
NODE_31753_length_2352_cov_35.533304_pilon_pilon_pilon_pilon_pilon_pilon_pilon_1	TonB-dependent receptor
NODE_31753_length_2352_cov_35.533304_pilon_pilon_pilon_pilon_pilon_pilon_pilon_1	TonB-dependent receptor
NODE_441_length_66817_cov_14.845496_pilon_pilon_pilon_pilon_pilon_pilon_pilon_47	Hypothetical protein
UBA7182 sp 009774145	
Assembly genome name	Gene name
NODE_10_length_95390_cov_67.928929_78	Transposase
NODE_10_length_95390_cov_67.928929_78	Transposase
NODE_10_length_95390_cov_67.928929_79	Hypothetical protein
NODE_101_length_1619_cov_7.217899_3	Hypothetical protein
NODE_130_length_919_cov_23.538005_1	Hypothetical protein
NODE_170_length_726_cov_1.257319_1	Group II intron-encoded protein LtrA
NODE_19_length_70290_cov_51.386595_16	Hypothetical protein

NODE_194_length_663_cov_2.522184_1	Tyrosine-type recombinase/integrase
NODE_194_length_663_cov_2.522184_1	Tyrosine-type recombinase/integrase
NODE_201_length_651_cov_4.379791_2	Hypothetical protein
NODE_231_length_602_cov_6.990458_1	Hypothetical protein
NODE_24_length_64193_cov_59.937223_20	Hypothetical protein
NODE_3_length_170980_cov_57.840167_1	SH3 domain-containing protein
NODE_34_length_50067_cov_43.997239_43	Aminoglycoside 6-adenylytransferase
NODE_35_length_49429_cov_39.877594_51	GHKL domain-containng protein
NODE_35_length_49429_cov_39.877594_52	Hypothetical protein
NODE_35_length_49429_cov_39.877594_52	Hypothetical protein
NODE_67_length_18825_cov_43.950555_1	ABC trasnporter ATP-binding protein
NODE_73_length_12790_cov_57.935892_1	Hypothetical protein
NODE_78_length_11458_cov_58.471575_12	Transposase from transposon Tn916
NODE_9_length_105304_cov_64.501060_77	Bacitracin export permease protein BceC
NODE_99_length_2742_cov_29.739212_3	ABC transporter permease
NODE_99_length_2742_cov_29.739212_2	IS66 family insertion sequence element accesory protein TnpB
NODE_99_length_2742_cov_29.739212_4	Transposase
Alistipes sp002362235	
Assembly genome name	Gene name
NODE_1678_length_35311_cov_24.016224_pilon_pilon_pilon_pilon_pilon_2	Hypothetical protein

Supplementary Table 8

dN/dS and pN/pS ratios of SNV changes, consensus SNV and variant SNV computed by inStrain of each selected microbial population

