

November 2016

Expansion of and reclassification within the family Lachnospiraceae

Kelly N. Haas
University of Massachusetts - Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2



Part of the Biodiversity Commons, Bioinformatics Commons, Environmental Microbiology and Microbial Ecology Commons, Evolution Commons, Genomics Commons, Microbial Physiology Commons, Organismal Biological Physiology Commons, Other Ecology and Evolutionary Biology Commons, Other Microbiology Commons, and the Systems Biology Commons

Recommended Citation

Haas, Kelly N., "Expansion of and reclassification within the family Lachnospiraceae" (2016). *Doctoral Dissertations*. 843.

https://scholarworks.umass.edu/dissertations_2/843

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

EXPANSION OF AND RECLASSIFICATION WITHIN THE FAMILY
LACHNOSPIRACEAE

A dissertation presented

by

KELLY NICOLE HAAS

Submitted to the Graduate School of the University of Massachusetts Amherst in partial
fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2016

Microbiology Department

© Copyright by Kelly Nicole Haas 2016

All Rights Reserved

EXPANSION OF AND RECLASSIFICATION WITHIN THE FAMILY

LACHNOSPIRACEAE

A Dissertation Presented

By

KELLY NICOLE HAAS

Approved as to style and content by:

Jeffrey L. Blanchard, Chair

Kristen M. DeAngelis, Member

Susan Leschine, Member

Li-Jun Ma, Member

Steven Sandler, Department Head
Microbiology

ACKNOWLEDGEMENTS

I would like to thank my advisor, Jeffrey Blanchard for his guidance and support as well as the freedom he has allowed me to pursue the science in and outside this thesis. My labmates, past and present, have offered moral support, advice, and friendship when I have needed it. This network includes Supratim Mukherjee, Amy Biddle, Farah Abdul-Rahman, William Rodriguez, and Lauren Alteio, Jesús Alvelo and Erin Amato in Sue Leschine's group, and Andrew Billings, Grace Pold, and Gina Chaput in Kristen DeAngelis's group. Without Amy's microcosms, I would certainly have a much different body of work. My thesis committee has been a pleasure to interact with whether in their expertise as committee members or as friends: Kristen DeAngelis, Sue Leschine, and Li-Jun Ma.

I could not have accomplished what I have without my parents, Bradley and Alina Haas, my siblings, Kristen and Kennen Haas, and Ben Cherniawski. They have offered me so much support throughout this journey, I cannot even find the words to express my gratitude.

ABSTRACT

EXPANSION OF AND RECLASSIFICATION WITHIN THE FAMILY

LACHNOSPIRACEAE

SEPTEMBER 2016

KELLY NICOLE HAAS, B.A., MIAMI UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Prof. Jeffrey L. Blanchard

The *Lachnospiraceae* is a recently approved taxonomic family, though its first representatives were discovered in the early 1900s. It is comprised of anaerobic saccharolytic bacteria, isolated from digestors, soil, the rumen, and the gut, and has become well-known for its many butyrate producers. Many of the early members were dubbed *Clostridium* spp. due to their ability to form spores. But in the 1990s, sequencing turned the taxonomy on its head: this genus crossed several families of distantly related bacteria.

Here, we attempt to rectify a large portion of this nomenclatural issue within the family *Lachnospiraceae*. Approximately one third of the family consists of *Clostridium* species who are still taxonomically recognized as *Clostridiaceae*, despite being phylogenetically distinct. The continued inclusion of these taxa in the *Clostridium* genus leads to confusion over the true phylogenetic breadth of this family, which is problematic for those who are not *Clostridial* taxonomy experts.

We begin with an in-depth analysis of *Clostridium methoxybenzovorans*, culminating in the assertion that it is a heterotype of *Clostridium indolis*, followed by reclassification of the broader group in which this organism resides. This clade comprises about one-eighth of the described species in the *Lachnospiraceae*. We propose two novel genera, *Lacriformis* and *Enterocloster*, to reclassify this clade, this includes reclassification of *Clostridium sphenoides*, *Clostridium indolis*, *Clostridium saccharolyticum*, *Clostridium celerecrescens*, *Clostridium xylanolyticum*, *Clostridium algidixylanolyticum*, *Clostridium aerotolerans*, *Clostridium amygdalinum*, and *Desulfotomaculum guttoideum* as *Lacriformis sphenoides*, comb. nov., *Lacriformis indolis*, comb. nov., *Lacriformis saccharolyticum*, comb. nov., *Lacriformis celerecrescens*, comb. nov., *Lacriformis xylanolyticum*, comb. nov., *Lacriformis algidixylanolyticum*, comb. nov., *Lacriformis aerotolerans*, comb. nov., *Lacriformis amygdalinum*, comb. nov., and *Lacriformis guttoideum*, comb. nov. A second genus, *Enterocloster*, includes *Clostridium clostridioforme*, *Clostridium bolteae*, *Clostridium citroniae*, *Clostridium lavalense*, *Clostridium aldenense*, and *Clostridium asparagiforme*, reclassified as *Enterocloster clostridioforme*, comb. nov., *Enterocloster bolteae*, comb. nov., *Enterocloster citroniae*, comb. nov., *Enterocloster lavalense*, comb. nov., *Enterocloster aldenense*, comb. nov., and *Enterocloster asparagiforme*, comb. nov.

Next we propose reclassification of another group of *Clostridium* species, composed of cellulolytics and hemicellulolytics. To properly house these four taxa, we propose two novel genera. One genus, *Cellulospecium*, gen. nov., is to include *Cellulospecium herbivorans*, comb. nov., *Cellulospecium populeti*, comb. nov., and *Cellulospecium polysaccharolyticum*, comb. nov., formerly *Clostridium herbivorans*, *Clostridium populeti*, and *Clostridium polysaccharolyticum*. Another genus, *Leschinia*, houses *Leschinia phytofermentans*, comb. nov.,

formerly *Clostridium phytofermentans*. We also describe a novel species, *Anaerocolumnna spermata*.

We then propose the classification of a novel genus, *Kineothrix*, with the novel species *Kineothrix alysoides*. The organism was isolated from an anaerobic switchgrass microcosm developed using forest soil. It is a highly motile anaerobic spore-former, with broad saccharolytic capabilities. It produces butyrate as a major fermentation product and is able to fix nitrogen.

We end with the microcosm from which *Anaerocolumnna spermata*, *Kineothrix alysoides*, and others novel taxa were isolated. Having isolated the most abundant members of this switchgrass-degrading community, we were able to approach the question of the rare biosphere through synthetic ecology. Rather than culling out diversity with toxins, we were able to remove diversity by simply leaving it out of a reconstituted community. These rare members had a significant effect on plant degradation, though a nutrient-limiting medium had a more pronounced effect. This approach is especially useful for comparative and perturbation studies.

While much of the *Lachnospiraceae* remains misnamed as *Clostridium*, *Ruminococcus*, or *Eubacterium* spp., this takes a significant step towards rectification, while also introducing new and exciting microbes to our lexicon. *Anaerocolumnna spermata* has already gained attention as a potential serotonin mediator in the gut. Hopefully the increase in accessibility that accompanies proper taxonomy will lead to more interest in and more research on these important bacteria.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xiv
CHAPTER	
1. A HISTORY OF THE LACHNOSPIRACEAE.....	1
2. NON-CONTIGUOUS GENOME SEQUENCE OF THE LIGNAN TRANSFORMING CLOSTRIDIUM METHOXYBENZOVORANS.....	11
2.1 Abstract.....	11
2.2 Introduction.....	12
2.3 Organism information.....	14
2.4 Genome sequencing information.....	18
2.5 Genome properties.....	21
2.6 Insights from the genome sequence.....	23
2.7 Conclusions.....	36
2.8 Taxonomic and nomenclatural proposals.....	37

3. RECLASSIFICATION OF THE CLOSTRIDIUM CLOSTRIDIOFORME AND
CLOSTRIDIUM SPHENOIDES GROUPS AS ENTEROCLOSTER, GEN. NOV. AND
LACRIFORMIS, GEN. NOV., INCLUDING RECLASSIFICATION OF FIFTEEN

TAXA.....	42
3.1 Abstract.....	42
3.2 Introduction.....	43
3.3 Results and Discussion.....	45
3.4 Description of <i>Enterocloster</i> , gen. nov.....	55
3.5 Description of <i>Enterocloster clostridioforme</i> , comb. nov.....	56
3.6 Description of <i>Enterocloster aldenense</i> , comb. nov.....	56
3.7 Description of <i>Enterocloster asparagiforme</i> , comb. nov.....	57
3.8 Description of <i>Enterocloster bolteae</i> , comb. nov.....	57
3.9 Description of <i>Enterocloster citroniae</i> , comb. nov.....	57
3.10 Description of <i>Enterocloster lavalense</i> , comb. nov.....	58
3.11 Description of <i>Lacriformis</i> , gen. nov.....	58
3.12 Description of <i>Lacriformis sphenoides</i> , comb. nov.....	58
3.13 Description of <i>Lacriformis aerotolerans</i> , comb. nov.....	59
3.14 Description of <i>Lacriformis algidixylanolyticum</i> , comb. nov.....	59
3.15 Description of <i>Lacriformis amygdalinum</i> , comb. nov.....	60
3.16 Description of <i>Lacriformis celerecrescens</i> , comb. nov.....	60
3.17 Description of <i>Lacriformis guttoideum</i> , comb. nov.....	60
3.18 Description of <i>Lacriformis indolis</i> , comb. nov.....	61
3.19 Description of <i>Lacriformis saccharolyticum</i> , comb. nov.....	61

3.20 Description of <i>Lacriformis xylanolyticum</i> , comb.nov.....	62
4. RECLASSIFICATION OF CLOSTRIDIUM POLYSACCHAROLYTICUM, CLOSTRIDIUM POPULETI, CLOSTRIDIUM HERBIVORANS, AND CLOSTRIDIUM PHYTOFERMENTANS AS CELLULOSPECIUM POLYSACCHAROLYTICUM, GEN. NOV., COMB. NOV., CELLULOSPECIUM POPULETI, COMB. NOV., CELLULOSPECIUM HERBIVORANS, COMB. NOV., AND LESCHINIA PHYTOFERMENTANS, GEN. NOV., COMB. NOV.....	63
4.1 Abstract.....	63
4.2 Introduction.....	64
4.3 Results.....	65
4.4 Description of <i>Cellulospecium</i> gen. nov.....	73
4.5 Description of <i>Cellulospecium herbivorans</i> , comb. nov.....	73
4.6 Description of <i>Cellulospecium populeti</i> , comb. nov.....	73
4.7 Description of <i>Cellulospecium polysaccharolyticum</i> , comb. nov.....	74
4.8 Description of <i>Leschinia</i> , gen. nov.....	74
4.9 Description of <i>Leschinia phytofermentans</i> , comb. nov.....	74
4.10 Description of <i>Anaerocolumna spermata</i> , sp. nov.....	75

5. KINEOTHRIX ALYSOIDES, GEN. NOV., SP. NOV., A SACCHAROLYTIC, BUTYRATE-PRODUCING NITROGEN FIXER WITHIN THE FAMILY LACHNOSPIRACEAE.....	77
5.1 Abstract.....	77
5.2 Introduction.....	78
5.3 Methods.....	78
5.4 Results.....	84
5.5 Description of <i>Kineothrix</i> , gen. nov.....	97
5.6 Description of <i>Kineothrix alysoides</i> , gen. nov. sp. nov.....	97
5.7 Discussion.....	98
6. PROBING THE IMPACT OF THE RARE BIOSPHERE THROUGH MICROCOSM DEVELOPMENT AND SYNTHETIC ECOLOGY.....	101
6.1 Abstract.....	101
6.2 Introduction.....	102
6.3 Methods.....	104
6.4 Results.....	108
6.5 Discussion.....	123
7. SUMMARY AND CONCLUSION.....	126
APPENDIX: SUPPLEMENTAL FIGURES.....	131

REFERENCES.....137

LIST OF TABLES

Table	Page
1. Classification and general features of <i>Clostridium methoxybenzovorans</i>	17
2. Project information.....	19
3. Genome statistics.....	22
4. Number of genes associated with the 25 general COG functional categories.....	23
5. Sialic acid utilization genes in <i>C. methoxybenzovorans</i>	29
6. Average Nucleotide Identity between <i>C. indolis</i> and <i>C. methoxybenzovorans</i>	39
7. GGDC values for <i>C. indolis</i> and <i>C. methoxybenzovorans</i>	40
8. ANIb for all sequenced type genomes in the clade spanning the <i>Clostridium sphenoides</i> and <i>Clostridium clostridioforme</i> groups, including the outliers <i>Hungatella hathewayi</i> and <i>Clostridium symbiosum</i>	52
9. Average nucleotide identities (ANIb) of taxa in the cluster.....	68
10. Substrate utilization test results for KNHs209 ^T	90
11. Sequencing project information.....	91
12. Nucleotide content and gene count levels of the genome.....	92
13. Number of genes associated with the 25 general COG functional categories.....	92
14. Comparison of KNHs209 ^T to closest relatives based on 16S rRNA gene sequence.....	96
15. Microcosm development sequencing metrics.....	108
16. Chimeras lead to overestimations of diversity.....	116

LIST OF FIGURES

Figure	Page
1. The phylogeny of the family <i>Lachnospiraceae</i>	5
2. O-demethylation of vanillic acid to protocatechuic acid.....	13
3. Photographs of <i>C. methoxybenzovorans</i>	14
4. Phylogeny of <i>C. methoxybenzovorans</i> and related species in the family <i>Lachnospiraceae</i> ...	15
5. Syntenic map of the genes responsible for o-demethylation in <i>Moorella thermoacetica</i> and their counterparts in the <i>C. methoxybenzovorans</i> genome.....	25
6. <i>C. indolis</i> and <i>C. methoxybenzovorans</i> growth on vanillic acid.....	26
7. <i>C. methoxybenzovorans</i> growth on vanillic acid with and without carbon dioxide.....	27
8. Phylogenetic tree of all copies of the 16S rRNA gene in <i>C. indolis</i> and <i>C. methoxybenzovorans</i>	38
9. Maximum likelihood method of <i>Lachnospiraceae</i> using the 16S rRNA gene.....	46
10. Neighbor-joining phylogeny using the 16S rRNA gene.....	47
11. <i>rpoβ</i> gene based phylogeny.....	49
12. <i>Lachnospiraceae</i> phylogeny based on concatenated alignment of 21 single-copy marker genes.....	50
13. 16S rRNA gene based phylogeny.....	67
14. <i>rpoβ</i> gene based phylogeny.....	70
15. Concatenated gene phylogeny using 22 conserved genes.....	72
16. Morphology of KNHs209.....	85
17. COG-based gene neighborhood maps comparing the nitrogenase loci.....	87
18. Phylogenetic tree showing the placement of KNHs209 within the <i>Lachnospiraceae</i>	94

19. Phylum-level diversity of the microcosm over time.....	109
20. Most abundant OTUs over time.....	111
21. OTUs shared across all replicates at transfer 62 (T62).....	112
22. Switchgrass degradation by medium and community type.....	120
23. Fermentation product analysis of the whole communities and rebuilt communities by media type, rich (A) or minimal (B).....	122

CHAPTER 1

A HISTORY OF THE LACHNOSPIRACEAE

In 1908, Henry Tissier published the description of a novel taxon isolated from toddler feces (Tissier, 1908). It was a rod-shaped anaerobe, presumed to be non-sporeforming, although Tissier describes the cells as sometimes having *paloton de jaridinier*, a “gardener’s ball of twine”, around their middles (Eggerth, 1935; Tissier, 1908), much like a developing endospore. At this early age of microbiology, these characteristics warranted classification as a *Bacillus*. Some decades later, Arnold Eggerth catalogued and reclassified several anaerobic bacilli from human feces, including those Tissier had isolated and described. Owing to the organism’s Gram positive staining, Eggerth reclassified it as a *Bacteroides*. In another decade it was once more reclassified as *Pseudobacterium*, an even less strict definition than *Bacillus* or *Bacteroides* in that it referred to any rod-shaped microbe

Prior to Eggerth’s work, it was believed that adult feces contained mainly aerobic organisms, and that these strict anaerobes were peculiar to a child’s digestive tract. Unfortunately, this organism, now called *Eubacterium ventriosum*, went largely ignored for almost a century – perhaps due to difficulty of maintenance *in vitro*. But in the last decade, it is being written about again, this time in the context of the human gut microbiome. Though not a dominant member, it does make up about 0.04% of the fecal flora (Eckburg, 2005), perhaps an underestimation due to a bias against *Clostridia* in

community sequencing methods, likely due to cell wall structure (Bahl et al., 2012; Cuív et al., 2011; Maukonen et al., 2012).

This story highlights the complicated taxonomic history of not only this organism, but many that were discovered during the first golden age of microbiology: when bacteriologists were sampling feces, and rumens, and dental plaque simply to discover what these microscopic cosmoses had been hiding. Many interesting but almost indistinguishable microbes were found and studied. Similar morphologies led to more detailed phenotypic and biochemical tests being employed to differentiate strains, something that has much been lost today due to our reliance on 16S rRNA sequencing.

Other early work on the human gut microbiome through the 1900s to the 1950s led to the discovery of several more non-proteolytic strict anaerobes, and dozens more when the 1960s and 1970s gave way to a boom in rumen microbiology. Round organisms were generally classified as *Ruminococci*, while rods were either *Clostridia* or *Eubacteria*, depending on the ability to form spores. Proteolysis, sulfate reduction, or other unique traits were sometimes the basis for development of a novel genus (Bryant and Small, 1956a, 1956b; Campbell and Postgate, 1965; Wachsman and Barker, 1954).

Several years after these organisms were discovered and classified, Frederick Sanger developed a technique that would change microbial classification forever (Sanger et al., 1977; Sanger and Coulson, 1975). Sanger sequencing enabled us to glimpse the code in a microbe's DNA. At about the same time, Carl Woese was developing the small ribosomal subunit as a molecular marker through oligonucleotide fingerprinting experiments (Woese and Fox, 1977). About a decade later, Carl Woese's work entitled "Bacterial Evolution" propelled the wide-spread use of 16S rRNA gene sequencing as a

chronometer of evolution and relatedness (Woese, 1987), spurring a sequencing frenzy that can be likened to whole genome sequencing today. This genomic fragment gave scientists phylogenetic resolution that they had never had access to before. But of course, with this groundbreaking approach came huge disruptions: not all anaerobic sporeformers were cousins.

By the mid-1990s, researchers had realized that there were serious problems with the taxonomy and phylogeny of the genus *Clostridium* (Collins et al., 1994; Stackebrandt et al., 1999). It spanned a huge phylogenetic breadth and was interleaved with various, phenotypically distinct genera. As originally defined, to be included an organism had to be a Gram positive anaerobe which formed spores and did not reduce sulfate. Today we realize the diversity present on the planet Earth, and that these four conditions are easily met by a great variety of bacteria spanning distant corners of the bacterial tree of life.

In order-level phylogenies, a natural tribe was formed by these mainly saccharolytic gut commensals, distinctive of other *Clostridium* spp. Collins *et. al* dubbed this clade *Clostridium* cluster XIVa (Collins et al., 1994). The relatively closely related cluster XIVb still was deep-branching and had few members. However, a contemporary restructuring of the taxonomy led to inclusion of both of these clusters into a single family, the *Lachnospiraceae*, though it is still arguable that these two groups be considered separate Families. The first usage that we can find of “*Lachnospiraceae*” is from 2003 (Janssen and Hugenholtz, 2003), with continued usage for the next several years (Carlier et al., 2007, 2004; Cotta and Forster, 2006) leading up to the official taxonomic acceptance of the family *Lachnospiraceae* (Rainey, 2010). Though the family name is relatively new, the first representative of the family was described by Tissier in

1908. Exploration of the breadth of this group of organisms is reaching its peak now, due to interest in the human gut microbiome. Approximately half of the family consists of taxa which were isolated and named since the year 2000.

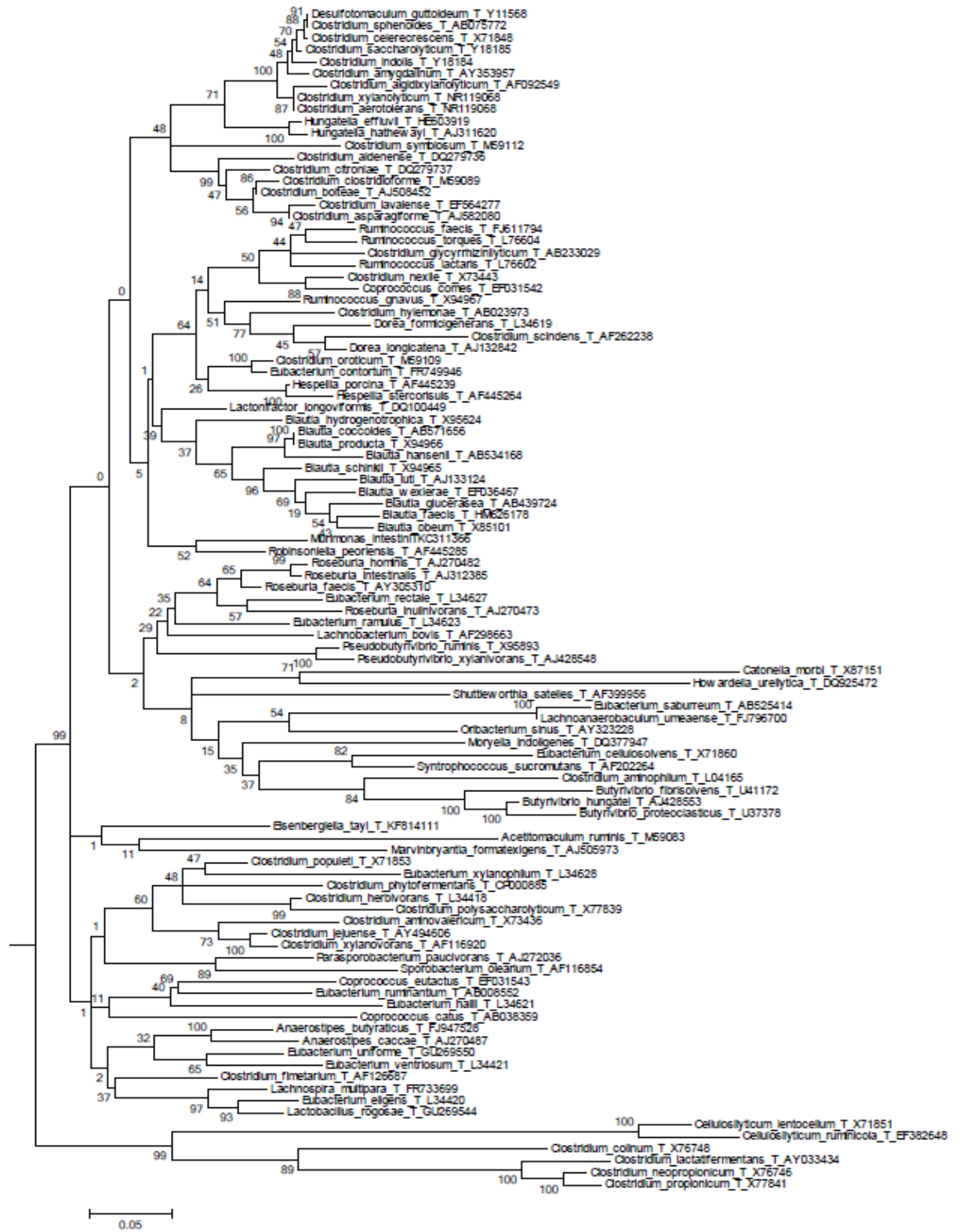


Figure 1. The Family *Lachnospiraceae*. This tree, based on the 16S rRNA gene shows the phylogenetic breadth of the family, including misclassified taxa.

The *Lachnospiraceae* consists of 26 different genera, including many *Clostridia* (Figure 1). They are some of the oldest members of the tribe and span the breadth of the phylogeny. *Clostridium sphenoides* is the second-oldest taxon in the family, having been described by Douglas in 1917. The organism was originally isolated from a gangrenous wound and later in human feces, soil, and marine sediment. Several morphologically similar taxa have been discovered that are close relatives, including *Clostridium indolis* (Gogotova and Vainshtein, 1983; Gylswyk and Toorn, 1987; McClung et al., 1957; Palop et al., 1989; Parshina et al., 2003; Rogers and Baecker, 1991). They are distinctive for their wedge shape during spore-formation (*spheno* Gr. n. “wedge”). *C. indolis* was named for its ability to produce indole, a product of tryptophan metabolism, now known to be a signaling molecule not only between microbes but between microbes and plants (Koul et al., 2015; Lee and Lee, 2010). This characteristic is actually quite uncommon among other taxa in the family. Though mostly unable to reduce sulfate, this *C. sphenoides* group seems to be some of the only H₂S producers in the family, reducing sulfite, cysteine, or thiosulfate instead (Broda et al., 2000; Douglas et al., 1917; Gogotova and Vainshtein, 1983; McClung et al., 1957; Parshina et al., 2003; our own work). This characteristic and perhaps confusion between reduction of the cysteine in the growth medium and reduction of added sulfate led to inclusion of *Desulfotomaculum guttoideum* in a sulfate-reducing genus (Gogotova and Vainshtein, 1983; Stackebrandt et al., 1997).

Clostridium oroticum, originally *Zymobacterium oroticum*, was described in 1954 and named for its ability to ferment orotic acid (Cato et al., 1968; Wachsman and Barker, 1954). Originally classified as a novel genus within the *Lactobacillaceae*, the organism was transferred to the *Clostridium* genus when heat-resistant spores were discovered

(Cato et al., 1968). It was isolated from bay mud enriched on orotic acid, and was also able both utilize and synthesize the compound. At the time, orotic acid was of interest in liver metabolism, uracil synthesis, and as a growth factor for some microbes. We now know that it is an intermediate in pyrimidine synthesis, explaining its activity as a growth factor. There is an extensive body of literature regarding its bioactivity in humans as well. It has been shown to be cardioprotective, specifically in post-ischemic hearts, increasing metabolism of fatty acids and glucose. Whether this effect is unique to orotic acid or is due to its pyrimidine end-products has not been shown.

Clostridium aminovalericum was isolated from sewage sludge and described in 1960 by Hardman and Stadtman; named for its ability to utilize aminovaleric acid as its sole carbon and nitrogen source (Hardman and Stadtman, 1960). Though the name conjures to mind a proteolytic organism, *C. aminovalericum* seems to be able only to use its eponymous amino acid for growth and cannot grow on whole protein. It can use a wide variety of mono-, di-, and polysaccharides, as well as sugar alcohols, and a later publication also revealed its ability to utilize cellulose (Jeong et al., 2004), a characteristic it shares with several relatives.

Clostridium phytofermentans was described by Warnick *et al.* in 2002. It is named for its ability to ferment plant material, including both the complex hemicellulolytic and the crystalline cellulolytic components. Its ability to do this while producing mainly ethanol made it an attractive catalyst for next generation biofuel production at a period when oil prices were at an all-time high. A company, Qteros, was developed around scaling up and commercializing the technology. Though sharing the same genus name, it

is phylogenetically quite distant from its closest cultured representative, *Clostridium populeti*, which falls into a small clade with *C. aminovalericum*.

In what used to be referred to as cluster XIVb is a very distant group of organisms consisting of the genera *Cellulosilyticum* and *Clostridium*. The first representative of this group, *Clostridium propionicum*, was characterized in 1946 by Cardon and Barker. This taxon was originally isolated from black San Francisco Bay mud enriched on alanine medium, and as the name denotes, produces mainly propionate from fermentation. A close relative was described in 1992 and due to a lack of phenotypic differences was aptly named *Clostridium neopropionicum*. They appear to be incapable of utilizing carbohydrates, but do ferment amino acids and short-chain fatty acids and alcohols.

Oddly enough, despite a distance of up to 15% in their 16S rRNA gene sequences, these organisms are all classified within the same taxonomic genus. While this discrepancy between taxonomy and phylogeny was well-acknowledged, it was a daunting task to reclassify the hundreds of misplaced *Clostridium* spp. In many novel species descriptions thereafter, authors addressed this issue but still went on to use the name *Clostridium* outside of the genus *sensu stricto*. In the last fifteen years, however, several researchers have set out to rectify the situation even if only one taxa at a time. In 2003, Taras et al. reclassified *Eubacterium formicigenerans*, another multi-family-spanning genus, as *Dorea formicigenerans*. In 2008, there were eight revisions within the family, in large part thanks to Liu et. al's creation of the genus *Blautia* and six coincident species revisions. In 2010, Cai et. al proposed the genus *Cellulosilyticum*, with the novel species *Cellulosilyticum ruminicola* and the reclassified *Cellulosilyticum lentocellum* (formerly *Clostridium lentocellum*). Another *Eubacterium* species gave way to the new genus

Lachnoanaerobaculum in 2013, including the novel species *Lachnoanaerobaculum orale* and the reclassified *Lachnoanaerobaculum saburreum*. However, approximately half of the family is still misclassified into polyphyletic genera including *Eubacterium*, *Clostridium*, and *Ruminococcus*, all which have their own eponymous *sensu stricto* Families.

Since these organisms are all taxonomically *Clostridium* spp., and therefore *Clostridiaceae*, taxonomic databases leave out all *Clostridium* spp. from the *Lachnospiraceae* (LSPN, NamesforLife, Taxinomicon, and Rainey's description of the *Lachnospiraceae*, fam. nov. in 2010). Researchers who do not know or understand the dichotomy between taxonomy and phylogeny use these sources as phylogenetic, rather than taxonomic, guides, leading to a continual development of misleading phylogenetic trees for new taxa, thereby perpetuating and even propagating the cycle. Even researchers who study *Clostridia* accidentally overlooked those in the *Lachnospiraceae* and *Ruminococcaceae* (along with other Families) in NCBI databases when the NCBI unassumingly changed the taxonomy of these genera to *Lachnoclostridium* and *Ruminoclostridium*. A serious oversight in work analyzing or comparing plant-degrading mechanisms. Stackebrandt (Stackebrandt, 2014) left out the *Clostridium* spp. from the family when discussing the capabilities of *Lachnospiraceae*, even though they make up approximately one third of the described species. Only including the taxonomic family *Lachnospiraceae* in the analysis led to the conclusion that spore-formation in this group is a rarity, though it is a trait shared by at least 50% of the phylogenetic family, likely even higher considering that many researchers did not test for it.

Official taxonomic revision for these organisms is integral to their visibility and accessibility, which is especially important given their phenotypic breadth and importance in industry and human health. This work attempts to disentangle at least the majority of this web, reclassifying several genera within the *Lachnospiraceae* using phenotypic and genomic data to inform proper placement. These genera comprise approximately two-thirds of the misplaced *Clostridium* spp. within the family, with the others largely dispersed as single representatives nested between other genera. The work is certainly not finished, but the reclassification of these singletons can be done with relative ease in the future. But of course the most difficult part of taxonomic revision is in the subsequent usage by other scientists.

CHAPTER 2

NON-CONTIGUOUS GENOME SEQUENCE OF THE LIGNAN TRANSFORMING CLOSTRIDIUM METHOXYBENZOVORANS

2.1 Abstract

Clostridium methoxybenzovorans is saccharolytic spore-former within the *Lachnospiraceae* family. It notably is capable of *O*-demethylation, enabling it to partially degrade methoxylated aromatics of which lignin and lignans are composed. These compounds are important in many biological and industrial processes, including digestion and bioprocessing, as they make up a large portion of plant biomass. Lignans are of recent interest in human health for their cardiovascular and cancer protective roles, particularly after their transformation to mammalian estrogens by gut microbes. To identify the genes related to lignin and lignan metabolism in *C. methoxybenzovorans*, we have sequenced and analyzed the genome. A large (2439 bp) gene is a fusion of the *mtvA* and *mtvC* genes in the *O*-demethylase system of *Moorella thermoacetica*. *C. methoxybenzovorans* has not been observed to be motile, but the genome includes genes for flagellar machinery as well as twenty uncharacterized chemoreceptors, suggesting wide substrate range. Close relative *Clostridium indolis* has this same chemotactic profile and *O*-demethylation genes as *C. methoxybenzovorans*. *O*-demethylation was observed *in vitro* by both *C. methoxybenzovorans* and *C. indolis*. *C. methoxybenzovorans* is only differentiated from *Clostridium indolis* by 600 genes (of 6,592 protein coding genes), half of which appear to be phage-related. Based on the average nucleotide identity (ANI)

(>99%), 16S rRNA sequence identity (>99%), and physiological characterization, we propose that *C. methoxybenzovorans* should be reclassified as *Clostridium indolis* strain methoxybenzovorans.

2.2 Introduction

C. methoxybenzovorans was named for its ability to *O*-demethylate phenyl methyl ethers (Fig. 2), also called methoxylated aromatics (Mechichi et al., 2005, 1999). These compounds are abundant in nature in the form of monolignols: precursors and breakdown products of lignin and lignans, two related but functionally distinct classes of plant-derived molecules (Lewis et al., 1998). Lignin is the heterogeneous polymer that confers structural rigidity to plant cell walls, while lignans are plant defense molecules, of recent interest for their oncoprotective role in humans (Adlercreutz, 1995; Arroo et al., 2014; Lewis et al., 1998; Mabrok et al., 2012; Struijs et al., 2009; Wang, 2002). Lignans, when metabolized by gut microbiota, are transformed from phytoestrogens into mammalian estrogens (Mabrok et al., 2012; Wang, 2002). These compounds have been shown to exhibit antagonistic effects on endogenous estrogens, counteracting the proliferative effect they have on breast cancer cells (Adlercreutz, 1995; Arroo et al., 2014; Wang, 2002). In human males, high serum levels of this same estrogen analogue are correlated with a significantly lower risk of acute coronary events (Vanharanta et al., 1999).

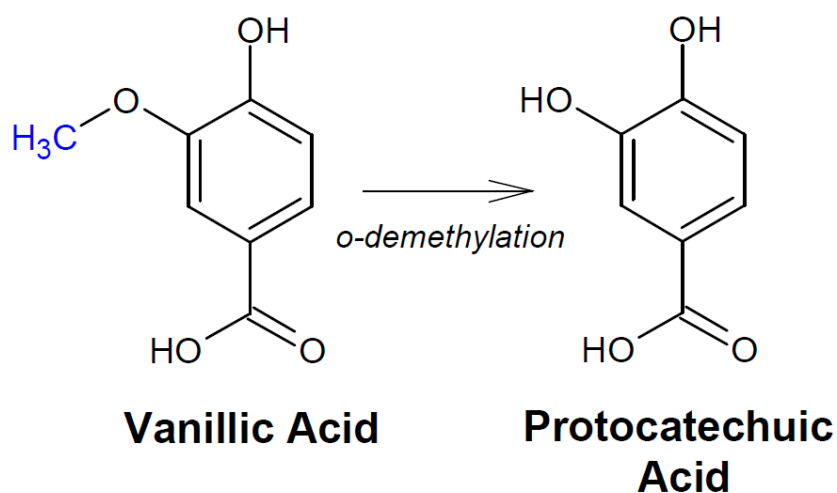


Figure 2. O-demethylation of vanillic acid to protocatechuic acid.

Lignin is often thought of as indigestible by anaerobic bacteria, although anaerobic lignin degradation has been observed in undefined microbial communities and more recently in pure cultures (DeAngelis et al., 2013, 2011; Sleat and Robinson, 1984). When *C. methoxybenzovorans* is grown on lignin-related compounds, the primary fermentation product is acetate, which can be further utilized as a carbon source by other microbes (McInerney et al., 2009; Mechichi et al., 1999; Mori et al., 2012; Morris et al., 2013; Müller et al., 2013). In mammalian hosts, acetate is absorbed through the intestinal lining, whereafter it promotes adipogenesis (Hong et al., 2005).

Here we analyze *C. methoxybenzovorans*'s genome, focusing on its metabolic capabilities, including the genes making up the *O*-demethylation pathway for which it is named. The genome is compared to the recently sequenced genome of *Clostridium indolis* (DSM 755), a very close relative to *C. methoxybenzovorans*, with greater than 99% shared identity between reference 16S rRNA gene sequences.

2.3 Organism information

Classification and features

Clostridium methoxybenzovorans (DSM 12182), initially described in 1999, was isolated from a methanogenic olive mill waste digester. It is a strictly anaerobic, Gram-positive, rod-shaped sporeformer in the family *Lachnospiraceae* (Fig. 3, Fig. 4).



Figure 3. Photographs of *C. methoxybenzovorans*. A) Phase contrast microscopy image of *C. methoxybenzovorans* in defined media with cellobiose without amino acids except cysteine B) Photograph of *C. methoxybenzovorans* colonies on GS2 media with cellobiose C) Phase contrast microscopy image composite of sporulating cells.

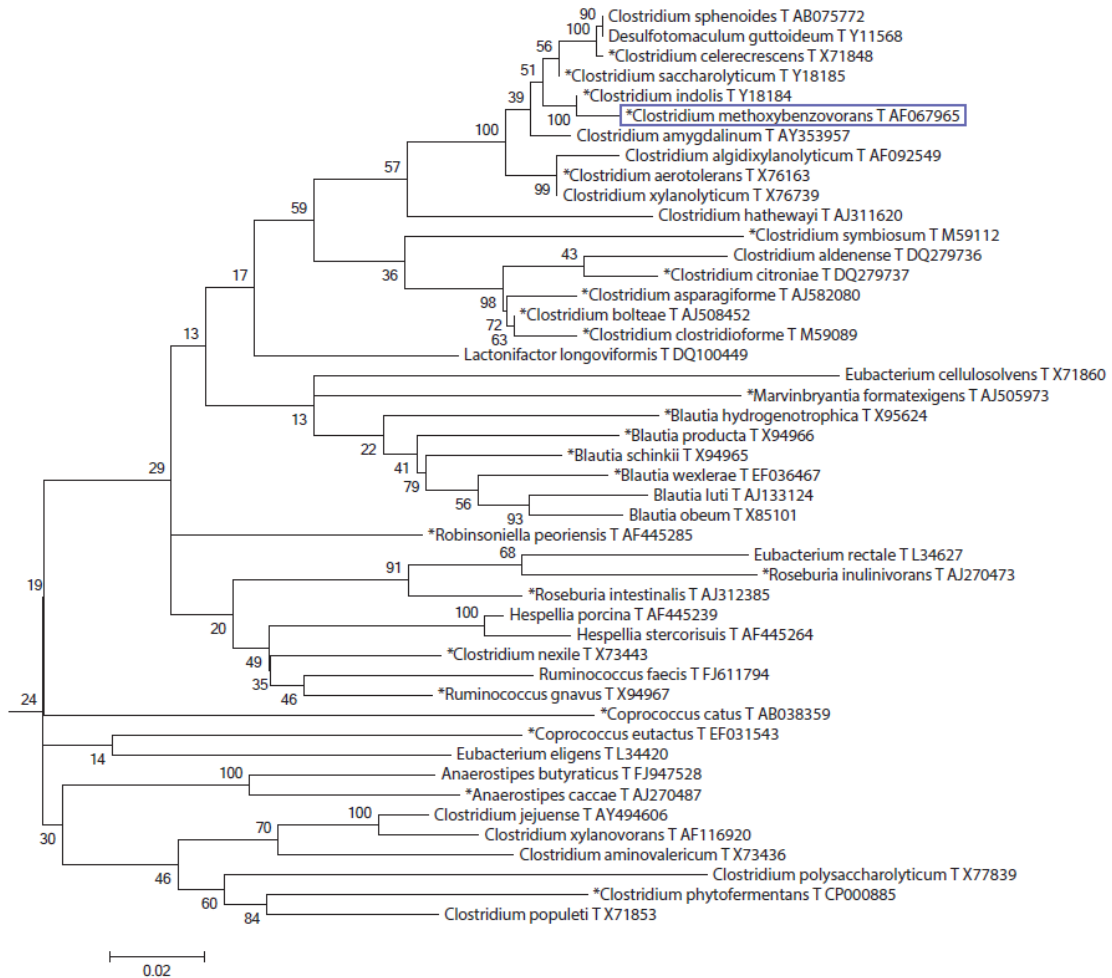


Figure 4. Phylogeny of *C. methoxybenzovorans* and related species in the family Lachnospiraceae. Not all taxa were included due to space. “T” denotes type strain; “*” denotes that the organism’s genome has been sequenced. *C. methoxybenzovorans* is highlighted by the blue box. The tree was rooted with *Bacillus subtilis*. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The tree with the highest log likelihood (-13903.9308) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2085)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.0000% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 101 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 999 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013, p. 6).

