

Strain Level Diversity of *Bacteroides* Between Different Host Species

Master Thesis

Kevin Bisdorf

Matr. No. 389120

First Examiner: Dr. Martin Zimmermann

Second Examiner: Prof. Dr. Thomas Clavel

Supervisors: M.Sc Afrizal, Dr. Thomas C.A Hitch

submitted

Aachen, 07.08.2020

Table of contents

List of Figures	IV
List of Tables	V
Abbreviations	VI
Abstract	VIII
1. Introduction.....	9
1.1. Gut Microbiome.....	9
1.2. Cultivation-Dependent and -Independent Approaches	11
1.3. Host Specificity	12
1.4. Strain Level Diversity	13
1.5. Aim of the Project.....	15
2. Material and Methods	16
2.1. Media Preparation.....	16
2.2. <i>B. vulgatus</i> Isolates.....	17
2.3. Sampling.....	17
2.4. Isolation of Bacteria	19
2.5. Maintenance and Storage of Bacteria.....	20
2.6. Identification of Isolates	21
2.7. MALDI-TOF-MS.....	21
2.8. 16S rRNA Gene Sequencing	22
2.10. HPLC-measurement of SCFAs	25
2.11. Bile Salt Hydrolase (BSH)-Assay	26
2.12. Sequencing	27
2.13. Genome Assembly and Plasmid Analysis	28
2.14. Published Genome Datasets.....	29
2.15. Ecological Analysis.....	30
2.16. Pangenomic Analysis with anvi'o	30

2.17.	<i>In silico</i> Carbohydrate- and Antibiotic Resistance Profiling	31
2.18.	Phenotypical Prediction by using Traitair	32
3.	Results	34
3.1.	Cultivation	34
3.2.	Ecological Analysis	35
3.3.	Curation of published genome datasets.....	36
3.4.	Pangenomic Analysis by anvi'o	37
3.6.	Phenotype Prediction by Traitair and PhenDB	43
3.7.	Plasmid Analysis.....	46
3.8.	Antibiotic Resistance Prediction.....	49
3.9.	BSH-Assay	50
3.10.	Production of SCFAs as measured by HPLC.....	51
3.11.	Carbon Source Utilization Analysis with the BIOLOG System	53
3.12.	Comparison between <i>in silico</i> Predicted and <i>in vitro</i> tested Phenotypes....	55
4.	Discussion	56
5.	Conclusion & Outlook	62
	Acknowledgement	63
	References	64
	Appendix 1.....	76
	Appendix 2.....	78
	Appendix 3.....	81

List of Figures

Figure 1: Exemplary workflow of the bacterial isolation	19
Figure 2: Spreading technique to ensure the generation of a pure culture	20
Figure 3: Exemplary pictures of the evaluation of AN MicroPlate™ BIOLOG tests	25
Figure 4: Exemplary evaluation of the BSH-activity assay	27
Figure 5: Filtering of the datasets of Poyet <i>et al.</i> (2019) and Forster <i>et al.</i> (2019)	29
Figure 6: Traitair workflow	32
Figure 7: Workflow of the PhenDB pipeline	33
Figure 8: <i>Bacteroides</i> species isolated from gut samples of the different host species	35
Figure 9: Occurrence of <i>B. vulgatus</i> in ten different types of samples	35
Figure 10: Distribution of isolates from the published datasets to the type strains of <i>B. vulgatus</i> and <i>B. dorei</i> based on ANI	36
Figure 11: Influence of plasmids within genome assemblies on Average-Nucleotide-Identities	37
Figure 12: Pangenome tree of 182 isolate genomes belonging to the <i>Bacteroides</i> genus	39
Figure 13: Number of CAZyme families and CAZymes within isolates from this study	41
Figure 14: MDS Jaccard dissimilarity of all <i>B. vulgatus</i> isolates (n=177) based on CAZyme families ..	42
Figure 15 MDS Bray-Curtis dissimilarity of all <i>B. vulgatus</i> isolates (n=177) based on CAZyme families	43
Figure 16: Predicted phenotypes by Traitair and PhenDB	45
Figure 17: SCFA profiles of the 19 isolates	52

List of Tables

Table 1: <i>B. vulgatus</i> isolates used in this study.....	17
Table 2: Standard primers for PCR reactions of the amplification of the 16S rRNA gene.....	23
Table 3: Composition of PCR reactions used for the amplification of 16S rRNA genes.....	23
Table 4: PCR Programm used for the amplification of 16S rRNA genes.....	23
Table 5: Numbers of isolates picked, identified, not yet identified, and lost during cultivation.	34
Table 6: Overview of the identified clusters and their content.....	40
Table 7: Plasmids in the genome of 19 own isolates	47
Table 8: Comparison of the gene content encoded on the plasmid for each isolate (n=19).....	48
Table 9: Genome-based antibiotic resistance potential of the 19 isolates from this study.	50
Table 10: Overview of the presence or absence of a BSH-enzyme in the 19 isolates	51
Table 11: Summary of strain-level differences in carbon source utilization	54
Table 12: Summary of the <i>in silico</i> predicted phenotypes compared to <i>in vitro</i> results	55

Abbreviations

ABC	ATP binding cassette
ANI	Average-Nucleotide Identity
AN-IF	AN Inoculating Fluid
ARG	Antibiotic-Resistance Genes
ATCC	American Type Culture Collection
bactNOG	Bacteria non-supervised orthologous groups
BBE	Bacteroides Bile Esculin
BHI	Brain Heart Infusion
BIO-ML	Broad Institute-OpenBiome Microbiome Library
BSH	Bile Salt Hydrolation
BUA	Biolog Universal Anaerobe
CAZyme	Carbohydrate-active Enzyme
CBM	Carbohydrate-Binding Module
CD	Cluster of Differentiation
CE	Carbohydrate Esterase
CONV-D	conventionalized
dDDh	digital-DNA-DNA Hybridization
DNA	Deoxyribonucleic Acid
DTT	1,4-Dithiothreitol
GAM (mod)	Gifu Anaerobic Media (modified)
GF	Germfree
GGDC	Genome-to-Genome Distance Calculator
GH	Glycoside Hydrolases
GIDEON	Global Infectious Disease and Epidemiology Online Network
GT	Glycosyl Transferases
HBC	Human Gastrointestinal Bacteria Culture Collection
HCCA	α -Cyano-4-hydroxycinnamic acid
HMb	Human Microbiota
HMM	Hidden Markov Model
HMP	Human Microbiome Project
HPLC	High-Pressure Liquid Chromatography
IMNGS	Integrated Microbial NGS platform
ITEP	Integrated Toolkit for Exploration
LS _A	Lincosamide, Streptogramin A
LS _{AP}	Lincosamide, Streptogramin A , Pleuromutilin
MALDI-TOF MS	Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry
MDS	Multi-Dimensional

MEGA	Molecular Evolutionary Genetics Analysis
MLS _B	Macrolide, Lincosamide, Streptogramin B
MMb	Mouse Microbiota
MS _B	Macrolide, Streptogramnin B
NGS	Next Generation Sequencing
NI	Not Identified
ORF	Open Reading Frames
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PL	Polysaccharide Lyase
Pfam	Protein family
QBiC	Quantitative Biology Center
rRna	ribosomale Ribonucleic Acid
SD	Standard Deviation
SFB	Segmented Filamentous Bacteria
TDCA	Taurodeoxycholic Acid
UV	Ultraviolet
WCA	Wilkins Chalgren Anaerobic
YCFA	Yeast extract, Casitone and Fatty Acid

Abstract

The human gut microbiome is dominated by highly diverse communities of bacteria that can degrade food components and influence the metabolism and health of their host. In this study, a combination of bioinformatic approaches and experimental techniques were used to assess the potential strain-level genomic and phenotypic variations *Bacteroides vulgatus* strains isolated across different mammalian hosts.

A pangenomic analysis demonstrated that *B. vulgatus* is not a monotypic species and allowed the identification of 21 clusters of genes which were specific for certain strains across 4 distinct phylogenetic clades.

The analysis of the plasmid repertoire identified 38 plasmids with a range of zero to seven per isolate. The investigation of the plasmid content revealed an antibiotic-resistance-associated protein leading us to further investigate the antibiotic resistance. A group of six isolates did not contain any antibiotic resistance-conferring proteins whereas another four isolates were predicted to have antibiotic resistance against a broad range of antibiotics.

As bile acids are known to be important signalling molecules for microbe-host interactions hence the bile salt hydrolysis activity was investigated. Strain-level variations were observed across the 19 tested isolates, of which 13 were positive and six were negative.

Further metabolic functions investigated by means of Biolog assays led to a screen of potential strain-level differences in carbon source utilization which can be further tested by using a minimal medium.

The comparative analysis of the *in silico* predicted and the confirmed *in vitro* phenotypes revealed limitations of the *in silico* prediction regarding BSH activity and sugar utilizations. This highlighted the importance of *in vitro* testing of isolates to complement genome-based analyses.

Overall, this study provides initial insights into strain-level variability within the important gut bacterial species *Bacteroides vulgatus* as well as the potential of combining bioinformatic and experimental techniques to determine phenotypic differences. Of particular interest would it be to confirm the predicted antibiotic-resistances via *in vitro* testing.

1. Introduction

1.1. Gut Microbiome

The term microbiome was re-visited and defined by Berg *et al.* (2020) as: ‘a characteristic microbial community occupying a reasonable well-defined habitat which has distinct physio-chemical properties. The microbiome not only refers to the microorganisms involved but also encompass their theatre of activity, which results in the formation of specific ecological niches. The microbiome, which forms a dynamic and interactive micro-ecosystem prone to change in time and scale, is integrated in macro-ecosystems including eukaryotic hosts, and here crucial for their functioning and health’.

Within the mammalian gastrointestinal tract, bacteria account for the majority of cells and diversity within the microbiome. According to Sender *et al.* (2016), an average 70 kg human male adult host carries 3.8×10^{13} bacterial cells which exceed the total number of human body cells (3.0×10^{13}).

Furthermore, the human gut microbiome has been shown to play an essential role in host metabolism (Rowland *et al.*, 2018; Visconti *et al.*, 2019), immunity (Iida *et al.*, 2013; Kelly *et al.*, 2005), drug response (Visconti *et al.*, 2019) and protection against pathogens (Ubeda *et al.*, 2017).

However, as the composition, both taxonomically and functionally, of the microbiome is influenced by external factors such as age, diet, geographic location, genetics and medical treatment, these can have an indirect effect on the mentioned host features via the microbiome (Hasan and Yang, 2019).

In adult humans, the dominant bacterial phyla within the gut are Firmicutes, Bacteroidetes, Actinobacteria, Fusobacteria, Verrucomicrobia and Proteobacteria (Arumugam *et al.*, 2011; Bäckhed *et al.*, 2012; Eckburg *et al.*, 2005; Segata *et al.*, 2012; Zoetendal *et al.*, 2008). The genus *Bacteroides* which belongs to the phylum Bacteroidetes consists of Gram-negative, obligate anaerobic bacteria (Whitman, 2011). Jacobson *et al.* (2018) showed that *Bacteroides* spp. contribute to colonization

resistance against *Salmonella Typhimurium* via propionate production, which can diffuse into the cytoplasm of Gram-negative bacteria leading to cellular acidification.

Bacteroides spp. are known to utilize simple sugars and polysaccharides for growth in the large intestine (Hooper *et al.*, 2002). For instance, *Bacteroides thetaiotaomicron* is known to degrade complex polysaccharides into monosaccharides through the use of SusC and SusD, which bind and transport polysaccharides into the cell, and glycosylhydrolases which degrade the polysaccharides into monosaccharides (Xu *et al.*, 2003). The released monosaccharides can then be used by bacterial species unable to degrade complex polysaccharides, such as *E. coli* (Hoskins *et al.*, 1985). Many species among the phyla present within the gut microbiome can interact with each other in order to create complex functional pathways for degradation of food components (Rubino *et al.*, 2017).

Bacteroides spp. commonly carry a variety of antibiotic-resistance genes (ARGs) (Sóki *et al.*, 2006). For instance, strains of *Bacteroides fragilis* were shown to have the potential to be resistant to imipenem, piperacillin-tazobactam, ampicillin-sulbactam and metronidazole (Pumbwe *et al.*, 2007; Schapiro *et al.*, 2004; Wareham *et al.*, 2005). *Bacteroides vulgatus* is another member of the *Bacteroides fragilis* group and showed to have a greater resistance to many antibiotics than *B. fragilis* (Hedberg and Nord, 2003). The antimicrobial resistance of 101 *B. vulgatus* isolates was examined nearly twenty years ago: 97 % of these isolates were resistant to ampicillin, 30 % to clindamycin, 19 % to moxifloxacin, 9 % to cefoxitin, 3 % to piperacillin/tazobactam, 2 % to metronidazole and 1 % to imipenem (Hedberg and Nord, 2003). Due to aminoglycosides requiring an oxygen- or nitrate-dependent electron transport chain to act upon, *Bacteroides* spp. are inherently resistant to aminoglycosides as they lack this pathway (Bryan *et al.*, 1979).

Nowadays, microbiome research and in particular studies on specific gut members such as *Bacteroides* has been revolutionized by high-throughput sequencing technology, permitting compositional and functional analyses that were previously an unrealistic undertaking (Arnold *et al.*, 2016).

1.2. Cultivation-Dependent and -Independent Approaches

Historically, culture-dependent methods dominated the field of microbiology and together with microscopy, have been long the only approach for studying microbiomes (Klieve and Bauchop, 1988). However, in recent years, the number of culture-independent metagenomic studies via sequencing has massively increased (Almeida *et al.*, 2019; Pasolli *et al.*, 2019). This has been driven by exponential decreasing sequencing costs via the rapid development of Next Generation Sequencing (NGS) technologies (Church, 2006).

Culture-independent-based approaches, such as shotgun metagenome sequencing, have enabled the investigation of both the taxonomic and functional composition of complex samples, facilitating the identification of yet unknown and so far uncultured bacterial species (Almeida *et al.*, 2019). These methods can also facilitate the isolation of uncultured species, as shown in Pope *et al.* (2011) where a metabolic model for a novel species was generated from metagenomic data. This model directed the design of a media-specific for the novel species, leading to its isolation.

However, the analysis of metagenomic datasets still relies heavily on high-quality reference databases (Quince *et al.*, 2017). Large-scale projects like Metagenomics of the Human Intestinal Tract (MetaHIT) or the Human Microbiome Project (HMP) provided such reference databases to help understand the functional interactions of microbes and the human host (Human Microbiome Project Consortium, 2012; Huttenhower *et al.*, 2012; Qin *et al.*, 2010). Therefore, culture-dependent approaches are still important and enable access to bacterial isolates that are sorely needed for *in vitro* experimental characterization and validation (Forster *et al.*, 2019; Poyet *et al.*, 2019).

Both culture-dependent and culture-independent approaches, have advantages but also limitations. This has led to their tandem use as mutually complementary methods which are necessary for further investigations of the human gut microbiome (Lagkouvardos *et al.*, 2019).

1.3. Host Specificity

The evolution of mammals and their gut microbes has become a topic of increasing interest in recent years (Ley *et al.*, 2008a; Shapira, 2016), for instance to study how specific bacterial species have evolved to colonise different host species, termed host specialisation (Frese *et al.*, 2011; Powell *et al.*, 2016). Although Huttenhower *et al.* (2012) showed that gut microbiomes can differ in diversity across adults, a core group of microbial species was identified for the human colon (Sekelja *et al.*, 2011), most of which belonged to *Clostridium* cluster XIVa (Cotta and Forster, 2006).

Geographic location and diet have been shown to drive interindividual variation in the microbiota composition (Ley *et al.*, 2008b; Tett *et al.*, 2019; Yatsunenko *et al.*, 2012). Tett *et al.* (2019) observed that *Prevotella copri* was more prevalent in populations with non-Westernized lifestyles including diets consisting of more fibre and complex carbohydrates and lower amounts of fat and animal proteins (Filippo *et al.*, 2010; Segata, 2015; Statovci *et al.*, 2017). They therefore proposed that *P. copri* has evolved in response to westernisation and the shift in both diet and overall lifestyle this encompasses.

The impact of a host-specific microbiota on the development of a functioning immune system was studied by Chung *et al.* (2012) in a small scale study in mice. Germfree (GF) mice were either colonized with mouse microbiota (MMb) or human microbiota (HMb) to determine whether small intestinal immune maturation depends on a coevolved host-specific microbiota. The HMb mouse intestines showed all characteristics of GF mice (e.g. low levels of CD4⁺ and CD8⁺ T cells and low antimicrobial peptide expression) despite gut bacterial numbers and phylum relative abundance being similar to those in the MMb group. Colonizing the GF mice or HMb mice with segmented filamentous bacteria (SFB) isolated from mouse led to a partial restoration of the T cell numbers suggesting that those SFB or MMb bacteria are crucial for the development of a healthy immune system.

Studies on *Lactobacillus reuteri* revealed that rodent-specific strains were able to form a biofilm in the mouse stomach whereas strains originating from other host species (human, pig, chicken) failed to do so. Frese *et al.* (2011) identified several genomic features which contributed to host specialization of *L. reuteri* within rodents. The

presence of surface proteins which are predicted to function as adhesins or mediators of biofilm formation, the SecA2 system likely involved in the secretion of these proteins and a urease cluster containing genes for urease production were major factors that play roles in biofilm formation and the mitigation of low pH. With these findings, Frese *et al.* (2011) confirmed that *L. reuteri* strains underwent an evolutionary process which has led to host specialization within rodents.

Cohousing gnotobiotic mice with various selected xenomicrobiota from human, zebrafish, and termite guts, human skin and tongue, soil, and estuarine microbial mats together with control “conventionalized” animals revealed that most bacterial phylotypes, including those selected from human gut microbiota, are not capable of realizing a niche in a gut harbouring an autochthonous microbiota (Seedorf *et al.*, 2014).

1.4. Strain Level Diversity

Whilst strain level variability of microbial species has been investigated between different hosts, recent studies have revealed a remarkable microbial strain variation between individuals of the same host species (Filippis *et al.*, 2019; Kujawska *et al.*, 2020; Sorbara *et al.*, 2020).

Nemergut *et al.* (2013) suggested that selective pressure by encountering other microorganisms, the host immune systems or variations in the diet requires the bacteria to adapt and drives strain-level variations. Zhao *et al.* (2019) identified parallel evolution in sixteen genes in intra-individual *B. fragilis* populations. These included genes involved in cell-envelope biosynthesis and polysaccharide utilization. Strains of *Bacteroides ovatus* have been shown to induce immunoglobulin A (IgA) production at varying levels, significantly altering the hosts IgA production levels when administered to mice (Yang *et al.*, 2020).

Tett *et al.* (2019) identified four distinct clades within the species *P. copri*, termed the *P. copri* complex. These clades are nearly ubiquitous and co-present in non-Westernized populations. The potential of carbohydrate utilization was significantly different between the four clades. The pectin-degrading carbohydrate-active enzyme (CAZyme)-families PL9 and PL10 were highly prevalent and nearly exclusive to one of

the four clades. Another clade had the GH9, a CAZy family of cellulases, enriched compared to the three other clades. The clades were also different in antimicrobial resistance gene content or their membrane transporter repertoire.

In the context of infectious disease clinical practice, the investigation of strain-level functional differences of pathogenic bacteria is important and already widespread in the field. Hall *et al.* (2017) identified two different clades of strains within the species *Ruminococcus gnavus*, one of which was specific for inflammatory bowel disease (IBD)-patients. This clade was identified to contain 199 IBD-specific genes which promoted colonization of the IBD gut, including genes involved in oxidative stress responses, adhesion, iron acquisition and mucus utilization.

A uniform manifold approximation and projection analysis (UMAP) on the presence or absence of annotated genes and protein clusters within all *Lachnospiraceae* revealed inter- and intra-species diversity (Sorbara *et al.*, 2020). The examination of the intra-species genomic diversity was performed by individual UMAP analysis on species with more than 10 representatives using altered UMAP parameters and showed that species including *Blautia luti*, *Dorea longicatena*, and *Fusicatenibacter saccharivorans* only formed one cluster whilst other species including *Blautia wexlerae*, *Ruminococcus gnavus*, and *Anaerostipes hadrus*, formed 2 or 3 distinct clusters of strains. For the investigation of the gene content which separate isolates of a species into distinct clusters, the core genomes of the distinct clusters were investigated. This revealed differences in their capacity for carbohydrate utilization, their antimicrobial production and defence and in the case of *R. gnavus* the ability to acidify rich culture media.

The aforementioned studies show that there can be substantial differences between strains within a species, e.g. in the use of carbon sources or antimicrobial defence (Sorbara *et al.*, 2020). However, these studies also indicate that many of these differences have not been adequately explored and there may be many more differences. In particular, comprehensive investigations of the congruence between genomic strain delineation (e.g. based on single nucleotide polymorphisms) and functional consequences for the bacteria are lacking. This applies to a large number of species that have not yet been considered under the focus of intra-species variations (Sorbara *et al.*, 2020).

1.5. Aim of the Project

This project aimed to study the inter- and intra-host diversity of *B. vulgatus*, both genome-wise and phenotypically. For this, the project was divided into two separate parts.

The first part included the isolation and cultivation of *B. vulgatus* strains from human, pig and mouse stool samples. Phenotypic testing of these isolates using the BIOLOG system, high-pressure liquid chromatography (HPLC)-short-chain fatty acid (SCFA) profiling and bile salt hydrolase (BSH) activity.

The second part studied genomic differences between the strains isolated in part one and others from the literature (Forster *et al.*, 2019; Poyet *et al.*, 2019). Anvi'o allowed pangenomic analysis to identify potential strain-level differences. Furthermore, CAZyme profiles, antibiotic resistances, plasmid repertoire, predicted growth phenotypes and bile salt resistance were studied.

2. Material and Methods

2.1. Media Preparation

Bacteria were cultured under anaerobic conditions since *Bacteroides* spp. are obligate anaerobes. A fibre medium with gentamicin and a modified version of Bacteroides Bile Esculine medium (BBE, Otto-Nordwald GmbH, Germany) (Becton, Dickinson and Company, 2003) were used for isolation from stool samples. In addition, rich media (e.g. BHI, WCA and GAMmod) were also used to maintain the isolates.

The recipe for all media is listed in Appendix 1. Plate-based cultivation in the anaerobic chamber and the Hungate technique were employed (Hungate, 1950; Hungate *et al.*, 1966). The transfer of isolates from agar plates into Hungate tubes containing liquid media was convenient to generate enough cell biomass for further analyses. The liquid medium for Hungate tubes was prepared in sterile Schott bottles as follows: All ingredients, except L-cysteine and 1,4-dithiothreitol (DTT), were dissolved in deionized water and heated to 100°C in a microwave to remove soluble oxygen. 1 ml of the redox potential indicator resazurin (1 mg/ml) was added before heating. L-cysteine and DTT, dissolved in deionized water were added after heating. Each Hungate tube was filled with 9 ml boiled medium and then gassed for 3 min with an oxygen-free gas mixture (89.3% N₂, 6% CO₂, 4.7% H₂). The tubes were then sealed with a rubber stopper and a screw cap and autoclaved at 121°C for 20 min. They were stored at room temperature when not in use. For the preparation of solid media, 1.5% (w/v) agar was added. Heat-sensitive ingredients such as L-cysteine and DTT were filter-sterilized and added to the medium (cooled down to 60°C) prior to pouring into Petri dishes. Agar plates were placed into the anaerobic chamber (MBRAUN, Germany) (89.3% N₂, 6% CO₂, 4.7% H₂) for at least 24h before usage.

2.2. *B. vulgatus* Isolates

Table 1: *B. vulgatus* isolates used in this study.