UBA7182 sp 009774145					Alistipes sp002362235					Alistipes sp002428825			
Gene name	dN/dS	pN/pS	Age of the Host		Gene name	dN/dS	pN/pS	Age of the host		Gene name	dN/dS	pN/pS	Age of the host
NODE_99_length_2742_cov_29.739212_3	NA	0.27433628	12m		NODE_1678_length_35311_cov_24.016224_pilon_pilon_pilon_pilon_pilon_2	NA	0.1124013		Young	NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_27	NA	0.07392543	12m
NODE_99_length_2742_cov_29.739212_3	NA	0	26m		NODE_1678_length_35311_cov_24.016224_pilon_pilon_pilon_pilon_pilon_2	NA	0.1124013	Host 1	Old	NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_27	NA	0.14785085	20m
NODE_99_length_2742_cov_29.739212_2	NA	NA	12m							NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_27	NA	0.14785085	26m
NODE_99_length_2742_cov_29.739212_2	NA	NA	26m							NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_38	NA	0	12m
NODE_98_length_2962_cov_20.065511_2	NA	0	12m							NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_38	NA	0	20m
NODE_98_length_2962_cov_20.065511_2	NA	0	26m		Muribaculum sp					NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_38	NA	0	26m
NODE_90_length_5835_cov_16.013894_3	NA	0.55460751	12m		Gene name	dN/dS	pN/pS	Age of the host		NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_45	NA	0.06012903	12m
NODE_90_length_5835_cov_16.013894_3	NA	0.09243458	26m		NODE_12_length_80848_cov_12.193621_12	NA	0	12m		NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_45	NA	0.03340502	20m
NODE_9_length_105304_cov_64.501060_79	0	0.09263352	12m		NODE_12_length_80848_cov_12.193621_12	NA	0.03863224	20m		NODE_100_length_85041_cov_80.163992_pilon_pilon_pilon_pilon_pilon_45	NA	0.05010753	26m
NODE_9_length_105304_cov_64.501060_79	NA	0.04631676	26m		NODE_16_length_70730_cov_10.934341_55	NA	NA	12m		NODE_1256_length_29122_cov_91.231534_pilon_pilon_pilon_pilon_pilon_21	NA	0	12m
NODE_9_length_105304_cov_64.501060_78	NA	0.1862069	26m		NODE_16_length_70730_cov_10.934341_55	NA	NA	20m		NODE_1256_length_29122_cov_91.231534_pilon_pilon_pilon_pilon_pilon_21	NA	0	20m
NODE_9_length_105304_cov_64.501060_78	NA	0.18416067	12m		NODE_21_length_55343_cov_9.899486_50	NA	NA	12m		NODE_1256_length_29122_cov_91.231534_pilon_pilon_pilon_pilon_pilon_21	NA	0	26m
NODE_9_length_105304_cov_64.501060_77	0	0.18101105	26m		NODE_21_length_55343_cov_9.899486_50	NA	NA	20m		NODE_158_length_72559_cov_71.662832_pilon_pilon_pilon_pilon_pilon_35	NA	0	12m
NODE_9_length_105304_cov_64.501060_77	NA	0.08818487	12m		NODE_33_length_29850_cov_9.283075_25	NA	0	12m		NODE_158_length_72559_cov_71.662832_pilon_pilon_pilon_pilon_pilon_35	NA	0	20m
NODE_88_length_6694_cov_53.535590_4	NA	0.16606398	12m		NODE_33_length_29850_cov_9.283075_25	NA	0	20m		NODE_158_length_72559_cov_71.662832_pilon_pilon_pilon_pilon_pilon_35	NA	0	26m
NODE_88_length_6694_cov_53.535590_4	NA	0.15815617	26m		NODE_5_length_119085_cov_10.114757_104	NA	1.57446809	12m	Transposase	NODE_1653_length_24982_cov_91.377743_pilon_pilon_pilon_pilon_pilon_13	NA	0.34800161	12m
NODE_85_length_7681_cov_47.716202_9	NA	0.14976959	12m		NODE_5_length_119085_cov_10.114757_104	NA	1.25957447	20m		NODE_1653_length_24982_cov_91.377743_pilon_pilon_pilon_pilon_pilon_13	NA	0.34800161	20m
NODE_85_length_7681_cov_47.716202_9	NA	0.14976959	26m		NODE_52_length_10155_cov_9.644374_1	NA	NA	12m		NODE_1653_length_24982_cov_91.377743_pilon_pilon_pilon_pilon_pilon_13	NA	0.34800161	26m
NODE_78_length_11458_cov_58.471575_12	NA	0.19966646	26m		NODE_52_length_10155_cov_9.644374_1	NA	NA	20m		NODE_16562_length_4175_cov_96.592233_pilon_pilon_pilon_pilon_4	NA	0.10823146	12m
NODE_78_length_11458_cov_58.471575_12	NA	0.12287167	12m		NODE_55_length_2544_cov_11.543981_2	NA	0.52758275	12m		NODE_16562_length_4175_cov_96.592233_pilon_pilon_pilon_pilon_4	NA	0.10823146	20m
NODE_78_length_11458_cov_58.471575_1	NA	NA	12m		NODE_55_length_2544_cov_11.543981_2	NA	0.52758275	20m		NODE_16562_length_4175_cov_96.592233_pilon_pilon_pilon_pilon_4	NA	0.10823146	26m
NODE_78_length_11458_cov_58.471575_1	NA	NA	26m		NODE_59_length_990_cov_4.694414_1	NA	0.25445705	12m		NODE_18581_length_3784_cov_53.473854_pilon_pilon_pilon_pilon_3	NA	0.03165468	12m
NODE_77_length_11460_cov_55.390934_12	NA	0.55772647	12m		NODE_59_length_990_cov_4.694414_1	NA	0.5089141	20m		NODE_18581_length_3784_cov_53.473854_pilon_pilon_pilon_pilon_3	NA	0.02877698	20m
NODE_77_length_11460_cov_55.390934_12	NA	NA	26m		NODE_6_length_105711_cov_11.523108_25	NA	NA	12m		NODE_18581_length_3784_cov_53.473854_pilon_pilon_pilon_pilon_3	NA	0.03956835	26m
