UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE HUMAN GASTROINTESTINAL MICROBIOME: INSIGHTS INTO THE MICROBIAL COMMUNITY STRUCTURE AND ANTIBIOTIC RESISTOME

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

LINDSEY L O'NEAL
Norman, Oklahoma
2017

THE HUMAN GASTROINTESTINAL MICROBIOME: INSIGHTS INTO THE MICROBIAL COMMUNITY STRUCTURE AND ANTIBIOTIC RESISTOME

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

____
Dr. Paul A. Lawson, Chair

____
Dr. Ralph S. Tanner

____
Dr. Bradley S. Stevenson

____
Dr. Cecil M. Lewis

____
Dr. Christina Warinner

© Copyright by LINDSEY L O'NEAL 2017
All Rights Reserved.

Table of Contents

List of Tables vi

List of Figures..viii

Abstract...ix

Chapter 1
Dominant groups of bacteria within the human colonic microbiota: Novel genera and species of the functionally important families *Lachnospiraceae* and *Ruminococcaceae*....................1

Abstract 1

Introduction 2

Materials and Methods..12

Results...18

Discussion 42

Chapter 2
***Clostridium amazonense* a novel obligate anaerobic bacterium isolated from a remote Amazonian community in Peru and antimicrobial resistance....45**

Abstract...45

Introduction...47

Materials and methods...51

Results...63

Discussion 83

Chapter 3
Insights into the antibiotic resistome of a traditional indigenous Peruvian community from cultivated representatives of *Bacillaceae, Paenibacillaceae, Lactobacillaceae, Streptococcaceae* and *Enterobacteriaceae* ………86

Abstract……………………………………………………………………...86

Introduction…………………………………………………………………….87

Materials and Methods…………………………………………………………92

Results…………………………………………………………………….....102

Discussion………………………………………………………………….....151

Conclusions……………..……………………………………………………..154

References……………..…………………..……………………………………..160

Appendix……………………………………………………………………………175

List of Tables

Table 1: Cellular fatty acid compostion of strain 80/3 19

Table 2: Characteristics of strain 80/3 20

Table 3: Comparison of group I ruminococci... 22

Table 4: Group I ruminococci genome summary 26

Table 5: *Ruminococcccus sensu stricto* and group I ANI values 26

Table 6: Group I ruminococci ANI data OrthoANIu.. 27

Table 7: Comparison of group II ruminococci (*Lachnospiracea*)............................ 27

Table 8: Group II (*Lachnospiraceae*) ruminococci genome information.................. 28

Table 9: Group II (*Lachnospiraceae*) misclassified ruminococci ANI values 28

Table 10: Group II (*Lachnospiraceae*) ANI data OrthoANIu values....................... 29

Table 11: API Rapid 32A differential reactions of *Anaerostipes* strains.................. 34

Table 12: API ZYM differential reactions of *Anaerostipes* strains 34

Table13: Cellular fatty acid composition of *Anaerostipes* strains 35

Table 14: ANI analysis of *Anaerostipes* strains... 38

Table 15: Antibiotics and concentrations for *Clostridium spp* 60

Table 16: Differential characteristics of strain NE08V^{T}.. 65

Table 17: Comparison of strain NE08VT to closely related species 67

Table 18: *Clostridium* species genome summary 68

Table 19: NE08V^{T} and species based on ANI analysis 69

Table 20: Fatty acid composition of NE08V^{T} 70

Table 21: EUCAST clinical breakpoints and MIC of strain NE08V^T 73

Table 22: Resistance genes of strain NE08V^T and related species 76

Table 23: Resistance genes of strain NE08V^T and related species 76

Table 24: Antibiotic resistance genes detected in *Clostridium* strains 77

Table 25: *Clostridium* strains SarpoB genes.. 79

Table 26: *Clostridium* strains SarpoB genes matrix 79

Table 27: *Clostridium* SagyrB resistance genes 80

Table 28: *Clostridium* SagyrB resistance genes matrix 80

Table 29: Antibiotic targets and modes of resistance 95

Table 30: Antibiotic concentrations for resistance screen by disc diffusion 96

Table 31: Antibiotics and concentrations selected for MIC determination 99

Table 32: Number of initial screen resistant strains.. 103

Table 33: Original disc resistance to antibiotics 104

Table 34: Species identity... 106

Table 35: OrthoANIu analysis of resitant strains 109

Table 36: Antibiotic, spectrum, and impact on GI tract 112

Table 37: MIC concentrations of resistant strains 116

Table 38: Resistant gene similarity to database references 128

List of Figures

Figure 1. Phylogenetic tree of *Lachnospiraceae*, *Ruminococcaceae* and *Clostridiaceae*.xi

Figure 2. Dominant genera of gastrointestinal tract...2

Figure 3. Members of group I and group II misclassified ruminococci…………………6

Figure 4. Images of strain 80/3………………………………………………………21

Figure 5. Phylogenetic tree of strain 80/3…………………………………………....23

Figure 6. Images of strain 5/1/63FAA……………………………….…………….…32

Figure 7. Phylogenetic tree of strain 5/1/63FAA...………………………..…………37

Figure 8. Shipibo community map…………………………………………..…….51

Figure 9. Phylogenetic tree of NE08V^T……..………………………………....…….66

Figure 10. San Mateo community map……………………………….…………...93

Figure 11. Phylogenetic tree of resistome isolates……………...……………………..108

Figure 12. Center for Disease Control initiative literature...……………..…………158

Figure A1. EUCAST data example...………………………………………………176

Figure A2. CARD data example……………………………………………………177

Abstract

The human gastrointestinal microbiome consists of approximately 10-100 trillion symbiotic microbial cells. Recent advances of molecular based approaches have enhanced the ease and rapidity with which the microbial community structure and function can be analyzed. However, a growing problem is the divide between molecular analyses and cultivation based methods of investigation of this complex microbial system. Cultivation and molecular methods should be utilized complementary. Whole genome sequencing and metagenomic data enable inference of growth conditions of fastidious gut microbes. Investigations of the complex microbial communities of the gastrointestinal tract can greatly benefit from a combination of culture-based and molecular approaches in order to determine function and diversity of the previously uncultured organisms.

Characterization of the gut microbiota has been of particular interest as it has such a tremendous amount of diversity and serves important roles in host health and disease. Data from gastrointestinal microbiome investigations is presented in the following chapters. Here I present research spanning organisms isolated from a Crohn's disease patient to key species isolated from the microbiome of remote, traditional, indigenous communities (Fig 1). The study of a novel species is especially important when it is an abundant organism with an unknown role within the gastrointestinal tract. Chapter one addresses the characterization of two novel numerically dominant organisms. Investigations of such organisms provide phylogenetic resolution within dominant taxa and insights into their function within

the gastrointestinal microbial community. Chapters two and three focus on the isolation and characterization of key species from remote, hunter-gatherer communities. Investigations with traditional, indigenous communities provides a unique opportunity to expand the dataset, which is mainly represented by studies from within industrialized countries. Sampling within traditional communities removes many of the modern cosmopolitan biases. The gastrointestinal microbial ecosystems of traditional Peruvian communities have been shown to be distinct when compared populations from the industrialized world and are more similar to those of ancient samples than those sampled from cosmopolitan groups. In comparison, the gastrointestinal system of Western, cosmopolitan populations have undergone many changes from the ancestral state microbiome. When sampling from traditional communities, changes from the ancestral state microbiome in response to urbanism and Westernization are anticipated. Indeed, such changes, especially the selective pressures from therapeutic antibiotic treatment, provide drivers for antibiotic resistance and thus, the gastrointestinal microbiome may serve as a for reservoir pathogen resistance gene mechanisms.

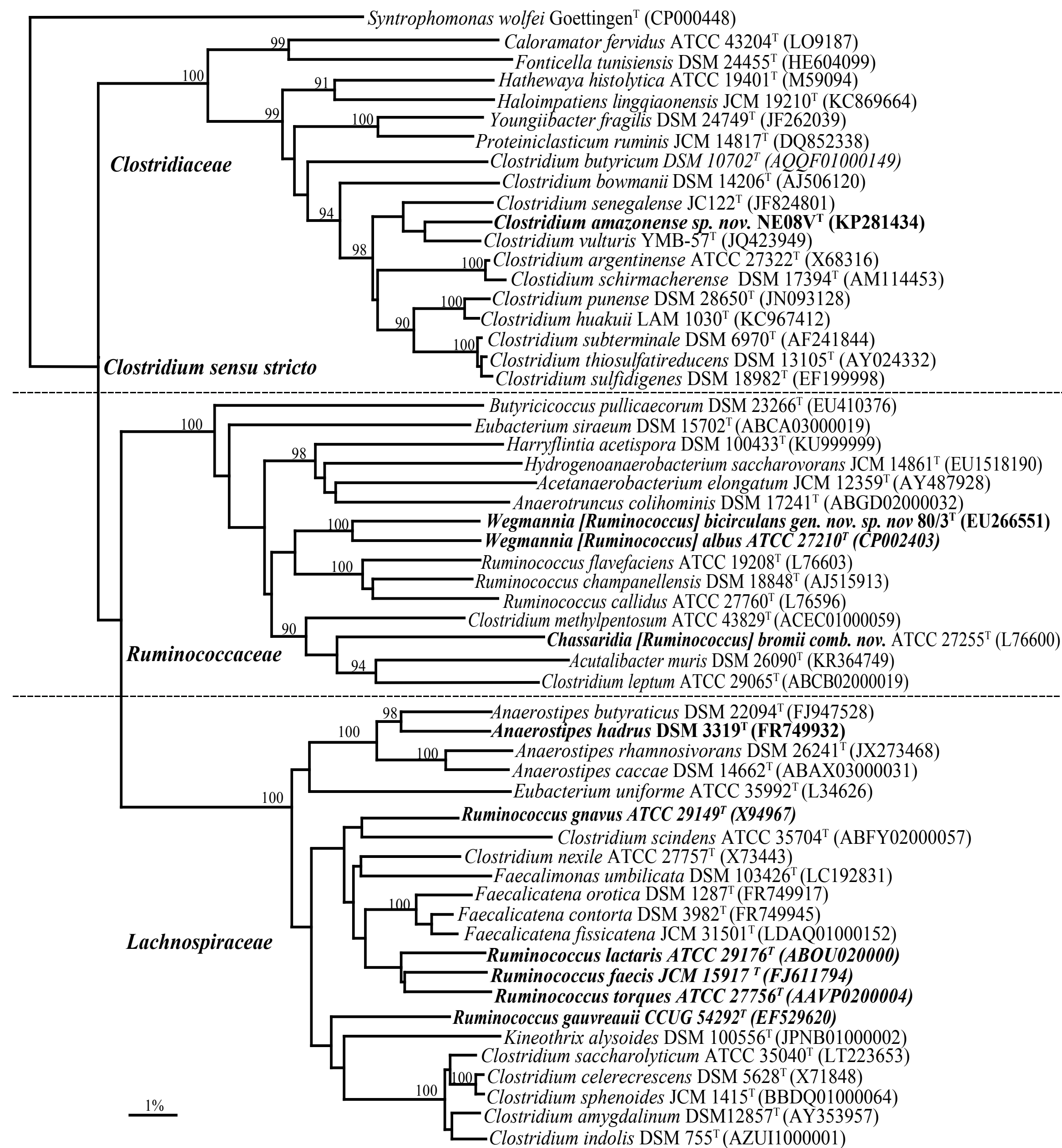

Figure 1. Phylogenetic tree of *Lachnospiraceae*, *Ruminococcaceae* and *Clostridiaceae*. All organisms in bold were characterized in this study and represent novel taxa.

Chapter 1

Dominant groups of bacteria within the human colonic microbiota: Novel genera and species of the functionally important families *Lachnospiraceae* and *Ruminococcaceae*

Abstract

The phyla *Actinobacteria*, *Bacteroidetes, Proteobacteria* and *Firmicutes* dominate the microbial community structure of the human gastrointestinal tract. This study aims to provide further insight into the members belonging to the *Firmicutes* that are numerically abundant within the gut microbiota and play an important role in host health and disease. Specifically, this chapter presents the taxonomic reorganization of key, misclassified organisms within the families *Ruminococcaceae* and *Lachnospiracea,* and the characterization of previously uncultured species that are among the most abundant 16S rRNA phylotypes of the human gut.

Introduction

The human colon harbors a diverse range of bacteria that form a complex community that has been subjected to both conventional cultivation and molecular-based studies (Cummings & Macfarlane, 1991, Eckburg *et al.*, 2005, Tap *et al.*, 2009, Moore *et al.*, 2011, Walker *et al.*, 2011). Understanding the role of the gastrointestinal microbial community and its complex relationship with the host is of great importance and has been a focus of many investigations for over a century (Suau *et al.*, 1999, Dethlefsen *et al.*, 2007, Bui *et al.*, 2016). Using molecular methods within this system, between 40-80% of the total 16S rRNA gene sequences represent yet-to-be cultivated species (Liu *et al.*, 2008). Discerning the relationship of key community members to specific aspects of host health is difficult with molecular methods alone. Characterization of novel bacterial functions and the description of the diversity of the trillions of bacterial cells that make up the gastrointestinal microbiome is more likely to occur by pairing culture-based with culture-independent approaches.

Dominant species are routinely identified through 16S rRNA gene sequencing to provide a roadmap for the application of classical and improved cultivation methods for distinguishing the diversity of novel species. Comparison with close neighbors may also provide insights into their function within the gastrointestinal tract environment. The continued development of complementary molecular and culture approaches is essential to identifying microbial taxa and genes necessary to maintain host health.

The complex environment is mainly dominated with organisms belonging within the phyla *Actinobacteria, Bacteroidetes* and *Firmicutes* and to a lesser degree *Proteobacteria* (Fig 2). Within the human gastrointestinal microbiome, members belonging to the *Firmicutes* phylum play a key role in health and disease through their involvement in nutrition, pathogenesis and immunology (Walker *et al.*, 2011, Flint *et al.*, 2012). Furthermore, it appears that the families *Ruminococcaceae* and *Lachnospiracea* contain many important organisms with respect to indicators with human health (Blaut & Clavel, 2007, Rainey, 2009, Flint *et al.*, 2012, Clavel *et al.*, 2014, Stackebrandt, 2014). Together these two families contain half of all known genera containing valid species.

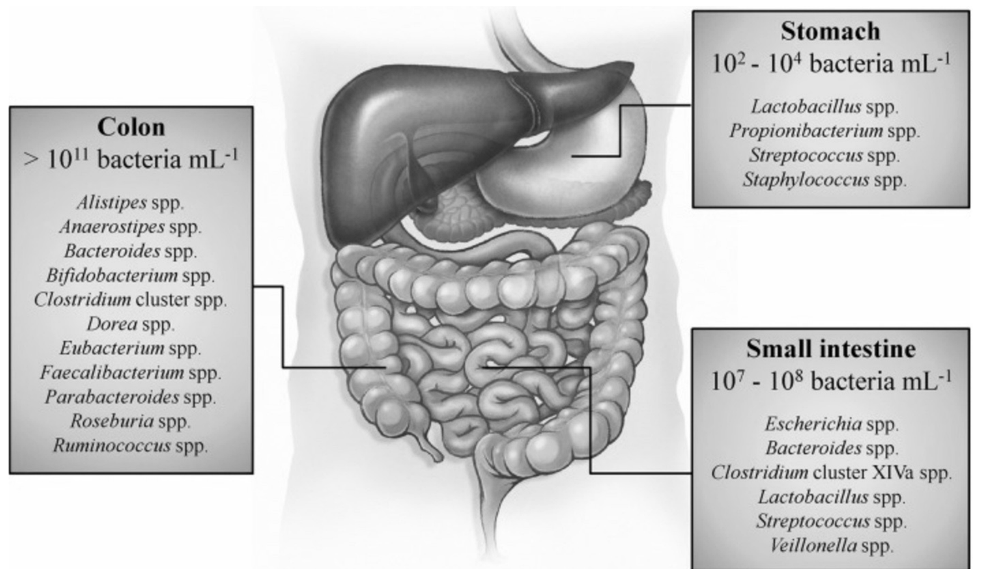

Figure 2. Dominant genera of gastrointestinal tract based on 16S rRNA gene sequence studies (Rivière *et al.*, 2016).

The abundance of an organism or group of organisms may indicate their functional role. The study of a novel species is especially important when it is an abundant organism with roles within the gastrointestinal tract. While the most abundant species are more likely to have cultured representatives, these groups remain inadequately represented (Walker *et al.*, 2011). In the present study, during the course of an ongoing microbiome investigation in collaboration with colleagues, members of these important groups have been phylogenetically, chemotaxonomically and biochemically characterized. Specifically, two novel organisms, shown to be important members of the human gastrointestinal tract and among the most abundant 16S rRNA phylotypes are presented here (Tap *et al.*, 2009, Qin *et al.*, 2010, Walker *et al.*, 2011). Additionally, the taxonomic restructuring of these organisms and closely related organisms is an important step to better understand the interrelationships between taxa. The aim of this study was the characterization and taxonomic reorganization of these abundant, previously uncultured species and will contribute to the overall understanding of their role and functional properties within the intestinal microbial community.

Overview and the need for taxonomic restructuring

Ruminococcaceae

Within the family *Ruminococcaceae*, the genus *Ruminococcus* represents an especially important taxon and corresponds to 5-15% of the total bacterial population in the colon (Chassard *et al.*, 2008, Ramirez-Farias *et al.*, 2008). A better understanding of the diversity and functionality of species within this group is of considerable importance. Few *Ruminococcaceae* are represented by cultured species although estimates indicate that *Firmicutes* account for up to 25% of bacteria within the intestine (Lay *et al.*, 2005). The application of 16S rRNA gene sequencing demonstrated that ruminococci are presently separated into two distinct, distantly related phylogenetic clades. Presently eleven species have been validly named and have standing in the nomenclature. Species are divided between the families *Ruminococcaceae* and *Lachnospiracea* (Fig 3) (Rainey & Janssen, 1995, Rainey, 2009, Stackebrandt, 2014).

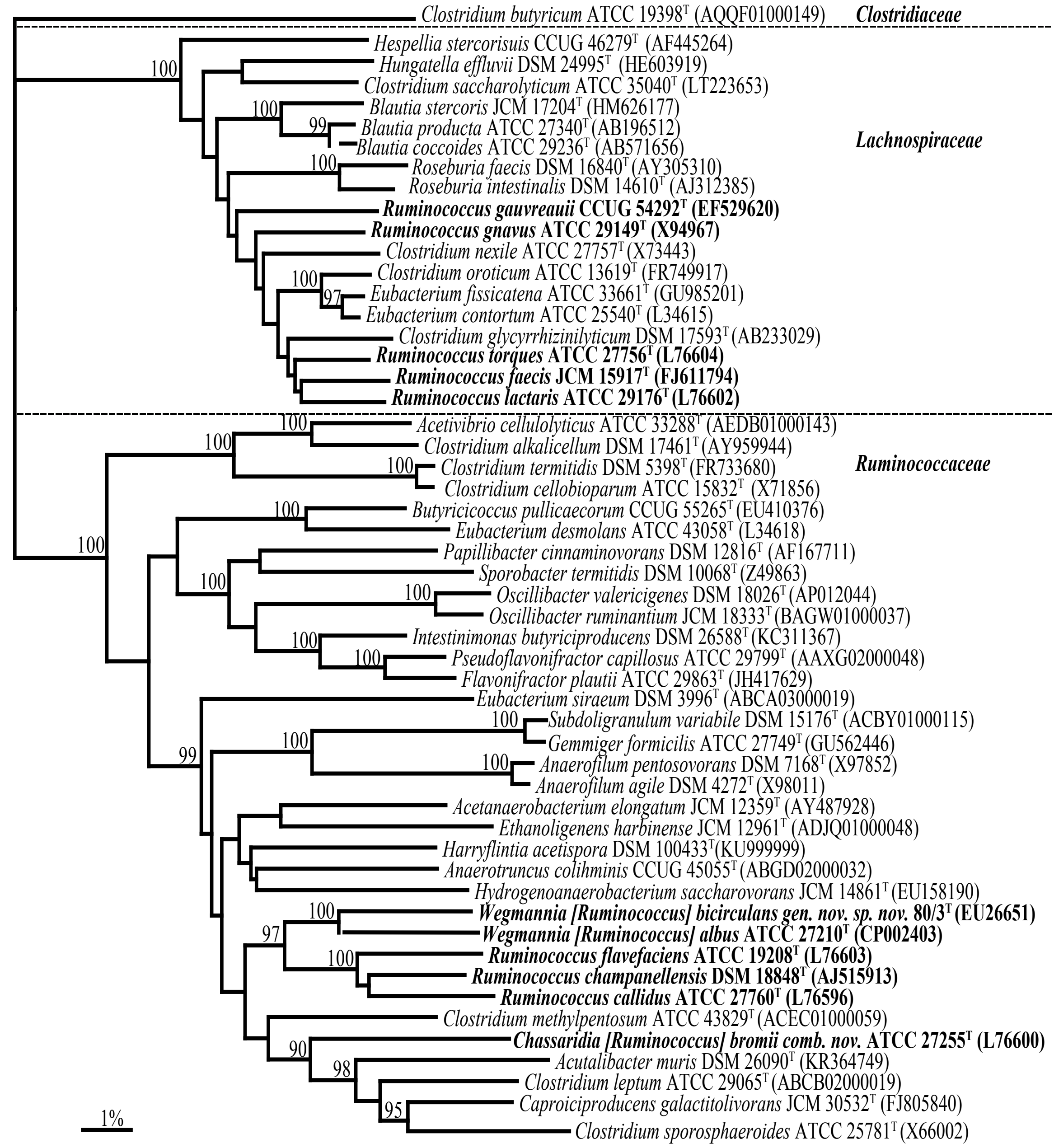

Figure 3. Members of group I (*Ruminococcus*, *Wegmannia* and *Chassardia*) and group II misclassified ruminococci within *Ruminococcaceae* and *Lachnospiracea*. *Clostridium butyricum* was used as the out-group.

The family *Ruminococcaceae* contains the genera *Ruminococcus* (type genus)*, Acetanaerobacterium, Acetivibrio, Anaerofilum, Anaerotruncus, Ethanoligenens, Faecalibacterium, Fastidiosipila, Oscillospira, Papilibacter, Sporobacter* and *Subdoligranulum*. Members of group I (*Ruminococcus)* are located within rRNA cluster IV (formally known as the *Clostridium leptum* cluster) as proposed by Collins *et al.* (1994) and include the type species *Ruminococcus flavefaciens* (Collins et al., 1994). Members of group I should be considered as *Ruminococcus sensu stricto*, i.e the true *Ruminococcus* species (Liu *et al.*, 2008, Rainey, 2009).

Members of group II ruminococci are located within rRNA cluster XIVa, of the family *Lachnospiraceae* which contains such genera as *Lachnospira* (type genus), *Acetitomaculum, Anaerostipes, Bryantella, Butyrivibrio, Catonella, Coprococcus, Dorea, Hespellia, Johnsonella, Lachnobacterium, Moryella, Oribacterium, Parasporobacterium, Pseudobutyrivibrio, Rusburia, Shuttleworthia, Sporobacterium* and *Syntrophococcus* among other misclassified species (Liu *et al.*, 2008, Stackebrandt, 2014). Therefore*,* members of group II should not be considered true representatives of the genus *Ruminococcus*, as such a number of former species have now been reclassified as *Blautia* and *Trichococcus* (Liu *et al.*, 2002, Liu *et al.*, 2008, Lawson & Finegold, 2015). Novel species and genera continue to be assigned to the ruminococci on a regular basis.

As part of continuing collaborations in the characterization of important novel strains of the human microbiome, a bacterial strain designated $80/3^T$, was

isolated from a human fecal sample by the group headed by Harry Flint (Dabek *et al.*, 2008, Wegmann *et al.*, 2014). Strain 80/3 was of unique interest as it was found to account for 0.7% of total sequences and on of the top 25 most dominant phylotypes detected in the fecal samples of six adult males (Walker *et al.*, 2011). Evidence also suggests that this novel organism and *Ruminococcus*-related strains are important members of the human GI tract microbiota being among the 10 most abundant 16S rRNA phylotypes present and indicating that they comprise approximately 2% of the colonic microbiota (Tap *et al.*, 2009, Walker *et al.*, 2011). In the present study phylogenetic, chemotaxonomic and biochemical analyses were performed to characterize this important member of the human GI ecosystem.

Lachnospiraceae

The gastrointestinal microbial community composition is fairly stable and distributed within the *Bacteroidetes* and *Firmicutes* among healthy individuals. While it is not yet proven whether changes in the gastrointestinal system are the cause or consequence of microbiota disorders, some members of the gut microbiota have been identified as significant for their central role in gut homeostasis and because the loss of these organisms adversely affect the microbial community and hosts' health. An imbalance within the gastrointestinal microbial community has been linked to conditions such as obesity, colorectal cancer, irritable bowel syndrome (IBS), and inflammatory bowel diseases (IBD) including ulcerative colitis (UC) and Crohn's disease (Sokol *et al.*, 2008, Round & Mazmanian, 2009, Sekirov *et al.*, 2010). This imbalance has also been generally associated with type 2 diabetes, and behavioural disorders like depression and regressive autism (Collins et al 2012, Dinan et al 2015).

Within the gastrointestinal ecosystem, it is recognized that butyric acid, produced in the intestinal lumen by microbial fermentation of dietary carbohydrates, can have a range of actions (Sokol *et al.*, 2008). Butyrate producing bacteria are an abundant and phylogenetically diverse group that play a key role in maintaining human gastrointestinal health (Wang *et al.*, 2014). These roles include health-promoting effects such as anti-inflammation, the stabilization of luminal pH and the enhanced proliferation of normal intestinal epithelial cells (Sokol *et al.*, 2008, Louis

& Flint, 2009). Among these species, decreased numbers of butyrate producers have been associated with gastrointestinal disorders. Thus, butyrate may have a protective role against a number of colonic diseases (Barcenilla *et al.*, 2000, Pryde *et al.*, 2002, Seksik *et al.*, 2003, Flint *et al.*, 2007, Louis & Flint, 2009).

Molecular studies have provided unprecedented insights into the microbiota of the human gastrointestinal tract and have demonstrated differences between healthy persons and patients with gastrointestinal disorders like Crohn's disease and ulcerative colitis (Seksik *et al.*, 2003, Zoetendal *et al.*, 2008). During a study of the microbiota of intestinal biopsy material taken from Crohn's disease patients, isolates were identified by amplification and sequencing of the 16S rRNA genes (strain 5/1/63FAA) and were shown to share more than 99% 16S rRNA sequence similarity to two butyrate-producing strains recovered from human fecal samples in previous studies (strains SS2/1 and SSC/2) (Duncan *et al.*, 2004). These strains were identified as belonging to an unknown species distantly related to *Anaerostipes caccae* NCIMB 13811^T, and were found to employ the butyryl CoA: acetate CoA transferase route for butyrate formation, either from sugars or from D-lactate and acetate (Duncan *et al.*, 2004). During the course of our investigations the 16S rRNA gene sequence of *Eubacterium hadrum* ATCC 29173^T was made publically available for the first time and it was apparent that this sequence was almost identical to that of the three isolates under investigation. According to Tap *et al.* (2009), dominant gastrointestinal organisms closely related to strain 5/1/63FAA were found to be

among the 10 most abundant OTUs in 50% of study participants. The abundance of these organisms indicates an important contribution to the core microbiome. Analysis of strain 5/1/63FAA was performed in order to resolve the relationships of these isolates and further characterize this important member of the human colonic microbial ecosystem.

Based on phylogenetic, phenotypic and chemotaxonomic evidence, strains 80/3^T and strains SS2/1, SSC/2 , 5/1/63FAA were found to constitute two novel species within the *Ruminococcaceae and Lachnospiraceae* families respectively. In addition to the characterization and description of these novel genera and species, the aim of the present work was to resolve the taxonomy of this group of organisms using genetic analyses. The further taxonomic restructuring of these species and closely related organisms was necessary in order to facilitate the reordering of those ruminococci species misclassified within *Ruminococcaceae and Lachnospiraceae.*

Materials and Methods

Strain 80/3 ^{T} belonging to the family Ruminococcaceae

Strain and culture conditions

Stain 80/3^T was isolated initially from a fecal sample from a female participant who had not received therapeutic antibiotics or other products likely to influence gastrointestinal composition in the previous six months as previously described (Dabek *et al.*, 2008). The isolate was purified by serial dilution in roll tubes. It was then maintained on M2GSC (Dabek *et al.*, 2008) at 37°C under anaerobic head space ($N_2/CO_2/H_2$, 85:10:5). Provisional phenotypic and genome analyses were performed in a recent study (Wegmann *et al.*, 2014) (Fig 3).

Phenotypic and biochemical characterization

The Lawson lab received the strain after initial isolation. Strain 80/3^T was grown anaerobically in M2GSC broth (Wegmann *et al.*, 2014) at 4-60°C. Growth in 0-5% (w/v) NaCl was measured in 0.5% increments in M2GSC broth. The pH range for growth was investigated in M2GSC broth adjusted to pH 5-8.5 in increments of 0.5 pH units. Utilization of substrates was performed in basal cellulolytic (BC) medium (Robert & Bernalier-Donadille, 2003) with the addition of 5 g/L arabinose or 10 g/l of glucose, lactose, cellobiose, mannose, maltose, mannitol, raffinose and sucrose. All were performed in duplicate.

Cellular fatty acid methyl ester (FAME) analysis was performed on 72 hour growth in M2GSC medium. The methyl esters were extracted and analyzed by gas chromatography according to the Sherlock Microbial Identification System protocol (MIDI) (Kämpfer & Kroppenstedt, 1996). Fatty acids were identified and expressed in the form of percentages using the QTSA peak naming database.

Imaging

All imaging was performed as previously described (Wegmann *et al.*, 2014). For scanning electron micrograph images, cells of strain 80/3^T were fixed with glutaraldehyde, dehydrated through an ethanol series, critical point dried, attached to SEM stubs and gold coated. For transmission electron micrograph images, cells were fixed in glutaraldehyde, embedded in low melting point (LMP) agarose, post-fixed in osmium tetroxide, dehydrated through an ethanol series, embedded in acryl resin LR white ultrathin sectioned to coated copper 200 hex grids and stained.

Phylogenetic analysis of strains

The complete 16S rRNA gene sequence was obtained from DNA extracted from a standard phenol-chloroform method. Primers corresponding to positions 8-27 (forward primer) and 1492-1510 (reverse primer) of the *Escherichia coli* numbering system were used for amplification of the 16S rRNA gene (Hutson *et al.*, 1993). The amplified products were treated with ExoSAP-IT (USB Corporation) and directly sequenced with primers directed towards the conserved positions of the rRNA gene. This sequence was submitted to the EZTaxon server (http://

www.ezbiocloud.net/taxonomy (Yoon *et al.*, 2016). The 16S rRNA gene sequences of stain 80/3^T and type strains of closely related genera, including representative type strains of *Ruminococcaceae* and *Lachnospiracea* were aligned using the Clustal X interface and manually corrected using GeneDoc (Nicholas *et al.*, 1997). A phylogenetic tree was constructed according to the neighbor-joining method using SEQtools and TreeView (Saitou & Nei, 1987, Page, 1996). The stability of the groupings was estimated by bootstrap analysis (1000 replications) using the same programs. All major branching nodes were confirmed by maximum parsimony (data not shown). The trees were rooted using the 16S rRNA gene sequence of *Clostridium butyricum.* Additional analysis of the whole genomes from group I and II *Ruminococcaceae* was performed with the EZTaxon Average Nucleotide Identity (ANI) calculator (Yoon *et al.*, 2017). Each of the five strains and strain 80/3^T type strains were downloaded from NCBI and compared to one another. The percent similarity and amount differences of the number of total nucleotides was determined.