No.	Strain designation	Study acronym	Origin	Host species
1*	TC-KB-H48.1	H-AC-1	Aachen	Human
2*	TC-KB-H118	H-AC-2	Aachen	Human
3*	TC-KB-H143	H-AC-3	Aachen	Human
4**	TC-JM-F11	H-AC-4	Aachen	Human
5**	4716a	H-BS-1	Braunschweig	Human
6**	4852a	H-BS-2	Braunschweig	Human
7**	4889a	H-BS-3	Braunschweig	Human
8**	4896a	H-BS-4	Braunschweig	Human
9**	4835b	H-BS-5	Braunschweig	Human
10**	4889b	H-BS-6	Braunschweig	Human
11**	d3	H-BS-7	Braunschweig	Human
12**	HDF	H-BS-8	Braunschweig	Human
13**	H-iso	H-VI-1	Vienna	Human
14*	TC-KB-M13	M-AC-1	Aachen	Mouse
15*	TC-KB-M20	M-AC-2	Aachen	Mouse
16*	TC-KB-M259	M-AC-3	Aachen	Mouse
17**	JM39a	M-FR-1	Freising (miBC)	Mouse
18**	M-iso	M-VI-1	Vienna	Mouse
19*	TC-KB-P90	P-AC-1	Aachen	Pig
20**	PiBac	P-FR-1	Freising (PiBAC)	Pig

*: strains isolated during the work; **: strains already available or isolated by others

2.3. Sampling

Human fecal material was collected after obtaining consent from the Ethics committee at the RWTH University Hospital (file no. EK 010/20). Samples from animals were collected either (i) for scientific purpose only under §4 and §7 of the Animal Welfare Act (Tierschutzgesetz), or (ii) in the context of otherwise ethically approved experiments performed by others (e.g. LANUV approval no. 84-02.04.2017.A296 and 84-02-04.2017.A307). Animals were thus not culled for the sole purpose of the present work and they did not undergo any specific treatment and were hosted at the Institute of Animal Science of the RWTH University Hospital under ethically approved conditions.

Human

Fresh faecal samples were collected in sterile buckets (VWR, USA). Under the laminar flow cabinet, the faecal material was taken with a sterile inoculation loop and transferred into a falcon tube. The transferred faecal material was diluted with 10 ml sterile anaerobic Phosphate-buffered saline (PBS) in the anaerobic chamber. This bacterial suspension was the starting point of isolation.

Pig

Eight sections of the gut from different pigs were received from the Institute of Animal Science (RWTH University Hospital) in sterile buckets (VWR, USA) which had an oxygen scavenger on the top to reduce the oxygen contamination level and were labelled with numbers from one to eight. The faecal content was extracted from the surrounding tissue using sterile dissection tools in the laminar flow cabinet. The extracted faecal content was then transferred into a falcon tube. The rest of the procedure was equivalent to sampling in humans.

Mice

Apart from frozen samples that were already available before project start, caecal content of mice were sampled shortly after the mouse was culled by neck dislocation. For the latter case, the carcass of the mice was pinned onto a styrofoam plate which was encased with the first layer of aluminium foil and a second layer of a plastic film. Before dissecting, the mouse body was washed with 70 % ethanol to reduce the risk of contamination with bacteria from the skin and environment. Using sterile tools, the caecum was extracted and the content was transferred into an Eppendorf tube in a laminar flow cabinet by squeezing the caecum using the lid of the tube. A sterile inoculation loop was used to either transfer the caecal content from the Eppendorf tube to a Hungate tube or sequencing cryo-tubes. The Hungate tube was prepared with 4 ml sterile, anaerobic PBS supplemented with DTT (0.2 g/l) and L-Cysteine (0.5 g/l) and containing glass beads to aid homogenisation. 2 ml of the mixture was sterilely transferred into another Hungate tube containing 8 ml sterile anaerobic PBS supplemented with L-Cysteine and DTT. The bacterial suspension in the latter Hungate tube was then used as a starting point for isolation. For the generation of frozen samples, 2 ml of the pre-dilution Hungate were taken and mixed in a ratio of 1:1 with

sterile, anaerobic PBS containing 40 % glycerol to create a frozen culture (4x) filled with 1 ml each. Those were placed on dry ice and later stored at -80°C.

2.4. Isolation of Bacteria

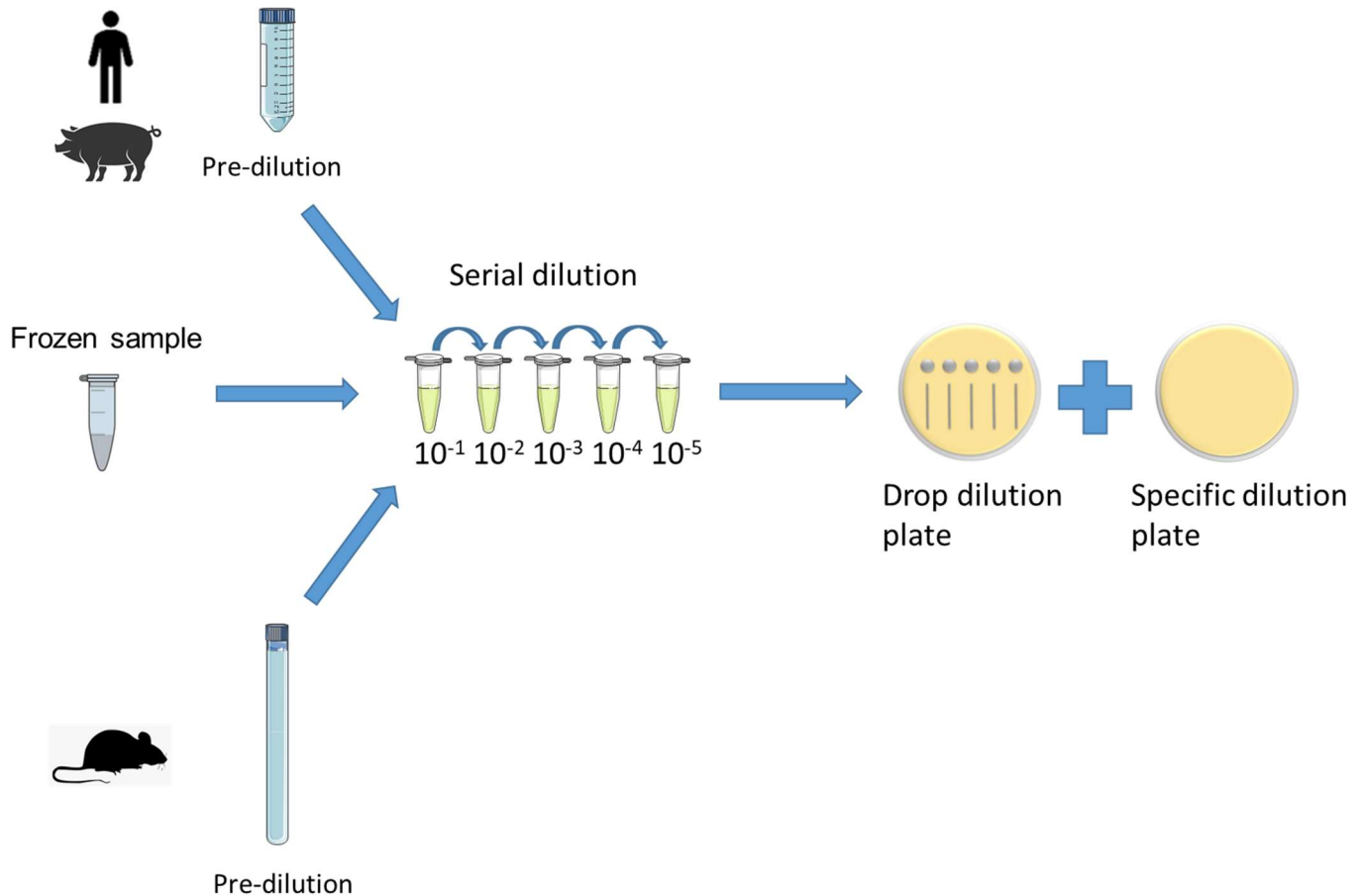


Figure 1: Exemplary workflow of the bacterial isolation

An overview of the workflow for bacterial isolation is shown in Figure 1. Serial dilutions (10^{-1} – 10^{-5}) were prepared either from the pre-dilutions (falcon tubes for humans and pigs or Hungate tubes for mice) containing faecal slurries or from a frozen sample stored in PBS containing 20% glycerol (prepared by another member of the lab before project start). From each dilution, a drop of 10 μ l was pipetted onto plates (BBE and Fibre medium) that were then tilted to spread the bacteria. On a second and third plate of each medium, 100 μ l of the dilutions 10^{-1} and 10^{-3} were spread with an L-Spatula according to the “spread-plate-method” (Sanders, 2012). All steps were performed in an MBraun anaerobic chamber (89.3% N₂, 6% CO₂ and 4.7% H₂) and the plates were left for growth at 37°C in an incubator inside the chamber.

Obtaining pure culture is a requirement for detailed investigation of microorganisms. In order to obtain pure cultures from complex samples, the colony picking technique was used which is based on the assumption that a bacterial colony arise from a single cell. Single colonies were picked and then re-streaked three times sequentially (after incubation times in between) onto a fresh plate as shown in Figure 2. Single colonies were picked based on their morphology and their clear distances to other colonies and plated either on rich agar medium (BHI, modified GAM or WCA) or BBE agar.



Figure 2: Spreading technique to ensure the generation of a pure culture

2.5. Maintenance and Storage of Bacteria

The Hungate technique was used at later stages after the pure isolate was obtained for more convenient work under a normal flow cabinet. The lid of the Hungates tubes was always flamed twice using absolute ethanol prior to aliquoting or sample transfer. Bacteria were subcultured in either BHI or modified GAM medium once before frozen stocks were generated.

Gram staining allowed visual testing for contamination before frozen stocks were prepared. Therefore, 1 ml of culture was centrifuged at 6,500 g for 5 minutes. The supernatant was removed and the pellet was resuspended in sterile PBS with a volume that depended on growth density. 10 μ l of the diluted culture was transferred onto a microscope slide. After the suspension dried under the laminar flow, the microscope slide was prepared for staining. Under the fume hood, the bacteria were fixed on the microscope slide by heat treatment. The fixed bacteria were then incubated with crystal violet. When the crystal violet was removed through water, the microscope slide was incubated with iodine for 1 minute. The iodine was removed by a short cleaning step with acetone followed by cleaning again with water. The final step was the incubation

of the fixed bacteria with carbolfuchsin for 2 minutes. The stained bacteria were visualized with the Axio Lab A1 (Zeiss, Germany) microscope using the 100x objective. Frozen stocks were prepared for long-term storage by mixing the cell suspension from Hungate tubes in a ratio of 1:1 with anaerobic medium containing 40 % glycerol. This happened under sterile, yet aerobic conditions under the laminar flow cabinet. 650 µl of the mixture was then rapidly aliquoted in a screw cap tube and placed on dry ice and then stored at -80°C.

2.6. Identification of Isolates

The isolates were first identified using Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) and then 16S rRNA gene sequencing for final confirmation, as MALDI can not differentiate between certain related bacteria such as *B. vulgatus* and *Bacteroides dorei* (Justesen *et al.*, 2011).

If the identified bacteria (other than *Bacteroides* spp.) were not yet included in our in-house bacterial collections they were stored as cryo-aliquots as described in section 2.5 “Cultivation and storage of bacteria”.

2.7. MALDI-TOF-MS

Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) allows rapid identification of bacteria and yeast. Bacteria identification using MALDI works by comparing the protein profile of the target organism with profiles of known bacteria stored in the reference database. The identification process of MALDI is split into three major steps. The identification process starts from subjecting the sample/matrix mix to a pulsed laser, resulting in the release of molecules as hot gas. This step is also called ablation or desorption. The produced ions are accelerated in an electric field and their time of flight is measured. The mass-to-charge ratio is calculated by their time of flight which also allows the generation of a specific mass spectrum (Karas *et al.*, 1987).

The generated mass spectrum is unique, thus it can be compared to the profiles in the database to identify the target organism (Wieser *et al.*, 2012). In this study, bacteria grown on agar plates were identified by directly applying cell material from a single colony onto the spots of the MALDI-target with sterile inoculation loops. Before measurement, the dried samples on the target plate were overlaid with 1 µl of the α-

Cyano-4-hydroxycinnamic acid (HCCA) matrix solution and allowed to dry under the laminar flow cabinet. The matrix solution was prepared by dissolving 2.5 mg HCCA in 0.25 ml solvent solution containing 47.5% H₂O, 50.0% acetonitrile and 2.5% trifluoroacetic acid. The MALDI Biotyper Microflex LT (Bruker, Germany) was used for measurement.

2.8. 16S rRNA Gene Sequencing

16S rRNA genes were sequenced to identify those bacteria that could not be identified by MALDI and to confirm the identity of those within the target species *B. vulgatus*. Two different approaches were used for DNA Isolation. The first approach was done with the Mericon DNA Bacteria Plus Kit (Qiagen, Netherlands) as per the manufacturer's instruction. The second DNA isolation method was done with a modified version of the protocol by Godon *et al.* (1997). 2 ml to 4 ml (depending on the turbidity) of liquid culture were used as starting material for the isolation. A bacterial pellet was obtained by centrifuging at 12,000 g for 10 minutes. It was then re-suspended in 600 µl Stool DNA Stabilizer (STRATEC, Germany). The suspension was transferred into screw cap tubes with 500 mg zirconium glass-beads (0.1 mm diameter; Carl Roth, Germany). 250 µl of 4M guanidinium thiocyanate and 500 µl 5% N-laurolysarcosine were added. Afterwards, the suspension was mixed by vortexing and then incubated at 70°C for 60 minutes. Next, the samples were subjected to mechanical disruption using a bead-beater FastPrep-24 5G (MP Biomedicals, USA) for three times with each cycle of 40 seconds and at a speed of 6.6 m/s. Dry ice was added into the cooling adaptor to prevent overheating of the samples between each step. After cell lysis, 15 mg polyvinylpolypyrrolidone was added to the samples followed by vortexing. Next, samples were centrifuged (15,000 g, 4°C, 3 min). The clear supernatants (500 µl) were transferred into new tubes and 5 µl RNase (10 mg/ml) was added prior to incubation at 37°C for 20-30 minutes with shaking (700 rpm). The isolated DNA was then purified using NucleoSpin® gDNA Clean-up kit (MacheryNagel, Germany) according to manufacturer's instruction. After DNA isolation, 16S RNA genes were amplified using Polymerase Chain Reaction (PCR). The primers (Baker *et al.*, 2003; Frank *et al.*, 2007) and reaction mixture are shown in Table 2 and Table 3, respectively.

Table 2: Standard primers for PCR reactions of the amplification of the 16S rRNA gene.

Primer description	Sequence
27F	5'-AGAGTTTGATCCTGGCTCAG-3'
1492r	5'-GGTTACCTTGTTACGACTT-3'

Table 3: Composition of PCR reactions used for the amplification of 16S rRNA genes.

Ingredients	Amount [μl]
Greentaq MM	20
H₂O	13
Primer 1 (1492R)	1
Primer 2 (27F)	1
+ DNA-Sample	5
Total Volume	40

DNA was amplified in a Biometra TAdvanced thermal cycler (Analytic Jena, Germany). The temperature and duration of the cycles are listed in Table 4.

Table 4: PCR Programm used for the amplification of 16S rRNA genes.

Step	Temp [°C]	Time	Cycles
Hot Start	95	3 min	
Denaturation	95	20 sec	30
Annealing	55	30 sec	
Elongation	72	90 sec	
Final Elongation	72	2 min	
Holding	12	∞	

Gel electrophoresis was done after the PCR step to confirm the presence of 16S rRNA gene amplicons. The agarose gel contained 1X TAE Buffer, 1,5 % (w/v) agarose and 16.7 μl/ml GelRed® Nucleic Acid Gel Stain (Biotium, USA). The latter was added to visualize the DNA under UV light. The gel was loaded with 4 μl of each sample and also with 4 μl of a standard ladder which served as a marker. The electrophoresis was carried out for 20 minutes at 110 V. The gel was visualized under UV light using the Gel Doc System (GelStudio SA, Analytic Jena, Germany). A band of around 1,5 kbs was visible if the amplification of the 16S rRNA gene was successful. The successfully amplified fragments were further cleaned with the MSB® Spin PCRapace kit (Stratec,

Germany) according to the manufacturer's instructions. The final products were sent to Eurofins Genomics for Sanger sequencing using primer 27f (see sequence above). The sequencing results were processed (e.g. manually truncated and aligned) with the Molecular Evolutionary Genetics Analysis (MEGA) software (Kumar *et al.*, 2018). Isolates were identified using the EzBioCloud database (Yoon *et al.*, 2017). For novel bacteria or those that were not yet in our strain collections, the same amplicons were further sequenced using primer 338r, 785f and 1492r to obtain nearly full-length 16S rRNA gene sequences.

2.9. BIOLOG Assays

BIOLOG (Hayward, CA) is a system to test for metabolic properties of strains. It consists of preconfigured sets of phenotypic tests in 96-well plates. Each well is designed to test one single substrate. Those specifically designed 96-well plates can be used for the identification of aerobic and anaerobic bacteria as well as yeasts and filamentous fungi. Furthermore, they can be used for Microbial Community Analysis and Phenotype MicroArray Analysis. In this study, the AN MicroPlate™ (BIOLOG Inc., 2007; Hayward, CA) was used to identify potential carbon sources. In each well, the metabolism of cells can be monitored using respiration as a reporter system. If the cells are metabolically active, there is a flow of electrons to NADH, which determines the reduction of a tetrazolium dye and the consequent production of purple colour (Bochner *et al.*, 2001). Out of the 20 *B. vulgatus* isolates from this study, 9 were selected for this experiment based on their diversity in the pangenomic tree created by anvi'o (2.16). Tests were performed in triplicates and according to the manufacturer's instruction with some modifications: (i) the cryo-stocks of the bacterial isolates were inoculated onto WCA instead of Biolog Universal Anaerobe (BUA) agar and subcultured once before usage; (ii) a final concentration of 0.75 mM potassium ferricyanide solution was used to reduce the high background signal produced by the *Bacteroides* species. Instead of using the manufacturers recommended turbidimeter to adjust the bacterial concentration, a turbidity standard (BLG3427) was used. Figure 3 shows pictures of wells after incubation to illustrate when a plate was deemed evaluable and when an isolate was positive for the utilization of a given substrate. A plate was considered evaluable when the A1 well was negative, i.e. translucent (panel A1). When A1 was positive (panel A2) or slightly positive (panel A3), the plate was discarded without recording results. Two or three evaluable plates were necessary for

an isolate to be included in the final analysis. An isolate was considered positive for a substrate when a purple colour formation appeared in the specific well (B panels)

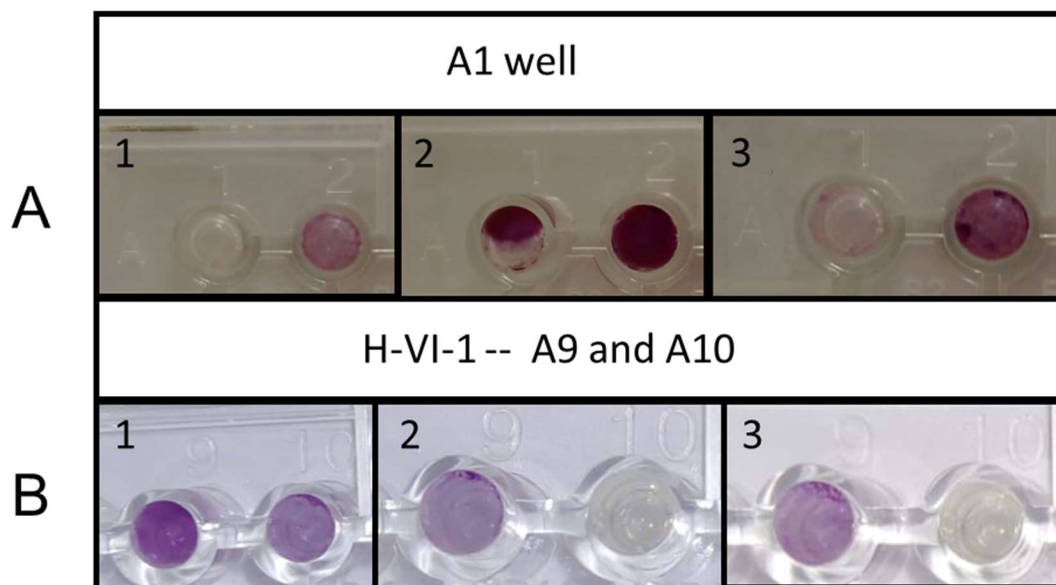


Figure 3: Exemplary pictures of the evaluation of AN MicroPlate™ BIOLOG tests.

(A1) Appropriate control, *i.e.* 'A1 well is translucent. In cases of a positive (A2) or slightly positive (A3) reaction in the well A1, plates were discarded. (B) Exemplary pictures of the A9 and A10 wells of isolate 'H-VI-1' is shown. The A9 well was positive, as indicated by a clear purple colour for all three plates (B1-3). Such a reaction was marked with '+++' in table 11-15. In contrast, the A10 well was only positive once (B1), and was marked with '+' in table 11-15.

2.10. HPLC-measurement of SCFAs

Measurements were done according to Rasmussen *et al.* (2019). The *B. vulgatus* isolates (Table 1) were grown from cryo-stocks. 300 µl of the stock was inoculated into a Hungate tube containing 9 ml YCFA medium. This medium was chosen due to its ability to support the growth of a wide range of bacteria while giving relatively lower background signal during analysis. The grown bacteria were later subcultured again in fresh YCFA medium in Hungate tubes. At Timepoint 0 (T0), 1 ml of the freshly inoculated medium was aliquoted into an Eppendorf tube and centrifuged (10 min, 10,000g, 4°C), the supernatant was then filtered through a 0.2 µm regenerated cellulose filter and transferred into an HPLC vial for analysis. After 24 hours of anaerobic incubation at 37°C, another sample (T24) was taken per in T0. YCFA medium without inoculum was used as negative control. The assay was done in triplicates for each *B. vulgatus* strain (Table 1) using one subcultured of each isolate as the same inoculum. The cut-off value for the inclusion of a measurement into the final analysis was 0.5 mM. The analysis setup consisted of a SUGAR SH1011 column (300 x 8.0mm) from Shodex (Showa Denko Europe, Munich, Germany) which was connected to a VWR Hitachi Chromaster 5450 Refractive Index Detector (VWR

International GmbH, Darmstadt, Germany). As a guard column, a Shodex SUGAR SH-G (6.0 x 50 mm) (Showa Denko Europe, Munich, Germany) was used. Injection volume was set to 40 µl and the operating temperature for the column and the guard column was set to 40°C and kept constant. The eluent was 10mM H₂SO₄ with a constant flow of 0.6ml/min. Concentration was analysed with commercially available external standards of all determined compounds with a comparison of the retention time (all chemical compounds: Sigma-Aldrich, Merck KGaA, Germany). Integration of the peaks was done using the Chromaster System Manager software (Version 2.0, Hitachi High-Tech Science Corporation, Japan).

2.11. Bile Salt Hydrolase (BSH)-Assay

A plate-based bile salt hydrolase (BSH)-assay (Dashkevicz and Feighner, 1989) was used to detect the presence of BSH activity. BSH catalyzes the deconjugation of primary bile acids by hydrolyzing the amide bond between the bile acid and amino acids by using a nucleophilic cysteine (Grill *et al.*, 1995). The products of the deconjugation are free primary bile acids (*e.g.* cholic and chenodeoxycholic acid). Deconjugation of bile salts allows further bioconversion of primary to secondary bile acids by the gut microbiota. For this assay, WCA-agar was supplemented with filter-sterilized 0.5 % w/v sodium taurodeoxycholate acid (TDCA) (Sigma-Aldrich, Missouri) after autoclaving. Moreover, phenosafranin was added as a redox potential indicator, which allows a better colour differentiation between colonies growing on control plate (in pinkish) and on TDCA plates (in white) (Figure 4, (A)). Normal WCA plates without TDCA but inoculated with each bacteria were used as controls. Plates were placed inside the anaerobic chamber at least 2 days before use. Each *B. vulgatus* isolate was tested once.