It grows optimally at 37°C and at a pH of 7.4, but is able to grow from 20°C to 45°C and pH 5.5 to 9.0 (Table 1). *C. methoxybenzovorans* is able to utilize carbohydrates including glucose, fructose, sorbose, galactose, *myo*-inositol, sucrose, lactose, and cellobiose; hydrogen gas, carbon dioxide, formate, acetate, and ethanol are produced (Mechichi et al., 1999). It is also able to utilize lactate, betaine, sarcosine, dimethylglycine, methanethiol, dimethylsulfide, methanol, and a variety of methoxylated aromatics, including vanillate and syringate (Mechichi et al., 1999). When grown on lactate, *C. methoxybenzovorans* produces methanol; when grown on methoxylated aromatics acetate and butyrate are produced, and on the other substrates acetate alone is produced (Mechichi et al., 1999). Utilization of methoxylated aromatics does not result in ring cleavage, but rather *O*-demethylation, where the methyl group is cleaved from the oxygen of the ether bond (Mechichi et al., 1999).

Table 1. Classification and general features of *Clostridium methoxybenzovorans*.

MIGS ID	Property	Term	Evidence code^a
	Current classification	Domain Bacteria Phylum Firmicutes Class Clostridiales Order Clostridia Family Lachnospiraceae Genus Clostridium Species Clostridium methoxybenzovorans Type strain yes Strain SR3	TAS ³
	Gram stain	Positive	TAS ³
	Cell shape	Rod	TAS ³
	Motility	Nonmotile	TAS ³
	Sporulation	Terminal	TAS ³
	Temperature range	20°C to 45°C	TAS ³
	Optimum temperature	37°C	TAS ³
	pH range; Optimum	5.5 to 9.0; 7.4	TAS ³
	Carbon source	Carbohydrates, fatty acids, CO ₂ /H ₂	TAS ³
	Energy source	Fermentation, sulfur compound respiration	TAS ³ , IDA
MIGS-6	Habitat	Isolated from methanogenic olive mill digester	TAS ³
MIGS-6.3	Salinity	Negative effect at >0.5%, inhibition at 3.5%	TAS ³
MIGS-22	Oxygen	Strict anaerobe	TAS ³
MIGS-15	Biotic relationship	Free living	TAS ³
MIGS-14	Pathogenicity	Not known, but close relatives <i>C. indolis</i> , <i>C. celerecrescens</i> , <i>C.</i> <i>sphenoides</i> and <i>D. guttoideum</i> nonpathogenic in animal tests	TAS
MIGS-4	Geographic location	Sfax, Tunisia	TAS ³
MIGS-5	Sample collection time	NR	
MIGS-4.1-2	Latitude – Longitude	34.748150, 10.745450	NAS
MIGS-4.3	Depth	NR	

(continued)

MIGS-4.4	Altitude	8 m above sea level	NAS
----------	----------	---------------------	-----

- a) Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from <http://www.geneontology.org/GO.evidence.shtml> of the Gene Ontology project (Ashburner et al., 2000).

2.4 Genome sequencing information

Genome project history

Sequencing was carried out by the JGI and funded through community sequencing proposal (CSP) 831, “Understudied soil microbes with underappreciated capabilities: Untangling the *Clostridium saccharolyticum* group.” Sequencing began in November of 2012 and was completed and released to authors in June of 2013. A hybrid of PacBio and Illumina technologies was used to achieve 1619 fold coverage of the genome. Assembly was done using AllPathsLG (Gnerre et al., 2011); gene calling was done in GenePRIMP (Pati et al., 2010) and Prodigal (Hyatt et al., 2010). The final genome is a high quality permanent draft in 14 scaffolds. The genome can be found in the IMG database (taxon ID 2526164529; locus tag H204DRAFT) and in Genbank (assembly ID GCA_000421505.1). The details are summarized in Table 2.

Table 2. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Permanent Draft
MIGS-28	Libraries used	Illumina standard shotgun, Illumina long insert mate pair, PacBio SMRTbell™
MIGS-29	Sequencing platforms	Illumina HiSeq 2000, PacBio RS
MIGS-31.2	Fold coverage	1619X
MIGS-30	Assemblers	AllpathsLG
MIGS-32	Gene calling method	Prodigal, GenePRIMP
	Locus Tag	H204DRAFT
	Genbank ID	729928, GCA_000421505.1
	Genbank Date of Release	2013/07/09
	GOLD ID	Gi22509
	BIOPROJECT	PRJNA187237
MIGS-13	Source Material Identifier	DSM 12182
	Project relevance	<i>O</i> -demethylation, plant degradation

Growth conditions and genomic DNA preparation

C. methoxybenzovorans SR3 DSM 12182 was obtained from the DSMZ culture collection. It was cultivated anaerobically on GS2 medium with cellobiose (McClung et al., 1957; Warnick et al., 2002) with a nitrogen gas headspace. DNA for genome sequencing was extracted using the DNA Isolation Bacterial Protocol available through the JGI with stock concentrations changed to 20mg/mL for both lysozyme and proteinase K. Genomic DNA fragment size was assessed by gel electrophoresis, and sample purity was assayed using a NanoDrop (ThermoScientific, Wilmington, DE). The Quant-iT™ Picogreen assay kit (Invitrogen, Carlsbad, CA) was used as directed in order to quantify the extracted genomic DNA.

Genome sequencing and assembly

The draft genome of *Clostridium methoxybenzovorans* SR3 was generated at the DOE Joint Genome Institute (JGI) using a hybrid of Illumina and Pacific Biosciences (PacBio) technologies. An Illumina standard shotgun library and long insert mate pair library was constructed and sequenced using the Illumina HiSeq 2000 platform (Bennett, 2004). 24,206,550 reads totaling 3,631.0 Mb were generated from the std shotgun and 82,680,164 reads totaling 7,441.2 Mb were generated from the long insert mate pair library. A PacBio SMRTbell™ library was constructed and sequenced on the PacBio RS platform. 132,270 raw PacBio reads yielded 166,032 adapter trimmed and quality filtered subreads totaling 399.6 Mb. All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts (Mingkun et al., 2011). Filtered Illumina and PacBio reads were assembled using AllpathsLG (PrepareAllpathsInputs: PHRED 64=1 PLOIDY=1 FRAG COVERAGE=50 JUMP COVERAGE=50; RunAllpathsLG: THREADS=8 RUN=std pairs TARGETS=standard VAPI WARN ONLY=True OVERWRITE=True) (Gnerre et al., 2011). The final draft assembly contained 14 contigs in 14 scaffolds. The total size of the genome is 7.1 Mb. The final assembly is based on 3,631.0 Mb of Illumina Std PE, 7,441.2 Mb of Illumina CLIP PE and 399.6 Mb of PacBio post filtered data, which provides an average 1559.5X Illumina coverage and 56.3X PacBio coverage of the genome, respectively.

Genome annotation

Genes were identified using Prodigal (Hyatt et al., 2010), followed by a round of manual curation using GenePRIMP (Pati et al., 2010). The predicted CDSs were

translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (Lowe and Eddy, 1997) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (Pruesse et al., 2007). Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (Nawrocki et al., 2009). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform (Markowitz et al., 2009; *The Integrated Microbial Genomes (IMG) platform.*, n.d.) developed by the Joint Genome Institute.

2.5 Genome properties

The genome is currently in fourteen scaffolds. Many putative phage genes are encoded in several of the shorter scaffolds, so it is possible that there are integrated prophages which may hinder concatenation of the remaining scaffolds. The total length of the fourteen scaffolds is 7,085,377 bp with an average GC content of 44.54%. A total of 6,705 genes were predicted, of which 6,592 are protein-coding. Of these protein-coding genes, 5,169 genes were assigned to a putative function with the remaining 1,423 genes annotated as hypothetical proteins. Other genome properties are summarized in Tables 3 and 4.

Table 3. Genome Statistics

Attribute	Value	% of total^a
Genome size (bp)	7085377	100.00%
DNA coding (bp)	6186117	87.31%
DNA G+C (bp)	3155934	44.54%
DNA scaffolds	14	100%
Total genes	6705	100.00%
Protein-coding genes	6592	98.31%
RNA genes	113	1.69%
Pseudo genes	196	2.92%
Genes in internal clusters	5153	76.85%
Genes with function prediction	5196	77.09%
Genes assigned to COGs	4968	74.09%
Genes with Pfam domains	5264	78.51%
Genes with signal peptides	366	5.46%
Genes with transmembrane helices	1555	23.19%
CRISPR repeats	1	00.00%

- a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

Table 4. Number of genes associated with the 25 general COG functional categories

Code	Value	%age of total^a	Description
J	186	3.40%	Translation, ribosomal structure and biogenesis
A	-	-	RNA processing and modification
K	568	10.38%	Transcription
L	370	6.76%	Replication, recombination and repair
B	2	0.04%	Chromatin structure and dynamics
D	43	0.79%	Cell cycle control, Cell division, chromosome partitioning
V	106	1.94%	Defense mechanisms
T	344	6.29%	Signal transduction mechanisms
M	240	4.39%	Cell wall/membrane biogenesis
N	71	1.30%	Cell motility
U	52	0.95%	Intracellular trafficking and secretion
O	128	2.34%	Posttranslational modification, protein turnover, chaperones
C	269	4.92%	Energy production and conversion
G	885	16.17%	Carbohydrate transport and metabolism
E	490	8.95%	Amino acid transport and metabolism
F	115	2.10%	Nucleotide transport and metabolism
H	157	2.87%	Coenzyme transport and metabolism
I	78	1.43%	Lipid transport and metabolism
P	331	6.05%	Inorganic ion transport and metabolism
Q	64	1.17%	Secondary metabolites biosynthesis, transport and catabolism
R	623	11.39%	General function prediction only
S	350	6.40%	Function unknown
-	1737	25.91%	Not in COGs

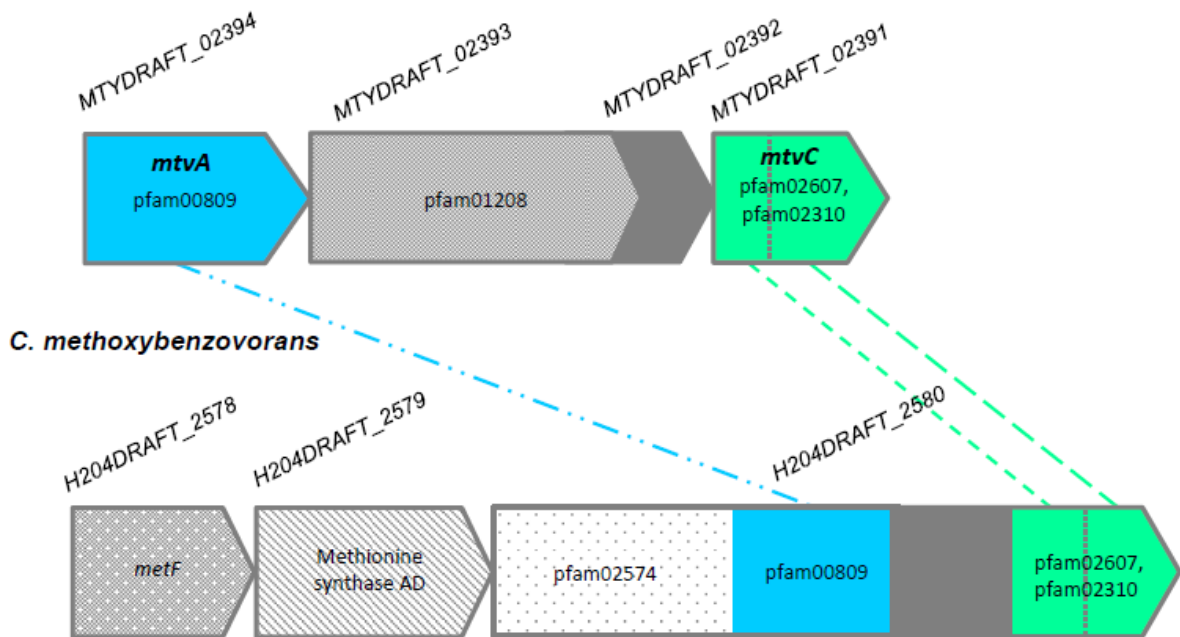
a) The total is based on the total number of protein coding genes in the annotated genome.

2.6 Insights from the genome sequence

One of the best characterized *O*-demethylase systems is from *Moorella thermoacetica*, also in the Class Clostridia. *M. thermoacetica* has a three-component system comprised of the proteins MtvA, MtvB, and MtvC (Naidu and Ragsdale, 2001). MtvA transfers the methyl group from the substrate to MtvC, which then is demethylated by MtvB. MtvB then transfers the methyl group onto tetrahydrofolate, from which it can be used to synthesize methionine from homocysteine or to produce acetate. In the article describing this system, the N-terminal sequences of MtvA and MtvC were published, but no sequence information for MtvB was given (Naidu and Ragsdale, 2001). We located *mtvA* and *mtvC* in the *M. thermoacetica* genome by querying the N-terminal amino acid sequences using BLASTP in IMG (Pierce et al., 2008). The MtvA N-terminal sequence matched MTYDRAFT_02394 (pfam00809/pterin-binding domain; KO term: *metH*); the MtvC hit belonged to MTYDRAFT_02391 (pfam02607 and pfam02310, both B₁₂ binding domains; KO term: *metH*). We used the pfam domains from these proteins to search for the matching proteins in *C. methoxybenzovorans*'s genome; interestingly, all three pfam domains were annotated within one *C. methoxybenzovorans* gene (H204DRAFT_2580), along with an additional S-methyl transferase domain (pfam02574) (Fig. 5). The H204DRAFT_2580 KEGG orthology (KO) term is also MetH, cobalamin dependent methionine synthase. Although no sequencing information was published for the third protein, MtvB, in *M. thermoacetica*, just upstream of H204DRAFT_2580 in *C. methoxybenzovorans*'s genome, there is a gene annotated as 5,10-methylenetetrahydrofolate reductase (MetF/MTHFR), followed by a gene annotated

as a methionine synthase activation domain. Together, MetF and MetH transfer methyl groups during methionine biosynthesis (Blanco et al., 1998; Rodionov et al., 2004). In *Sphingomonos paucimobilis* SYK-6, a *metF* deletion was sufficient to inhibit growth on syringic acid and vanillic acid (Sonoki et al., 2002), coinciding with a buildup of methyl-tetrahydrofolate. It seems likely that MtvB is MetF, in *C. methoxybenzovorans* (Fig. 5).

Figure 5. Syntenic map of the genes responsible for O-demethylation in Moorella *M. thermoacetica*



thermoacetica and their counterparts in the *C. methoxybenzovorans* genome. Shared pfams are color-coded; dotted lines show the corresponding locations of the pfam domains found in *mtvA* and *mtvC*, highlighting that both genes may be encompassed within a single gene in *C. methoxybenzovorans*. Relative lengths are shown to scale.

While *C. methoxybenzovorans* has the ability to *O*-demethylate methoxylated aromatics,

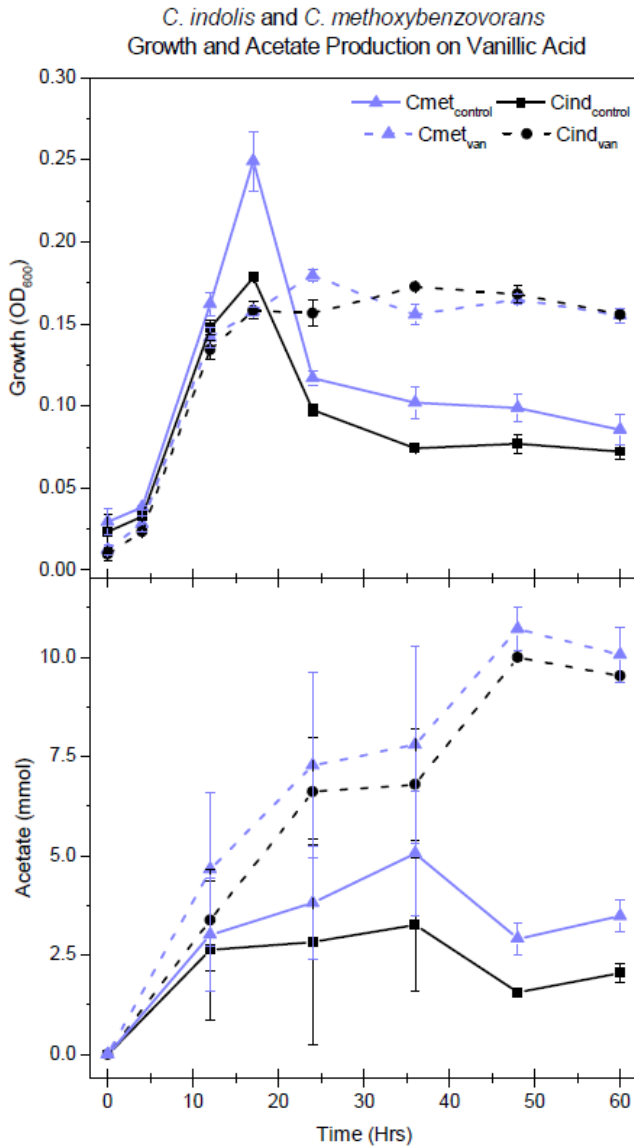


Figure 6. *C. indolis* and *C. methoxybenzovorans* growth on vanillic acid.

it likely does not use them as a carbon source: when comparing inoculated basal medium lacking growth substrate, there is actually a decrease in the maximum turbidity (OD₆₀₀) when cultured with vanillic acid. On basal medium without substrate, *C. methoxybenzovorans* produces about 5mM of acetate, which doubles to 10mM when grown on 5mM vanillic acid (Fig. 6).

The *O*-demethylase pathway in *M. thermoacetica* is known to utilize CO₂ (Naidu and Ragsdale, 2001), and since we could not confidently locate all three components in the genome,

we tested whether addition of CO₂ would enhance acetate production (Fig. 7). While there was no growth benefit to adding CO₂, we did observe an increase in acetate production at and beyond 15 hours. It seems that the *O*-demethylase system in *C.*

methoxybenzovorans functions very similarly to those in *M. thermoacetica* and *S. paucimobilis* and is linked to methionine synthesis (Sonoki et al., 2002).

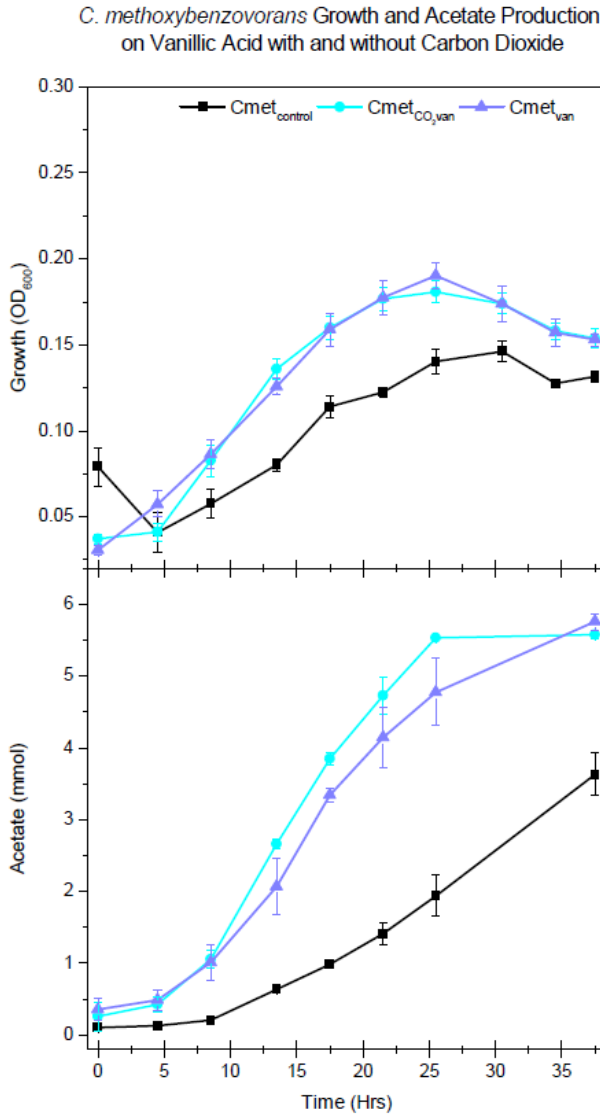


Figure 7. *C. methoxybenzovorans* growth on vanillic acid with and without carbon dioxide.

C. methoxybenzovorans's genome contains two loci that encode BMC proteins. BMCs are proteinaceous organelles descendent from carboxysomes present in Cyanobacteria. In non-photosynthetic bacteria BMCs are thought to be necessary for sequestering toxic intermediates and/or for increasing processivity of unfavorable reactions by concentrating reactants (Cannon et al., 2001; Kaplan and Reinhold, 1999; Penrod and Roth, 2006). There are several different types of BMCs, some capable of utilizing the nitrogen-containing alcohols ethanolamine or choline while others can utilize the methyl-pentoses fucose and rhamnose (Kofoid et al., 1999; Kuehl et al., 2014; Petit et al., 2013; Pitts et al., 2012).

To determine which BMC types *C. methoxybenzovorans* encodes, we searched the genome for marker protein families (pfams), as described by Abdul-Rahman et al.

2013 (AbdulRahman, 2013) and using the pfam identifiers for the choline BMC as detailed in Kuehl et al. 2014 (Kuehl et al., 2014). There are two putative BMC loci based on the prevalence of the vertex protein gene (CcmL/EutN/pfam03319), from H204DRAFT_2037-2057 and from H204DRAFT_3512-3526. The BMCs encoded are the same as in *C. indolis*, including a choline utilization (Cut) BMC and a completely novel type of BMC, as discussed by Biddle et al. 2014, which was named the CoAT BMC (Biddle et al., 2014). This CoAT BMC contains a few genes typically found in BMC operons, such as acetaldehyde dehydrogenase (pfam00171), a class IV alcohol dehydrogenase (pfam00465), and *pduL* (pfam06130), an uncharacterized protein involved in propanediol utilization. There is also a *fucA* homologue (pfam00596), which is present in the Grp locus responsible for fucose/rhamnose metabolism in *Clostridium phytofermentans* and close relative *Clostridium* sp. KNHs212. These species are in the same family, *Lachnospiraceae*, as *C. methoxybenzovorans*. In addition to the two coenzyme A transferases (H204DRAFT_3515-3516/pfam01144), there are two other genes unique to known BMCs: a dihydroxyacetone kinase domain gene (H204DRAFT_3513/pfam02733) and a dihydroxyacetone kinase phosphatase domain gene (H204DRAFT_3512/pfam02734). These genes are conserved in the CoAT operons of the close relatives *C. indolis*, *Clostridium* sp. ASB410, *Clostridium* sp. KNHs206, *Clostridium aerotolerans*, *Clostridium saccharolyticum*, and *Clostridium celerecrescens*. Dihydroxyacetone kinase is involved in glycerol degradation, substituting one of the two hydroxyl groups with a phosphate from ATP.

C. methoxybenzovorans has genes associated with metabolism of isomaltose, cellobiose, α - and β -xylosides, pectin, trehalose, mannose, sucrose, β -glucosides (e.g. salicin, esculin), α -fucosides, α -rhamnosides, arabinofuranosides, mannosylglycerate, and maltose/starch. It does not appear to be capable of cellulolysis. There is a *pulA* homologue (GH13), which hydrolyzes 1-6- α -glucose linkages in polysaccharides such as pullulan, amylopectin, and glycogen (Kornacker and Pugsley, 1990).

There are two encoded chitinases (EC:3.2.1.14, GH18), though enzymes in this class may also act as lysozymes (Ghasemi et al., 2011; Horn et al., 2006). *C.*

methoxybenzovorans has transporters for N-acetylglucosamine, the monomer that makes up chitin and alternates with N-acetylmuramic acid in peptidoglycan. However, lysozyme action leads to a disaccharide of N-acetylmuramic acid and N-acetylglucosamine (Sharon, 1967), so the transporters suggest that the organism is able to utilize chitin as a carbon source.

Table 5. Sialic acid utilization genes in *C. methoxybenzovorans*.

Gene	EC Number	COG Number	Gene locus
sialidase	–	COG4692	H204DRAFT_4826
<i>nanA</i>	4.3.3.1	COG0329	H204DRAFT_2093 H204DRAFT_4229
<i>nanE</i>	5.1.3.9	COG3010	H204DRAFT_2794 H204DRAFT_5889
<i>nanK</i>	2.7.1.60	COG1940	–
<i>nagA</i>	3.5.1.25	COG1820	H204DRAFT_0124 H204DRAFT_0166 H204DRAFT_3534
<i>nagB</i>	3.5.99.6	COG0363	H204DRAFT_0173 H204DRAFT_5987

The EC numbers, COG numbers, and locus ID in C. methoxybenzovorans are shown. We could find homologues for all except nanK.

C. methoxybenzovorans's genome contains a sialidase (H204DRAFT_4826), as well as a nearly complete sialic acid utilization gene set including multiple homologues of *nanAE* and *nagAB* (Table 5) (Li and Chen, 2012; Vimr et al., 2004). The only gene without an obvious potential homologue is *nanK*, an ROK (Repressor, Open reading frame, Kinase) superfamily protein (EC 2.7.1.60). An “incomplete” pathway like this is not uncommon in the genetic profile of sialic acid utilizers, as the operons seem to be very personalized and are often missing distinct homologues of the canonical genes in *Escherichia* and *Salmonella* (Vimr et al., 2004). Given the sialidases and other catabolic enzymes encoded, it is likely that *C. methoxybenzovorans* is able to utilize sialic acids as a source of carbon and nitrogen.

The genome contains at least sixteen different PTS transporters and at least seven complete ABC transporters. PTS transporters (EC:2.7.1.69) encode for cellobiose, mannose, fructose, mannitol, ascorbate, galactitol, arbutin, trehalose, sucrose, glucose, maltose, D-glucosamine, β -glucosides, N-acetyl-D-glucosamine, galactosamine, 2-O- α -mannosyl-D-glycerate, and D-glucosamininate. There are nine copies of the cellobiose PTS system (COG1440 and COG1447), as well as for a mannose/fructose specific PTS system (COG2893 and COG3444).

The encoded ABC transporters are potentially specific for arabinooligosaccharide, ribose, xylose, rhamnose, methyl-galactoside, and aldouronate. Aldouronates, or aldouronic acids, are polymers resulting from the acidic breakdown of glucurono-xylan, which is often methylated (Chow et al., 2007; Komiyama et al., 2009). These compounds can make up as much as ~30% of hemicellulose yet are not ubiquitously fermentable (Bi

et al., 2009). The genome encodes an MsmX homologue, which is a multitasking ABC transporter associated ATPase, known to interact with multiple oligosaccharide transporters (Ferreira and Sá-Nogueira, 2010). *OpuBA*, *opuBB*, and *opuBC* are present and may be choline transporters, feeding into the choline-degrading microcompartment (Kappes et al., 1999).

The original characterization of *C. methoxybenzovorans* by Mechichi et al. reported growth on cellobiose, fructose, galactose, glucose, lactose, sorbose, and sucrose (Mechichi et al., 1999). Of these, it is missing genes only for sorbose transport, though there may be uncharacterized genes responsible for this activity. There is a putative multiple sugar ABC transporter whose homologue is involved in transport of L-arabinose, D-fucose, D-galactose, D-glucose, and D-xylose, though *C. methoxybenzovorans* was reported not to utilize L-arabinose, xylan, or xylose (Mechichi et al., 1999; Zhao and Binns, 2011). *C. methoxybenzovorans* is likely able to grow on rhamnose, mannose, and mannitol, based on transport genes, though these substrates were not originally tested.

According to Mechichi et al., *C. methoxybenzovorans* is not motile (Mechichi et al., 1999). However, *C. indolis* is motile and has the same suite of flagellar genes, as do other motile *Clostridium* species. Both *C. indolis* and *C. methoxybenzovorans* are missing genes for FlgF (the proximal rod), FlgH (the L ring), and FlgI (the P ring), which are considered essential genes in the KEGG pathway for flagellar assembly (Pallen et al., 2005). The P ring and L ring appear to interact with the outer membrane, based on electron micrographs of structurally Gram-negative cells (Chen et al., 2011). These

proteins may not have a purpose in Gram-positive bacteria since hydrophobic proteins would not be necessary in bacteria lacking an outer membrane. In Gram-positive cells, it seems likely that the flagellar hook is directly attached to the proximal rod of the basal body, which is integrated in the peptidoglycan of the cell wall. In support of this, electron micrographs of the flagellar assembly show that *B. subtilis* has only two rings, while *E. coli* has four. *B. subtilis* has the same profile of flagellar genes as *C. methoxybenzovorans*, notably lacking FlgF, FlgH, and FlgI (Kubori et al., 1997; Pallen et al., 2005).

In fact, when we compared all “Clostridium” genomes in IMG, totaling 483: FlgF, FlgH, and FlgI were missing in every one, though many *Clostridium* species are motile. Other flagellar genes that were not present in the *Clostridium* genomes were FlhC, FlhD, FliT, FlgN, and FlgA. So, while *C. methoxybenzovorans* appears to be missing several canonical flagellar genes, this assembly is typical of Gram-positive bacteria.

Consistent with a motile organism, the KEGG pathway for chemotaxis is nearly complete. However, as with the flagellar apparatus, the chemotaxis machinery differs across bacterial phyla. Though the paradigm of chemotaxis is built around *E. coli*, most bacteria do not have CheZ, which is exclusively found in β and γ proteobacteria (Szurmant and Ordal, 2004). Both *C. methoxybenzovorans* and *C. indolis* are missing a homologue for *CheZ* (COG3143/pfam04344), a protein phosphatase that enhances the rate of CheY-P dephosphorylation (Bren et al., 1996; Surmant and Ordal, 2004). Phosphorylated CheY promotes tumbling (clockwise movement), and *E. coli*

mutants lacking CheZ are still motile, but have a tumbling phenotype (Parkinson, 1978). *B. subtilis* is still chemotactic due to a synonymous gene, *FliY* (Bischoff and Ordal, 1992; Szurmant and Ordal, 2004), which also functions as the flagellar switch responsible for directional changes (Szurmant et al., 2003). BLASTP was used to query the *FliY* amino acid sequence from *B. subtilis* (gi: 142926) against *C. methoxybenzovorans*'s genome, resulting in a 40% identity match to H204DRAFT_1262 which has hits to the domains: pfam01052 (surface presentation of antigens, SPOA), pfam04509 (CheC), and TIGR02480 (FliN). These protein domains are present in *FliY* as well, as described on UniProtKB protein knowledgebase (P24073)(Magrane and Consortium, 2011).

C. methoxybenzovorans has twenty chemoreceptor proteins (COG0840), the transmembrane proteins that initiate chemotaxis based on interaction with a specific ligand. The N terminus/periplasmic end of the protein is highly varied, since each is ligand-specific (Zhulin, 2001). Genomes within the taxonomic genus of *Clostridium* contain between 1 and 85 such proteins; close relatives *C. saccharolyticum*, *C. indolis*, and *C. celerescens* have 20, 20, and 18, respectively. Little is known about the specificity of these proteins, so we could not determine to which compounds the cell is attracted based on the sequences. Given the presence of chemotactic genes, the complete Gram-positive flagellar apparatus, and *C. indolis*'s motility, *C. methoxybenzovorans* should be motile as well, though this has not yet been observed. As such, further work is necessary to test for motility in *C. methoxybenzovorans*.

C. methoxybenzovorans does not appear to have a complete citric acid cycle: genes encoding malate dehydrogenase, succinic dehydrogenase, and succinyl-CoA

synthetase are not present in the genome. Moreover, the pathway for producing acetyl-CoA is incomplete, lacking pyruvate dehydrogenase which would convert pyruvate to acetyl-CoA. While missing genes may be in sequencing gaps, the same genetic profile is found in *C. indolis*.

C. methoxybenzovorans has a full suite of citrate utilization genes (*citD*, *citE*, *citF*, *citG*, *citX*, and *citMHS*, and putative analogues of *citAB*; H204DRAFT_1133-1142), yet its growth on citrate is impaired compared to close relatives *C. sphenoides*, *D. guttoideum*, and *C. celerecrescens* (data not shown). Closer inspection of these genes shows that the operon has been split by a transposase insertion between *citE* and *citF*, citrate lyase subunits β and α respectively. This insertion appears to interrupt the stop codon within *citE*, leading to an extra 7 amino acids at the C terminus, possibly inhibiting its binding to CitF and causing the observed phenotype.

In the *C. methoxybenzovorans* characterization paper, Mechichi et al. (Mechichi et al., 1999) reported that tests for dissimilatory sulfur metabolism, including reduction of sulfate, sulfite, and thiosulfate, were all negative. However, *C. methoxybenzovorans* has genes linked to assimilatory and dissimilatory sulfate reduction. Import of sulfate into the cytoplasm via an ABC transporter (CysPUWA/H204DRAFT_5788-5791), production of 5'-adenylylsulfate (APS) from sulfate and AMP (CysND/H204DRAFT_5780-5781), and production of sulfite and AMP from APS (CysH/H204DRAFT_0838, AprAB/H204DRAFT_5782-5783) can be easily accounted for based on *C.*

methoxybenzovorans's genes (Bradley et al., 2011). However, the KEGG pathway did not map a homologue for sulfite reductases, which would catalyze the final reduction of sulfite to sulfide.

Since most of the sulfate reduction genes are near each other in the genome, it seemed possible that they may be in a single operon. So, we scanned the flanking genes and found one (H205DRAFT_5778) annotated as a dissimilatory sulfite reductase (desulfoviridin AB), which is not included in the KEGG pathway for sulfur metabolism. This enzyme catalyzes the reduction of sulfite to sulfide. To double check the annotation of this gene, we retrieved the amino acid sequence for desulfoviridin from its host, *Desulfovibrio desulfuricans* (F461DRAFT_02952) and used TBLASTN to locate homologues in *C. methoxybenzovorans*. The best hit was H205DRAFT_5778; both are called as COG2221. *C. indolis* is characterized as producing H₂S (McClung et al., 1957) and has an identical dissimilatory sulfite reductase. The genes just upstream of H205DRAFT_5778 are cysteine synthase and o-acetylhomoserine sulphydrylase, both involved in biosynthesis of the sulfur-containing amino acids methionine and cysteine. Additionally, the *C. methoxybenzovorans* genome encodes a putative sulfurtransferase (EC 2.8.1.1/H204DRAFT_4209) responsible for reducing thiosulfate to sulfite.

The KEGG pathway map indicates an inability to produce tyrosine and phenylalanine, due to lack of a transaminase (EC 2.6.1.5) that would catalyze the amination of the carbonyl in phenylpyruvate or 4-hydroxypyruvate to produce phenylalanine or tyrosine, respectively. Yet, *in vitro*, *C. methoxybenzovorans* grows well on minimal media, with the only amino acid source present being cysteine, a reducing

agent for anaerobic media. The organism took slightly longer to grow than it did on rich media, but ultimately reached an absorbance of ~1.2 (optical density at 600nm) (data not shown). The enzymes catalyzing this reaction are known to have broad specificities, and the genome contains an aspartate transaminase (EC 2.6.1.1) that could also function on tyrosine, phenylalanine, and tryptophan (Artimo et al., 2012). According to its genome, it is also a selenocysteine synthesizer, encoding selenophosphate synthase (H204DRAFT_5628) and L-seryl-tRNA selenium transferase (H205DRAFT_5619).

The genome encodes ABC transporters for lysine, methionine, glutamine, and branched amino acid uptake. The support for cystine transport is ambiguous: the ABC transport system is lacking *tcyJ*, similar to *tcyK*, which is present. The amino acid sequences for these two proteins are 57% identical, but each's specificity is unknown, so it is not known how this affects transport (Burguière et al., 2004).

The genome seems to encode pathways for synthesis of nucleotides from amino acids, except for thymine. This may be one reason that yeast extract is required for growth.

2.7 Conclusions

C. methoxybenzovorans is an interesting microbe with a varied metabolism, probably capable of sulfate reduction, transformation of lignin/lignan compounds via *O*-demethylation, and utilization of a wide array of carbohydrates. Anaerobic respiration of sulfur compounds likely enhances the growth rate of this organism, while its broad substrate repertoire, which includes rare or difficult to degrade molecules, like trehalose,

lignin derivatives, and aldouronic acids, provides it with a competitive advantage. The ability to degrade these ubiquitous plant-derived molecules makes *C. methoxybenzovorans* a potential asset within a bioprocessing consortium, given that it produces acetate, CO₂, and H₂ as major endproducts.