Strain 5/1/63FAA belonging to the family Lachnospiraceae

Strain and culture conditions

Strains SSC/2 and SS2/1 were isolated as previously described from one healthy female who had not taken any antibiotics or other medication known to influence colonic microbiota for a period of three months or more (Duncan *et al.*, 2004). Strain 5/1/63FAA was isolated from biopsy material taken from the ascending colon of a 51 year old Canadian female, diagnosed as suffering from Crohn's disease. The biopsy specimen showed no sign of inflammation at the time of collection. Samples were streaked to purity by continued passages on Fastidious Anaerobe Agar (Oxoid) supplemented with 5% sheep blood until isolated colonies were obtained. The strain was received by the Lawson lab after initial isolation.

Phenotypic and biochemical characterization

The strain was characterized biochemically by using the commercially available biochemical systems API Rapid ID 32A and API ZYM (bioMérieux) according to the manufacturers' instructions. All biochemical tests were performed in duplicate. Substrate utilization was determined by adding a final concentration of 0.5% filter-sterilized sugar stock solutions (10%, w/v) to yeast extract, casitone, fatty acid (YCFA) medium dispensed in 7.5 ml aliquots in Hungate-type tubes, inoculated with test strains and incubated for 48 h at 37°C at a pH of 6.5 under anaerobic conditions. YCFA medium consisted of (per 100 ml): 1 g casitone, 0.25 g yeast extract, 0.4 g $NaHCO_3$, 0.1 g cysteine, 0.045 g K_2HPO_4, 0.045 g KH_2PO_4, 0.09 g

$(NH_4)_2SO_4$, 0.09 g NaCl, 0.009 g $MgSO_4$ 7H2O, 0.009 g $CaCl_2$, 0.1 mg resazurin, 1 mg haemin, 1 mg biotin, 1 mg cobalamin, 3 mg p-aminobenzoic acid, 5 mg folic acid and 15 mg pyridoxamine (Lopez-Siles *et al.*, 2012). Tested sugar substrates were arabinose, cellobiose, fructose, glucose, maltose, melezitose, raffinose, rhamnose, ribose, sucrose and xylose. Acid end products from glucose metabolism were determined by capillary GC (Richardson *et al.*, 1989).

To further characterize the strain, analysis of fatty acids and mol% G + C was performed on 48 h cultures grown on Brucella blood agar plates. Cellular FAME products were extracted using the Sherlock Microbial Identification System (MIDI Labs Inc, Newark, DE) version 6.1 and analyzed by gas chromatography as described previously (Kämpfer & Kroppenstedt, 1996). Identification of fatty acids was determined using the QBA1 peak naming database. The G + C content (mol%) was determined by HPLC according to the method of Mesbah et al. (Mesbah *et al.*, 1989)

Imaging

Phase contrast images of strain 5/1/63FAA were taken using Leica Applications Suite (LAS EZ Version 1.7.0) software as previously described (Allen-Vercoe *et al.*, 2012). Transmission electron micrographs were imaged on a Philips CM10 electron microscope. Cells were stained with 0.5% uranyl acetate.

Phylogenetic analysis of strains

Chromosomal DNA was purified by a standard phenol-chloroform method. For amplification of the 16S rRNA gene, a universal primer, corresponding to positions 8-27 (forward primer) and 1492-1510 (reverse primer) of the *Escherichia coli* numbering system was used (Hutson *et al.*, 1993). The amplified products were treated with ExoSAP-IT to remove unincorporated primers and dNTPs (USB Corporation) and directly sequenced with primers directed towards the conserved positions of the rRNA gene. Approximately 1425 bases of the 16S rRNA gene were determined.

The closest known related strains were retrieved from GenBank and EzTaxon databases and aligned with the newly determined sequences using the program SEQtools (http://www.ezbiocloud.net/eztaxon; Yoon *et al*., 2017, Kim *et al*., 2012, Rasmussen, 2002). The resulting multiple sequence alignment was corrected manually using the program GeneDoc (Nicholas *et al.*, 1997), and a phylogenetic tree was constructed according to the neighbor-joining method (Saitou & Nei, 1987) with the programs SEQtools and TREEVIEW (Page, 1996). The stability of the groupings was estimated by bootstrap analysis (1000 replications). All major branching nodes were confirmed by maximum parsimony (data not shown). Additional analysis of the genomes belonging to the strains and additional strains was performed using the EZTaxon Average Nucleotide Identity (ANI) calculator (Yoon *et al.*, 2017).

Results

Strain 80/3^T belonging to the family Ruminococcaceae

Growth occurred in pH 7-8, NaCl 0-3.0% and 30-37°C. The ability of strain 80/3^T, to utilize substrates, in addition to those carbohydrates previously tested, was determined. Utilization of glucose, lactose, mannose and cellobiose was observed, but no growth was observed with other carbohydrates, including cellulose, xylan, starch, maltose, raffinose, sucrose, and xylose.

Acetate, ethanol and formate were found to be the major end products as described previously (Wegmann *et al.*, 2014). Interestingly, unlike similar species of ruminococci found in the bovine rumen and human colon, strain 80/3^T did not utilize cellulose or xylan, but did have the specialized ability to utilize β-glucans and xyloglucan (Wegmann *et al.*, 2014). The fatty acid profile included major fatty acids $C_{16:0}$ and $C_{17:0\ \omega7c}$ *cyclo* (Table 1). The phenotypic characteristics of strain 80/3^T were compared with those of its closely related taxa in Table 2. Unlike these closely related species, strain 80/3^T did not to ferment cellulose and xylan and also did not ferment sucrose as seen in *R. albus*. Major fermentation product differences were also observed among related species. All strains compared here produced acetate but only strain 80/3^T, *R. albus* and *R. flavefaciens* produced ethanol. Of the group I *Ruminococcus* species, only strain 80/3^T and *R. flavefaciens* produced formate.

Table 1. Cellular fatty acid composition of strain $80/3^T$.

Cellular fatty acid	**$80/3^T$**
Saturated	
$C_{12:0}$	3.2
$C_{13:0}$	5.4
$C_{14:0}$	4.5
$C_{14:0}$ iso	1.9
$C_{15:0}$ iso	3.9
$C_{15:0}$ anteiso	6.1
$C_{15:0}$	4.3
$C_{16:0}$ iso	2.3
$C_{16:0}$	**24.8**
$C_{17:0}$	1.4
Unsaturated	
$C_{17:0\ \omega 7c}$ *(cyclo)*	**18.1**
$C_{19:1}$ iso	1.9
Hydroxy acid	
$C_{16:0}$ 3-OH	2.0
Summed features[a]	
Summed feature 3	1.7
Summed feature 5	3.2
Summed feature 8	6.8
Summed feature 11	1.0

Only those fatty acids that make up >1% are shown

[a] Summed feature 3 is comprised of $C_{16:1}$ ω7c/$C_{16:1}$ ω6c; summed feature 5 is comprised of $C_{18:0}$ ante/$C18_{:2}$ *ω6,*9c; summed feature 8 is comprised of $C_{18:1}$ ω9c/$C_{18:1}$ ω7c/$C_{18:1}$ ω6c; summed feature 11 is comprised of $C_{19:0}$ cyclo ω8c/$C_{20:0.}$

Products in bold denote predominant fatty acids.

Table 2. Comparison of differential characteristics of strain 80/3^{T} and closely related group I *Ruminococcus* species

Characteristic	80/3^{T}	*R. albus*	*R. champanellensis*	*R. flavefaciens*	*R. callidus*	*R. bromii*
Source	**Human**	**Human**	**Human**	**Ruminants**	**Human**	**Human, cow, pig**
Major Fermentation products*	**A,E,F**	**A,E**	**A,S**	**A,F,E,S**	**S,A**	**A**
G+C (mol%)	**43**	**43-46**	**53**	**39-44**	**43**	**39-40**
Fermentation of:						
Cellulose	-	+	+	+	-	**ND**
Xylan	-	+	+	+	+	**ND**
Arabinose	-	-	-	-	-	-
Starch	-	-	-	-	+	+
Glucose	+	+	-	-	+	+
Cellobiose	+	+	+	+	+	-
Lactose	+	+	-	+	+	-
Mannose	+	+	-	-	-	**w/-**
Maltose	-	-	-	-	+	+
Mannitol	-	-	-	-	-	**ND**
Raffinose	-	-	-	-	+	-
Sucrose	-	+	-	-	+	-
Xylose	-	-	-	-	**w/-**	-

*Major fermentation products: A, acetate; E, ethanol; F, formate; S, succinate
Data from this study, Chassard *et al.* (2012) and Wegmann *et al.* (2014)

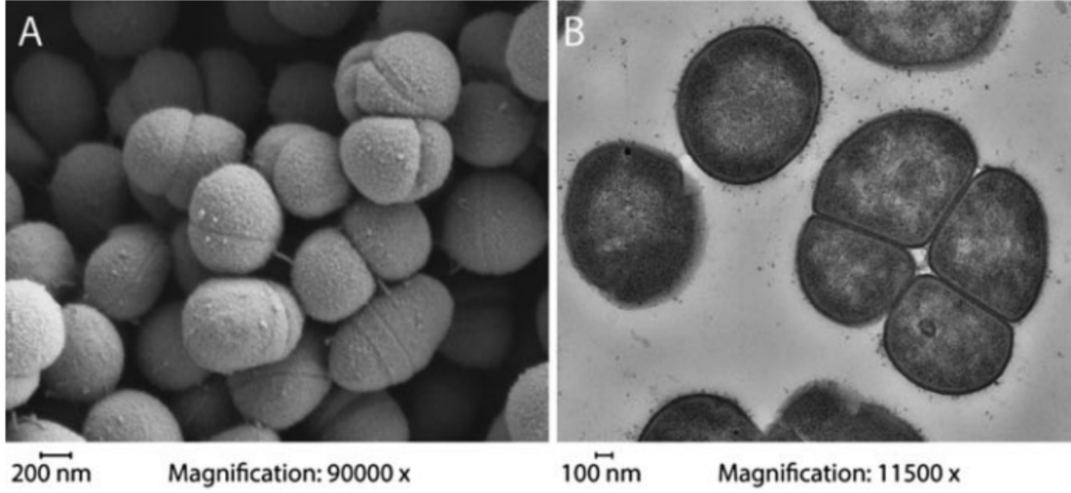

Figure 4. A) Scanning electron micrograph image of strain 80/3^T B) Transmission electron micrograph image of strain 80/3^T cells grown in YCFAG medium (Wegmann *et al.*, 2014).

The results of the 16S rRNA based analyses clearly identify strain **80/3^T** as a member of the family *Ruminococcaceae* and is most closely related to *Ruminococcus albus* with 94.2% 16S rRNA gene sequence similarity (Table 3). Lower sequence similarities were obtained with members of a separate lineage that contain the type strain *R. flavefaciens* (91.8%), *R. champanellensis* (91.8%) and *R. callidus* (90.7%). Futhermore, *R. bromii* is located in a separate line of decent with only 88.1% sequence similarity. Based on recommendations of Yarza *et al.* (2014) a sequence similarity of less than 94% to delineate genera is now regularly used in taxonomy.

Table 3. Comparison of 16S rRNA gene sequence similarity between group I ruminococci type strains.

Strain Similarity	Strain 80/3^T	*R. flavefaciens*	*R. albus*	*R. champanellensis*	*R. callidus*
Strain 80/3^T	-				
R. flavefaciens	91.8%	-			
R. albus	94.2%	91.6%	-		
R. champanellensis	91.9%	94.9%	92.0%	-	
R. callidus	90.7%	93.3%	89.8%	94.1%	-
R. bromii	88.1%	89.1%	89.0%	89.8%	89.0%

Wegmann *et al*. 2014 determined the complete genome of strain 80/3^T. The DNA G+C of strain 80/3^T was 43 mol%. The genome consisted of two circular chromosomes. The split of the genome into two chromosomes is thought to shorten the time needed for DNA replication supporting faster growth and could provide a selective advantage favoring the multireplicon genome architecture (Wegmann *et al.,* 2014).

The phylogenetic analyses performed in this study revealed not only that strain 80/3 represented a novel species of a novel genus, but also demonstrated that the genus *Ruminococcus* was in need of restructuring. It is proposed that strain 80/3^T and *R. albus* be classified, along with another member of misclassified ruminococci, to two novel genera. Together, strain 80/3^T and *R. albus* form a distinct lineage within the subcluster of several recognized group I *Ruminococcus* species that are closely related to each other but distinct enough from *R. flavefaciens* genera to merit the formation of a novel genera. Likewise, strain 80/3^T was found to have 16S rRNA

gene similarities with the true ruminococci and was only distantly related to *R. bromii*.

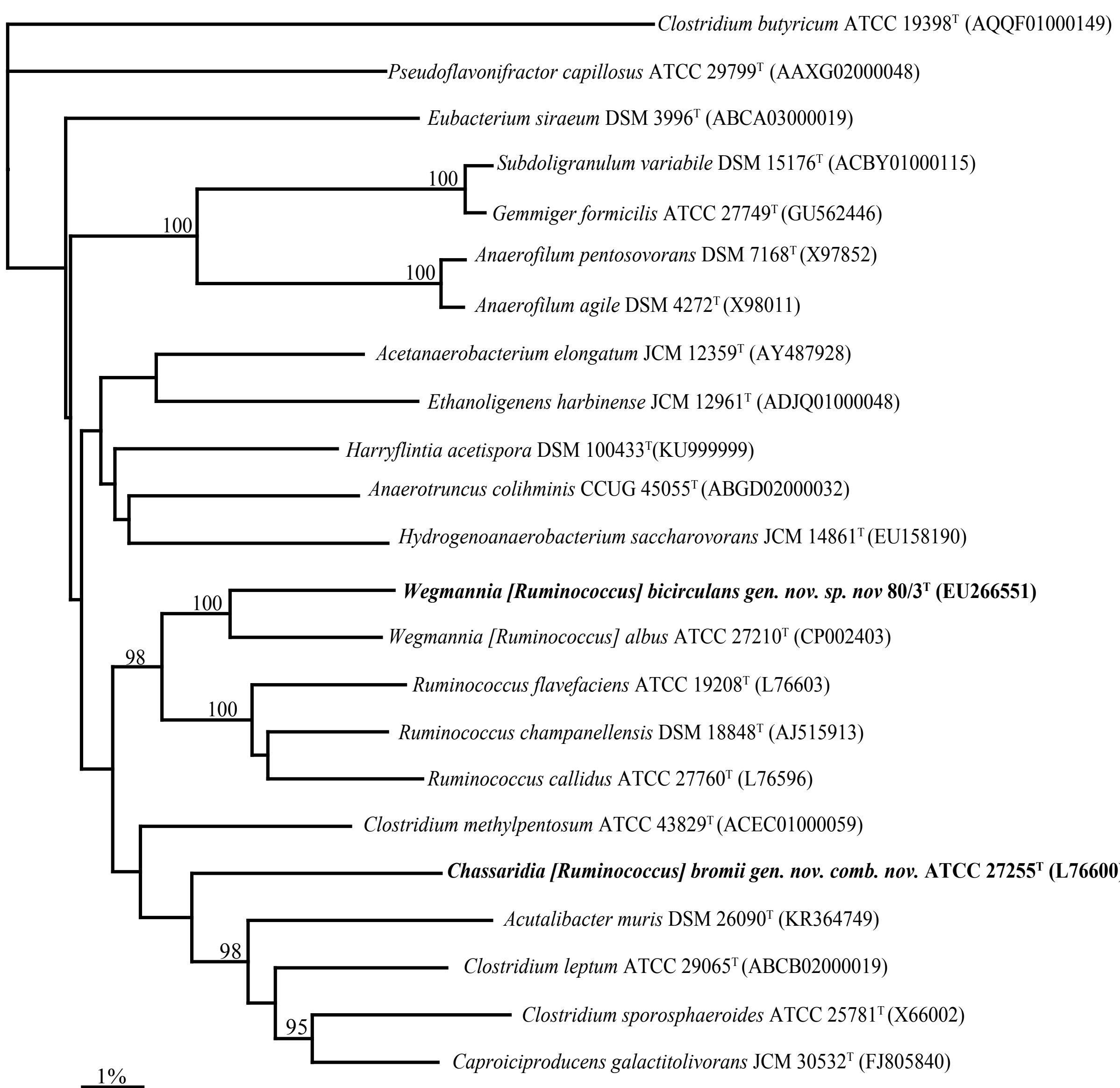

Figure 5. Phylogenetic tree showing the relationship of strain 80/3^{T} to closely related members of the family *Ruminococcaceae*. *C. butyricum* was used as the out-group. The scale bar indicates 1% sequence divergence.

As previously suggested (La Reau *et al.*, 2016), *R. bromii* is not closely related to the true ruminococci, but is more closely related to *Clostridium* spp. and forms an entirely separate genus (Fig 5). At the time when *R. bromii* was first described, the strain was placed incorrectly within the genus *Ruminococcus.* As this was prior to the development of molecular methods *R. bromii* was assigned to the *Ruminococcus* genus based on its culture and metabolic characteristics although the authors acknowledged the differences between the strain and *Ruminococcus* species (Moore *et al*., 1972).

Based on recent recommendations, it is thus proposed that those species that do not share close phylogenetic relationship (~89-91%) to *R. flavefaciens* should be reclassified into two novel genera within the *Ruminococcaceae.* In addition to initial genome description data, this polyphasic taxonomic approach utilizing 16S rRNA gene sequence analysis indicates that strain $80/3^{T}$ should be placed in a novel genus with *R. albus,* and that *R. bromii* should be placed in a separate but related genus to which the names *Wegmannia* and *Chassardia* are respectively proposed.

The remaining members of group I ruminococci are those species limited to *Ruminococcus sensu stricto R. flavefaciens*, *R. champanellensis* and *R. callidus*. These strains form their own branch within the group I ruminococci with a ~94% 16S rRNA similarity (Table 3). This value is on the cusp of the values used for the delineation of genera and arguments can be made for inclusion into a single genus or the separation into novel genera.

The average nucleotide identity (ANI) value was determined to be 68-70% and G+C content of the DNA is 39-53 mol% (Table 4-6). While there is no ANI % cutoff for species within genera description consistent with the 95 % ANI value corresponding to the 70 % DDH recommendation for species cutoff, the low ANI values do support the separation of these strains into novel genera. However, at this time and until additional species are described, these species should be retained in the genus *Ruminococcus*. Additional strains may reveal diagnostic and unique traits that would allow further restructuring of this genus.

Table 4. Group I ruminococci genome summary

Species	Strain	Contigs	Total length (bp)	A	C	G	T	N	GC ratio
Strain 80/3^{T}	-	2	2,968,500	857,135	610,929	654,375	846,061	0	42.6%
R. flavefaciens	ATCC 19208	25	3,587,8031	1,019,835	762,688	783,705	1,015,150	6,425	43.1%
R. callidus	ATCC 27760	226	3,096,487	779,775	752,637	754,474	783,801	25,800	49.1%
R. chapanellensis	18P13	1	2,573,208	585,343	649,573	693,827	584,638	59,827	53.5%
R. albus	7	5	4,482,087	1,280,014	961,345	993,531	1,247,197	0	43.6%
R. bromii	L2-63	1	2,249,085	656,728	458,441	464,908	650,960	18,048	41.4%

Table 5.1. *Ruminococcus sensu stricto* whole genome sequence comparison

Strain 1	Strain 2	Average aligned length (bp)
Ruminococcus flavefaciens	*Ruminococcus callidus*	347,274
Ruminococcus flavefaciens	*Ruminococcus champanellensis*	340, 093
Ruminococcus callidus	*Ruminococcus champanellensis*	448, 485

Table 5.2. *R. flavefaciens* and group I misclassified ruminococci whole genome sequence comparison

Strain 1	Strain 2	Average aligned length (bp)
Ruminococcus flavefaciens	Strain 80/3^{T}	315,687
Ruminococcus flavefaciens	*Ruminococcus albus*	438,902
Ruminococcus flavefaciens	*Ruminococcus bromii*	185,564
Strain 80/3^{T}	*Ruminococcus albus*	587,380
Strain 80/3^{T}	*Ruminococcus bromii*	263,381
Ruminococcus albus	*Ruminococcus bromii*	169,531

Table 6. Comparison of group I ruminococci strains ANI data OrthoANIu values

Strain similarity	Strain 80/3^T	*R. flavefaciens*	*R. albus*	*R. champanellensis*	*R. callidus*
Strain 80/3^T	-				
R. flavefaciens	**69.4%**	-			
R. albus	**71.8%**	**70.6%**	-		
R. champanellensis	**68.4%**	**68.7%**	**68.4%**	-	
R. callidus	**68.2%**	**68.3%**	**67.9%**	**70.2%**	-
R. bromii	**70.4%**	**67.5%**	**67.3%**	**67.1%**	**67.8%**

The misplaced species in group II ruminococci - *R. gnavus, R. gaureauii, R. lactaris, R. faecis* and *R. torques* form multiple distinct novel genera based on phylogenetic analysis (Fig 2) and shown in Table 7. Likewise, the ANI values of whole genome analysis of the group II indicated multiple genera (Table 8 and 9). *R. lactaris, R. faecis* and *R. torques* likely form one novel genus (misclassified I), sharing 96.2-96.3% similarity of 16S rRNA genes and 73.4-77.5% similarity of ANI (Table 10). Likewise, phylogenetic analysis suggests that *R. gnavus* and *R. gaureauii* warrant two novel genera, misclassified II and misclassified III, respectively. Misclassified II and misclassified III have distinct lineages from misclassified I and each other.

Table 7. Comparison of 16SrRNA gene of misclassified *Lachnospiraceae*

Strain Similarity	*R. faecis*	*R. torques*	*R. lactaris*	*R. gauvreauii*
R. faecis	-			
R. torques	96.3%	-		
R. lactaris	96.2%	96.3%	-	
R. gauvreauii	93.2%	93.0%	92.8%	-
R. gnavus	94.4%	93.9%	94.8%	94.1%

Table 8. Group II (*Lachnospiraceae*) ruminococci genome information

Species	Strain (Type)	Contigs	Total length (bp)	A	C	G	T	N	GC ratio
R. torques	ATCC 27756	41	2,741,706	810,799	553,228	596,540	778,839	2,300	41.9%
R. lactaris	ATCC 29176	47	2,731,235	799,145	559,932	605,584	765,074	1,500	42.7%
R. faecis	JCM 15917	45	3,259,449	1,016,649	589,366	752,259	901,175	0	41.2%
R. gauvreauii	DSM 19829	70	4,094,708	1,089,022	933,073	1,016,031	1,056,152	430	47.6%
R. gnavus	ATCC 29149	43	3,501,911	994,994	764,635	736,886	1,005,396	0	42.9%

Table 9. Group II misclassified ruminococci whole genome sequence comparison

Strain 1	Strain 2	Average aligned length (bp)
Ruminococcus faecis	*Ruminococcus torques*	672,774
Ruminococcus faecis	*Ruminococcus lactaris*	867,818
Ruminococcus faecis	*Ruminococcus gauvreauii*	364,715
Ruminococcus faecis	*Ruminococcus gnavus*	810,075
Ruminococcus torques	*Ruminococcus lactaris*	738,618
Ruminococcus torques	*Ruminococcus gauvreauii*	374,106
Ruminococcus torques	*Ruminococcus gnavus*	774,508
Ruminococcus lactaris	*Ruminococcus gauvreauii*	376,641
Ruminococcus lactaris	*Ruminococcus gnavus*	729,932
Ruminococcus gauvreauii	*Ruminococcus gnavus*	458,712

Table 10. Comparison of Group II ruminococci strains ANI data OrthoANIu values

Strain Similarity	*R. flavefaciens*	*R. faecis*	*R. torques*	*R. lactaris*	*R. gauvreauii*
R. flavefaciens	-				
R. faecis	**66.1%**	-			
R. torques	**65.3%**	**73.8%**	-		
R. lactaris	**65.7%**	**77.5%**	**73.4%**	-	
R. gauvreauii	**64.4%**	**68.8%**	**68.7%**	**69.5%**	-
R. gnavus	**65.6%**	**74.2%**	**73.7%**	**73.9%**	**69.1%**

These genera divisions are not well resolved phylogenetically as the later genera are only represented by a single species and should serve only as guidance for future studies. An emphasis should be placed on these species as additional novel strains within this group are isolated and the phylogeny of the genera can be further resolved as additional strains within the genera are characterized.

Genus and species descriptions

Description of *Wegmannia* gen. nov.

Wegmannia (Weg.man.nia N.L. fem. n. Wegmannia in honor of Udo Wegmann, a German microbiologist for his contributions to the colonic microbiota). Cells are Gram-positive non motil cocci. Bacteria are anaerobic, optimal growth is observed at 37°C and pH tolerance ranges from 6-8. Fermentation of glucose, cellobiose, lactose and mannose. The G+C content of the DNA is 43-46 mol%. Strains are isolated form human feces. The type species of the genus is *Wegmannia albus*.

Description of *Wegmannia bicirculans* sp. nov.

Wegmannia bicirculans (bi.circ'.u.lans) The description of *Wegmannia birciculans* is identical to that proposed previously (Fig 4) (Dabek *et al.*, 2008, Wegmann *et al.*, 2014). In addition to the characteristics given in the genus description, colonies grown on M2 agar plates were opaque grey in color. Its predominant cellular fatty acids are $C_{16:0}$ and $C_{17:0\ \omega 7c}$ *cyclo*. Positive for fermentation of glucose, cellobiose, lactose and mannose. The type strain was isolated from the fecal sample of a healthy, adult female. The DNA G+C content of the type species is 43%. The whole genome sequence was deposited in the EMBL database with the accession numbers HF545616 and HF545617 for chromosome I and chromosome II, and the 16S rRNA sequence deposited with the number EU26655. The type strain is $80/3^T$.

Description of *Wegmannia albus* comb. nov.

Wegmannia albus

Basonym: R. albus (al'bus. L. adj. albus white.) (Hungate, 1957).

The description of *Wegmannia albus* is identical to that proposed for *Ruminococcus albus* (Hungate, 1957). Cells are Gram-poisitive-staining and non motile cocci. Positive for fermentation of cellulose, xylan, glucose, cellobiose, lactose, mannose and sucrose. Major fermentation products are acetate and ethanol. The DNA G+C content of the type species is 43-46%. The type strain is 7 (= ATCC 27210^T = DSM 20455^T = JCM 14654^T) isolated from human feces.

Description of *Chassaridia bromii* gen. nov. comb. nov.

Chassardia bromii (Chass.ar.di'a. M.L. fem. n. Chassardia in honor of Christopher Chassard, a French microbiologist, in recognition of his contributions to human gastrointestinal microbiology). The description of *Chassardia bromii* is identical to *R. bromii* (Moore *et al.*, 1972). The type strain is ATCC 27255^T isolated from the feces of humans, cows and pigs.

Strain 5/1/63FAA belonging to the family Lachnospiraceae

Colonies of strains ATCC 29173^{T} , SS2/1^{T} and SSC/2^{T} appeared white-opaque with a translucent, irregular margin, flat, rough and 1-3 mm diameter in size when grown on Brucella blood agar after 24 h at 37°C. Under the same incubation conditions, colonies of 5/1/63FAA were grey, round, convex and glossy, with a ridged margin, and 1-3 mm diameter in size. The cells of all strains consisted of Gram-positive-staining, non-motile, non-spore-forming, rod shaped cells (on average, 4.75 x 0.75 mm in size) (Fig 6). Growth was observed with glucose, cellobiose, maltose, fructose or sucrose, and weak growth with melezitose or raffinose, as energy sources. No growth was obtained with arabinose and rhamnose, while all strains grow well on xylose and none used ribose.

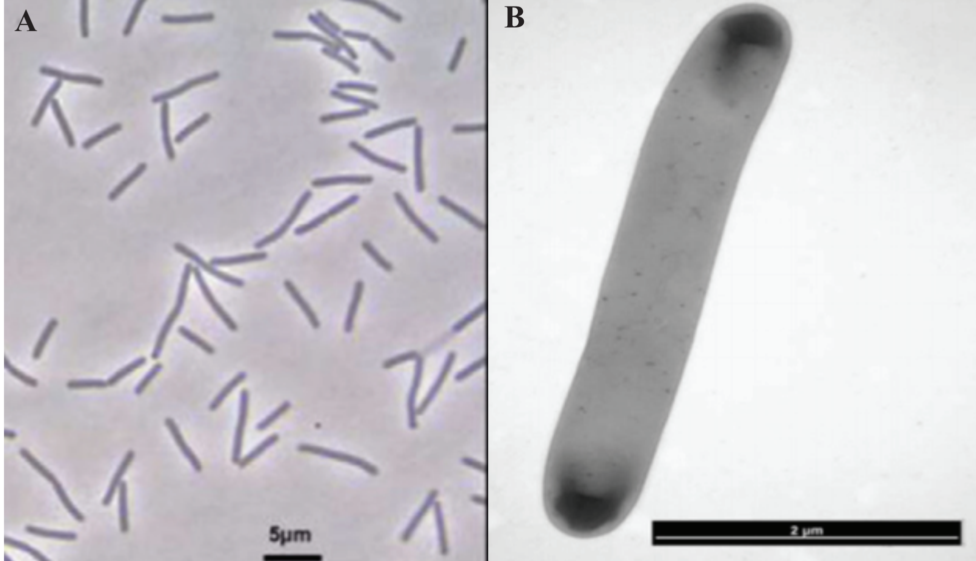

Figure 6. A) Phase contrast image of strain 5/1/63FAA. B) Transmission electron micrograph image of strain 5/1/63FAA (Allen-Vercoe *et al.*, 2012).

Strains SS2/1^T and SSC/2^T have been shown to utilize D-lactate and acetate to form butyrate (Duncan *et al.*, 2004). *E. hadrum* ATCC 29173^T was shown here to have the same ability to utilize acetate, as well as the D- but not the L- isomer of lactate. This contrasts with the current description of *E. hadrum*, possibly because lactate utilization requires the presence of acetate in the medium. D-lactate was estimated by measuring total DL-lactate and L-lactate (Schwiertz *et al.*, 2002, Duncan *et al.*, 2004).

Using the rapid ID 32A, almost identical profiles were obtained for the four strains, which were generally unreactive (Table 11). Positive reactions were obtained for β-galactosidase, indole (weak) and alkaline phosphatase (weak). In addition, strain 5/1/63FAA showed a positive reaction for α-galactosidase (weak).

All other tests were negative. Likewise, the strains also gave mostly negative reactions using the API ZYM system (Table 12), only giving weakly positive reactions for acid phosphatase, naphthol-AS-BI-phosphohydrolase, and strain 5/1/63FAA also gave a weak reaction for α-galactosidase. The end products of glucose metabolism were butyrate (11.07± 0.95mM) and formate (7.61± 1.48mM) (mean of triplicate incubations. All strains possessed both the ability to produce and utilize acetate and net utilization of acetate is often observed, dependent on growth conditions.

Table 11. API Rapid 32A differential reactions of strains *E. hadrum* ATCC 29173^T and strains SS2/1 ^{T}, SSC/2^T and 5/1/63FAA

Reactions/Enzymes	ATCC 29173^T	SS2/1	SSC/2	5/1/63FAA
Arginine dihydrolase	$+^w$	$+^w$	-	-
α-Galactosidase	-	-	-	$+^w$
β-galactosidase-6-phosphate	$+^w$	-	-	-
α-Glucosidase	-	-	-	$+^w$
Indole production	-	$+^w$	$+^w$	$+^w$

+, positive reaction; - negative reaction; w, weak reaction.