Fusobacterium mortiferum 'FSA-380-WT-2B' (DSM108838, pig origin) and *E. coli* 'McC-252-APC-1B1' (DSM106279, pig origin) were used as a positive and negative control, respectively. The plates were incubated at 37°C under anaerobic conditions until growth was visible.

Figure 4 illustrates the evaluation of BSH assays. When precipitation around the colonies on WCA agar plates containing 0.5 % TDCA, resulting in the agar becoming more opaque (A and B), or a distinct difference in the colour of arising colony

compared to the WCA plates (A) could be observed, the isolate was considered positive for the presence of a BSH-enzyme. When the colonies had the same pink colour and no precipitation was seen, the isolate was considered negative for the presence of a BSH-enzyme (C).

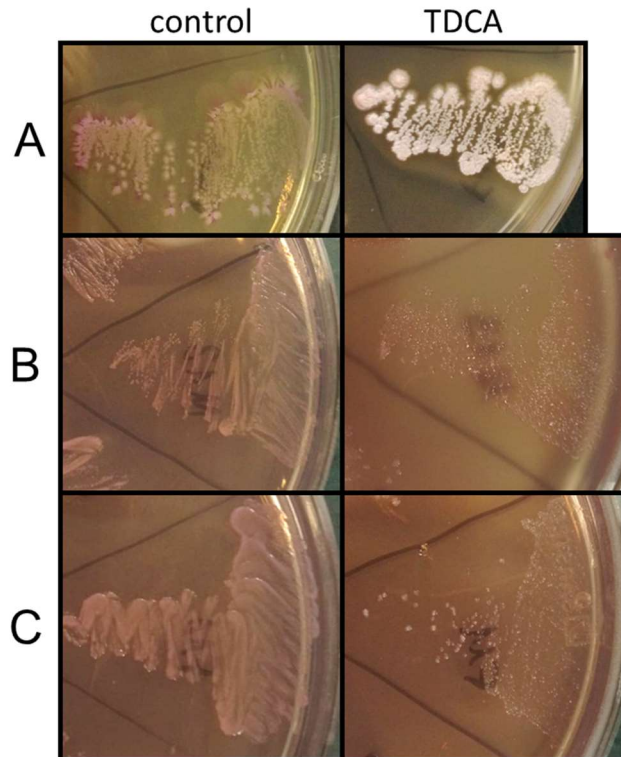


Figure 4: Exemplary evaluation of the BSH-activity assay

Pure cultures were transferred on WCA agar plates containing 0.5 % TDCA or WCA control plates and incubated for six days at 37°C. (A) *Fusobacterium mortiferum* produced pinkish colonies without signs of precipitation on the control plate (left) whereas it produced white colonies surrounded by a halo of precipitation on the WCA-TDCA plate (right). (B) The own isolate *B. vulgatus* 'M-AC-1' produced pinkish colonies without signs of precipitation on the control plate (left) whereas the WCA-TDCA test agar became opaque due to precipitation (C) *B. vulgatus* H-AC-4 did not show any differences in color of the biomass and appearance of the agar between the two media.

2.12. Sequencing

The genomes of all *B. vulgatus* isolates (Table 1) were sequenced at the Quantitative Biology Center (QBiC) (Tübingen). Therefore, the *B. vulgatus* isolates were revived from cryostocks by transferring 300 µl into a Hungate tube containing 9 ml modified GAM medium. The bacterial cultures were transferred once in a new Hungate tube, to reduce the amount of glycerol in the medium and the DNA was extracted as described in 16S rRNA gene sequencing (modified version of the protocol by Godon *et al.* (1997)). The extracted DNA was then measured by QUBIT (Thermo Fisher Scientific, USA). If the amount of DNA was sufficient (>50 ng/µl) 1 µl of each isolate's DNA was loaded together with 1:6 Loading Dye (Thermo Fisher Scientific, USA) on a 1,5 % (w/v)

agarose gel to check the DNA for degradation. Undegraded DNA was then stored at -80°C until shipment on dry ice.

2.13. Genome Assembly and Plasmid Analysis

Genome Assembly and Plasmid Extraction

Raw sequence files were quality filtered and have Illumina adapters removed (parameters: PE -phred33 ILLUMINACLIP:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:5:20 MINLEN:50) using Trimmomatic v0.39 (Bolger *et al.*, 2014). In order to remove phiX (Sanger *et al.*, 1977) contamination (Mukherjee *et al.*, 2015) reads were aligned to the phiX genome using bbduk (Bushnell, 2014) and the unaligned reads were used in downstream analyses. Prior to genome assembly, a pipeline of plasmidSPAdes (Antipov *et al.*, 2016), SAMtools (Li *et al.*, 2009) and Recycler (Rozov *et al.*, 2017) were used to reconstruct plasmids. The reconstructed plasmids sequences were filtered from the reads by using bbduk. In the final step, the genomes were assembled using SPAdes v3.13.1 (Bankevich *et al.*, 2012). The completeness and presence of contamination within the assembled genomes were tested using CheckM (Parks *et al.*, 2015). The isolate 'M-VI-1' with 27.24 % contamination was excluded from the analyses.

Plasmid Identification and Content Analysis

In order to identify plasmids within the genomes of isolates, their sequences were searched against the bacterial plasmid database PLSDB (Galata *et al.*, 2019). The nucleotide sequences of the plasmids were compared using Mash (Ondov *et al.*, 2016) with a max. p-value of 0.1 and a min. identity of 0.9. Plasmids that could not be identified by PLSDB were described as 'Not Identified (NI)'. To study the gene repertoire of the plasmids, Prodigal (Hyatt *et al.*, 2010) was utilized (with default parameters) to predict protein sequences. The protein sequences were then searched against the protein family (Pfam) database (Finn *et al.*, 2014) using the HMMER web service (Finn *et al.*, 2011).

2.14. Published Genome Datasets

For a comprehensive analysis, additional genomes from *B. vulgatus* isolates obtained in two published large-scale cultivation studies were processed. The first dataset is the Human Gastrointestinal Bacteria Culture Collection (HBC) by Forster *et al.* (2019). The second dataset is the Broad Institute-OpenBiome Microbiome Library (BIO-ML) by Poyet *et al.* (2019). In total, both datasets contained more than 4,000 genomes of cultured human gut bacteria (Figure 5).

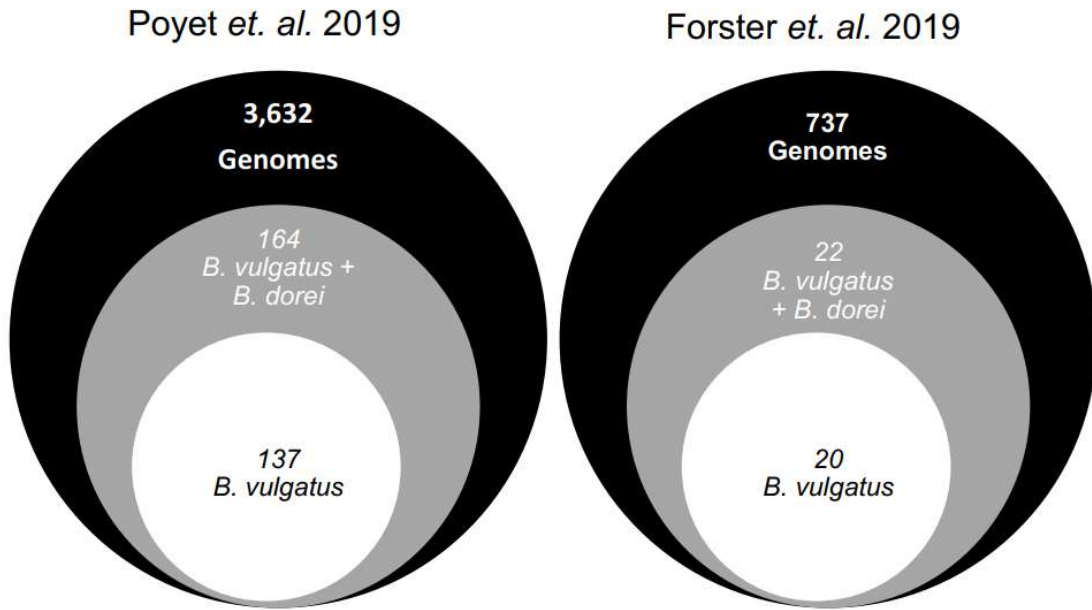


Figure 5: Filtering of the datasets of Poyet *et al.* (2019) and Forster *et al.* (2019)

Average Nucleotide Identity values (ANI-values) calculated via fastANI (Jain *et al.*, 2018) were used to delineate species boundaries using the genome of *B. vulgatus* ATCC 8482^T (CP000139) as a reference genome. The threshold-value for species-level of ANI-values is >95 % (Konstantinidis and Tiedje, 2005). Genomes meeting this criterion were then compared to the genome of *B. dorei* CL03T12C01^T (CP011531) due to the phylogenetic proximity between *B. vulgatus* and *B. dorei*. Isolates with ANI values higher to the *B. dorei* reference genome were compared to both references using digital DNA-DNA hybridization (dDDH). dDDH values were calculated by using the Genome-to-Genome Distance Calculator (GGDC, (Meier-Kolthoff *et al.*, 2013). The dDDH threshold-value for species-level is >70% (Moore *et al.*, 1987; STACKEBRANDT and GOEBEL, 1994). This process was done to confirm the identification of *B. dorei* genomes within the current dataset. All downstream analysis was conducted using only the genomes assigned strictly to *B. vulgatus*.

2.15. Ecological Analysis

The prevalence and average relative abundance of *B. vulgatus* in human body microbiomes and the pig, mice, chicken and bovine gut was calculated using Protologger (<http://protologger.de/>) based on 1,000 16S rRNA gene amplicon samples from each sample type downloaded from IMNGS (www.imngs.org) (Lagkouravdos *et al.*, 2016a). The input was the genome of *B. vulgatus* ATCC 8482^T (CP000139) and its 16S rRNA gene (NR_074515).

2.16. Pangenomic Analysis with anvi'o

Anvi'o is an analysis and visualization platform for 'omics data which allows the user to conduct multiple comparative genome-based analyses, including pangenomics (Eren *et al.*, 2015). The pangenomic workflow of anvi'o (Delmont and Eren, 2018) was utilised to generate a pangenomic tree of the 176 sequenced *B. vulgatus* strains (20 own isolates and else from the literature) and six additional other *Bacteroides* species as an out-group. Before processing in anvi'o, the genome files had to be reformatted and converted into an anvi'o contigs database. The reformatting was done using the program 'anvi-script-reformat-fasta' with the --simplify-names flag. Anvi'o needs simplified deflines in order to create an anvi'o contigs database. The reformatted genome files were then converted by using the program 'anvi-gen-contigs-database'. This program utilizes Prodigal v2.6.3 (Hyatt *et al.*, 2010) with default settings to identify open reading frames. The command 'anvi-run-hmms' was used to identify single-copy core genes within each genome for the generation of the pangenomic tree. Protein sequences were assigned function via comparison against the Pfam database (Finn *et al.*, 2014) using the command program 'anvi-run-pfams'. Both annotations utilised Hidden Markov Models (HMMs), implemented in HMMER (Eddy, 2009, 2011). The command, 'anvi-gen-genomes-storage', with the '--external-genomes' flag was used to store DNA and amino acid sequences, as well as functional annotations of each genome in a single database. The pangenome was then identified using the 'anvi-pangenome' command with the flags; '--use-ncbi-blast', the parameters '--mcl-inflation 10' '--min-occurrence 2' and the default parameter '--minbit 0.5'. This command compared all proteins within each genome to each other via blastp (Altschul *et al.*, 1990). Weak hits, based on the aligned fraction between the two reads, were removed by the program utilizing the 'minbit heuristic', originally described in an integrated toolkit for

exploration of microbial pan-genomes (ITEP, Benedict *et al.*, 2014). The program used the MCL algorithm (van Dongen and Abreu-Goodger, 2012) to identify gene clusters in the remaining blastp search results. For comparing very closely related strains, an MCL-inflation coefficient of 10 was used. Once identified, the gene clusters occurrence across the genomes was used to conduct hierarchical clustering analyses based on an agglomerative strategy using the euclidean distance (Anton, 1994) and wards linkage algorithm (Ward, 1963). The flag 'min-occurrence 2' was used to prevent the inclusion of gene-clusters present in less than 2 genomes.

Phylogenomic analysis of the *Bacteroides* genomes was conducted using the command 'anvi-get-sequences-for-gene-clusters' and 'anvi-gen-phylogenomic-tree'. The first command collected the required gene sequences as determined by; '--min-num-genomes-gene-cluster-occurs 182', '--max-num-genes-from-each-genome 1', '--min-geometric-homogeneity-index 0.95' and '--concatenate-gene-clusters' to concatenate and align target genes from all genomes. Then the command 'anvi-gen-phylogenomic-tree' generate the phylogenomic tree from the concatenated FASTA file. The phylogenomic tree was rerooted using *Bacteroides fragilis* as the outgroup. Final visualisation was conducted using the 'anvi-display-pan' (Delmont and Eren, 2018).

2.17. *In silico* Carbohydrate- and Antibiotic Resistance Profiling

For the *in silico* prediction of carbohydrate- and antibiotic resistance-profiles, Protologger (<http://protologger.de/>) was used. Protologger performs diamond (Buchfink *et al.*, 2015) searches against the Carbohydrate-active enzymes database (Lombard *et al.*, 2014) (with parameters -blastp --query-cover 40.0 --subject-cover 40.0 --min-score 100.0) and the comprehensive antibiotic resistance database (Jia *et al.*, 2017; McArthur *et al.*, 2013) (with parameters --query-cover 80.0 --subject-cover 80.0 --id 80.0) in order to identify proteins that match against proteins in the respective databases. Annotated proteins are filtered according to their best-hit, as determined by bitscore.

Multi-dimensional plotting of the presence and absence of CAZyme families using both the binarized and occurrence profiles was conducted by calculating their dissimilarities via the Jaccard (Jaccard, 1902) and Bray-Curtis (Bray and Curtis, 1957) indices.

Plotting was done in R using the *vegan* (Jari Oksanen *et al.*, 2019), *dplyr* (Andy Bunn *et al.*, 2020) and *RcolorBrewer* (Erich Neuwirth, 2014) packages.

2.18. Phenotypical Prediction by using Traitair

Traitair is a microbial trait analyzer that utilises the genome sequence of an isolate to predict potential phenotypes (Weimann *et al.*, 2016). The online version of Traitair was used at <https://research.bifo.helmholtz-hzi.de/webapps/wa-webservice/pipe.php?pr=traitair>. Traitair can predict 67 phenotypes directly from a genome sequence. The standard workflow of Traitair is shown in Figure 6.

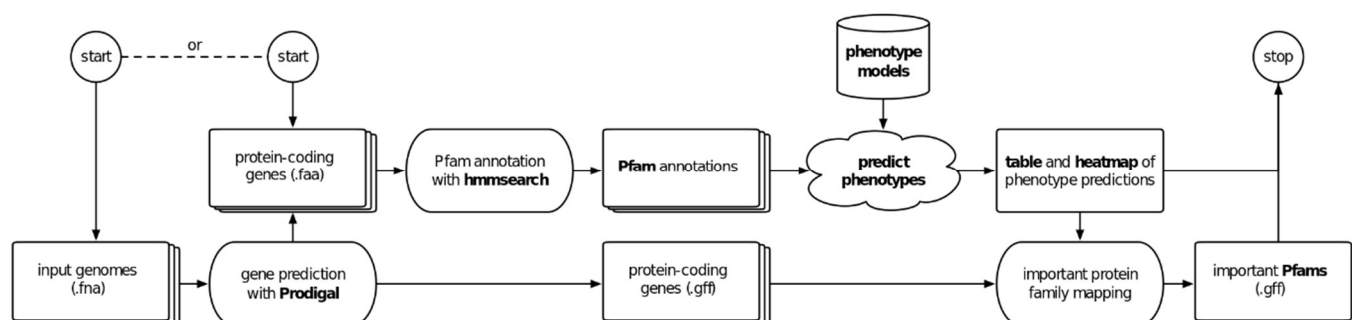


Figure 6: Traitair workflow

Taken from github.com/hzi-bifo/traitair/

Prodigal (Hyatt *et al.*, 2010) predicts open reading frames (ORFs) of which Pfam annotation via the Pfam database (Finn *et al.*, 2014) is done using HMMER 3.0 (Eddy, 2009, 2011). Traitair utilises the Pfam annotations to predict phenotypes for the 67 traits by the use of one of two different classification models. The first model, called ‘phypat classifier’, trained on the protein and phenotype presence and absence labels from 234 bacterial species whereas the second classifier, ‘phypat+PGL classifier’ was trained by using the same data and additional information on evolutionary protein family and phenotype gains and losses (Weimann *et al.*, 2016). Phenotype data used for machine learning originated from the microbiology section of the Global Infectious Disease and Epidemiology Online Network (GIDEON) (Berger, 2005).

2.19. Phenotypical Prediction by using PhenDB

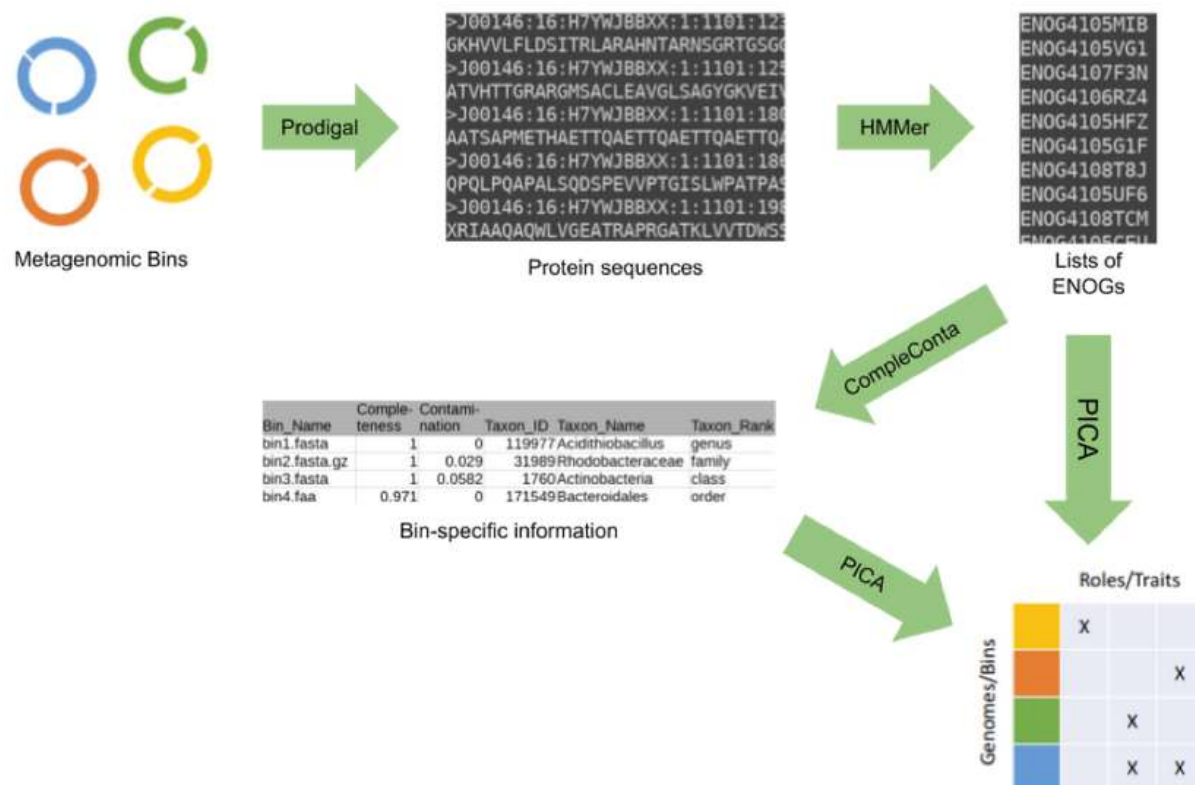


Figure 7: Workflow of the PhenDB pipeline
(<https://phenodb.csb.univie.ac.at/methods>)

PhenDB predictions are performed on orthologous groups of proteins (ENOGs), as shown in Figure 7. The uploaded bins/genomes predicted proteins are represented in those ENOGs. The first step of the pipeline utilizes Prodigal (Hyatt *et al.*, 2010) to predict protein-coding genes in the bins/genomes. For annotation, the predicted proteins are compared to the bactNOG database of eggNOG 4.5.1 (Huerta-Cepas *et al.*, 2016) using the ‘-hmmsearch’ command (E-value: $1e^{-30}$) of HMMER (v.3.1b2, Eddy, 2009, 2011). To predict phenotypes, the PhenDB pipeline uses the PICA software (Feldbauer *et al.*, 2015) based on models trained on a dataset of genomes which belong to organisms with known traits. In this study, the PhenDB pipeline was used to predict the presence of BSH activity only.

3. Results

3.1. Cultivation

Gut bacteria were cultured in a targeted manner with the aim of obtaining multiple strains of any given *Bacteroides* species from humans, mice and pigs. In total, 571 colonies were picked, of which 382 could be identified by either MALDI-TOF-MS or 16S-rRNA gene sequencing. Out of the remaining isolates, 22 isolates could not be identified in due time and were stored as cryo-cultures at -80°C. A further 167 isolates were lost during the identification process (*i.e.* did not grow after sub-cultivation). From all colonies picked, 48.3 % were from mice (276), 32.2% from pigs (184), and 19.4 % from humans (111) (Table 5).

Table 5: Numbers of isolates picked, identified, not yet identified, and lost during cultivation.

Host species	Colonies picked	Identified isolates	Not yet identified	Lost
Mice	276	163	12	101
Pig	184	122	6	56
Human	111	97	4	10
Total	571	382	22	167

Bacteroides acidifaciens was isolated in seven out of fifteen mice. However, this species was not found in either human or pig samples and was thus excluded from downstream analyses. *B. ovatus* was very prevalent in human (five of six samples) but was found only once in each pigs and mice. *B. vulgatus* was the only species for which several strains could be isolated from all three mammalian hosts, including three human, two pig and three mice samples (Figure 8). However, one of the pig strains could not be maintained in culture. Additionally, human *B. vulgatus* strains were obtained from our collaborator Till Strowig in Braunschweig (n=8) and David Berry in Vienna (n=1). The *B. vulgatus* strains from miBC (mice) (Lagkouvardos *et al.*, 2016b) and PiBAC (pig) (www.dsmz.de/pibac) (unpublished) were also included.

Bacteria / Samples	PG 1	PG 2	PG 3	PG 4	PG 5	PG 6	PG 7	PG 8	MICE 1	MICE 2	MICE 3	MICE 4	MICE 5	MICE 6	MICE 8	MICE 9	MICE 10	MICE 11	MICE 12	MICE 13	MICE 14	MICE 15	MICE 16	HUMAN 1	HUMAN 2	HUMAN 3	HUMAN 4	HUMAN 5	HUMAN 6
Bacteroides acidifaciens																													
Bacteroides caccae																													
Bacteroides caecimuris																													
Bacteroides dorei																													
Bacteroides faecis																													
Bacteroides fragilis																													
Bacteroides ovatus																													
Bacteroides stercoris																													
Bacteroides thetaiotaomicron																													
Bacteroides vulgatus																													
Bacteroides xylanisolvens																													
Non-Bacteroidaceae																													

Figure 8: *Bacteroides* species isolated from gut samples of the different host species

Green boxes indicate that a sample was positive for the respective *Bacteroides* species. *B. vulgatus* is highlighted in light green as this is species of choice.