While these attributes benefit a soil-dwelling *C. methoxybenzovorans*, they may also provide a competitive advantage in a gut system. It has the ability to utilize sialic acid residues and conjugated sugars (e.g. fucosides and galactosides), which are presented on mucins which line the gut, acting as endogenous prebiotics that shape the microflora (Pacheco et al., 2012; Stahl et al., 2011; Tailford et al., 2015; Vimr et al., 2004). *C. methoxybenzovorans* has been shown to aid in the conversion of the phytoestrogen secoisolariciresinol (SECO) to the mammalian lignans enterodiol (END) and enterolactone (ENL) and to produce caffeic acid from ferulic acid (Chamkha et al., 2001; Micard et al., 2002) through its *O*-demethylation activity. Flavones with more hydroxyl groups (produced by *O*-demethylation) are more potent inhibitors of mutagenesis that can initiate cancer (Arroo et al., 2014). These polymethoxylated flavones have also been shown to be selectively cytostatic and apoptotic in breast cancer cells and to lessen the risk for acute coronary events (Arroo et al., 2014; Mabrok et al., 2012; Peterson et al., 2010; Vanharanta et al., 1999). These traits may make *C. methoxybenzovorans* attractive as a probiotic accompanied by a high lignan diet.

2.8 Taxonomic and nomenclatural proposals

Given the great similarity (~99.36%) between the published representative 16S rDNA sequences (in RDP) of *C. indolis* and *C. methoxybenzovorans*, we built a phylogenetic tree with all six copies of the 16S rRNA gene from both genomes (Fig. 8).

Two of the copies are identical to each other, while the other four copies have no exact match in the other organism. It is interesting to note that neither of the published sequences exactly matches any of the genomic copies, with signals from heterogeneous amplicons used in the original Sanger sequencing leading to a sequencing-based chimera or with physical chimeras forming during PCR amplification of the 16S rRNA gene.

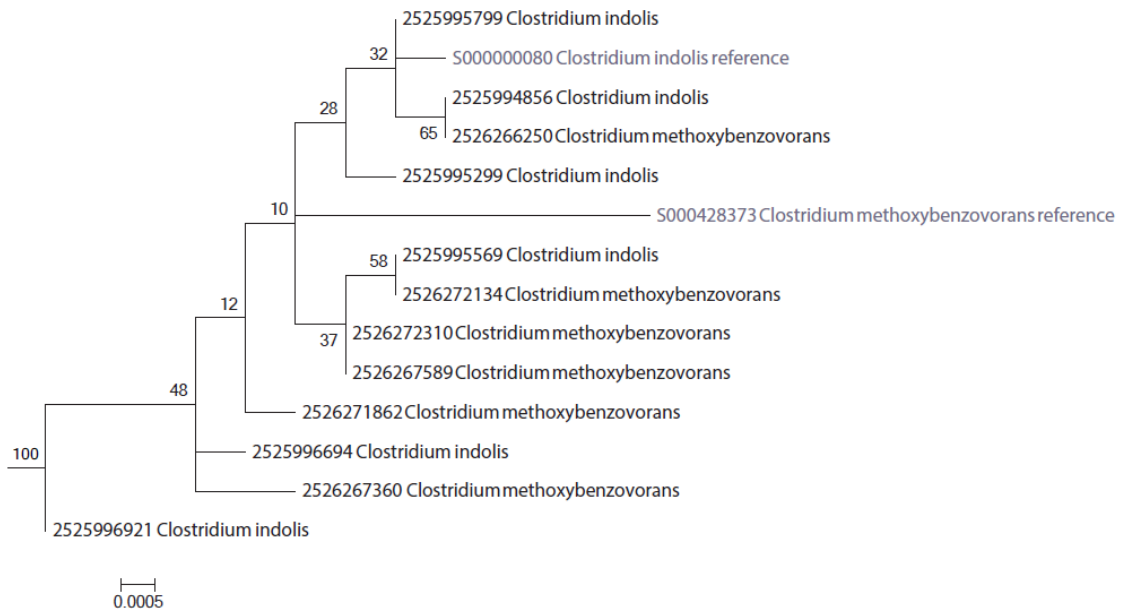


Figure 8. Phylogenetic tree of all copies of the 16S rRNA gene in *C. indolis* and *C. methoxybenzovorans*, with reference sequences taken from the RDP (Cole et al., 2009). *Clostridium saccharolyticum* 16S rDNA sequences were used as the outgroup. Numbers preceding the taxa name are the gene numbers in IMG. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model⁸⁰. The tree with the highest log likelihood (-2263.7911) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.0500)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 2.1033% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 20 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1413 positions in

the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013, p. 6).

We then compared the genome of *C. methoxybenzovorans* to *C. indolis* in order to obtain better resolution than is possible with a lone marker gene. We determined the ANI shared between the two taxa (Goris et al., 2007) and also employed the DSMZ’s GGDC (Meier-Kolthoff et al., 2013a) to understand their whole genome similarity. The ANI values were consistently ~99.5% between *C. indolis* and *C. methoxybenzovorans* regardless of method (Table 6). Based on the DSMZ’s GGDC, the estimated DNA-DNA hybridization values were between 82.00% and 94.50% (Table 7), which are well above the traditional 70% species delineation cut-off.

Table 6. Average Nucleotide Identity between *C. indolis* and *C. methoxybenzovorans*

1-Way ANI 1	1-Way ANI 2	2-Way ANI
99.46%	99.45%	99.53%

Calculated using the method published by Goris et al. 2007 (Goris et al., 2007) with the online tool developed by the Konstantinidis group (“Kostas lab | ANI calculator,” n.d.). Values above ~95% identity suggest the two organisms are the same species.

Table 7. GGDC values for *C. indolis* and *C. methoxybenzovorans*.

Method:	HSP length / total length	identities / HSP length	identities / total length
DDH Estimate	82.00% +/- 3.59	94.50% +/- 1.48	87.00% +/- 2.79
Probability Same			
Species	92.89%	97.09%	98.88%
Probability Same			
Subspecies	67.68%	72.08%	78.49%

Each genome was uploaded onto the DSMZ's Genome-to-Genome Distance Calculator website, which mimics DNA-DNA hybridization (DDH) in silico (Meier-Kolthoff et al., 2013b). A 70% hybridization value is the lower limit for two organisms to be considered the same species; a value of 79% is the lower limit for grouping into the same subspecies.

A pairwise BLASTP comparison was performed on the predicted genes in *C. indolis* and *C. methoxybenzovorans*. Of genes binned into a COG category, there were 600 without homologues in one or the other genome; 236 unique genes in *C. indolis* and 364 in *C. methoxybenzovorans*. Many are explicitly phage-related, encoding phage holins, Pin-like DNA invertases, reverse transcriptases, and transposases. Many others are involved in transcription (transcription factors, helicases, polymerases, and recombinases), which may also be phage-encoded. It is difficult to tell whether these genes confer a phenotypic difference between these organisms, especially given the lack of comparable phenotypic data.

Neither organism's namesake biochemistry is unique to it: *C. methoxybenzovorans*'s genome contains a tryptophanase gene (H204DRAFT_1573) and produces indole in culture (data not shown), the compound which conferred *C. indolis* its name. And when *C. indolis* was inoculated alongside *C. methoxybenzovorans* in basal media supplemented with 5mM vanillic acid, the two strains' growth rates and maxima were indistinguishable, as well as their production of acetate (Fig. 6). *C. indolis* and *C. methoxybenzovorans* produce the same fermentation products: ethanol, acetate, formate, and butyrate (McClung et al., 1957; Mechichi et al., 1999). As mentioned above *C. indolis* and *C. methoxybenzovorans*, have the same genetic profile for systems about which the phenotypic characterizations disagree, including motility and sulfur reduction. Many other phenotypes are not comparable due to lack of data on one or the other organism, including utilization of certain sugars (sorbose, sucrose, galactose, fructose, etc.), pH range for growth, and growth on proteins/amino acids.

Based on genomic, phylogenetic, and phenotypic considerations, and given that *C. indolis* was characterized prior to *C. methoxybenzovorans*, *C. methoxybenzovorans* is a later heterotype of *C. indolis*.

Genome sequencing of other taxa from the same phylogenetic group (*Clostridium sphenoides*, *Desulfotomaculum guttoideum*, and *Clostridium celerecrescens*) from the same phylogenetic group is currently underway. Genomic characterization of these species is likely to shed light on their shared evolutionary history, physiologies, and commonalities of the roles they play in their environments.

CHAPTER 3

RECLASSIFICATION OF THE CLOSTRIDIUM CLOSTRIDIOFORME AND CLOSTRIDIUM SPHENOIDES GROUPS AS ENTEROCLOSTER, GEN. NOV., and LACRIFORMIX, GEN. NOV., INCLUDING RECLASSIFICATION OF FIFTEEN TAXA

3.1 Abstract

The taxonomy and phylogeny of the genus *Clostridium* are at odds, with representatives crossing several families and even into a different phylum. While this issue is well-known within the field, formal revision requires systematic studies of the individual misclassified taxa. Motivated by recently completed genome sequences, here we propose reclassification of two separate clades that include misclassified *Clostridium* species which phylogenetically lie within *Lachnospiraceae* family, now known for being benign members of human and animal gut microbiomes and for their plant-degrading capabilities. We use several phylogenetic perspectives as well as phenotypic comparisons to gain insight into the evolutionary history of these taxa and support for their reclassification. One clade which includes *Clostridium clostridioforme*, *Clostridium bolteae*, *Clostridium lavalense*, *Clostridium asparagiforme*, *Clostridium aldenense*, and *Clostridium citroniae* we propose to reclassify as *Enterocloster*, gen. nov., and reclassification of the species as *Enterocloster clostridioforme*, comb. nov., *Enterocloster bolteae*, comb. nov., *Enterocloster lavalense*, comb. nov., *Enterocloster asparagiforme*, comb. nov., *Enterocloster aldenense*, comb. nov., and *Enterocloster citroniae*, comb.

nov. The other clade comprises *Clostridium sphenoides*, *Clostridium indolis*, *Clostridium saccharolyticum*, *Clostridium celerecrescens*, *Clostridium aerotolerans*, *Clostridium xylanolyticum*, *Clostridium algidixylanolyticum*, *Clostridium amygdalinum*, and *Desulfotomaculum guttoideum*, and we propose to reclassify as *Lacriformis*, gen. nov., including reclassification of the members as *Lacriformis indolis*, comb. nov., *Lacriformis sphenoides*, comb. nov., *Lacriformis saccharolyticum*, comb. nov., *Lacriformis celerecrescens*, comb. nov., *Lacriformis aerotolerans*, comb. nov., *Lacriformis xylanolyticum*, comb. nov., *Lacriformis algidixylanolyticum*, comb. nov., *Lacriformis amygdalinum*, comb. nov., and *Lacriformis guttoideum*, comb. nov.

3.2 Introduction

Though the first *Lachnospiraceae* was isolated from human feces (Tissier, 1908), many members were subsequently isolated from rumens or other animals (Berkhoff, 1985; Bryant, 1959; Bryant et al., 1958a, 1958b, Bryant and Small, 1956a, 1956b; Cai and Dong, 2010; Greening and Leedle, 1989; Gylswyk, 1980; Gylswyk and Toorn, 1987, 1985; Kotsyurbenko et al., 1995; Varel et al., 1995). Renewed interest in the human gut microbiome over the last decade has led to the isolation and description of about half of the taxa currently in the *Lachnospiraceae* (Amir et al., 2014; Broda et al., 2000; Cai and Dong, 2010; Carlier et al., 2007, 2004; Clavel et al., 2007; Cook et al., 2007; Cotta et al., 2009; Domingo et al., 2009; Downes et al., 2002; Duncan et al., 2006, 2002; Eeckhaut et al., 2010; Furuya et al., 2010; Hedberg et al., 2012; Jeong et al., 2004; Kaur et al., 2014; Kim et al., 2011; Kitahara et al., 2000; Klaring et al., 2015; Kopečný et al., 2003, 2003; Liu et al., 2008; Lomans et al., 2001; Mohan et al., 2006; Park et al., 2013; Parshina et

al., 2003; Sakuma K et al., 2006; Schwiertz et al., 2002; Song et al., 2003; Steer et al., 2001; Taras et al., 2002; Warnick et al., 2002; Warren et al., 2006; Whitehead et al., 2004; Whitford et al., 2001; Wielen et al., 2002; Wolin et al., 2003). Approximately 10-30% of the average human gut microbiome is composed of *Lachnospiraceae* (Eckburg, 2005; Ross et al., 2015), with butyrate producers like *Roseburia* spp. gaining attention for their role in attenuating and preventing colon cancer (Encarnação et al., 2015; Louis et al., 2007).

Prior to the creation of the Family *Lachnospiraceae*, this large group was known as *Clostridium* cluster XIVa, due to the large number of misclassified *Clostridium* species (Collins et al., 1994). These taxa make up a significant portion of the phylogenetic breadth of within the family, though they are taxonomically still included in the *Clostridiaceae*. It has been clear for the past two decades that *Clostridium* species that fall outside the *Clostridium sensu stricto* genus are in need of reclassification (Collins et al., 1994; Lawson and Rainey, 2016; Stackebrandt et al., 1999; Yutin and Galperin, 2013), and now with the interest in the human microbiome and attempts to isolate and characterize these novel taxa, this task is increasingly relevant. To enhance the visibility and accessibility of this group of *Clostridia*, we aim to untangle the taxonomy and phylogeny of a large group of misclassified taxa which phylogenetically belong to the *Lachnospiraceae*.

Based on 16S rRNA gene phylogeny, there are three main subclades within the *Lachnospiraceae* comprising misclassified *Clostridium* spp., 1) one containing *Clostridium phytofermentans*, *Clostridium populeti*, *Clostridium polysacchrolyticum*, and *Clostridium herbivorans*, and the recently proposed genus *Anaerocolumna* (previously

Clostridium xylanovorans, *Clostridium aminovalericum*, and *Clostridium jejuense*) 2) another containing *Clostridium clostridioforme*, *Clostridium bolteae*, *Clostridium citroniae*, *Clostridium asparagiforme*, *Clostridium lavalense*, and *Clostridium aldenense*, and 3) a third containing *Clostridium xylanolyticum*, *Clostridium indolis*, *Clostridium saccharolyticum*, *Clostridium celerecrescens*, *Clostridium amygdalinum*, *Clostridium algidixylanolyticum*, *Clostridium aerotolerans*, *Clostridium sphenoides*, and *Desulfotomaculum guttoideum*. Together these subclades comprise two-thirds of the *Clostridium* species in the *Lachnospiraceae*, with the remaining taxa interspersed throughout the family or in the deep-branching clade previously known as Cluster XIVb (Collins et al., 1994) or *Tyzzarella* (Yutin and Galperin, 2013). Recent genomic sequencing of many type strains within the *C. sphenoides* and *C. clostridioforme* (groups 2 and 3, above) clades have made higher resolution comparisons possible in order to better classify these taxa.

3.3 Results and Discussion

The *C. clostridioforme* and *C. sphenoides* groups are closely related but 16S rRNA gene based phylogenies place these two groups in two clades, separated by deeper-branching outliers (Figs. 9-10), two of which have already been placed in a novel genus (*Hungatella*) (Kaur et al., 2014). The general arrangement of the clades is conserved between the neighbor-joining and maximum likelihood topologies of the Family, though there are slight differences in intraclade placements. Using 16S rRNA gene reference sequences as a phylogenetic marker, distances within the *C. sphenoides* group range from 0.00% to 2.56% and within the *C. clostridioforme* group from 0.61% to 2.67%.

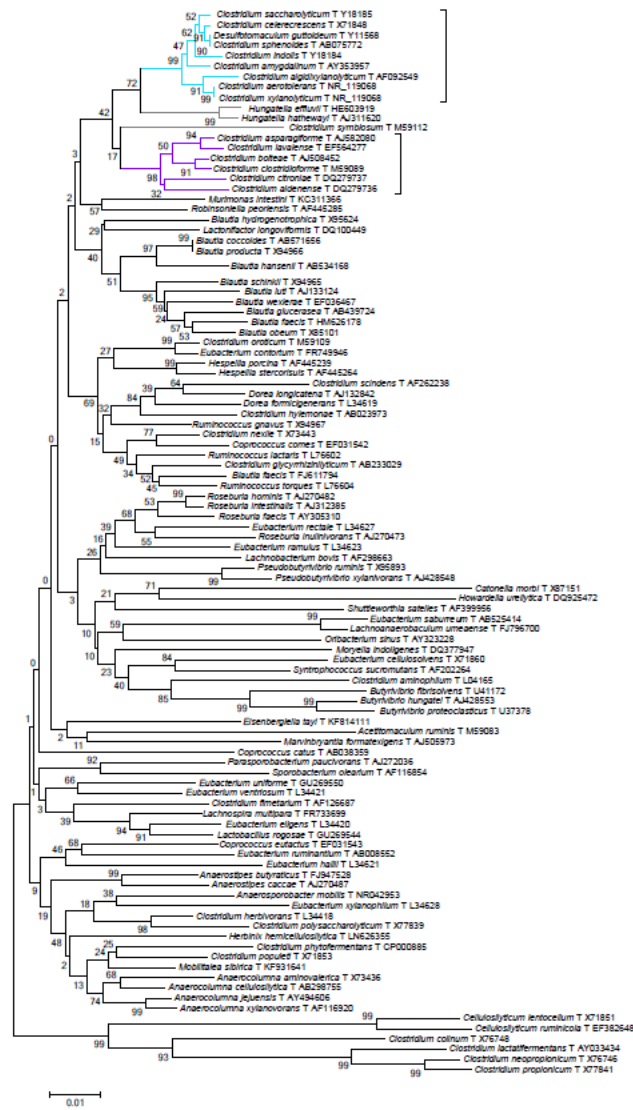


Figure 9. Neighbor-joining phylogeny using the 16S rRNA gene. The optimal tree with the sum of branch length = 2.37671466 is shown. The fraction of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Saitou and Nei, 1987). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 4). The analysis involved 104 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 994 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013, p. 6).

The distances between these two groups (3.95% to 5.49%) are typical of a single genus within the *Lachnospiraceae*, rather than disparate genera.

To test the robustness of these relationships, we developed phylogenies with the *rpoB* gene (Fig. 11) and with a concatenation of 21 conserved single-copy genes (Lang et al., 2013) (Fig. 12) for greater resolution. *C. xylanolyticum*, *C. amygdalinum*, *C. algidixylanolyticum*, and *D. guttoideum* lack genome sequences and thus were not included in the analysis. The results are generally congruent with the 16S rRNA phylogeny (excepting a few differences in branching order within the groups) with separate clades corresponding to the two groups, and *C. symbiosum* being the last common ancestor for both species groups as well as *Hungatella*.

These higher resolution phylogenies hinted at genomic differences that are not reflected in the very low 16S rRNA interspecies distances. To gain a better perspective on whole genome similarity, we performed BLAST-based ANI (ANiB) analyses (Goris et al., 2007; Richter et al., 2015) (Table 8). Resulting ANI values below 95% (equivalent to 70% DDH) support the delineation of these taxa as separate species, even those that are less than 1% apart based on the 16S rRNA gene. Between clades, the genomes are very dissimilar, with values often falling below the effective limit (~70% ANI) of this method (Goris et al., 2007). Compounding these low similarity values are even smaller fractions of the genomes that could be aligned, denoting large genome coding/content differences.

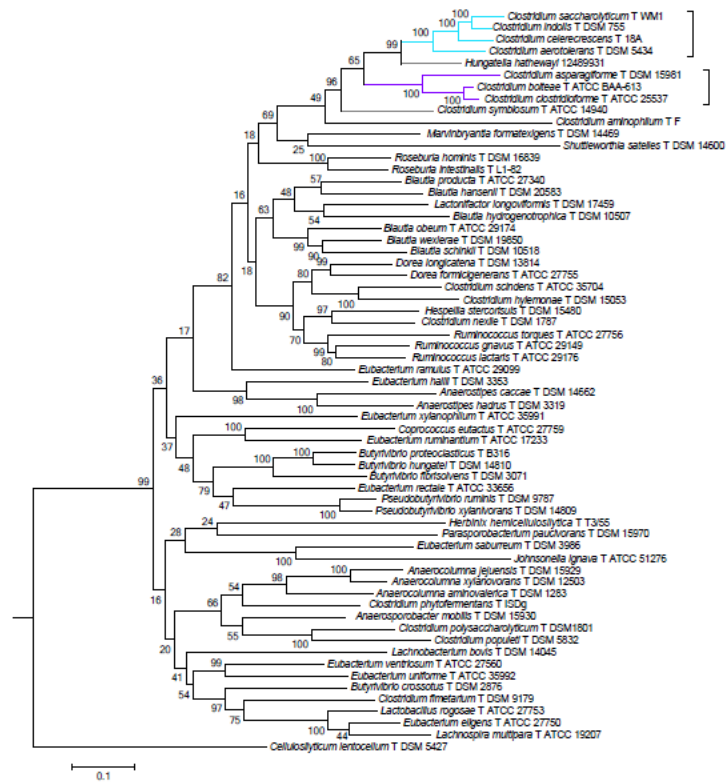


Figure 11. *rpoβ* gene based phylogeny. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The tree with the highest log likelihood (-95530.9770) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (4 categories (+G, parameter = 0.7955)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 28.9117% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 65 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 2874 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013).

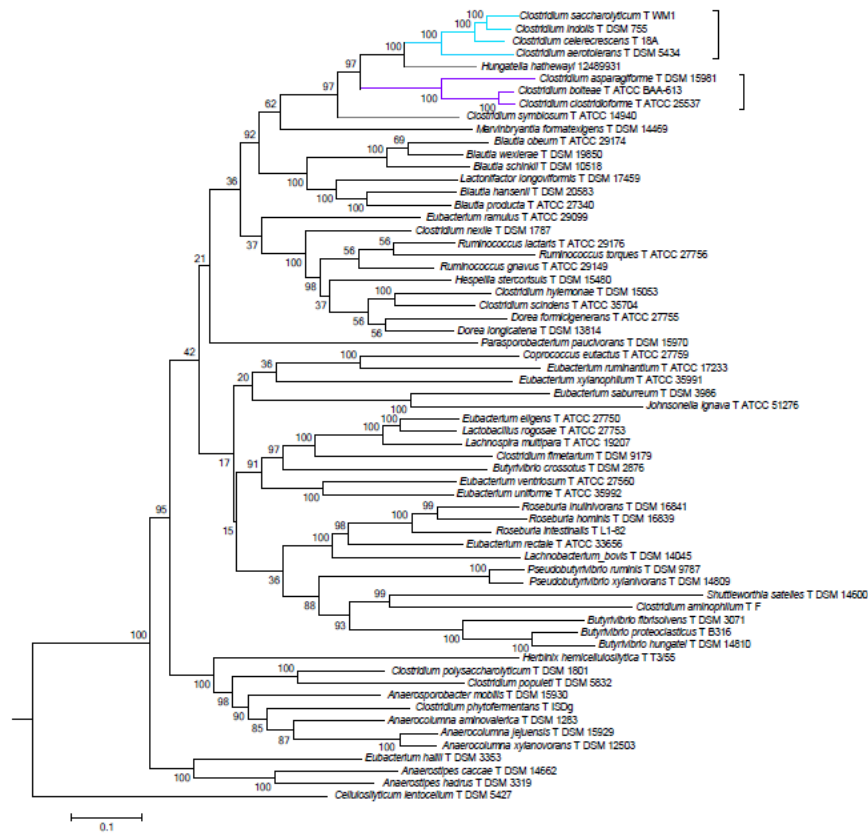


Figure 12. *Lachnospiraceae* phylogeny based on concatenated alignment of 21 single-copy marker genes. Genomes of *Lachnospiraceae* type species were curated to a set of 65. Due to multiple marker genes missing, two genomes were excluded. Two genes were left out of the analysis, since they were not present in several of our genomes of interest, leading to 22 rather than 24 genes being used (Lang et al., 2013). Each gene set was aligned separately in MEGA6 using ClustalW1.6 (Larkin et al., 2007; Tamura K et al., 2013), before concatenating all alignments. The phylogeny was built using the Maximum Likelihood method and the General Time Reversible (GTR) model in MEGA6 with 500 bootstrap replicates (Nei and Kumar, 2000). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value (-427761.6396). A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.7351)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 25.0146% sites). Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 12714 positions in the final dataset.

The species groups can also be separated based on G + C content. The *C. sphenoides* group has G + C content ranging from 42 mol% to 45 mol%, with the deepest branch being a taxonomically distinct genus: *Hungatella*, the type species *hathewayi* containing 49 mol% G + C. The *C. clostridioforme* group's G + C content ranges from 49 mol% to 56 mol%, with another deeper branching relative, *Clostridium symbiosum*, containing 48 mol% G + C.

Beyond genomic differences, there are phenotypic and ecologic features that aid in differentiating one clade from the other. Based on the isolation source for type strains, the *C. sphenoides* group is mainly plant/soil associated while the *C. clostridioforme* group are often implicated in human disease (Table S1). *C. aldenense* and *C. citroniae* were isolated from human peritoneal fluid; *C. lavalense*, *C. asparagiforme*, and *C. boltea* were isolated from human feces. And while the type species of the group *C. clostridioforme* was originally isolated from calf rumen, it has more recently and more often been discovered in various human clinical specimens based on 16S rRNA datasets and literature searches.

Table 8. ANIb for all sequenced type genomes in the clade spanning the *Clostridium sphenoides* and *Clostridium clostridioforme* groups, including the outliers *Hungatella hathewayi* and *Clostridium symbiosum*. The outliers are denoted by grey text, the *C. sphenoides* group is shaded light grey, and the *C. clostridioforme* group is shaded medium grey. The ANIb is shown with the percentage of the genome aligned in brackets. Values in red are below the confidence threshold for this method. None of the compared genomes are above the ~95% similarity threshold which denotes membership in a single species.

	<i>C. aerotolerans</i>	<i>C. saccharo-lyticum</i>	<i>C. indolis</i>	<i>C. sphenoides</i>	<i>C. celerecrecens</i>	<i>H. hathewayi</i>	<i>C. bolteae</i>	<i>C. asparagiforme</i>	<i>C. clostridioforme</i>
<i>C. aerotolerans</i>	-	74.81 [42.21]	74.93 [50.63]	74.87 [48.59]	75.00 [47.58]	71.36 [29.69]	68.54 [24.46]	67.86 [21.88]	68.60 [24.41]
<i>C. saccharolyticum</i>	74.79 [43.82]	-	86.05 [66.62]	84.30 [63.41]	84.26 [65.29]	72.37 [33.76]	69.67 [29.32]	69.37 [26.18]	69.54 [29.54]
<i>C. indolis</i>	74.69 [38.57]	85.22 [49.86]	-	84.47 [58.12]	84.59 [59.83]	71.95 [28.73]	69.41 [24.99]	68.70 [21.97]	69.23 [24.45]
<i>C. sphenoides</i>	74.63 [44.94]	83.77 [57.23]	84.86 [69.23]	-	93.82 [80.68]	72.11 [31.11]	69.31 [26.58]	68.79 [23.70]	69.14 [26.63]
<i>C. celerecrecens</i>	74.79 [44.21]	83.79 [59.28]	85.10 [71.71]	93.84 [81.09]	-	71.92 [32.13]	69.37 [28.10]	68.73 [24.76]	69.30 [27.58]
<i>H. hathewayi</i>	71.26 [21.23]	72.22 [23.50]	72.15 [26.52]	71.98 [24.08]	71.87 [24.50]	-	71.24 [24.67]	70.34 [22.02]	70.11 [20.91]
<i>C. bolteae</i>	68.19 [18.98]	68.96 [21.38]	69.12 [24.31]	68.84 [22.05]	68.83 [22.81]	70.60 [25.46]	-	72.05 [31.16]	89.29 [53.93]
<i>C. asparagiforme</i>	67.72 [16.72]	68.79 [19.59]	68.73 [21.37]	68.46 [19.63]	68.56 [20.17]	70.17 [23.29]	72.10 [31.62]	-	72.71 [28.29]
<i>C. clostridioforme</i>	68.43 [22.30]	69.30 [26.01]	69.46 [28.16]	69.26 [25.82]	69.20 [26.59]	70.06 [26.10]	89.76 [63.16]	72.90 [32.75]	-
<i>C. symbiosum</i>	68.11 [21.54]	68.70 [23.02]	68.49 [24.15]	68.50 [23.07]	68.48 [23.09]	69.68 [24.15]	68.98 [25.46]	68.76 [23.89]	70.02 [25.63]

Butyrate production and spore-formation have been proposed as indicators for gut-association within the *Lachnospiraceae* (Meehan and Beiko, 2014). Fermentation product data is missing for some members of the *C. clostridioforme* group, but none of those with product data have been shown to produce butyrate. This endproduct, however, is related to the carbon source provided, and these taxa were generally maintained on blood agar plates which may not provide appropriate feedstocks for butyrate production (Duncan et al., 2004; Encarnaç o et al., 2015; Falony et al., 2006; Van den Abbeele et al., 2013). In contrast, most of the *C. sphenoides* organisms have been shown to produce butyrate. We screened all available genomes from both groups in order to determine butyrate producing capability, using the marker genes from the four validated common pathways used by Vital *et. al* (Vital et al., 2014). Only *C. saccharolyticum* (a member of the *C. sphenoides* group) and *C. asparagiforme* (a member of the *C. clostridioforme* group) lack all four butyrate pathways. Interestingly, the *C. sphenoides* group (four of five genomes) contained only the 4-aminobutyrate pathway, while the *C. clostridioforme* group (three of four genomes) contained only the acetyl-CoA pathway. While butyrate production has been linked to gut inhabitation (Meehan and Beiko, 2014), which pathway is used may be a better indicator. Meehan and Beiko’s work focused on only the acetyl-CoA pathway for butyrate formation, using the final enzymes (butyrate kinase or butyryl-CoA:acetyl-CoA transferase) as proxies. While this is the most common pathway to butyrate, there are several alternative known pathways. Vital *et al.* performed a large meta-analysis of butyrate pathways, also sifting through human fecal metagenomic data to assay the incidence of each pathway in the gut. Alternative pathways represented a very small portion of butyrate production pathways in the human gut, while the acetyl-

CoA marker genes represented the vast majority, strengthening the evidence that the *C. clostridioforme* group members are likely to be human commensals and that the *C. sphenoides* group members are not usually human gut inhabitants.

Patterns in cellobiose utilization begin to emerge when comparing species description papers, but are not robust differentiating factors between the two groups. All taxa in the *C. sphenoides* group ferment cellobiose, whereas all of the species in the *C. clostridioforme* group do not, save for *C. clostridioforme* itself. Trying to tease this apart through genomic analysis proved no more fruitful: *C. citroniae* and *C. clostridioforme* both contained genes annotated as cellobiose phosphorylases. Interestingly, all HMP isolates of *C. clostridioforme* lack a cellobiose phosphorylase, suggesting this may be a niche-specific (e.g. rumen) characteristic for this clade.

A few of the characterizations of *C. sphenoides* group members include a note on or images of multilayer cell walls (Broda et al., 2000; Rogers and Baecker, 1991). In our own experience with genomic DNA extraction of these species, cell lysis necessitated a heat lysis step and long lysozyme incubations (~24hrs) followed by multiple freeze-thaws, which may be explained by this cell wall structure. There are no TEM images from the *C. clostridioforme* group, so the uniqueness of this structure should be further explored. Based on written descriptions, both groups seem to have similar morphologies: fusiform rods, with slight variations in length to width proportion. In the *C. sphenoides* group, the morphology of sporulating cells is distinctive as well as mostly conserved among the group. Spores tend to be subterminal and swell the cell into a wedge-shaped, or ovate, morphology. A more rigid cell wall may be the reason for this tenting effect, rather than the rounded protrusion that usually accompanies spore formation. The *C.*

clostridioforme group rarely forms spores and no mention is made of oddly shaped or swollen cells, though growth medium may be a confounding factor in sporulation.

The most robust differentiator is the presence or absence of a particular bacterial microcompartment in the genome: the CoAT microcompartment (Biddle et al., 2014). Only the *C. sphenoides* group members have this operon, and neither group's outlier, *H. hathewayi* nor *C. symbiosum*, has this microcompartment. This type can most easily be found by locating the vertex shell gene (pfam03319) and searching genes within the operon for coenzyme A transferase (pfam01144), or visa versa. Unfortunately, as the function of this microcompartment is unknown, a simple growth-based assay is not currently possible. But upon elucidation, this could be a quick and uncomplicated test to classify isolates in a laboratory or clinical setting.

Given these phenotypic differences, the phylogenetic distinctness, and the *C. clostridioforme* group's apparent role in human disease, each species group warrants a distinct genus classifier (Marvaud et al., 2011; Ogah et al., 2012; Pequegnat et al., 2013; Song et al., 2003; Warren et al., 2006; Yuli et al., 2005). We therefore propose the creation of two novel genera: *Enterocloster*, formerly the *C. clostridioforme* group, and *Lacriformis*, formerly the *C. sphenoides* group.

3.4 Description of a *Enterocloster*, gen. nov.

Enterocloster (Gr. n. *enteron*, intestine Gr. n. *klôstêr*, spindle, referring to these fusiform rods which inhabit the human gut) refers to a genus of anaerobic fusiform rods which are able to but rarely do form spores. They appear to be nonproteolytic, utilizing simple

carbohydrates for growth. They commonly inhabit the human gut and possibly opportunistic pathogens based on their presence in diseased fluids and tissues.

3.5 Description of *Enterocloster clostridioforme*, comb. nov

Enterocloster clostridioforme (Gr. n. *klôstêr*, a spindle; Gr. dim. suff. *-idion*; N.L. neut. n. *clostridium*, a small spindle; L. suff. *-formis -is -e* (from L. n. *forma*, figure, shape, appearance), -like, in the shape of; N.L. neut. adj. *clostridioforme*, in the form of a small spindle, spindle-shaped) (“LPSN - List of Prokaryotic names with Standing in Nomenclature,” n.d.)

Basonym: *Clostridium clostridioforme*

Other synonym: *Eggerthella clostridiiformis*; *Ristella clostridiiformis*; *Bacteroides clostridiiformis*

The description of *Enterocloster clostridioforme* is identical to that for *Clostridium clostridioforme* (Bryant et al., 1958a; Kaneuchi et al., 1976). The type strain is ATCC 25537^T ([DSM 933](#)^T; JCM 1291^T) (Kaneuchi et al., 1976).

3.6 Description of *Enterocloster aldenense*, comb. nov. *Enterocloster aldenense*

(Al.de.nen'se. N.L. neut. adj. *aldenense*, pertaining to R. M. Alden Research Laboratory and its first patron, Rose M. Alden Goldstein) (Warren et al., 2006)

Basonym: *Clostridium aldenense*

The description of *Enterocloster aldenense* is identical to that of *Clostridium aldenense*.

The type strain is RMA 9741^T (ATCC BAA-1318^T; CCUG 52204^T) (Warren et al., 2006).

3.7 Description of *Enterocloster asparagiforme*, comb. nov. *Enterocloster*

asparagiforme (as.pa.ra.gi.for' me. L. masc.n. asparagus asparagus, L. neut. suffix -forme having the shape of, N.L. neut. adj. asparagiforme having the shape of asparagus stems) (Mohan et al., 2006)

Basonym: *Clostridium asparagiforme*

The description of *Enterocloster asparaagiforme* is identical to the one proposed for *Clostridium asparagiforme*. The type strain is N6^T (DSM 15981^T; CCUG 48471^T) (Mohan et al., 2006).

3.8 Description of *Enterocloster bolteae*, comb. nov. *Enterocloster bolteae* (bolt'ã.e, to honour the American Ellen Bolte who first proposed a bacterial role in late onset autism and stimulated our work in this area) (Song et al., 2003)

Basonym: *Clostridium bolteae*

The description of *Enterocloster bolteae* is identical to the description of *Clostridium bolteae*. The type strain is WAL 16351^T (ATCC BAA-613^T; CCUG 46953^T) (Song et al., 2003).

3.9 Description of *Enterocloster citroniae*, comb. nov. *Enterocloster citroniae*

(Ci.tro'ni.i. N.L. gen. n. *citronii*, named after Diane M. Citron for numerous contributions to clinical anaerobic bacteriology as a clinical microbiologist and educator) (Warren et al., 2006)

Basonym: *Enterocloster citroniae*

The description for *Enterocloster citroniae* is identical to that of *Clostridium citroniae*. The type strain of *C. citroniae* is RMA 16102^T (ATCC BAA-1317^T; CCUG 52203^T) (Warren et al., 2006).

3.10 Description of *Enterocloster lavalense*, comb. nov *Enterocloster lavalense*

(la.va.len'se. N.L. neut. adj. lavalense pertaining to the institution, Université Laval, Québec, Canada) (Domingo et al., 2009)

Basonym: *Clostridium lavalense*

The description for *Enterocloster lavalense* is identical to that of *Clostridium lavalense*. The type strain is CCRI-9842^T (CCUG 54291^T; JCM 14986^T; NML 03-A-015^T) (Domingo et al., 2009).

3.11 Description of *Lacriiformis*, gen. nov.

Lacriiformis (la.cri.for'mis L. n. *lacrima* tear; L.n. *formis* shape, tear-shaped) is a genus of structurally Gram positive, spore-forming anaerobes, becoming characteristically teardrop or wedge shaped upon endospore development. They commonly stain Gram negative, though they often have thick multi-layered cells walls. Anaerobic respiration of sulfurous compounds is common, though none are able to reduce sulfate. The CoAT microcompartment locus is unique to this group of organisms (Biddle et al., 2014). The type species of the genus is *Lacriiformis sphenoides*, originally described in 1917 by Douglas and Hall (Douglas et al., 1917).

3.12 Description of *Lacriiformis sphenoides*, comb. nov.

Lacriformis sphenoides (Gr. n. *sphên sphênos*, wedge; L. suff. *-oides* (from Gr. suff. *eides*, from Gr. n. *eidos*, that which is seen, form, shape, figure), resembling, similar; N.L. neut. adj. *sphenoides*, wedge-shaped.)

Basonym: *Clostridium sphenoides*

Synonyms: *Bacillus sphenoides* Douglas 1917, *Douglasillus sphenoides* Heller 1922, *Plectridium sphenoides* Prévot 1938

The description for *Lacriformis sphenoides* is identical to that of *Clostridium sphenoides* (Douglas et al., 1917; Hall, 1922; McClung et al., 1957; Walther et al., 1977). The type strain is Thorlby^T (ATCC 19403^T; DSM 632^T; NCIB 10627^T; NCTC 507^T). *Lacriformis sphenoides* is the type species of the genus.

3.13 Description of *Lacriformis aerotolerans*, comb. nov.

Lacriformis aerotolerans (Gr. n. *aer aeros*, air, gas; L. part. adj. *tolerans*, tolerating; N.L. part. adj. *aerotolerans*, air-tolerating)

Basonym: *Clostridium aerotolerans*

The description for *Lacriformis aerotolerans* is identical to that of *Clostridium aerotolerans* (Gylswyk and Toorn, 1987). The type strain is XSA62^T (ATCC 43524^T; DSM 5434^T).