Table 12. API ZYM differential reactions of strains *E. hadrum* ATCC 29173^T and strains SS2/1^T, SSC/2^T and 5/1/63FAA

Reactions/Enzymes	ATCC 29173^T	SS2/1	SSC/2	5/1/63FAA
Alkaline phosphatase	$+^w$	-	-	-
Esterase	$+^w$	-	-	-
Esterase lipase	$+^w$	-	-	-
Acid phosphatase	+	+	$+^w$	$+^w$
Naphthol-AS-BI-phosphohydrolase	+	$+^w$	$+^w$	$+^w$
α-Galactosidase	-	-	-	$+^w$

+, positive reaction; - negative reaction; w, weak reaction.

.

Predominant cellular fatty acids were $C_{12:0}$, $C_{17:0}$ and summed features 2 and 10 (Table 13). Strain differences may reflect real differences in cell membrane composition. The G + C mol% was determined to be 37.0-41.8%. In the original description the G + C mol% of *E. hadrum* ATCC 29173^T was reported as 32-33% (Moore *et al.*, 1976). This difference is likely due to determination of G + C mol% by biochemical means and the recognition that such methods are subject to experimental error, especially between different laboratories. The initial value of 32-33% is therefore likely to be erroneous.

Table 13. Cellular fatty acid composition of *E. hadrum* ATCC 29173[T] and strains SS2/1 [T], SSC/2[T] and 5/1/63FAA

Cellular fatty acid	ATCC 29173[T]	SS2/1 [T]	SSC/2[T]	5/1/63FAA
Saturated straight chain				
$C_{10:0}$	7.7	2.2	2.4	1.5
$C_{12:0}$	**23.5**	**41.4**	**33.5**	**24.0**
$C_{13:0}$	2.2	-	-	-
$C_{14:0}$	4.5	2.8	2.8	2.0
$C_{16:0}$	4.6	5.7	7.8	8.5
$C_{17:0}$	3.6	8.4	**12.7**	**14.2**
Unsaturated straight chain				
$C_{17:1\ \omega 8c}$	-	1.1	-	2.0
$C_{17:1\ \omega 6c}$	-	2.8	2.1	2.3
Hydroxy acid				
$C_{11:0}$ 3-OH	1.1	5.1	2.2	-
Summed features[a]				
Summed feature 2	**20.2**	**12.0**	**13.4**	6.7
Summed feature 4	3.3	-	-	-
Summed feature 8	-	-	-	2.4
Summed feature 10	4.5	7.2	12.7	**27.6**
Summed feature 11	4.3	-	-	1.6

Only those fatty acids that make up >1% are shown.
[a] Summed feature 2 is comprised of $C_{14:0}$ 3-OH, $C_{16:1}$ iso l; summed feature 4 is an unknown (likely $C_{15:0}$); summed feature 8 is comprised of $C_{18:1\ \omega 9c}$, C18:1 *ω6c;* summed feature 10 is comprised of $C_{17:0}$ cyclo and/or $C_{18:0}$, $C_{17:0}$ cyclo. Products in bold denote predominant fatty acids.

Phylogenetic analysis of the strains demonstrated they were almost identical based on 16S rRNA gene sequence similarity. A tree constructed using the neighbor-joining method demonstrated that the organism is related to members of the genus *Anaerostipes* of the family *Lachnospiraceae* (formerly clostridial cluster XIVa) within the phylum *Firmicutes* (Schwiertz *et al.*, 2002). The most closely related species are *A. butyraticus* LMG 24724[T] (94.7% sequence similarity), *A. caccae* NCIMB 13811[T] (93.5% sequence similarity), and *A. rhamnosivorans* (93.4% sequence similarity), with the latter two organisms sharing a branching node (Fig 7).

This cluster was supported by a significant bootstrap value of 100%. While there is no precise correlation between 16S rRNA sequence divergence and species delineation, it is generally recognized that divergence values of 3% or more are significant (Stackebrandt & Ebers, 2006, Tindall *et al.*, 2010). More recently, Kim *et al.* (2014) suggests that a 98.7 % 16S rRNA gene sequence similarity can be used as the threshold for differentiating two species. Additional ANI analysis of the strains supported the strains belonging to one species (Table 14). The strains described here should therefore be regarded as belonging to a novel species within the genus *Anaerostipes,* that should now be designated *Anaerostipes hadrus*.

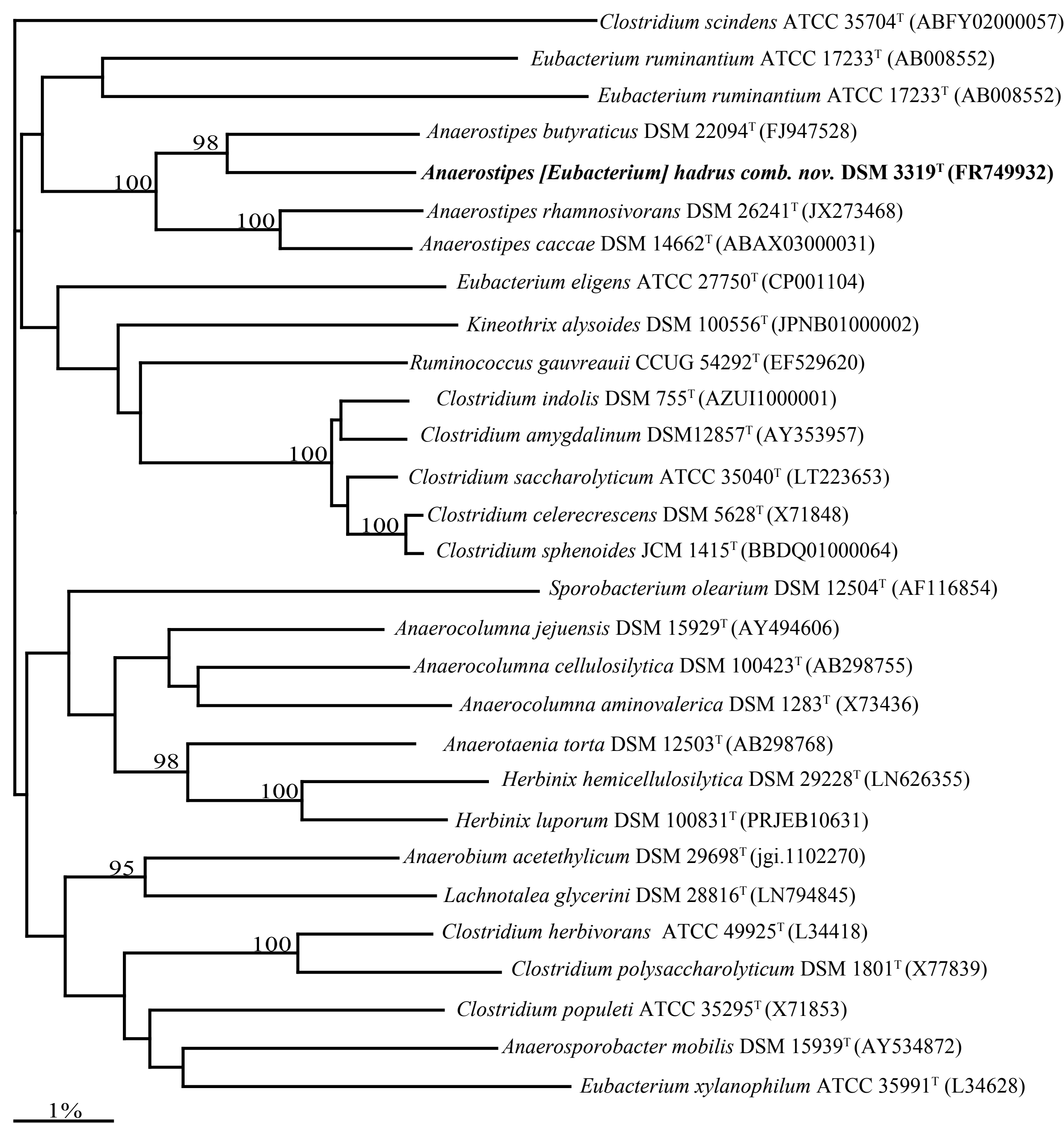

Figure 7. Phylogenetic relationship of *Anaerostipes hadrus* comb. nov. with closely related members of the family *Lachnospiraceae.* Bootstrap values, expressed as a percentage of 1000 replications, are given at branching nodes. Only values above 90% are shown. Scale bar = 1% sequence divergence.

Table 14. Comparison of *Anaerostipes hadrus* strains by ANI analysis

Strain Similarity	**DSM 3319**	**5/1/63FAA**	**SS2/1**	**SSC/2**	**BPB5**	**PEL 85**	**AH2**	**AH3**	**AH4**
DSM 3319^T	**-**								
5/1/63FAA	**98.5%**	**-**							
SS2/1	**98.6%**	**98.7%**	**-**						
SSC/2	**98.6%**	**98.7%**	**99.8%**	**-**					
BPB5	**98.3%**	**98.6%**	**98.4%**	**98.5%**	**-**				
PEL 85	**98.5%**	**98.6%**	**98.6%**	**98.7%**	**98.4%**	**-**			
AH2	**98.3%**	**98.6%**	**98.6%**	**98.7%**	**98.4%**	**97.8%**	**-**		
AH3	**98.7%**	**98.7%**	**98.9%**	**98.8%**	**98.5%**	**98.6%**	**98.3%**	**-**	
AH4	**98.7%**	**98.6%**	**98.7%**	**98.7%**	**98.4%**	**98.5%**	**98.4%**	**98.7%**	**-**
AH5	**98.5%**	**98.6%**	**98.7%**	**98.6%**	**98.4%**	**98.6%**	**98.1%**	**98.7%**	**98.4%**

Anaerostipes hadrus strains: AH2, 3, 4 and 5 are 2789STDY5834908, 2789STDY5834860, 2789STDY5834959, 2789STDY5608868 respectively.

In addition to its unique 16S rRNA gene sequence, A. *hadrus* can clearly be distinguished phenotypically from previously described *Anaerostipes* species. Biochemically *A. hadrus* can be distinguished from *A. butyraticus* LMG 24724^T and *A. caccae* NCIMB 13811^T by the differences in the reactions to β-galactosidase, indole production and alkaline phosphatase. In addition, while both *A. caccae* NCIMB 13811^T and *A. hadrus* are known to utilize lactate, *A. hadrus* differs from *A. caccae* L1- 92^T in using only the D-isomer while the latter can use both the D and L-isomers (Duncan *et al.*, 2004).

Since its original isolation, the significance of strains belonging to *A. hadrus* within the human GI tract has been reported. In the healthy colon, this newly defined species may represent more than 2% of the total colonic microbiota (Flint *et al.*, 2012). Walker *et al.* (2011) reported in studies using 16S rRNA sequencing to examine the response of different bacterial groups to changes in diet, *A. hadrus* was one of the 10 most abundant species in fecal samples from 6 overweight human volunteers (Walker *et al.*, 2011). Likewise, *A. hadrus* represented one of the four most prevalent OTUs among butyryl CoA: acetate CoA transferase and 16S rRNA genes amplified from samples from 10 non-obese human volunteers (Louis *et al.*, 2010). This species has been shown to use D-lactic acid at mildly acidic conditions to form butyric acid, which is the major energy source for colonocytes (Belenguer *et al.*, 2011). It has been estimated that approximately 20% of the butyric acid pool in the large intestine is derived from lactic acid metabolism. The ability to produce

butyric acid from lactic acid and the high abundance level indicate that the organism likely plays an important functional role in this complex ecosystem.

Description of *Anaerostipes hadrus* comb. nov.

Anaerostipes hadrus (had'rus. N.L. adj. *hadrus* (from Gr. N. adj. hadron thick, bulky referring to the relatively large size of the cell).

Physiological data and phylogenetic analysis based on 16S rRNA genes sequences support the reclassification of *E. hadrum* as *A. hadrus* comb. nov. In addition to the description given by Moore et al. (1976), cells were Gram-positive staining rods, non-motile and that are approximately 3 x 0.5 mm in size (Moore *et al.*, 1976). Endospores were not observed. On Brucella agar supplemented with blood after 24h at 37°C the colonies of strains SS2/1, and SSC/ 2 appear white-opaque with a translucent, irregular margin, flat, rough and 1-3 mm diameter in size. Under the same cultivation conditions, colonies of 5/1/63FAA are grey, round, convex and glossy, with a ridged margin, and 1-3 mm diameter in size. In broth cultures better growth is sometimes observed at pH 5.5 than either pH 6.2 and 6.7. Indole produced (weak). Nitrate reduced to nitrite. Strains produced formate (2-8 mM) and butyrate (8-12 mM) on YCFA medium containing 0.2% glucose. The species grew and produced butyrate from acetate and lactate in YCFA medium in the absence of carbohydrate.

The species showed positive reactions of API Rapid ID 32A for β-galactosidase, indole (weak), alkaline phosphatase (weak). In addition, strain 5/1/63FAA showed a positive reaction for α-galactosidase (weak). No activity was detected for urease, arginine dihydrolase, β-galactosidase-6-phosphate, β-glucosidase, α-arabinosidase, β-glucuronidase, N-acetyl-β-glucosaminidase, mannose, raffinose, glutamic acid decarboxylase, α-fucosidase, arginine arylamidase, proline arylamidase, leucyl glycine arylamidase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase, tyrosine arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase and serine arylamidase

The API ZYM system only gave weakly positive reactions for acid phosphatase, naphthol-AS-BI-phosphohydrolase, and strain 5/1/63FAA also gave a weak reaction for α-galactosidase. Enzyme activity was not detected for alkaline phosphatase, for esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, β-galactosidase, β-glucuronidase, α-glucosidase, b-glucosidase, N-acteyl-β-glucosaminidase, α-mannosidase or α-fucosidase. A variable reaction is obtained with α-galactosidase. The major long-chain fatty acids are $C_{12:0}$, $C_{17:0}$, and summed feature 2. The G+C content is 37-42 mol%.The type strain is ATCC 29173^{T} (=DSM 3319^{T} =VP 82-52).

Discussion

As the majority of organisms within the human gastrointestinal system are uncultivated species, understanding the role novel organisms play within the gastrointestinal microbial community is essential to determine the relationship with host health. The characterization and taxonomic restructuring of the numerically dominant organisms is necessary to determine their importance within the system. Within the gut microbiome, it is recognized that the limitations of the hosts' ability to produce butyrate and degrade plant polysaccharides requires a diverse array of gut bacteria with the ability to perform these functions.

The organisms described here are among the most abundant 16S rRNA phylotypes of the human gut (Tap *et al.*, 2009, Qin *et al.*, 2010, Walker *et al.*, 2011). On the basis of phenotypic, genotypic and phylogenetic differences, the strains were clearly distinguished from all previously described species of the *Lachnospiraceae and Ruminococcaceae*. Previous studies have reported that *A. hadrus* was one of the 10 most abundant species in fecal samples (Louis *et al.*, 2010, Walker *et al.*, 2011). The ability of *A. hadrus* to produce butyrate from lactic acid provides a major energy source for colonocytes. Approximately 20% of the butyric acid pool in the large intestine is derived from lactic acid metabolism, indicating the essential role *A. hadrus* and other butyrate producers play in the system.

Similarly, the role of *Wegmannia spp.* and closely related *Ruminococcus spp.* and *Chassardia spp*. play a vital role within the gastrointestinal microbial community and to host health. The polysaccharide components of plant cell walls are

insoluble and resist degradation by human enzymes when consumed as part of the diet. However, dietary components that are indigestible to humans are fermented by the colonic microbiota. Degradation of plant polysaccharides in the colon requires the action of a diverse array of gut bacteria enriched with genes encoding fibrolytic enzymes that cleave specific glycosidic linkages (Martens *et al.*, 2011).

The genome of *W. bicirculans* was the first completed genome sequence for a human colonic representative of the *Ruminococcaceae* and one of the first completed genome sequences for a strictly anaerobic Firmicutes bacterium from the human GI tract. Genomic data is available for only a small number of species of *Ruminococcaceae* isolated from the rumen and human colon. Similarly, as few species of *Ruminococcaceae* are represented by cultured species, the characterization of strain *Wegmannia bicirculans* presented here contributes to an improved understanding of the role these organisms play within the human GI tract and provides further taxonomic resolution of the misclassified ruminococci and related phylogenetic clades. As several former *Ruminococcus* spp. have been reclassified to other genera, the reclassification of the remaining *Ruminococcus* spp. is needed to avoid confusion and to resolve phylogeny. As no novel organisms representative of the group II misclassified ruminococci are being classified in this study, reclassification of the these *Lachnospiracea* species is beyond the scope of the work presented here.

Molecular profiles based on 16S rRNA gene sequences provide tremendous insights into the diversity present in the human GI tract. However, it its still

important to culture and isolate the organisms that constitute these microbial communities. Cultivation and molecular methods are complementary, as metagenomic data can allow for the inference of the necessary growth conditions of uncultivated bacteria and characterization of their functions (Lagkouvardos *et al.*, 2016). Molecular methodology continues to drive and aid novel approaches for the cultivation of these uncultivated organisms. As the organisms representing the most abundant phylotypes of the human gut continue to be cultivated and characterized, the diversity of taxa within the system and the interactions between specific strains of the microbiome is better understood. Further experiments with the microbial consortia can be constructed to determine interrelationships between different taxa and the host.

Chapter 2

Clostridium amazonense a novel obligate anaerobic bacterium isolated from a remote Amazonian community in Peru and antimicrobial resistance

Abstract

Initial studies of the human gastrointestinal tract have revealed that the microbiome of traditional, indigenous communities are significantly different than those derived from Westernized and urbanized communities. Investigations of these communities have revealed microbes atypical of those found in industrialized nations and may provide insights into how the gastrointestinal microbiome responds to urbanism and Westernization. Examination of the literature reveals that the majority of investigations are based on molecular methods providing important information on the diversity present via community profiles employing 16S rRNA gene sequence analysis. However, cultivation and characterization of representative taxa from non-Western groups is necessary to gain a more complete understanding of the ecology and functional diversity of the gastrointestinal tract.

A number of novel taxa including a novel genus and two novel species from the family *Peptoniphilaceae* and a novel species from *Clostridium* cluster I (*Clostridium sensu stricto*) were discovered from geographically remote Amazonian communities located in Peru. The focus of this chapter is the isolation and characterization of an organism that belongs to the genus *Clostridium*, based on the comparative

phylogenetic analysis of the 16S rRNA gene sequence*,* for which the name *Clostridium amazonense* sp. nov. is proposed. Unlike the organisms reported in the previous chapter, the novel *Clostridium* species presented here is not an abundant phylotype according to molecular surveys based on 16S rRNA gene analysis. The organism was characterized using biochemical and chemotaxonomic methods. In addition, antimicrobial susceptibility testing was performed against a panel of 15 antibiotics to determine resistance to antibiotic classes commonly used to treat infections. To date strain NE08V^{T} has only been recovered from Community 1 and has yet to be reported in over 8,000 studies human microbiome studies performed to date, demonstrating the importance of the continued inclusion of groups removed from industrialized biases.

Introduction

The human body is host to tens of trillions of microbial cells that form highly complex microbial communities. These microbial symbionts are referred to as the human microbiome and are the collection of all the microorganisms – including eukaryotes, archaea, bacteria and viruses - present in the nasal, oral, skin, gastrointestinal tract and urogenital tract. The characterization of these microorganisms is essential to understanding not only the diversity and stability of the microbial communities within our bodies but also the extent of the role they play in human health and disease states.

There has been a growing interest in the commensal human microbiome and in 2007, the National Institutes of Health (NIH) launched their roadmap initiative – the Human Microbiome Project (HMP). The major aims of the HMP included determining if individuals across the world share a core human microbiome and whether changes in the human microbiome can be correlated with changes in human health (Turnbaugh *et al.*, 2007, Peterson *et al.*, 2009, Consortium, 2012). Molecular inventories based mainly on the use of 16S rRNA gene sequencing have provided tremendous insights into the microbial diversity and the richness of taxa present in the human gastrointestinal tract. Indeed, such studies show that approximately 40-80% of the phylotypes represent uncultured bacteria (Liu *et al.*, 2008, Zhou *et al.*, 2014). However, the majority of these studies have focused on Western populations using 16S rRNA gene sequence libraries derived from culture-independent methods (Flint *et al.*, 2012, Gordon, 2012, Kim *et al.*, 2014, Zhou *et al.*, 2014). Since the

1940s, the widespread consumption of processed and chemically treated food and water sources of Western populations has led to a tremendous increase in the number of diseases affecting the human gastrointestinal tract. Certain organisms that have previously played a protective role in the gastrointestinal tract have been lost due to the modern diet and changes in lifestyle (Blaser, 2014). In order to truly evaluate the core microbiome and elucidate the complex interactions between microorganisms, individuals from a variety of geographic regions with diverse diets must be included in these investigations (Hattori & Taylor, 2009, Gordon, 2012, Lewis *et al.*, 2012).

Studies of the gastrointestinal microbiome are revealing that the microbiome of indigenous communities may be significantly different than those derived from Westernized and urbanized communities (Contreras *et al.*, 2010, Blaser *et al.*, 2013, Schnorr *et al.*, 2014, Obregon-Tito *et al.*, 2015). Similarly, gastrointestinal microbes, atypical of industrialized nations, may have the ability to perform functional roles within the host microbiome of remote communities that are normally undertaken by different taxa or are absent in Westernized communities.

Culture-dependent approaches and the application of methods to characterize the microbial physiological and metabolic properties can be used to better understand the ecology of these microbial communities and the interaction between individual species that play a role in both health and disease processes (Lagier *et al.*, 2012, Wylie *et al.*, 2012, Clavel *et al.*, 2014).

An increased international effort is now underway to include indigenous communities outside of the commonly sampled urbanized and Western populations.

This effort led our team to work with communities in Peru (Obregon-Tito *et al.*, 2015). These communities include two traditional hunter-gatherer groups deep within the Amazon River basin as well as a coastal community that maintains traditional subsistence fishing and farming practices. Studies of communities that continue subsistence and nomadic traditional lifestyles are critical for providing a foundation for understanding how the human microbiome responds to urbanism and Westernization and may also yield insights into the ancestral state of the human microbiome.

We hypothesized that these indigenous communities were host to a distinct gastrointestinal microbiome, due to their geographically remote location and traditional lifestyle, when compared to Western populations and harbor novel intestinal microorganisms. Furthermore, the antibiotic resistome of the isolates from these communities will differ from those isolated from Western samples in the absence of the selective pressures associated with antibiotic therapy. Sampling within traditional communities removes many possible effects of modern-day lifestyles placed upon the gastrointestinal tract of many urbanized populations. Evidence for such changes is the gastrointestinal microbial ecosystems of traditional communities previously sampled in Peru have been shown to be more similar to those of ancient samples than those sampled from cosmopolitan groups (Obregon-Tito *et al.*, 2015). In comparison, the organisms composing the microbial communities of the GI tract of Western, cosmopolitan populations have undergone many changes from the ancestral state microbiome. These changes may be due to the

ingestion of processed food and water, improved sanitation technologies, globalization of urban microbiomes and changes in the GI microbial community due to the common prescription of antibiotics (Blaser, 2014). Ultimately, by sampling from traditional communities we anticipate that changes from the ancestral state microbiome in response to urbanism and Westernization will be observed.

In our continuing studies of traditional communities in Peru, a novel, Gram-stain-positive, strictly anaerobic bacterium was isolated from a fecal sample. Based on the phylogenetic, phenotypic and chemotaxonomic evidence, strain NE08V^T represents a novel species within the genus *Clostridium* for which the name *Clostridium amazonense* sp. nov. is proposed.

Materials and methods

Community and participant selection

A small, remote traditional Shipibo community (Community 1), from the Loreto region, Peru was selected for a collaborative research effort after careful planning and consideration of both national and traditional regional authorities (Fig 8). As with other indigenous populations, the Shipibo are generally distrustful of outsiders. Following in-depth consultations with local leaders and governmental health institutes, the protocol for sampling within this traditional Amazonian community was introduced to the community and volunteer participants consented to involvement upon deposition of fecal samples (Obregon-Tito *et al.*, 2015).

Figure 8. Location of Shipibo community in Peru

The remoteness of the communities is one of the reasons their unique culture has survived with little outside influence. The community members maintain their cultural identity and for the most part live like the many generations before them, relying mainly on subsistence strategies including hunting and fishing. There is a strong seasonal difference in nutrition and health as subsistence strategies greatly vary with the seasonal changes. Each participant self-identified as a permanent member of the community and furthermore, identified all four grandparents as permanent members of the community. All participants completed a detailed diet, health and medical history questionnaires during public, community engagement meetings. Although rare, participants medically treated outside of the community or those that recently travelled outside of the community were excluded for this study. None of the participants presented here had been treated for any illness requiring therapeutic antibiotics (Obregon-Tito *et al.*, 2015).

Enrichment and cultivation

Freshly voided fecal material was collected and processed anaerobically on site prior to being transported on ice to the University of Oklahoma, Norman, Oklahoma for further processing. Solid material from each sample was stored in three ways (i) swabs were used to collect fecal material then added to pre-reduced, sterile transport media, (ii) solid material was added directly to microcentrifuge tubes and (iii) to tubes containing pre-reduced glycerol. All samples were then stored in an

anaerobic jar and remained on ice or were refrigerated until arrival at the Lawson Microbial Systematics Laboratory.

Isolation and identification

Anaerobic enrichments for the cultivation and isolation of dominant members of the gastrointestinal tract were prepared based on their individual growth requirements and inoculated with 1 ml of fecal slurry. Media enrichments included liquid and solid MRS (Difco), Brucella (Difco), Brain Heart Infusion (BD), Schaedler medium (Remel), Columbia medium (BD), Fastidious Anaerobic Agar (Oxoid) and Cooked Meat Medium (BD). Media were additionally supplemented with 5% defibrinated blood, rumen fluid, vitamin K, hemin or horse serum. Strains were isolated from enrichments using the roll tube method, dilution to extinction, and repeated sub-cultivation on solid agar. DNA of strains was extracted and identified by 16S rRNA gene sequencing as described below in DNA extraction and phylogenetic analysis section. While a number of strains were identified from the enrichments as those previously described members belonging to dominant phyla of the gastrointestinal microbiome, the focus of this chapter is a previously uncharacterized bacterial strain, NE08V^T.

Strain NE08V^T

Strain NE08V^T was isolated from an enrichment (modified MRS) that contained (per Liter): casein peptone, tryptic digest 10 g, meat extract 10 g, yeast extract 5 g, glucose 20 g, Tween 80 1 g, potassium phosphate 2 g, sodium acetate 5 g, ammonium citrate 2 g, magnesium sulfate 0.2 g, manganese sulfate 0.05 g with 20 % ethanol and 0.1 % hemin solution, pH 6.5. Enrichments were incubated under an anaerobic gas mix of 85 % nitrogen, 10 % carbon dioxide, and 5 % hydrogen at 37°C. After incubation for 7 days, serial dilution and repeated sub-culturing onto fresh modified MRS agar was performed until pure colonies were obtained. The strain was then transferred to and maintained on Brucella (BD/BBL) agar supplemented with 5 % defibrinated blood.

Phenotypic and biochemical characterization

For morphological observations, strain NE08V^T was grown on Brucella (BD/BBL) agar supplemented with 5 % defibrinated blood at 37°C for 48 h. Cell characterized with an Olympus CX41 light microscope. The pH range, temperature range and salinity range were determined using peptone-yeast extract (PY) broth (DSMZ 104 medium). The pH range for growth was assessed over the range of pH 5.0-9.0 (increments of 0.5 pH units) using sodium citrate buffer for pH 5.0-6.0, potassium phosphate buffer for pH 6.0-8.0, and Tris buffer for pH 8.0-9.0. The salinity range for growth was tested with NaCl concentrations of 0 - 9% (w/v). The temperature range was determined by incubating strain NE08V^T at 4, 10, 15, 20, 25,

30, 37, 43, 47, and 60 °C. Optimum growth conditions were determined by monitoring the optical density using a spectrophotometer at 600 nm (Spectronic 20D, Milton Roy). Additional biochemical characterization was determined using the Rapid ID32A system (bioMeriux) and all reactions were performed in duplicate. Metabolic characterization was performed by use of the BIOLOG system (BIOLOG, Hayward, California) following the supplied manufacturer's instructions. Briefly, strain NE08V^T was cultivated on BUA^{TM} agar supplemented with 5 % defibrinated blood for 48 hrs and used to inoculate an AN MicroPlate^{TM}. The AN MicroPlate^{TM} was incubated at 37°C under anaerobic conditions for 72 hr. All test were performed in duplicate.

Metabolic end products were determined from cultures grown under anaerobic conditions in PYG broth. Sample analyses were carried out in duplicate on an Aminex HPX-87H organic acid analysis column (Bio-Rad), using ion-exclusion HPLC with 0.015 HCl running buffer at a flow rate of 0.9 mL/min. Retention times and peak areas of fermentation products were compared to standards of acetate, butyrate, lactate, succinate, formate, and propionate.

DNA extraction and phylogenetic analysis

For phylogenetic analysis, DNA of strain NE08V^T was extracted using the UltraClean® Microbial DNA Isolation Kit (MoBio Laboratories, Inc.) following manufacturer's instructions. 16S rRNA gene fragments were generated by PCR

using universal primers 8f (positions 8 to 28, *Escherichia coli* numbering) and pH* (1542 to 1522) (Hutson *et al.*, 1993). The amplicon was purified using Exo-SapIt (USB Corporation) to remove unincorported primers and dNTPs and the sequence determined. The closest known relatives of strain NE08V^T based on the 16S rRNA gene sequence were determined by performing database searches using the program EzBioCloud's Identify (https://www.ezbiocloud.net/) (Kim *et al.*, 2012, Yoon *et al.*, 2016). These sequences and those of other related strains were aligned with the sequence derived from NE08V^T using the program ClustalX. Phylogenetic reconstructions were performed in MEGA (version 4) (Tamura *et al.*, 2011) using the neighbour-joining method (Saitou & Nei, 1987), applying evolutionary genetic distances that had been calculated by the Kimura two-parameter model (Kimura, 1980).

Chemotaxonomic methods

Analysis of fatty acids and mol% G+C content was performed under conditions chosen to directly compare strain NE08V^T with its closest relative, *C. vulturis*. Biomass for fatty acid analysis was collected from Brucella (BD/BBL) agar supplemented with 5% defibrinated blood at 37°C for 48 h. Fatty acid methyl esters were extracted from isolated colonies using the Sherlock Microbial Identification System (MIDI) version 6.1 as described previously (Sasser, 1990, Kämpfer & Kroppenstedt, 1996). Analysis was performed using an Agilent Technologies 6890N gas chromatograph equipped with a phenyl methyl silicone fused silica capillary

column (HP-2 25m x 0.2 mm x 0.33 µm film thickness) and a flame ionization detector with hydrogen used as the carrier gas. The temperature program was initiated at 170°C and increased at 5°C min^{-1} reaching a final temperature of 270°C. Fatty acids were identified and expressed in the form of percentages using the QBA1 peak naming database. The mol% G+C was determined according to the method of Mesbah et al. (Mesbah *et al.*, 2011).

Antimicrobial susceptibility testing

All antimicrobial susceptibility testing was performed according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) (https://mic.eucast.org/Eucast2/) and the Clinical and Laboratory Standards Institute and International Organization for Standardization (ISO) guidelines (https://clsi.org/standards/products/iso-documents/). Stock solutions were prepared for 15 antibiotics and filter sterilized. These compounds included β-lactams (ampicillin, carbenicillin, cefaclor, and ceftriaxone), aminoglycosides (gentamycin and kanamycin), tetracyclines (doxycycline, oxycycline and tetracycline), quinolones (ciprofloxacin and nalidixic acid), macrolides (erythromycin), lincosamides (clindamycin), glycopeptides (vancomycin), and chloramphenicol. Working solutions were created in Mueller Hinton (MH) broth from each stock solution according to EUCAST break points and the upper limit growth (Table 15).