3.2. Ecological Analysis

We then assessed the habitat preference of *B. vulgatus* by analyzing thousands of amplicon data from different host species at a threshold of 97 % sequence identity. There was a strong association with gut environments in several host species including human (77.4 % prevalence), pig (52.5 %), chicken (35.5 %), mouse (31.1 %) and bovine (29.2 %). *B. vulgatus* was less prevalent in the insect gut (5.4 %) and also other human body habitats: 34.5 %, 20.2 %, 10.0 % and 4.8 % in vaginal, skin, lung and, oral samples, respectively (Figure 9). The mean relative abundance of *B. vulgatus*-like sequences in gut environments was above 2%, except for the insect gut (0.85 %). While the mean relative abundance in human gut samples was 7.3 % (SD: 10.6 %), it was below 1 % in the other human-environment samples (Figure 9).

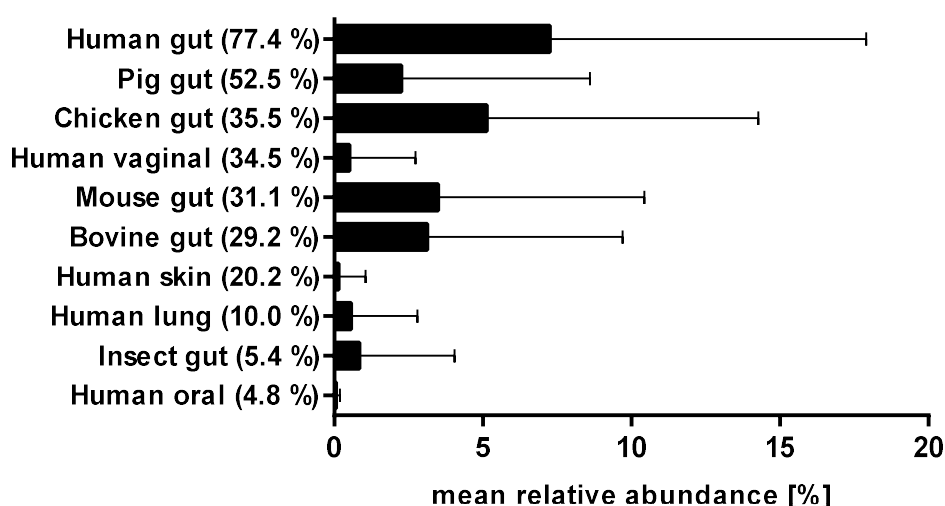


Figure 9: Occurrence of *B. vulgatus* in ten different types of samples

The mean relative abundance (97 % 'species-level' similarity) was calculated across 1,000 samples per environment with only samples positive for *B. vulgatus* being considered. The prevalence of *B. vulgatus*, i.e. percentage of positive samples, is indicated in parenthesis next to the corresponding sample type (in decreasing order).

3.3. Curation of published genome datasets

As the two species, *B. vulgatus* and *B. dorei* are phylogenetically closely related, the affiliation of the isolates of the published genome datasets (Figure 5) (Forster *et al.*, 2019; Poyet *et al.*, 2019) was investigated by comparing their ANI to the type strains *B. vulgatus* ATCC 8482^T and *B. dorei* CL03T12C01^T (Figure 10). This filtering of the was a requirement to ensure a valid comparison of strains solely within the species *B. vulgatus*.

The importance of the filtering is highlighted in Figure 10, showing that 28 out of the 186 genomes originally selected because of ANI values >95 % (species level delineation) to *B. vulgatus* ATCC 8482^T had a greater ANI to *B. dorei* CL03T12C01^T (>98.5 %) than to *B. vulgatus* (<96.0 %). Those genomes were thus excluded from all analyses.

Furthermore, 150 out of the 186 genomes had a greater ANI to *B. vulgatus* ATCC 8482^T (>98.0 %) than to *B. dorei* CL03T12C01^T (95 -96 %). A further six genomes had an ANI of ~97.0 % to *B. vulgatus* ATCC 8482^T and ~96.0 % to *B. dorei* CL03T12C01^T (Appendix 3, Supp. Table S7). Those 156 genomes were then used as a clean comparison dataset in downstream analyses.

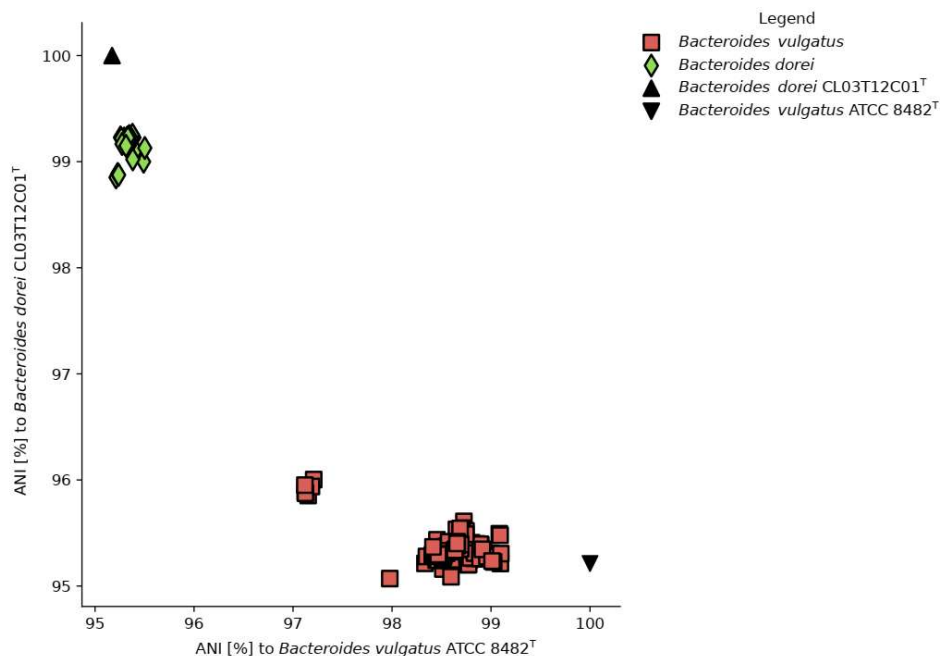


Figure 10: Distribution of isolates from the published datasets to the type strains of *B. vulgatus* and *B. dorei* based on ANI

ANI [%] of the isolates to *B. vulgatus* ATCC 8482^T and *B. dorei* CL03T12C01^T are shown on the x- and y-axis, respectively.

As ANI values are used to delineate species, we looked at a potential confounding contribution of plasmidic DNA. Comparisons of ANI values to *B. vulgatus* ATCC 8482^T calculated with or without removal of plasmid sequences as identified by a pipeline of plasmidSPAdes and Recycler revealed no significant differences (Figure 11).

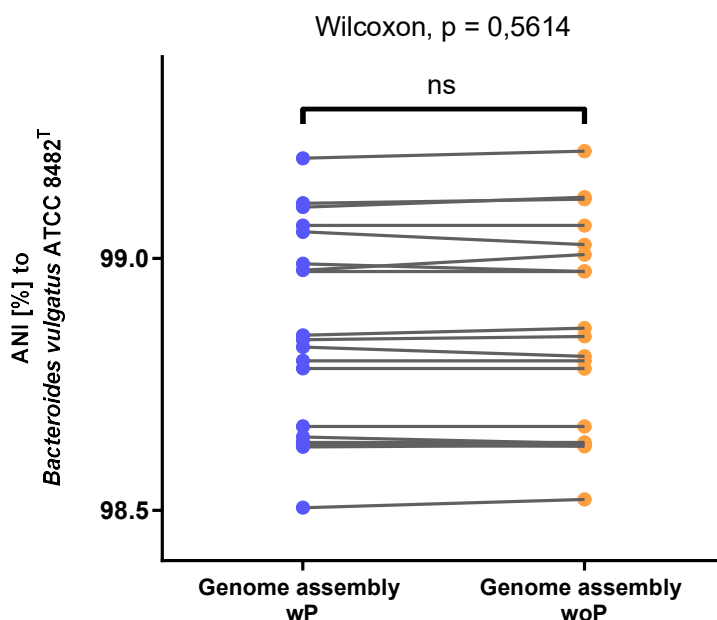


Figure 11: Influence of plasmids within genome assemblies on Average-Nucleotide-Identities

The different genome assembly methods are shown on the x-axis whilst the ANI [%] to *B. vulgatus* ATCC8482^T is plotted on the y-axis. Wilcoxon matched-pairs signed rank test was used to calculate statistical significance. Abbreviations: wP, with Plasmid; woP, without Plasmid

3.4. Pangenomic Analysis by anvio

The pangenome analysis of 182 isolates was conducted to identify gene assemblages which are unique for specific strains. These genomes included our own 19 isolates, the 156 genomes from the literatures, and six additional *Bacteroides* species other than *B. vulgatus* to allow the distinct determination of the core pangenome of *B. vulgatus*.

User-directed merging of gene assemblages linked to taxonomic clades led to the identification of 19 distinct clusters of genes specific for clades or even subclades of strains. As expected, the inclusion of related *Bacteroides* species allowed separation of a *Bacteroides* core genome from a core specific to *B. vulgatus* (Figure 12). The identification of the ‘*Bacteroides dorei* / *vulgatus* core’ cluster and other more specific clusters of genes (e.g. corresponding to host species-specific isolates) are the major

findings of this analysis and are therefore further described in this section. A very broad summary of all bacterial clade-specific clusters of genes can be found in Table 6.

Across 176 genomes, a *B. vulgatus* core pangenome was identified and also shared with the *B. dorei* reference genome (Figure 12). This 'Bacteroides dorei / vulgatus core' cluster consisted of 449 gene clusters, of which 337 were assigned to known pfams including PF07944 (β -L-arabinofuranosidase), PF00557 (Metallopeptidase family M24 protein), multiple times PF7980 (SusD family protein), PF10566 (Glycoside hydrolase 97), PF01112 (Asparaginase), PF00534 (Glycosyl transferases group 1).

The clade cluster 'cluster1b sub 1c mouse' was of special interest since it included all mouse strains from this study. This cluster was identified to be unique for those four mouse isolates and another human isolate from the literature genome datasets (Figure 12) In total, 181 gene clusters were specific of this bacterial clade, of which 81 were assigned to known pfams including PF13354 (β -lactamase family protein), PF00535 (Glycosyltransferase family 2 protein), PF00753 (Metallo- β -lactamase family protein) and PF00534 (Glycosyltransferase family 1 protein). 'Cluster4a' and 'cluster4b' were specific for the clade containing the six genomes of those strains that were only slightly more similar to *B. vulgatus* than to *B. dorei* based on ANI values. Five of these six genomes contributed genes to 'cluster4a' including functions such as PF07980 (SusD family protein), PF04616 (Glycosylhydrolase family 43 protein), PF00295 (Glycosylhydrolase family 28 protein) and PF15979 (Glycosylhydrolase family 115 protein). Cluster4b consists of a single *B. vulgatus* isolate that shared 11 gene clusters with the *B. dorei* reference genome. Seven out of these 11 gene clusters were assigned to known pfams, including PF10566 (Glycosylhydrolase 97 family protein), PF00534 (Glycosyltransferase family 1 protein) and PF00295 (Glycosylhydrolase family 28 protein).

The identification of four major distinct clades of strains, which contained further subclades with unique assemblages of genes is promising. The identification of a cluster of almost only mouse strains supports the concept of host specificity mentioned in the introduction. Furthermore, in both Cluster4a and Cluster4b, a high number of CAZymes were identified, suggesting that the carbohydrate-degrading capacity of an isolate may be related to its phylogenetic placement.

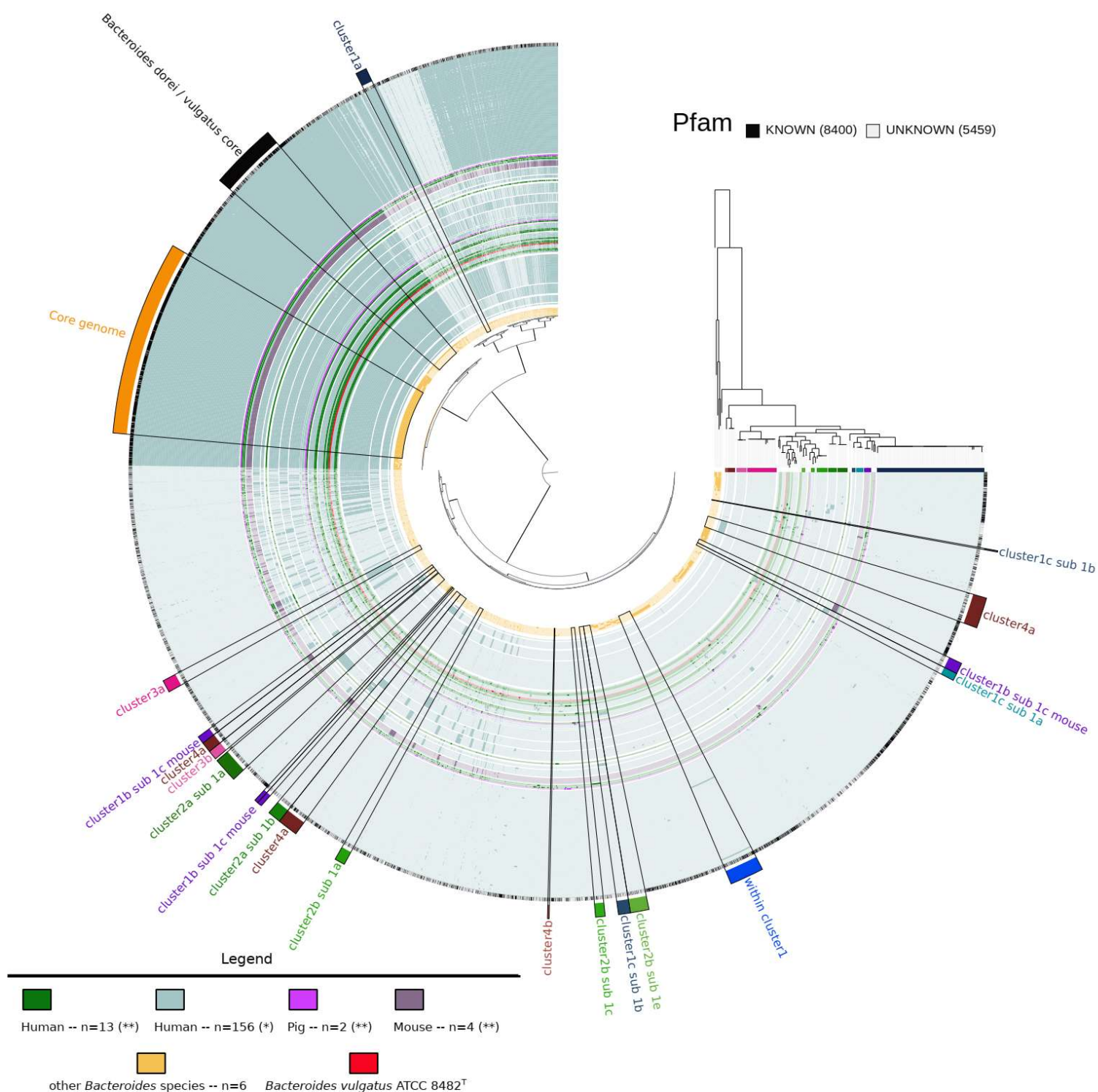


Figure 12: Pangenome tree of 182 isolate genomes belonging to the *Bacteroides* genus

177 isolates belonging to the *B. vulgatus* species were studied. Of these, 156 isolates (light blue) are from the Forster *et al.* (2019) and Poyet *et al.* (2019) datasets (indicated by * in the legend), 13 own isolates (dark green) from human study, 2 isolates (pink) from pig, 4 isolates (purple) from mice, and the type strain *B. vulgatus* ATCC 8482^T (red). Own isolates are marked with ** in the legend. Six additional genomes (orange) representing *Bacteroides* species other than *B. vulgatus* were also included. Clusters of genes represented by different colours around the tree were specific to (sub)clades of strains across the phylogenomic tree. The colours of clades under the phylogenomic tree are consistent with the colours of the clusters of genes. The phylogenomic tree was rooted on *B. fragilis* as an outgroup. The outer circle in black indicates whether the assigned pfam to a gene cluster were known (black) or unknown (white). The cluster 'Core genome' (orange) includes gene clusters which were present in all 182 isolates. The cluster 'Bacteroides dorei / vulgatus core' (black) includes genes that were exclusive to the 177 *B. vulgatus* genomes and the *B. dorei* reference genome. Hierarchical clustering was performed on the presence/absence of gene clusters using Euclidean distance and Ward linkage.

Table 6: Overview of the identified clusters and their content.

Cluster	Gene Clusters	Pfam known	Pfam unknown	Genomes contributing genes to cluster
Core genome	1305	1260	45	182
<i>Bacteroides dorei</i> / <i>vulgatus</i> core	449	337	112	177
cluster1a	74	40	34	79
within cluster1	227	151	76	2
cluster1b sub 1c mouse	181	81	100	5
cluster1c sub 1a	50	16	34	5
cluster1c sub 1b	89	29	60	2
cluster2a sub 1a	165	82	83	7
cluster2a sub 1b	88	38	50	6
cluster2b sub 1a	69	30	39	8
cluster2b sub 1c	64	27	37	2
cluster2b sub 1e	124	52	72	2
cluster3a	83	45	38	21
cluster3b	63	24	39	7
cluster4a	413	236	177	5
cluster4b	11	7	4	1+ <i>B. dorei</i>

3.5. CAZyme Analysis

The pangenomic analysis revealed strain-level differences in the CAZyme content of strains, therefore the investigation of the carbohydrate-degrading capacity of all own isolates was conducted. For comparison, the number of present CAZyme families and the number of total CAZymes were plotted in Figure 13.

Minor differences between the 19 isolates were identified. The isolates originated from mice consistently lacked 2 CBM's (CBM2 and CBM35) compared to the majority of the other isolates. They were also characterized by up to 10 % lower total number of CAZymes. The identified differences during the pangenomic analysis and those shown in Figure 13 led to the further investigation of the dissimilarities between the isolates. A Jaccard and Bray-Curtis dissimilarity plot with all isolates was created (Figure 14 and Figure 15, respectively) to illustrate those dissimilarities of the isolates due to the presence/absence of CAZyme families and the number of enzymes within each family.

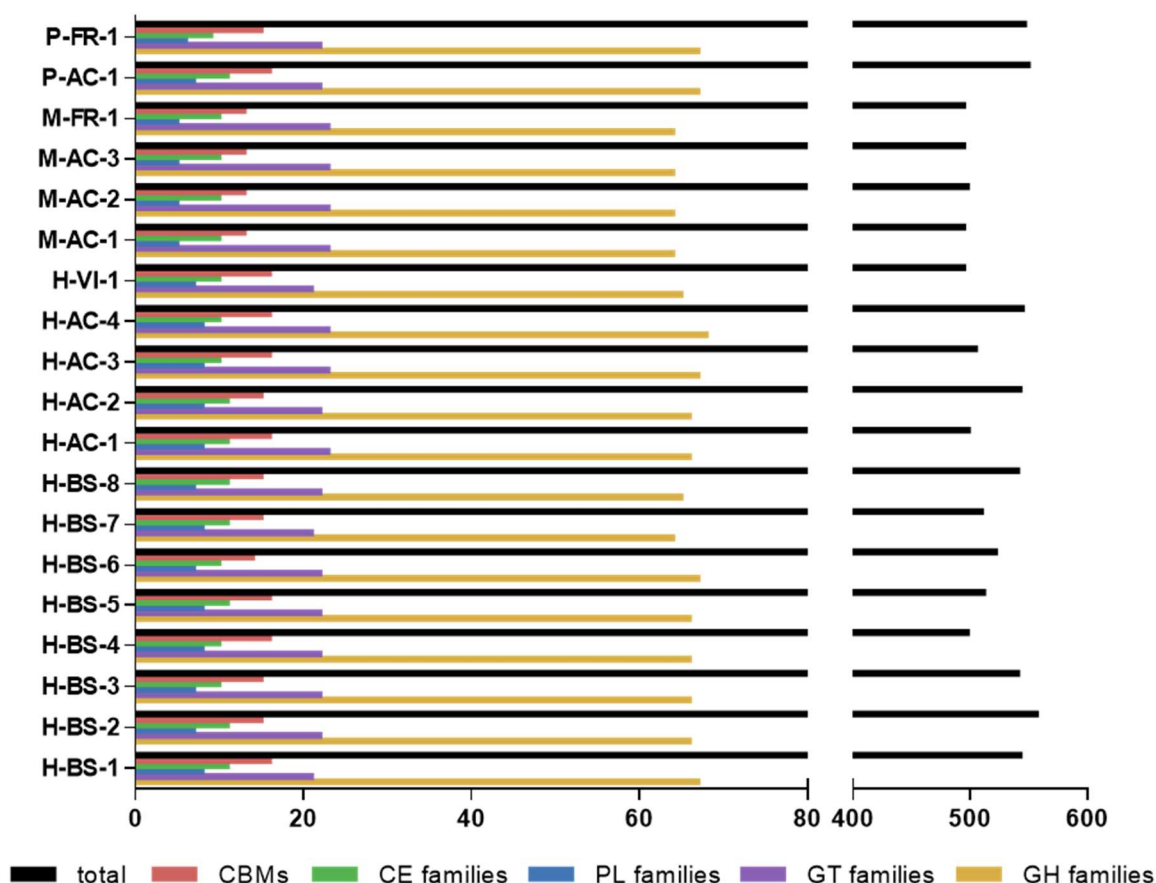


Figure 13: Number of CAZyme families and CAZymes within isolates from this study

The number of CAZymes and families thereof are shown on the x-axis. The corresponding isolates are distributed on the y-axis according to their origin: host species (P, pig; M, mouse; H, human); city (Fr, Freising; AC, Aachen; VI, Vienna; BS, Braunschweig). Abbreviations: Carbohydrate-Binding Modules (CBMs), Carbohydrate Esterases (CEs), Polysaccharide Lyases (PLs), Glycosyl Transferases (GTs) and Glycoside Hydrolases (GHs).

The isolates 'P-AC-1' and 'P-FR-1' from pigs were compared because they were slightly separated from each other based on their presence and absence of their CAZyme families (Figure 14, orange dots). The CAZyme families CE15 (glucuronoyl esterases), GH136 (lacto-N-biosidase), GH27 (α -Galactosidase), GT23 (N-acetyl- β -D-glucosaminide α -1,6-L-fucosyltransferase and chitin-oligosaccharide α -1,6-L-fucosyltransferase) and PL27 (L-rhamnose- α -1,4-D-glucuronate lyase) family were identified to be present in 'P-AC-1' but not in 'P-FR-1' while GH19 (chitinase and lysozyme), GH30_2 (β -glucosylceramidase, β -1,6-glucanase and β -xylosidase), GH8 (chitosanase, cellulase, licheninase and endo-1,4- β -xylanase) and GH101 (endo- α -N-acetylgalactosaminidase) were present in 'P-FR-1' but not in 'P-AC-1'.

The Jaccard dissimilarity plot suggested that the isolates belonging to mice do not harbour different CAZyme families (Figure 14, all four dots overlapped). This was investigated further and no differences in the content of CAZyme families could be identified.

In order to investigate the visible dissimilarity in both plots (Figure 14 and Figure 15), the mouse isolates (purple) were compared to the pig isolates (orange). This revealed that the CAZyme family GT10 (galactoside α -1,3/1,4-L-fucosyltransferase and galactoside α -1,3-L-fucosyltransferase activity) was absent in the two pig isolates.