3.14 Description of *Lacriformis algidixylanolyticum*, comb. nov.

Lacriformis algidixylanolyticum (al.gi.di.xy.la.no.ly'ti.cum. L. adj. *algidus* cold; Gr. derived L. masc. n. *xylanum* xylan; Gr. adj. *lyticus* dissolving; L. neut. gen. n. *algidixylanolyticum* cold xylan-dissolving) (Broda et al., 2000)

Basonym: *Clostridium algidixylanolyticum*

The description for *Lacriformis algidixylanolyticum* is identical to that of *Clostridium algidixylanolyticum* (Broda et al., 2000). The type strain is SPL73^T (DSM 12273^T; ATCC BAA-156^T).

3.15 Description of *Lacriformis amygdalinum*, comb. nov.

Lacriformis amygdalinum (a.myg.da.li'num. L. neut. adj. amygdalinum made from almonds, referring to the smell of benzaldehyde, which is reduced by the type strain)(Parshina et al., 2003)

Basonym: *Clostridium amygdalinum*

The description for *Lacriformis amygdalinum* is identical to that of *Clostridium amygdalinum* (Parshina et al., 2003). The type strain is BR-10^T (DSM 12857^T; ATCC BAA-501^T).

3.16 Description of *Lacriformis celerecrescens*, comb. nov.

Lacriformis celerecrescens (ce'le. re. cres. cens. L. adv. *celere*, fast; *crescens*, L. pres. part. of verb *crescere*, to grow; M.L. adj. *celerecrescens* reflecting the fast growth of the organism) (Palop et al., 1989)

Basonym: *Clostridium celerecrescens*

The description for *Lacriformis celerecrescens* is identical to that for *Clostridium celerecrescens* (Palop et al., 1989). The type strain is 18A^T (CECT 954^T; DSM 5628^T).

3.17 Description of *Lacriformis guttoideum*, comb. nov.

Lacriformis guttoideum (gut.to.i'de.um L. fem. n. gutta drop, L. suf. -ide, drop-like)
(Gogotova and Vainshtein, 1983)

Basonym: *Desulfotomaculum guttoideum*

The description for *Lacriformis guttoideum* is identical to that of *Desulfotomaculum guttoideum* (Gogotova and Vainshtein, 1983; Stackebrandt et al., 1997), emphasizing that *L. guttoideum* does not reduce sulfate, though it does produce H₂S from other sulfurous compounds (Stackebrandt et al., 1997). The type strain is 50^T (DSM 4024^T; VKM B-1591^T).

3.18 Description of *Lacriformis indolis*, comb. nov.

Lacriformis indolis (N.L. n. *indol*, indole; N.L. gen. n. *indolis*, of indole)

Basonym: *Clostridium indolis*

The description for *Lacriformis indolis* is identical to that of *Clostridium indolis* (McClung et al., 1957). The type strain is 7^T (DSM 755^T; ATCC 25771^T; NCIB 9731^T).

3.19 Description of *Lacriformis saccharolyticum*, comb. nov.

Lacriformis saccharolyticum (Gr. adj. *polu*, many; Gr. n. *sakchâr*, sugar; N.L. neut. adj. *lyticum* (from Gr. neut. adj. *lutikon*), able to loosen, able to dissolve; N.L. neut. adj. *polysaccharolyticum*, degrading several polysaccharides) (Murray et al., 1982)

Basonym: *Clostridium saccharolyticum*

The description for *Lacriformis saccharolyticum* is identical to that for *Clostridium saccharolyticum*. The type strain is WM1 (NRC 2533^T; DSM 2544^T; ATCC 35040^T) (Murray et al., 1982).

3.20 Description of *Lacriiformis xylanolyticum*, comb. nov.

Lacriiformis xylanolyticum (xy.lan.o.ly' ti. cum. Gr. n. *xylanosum*, xylan; Gr. adj. *lyticus*, dissolving; L. neut. adj. *xylanolyticum*, xylan dissolving) (Rogers and Baecker, 1991)

Basonym: *Clostridium xylanolyticum*

The description for *Lacriiformis xylanolyticum* is identical to that for *Clostridium xylanolyticum* (Rogers and Baecker, 1991). The type strain is ATCC 49623^T (DSM 6555^T).

CHAPTER 4

RECLASSIFICATION OF CLOSTRIDIUM POLYSACCHAROLYTICUM, CLOSTRIDIUM POPULETI, CLOSTRIDIUM HERBIVORANS, AND CLOSTRIDIUM PHYTOFERMENTANS AS CELLULOSPECIUM POLYSACCHAROLYTICUM, GEN. NOV., COMB. NOV., CELLULOSPECIUM POPULETI, COMB. NOV., CELLULOSPECIUM HERBIVORANS, COMB. NOV., AND LESCHINIA PHYTOFERMENTANS, GEN. NOV., COMB. NOV

4.1 Abstract

Incongruence between taxonomic classification of *Clostridium* species and sequence-based phylogenies has left many *Clostridia* lying outside of the genus *Clostridium sensu stricto*, and even the Family *Clostridiaceae*. Misclassified *Clostridium* spp. make up approximately one-third of the described species in the *Lachnospiraceae* and are in need of taxonomic revision to improve scientific communication. With a large recent expansion of the Family *Lachnospiraceae* and the increasing interest in these taxa as plant-degraders in the digestive tract and in soil or bioprocessing contexts, it is increasingly important to rectify the systematics of this group. Here we propose creation of two novel genera to house four misclassified *Clostridium* spp. One genus, *Cellulospecium*, gen. nov., is to include *Cellulospecium herbivorans*, comb. nov., *Cellulospecium populeti*, comb. nov., and *Cellulospecium polysaccharolyticum*, comb. nov. The second genus, *Leschinia*, gen. nov., houses *Leschinia phytofermentans*, comb.

nov. We also describe a novel species within the recently reclassified genus *Anaerocolumna*, *Anaerocolumna spermata*, sp. nov.

4.2 Introduction

Many of the species classified as *Clostridia* were described prior to DNA sequencing technologies, leading to a disconnect between taxonomy and phylogeny once sequencing was undertaken. The taxonomic genus *Clostridium* has representatives in several families based on phylogeny, though they all belong to the family *Clostridiaceae* due to their taxonomy. Some researchers have taken steps to correct this issue (Cai and Dong, 2010; Kaur et al., 2014; Liu et al., 2008), but the rectification of this genus's taxonomy is far from over.

Phylogenetically, the *Lachnospiraceae* family contains the genera *Roseburia*, *Marvinbryantia*, *Cellulosyliticum*, *Coprococcus*, *Blautia*, *Hungatella*, *Lactonifactor*, *Hespellia*, *Robinsoniella*, *Pseudobutyrvibrio*, *Lachnobacterium*, *Dorea*, *Moryella*, *Shuttleworthia*, *Oribacterium*, *Lachnoanaerobaculum*, *Butyrvibrio*, *Syntrophococcus*, *Lactobacillus*, *Lachnospira*, *Parasporobacterium*, *Anaerostipes*, *Sporobacterium*, *Howardella*, *Catonella*, *Acetitomaculum*, *Anaerosporobacter*, *Anaerocolumna*, *Anaerobium*, *Fusicatena*, *Herbinix*, *Murimonas*, *Eisenbergiella*, and the misclassified genera *Eubacterium*, *Ruminococcus*, and *Clostridium*. While this is a rather genera-rich family, the issue is in the interleaving of disparate taxonomic genera. Moreover, the genera *Clostridium*, *Eubacterium*, and *Ruminococcus* are not unique to the *Lachnospiraceae*, and are problematic due to their own eponymous families.

In medical microbiology and human microbiome research, the *Lachnospiraceae* are still often referred to as the *Clostridium coccooides* group, despite *C. coccooides* reclassification in 2008 as *Blautia coccooides* (Kurakawa et al., 2015; Liu et al., 2008; Touyama et al., 2015; Vermeiren et al., 2012). The NCBI has, to the confusion and dismay of many users, altered the taxonomy of the *Clostridia* in their databases (Yutin and Galperin, 2013), due to the taxonomic problems with the group. The *Clostridium* species in the *Lachnospiraceae* are now referred to as *Lachnoclostridium*, while other *Clostridia* may be called *Peptoclostridium*, *Ruminiclostridium*, or *Erysipelatoclostridium*, depending on their phylogenetic placement. However, these revisions were never approved, so the same organism can now have two contemporary names, depending on the source. Between the taxonomy, vernacular, and database specific revisions, there is an even less navigable taxonomic space for these organisms.

4.3 Results

There are approximately 30 misclassified *Clostridium* spp. within the family *Lachnospiraceae*. Two-thirds of these fall into three clusters, one which includes *Clostridium clostridioforme*, one which contains *Clostridium sphenoides*, and another which is much broader phylogenetically that contains *Clostridium aminovalericum* (Figure 13). This last cluster is comprised of *Clostridium phytofermentans*, *Clostridium polysaccharolyticum*, *Clostridium populeti*, *Clostridium herbivorans*, *Eubacterium xylanohilum*, *Anaerosporobacter mobilis*, *Clostridium xylanolyticum*, *Clostridium jejuense*, and *Clostridium aminovalericum*. Very recently, the latter three species have been placed in a novel genus, *Anaerocolumna* (Ueki et al., 2016). And within the last two years, the genera *Herbinix* and *Mobilitalea* have been added to this cluster as well. The

Clostridium spp. in this clade were originally classified based on the classical *Clostridial* definition: non-sulfate reducing, anaerobic sporeformers. Because of DNA sequencing, today we know that this definition is inclusive of several families and is not phylogenetically restricted.

While the *Clostridium* species in this clade were placed in the same taxonomic genus, upon careful examination, it becomes clear that these organisms are not so closely related phylogenetically or phenotypically. There are differences in cell morphology, substrate utilization categories, Gram staining, sulfur utilization, and fermentation products. The genetic distance between these taxa lends little clarity, as ANI values are below the method's threshold (Goris et al., 2007) (Table 9).

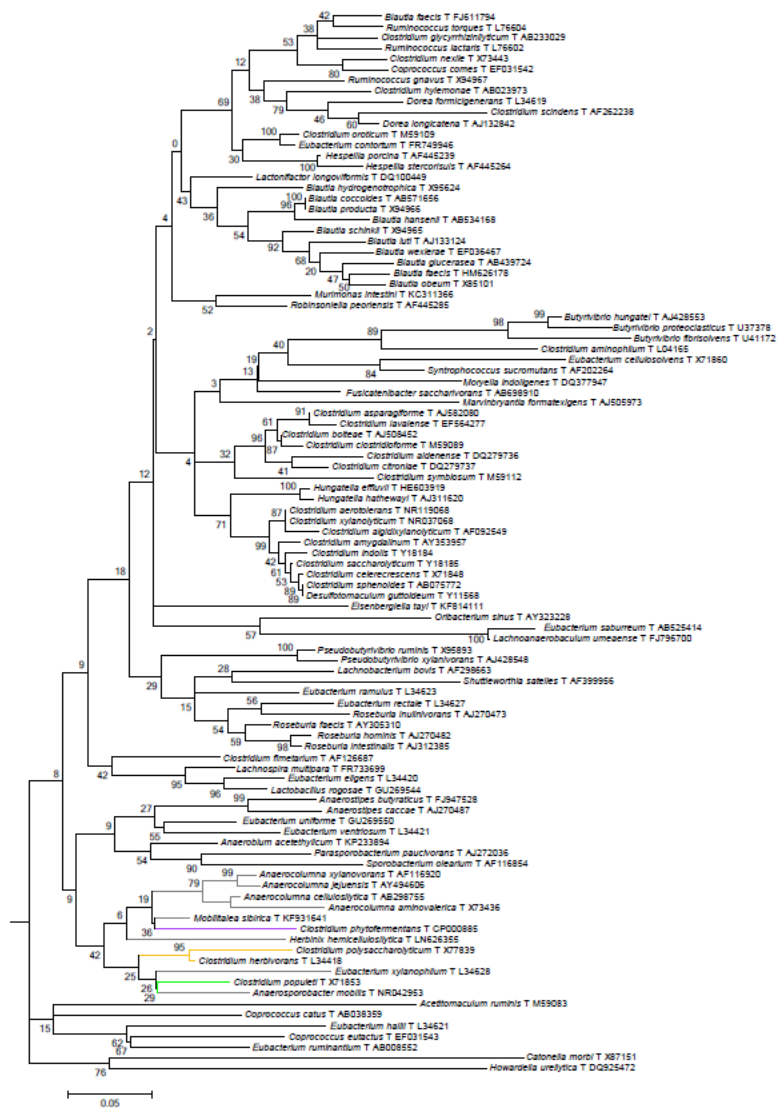


Figure 13. 16S rRNA gene based phylogeny. The taxa of interest are colored based on their grouping. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The tree with the highest log likelihood (-13851.9204) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.6424)). The rate variation model allowed for some sites to be evolutionarily ([+I], 62.1370% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 105 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1001 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013). Not all sequences used in making the tree are shown.

Table 9. Average nucleotide identities (ANIb) of taxa in the cluster. Most pairs have very low ANIb values (<70%) as well as small fractions of the genomes that could be aligned (in brackets). *A. jejuensis* and *A. xylanovorans* have much higher ANIb values and percent of genomes aligned, supporting their inclusion in the same genus. *C. populeti* and *C. polysaccharolyticum* have genomes which are ~35% analogous, highlighting their relative similarity to each other.

	<i>Clostridium phytofermentans</i>	<i>Anaerosporbacter mobilis</i>	<i>Anaerocolu mna aminovaleric a</i>	<i>Anaerocolu mna jejuensis</i>	<i>Clostridium polysaccharolyti cum</i>	<i>Clostridium populeti</i>	<i>Anaerocolu mna xylanovorans</i>	<i>Herbinix hemicellulosilytica</i>
<i>Clostridium phytofermentans</i>	-	68.93 [23.27]	70.51 [25.80]	67.71 [21.42]	68.02 [15.43]	67.68 [17.06]	67.81 [20.82]	67.13 [15.99]
<i>Anaerosporbacter mobilis</i>	68.97 [20.90]	-	68.77 [21.09]	67.04 [20.38]	68.35 [17.44]	67.77 [19.95]	67.20 [20.33]	66.44 [14.41]
<i>Anaerocolumna aminovaleric a</i>	70.39 [26.12]	69.02 [23.33]	-	69.35 [28.04]	68.42 [16.63]	68.08 [19.16]	69.57 [28.46]	67.80 [18.84]
<i>Anaerocolumna jejuensis</i>	67.31 [16.25]	66.77 [16.89]	69.13 [20.40]	-	66.94 [11.74]	67.04 [14.21]	80.18 [42.44]	67.38 [12.66]
<i>Clostridium polysaccharolyti cum</i>	68.12 [20.95]	68.38 [24.86]	68.37 [20.89]	67.09 [21.14]	-	69.66 [35.47]	67.30 [20.52]	66.53 [17.66]
<i>Clostridium populeti</i>	67.94 [18.35]	68.01 [23.25]	68.16 [20.13]	67.32 [20.52]	69.41 [28.44]	-	67.72 [21.03]	66.64 [15.83]
<i>Anaerocolumna xylanovorans</i>	67.62 [18.66]	67.07 [19.09]	69.40 [24.96]	80.39 [50.50]	67.07 [13.58]	67.63 [16.49]	-	67.21 [14.84]
<i>Herbinix hemicellulosilytica</i>	66.84 [23.84]	66.69 [23.73]	67.88 [27.36]	67.58 [24.93]	66.42 [19.71]	66.72 [21.96]	67.47 [24.98]	-

The phylogeny based on the 16S rRNA gene shows how distantly related most of the taxa in this clade are, with *Anaerocolumna xylanovorans* and *Anaerocolumna jejuensis* being the closest relatives in this clade at a distance of ~3.0%. The next closest relatives are *C. herbivorans* and *C. polysaccharolyticum* at a distance of 3.4%. The remaining taxa are between ~4.7% and ~11.0% different from each other, with the mean distance being 8.01% (median 8.23%). Most of these are arguably distinct enough to be classified as separate genera.

Thanks to the recent sequencing of several of these genomes, we are able to do much more in-depth analyses of the evolutionary history and relatedness of these taxa. A comparison of the *rpoβ* gene sequences highlights the distances between these organisms better than the 16S rRNA gene (Figure 14). The *rpoβ* gene has been shown to be a robust proxy for whole genome similarity (Adékambi et al., 2008). Quite surprisingly this method places *E. xylanophilum* with the *Coprococcus* spp. though this organism does not fit the phenotypic description of this genus. There is a rearrangement of *C. populeti* from branching with *C. phytofermentans* to branching with *C. polysaccharolyticum*, while *A. xylanovorans*, *A. jejuense*, and *A. aminovalericum* are still clustering together regardless of methodology.

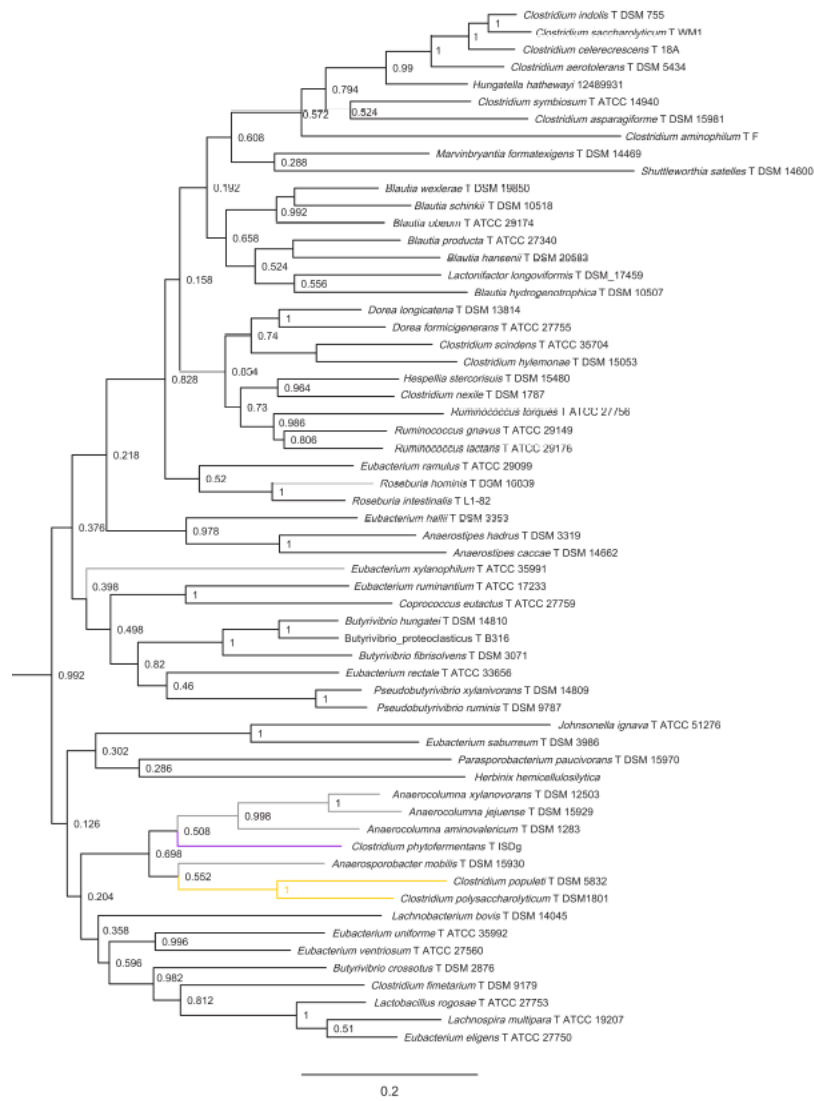


Figure 14. rpoB gene based phylogeny. Taxa are colored based on grouping, retaining the colors as in Figure 13, except that *C. populeti* is now grouping with *C. polysaccharolyticum* (gold). The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The tree with the highest log likelihood (-71686.8892) is shown. The fraction of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.6590)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.0000% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 48 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 2933 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013, p. 6).

For greater resolution, we also constructed a phylogeny from a 22 conserved genes (Figure 15), most of them encoding ribosomal proteins (Lang et al., 2013). This concatenated gene tree is in perfect agreement with the *rpoβ* based tree, sharing equally strong bootstrap values. Together these two methods help clarify some of the weaker relationships shown in the 16S rRNA based tree. *C. populeti* is most affected by this greater resolution, being grouped with *C. polysaccharolyticum* rather than with *E. xylanophilum*. The position of *C. aminovalericum* remains with *C. xylanovorans* and *C. jejuense* regardless of method, but the higher resolution phylogenies strengthen this placement. Based on these placements, we propose the development of three novel genera to rectify the taxonomy of these species.

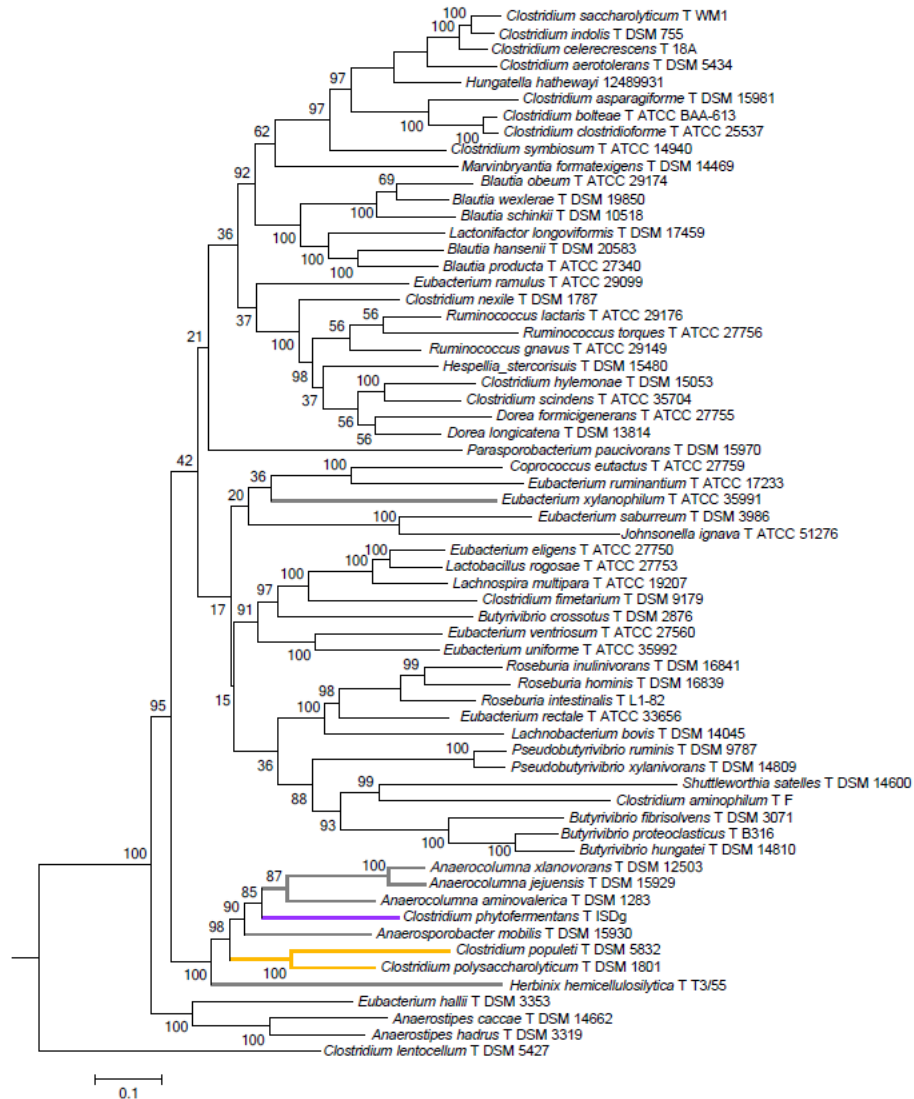


Figure 15. Concatenated gene phylogeny using 22 conserved genes (Lang et al., 2013). Due to some genomes missing genes, ribosomal proteins S2p and S11p were left out of the analysis, leaving 22 rather than 24 genes in the final alignment. Branches are colored based on grouping of taxa. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The tree with the highest log likelihood (-427761.6396) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.7887)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 24.8442% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 66 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 12714 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013, p. 6).

4.4 Description of *Cellulospecium* gen. nov.

Cellulospecium (L. n. *cellulo* cellulose; L. n. *specialita* specialist; cellulose-specialists, referring to the ability to degrade cellulose and a limited range of other carbohydrates) is a genus of motile sporeforming obligate anaerobes. They are thick short rods which produce butyrate and degrade cellulose, though they are able to utilize hemicellulolytic components as well. They have a Gram positive cell wall structure, though they may stain negative. The type species is *Cellulospecium polysaccharolyticum*^T.

4.5 Description of *Cellulospecium herbivorans*^T, comb. nov.

Cellulospecium herbivorans^T (L. fem. n. *herba*, grass, herb, a green plant; L. v. *vorare*, to devour; N.L. part. adj. *herbivorans*, devouring plants.) (Varel et al., 1995)

Basonym: *Clostridium herbivorans*^T

The description for *Cellulospecium herbivorans*^T is identical to that for *Clostridium herbivorans*^T (Varel et al., 1995). The type strain is 54408^T (ATCC 49925^T; DSM 14428^T).

4.6 Description of *Cellulospecium populeti*^T, comb. nov.

Cellulospecium populeti^T (po.pu'le.ti. L. n. *populetum* poplar wood; L. gen. n. *populeti* of poplar wood) (Sleat and Mah, 1985)

Basonym: *Clostridium populeti*^T

The description of *Cellulospecium populeti*^T is identical to that for *Clostridium populeti*^T.

The type strain is 743A^T (ATCC 35295^T; DSM 5832^T).

4.7 Description of *Cellulospecium polysaccharolyticum*^T, comb. nov.

Cellulospecium polysaccharolyticum^T (Gr. adj. *polu*, many; Gr. n. *sakchâr*, sugar; N.L. neut. adj. *lyticum* (from Gr. neut. adj. *lutikon*), able to loosen, able to dissolve; N.L. neut. adj. *polysaccharolyticum*, degrading several polysaccharides.) (“LPSN - List of Prokaryotic names with Standing in Nomenclature,” n.d.)

Basonym: *Clostridium polysaccharolyticum*^T

Other synonym: *Fusobacterium polysaccharolyticum*^T

The description for *Cellulospecium polysaccharolyticum*^T is identical to that for *Clostridium polysaccharolyticum*^T (Gylswyk, 1980; Gylswyk et al., 1980). The type strain is B^T (ATCC 33142^T; DSM 1801^T).

4.8 Description of *Leschinia*, gen. nov.

Leschinia (N. L. *Leschinia*, referring to the distinguished scientist Dr. Susan B. Leschine, who has made significant contributions to our knowledge of microbial breakdown of complex polysaccharides) describes a genus of thing rod-shaped, anaerobic sporeformers with broad substrate utilization capabilities, including simple and complex carbohydrates. The type species is *Leschinia phytofermentans*^T.

4.9 Description of *Leschinia phytofermentans*^T, comb. nov.

Leschinia phytofermentans^T (phy.to.fer.men«tans. Gr. n. *phyton* plant; L. part. adj. *fermentans* fermenting;

N.L. part. adj. *phytofermentans* plant-fermenting, referring to the wide range of plant polysaccharides that this organism is capable of utilizing as growth substrate) (Warnick et al., 2002)

Basonym: *Clostridium phytofermentans*^T

The description for *Leschinia phytofermentans*^T is identical to that for *Clostridium phytofermentans*^T (Warnick et al., 2002). The type strain is ISDg^T (ATCC 700394^T; DSM 18823^T).

4.10 Description of *Anaerocolumnna spermata*, sp. nov.

Anaerocolumnna spermata (L. n. *sperma*, sperm cell; *spermata* resembling a sperm cell) is an motile, spore-forming obligate anaerobe. *A. spermata* was isolated from a forest-soil seeded microcosm in Montague, MA, USA, which was maintained anaerobically on switchgrass at 30 °C. *A. spermata* is Gram variable, but when fixed with methanol, is evenly Gram negative (early log phase). It is a fast-growing rod-shaped bacterium, usually forming filamentous chains which form tangled nets microscopically. Subterminal oval to ovoid spores develop which swell the wall, leading to a sperm-like morphology. It can utilize a wide variety of carbohydrates for growth: DL-arabinose, cellobiose, chitin, fructose, galactose, gentiobiose, glucose, inulin, laminarin, mannose, maltose, maltodextrin, maltotriose, melibiose, pectin, raffinose, rhamnose, salicin, starch, sucrose, trehalose, xylan, and xylose, as well as the methylated sugars B-Methyl-D-Xylose and 10 methyl α -D-Glucopyranoside. Ribose, sorbose, melezitose did not support growth. It was unable to grow using short chain fatty acids: acetate, butyrate, formate, lactate, propionate, pyruvate, succinate, or sugar alcohols: adonitol, dulcitol, sorbitol,

xylitol, except mannitol. Casamino acids did not support growth. It grows over a temperature range from as low as 10 °C to 40 °C, growing optimally at 37-40 °C. Its pH range for growth is 6 to 9, with an optimum of 8. It does not reduce sulfur compounds (sulfite, sulfate, thiosulfate, cysteine) or nitrate. It is urease, catalase, oxidase, gelatinase, and indole negative. The closest relative, *Clostridium jejuense*, shares 96.8% 16S rRNA gene identity, though this is a partial sequence, so the true similarity may be lower. The type strain is KNHs205^T (ATCC TSD-29^T; DSM 100533^T).

CHAPTER 5

KINEOTHRIX ALYSOIDES, GEN. NOV., SP. NOV., A SACCHAROLYTIC, BUTYRATE-PRODUCING NITROGEN FIXER WITHIN THE FAMILY LACHNOSPIRACEAE

5.1 Abstract

Kineothrix alysoides strain KNHs209^T (ATCC TSD-26^T, DSMZ 100556^T) was isolated from a forest soil-seeded microcosm, maintained anaerobically on switchgrass. The organism is a highly motile rod, often forming long filamentous chains which are easily observed moving under the microscope. Its closest relatives lie within the *Lachnospiraceae*/Cluster XIVa of the Clostridia, though it is easily distinguishable based on its morphology and flagellar arrangement. Whole genome sequencing enabled development of minimal medium, and also suggested that the organism is capable of fixing nitrogen. Its wide variety of growth substrates is highlighted in its concomitantly high number of encoded chemotaxis receptors (45, the highest in the family *Lachnospiraceae*). It is capable of utilizing a wide variety of carbohydrates, but not cellulose or xylan. Fermentation products include formate, acetate, and butyrate; it does not reduce sulfur compounds or nitrate. *K. alysoides* grows optimally at 35-40 °C and pH 7. Based on its low abundance in the microcosm and evasion of discovery until now, it seems that it is a member of the rare biosphere. Its placement in the well-known gut commensal family *Lachnospiraceae*, coupled with its saccharolytic capabilities and butyrate production hint that it may even be a rare member of the gut microbiome.

5.2 Introduction

It is widely accepted that most of the microbial diversity on the planet is yet to be discovered (Rappé and Giovannoni, 2003; Sogin et al., 2006; Stevenson et al., 2004), with approximately 300,000 species of bacteria estimated from current rarefaction curves and only about 4,000 characterized taxa. Much of this diversity is thought to be present in the rare biosphere, the low-abundance members that constitute most of the richness in a community. While we can make estimates of global bacterial diversity based on sequencing surveys, the function of these microbes cannot be determined without isolating and observing them.

We developed and maintained a soil-seeded microcosm on switchgrass for approximately three years, sequencing the progenitor soil sample and subsequent transfers to determine the stability and structure of the microcosm. After sequencing data revealed bacteria from unexplored genomic space, we began isolations. One of the isolates appeared to represent a novel genus, based on 16S rDNA sequence similarity to known taxa, and deep branching on phylogenetic trees. The organism, KNHs209, is a member of the rare biosphere of our microcosm, at ~0.04% relative abundance based on phylotag sequencing. In the initial soil sample, however, only one matching sequencing was found (~0.001% of all reads). Here we describe the phenotypic characterization alongside a genomic analysis.

5.3 Methods

Isolation and Maintenance

One-week old Microcosms were vortexed to mix; serial dilutions were made in sterile water, which were then streaked onto GS2 plates. Colonies were restreaked until they appeared pure by colony morphology, then observed microscopically to ensure one cell morphotype was present. The 16S rRNA gene was amplified and sequenced using 27F and 1492R universal bacterial primers to test for purity and to analyze phylogeny.

KNHs209 was isolated on GS2 agar medium supplemented with 0.3% cellobiose. The media contained, per liter: 3 g Na₂HPO₄, 1.77 g NaH₂PO₄, 2.18g K₂HPO₄, 3.4 g KH₂PO₄, 2.1 g urea, 3 g sodium citrate•2H₂O, 6 g yeast extract (Fisher BP9727-500), 2 g L-cysteine hydrochloride monohydrate, 1 mL 0.1% Resazurin, 100 mL GS2 salt solution, 900 mL ultrapure water, 3 g cellobiose. 100 mL of GS2 salts contained: 1 g MgCl₂•6H₂O, 0.15 g CaCl₂•2H₂O, 0.00125 g FeSO₄•7H₂O in ultrapure H₂O. Solidified plates were equilibrated in the anaerobic chamber (atmosphere containing ~3% H₂, ~10% CO₂, and ~87% N₂) at least 8 hours before use.

Growth based assays were carried out in liquid GS2 medium. Medium was prepared per the Hungate technique (HUNGATE, 1950). Unless otherwise stated, growth assays were carried out at pH 7 and 30 °C. For pH alterations, HCl was used to acidify, NaOH was used to alkalify up to pH 8.5, 10% NaCO₃ was to alkalify for pHs 9.0 and 9.5 (Mitsui et al., 2010). Temperatures were tested in 5 °C increments except for the addition of 37 °C; pH values were tested in increments of 0.5.

Freezer stocks were made from a mid-late log phase culture, first iced for ~10 minutes, then combined in equal parts (0.75 mL) with 30% sterile glycerol in 2.0 mL cryovials and stored at -80 °C. Viability and purity of freezer stocks were checked by

inoculating into GS2 with cellobiose and observing the resultant culture microscopically with a phase contrast microscope at 400X and 1000X magnification.

The defined medium developed for KNHs209 was dubbed MM9 and contained (per L): 2 g NaH₂PO₄, 10 g K₂HPO₄, 1 g L-cysteine hydrochloride monohydrate, 0.05 g L-proline, 40 mL XT2 solution, 10 mL BTE solution, and 1 mL 0.1% Resazurin. 9.4 mL were aliquoted per tube. After autoclaving, 0.3 mL 10% cellobiose and 0.2 mL CPV4 were added to each tube. XT2 contained (per 100 mL): 0.25 g xanthine, 0.25 g thymine, and 1 mL 6N NaOH. BTE contained (per L): nitriolotriacetic acid, 3 g MgSO₄·7H₂O, 0.5 g MnSO₄·4H₂O, 1 g NaCl, 0.1 g FeSO₄·7H₂O, 0.1 g CoCl₂·6H₂O, 0.1 g CaCl₂, 0.1 g ZnSO₄·7H₂O, 0.01 g CuSO₄·5H₂O, 0.01 g AlK(SO₄)₂, 0.01 g H₃BO₃, 0.01 g Na₂MoO₄, 0.03 g NiSO₄·6H₂O, 0.02 g Na₂SeO₃, and 0.02 g Na₂WO₄. CPV4 contained (per 100 mL): 4 mg p-aminobenzoic acid, 0.1 mg biotin, 0.6 mg folinic acid, 8 mg nicotinamide, 0.5 g pantethine, 0.4 mg pyridoxal, 3.0 mg riboflavin, and 1.0 mg thiamine.

Biochemical Assays

Testing for catalase was performed by adding 3% hydrogen peroxide to a colony which had been transferred to a clean microscope slide. The oxidase test was performed using BBL Oxidase (Becton, Dickinson and Company, item 261181) as per the instructions. The indole test was performed using Kovac's aldehyde reagent (Ricca Chemical, item 4260-16); Resazurin was left out of the medium for this test, so that its color did not confound test results.

For assaying anaerobic respiration of sulfur compounds, GS2CB (1.5% agar) was supplemented with either cysteine, sulfate, sulfite, or thiosulfate, and ferric ammonium

citrate. The isolate was streaked onto the plates in the anaerobic chamber, which were sealed with strips of Parafilm® to minimize dehydration.

Nitrate reduction was also tested as per manufacturer's instructions (Becton, Dickinson and Company, item 211299), followed by a final step where elemental zinc crystals were added to check for the presence of unreduced nitrates by reducing them to nitrite. Resazurin was also left out of this medium, to prevent confounding color changes.

Motility was tested in sulfate indole motility (SIM) semisolid agar stabs and also observed microscopically. The Gram stain was carried out as directed (Becton, Dickinson and Company, item 212539) but with methanol rather than heat fixation (Mangels et al., 1984; Minnerath et al., 2009). The KOH test was performed according to (Halebian et al., 1981).

Fermentation Product Analysis

Samples were filtered through a 0.22 µm PVDF syringe-driven filter (Millipore SLGVX13NL) then frozen at -20 °C. Samples were run through a Shimadzu Prominence liquid chromatograph (LC-20AD) equipped with a refractive index detector (RID-10A) and a Biorad Aminex HPX-87 ion exclusion column (125-0140). 20 µL injection volumes were run through the system in 5 mM H₂SO₄ running buffer at a flow of 0.6 mL per minute for 30 minutes per sample. Standards included ethanol, acetate, propanol, butyrate, lactate, propionate, succinate, and formate.

Gas chromatography was run on a Shimadzu GC-8A with a TCD detector and Supelco 60/80 Carboxen 1000 column, with argon as the carrier gas. Standards included carbon dioxide, methane, nitrogen, and hydrogen.

Electron Microscopy

Scanning electron microscopy (SEM) was done on a JEOL JCM-5000; samples were prepared as described elsewhere (Fox and Demaree, 1999) and sputter coated with gold. Transmission electron microscopy (TEM) samples were prepared by drying then negative staining culture with uranyl acetate. Prepared samples were visualized on a JEOL 100X transmission electron microscope.

Phylogenetic Analyses

Lachospiraceae 16S rDNA sequences were taken from the RDP, using only those that represented type strains with long, good quality sequences. The species *Desulfotomaculum guttoideum* is not included as a type strain in the RDP, but it was included as well. The 16S rDNA sequence for *Eisenbergiella tayi* was retrieved from NCBI. The 16S rDNA sequence for KNHs209 was taken directly from the genome (CC90DRAFT_4253).

Sequences were aligned in MEGA6 using ClustalW1.6 with the default parameters. To determine the best probable method for phylogenetic tree construction of this group, JModelTest was used. Using the resulting parameters, a phylogenetic tree was constructed in MEGA6 using the Maximum Likelihood method and the General Time Reversible (GTR) model with gamma distributed rates (4 categories) and invariant sites (G+I). Gaps were treated as complete deletions, and 500 bootstrap replicates were performed.