NODE_73_length_12790_cov_57.935892_1	NA	0	12m		NODE_6_length_105711_cov_11.523108_25	NA	NA	20m		NODE_2034_length_22041_cov_64.834713_pilon_pilon_pilon_pilon_11	NA	0.0775452	12m
NODE_73_length_12790_cov_57.935892_1	NA	0	26m							NODE_2034_length_22041_cov_64.834713_pilon_pilon_pilon_pilon_11	NA	0.06203616	20m
NODE_71_length_15077_cov_41.725533_17	NA	NA	12m							NODE_2034_length_22041_cov_64.834713_pilon_pilon_pilon_pilon_11	NA	0.06203616	26m
NODE_71_length_15077_cov_41.725533_17	NA	NA	26m							NODE_2210_length_20971_cov_78.728294_pilon_pilon_pilon_pilon_19	NA	0.17147335	12m
NODE_7_length_106511_cov_47.870568_1	NA	0.18722182	12m							NODE_2210_length_20971_cov_78.728294_pilon_pilon_pilon_pilon_19	NA	0.22863114	20m
NODE_7_length_106511_cov_47.870568_1	NA	0.13372987	26m							NODE_2210_length_20971_cov_78.728294_pilon_pilon_pilon_pilon_19	NA	0.17147335	26m
NODE_67_length_18825_cov_43.950555_1	0.2902655	0.14513274	12m		Lachnospiraceae 14-2					NODE_2223_length_20893_cov_80.945772_pilon_pilon_pilon_pilon_14	NA	0	12m
NODE_67_length_18825_cov_43.950555_1	0.2902655	0.14513274	26m		Gene name	dN/dS	pN/pS	Age of host		NODE_2223_length_20893_cov_80.945772_pilon_pilon_pilon_pilon_14	NA	0	20m
NODE_66_length_19507_cov_51.754812_1	NA	0.44666667	26m		NODE_441_length_66817_cov_14.845496_pilon_pilon_pilon_pilon_pilon_pilon_47	NA	NA	12m		NODE_2223_length_20893_cov_80.945772_pilon_pilon_pilon_pilon_14	NA	0	26m
NODE_66_length_19507_cov_51.754812_1	NA	0.1675	12m		NODE_441_length_66817_cov_14.845496_pilon_pilon_pilon_pilon_pilon_pilon_47	NA	0.2775715	20m		NODE_29778_length_2558_cov_57.658010_pilon_pilon_pilon_pilon_2	NA	0	12m
NODE_64_length_20532_cov_63.479443_19	NA	NA	12m		NODE_31753_length_2352_cov_35.533304_pilon_pilon_pilon_pilon_pilon_1	NA	0.2146549	12m		NODE_29778_length_2558_cov_57.658010_pilon_pilon_pilon_pilon_2	NA	0	20m
NODE_64_length_20532_cov_63.479443_19	NA	NA	26m		NODE_31753_length_2352_cov_35.533304_pilon_pilon_pilon_pilon_pilon_1	0	0.2253877	20m		NODE_29778_length_2558_cov_57.658010_pilon_pilon_pilon_pilon_2	NA	0.02456025	26m
NODE_62_length_21107_cov_40.466049_21	NA	0.15530547	26m		NODE_31071_length_2399_cov_14.620307_pilon_pilon_pilon_pilon_pilon_pilon_3	NA	0	12m		NODE_3553_length_15080_cov_80.518270_pilon_pilon_pilon_pilon_9	NA	0	12m
NODE_62_length_21107_cov_40.466049_21	NA	0.12942122	12m		NODE_31071_length_2399_cov_14.620307_pilon_pilon_pilon_pilon_pilon_pilon_3	NA	0	20m		NODE_3553_length_15080_cov_80.518270_pilon_pilon_pilon_pilon_9	NA	0	20m
NODE_6_length_107280_cov_47.321176_103	NA	0.4330617	12m		NODE_3064_length_22622_cov_18.802987_pilon_pilon_pilon_pilon_pilon_pilon_13	NA	0.1384298	12m		NODE_3553_length_15080_cov_80.518270_pilon_pilon_pilon_pilon_9	NA	0	26m
NODE_6_length_107280_cov_47.321176_103	NA	0.21653085	26m		NODE_3064_length_22622_cov_18.802987_pilon_pilon_pilon_pilon_pilon_pilon_13	NA	0.1384298	20m		NODE_406_length_49882_cov_69.203625_pilon_pilon_pilon_pilon_14	NA	0.08552205	12m
NODE_58_length_24492_cov_63.543641_33	NA	1.09004739	26m	IS1182 family transposase	NODE_2088_length_29727_cov_17.761627_pilon_pilon_pilon_pilon_pilon_pilon_15	NA	0	12m		NODE_406_length_49882_cov_69.203625_pilon_pilon_pilon_pilon_14	NA	0.07126838	20m
NODE_58_length_24492_cov_63.543641_33	NA	NA	12m		NODE_2088_length_29727_cov_17.761627_pilon_pilon_pilon_pilon_pilon_pilon_15	NA	0	20m		NODE_406_length_49882_cov_69.203625_pilon_pilon_pilon_pilon_14	NA	0.03167483	26m
NODE_54_length_28751_cov_61.213643_5	NA	0.9273399	26m		NODE_1406_length_38287_cov_13.853447_pilon_pilon_pilon_pilon_pilon_5	NA	0.2176678	12m		NODE_406_length_49882_cov_69.203625_pilon_pilon_pilon_pilon_33	NA	0.05063343	12m
NODE_54_length_28751_cov_61.213643_5	NA	0.25759442	12m		NODE_1406_length_38287_cov_13.853447_pilon_pilon_pilon_pilon_pilon_5	NA	0.2720848	20m		NODE_406_length_49882_cov_69.203625_pilon_pilon_pilon_pilon_33	NA	0.06076012	20m
NODE_54_length_28751_cov_61.213643_1	NA	0.29760609	12m		NODE_1406_length_38287_cov_13.853447_pilon_pilon_pilon_pilon_pilon_17	NA	NA	12m		NODE_406_length_49882_cov_69.203625_pilon_pilon_pilon_pilon_33	NA	0.05063343	26m
NODE_54_length_28751_cov_61.213643_1	NA	0.29760609	26m		NODE_1406_length_38287_cov_13.853447_pilon_pilon_pilon_pilon_pilon_17	NA	NA	20m		NODE_439_length_48422_cov_89.980793_pilon_pilon_pilon_pilon_46	NA	0	12m
NODE_49_length_31401_cov_57.633987_4	NA	0.09090909	12m		NODE_10935_length_6698_cov_10.942797_pilon_pilon_pilon_pilon_pilon_pilon_7	NA	NA	12m		NODE_439_length_48422_cov_89.980793_pilon_pilon_pilon_pilon_46	NA	0	20m