The mouse isolates, similar to the 'P-AC-1' isolate, were identified to lack GT23. Single comparisons between each mouse isolate to the 'P-AC-1' revealed that the families CBM2 (bind crystalline cellulose, insoluble chitin or xylan), CBM35 (bind uronic acids), CE15, GH108, GH136, GH27, GH43_9 (major activities of α -L-arabinofuranosidases, endo- α -L-arabinanases and, β -D-xylosidases) PL27 and PL29 (chondroitin-sulfate ABC endolyase) were only present in the 'P-AC-1' isolate. For the 'P-FR-1' isolate, the families CBM2, CBM35, GH108 (N-acetylmuramidase), GH19, GH30_2, GH43_9, GH8, GT101 (β -glucosyltransferase) and PL29 were identified to be different from the mouse isolates.

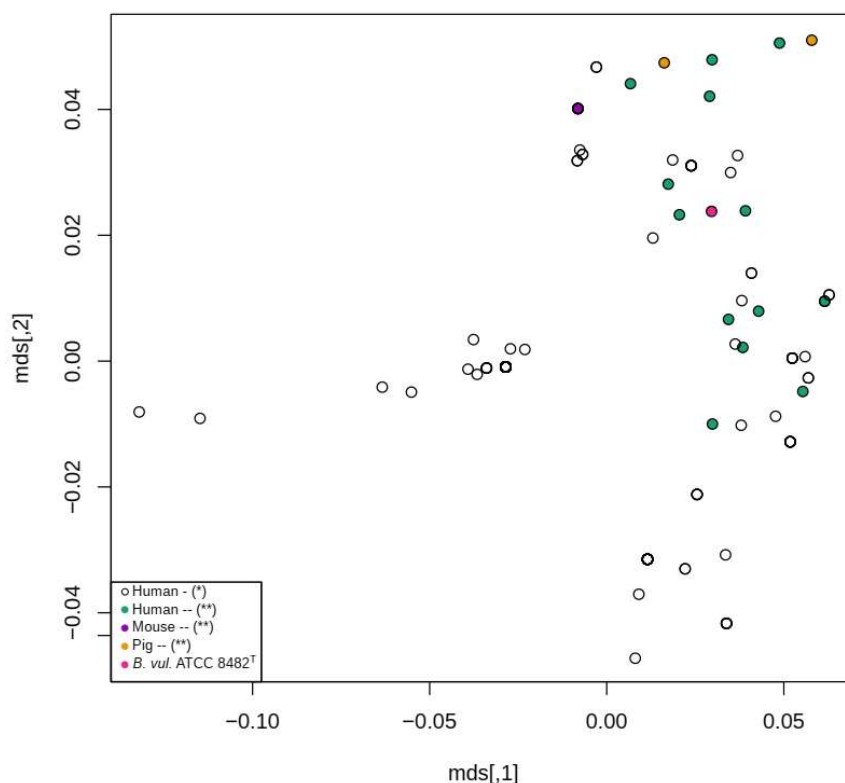


Figure 14: MDS Jaccard dissimilarity of all *B. vulgatus* isolates (n=177) based on CAZyme families

Multi-dimensional plotting (MDS) of the presence and absence of CAZyme families using the binarized profile was conducted by calculating the dissimilarity between each isolate via the Jaccard index. Plotting was done in R, using the vegan, dplr and RcolorBrewer packages.

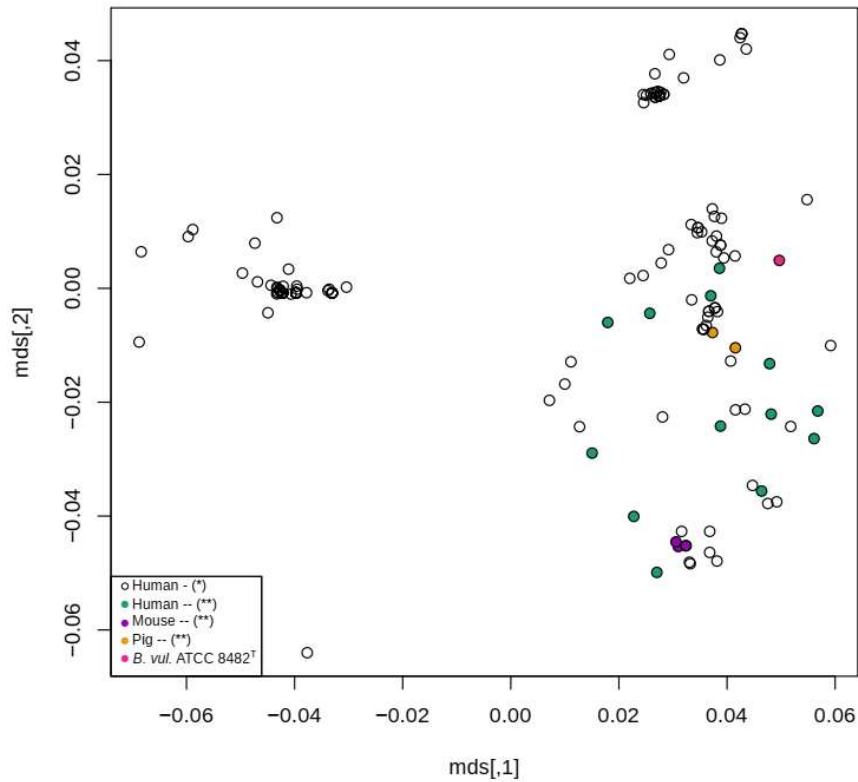


Figure 15 MDS Bray-Curtis dissimilarity of all *B. vulgatus* isolates (n=177) based on CAZyme families
MDS plotting of the presence and absence of CAZyme families using the occurrence profile was conducted by calculating the dissimilarity between each isolate via Bray-Curtis index. Plotting was done in R using the vegan, dplr and RcolorBrewer packages.

3.6. Phenotype Prediction by Traitair and PhenDB

Traitair and PhenDB were used to predict the phenotype of all literature and own strains (Figure 16), of which the latter were then tested during *in vitro* experiments (next sections). Within this section, only major differences between the bacterial clades are described. For individual differences between isolates or for predicted phenotypes that were not different between the clades, readers are referred to Figure 16.

Clade1a ('cluster1a'), which contained 79 genomes only from the literature datasets (Table 6), was characterized by the prediction of susceptibility to bile by either one or both classifiers and by the presence of enzymes hydrolyzing pyrrolidonyl- β -naphthylamide. Susceptibility to bile was also predicted to be positive for clade3b ('cluster3b') which contained seven genomes only from the literature datasets. In contrast, clade3b was predicted to tartrate utilization, which was very common across all the isolates.

The subclade 1c within clade1b ('cluster1b_sub1c'), which contained the four mouse isolates from this study and another human isolate from the literature datasets, was predicted to be positive for a DNase. For the subclade1b within clade2a

('cluster2a_sub1b'), which contained six genomes from the literature datasets, this DNase was also predicted to be present. Additional to the DNase, the latter clade was predicted to be able to grow on agar containing lysine through the presence of a lysine decarboxylase and be able to utilize acetate and citrate. Subclade1a of clade2a, which contained 7 genomes from the literature datasets was only predicted positive for citrate utilization. Isolates belonging to this cluster were negative for the DNase, the lysine decarboxylase, and acetate utilization. Isolates belonging to clade4a ('cluster4a'), which consisted of the *B. vulgatus* isolates with closest similarity to *B. dorei* were also only positive for citrate utilization and did not show any other major differences compared to other (sub)clades.

3.7. Plasmid Analysis

Plasmids are important “vehicles” for the exchange of genetic information between bacteria, which may provide advantages to the host, including additional metabolic capacities or resistance to antibiotics. Therefore, the plasmid repertoire of the 19 own isolates, for which a genome was generated, was investigated (Table 7).

In total, 38 plasmids were identified, with a range of zero to seven per isolate. Six of these 38 plasmids could not be identified with a MASH identity greater than 0.90 and were therefore named ‘Not Identified’ (NI). Only Plasmids whose MASH identity was greater than 0.99 (bold in the table) were considered to be validly identified by PLSDB. In 30 % of the human isolates (4 out of 13), the plasmid ‘pBFO18_2’ and a plasmid called ‘plasmid unnamed1’ were present. The plasmid ‘pBFO42_2’ was identified only in 15 % of the human isolates (2 out of 13). For the human isolate ‘H-VI-1’ the plasmid ‘p5482’ was identified. All mouse isolates contained the same two plasmids, one of which was unknown (‘plasmid unnamed2’). Additionally, isolate ‘M-AC-2’ contained a further plasmid.

To further investigate strain-level diversity within this plasmid repertoire, the protein content of each plasmid sequence was annotated. All functions not involved in replication were gathered and compared between the 19 isolates and summarized in Table 8.

A frequent function was ‘YoeB-like toxin of bacterial type II toxin-antitoxin system’ (12 out of 19 isolates), which has been proposed to be an mRNA interferase but also an inhibitor of translation initiation.

Moreover, a β -lactamase family protein was found on the plasmid of the isolate ‘H-AC-2’, an interesting finding that led us to further investigate differences between the isolates in terms of antibiotic sensitivity (next section).

Table 7: Plasmids in the genome of 19 own isolates

Isolate	Host species	Plasmids [#]	Plasmid ID	length extracted plasmid [bp]	identified by PLSDB as	Mash identity	length (PLSDB) [bp]
H-BS-1	Human	2	H-BS-1_p1	4138	pBFO18_2	0,9987	4137
			H-BS-1_p2	5325	pBFO67_1	0,95533	6129
H-BS-2	Human	1	H-BS-2_p1	4138	pBFO18_2	0,9985	4137
H-BS-3	Human	3	H-BS-3_p1	6500	pRGF00040	0,988957	6931
			H-BS-3_p2	5909	pRGF00015	0,976965	5934
			H-BS-3_p3	2684	NI**	-	-
H-BS-4	Human	7	H-BS-4_p1	11828	NI**	-	-
			H-BS-4_p2	5594	pBFO42_2	0,998649	5594
			H-BS-4_p3	7584	pBUN3	0,978386	6783
			H-BS-4_p4	4138	pBFO18_2	0,999281	4137
			H-BS-4_p5	5397	pBFO67_1	0,93222	6129
			H-BS-4_p6	3234	NI**	-	-
			H-BS-4_p7	2483	NI**	-	-
H-BS-5	Human	3	H-BS-5_p1	4981	pBFO67_1	0,908727	6129
			H-BS-5_p2	2756	plasmid unnamed1	0,998058	2604
			H-BS-5_p3	6748	pBUN3	0,970481	6783
H-BS-6	Human	2	H-BS-6_p1	9472	plasmid unnamed2	0,966129	4148
			H-BS-6_p2	2751	plasmid unnamed1	0,998551	2604
H-BS-7	Human	1	H-BS-7_p1	3043	NI**	-	-
H-BS-8	Human	2	H-BS-8_p1	9472	plasmid unnamed2	0,966129	4148
			H-BS-8_p2	2751	plasmid unnamed1	0,998551	2604
H-AC-1	Human	0	-	-	-	-	-
H-AC-2	Human	1	H-AC-2_p1	6197	pRGF00015	0,956268	5934
H-AC-3	Human	4	H-AC-3_p1	4138	pBFO42_2	0,999281	4137
			H-AC-3_p2	2749	plasmid unnamed1	0,998354	2604
			H-AC-3_p3	4138	pBFO18_2	0,999377	4137
			H-AC-3_p4	2817	pBFO67_1	0,932215	6129
H-AC-4	Human	1	H-AC-1_p1	5397	pBFO67_1	0,932408	6129
H-VI-1	Human	1	H-VI-1_p1	33145	p5482	0,99736	33038
M-AC-1	Mouse	2	M-AC-1_p1	4971	pBFO67_1	0,938887	6129
			M-AC-1_p2	4148	plasmid unnamed2	0,999809	4148
M-AC-2	Mouse	3	M-AC-2_p1	4971	pBFO67_1	0,938887	6129
			M-AC-2_p2	4148	plasmid unnamed2	0,999809	4148
			M-AC-2_p3	18567	NI**	-	-
M-AC-3	Mouse	2	M-AC-3_p1	4971	pBFO67_1	0,938887	6129
			M-AC-3_p2	4148	plasmid unnamed2	0,999809	4148
M-FR-1	Mouse	2	M-FR-1_p1	4971	pBFO67_1	0,938887	6129
			M-FR-1_p2	4148	plasmid unnamed2	0,999809	4148
P-AC-1	Pig	-	-	-	-	-	-
P-FR-1	Pig	1	P-FR-1_p1	5921	pBFO67_1	0,981458	6129

*Mash identities in bold indicate that it is the identified plasmid by PLSDB.

**NI = 'Not Identified'

Table 8: Comparison of the gene content encoded on the plasmid for each isolate (n=19)

Protein	H-BS-1	H-BS-2	H-BS-3	H-BS-4	H-BS-5	H-BS-6	H-BS-7	H-BS-8	H-AC-1	H-AC-2	H-AC-3	H-AC-4	H-VI-1	M-AC-1	M-AC-2	M-AC-3	M-FR-1	P-AC-1	P-FR-1
Peptidase family M23													+						
Transglycosylase associated protein															+				
FAD dependent oxidoreductase															+				
YoeB-like toxin of bacterial type II toxin-antitoxin system	+	+	+	+		+		+		+	+			+	+	+	+		
PemK-like, MazF-like toxin of type II toxin-antitoxin system				+							+								
Manganese containing catalase															+				
CRISPR associated protein Cas2					+														
Beta-lactamase enzyme family										+									
Transposase IS66 family			+	+			+												
Putative transposase of IS4/5 family			+																
IS66 Orf2 like protein			+	+			+												
Conjugative transposon, TraM													+						
Conjugative transposon protein TraO													+						
Transposase DDE domain															+				
Fimbrillin-like													+						
Virulence protein RhuM family													+						
OmpA family															+				
Outer membrane protein beta-barrel domain															+				
PLAT/LH2 and C2-like Ca ²⁺ -binding lipoprotein													+						
N-6 DNA Methylase															+				
Phage integrase family				+															
Resolvase, N terminal domain													+						
YtxH-like protein															+				
SprT-like family													+						
Fic/DOC family				+									+						
Two-component regulator propeller															+				

light blue = Metabolism, green= Resistance, orange= mobile elements, light orange= Virulence factors, pink= Membrane interaktions, purple= DNA-Interactions, white= without category

3.8. Antibiotic Resistance Prediction

The identification of antibiotic resistance-associated proteins on a plasmid hint at potential strain-level differences in antimicrobial defence, as investigated recently in within the family *Lachnospiraceae* (Sorbara *et al.*, 2020). To further identify potential differences between the strains, all genomes and plasmids were searched against the CARD database (Table 9). Proteins that may confer antibiotic resistance against antibiotics within the following classes were detected: cephamycin, lincosamide, macrolide, oxazolidinone, phenicol, pleuromutilin, streptogramin and tetracycline.

More than half (n=11) of the isolates contained proteins conferring resistance to tetracyclines resistance (TetQ or Mel). Three out of 19 isolates encoded multiple antibiotic resistance proteins. One of these proteins was Mel which was predicted to confer antibiotic resistance to phenicol, lincosamides, macrolides, oxazolidinones, pleuromutilins and streptogramins by antibiotic target protection. Isolate 'H-BS-4' additionally contained the TetQ protein predicted to confer antibiotic resistance to tetracycline antibiotics by antibiotic target protection. A further isolate 'H-BS-7' contained next to Mel and TetQ a third antibiotic resistance-conferring protein; Mef(En2). This gene was predicted to confer macrolide antibiotic resistance by antibiotic efflux. The isolate 'P-FR-1' contained next to Mel, TetQ and Mef(En2) another antibiotic resistance protein called ErmB. ErmB confers macrolide resistance by antibiotic target alteration similar to ErmF, which was the detected antibiotic resistance-conferring protein in isolate 'H-BS-5'.

Only two isolates contained proteins conferring resistance to cephamycin. The isolate 'H-AC-1' had the CfxA6 protein encoded on the plasmid. Out of the 19 isolates, six isolates did not contain any antibiotic resistance-conferring proteins.

Table 9: Genome-based antibiotic resistance potential of the 19 isolates from this study.

Isolate	Cephamycin	Lincosamide	Macrolide	Oxazolidinone	Phenicol	Pleuromutilin	Streptogramin	Tetracycline			
H-BS-1											
H-BS-2								+	(TetQ)		
H-BS-3	+							+	(TetQ)	(CfxA-CfxA6)	
H-BS-4		+	+	+	+	+	+	+	+	(Mel) (TetQ)	(Mel)
H-BS-5		+	+				+			(ErmF)	(ErmF)
H-BS-6									+	(TetQ)	
H-BS-7		+	+	+	+	+	+	+	+	(Mel) (TetQ)	(Mel) (Mef(En2))
H-BS-8											
H-AC-1											
H-AC-2	(+)*									(CfxA6)	
H-AC-3		+	+	+	+	+	+	+	+	(Mel)	(Mel)
H-AC-4									+	(TetQ)	
H-VI-1											
M-AC-1									+	(TetQ)	
M-AC-2									+	(TetQ)	
M-AC-3									+	(TetQ)	
M-FR-1											
P-AC-1											
P-FR-1		+	+	+	+	+	+	+	+	(Mel) (TetQ)	(ErmB) (Mel) (Mef(En2))

in parenthesis = antibiotic resistance-conferring protein(s) detected

(+)* = antibiotic resistance-conferring protein is encoded on plasmid

3.9. BSH-Assay

The deconjugation of primary bile acids by BSH-enzymes, making them available for further bioconversion by bacteria of the gut microbiome is an important metabolic function as bile acids are key signalling molecules which for instance play a role in the regulation of energy metabolism in the host. Therefore, the presence of BSH activity in the own isolates was investigated (Table 10).

In total, 12 of the 19 isolates were BSH positive. One additional isolate was considered to be only weakly positive due to the agar being less opaque as that for the other positive isolates after six days. Six isolates were BSH-negative. Notably, the growth of

those isolates was strongly reduced or inhibited in the presence of TDCA (Appendix 2, Supp. Table S6).

Table 10: Overview of the presence or absence of a BSH-enzyme in the 19 isolates

Isolate	bsh
H-BS-1	-
H-BS-2	+
H-BS-3	-
H-BS-4	(+)
H-BS-5	+
H-BS-6	+
H-BS-7	+
H-BS-8	+
H-AC-1	+
H-AC-2	-
H-AC-3	-
H-AC-4	-
H-VI-1	+
M-AC-1	+
M-AC-2	+
M-AC-3	+
M-FR-1	+
P-AC-1	+
P-FR-1	-

3.10. Production of SCFAs as measured by HPLC

Short-chain fatty acid (SCFA) analysis was performed because SCFA's are known to play an important role in regulating host responses. Therefore, the production of SCFAs by the 19 isolates was quantified (Figure 17) in triplicate by HPLC after 24h of growth in YCFA medium. We observed similar levels of succinate (16-18 mM) and acetate (10-12mM) produced by all isolates. Notably, none of the isolates produced propionate although *B. vulgatus* is reported to use the succinate pathway to produce propionate (Louis *et al.*, 2014). Whilst lactate was also produced at levels ranging from 1.8-2.9 mM, the production of iso-butyrate was observed for only 11 isolates at levels slightly above the cutoff value of 0.5-0.6 mM.

The reasons underlying the absence of propionate production and low levels of isobutyrate were interesting and need to be investigated.

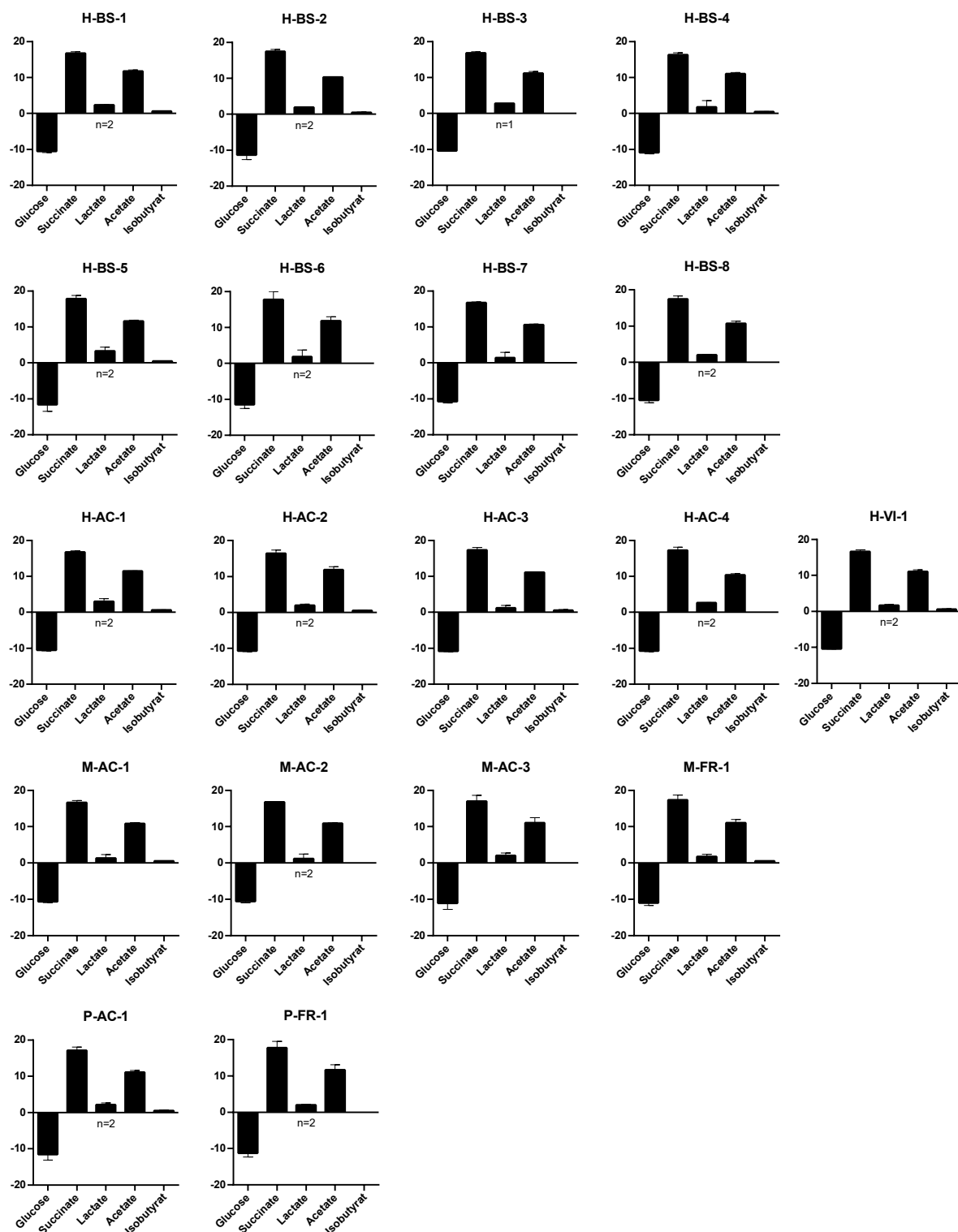


Figure 17: SCFA profiles of the 19 isolates

The SCFA-profiles were generated in triplicates at timepoints t0 (immediately after subculturing) and t24 (after 24h of growth) in YCFA medium. If the cut-off value of 0.5 mM was exceeded, SCFAs were included in the plots. In case of the lactate peak evaluation, a second peak very close to that of lactate made the evaluation difficult, thus when only one or two out of the three measurements were usable, the n-value is given. The human isolates are shown in row 1-3, the mouse isolates in row 4 and the pig isolates in row 5.

3.11. Carbon Source Utilization Analysis with the BIOLOG System

The Carbon source utilization of 95 substrates was tested with the BIOLOG system on nine representative isolates. The representatives were chosen based on their diversity in the *anvi'o* pangenome tree and belonged to different subclades within clade2 and clade1. Only the substrates for which potential differences between the strains were observed are summarized in Table 11.