Average nucleotide identity (ANI) was calculated using JSpecies blast-based ANI (ANiB) calculation (Goris et al., 2007; Richter et al., 2015; Richter and Rosselló-Móra, 2009).

Genome Extraction

KNHs209 was grown in 50mL of GS2 medium with 0.5% cellobiose. The culture was harvested for genomic DNA in early stationary phase: the flask was put on ice for ~10 minutes prior to pelleting. The JGI DNA extraction protocol was followed thereafter, except that the concentrations of both lysozyme and proteinase K were reduced to 20 mg/mL and incubated 1 hour each.

Genome Sequencing and Assembly

The draft genome of KNHs209 was generated at the DOE Joint genome institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology (Eid et al., 2009). A Pacbio SMRTbell™ library was constructed and sequenced on the PacBio RS platform, which generated 145,422 filtered subreads totaling 393.4 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov>. The raw reads were assembled using HGAP (version: 2.1.1)(Chin et al., 2013). The final draft assembly contained 3 contigs in 3 scaffolds, totalling 4.7 Mbp in size. The input read coverage was 51.7X.

Genome Annotation

Genes were identified using Prodigal (Hyatt et al., 2010), followed by a round of manual curation using GenePRIMP (Pati et al., 2010). The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (Lowe and Eddy, 1997) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (Pruesse et al., 2007). Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (Nawrocki et al., 2009). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform (*The Integrated Microbial Genomes (IMG) platform.*, n.d.) developed by the Joint Genome Institute, Walnut Creek, CA, USA (Markowitz et al., 2009).

5.4 Results

KNHs209 stains Gram negative (Figure 16A), though the KOH test (Halebian et al., 1981) suggests a Gram positive cell wall structure, a common conundrum within the *Lachnospiraceae*. By 96 hours, colonies on GS2-cellobiose are approximately 1mm in diameter; round, translucent off-white, and glossy with entire margins, taking on a micro-undulating margin as the colonies age. Slightly sub-terminal spores that swell the wall are produced, (Figure 16G), which lend the cells a sperm-like morphology. Single cells are straight rods (0.5 to 0.6 μm by 3.3 to 6.7 μm) with rounded ends (Figure 16). When cells

are in chains or dividing, the length can exceed this however. Chains are often observed, and regularly reach ~30 μm in length, becoming somewhat filamentous (Figure 16).

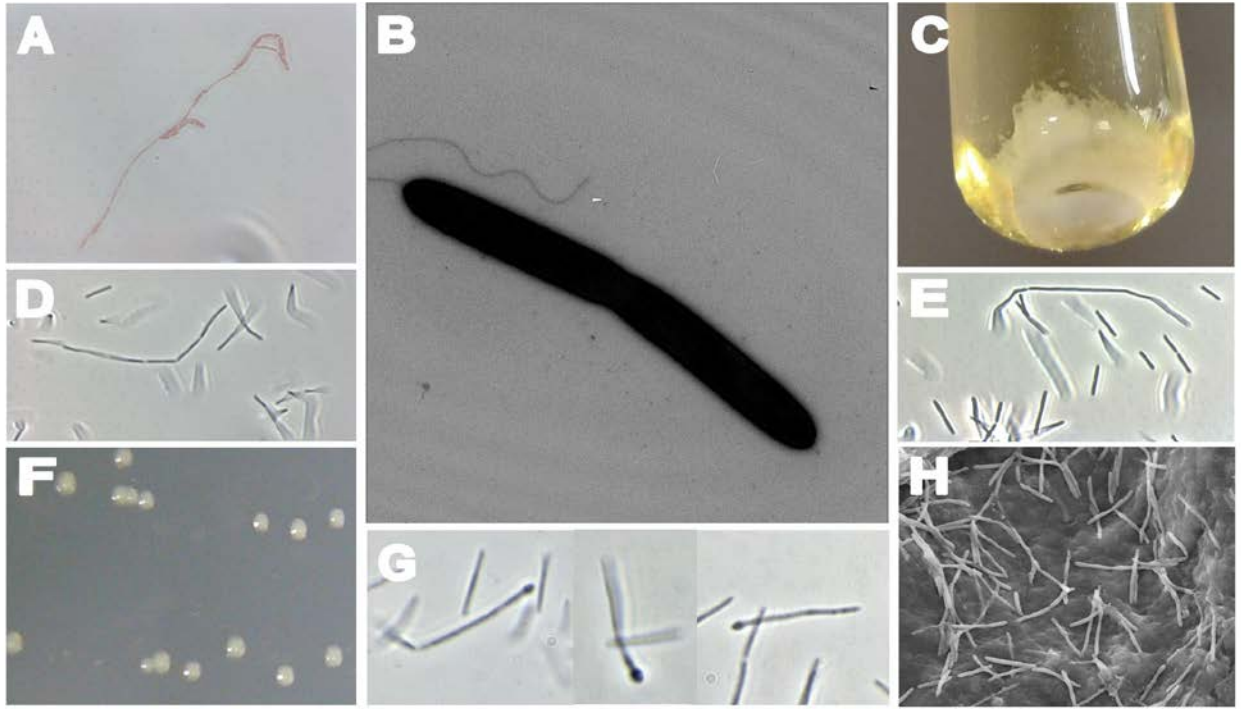


Figure 16. Morphology of KNHs209. All cultures were grown in GS2 with cellobiose at 30 °C. A) Gram stain (1000X) B) negative stained TEM showing a single polar flagellum C) an unagitated culture, showcasing the tendency to form a biofilm D-E) Vegetative cells (1000X) where long chains are a common occurrence F) Colonies after 48 hours G) Typical morphology of sporulating cells H) SEM of vegetative cells

Cells are highly motile, easily dispersing through 0.5% agar stabs. Motility is conferred by a single polar flagellum. Interestingly, motility is maintained even when in chain formation. Despite this, in broth culture, KNHs209 tends to grow in a robust biofilm (Figure 16C), which sometimes floats (possibly due to gas production). Vortexing results in a heterogeneous mixture of flocs and planktonic culture. KNHs209 has a typical Gram-positive flagellar assembly, notably lacking the outer-membrane-traversing FlgF, FlgH, and FlgI proteins, which make up the proximal rod, the L ring, and the P ring respectively (Chen et al., 2011; Pallen et al., 2005). The genome highlights the

organism's broad substrate repertoire and motility with over 3% of genes related to chemotaxis and motility. Forty-five of these genes are chemoreceptors (COG0840, methyl-accepting chemotaxis protein), the most of any sequenced *Lachnospiraceae* (exclusive range 0-32). KNHs209 also has a locus encoding a type IV pilus assembly (CC90DRAFT_1101-1108).

Catalase, indole, oxidase, urease, gelatinase, and citrate tests were all negative. KNHs209 does not encode a tryptophanase, which is as expected based on its negative indole reaction. Anaerobic respiration of nitrate, sulfate, sulfite, and thiosulfate was not observed.

The genome sequence suggested that KNHs209 can fix nitrogen gas due to the presence of *nifDHK* (CC90DRAFT_0542-0543, CC90DRAFT_0550, CC90DRAFT_0855-0856). Dos Santos et al. (Dos Santos et al., 2012) recommended a minimum set of six genes be present when scanning genomes for nitrogen fixation: *nifHDK* and *nifBEN*. The overall geneset conferring nitrogen fixation to an organism varies from species to species, but these genes are mostly conserved. NifDK make up the dinitrogenase enzyme, while NifH is the dinitrogenase reductase. Typically, an iron-molybdenum cofactor (FeMoco) is required for function of dinitrogenase; this cofactor is in part produced by NifBEN. However, some genomes lack *nifEN* and an FeMoco binding site within the nitrogenase, indicating an iron-sulfur (FeS) active site, rather than one including molybdenum (Dos Santos et al., 2012). Interestingly, many of the genes flanking the two *nifDK* sets in KNHs209 are involved in transport of sulfur containing organic compounds, such as taurine, methionine, and cysteine, which may contribute to the synthesis of FeS clusters.

Comparison of nitrogen fixation loci in KNHs209 to other organisms

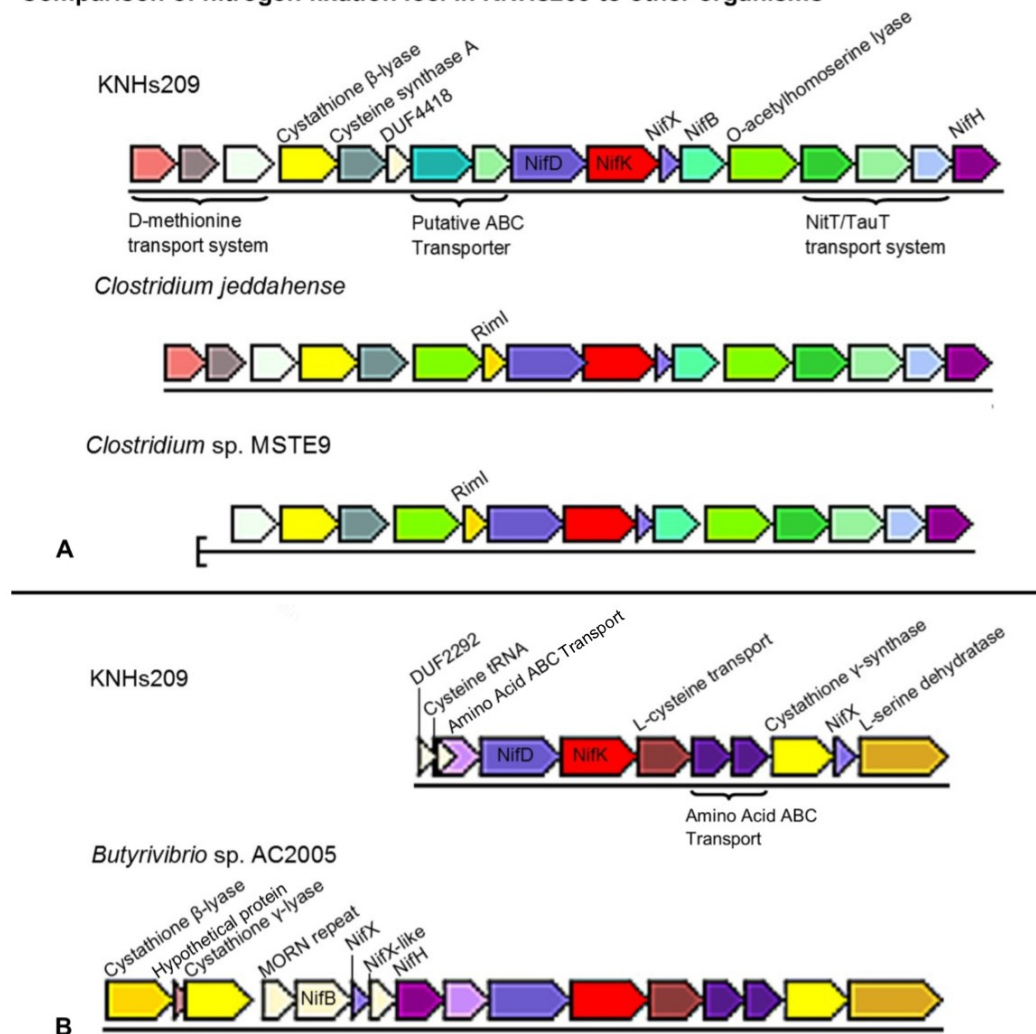


Figure 17. COG-based gene neighborhood maps comparing the nitrogenase loci, based on BLASTP (NR) best hits of *nifD* and *nifK* amino acid sequences. A) Comparison between CC90DRAFT_0534-0550 (KNHs209), Ga0055386_102436-102421 (*C. jeddahense*), and HMPREF1141_2537-2550 (*C. MSTE9*). B) Comparison between CC90DRAFT_0851-0862 (KNHs209) and G602DRAFT_03543-3558 (*B. AC2005*). After the reference map for KNHs209, only genes which differ among the analogous loci are labelled.

There is surprisingly low sequence similarity between the copies of *nifD* and *nifK* in the genome. Pairwise BLASTP results show a 45% identity between the *nifD* copies and 42% identity between the *nifK* copies. A BLASTP query against all recorded protein

sequences resulted in much higher hits, however: the *nifDK* at CC90DRAFT_0855-0856 showed the highest similarity (65-70%) to *Butyrivibrio* strain AC2005 from the Hungate 1000 project, while the *nifDK* at CC90DRAFT_0542-0543) was most similar to human gut microbes *Clostridium* spp. JCD (now *Clostridium jeddahense*) and MSTE9. Syntenic maps of these neighborhoods are shown in Figure 17. Unfortunately, little to no phenotypic data is available for most of the strains; the description of *C. jeddahense* does not include any mention of nitrogen fixation (Lagier et al., 2014).

We tested KNHs209's ability to fix nitrogen by growing it in minimal medium (MM9) with N₂ provided in the headspace as the only nitrogen source, save for L-proline and the reducing agent cysteine. It continually grew in the medium over serial transfers, though it reached only about 1/2 of its normal turbidity. Maximal turbidity was increased by supplementing with 0.1% (NH₄)₂SO₄. This suggests some limiting factor in nitrogen fixation, perhaps iron or sulfur.

HPLC analysis on spent GS2 cellobiose media revealed production of formate, acetate, and butyrate with butyrate concentration increasing during late log phase. The ratio of product concentrations was about 3 formate to 2 acetate to 1 butyrate at log phase. A locus (CC90DRAFT_2157-2158) containing phosphate butyryltransferase and butyrate kinase (EC:2.7.2.7) were discovered by querying the genome for the EC numbers associated with both known pathways for butyrate production.

In stab cultures, small diffuse bubbles are evident in the media. Gas chromatography has revealed only small amounts of carbon dioxide production. This product is present in such low concentrations, however, that it cannot be reliably found from run to run.

KNHs209 grows on a wide variety of substrates (Table 10). Substrates utilized include L-arabinose, arbutin, cellobiose, chitin, fructose, D-galactose, β -gentiobiose, D-glucose, α -lactose, laminarin, levulose, D-lyxose, maltose, maltodextrin, maltotriose, D-mannose, D-ribose, salicin, sorbose, starch, sucrose, D-trehalose, D-xylose, and melezitose. Despite this wide repertoire, it was unable to utilize the abundant plant polysaccharides xylan and cellulose, even though the mono or disaccharides xylose, cellobiose, and glucose were readily metabolized. It seems unable to utilize methylated sugars, (fucose, rhamnose, and methylated xylose) and sugar alcohols (sorbitol, dulcitol, xylitol, adonitol, and mannitol), except for erythritol, on which it grows poorly. It is unable to utilize proteins, amino acids or fatty acids as carbon sources.

KNHs209's optimal temperature for growth was 40 °C, though it grew almost as well at 37 °C and 35 °C. It could not grow at 45 °C. Increasingly slower growth was observed from 30 to 15 °C.

Acceptable pH for growth ranged from 5.5 to 8.0, with no growth at 5.0 or 8.5. Optimal pH for growth was 7.0, and though it grew equally fast at 6.5, it reached a slightly lower maximum OD₆₀₀. It grew well at pH 7.5, though at pH 8 and pH 6.5 growth was impaired. Growth was severely impaired at pH 5.5, and did not grow at pH 5.0 or 8.5.

Table 10. Substrate utilization test results for KNHs209. Substrate utilization tests were carried out in GS2 liquid medium at 30 °C. Duplicate tests were run in 96 well microplates, at a final volume of 200uL and concentration of 1.5% substrate per well. Ambiguous or unclear results were redone in 10mL cultures. Growth was determined by changes in optical density at 600 nm compared to substrate-free wells and uninoculated control plates. ± denotes poor growth ($OD_{600} \geq 0.3$), - denotes no growth ($OD_{600} < 0.1$); + denotes good growth ($OD_{600} > 0.3$, usually between 0.5 and 1.0).

Acetate	-	D-Lyxose	+
Adonitol	-	Maltodextrin	+
Amygdalin	-	Maltose	+
D-Arabinose	-	Maltotriose	+
L-Arabinose	+	Mannitol	-
Arbutin	+	D-Mannose	+
Butyrate	-	Melezitose	+
Casamino acids	-	Melibiose	-
Casein	-	10 Methyl- α -D-glucopyranoside	-
Cellobiose	+	β -Methyl-D-xylose	-
Cellulose	-	Pectin	±
Chitin	+	Peptone	-
Dulcitol	-	Polygalacturonic acid	+
Erythritol	±	Propionate	-
Formate	-	Pyruvate	-
Fructose	+	Raffinose	-
Fuoidan	-	Rhamnose	-
L-Fucose	-	D-Ribose	+
D-Galactose	+	Salicin	+

D-Galacturonic acid	-	D-Sorbitol	-
β -Gentiobiose	-	Sorbose	+
D-Glucono- γ -lactone	-	Starch	+
D-Glucosamine	-	Succinate	-
Glucose	+	Sucrose	+
D-Glucuronic acid	-	D-Trehalose	+
Inulin	-	Xylan	-
Lactate	-	Xylitol	-
α -Lactose	+	D-Xylose	+
Laminarin	+		

Genome Summary

The genome is 4,676,087 kb in length, across three scaffolds. There are 4,176 protein coding genes, five copies of the 16S rRNA gene (1410 to 1412 bp), and two CRISPR regions. G + C content is 42.74 mol%. A genome summary and statistics can be found in tables 11 through 13.

Table 11. Sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Permanent Draft
MIGS-28	Libraries used	Pacbio SMRTbell™
MIGS-29	Sequencing platform	PacBio RS
MIGS-31.2	Fold coverage	51.7X
MIGS-30	Assemblers	HGAP 2.1.1
MIGS-32	Gene calling method	Prodigal 2.5
	Genbank ID	NZ_JPNB01000001
	GOLD ID	Gi0049735

Project relevance	Biotechnology, Digestion, Anaerobic Plant Degradation
-------------------	--

Table 12. Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	% of total ^a
Size (bp)	4,676,087	100.00%
G+C content (bp)	1,998,618	42.74%
Coding region (bp)	4,138,826	88.51%
Total genes ^b	4,263	100.00%
RNA genes	87	2.04%
Protein-coding genes	4,176	97.96%
Genes in paralog clusters	3,172	74.41%
Genes assigned to COGs	2,699	63.31%
Genes with signal peptides	155	3.64%
Genes with transmembrane helices	1,133	26.58%
Paralogous groups	3,172	74.41%

Table 13. Number of genes associated with the 25 general COG functional categories

Code	Value	% of total ^a	Description
J	157	5.23%	Translation
K	336	11.19%	Transcription
L	148	4.93%	Replication, recombination and repair
D	24	0.80%	Cell cycle control, mitosis and meiosis
V	112	3.73%	Defense mechanisms
T	204	6.79%	Signal transduction mechanisms
M	164	5.46%	Cell wall/membrane biogenesis
N	103	3.43%	Cell motility
U	42	1.40%	Intracellular trafficking and secretion
O	78	2.60%	Posttranslational modification, protein turnover, chaperones
C	106	3.53%	Energy production and conversion
G	382	12.72%	Carbohydrate transport and metabolism
E	233	8.63%	Amino acid transport and metabolism
F	71	2.36%	Nucleotide transport and metabolism
H	89	2.96%	Coenzyme transport and metabolism
I	69	2.30%	Lipid transport and metabolism
P	131	4.36%	Inorganic ion transport and metabolism

Q	31	1.03%	Secondary metabolites biosynthesis, transport and catabolism
R	326	10.86%	General function prediction only
S	197	6.56%	Function unknown
-	1,564	36.69%	Not in COGs

^aThe total is based on the total number of protein coding genes in the annotated genome.

Comparison to known taxa

Phylogenetically, KNHs209 forms a rather deep branch in the *Lachnospiraceae* (Figure 18) most closely related to another recently describe novel genus *Eisenbergiella*. *Eisenbergiella*, that has one published representative, *Eisenbergiella tayi*, and close relatives have been discovered in human gut microbiome sequencing surveys (Amir et al., 2014). While no genome sequence exists for the type organism, a comparison of the 16S rDNA sequence shows only a 94.2% sequence similarity with KNHs209. Both are strict anaerobes forming rod-shaped cells that stain Gram negative, are structurally Gram positive and that produce butyrate (Amir et al., 2014). A further phenotypic comparison shows stark differences in the lifestyle of these taxa (Table 14), with KNHs209 being a highly motile, saccharolytic spore former and *E. tayi* being a non-motile, non-saccharolytic non-sporeformer. Colony morphologies are are also strikingly different between the two, as well as a positive catalase reaction by *E. tayi*, an unusual enzyme to be produced by strict anaerobes.

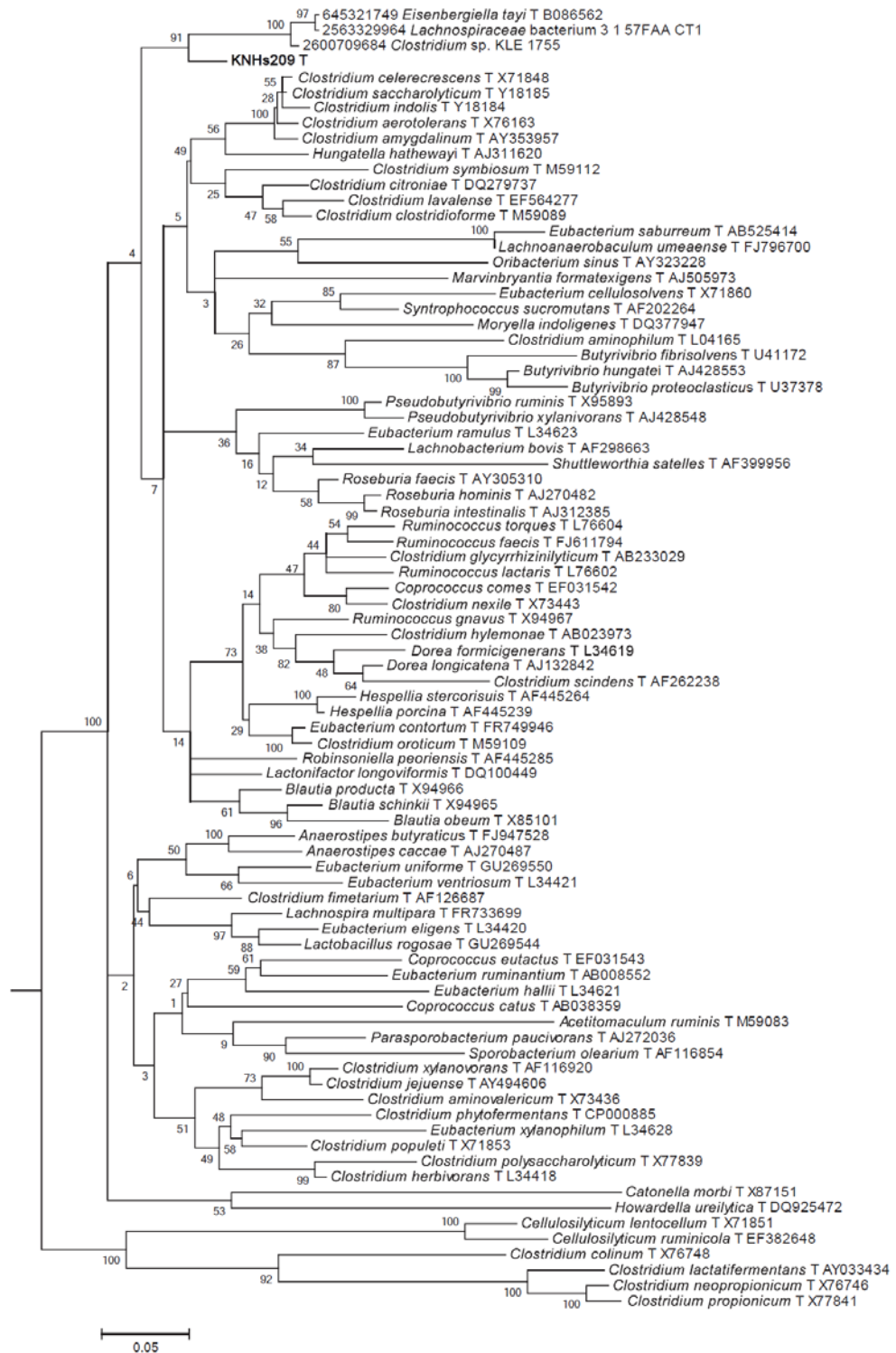


Figure 18. Phylogenetic tree showing the placement of KNHs209 within the *Lachnospiraceae*. (continued) Sequences meeting the following conditions were downloaded from the Ribosomal Database Project (RDP): *Lachnospiraceae* family, type strains, isolates, good quality, length >1200 bp. The best BLAST hits of environmental sequences are also included, though they cluster with the new genus *Eisenbergiella*. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The tree with the highest log likelihood (-13360.1513) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.6261)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 61.4081% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 107 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 997 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura K et al., 2013). Branches were pruned in R using the package Ape to condense the tree.

Beyond this genus, the next closest relatives are scattered among the large *Lachnospiraceae* subclade which includes the *Blautia* spp., *Roseburia* spp., and *Clostridium indolis* group, showcasing the uniqueness of KNHs209 from other described species. 16S rDNA identities for these members (type species/strains) are at most 92.6%, and average nucleotide identities (ANIs) range from 64.35% to 67.95% across all *Lachnospiraceae* type strains, with the median and mean ANI being ~66.5%. These taxa are so divergent, that these ANI values are beyond the limits of the method, with correlating DNA-DNA hybridization value being negative (Goris et al., 2007). It is clear that this organism represents a genetically and phenotypically distinct group, of which other species have yet to be found. We propose the introduction of KNHs209 as a novel species, *alysoides*, within the novel genus, *Kineothrix*.

Table 14. Comparison of KNHs209 to closest relatives based on 16S rRNA gene sequence.

	<i>KNHs209^a</i>	<i>Eisenbergiella tayi^b</i>
<i>16S rDNA sequence identity with KNHs209</i>	-	94.20% †
<i>Isolation</i>	forest soil	blood culture from 86 year old man with bacteremia
<i>%GC</i>	42.74% ‡	46.00%
<i>temperature opt</i>	40 °C (35 °C, 37 °C)	30 to 37 °C
<i>temperature range</i>	≤15 to 40 °C, not at 45 °C	15 to 45 °C
<i>morphology (colony)</i>	circular, white, translucent shiny, convex, entire margins, ~1 mm after 72 hrs	flat, opaque, irregular, rhizoid margins, 0.17-0.6 mm in diameter after 48-72 hrs
<i>morphology (cell)</i>	rods often singles or in filamentous chains reaching 30 µm in length; 3.3-6.7 µm X 0.5-0.6 µm	rods with slightly tapered ends, in singles or pairs, 3.4-7.3 µm X 0.4-0.7 µm
<i>fermentation products^c</i>	formate, acetate, butyrate	butyrate, lactate, acetate, succinate
<i>spores</i>	round subterminal, resulting in a lightbulb shape	none observed
<i>flagella</i>	single polar	none observed
<i>motile</i>	+	-
<i>catalase</i>	-	+
<i>substrates:</i>		
<i>arabinose</i>	L+,D-	-
<i>cellobiose</i>	+	-
<i>fructose</i>	+	-
<i>galactose</i>	+	-
<i>glucose</i>	+	-
<i>lactose</i>	+	-
<i>maltose</i>	+	-
<i>mannose</i>	+	-
<i>melezitose</i>	+	-
<i>ribose</i>	+	-
<i>salicin</i>	+	-
<i>starch</i>	+	-

(continued)

<i>sucrose</i>	+	-
<i>trehalose</i>	+	-
<i>xylose</i>	+	-

^aThis study, ^b(Amir et al., 2014), ^cin order of relative abundance, [†]IMG, [‡]16S rDNA sequences were aligned and then trimmed in MEGA6 prior to exporting into Emboss for Needleman-Wunsch global alignments. Untrimmed sequences artificially inflated distances with this algorithm, though trimming likely artefactually inflates similarity. Genome sequences and gene calling with the same method would lead to more accurate values, probably lower than those reported here.

5.5 Description of gen. nov., *Kineothrix*

Kineothrix: trich'o.pos.i.a Gr. n. *trícha* a single hair; Gr. adj. *ópos* like, hair like.

Given the relative conservation of cellular morphology between closely related bacteria, the name focuses on this trait. The long filamentous chains of cells appear like strands of hair under the microscope. The genus consists of anaerobic, motile, spore-forming rods, which ferment sugars and produce butyrate. They do not produce H₂S or reduce nitrate. They stain Gram negative but have a Gram positive cell wall structure. The DNA G + C content is approximately 43 mol%. The type species is *Kineothrix alysoides*.

5.6 Description of sp. nov., *Kineothrix alysoides*

Kineothrix alysoides (Ki'ne.o.thix. Gr. v. kineo to move; Gr. fem. n. thrix, trichos hair; N.L. fem. n. Kineothrix a moving hair; a.lys.o'i.des. Gr. n. alysis chain; L. suff. -oides (from Gr. suff. -eides, from Gr. neut. n. eidos, that which is seen, form, shape, figure) resembling, similar; N.L. neut. n. (nominative in apposition) a shape similar to a chain) is an obligate anaerobe within the family *Lachnospiraceae*. Cells are generally slim straight rods 3.3 - 6.7 µm in length by 0.5 - 0.6 µm in width, often longer than this in chains and when dividing, with chains commonly reaching 30 µm in length and

becoming filamentous. *K. alysoides* forms round terminal spores; swollen cells are not observed. Cells are highly motile, even in chains directional movement can be observed microscopically. This movement appears to be powered a single polar flagellum. Its optimal pH is 7, though it can grow at pH values down to 5.5 and up to 8.0. Growth is observed at least at low as 15 °C, but cannot occur at 45 °C, while it grows optimally from 35 to 40 °C. It does not seem to be capable of anaerobic respiration, based sulfite, sulfate, thiosulfate, and nitrate reduction assays. It does appear to be capable of nitrogen fixation, though it grows most robustly with urea in the medium. On cellobiose, *K. alysoides* produces formate, acetate, and butyrate as fermentation end products. It is urease, gelatinase, indole, catalase, and oxidase negative. A wide range of mono-, di-, and poly- saccharides can be fermented: L-arabinose, arbutin, cellobiose, chitin, fructose, galactose, glucose, lactose, laminarin, lyxose, maltodextrin, maltose, maltotriose, mannose, melezitose, polygalacturonic acid, ribose, salicin, sorbose, starch, sucrose, trehalose, and xylose. It is not able to grow on SCFAs (acetate, butyrate, formate, lactate, propionate, succinate), adonitol, amygdalin, D-arabinose, amino acids, cellulose, dulcitol, fucoidan, fucose, B-gentiobiose, D-glucono-γ-lactone, D-glucosamine, inulin, mannitol, melibiose, 10-methyl-α-D-glucopyranoside, B-methyl-D-xylose, pyruvate, raffinose, rhamnose, sorbitol, xylan, or xylitol. It was able to utilize pectin, but it was difficult to determine to what degree due to the turbidity of the polysaccharide. *K. alysoides* also utilized erythritol but did not grow well on it. The organism is available from the ATCC (TSD-26) and the DSMZ (100556).

5.7 Discussion

K. alysoides has presumably never been isolated or sequenced before, perhaps in part due to its rarity in the natural environment, but the closest relatives (~96% 16S rDNA identity) have been found in sequencing surveys of insect and mammalian guts. The discovery of this taxon further demonstrates that much microbial diversity remains unexplored, even in soil, which has been heavily sampled in environmental sequencing surveys, but also that this endeavor is attainable. The role this organism plays in a complex community as well as its trove of uncharacterized chemotaxis receptors deserve further exploration.

The family, *Lachnospiraceae*, in which *K. alysoides* resides, is one of the most abundant in the human gut microbiome. This family contains many known plant-degraders and most of the butyrate producers in the gut. Higher relative abundances of *Lachnospiraceae* have been associated with lower colon cancer incidences (Abreu and Peek Jr., 2014), likely due to a high plant matter/fiber-rich diet and the resultant butyrate production (Encarnaç o et al., 2015; Zhang et al., 2010). It is possible that this organism resides in the gut or rumen naturally, especially given its optimal temperature range, which matches the internal temperature of humans and large animals. In a study examining the relationship between butyrate production and habitat, *Lachnospiraceae* producing butyrate and forming spores tended to be gut inhabitants, while those inhabiting other sites seem to have lost this capability (Meehan and Beiko, 2014). Given that mucins lining the colon enrich for this family and butyrate production (Van den Abbeele et al., 2013), these members could easily be missed by the usual fecal material sequencing surveys. But, not surprisingly, nitrogen-fixing bacteria seem to be enriched in environments low in available nitrogen (urea, ammonium, amino acids), like the termite

gut (Shi et al., 2013). And, in the termite gut, nitrogen-fixation has been linked to cellulolysis by a protist endosymbiont (Hongoh et al., 2008), due mostly to production of amino acids from atmospheric nitrogen. Whether in mammals or insects, it is plausible that *K. alysoides* is a gastrointestinal tract-associated bacterium.

CHAPTER 6

PROBING THE IMPACT OF THE RARE BIOSPHERE THROUGH MICROCOSM DEVELOPMENT AND SYNTHETIC ECOLOGY

6.1 Abstract

Substrate utilization in nature results from metabolic and spatial interactions between microbes that are more easily teased apart in simplified communities. While these can be established by intentionally combining bacteria of interest, it is not clear how stable or ecologically-relevant the resulting communities would be.

Here we let the environment select for a simplified and stable community by introducing a soil sample to an anaerobic habitat containing switchgrass. The resultant microcosm was stable over several years of serial transfers and highly replicable. A small number of OTUs made up the majority of the reads, comprised entirely of the phylum Firmicutes, and the commonly gut-associated families *Enterococcaceae*, *Clostridiaceae*, *Lachnospiraceae*, and *Ruminococcaceae*. However, most of the diversity in the microcosm was due to thousands of low-abundance taxa, which persisted for the course of the experiment.

To test the role of these rare members in the microcosm, the most abundant members were isolated and reconstituted creating a community with only 10 isogenic strains. This synthetic community was directly compared with the evolved community

containing rarer members. Rare members led to a small though significant increase in degradation rate, though media formulation had a greater effect on productivity. This approach could be a very valuable tool for studying the contribution of genetic diversity and a community's rare biosphere to ecosystem function, resilience, and resistance.

6.2 Introduction

Soil is one of the most microbially-diverse habitats on Earth, containing thousands of species and between 10^6 and 10^{10} cells per gram (Crecchio et al., 2004; Curtis et al., 2002; Torsvik et al., 2002; Vieira and Nahas, 2005; Wang et al., 2016; Whitman et al., 1998). Less than a few percent of these species have been isolated, and those cultured in the laboratory are assayed outside the context of community dynamics and other ecological interactions. The discovery of the rare biosphere (Sogin et al., 2006), consisting predominately of uncultured members, over the last decade has only made the issue more confusing. When the majority of the diversity is due to low abundance taxa, we assume that they are playing a role in ecosystem function. Conversely, if something is in such low abundance, is it really able to contribute in a significant way?

Most of our taxonomic data is from amplicon libraries using the 16S rRNA gene for bacteria. These methods have problems accurately reflecting abundances (Gonzalez et al., 2012; Spiegelman et al., 2005) and can lead to completely new "taxa" due to chimera formation during amplification steps (Gonzalez et al., 2005; Kunin et al., 2010; Zajec et al., 2012). And, in the soil, inert DNA can remain viable for sequencing for decades (Trevors, 1996), likely leading to non-soil microbes to be labelled as soil-associated. These issues lead some researchers to believe that the rare biosphere is may

be comprised predominately of artefact sequencing artefacts(Kunin et al., 2010).

Most of these rare organisms have yet to be isolated and studied in pure culture(Albertsen et al., 2013; Neufeld et al., 2008; Shade et al., 2012; Sogin et al., 2006), so outside of broad classification into phylogenetic-functional guilds, their contribution to ecosystem processes and community function is largely unknown. This diversity makes modelling soil processes very difficult. Diversity is assumed to be beneficial, especially to ecosystem stability. Diversity leads to niche fulfillment, resistance, and resilience(Chapin III et al., 2000; Cook et al., 2006; Eisenhauer et al., 2013; Reich et al., 2012). *Clostridium difficile* infections are the prototype for negative consequences of low microbial diversity: antibiotics remove diversity, clearing the way for mass proliferation of an invasive microbe, which leads to ecosystem collapse.

While macroecological research has long focused on the effect of diversity on community stability and function, microbial ecology has struggled with effective experimental design. Sheer diversity can be manipulated *in situ* by adding toxins to cull whole populations, though these invariably target specific cell wall or membrane types or metabolic capabilities (Aguayo et al., 2013; Girvan et al., 2005; Müller et al., 2002). Not only does this add the variable of cellular stress to the experiment, it also removes taxa irrespective of their abundance or role in the community. This approach treats microbes as mere numbers rather than interactive and unique pieces of the ecosystem.

Targeted removal of only the rare members is even more complicated due to our lack of knowledge about their physiology. Molecular techniques have also been show to miss rare members that culturing can capture (Shade et al., 2012). How can we assay the effect of diversity alone, without removing whole classes of organisms? Synthetic

ecology has accomplished this by rebuilding environments and inputting particular organisms, while leaving out others that are present in the natural environment. This approach has been mainly used in plant ecology, and is obviously more difficult with microbial populations. But it is a promising, non-destructive approach for understanding the core tenants of microbial ecology.

In this work, we aimed to develop a simplified and stable community mirroring the anaerobic plant-degrading communities in decaying leaf litter in the soil. With fewer organisms overall and only the active members present, we could better understand which microbes are involved in the process. Moreover, maintaining a stable community *in vitro* allows us to probe the effect of perturbations to the environment, such as changes in temperature or substrate. We developed this microcosm by evolving a soil community through serial transfers covering approximately three years, while tracking the community structure to assay stability over time. We isolated the most abundant OTUs from the microcosm, enabling us to probe the effect of lower diversity in the context of the rare biosphere by reconstituting the community without these rare members.