The range of concentrations were selected based on similar reference strains in EUCAST database, when available (https://mic.eucast.org/Eucast2/). The range allowed for endpoint minimum inhibitory concentration (MIC) determination. Each

antibiotic solution was dispensed into 96-well plates at 50 µl per well. An inoculum was prepared in anaerobic MH broth from colonies grown on BHI + 5% blood from 24 hour growth at 37°C at a turbidity equivalent to a 0.5 McFarland standard. To each antibiotic solution, 50 µl of inoculum was added for a final volume of 100 µl /well. Each antibiotic solution was tested in triplicate.

The 96-well plates were incubated for 24 hours at 37°C under anaerobic conditions. Each antibiotic was tested in triplicate for all concentrations. When not in agreement, the isolate was retested. All plates were graded by visual confirmation of growth, without strain information attached to each plate to reduce biases. Results were determined and compared to EUCAST breakpoints.

A volume of 100 µl inoculum was added to the wells without antibiotic solution and an aliquot was plated onto BHI + 5% blood agar after incubation of 96-well plates to ensure purity. Sterility of uninoculated antibiotic solution wells were verified after incubation. Each well was compared to the growth control to determine MIC.

Whole genome sequencing and assembly

Whole genome sequencing (WGS) was employed to determine functional properties present in the genome and other phylogenetic relationships to close relatives. The genome of strain NE08V^T was sequenced using an Illumina MiSeq sequencing platform (Illumina, CA). Library preparation, sequencing, and runs were performed in accordance with the manufacturer's instructions. Genomic library

construction was performed by the Laboratories of Molecular Anthropology and Microbiome Research under the supervision of Dr. Krithi Sankaranarayanan.

Table 15. Antibiotics and concentrations selected for MIC determination

Antibiotic (mg/L)	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Gentamycin	Kanamycin	Ciprofloxacin	Nalidixic Acid	Doxycycline	Oxycycline	Tetracycline	Vancomycin	Erythromycin	Clindamycin	Chloramphenicol
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	64	32	16	16	16	32	8	16	16	16	16	16	64	8	16
	128	64	32	32	32	64	16	32	32	32	32	32	128	16	32
	256	128	64	64	64	128	32	64	64	64	64	64	256	32	64
	512	256	128	128	128	256	64	128	128	128	128	128	512	64	128
	1024	512	256	256	256	512	128	256	256	256	256	256	1024	128	256
	2048	1024	512	512	512	1024	256	512	512	512	512	512	2048	256	512
	4096	2048	1024	1024	1024	2048	512	1024	1024	1024	1024	1024	3072	512	768

Comparison strains

As strain NE08V^T is the only representative of the species, the closest relatives based on 16S rRNA gene sequence similarity were selected for comparison of antibiotic resistance gene (ARG) determinants. *C. sulfidigenes* strain 113A, *C. senegalense* strain JC122^T, *C. argentinense* strain CDC 2741 and C*lostridium* cluster I type species *C. butyricum* strains KNUL09 and CDC 51208 were chosen as both clinical and non-clinical representatives. All genomes were downloaded from National Center for Biotechnology Information (NCBI). Average Nucleotide Identities (ANI) of NE08V^T and related taxa were calculated using the OrthoANIu algorithm (https://www.ezbiocloud.net/tools/ani) (Lee *et al.*, 2016, Yoon *et al.*, 2017).

Resistome analysis

Genomes of the *Clostridium* species were investigated for the presence of ARGs using the Comprehensive Antibiotic Resistance Database (CARD) (version 1.1.8) (McArthur *et al.*, 2013, Jia *et al.*, 2017). The assembled genomes were annotated using Prokaa version 1.12 (beta) (Seemann, 2014), output files were analyzed using the Resistance Gene Identifier (version 3.2.0) to predict resistance genes based on homology and single-nucleotide polymorphism (SNP) models. Additionally, results were verified by standard BLASTP search against the CARD reference sequences and manually compared. Only those predicted AR genes with

≥60% similarity to the reference sequences were included as a cutoff as AR genes <60% could be the result of novel genes or significant divergence from known mechanisms available in the database that CARD is able to recognize. This cut-off restricts the detection of more distant homologs and ensures the detected variant is likely a functional AR gene. The resistance gene analysis of the clostridial genomes was then compared between strains. Predicted AR proteins occurring in multiple species were aligned with one another and with the CARD reference sequences. Additional sequence similarity of the aligned resistance proteins was calculated using Clustal X and GeneDoc. The results of the predicted AR proteins were compared with the results of strain NE08V^T susceptibility testing.

Results

Phenotypic and biochemical characterization

NE08V^T was determined to be a Gram-stain positive, strictly anaerobic, spore-forming and non-hemolytic rod-shaped bacterium. Cells occurred as single cells or in pairs. Subterminal to terminal spores were observed. Colonies grown on Brucella (BD/BBL) agar supplemented with 5 % defibrinated blood at 37°C for 48 h were observed to be yellow brown in color, irregular with undulate margins. Oxidase and catalase activities were not observed. The pH and temperature ranges for growth were pH 5.0 - pH 9.0 and 30 - 47 °C with optimal growth at pH 7.0 and temperature 37 °C. The strain grew well in NaCl concentrations between 0-3 % (w/v). The main fermentation end-products detected from PYG media were acetate and butyrate. Positive reactions of the AN MicroPlate™ indicate the organism's ability to utilize L-fucose, palatinose, L-rhamnose, L-glutamine, and L-serine as sole carbon sources.

Using the API Rapid ID 32A system positive reaction were obtained with arginine dehydrogenase, alkaline phosphatase, arginine arylamidase, glycine arylamidase, histidine arylamidase, leucine arylamidase, leucyl glycine arylamidase, mannose, and pyroglutamic acid arylamidase. No activity was observed for urease, α-galactosidase, β-galactosidase, β-galactosidase-6 phosphate, α-gucosidase, α-arabinosidase- β-glucuronidase, N-acetyl- β-glucosidase, raffinose, proline arylamidase, phenyl arylmidase, tyrosine arylamidase, alanine arylamidase, glutamic acid decarboxylase, α-fucosidase, glutamyl glutamic acid and serine arylamidase.

Morphological, physiological and biochemical properties are provided in the species description and Table 16.

Phylogenetic analysis

Phylogenetic analysis demonstrated that the organism was a member of the *Firmicutes* sharing a relationship with members of the genus *Clostridium sensu stricto*, also recognized as *Clostridium* rRNA group I (Collins *et al.*, 1994, Wiegel *et al.*, 2006) (Fig 9). The pairwise comparisons of trimmed sequences showed that strain NE08V^T was most closely related to *C. vulturis* (97.4% 16S rRNA sequence similarity) with the two organisms forming a cluster in the phylogenetic tree. A number of organisms demonstrated lower sequence similarity values revealing a more loose relationship; these included *C. senegalense*, *C. subterminale*, *C. argentinense*, *C. thiosulfatireducens, C. sulfidigenes, C. schirmacherense*, *C. punense* and *C. huakuii* (Table 17). All the major groupings in the neighbor-joining tree were confirmed using the maximum parsimony program.

Table 16. Comparison of differential characteristics of strain NE08V^T and closely related *Clostridium* species. Strains: 1, strain NE08V^T 2, *C. vulturis* 3, *C. subterminale* 4, *C. thiosulfatireducesns* 5, *C. sulfidigenes*

Characteristic	1	2	3	4	5
Temperature range °C	30-47	30-45	25-45	18-45	18-48
Optimal temperature °C	37	37	37	37	34
NaCl range (%, w/v)	0-3%	0-3%	0-80	0-20	0-60
Optimum NaCl	0	ND	0-10	0	0
pH range	5.5-9	5-10	5.4-9.3	6-9.8	5.5-9.0
Optimal pH	7.0	7.5	7.5	6.0	6.6
β-hemolysis	-	-	-	+	+
Enzyme activity:					
Alanine arylamidase	-	-	+	+	+
Arginine arylamidase	+	-	+	+	+
α-Fucosidase	-	-	+	+	-
Glutamic acid arylamidase	-	-	+	-	-
Glutamic acid decarboxylase	-	+	-	-	-
Glycine arylamidase	+	-	+	+	+
Leucyl glycine arylamidase	+	-	+	+	+
Mannose fermentation	+	-	+	-	+
Pyroglutamic acid arylamidase	+	-	+	-	-
Tyrosine arylamidase	-	-	-	+	-
Major fatty acids	$C_{13:0}$ anteiso, $C_{16:0}$.	$C_{16:0}$, $C_{18:1\ \omega9c}$, $C_{18:1\ \omega9c\ DMA}$	$C_{14:0}$, $C_{16:0}$, $C_{18:1\ \omega9c}$, $C_{18:1\ \omega9c\ DMA}$	$C_{14:0}$, $C_{16:0}$, $C_{18:1\ \omega9c}$, $C_{18:1\ \omega9c\ DMA}$	$C_{14:0}$, $C_{16:0}$, $C_{18:1\ \omega9c}$, $C_{18:1\ \omega9c\ DMA}$
DNA G+C content (mol%)	30.5	34.0	28- 30.8	31.4	32.3
Source	Human feces	Vulture intestine	Human infection	Cheese factory wastewater	Pond sediment

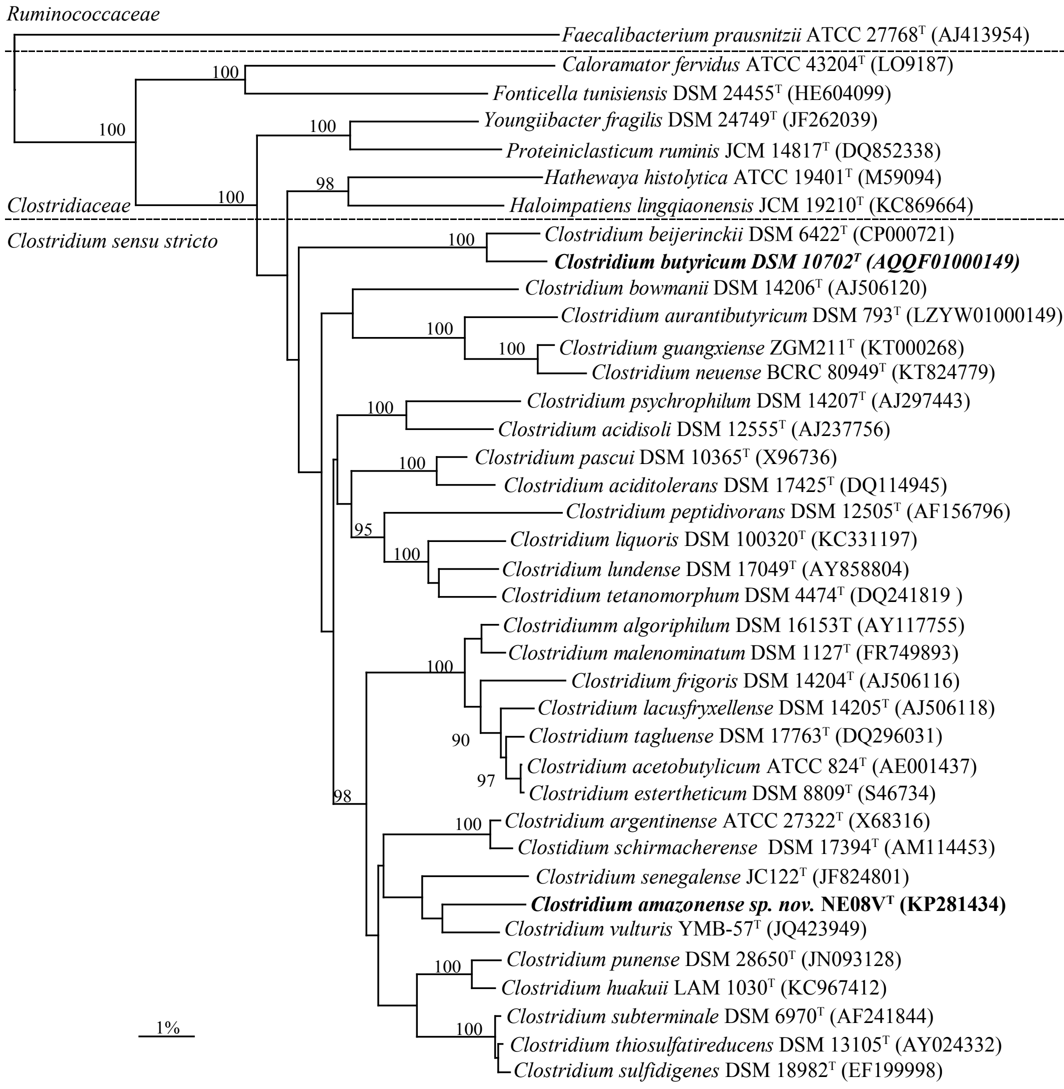

Figure 9. Phylogenetic tree showing the relationship of NE08V^{T}to the most closely related species within *Clostridium sensu stricto* and other representative members of the *Clostridiaceae*. The tree was constructed using the neighbor-joining method based on the pairwise comparison of approximately 1,370 nucleotides. *Faecalibacterium prausnitzii* was used as the outgroup. Bootstrap values (>90%), expressed as a percentage of 1000 replications. The scale bar indicates 1% sequence divergence.

Table 17. Comparison of strain NE08V^T to closely related species.

Strain Similarity	16S rRNA gene sequence similarity	Length of sequence (bp)
C. amazonense	-	1423
C. vulturis	97.4%	1409
C. senegalense	96.9%	1432
C. subterminale	95.5%	1431
C. argentinense	95.3%	1432
C. thiosulfatireducens	95.2%	1431
C. schirmacherense	95.1%	1431
C. sulfidigenes	95.1%	1430
C. punense	95.0%	1429
C. huakuii	94.9%	1385

Although no precise correlation exists between % 16S rRNA sequence divergence and species delineation, for many years 3% divergence values were generally considered adequate for the separation of species (Stackebrandt & Goebel, 1994). However, studies by Stackebrandt and Ebers demonstrated that this value could be lowered to 1.3% (Stackebrandt & Ebers, 2006). Their exhaustive investigation determined that where 16S sequence similarities values of 98.7% and below were obtained between strains, DNA-DNA hybridizations could be omitted without a risk for incorrect species differentiation (Meier-Kolthoff *et al.*, 2013). This cutoff of 98.7% was verified by Kim *et al.* (2014) for differentiating two species. Similarly, the cutoff value of <94-95% 16S sequence similarity is now routinely used in the delineation of genera when considered in tandem with appropriate phenotypic and chemotaxonomic information.

The draft genome assembly of strain NE08V^{T} consisted of 447 contigs and 4,409,344 bp. Genome analysis using the CARD database revealed genes coding for resistance to the antibiotics vancomycin, elfamycin, rifampicin, kirromycin and isoniazid. A summary of the sequencing statistics for the related *Clostridium* strains can be found in Table 18. The genomes ranged in G+C ratio from 26.9% (*C. senegalense*) to 30.0% (*C. sulfigenes*). Average aligned length of the *Clostridium* genomes to strain NE08V^{T} is shown in Table 19. ANI analysis of draft genomes of most closely related *Clostridium* species based on 16S rRNA sequence similarity (*C. vulturis, C. senegalense,* and *C. subterminale*) was not calculated due to lack of available genomes. The ANI analysis of strain NE08V^{T} with the genomes of closely relates species revealed a similarity index with *C. argentinense* CDC 2741 of 75.2% of 880,713 aligned base pairs.

Table 18. *Clostridium* species genome summary

Strain	Contigs	Total length (bp)	GC ratio
Strain NE08V^{T}	447	4,409,344	27.0%
C. argentinense Strain CDC 2741	2	4,803,058	28.8%
C. sulfidigenes Strain 113A	96	3,717,420	30.0%
C. senegalense Strain JC112	83	3,925,888	26.9%
C. butyricum Strain KNULO9	2	4,627,894	28.7%
C. butyricum Strain CDC 51208	3	4,639,914	28.7%

Table 19. Comparison of strain NE08V^T to related species based on ANI analysis

Strain similarity	OrthoANIu value	Average aligned length (bp)
C. argentinense Strain CDC 2741	75.2%	880,713
C. sulfidigenes Strain 113A	72.8 %	703,087
C. senegalense Strain JC 112	77.1%	1,216,165
C. butyricum Strain KNUL09	71.1%	440,104
C. butyricum Strain CDC 51208	71.0%	447,850

Chemotaxonomic analysis

The cellular fatty acid profile of strain NE08V^T determined that the predominant fatty (10 % of total fatty acids or higher) were $C_{13:0}$ anteiso (19.7%), and $C_{16:0}$ (26.4%) with lower amounts of $C_{16:1}$ ω9c (5.6%), $C_{14:0}$ (6.0%), and $C_{18:1\ \omega9c}$ (8.2%) (Table 6). The profile differs significantly from *C. vulturis* in that this organism does not produce $C_{13:0}$ anteiso but produces major amounts of $C_{18:1\ \omega9c}$ in comparison to strain NE08V^T. *C. vulturis* also produces $C_{18:1\ \omega9c\ DMA}$ which strain NE08V^T does not (Paek *et al.*, 2014). The lack of $C_{18:1}$ ω9c DMA in strain NE08V^T appears to be an important differential characteristic with other members of this loose group of organisms that includes *C. vulturis*, *C. subterminale*, *C. sulfidigenes* and *C. thiosulfatireducens* all of which synthesize this particular fatty acid (Table 20). The DNA G + C content of the strain was determined to be 30.5 mol%.

Table 20. Fatty acid composition of NE08V[T] and its closest relative *C. vulturis*

Fatty Acids	1	2
$C_{12:0}$	1.9	-
$C_{13:0}$ anteiso	**19.7**	-
$C_{14:0}$	6.0	2.4
$C_{15:0}$ anteiso	1.3	-
$C_{15:1}$ ω8c	4.6	-
$C_{16:0}$	**26.4**	**31.9**
$C_{16:0}$ iso	1.9	-
$C_{16:1}$ ω9c	5.6	-
$C_{16:1}$ ω5c	1.2	-
$C_{17:0}$	-	1.4
$C_{17:1}$ ω8c	1.7	-
$C_{17:1}$ ω9c	1.7	-
$C_{18:0}$	-	8.7
$C_{18:1}$ ω9c	**8.2**	**27.0**
$C_{18:1}$ t 11	-	2.5
$C_{18:2}$ ω9,12c	-	3.1
$C_{18:0\ DMA}$	-	1.1
$C_{18:1}$ ω9c DMA	-	**10.3**
$C_{19:0}$ cyclo 9, 10/:1	-	1.6
$C_{19:0}$ cyclo 9, 10 DMA	-	1.2
Summed feature 3	3.9	-
Unknown 14.969	2.4	-
Unknown 15.301	2.5	-

Predominant products are shown in bold, products that represented <1.0 % were omitted; cyclo, cyclopropane; DMA, dimethylacetal

Determination of breakpoints

Few cultured *Clostridium* spp. are included in the EUCAST database and there are no genus specific breakpoints. When available, Gram-positive anaerobe breakpoints were used (Table 21). All other breakpoints are non-species specific and incorporate the EUCAST pharmacokinetic (drug concentration over time) and pharmacodynamics (drug effect over time) (PK/PD) indices as previously reported (Mouton *et al.*, 2012). Breakpoints and results of susceptibility testing are reported in Table 21.

The MIC for strain NE08V^T was determined to be resistant for all antibiotics. The organism was not expected to have such high levels of resistance across all classes of antibiotics, especially as the community from which the organism was isolated has not been exposed to the selective pressures normally associated with antibiotic treatment. While ARG are naturally occurring and expected to be present to some degree within gastrointestinal communities, resistance to concentrations at this level are especially concerning. These organisms may act as a reservoir of resistance determinants that can be mobilized throughout the gastrointestinal microbial community. Further analysis of ARG of closely related species and related taxa of clinical significance should be performed. Unfortunately, no additional strains from this traditional community are available for analysis. However, the ability of ARG to be established and maintained within such remote communities remains an area of continued investigations (Chapter 3). Resistance genes shared

between taxa may serve as an indicator of how they are spread and how the genes have emerged from environmental sources (Jia *et al.*, 2017).

Table 21. EUCAST clinical breakpoints and MIC values of strain NE08V^T.

	Clindamycin	**Erythromycin**	**Chloramphenicol**	**Doxycycline**	**Oxycycline**	**Tetracycline**	**Gentamicin**	**Kanamycin**	**Vancomycin**	**Ampicillin**	**Carbenicillin**	**Cefaclor**	**Ceftriaxone**	**Nalidixic acid**	**Ciprofloxacin**
EUCAST breakpoints mg/L															
Breakpoint	>4*	ND	>8*	ND	ND	ND	ND	ND	>2*	>8	ND	ND	>2**	ND	>0.5**
Determined MIC values mg/L															
C. amazonense	32	2048	>768	64	256	>512	32	>2048	256	512	256	1024	>1024	512	32

Breakpoints were in agreement unless denoted *, G+ anaerobe specific breakpoint; **, PK/PD specific breakpoint; ND, not determined as there is not enough information to determine breakpoint.

Resistome analysis

The CARD was chosen because it is one of the most continuously updated, comprehensive resources. The CARD is a curated collection of antimicrobial resistance gene and mutation sequences that uses the Resistance Gene Identifier (RGI) software (https://card.mcmaster.ca/analyze/rgi) for the detection of AR in protein sequences. The CARD then catalogs resistance predictions by drug class and resistance mechanisms. Sequences are analyzed under three paradigms – Perfect, Strict and Loose (a.k.a. Discovery). Perfect classification detects perfect matches to the curated reference sequences and mutations in CARD, Strict classification detects previously unknown variants of known ARGs with curated similarity cut-offs to ensure the variant is most likely a functional ARG and Loose classification predicts potentially new ARG and catalogs homologous sequences that may not have a role in AR. All of the reported ARG for *Clostridium* strains were classified as Strict protein variants. The resistance gene % sequence similarity of detected resistance genes and CARD reference genes is summarized in Tables 22 and 23. A total of 29 resistance genes were detected between the six clostridial genomes. The summary of antibiotic resistance ontology (ARO) accession number, antibiotic class resistance, resistance mechanisms and the organism of which the reference sequence was curated are displayed in Table 24. Of these 29 genes 22 were detected within the *C. argentinense* genome. Multiple tetracycline resistance determinants, including ribosomal protection proteins (tetM, tetB(P) and tetO) and an efflux protein tetA(P), and six vanF gene cluster variants and vancomycin target alterations were identified within

the *C. argentinense* genome. Strain NE08V^{T} shared two ARGs in common with *C. argentinense* (SarpoB and ScEFTu). Both genes shared 91% sequence similarity with *C. argentinense* ARGs. These shared a lower percent similarity to database references than when compared to other resistant gene variants from within *Clostridium sensu stricto*, but are of interest in the future as they were detected within organisms of clinical significance. Strain NE08V^{T} contained only 6 genes SarpoB, ScEFTu, EcEFTu, MTkatG, vanRG and vanRI that confer resistance rifampicin, elfamycins, isoniazid and vancomycin respectively. Interestingly, two genes, ECEFtu and MTkatG, were found exclusively in strain NE08V^{T}. The high level of resistance is not determined by the detected genes, suggesting additional resistance genes not detected here, or potential novel resistance mechanisms. A further investigation of this organism and the resistome is necessary in order to identify these determinants. Excluding some ARGs of *C. argentinense*, lower % similarity was detected from the CARD reference genes than those determined between the clostridial variants. This could be due to the variants originating from distantly related genomes, and highlights the need to include more resistant determinants from other taxa. The two genes shared between all clostridial species but strain NE08V^{T} (SarpoB and SagyrB), shared a higher % sequence similarity when compared to each other than with the CARD reference genes (Table 25 -28).

Table 22. Resistance gene % similarity to reference strains of strain NE08V^T and related species

Resistance gene	cat	SagyrB	SagyrB	SarpoB	ScEFTu	ScEFTu	ScEFTu	ECEFtu	tetA(P)	tetB(P)	tetM	tetO
C. amazonense Strain NE08V^T				64%*	68%**			75%*				
C. argentinense Strain CDC 2741		63%*		65%*	66%**	67%**	75%*			75%		
C. sulfidigenes Strain 113A		62%*		65%*	66%**	67%**			98%	81%**	92%**	92%*
C. senegalense Strain JC 112		64%*	64%*		65%**							
C. butyricum Strain KNULO9	67%	64%*		65%*	69%**	69%**						
C. butyricum Strain CDC 51208		62%*		65%*	69%**	69%**						

Table 23. Resistance gene % strain NE08V^T and related species

Resistance gene	MTkatG	vanB	vanF	vanHF	vanHM	vanRF	vanRG	vanRI	vanSF	vanV	vanXF	vanYF
C. amazonense Strain NE08V^T	60%*						75%**	72%**				
C. argentinense Strain CDC 2741		95%*	96%**	94%*	95%**	90%			94%		92%	86%
C. senegalense Strain JC 112										64%**		

Resistance gene	bcrA	bcrB	ANT(6)-la	ANT(9)-la	CIPa	IsaE	clbB	vatB
C. argentinense Strain CDC 2741	65%	68%**	79%**	79%	63%**	70%	62%*	81%

Values represent % similarity to resistance gene in database; * Only in CARD; **Only in blastP database

Table 24. Antibiotic resistance genes detected in *Clostridium* strains

Gene	Resistance	ARO	Category	Database Strain
cat	chloramphenicol	3002670	chloramphenicol inactivation enzyme	*Enterococcus faecalis*
SagyrB	aminocoumarin	3003301	aminocoumarin resistance gene	*Staphylococcus aureus*
SarpoB	rifampicin	3003285	rifamycin resistance protein	*Staphylococcus aureus*
SCEFTu	elfamycin	3003359	elfamycin resistance protein	*Streptomyces cinnamoneus*
ECEFtu	kirromycin (elfamycin)	3003368	elfamycin resistance protein	*Escherichia coli*
tetA(P)	tetracycline	3000180	efflux pump conferring antibiotic resistance	*Clostridium perfringens*
tetB(P)	tetracycline	3000195	Target protection protein	*Clostridium perfringens*
tetM	tetracycline	3000186	Target protection protein	*Clostridium difficile 630*
tetO	tetracycline	3000190	Target protection protein	*Campylobacter jejuni*
MtkatG	isoniazid	3003392	isoniazid resistance protein	*Mycobacterium tuberculosis*
bcrA	bacitracin	3002987	efflux pump complex or subunit	*Bacillus licheniformis*
bcrB	bacitracin	3002988	efflux pump complex or subunit	*Bacillus licheniformis*
IsaE	pleuromutilin, lincosamide, and streptogramin	3003206	efflux pump complex or subunit	*Staphylococcus aureus*
ANT(6)-la	aminoglycoside	3002626	aminoglycoside nucleotidyltransferase (ANT), antibiotic inactivation enzyme	*Exiguobacterium sp. S3-2*
ANT(9)-la	aminoglycoside	3002630	aminoglycoside nucleotidyltransferase (ANT), antibiotic inactivation enzyme	*Staphylococcus aureus*

Table 24. contd. Antibiotic resistance genes detected in *Clostridium* strains

Gene	Resistance	ARO	Category	***Database Strain***
CIPa	Clindamycin, lincomycin chloramphenicol	3003907	23S ribosomal RNA methyltransferase	*Paenibacillus sp. Y412MC10*
clbB	Macrolide, phenicol, lincosamide, streptogramin linezolid	3002815	antibiotic target modifying enzyme	*Brevibacillus brevis NBRC 100599*
vatB	streptogramin	3002841	antibiotic inactivation enzyme	*Staphylococcus aureus*
vanB	glycopeptides	3000013	antibiotic target alteration	*Enterococcus faecium*
vanF	glycopeptides	3002908	antibiotic target alteration	*Paenibacillus popilliae ATCC 14706*
vanHF	glycopeptides	3002945	antibiotic target alteration	*Paenibacillus popilliae ATCC 14706*
vanHM	glycopeptides	3002947	antibiotic target alteration	*Enterococcus faecium*
vanRF	glycopeptides	3002925	antibiotic target alteration	*Paenibacillus popilliae ATCC 14706*
vanRG	glycopeptides	3002926	antibiotic target alteration	*Enterococcus faecalis*
vanRI	glycopeptides	3003728	antibiotic target alteration	*Desulfitobacterium hafniense*
vanSF	glycopeptides	3002936	antibiotic target alteration	*Paenibacillus popilliae ATCC 14706*
vanV	glycopeptides	3002916	antibiotic target alteration	*Enterococcus faecalis V583*
vanXF	glycopeptides	3002952	antibiotic target alteration	*Paenibacillus popilliae ATCC 14706*
vanYF	glycopeptides	3002958	antibiotic target alteration	*Paenibacillus popilliae ATCC 14706*

Table 25. Clostridium strains % similarity of shared SarpoB resistance genes

SarpoB	NE08V[T]	*C. argentinense*[1]	*C. butyricum*[2]	*C. butyricum*[3]
NE08V[T]	-	-	-	-
***C. argentinense*[1]**	91%	-	-	-
***C. butyricum*[2]**	82%	81%	-	-
***C. butyricum*[3]**	82%	82%	99%	-
***CARD*[4]**	64%	64%	64%	64%

1, Strain CDC 2741;2, Strain KNULO9; 3, strain CDC 51208; 4, CARD comparison strain from *S. aureus*

Table 26. Clostridium strains similarity of shared SarpoB resistance genes

SarpoB	NE08V[T]	*C. argentinense*[1]	*C. butyricum*[2]	*C. butyricum*[3]	*CARD*[4]
NE08V[T]	1187	91%	82%	82%	64%
	0	96%	91%	91%	78%
	0	0	0	0	2%
***C. argentinense*[1]**	1085	1187	81%	82%	64%
	1146	0	91%	91%	78%
	0	0	0	0	2%
***C. butyricum*[2]**	984	973	1186	99%	64%
	1091	1091	0	100%	79%
	1	1	0	0	2%
***C. butyricum*[3]**	985	974	1177	1186	64%
	1091	1091	1186	0	79%
	1	1	0	0	2%
***CARD*[4]**	768	774	771	769	1172
	942	942	951	951	0
	33	33	32	32	0

1, Strain CDC 2741;2, Strain KNULO9; 3, strain CDC 51208; 4, CARD comparison strain from *S. aureus*.

Table 27. Clostridium strains % similarity of shared SagyrB resistance genes

SagyrB	*C. butyricum*[1]	*C. butyricum*[2]	*C. senegalense*[3]	*C. argentinense*[4]	*C. sulfidigenes*[5]
C. butyricum[1]	-	-	-	-	-
C. butyricum[2]	99%	-	-	-	-
C. senegalense[3]	77%	76%	-	-	-
C. argentinense[4]	76%	76%	85%	-	-
C. sulfidigenes[5]	76%	76%	82%	81%	-
CARD[6]	62%	62%	64%	63%	62%

1, Strain KNULO9; 2, strain CDC 51208; 3, strain JC 112; 4, strain CDC 2741; 5, strain 113A; 6, CARD comparison gene from *S. aureus*

Table 28. Clostridium strains % similarity of shared SagyrB resistance genes

SagyrB	*C. butyricum*[1]	*C. butyricum*[2]	*C. senegalense*[3]	*C. argentinense*[4]	*C. sulfidigenes*[5]	CARD[6]
	636	99%	77%	76%	76%	62%
C. butyricum[1]	0	100%	88%	86%	88%	78%
	0	0	0	0	0	0
	632	636	76%	76%	76%	62%
C. butyricum[2]	636	0	88%	86%	88%	78%
	0	0	0	0	0	0
	491	490	635	85%	82%	64%
C. senegalense[3]	564	564	0	94%	91%	78%
	3	3	0	0	0	0
	448	487	541	635	81%	63%
C. argentinense[4]	554	554	597	0	91%	77%
	3	3	0	0	0	0
	488	489	525	518	635	62%
C. sulfidigenes[5]	564	564	580	583	0	75%
	3	3	0	0	0	0
	398	399	409	407	398	637
CARD[6]	499	499	499	494	485	0
	5	5	6	6	6	0

1, Strain KNULO9; 2, strain CDC 51208; 3, strain JC 112 4, strain CDC 2741; 5, strain 113A 6, CARD comparison gene from *S. aureus*

Description of *Clostridium amazonense* sp. nov.