All tested isolates were positive for the following substrates: D-cellobiose, D-fructose, L-fructose, D-galactose, gentiobiose, α -D-glucose, D-glucose-6-phosphate, α -D-lactose, lactulose, maltose, maltotriose, D-mannose, D-melibiose, 3-methyl-D-glucose, palatinose, L-rhamnose, turanose, glyoxylic acid, α -hydroxybutyric acid, α -ketobutyric acid, α -ketovaleric acid, D, L-lactic acid, D-malic acid, pyruvic acid, pyruvic acid methyl ester, succinamic acid, succinic acid, succinic acid mono-methyl ester.

They were negative for the following substrates: adonitol, D-arabitol, dulcitol, i-erythritol, glycerol, D, L- α -glycerol-phosphate, Myo-inositol, D-mannitol, α -methyl-D-glucoside, β -methyl-D-glucoside, salicin, acetic acid, formic acid, fumaric acid, β -hydroxybutyric acid, itaconic acid, propionic acid, D-saccharic acid, glycyl-L-methionine, glycyl-L-proline, L-phenylalanine and L-serin.

Interestingly, the isolates 'H-AC-4', 'M-AC-2', 'M-FR-1' and 'P-FR-1' could not utilize α -/ β -cyclodextrin and dextrin whilst the 'H-BS-', 'H-BS-2', 'H-BS-5', 'H-BS-7' and 'H-VI-1' could do. For the sugars stachyose, sucrose and D-trehalose, only the isolates 'H-BS-1', 'H-BS-2', 'H-BS-5' and 'H-VI-1' were positive.

A potential host-specific pattern could be observed related to the utilization of amino acids and their modified version (Table 11). The pig isolate 'P-FR-1' was negative for all corresponding 24 substrates whilst the human and mouse isolates could utilize at least two out of the 24 substrates.

Table 11: Summary of strain-level differences in carbon source utilization

Positive Isolates	Substrate	H-BS-1	H-BS-2	H-BS-5	H-BS-7	H-VI-1	H-AC-4	M-AC-2	M-FR-1	P-FR-1
8	Arbutin	+++	+	+++	+++	-	+++	+++	+++	+++
8	α -D-Glucose-1-Phosphate	+++	+++	+++	+++	+++	-	++	+++	+
8	Urocanic Acid	+++	+++	++	+++	+++	-	+++	+++	+
7	D-Galacturonic Acid	++	-	++	-	+++	+++	+++	+++	+++
7	L-Malic Acid	++	+++	+++	+++	+++	-	+	-	+
6	N-Acetyl-D-Galactosamine	+	+++	+++	+	+++	-	+	-	-
6	N-Acetyl-D-Glucosamine	+	+++	+++	+	+++	-	+	-	-
6	N-Acetyl- β -D-Mannosamine	+	++	+++	+	+++	-	+	-	-
5	α -Cyclodextrin	+	+++	+++	+	+	-	-	-	-
5	β -Cyclodextrin	++	+++	+++	+	+	-	-	-	-
5	Dextrin	+	+++	+++	+	+++	-	-	-	-
5	D-Raffinose	+	++	+++	-	+++	-	-	-	-
5	L-Alanyl-L-Histidine	+++	++	+++	+++	+++	-	-	-	-
5	Thymidine	+	++	+++	-	+++	-	+	-	-
4	β -Methyl-D-Galactoside	++	+++	+++	-	+++	-	+	-	-
4	Stachyose	+	++	+++	-	+++	-	-	-	-
4	Sucrose	++	+++	+++	-	+++	-	-	-	-
4	D-Trehalose	+	++	++	-	+	-	-	-	-
4	m-Tartaric Acid	+	+	-	+	+	-	-	-	-
4	L-Glycyl-L-Aspartic Acid	+	++	+++	-	+++	-	-	-	-
4	2'-Deoxy Adenosine	+	++	+++	-	+++	-	-	-	-
4	Uridine	-	++	+++	-	+++	-	+	-	-
3	Alaninamide	-	++	++	-	++	-	-	-	-
3	L-Alanine	+	++	+++	-	+++	-	-	-	-
3	L-Alanyl-L-Glutamine	+	-	+++	-	+++	-	-	-	-
3	L-Alanyl-L-Threonine	-	-	+++	-	+++	-	+	-	-
3	L-Asparagine	+	-	+++	-	+	-	-	-	-
3	L-Valine plus L-Aspartic Acid	+	-	++	-	+++	-	-	-	-
3	Uridine-5'-Monophosphate	-	++	+++	-	+++	-	-	-	-
2	D-Gluconic Acid	+	-	-	+	-	-	-	-	-
2	L-Glutamic Acid	+	-	+++	-	-	-	-	-	-
2	L-Glutamine	-	-	+++	-	+	-	-	-	-
2	Glycyl-L-Glutamine	-	-	+++	-	+	-	-	-	-
2	L-Theonine	+	-	+++	-	-	-	-	-	-
2	Inosine	-	-	+++	-	+++	-	-	-	-
2	Thymidine-5'-Mono-phosphate	-	-	+++	-	++	-	-	-	-
1	Amygdalin	-	-	++	-	-	-	-	-	-
1	D-Gluconic Acid	+	-	-	+	-	-	-	-	-
1	D-Glucosaminic Acid	+	-	-	-	-	-	-	-	-
1	D-Melezitose	-	+	-	-	-	-	-	-	-
1	α -Methyl-D-Galactoside	-	-	+++	-	-	-	-	-	-
1	D-Sorbitol	-	-	-	-	+	-	-	-	-
1	L-Methionine	-	-	+	-	-	-	-	-	-
1	L-Valine	-	-	+	-	-	-	-	-	-

dark grey (+++): positive on all three BIOLOG plates; medium grey(++): positive on two plates

light grey(+): positive on one plate only

3.12. Comparison between *in silico* Predicted and *in vitro* tested Phenotypes

Since the number of sequenced genomes is growing exponentially, phenotype prediction software becomes very valuable. In this study, the *in silico* predicted phenotypes were compared to the confirmed *in vitro* phenotypes in order to gain insights into the accuracy of those predictions.

The utilization of D-mannose, galactose, glycerol, lactose, L-rhamnose, maltose, melibiose and Myo-inositol as carbon sources were correctly predicted for all isolates tested. However, the ability to metabolize cellobiose and D-mannitol as carbon sources showed conflicting results for all tested isolates. Whilst utilization of cellobiose was predicted to be negative *in silico* but tested positive *in vitro* across all isolates, the utilization of D-mannitol was predicted to be positive (Figure 16) but was negative *in vitro*. The genome-based prediction of the presence of BSH failed for the isolates 'H-BS-2', 'H-BS-7', 'H-VI-1', 'M-AC-2', 'M-FR-1' as it was predicted negative (Figure 16) but positive *in vitro* (Table 10).

The utilization of raffinose, sucrose and trehalose as carbon sources showed conflicting results for the isolates 'H-BS-7', 'H-AC-4', 'M-AC-2', 'M-FR-1' and 'P-FR-1'. These isolates were predicted positive for the utilization of those sugars (Figure 16) but tested negative *in vitro* (Table 11).

Table 12: Summary of the *in silico* predicted phenotypes compared to *in vitro* results

Phenotype	H-BS-1	H-BS-2	H-BS-5	H-BS-7	H-VI-1	H-AC-4	M-AC-2	M-FR-1	P-FR-1
bsh									
Cellobiose									
D-Mannitol									
D-Mannose									
D-Sorbitol									
Galactose									
Glycerol									
Lactose									
L-Rhamnose									
Maltose									
Melibiose									
Myo-Inositol									
Raffinose									
Salicin									
Sucrose									
Trehalose									

green: matched; red: *in silico* negative, *in vitro* positive; orange: *in silico* positive, *in vitro* negative

4. Discussion

In recent years, high genomic diversity both between and within bacterial species has been identified via sequencing studies of isolates and large metagenomic datasets (Poyet *et al.*, 2019; Tett *et al.*, 2019). Even though members of the genus *Bacteroides* are common and dominant members of mammalian gut microbiomes and have also been linked to the development of several non-communicable diseases, our knowledge about these species is still limited. Certain species within the *Bacteroides* genus have been studied in greater details. For instance, interactions between *B. fragilis* and the immune systems have been studied extensively and *B. thetaiotaomicron* is well known for its ability to degrade a wide range complex sugars (Martens *et al.*, 2008; Xu *et al.*, 2003). Although *B. vulgatus* is quite prevalent in the mammalian gut, however, the metabolic and functional diversity within this species is largely unexplored. Here, we studied strain-level diversity within this species by combining *in silico* predictions and *in vitro* testing.

As expected, *B. vulgatus* showed a strong association with gut environments in several host species based on 16S rRNA gene amplicon data. This was also observed in the *Bacteroides*-directed cultivation in which *B. vulgatus* was present across all three mammalian host species whilst it was found in 50% of the human samples. This contrasts with the prevalence of 77 % from the sequence-based analysis. However, the presence of *B. vulgatus* was originally detected by MALDI-TOF MS in all human samples except one but not all isolates were confirmed by 16S gene sequencing due to the lack of time. The isolates from Human subject 3 and 5 turned out to be *B. dorei* after sequencing. Taking these points into account may explain the lower prevalence of *B. vulgatus* after cultivation vs. sequencing. In contrast to human, the isolation of *B. vulgatus* from the pig or mouse samples were less frequent. In the pig, *B. vulgatus* was only present in 25 % of the samples (taking the lost isolate into account); for mice, the cultivation based prevalence was 20 %.

Strikingly, this study revealed 21 unique clusters of genes within 4 distinct phylogenetic clades of strains. Parallels can be drawn to a study on *P. copri*, where this species was split into 4 distinct clusters (Tett *et al.*, 2019), suggesting that *B. vulgatus* is also not a monotypic species.

The additional six *Bacteroides* genomes used, namely *B. acidifaciens*, *B. dorei*, *B. fragilis*, *B. ovatus*, *B. thetaiotaomicron* and *B. xylanisolvens*, allowed the identification of the core pangenome of *B. vulgatus*. This pangenome was actually shared by *B.*

dorei, suggesting that these two species are very close to each other compared with other known *Bacteroides*. This was also observed when filtering the publicly available datasets of *B. vulgatus* genomes (Forster *et al.*, 2019; Poyet *et al.*, 2019). After initial ANI-based filtering (>95%) against the type strain of *B. vulgatus*, *B. dorei* isolates remained in the dataset. Further Digital-DNA-DNA-hybridization (dDDH) analysis confirmed this result and allowed confident exclusion of those isolates from further analyses. These findings clearly indicate that *B. vulgatus* and *B. dorei* represent an exception to established procedures for species delineation.

The finding of a shared core pangenome for *B. dorei* and *B. vulgatus* is in line with a Bacteroidetes taxonomic analysis study on 1,000 type-strain genomes, which recommended a reclassification of *B. dorei* and *B. vulgatus* to *Phocaeicola dorei* and *Phocaeicola vulgatus* (García-López *et al.*, 2019). Please note that these new combinations of names have been validated during the course of this master thesis. We acknowledge this new status of valid names but kept on referring to *B. vulgatus* throughout the thesis.

The observation of many CAZymes being assigned to unique gene clusters specific to certain clades of strains (e.g. cluster1b sub1c mouse, cluster4a and cluster4b) led us to investigate CAZyme profiles. We found strain-level differences of (max. 10%) in terms of CAZyme numbers and identified specific groups of strains based CAZyme family profiles.

Mouse strains were identified to lack the two carbohydrate-binding modules 2 and 35 which are associated with xylan (Bolam *et al.*, 2004; Simpson *et al.*, 1999) indicating that the mouse isolates might not be able to degrade xylan. This was however not yet tested *in vitro*.

The absence of glycoside hydrolase family 27 in all mouse strains and the pig isolate 'P-FR-1', which contains α -D-galactosidases, known to cleave α -linked galactose residues from carbohydrates such as melibiose, raffinose, stachyose, sucrose and trehalose (Meier and Reid, 1982) would explain the incapacity of those strains to utilize these sugars in the Biolog assays. Notably, these findings show that connections can be drawn from the presence or absence of CAZyme families to confirmed *in vitro* phenotypes. However, in-depth investigations using a greater number of isolates is required, as previously for the species *Bifidobacterium longum* (Kujawska *et al.*, 2020). Next to CAZyme analysis, the potential for antibiotic resistance was predicted by searching for the presence of proteins within the CARD database in the genomes of

own isolates. The 19 *B. vulgatus* strains revealed differences in the content of antibiotic resistance-conferring proteins. In total four of the 19 isolates were predicted to possess resistance against a broad range of antibiotic classes. Mel, a protein belonging to the ATP-binding cassette (ABC)-F subfamily, was frequently detected within the genomes. ABC-F proteins consist of a single polypeptide containing two ATP-binding cassette domains separated by a linker of ~80 amino acids (Sharkey *et al.*, 2016) and confer resistance to the majority of antibacterial drug classes that target the large (50S) subunit of the bacterial ribosome, including the ketolides (Reynolds and Cove, 2005), lincosamides (Novotna and Janata, 2006; Singh *et al.*, 2002), macrolides (Ross *et al.*, 1990), oxazolidinones (Wang *et al.*, 2015), phenicols (Wang *et al.*, 2015), pleuromutilins (Gentry *et al.*, 2008), and streptogramins of group A (Singh *et al.*, 2002) and B (Ross *et al.*, 1990).

So far known from the literature, no individual ABC-F determinant (e.g. Mel) confers resistance to all of the listed classes. Three phenotypic resistance profiles are distinguished while the gene types *vga-*, *lsa-*, and *sal* (LS_A or LS_{AP} phenotype) confer resistances to lincosamide, streptogramin A and sometimes pleuromutilin (Hot *et al.*, 2014; Ross *et al.*, 1990; Tessé *et al.*, 2013).

The *msr*-gene types including *mel* (MS_B phenotype) confer resistances to macrolides, group B streptogramins and sometimes ketolides (Reynolds and Cove, 2005; Ross *et al.*, 1990).

The resistance to phenicols and oxazolidinones is conferred by the *optrA* gene (Wang *et al.*, 2015). Furthermore, Mel alone was found to be capable of conferring macrolide resistance in a susceptible *Streptococcus pneumoniae* strain by transformation but did not fully restore the minimum inhibitory concentration (MIC) resistance of the donor strain (Daly *et al.*, 2004). In the co-presence of MefE, both proteins synergize (Ambrose *et al.*, 2005). In this study, the Mef(En2) protein, a MefE homolog in the *Bacteroides* species, was identified in the isolates 'H-BS-7' and 'P-FR-1' suggesting that their macrolide resistance will be higher compared to those that can only express the Mel protein.

Considering this, the CARD-based predictions cover the antibiotic resistances of all three known phenotypes (LS_A or LS_{AP}, MS_B and PO) whereas the presence of *mel* could only confer one of those phenotypes. This highlights the importance of literature-based research and *in vitro* testing as prediction software and databases are required to improve.

In the isolate 'P-FR-1', the ErmB protein was detected, a methylase which catalyzes the demethylation of a single adenine in 23S rRNA leading to high-level cross-resistance due to the overlapping binding sites of macrolides, lincosamide and streptogramin B (MLS_B) antibiotics (Weisblum, 1995). Therefore, the isolate 'P-FR-1' is likely to have a high-level resistance to those antibiotics. The ErmF protein, another 23S rRNA methyltransferase which confers the MLS_B phenotype (Kangaba *et al.*, 2015) was detected in the isolate 'H-BS-5'. The *erm* gene presence is often associated with the *tetQ* gene on tetracycline resistance conjugative transposons (Wang *et al.*, 2000). The isolate 'H-BS-5' was also predicted to contain the TetQ protein (Leng *et al.*, 1997; Tribble *et al.*, 2010) which is known to confer tetracycline resistance (Table 9). This isolate potentially will show high-level resistances to tetracycline antibiotics and the MLS_B antibiotics. Whereas further seven isolates only contained the TetQ protein indicating that those isolates will show resistance to tetracycline antibiotics.

The isolate 'H-BS-4' was predicted to contain 6 CfxA proteins, which encode class A β -lactamases found in *B. vulgatus* (Parker and Smith, 1993). The isolate 'H-AC-2' had the genetic information for such a protein encoded on its plasmid. The predicted presence of this beta-lactamase suggests resistance to a broad range of beta-lactam antibiotics.

As mentioned above, prediction software is still required to further improve which was also observed in our comparison of the *in silico* predicted and *in vitro* confirmed phenotypes. Those revealed contradictions with respect to BSH activity and in some cases the utilization of D-sorbitol, raffinose, sucrose and trehalose. Whilst the CAZyme analysis already exposed that the mouse isolates and the pig isolate might lack an urgently needed group of enzymes for utilization of raffinose, sucrose and trehalose, it is possible that traitar was not able to identify that those isolates lack enzymes which are required for the complete pathway. This could be because the classifiers are trained on datasets containing *Bacteroides* regardless the completeness of the pathway. The prediction of the utilization of D-mannitol needs to be further investigated since *B. vulgatus* was reported to be unable to utilize mannitol in a carbon-free mannitol medium (Tan *et al.*, 2018). The results regarding the D-sorbitol utilization of the isolate 'H-VI-1' and the utilization of cellobiose across all the isolates was another observed contradictory result.

B. vulgatus was known to be unable to utilize cellobiose (McCarthy *et al.*, 1988) and was also predicted to be unable to utilize cellobiose in our study. D-sorbitol utilization

of the isolate 'H-VI-' appeared only on one of the three, which could be due to a false positive signal. The *Bacteroides* spp. are known to produce a high background signal in the Biolog system whereby a false positive signal could easily be achieved. According to the manufacturer, the cell count is crucial for achieving good results. However, the adjustment of the cell density in the AN inoculating fluid (AN-IF)-solution is done via visual comparison of the turbidity of the AN-IF-solution against a turbidity standard by eye which is approximate. This could be avoided by the use of the Biolog turbidimeter which ensures a standardized turbidity based measurement of cell densities in the AN-IF-solution.

In the case of BSH, the prediction might have failed due to high levels of sequence homology with the closely related penicillin-V acylase (PVA) enzymes, which hydrolyze penicillin-V (O'Flaherty *et al.*, 2018). This could lead to a wrong assignment of eggNOGs which ended in a false negative prediction of the presence of this enzyme. The plasmid repertoire was found to differ for the 19 isolates while several contained only one to two plasmids, others were found to have up to 7 plasmids. The plasmids pBFO18_2 and pBFO42_2 were both identified in *B. fragilis* strains (NZ_CP036544 and NZ_CP036552). The unnamed_plasmid1 was identified in a *B. xylanisolvens* strain (NZ_CP041232). Whilst the plasmid_unnamed2 was identified in a *Citrobacter portucalensis* strain (NZ_CP044099), the plasmid p5482 was identified in *B. thetaiotaomicron* (NC_004703). This accounts for an active horizontal gene transfer between *B. vulgatus* and other species present in the gut. The analysis of the plasmid content revealed that many isolates contained a YoeB-like toxin. An *S. pneumoniae* expressed YoeB was shown to severe inhibition of cell growth and a reduction in cell viability of *E. coli* (Nieto *et al.*, 2007). This may imply a mechanism *B. vulgatus* is using to compete against in the gut present *E. coli* strains.

The measurement of SCFA showed that none of the isolates produced propionate although *B. vulgatus* has been previously reported to use the succinate pathway to convert succinate in propionate (Louis *et al.*, 2014). Propionate production via the succinate pathway can be affected by environmental conditions. At high pCO₂ and high dilution rates, succinate accumulates in cultures of *Bacteroides* spp. (Caspari and Macy, 1983). Conversion of succinate to propionate also requires vitamin B12 and succinate has been shown to accumulate in B12-depleted cultures of *Prevotella ruminicola* (Strobel, 1992). As the growth media was supplemented with a vitamin

solution containing vitamin B12, this is unlikely to be the cause, suggesting environmental conditions may be the driver of succinate accumulation. The few isolates that were positive for iso-butyrate, produced only minor amounts that were slightly above the cut-off value of 0.5 mM. Hence, confirmation will be needed prior to drawing conclusion on strain-level differences. The production of acetate and succinate were confirmed by the literature (Gutiérrez and Garrido, 2019; Louis and Flint, 2017).

A combination of bioinformatic approaches and experimental techniques was used to assess the potential strain-level genomic variations and phenotypic abilities of the gut-derived *B. vulgatus* isolated across different mammalian hosts. However, this study is not without limitations. One important caveat is the small number of *B. vulgatus* isolates (n=19) available for analysis. Inclusion of additional strains of pig or mouse origin could contribute to a further investigation on the inter-host individual diversity of *B. vulgatus* and the reliability of the study. Moreover, the inclusion of additional human strains could aid in finding representative isolates for each identified clade or even subclade. The majority of isolates belonged primarily to clade 2, which prohibits the analysis of clades not represented by own isolates. Furthermore, phenotypic investigation of carbohydrate metabolism properties of the representative *B. vulgatus* isolates revealed inconsistencies in growth even for a given strain. Those inconsistencies need to be further investigated by optimizing the Biolog system via the use of another approach. The investigation of CAZyme-associated metabolic functions was only done within the mouse and pig strains and between them. An in-depth investigation of the human samples is required and could contribute to further interesting observations between individual strain diversity. The small addition of only six additional *Bacteroides* species leads to the problem that a *Bacteroides* core pangenome could not be identified. Inclusion of all *Bacteroides* species with a valid name and even isolates belonging to the newly proposed genus *Phocaeicola* could aid in understanding the phylogenomy and functional diversity of *B. vulgatus*.

5. Conclusion & Outlook

The current work identified distinct strain-level clades within the dominant gut bacterial species *B. vulgatus*. Strain variations in carbon source utilization predicted antibiotic resistance and BSH activity were also studied. By combining *in silico* prediction of CAZymes with *in vitro* testing of specific carbohydrate utilization, variability between strains was confirmed

Future studies should include additional genomes of the genera *Bacteroides* and *Phocaeicola* for a comprehensive comparison, as well as greater numbers of isolates from each mammalian host and strains from the different (sub)clades identified. The *B. vulgatus* isolates should be tested for their antibiotic resistance in the lab for accurate determination of the predicted resistances. The differences observed in CAZyme machineries should be further investigated between all isolates including a cluster-based analysis as well as *in vitro* metabolic differences between isolates based on the screening of single substrates using a minimal medium. Altogether, the field will greatly benefit from a comprehensive analysis of how genome-based variations translate into real functional differences between strains.

Acknowledgement

I want to thank the following people who supported me during the work on this thesis:

Prof. Dr. Clavel for providing me the opportunity to do my master thesis in his research group and for his constant support and advice.

All members in the Clavel lab for the pleasant atmosphere and the possibility to ask questions to everyone at any time. Special thanks go to my supervisors, Afrizal and Thomas Hitch who always guided and supported me patiently and to Esther Wortmann for the help with measurements of short-chain fatty acids.

Furthermore, I want to express my gratitude to Dr. Martin Zimmermann for accepting to act as the first examiner of this thesis.

Finally, I would like to thank the collaboration partners from Braunschweig and Vienna for providing additional isolates.