6.3 Materials and Methods

Detailed methods for microcosm development and maintenance, HPLC analysis, and amplicon sequencing are as previously described.(Biddle, 2014)

Isolations and maintenance

Isolates were begun from the microcosm (replicate S2) at Transfer 43. All isolates were isolated on GS2 agar (1.5%) medium supplemented with 0.3% cellobiose. The

media contained, per liter: 3 g Na₂HPO₄, 1.77 g NaH₂PO₄, 2.18g K₂HPO₄, 3.4 g KH₂PO₄, 2.1 g urea, 3 g sodium citrate•2H₂O, 6 g yeast extract (Fisher BP9727-500), 2 g L-cysteine hydrochloride monohydrate, 1 mL 0.1% Resazurin, 100 mL GS2 salt solution, 900 mL ultrapure water, 3 g cellobiose. 100 mL of GS2 salts contained: 1 g MgCl₂•6H₂O, 0.15 g CaCl₂•2H₂O, 0.00125 g FeSO₄•7H₂O in ultrapure H₂O. Solidified plates were put into the anaerobic chamber (atmosphere ~3% H₂, ~10% CO₂, and balance N₂) at least 8 hours before use. Microcosm tubes were vortexed to mix; serial dilutions were made in sterile water, which were then streaked onto GS2 plates. Colonies were restreaked until they appeared pure by colony morphology, then observed microscopically to ensure a single cell morphology was present. The 16S rRNA gene was amplified and then sequenced using 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-ACGGCTACCTTGTTACGACTT-3') universal bacterial primers to test for purity and to classify the organism.

The Ribosomal Database Project (RDP) website was used to classify taxa based on 16S rRNA gene sequences (Cole et al., 2009; Wang et al., 2007). More accurate taxonomic and phylogenetic placements were determined via phylogenetic tree development using all available type species 16S rRNA gene sequences in the RDP database (Tamura et al., 2011). Trees were built using the Neighbor-Joining method in MEGA5 (Tamura et al., 2011, p. 5).

Multiple rounds of isolations were carried out using various medium formulations, including the substrates acetate, raffinose, rhamnose, and xylose, and on both GS2 and Mic base media. Both spreads and streaks of serial dilutions were performed. In an effort to obtain more novel isolates, a PCR-based assay was developed

to check for the next top three abundant bacteria in the communities. Primers were built based on the Illumina phylotag sequencing data: a consensus sequence, including degenerate sites, was developed for each of the three OTU clusters using in-house R scripts. Thereafter, general guidelines for primer design were followed. These primers were used to perform PCR, testing if these OTUs were present in the whole communities, enrichments, and in isolated cultures.

Pure cultures of isolates were grown anaerobically in GS2 liquid medium supplemented with 0.3% cellobiose. Medium was prepared per the Hungate technique (Hungate, 1969), flushing warm medium with nitrogen gas (N₂) while stirring until Resazurin became colorless. 9.8mL was then aliquotted into tubes being flushed with nitrogen gas, and flushed until colorless again. With nitrogen flowing, tubes were placed on ice till cooled below room temperature. Rubber stoppers were firmly inserted after removing from nitrogen gas, and tubes were autoclaved at 121 °C for 15 minutes.

Freezer stocks were made upon confirmation of purity by 16S rRNA gene sequence analysis. A sterile solution of 30% glycerol and a mid-late log phase culture were iced for ~10 minutes, then 0.75mL of glycerol was added to a 2.0mL cryovial, followed by 0.75mL of the culture. Cryovial caps were tightly closed, and freezer stocks were immediately transferred to a -80 °C freezer. Viability and purity of freezer stocks were checked by inoculating into GS2 with cellobiose and observing the resultant culture microscopically with a phase contrast microscope at 400X and 1000X magnification.

Reconstituting communities

Isolates were grown individually from freezer stocks in GS2 with 0.3% cellobiose and combined during each's mid-to-late log phase in anaerobically prepared liquid Mic_{peb}

or MQM_{peb} medium, containing 3mL of 3% pebble-milled switchgrass (see next section). Once inoculated, tubes were agitated to disperse and mix the cells. Tubes were kept at 30 °C for the duration of the experiment. Two transfers of the rebuilt communities were performed before beginning the experiment. Triplicate samples were sacrificed at each timepoint for HPLC analysis.

Switchgrass Degradation Measurements

15g of dried switchgrass was homogenized in a Waring blender for two minutes before being combined with 300mL of deionized distilled and filtered water in a pebble milling jar. The slurry was milled for 1 week (± 30 minutes) before being diluted with another 200mL of water. This slurry was then immediately used to make liquid Mic_{peb} or MQM_{peb} medium. Switchgrass degradation was measured by switchgrass height every three days. After each reading, the tubes were vortexed for two seconds to allow release of gas bubbles.

MQM contained (per 690 mL): 2 g NaH₂PO₄, 10 g K₂HPO₄, 1 g L-cysteine hydrochloride monohydrate, 0.05 g L-proline, 40 mL XT2 solution, 10 mL BTE solution, and 1 mL 0.1% Resazurin. 3 mL pebble-milled switchgrass was aliquoted per tube, followed by 6.9 mL of medium. After autoclaving, 0.2 mL CPV4 was added to each tube. XT2 contained (per 100 mL): 0.25 g xanthine, 0.25 g thymine, and 1 mL 6N NaOH. BTE contained (per L): nitriolotriacetic acid, 3 g MgSO₄·7H₂O, 0.5 g MnSO₄·4H₂O, 1 g NaCl, 0.1 g FeSO₄·7H₂O, 0.1 g CoCl₂·6H₂O, 0.1 g CaCl₂, 0.1 g ZnSO₄·7H₂O, 0.01 g CuSO₄·5H₂O, 0.01 g AlK(SO₄)₂, 0.01 g H₃BO₃, 0.01 g Na₂MoO₄, 0.03 g NiSO₄·6H₂O, 0.02 g Na₂SeO₃, and 0.02 g Na₂WO₄. CPV4 contained (per 100 mL): 4 mg p-

aminobenzoic acid, 0.1 mg biotin, 0.6 mg folinic acid, 8 mg nicotinamide, 0.5 g pantethine, 0.4 mg pyridoxal, 3.0 mg riboflavin, and 1.0 mg thiamine.

6.4 Results

Microcosm Community structure

Sequencing and diversity metrics of microcosm samples are as previously described by Biddle (Biddle, 2014) (Table 15). Briefly, there were 4.5 million paired-end reads representing 99,558 OTUs. Chimera Slayer (Haas et al., 2011) removed approximately 300,000 sequences and 7,212 OTUs. There was a dramatic decrease in diversity between the initial soil sample and all subsequent transfers, following relative stability which was consistent across replicates.

Table 15. Microcosm development sequencing metrics. All sequenced replicates and transfers are included in these values.

	Before Chimera Check	After Chimera Check	Chimera Checked > 1.5% Abundance
Total number of sequences	4554420	4204943	3328037
Sequences per sample (mean)	114734.5	105123.575	83200.925
Sequences per sample (min)	49400	45989	37
Number of OTUs	99558	92346	18

Total number of sequences, sequences per sample, and OTUs before and after chimera detection using Chimera Slayer (Haas et al., 2011).

The original soil sample was dominated by the phyla *Acidobacteria*, *Proteobacteria*, and *Verrucomicrobia*, with 10 phyla having a relative abundance of

greater than 1.5% (Figure 19). Together these 10 phyla comprised 93% of all sequences. Between Time 0 and Transfer 2, these 10 phyla became only two: the *Firmicutes* and the *Proteobacteria*. Though at Time 0 the *Firmicutes* made up only 0.47% of sequences, they quickly became dominant, making up between 74% and >99% of sequences for all other transfers.

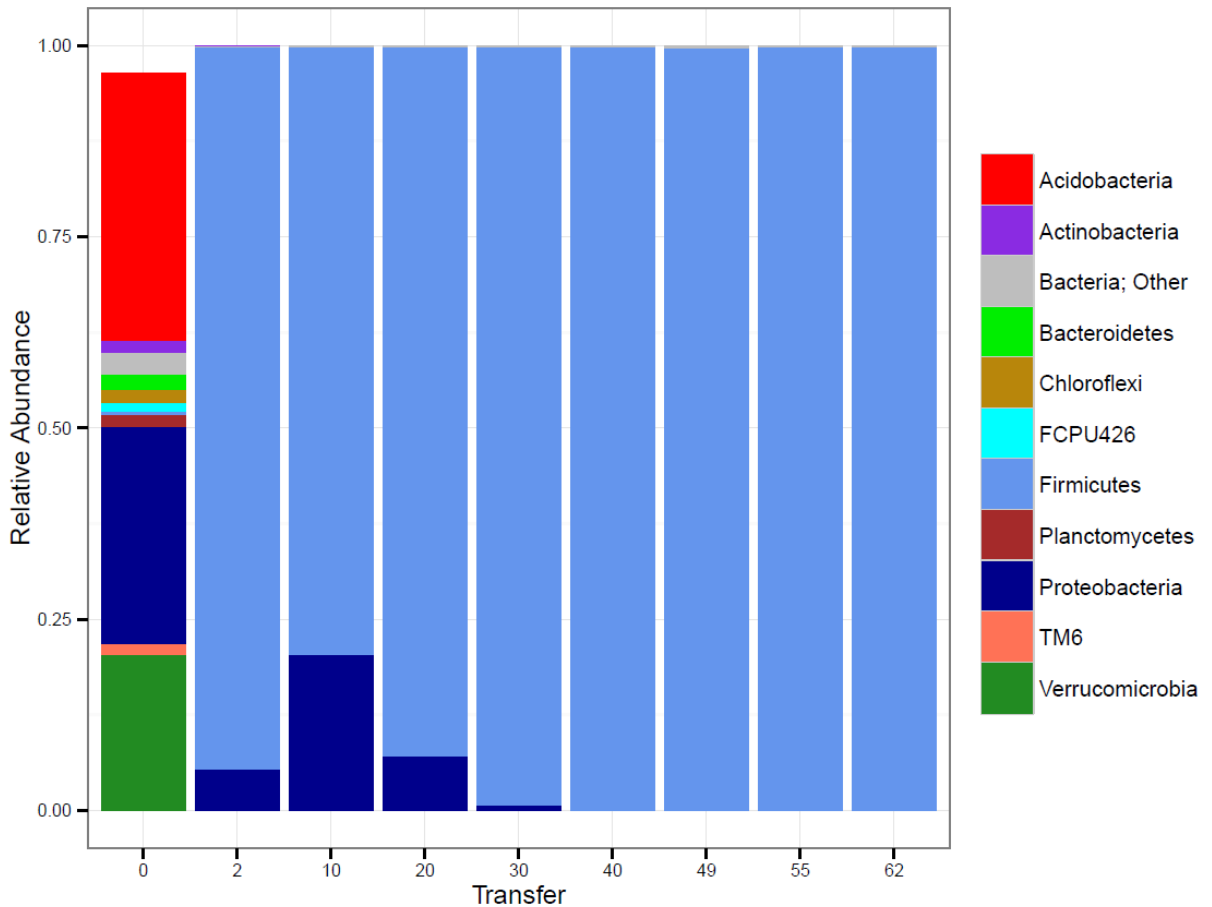


Figure 19. Phylum-level diversity of the microcosm over time. Transfer 0 is the initial soil sample from which the microcosms were developed. The soil community is typical, with Acidobacteria, Proteobacteria, and Verrucomicrobia dominating and other soil-associated phyla like Actinobacteria, TM6, and Planctomycetes making up significant fractions. This timecourse is representative of all five replicates, with the Firmicutes being the most abundant microcosm member, and the Proteobacteria being lost as transfer intervals increased from 1 week to 4 weeks.

Of the three dominant phyla at Time 0, only the *Proteobacteria* were still present greater than 1.5% at Transfer 2, though the identity of the OTUs changed dramatically (Figure 20). The two most abundant *Proteobacteria* OTUs at Time 0 were no longer abundant by Transfer 2, while three other *Proteobacteria* OTUs which were low-abundance at Time 0 became abundant by Transfer 2. These three OTUs were members of the *Enterobacteriaceae* and persisted until Transfer 20, after which they declined (Figure 20). None of the abundant OTUs from Time 0 remained abundant at or beyond Transfer 2. Conversely, none of the OTUs at 1.5% or greater abundance in Transfer 2 were found at greater than 0.09% at Time 0.

By Transfer 2 there were eighteen OTUs from five families at relative abundances greater than 1.5% (Figure 20), including the three *Enterobacteriaceae* species. The remaining OTUs belonged to the *Enterococcaceae*, *Lachnospiraceae*, *Clostridiaceae*, and *Ruminococcaceae* families (Figures S2-S4). By Transfer 62, ten OTUs were present at relative abundances greater than 1.5%, three of which were common to all five replicates (Figure 21). Entcoc1 and Lachno6 were abundant across every replicate and time point after Time 0. Together, Entcoc1, Lachno6, and Lachno4 comprised 48-67% of the total relative abundance in all five replicates at Transfer 62 (Figure 21). At this timepoint, all five replicates shared 265 OTUs which comprised 83-93% of the reads, and most of the diversity (>99%) in the microcosms was represented by a relatively small portion of the reads (7-17%).

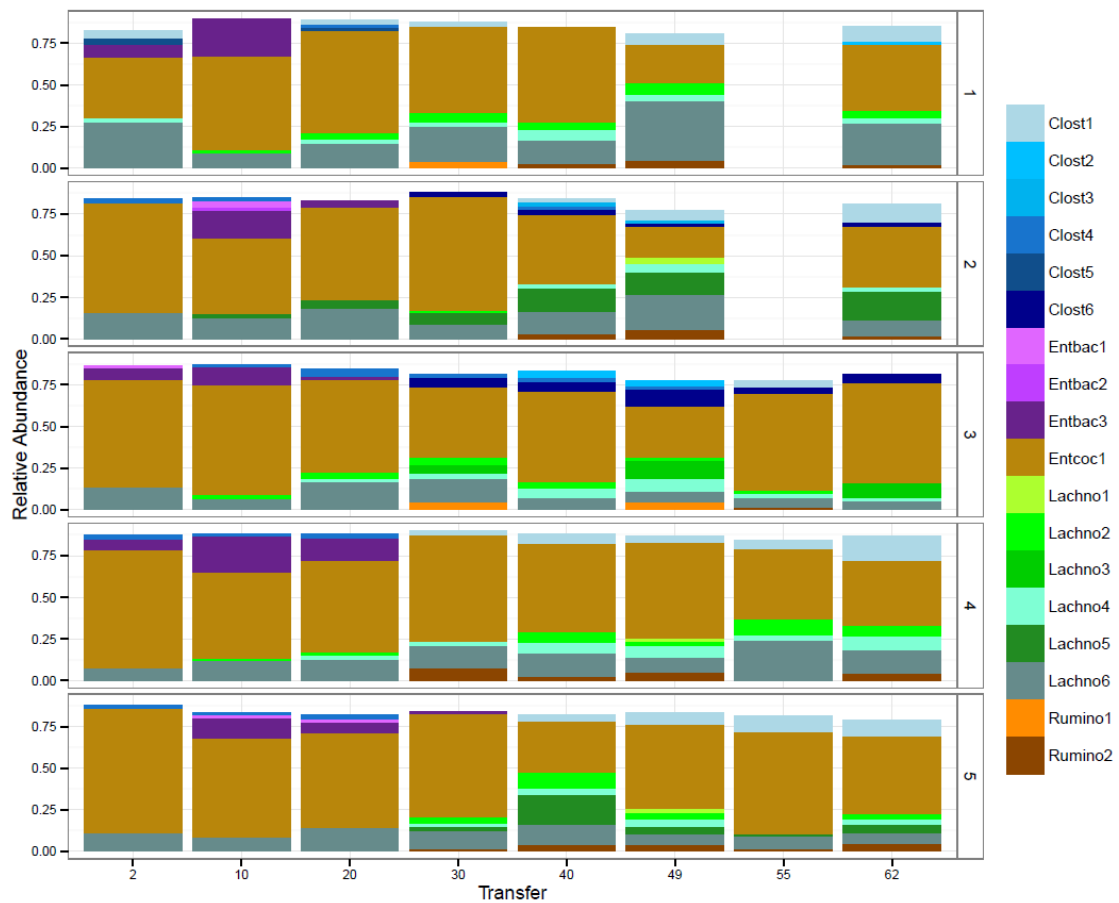


Figure 20. Most abundant OTUs over time. Abundant OTUs ($\geq 1.5\%$ relative abundance) are shown by replicate (1-5) and over time (transfers 2-62). OTUs are named by their family: *Lachnospiraceae* (Lachno1-6), *Clostridiaceae* (Clost1-6), *Ruminococcaceae* (Rumino1-2), *Enterobacteriaceae* (Entbac1-3) and *Enterococcaceae* (Entcoc1). None were considered abundant members of the initial soil sample (transfer 0).

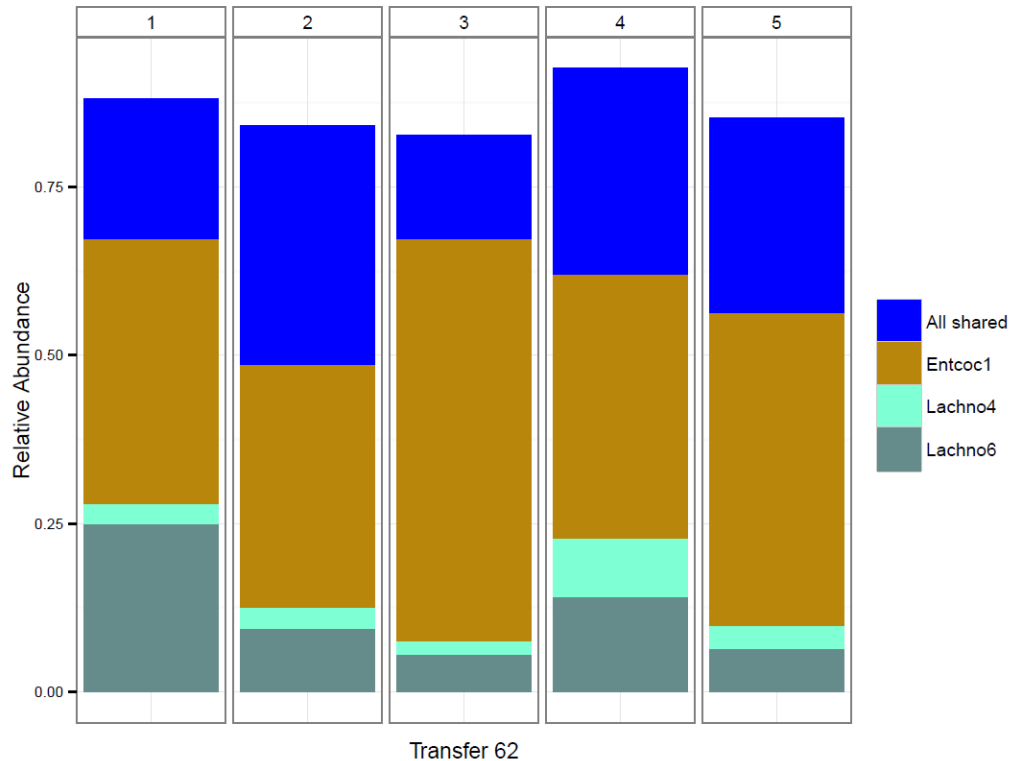


Figure 21. OTUs shared across all replicates at transfer 62 (T62). Three OTUs comprised most of the reads amongst all replicates at T62. There were several other lower abundance OTUs that were shared across all replicates as well (“All shared”), making up >80% of the reads in any given replicate. This highlights the replicability of this approach.

Enterococcus faecalis had maintained its rank as the most abundant microbe across all replicates, though *E. faecalis* is known for its role as a lactic acid bacterium and as a pathogen (Callewaert et al., 2000; Corredoira et al., 2015; Foulquié Moreno et al., 2003; Harwood et al., 2004), not as a plant degrader. It does have varied metabolic capabilities (McBride et al., 2007; Ryu et al., 2001), allowing it to grow very robustly under myriad conditions. This coupled with its rapid growth rate, made it seem like a weed rather than a workhorse in the system. We hypothesized that *Enterococcus* blooms were outcompeting the microbes capable of complex carbohydrate degradation. In order to remove the *Enterococcus faecalis* from the microcosm, which we presumed was not

contributing to switchgrass degradation, we transferred the community to a minimal medium, referred to as MQM. After two transfers on this medium, we re-sequenced the community and found that *E. faecalis* had indeed lost its foothold in the microcosm, while allowing rarer members to become more abundant (Figure S5). There is likely a confounding effect between the loss of *E. faecalis* and the change in media on the shifting community. More importantly, this led to an observable increase in rate of switchgrass degradation.

Microcosm Isolates

We were able to isolate the most abundant bacteria in the community and two rather rare members. *Lachnospiraceae* bacterium sp. KNHs209 (GenBank: GCA_000732725.1) comprised only 0.034% of reads on average by T62. *Robinsoniella peoriensis* strain KNHs210 comprised only 0.322% of reads by T62. *E. faecalis* was very easily isolated from the microcosm, and often appeared in co-culture with other species. Heat or ethanol treatment allowed us to remove it from mixed cultures with spore-formers. Except for *E. faecalis*, all of the isolates are *Clostridia*, with five in the family *Lachnospiraceae*, one in the *Clostridiaceae*, and one in the *Ruminococcaceae*.

Four strains of what appears to be a novel species based on *rpoβ* gene and ANI comparisons were discovered and initially mistaken for *Clostridium phytofermentans*, which is noted for its ability to degrade minimally processed plant material and cellulose while producing mainly ethanol. These strains do not perform as well as *C. phytofermentans* does in utilizing plant fibers, but they were the only isolates capable of

visibly degrading switchgrass. In certain replicates, this species is abundant (16.949% of S2 at T62), and in others it is found at very low abundances (0.003% of S4 at T62; 0.005% of S1 at T62).

The *Clostridiaceae* species (*Clostridiaceae* sp. KNHs214, GenBank: GCF_000744935.1) forms a very deep branch within the family, perhaps justifying classification as a novel genus. It appears to be responsible for protein turnover in the system, being the only one capable of utilizing proteins and amino acids for growth. Though plant material is often thought of as a carbohydrate source, leaves are also high in protein (Ghaly, 2010; Pandey and Srivastava, 1991; *Plants*, 1983), so dominance of a proteolytic taxon is not surprising. The others are all saccharolytic, though their substrates do vary, likely allowing them to specialize somewhat.

We obtained two isolates of the *Ruminococcaceae* species (*Ruminococcaceae* sp. KNHs216, GenBank: PRJNA262381), but due to lack of phenotypic distinctiveness, we continued working with a single isolate. Like the *Clostridiaceae* sp., this taxon is also phylogenetically very distinct from described relatives, likely classifiable as a novel genus.

Due to the apparent novelty of these isolates and in order to better understand their roles, we are completing whole genome sequencing and characterizations. The current strain names are KNHs205 (*Lachnospiraceae*), KNHs206 (*Lachnospiraceae*), KNHs209 (*Lachnospiraceae*), KNHs210 (*Lachnospiraceae*), KNHs212 (with strains KNHs2131, KNHs2132, and KNHs219) (*Lachnospiraceae*), KNHs214 (*Clostridiaceae*), and KNHs216 (*Ruminococcaceae*). Phylogenies of these taxa are in supplemental figures S2-S4.

Searching for the Rarer Bacteria

Though we had the seven most abundant taxa isolated, there were a few that had evaded us, and so we returned to isolations, this time with different media, temperatures, and substrates. Mostly appearing as nondescript rods under the microscope, we utilized a PCR based assay to determine whether an isolate was one of these taxa. While whole community DNA gave positive results (bands), no isolate did. It is very likely that, even though they were so abundant, these organisms were chimeras. Of course, it is also possible that we simply could not isolate them. But, to test the degree of chimera formation in our communities, we performed deep sequencing on pooled isolate gDNA (from our isolates corresponding to Enterococ1, Clost1, Rumino2, Lachno1, Lachno2, Lachno4, Lachno5, and Lachno6).

There were ~334,000 sequences with an average length of 254 bp and 5,900 OTUs after quality filtering and initial OTU picking in QIIME (Caporaso et al., 2010) (Table 16). Chimera Slayer (Haas et al., 2011) removed approximately 1,000 sequences, leaving 4,131 OTUs for further analysis.

Table 16. Chimeras lead to overestimations of diversity. Eight distinct species' genomes were extracted then pooled in equal masses. This milieu was sequenced using the same methods as were used for initial microcosm sequencing. These eight taxa led to almost 6,000 OTUs, highlighting the great impact of chimeric sequences on diversity metrics. Chimera Slayer was able to remove some of these sequences, but likely due to the closely-relatedness of sequences within a chimera, could not identify all of them.

	Before Chimera Check	After Chimera Check	Isolate Sequences
Total number of sequences	334,881	333,086	305,356
Number of OTUs	5918	4131	8

The sequence identities of the eight, pooled isolates were matched against the OTU sequences in the dataset using local BLASTN to verify recovery of the known sequences. Percent identity to each OTU representative sequence ranged from 97% to 100%, with 100% coverage. Phylogenies between the isolate sequences, OTU representatives, and selected type species verify of relationships (Fig S2-S4). The eight isolates corresponded to the top eight OTUs and represented 92% of the total reads. The remaining reads all classified as Firmicutes.

After a thorough search for the taxa lying just below 1% relative abundance, we believe that these may be artefacts of amplicon sequencing. Our pooled DNA experiment did reveal however, that there are certainly more than the eight taxa we isolated: our original sequencing effort revealed 92,346 OTUs while our mainly chimeric results gave only 4,900 OTUs. The putatively low diversity of these samples may have benefitted from use of PhiX as an internal control, though this would have only mitigated sequencing errors, not chimera formation during the PCR step. While our 8-isolate sequencing data still had a great diversity (4,900 OTUs), it shows that there must be other

organisms in the community due to the ~92,000 OTUs present in the whole microcosms. In order to determine which taxa were most likely to be representative of real rare microcosm members, we scanned the data for recurring taxa which were classified as distinct species, genera, classes, etc. from our isolates. Some of these persistent rare members were classified as *Cupriavidis* spp. within the β -Proteobacteria, *Bradyrhizobium* within the α -Proteobacteria, and *Bacteroides* spp. within the Bacteroidetes. Given their phylogenetic distance from our isolates, these are not chimeras. There were also low-abundance taxa within the *Ruminococcaceae* (*Oscillospira*), the *Clostridiaceae* (*Sporoanaerobacter*), and the *Lachnospiraceae* (*Anaerostipes*), though we had isolated some taxa from these families, but not these genera, as well.

PCR and sequencing artefacts have been shown to lead to overestimations of diversity which can be improved through strictly quality filtering (Jeon et al., 2015; Krohn et al., 2016; Kunin et al., 2010). We re-performed the data analysis with stricter parameters during read merging and filtering steps. Merging was done using Flash with a minimum overlap of 20 bases. Read filtering used a minimum Phred score of 30 across at least 80% of the read, and the first 10 bases of each sequence were trimmed. When we used stricter quality parameters for read concatenation, we lost approximately 1,696,035 reads (59.7%) from the original sequencing dataset. We were left with 2,508,908 paired-end sequences and 2,141 OTUs, a stark contrast to the initial analysis. 1,701 of these OTUs were only present in the soil samples (T0), with 440 OTUs remaining across the microcosm datasets. There were, however, still low-abundance OTUs which were present across transfers and replicates, suggesting that there are certainly rare taxa which have

persisted in these microcosms. It is also possible that quality filtering led to removal of other low-abundance taxa, given that approximately 3 of every 5 reads were left out of the analysis.

Our isolates were still the most abundant taxa in later transfers, though abundances had changed slightly. Isolated taxa comprised between 77.56% and 89.55% of T62 replicate reads. All taxa over 1.5% relative abundance accounted for between 89.96% and 93.36% of T62 replicate reads, with a few unisolated taxa appearing to be more abundant.

The pooled-isolate sequencing results also changed considerably, leaving only 251 OTUs (originally 4,900). A smaller fraction of reads was lost from this dataset (~5.8%), which was sequenced separately from the initial microcosm sequencing run, due to longer and higher quality initial reads. The difference between the whole community and the pooled-isolate sequencing results (440 vs 251 OTUs) still suggests that there are more taxa in the whole community than just the isolates, but that there is also an overestimation of distinct taxa either due to sequencing errors and/or chimera formation. Many of the same taxa from the original analysis were still present as lower-abundance taxa in the stricter analysis: *Sporoanaerobacter*, *Oscillospira*, *Anaerostipes*. *Anaerovorax*, *Soehngenia*, and an unknown *Clostridium* species were also present.

Probing the Significance of Rare Members

To test the significance of these evasive rare members on community function, we reconstituted the communities on both the original Mic medium and on the defined MQM

medium, using pebble-milled switchgrass as the carbon source. The communities differed based on the growth medium, and these differences were reflected in rebuilding them (Supplemental Figure S5). For example, *E. faecalis* was only present in the community grown in Mic medium, while KNHs209 was only abundant in the community on MQM medium. We did include all four strains of the cellulolytic, KNHs212, in the rebuilt communities since they are indistinguishable at the OTU-level. However, there is certainly some microdiversity within species in the community that we could not account for in this experiment.

Both the whole and rebuilt Mic communities degraded poorer than those in MQM, with the whole community in MQM degrading switchgrass the fastest (Figure 22). All switchgrass columns reached a nadir at day 18, but began to get taller again. This may be due to an increase in cell biomass as plant material is consumed, which is a common problem in mass-based measurements of plant matter degradation. Gas production also leads to fluffing of the substrate. The greatest switchgrass loss was 25% by the rebuilt community in Mic, 24% by the whole community in Mic, 30.1% by the rebuilt community in MQM, and 35.9% by the whole community in MQM. Two-way Anovas ($\alpha=0.05$) confirmed that there were significant differences in switchgrass degradation between the whole and rebuilt communities in both Mic and MQM media ($p<0.05$).

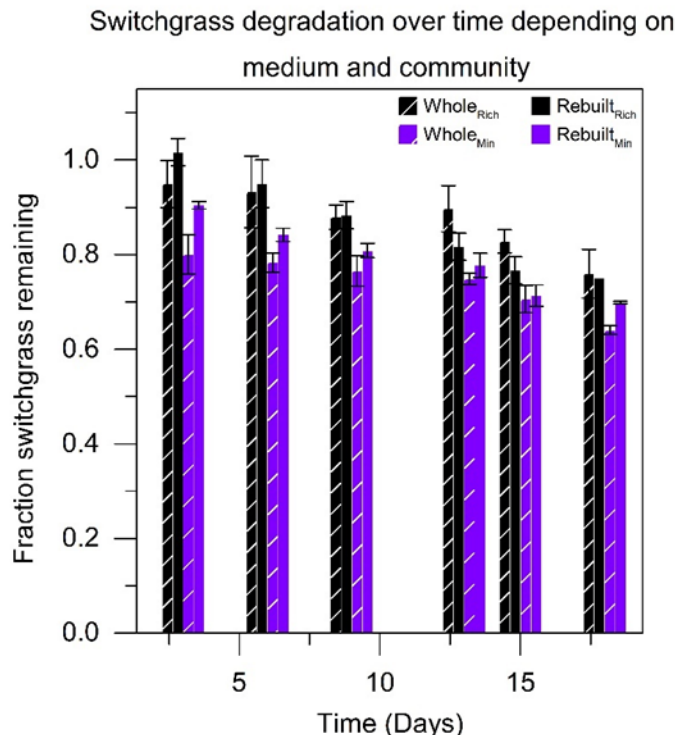


Figure 22. Switchgrass degradation by medium and community type. Shown as fraction of switchgrass remaining compared to the height (in mm) just prior to inoculation. MQM was the minimal medium (“Min”) and Mic (“Rich”) was the undefined, richer medium. The whole community (“Whole”) is a continuation of the microcosm whereas the rebuilt community (“Rebuilt”) was synthesized by building a community from only the most abundant isolates, thereby removing the rare members. The striped pattern denotes the whole community, including rare members. The bars are color-coded by medium, with black being the richer and violet being the defined medium. Three-way ANOVAs show significance ($p < 0.05$) by medium and by community. Gas production or bacterial biomass led to an increase in switchgrass height after day 18, so the remaining datapoints are not shown.

Acetate and ethanol were the major fermentation products in all four communities (Figure 23). However, productivity appears to be higher in Mic medium than in MQM, with approximately three times as much acetate produced. For the whole community in Mic medium, acetate had already reached 72.3 mmol by day 3, and 3.0 mmol of ethanol.

Ethanol may have been used by other community members, as it decreased slowly thereafter. Acetate peaked at day 13 (91.1 mmol) then decreased down to 72.7 mmol by day 24. The lag in degradation by the communities in Mic medium coupled with this quick initial rise in acetate levels seems to show that an initial bloom, probably powered by the small amount of yeast extract in the medium, outcompeted plant degraders and delayed their growth. This yeast extract may also have been preferentially used as easy carbon by the plant degraders, delaying plant matter utilization.

Interestingly, the community which showed the fastest plant degradation (“Whole_{MQM}”) produced the least acetate (between 10.9 and 17.5 mmol). The rebuilt community on MQM had between 16.9 and 28.0 mmol of acetate, first rising steadily, then apparently being utilized after day 13. And while its ethanol production tracked with the ethanol produced by the rebuilt community in MQM, after day 9 the ethanol in the rebuilt microcosm began to decrease slowly and the ethanol in the whole microcosm slowly increased. Here it seems there were rare members who somehow led to a small but significant increase in plant degradation.

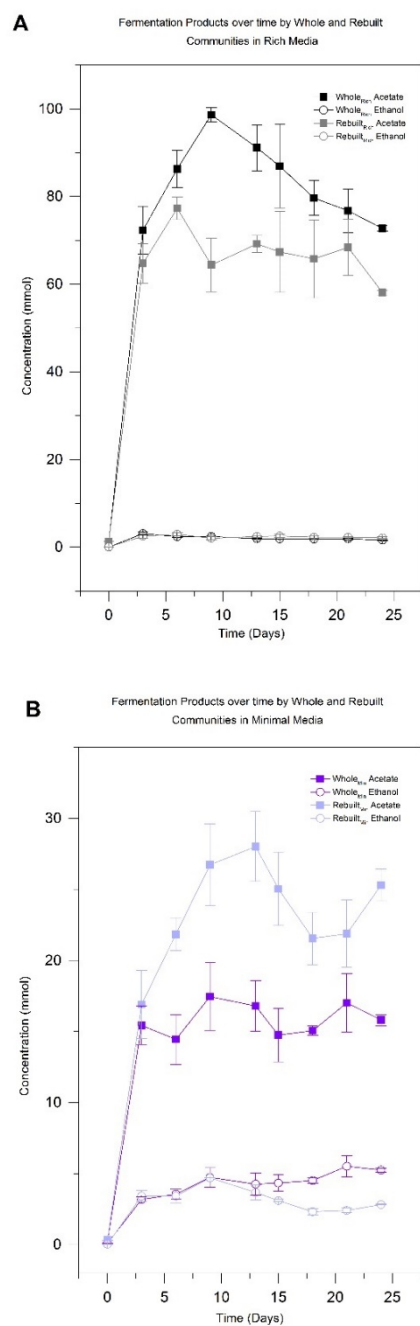


Figure 23. Fermentation product analysis of the whole communities and rebuilt communities by media type, rich (A) or minimal (B). HPLC was used to assay production of short chain fatty acids and alcohols, though only acetate and ethanol were found. Isolates produce a greater variety of fermentation products (formate, lactate, succinate, and butyrate), but they are likely reused as electron acceptors or carbon sources for other community members.

6.5 Discussion

While soil communities are predominantly *Proteobacteria*, *Acidobacteria*, and *Actinobacteria*, many of these organisms are oligotrophic root-associated aerobes or microaerophiles (Janssen, 2006; Miao and Davies, 2010), becoming outcompeted by the more opportunistic copiotrophs when conditions become more nutrient rich or anaerobic. The shift from a typical soil microbiome to a typical fecal microbiome has been seen in cattle grazing grounds, where a continued input of cattle manure led to a soil community comprised mainly of *Firmicutes* (Chroňáková et al., 2015). *Firmicutes* are well-known as dominant members of mammalian gut microbiomes, as well as the foreguts of ruminants, though they do make up a small but significant portion of soil communities (Janssen, 2006). These organisms also make up large portions of waste reactors and landfills (Franklin and Mills, 2006; McDonald et al., 2012; Parshina et al., 2000; Sun et al., 2015), likely due to a combination of anaerobic conditions, high plant material content (e.g. paper and cotton products), and food waste. Similarly, in this work we adapted the soil sample input into an anaerobic plant-degrading community predominated by *Firmicutes*.

The *Lachnospiraceae*, *Ruminococcaceae* and some *Clostridiaceae* are well-known as plant-degrading bacteria, often inhabiting the gut (Arumugam et al., 2011; Biddle et al., 2013; Brulc et al., 2009). Several OTUs are related to the cellulose degraders *Clostridium phytofermentans*, *Clostridium sporosphaeroides*, and *Clostridium celerecrescens* (Fig S2-S4). These species may break down cellulose, enabling the noncellulolytic organisms (e.g. *Enterococcus spp.*) to utilize the resultant simple sugars or more easily access hemicellulolytic components. *Enterococcaceae* are not known to be

cellulolytic, though they have been shown to ferment a wide variety of substrates (Byappanahalli et al., 2012; *The enterococci*, 2002). This coupled with their quick growth explains their dominance in the microcosms. *Enterococci* are used in fermentations of fermented dairy and meat products due to their production of bacteriocins which prevent colonization and spoilage by other organisms (Callewaert et al., 2000) (Foulquié Moreno et al., 2003), but relatively little else is known about their ecology and interspecies interactions.

Individually, these microbes produce a variety of fermentation products, including lactate, formate, butyrate, and succinate, though none of these fatty acids were visible by HPLC when the isolates were combined. The communities seem to be evolved for efficiency, meaning complete utilization of every resource in the system; where end-products are recycled by other microbes. This is likely at the cost of plant degradation, which is a physiologically expensive microbial process, especially when extracellular enzymes must be produced and exported and then benefit not just the enzyme-producer but the whole community. In a richer environment where alternative carbon sources are available, plant degradation would likely be a last resort. This may be why the minimal media produced the most favorable results in terms of plant degradation. In future work that aims to enhance plant utilization rates, removing all alternative carbon sources would be essential.

Excluding rare members had a small negative effect on overall productivity, leading to slightly slower degradation of switchgrass. Based on these results, the rare members, while able to persist, are not integral to microcosm function. It seems logical that the most abundant members would be performing most of the work in a community,

while rarer members are scraping by in whatever niche they can fill. These rare members do, however, add to the efficiency of the system by increasing functional diversity. There are notable exceptions where rare taxa have been dubbed keystone species (Neufeld et al., 2007; Pester et al., 2010), but more often keystone species are relatively abundant members (Fisher and Mehta, 2014; Power et al., 1996; Rafrafi et al., 2013; Ze et al., 2012). It is believable that most confer the almost negligible benefit seen here, where the relatively abundant cellulolytic (KNHs212 strains) was arguably the keystone. This does not preclude the role of rare members as a seed bank, but in a stable environment, the community can be assumed to be equally stable. Shaping this community for maximum desired functionality (e.g. bioprocessing) would include a diversity of taxa, along with nutrient-limiting pressure to focus the unique talents of each member onto the task at hand.