Clostridium amazonense (am.a.zon,en'se.) N.L. neut. adj. *amazonense* pertaining to the community from where the organism was first isolated. Cells are Gram-positive obligately anaerobic rods occurring as single cells or in pairs. Spores are formed and located either subterminally or terminally. When grown anaerobically on Brucella blood agar for 48 h, colonies are irregular, 1 mm in diameter with undulate margins and have a yellow brown color. Catalase and oxidase negative. Optimal growth temperature is 37°C, and growth occurs between 30-47°C. Growth is supported at pH 5.5-9.0, and is optimal at pH 7.0. Growth occurs in the presence of 0-3% (w/v) NaCl. Using the ID 32A system, positive reactions are obtained for arginine dehydrogenase, alkaline phosphatase, arginine arylamidase, glycine arylamidase, histidine arylamidase, leucine arylamidase, leucyl glycine arylamidase, mannose, and pyroglutamic acid arylamidase. Negative reactions are obtained for α-galactosidase, β-galactosidase, β-galactosidase-6 phosphate, α-gucosidase, α-arabinosidase- β-glucuronidase, N-acetyl- β-glucosidase, raffinose, proline arylamidase, phenyl arylmidase, tyrosine arylamidase, alanine arylamidase, glutamic acid decarboxylase, α-fucosidase, glutamyl glutamic acid and serine arylamidase. The main fermentation end products from PYG medium were acetate and butyrate. Antimicrobial susceptibility testing revealed the strain was likely resistant to β-lactams (ampicillin, carbenicillin, cefaclor, and ceftriaxone), aminoglycosides (gentamycin and kanamycin), tetracyclines (doxycycline, oxycycline and tetracycline), quinolones (ciprofloxacin and nalidixic acid),

macrolides (erythromycin), lincosamides (clindamycin), glycopeptides (vancomycin), and chloramphenicol. Six resistance gene variants were determined corresponding to vancomycin, elfamycin, rifampicin, kirromycin and isoniazid. The major fatty acids (>10%) are $C_{13:0}$ anteiso and $C_{16:0}$. The DNA G + C content of the type strain is 30.5 mol%. The type strain is NE08V^{T} (=DSM 23598^{T} = CCUG 59712^{T}) is isolated from a fecal sample of an individual in the province of Loreto, Peru. The 16S rRNA gene sequence for strain NE08V^{T} was deposited with the GenBank Sequence Database under the following accession number KP281434.

Discussion

In addition to its unique 16S rRNA gene sequence, strain NE08V^{T} was found to possess biochemical and chemotaxonomic traits consistent with organisms belonging to members of the genus *Clostridium* but could clearly be distinguished from its nearest phylogenetic relatives using characteristics shown in Tables 2 and 3. Although there is much discussion on the validity of describing novel genera and species on a single strain, this continues to be a common practice and allows further strains or related species to be identified in subsequent studies by other researchers.

As the number of publically available genomes continues to increase, there is a need to critically evaluate the relatedness between strains in order to resolve phylogeny and correctly assign these strains to the correct taxa. The DNA-DNA method (DDH) has been regarded as the gold standard (despite evidence that it is error-prone and labor-intensive) to demo

nstrate the relatedness between strains as an indirect measure of sequence similarity (Rosselló-Móra & Amann, 2015). The species boundary DDH value of 70% equates to ANI values of 95-96%. Similarly, the ANI value of 96% corresponds with the 16S rRNA sequence similarity of 98.65% (Kim *et al.*, 2014).

Taxonomic revisions to the classification of *Clostridium sensu stricto* continue to reassign misclassified species to other taxa based on 16S rRNA sequences (Yutin & Galperin, 2013, Lawson & Rainey, 2016). While the state of classification of the genus *Clostridium* and related taxa is far from perfect, recent concerted efforts have led to the taxonomy of *Clostridium sensu stricto* being improved upon (Gupta & Gao, 2009, Vos *et al.*, 2011). In addition to the 16S rRNA analysis, additional molecular markers have been identified that appear to be conserved in *Clostridium* cluster I and suggest that the genus is phylogenetically distinct from other phylogenetic groups within the clostridia (Gupta & Gao, 2009).

At the time of publication >100 species are assigned to *Clostridium sensu stricto* (http://www.bacterio.net/clostridium.html*).* WGS prediction of an organism's phenotype may aid in differentiation between organisms, but at present should be combined in a polyphasic approach. Large, independent studies have yet to accurately demonstrate that the estimated >37,000 available bacterial genomes provide precise phenotypic predictions in studies of specific organisms (Fournier *et al.*, 2015).

Contributions to EUCAST (antibiotic break point information, addition to few studies of non-clinical strains/strains removed from therapeutic antibiotic

selection) and CARD (novel, non-clinical strains and non-Western strains for comparison to clinical strains and strains with industrial biases, potential detection of novel genes) from this study are especially important at present as the widespread antibiotic resistance in bacterial populations is particularly troubling. As more strains and resistant determinants are added to CARD this issue will likely be resolved. With continued curation of resistance genotypes, the accurate prediction of resistance phenotypes has the potential to become a valuable tool for the surveillance of emerging and established pathogens.

Furthermore, the increasing number of cultured strains recovered from individuals from indigenous communities will augment the insights gained from the molecular tools commonly used in these studies. Studies of hunter-gatherers maintaining traditional dietary practices are a powerful resource for our understanding of the ancestral state of the human microbiome and provide a foundation for understanding how the human microbiome responds to urbanism and Westernization. Though the exact role of $NE08V^T$ in the gastrointestinal tract remains unclear, this study confirmed that the complex ecosystem within the gastrointestinal microbiome of remote, indigenous Peruvian communities harbors novel bacterial species previously not identified within Western, urbanized populations. Loss of key microbial species within urbanized populations compared to traditional indigenous communities is an area of increased scientific interest and studies like the one reported here are essential for developing our understanding of the roles the microbes play within our gastrointestinal tract (Blaser, 2014).

Ultimately, the identification and characterization of novel gastrointestinal species will allow us to form a clearer picture of the interrelationships between organisms that constitute this ecosystem and better understand their role in human health and disease. The characterization and molecular insights of novel human associated microbes from these indigenous communities may aid in shaping our understanding of the human microbiome and the antibiotic resistome, and direct the development of tools to treat infections and emerging diseases.

Chapter 3

Insights into the antibiotic resistome of a traditional indigenous Peruvian community from cultivated representatives of *Bacillaceae, Paenibacillaceae, Lactobacillaceae, Streptococcaceae* and *Enterobacteriaceae*

Abstract

Presented here are the results of an initial screening and subsequent susceptibility testing and determination of resistance genes by whole genome sequencing of a number of organisms isolated from a remote, traditional indigenous Peruvian community. Following the initial resistance screen, bacterial strains were identified and resistance genes were determined using a curated resistance database. Resistant organisms were found to contain resistance genes for all classes of antibiotics tested. The community has had limited or no exposure to modern therapeutic antibiotics and therefore represents an ideal cohort in which to work with to determine if wide spread resistance is a modern phenomenon linked to human overuse and abuse of antibiotics or a natural phenomenon of the ancient and pervasive spread of genetically diverse resistance determinants.

Introduction

In the past several decades, the rise and spread of antibiotic resistance has become alarming and has led to a worldwide health concern (World Health Organization, 2014). Antibiotic resistance has become a threat to the effective treatment of a growing number of bacteria, especially among those bacteria responsible for common health-care associated infections. In a recent report, the United States *Center for Disease Control and Prevention* gave a conservative estimate of 2 million illnesses and over 20,000 deaths directly caused by antibiotic resistant bacteria each year (Centers for Diseae Control & Prevention, 2013). In addition to health concerns, the costs of bacterial infections total over $23 billion in societal and healthcare costs annually in the United States. Resistant pathogens lead to higher health care costs because they often require more expensive drugs and extended hospital stays.

The recent increase and spread of antibiotic resistant bacteria is most often attributed to the overuse via over prescription and misuse of antibiotics in clinical, agricultural, and veterinary practices as therapeutic agents, growth promoters and as prophylactics (Levy & Marshall, 2004, Cabello, 2006, Koike *et al.*, 2007, Manuzon *et al.*, 2007, Forslund *et al.*, 2013, Spellberg *et al.*, 2013). The hypothesis that these selective pressures provide drivers for antibiotic resistance is consistent with studies that have demonstrated that when a microbial community lack such selective pressures, antibiotic resistance genes (ARG) were absent (Österblad *et al.*, 2001, Thaller *et al.*, 2010). However, other, contrasting reports have revealed widespread,

multidrug resistance from isolated, natural habitats such as isolated cave systems, Alaskan soils and glacier environments (Allen *et al.*, 2009, Bhullar *et al.*, 2012, Segawa *et al.*, 2013). In this context, we sought to examine the prevalence of antibiotic resistance genes and the minimum inhibitory concentration of a panel of antibiotics of strains cultured from the gastrointestinal microbiome of an isolated, traditional indigenous Peruvian community with limited or no exposure to modern therapeutic antibiotics. Working with such groups will contribute to our understanding of how wide spread ARGs are and how insights into how the genetic diversity of the resistant determinants has developed. The high level of diversity and global distribution of resistance genes is of particular concern as antimicrobial resistance (AMR) mechanisms have been reported for all known antibacterial drugs that are currently available for clinical use in both human and veterinary medicine (World Health Organization, 2014). Understanding the origin of ARGs and the mechanisms of emergence, prevalence, and their evolution over time is fundamental in our efforts to prevent, manage, and ultimately the potential to predict genes of concern.

Antibiotic resistance can be due to innate susceptibility, intrinsic resistance, or acquired resistance (D'Costa *et al.*, 2011). Even in the absence of previous or ongoing antibiotic treatment, antibiotic resistance genes from other sources can establish within the body's commensal microbiome (Arason *et al.*, 1996, Lancaster *et al.*, 2003, Bartoloni *et al.*, 2004, Jernberg *et al.*, 2010). The term "Antibiotic Resistome" was first coined by D'Costa *et al.* (2006) in a survey of resistance genes

within a microbial soil community. The resistome serves as a reservoir of resistance determinants that can be mobilized into the microbial community (D'Costa *et al.*, 2006). Resistome machinery likely originated from within soil communities as antibiotic producing strains possess self-immunity and eventually transferred this property to animal and human pathogens (D'Costa *et al.*, 2011, Nesme & Simonet, 2015). Soil communities are one of the richest habitats in terms of microbial diversity and contain most of the known bacterial producers of antibiotic molecules. Furthermore these producers contain ARGs identical to those found in clinical samples (Nesme & Simonet, 2015).

Another likely reservoir is the human gastrointestinal tract, that essentially is a vessel facilitating the exchange of genetic material between the enormous number of microorganisms that play a vital role in health and disease processes. The interactions between host and the diverse gastrointestinal microbial communities aid in nutrition, development of the immune system and protection against pathogens (Xu & Gordon, 2003, Cho & Blaser, 2012). In addition to playing these roles in host health and development, the gastrointestinal microbiome may serve as a reservoir of ARGs accessible to commensal and pathogenic organism (Wang *et al.*, 2006, Sommer *et al.*, 2009, Penders *et al.*, 2013, Rolain, 2013). In an investigation into the role of the human gastrointestinal microbiome as a reservoir of resistance, Sommer *et al.* (2009) determined that most of the resistance genes identified from using culture-independent sampling, had not been previously identified. Furthermore, these genes were evolutionarily distant from known resistance genes, but fully functional

when cloned into host cells. In contrast, nearly half of the resistance genes identified in cultured isolates were identical to resistance genes of major pathogens (Sommer *et al.*, 2009).

Ongoing investigations of the ancestral state of the gastrointestinal microbiome may aid in expanding our knowledge of how modern, cosmopolitan biases have impacted the relationships between host animals and microbes and serve in determining the role of the gut microbiome as a reservoir for resistance and possible origin of resistance genes. Efforts by collaborators have determined that the ancestral microbiome is more similar to traditional communities than cosmopolitan populations (Tito *et al.*, 2012, Obregon-Tito *et al.*, 2015). Novel taxa, atypical of cosmopolitan groups were reported from previous investigations of traditional communities (O'Neal *et al.*, 2015, Patel *et al.*, 2015). Given that these communities harbor unique microbial communities and novel taxa, I hypothesized that such groups may be host to a distinct antibiotic resistome, and may contain novel resistance mechanisms in addition to those previously identified in the microbial communities encountered in industrialized nations.

Here, data is presented from San Mateo, a traditional, hunter-gatherer group residing along the Galves River deep within the Amazonian jungle of Peru (Fig 1) (this is a different community than investigated in previous chapter). This Matses community is geographically remote and has maintained a traditional lifestyle with little contact with outside groups. A recent investigation determined that the gastrointestinal microbial communities of the Matses exhibited a higher level of

diversity when compared to industrialized nations (Obregon-Tito *et al.*, 2015). Gut microbiome of indigenous populations may provide insights into the spread and evolution of ARGs, independent of selective pressures of cosmopolitan biases.

Materials and Methods

Sampling from San Mateo

The community of San Mateo is composed of 14 hunter-gatherer families residing within the Matses National Reserve, along the Peruvian-Bolivian border (Fig 10). A total of 25 participants between the ages of 2 and 52 volunteered to participate in the study. This study was conducted in collaboration with the Matses as previously described in Obregon-Tito *et al.* (2015). Briefly, members have been geographically, historically and socially isolated and are traditional hunter-gatherers. Before sampling, community meetings and parasite screening were offered to all community members. All participants filled out detailed diet, health and medical history questionnaires. Although rare, participants medically treated outside of the community or those that recently travelled outside of the community were excluded for this study. None of the participants presented here had been treated for any illness requiring therapeutic antibiotics. Each volunteer participant in the study was assigned a number to remove personal data from the samples (Obregon-Tito *et al.*, 2015).

Freshly voided fecal samples were collected from members of the community and immediately processed. Solid fecal samples collected in microcentrifuge tubes and swabs were processed into anaerobic transport media. Additionally, fecal slurry was processed onsite and added to pre-reduced sterile media and glycerol. All samples were transported in anaerobic jars on ice, down the river over a three-day period. Samples were stored on ice until arrival.

Figure 10. Location of San Mateo community, located on Peruvian-Bolivian border

Initial resistance screen

All antibiotics used throughout this study, corresponding molecular targets and mode of resistance are shown in Table 29. Antibiotic resistant isolates were

selected by screening, utilizing a disk diffusion method developed in collaboration with the Dr. Lars Hansen Laboratory, Aarhus University, Denmark. Samples from four healthy individuals were used in this investigation. In consultation with the European Committee on Antimicrobial Susceptibility Testing (EUCAST) (https://mic.eucast.org/Eucast2/) and the Clinical and Laboratory Standards Institute (CLSI) (https://clsi.org/standards/products/iso-documents/), the screening process was standardized to enhance current databases. Dilutions of 1 g solid fecal material (1/5, 1/10 and 1/100) were inoculated on Mueller Hinton (MH) agar (Oxoid) amended with 5% blood. Antibiotic discs (BD BBL) were applied to the agar and incubated for 24 hours at 37°C under aerobic conditions or anaerobic conditions using a gas mix of 85 % nitrogen, 10 % carbon dioxide, and 5 % hydrogen at 37°C. All plates were prepared in triplicate.

The antibiotic discs used for the screening process included β-lactams, aminoglycosides, tetracyclines, quinolones, macrolides, lincosamides, and chloramphenicol (Table 30). Resistant colonies were selected from within the zone of inhibition that surrounded each disc. Resistant organisms were plated onto MH agar supplemented with 5% blood at 37°C and repeated until a pure culture was obtained. Isolates were stored in glycerol stocks (-80°C) after pure cultures were obtained to maintain antibiotic resistance.

Table 29. Antibiotics in this study, targets and modes of resistance

Antibiotic Class	Members	Target	Mode of resistance
β-lactams	Ampicillin Amoxicillin Carbenicillin Cefaclor Ceftriaxone Penicillin	Peptidoglycan biosynthesis Penicillin binding proteins	Hydrolysis Efflux Altered target
Aminoglycosides	Gentamycin Kanamycin	Translation 30S ribosomal subunit	Phosphorylation Acetylation Nucleotidylation Efflux Altered target
Glycopeptides	Vancomycin	Peptidoglycan biosynthesis	Reprogramming of peptidoglycan biosynthesis
Tetracyclines	Doxycycline Oxycycline Minocycline Tetracycline	Translation 30S ribosomal subunit	Monooxygenation Efflux Altered target
Macrolides	Erythromycin	Translation 50S ribosomal subunit	Hydrolysis Glycosylation Phosphorylation Efflux Altered target
Lincosamides	Clindamycin	Translation Bacterial 50S rRNA	Nucleotidylation Efflux Altered target
Phenicols	Chloramphenicol	Translation	Acetylation Efflux Altered target
Quinolones	Ciprofloxacin Nalidixic Acid	DNA replication Topoisomerase II and IV	Acetylation Efflux Altered target

Table 30. Antibiotic concentrations for resistance screen by disc diffusion

Antibiotic	Concentration	Incubation Conditions
Ampicillin	2 µg	Anaerobic
Ampicillin	10 µg	Aerobic
Amoxicillin	30 µg	Anaerobic
Carbenicillin	100 µg	Aerobic
Penicillin	2 IU	Anaerobic
Cefaclor	30 µg	Anaerobic
Cefaclor	30 µg	Aerobic
Ceftriaxone	30 µg	Aerobic
Kanamycin	30 µg	Aerobic
Doxycycline	30 µg	Anaerobic
Minocycline	30 µg	Anaerobic
Oxytetracycline	30 µg	Anaerobic
Tetracycline	5 µg	Aerobic
Ciprofloxacin	5 µg	Anaerobic
Ciprofloxacin	5 µg	Aerobic
Nalidixic Acid	30 µg	Aerobic
Erythromycin	2 µg	Anaerobic
Erythromycin	10 µg	Aerobic
Clindamycin	5 µg	Anaerobic
Chloramphenicol	30 µg	Aerobic

Phylogenetic analysis of strains

Upon isolation of resistant strains, chromosomal DNA was extracted by UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Inc). Amplification of the 16S rRNA gene was achieved using universal primers, corresponding to positions 8-27 (forward primer) and 1492-1510 (reverse primer) of the *Escherichia coli* numbering system (Hutson *et al.*, 1993). The amplified products were treated with ExoSAP-IT to remove unincorporated primers and dNTPs (USB Corporation) and directly sequenced with primers directed towards the conserved positions of the rRNA gene.

The closest known related strains were retrieved from GenBank and EzTaxon databases and aligned with the newly determined sequences using the program SEQtools (https://www.ncbi.nlm.nih.gov/genbank/; http://www.ezbiocloud.net/eztaxon; Yoon *et al*., 2017, Kim *et al*., 2012, Rasmussen, 2002). The resulting multiple sequence alignment was corrected manually using the program GeneDoc (Nicholas *et al.*, 1997). A phylogenetic tree was constructed according to the neighbor-joining method (Saitou & Nei, 1987) with the programs SEQtools and TREEVIEW (Page, 1996). The stability of the groupings was estimated by bootstrap analysis (1000 replications).

Antimicrobial susceptibility testing

All antimicrobial susceptibility testing was performed according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the International Organization for Standardization (ISO) guidelines. Stock solutions were prepared for the 15 antibiotics (Table 31) and filter sterilized. Working solutions were created in Mueller Hinton (MH) broth from each stock solution according to EUCAST break points and the upper limit growth (Table 31).

The range of concentrations selected was compared to reference strains in EUCAST database, when available (https://mic.eucast.org/Eucast2/) (Fig A1). The range allowed for endpoint minimum inhibitory concentration (MIC) determination. Each antibiotic solution was dispensed into 96-well plates with 50 µl per well.

An inoculum was prepared in anaerobic or aerobic MH broth based on isolation conditions, from colonies grown on BHI + 5% blood from 24-hour growth at 37°C. The inoculum was prepared to a turbidity equivalent of 0.5 McFarland standard. To each antibiotic solution, 50 µl of inoculum was added for a final volume of 100 µl /well. Each antibiotic solution was tested in triplicate. *Escherichia coli* strain ATCC 25922 was also tested as a control and compared to previously reports and CLSI and EUCAST databases.

The 96-well plates were incubated for 24 hours at 37°C under anaerobic or aerobic conditions corresponding to the original isolation conditions. A volume of 100 µl inoculum was added to the wells without antibiotic solution and an aliquot was plated onto BHI + 5% blood agar after incubation to ensure purity. Sterility of uninoculated antibiotic solution wells was verified after incubation by plating onto BHI + 5% blood agar. Each well was compared to the growth control to determine MIC.

Table 31. Antibiotics and concentrations selected for MIC determination

Antibiotic (mg/L)	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Gentamycin	Kanamycin	Ciprofloxacin	Nalidixic Acid	Doxycycline	Oxycycline	Tetracycline	Vancomycin	Erythromycin	Clindamycin	Chloramphenicol
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.5	0.5	0.5	0.5	0.5	0.5	0.125	1	0.125	0.5	0.5	0.5	0.5	0.5	0.5
	1	1	1	1	1	1	0.25	2	0.25	1	1	1	1	1	1
	2	2	2	2	2	2	0.5	4	0.5	2	2	2	2	2	2
	4	4	4	4	4	4	1	8	1	4	4	4	4	4	4
	8	8	8	8	8	8	2	16	2	8	8	8	8	8	8
	16	16	16	16	16	16	4	32	4	16	16	16	16	16	16
	32	32	32	32	32	32	8	64	8	32	32	32	32	32	32
	64	64	64	64	64	64	16	128	16	64	64	64	64	64	64
	128	128	128	128	128	128	32	256	32	128	128	128	128	128	128
	256	256	256	256	256	256	64	512	64	256	256	256	256	256	256
	512	512	512	512	512	512	128	1024	128	512	512	512	512	512	512
	1024	1024	1024	1024	1024	1024	256	-	256	1024	1024	1024	1024	-	768*
	2048	2048	-	-	-	2048	512	-	512	-	-	-	2048	-	-
	4096	-	-	-	-	-	-	-	1024	-	-	-	3072*	-	-

** Saturation point*

Whole genome sequencing and assembly

Whole genome sequencing (WGS) was employed for functional genomic analysis and to resolve phylogenetic relationships. The genome of resistant isolates was sequenced using an Illumina MiSeq sequencing platform (Illumina, CA). Library preparation, sequencing, and runs were performed in accordance with the manufacturer's instructions. Genomic library construction was performed by by the Laboratories of Molecular Anthropology and Microbiome Research under the supervision of Dr. Krithi Sankaranarayanan.

Genome analyses

Draft genomes of closely related species, based on 16S rRNA sequence % similarity, were selected as comparison strains. All microbial genomes were downloaded from the National Center for Biotechnology Information (NCBI). Average Nucleotide Identities (ANI) of all resistant strains and related taxa were calculated using the OrthoANIu algorithm (https://www.ezbiocloud.net/tools/ani) (Lee *et al.*, 2016, Yoon *et al.*, 2017). The program RNAmmer (1.2 server) was used to predict 16S rRNA genes from genome sequences to confirm the initial identities of these organisms (Lagesen *et al.*, 2007). Draft genomes sequences were investigated for the presence of ARGs using the Comprehensive Antibiotic Resistance Database (CARD) (version 1.1.8) (McArthur *et al.*, 2013, Jia *et al.*, 2017). The assembled genomes were annotated using Prokaa version 1.12 (beta)

(Seemann, 2014). Output files was analyzed using the Resistance Gene Identifier (version 3.2.0) to predict resistance genes based on homology and single-nucleotide polymorphism (SNP) models. Additionally, results were verified by standard BLASTP search against the CARD reference sequences and manually compared. Only those predicted AR genes with ≥60% similarity to the reference sequences were included (Fig A2). This value was chosen as a cutoff as AR genes <60% could be the result of novel genes or significant divergence from known mechanisms available in the database that CARD is able to recognize. This cut-off restricts the detection of more distant homologs and ensures the detected variant is likely a functional ARG. The resistance gene analysis of each strain was compared between taxa. The results of the predicted AR proteins were compared within taxa.

Results

Initial resistance screen and phylogenetic analysis

Resistant colonies were observed to all antibiotics in the initial disc diffusion resistance screen with the exception of minocycline. Only those organisms from well inside the zone of inhibition were selected. Of the 200 organisms originally isolated, the focus of this chapter is restricted to 64 strains based on phylogenetic analyses so representatives of different taxa were represented. Resistant organisms were isolated from both anaerobic and aerobic cultivations from each of the four participants, with a total of 15 strains from Matses1, 17 strains from Matses20, 15 strains from Matses31 and 17 strains from Matses33 (Table 32). Initial partial 16S rRNA gene sequencing revealed these resistant strains as species belonging to *Bacillales* and *Lactobacillales* including *Bacillus spp*., *Paenibacillus spp*., *Lactobacillus spp*., *Pediococcus* spp. and *Lactococcus spp*., and species of the *Escherichia* and *Shigella* cluster.

Bacillus species were resistant to β-lactam, chloramphenicol and quinolone antibiotics. *Paenibacillus* species were resistant to β-lactam and chloramphenicol antibiotics. *Lactobacillus* species were resistant to β-lactam and tetracycline antibiotics. *Pediococcus* species were resistant to β-lactam antibiotics. *Lactococcus* species were resistant to quinolone, lincosamide, and βeta-lactam antibiotics. *Escherichia* and *Shigella* cluster species were resistant to quinolone, lincosamide, and βeta-lactam and macrolide antibiotics. Each resistant organism is reported by

antibiotic disc resistant and individual in Tables 33.1-33.5. No resistant isolates were recovered from amoxicillin, minocycline or gentamycin. This is not indicative of lack of resistance to these antibiotics as the use of this methodology was conducted as a preliminary resistance screen using a fecal slurry rather than determination of resistance of pure cultures. The use of pure culture disc diffusion was not used in this investigation but instead MIC was determined to a range of antibiotics and concentrations. Prescreening the isolates ensured all strains were resistant to at least one antibiotic class. Determination of MIC was performed by antibiotic susceptibility testing and should indicate resistance determinants in the CARD resistome analyses.

Table 32. Number of initial screen resistant strains by indigenous individual

Individual	Matses 1	Matses 20	Matses 31	Matses 33	Total isolates*
Strains	**15**	**17**	**15**	**17**	**64**
Bacillus altitudinis	-	3	-	1	4 (A)
Bacillus safensis	-	1	-	1	2 (A)
Bacillus thuringiensis	-	2	-	-	2 (A, B)
Bacillus xiamenensis	-	2	-	-	2 (A)
Escherichia/Shigella spp	12	1	9	1	23(11A, 12B)
Lactobacillus ruminis	-	5		-	5 (B)
Lactococcus lactis	3	-	5	8	16 (13A 3B)
Lactococcus taiwanensis	-	2	1	-	3 (A)
Paenibacillus lactis	-	-	-	6	6 (A)
Pediococcus pentosaceus	-	1	-	-	1 (A)

* A, aerobic conditions; B, anaerobic conditions

Table 33.1 Original disc resistance to antibiotics

Strain	Lincosamide Clindamycin	Macrolide Erythromycin	Chloramphenicol Chloramphenicol
Total	**7**	**1**	**5**
Bacillus altitudinis	-	-	3
Bacillus safensis	-	-	1
Escherichia/Shigella	5	1	2
Lactococcus lactis	2	-	-

Table 33.2. Original disc resistance to β-lactams antibiotics

Table 33.2. Original disc resistance to β-lactams antibiotics

Strain	Penicillin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone
Total	**4**	**10**	**4**	**4**	**5**
Bacillus altitudinis	-	-	-	-	1
Bacillus thuringiensis	1	-	-	-	-
Bacillus xiamenensis	-	-	-	-	2
Escherichia/Shigella	2	7	1	-	-
Lactobacillus ruminis	1	2	-	1	-
Lactococcus lactis	-	-	1	1	-
Paenibacillus lactis	-	-	2	2	2
Pediococcus pentosaceus	-	1	-	-	-

Table 33.3. Original disc resistance to quinolone antibiotics

Strain	Nalidixic acid	Ciprofloxacin
Total	**10**	**7**
Bacillus safensis	-	1
Escherichia/Shigella	1	-
Lactococcus lactis	8	4
Lactococcus taiwanensis	1	2

Table 33.4. Original disc resistance to tetracycline antibiotics

Strain	Doxycycline	Oxycycline	Tetracycline
Total	**3**	**1**	**1**
Escherichia/Shigella	3	1	-
Lactobacillus ruminis	-	-	1

Table 33.5. Original disc resistance to aminoglycoside

Strain	Kanamycin
Bacillus thuringiensis	1

Phylogenetic and genomic analysis of strains

Of the 64 strains from the initial disc resistance screen, 50 genomes were suitable for further genomic analyses. Two organisms were unable to revive from glycerol stocks and 12 genomes were excluded because of poor sequence assembly. The original identities were confirmed by the full 16S rRNA analysis predicted by RNAmmer from genome sequences (Table 34.1-34.5). All organisms were determined to be strains of previously described species. The relationships of all strains and closely related species are shown in the phylogenetic tree (Fig 11). ANI analysis was performed between resistant strains and the closest relatives based on 16S rRNA sequencing similarity results, between the resistant strains and the related comparison strains, and between members of the same taxa (Table 35.1-35.4). Due to their very close relationships, *E. coli and Shigella* species were not well resolved based on 16S rRNA, ANI and additional analysis (not shown). Continuing investigations into differentiation of this group of organisms remains taxing to molecular microbiologists and is beyond the scope of this investigation for the purposes of this investigation and will be grouped together and referred to throughout as *E. coli/Shigella* cluster species (Zuo *et al.*, 2013).