References

- Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. *Nature* 2019;568(7753):499–504. <https://doi.org/10.1038/s41586-019-0965-1>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* 1990;215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Ambrose KD, Nisbet R, Stephens DS. Macrolide efflux in *Streptococcus pneumoniae* is mediated by a dual efflux pump (mel and mef) and is erythromycin inducible. *Antimicrobial agents and chemotherapy* 2005;49(10):4203–9. <https://doi.org/10.1128/AAC.49.10.4203-4209.2005>.
- Andy Bunn, Mikko Korpela, Franco Biondi, Filipe Campelo, Pierre Mérian, Fares Qeadan, *et al.* dplR: Dendrochronology Program Library in R, 2020. <https://CRAN.R-project.org/package=dplR>.
- Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics (Oxford, England)* 2016;32(22):3380–7. <https://doi.org/10.1093/bioinformatics/btw493>.
- Anton H. Elementary linear algebra. 7th ed. New York: Wiley; 1994.
- Arnold JW, Roach J, Azcarate-Peril MA. Emerging Technologies for Gut Microbiome Research. *Trends in microbiology* 2016;24(11):887–901. <https://doi.org/10.1016/j.tim.2016.06.008>.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473(7346):174–80. <https://doi.org/10.1038/nature09944>.
- Bäckhed F, Fraser CM, Ringel Y, Sanders ME, Sartor RB, Sherman PM, et al. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell host & microbe* 2012;12(5):611–22. <https://doi.org/10.1016/j.chom.2012.10.012>.
- Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. *Journal of microbiological methods* 2003;55(3):541–55. <https://doi.org/10.1016/j.mimet.2003.08.009>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
- Becton, Dickinson and Company. GEBRAUCHSANWEISUNG –GEBRAUCHSFERTIGE PLATTENMEDIEN: BD Bacteroides Bile Esculin Agar with Amikacin, 2003. <https://legacy.bd.com/europe/regulatory/Assets/IFU/HB/CE/PA/DE-PA-254480.pdf> (accessed April 01, 2020).
- Benedict MN, Henriksen JR, Metcalf WW, Whitaker RJ, Price ND. ITEP: an integrated toolkit for exploration of microbial pan-genomes. *BMC genomics* 2014;15:8. <https://doi.org/10.1186/1471-2164-15-8>.

- Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, et al. Microbiome definition revisited: old concepts and new challenges. *Microbiome* 2020;8(1):103. <https://doi.org/10.1186/s40168-020-00875-0>.
- Berger SA. GIDEON: a comprehensive Web-based resource for geographic medicine. *International journal of health geographics* 2005;4(1):10. <https://doi.org/10.1186/1476-072X-4-10>.
- BIOLOG Inc. Anaerobe Identification Test Panel, 2007. <https://www.biolog.com/wp-content/uploads/2020/04/00A-006rB-AN-Sell-Sheet-Jul07.pdf> (accessed June 18, 2020).
- Bochner BR, Gadzinski P, Panomitros E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome research* 2001;11(7):1246–55. <https://doi.org/10.1101/gr.186501>.
- Bolam DN, Xie H, Pell G, Hogg D, Galbraith G, Henrissat B, et al. X4 modules represent a new family of carbohydrate-binding modules that display novel properties. *The Journal of biological chemistry* 2004;279(22):22953–63. <https://doi.org/10.1074/jbc.M313317200>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* 1957;27(4):325–49. <https://doi.org/10.2307/1942268>.
- Bryan LE, Kowand SK, van den Elzen HM. Mechanism of aminoglycoside antibiotic resistance in anaerobic bacteria: *Clostridium perfringens* and *Bacteroides fragilis*. *Antimicrobial agents and chemotherapy* 1979;15(1):7–13. <https://doi.org/10.1128/aac.15.1.7>.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature methods* 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
- Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner 2014.
- Caspari D, Macy JM. The role of carbon dioxide in glucose metabolism of *Bacteroides fragilis*. *Archives of microbiology* 1983;135(1):16–24. <https://doi.org/10.1007/BF00419476>.
- Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* 2012;149(7):1578–93. <https://doi.org/10.1016/j.cell.2012.04.037>.
- Church GM. Genomes for all. *Scientific American* 2006;294(1):46–54. <https://doi.org/10.1038/scientificamerican0106-46>.
- Cotta M, Forster R. The Family Lachnospiraceae, Including the Genera *Butyrivibrio*, *Lachnospira* and *Roseburia*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. *The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria*. New York, NY: Springer US; 2006. p. 1002–1021.

- Daly MM, Doktor S, Flamm R, Shortridge D. Characterization and prevalence of MefA, MefE, and the associated msr(D) gene in *Streptococcus pneumoniae* clinical isolates. *Journal of clinical microbiology* 2004;42(8):3570–4. <https://doi.org/10.1128/JCM.42.8.3570-3574.2004>.
- Dashkevich MP, Feighner SD. Development of a differential medium for bile salt hydrolase-active *Lactobacillus* spp. *Applied and Environmental Microbiology* 1989;55(1):11–6.
- Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 2018;6:e4320. <https://doi.org/10.7717/peerj.4320>.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)* 2005;308(5728):1635–8. <https://doi.org/10.1126/science.1110591>.
- Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome informatics. International Conference on Genome Informatics* 2009;23(1):205–11.
- Eddy SR. Accelerated Profile HMM Searches. *PLoS computational biology* 2011;7(10):e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
- Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 2015;3:e1319. <https://doi.org/10.7717/peerj.1319>.
- Erich Neuwirth. RColorBrewer: ColorBrewer Palettes, 2014. <https://CRAN.R-project.org/package=RColorBrewer>.
- Feldbauer R, Schulz F, Horn M, Rattei T. Prediction of microbial phenotypes based on comparative genomics. *BMC bioinformatics* 2015;16 Suppl 14:S1. <https://doi.org/10.1186/1471-2105-16-S14-S1>.
- Filippis F de, Pasolli E, Tett A, Tarallo S, Naccarati A, Angelis M de, et al. Distinct Genetic and Functional Traits of Human Intestinal *Prevotella copri* Strains Are Associated with Different Habitual Diets. *Cell host & microbe* 2019;25(3):444-453.e3. <https://doi.org/10.1016/j.chom.2019.01.004>.
- Filippo C de, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107(33):14691–6. <https://doi.org/10.1073/pnas.1005963107>.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic acids research* 2014;42(Database issue):D222-30. <https://doi.org/10.1093/nar/gkt1223>.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* 2011;39(Web Server issue):W29-37. <https://doi.org/10.1093/nar/gkr367>.
- Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nature biotechnology* 2019;37(2):186–92. <https://doi.org/10.1038/s41587-018-0009-7>.

- Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104(34):13780–5. <https://doi.org/10.1073/pnas.0706625104>.
- Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, et al. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS genetics* 2011;7(2):e1001314. <https://doi.org/10.1371/journal.pgen.1001314>.
- Galata V, Fehlmann T, Backes C, Keller A. PLSDb: a resource of complete bacterial plasmids. *Nucleic acids research* 2019;47(D1):D195-D202. <https://doi.org/10.1093/nar/gky1050>.
- García-López M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, Kyrpides NC, et al. Analysis of 1,000 Type-Strain Genomes Improves Taxonomic Classification of Bacteroidetes. *Frontiers in microbiology* 2019;10:2083. <https://doi.org/10.3389/fmicb.2019.02083>.
- Gentry DR, McCloskey L, Gwynn MN, Rittenhouse SF, Scangarella N, Shawar R, et al. Genetic characterization of Vga ABC proteins conferring reduced susceptibility to pleuromutilins in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* 2008;52(12):4507–9. <https://doi.org/10.1128/AAC.00915-08>.
- Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Applied and Environmental Microbiology* 1997;63(7):2802–13.
- Grill J, Schneider F, Crociani J, Ballongue J. Purification and Characterization of Conjugated Bile Salt Hydrolase from *Bifidobacterium longum* BB536. *Applied and Environmental Microbiology* 1995;61(7):2577–82. <https://doi.org/10.1128/AEM.61.7.2577-2582.1995>.
- Gutiérrez N, Garrido D. Species Deletions from Microbiome Consortia Reveal Key Metabolic Interactions between Gut Microbes. *mSystems* 2019;4(4). <https://doi.org/10.1128/mSystems.00185-19>.
- Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome medicine* 2017;9(1):103. <https://doi.org/10.1186/s13073-017-0490-5>.
- Hasan N, Yang H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* 2019;7:e7502. <https://doi.org/10.7717/peerj.7502>.
- Hedberg M, Nord CE. Antimicrobial susceptibility of *Bacteroides fragilis* group isolates in Europe. *Clinical microbiology and infection the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2003;9(6):475–88. <https://doi.org/10.1046/j.1469-0691.2003.00674.x>.
- Hooper LV, Midtvedt T, Gordon JI. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annual review of nutrition* 2002;22:283–307. <https://doi.org/10.1146/annurev.nutr.22.011602.092259>.

- Hoskins LC, Agustines M, McKee WB, Boulding ET, Kriaris M, Niedermeyer G. Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *Journal of Clinical Investigation* 1985;75(3):944–53. <https://doi.org/10.1172/JC1111795>.
- Hot C, Berthet N, Chesneau O. Characterization of sal(A), a novel gene responsible for lincosamide and streptogramin A resistance in *Staphylococcus sciuri*. *Antimicrobial agents and chemotherapy* 2014;58(6):3335–41. <https://doi.org/10.1128/AAC.02797-13>.
- Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research* 2016;44(D1):D286-93. <https://doi.org/10.1093/nar/gkv1248>.
- Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012;486(7402):215–21. <https://pubmed.ncbi.nlm.nih.gov/22699610>. <https://doi.org/10.1038/nature11209>.
- Hungate RE. The anaerobic mesophilic cellulolytic bacteria. *Bacteriological Reviews* 1950;14(1):1–49.
- Hungate RE, Smith W, Clarke RT. Suitability of butyl rubber stoppers for closing anaerobic roll culture tubes. *Journal of Bacteriology* 1966;91(2):908–9.
- Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486(7402):207–14. <https://doi.org/10.1038/nature11234>.
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* 2010;11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, et al. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science (New York, N.Y.)* 2013;342(6161):967–70. <https://doi.org/10.1126/science.1240527>.
- Jaccard P. Lois de distribution florale dans la zone alpine 1902. <https://doi.org/10.5169/SEALS-266762>.
- Jacobson A, Lam L, Rajendram M, Tamburini F, Honeycutt J, Pham T, et al. A Gut Commensal-Produced Metabolite Mediates Colonization Resistance to Salmonella Infection. *Cell host & microbe* 2018;24(2):296-307.e7. <https://doi.org/10.1016/j.chom.2018.07.002>.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications* 2018;9(1):5114. <https://doi.org/10.1038/s41467-018-07641-9>.
- Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, et al. *vegan: Community Ecology Package*, 2019. <https://CRAN.R-project.org/package=vegan>.

- Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic acids research* 2017;45(D1):D566-D573. <https://doi.org/10.1093/nar/gkw1004>.
- Justesen US, Holm A, Knudsen E, Andersen LB, Jensen TG, Kemp M, et al. Species identification of clinical isolates of anaerobic bacteria: a comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry systems. *Journal of clinical microbiology* 2011;49(12):4314–8. <https://doi.org/10.1128/JCM.05788-11>.
- Kangaba AA, Saglam FY, Tokman HB, Torun M, Torun MM. The prevalence of enterotoxin and antibiotic resistance genes in clinical and intestinal *Bacteroides fragilis* group isolates in Turkey. *Anaerobe* 2015;35(Pt B):72–6. <https://doi.org/10.1016/j.anaerobe.2015.07.008>.
- Karas M, Bachmann D, Bahr U, Hillenkamp F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes* 1987;78:53–68. [https://doi.org/10.1016/0168-1176\(87\)87041-6](https://doi.org/10.1016/0168-1176(87)87041-6).
- Kelly D, Conway S, Aminov R. Commensal gut bacteria: mechanisms of immune modulation. *Trends in immunology* 2005;26(6):326–33. <https://doi.org/10.1016/j.it.2005.04.008>.
- Klieve AV, Bauchop T. Morphological diversity of ruminal bacteriophages from sheep and cattle. *Applied and Environmental Microbiology* 1988;54(6):1637–41.
- Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102(7):2567–72. <https://doi.org/10.1073/pnas.0409727102>.
- Kujawska M, La Rosa SL, Roger LC, Pope PB, Hoyles L, McCartney AL, et al. Succession of *Bifidobacterium longum* strains in response to a changing early life nutritional environment reveals dietary substrate adaptations. *iScience* 2020:101368. <https://doi.org/10.1016/j.isci.2020.101368>.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* 2018;35(6):1547–9. <https://doi.org/10.1093/molbev/msy096>.
- Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, et al. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Scientific reports* 2016a;6:33721. <https://doi.org/10.1038/srep33721>.
- Lagkouvardos I, Lesker TR, Hitch TCA, Gálvez EJC, Smit N, Neuhaus K, et al. Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome* 2019;7(1):28. <https://doi.org/10.1186/s40168-019-0637-2>.
- Lagkouvardos I, Pukall R, Abt B, Foesel BU, Meier-Kolthoff JP, Kumar N, et al. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nature microbiology* 2016b;1(10):16131. <https://doi.org/10.1038/nmicrobiol.2016.131>.

- Leng Z, Riley DE, Berger RE, Krieger JN, Roberts MC. Distribution and mobility of the tetracycline resistance determinant tetQ. *The Journal of antimicrobial chemotherapy* 1997;40(4):551–9. <https://doi.org/10.1093/jac/40.4.551>.
- Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. *Science (New York, N.Y.)* 2008a;320(5883):1647–51. <https://doi.org/10.1126/science.1155725>.
- Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature reviews. Microbiology* 2008b;6(10):776–88. <https://doi.org/10.1038/nrmicro1978>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* 2014;42(Database issue):D490-5. <https://doi.org/10.1093/nar/gkt1178>.
- Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. *Environmental microbiology* 2017;19(1):29–41. <https://doi.org/10.1111/1462-2920.13589>.
- Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nature reviews. Microbiology* 2014;12(10):661–72. <https://doi.org/10.1038/nrmicro3344>.
- Martens EC, Chiang HC, Gordon JI. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell host & microbe* 2008;4(5):447–57. <https://doi.org/10.1016/j.chom.2008.09.007>.
- McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy* 2013;57(7):3348–57. <https://doi.org/10.1128/AAC.00419-13>.
- McCarthy RE, Pajeau M, Salyers AA. Role of starch as a substrate for *Bacteroides vulgatus* growing in the human colon. *Applied and Environmental Microbiology* 1988;54(8):1911–6. <https://doi.org/10.1128/AEM.54.8.1911-1916.1988>.
- Meier H, Reid JSG. Reserve polysaccharides other than starch in higher plants; 1982.
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics* 2013;14:60. <https://doi.org/10.1186/1471-2105-14-60>.
- Moore WEC, STACKEBRANDT E, Kandler O, Colwell RR, Krichevsky MI, Truper HG, et al. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International journal of systematic and evolutionary microbiology* 1987;37(4):463–4. <https://doi.org/10.1099/00207713-37-4-463>.

- Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Standards in Genomic Sciences* 2015;10:18. <https://doi.org/10.1186/1944-3277-10-18>.
- Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews* MMBR 2013;77(3):342–56. <https://doi.org/10.1128/MMBR.00051-12>.
- Nieto C, Cherny I, Khoo SK, Lacoba MG de, Chan WT, Yeo CC, et al. The yefM-yoeB toxin-antitoxin systems of *Escherichia coli* and *Streptococcus pneumoniae*: functional and structural correlation. *Journal of Bacteriology* 2007;189(4):1266–78. <https://doi.org/10.1128/JB.01130-06>.
- Novotna G, Janata J. A new evolutionary variant of the streptogramin A resistance protein, Vga(A)LC, from *Staphylococcus haemolyticus* with shifted substrate specificity towards lincosamides. *Antimicrobial agents and chemotherapy* 2006;50(12):4070–6. <https://doi.org/10.1128/AAC.00799-06>.
- O'Flaherty S, Briner Crawley A, Theriot CM, Barrangou R. The *Lactobacillus* Bile Salt Hydrolase Repertoire Reveals Niche-Specific Adaptation. *mSphere* 2018;3(3). <https://doi.org/10.1128/mSphere.00140-18>.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology* 2016;17(1):132. <https://doi.org/10.1186/s13059-016-0997-x>.
- Parker AC, Smith CJ. Genetic and biochemical analysis of a novel Ambler class A beta-lactamase responsible for cefoxitin resistance in *Bacteroides* species. *Antimicrobial agents and chemotherapy* 1993;37(5):1028–36. <https://doi.org/10.1128/aac.37.5.1028>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 2015;25(7):1043–55. <https://doi.org/10.1101/gr.186072.114>.
- Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 2019;176(3):649-662.e20. <https://doi.org/10.1016/j.cell.2019.01.001>.
- Pope PB, Smith W, Denman SE, Tringe SG, Barry K, Hugenholtz P, et al. Isolation of *Succinivibrionaceae* implicated in low methane emissions from Tammar wallabies. *Science (New York, N.Y.)* 2011;333(6042):646–8. <https://doi.org/10.1126/science.1205760>.
- Powell E, Ratnayeke N, Moran NA. Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. *Molecular ecology* 2016;25(18):4461–71. <https://doi.org/10.1111/mec.13787>.
- Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nature medicine* 2019;25(9):1442–52. <https://doi.org/10.1038/s41591-019-0559-3>.

- Pumbwe L, Wareham DW, Aduse-Opoku J, Brazier JS, Wexler HM. Genetic analysis of mechanisms of multidrug resistance in a clinical isolate of *Bacteroides fragilis*. *Clinical microbiology and infection the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2007;13(2):183–9. <https://doi.org/10.1111/j.1469-0691.2006.01620.x>.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464(7285):59–65. <https://doi.org/10.1038/nature08821>.
- Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nature biotechnology* 2017;35(9):833–44. <https://doi.org/10.1038/nbt.3935>.
- Rasmussen TS, Streidl T, Hitch TCA, Wortmann E, Deptula P, Hansen M, et al. *Sporaefaciens musculi* gen. nov., sp. nov., a novel bacterium isolated from the caecum of an obese mouse; 2019.
- Reynolds ED, Cove JH. Resistance to telithromycin is conferred by *msr(A)*, *msrC* and *msr(D)* in *Staphylococcus aureus*. *The Journal of antimicrobial chemotherapy* 2005;56(6):1179–80. <https://doi.org/10.1093/jac/dki378>.
- Ross JI, Eady EA, Cove JH, Cunliffe WJ, Baumberg S, Wootton JC. Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super-gene family. *Molecular microbiology* 1990;4(7):1207–14. <https://doi.org/10.1111/j.1365-2958.1990.tb00696.x>.
- Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, et al. Gut microbiota functions: metabolism of nutrients and other food components. *European journal of nutrition* 2018;57(1):1–24. <https://doi.org/10.1007/s00394-017-1445-8>.
- Rozov R, Brown Kav A, Bogumil D, Shterzer N, Halperin E, Mizrahi I, et al. Recycler: an algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics (Oxford, England)* 2017;33(4):475–82. <https://doi.org/10.1093/bioinformatics/btw651>.
- Rubino F, Carberry C, M Waters S, Kenny D, McCabe MS, Creevey CJ. Divergent functional isoforms drive niche specialisation for nutrient acquisition and use in rumen microbiome. *The ISME journal* 2017;11(4):932–44. <https://doi.org/10.1038/ismej.2016.172>.
- Sanders ER. Aseptic laboratory techniques: plating methods. *Journal of Visualized Experiments JoVE* 2012(63):e3064. <https://doi.org/10.3791/3064>.
- Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977;265(5596):687–95. <https://doi.org/10.1038/265687a0>.
- Schapiro JM, Gupta R, Stefansson E, Fang FC, Limaye AP. Isolation of metronidazole-resistant *Bacteroides fragilis* carrying the *nimA* nitroreductase gene from a patient in Washington State. *Journal of clinical microbiology* 2004;42(9):4127–9. <https://doi.org/10.1128/JCM.42.9.4127-4129.2004>.
- Seedorf H, Griffin NW, Ridaura VK, Reyes A, Cheng J, Rey FE, et al. Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* 2014;159(2):253–66. <https://doi.org/10.1016/j.cell.2014.09.008>.

- Segata N. Gut Microbiome: Westernization and the Disappearance of Intestinal Diversity. *Current biology* CB 2015;25(14):R611-3. <https://doi.org/10.1016/j.cub.2015.05.040>.
- Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome biology* 2012;13(6):R42. <https://doi.org/10.1186/gb-2012-13-6-r42>.
- Sekelja M, Berget I, Næs T, Rudi K. Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. *The ISME journal* 2011;5(3):519–31. <https://doi.org/10.1038/ismej.2010.129>.
- Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS biology* 2016;14(8):e1002533. <https://doi.org/10.1371/journal.pbio.1002533>.
- Shapira M. Gut Microbiotas and Host Evolution: Scaling Up Symbiosis. *Trends in ecology & evolution* 2016;31(7):539–49. <https://doi.org/10.1016/j.tree.2016.03.006>.
- Sharkey LKR, Edwards TA, O'Neill AJ. ABC-F Proteins Mediate Antibiotic Resistance through Ribosomal Protection. *mBio* 2016;7(2):e01975. <https://doi.org/10.1128/mBio.01975-15>.
- Simpson PJ, Bolam DN, Cooper A, Ciruela A, Hazlewood GP, Gilbert HJ, et al. A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. *Structure* 1999;7(7):853–64. [https://doi.org/10.1016/S0969-2126\(99\)80108-7](https://doi.org/10.1016/S0969-2126(99)80108-7).
- Singh KV, Weinstock GM, Murray BE. An *Enterococcus faecalis* ABC homologue (Lsa) is required for the resistance of this species to clindamycin and quinupristin-dalfopristin. *Antimicrobial agents and chemotherapy* 2002;46(6):1845–50. <https://doi.org/10.1128/aac.46.6.1845-1850.2002>.
- Sóki J, Gal M, Brazier JS, Rotimi VO, Urbán E, Nagy E, et al. Molecular investigation of genetic elements contributing to metronidazole resistance in *Bacteroides* strains. *The Journal of antimicrobial chemotherapy* 2006;57(2):212–20. <https://doi.org/10.1093/jac/dki443>.
- Sorbara MT, Littmann ER, Fontana E, Moody TU, Kohout CE, Gjonbalaj M, et al. Functional and Genomic Variation between Human-Derived Isolates of *Lachnospiraceae* Reveals Inter- and Intra-Species Diversity. *Cell host & microbe* 2020;28(1):134-146.e4. <https://doi.org/10.1016/j.chom.2020.05.005>.
- STACKEBRANDT E, GOEBEL BM. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International journal of systematic and evolutionary microbiology* 1994;44(4):846–9. <https://doi.org/10.1099/00207713-44-4-846>.
- Statovci D, Aguilera M, MacSharry J, Melgar S. The Impact of Western Diet and Nutrients on the Microbiota and Immune Response at Mucosal Interfaces. *Frontiers in Immunology* 2017;8:838. <https://doi.org/10.3389/fimmu.2017.00838>.
- Strobel HJ. Vitamin B12-dependent propionate production by the ruminal bacterium *Prevotella ruminicola* 23. *Applied and Environmental Microbiology* 1992;58(7):2331.