In this system, a dramatic change in microbial community composition occurred during the initial serial transfers followed by relative stability, in a process that results in quite similar communities. The stability over time shows that this is a robust method for studying *in vitro* communities especially in response to environmental perturbations, assuming that the growth conditions accurately reflect the natural system, including temperature, turnover rate, and nutrient availability. Here an evolved microcosm approach also enabled the discovery of novel microbes in a community context that might otherwise evade detection due to low abundance in the natural environment.

CHAPTER 7

SUMMARY AND CONCLUSION

In this body of work, we have worked towards a taxonomy within the family *Lachnospiraceae* which is in line with phylogenetic and physiological data, while adding new and possibly rare taxa to our body of knowledge. An anaerobic microcosm evolved and maintained on switchgrass was the beginning of this journey. From this system, we were able to isolate several *Clostridia* that were in very low abundance in the initial soil sample. While the very first isolate was an environmental strain of *Enterococcus faecalis* ORF1, each other isolate was unique from previously characterized taxa.

And while we had isolated the most abundant organisms from the microcosm, our community sequencing data showed that there were still taxa present that we were unable to isolate. Surprised that low-abundance members were able to persist for so long, we reasoned that they must be playing a role in the community function. The role of these rare members in the environment is largely unknown, though in some studies they have been shown to act as a seed bank which flourishes when new nutrient are available, presumably adding to the stability and resistance of the ecosystem. In order to test their contribution to ecosystem function, we rebuilt the communities without these rarer members. In comparison to the whole communities, the rebuilt communities degraded switchgrass slower, though a nutrient-limiting habitat (minimal medium) was actually more responsible for quick degradation than was the community composition.

One of the taxa which became abundant when the microcosm was grown on

minimal media was a butyrate producer with a wide substrate range. At the time of isolation, the closest relatives shared only about 90% identity at the 16S rRNA gene level, leading us to classify this a novel genus. Though within the year prior to publication, *Eisenbergiella tayi* was described and became the closest known relative, sharing a 16S rRNA gene identity of ~94% or less. This isolate was found in a human blood specimen and was strictly proteolytic with an interestingly irregular colony morphology. Very close relatives of *E. tayi* had been found in human fecal sequencing surveys, hinting that it had made its way into the blood by way of a leaky colon. While our organism, *Kineothrix alysoides*, is not too distantly related to these, its ecology and lifestyle seem to be quite different. *K. alysoides* is a saccharolytic butyrate producer, also capable of nitrogen fixation. Butyrate production seems mostly relegated to gut and rumen mutualists (Duncan et al., 2004, 2002; Eeckhaut et al., 2010; Meehan and Beiko, 2014), even having been shown to correlate with mucin-degrading commensals (Van den Abbeele et al., 2013). The need for nitrogen fixation pathways, however, seems unnecessary in the protein-rich human gut. This metabolism has been well studied in the gut of foraging insects: the relatively nutrient poor woody biomass that termites consume leads to a specialized consortium capable of nitrogen fixation which enables amino acid production. Another habitat often associated with nitrogen fixation is the soil associated with plant roots, the rhizosphere. Its evasion until now makes it difficult to say what its natural habitat is, but its rarity in the soil does not seem to be a lifestyle choice, since better growth conditions allow it to flourish.

In our analysis of the other microbes present in the microcosm and their possible role in the community function and dynamics, we were brought to a group of closely

related taxa in the *Lachnospiraceae*. The oldest representative of the group was one of the first representatives of the whole family, *Clostridium sphenoides*. It was so-named because of its distinctive shape during sporulation, with the name being Greek for “wedge-like.” This characteristic seemed to be shared by the clade, but otherwise the descriptions were sparse and not standardized. Misclassification made the situation even more convoluted, with some of the taxa being more difficult to find and subsequently left out of many phylogenies or comparisons. A sequencing grant through the Joint Genome Institute (JGI) allowed us to do better comparisons and ultimately more accurately classify them. One of the taxa was originally named *Clostridium methoxybenzovorans* by Mechichi *et al.* in 1999. Mechichi added several taxa to the bacterial tree of life and brought an interesting perspective to the field (Mechichi *et al.*, 2005; Tahar Mechichi *et al.*, 1999; T. Mechichi *et al.*, 1999; Mechichi T *et al.*, 1999), but unfortunately *Clostridium indolis* was overlooked in the description of *C. methoxybenzovorans*. The published 16S rRNA sequences for the two organisms are almost identical, though these are not always representative of genomic similarity. Yet in this case, the genomes were also almost identical, sharing >99% average nucleotide identity and seemingly all phenotypic traits, including the metabolism that *C. methoxybenzovorans* was named for. Given the lack of distinguishing features, we have emended the description of *C. methoxybenzovorans* to state that it is a later heterotypic synonym of *C. indolis*. This work also shed light on an interesting characteristic that may be shared by this group, but that could easily go unnoticed due its lack of a unique genetic basis: the *O*-demethylation of aromatic compounds. This is of great interest to field of nutrition and cancer biology, since conversion of methyl esters to hydroxyl groups enhances the estrogenic effects of

phytoestrogens in edible plants like beans and seeds. The greater the potency of the phytoestrogens, the greater the antioncogenic effect.

While a step in the right direction, this did not fix the nomenclatural issues with the rest of the clade. And, due to their taxonomic standing, these species were being left out of family-level analysis altogether. In working on genus level reclassification for the *C. sphenoides* clade, we realized that a closely related clade was also misnamed, though distinct due to phylogeny and ecology. This group's type species is *Clostridium clostridioforme*, originally described in 1958 and renamed several times since. Most of the other taxa in the group were discovered in the 2000s, isolated from clinical specimens, whereas the *C. sphenoides* group tended to be environmental isolates. These and other genomic characteristics and phenotypic differences led to the creation of two novel genera to house these taxa: *Lacriformis* and *Enterocloster*.

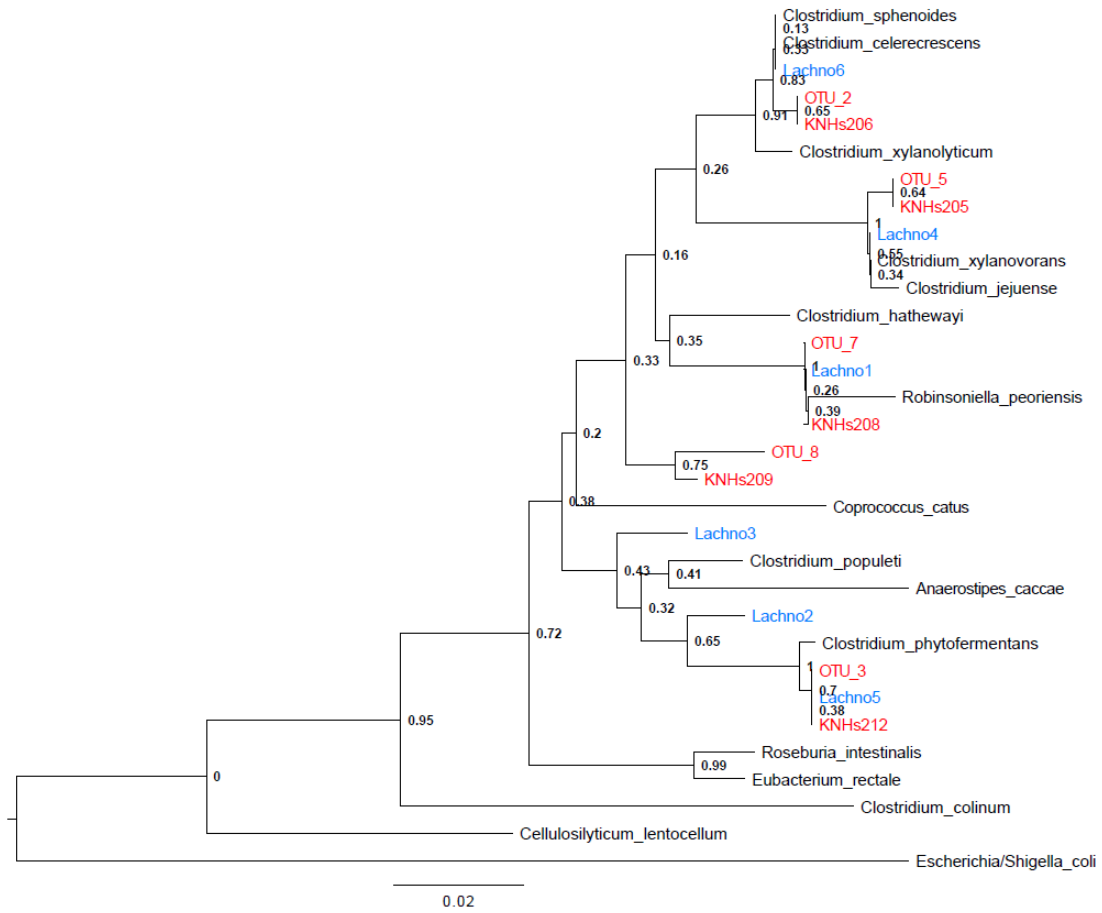
Another abundant taxon from the microcosm, initially thought to be a strain of *Clostridium phytofermentans*, turned out to be the only isolate capable of plant degradation and cellulolysis. Differences in fermentation products and cellulose degradation rates led us to a comparative genomic project between our four isolated strains and the type strain, *C. phytofermentans* ISDg. However, after genomic comparisons, we realized that these strains represented a novel species with a misleadingly similar 16S rRNA gene sequence. The genomic synteny was generally conserved, but the gene-level similarity was consistently ~85%, while hundreds of genes were present only in ISDg. The characterization of this novel taxon led us to reexamine the taxonomy of the Clostridial clade in which *C. phytofermentans* resided. While these taxa all were capable of plant degradation, either through ready use of xylan or cellulose,

they did not share many characteristics or genomic similarity. An in-depth phylogenetic analysis led to a restructuring of the clade into two novel genera: *Cellulospecium*, and *Leschinia*. Future work includes publication of the comparison between our isolates and *L. phytofermentans*, with a special emphasis on the bacterial microcompartments and choline degradation.

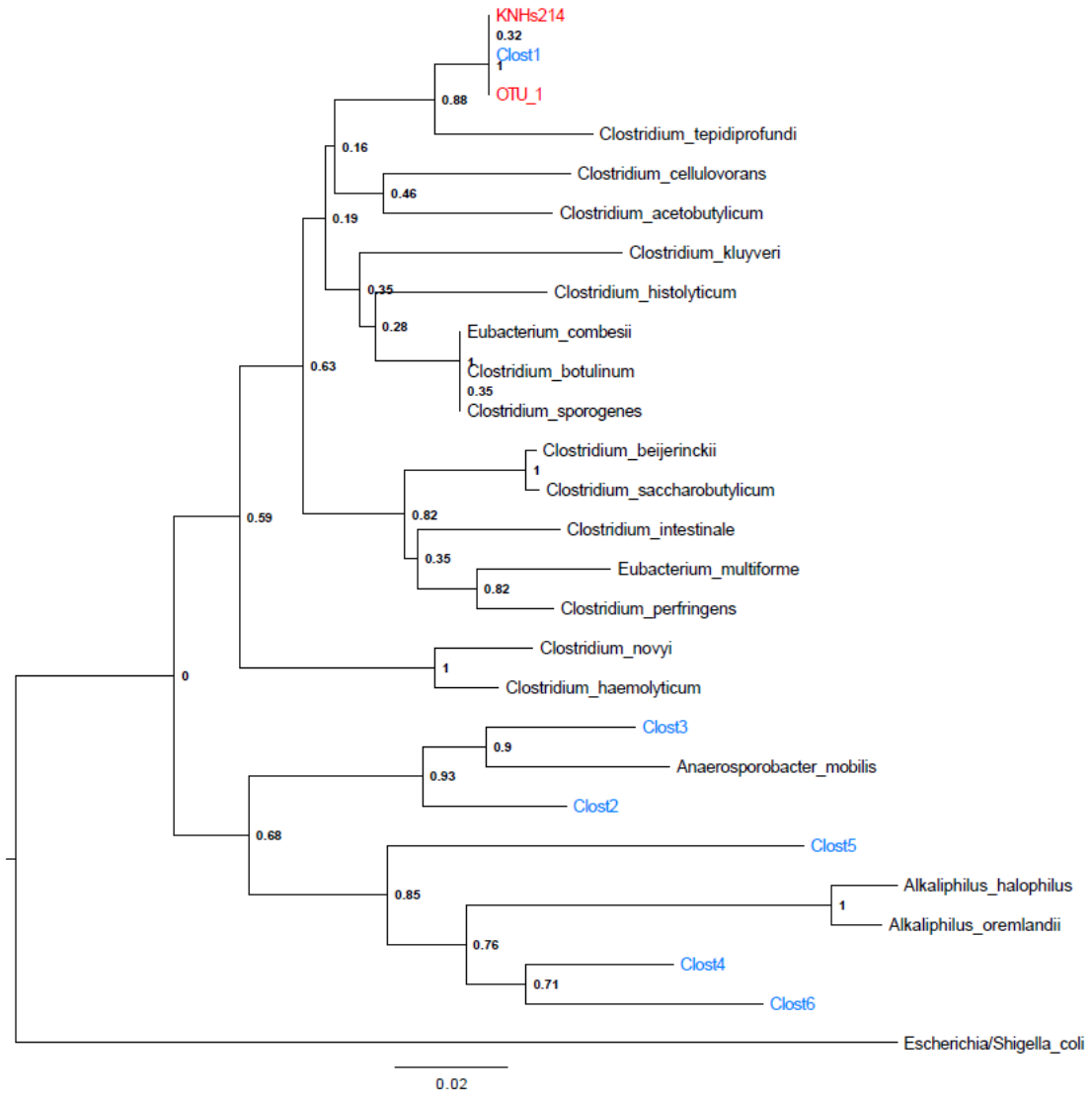
While this work has attempted to rectify the majority of the Clostridial taxonomy within the *Lachnospiraceae*, there are still many misnamed organisms. Classification within monophyletic groups is essential to the accessibility of these taxa, given the way that misclassification is currently handled. Once all *Clostridium* spp. are reclassified, the *Ruminococci* and *Eubacteria* are next in line. Both have their own *sensu stricto* families, precluding inclusion into other families, leading to their stark absence from many current taxonomies and phylogenies as well. We owe it to these forefathers of the family - the researchers and these interesting microbes - to not let them be lost.

APPENDIX

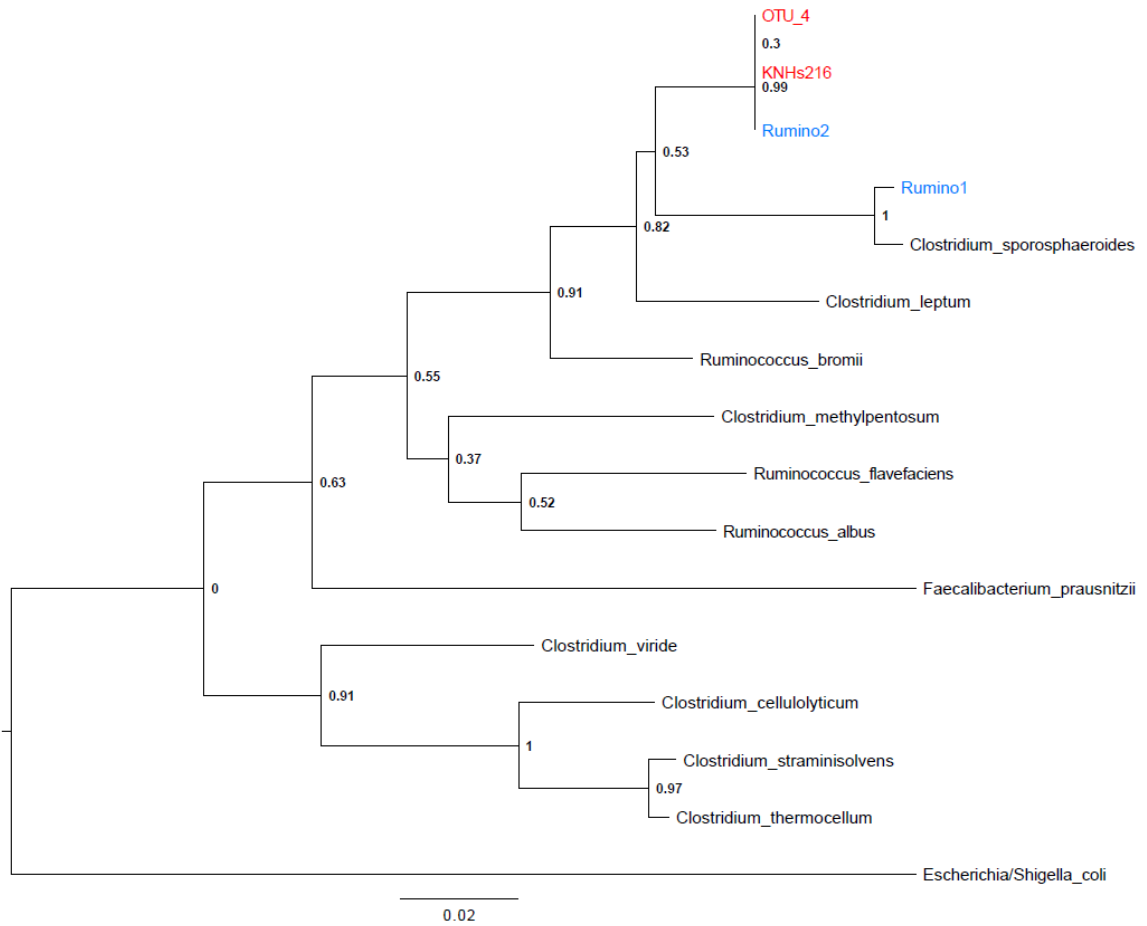
SUPPLEMENTAL FIGURES



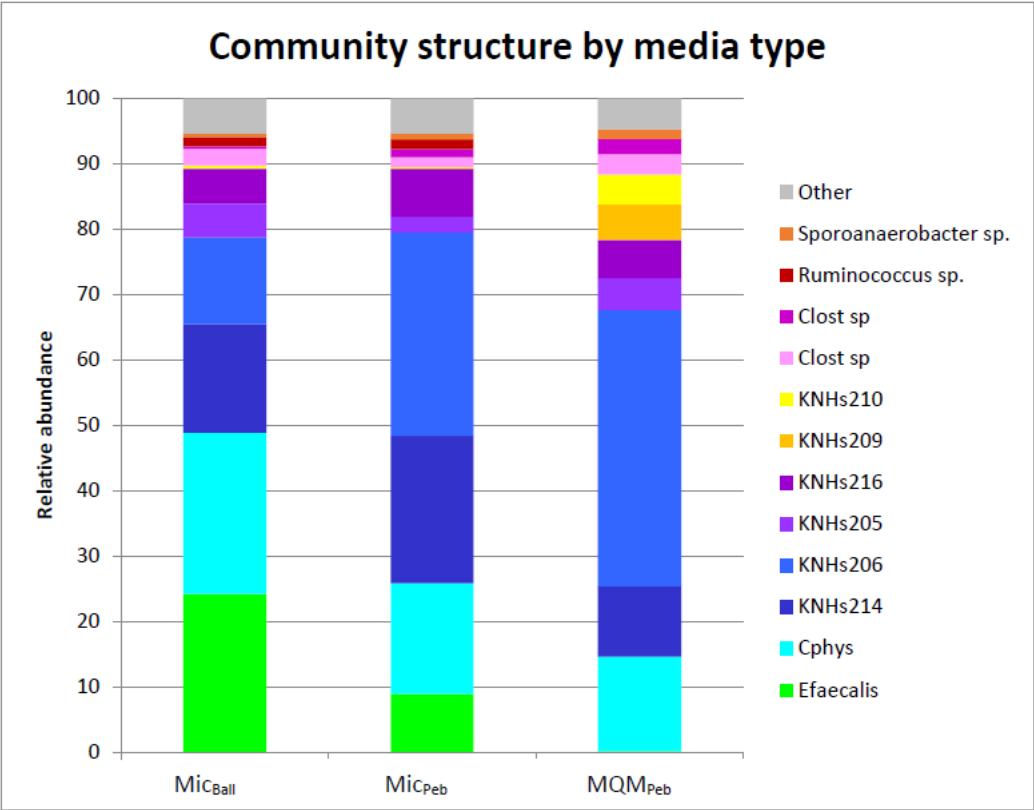
Phylogenetic tree of the Lachnospiraceae OTUs, relevant type species, and isolate 16S rRNA gene sequences using the Neighbor-joining method.



Phylogenetic tree of the Clostridiaceae OTUs, relevant type species, and isolate 16S rRNA gene sequences using the Neighbor-joining method.



Phylogenetic tree of the Ruminococcaceae OTUs, relevant type species, and isolate 16S rRNA gene sequences using the Neighbor-joining method.



Media effect on community structure. Isolates begin with “KNH” but also include “Cphys” and “Efaecalis.” Mic medium contains a small amount of yeast extract, while MQM is completely defined. “Ball” refers to dry ball-milled switchgrass, whereas “Peb” refers to wet pebble-milled switchgrass.

REFERENCES

- AbdulRahman, F., 2013. The Distribution of Polyhedral Bacterial Microcompartments Suggests Frequent Horizontal Transfer and Operon Reassembly. *J. Phylogenetics Evol. Biol.* 1. doi:10.4172/2329-9002.1000118
- Abreu, M.T., Peek Jr., R.M., 2014. Gastrointestinal Malignancy and the Microbiome. *Gastroenterology, The Gut Microbiome in Health and Disease* 146, 1534–1546.e3. doi:10.1053/j.gastro.2014.01.001
- Adékambi, T., Shinnick, T.M., Raoult, D., Drancourt, M., 2008. Complete *rpoB* gene sequencing as a suitable supplement to DNA–DNA hybridization for bacterial species and genus delineation. *Int. J. Syst. Evol. Microbiol.* 58, 1807–1814. doi:10.1099/ijs.0.65440-0
- Adlercreutz, H., 1995. Phytoestrogens: Epidemiology and a Possible Role in Cancer Protection. *Environ. Health Perspect.* 103, 103–112. doi:10.2307/3432518
- Aguayo, P., González, C., Barra, R., Becerra, J., Martínez, M., 2013. Herbicides induce change in metabolic and genetic diversity of bacterial community from a cold oligotrophic lake. *World J. Microbiol. Biotechnol.* 30, 1101–1110. doi:10.1007/s11274-013-1530-y
- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi:10.1038/nbt.2579
- Amir, I., Bouvet, P., Legeay, C., Gophna, U., Weinberger, A., 2014. *Eisenbergiella tayi* gen. nov., sp. nov., isolated from human blood. *Int. J. Syst. Evol. Microbiol.* 64, 907–914. doi:10.1099/ijs.0.057331-0
- Arroo, R.R.J., Beresford, K., Bhabra, A.S., Boarder, M., Budriesi, R., Cheng, Z., Micucci, M., Ruparelia, K.C., Surichan, S., Androustopoulos, V.P., 2014. Phytoestrogens as natural prodrugs in cancer prevention: towards a mechanistic model. *Phytochem. Rev.* 13, 853–866. doi:10.1007/s11101-014-9355-3
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E. de, Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi:10.1093/nar/gks400

- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., Weissenbach, J., Ehrlich, S.D., Bork, P., 2011. Enterotypes of the human gut microbiome. *Nature* 473, 174–180. doi:10.1038/nature09944
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi:10.1038/75556
- Bahl, M.I., Bergström, A., Licht, T.R., 2012. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. *FEMS Microbiol. Lett.* 329, 193–197. doi:10.1111/j.1574-6968.2012.02523.x
- Bennett, S., 2004. Solexa Ltd. *Pharmacogenomics* 5, 433–438. doi:10.1517/14622416.5.4.433
- Berkhoff, H.A., 1985. *Clostridium colinum* sp. nov., nom. rev., the Causative Agent of Ulcerative Enteritis (Quail Disease) in Quail, Chickens, and Pheasants. *Int. J. Syst. Bacteriol.* 35, 155–159. doi:10.1099/00207713-35-2-155
- Bi, C., Zhang, X., Ingram, L.O., Preston, J.F., 2009. Genetic Engineering of *Enterobacter asburiae* Strain JDR-1 for Efficient Production of Ethanol from Hemicellulose Hydrolysates. *Appl. Environ. Microbiol.* 75, 5743–5749. doi:10.1128/AEM.01180-09
- Biddle, A., 2014. Anaerobic microbes and communities in the context of soil and the equine digestive tract.
- Biddle, A., Stewart, L., Blanchard, J., Leschine, S., 2013. Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. *Diversity* 5, 627–640. doi:10.3390/d5030627
- Biddle, A.S., Leschine, S., Huntemann, M., Han, J., Chen, A., Kyrpides, N., Markowitz, V., Palaniappan, K., Ivanova, N., Mikhailova, N., Ovchinnikova, G., Schaumberg, A., Pati, A., Stamatis, D., Reddy, T., Lobos, E., Goodwin, L., Nordberg, H.P., Cantor, M.N., Hua, S.X., Woyke, T., Blanchard, J.L., 2014. The complete

- genome sequence of *Clostridium indolis*. *Stand. Genomic Sci.* 9, 1089–1104. doi:10.4056/528
- Bischoff, D.S., Ordal, G.W., 1992. Identification and characterization of FliY, a novel component of the *Bacillus subtilis* flagellar switch complex. *Mol. Microbiol.* 6, 2715–2723. doi:10.1111/j.1365-2958.1992.tb01448.x
- Blanco, J., Coque, J.J.R., Martin, J.F., 1998. The Folate Branch of the Methionine Biosynthesis Pathway in *Streptomyces lividans*: Disruption of the 5,10-Methylenetetrahydrofolate Reductase Gene Leads to Methionine Auxotrophy. *J. Bacteriol.* 180, 1586–1591.
- Bradley, A.S., Leavitt, W.D., Johnston, D.T., 2011. Revisiting the dissimilatory sulfate reduction pathway. *Geobiology* 9, 446–457. doi:10.1111/j.1472-4669.2011.00292.x
- Bren, A., Welch, M., Blat, Y., Eisenbach, M., 1996. Signal Termination in Bacterial Chemotaxis: CheZ Mediates Dephosphorylation of Free Rather than Switch-Bound CheY. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10090–10093.
- Broda, D.M., Saul, D.J., Bell, R.G., Musgrave, D.R., 2000. *Clostridium algidixylanolyticum* sp. nov., a psychrotolerant, xylan-degrading, spore-forming bacterium. *Int. J. Syst. Evol. Microbiol.* 50, 623–631. doi:10.1099/00207713-50-2-623
- Brulc, J.M., Antonopoulos, D.A., Miller, M.E.B., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A., Edwards, R.E., Frank, E.D., Emerson, J.B., Wacklin, P., Coutinho, P.M., Henrissat, B., Nelson, K.E., White, B.A., 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci.* 106, 1948–1953. doi:10.1073/pnas.0806191105
- Bryant, M.P., 1959. BACTERIAL SPECIES OF THE RUMEN. *Bacteriol. Rev.* 23, 125–153.
- Bryant, M.P., Small, N., 1956a. THE ANAEROBIC MONOTRICHOUS BUTYRIC ACID-PRODUCING CURVED ROD-SHAPED BACTERIA OF THE RUMEN. *J. Bacteriol.* 72, 16–21.
- Bryant, M.P., Small, N., 1956b. Characteristics of Two New Genera of Anaerobic Curved Rods Isolated from the Rumen of Cattle. *J. Bacteriol.* 72, 22–26.
- Bryant, M.P., Small, N., Bouma, C., Robinson, I., 1958a. Studies on the Composition of the Ruminal Flora and Fauna of Young Calves. *J. Dairy Sci.* 41, 1747–1767. doi:10.3168/jds.S0022-0302(58)91160-3

- Bryant, M.P., Small, N., Bouma, C., Robinson, I.M., 1958b. CHARACTERISTICS OF RUMINAL ANAEROBIC CELLULOLYTIC COCCI AND CILLOBACTERIUM CELLULOSOLVENS N. SP. J. Bacteriol. 76, 529–537.
- Burguière, P., Auger, S., Hullo, M.-F., Danchin, A., Martin-Verstraete, I., 2004. Three Different Systems Participate in l-Cystine Uptake in *Bacillus subtilis*. J. Bacteriol. 186, 4875–4884. doi:10.1128/JB.186.15.4875-4884.2004
- Byappanahalli, M.N., Nevers, M.B., Korajkic, A., Staley, Z.R., Harwood, V.J., 2012. Enterococci in the Environment. Microbiol. Mol. Biol. Rev. MMBR 76, 685–706. doi:10.1128/MMBR.00023-12
- Cai, S., Dong, X., 2010. *Cellulosilyticum ruminicola* gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of *Clostridium lentocellum* as *Cellulosilyticum lentocellum* comb. nov. Int. J. Syst. Evol. Microbiol. 60, 845–849. doi:10.1099/ij.s.0.014712-0
- Callewaert, R., Hugas, M., Vuyst, L.D., 2000. Competitiveness and bacteriocin production of Enterococci in the production of Spanish-style dry fermented sausages. Int. J. Food Microbiol. 57, 33–42. doi:10.1016/S0168-1605(00)00228-2
- Campbell, L.L., Postgate, J.R., 1965. Classification of the spore-forming sulfate-reducing bacteria. Bacteriol. Rev. 29, 359–363.
- Cannon, G.C., Bradburne, C.E., Aldrich, H.C., Baker, S.H., Heinhorst, S., Shively, J.M., 2001. Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. Appl. Environ. Microbiol. 67, 5351–5361. doi:10.1128/AEM.67.12.5351-5361.2001
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. doi:10.1038/nmeth.f.303
- Carrier, J.-P., K'ouas, G., Bonne, I., Lozniewski, A., Mory, F., 2004. *Oribacterium sinus* gen. nov., sp. nov., within the family “Lachnospiraceae” (phylum Firmicutes). Int. J. Syst. Evol. Microbiol. 54, 1611–1615. doi:10.1099/ij.s.0.63060-0
- Carrier, J.-P., K'ouas, G., Han, X.Y., 2007. *Moryella indoligenes* gen. nov., sp. nov., an anaerobic bacterium isolated from clinical specimens. Int. J. Syst. Evol. Microbiol. 57, 725–729. doi:10.1099/ij.s.0.64705-0

- Chamkha, M., Garcia, J.L., Labat, M., 2001. Metabolism of cinnamic acids by some Clostridiales and emendation of the descriptions of *Clostridium aerotolerans*, *Clostridium celerecrescens* and *Clostridium xylanolyticum*. *Int. J. Syst. Evol. Microbiol.* 51, 2105–2111. doi:10.1099/00207713-51-6-2105
- Chapin III, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C., Diaz, S., 2000. Consequences of changing biodiversity. *Nature* 405, 234.
- Chen, S., Beeby, M., Murphy, G.E., Leadbetter, J.R., Hendrixson, D.R., Briegel, A., Li, Z., Shi, J., Tocheva, E.I., Müller, A., Dobro, M.J., Jensen, G.J., 2011. Structural diversity of bacterial flagellar motors. *EMBO J.* 30, 2972–2981. doi:10.1038/emboj.2011.186
- Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi:10.1038/nmeth.2474
- Chow, V., Nong, G., Preston, J.F., 2007. Structure, Function, and Regulation of the Aldouronate Utilization Gene Cluster from *Paenibacillus* sp. Strain JDR-2. *J. Bacteriol.* 189, 8863–8870. doi:10.1128/JB.01141-07
- Chroňáková, A., Schloter-Hai, B., Radl, V., Endesfelder, D., Quince, C., Elhottová, D., Šimek, M., Schloter, M., 2015. Response of Archaeal and Bacterial Soil Communities to Changes Associated with Outdoor Cattle Overwintering. *PLOS ONE* 10, e0135627. doi:10.1371/journal.pone.0135627
- Clavel, T., Lippman, R., Gavini, F., Doré, J., Blaut, M., 2007. *Clostridium saccharogumia* sp. nov. and *Lactonifactor longoviformis* gen. nov., sp. nov., two novel human faecal bacteria involved in the conversion of the dietary phytoestrogen secoisolariciresinol diglucoside. *Syst. Appl. Microbiol.* 30, 16–26. doi:10.1016/j.syapm.2006.02.003
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37, D141–D145. doi:10.1093/nar/gkn879
- Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., Farrow, J.A.E., 1994. The Phylogeny of the Genus *Clostridium*: Proposal of Five New Genera and Eleven New Species Combinations. *Int. J. Syst. Evol. Microbiol.* 44, 812–826. doi:10.1099/00207713-44-4-812

- Cook, A.R., Riley, P.W., Murdoch, H., Evans, P.N., McDonald, I.R., 2007. *Howardella ureilytica* gen. nov., sp. nov., a Gram-positive, coccoid-shaped bacterium from a sheep rumen. *Int. J. Syst. Evol. Microbiol.* 57, 2940–2945.
doi:10.1099/ijs.0.64819-0
- Cook, K.L., Garland, J.L., Layton, A.C., Dionisi, H.M., Levine, L.H., Sayler, G.S., 2006. Effect of Microbial Species Richness on Community Stability and Community Function in a Model Plant-Based Wastewater Processing System. *Microb. Ecol.* 52, 725–737.
- Corredoira, J., García-País, M.J., Coira, A., Rabuñal, R., García-Garrote, F., Pita, J., Rodríguez-Macías, A., Blanco, M., Lopez-Roses, L., López-Álvarez, M.J., Alonso-García, M.P., 2015. Differences between endocarditis caused by *Streptococcus bovis* and *Enterococcus* spp. and their association with colorectal cancer. *Eur. J. Clin. Microbiol. Infect. Dis.* 34, 1657–1665. doi:10.1007/s10096-015-2402-1
- Cotta, M.A., Whitehead, T.R., Falsen, E., Moore, E., Lawson, P.A., 2009. *Robinsoniella peoriensis* gen. nov., sp. nov., isolated from a swine-manure storage pit and a human clinical source. *Int. J. Syst. Evol. Microbiol.* 59, 150–155.
doi:10.1099/ijs.0.65676-0
- Crecchio, C., Gelsomino, A., Ambrosoli, R., Minati, J.L., Ruggiero, P., 2004. Functional and molecular responses of soil microbial communities under differing soil management practices. *Soil Biol. Biochem.* 36, 1873–1883.
doi:10.1016/j.soilbio.2004.05.008
- Cuív, P.Ó., Career, D.A. de, Jones, M., Klaassens, E.S., Worthley, D.L., Whitehall, V.L.J., Kang, S., McSweeney, C.S., Leggett, B.A., Morrison, M., 2011. The Effects from DNA Extraction Methods on the Evaluation of Microbial Diversity Associated with Human Colonic Tissue. *Microb. Ecol.* 61, 353–362.
- Curtis, T.P., Sloan, W.T., Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10494–10499.
doi:10.1073/pnas.142680199
- DeAngelis, K.M., Allgaier, M., Chavarria, Y., Fortney, J.L., Hugenholtz, P., Simmons, B., Sublette, K., Silver, W.L., Hazen, T.C., 2011. Characterization of Trapped Lignin-Degrading Microbes in Tropical Forest Soil. *PLoS ONE* 6, 1–9.
doi:10.1371/journal.pone.0019306
- DeAngelis, K.M., Sharma, D., Varney, R., Simmons, B.A., Isern, N.G., Markillie, L.M., Nicora, C.D., Norbeck, A.D., Taylor, R.C., Aldrich, J.T., Robinson, E.W., 2013. Evidence supporting dissimilatory and assimilatory lignin degradation in *Enterobacter lignolyticus* SCF1. *Microb. Physiol. Metab.* 4, 280.
doi:10.3389/fmicb.2013.00280

- Domingo, M.-C., Huletsky, A., Boissinot, M., Hélie, M.-C., Bernal, A., Bernard, K.A., Grayson, M.L., Picard, F.J., Bergeron, M.G., 2009. *Clostridium lavalense* sp. nov., a glycopeptide-resistant species isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 59, 498–503. doi:10.1099/ijs.0.001958-0
- Dos Santos, P.C., Fang, Z., Mason, S.W., Setubal, J.C., Dixon, R., 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* 13, 162. doi:10.1186/1471-2164-13-162
- Douglas, S.R., Fleming, A., Colebrook, L., 1917. STUDIES IN WOUND INFECTIONS.: ON THE GROWTH OF ANAEROBIC BACILLI IN FLUID MEDIA UNDER APPARENTLY AEROBIC CONDITIONS. *The Lancet*, Originally published as Volume 2, Issue 4910 190, 530–532. doi:10.1016/S0140-6736(01)52070-X
- Downes, J., Munson, M.A., Radford, D.R., Spratt, D.A., Wade, W.G., 2002. *Shuttleworthia satelles* gen. nov., sp. nov., isolated from the human oral cavity. *Int. J. Syst. Evol. Microbiol.* 52, 1469–1475. doi:10.1099/ijs.0.02122-0
- Duncan, S.H., Aminov, R.I., Scott, K.P., Louis, P., Stanton, T.B., Flint, H.J., 2006. Proposal of *Roseburia faecis* sp. nov., *Roseburia hominis* sp. nov. and *Roseburia inulinivorans* sp. nov., based on isolates from human faeces. *Int. J. Syst. Evol. Microbiol.* 56, 2437–2441. doi:10.1099/ijs.0.64098-0
- Duncan, S.H., Hold, G.L., Barcenilla, A., Stewart, C.S., Flint, H.J., 2002. *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* 52, 1615–1620. doi:10.1099/ijs.0.02143-0
- Duncan, S.H., Louis, P., Flint, H.J., 2004. Lactate-Utilizing Bacteria, Isolated from Human Feces, That Produce Butyrate as a Major Fermentation Product. *Appl. Environ. Microbiol.* 70, 5810–5817. doi:10.1128/AEM.70.10.5810-5817.2004
- Eckburg, P.B., 2005. Diversity of the Human Intestinal Microbial Flora. *Science* 308, 1635–1638. doi:10.1126/science.1110591
- Eeckhaut, V., Immerseel, F.V., Pasmans, F., Brandt, E.D., Haesebrouck, F., Ducatelle, R., Vandamme, P., 2010. *Anaerostipes butyraticus* sp. nov., an anaerobic, butyrate-producing bacterium from *Clostridium* cluster XIVa isolated from broiler chicken caecal content, and emended description of the genus *Anaerostipes*. *Int. J. Syst. Evol. Microbiol.* 60, 1108–1112. doi:10.1099/ijs.0.015289-0

- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323, 133–138. doi:10.1126/science.1162986
- Eisenhauer, N., Schulz, W., Scheu, S., Jousset, A., 2013. Niche dimensionality links biodiversity and invasibility of microbial communities. *Funct. Ecol.* 27, 282–288. doi:10.1111/j.1365-2435.2012.02060.x
- Encarnação, J.C., Abrantes, A.M., Pires, A.S., Botelho, M.F., 2015. Revisit dietary fiber on colorectal cancer: butyrate and its role on prevention and treatment. *Cancer Metastasis Rev.* 34, 465–478. doi:10.1007/s10555-015-9578-9
- Falony, G., Vlachou, A., Verbrugghe, K., De Vuyst, L., 2006. Cross-feeding between *Bifidobacterium longum* BB536 and acetate-converting, butyrate-producing colon bacteria during growth on oligofructose. *Appl. Environ. Microbiol.* 72, 7835–7841. doi:10.1128/AEM.01296-06
- Ferreira, M.J., Sá-Nogueira, I. de, 2010. A Multitask ATPase Serving Different ABC-Type Sugar Importers in *Bacillus subtilis*. *J. Bacteriol.* 192, 5312–5318. doi:10.1128/JB.00832-10
- Fisher, C.K., Mehta, P., 2014. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PloS One* 9, e102451. doi:10.1371/journal.pone.0102451
- Foulquié Moreno, M.R., Rea, M.C., Cogan, T.M., De Vuyst, L., 2003. Applicability of a bacteriocin-producing *Enterococcus faecium* as a co-culture in Cheddar cheese manufacture. *Int. J. Food Microbiol.* 81, 73–84. doi:10.1016/S0168-1605(02)00167-8
- Fox, N.E., Demaree, R.S., 1999. Quick bacterial microwave fixation technique for scanning electron microscopy. *Microsc. Res. Tech.* 46, 338–339. doi:10.1002/(SICI)1097-0029(19990815/01)46:4/5<338::AID-JEMT12>3.0.CO;2-L
- Franklin, R.B., Mills, A.L., 2006. Structural and Functional Responses of a Sewage Microbial Community to Dilution-Induced Reductions in Diversity. *Microb. Ecol.* 52, 280–288.