NODE_49_length_31401_cov_57.633987_4	NA	0.06818182	26m		NODE_10935_length_6698_cov_10.942797_pilon_pilon_pilon_pilon_pilon_pilon_7	NA	NA	20m		NODE_439_length_48422_cov_89.980793_pilon_pilon_pilon_pilon_46	NA	0	26m
NODE_48_length_32954_cov_55.446999_15	NA	0.97225942	12m		NODE_10935_length_6698_cov_10.942797_pilon_pilon_pilon_pilon_pilon_pilon_6	NA	NA	12m		NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_13	NA	0.30276243	12m
NODE_48_length_32954_cov_55.446999_15	NA	0.48612971	26m		NODE_10935_length_6698_cov_10.942797_pilon_pilon_pilon_pilon_pilon_pilon_6	NA	NA	20m		NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_13	NA	0.30276243	20m
NODE_43_length_36908_cov_41.273221_1	0	0.19225716	26m							NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_13	NA	0.30276243	26m
NODE_43_length_36908_cov_41.273221_1	NA	0.11215001	12m							NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_34	NA	0.40677966	12m
NODE_38_length_44970_cov_56.672933_1	NA	NA	12m		Lachnospiraceae 14-2					NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_34	NA	0.50847458	20m
NODE_38_length_44970_cov_56.672933_1	NA	NA	26m		Gene name	dN/dS	pN/pS	Age of host		NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_34	NA	0.50847458	26m
NODE_37_length_46007_cov_52.524211_36	NA	1.06532663	12m		NODE_8668_length_8461_cov_15.564240_pilon_pilon_pilon_pilon_pilon_pilon_11	NA	0.2655022	12m		NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_6	NA	0	12m
NODE_37_length_46007_cov_52.524211_36	NA	0.39949749	26m		NODE_8668_length_8461_cov_15.564240_pilon_pilon_pilon_pilon_pilon_pilon_11	NA	0	20m		NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_6	NA	0	20m
NODE_35_length_49429_cov_39.877594_52	NA	0	12m		NODE_8668_length_8461_cov_15.564240_pilon_pilon_pilon_pilon_pilon_pilon_11	0	NA	26m		NODE_456_length_47370_cov_71.541646_pilon_pilon_pilon_pilon_6	NA	0	26m
NODE_35_length_49429_cov_39.877594_52	NA	0	26m							NODE_4872_length_11973_cov_73.590955_pilon_pilon_pilon_pilon_5	NA	0.07853982	12m
NODE_35_length_49429_cov_39.877594_51	NA	0.11439114	12m		Duncaniella freteri					NODE_4872_length_11973_cov_73.590955_pilon_pilon_pilon_pilon_5	NA	0.07853982	20m
NODE_35_length_49429_cov_39.877594_51	NA	0.11439114	26m		Gene name	dN/dS	pN/pS	Age of host		NODE_4872_length_11973_cov_73.590955_pilon_pilon_pilon_pilon_5	NA	0.07853982	26m
NODE_35_length_49429_cov_39.877594_50	0	0.08241445	26m		NODE_9977_length_6516_cov_39.052933_pilon_pilon_pilon_pilon_2	NA	0	12m		NODE_511_length_45190_cov_89.192179_pilon_pilon_pilon_pilon_4	NA	0.30735931	12m
NODE_35_length_49429_cov_39.877594_50	NA	0.06339573	12m		NODE_9977_length_6516_cov_39.052933_pilon_pilon_pilon_pilon_2	NA	0.6478076	26m		NODE_511_length_45190_cov_89.192179_pilon_pilon_pilon_pilon_4	NA	0.15367965	20m
NODE_34_length_50067_cov_43.997239_43	NA	0	12m		NODE_971_length_33758_cov_39.519746_pilon_pilon_pilon_pilon_23	NA	NA	12m		NODE_511_length_45190_cov_89.192179_pilon_pilon_pilon_pilon_4	NA	0.30735931	26m
NODE_34_length_50067_cov_43.997239_43	NA	0	26m		NODE_971_length_33758_cov_39.519746_pilon_pilon_pilon_pilon_23	NA	NA	26m		NODE_5411_length_11012_cov_71.679839_pilon_pilon_pilon_pilon_5	NA	0	12m
NODE_3_length_170980_cov_57.840167_51	NA	0.23608505	12m		NODE_8939_length_7215_cov_40.918017_pilon_pilon_pilon_pilon_1	NA	NA	12m		NODE_5411_length_11012_cov_71.679839_pilon_pilon_pilon_pilon_5	NA	0	20m
NODE_3_length_170980_cov_57.840167_51	NA	0.20235862	26m		NODE_8939_length_7215_cov_40.918017_pilon_pilon_pilon_pilon_1	NA	NA	26m		NODE_5411_length_11012_cov_71.679839_pilon_pilon_pilon_pilon_5	NA	0	26m
NODE_3_length_170980_cov_57.840167_50	NA	0.0558235	12m		NODE_893_length_35356_cov_38.321351_pilon_pilon_pilon_pilon_pilon_18	NA	NA	12m		NODE_591_length_42685_cov_71.541239_pilon_pilon_pilon_pilon_14	NA	NA	12m
NODE_3_length_170980_cov_57.840167_50	NA	0.0558235	26m		NODE_893_length_35356_cov_38.321351_pilon_pilon_pilon_pilon_pilon_18	NA	NA	26m		NODE_591_length_42685_cov_71.541239_pilon_pilon_pilon_pilon_14	NA	NA	20m
NODE_3_length_170980_cov_57.840167_1	0	0.54748603	12m		NODE_772_length_37656_cov_38.382117_pilon_pilon_pilon_pilon_8	NA	NA	12m		NODE_591_length_42685_cov_71.541239_pilon_pilon_pilon_pilon_14	NA	NA	26m
NODE_3_length_170980_cov_57.840167_1	0	NA	26m		NODE_772_length_37656_cov_38.382117_pilon_pilon_pilon_pilon_8	NA	NA	26m		NODE_593_length_42643_cov_77.871959_pilon_pilon_pilon_pilon_3	NA	0	12m
NODE_27_length_59548_cov_44.213465_42	NA	0	12m		NODE_508_length_45345_cov_37.865599_pilon_pilon_pilon_pilon_pilon_12	NA	0	12m		NODE_593_length_42643_cov_77.871959_pilon_pilon_pilon_pilon_3	NA	0	20m