Table 34.1. *Bacillus* species identity and 16SrRNA gene sequence similarity %

Strain designation		Identity	Similarity (%)
34	IM20A11	*Bacillus altitudinis*	99.9
117	M33A35	*Bacillus altitudinis*	100
45	IM20A23	*Bacillus altitudinis*	100
46	IM20A24	*Bacillus altitudinis*	100
116	M33A32	*Bacillus safensis*	99.9
52	IM20A30	*Bacillus xiamenensis*	100
54	IM20A32	*Bacillus xiamenensis*	100

Table 34.2. Species identity and 16S rRNA gene sequence similarity

Strain designation		Identity	Similarity (%)
132	M33A34	*Paenibacillus lactis*	100
136	M33A47	*Paenibacillus lactis*	99.8
137	M33A48	*Paenibacillus lactis*	99.9
139	M33A50	*Paenibacillus lactis*	99.9

Table 34.3. Species identity and 16SrRNA gene sequence similarity %

Strain designation		Identity	Similarity (%)
8	IM01A8	*E. coli/Shigella spp*	99.9
17	IM01A17	*E. coli/Shigella spp*	99.9
20	IM01A21	*E. coli/Shigella spp*	99.9
21	IM01A22	*E. coli/Shigella spp*	99.8
41	IM20A19	*E. coli/Shigella spp*	99.9
58	IM01B1	*E. coli/Shigella spp*	99.9
59	IM01B2	*E. coli/Shigella spp*	99.9
63	IM01B8	*E. coli/Shigella spp*	99.9
70	IM01B19	*E. coli/Shigella spp*	99.9
74	IM01B26	*E. coli/Shigella spp*	99.9
77	IM01B30	*E. coli/Shigella spp*	99.9
133	M33A40	*E. coli/Shigella spp*	99.9
163	M31A11	*E. coli/Shigella spp*	99.9
169	M31A14	*E. coli/Shigella spp*	99.9
193	M31An18	*E. coli/Shigella spp*	99.9

Table 34.4. *Lactococcus* species identity and 16SrRNA gene sequence similarity %

Strain designation		Identity	Similarity (%)
4	IM01A5	*Lactococcus lactis*	99.8
16	IM01A16	*Lactococcus lactis*	99.9
78	IM01B3	*Lactococcus lactis*	99.8
118	M33A36(3)	*Lactococcus lactis*	99.9
120	M33A38	*Lactococcus lactis*	100
121	M33A39	*Lactococcus lactis*	99.9
125	M33A46	*Lactococcus lactis*	100
141	M33A57	*Lactococcus lactis*	99.8
143	M33A59	*Lactococcus lactis*	100
145	M33An25	*Lactococcus lactis*	99.9
160	M31A08	*Lactococcus lactis*	99.9
164	M31A12	*Lactococcus lactis*	99.9
173	M31A18	*Lactococcus lactis*	99.8
179	M31An02	*Lactococcus lactis*	99.8
33	IM20A10	*Lactococcus taiwanensis*	99.9
37	IM20A15	*Lactococcus taiwanensis*	99.9
167	M31A13-B	*Lactococcus taiwanensis*	99.9

Table 34.5. Species identity and 16S rRNA gene sequence similarity

Strain designation		Identity	Similarity (%)
82	IM20B2	*Lactobacillus ruminis*	99.6
85	IM20B6	*Lactobacillus ruminis*	99.9
101	IM20B25	*Lactobacillus ruminis*	99.7
105	IM20B29	*Lactobacillus ruminis*	99.6
87	IM20B9	*Pediococcus pentosaceus*	100

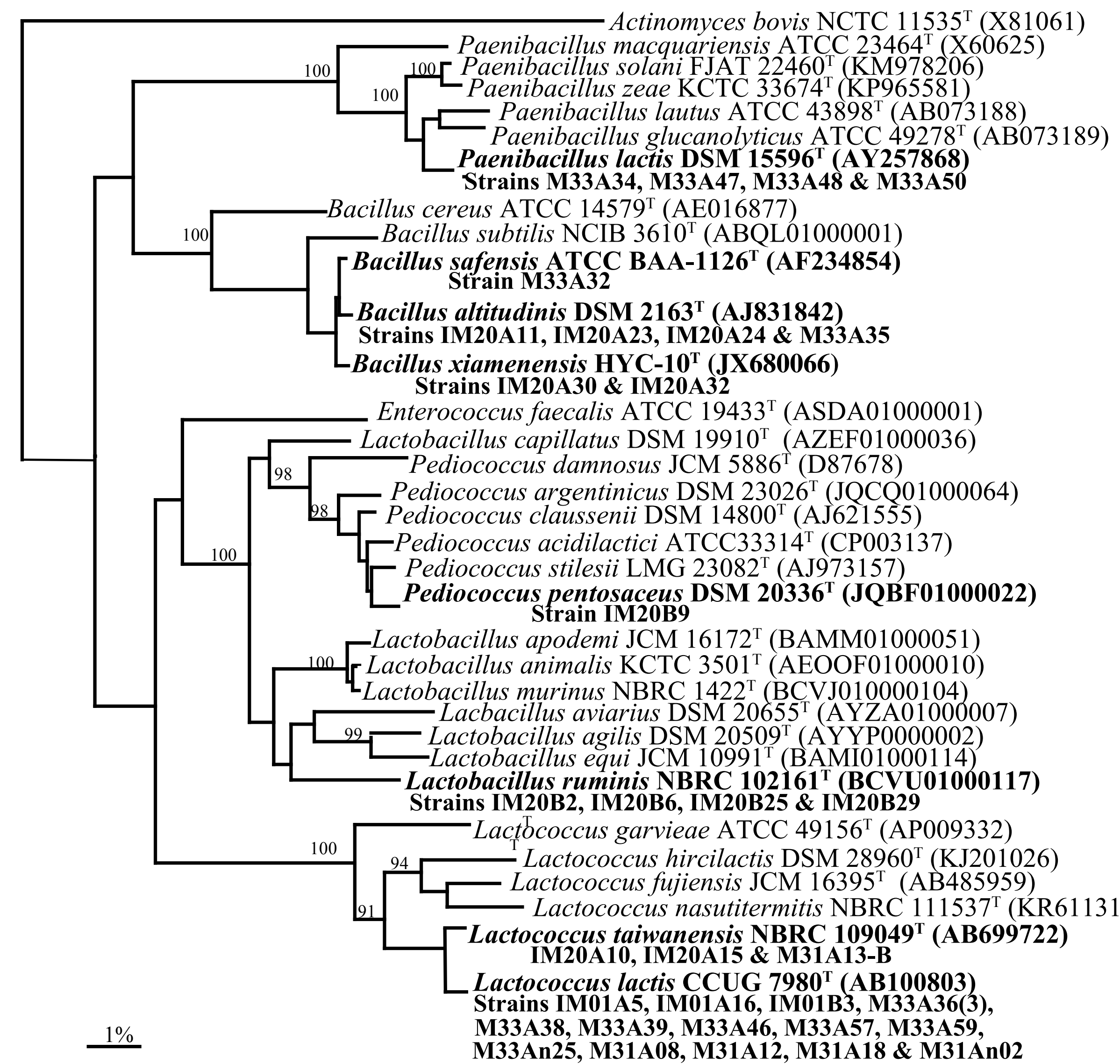

Figure 11. Phylogenetic tree showing the relationship of resistant organisms to closely related species. Bootstrap values (>90%), expressed as a percentage of 1000 replications. The scale bar indicates 1% sequence divergence.

Table 35.1. Resistant strains of *Lactobacillaceae* and type strains *OrthoANIu* analysis

Strain	ATCC 27782	82	85	101	105	ATCC 25745
Lactobacillus ruminis						
ATCC 27782	-					
82 IM20B2	96.7	-				
85 IM20B6	96.6	99.9	-			
101 IM20B25	96.6	99.9	99.9	-		
105 IM20B29	96.7	99.9	99.9	99.9	-	
Pediococcus pentosaceus						
ATCC 25745	67.9	66.8	67.3	67.3	67.0	-
87 IM20B9	67.4	68.2	68.5	67.4	68.4	99.2

Table 35.2. Resistant strains of *Paenibacillus lactis* and type strains *OrthoANIu* analysis

Strain	*P. lactis* 154	TW132	TW136	TW137
Paenibacillus lactis				
132 M33A34	97.2	-		
136 M33A47	97.2	100	-	
137 M33A48	97.3	100	100	-
139 M33A50	97.2	100	100	100

Table 35.3. Resistant strains of *Bacillus species* and type strains *OrthoANIu* analysis

Strain	YNP4-TSU	34	45	46	117	KCTC12796BP	116	VV3	052
Bacillus altitudinis									
YNP4-TSU	-								
34 IM20A11	98.5	-							
45 IM20A23	98.3	99.0	-						
46 IM20A24	98.3	98.6	98.6	-					
117 M33A35	98.2	98.5	98.5	98.6	-				
Bacillus safensis									
KCTC 12796BP	89.0	89.1	89.0	89.0	89.0	-			
116 M33A32	89.0	89.1	89.0	88.9	88.9	96.4	-		
Bacillus xiamenensis									
VV3	91.5	91.4	91.5	91.7	91.62	89.4	89.3	-	
52 IM20A30	91.5	91.4	91.3	91.5	91.58	89.4	89.3	99.5	-
54 IM20A32	91.6	91.5	91.4	91.6	91.54	89.4	89.4	99.5	100

Table 35.4. Resistant strains of *Lactococcus species* and type strains *OrthoANIu* analysis

	Strain	Il1403	4	16	78	118	120	121	125
					Lactococcus lactis				
	Il1403	-							
4	IM01A5	95.8	-						
16	IM01A16	98.0	95.8	-					
78	IM01B3	95.8	100	95.8	-				
118	M33A36(3)	98.5	95.8	98.1	95.7	-			
120	M33A38	98.6	95.8	98.1	95.8	100	-		
121	M33A39	97.8	95.7	99.7	95.7	98.1	98.2	-	
125	M33A46	98.6	95.9	98.1	95.7	100	100	98.1	-
141	M33A57	95.9	99.8	95.7	99.8	95.9	95.7	95.8	96.0
143	M33A59	98.6	95.8	98.1	95.9	100	100	98.1	100
145	M33An25	95.7	99.8	95.8	99.7	95.8	95.9	95.7	95.9
160	M31A08	97.9	95.8	99.9	95.7	98.1	98.2	99.8	98.1
164	M31A12	98.0	95.7	99.8	95.6	98.2	98.1	99.6	98.1
179	M31An02	95.8	99.9	95.7	99.9	95.9	95.8	95.8	95.8
					Lactococcus taiwanensis				
33	IM20A10	79.5	79.7	80.0	79.7	79.8	79.1	79.9	79.3
37	IM20A15	79.4	80.0	79.8	79.8	80.0	79.3	79.7	79.3
167	M31A13-B	79.3	79.8	79.7	79.9	79.7	79.3	79.7	79.3

Table 35.4.cont. Resistant strains of *Lactococcus species* and type strains *OrthoANIu* analysis

	Strain	141	143	145	160	164	179	33	37
					Lactococcus lactis				
141	M33A57	-							
143	M33A59	95.8	-						
145	M33An25	99.7	96.0	-					
160	M31A08	95.8	98.1	95.8	-				
164	M31A12	95.8	98.1	95.6	99.7	-			
166	M31An02	99.7	95.9	99.9	95.7	95.8			
179	M33A57	99.8	95.8	99.8	95.8	95.5	-		
					Lactococcus taiwanensis				
33	IM20A10	79.9	79.4	79.4	79.9	79.5	79.3	-	
37	IM20A15	79.9	79.2	79.4	79.7	79.7	79.3	100	-

Table 36. Antibiotics used in this study, spectrum, and impact on GI tract

Class	Spectrum	Impact on GI tract[a]
β-lactams	Narrow and broad	Decreased bacterial diversity, decreased abundance of anaerobes, greater prevalence of *Enterobacter* spp., decreased abundance of enterobacteria
Aminoglycosides	Broad	Decreased bacterial diversity, greater prevalence of *Enterobacter* spp.; abundance of *Ruminococcaceae* and *Bacteroidaceae* increases
Glycopeptides	Narrow	Decreased bacterial diversity
Tetracyclines	Broad	Reduction in abundance of *Bacteroidetes* enterococci, *E. coli*, lactobacilli, and bifidobacteria and increases in other enterobacter and yeasts; increase in Proteobacteria
Macrolides	Narrow	Decreased abundance of Streptococci, enterococci, and enerobacteria; increase in abundance of staphylococci; alteration in abundance of anaerobes
Lincosamides	Moderate	Decreased abundance of enterococci, streptococci and anaerobic bacteria then recovery of abundance of streptococci and anaerobic bacteria, reduced diversity of *Bacteroides* spp., decrease in SCFA producers
Phenicols	Broad	Lower bacterial diversity, increases in yeast
Quinolones	Narrow and broad	Lower bacterial diversity, decrease in short-chain fatty acid producers, decreased abundance of enterobacteria

[a] Data from Langdon et al 2016.

Antimicrobial susceptibility testing

Eight classes of antibiotics were used in the susceptibility test. The classes of antibiotics and their attributed effect on the gastrointestinal microbial community are shown in Table 36. Each organism was tested in triplicate to the full panel of 15 antibiotics. When the panel was not in agreement, the isolate was re-tested against the panel with fresh growth from BHI blood plates. All 96-well plates were graded by visual confirmation of growth, without strain information to reduce biases. MIC were determined and compared to EUCAST breakpoints for the organism or the general anaerobic breakpoint to determine resistant or susceptible classifications when available (Table 37.1-37.6). Many antibiotics do not have data for any of the screened organism. The efforts described here will benefit the EUCAST database by extending the MIC breakpoints of additional organisms and to additional antibiotics (Jorgensen & Turnidge, 2015). When no MIC breakpoints were recorded in databases, values ≥64 are likely indicative of resistance and are highlighted for further investigation.

All *L. lactis* isolates displayed resistance to nalidixic acid. Overall strains *L. lactis* strains M22A46, M33A57 and M33A59 displayed the least resistance to all antibiotics tested. *L. lactis* strains IM01A15, IM01A16, M33An25 and M31A12-B were resistant to all antibiotic tested. Similarly, *L. taiwanensis*, strain M31A13-B, displayed resistance to all antibiotics tested.

Bacillus species displayed the least resistant to any antibiotics tested of all taxa. All *B. altitudinis* and *B. xiamenensis* strains were resistant to chloramphenicol.

All *Bacillus* isolates were susceptible to tetracyclines and vancomycin. Overall*,* *Escherichia/Shigella* species were found to display the most resistance compared to isolates from other taxa. All strains displayed resistance to cefalclor, ceftriaxone and vancomycin*,*

E. coli/Shigella were tested under anaerobic and aerobic conditions to determine differences in MIC between conditions. Under anaerobic growth conditions, all strains were resistant to clindamycin, kanamycin, vancomycin, cefaclor and ceftriaxone. All strains but IM20A19 were resistant to erythromycin and carbenicillin under anaerobic conditions. All *E. coli/Shigella* species but strain M31An17 were resistant to chloramphenicol under aerobic conditions. All *Escherichia/Shigella* species but strain M31A14 were resistant to clindamycin under aerobic conditions. Alternatively, *Escherichia/Shigella* species were susceptible to tetracycline and oxytetracycline but strains IMO1A21, IMO1A22, M31An09, M31An11, M31An16 and M31A11 when grown under aerobic conditions. Differences in resistance between anaerobic and aerobic conditions were likely observed due to requirements for an oxygen-dependent carrier in order to enter the cell as is the case of aminoglycoside uptake.

Paenibacillus isolates were susceptible to doxycycline, tetracycline and ceftriaxone. All *Paenibacillus* isolates were susceptible to erythromycin but strain M33A48, susceptible to vancomycin but strain M33A34 and susceptible to ampicillin but strain M33A51. All *Lactobacillus* strains were susceptible to doxycycline, oxycycline and tetracycline. All *Lactobacillus* strains but IM20B28

were susceptible to clindamycin. All strains were resistant to vancomycin and all strains but IM20B28 were resistant to chloramphenicol, ampicillin, carbenicillin and nalidixic acid. *P. pentosaceus* was resistant to all antibiotics.

Table 37.1. MIC concentrations of *Bacillus* species.

	Strain designations	Clindamycin	Erythromycin	Chloramphenicol	Doxycycline	Oxycycline	Tetracycline	Gentamicin	Kanamycin	Vancomycin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Nalidixic acid	Ciprofloxacin
							Bacillus altitudinis									
	EUCAST G+	>4R	ND	>8R	ND	ND	ND	ND	ND	>2R	>8R	ND	ND	ND	ND	ND
34	IM20A11	64	1	16	0.125	4	2	0.5	4	1	2	4	4	32	4	0.25
45	IM20A23	1	0.5	32	<16	16	2	64	128	0.5	1024	128	64	32	1	32
46	IM20A24	16	0.5	64	0.125	0.5	2	0.5	4	2	512	128	4	32	32	0.5
117	M33A35	4	128	32	32	4	4	8	4	2	2	4	4	32	64	64
							Bacillus safensis									
116	M33A32	2	2	4	2	2	2	0.5	4	0.5	256	128	4	32	32	16
							Bacillus thuringiensis									
106	IM20B3	0.25	<64	<16	0.125	0.5	2	0.5	2	<16	256	2048	512	128	256	<8
							Bacillus xiamenensis									
52	IM20A30	0.125	0.5	32	2	2	2	1024	8	0.5	1	16	32	32	512	32
54	IM20A32	1	0.5	32	2	0.5	2	0.5	8	0.5	0.5	0.5	4	32	32	16

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint

Table 37.2. MIC concentrations of *Lactococcus* species.

Strain designations		Clindamycin	Erythromycin	Chloramphenicol	Doxycycline	Oxycycline	Tetracycline	Gentamicin	Kanamycin	Vancomycin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Nalidixic acid	Ciprofloxacin
Lactococcus lactis																
EUCAST G+		>4R	ND	>8R	ND	ND	ND	ND	ND	>2R	>8R	ND	ND	ND	ND	ND
4	IMO1A5	>512	1024	>768	>1024	128	>512	>1024	>2048	64	1024	>2048	128	32	512	512
16	IMO1A16	128	512	64	1024	>1024	>512	256	>2048	1024	512	256	256	1024	512	128
118	M33A36	16	64	64	16	64	64	64	128	64	64	64	64	64	128	64
119	M33A37	16	64	8	16	64	64	4	16	>1024	64	>2048	1024	256	256	32
120	M33A38	16	16	128	256	128	64	32	512	2	2048	>2048	128	32	512	128
121	M33A39	64	256	256	64	16	>512	512	>2048	>1024	1024	>2048	>1024	>1024	512	128
125	M33A46	8	16	32	8	8	2	512	64	2	512	128	128	128	512	16
141	M33A57	1	4	8	1	8	2	512	8	1	512	128	64	64	512	16
143	M33A59	256	32	64	16	64	32	64	32	2	64	256	128	32	128	256
145	M33An25	512	3072	>768	128	512	>512	1024	>2048	>1024	256	>2048	>1024	>1024	>1024	>512
164	M31A12	64	64	32	>1024	128	512	>1024	>2048	>1024	512	256	128	32	256	>512
166	M31A12-B	64	64	16	>1024	128	512	>1024	>2048	>1024	512	128	512	>1024	1024	>512
179	M31An02	16	32	16	8	128	64	>1024	128	64	512	64	128	64	256	>512
Lactococcus taiwanensis																
167	M31A13-B	16	64	16	>1024	128	512	>1024	>2048	>1024	512	128	512	>1024	1024	>512

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint

Table 37.3. MIC concentrations of *E. coli*/*Shigella* species (aerobic conditions)

Strain designations		Clindamycin	Erythromycin	Chloramphenicol	Doxycycline	Oxycycline	Tetracycline	Gentamicin	Kanamycin
				Escherichia/Shigella					
EUCAST MIC		ND	ND	>16R	>4R	ND	ND	>2R	>8R
G- anaerobe MIC		>4R	ND	>8R	ND	ND	ND	≥4	ND
8	IM01A8	>512	128	256	128	8	4	2	256
15	IM01A15	256	128	256	128	16	32	2	128
17	IM01A17	256	128	256	128	8	32	4	512
20	IM01A21	256	>32	64	128	512	256	4	16
21	IM01A22	>32	>1024	128	>64	>128	>32	>512	64
41	IM20A19	256	1024	64	512	32	32	512	256
58	IM01B1	256	1024	16	1024	32	32	512	256
59	IM01B2	>512	64	128	512	8	16	2	16
63	IM01B8	512	64	16	8	32	4	8	16
64	IM01B9	256	1024	256	64	16	4	1	512
70	IM01B19	ND	8	16	ND	4	4	ND	ND
74	IM01B26	256	128	128	8	16	16	2	8
77	IM01B30	>512	64	128	4	4	32	2	8
133	M33A40	512	512	256	64	16	32	512	256
163	M31A11	64	<0.5	32	2	4	64	512	8
169	M31A14	32	64	16	1	2	4	2	16
171	M31A16	>512	64	16	0.5	16	4	8	32
186	M31An09	>512	64	16	16	512	128	2	8
188	M31An11	>512	64	16	16	512	128	2	8
191	M31An16	128	64	16	16	>1024	128	2	8
192	M31An17	64	64	8	2	4	4	2	16
193	M31An18	64	64	16	2	4	4	4	16

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint

Table 37.3 contd. MIC concentrations of *E. coli*/*Shigella* species (aerobic conditions)

Strain designations		Vancomycin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Nalidixic acid	Ciprofloxacin
				E. coli/Shigella				
EUCAST MIC		ND	>8R	ND	>4R	>0.125R	ND	>0.64R
CLSI MIC		≥32	≥32	ND	ND	≥4	ND	≥4
8	IM01A8	64	4096	> 2048	512	32	256	2
15	IM01A15	32	8	256	256	32	256	0.5
17	IM01A17	64	8	256	256	32	128	0.5
20	IM01A21	64	4096	> 2048	>1024	32	128	2
21	IM01A22	ND	256	256	256	>32	>256	>16
41	IM20A19	256	256	32	256	32	128	8
58	IM01B1	256	256	256	256	32	256	1
59	IM01B2	128	4096	>2048	512	64	256	64
63	IM01B8	1024	1024	2048	1024	32	256	1
64	IM01B9	1024	1024	256	>1024	128	256	32
70	IM01B19	256	ND	64	256	ND	ND	ND
74	IM01B26	64	16	256	256	32	256	64
77	IM01B30	32	8	256	256	32	256	128
133	M33A40	256	32	256	128	32	256	16
163	M31A11	>1024	256	1024	256	512	64	4
169	M31A14	64	4	32	128	32	32	0.5
171	M31A16	64	8	32	256	>1024	512	>512
186	M31An09	64	4096	>2048	512	32	16	0.125
188	M31An11	64	4096	>2048	512	32	32	0.125
191	M31An16	64	>4096	>2048	512	32	2	0.125
192	M31An17	64	8	32	128	32	2	0.125
193	M31An18	64	8	32	128	32	2	0.125

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint

Table 37.4. MIC concentrations of *E. coli/Shigella* species (anaerobic conditions)

Strain designations		Clindamycin	Erythromycin	Chloramphenicol	Doxycycline	Oxycycline	Tetracycline	Gentamicin	Kanamycin
					E. coli/Shigella				
EUCAST MIC		ND	ND	>16R	>4R	ND	ND	>2R	>8R
G- anaerobe MIC		>4R	ND	>8R	ND	ND	ND	≥4R	ND
8	IM01A8	512	128	16	16	64	16	4	16
15	IM01A15	512	256	128	>1024	1024	512	4	16
17	IM01A17	>512	256	32	1	32	32	512	256
20	IM01A21	256	64	8	1	8	4	512	128
21	IM01A22	>512	256	128	256	1024	256	4	16
41	IM20A19	ND	<64	32	256	128	32	128	16
58	IM01B1	512	128	64	64	256	128	8	32
59	IM01B2	128	3072	32	64	32	>512	128	16
63	IM01B8	128	256	128	256	64	4	4	32
64	IM01B9	256	128	8	128	64	>512	8	128
70	IM01B19	256	256	128	1024	1024	32	8	32
74	IM01B26	256	128	16	0.5	1	64	8	32
77	IM01B30	512	128	16	256	2	>512	>3072	32
133	M33A40	>512	64	8	1	0.5	4	4	16
163	M31A11	256	64	16	1	0.5	4	64	256
169	M31A14	256	64	8	16	512	128	8	32
171	M31A16	>512	64	16	16	512	128	8	32
186	M31An09	256	64	8	16	512	128	8	16
188	M31An11	128	64	8	1	1	4	8	32
191	M31An16	128	64	8	1	1	4	8	32
192	M31An17	>512	64	8	1	1	4	8	32
193	M31An18	256	64	16	0.25	0.5	4	4	32

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint.

Table 37.4 contd. MIC concentrations of *E. coli/Shigella* species (anaerobic conditions)

Strain designations		Vancomycin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Nalidixic acid	Ciprofloxacin
				E. coli/Shigella				
EUCAST MIC		ND	>8R	ND	>4R	>0.125R	ND	>0.64R
CLSI MIC		≥32	≥32	ND	ND	≥4	ND	≥4
8	IM01A8	128	4096	>2048	1024	32	32	1
15	IM01A15	128	4096	>2048	1024	32	2	<0.125
17	IM01A17	512	4	64	128	16	256	128
20	IM01A21	128	4	64	128	16	512	<0.125
21	IM01A22	128	4096	>2048	1024	32	4	<0.125
41	IM20A19	128	<16	16	128	16	128	8
58	IM01B1	512	4	128	128	32	32	32
59	IM01B2	512	64	64	256	32	1024	1
63	IM01B8	512	4096	>2048	1024	32	4	1
64	IM01B9	>1024	512	256	1024	32	512	256
70	IM01B19	128	4	32	1024	32	8	>512
74	IM01B26	64	8	32	1024	32	32	8
77	IM01B30	128	128	32	128	32	256	64
133	M33A40	64	8	32	128	32	8	0.125
163	M31A11	128	4	128	256	>1024	512	64
169	M31A14	64	2048	64	1024	32	4	0.125
171	M31A16	512	2048	>2048	512	32	4	0.125
186	M31An09	64	2048	>2048	512	32	1	0.125
188	M31An11	64	8	128	128	32	1	0.125
191	M31An16	64	8	32	128	32	1	0.125
192	M31An17	64	1024	>2048	256	32	1	0.125
193	M31An18	512	2048	>2048	512	32	2	0.125

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint.

Table 37.5. MIC concentrations of *Paenibacillus lactis*

Strain designations		Clindamycin	Erythromycin	Chloramphenicol	Doxycycline	Oxycycline	Tetracycline	Gentamicin	Kanamycin	Vancomycin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Nalidixic acid	Ciprofloxacin
Paenibacillus lactis																
EUCAST G+		>4R	ND	>8R	ND	ND	ND	ND	ND	>2 R	>8R	ND	ND	ND	ND	ND
132	M33A34	128	16	128	16	64	2	256	4	64	32	256	128	32	64	256
136	M33A47	0.5	32	8	4	16	8	32	64	32	2	128	2	32	16	16
137	M33A48	16	64	64	16	64	2	64	>2048	0.5	32	32	64	32	64	16
138	M33A49	128	4	128	16	32	2	256	4	0.5	32	256	256	32	64	256
139	M33A50	32	0.5	8	0.125	0.5	2	0.5	16	0.5	32	32	128	32	64	0.5
140	M33A51	8	16	16	4	2	2	0.5	32	0.5	512	128	128	32	64	2

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint.

Table 37.6. MIC concentrations of *Lactobacillaceae* species

Strain designations		Clindamycin	Erythromycin	Chloramphenicol	Doxycycline	Oxycycline	Tetracycline	Gentamicin	Kanamycin	Vancomycin	Ampicillin	Carbenicillin	Cefaclor	Ceftriaxone	Nalidixic acid	Ciprofloxacin
								Lactobacillus ruminis								
EUCAST G+		>4R	ND	>8R	ND	ND	ND	ND	ND	>2R	>8R	ND	ND	ND	ND	ND
82	IM20B2	4	512	16	16	8	16	64	8	64	256	64	32	128	64	64
85	IM20B6	1	128	32	16	8	16	16	32	128	512	128	128	64	256	32
104	IM20B28	64	2	1	2	4	2	2	1	4	4	8	256	32	4	8
105	IM20B29	1	1	32	0.125	0.5	2	16	64	32	1024	>1024	64	32	128	0.5
82	IM20B2	4	512	16	16	8	16	64	8	64	256	64	32	128	64	64
								Pediococcus pentosaceus								
87	IM20B9	64	1024	512	64	128	128	1024	256	1024	1024	1024	512	>1024	512	256

All concentrations in mg/L; ND, not determined in database; highlighted cells exhibit values above resistance breakpoint.

Resistome analysis

Resistance determinants were identified by CARD RGI and verified by CARD curated BlastP. RGI catalogs resistome predictions by both drug class and resistance mechanisms. CARD utilizes two models for the detection of AR, protein homolog models and protein variant models. Resistance genes are classified as protein homolog models based on similarity to a curated reference sequence of functional homologs. Protein homolog models are categorized by three criteria: perfect, strict and loose. All protein homolog models predicted for the taxa reported here were classified as perfect or strict. Perfect classification detects matches 100% identical to curated reference sequences and mutations. Strict classification detects homolog models similar to those previously to the curated reference sequences and mutations with <100% matches.

Resistant genes are classified as protein variant models when resistance genes may not necessarily be functional. Submitted sequences are additionally searched for curated sets of mutations shown clinically to confer resistance relative to wild-type. Protein variant models are categorized as strict or loose. All protein variant models for the resistant taxa were categorized as strict. Strict matches contain a previously detected mutation of resistance variants. The CARD RGI and BlastP investigate the presence of resistance genes yet the presence alone of such genes does not indicate with certainty whether they are expressed or functional (protein variant models). Therefore, alternative methods should be utilized to determine if the genes are functional. Additionally, multiple genes that can be impacted by environmental

conditions, influence many resistance phenotypes. Efflux systems are especially influenced in this manner and CARD acknowledges this with continuous curation and is currently developing a meta-model to detect efflux pumps and mutations that may indicate over-expression.

There were differences detected in the number and type of resistance genes in the CARD curated RGI and BlastP databases. RGI detected both protein homolog and protein variant models where BlastP only detected homolog models. Resistance genes were examined within taxa and between taxa. The total number of genes for all resistant isolates and comparison strains in the RGI database is 178 and in BlastP detected 131 resistance genes.

Four resistance genes were detected in the *Lactobacillus* and *Pediococcus* species (Table 38.1). All were strict protein homolog models except SarpoB, which is a strict protein variant model. When compared to *L. ruminis* strain ATCC27782, Matses *L. ruminis* strains genomes IM20B2, IM20B6 and IM20B25 contained 2 additional PmrE genes. PmrE is a gene that alters cell wall charge to confer resistance to polymyxin. All strains shared resistance genes ScEFTU, SarpoB and dfrE conferring resistance to elfamycin, rifamycin and trimethoprim respectively. No resistance genes were detected for any of the antibiotics tested for MIC in this study. It is possible that novel machinery is present for the resistance displayed here within *L. ruminis* strains and will be further investigated.

CARD detected 26 resistance genes among the *P. lactis* strains (Table 38.2). All Peruvian *P. lactis* strains had an additional VatI gene when compared to the comparison strain and two additional resistance genes (otrA and vanYM). Comparison strain *P. lactis* 154 contained resistance genes, vanRM, vanSM, vanYF, mel, CIPa, msrE SAgyrB and clbB conferring resistance for vancomycin, macrolide, streptogramin, aminocoumarin and multidrug resistance.

All Matses *L. lactis* genomes contained six resistance genes (Table 38.3). *L. taiwanensis* genomes contained 6 resistance genes conferring resistance to elfamycin, chloramphenicol and lincosamide and general efflux pumps. Only the comparison strain *L. lactis* strain II1403 contained *Mycobacterium tuberculosis* rpoB mutants conferring resistance to rifampicin. Only resistant Peruvian strains contained *E. coli* EF-Tu mutants conferring resistance to elfamycin.

All resistant *Bacillis* species contained ScEFTU, ykkC, ykkD, blt and lmrB conferring resistance to elfamycin, chloramphenicol and tetracycyline, and multidrug resistance (Table 38.4). Only comparison stains contained SagyrB and an additional copy of Bmph gene. A total of 12 types of resistance genes were detected for the *Bacillus* species.

The *E. coli*/*Shigella* cluster had ~130 resistant genes (Table 38.5). Most of these genes confer multi-drug resistance by efflux pumps. Genes were classified as perfect or strict protein homologs or protein variant models as *E. coli* genes are well represented within the CARD curated RGI and BlastP datasets. Also present were resistance genes conferring resistance by inactivation enzymes, target replacement

protein and cell wall charge alteration. There were many observed genes that were shared between all strains. Some differences between comparison and Peruvian strains were observed. Genes unique to comparison strains included PmrA, PmrB, mexB and mexD (3 copies). Genes unique to Peruvian strains included APH(6)-Id, APH(3”)-Ib, dfrA8, afrA15, mtrD (3 copies), AAC(6’)-On8, efrB, cprR, oqxA, TEM-1, sul1 &2, ECUhpT, lsaA, and a second copy of EcacrA, CRP, EcemrE, mdtH, mdtK, mdtL, dfrE and tolC. As most of the genes are not antibiotic specific but multidrug resistance efflux pumps, it is difficult to determine which genes are responsible for displaying resistance to antibiotics tested here. To further complicate the issue, not all genes are well described and do not have antibiotics assigned specifically to the gene. The *E. coli* strains are of special interest as they are so well characterized and the genomes contained resistance genes to all classes of antibiotics tested, although no single organism was resistant to all antibiotics. Resistance genes were detected for all antibiotic classes testing in this study, accounting for the high level of resistance observed. Further investigation of these strains is currently underway. Transcriptome data will allow for determination of genes activity or expression in the cells in differing concentrations of each class of antibiotics.