- Tan H, Zhao J, Zhang H, Zhai Q, Chen W. Isolation of Low-Abundant Bacteroidales in the Human Intestine and the Analysis of Their Differential Utilization Based on Plant-Derived Polysaccharides. *Frontiers in microbiology* 2018;9:1319. <https://doi.org/10.3389/fmicb.2018.01319>.
- Tessé S, Trueba F, Berthet N, Hot C, Chesneau O. Resistance Genes Underlying the LSA Phenotype of Staphylococcal Isolates from France. *Antimicrobial agents and chemotherapy* 2013;57(9):4543–6. <https://doi.org/10.1128/AAC.00259-13>.
- Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, et al. The *Prevotella copri* Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell host & microbe* 2019;26(5):666-679.e7. <https://doi.org/10.1016/j.chom.2019.08.018>.
- Tribble GD, Garza JJ, Yeung VL, Rigney TW, Dao D-HV, Rodrigues PH, et al. Genetic analysis of mobile tetQ elements in oral *Prevotella* species. *Anaerobe* 2010;16(6):604–9. <https://doi.org/10.1016/j.anaerobe.2010.08.008>.
- Ubeda C, Djukovic A, Isaac S. Roles of the intestinal microbiota in pathogen protection. *Clinical & translational immunology* 2017;6(2):e128. <https://doi.org/10.1038/cti.2017.2>.
- van Dongen S, Abreu-Goodger C. Using MCL to Extract Clusters from Networks. In: van Helden J, Toussaint A, Thieffry D, editors. *Bacterial Molecular Networks: Methods and Protocols*. New York, NY: Springer New York; 2012. p. 281–295.
- Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohny RP, et al. Interplay between the human gut microbiome and host metabolism. *Nature communications* 2019;10(1):4505. <https://doi.org/10.1038/s41467-019-12476-z>.
- Wang J, Shoemaker NB, Wang GR, Salyers AA. Characterization of a *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. *Journal of Bacteriology* 2000;182(12):3559–71. <https://doi.org/10.1128/jb.182.12.3559-3571.2000>.
- Wang Y, Lv Y, Cai J, Schwarz S, Cui L, Hu Z, et al. A novel gene, *optrA*, that confers transferable resistance to oxazolidinones and phenicols and its presence in *Enterococcus faecalis* and *Enterococcus faecium* of human and animal origin. *The Journal of antimicrobial chemotherapy* 2015;70(8):2182–90. <https://doi.org/10.1093/jac/dkv116>.
- Ward JH. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 1963;58(301):236–44. <https://doi.org/10.1080/01621459.1963.10500845>.
- Wareham DW, Wilks M, Ahmed D, Brazier JS, Millar M. Anaerobic sepsis due to multidrug-resistant *Bacteroides fragilis*: microbiological cure and clinical response with linezolid therapy. *Clinical infectious diseases an official publication of the Infectious Diseases Society of America* 2005;40(7):e67-8. <https://doi.org/10.1086/428623>.
- Weimann A, Mooren K, Frank J, Pope PB, Bremges A, McHardy AC. From Genomes to Phenotypes: TraitR, the Microbial Trait Analyzer. *mSystems* 2016;1(6). <https://doi.org/10.1128/mSystems.00101-16>.

- Weisblum B. Erythromycin resistance by ribosome modification. *Antimicrobial agents and chemotherapy* 1995;39(3):577–85. <https://doi.org/10.1128/aac.39.3.577>.
- Whitman WB. *Bergey's Manual of Systematic Bacteriology: The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*. New York, NY: Bergey's Manual Trust; 2011.
- Wieser A, Schneider L, Jung J, Schubert S. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). *Applied microbiology and biotechnology* 2012;93(3):965–74. <https://doi.org/10.1007/s00253-011-3783-4>.
- Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, et al. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science (New York, N.Y.)* 2003;299(5615):2074–6. <https://doi.org/10.1126/science.1080029>.
- Yang C, Mogno I, Contijoch EJ, Borgerding JN, Aggarwala V, Li Z, et al. Fecal IgA Levels Are Determined by Strain-Level Differences in *Bacteroides ovatus* and Are Modifiable by Gut Microbiota Manipulation. *Cell host & microbe* 2020;27(3):467-475.e6. <https://doi.org/10.1016/j.chom.2020.01.016>.
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* 2012;486(7402):222–7. <https://doi.org/10.1038/nature11053>.
- Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International journal of systematic and evolutionary microbiology* 2017;67(5):1613–7. <https://doi.org/10.1099/ijsem.0.001755>.
- Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, et al. Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell host & microbe* 2019;25(5):656-667.e8. <https://doi.org/10.1016/j.chom.2019.03.007>.
- Zoetendal EG, Rajilic-Stojanovic M, Vos WM de. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* 2008;57(11):1605–15. <https://doi.org/10.1136/gut.2007.133603>.

Appendix 1

Supplementary Table S1: Fibre Medium composition

Ingredients	Concentrations
Yeast	0,5 g/l
Meat Extract	0,5 g/l
Peptone	0,5 g/l
K ₂ HPO ₄	0,45 g/l
KH ₂ PO ₄	0,45 g/l
MgSO ₄ x 7 H ₂ O	45 mg/l
CaCl ₂ x 2 H ₂ O	90 mg/l
NaHCO ₃	4 g/l
Konjac Glucomannan	1 g/l
Inulin	1 g/l
Xylan	1 g/l
Hemin	10 mg/l
Vitamin Mix*	10 ml/l
Trace Element Mix*	1 ml/l
Vitamin K*	1 ml/l
L-cysteine*	500 mg/l
1,4-Dithiothreitol (DTT) *	200 mg /l
Gentamicin	100 mg/l

* added after autoclaving

Supplementary Table S2: Modified Gifu Anaerobic Medium (GAM) Modified composition

Ingredients	Concentrations
GAM broth base (Hyserver, Germany)	41.7 g/l
L-cysteine*	500 mg/l
1,4-Dithiothreitol (DTT) *	200 mg /l

Supplementary Table S3: Brain Heart Infusion Medium (BHI) composition

Ingredients	Concentrations
BHI broth (Thermo Fisher Scientific, USA)	37 g/l
L-cysteine*	500 mg/l
1,4-Dithiothreitol (DTT) *	200 mg /l

* added after autoclaving

Supplementary Table S4: Wilkins Chalgren Anaerobic Medium (WCA) composition

Ingredients	Concentrations
WCA broth (Thermo Fisher Scientific, USA)	33 mg/l
Phenosafranin	2,5 mg/l
L-cysteine*	500 mg/l
1,4-Dithiothreitol (DTT) *	200 mg/l



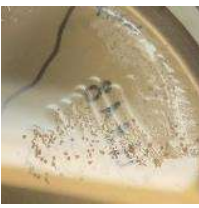


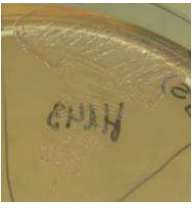


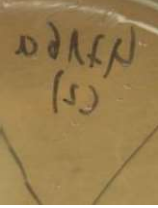


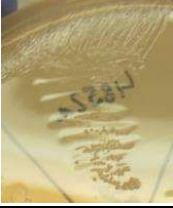
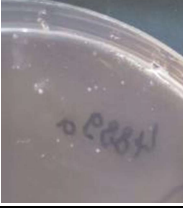

Supplementary Table S5: Yeast extract, casitone and fatty acid (YCFA) medium composition


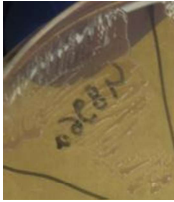






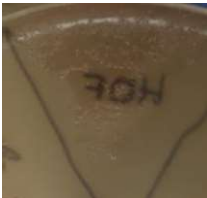




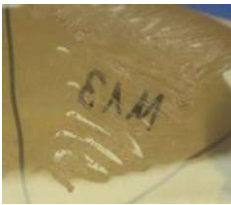


Ingredients	Concentrations
Casitone	10 g/l
Yeast extract	2.5 g/l
Glucose	5 g/l
MgSO₄ x 7 H₂O	45 mg/l
CaCl₂ x 2 H₂O	90 mg/l
K₂HPO₄	0.45 g/l
KH₂PO₄	0.45 g/l
NaCl	0.9 g/l
NaHCO₃	4 g/l
L-cysteine	1 g/l
Hemin	10 mg/l
Acetic acid	1.9 ml/l
Propionic acid	0.7 ml/l
iso-Butyric acid	90 µl/l
n-Valeric acid	100 µl/l
iso-Valeric acid	100 µl/l
Vitamin solution*	10 ml/l









* added after autoclaving

Appendix 2

Supplementary Table S6: Pictures of the BSH-Assay

Isolate	WCA with 0.5 % tdca	WCA control	Result
H-AC-1			positive
H-AC-2			negative
H-AC-3			negative
H-AC-4			negative
H-BS-1			negative
H-BS-2			positive
H-BS-3			negative

H-BS-4			weakly positive
H-BS-5			positive
H-BS-6			positive
H-BS-7			positive
H-BS-8			positive
H-VI-1			positive
M-AC-1			positive
M-AC-2			positive

M-AC-3			positive
M-FR-1			positive
P-AC-1			positive
P-FR-1			negative

Appendix 3

Supplementary Table S7: Summary of the ANI and DDH-values of the uncurated literature dataset (3.3)

Genome Accession Number	ANI [%] to <i>B. vulgatus</i> ATCC 8482 ^T	ANI [%] to <i>B. dorei</i> CL03T12C01 ^T	DDH [%] to <i>B. vulgatus</i> ATCC 8482 ^T	DDH [%] to <i>B. dorei</i> CL03T12C01 ^T
GCA_009017515.1_ASM901751v1	98,4954	95,3508	89,4	63,5
GCA_009017525.1_ASM901752v1	98,4815	95,2924	89,4	63,5
GCA_009017535.1_ASM901753v1	98,5355	95,285	89,4	63,5
GCA_009017545.1_ASM901754v1	98,4723	95,2761	89,2	63,4
GCA_009017555.1_ASM901755v1	98,5162	95,2631	89,2	63,4
GCA_009017615.1_ASM901761v1	98,3326	95,2138	89,3	63,5
GCA_009017625.1_ASM901762v1	98,5952	95,2528	90,6	62,6
GCA_009017635.1_ASM901763v1	98,4864	95,273	89,2	63,4
GCA_009017645.1_ASM901764v1	98,475	95,2962	89,4	63,5
GCA_009017655.1_ASM901765v1	98,4578	95,2751	89,4	63,4
GCA_009017715.1_ASM901771v1	98,4602	95,3177	89,4	63,5
GCA_009017725.1_ASM901772v1	98,5718	95,3099	89,4	63,4
GCA_009017735.1_ASM901773v1	98,5305	95,2275	89,3	63,4
GCA_009017745.1_ASM901774v1	98,4642	95,3162	89,5	63,5
GCA_009017755.1_ASM901775v1	98,5166	95,2865	89,4	63,4
GCA_009017815.1_ASM901781v1	98,3476	95,2789	89,8	64,8
GCA_009017825.1_ASM901782v1	98,5131	95,1603	89,1	63,2
GCA_009017835.1_ASM901783v1	98,4957	95,2845	89,4	63,4
GCA_009017845.1_ASM901784v1	98,4921	95,328	89,4	63,4
GCA_009017855.1_ASM901785v1	98,4858	95,3234	89,4	63,4
GCA_009017905.1_ASM901790v1	98,4453	95,2437	89,4	64,3
GCA_009017925.1_ASM901792v1	98,454	95,3571	89,4	63,5
GCA_009017935.1_ASM901793v1	98,4063	95,2929	89	63,3
GCA_009017945.1_ASM901794v1	98,5845	95,3213	89,4	63,4
GCA_009017995.1_ASM901799v1	98,519	95,2997	89,4	63,5
GCA_009018005.1_ASM901800v1	98,4944	95,327	89,4	63,4
GCA_009018035.1_ASM901803v1	98,5159	95,32	89,4	63,5
GCA_009018045.1_ASM901804v1	98,5084	95,3358	89,4	63,4
GCA_009018065.1_ASM901806v1	98,5783	95,2961	90,7	62,7
GCA_009018095.1_ASM901809v1	98,5981	95,2615	90,6	62,7
GCA_009018105.1_ASM901810v1	98,6077	95,2486	90,6	62,6
GCA_009018135.1_ASM901813v1	98,5524	95,3093	90,6	62,6
GCA_009018145.1_ASM901814v1	98,5411	95,3307	90,5	62,6
GCA_009018155.1_ASM901815v1	98,5226	95,3581	89,5	63,5
GCA_009018195.1_ASM901819v1	98,5144	95,3026	89,5	63,5
GCA_009018215.1_ASM901821v1	98,4619	95,2633	89,4	63,4
GCA_009018225.1_ASM901822v1	98,4822	95,331	89,4	63,4
GCA_009018235.1_ASM901823v1	97,9773	95,0705	90	65
GCA_009018265.1_ASM901826v1	98,511	95,2973	89,3	63,5
GCA_009018285.1_ASM901828v1	98,4752	95,3275	89,2	63,4
GCA_009018315.1_ASM901831v1	98,4894	95,3051	89,4	63,5
GCA_009018325.1_ASM901832v1	98,467	95,3171	89,4	63,5

GCA_009018355.1_ASM901835v1	98,4594	95,3382	89,4	63,4
GCA_009018365.1_ASM901836v1	98,505	95,3009	89,4	63,4
GCA_009018395.1_ASM901839v1	98,508	95,2794	89,4	63,5
GCA_009018415.1_ASM901841v1	98,4537	95,4359	89,5	64,3
GCA_009018425.1_ASM901842v1	98,4453	95,3151	89,2	63,5
GCA_009018435.1_ASM901843v1	98,4799	95,28	89,2	63,4
GCA_009018465.1_ASM901846v1	98,4991	95,3113	89,3	63,5
GCA_009018495.1_ASM901849v1	98,4595	95,2765	89,3	63,5
GCA_009018505.1_ASM901850v1	98,4866	95,318	89,4	63,4
GCA_009018515.1_ASM901851v1	98,4866	95,2829	89,3	63,5
GCA_009018555.1_ASM901855v1	98,4996	95,2613	89,3	63,5
GCA_009018565.1_ASM901856v1	98,5137	95,3336	89,3	63,4
GCA_009018595.1_ASM901859v1	98,5106	95,3174	89,3	63,5
GCA_009018605.1_ASM901860v1	98,4148	95,3089	89,3	63,5
GCA_009018615.1_ASM901861v1	98,5503	95,3258	89,3	63,5
GCA_009018655.1_ASM901865v1	98,4964	95,2699	89,3	63,5
GCA_009018665.1_ASM901866v1	98,4646	95,2933	89,3	63,5
GCA_009018695.1_ASM901869v1	98,4469	95,3379	89,4	63,4
GCA_009018705.1_ASM901870v1	98,5176	95,3405	89,4	63,4
GCA_009018715.1_ASM901871v1	98,4719	95,279	89,3	63,7
GCA_009018755.1_ASM901875v1	98,4579	95,2962	89,1	63,4
GCA_009018765.1_ASM901876v1	98,5425	95,3728	89,2	63,4
GCA_009018785.1_ASM901878v1	98,4655	95,2428	89,5	63,5
GCA_009018815.1_ASM901881v1	98,5638	95,3781	89,4	63,4
GCA_009018825.1_ASM901882v1	98,4789	95,2795	89,4	63,4
GCA_009018855.1_ASM901885v1	98,5218	95,2852	89,4	63,5
GCA_009018875.1_ASM901887v1	98,6504	95,5348	91,3	65,3
GCA_009018885.1_ASM901888v1	98,7229	95,4336	91,6	65,3
GCA_009018915.1_ASM901891v1	98,7566	95,2719	92,6	63,5
GCA_009018935.1_ASM901893v1	98,7858	95,2487	92,4	63,5
GCA_009018955.1_ASM901895v1	98,8606	95,2567	92,5	63,5
GCA_009018965.1_ASM901896v1	98,8499	95,288	92,5	63,6
GCA_009018975.1_ASM901897v1	98,8054	95,3115	92,4	63,4
GCA_009018995.1_ASM901899v1	98,7674	95,2004	92,2	63,5
GCA_009019035.1_ASM901903v1	98,7913	95,2768	92,2	63,5
GCA_009101525.1_ASM910152v1	98,7739	95,2029	93,6	63,9
GCA_009101815.1_ASM910181v1	98,8223	95,2611	93,6	64
GCA_009101825.1_ASM910182v1	98,7867	95,3244	93,3	63,6
GCA_009101845.1_ASM910184v1	98,8591	95,3609	93,4	63,4
GCA_009101865.1_ASM910186v1	98,8272	95,3596	93,4	63,4
GCA_009101885.1_ASM910188v1	98,8024	95,4083	93,3	63,6
GCA_009101915.1_ASM910191v1	98,895	95,3452	93,3	63,4
GCA_009101925.1_ASM910192v1	98,9933	95,3092	93,5	63,4
GCA_009102345.1_ASM910234v1	98,5774	95,4141	90,4	64,3
GCA_009102865.1_ASM910286v1	98,9159	95,2974	93,4	63,4
GCA_009102885.1_ASM910288v1	98,8838	95,3336	93,3	63,4
GCA_009102905.1_ASM910290v1	98,8555	95,3974	93,4	63,4
GCA_009102955.1_ASM910295v1	98,8959	95,3892	93,4	63,4
GCA_009102985.1_ASM910298v1	98,923	95,3275	93,4	63,4

GCA_009102995.1_ASM910299v1	98,8245	95,3881	93,4	63,4
GCA_009103005.1_ASM910300v1	98,7905	95,2658	93	63,4
GCA_009103025.1_ASM910302v1	98,8288	95,3574	93,4	63,4
GCA_009103055.1_ASM910305v1	98,8528	95,3556	93,4	63,4
GCA_009103085.1_ASM910308v1	98,8589	95,365	93,3	63,4
GCA_009103095.1_ASM910309v1	98,8691	95,3917	93,4	63,4
GCA_009103105.1_ASM910310v1	98,8184	95,3802	93,3	63,4
GCA_009103135.1_ASM910313v1	98,8199	95,3401	93,3	63,4
GCA_009103155.1_ASM910315v1	98,8331	95,3105	93,3	63,4
GCA_009103185.1_ASM910318v1	98,8877	95,3159	93,3	63,5
GCA_009107465.1_ASM910746v1	98,4855	95,2952	89,4	63,5
GCA_009107475.1_ASM910747v1	98,5077	95,3592	89,4	63,5
GCA_009107485.1_ASM910748v1	98,8885	95,2607	93,1	63,6
GCA_009107495.1_ASM910749v1	98,509	95,2524	89,3	63,5
GCA_009107545.1_ASM910754v1	98,5185	95,2709	89,3	63,5
GCA_009107565.1_ASM910756v1	98,7259	95,6093	92,4	64,2
GCA_009107575.1_ASM910757v1	98,6878	95,4839	92,4	64,2
GCA_009107585.1_ASM910758v1	98,9583	95,278	94,5	63,9
GCA_009107605.1_ASM910760v1	98,7224	95,5455	92,4	64,2
GCA_009107625.1_ASM910762v1	98,4366	95,3276	89,3	63,4
GCA_009107665.1_ASM910766v1	98,5475	95,3235	89,4	63,4
GCA_009107685.1_ASM910768v1	98,6216	95,3882	90,3	64,2
GCA_009107695.1_ASM910769v1	98,5959	95,3764	90,5	64,4
GCA_009107705.1_ASM910770v1	98,6285	95,4018	90,9	64,1
GCA_009107745.1_ASM910774v1	98,5721	95,4157	90,2	64,1
GCA_009107765.1_ASM910776v1	99,0837	95,4932	94,7	64
GCA_009107785.1_ASM910778v1	98,8917	95,3945	93,7	63,5
GCA_009107795.1_ASM910779v1	99,0947	95,2127	95,2	63,1
GCA_009107815.1_ASM910781v1	99,0915	95,4792	94,7	63,9
GCA_009107825.1_ASM910782v1	99,0988	95,3055	94,5	63,9
GCA_009107865.1_ASM910786v1	98,747	95,5202	92,4	64,1
GCA_009107885.1_ASM910788v1	98,7077	95,5278	92,4	64,2
GCA_009107895.1_ASM910789v1	98,7459	95,4886	92,3	64,2
GCA_009107915.1_ASM910791v1	98,6874	95,5429	92,4	64,2
GCA_009107935.1_ASM910793v1	98,4819	95,2955	89,2	63,4
GCA_009107965.1_ASM910796v1	98,4917	95,2979	89,4	63,4
GCA_009107985.1_ASM910798v1	98,5122	95,2987	89,3	63,5
GCA_009107995.1_ASM910799v1	98,52	95,3272	89,4	63,4
GCA_009108005.1_ASM910800v1	98,4384	95,3264	89,3	63,4
GCA_009108045.1_ASM910804v1	98,455	95,2885	89,1	63,4
GCA_009108065.1_ASM910806v1	98,4673	95,2919	89,1	63,4
GCA_009108085.1_ASM910808v1	98,4956	95,2831	89,4	63,4
GCA_009108105.1_ASM910810v1	98,4961	95,3058	89,3	63,5
GCA_009108115.1_ASM910811v1	98,4881	95,3322	89,2	63,3
GCA_009108145.1_ASM910814v1	98,5123	95,2917	89,3	63,4
GCA_009108165.1_ASM910816v1	98,4887	95,3042	89,3	63,4
GCF_008571705.1_ASM857170v1	95,2962	99,1791	63,8	94
GCF_008571715.1_ASM857171v1	95,3929	99,2271	63,8	94
GCF_008571725.1_ASM857172v1	95,257	99,2285	63,8	94

GCF_008571735.1_ASM857173v1	95,4905	98,9996	66,6	91,3
GCF_008571795.1_ASM857179v1	95,3294	99,2284	63,8	94
GCF_008571815.1_ASM857181v1	95,279	99,2072	63,8	94,1
GCF_008571825.1_ASM857182v1	95,3244	99,2058	63,8	94
GCF_008571855.1_ASM857185v1	95,3282	99,2296	63,8	93,9
GCF_008571875.1_ASM857187v1	95,5027	99,1286	65,2	93,1
GCF_008571885.1_ASM857188v1	95,2947	99,2181	63,8	94
GCF_008571915.1_ASM857191v1	95,3364	99,153	63,9	94
GCF_008571925.1_ASM857192v1	95,378	99,2533	63,8	94
GCF_008571955.1_ASM857195v1	95,3451	99,2418	64	94
GCF_008571975.1_ASM857197v1	95,3543	99,2026	63,8	93,9
GCF_008571985.1_ASM857198v1	95,3723	99,2061	64	94
GCF_008572005.1_ASM857200v1	95,3512	99,1987	63,8	94
GCF_008572035.1_ASM857203v1	95,3331	99,2015	63,8	94
GCF_008572045.1_ASM857204v1	95,2645	99,2171	63,9	94
GCF_008572075.1_ASM857207v1	95,3317	99,2263	63,8	94,1
GCF_008572085.1_ASM857208v1	95,2728	99,1665	64,1	94,1
GCF_008572095.1_ASM857209v1	95,3125	99,1496	63,9	93,8
GCF_008572125.1_ASM857212v1	95,3812	99,0214	63,5	93,9
GCF_008572145.1_ASM857214v1	95,2115	98,8517	63,3	91,7
GCF_008572155.1_ASM857215v1	95,2288	98,8842	63,4	91,9
GCF_008572205.1_ASM857220v1	95,2381	98,8749	63,6	91,8
ERR1022358_13470_2#61	99,0265	95,2269	94,2	63,5
ERR1022411_14207_7#20	98,9128	95,346	93,4	63,8
ERR1022458_14207_7#67	99,0103	95,2357	93,9	63,3
ERR1203951_18048_1#68	98,4669	95,2971	89,2	62,9
ERR1204045_18048_2#68	98,4133	95,3689	89,1	62,8
ERR2221206_14207_6#8	98,5957	95,0874	91,2	63,2
ERR2221252_20287_6#9	95,3893	99,1033	63,9	93,5
ERR2221254_20287_6#14	98,6564	95,418	92,1	64,6
ERR2221256_20287_6#19	98,6934	95,3504	92,3	64,4
ERR2221259_20287_6#40	98,6943	95,3912	92,2	64,5
ERR2221267_20287_6#65	97,2088	96,0039	78,8	67,7
ERR2221272_20287_6#81	98,6372	95,3448	92,1	64,5
ERR2221273_20287_6#82	97,1525	95,8533	77,6	67,9
ERR2221276_20287_6#85	97,124	95,8853	77,6	67,9
ERR2230079_20287_6#79	98,6274	95,3245	92,2	64,5
ERR2230082_21673_4#2	95,4089	99,1058	63,9	93,5
ERR2230086_21673_4#6	98,6595	95,3815	92,2	64,5
ERR2230091_21673_4#11	97,1862	95,9383	77,7	68,2
ERR2230095_21673_4#15	97,1248	95,8747	77,7	68,2
ERR2230128_21673_4#50	98,6346	95,3382	92,2	64,5
ERR2230134_21673_4#58	98,6565	95,402	92,3	64,5
ERR2230150_21673_4#79	97,1192	95,9514	77,7	68,2

green=above species-level threshold, red=below threshold, orange= subspecies threshold (DDH)