NODE_27_length_59548_cov_44.213465_42	NA	0	26m		NODE_508_length_45345_cov_37.865599_pilon_pilon_pilon_pilon_12	NA	0	26m			NODE_593_length_42643_cov_77.871959_pilon_pilon_pilon_pilon_3	NA	0	26m
NODE_241_length_586_cov_13.267191_1	NA	NA	12m		NODE_4945_length_11834_cov_39.761185_pilon_pilon_pilon_pilon_6	NA	NA	12m			NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_24	NA	0.04265126	12m
NODE_241_length_586_cov_13.267191_1	NA	NA	26m		NODE_4945_length_11834_cov_39.761185_pilon_pilon_pilon_pilon_6	NA	NA	26m			NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_24	NA	0.0663464	20m
NODE_24_length_64193_cov_59.937223_20	NA	0.09761355	26m		NODE_4047_length_13678_cov_39.049108_pilon_pilon_pilon_pilon_pilon_5	NA	0	12m			NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_24	NA	0.0331732	26m
NODE_24_length_64193_cov_59.937223_20	NA	0.07986563	12m		NODE_4047_length_13678_cov_39.049108_pilon_pilon_pilon_pilon_pilon_5	0	NA	26m			NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_37	NA	0	12m
NODE_239_length_587_cov_45.403922_1	NA	0.74039882	26m		NODE_4047_length_13678_cov_39.049108_pilon_pilon_pilon_pilon_4	NA	0	12m			NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_37	NA	0	20m
NODE_239_length_587_cov_45.403922_1	NA	0.59231905	12m		NODE_4047_length_13678_cov_39.049108_pilon_pilon_pilon_pilon_4	NA	0	26m			NODE_600_length_42397_cov_65.448822_pilon_pilon_pilon_pilon_37	NA	0	26m
NODE_231_length_602_cov_6.990458_1	NA	0.28	26m		NODE_21242_length_3379_cov_38.154633_pilon_pilon_pilon_pilon_1	NA	0.2903178	12m			NODE_615_length_41917_cov_91.941188_pilon_pilon_pilon_pilon_26	NA	0.04682515	12m
NODE_231_length_602_cov_6.990458_1	NA	0.14	12m		NODE_21242_length_3379_cov_38.154633_pilon_pilon_pilon_pilon_1	NA	0	26m			NODE_615_length_41917_cov_91.941188_pilon_pilon_pilon_pilon_26	NA	0.04682515	20m
NODE_201_length_651_cov_4.379791_2	NA	0.71938776	12m		NODE_2068_length_21847_cov_40.153910_pilon_pilon_pilon_pilon_5	NA	NA	12m			NODE_615_length_41917_cov_91.941188_pilon_pilon_pilon_pilon_26	NA	0.06555521	26m
NODE_201_length_651_cov_4.379791_2	NA	0.47959184	26m		NODE_2068_length_21847_cov_40.153910_pilon_pilon_pilon_pilon_5	NA	NA	26m			NODE_67_length_93862_cov_60.913343_pilon_pilon_pilon_pilon_67	NA	0	12m
NODE_199_length_653_cov_1.529514_2	NA	NA	12m		NODE_1996_length_22309_cov_44.583086_pilon_pilon_pilon_pilon_2	NA	0	12m			NODE_67_length_93862_cov_60.913343_pilon_pilon_pilon_pilon_67	NA	0	20m
NODE_199_length_653_cov_1.529514_2	NA	NA	26m		NODE_1996_length_22309_cov_44.583086_pilon_pilon_pilon_pilon_2	NA	0	26m			NODE_67_length_93862_cov_60.913343_pilon_pilon_pilon_pilon_67	NA	0	26m
NODE_194_length_663_cov_2.522184_1	NA	0.40471785	26m		NODE_1799_length_23938_cov_38.620358_pilon_pilon_pilon_pilon_4	NA	0	12m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_25	NA	0.07625444	12m
NODE_194_length_663_cov_2.522184_1	NA	0.21584952	12m		NODE_1799_length_23938_cov_38.620358_pilon_pilon_pilon_pilon_4	NA	0	26m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_25	NA	0.0508363	20m
NODE_19_length_70290_cov_51.386595_65	NA	0.17448117	12m		NODE_1783_length_24024_cov_39.683216_pilon_pilon_pilon_pilon_9	NA	NA	12m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_25	NA	0.15250889	26m
NODE_19_length_70290_cov_51.386595_65	NA	0.17448117	26m		NODE_1783_length_24024_cov_39.683216_pilon_pilon_pilon_pilon_9	NA	NA	26m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_27	NA	0.24492187	12m
NODE_19_length_70290_cov_51.386595_6	NA	0.38907743	12m		NODE_13764_length_4919_cov_73.604235_pilon_pilon_pilon_pilon_5	NA	0.17261	12m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_27	NA	0.26125	20m
NODE_19_length_70290_cov_51.386595_6	NA	0.38907743	26m		NODE_13764_length_4919_cov_73.604235_pilon_pilon_pilon_pilon_5	NA	0.2301467	26m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_27	NA	0.20410156	26m
NODE_19_length_70290_cov_51.386595_3	NA	0	12m								NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_62	NA	0	12m
NODE_19_length_70290_cov_51.386595_3	NA	0	26m								NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_62	NA	0	20m
NODE_19_length_70290_cov_51.386595_28	NA	0.51974156	12m		Alistipes sp002358415						NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_62	NA	0	26m
NODE_19_length_70290_cov_51.386595_28	NA	0.51974156	26m		Gene name	dN/dS	pN/pS	Age of host			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_66	NA	0.14140731	12m
NODE_19_length_70290_cov_51.386595_2	NA	NA	12m		NODE_123_length_543_cov_1.873391_1	NA	0	12m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_66	NA	0.21211096	20m
NODE_19_length_70290_cov_51.386595_2	NA	NA	26m		NODE_123_length_543_cov_1.873391_1	NA	0	20m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_66	NA	0.07070365	26m
NODE_19_length_70290_cov_51.386595_16	NA	NA	12m		NODE_123_length_543_cov_1.873391_1	0	NA	26m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_67	NA	NA	12m
NODE_19_length_70290_cov_51.386595_16	NA	NA	26m		NODE_13_length_38799_cov_17.086592_13	NA	0.1478845	12m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_67	NA	NA	20m
NODE_19_length_70290_cov_51.386595_1	NA	0.41796517	12m		NODE_13_length_38799_cov_17.086592_13	NA	0.1971793	20m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_67	NA	NA	26m
NODE_19_length_70290_cov_51.386595_1	NA	0.41796517	26m		NODE_13_length_38799_cov_17.086592_13	NA	0.1971793	26m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_68	NA	0.07715134	12m
NODE_170_length_726_cov_1.257319_1	NA	0.12738854	12m		NODE_47_length_12296_cov_19.285048_1	NA	0	12m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_68	NA	0	20m
NODE_170_length_726_cov_1.257319_1	NA	0.06369427	26m		NODE_47_length_12296_cov_19.285048_1	NA	0.1016975	20m			NODE_79_length_89747_cov_63.732418_pilon_pilon_pilon_pilon_68	NA	0	26m
NODE_17_length_77453_cov_64.015379_54	NA	0	12m		NODE_47_length_12296_cov_19.285048_1	NA	0	26m			NODE_915_length_34877_cov_78.230946_pilon_pilon_pilon_pilon_22	NA	0.28717949	12m
NODE_17_length_77453_cov_64.015379_54	NA	0	26m		NODE_5_length_68068_cov_15.929299_37	NA	0.1519791	12m			NODE_915_length_34877_cov_78.230946_pilon_pilon_pilon_pilon_22	NA	0.21538462	20m
NODE_169_length_727_cov_10.484615_1	NA	0.38265306	12m		NODE_5_length_68068_cov_15.929299_37	NA	0.1519791	20m			NODE_915_length_34877_cov_78.230946_pilon_pilon_pilon_pilon_22	NA	0.35897436	26m
NODE_169_length_727_cov_10.484615_1	NA	0.25510204	26m		NODE_5_length_68068_cov_15.929299_37	NA	0.3039582	26m			NODE_9257_length_6998_cov_74.763791_pilon_pilon_pilon_pilon_6	NA	0.49922384	12m
NODE_163_length_744_cov_6.772114_2	NA	0.23272727	12m		NODE_6_length_60732_cov_16.918045_5	NA	NA	12m			NODE_9257_length_6998_cov_74.763791_pilon_pilon_pilon_pilon_6	NA	0.3328159	20m
NODE_163_length_744_cov_6.772114_2	NA	0.23272727	26m		NODE_6_length_60732_cov_16.918045_5	NA	NA	20m			NODE_9257_length_6998_cov_74.763791_pilon_pilon_pilon_pilon_6	NA	0.49922384	26m
NODE_163_length_744_cov_6.772114_1	0	NA	26m		NODE_6_length_60732_cov_16.918045_5	NA	NA	26m						
NODE_163_length_744_cov_6.772114_1	NA	0	12m		NODE_8_length_54086_cov_16.410598_2	NA	NA	12m						
NODE_15_length_80049_cov_50.213062_41	NA	0.10005621	12m		NODE_8_length_54086_cov_16.410598_2	NA	NA	20m						
NODE_15_length_80049_cov_50.213062_41	NA	0.06003373	26m		NODE_8_length_54086_cov_16.410598_2	NA	NA	26m						
NODE_148_length_836_cov_32.976285_2	NA	0.53283302	26m		NODE_8_length_54086_cov_16.410598_3	NA	NA	12m						
NODE_148_length_836_cov_32.976285_2	NA	0.17761101	12m		NODE_8_length_54086_cov_16.410598_3	NA	NA	20m						
NODE_148_length_836_cov_32.976285_1	NA	0.14285714	12m		NODE_8_length_54086_cov_16.410598_3	NA	NA	26m						
NODE_148_length_836_cov_32.976285_1	NA	0.07142857	26m											
NODE_147_length_839_cov_3.463255_1	0.1412155	0.28243105	12m											
NODE_147_length_839_cov_3.463255_1	NA	0.14121553	26m											
NODE_14_length_81495_cov_42.970522_6	NA	0.26385809	12m											
NODE_14_length_81495_cov_42.970522_6	NA	0.26385809	26m											
NODE_14_length_81495_cov_42.970522_5	NA	0.54054054	12m											
NODE_14_length_81495_cov_42.970522_5	NA	0.33783784	26m											
NODE_137_length_875_cov_11.733750_1	NA	0.12666667	26m											
NODE_137_length_875_cov_11.733750_1	NA	0	12m											
NODE_130_length_919_cov_23.538005_1	NA	0.49848024	26m											
NODE_130_length_919_cov_23.538005_1	NA	0.37386018	12m											
NODE_110_length_1103_cov_8.850877_1	NA	0.20374682	26m											
NODE_110_length_1103_cov_8.850877_1	NA	0.16299746	12m											
NODE_101_length_1619_cov_7.217899_3	0	0	26m											
NODE_101_length_1619_cov_7.217899_3	NA	0	12m											
NODE_10_length_95390_cov_67.928929_79	0	NA	12m											
NODE_10_length_95390_cov_67.928929_79	0	NA	26m											
NODE_10_length_95390_cov_67.928929_78	NA	1.92467452	26m											
NODE_10_length_95390_cov_67.928929_78	NA	1.37476751	12m											