Table 38.1. *Lactobacillaceae* resistance gene % similarity to database

Resistance gene	ScEFTU	PmrE	PmrE	SarpoB	dfrE
Strain					
Lactobacillus ruminis ATCC 27782	70**			71*	76**
Lactobacillus ruminis 82 IM20B2	69	60	60	71*	76**
Lactobacillus ruminis 85 IM20B6	69	60	60	71*	76**
Lactobacillus ruminis 101 IM20B25	69	63*	64*	71*	76**
Lactobacillus ruminis 105 IM20B29	69			71*	69*
Pediococcus pentosaceus ATCC 25745	69**			69*	70**
Pediococcus pentosaceus 87 IM20B9	69			69*	70**

* Only in Card database

**Only in blastP database

Table 38.2. *P. lactis* resistance gene % similarity to database

Resistance gene Strain	VgbC	mphI	lsaB	tetA(48)	BahA	VatI	VatI	rphB	TaeA
Paenibacillus lactis 154	61**	74**	80**	81**	81**	94**		82**	87**
Paenibacillus lactis 132 M33A34	60	75	79	80	80	81	96	82	86
Paenibacillus lactis 136 M33A47	60	75	79	80	80	81	96	82	86
Paenibacillus lactis 137 M33A48	60	75	79	80	80	81	96	82	86
Paenibacillus lactis 139 M33A50	60	75	79	80	80	81	96	82	86
Resistance gene Strain	**BLA1**	**LlmA**	**otrA**	**tetB(48)**	**vanRM**	**vanSM**	**vanYF**	**vanYM**	**SarpoB**
Paenibacillus lactis 154	64	94**		88**	70	65	60		74*
Paenibacillus lactis 132 M33A34	65**	94	65**	88				60*	74*
Paenibacillus lactis 136 M33A47	65**		65**					60*	74*
Paenibacillus lactis 137 M33A48	65**		65**					60*	74*
Paenibacillus lactis 139 M33A50	65**		65**					60*	74*

Table 38.2. contd. *P. lactis* resistance gene % similarity to database

Resistance gene	blaI	mel	ScEFTU	ScEFTU	CIPa	CIPa	msrE	SAgyrB	clbB
Strain									
Paenibacillus lactis 154	65**	65**	69**		93**	80*	65*	63*	70*
Paenibacillus lactis 132 M33A34	65**		69**	70**					
Paenibacillus lactis 136 M33A47	65**		69**						
Paenibacillus lactis 137 M33A48	65**		69**						
Paenibacillus lactis 139 M33A50	65**		69**						

Resistance gene	ECEFTU	ECEFTU
Paenibacillus lactis 132 M33A34	74*	76*
Paenibacillus lactis 136 M33A47	76*	
Paenibacillus lactis 137 M33A48	76*	
Paenibacillus lactis 139 M33A50	76*	

* Only in Card database
**Only in blastP database

Table 38.3. *L. lactis* resistance gene % similarity to database

Resistance gene	emeA	lmrC	lmrD	ScEFTU	patB	EcEFTU	MtrpoB
Lactococcus lactis Il1403	80	100	100	69**	62**		60
Lactococcus lactis 4 IM01A5	79	100	100	69**	62	71*	
Lactococcus lactis 16 IM01A16	79	95	100	69**	62	71*	
Lactococcus lactis 78 IM01B3	79	100	100	69**	62	71*	
Lactococcus lactis 118 M33A36(3)	79	100	100	69**		71*	
Lactococcus lactis 120 M33A38	79	100	100	69**	62	71*	
Lactococcus lactis 121 M33A39	79	100	95	69**	62	71*	
Lactococcus lactis 125 M33A46	79	100	100	69**	62	71*	
Lactococcus lactis 141 M33A57	79	100	100	69**	62	71*	
Lactococcus lactis 143 M33A59	79	100	100	69**	62	71*	
Lactococcus lactis 145 M33An25	79	100	100	69**	62	71*	
Lactococcus lactis 160 M31A12	79	100	95	69**	62	71*	
Lactococcus lactis 164 M31A12	79	100	95	69**	62	71*	
Lactococcus lactis 179 M31An02	79	100	100	69**	62	71*	

Table 38.3. contd. *L. taiwanensis* resistance gene % similarity to database

Resistance gene	emeA	lmrC	lmrD	ScEFTU	patB	EcEFTU
Strain						
Lactococcus taiwanensis 33 IM20A10	78	94	96	69**	62	71*
Lactococcus taiwanensis 37 IM20A15	78	94	96	69**	62	71*
Lactococcus taiwanensis 167 M31A13-B	78			69**	62	71*

* Only in Card database

**Only in blastP database

Table 38.4. *Bacillus* species resistance gene % similarity to database

Resistance gene	sav1866	ykkC	ykkD	blt	cat86	cat86	lmrB
Strain							
Bacillus altitudinis YNP4	60	64**	62	76	87		65**
Bacillus altitudinis 34 IM20A11	60	64**	62**	76	88		65
Bacillus altitudinis 45 IM20A23	60	65**	63**	76	87	87**	65
Bacillus altitudinis 46 IM20A24	69	65**	63**	76	92		65
Bacillus altitudinis 117 M33A35	60	66**	62**	76	89		65
Bacillus safensis KCTC12796BP		64**	64	76	92		65**
Bacillus safensis 116 M33A32	60	63**	63**	76	89		65
Bacillus xiamenensis VV3	60	66	63	77			65**
Bacillus xiamenensis 52 IM20A30	60	66**	63**	77			64
Bacillus xiamenensis 54 IM20A32	60	66**	63**	77			64

Table 38.4. contd. *Bacillus* species resistance gene % similarity to database

Resistance gene	BsmprF	Bmph	Bmph	ScEFTU	ScEFTU	SagyrB	SarpoB	EcEFTU
Strain								
Bacillus altitudinis YNP4	60	68**	68**	71**	63**	70*	80*	
Bacillus altitudinis 34 IM20A11	60	68		71**	63**		80*	74*
Bacillus altitudinis 45 IM20A23	60			71**	63**		80*	74*
Bacillus altitudinis 46 IM20A24	61			71**	63**		80*	74*
Bacillus altitudinis 117 M33A35	60	67		71**	63**		80*	74*
Bacillus safensis KCTC12796BP	60	67**	69*	71**	65**	70*	80*	
Bacillus safensis 116 M33A32	60	67		71**	63**			
Bacillus xiamenensis VV3				71**	64**	70*	80*	
Bacillus xiamenensis 52 IM20A30				71**	64**		80*	75*
Bacillus xiamenensis 54 IM20A32				71**	64**		80*	75*

* Only in Card database
**Only in blastP database

Table 38.5. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	EcacrA	EcacrA	acrB	acrB	acrD	acrD	acrE	acrE
Strain								
Escherichia coli 0157: H7	100**		92**		100**		100	
Escherichia coli str. K-12	100**		92**		100		100	
Escherichia coli ATCC 25922	100		92**		100**		99	
Shigella boydii ATCC 9210	100		92**		99	100**	99	85*
Shigella dysenteriae Sd197	100**		92**		99**		99*	85*
Shigella flexneri 2a str. 301	100**		92**		100*		86*	
Shigella sonnei Ss046	100**		92**		100**		99	86*
8 IM01A8	100		92**		100**		100	
17 IM01A17	100		92**		100**		100	
20 IM01A21	100	99*	92**		100**		100	
21 IM01A22	100		92**		100**		100	
41 IM20A19	100		92**		100**		99**	
58 IM01B1	100		92**		100**		100	
59 IM01B2	100		92**		100**		100	
63 IM01B8	100		100**		100**		100	
70 IM01B19	100	100*	92**		100**		100	
74 IM01B26	100		92**		100**		100	
77 IM01B30	100	90*	92**	92**	100**		77**/100 *	
133 M33A40	100	99*	92**		100**		100	
163 M31A11	100		92**		100**		99**	100
169 M31A14	100		92**		100**		99**	
193 M31An18	100		92**		100**	100	100	

Table 38.5. contd *E .coli/Shigella* species resistance gene % similarity to database

Resistance gene	acrF	acrS	KpacrR	aadA	aminoalaS	aminocysB
Strain						
Escherichia coli 0157: H7		100	80*		100**	100**
Escherichia coli str. K-12	100	100	80*		100**	100**
Escherichia coli ATCC 25922		99	81*		100**	100
Shigella boydii ATCC 9210	100**	100	80*		100**	100**
Shigella dysenteriae Sd197			80*		100**	99**
Shigella flexneri 2a str. 301			80*		99**	100**
Shigella sonnei Ss046	98*	99	80*	100	100**	100**
8 IM01A8	100**	100	80*		100	100
17 IM01A17	100**	100	80*		100	100
20 IM01A21	100**	100	80*		100	100
21 IM01A22	100**	100	80*	96	100	100
41 IM20A19	100**	98	80*		99	100
58 IM01B1	100**	100	80*		100	100
59 IM01B2	100**	100	80*		100	100
63 IM01B8	100**	100	80*		100	100
70 IM01B19	100**	100	80*		100	100
74 IM01B26	100**	100	80*		100	100
77 IM01B30	100**	100	80*		100	100
133 M33A40	100**	100	80*		100	100
163 M31A11	100**		80*		100	100
169 M31A14	99**	99	80*		100	100
193 M31An18	100**					

Table 38.5. contd. *E .coli/Shigella* species resistance gene % similarity to database

Resistance gene	APH(6)-Id	APH(3")-Ib	bacA	cpxA	cpxR	ACT-7	amrB
Strain							
Escherichia coli 0157: H7			100	100	100	76*	99*
Escherichia coli str. K-12			100	100	100	77*	
Escherichia coli ATCC 25922			100	100	100		
Shigella boydii ATCC 9210			100	100	100	77*	99*
Shigella dysenteriae Sd197				99	100	75*	98*
Shigella flexneri 2a str. 301			100	100	100	76*	99*
Shigella sonnei Ss046			99	100	100	77*	
8 IM01A8	99	100	100	100	100		
17 IM01A17			100	100	100		
20 IM01A21			100	100	100		
21 IM01A22			100	100	100		
41 IM20A19			100	100	99		
58 IM01B1			100	100	100		
59 IM01B2			100	100	100		
63 IM01B8	99	100	100	100	100		
70 IM01B19			100	100	100		
74 IM01B26			100	100	100		
77 IM01B30			100	100	100		
133 M33A40			100	100	100		
163 M31A11			100	100	100		
169 M31A14	99	100	100	100	99		
193 M31An18			100	100	100		

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	arnA	arnA	baeR	baeS	CMY-47	CMY-69
Strain						
Escherichia coli 0157: H7	99		100	99	76**	
Escherichia coli str. K-12	100		100	100	77**	
Escherichia coli ATCC 25922	99		100	96	77**	77*
Shigella boydii ATCC 9210	99		100	99	77**	
Shigella dysenteriae Sd197	99	100*	100	99	75**	
Shigella flexneri 2a str. 301		98*	99	99	77**	
Shigella sonnei Ss046		99	100	99	77**	
8 IM01A8	100		100	99	76**	99*
17 IM01A17	100		99	100	76**	100*
20 IM01A21	100		99	100	76**	100*
21 IM01A22	99		100	100	77**	100*
41 IM20A19	98		100	96	76**	95*
58 IM01B1	100		99	100	76**	100*
59 IM01B2	100		100	99	76**	99*
63 IM01B8	100		100	99	76**	99*
70 IM01B19	100		99	100	76**	100*
74 IM01B26	100		99	100	76**	100*
77 IM01B30	77	100	99	100	76**	100*
133 M33A40	100		100	100	76**	99*
163 M31A11	100		100	100	76**	99*
169 M31A14	98		100	96	77**	96*
193 M31An18	100		100	100	76**	100*

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	CRP	CRP	emrA	emrB	emrD	EcemrE	EcemrE
Strain							
Escherichia coli 0157: H7	100		99	88*/100**	99	98	
Escherichia coli str. K-12	100		100	100	99	100	
Escherichia coli ATCC 25922	100		100	100**	99	98	
Shigella boydii ATCC 9210	100		100	100**	99		
Shigella dysenteriae Sd197	100		99	100**	100		
Shigella flexneri 2a str. 301	100			99**	100	98	
Shigella sonnei Ss046	100		100	100**	100	98	
8 IM01A8	100		100	100	99	97	98
17 IM01A17	100		100	100	99	99	
20 IM01A21	100		100	100	99	99	
21 IM01A22	100		99	100	99	98	98
41 IM20A19	100		100	100**	99	98	
58 IM01B1	100		100	100	99	99	
59 IM01B2	100		100	100	99	97	98
63 IM01B8	100		100	100	99	97	98
70 IM01B19	100		100	100	99	99	
74 IM01B26	100		100	100	99	99	
77 IM01B30	100	99	100	100	99**	99	
133 M33A40	100		100	100	99	98	
163 M31A11	100		100	100	99		
169 M31A14	100		100	100	99	98	99
193 M31An18	100		100	100	99		

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	mdfA	mdtA	mdtB	mdtB	mdtC	mdtD	mdtD	mdtE
Strain								
Escherichia coli 0157: H7	97	99	100**		100**	99		99
Escherichia coli str. K-12	97	100	100		100**	100		100
Escherichia coli ATCC 25922	97	99	100		99**	98		100
Shigella boydii ATCC 9210	96	99	99**		100*	100	100	100
Shigella dysenteriae Sd197	99**/97*					97*		99
Shigella flexneri 2a str. 301	96	99	100**		99**	99		99*
Shigella sonnei Ss046	95*	98	99		100	99		100
8 IM01A8	97	100	100		100*	99		100
17 IM01A17	96	99	100**		100**	100		100
20 IM01A21	96	99	100**		100**	100		100
21 IM01A22	97	99**	99**		100**	100		100
41 IM20A19	97	99	100**		99**	99		100
58 IM01B1	96	99	100**		100**	100		100
59 IM01B2	97	100	100**		100**	99		100
63 IM01B8	97	100	100**		100**	99		100
70 IM01B19	96	99	100**		100**	100		100
74 IM01B26	96	99	100**		100**	100		100
77 IM01B30	96	99	89**	100**	100**	100		100
133 M33A40	97	100	100**		100**	100		100
163 M31A11	97	99	100**		100**	100		100
169 M31A14	97	99	100**		99**	98		100
193 M31An18	97	99	100**		100**	100		100

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	mdtF	mdtG	mdtH	mdtH	mdtK	mdtK	mdtL	mdtL
Strain								
Escherichia coli 0157: H7	100	100*	100		99*		98	
Escherichia coli str. K-12	100	100	100		82**/100*		100	
Escherichia coli ATCC 25922	100**	100	100		81		99	
Shigella boydii ATCC 9210	100**	99	100		82**/99*		100	
Shigella dysenteriae Sd197	100**	99*			99*		99	
Shigella flexneri 2a str. 301	100*		100				99	
Shigella sonnei Ss046	100	99*	99		82		100	
8 IM01A8	100*	100	100		82		100	
17 IM01A17	100**	100	100		82		99	
20 IM01A21	100**	100	100		82		99	
21 IM01A22	100**	100	100		82		99	
41 IM20A19	100**	100	100		81		99	
58 IM01B1	100**	100	100		82		99	
59 IM01B2	100**	100	100		82		100	
63 IM01B8	100**	100	100		82		100	
70 IM01B19	100**	100	100		82		99	99
74 IM01B26	100**	100	100		82		99	
77 IM01B30	100**	100	85	100	82	84	75	
133 M33A40	100**	100	100		82		100	
163 M31A11	100**	100	100		82		100	
169 M31A14	99**	100	100		82		99	
193 M31An18	100	100	100		82		100	

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	mdtM	mdtN	mdtO	mdtP	gadE	gadW	gadX
Strain							
Escherichia coli 0157: H7	99	99	99	98	100**	95**	98
Escherichia coli str. K-12	100	100	100	100	100**	95**	100
Escherichia coli ATCC 25922	97	99	97	98	100	100	93
Shigella boydii ATCC 9210	98	99	99	98	99**	93**	99
Shigella dysenteriae Sd197	97				99**		97
Shigella flexneri 2a str. 301		99	98	98	100**	95**	98
Shigella sonnei Ss046		100	98	98	100**	94**	99**
8 IM01A8	97	100	99	98	100	94	99
17 IM01A17	98	100	99	98	100	94	99
20 IM01A21	98	100	99	98		94	99
21 IM01A22	98	99	99	98	100	95	99
41 IM20A19		99	98	98	100	100	93
58 IM01B1	98	100	99	98	100	94	99
59 IM01B2	97	100	99	98	100	94	99
63 IM01B8	97	100	99	98	100	94	99
70 IM01B19	98	100	99	98	100	94	99
74 IM01B26	98	100	99	98	100	94	99
77 IM01B30	98	100	99	98	100	94	99
133 M33A40	98	100	100	100	100	95	100
163 M31A11	98	100	100	100	100	95	100
169 M31A14		99	97	98	99	100	93
193 M31An18	98	100	100	100	100	95	100

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	evgA	evgA	evgS	evgS	dfrA1	dfrA8	dfrA15
Strain							
Escherichia coli 0157: H7	100		98				
Escherichia coli str. K-12	100		100				
Escherichia coli ATCC 25922	100		97				
Shigella boydii ATCC 9210	100		98				
Shigella dysenteriae Sd197	100		97	96*			
Shigella flexneri 2a str. 301	100		99				
Shigella sonnei Ss046	100		98		84**/100*		
8 IM01A8	100			99		100	
17 IM01A17	100		99				
20 IM01A21	100		99				
21 IM01A22	100		98				100
41 IM20A19	100		97				
58 IM01B1	100		99				
59 IM01B2	100		99				
63 IM01B8	100		99			100	
70 IM01B19	100		99				
74 IM01B26	100		99				
77 IM01B30	68	100**	99				
133 M33A40	100		100				
163 M31A11			97**	100**			
169 M31A14	100		97			100	
193 M31An18							

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	emrK	emrR	emrY	emrY	dfrE	dfrE
Strain						
Escherichia coli 0157: H7	100	100	99		60**	
Escherichia coli str. K-12	100	100	100		60**	
Escherichia coli ATCC 25922	99	100	100	62*	60**	
Shigella boydii ATCC 9210	100	100	100	98*		
Shigella dysenteriae Sd197	99	99	100	99*	60**	
Shigella flexneri 2a str. 301	99	97	99	97*		
Shigella sonnei Ss046	99	100	100	98*	60**	
8 IM01A8	100	100	100		60**	
17 IM01A17	100	100	100		60**	
20 IM01A21	100	100	100		60**	
21 IM01A22	100	100	100		60**	
41 IM20A19	99	99	100	99*	60**	
58 IM01B1	100	100	100		60**	
59 IM01B2	100	100	100		60**	
63 IM01B8	100	100	100		60**	
70 IM01B19	100	100	100		60**	
74 IM01B26	100	100	100		60**	
77 IM01B30	100	100	100		60**	61**
133 M33A40	100	100	100**		60**	
163 M31A11	100	100	99		99**	60**
169 M31A14	99	100	100	98	60**	
193 M31An18	100	100	99		60**	

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	H-NS	H-NS	PmrA	PmrB	PmrC	PmrE	PmrF	mfd
Strain								
Escherichia coli 0157: H7	100		99*	99*	100	99	99	100**
Escherichia coli str. K-12	100		100*	100*	100	100	100	100**
Escherichia coli ATCC 25922	100	63**			96	99	99	99
Shigella boydii ATCC 9210	100		99*	99*	99	84	100	100**
Shigella dysenteriae Sd197	100		100*	98*	99	99	100	
Shigella flexneri 2a str. 301	100		99*	99*	99	97*	100	100**
Shigella sonnei Ss046	100		99*	99*	99	99	100	100**
8 IM01A8	100				100	89	100	100
17 IM01A17	100				100	84	100	100
20 IM01A21	100				100	84	100	100
21 IM01A22	100				100	89	100	100
41 IM20A19	100				96	100	99	99
58 IM01B1	100				100	84	100	100
59 IM01B2	100				100	89	100	100
63 IM01B8	100				100	89	100	100
70 IM01B19	100				100	84	100	100
74 IM01B26	100				100	84	100	100
77 IM01B30	100				100	84	100	100
133 M33A40	100				100	89	100	100
163 M31A11	100				100	99	100	100
169 M31A14	100				96	99	99	99
193 M31An18	100				100	84	100	100

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	ScEFTu	ScEFTu	EcEFtu	marA	EcmarR	msbA	msbA	msrB
Strain								
Escherichia coli 0157: H7	72**	72**		100	99*	100**	100**	100**
Escherichia coli str. K-12	72**	72**		100			100**	100**
Escherichia coli ATCC 25922	72	72		99	99*	100		100
Shigella boydii ATCC 9210	72**	72**		100	99*		100**	99**
Shigella dysenteriae Sd197	72**	72**		99	98*	100**		99**
Shigella flexneri 2a str. 301	72**	72**		100	87*		100**	99**
Shigella sonnei Ss046	72**	72**		100	99*		100**	100**
8 IM01A8				100	100*	100		100
17 IM01A17				100	99*	100		99**
20 IM01A21				100	99*	100		99**
21 IM01A22				100	100*		100	100
41 IM20A19				99	99*		100	100
58 IM01B1	73	76**		100	99*	100		99**
59 IM01B2				100	100*	100		100
63 IM01B8				100	100*	100		100
70 IM01B19				100	99*	100		99**
74 IM01B26				100	99*	100	100	99**
77 IM01B30				100	100*	100		99**
133 M33A40				100	99*	100		100
163 M31A11			100	100	99*	100		100
169 M31A14				99	99*	100		100
193 M31An18				100	100*	100		100

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	mexB	mexD	mexD	mexD	mexN	mexN	robA
Strain							
Escherichia coli 0157: H7	99*	92*	99*	99*	99*	99*	83
Escherichia coli str. K-12		92*					83
Escherichia coli ATCC 25922							83
Shigella boydii ATCC 9210		92*	99*	99*	99*	99*	83**
Shigella dysenteriae Sd197		91*	99*	99*			82**
Shigella flexneri 2a str. 301		92*	99*		99*	99*	83*
Shigella sonnei Ss046		92*	99*	99*	99*	99*	83
8 IM01A8					99*	99*	83
17 IM01A17					99*	99*	83
20 IM01A21					99*	99*	83
21 IM01A22					99*	99*	83
41 IM20A19					98*	99*	83
58 IM01B1					99*	99*	83
59 IM01B2					99*	99*	83
63 IM01B8					99*	99*	83
70 IM01B19					99*	99*	83
74 IM01B26					99*	99*	83
77 IM01B30					99*	99*	72/83**
133 M33A40					99*	99*	83
163 M31A11					99*	99*	83
169 M31A14					98*	99*	83
193 M31An18					99*	99*	83

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	phoB	phoQ	phoP	leuO	patA	tolC	tolC	YojI
Strain								
Escherichia coli 0157: H7	99*	99*		99**	100**	100		
Escherichia coli str. K-12		100*	100*	100**	100**	100		
Escherichia coli ATCC 25922				99	100	100		
Shigella boydii ATCC 9210		100*	99*	100**	100**	100		
Shigella dysenteriae Sd197		99*	99*			100		
Shigella flexneri 2a str. 301		99*	99*	100**	99**	100		
Shigella sonnei Ss046		100*	99*	100**	99**	100		
8 IM01A8				100	100	100		
17 IM01A17				100	99	100		
20 IM01A21				100	99	100		
21 IM01A22				99	100	100		
41 IM20A19				99	100	100		
58 IM01B1				100	99	100		
59 IM01B2				100	100	100		
63 IM01B8				100	100	100		
70 IM01B19				100	99	100		
74 IM01B26				100	99	100		
77 IM01B30				100	99	82		100
133 M33A40				99	100	100		
163 M31A11				100	100	100	100	
169 M31A14				99	100	100		
193 M31An18				100	100	100		

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	kdpE	ECGlpt	smeE	smeB	mtrD	mtrD	mtrD
Strain							
Escherichia coli 0157: H7	99**						
Escherichia coli str. K-12	100**						
Escherichia coli ATCC 25922	100	100*	62*				
Shigella boydii ATCC 9210	99**						
Shigella dysenteriae Sd197	98**						
Shigella flexneri 2a str. 301							
Shigella sonnei Ss046	99**			99*			
8 IM01A8	99	99*	99*		92*	99*	99*
17 IM01A17	99	99*	99*		92*	99*	99*
20 IM01A21	99	99*	99*		92*	99*	99*
21 IM01A22	99	100*	99*		92*	99*	99*
41 IM20A19	100	99*	99*		92*	99*	99*
58 IM01B1	99	99*	72*		92*	99*	99*
59 IM01B2	99	99*	99		92*	99*	99*
63 IM01B8	99*	99*	99*		92*	99*	99*
70 IM01B19	99	99*	99*		92*	99*	99*
74 IM01B26	99	99*	99*		92*	99*	99*
77 IM01B30	99	99*	99*		92*	99*	99*
133 M33A40	100		99*		92*	99*	99*
163 M31A11	100		99*		92*	99*	99*
169 M31A14	100	99*	99*		92*	99*	99*
193 M31An18	100				92*	99*	99*

Table 38.5. contd. *E. coli/Shigella* species resistance gene % similarity to database

Resistance gene	sat-1	sdiA	sdiA	AAC(6')- Ib8	efrB	cprR	oqxA
Strain							
Escherichia coli 0157: H7		71**					
Escherichia coli str. K-12		71**					
Escherichia coli ATCC 25922		71**					
Shigella boydii ATCC 9210		66**	75**				
Shigella dysenteriae Sd197							
Shigella flexneri 2a str. 301		67**					
Shigella sonnei Ss046	100	72**					
8 IM01A8		72**					
17 IM01A17		72**					
20 IM01A21		72**					
21 IM01A22		71**					
41 IM20A19		71**					
58 IM01B1		72**		68**			
59 IM01B2		72**		68**			
63 IM01B8		72**		68**			
70 IM01B19		72**		68**			
74 IM01B26		72**		68**			
77 IM01B30		67**	72**			95	97
133 M33A40		73**					
163 M31A11		73**			100		
169 M31A14		71**					
193 M31An18		72**					

Table 38.5. contd. *E .coli/Shigella* species resistance gene % similarity to database

Resistance gene	TEM-1	sul1	sul2	ECUhpT	lsaA
Strain					
Escherichia coli 0157: H7					
Escherichia coli str. K-12					
Escherichia coli ATCC 25922					
Shigella boydii ATCC 9210					
Shigella dysenteriae Sd197					
Shigella flexneri 2a str. 301					
Shigella sonnei Ss046					
8 IM01A8	100		100		
17 IM01A17					
20 IM01A21					
21 IM01A22	100	100		100*	
41 IM20A19				99*	
58 IM01B1					
59 IM01B2					
63 IM01B8	100		100		
70 IM01B19					
74 IM01B26					
77 IM01B30				99*	
133 M33A40					
163 M31A11					99
169 M31A14	100		100		
193 M31An18	100				

* Only in Card database

**Only in blastP database

Discussion

Genome analysis of the Matses resistome using the CARD databases revealed genes encoding for resistance to the all classes of antibiotics in the study. *E. coli/Shigella* cluster contained by far the most genes as expected as these resistance genes were best classified and described within the CARD curated databases. The continued development of the CARD database will aid in the identification of additional resistance genes through the meta model functions. This study is but a first step in the investigations of the immense diversity of antibiotic resistance mechanisms, in the Matses microbiome. Differences between taxa specific resistance genes and MIC resistance determination varied greatly suggesting novel machinery or novel functions for previously described genes. An example of this is *P. pentosaceus* strain IM20B9 only containing three resistance determinants but displaying high levels of resistance to all antibiotics in susceptibility testing. Additionally, all *Bacillus* strains contained multiple chloramphenicol resistance strains but only *B. altitudensis* strains displayed resistance suggesting additional required mechanisms.

Paenibacillus strain M33A34 was found to be 100% similar based on ANI analysis but two additional genes, TetB and LlmA, were detected when compared to other Matses resistance strains suggesting ANI analysis is not an adequate prediction of resistance gene detection between strains. Of particular interest of strain M33A34 is gene LlmA, conferring resistance to clindamycin, which was only recently discovered in an isolated cave system from a *Paenibacillus* strain (Pawlowski *et al.*,

2016). Similarly, the Matses resistome, also from an isolated area, has the potential to harbor novel machinery. Examples of novel machinery, such as gene LlmA, conferring resistance to clindamycin, will likely be revealed upon further transcriptome analyses. Strain M33AA34 did exhibit resistance to clindamycin but this was not unique to the strain as strain M33A49 also was resistant. Similarly, MIC results showed that TetB did not uniquely confer resistance to tetracycline antibiotics as strains M33A34 and M33A49 were resistant
to oxytetracycline and no strains were resistant to doxytetracycline or tetracycline.

Lactococcus strains genomes were interesting as only 6 genes were detected within all genomes except *L. taiwanensis* strain M31A13-B which did not contain missing LmrC and LmrD conferring resistance to lincosamides. The four genes shared by all strains, EcEFTu aand ScEFTu, patB and emeA, confer resistance to elfamycin, rifampcin, fluoroquinone and multidrug efflux pumps respectively. Although strain M31A13-B contained only four resistance genes it displayed resistance to all antibiotics tested. Of the *L. lactis* strains, IM01A15, IM01A16, M33An25 and M31A12-B displayed resistance to all antibiotics tested as well. The patB confers resistance to fluoroquinone antibiotics and all strains were found to be highly resistant to nalidixic acid and most were resistant to ciprofloxacin. The occurrence of genes responsible for synthetic antibiotics is especially interesting as the Matses community groups have not been exposed to therapeutic antibiotics. The occurrence of these genes and the resistance levels of fluoroquninones suggests a natural reservoir for such resistance determinants. If emeA was responsible for the

remaining antibiotic class resistance, it would be predicted that all *Lactococcus* strains would be resistant, and as some strains are susceptible, there is likely additional resistant machinery present but not detected and/or the current machinery is poorly described.

The *E. coli/Shigella* organisms contained resistance genes for all tested classes of antibiotics and overall, the genes occurred in all strains with few exceptions. However MIC revealed patterns of sporadic resistance that were not well correlated to the genes. This is most likely due to the regulation of efflux mechanisms (Blair *et al.*, 2015). This study has only begun to scratch the surface of the immense diversity of antibiotic resistance machinery in the Matses microbiome. Finally, the diversity in the resistome also suggests that there are a myriad of bioactive molecules with antibiotic properties waiting to be discovered. Some of these may have the potential to be productive leads as new antibiotics.

Conclusions

Moving forward to reduce resistance

Historically, it is commonly accepted that antibiotics have been over prescribed and used where not appropriate. Even when antibiotics are prescribed appropriately, often they are not taken as directed with patients opting to stop taking an antibiotic when they feel better and therefore not completing treatment regiments. The consequence is that organisms are not completely killed off, leading to resistant organisms that proliferate. The widespread use of antibiotics in agriculture can serve as a driver for evolution and spread of bacterial resistance.