Supplementary Table 9

Annotation of two circular contigs usint the RAST-tk in the PATRIC Bacterial Bioinformatics Resource CEnter

Genome	Genome ID	Accession	BRC ID	RefSeq Locus Tag	Alt Locus Tag	Feature ID	Annotation	Feature Type	Start	End	Length	Strand	FIGfam ID	PATRIC genus-spec	PATRIC cross-genu	Protein ID	AA Length	Gene Symbol	Product	GO
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.55			PATRIC.10686.3.106	PATRIC	CDS	15803	16234	432	-			PGF_00179802		143		Phage protein Mup36, J	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.57			PATRIC.10686.3.106	PATRIC	CDS	16662	16835	174	-			PGF_03302519		57		Phage major head subunit Mup34, T	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.59			PATRIC.10686.3.106	PATRIC	CDS	17228	17434	207	-					68		Phage major capsid protein of Caudovirales	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.77			PATRIC.10686.3.106	PATRIC	CDS	21571	21756	186	-			PGF_00119183		61		Phage virion morphogenesis protein Mup31, G	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.82			PATRIC.10686.3.106	PATRIC	CDS	22959	23567	609	-			PGF_04567559		202		Phage virion morphogenesis protein Mup30, F	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.84			PATRIC.10686.3.106	PATRIC	CDS	24452	24916	465	-			PGF_01645285		154		Phage protein Mup29, H	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.85			PATRIC.10686.3.106	PATRIC	CDS	25045	25467	423	-			PGF_01645285		140		Phage protein Mup29, H	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.86			PATRIC.10686.3.106	PATRIC	CDS	25440	25604	165	-			PGF_00389737		54		Phage terminase, large subunit Mup28, E	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.88			PATRIC.10686.3.106	PATRIC	CDS	25729	26808	1080	-			PGF_00389737		359		Phage terminase, large subunit Mup28, E	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.89			PATRIC.10686.3.106	PATRIC	CDS	26808	26999	192	-			PGF_06493692		63		Phage terminase, small subunit Mup27, D	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.92			PATRIC.10686.3.106	PATRIC	CDS	27396	27692	297	-			PGF_01809937		98		Phage protein Mup26	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.93			PATRIC.10686.3.106	PATRIC	CDS	27689	28117	429	-			PGF_03348458		142		Phage protein Mup25, orf4	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.113			PATRIC.10686.3.106	PATRIC	CDS	35073	35795	723	-			PGF_12761319		240		Phage replication initiation ATPase Mup04, B	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.114			PATRIC.10686.3.106	PATRIC	CDS	35838	37001	1164	-			PGF_07691576		387		DNA transposition protein Mup03, A	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.115			PATRIC.10686.3.106	PATRIC	CDS	37001	37531	531	-			PGF_07691576		176		DNA transposition protein Mup03, A	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.119			PATRIC.10686.3.106	PATRIC	CDS	38473	39207	735	#ERROR!			PGF_00091980		244		Phage repressor protein cI	
Circular_contig_395Kb	10686.3	10686.3.con.0001	fig|10686.3.peg.21			PATRIC.10686.3.106	PATRIC	CDS	6090	6794	705	-			PGF_08225224		234		Phage protein	

Genome	Genome ID	Accession	BRC ID	RefSeq Locus Tag	Alt Locus Tag	Feature ID	Annotation	Feature Type	Start	End	Length	Strand	FIGfam ID	PATRIC genus-spec	PATRIC cross-genu	Protein ID	AA Length	Gene Symbol	Product	GO
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.34			PATRIC.10738.3.107	PATRIC	CDS	10135	10305	171	#ERROR!			PGF_08990978		56		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.35			PATRIC.10738.3.107	PATRIC	CDS	10292	10624	333	#ERROR!					110		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.36			PATRIC.10738.3.107	PATRIC	CDS	10658	10813	156	#ERROR!			PGF_00559796		51		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.37			PATRIC.10738.3.107	PATRIC	CDS	10823	10957	135	#ERROR!					44		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.38			PATRIC.10738.3.107	PATRIC	CDS	10963	11157	195	-			PGF_00032855		64		Phage protein Ogr	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.39			PATRIC.10738.3.107	PATRIC	CDS	11195	11503	309	-			PGF_02383496		102		Phage portal vertex protein GpQ	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.40			PATRIC.10738.3.107	PATRIC	CDS	11547	11702	156	-					51		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.41			PATRIC.10738.3.107	PATRIC	CDS	11689	12276	588	-					195		Phage capsid packaging protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.3			PATRIC.10738.3.107	PATRIC	CDS	1188	1832	645	-			PGF_08249353		214		involved in phage exclusion	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.42			PATRIC.10738.3.107	PATRIC	CDS	12276	12815	540	-			PGF_01647629		179		Phage terminase, ATPase subunit GpP	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.43			PATRIC.10738.3.107	PATRIC	CDS	12784	12963	180	-			PGF_01647629		59		Phage terminase, ATPase subunit GpP	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.44			PATRIC.10738.3.107	PATRIC	CDS	12960	14720	1761	-			PGF_01647629		586		Phage terminase, ATPase subunit GpP	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.45			PATRIC.10738.3.107	PATRIC	CDS	14684	14773	90	-					29		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.46			PATRIC.10738.3.107	PATRIC	CDS	14896	15801	906	#ERROR!			PGF_00078035		301		Phage capsid scaffolding protein GpO	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.47			PATRIC.10738.3.107	PATRIC	CDS	15795	15890	96	#ERROR!					31		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.48			PATRIC.10738.3.107	PATRIC	CDS	15910	16116	207	#ERROR!					68		Phage major capsid protein of Caudovirales	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.49			PATRIC.10738.3.107	PATRIC	CDS	16239	16727	489	#ERROR!			PGF_01555009		162		Phage major capsid protein GpN	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.50			PATRIC.10738.3.107	PATRIC	CDS	16765	16860	96	#ERROR!					31		Phage major capsid protein of Caudovirales	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.51			PATRIC.10738.3.107	PATRIC	CDS	16864	17043	180	#ERROR!					59		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.52			PATRIC.10738.3.107	PATRIC	CDS	17086	17367	282	#ERROR!			PGF_00218886		93		Phage terminase, endonuclease subunit GpM	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.53			PATRIC.10738.3.107	PATRIC	CDS	17318	17596	279	#ERROR!					92		Phage terminase, small subunit	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.54			PATRIC.10738.3.107	PATRIC	CDS	17662	17763	102	#ERROR!					33		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.55			PATRIC.10738.3.107	PATRIC	CDS	17760	18281	522	#ERROR!			PGF_00118026		173		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.56			PATRIC.10738.3.107	PATRIC	CDS	18321	18653	333	#ERROR!			PGF_00118026		110		Phage protein (ACLAME 316)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.57			PATRIC.10738.3.107	PATRIC	CDS	18673	19383	711	-					236		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.58			PATRIC.10738.3.107	PATRIC	CDS	19364	19516	153	#ERROR!					50		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.4			PATRIC.10738.3.107	PATRIC	CDS	1941	2216	276	-			PGF_00106638		91		Phage repressor	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.59			PATRIC.10738.3.107	PATRIC	CDS	19629	19970	342	#ERROR!					113		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.60			PATRIC.10738.3.107	PATRIC	CDS	19977	20114	138	#ERROR!					45		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.61			PATRIC.10738.3.107	PATRIC	CDS	20095	20508	414	#ERROR!			PGF_10331066		137		Phage tail fiber protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.62			PATRIC.10738.3.107	PATRIC	CDS	20512	20940	429	#ERROR!			PGF_10057187		142		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.63			PATRIC.10738.3.107	PATRIC	CDS	20969	21196	228	#ERROR!					75		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.64			PATRIC.10738.3.107	PATRIC	CDS	21202	21513	312	#ERROR!			PGF_12845357		103		Phage holin #Lambda-like group I holin	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.65			PATRIC.10738.3.107	PATRIC	CDS	21515	22384	870	#ERROR!			PGF_08225224		289		Phage tail fiber protein (long tail fiber)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.5			PATRIC.10738.3.107	PATRIC	CDS	2156	2656	501	-			PGF_00106638		166		Phage repressor	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.66			PATRIC.10738.3.107	PATRIC	CDS	22384	22560	177	#ERROR!					58		Putative endolysin or lysozyme (ACLAME 334)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.67			PATRIC.10738.3.107	PATRIC	CDS	22615	22809	195	#ERROR!			PGF_03839527		64		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.68			PATRIC.10738.3.107	PATRIC	CDS	22927	23238	312	#ERROR!					103		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.69			PATRIC.10738.3.107	PATRIC	CDS	23449	23964	516	#ERROR!			PGF_09945558		171		Phage tail length tape-measure protein GpT	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.70			PATRIC.10738.3.107	PATRIC	CDS	23957	24364	408	#ERROR!					135		Putative tail protein (ACLAME 1)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.71			PATRIC.10738.3.107	PATRIC	CDS	24319	25107	789	#ERROR!			PGF_05510527		262		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.72			PATRIC.10738.3.107	PATRIC	CDS	25176	25355	180	#ERROR!					59		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.73			PATRIC.10738.3.107	PATRIC	CDS	25352	25678	327	#ERROR!					108		Phage baseplate protein (ACLAME 37)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.74			PATRIC.10738.3.107	PATRIC	CDS	25797	26075	279	#ERROR!			PGF_03828108		92		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.75			PATRIC.10738.3.107	PATRIC	CDS	26086	26292	207	#ERROR!					68		Phage baseplate protein (ACLAME 37)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.76			PATRIC.10738.3.107	PATRIC	CDS	26285	26572	288	#ERROR!					95		Phage hemolysin (ACLAME 315)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.77			PATRIC.10738.3.107	PATRIC	CDS	26569	26742	174	#ERROR!			PGF_03340292		57		Phage tail formation protein GpI	

Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.78			PATRIC.10738.3.107:	PATRIC	CDS	26742	27299	558	#ERROR!					185		Phage tail formation protein Gpl	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.79			PATRIC.10738.3.107:	PATRIC	CDS	27396	27860	465	#ERROR!					154		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.80			PATRIC.10738.3.107:	PATRIC	CDS	27881	28048	168	#ERROR!					55		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.6			PATRIC.10738.3.107:	PATRIC	CDS	2793	2945	153	#ERROR!			PGF_00550289		50		Phage transcriptional regulator (ACLAME 435)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.81			PATRIC.10738.3.107:	PATRIC	CDS	28142	28246	105	#ERROR!					34		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.82			PATRIC.10738.3.107:	PATRIC	CDS	28322	28630	309	#ERROR!					102		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.83			PATRIC.10738.3.107:	PATRIC	CDS	28609	29109	501	#ERROR!					166		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.84			PATRIC.10738.3.107:	PATRIC	CDS	29090	29683	594	#ERROR!					197		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.85			PATRIC.10738.3.107:	PATRIC	CDS	29686	29790	105	#ERROR!					34		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.86			PATRIC.10738.3.107:	PATRIC	CDS	29818	29943	126	#ERROR!					41		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.87			PATRIC.10738.3.107:	PATRIC	CDS	29944	30036	93	#ERROR!					30		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.88			PATRIC.10738.3.107:	PATRIC	CDS	30078	30254	177	#ERROR!					58		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.7			PATRIC.10738.3.107:	PATRIC	CDS	3013	3189	177	#ERROR!					58		Phage repressor protein CII	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.89			PATRIC.10738.3.107:	PATRIC	CDS	30277	30711	435	#ERROR!					144		Phage protein (ACLAME 343)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.90			PATRIC.10738.3.107:	PATRIC	CDS	30754	31035	282	#ERROR!					93		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.91			PATRIC.10738.3.107:	PATRIC	CDS	31040	31600	561	#ERROR!			PGF_10560543		186		Phage protein (ACLAME 283)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.92			PATRIC.10738.3.107:	PATRIC	CDS	31597	31695	99	#ERROR!					32		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.93			PATRIC.10738.3.107:	PATRIC	CDS	31779	31910	132	#ERROR!					43		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.94			PATRIC.10738.3.107:	PATRIC	CDS	31877	32254	378	#ERROR!					125		Phage protein (ACLAME 180)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.95			PATRIC.10738.3.107:	PATRIC	CDS	32251	32430	180	#ERROR!					59		Phage protein (ACLAME 180)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.96			PATRIC.10738.3.107:	PATRIC	CDS	32431	32532	102	#ERROR!					33		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.8			PATRIC.10738.3.107:	PATRIC	CDS	3252	3545	294	#ERROR!			PGF_00080423		97		Phage repressor protein CII	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.97			PATRIC.10738.3.107:	PATRIC	CDS	32532	32822	291	#ERROR!			PGF_04586255		96		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.98			PATRIC.10738.3.107:	PATRIC	CDS	32803	32898	96	#ERROR!					31		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.99			PATRIC.10738.3.107:	PATRIC	CDS	32947	33252	306	#ERROR!					101		Phage protein (ACLAME 180)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.100			PATRIC.10738.3.107:	PATRIC	CDS	33265	33516	252	#ERROR!					83		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.101			PATRIC.10738.3.107:	PATRIC	CDS	33627	33719	93	#ERROR!					30		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.102			PATRIC.10738.3.107:	PATRIC	CDS	33697	33819	123	#ERROR!					40		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.103			PATRIC.10738.3.107:	PATRIC	CDS	33819	33947	129	#ERROR!					43		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.9			PATRIC.10738.3.107:	PATRIC	CDS	3557	4015	459	#ERROR!			PGF_00205452		152		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.10			PATRIC.10738.3.107:	PATRIC	CDS	4077	4319	243	#ERROR!			PGF_00558614		80		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.11			PATRIC.10738.3.107:	PATRIC	CDS	4378	4545	168	#ERROR!					55		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.12			PATRIC.10738.3.107:	PATRIC	CDS	4609	4761	153	#ERROR!					50		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.13			PATRIC.10738.3.107:	PATRIC	CDS	4814	5056	243	#ERROR!					80		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.14			PATRIC.10738.3.107:	PATRIC	CDS	5083	5238	156	#ERROR!					51		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.15			PATRIC.10738.3.107:	PATRIC	CDS	5235	5333	99	#ERROR!					32		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.16			PATRIC.10738.3.107:	PATRIC	CDS	5395	5793	399	#ERROR!					132		Phage protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.17			PATRIC.10738.3.107:	PATRIC	CDS	5790	5891	102	#ERROR!					33		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.18			PATRIC.10738.3.107:	PATRIC	CDS	5931	6122	192	#ERROR!					63		Phage replication initiation protein (gene A) (ACLAME 154)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.19			PATRIC.10738.3.107:	PATRIC	CDS	6143	6268	126	#ERROR!					41		Phage replication initiation protein (gene A) (ACLAME 154)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.20			PATRIC.10738.3.107:	PATRIC	CDS	6287	6382	96	#ERROR!					31		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.21			PATRIC.10738.3.107:	PATRIC	CDS	6428	6523	96	#ERROR!					31		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.22			PATRIC.10738.3.107:	PATRIC	CDS	6520	6984	465	#ERROR!					154		Phage replication initiation protein (gene A) (ACLAME 154)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.23			PATRIC.10738.3.107:	PATRIC	CDS	7043	7294	252	#ERROR!			PGF_10357966		83		Phage replication protein A, endonuclease	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.24			PATRIC.10738.3.107:	PATRIC	CDS	7388	7642	255	#ERROR!					84		Phage replication initiation protein (gene A) (ACLAME 154)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.25			PATRIC.10738.3.107:	PATRIC	CDS	7740	7913	174	#ERROR!					57		Phage replication initiation protein (gene A) (ACLAME 154)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.26			PATRIC.10738.3.107:	PATRIC	CDS	7948	8172	225	#ERROR!					74		Phage replication initiation protein (gene A) (ACLAME 154)	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.27			PATRIC.10738.3.107:	PATRIC	CDS	8142	8318	177	-					58		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.28			PATRIC.10738.3.107:	PATRIC	CDS	8343	9053	711	-					236		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.1			PATRIC.10738.3.107:	PATRIC	CDS	88	747	660	-			PGF_01615525		219		Phage integrase	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.29			PATRIC.10738.3.107:	PATRIC	CDS	9120	9278	159	#ERROR!					52		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.30			PATRIC.10738.3.107:	PATRIC	CDS	9275	9607	333	#ERROR!					110		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.31			PATRIC.10738.3.107:	PATRIC	CDS	9632	9721	90	#ERROR!					29		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.32			PATRIC.10738.3.107:	PATRIC	CDS	9738	9893	156	#ERROR!					51		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.2			PATRIC.10738.3.107:	PATRIC	CDS	979	1110	132	-					43		hypothetical protein	
Circular_contigs_33.4Kb	10738.3	10738.3.con.0001	fig|10738.3.peg.33			PATRIC.10738.3.107:	PATRIC	CDS	9999	10106	108	#ERROR!					35		hypothetical protein	