In 2016, only five of the top 50 drug companies were actively developing new antibiotics. While the reasons for the sharp decline in antibiotic trials are numerous, one glaringly, obvious reason is the expense. Since 1995, one company still actively pursuing antibiotic research and design, GlasxoSmithKline analyzed some 300 new antibiotic targets at an estimated cost of $1.6 billion. Of the 300 targets, only 7 percent resulted in initially promising leads but all ultimately failed in clinical trials (Jones, 2010, Czaplewski *et al.*, 2016). Other key reasons for the lack of novel antibiotics are the time intensive research and development process and safety considerations of the compounds. In many cases, promising compounds are not considered tolerable or safe at the high concentration required to work effectively. In instances where a patient is already very ill, the drug candidates must be both safe for the patient and effective against the resistant infection.

Yet another challenge to the development of new antibiotics is the lack of business incentives. For years there has been an inverse relationship between the challenges and rising occurrence of antibiotic resistance and the incentives and rewards to address the issue. Only recently, the federal government passed rules that have ushered in the development of dozens of new drugs by extending market exclusivity to new antibiotics (Brogan & Mossialos, 2013).

Antibiotic stewardship

Antibiotic stewardship programs and education workshops help to preserve the effectiveness of antibiotic treatments. Recently, former President Barack Obama proclaimed November 13 – November 19 as Get Smart About Antibiotic Week and called for medical professionals and scientists to observe the week by "promoting the responsible use of antibiotics and raising awareness of the dangers inherent in their misuse and overuse". Furthermore, at a White House hosted forum, a directive was issued that directed the federal government to purchase meats from sources that follow responsible antibiotic use and that have raised animals in accordance with the same responsible-use policies. Programs like the National Action Plan for Combating Antibiotic-resistant bacteria (https://www.cdc.gov/drugresistance) and the WHO Global Action Plan on Antimicrobial Resistance (http://www.wpro.who.int/topics/drug_resistance/en) highlight the issue in a collaborative effort to not only to better inform the public but also to identify key actions that can combat the rise of antibiotic resistant organisms. The National

Action Plan also provides an improved Federal fund set aside to combat and prevent antibiotic resistance totaling more than $1.2 billion. Programs like the FDA and CDC's antibiotic stewardship program - the Get Smart Campaign (https://www.cdc.gov/getsmart/) – help ensure that patients better understand their antibiotic treatment program. These programs provide information for the patient about taking the right antibiotic at the right dosage for the prescribed amount of time, which improves consumer and provider education around appropriate antibiotic usage and the spread of resistance (Fig 12).

Ongoing stewardship activity has been shown to have the most measureable effects on inappropriate antibiotic use in a recent report by the Pew Charitable Trusts (PCT). Understanding the origin and mechanisms of ARGS can further contribute to the development and expansion of promising new antibacterial drug candidates and new combination therapies.

Figure 12. Examples of the Center for Disease Control initiative Get Smart About Antibiotic Use and Resistance program print materials (www.cdc.gov/getsmart)

Understanding the role of the intestinal microbial community and its complex relationship with the host is of great importance. The diverse taxa that reside within the gastrointestinal microbial community are routinely identified through gene sequencing and cultivation methods that allow for discerning novel species diversity and provide insights into their function within the gastrointestinal tract environment. Despite large scale sequencing efforts, the full extent of gut microbiome remains unexplored. As the organisms representing most abundant phylotypes of the human gut continue to be cultivated and characterized, isolated microbes can provide a better understanding the diversity and functionality of species within this group. A combination of bacterial isolation and cultivation and molecular sequencing methods has resulted in the comprehensive classification of four new taxa. The study of novel species with an unknown role within the gastrointestinal tract is of considerable importance. The microbiome includes a community of symbionts and commensals that have effect on immune function and nutrient processing and the role of novel taxa in health or disease functions has led to unprecedented sequencing and cultivation efforts. Globalization affects potentially beneficial bacteria harbored in the human body. The developing world harbors a number of biological biases that are absent in the developed world. Novel taxa and novel microbial community profiles have been reported when compared to Western populations. Furthermore, isolates from these communities displayed a high level of resistance to a number of antibiotic classes suggesting that while antibiotic therapy is a driver of selection for resistance genes within the microbial community, antibiotic resistance machinery is

present even when such drivers are absent. Further investigations of these organisms may have implications of drug discovery. The study of traditional, indigenous communities provides a unique opportunity to expand investigations of ‘missing’ gastrointestinal microbes when modern populations are compared to ancient populations.

References

Allen HK, Moe LA, Rodbumrer J, Gaarder A & Handelsman J (2009) Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *The ISME journal* **3**: 243-251.

Allen-Vercoe E, Daigneault M, White A, Panaccione R, Duncan SH, Flint HJ, O'Neal L & Lawson PA (2012) Anaerostipes hadrus comb. nov., a dominant species within the human colonic microbiota; reclassification of Eubacterium hadrum Moore et al. 1976. *Anaerobe* **18**: 523-529.

Arason VA, Kristinsson KG, Sigurdsson JA, Stefansdottir G, Mölstad S & Gudmundsson S (1996) Do antimicrobials increase the carriage rate of penicillin resistant pneumococci in children? cross sectional prevalence study. *BMJ: British Medical Journal* **313**: 387.

Barcenilla A, Pryde SE, Martin JC, Duncan SH, Stewart CS, Henderson C & Flint HJ (2000) Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Applied and environmental microbiology* **66**: 1654-1661.

Bartoloni A, Bartalesi F, Mantella A, Dell'Amico E, Roselli M, Strohmeyer M, Barahona HG, Barrón VP, Paradisi F & Rossolini GM (2004) High prevalence of acquired antimicrobial resistance unrelated to heavy antimicrobial consumption. *Journal of Infectious Diseases* **189**: 1291-1294.

Belenguer A, Holtrop G, Duncan SH, Anderson SE, Calder AG, Flint HJ & Lobley GE (2011) Rates of production and utilization of lactate by microbial communities from the human colon. *FEMS microbiology ecology* **77**: 107-119.

Bhullar K, Waglechner N, Pawlowski A, Koteva K, Banks ED, Johnston MD, Barton HA & Wright GD (2012) Antibiotic resistance is prevalent in an isolated cave microbiome. *PloS one* **7**: e34953.

Blair JM, Webber MA, Baylay AJ, Ogbolu DO & Piddock LJ (2015) Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology* **13**: 42-51.

Blaser MJ (2014) *Missing microbes: how the overuse of antibiotics is fueling our modern plagues*. Macmillan.

Blaser MJ, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Estrada I, Gao Z, Clemente JC, Costello EK & Knight R (2013) Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and US residents. *The ISME journal* **7**: 85.

Blaut M & Clavel T (2007) Metabolic diversity of the intestinal microbiota: implications for health and disease. *The Journal of nutrition* **137**: 751S-755S.

Brogan DM & Mossialos E (2013) Incentives for new antibiotics: the Options Market for Antibiotics (OMA) model. *Globalization and health* **9**: 58.

Bui TPN, Shetty SA, Lagkouvardos I, Ritari J, Chamlagain B, Douillard FP, Paulin L, Piironen V, Clavel T & Plugge CM (2016) Comparative genomics and physiology of the butyrate‐producing bacterium Intestinimonas butyriciproducens. *Environmental microbiology reports* **8**: 1024-1037.

Cabello FC (2006) Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. *Environmental microbiology* **8**: 1137-1144.

Chassard C, Scott KP, Marquet P, Martin JC, Del'homme C, Dapoigny M, Flint HJ & Bernalier‐Donadille A (2008) Assessment of metabolic diversity within the intestinal microbiota from healthy humans using combined molecular and cultural approaches. *FEMS microbiology ecology* **66**: 496-504.

Cho I & Blaser MJ (2012) The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* **13**: 260-270.

Clavel T, Desmarchelier C, Haller D, Gérard P, Rohn S, Lepage P & Daniel H (2014) Intestinal microbiota in metabolic diseases: from bacterial community structure and functions to species of pathophysiological relevance. *Gut microbes* **5**: 544-551.

Collins M, Lawson P, Willems A, Cordoba J, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H & Farrow J (1994) The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. *International Journal of Systematic and Evolutionary Microbiology* **44**: 812-826.

Consortium HMP (2012) Structure, function and diversity of the healthy human microbiome. *nature* **486**: 207.

Contreras M, Costello EK, Hidalgo G, Magris M, Knight R & Dominguez-Bello MG (2010) The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology* **156**: 3282-3287.

Control CfD & Prevention (2013) *Antibiotic resistance threats in the United States, 2013*. Centres for Disease Control and Prevention, US Department of Health and Human Services.

Cummings J & Macfarlane G (1991) The control and consequences of bacterial fermentation in the human colon. *Journal of Applied Microbiology* **70**: 443-459.

Czaplewski L, Bax R, Clokie M, Dawson M, Fairhead H, Fischetti VA, Foster S, Gilmore BF, Hancock RE & Harper D (2016) Alternatives to antibiotics—a pipeline portfolio review. *The Lancet Infectious Diseases* **16**: 239-251.

D'Costa VM, McGrann KM, Hughes DW & Wright GD (2006) Sampling the antibiotic resistome. *Science* **311**: 374-377.

D’Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula G, Calmels F & Debruyne R (2011) Antibiotic resistance is ancient. *Nature* **477**: 457-461.

Dabek M, McCrae SI, Stevens VJ, Duncan SH & Louis P (2008) Distribution of β-glucosidase and β-glucuronidase activity and of β-glucuronidase gene gus in human colonic bacteria. *FEMS microbiology ecology* **66**: 487-495.

Dethlefsen L, McFall-Ngai M & Relman DA (2007) An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**: 811-818.

Duncan SH, Louis P & Flint HJ (2004) Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Applied and environmental microbiology* **70**: 5810-5817.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE & Relman DA (2005) Diversity of the human intestinal microbial flora. *science* **308**: 1635-1638.

Flint HJ, Duncan SH, Scott KP & Louis P (2007) Interactions and competition within the microbial community of the human colon: links between diet and health. *Environmental microbiology* **9**: 1101-1111.

Flint HJ, Scott KP, Louis P & Duncan SH (2012) The role of the gut microbiota in nutrition and health. *Nature Reviews Gastroenterology and Hepatology* **9**: 577-589.

Flint HJ, Scott KP, Duncan SH, Louis P & Forano E (2012) Microbial degradation of complex carbohydrates in the gut. *Gut microbes* **3**: 289-306.

Forslund K, Sunagawa S, Kultima JR, Mende DR, Arumugam M, Typas A & Bork P (2013) Country-specific antibiotic use practices impact the human gut resistome. *Genome research* **23**: 1163-1169.

Fournier P-E, Lagier J-C, Dubourg G & Raoult D (2015) From culturomics to taxonomogenomics: A need to change the taxonomy of prokaryotes in clinical microbiology. *Anaerobe*.

Gordon JI (2012) Honor thy gut symbionts redux. *Science* **336**: 1251-1253.

Gupta RS & Gao B (2009) Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridiumsensu stricto (cluster I). *International journal of systematic and evolutionary microbiology* **59**: 285-294.

Hattori M & Taylor TD (2009) The human intestinal microbiome: a new frontier of human biology. *DNA research* **16**: 1-12.

Hungate RE (1957) Microoganisms in the rumen of cattle fed a constant ration. *Canadian Journal of Microbiology* **3**: 289-311.

Hutson R, Thompson D & Collins M (1993) Genetic interrelationships of saccharolytic Clostridium botulinum types B, E and F and related clostridia as

revealed by small-subunit rRNA gene sequences. *FEMS microbiology letters* **108**: 103-110.

Hutson R, Thompson D, Lawson P, Schocken-Itturino R, Böttger E & Collins M (1993) Genetic interrelationships of proteolyticClostridium botulinum types A, B, and F and other members of theClostridium botulinum complex as revealed by small-subunit rRNA gene sequences. *Antonie Van Leeuwenhoek* **64**: 273-283.

Jernberg C, Löfmark S, Edlund C & Jansson JK (2010) Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* **156**: 3216-3223.

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S & Sharma AN (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic acids research* **45**: D566-D573.

Jones D (2010) The antibacterial lead discovery challenge. *Nat Rev Drug Discov* **9**: 751-752.

Jorgensen JH & Turnidge JD (2015) Susceptibility test methods: dilution and disk diffusion methods. *Manual of Clinical Microbiology, Eleventh Edition,*p.^pp. 1253-1273. American Society of Microbiology.

Kämpfer P & Kroppenstedt RM (1996) Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Canadian journal of microbiology* **42**: 989-1005.

Kim M, Oh H-S, Park S-C & Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International journal of systematic and evolutionary microbiology* **64**: 346-351.

Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, Park S-C, Jeon YS, Lee J-H & Yi H (2012) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International journal of systematic and evolutionary microbiology* **62**: 716-721.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution* **16**: 111-120.

Koike S, Krapac I, Oliver H, Yannarell A, Chee-Sanford J, Aminov R & Mackie R (2007) Monitoring and source tracking of tetracycline resistance genes in lagoons and groundwater adjacent to swine production facilities over a 3-year period. *Applied and Environmental Microbiology* **73**: 4813-4823.

La Reau AJ, Meier-Kolthoff JP & Suen G (2016) Sequence-based analysis of the genus Ruminococcus resolves its phylogeny and reveals strong host association. *Microbial Genomics* **2**.

Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T & Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* **35**: 3100-3108.

Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G & Maraninchi M (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection* **18**: 1185-1193.

Lagkouvardos I, Pukall R, Abt B, Foesel BU, Meier-Kolthoff JP, Kumar N, Bresciani A, Martínez I, Just S & Ziegler C (2016) The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nature microbiology* **1**: 16131.

Lancaster H, Ready D, Mullany P, Spratt D, Bedi R & Wilson M (2003) Prevalence and identification of tetracycline-resistant oral bacteria in children not receiving antibiotic therapy. *FEMS microbiology letters* **228**: 99-104.

Lawson PA & Finegold SM (2015) Reclassification of Ruminococcus obeum as Blautia obeum comb. nov. *International journal of systematic and evolutionary microbiology* **65**: 789-793.

Lawson PA & Rainey FA (2016) Proposal to restrict the genus Clostridium Prazmowski to Clostridium butyricum and related species. *International journal of systematic and evolutionary microbiology* **66**: 1009-1016.

Lay C, Sutren M, Rochet V, Saunier K, Dore J & Rigottier-Gois L (2005) Design and validation of 16S rRNA probes to enumerate members of the Clostridium leptum subgroup in human faecal microbiota. *Environmental microbiology* **7**: 933-946.

Lee I, Kim YO, Park S-C & Chun J (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International journal of systematic and evolutionary microbiology* **66**: 1100-1103.

Levy SB & Marshall B (2004) Antibacterial resistance worldwide: causes, challenges and responses. *Nature medicine* **10**: S122-S129.

Lewis CM, Obregón-Tito A, Tito RY, Foster MW & Spicer PG (2012) The Human Microbiome Project: lessons from human genomics. *Trends in microbiology* **20**: 1-4.

Liu C, Finegold SM, Song Y & Lawson PA (2008) Reclassification of Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus and Ruminococcus schinkii as Blautia coccoides gen. nov., comb. nov., Blautia hansenii comb. nov., Blautia hydrogenotrophica comb. nov., Blautia luti comb. nov., Blautia producta comb. nov., Blautia schinkii comb. nov. and description of Blautia wexlerae sp. nov., isolated from human faeces. *International journal of systematic and evolutionary microbiology* **58**: 1896-1902.

Liu J-R, Tanner RS, Schumann P, Weiss N, McKenzie CA, Janssen PH, Seviour EM, Lawson PA, Allen TD & Seviour RJ (2002) Emended description of the genus Trichococcus, description of Trichococcus collinsii sp. nov., and reclassification of Lactosphaera pasteurii as Trichococcus pasteurii comb. nov. and of Ruminococcus palustris as Trichococcus palustris comb. nov. in the low-G+ C gram-positive bacteria. *International journal of systematic and evolutionary microbiology* **52**: 1113-1126.

Lopez-Siles M, Khan TM, Duncan SH, Harmsen HJ, Garcia-Gil LJ & Flint HJ (2012) Cultured representatives of two major phylogroups of human colonic Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Applied and environmental microbiology* **78**: 420-428.

Louis P & Flint HJ (2009) Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS microbiology letters* **294**: 1-8.

Louis P, Young P, Holtrop G & Flint HJ (2010) Diversity of human colonic butyrate‐producing bacteria revealed by analysis of the butyryl‐CoA: acetate CoA‐transferase gene. *Environmental microbiology* **12**: 304-314.

Manuzon MY, Hanna SE, Luo H, Yu Z, Harper WJ & Wang HH (2007) Quantitative assessment of the tetracycline resistance gene pool in cheese samples by real-time TaqMan PCR. *Applied and environmental microbiology* **73**: 1676-1677.

Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ & Bolam DN (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS biology* **9**: e1001221.

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G & Ejim L (2013) The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy* **57**: 3348-3357.

Meier-Kolthoff JP, Göker M, Spröer C & Klenk H-P (2013) When should a DDH experiment be mandatory in microbial taxonomy? *Archives of microbiology* **195**: 413-418.

Mesbah M, Premachandran U & Whitman WB (1989) Precise measurement of the G+ C content of deoxyribonucleic acid by high-performance liquid chromatography. *International Journal of Systematic and Evolutionary Microbiology* **39**: 159-167.

Mesbah NM, Whitman WB & Mesbah M (2011) 14- Determination of the G+ C Content of Prokaryotes. *Methods in Microbiology* **38**: 299-324.

Moore AM, Munck C, Sommer MO & Dantas G (2011) Functional metagenomic investigations of the human intestinal microbiota. *Frontiers in microbiology* **2**.

Moore W, Cato EP & Holderman LV (1972) Ruminococcus bromii sp. n. and emendation of the description of Ruminococcus Sijpestein. *International Journal of Systematic and Evolutionary Microbiology* **22**: 78-80.

Moore W, Johnson J & Holdeman LV (1976) Emendation of Bacteroidaceae and Butyrivibrio and descriptions of Desulfomonas gen. nov. and ten new species in the genera Desulfomonas, Butyrivibrio, Eubacterium, Clostridium, and Ruminococcus. *International Journal of Systematic and Evolutionary Microbiology* **26**: 238-252.

Mouton JW, Brown D, Apfalter P, Canton R, Giske C, Ivanova M, MacGowan A, Rodloff A, Soussy CJ & Steinbakk M (2012) The role of pharmacokinetics/pharmacodynamics in setting clinical MIC breakpoints: the EUCAST approach. *Clinical Microbiology and Infection* **18**.

Nesme J & Simonet P (2015) The soil resistome: a critical review on antibiotic resistance origins, ecology and dissemination potential in telluric bacteria. *Environmental microbiology* **17**: 913-930.

Nicholas KB, Nicholas Jr HB & Deerfield II DW (1997) embnet. news. *GeneDoc: Analysis and Visualization of Genetic Variation* **4**: 14.

Nicholas KB, Nicholas HBJ & Deerfield DWI (1997) GeneDoc, analysis and visualization of genetic variation. *EMBNEW News* **4**: 14.

O'Neal L, Obregón-Tito AJ, Tito RY, Ozga AT, Polo SI, Lewis CM & Lawson PA (2015) Clostridium amazonense sp. nov. an obliqately anaerobic bacterium isolated from a remote Amazonian community in Peru. *Anaerobe* **35**: 33-37.

Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, Xu ZZ, Van Treuren W, Knight R & Gaffney PM (2015) Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature communications* **6**.

Organization WH (2014) *Antimicrobial resistance: global report on surveillance*. World Health Organization.

Österblad M, Norrdahl K, Korpimäki E & Huovinen P (2001) Antibiotic resistance: How wild are wild mammals? *Nature* **409**: 37-38.

Paek J, Lee M-H, Kim B-C, Sang B-I, Paek WK, Jin T-E, Shin Y, Park I-S & Chang Y-H (2014) Clostridium vulturis sp. nov., isolated from the intestine of the cinereous vulture (Aegypius monachus). *Antonie van Leeuwenhoek* **106**: 577-583.

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357-358.

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357-358.

Patel NB, Tito RY, Obregón-Tito AJ, O'Neal L, Trujillo-Villaroel O, Marin-Reyes L, Troncoso-Corzo L, Guija-Poma E, Hamada M & Uchino Y (2015) Ezakiella peruensis gen. nov., sp. nov. isolated from human fecal sample from a coastal traditional community in Peru. *Anaerobe* **32**: 43-48.

Pawlowski AC, Wang W, Koteva K, Barton HA, McArthur AG & Wright GD (2016) A diverse intrinsic antibiotic resistome from a cave bacterium. *Nature communications* **7**: 13803.

Penders J, Stobberingh EE, Savelkoul PH & Wolffs PF (2013) The human microbiome as a reservoir of antimicrobial resistance. *Frontiers in microbiology* **4**.

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA & Deal C (2009) The NIH human microbiome project. *Genome research* **19**: 2317-2323.

Pryde SE, Duncan SH, Hold GL, Stewart CS & Flint HJ (2002) The microbiology of butyrate formation in the human colon. *FEMS microbiology letters* **217**: 133-139.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F & Yamada T (2010) A human gut microbial gene catalog established by metagenomic sequencing. *nature* **464**: 59.

Rainey FA (2009) Family VIII. Ruminococcaceae fam. nov. *Bergey's Manual of Systematic Bacteriology* **3**: 1016-1043.

Rainey FA & Janssen PH (1995) Phylogenetic analysis by 16S ribosomal DNA sequence comparison reveals two unrelated groups of species within the genus Ruminococcus. *FEMS microbiology letters* **129**: 69-73.

Ramirez-Farias C, Slezak K, Fuller Z, Duncan A, Holtrop G & Louis P (2008) Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii. *British Journal of Nutrition* **101**: 541-550.

Rasmussen S (2002) SEQtools, a software package for analysis of nucleotide and protein sequences. p.^pp.

Richardson A, Calder A, Stewart C & Smith A (1989) Simultaneous determination of volatile and non-volatile acidic fermentation products of anaerobes by capillary gas chromatography. *Letters in Applied Microbiology* **9**: 5-8.

Rivière A, Selak M, Lantin D, Leroy F & De Vuyst L (2016) Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Frontiers in microbiology* **7**.

Robert C & Bernalier-Donadille A (2003) The cellulolytic microflora of the human colon: evidence of microcrystalline cellulose-degrading bacteria in methane-excreting subjects. *FEMS microbiology ecology* **46**: 81-89.

Rolain J-M (2013) Food and human gut as reservoirs of transferable antibiotic resistance encoding genes. *Frontiers in microbiology* **4**.

Rosselló-Móra R & Amann R (2015) Past and future species definitions for Bacteria and Archaea. *Systematic and applied microbiology* **38**: 209-216.

Round JL & Mazmanian SK (2009) The gut microbiome shapes intestinal immune responses during health and disease. *Nature reviews Immunology* **9**: 313.

Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406-425.

Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.

Sasser M (1990) Identification of bacteria by gas chromatography of cellular fatty acids.

Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turroni S, Biagi E, Peano C & Severgnini M (2014) Gut microbiome of the Hadza hunter-gatherers. *Nature communications* **5**.

Schwiertz A, Hold GL, Duncan SH, Gruhl B, Collins MD, Lawson PA, Flint HJ & Blaut M (2002) Anaerostipes caccae gen. nov., sp. nov., a new saccharolytic, acetate-utilising, butyrate-producing bacterium from human faeces. *Systematic and applied microbiology* **25**: 46-51.

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

Segawa T, Takeuchi N, Rivera A, Yamada A, Yoshimura Y, Barcaza G, Shinbori K, Motoyama H, Kohshima S & Ushida K (2013) Distribution of antibiotic resistance genes in glacier environments. *Environmental microbiology reports* **5**: 127-134.

Sekirov I, Russell SL, Antunes LCM & Finlay BB (2010) Gut microbiota in health and disease. *Physiological reviews* **90**: 859-904.

Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, Jian R & Dore J (2003) Alterations of the dominant faecal bacterial groups in patients with Crohn9s disease of the colon. *Gut* **52**: 237-242.

Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux J-J, Blugeon S, Bridonneau C, Furet J-P & Corthier G (2008) Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences* **105**: 16731-16736.

Sommer MO, Dantas G & Church GM (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *science* **325**: 1128-1131.

Spellberg B, Bartlett JG & Gilbert DN (2013) The future of antibiotics and resistance. *New England Journal of Medicine* **368**: 299-302.

Stackebrandt E (2014) The Family Lachnospiraceae. *The Prokaryotes,*p.^pp. 197-201. Springer.

Stackebrandt E & Goebel B (1994) 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* **44**: 846-849.

Stackebrandt E & Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today Nov* **6**: 152-155.

Suau A, Bonnet R, Sutren M, Godon J-J, Gibson GR, Collins MD & Doré J (1999) Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Applied and environmental microbiology* **65**: 4799-4807.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**: 2731-2739.

Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP, Ugarte E, Muñoz-Tamayo R, Paslier DL & Nalin R (2009) Towards the human intestinal microbiota phylogenetic core. *Environmental microbiology* **11**: 2574-2584.

Thaller MC, Migliore L, Marquez C, Tapia W, Cedeño V, Rossolini GM & Gentile G (2010) Tracking acquired antibiotic resistance in commensal bacteria of Galapagos land iguanas: no man, no resistance. *PloS one* **5**: e8989.

Tindall BJ, Rosselló-Mora R, Busse H-J, Ludwig W & Kämpfer P (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology* **60**: 249-266.

Tito RY, Knights D, Metcalf J, Obregon-Tito AJ, Cleeland L, Najar F, Roe B, Reinhard K, Sobolik K & Belknap S (2012) Insights from characterizing extinct human gut microbiomes. *PloS one* **7**: e51146.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R & Gordon JI (2007) The human microbiome project. *Nature* **449**: 804-810.

Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H & Whitman W (2011) *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*. Springer Science & Business Media.

Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, Brown D, Stares MD, Scott P & Bergerat A (2011) Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME journal* **5**: 220.

Wang HH, Manuzon M, Lehman M, Wan K, Luo H, Wittum TE, Yousef A & Bakaletz LO (2006) Food commensal microbes as a potentially important avenue in transmitting antibiotic resistance genes. *FEMS microbiology letters* **254**: 226-231.

Wang W, Chen L, Zhou R, Wang X, Song L, Huang S, Wang G & Xia B (2014) Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *Journal of clinical microbiology* **52**: 398-406.

Wegmann U, Louis P, Goesmann A, Henrissat B, Duncan SH & Flint HJ (2014) Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('Ruminococcus bicirculans') reveals two chromosomes and a selective capacity to utilize plant glucans. *Environmental Microbiology* **16**: 2879-2890.

Wiegel J, Tanner R & Rainey FA (2006) An introduction to the family Clostridiaceae. *The prokaryotes,*p.^pp. 654-678. Springer.

Wylie KM, Truty RM, Sharpton TJ, Mihindukulasuriya KA, Zhou Y, Gao H, Sodergren E, Weinstock GM & Pollard KS (2012) Novel bacterial taxa in the human microbiome. *PloS one* **7**: e35294.

Xu J & Gordon JI (2003) Honor thy symbionts. *Proceedings of the National Academy of Sciences* **100**: 10452-10459.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R & Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**.

Yoon S-H, Ha S-m, Lim J, Kwon S & Chun J (2017) A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 1-6.

Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H & Chun J (2016) Introducing EzBioCloud: A taxonomically united database of 16S rRNA and whole genome assemblies. *International Journal of Systematic and Evolutionary Microbiology*.

Yutin N & Galperin MY (2013) A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental microbiology* **15**: 2631-2641.

Zhou Y, Mihindukulasuriya KA, Gao H, La Rosa PS, Wylie KM, Martin JC, Kota K, Shannon WD, Mitreva M & Sodergren E (2014) Exploration of bacterial community classes in major human habitats. *Genome Biol* **15**: R66.

Zoetendal E, Rajilić-Stojanović M & De Vos W (2008) High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* **57**: 1605-1615.

Zuo G, Xu Z & Hao B (2013) Shigella strains are not clones of Escherichia coli but sister species in the genus Escherichia. *Genomics, proteomics & bioinformatics* **11**: 61-65.

Appendix

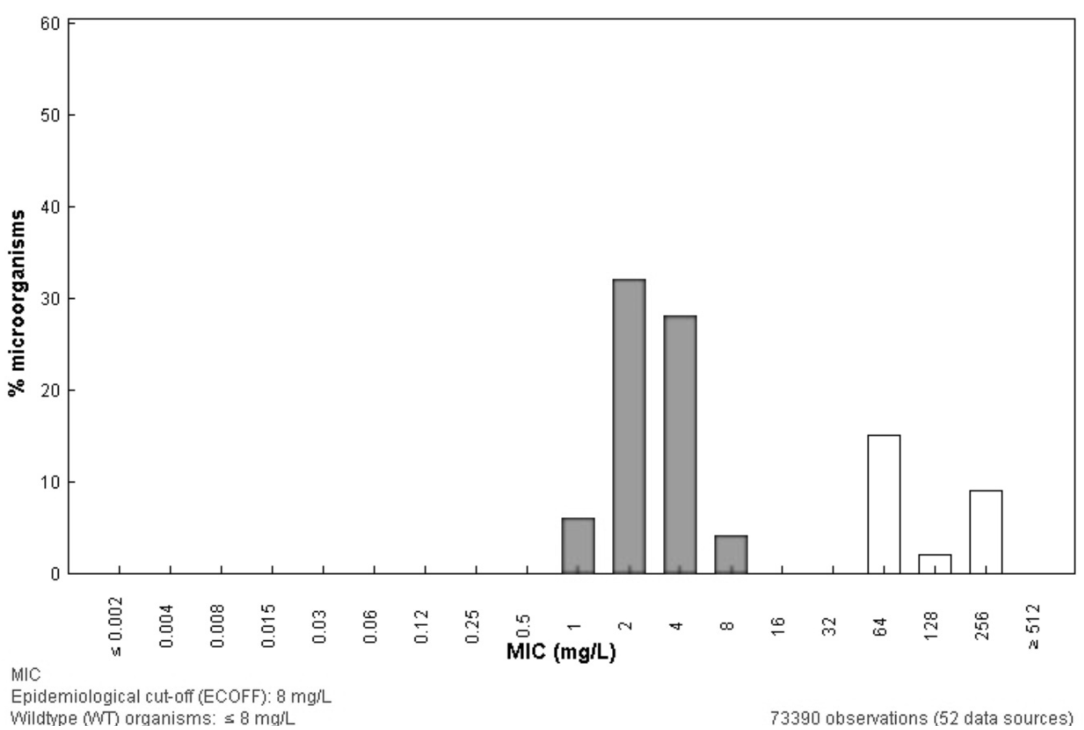

Figure A1: EUCAST example histogram

VatI

Download Sequences

Accession	ARO:3003987
Synonym(s)	
Definition	Streptogramin A acetyltransferase found in Paenibacillus sp. LC231, isolated from Lechuguilla Cave, NM, USA. Confers resistance to streptogramin A antibiotics. Described by Pawlowski et al. 2016.
Classification	+ process or component of antibiotic biology or chemistry + antibiotic molecule + mechanism of antibiotic resistance + determinant of antibiotic resistance + streptogramin antibiotic [Drug Class] + antibiotic inactivation enzyme + determinant of streptogramin resistance + streptogramin A antibiotic + antibiotic inactivation [Resistance Mechanism] + acylation of antibiotic conferring resistance + madumycin II [Antibiotic] + streptogramin inactivation enzyme + dalfopristin [Antibiotic] + griseoviridin [Antibiotic] + pristinamycin IIA [Antibiotic]
Parent Term(s)	+ streptogramin vat acetyltransferase (is_a) [AMR Gene Family]
Sub-Term(s)	
Publications	Pawlowski AC, et al. 2016. Nat Commun 7:13803 A diverse intrinsic antibiotic resistome from a cave bacterium.. (PMID 27929110)

Figure A2: CARD gene information example