Supplementary Table 10

Bacterial Isolate	completeness	contamination	strain_heteroge	length	N50	Assembly
BC01	83.82	33.05	8.62	8672352	4540248	Long-reads only
BC02	72.07	2.32	60	2624308	2603749	Long-reads only
BC03	99.45	4.19	18.18	3616041	3444714	Hybrid
BC04	75.1	44.36	0	6876170	95225	Long-reads only
BC05	91.28	13.64	1.24	6189984	4965937	Hybrid
BC06	97.41	153.12	0.89	15467599	1127517	Hybrid
BC07	98.65	0.28	0	6054259	5591092	Hybrid
BC08	80.98	0.99	0	5155983	5139781	Hybrid
BC09	99.03	0.55	0	3154845	3154845	Hybrid
BC10	100	1.46	0	3839305	409731	Long-reads only

Supplementary Table 11

user_genome	classification	fastani_r	fastani_r	fastani_t	fastani_a	fastani_a	closest_p	closest_p	closest_p	closest_p	closest_p	placer_t	classification_method
BC01.fasta	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Aeromonadaceae;g__Aeromonas;s__Aeromona	GCF_00082	95	d__Bacter	96.03	0.91	GCF_00082	95	d__Bacter	96.03	0.91	d__Bacter	taxonomic classification defined by topology and ANI
BC02.fasta	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus_I;s__Enterococcus_I aqui	GCF_00188	95	d__Bacter	97.73	0.87	GCF_00188	95	d__Bacter	97.73	0.87	d__Bacter	taxonomic classification defined by topology and ANI
BC03.fasta	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Carnobacterium;s__Carnobacterium ma	GCF_00074	95	d__Bacter	98.91	0.89	GCF_00074	95	d__Bacter	98.91	0.89	d__Bacter	taxonomic classification defined by topology and ANI
BC04.fasta	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pse	GCF_00104	95	d__Bacter	97.89	0.69	GCF_00104	95	d__Bacter	97.89	0.69	d__Bacter	taxonomic classification defined by topology and ANI
BC05.fasta	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas_E;s__Pse	GCF_00104	95	d__Bacter	98.59	0.89	GCF_00104	95	d__Bacter	98.59	0.89	d__Bacter	taxonomic classification defined by topology and ANI
BC06.fasta	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Vibrionaceae;g__Vibrio;s__Vibrio metsch	GCF_00017	95	d__Bacter	97.4	0.89	GCA_00712	95	d__Bacter	82.53	0.71	d__Bacter	ANI
BC07.fasta	d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_G;g__Bacillus_A;s__Bacillus_A thuringiensis_S	GCF_00002	96.4316	d__Bacter	98.94	0.89	GCF_00002	96.4316	d__Bacter	98.94	0.89	d__Bacter	taxonomic classification defined by topology and ANI
BC08.fasta	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Shewanellaceae;g__Shewanella;s__Shewane	GCF_90045	95	d__Bacter	95.52	0.86	GCF_90045	95	d__Bacter	95.52	0.86	d__Bacter	taxonomic classification defined by topology and ANI
BC09.fasta	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Vagococcaceae;g__Vagococcus_D;s__Vagococcus_D salmonar	GCF_00398	95	d__Bacter	97.3	0.88	GCF_00398	95	d__Bacter	97.3	0.88	d__Bacter	taxonomic classification defined by topology and ANI
BC10.fasta	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Leucobacter;s__Leucobac	GCF_00421	95	d__Bacter	97.88	0.89	GCF_00421	95	d__Bacter	97.88	0.89	d__Bacter	taxonomic classification defined by topology and ANI

Erklärung zur Dissertation
gemäß der Promotionsordnung vom 12. März 2020

Diese Erklärung muss in der Dissertation enthalten sein.
(*This version must be included in the doctoral thesis*)

„Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht."

Teilpublikationen:

1) Aleman FDD, Valenzano DR (2019) Microbiome evolution during host aging. PLoS Pathog 15(7): e1007727. https://doi.org/10.1371/journal.ppat.1007727

2) Davila-AlemanFD, Popkes M, Valenzano DR(2021). Characterization and genome assembly of the gut microbiota model organism Nothobranchius furzeri using short- and long-read sequencing. (In preparation, PeerJ)

Datum, Name und Unterschrift

30.11.2021 Francisco Daniel Davila Aleman