- Furuya, H., Ide, Y., Hamamoto, M., Asanuma, N., Hino, T., 2010. Isolation of a novel bacterium, *Blautia glucerasei* sp. nov., hydrolyzing plant glucosylceramide to ceramide. *Arch. Microbiol.* 192, 365–72.
doi:<http://dx.doi.org/silk.library.umass.edu/10.1007/s00203-010-0566-8>
- Ghaly, 2010. Extraction of Protein from Common Plant Leaves for Use as Human Food. *Am. J. Appl. Sci.* 7, 331–342. doi:10.3844/ajassp.2010.331.342
- Ghasemi, S., Ahmadian, G., Sadeghi, M., Zeigler, D.R., Rahimian, H., Ghandili, S., Naghibzadeh, N., Dehestani, A., 2011. First report of a bifunctional chitinase/lysozyme produced by *Bacillus pumilus* SG2. *Enzyme Microb. Technol.* 48, 225–231. doi:10.1016/j.enzmictec.2010.11.001
- Girvan, M.S., Campbell, C.D., Killham, K., Prosser, J.I., Glover, L.A., 2005. Bacterial diversity promotes community stability and functional resilience after perturbation. *Environ. Microbiol.* 7, 301–313. doi:10.1111/j.1462-2920.2005.00695.x
- Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S., Jaffe, D.B., 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1513–1518. doi:10.1073/pnas.1017351108
- Gogotova, G.I., Vainshtein, M.B., 1983. Spore-forming sulfate reducing bacterium *desulfotomaculum guttoideum* sp nov. *Mikrobiologiya* 52, 789–793.
- Gonzalez, J.M., Portillo, M.C., Belda-Ferre, P., Mira, A., 2012. Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities. *PLoS ONE* 7, e29973. doi:10.1371/journal.pone.0029973
- Gonzalez, J.M., Zimmermann, J., Saiz-Jimenez, C., 2005. Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics* 21, 333–337. doi:10.1093/bioinformatics/bti008
- Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi:10.1099/ijs.0.64483-0
- Greening, R.C., Leedle, J.A.Z., 1989. Enrichment and isolation of *Acetitomaculum ruminis*, gen. nov., sp. nov.: acetogenic bacteria from the bovine rumen. *Arch. Microbiol.* 151, 399–406.

- Gylswyk, N.O.V., 1980. *Fusobacterium polysaccharolyticum* sp.nov., a Gram-negative Rod from the Rumen that Produces Butyrate and Ferments Cellulose and Starch. *J. Gen. Microbiol.* 116, 157–163. doi:10.1099/00221287-116-1-157
- Gylswyk, N.O.V., Morris, E.J., Els, H.J., 1980. Sporulation and Cell Wall Structure of *Clostridium polysaccharolyticum* comb.nov. (Formerly *Fusobacterium polysaccharolyticum*). *J. Gen. Microbiol.* 121, 491–493. doi:10.1099/00221287-121-2-491
- Gylswyk, N.O.V., Toorn, J.J.T.K.V.D., 1987. *Clostridium aerotolerans* sp. nov., a Xylanolytic Bacterium from Corn Stover and from the Rumina of Sheep Fed Corn Stover. *Int. J. Syst. Bacteriol.* 37, 102–105. doi:10.1099/00207713-37-2-102
- Gylswyk, N.O.V., Toorn, J.J.T.K.V.D., 1985. *Eubacterium uniforme* sp. nov. and *Eubacterium xylanophilum* sp. nov., Fiber-Digesting Bacteria from the Rumina of Sheep Fed Corn Stover. *Int. J. Syst. Bacteriol.* 35, 323–326. doi:10.1099/00207713-35-3-323
- Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methé, B., DeSantis, T.Z., Human Microbiome Consortium, Petrosino, J.F., Knight, R., Birren, B.W., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi:10.1101/gr.112730.110
- Haleblian, S., Harris, B., Finegold, S.M., Rolfe, R.D., 1981. Rapid method that aids in distinguishing Gram-positive from Gram-negative anaerobic bacteria. *J. Clin. Microbiol.* 13, 444–448.
- Hall, I.C., 1922. Differentiation and Identification of the Sporulating Anaerobes. *J. Infect. Dis.* 30, 445–504.
- Harwood, V. j., Delahoya, N. c., Ulrich, R. m., Kramer, M. f., Whitlock, J. e., Garey, J. r., Lim, D. v., 2004. Molecular confirmation of *Enterococcus faecalis* and *E. faecium* from clinical, faecal and environmental sources. *Lett. Appl. Microbiol.* 38, 476–482. doi:10.1111/j.1472-765X.2004.01518.x
- Hedberg, M.E., Moore, E.R.B., Svensson-Stadler, L., Horstedt, P., Baranov, V., Hernell, O., Wai, S.N., Hammarstrom, S., Hammarstrom, M.-L., 2012. *Lachnoanaerobaculum* gen. nov., a new genus in the Lachnospiraceae: characterization of *Lachnoanaerobaculum umeaense* gen. nov., sp. nov., isolated from the human small intestine, and *Lachnoanaerobaculum orale* sp. nov., isolated from saliva, and reclassification of *Eubacterium saburreum* (Prevot 1966) Holdeman and Moore 1970 as *Lachnoanaerobaculum saburreum* comb. nov. *Int. J. Syst. Evol. Microbiol.* 62, 2685–2690. doi:10.1099/ijs.0.033613-0

- Hong, Y.-H., Nishimura, Y., Hishikawa, D., Tsuzuki, H., Miyahara, H., Gotoh, C., Choi, K.-C., Feng, D.D., Chen, C., Lee, H.-G., Katoh, K., Roh, S.-G., Sasaki, S., 2005. Acetate and Propionate Short Chain Fatty Acids Stimulate Adipogenesis via GPCR43. *Endocrinology* 146, 5092–5099. doi:10.1210/en.2005-0545
- Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M., Ohkuma, M., 2008. Genome of an Endosymbiont Coupling N₂ Fixation to Cellulolysis Within Protist Cells in Termite Gut. *Science* 322, 1108–1109. doi:10.1126/science.1165578
- Horn, S.J., Sørbotten, A., Synstad, B., Sikorski, P., Sørli, M., Vårum, K.M., Eijsink, V.G.H., 2006. Endo/exo mechanism and processivity of family 18 chitinases produced by *Serratia marcescens*. *FEBS J.* 273, 491–503. doi:10.1111/j.1742-4658.2005.05079.x
- Hungate, R.E., 1969. Chapter IV A Roll Tube Method for Cultivation of Strict Anaerobes, in: Ribbons, J.R.N. and D.W. (Ed.), *Methods in Microbiology*. Academic Press, pp. 117–132.
- HUNGATE, R.E., 1950. The anaerobic mesophilic cellulolytic bacteria. *Bacteriol. Rev.* 14, 1–49.
- Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi:10.1186/1471-2105-11-119
- Janssen, P.H., 2006. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi:10.1128/AEM.72.3.1719-1728.2006
- Jeon, Y.-S., Park, S.-C., Lim, J., Chun, J., Kim, B.-S., 2015. Improved pipeline for reducing erroneous identification by 16S rRNA sequences using the Illumina MiSeq platform. *J. Microbiol.* 53, 60–69. doi:10.1007/s12275-015-4601-y
- Jeong, H., Yi, H., Sekiguchi, Y., Muramatsu, M., Kamagata, Y., Chun, J., 2004. *Clostridium jejuense* sp. nov., isolated from soil. *Int. J. Syst. Evol. Microbiol.* 54, 1465–1468. doi:10.1099/ijs.0.63128-0
- Kaneuchi, C., Watanabe, K., Terada, A., Benno, Y., Mitsuoka, T., 1976. Taxonomic Study of *Bacteroides clostridiiformis* subsp. *clostridiiformis* (Burri and Ankersmit) Holdeman and Moore and of Related Organisms: Proposal of *Clostridium clostridiiformis* (Burri and Ankersmit) comb. nov. and *Clostridium symbiosum* (Stevens) comb. nov. *Int. J. Syst. Bacteriol.* 26, 195–204. doi:10.1099/00207713-26-2-195

- Kaplan, A., Reinhold, L., 1999. Co₂ Concentrating Mechanisms in Photosynthetic Microorganisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 539–570. doi:10.1146/annurev.arplant.50.1.539
- Kappes, R.M., Kempf, B., Kneip, S., Boch, J., Gade, J., Meier-Wagner, J., Bremer, E., 1999. Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Mol. Microbiol.* 32, 203–216.
- Kaur, S., Yawar, M., Kumar, P.A., Suresh, K., 2014. *Hungatella effluvii* gen. nov., sp. nov., an obligately anaerobic bacterium isolated from an effluent treatment plant, and reclassification of *Clostridium hathewayi* as *Hungatella hathewayi* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 64, 710–718. doi:10.1099/ijs.0.056986-0
- Kim, M., Roh, S.W., Bae, J., 2011. *Ruminococcus faecis* sp. nov., isolated from human faeces. *J. Microbiol.* 49, 487–91. doi:http://dx.doi.org.silk.library.umass.edu/10.1007/s12275-011-0505-7
- Kitahara, M., Takamine, F., Imamura, T., Benno, Y., 2000. Assignment of *Eubacterium* sp. VPI 12708 and related strains with high bile acid 7 α -dehydroxylating activity to *Clostridium scindens* and proposal of *Clostridium hylemonae* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 50, 971–978. doi:10.1099/00207713-50-3-971
- Klaring, K., Just, S., Lagkouvardos, I., Hanske, L., Haller, D., Blaut, M., Wenning, M., Clavel, T., 2015. *Murimonas intestini* gen. nov., sp. nov., an acetate-producing bacterium of the family Lachnospiraceae isolated from the mouse gut. *Int. J. Syst. Evol. Microbiol.* 65, 870–878. doi:10.1099/ijs.0.000030
- Kofoed, E., Rappleye, C., Stojiljkovic, I., Roth, J., 1999. The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J. Bacteriol.* 181, 5317–5329.
- Komiyama, H., Enomoto, A., Sueyoshi, Y., Nishio, T., Kato, A., Ishii, T., Shimizu, K., 2009. Structures of aldouronic acids liberated from kenaf xylan by endoxylanases from *Streptomyces* sp. *Carbohydr. Polym.* 75, 521–527. doi:10.1016/j.carbpol.2008.08.021
- Kopečný, J., Zorec, M., Mrázek, J., Kobayashi, Y., Marinšek-Logar, R., 2003. *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrvibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. *Int. J. Syst. Evol. Microbiol.* 53, 201–209. doi:10.1099/ijs.0.02345-0
- Kornacker, M.G., Pugsley, A.P., 1990. Molecular characterization of *pulA* and its product, pullulanase, a secreted enzyme of *Klebsiella pneumoniae* UNF5023. *Mol. Microbiol.* 4, 73–85. doi:10.1111/j.1365-2958.1990.tb02016.x

- Kostas lab | ANI calculator [WWW Document], n.d. URL <http://enve-omics.ce.gatech.edu/ani/> (accessed 5.23.16).
- Kotsyurbenko, O.R., Nozhevnikova, A.N., Osipov, G.A., Kostrikina, N.A., Lysenko, A.M., 1995. A New Psychroactive Bacterium *Clostridium fimetarium*, Isolated from Cattle Manure Digested at Low Temperature. *Microbiology*. 64, 681.
- Krohn, A., Stevens, B., Robbins-Pianka, A., Belus, M., Allan, G.J., Gehring, C., 2016. Optimization of 16S amplicon analysis using mock communities: implications for estimating community diversity. *PeerJ Preprints*.
- Kubori, T., Okumura, M., Kobayashi, N., Nakamura, D., Iwakura, M., Aizawa, S.-I., 1997. Purification and characterization of the flagellar hook–basal body complex of *Bacillus subtilis*. *Mol. Microbiol.* 24, 399–410. doi:10.1046/j.1365-2958.1997.3341714.x
- Kuehl, J.V., Price, M.N., Ray, J., Wetmore, K.M., Esquivel, Z., Kazakov, A.E., Nguyen, M., Kuehn, R., Davis, R.W., Hazen, T.C., Arkin, A.P., Deutschbauer, A., 2014. Functional Genomics with a Comprehensive Library of Transposon Mutants for the Sulfate-Reducing Bacterium *Desulfovibrio alaskensis* G20. *mBio* 5. doi:10.1128/mBio.01041-14
- Kunin, V., Engelbrekton, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118–123. doi:10.1111/j.1462-2920.2009.02051.x
- Kurakawa, T., Ogata, K., Matsuda, K., Tsuji, H., Kubota, H., Takada, T., Kado, Y., Asahara, T., Takahashi, T., Nomoto, K., 2015. Diversity of Intestinal *Clostridium coccoides* Group in the Japanese Population, as Demonstrated by Reverse Transcription-Quantitative PCR. *PLoS ONE* 10, e0126226. doi:10.1371/journal.pone.0126226
- Lagier, J.-C., Bibi, F., Ramasamy, D., Azhar, E.I., Robert, C., Yasir, M., Jiman-Fatani, A.A., Alshali, K.Z., Fournier, P.-E., Raoult, D., 2014. Non contiguous-finished genome sequence and description of *Clostridium jeddahense* sp. nov. *Stand. Genomic Sci.* 9, 1003–1019. doi:10.4056/sigs.5571026
- Lang, J.M., Darling, A.E., Eisen, J.A., 2013. Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes: Supertrees and Supermatrices. *PLoS ONE* 8, 1–15. doi:10.1371/journal.pone.0062510
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinforma. Oxf. Engl.* 23, 2947–2948. doi:10.1093/bioinformatics/btm404

- Lawson, P.A., Rainey, F.A., 2016. Proposal to restrict the genus *Clostridium* Prazmowski to *Clostridium butyricum* and related species. *Int. J. Syst. Evol. Microbiol.* 66, 1009–1016. doi:10.1099/ijsem.0.000824
- Lewis, N.G., Sarkanen, S., American Chemical Society., Cellulose, P., and Textile Division., American Chemical Society., Meeting, 1998. Lignin and lignan biosynthesis. American Chemical Society, Washington, DC.
- Li, Y., Chen, X., 2012. Sialic acid metabolism and sialyltransferases: natural functions and applications. *Appl. Microbiol. Biotechnol.* 94, 887–905. doi:10.1007/s00253-012-4040-1
- Liu, C., Finegold, S.M., Song, Y., Lawson, P.A., 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. *Int. J. Syst. Evol. Microbiol.* 58, 1896–1902. doi:10.1099/ijms.0.65208-0
- Lomans, B.P., Leijdekkers, P., Wesselink, J.-J., Bakkes, P., Pol, A., Drift, C. van der, Camp, H.J.M.O. den, 2001. Obligate Sulfide-Dependent Degradation of Methoxylated Aromatic Compounds and Formation of Methanethiol and Dimethyl Sulfide by a Freshwater Sediment Isolate, *Parasporobacterium paucivorans* gen. nov., sp. nov. *Appl. Environ. Microbiol.* 67, 4017–4023. doi:10.1128/AEM.67.9.4017-4023.2001
- Louis, P., Scott, K.P., Duncan, S.H., Flint, H.J., 2007. Understanding the effects of diet on bacterial metabolism in the large intestine. *J. Appl. Microbiol.* 102, 1197–1208. doi:10.1111/j.1365-2672.2007.03322.x
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
- LPSN - List of Prokaryotic names with Standing in Nomenclature [WWW Document], n.d. URL <http://www.bacterio.net/> (accessed 6.13.16).
- Mabrok, H.B., Klopffleisch, R., Ghanem, K.Z., Clavel, T., Blaut, M., Loh, G., 2012. Lignan transformation by gut bacteria lowers tumor burden in a gnotobiotic rat model of breast cancer. *Carcinogenesis* 33, 203–208. doi:10.1093/carcin/bgr256
- Magrane, M., Consortium, U., 2011. UniProt Knowledgebase: a hub of integrated protein data. Database 2011, bar009. doi:10.1093/database/bar009

- Mangels, J.I., Cox, M.E., Lindberg, L.H., 1984. Methanol fixation an alternative to heat fixation of smears before staining. *Diagn. Microbiol. Infect. Dis.* 2, 129–137. doi:10.1016/0732-8893(84)90008-7
- Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I.-M.A., Chu, K., Kyrpides, N.C., 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25, 2271–2278. doi:10.1093/bioinformatics/btp393
- Marvaud, J.-C., Mory, F., Lambert, T., 2011. *Clostridium clostridioforme* and *Atopobium minutum* clinical isolates with vanB-type resistance in France. *J. Clin. Microbiol.* 49, 3436–3438. doi:10.1128/JCM.00308-11
- Maukonen, J., Simões, C., Saarela, M., 2012. The currently used commercial DNA-extraction methods give different results of clostridial and actinobacterial populations derived from human fecal samples. *FEMS Microbiol. Ecol.* 79, 697–708. doi:10.1111/j.1574-6941.2011.01257.x
- McBride, S.M., Fischetti, V.A., LeBlanc, D.J., Jr, R.C.M., Gilmore, M.S., 2007. Genetic Diversity among *Enterococcus faecalis*. *PLOS ONE* 2, e582. doi:10.1371/journal.pone.0000582
- McClung, L., McCoy, E., Prazmovski, A., 1957. Bergey's Manual of Determinative Bacteriology, in: *Bergey's Manual of Determinative Bacteriology*. Williams and Wilkins, Baltimore, pp. 634–693.
- McDonald, J.E., Houghton, J.N.I., Rooks, D.J., Allison, H.E., McCarthy, A.J., 2012. The microbial ecology of anaerobic cellulose degradation in municipal waste landfill sites: evidence of a role for fibrobacters. *Environ. Microbiol.* 14, 1077–1087. doi:10.1111/j.1462-2920.2011.02688.x
- McInerney, M.J., Sieber, J.R., Gunsalus, R.P., 2009. Syntrophy in anaerobic global carbon cycles. *Curr. Opin. Biotechnol., Chemical biotechnology • Pharmaceutical biotechnology* 20, 623–632. doi:10.1016/j.copbio.2009.10.001
- Mechichi, T., Labat, M., Garcia, J.-L., Thomas, P., Patel, B.K.C., 1999. Characterization of a New Xylanolytic Bacterium, *Clostridium xylanovorans* sp. nov. *Syst. Appl. Microbiol.* 22, 366–371. doi:10.1016/S0723-2020(99)80044-7
- Mechichi, T., Labat, M., Patel, B.K.C., Woo, T.H.S., Thomas, P., Garcia, J.-L., 1999. *Clostridium methoxybenzovorans* sp. nov., a new aromatic o-demethylating homoacetogen from an olive mill wastewater treatment digester. *Int. J. Syst. Bacteriol.* 49, 1201–1209. doi:10.1099/00207713-49-3-1201

- Mechichi, T., Patel, B.K.C., Sayadi, S., 2005. Anaerobic degradation of methoxylated aromatic compounds by *Clostridium methoxybenzovorans* and a nitrate-reducing bacterium *Thauera* sp. strain Cin3,4. *Int. Biodeterior. Biodegrad.* 56, 224–230. doi:10.1016/j.ibiod.2005.09.001
- Mechichi T, Labat M, Garcia JL, Thomas P, Patel BK, 1999. *Sporobacterium olearium* gen. nov., sp. nov., a new methanethiol-producing bacterium that degrades aromatic compounds, isolated from an olive mill wastewater treatment digester. *Int. J. Syst. Bacteriol.* 49, 1741–8.
- Meehan, C.J., Beiko, R.G., 2014. A Phylogenomic View of Ecological Specialization in the Lachnospiraceae, a Family of Digestive Tract-Associated Bacteria. *Genome Biol. Evol.* 6, 703–713. doi:10.1093/gbe/evu050
- Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013a. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14, 60. doi:10.1186/1471-2105-14-60
- Meier-Kolthoff, J.P., Göker, M., Spröer, C., Klenk, H.-P., 2013b. When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* 195, 413–418. doi:10.1007/s00203-013-0888-4
- Miao, V., Davies, J., 2010. Actinobacteria: the good, the bad, and the ugly. *Antonie Van Leeuwenhoek* 98, 143–150. doi:10.1007/s10482-010-9440-6
- Micard, V., Landazuri, T., Surget, A., Moukha, S., Labat, M., Rouau, X., 2002. Demethylation of Ferulic Acid and Feruloyl-arabinoxylan by Microbial Cell Extracts. *LWT - Food Sci. Technol.* 35, 272–276. doi:10.1006/fstl.2001.0856
- Mingkun, L., Copeland, A., Han, J., 2011. DUK.
- Minnerath, J.M., Roland, J.M., Rossi, L.C., Weishalla, S.R., Wolf, M.M., 2009. A Comparison of Heat versus Methanol Fixation for Gram Staining Bacteria. *Bioscene J. Coll. Biol. Teach.* 35, 36–41.
- Mitsui, R., Hirota, M., Tsuno, T., Tanaka, M., 2010. Purification and characterization of vanillin dehydrogenases from alkaliphile *Micrococcus* sp. TA1 and neutrophile *Burkholderia cepacia* TM1. *FEMS Microbiol. Lett.* 303, 41–47. doi:10.1111/j.1574-6968.2009.01859.x
- Mohan, R., Namsolleck, P., Lawson, P.A., Osterhoff, M., Collins, M.D., Alpert, C.-A., Blaut, M., 2006. *Clostridium asparagiforme* sp. nov., isolated from a human faecal sample. *Syst. Appl. Microbiol.* 29, 292–299. doi:10.1016/j.syapm.2005.11.001

- Mori, K., Iino, T., Suzuki, K.-I., Yamaguchi, K., Kamagata, Y., 2012. Aceticlastic and NaCl-Requiring Methanogen “*Methanosaeta pelagica*” sp. nov., Isolated from Marine Tidal Flat Sediment. *Appl. Environ. Microbiol.* 78, 3416–3423. doi:10.1128/AEM.07484-11
- Morris, B.E.L., Henneberger, R., Huber, H., Moissl-Eichinger, C., 2013. Microbial syntrophy: interaction for the common good. *FEMS Microbiol. Rev.* 37, 384–406. doi:10.1111/1574-6976.12019
- Müller, A.K., Westergaard, K., Christensen, S., Sørensen, S.J., 2002. The Diversity and Function of Soil Microbial Communities Exposed to Different Disturbances. *Microb. Ecol.* 44, 49–58.
- Müller, B., Sun, L., Schnürer, A., 2013. First insights into the syntrophic acetate-oxidizing bacteria – a genetic study. *MicrobiologyOpen* 2, 35–53. doi:10.1002/mbo3.50
- Murray, W.D., Khan, A.W., Berg, L. van den, 1982. *Clostridium saccharolyticum* sp. nov., a Saccharolytic Species from Sewage Sludge. *Int. J. Syst. Bacteriol.* 32, 132–135. doi:10.1099/00207713-32-1-132
- Naidu, D., Ragsdale, S.W., 2001. Characterization of a Three-Component Vanillate O-Demethylase from *Moorella thermoacetica*. *J. Bacteriol.* 183, 3276–3281. doi:10.1128/JB.183.11.3276-3281.2001
- Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25, 1335–1337. doi:10.1093/bioinformatics/btp157
- Nei, M., Kumar, S., 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford; New York.
- Neufeld, J.D., Li, J., Mohn, W.W., 2008. Scratching the surface of the rare biosphere with ribosomal sequence tag primers. *FEMS Microbiol. Lett.* 283, 146–153. doi:10.1111/j.1574-6968.2008.01124.x
- Neufeld, J.D., Schäfer, H., Cox, M.J., Boden, R., McDonald, I.R., Murrell, J.C., 2007. Stable-isotope probing implicates *Methylophaga* spp and novel Gammaproteobacteria in marine methanol and methylamine metabolism. *ISME J.* 1, 480–491. doi:10.1038/ismej.2007.65
- Ogah, K., Sethi, K., Karthik, V., 2012. *Clostridium clostridioforme* liver abscess complicated by portal vein thrombosis in childhood. *J. Med. Microbiol.* 61, 297–299. doi:10.1099/jmm.0.031765-0

- Pacheco, A.R., Curtis, M.M., Ritchie, J.M., Munera, D., Waldor, M.K., Moreira, C.G., Sperandio, V., 2012. Fructose sensing regulates bacterial intestinal colonization. *Nature* 492, 113–117. doi:10.1038/nature11623
- Pallen, M.J., Penn, C.W., Chaudhuri, R.R., 2005. Bacterial flagellar diversity in the post-genomic era. *Trends Microbiol.* 13, 143–149. doi:10.1016/j.tim.2005.02.008
- Palop, M.L., Valles, S., Piñaga, F., Flors, A., 1989. Isolation and Characterization of an Anaerobic, Cellulolytic Bacterium, *Clostridium celerecrescens* sp. nov. *Int. J. Syst. Bacteriol.* 39, 68–71. doi:10.1099/00207713-39-1-68
- Pandey, V.N., Srivastava, A.K., 1991. Yield and nutritional quality of leaf protein concentrate from *Eleocharis dulcis* (Burm. f.) Hensch. *Aquat. Bot.* 41, 369–374. doi:10.1016/0304-3770(91)90054-9
- Park, S.-K., Kim, M.-S., Bae, J.-W., 2013. *Blautia faecis* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 63, 599–603. doi:10.1099/ijs.0.036541-0
- Parkinson, J.S., 1978. Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. *J. Bacteriol.* 135, 45–53.
- Parshina, S.N., Kleerebezem, R., Sanz, J.L., Lettinga, G., Nozhevnikova, A.N., Kostrikina, N.A., Lysenko, A.M., Stams, A.J.M., 2003. *Soehngenia saccharolytica* gen. nov., sp. nov. and *Clostridium amygdalinum* sp. nov., two novel anaerobic, benzaldehyde-converting bacteria. *Int. J. Syst. Evol. Microbiol.* 53, 1791–1799. doi:10.1099/ijs.0.02668-0
- Parshina, S.N., Kleerebezem, R., van Kempen, E., Nozhevnikova, A.N., Lettinga, G., Stams, A.J.M., 2000. Benzaldehyde conversion by two anaerobic bacteria isolated from an upflow anaerobic sludge bed reactor. *Process Biochem.* 36, 423–429. doi:10.1016/S0032-9592(00)00230-2
- Pati, A., Ivanova, N.N., Mikhailova, N., Ovchinnikova, G., Hooper, S.D., Lykidis, A., Kyrpides, N.C., 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* 7, 455–457. doi:10.1038/nmeth.1457
- Penrod, J.T., Roth, J.R., 2006. Conserving a Volatile Metabolite: a Role for Carboxysome-Like Organelles in *Salmonella enterica*. *J. Bacteriol.* 188, 2865–2874. doi:10.1128/JB.188.8.2865-2874.2006
- Pequegnat, B., Sagermann, M., Valliani, M., Toh, M., Chow, H., Allen-Vercoe, E., Monteiro, M.A., 2013. A vaccine and diagnostic target for *Clostridium bolteae*, an autism-associated bacterium. *Vaccine* 31, 2787–2790. doi:10.1016/j.vaccine.2013.04.018

- Pester, M., Bittner, N., Deevong, P., Wagner, M., Loy, A., 2010. A “rare biosphere” microorganism contributes to sulfate reduction in a peatland. *ISME J.* 4, 1591–1602. doi:10.1038/ismej.2010.75
- Peterson, J., Dwyer, J., Adlercreutz, H., Scalbert, A., Jacques, P., McCullough, M.L., 2010. Dietary lignans: physiology and potential for cardiovascular disease risk reduction. *Nutr. Rev.* 68, 571–603. doi:10.1111/j.1753-4887.2010.00319.x
- Petit, E., LaTouf, W.G., Coppi, M.V., Warnick, T.A., Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J.G., Wardman, C., Schnell, D.J., Leschine, S.B., Blanchard, J.L., 2013. Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*. *PLoS ONE* 8, e54337. doi:10.1371/journal.pone.0054337
- Pierce, E., Xie, G., Barabote, R.D., Saunders, E., Han, C.S., Detter, J.C., Richardson, P., Brettin, T.S., Das, A., Ljungdahl, L.G., Ragsdale, S.W., 2008. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ. Microbiol.* 10, 2550–2573. doi:10.1111/j.1462-2920.2008.01679.x
- Pitts, A.C., Tuck, L.R., Faulds-Pain, A., Lewis, R.J., Marles-Wright, J., 2012. Structural insight into the *Clostridium difficile* ethanolamine utilisation microcompartment. *PLoS One* 7, e48360. doi:10.1371/journal.pone.0048360
- Plants :the potentials for extracting protein, medicines, and other useful chemicals : workshop proceedings., 1983. . Washington, D.C. :
- Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S., Daily, G., Castilla, J.C., Lubchenco, J., Paine, R.T., 1996. Challenges in the Quest for Keystones. *BioScience* 46, 609–620. doi:10.2307/1312990
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi:10.1093/nar/gkm864
- Rafrafi, Y., Trably, E., Hamelin, J., Latrille, E., Meynial-Salles, I., Benomar, S., Giudici-Orticoni, M.-T., Steyer, J.-P., 2013. Sub-dominant bacteria as keystone species in microbial communities producing bio-hydrogen. *Int. J. Hydrog. Energy* 38, 4975–4985. doi:10.1016/j.ijhydene.2013.02.008
- Rappé, M.S., Giovannoni, S.J., 2003. The Uncultured Microbial Majority. *Annu. Rev. Microbiol.* 57, 369–394. doi:10.1146/annurev.micro.57.030502.090759
- Reich, P.B., Tilman, D., Isbell, F., Mueller, K., Hobbie, S.E., Flynn, D.F.B., Eisenhauer, N., 2012. Impacts of Biodiversity Loss Escalate Through Time as Redundancy Fades. *Science* 336, 589–592. doi:10.1126/science.1217909

- Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* 106, 19126–19131. doi:10.1073/pnas.0906412106
- Richter, M., Rosselló-Móra, R., Glöckner, F.O., Peplies, J., 2015. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 31, 1654–1658. doi:10.1093/bioinformatics/btv681
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., Gelfand, M.S., 2004. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Res.* 32, 3340–3353. doi:10.1093/nar/gkh659
- Rogers, G.M., Baecker, A. a. W., 1991. *Clostridium xylanolyticum* sp. nov., an Anaerobic Xylanolytic Bacterium from Decayed *Pinus patula* Wood Chips. *Int. J. Syst. Bacteriol.* 41, 140–143. doi:10.1099/00207713-41-1-140
- Ross, M.C., Muzny, D.M., McCormick, J.B., Gibbs, R.A., Fisher-Hoch, S.P., Petrosino, J.F., 2015. 16S gut community of the Cameron County Hispanic Cohort. *Microbiome* 3, 7. doi:10.1186/s40168-015-0072-y
- Ryu, H.-W., Kang, K.-H., Pan, J.-G., Chang, H.-N., 2001. Characteristics and glycerol metabolism of fumarate-reducing *Enterococcus faecalis* RKY1. *Biotechnol. Bioeng.* 72, 119–124. doi:10.1002/1097-0290(20010105)72:1<119::AID-BIT15>3.0.CO;2-0
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sakuma K, Kitahara M, Kibe R, Sakamoto M, Benno Y, 2006. *Clostridium glycyrrhizinilyticum* sp. nov., a glycyrrhizin-hydrolysing bacterium isolated from human faeces. *Microbiol. Immunol.* 50, 481–5.
- Schwartz, A., Hold, G.L., Duncan, S.H., Gruhl, B., Collins, M.D., Lawson, P.A., Flint, H.J., Blaut, M., 2002. *Anaerostipes caccae* gen. nov., sp. nov., a New Saccharolytic, Acetate-utilising, Butyrate-producing Bacterium from Human Faeces. *Syst. Appl. Microbiol.* 25, 46–51. doi:10.1078/0723-2020-00096
- Shade, A., Hogan, C.S., Klimowicz, A.K., Linske, M., McManus, P.S., Handelsman, J., 2012. Culturing captures members of the soil rare biosphere. *Environ. Microbiol.* 14, 2247–2252. doi:10.1111/j.1462-2920.2012.02817.x
- Sharon, N., 1967. The Chemical Structure of Lysozyme Substrates and Their Cleavage by the Enzyme. *Proc. R. Soc. Lond. B Biol. Sci.* 167, 402–415.

- Shi, W., Xie, S., Chen, X., Sun, S., Zhou, X., Liu, L., Gao, P., Kyrpides, N.C., No, E.-G., Yuan, J.S., 2013. Comparative Genomic Analysis of the Endosymbionts of Herbivorous Insects Reveals Eco-Environmental Adaptations: Biotechnology Applications. *PLoS Genet* 9, e1003131. doi:10.1371/journal.pgen.1003131
- Sleat, R., Mah, R.A., 1985. *Clostridium populeti* sp. nov., a Cellulolytic Species from a Woody-Biomass Digester. *Int. J. Syst. Bacteriol.* 35, 160–163. doi:10.1099/00207713-35-2-160
- Sleat, R., Robinson, J. p., 1984. The bacteriology of anaerobic degradation of aromatic compounds. *J. Appl. Bacteriol.* 57, 381–394. doi:10.1111/j.1365-2672.1984.tb01404.x
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci.* 103, 12115–12120. doi:10.1073/pnas.0605127103
- Song, Y., Liu, C., Molitoris, D.R., Tomzynski, T.J., Lawson, P.A., Collins, M.D., Finegold, S.M., 2003. *Clostridium bolteae* sp. nov., Isolated from Human Sources. *Syst. Appl. Microbiol.* 26, 84–89. doi:10.1078/072320203322337353
- Sonoki, T., Otsuka, Y., Ikeda, S., Masai, E., Kajita, S., Katayama, Y., 2002. Tetrahydrofolate-dependent vanillate and syringate O-demethylation links tightly to one-carbon metabolic pathway associated with amino acid synthesis and DNA methylation in the lignin metabolism of *Sphingomonas paucimobilis* SYK-6. *J. Wood Sci.* 48, 434–439. doi:10.1007/BF00770706
- Spiegelman, D., Whissell, G., Greer, C.W., 2005. A survey of the methods for the characterization of microbial consortia and communities. *Can. J. Microbiol.* 51, 355–386.
- Stackebrandt, E., Kramer, I., Swiderski, J., Hippe, H., 1999. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. *FEMS Immunol. Med. Microbiol.* 24, 253–258. doi:10.1111/j.1574-695X.1999.tb01291.x
- Stackebrandt, E., Sproer, C., Rainey, F.A., Burghardt, J., Päuker, O., Hippe, H., 1997. Phylogenetic Analysis of the Genus *Desulfotomaculum*: Evidence for the Misclassification of *Desulfotomaculum guttoideum* and Description of *Desulfotomaculum orientis* as *Desulfosporosinus orientis* gen. nov., comb. nov. *Int. J. Syst. Bacteriol.* 47, 1134–1139. doi:10.1099/00207713-47-4-1134
- Stahl, M., Friis, L.M., Nothaft, H., Liu, X., Li, J., Szymanski, C.M., Stintzi, A., 2011. 1-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc. Natl. Acad. Sci.* 108, 7194–7199. doi:10.1073/pnas.1014125108

- Steer, T., Collins, M.D., Gibson, G.R., Hippe, H., Lawson, P.A., 2001. *Clostridium hathewayi* sp. nov., from Human Faeces. *Syst. Appl. Microbiol.* 24, 353–357. doi:10.1078/0723-2020-00044
- Stevenson, B.S., Eichorst, S.A., Wertz, J.T., Schmidt, T.M., Breznak, J.A., 2004. New Strategies for Cultivation and Detection of Previously Uncultured Microbes. *Appl. Environ. Microbiol.* 70, 4748–4755. doi:10.1128/AEM.70.8.4748-4755.2004
- Struijs, K., Vincken, J.-P., Gruppen, H., 2009. Bacterial conversion of secoisolariciresinol and anhydrosecoisolariciresinol. *J. Appl. Microbiol.* 107, 308–317. doi:10.1111/j.1365-2672.2009.04209.x
- Sun, L., Pope, P.B., Eijsink, V.G.H., Schnürer, A., 2015. Characterization of microbial community structure during continuous anaerobic digestion of straw and cow manure. *Microb. Biotechnol.* 8, 815–827. doi:10.1111/1751-7915.12298
- Szurmant, H., Bunn, M.W., Cannistraro, V.J., Ordal, G.W., 2003. *Bacillus subtilis* Hydrolyzes CheY-P at the Location of Its Action, the Flagellar Switch. *J. Biol. Chem.* 278, 48611–48616. doi:10.1074/jbc.M306180200
- Szurmant, H., Ordal, G.W., 2004. Diversity in Chemotaxis Mechanisms among the Bacteria and Archaea. *Microbiol. Mol. Biol. Rev.* 68, 301–319. doi:10.1128/MMBR.68.2.301-319.2004
- Tailford, L.E., Crost, E.H., Kavanaugh, D., Juge, N., 2015. Mucin glycan foraging in the human gut microbiome. *Nutrigenomics* 6, 81. doi:10.3389/fgene.2015.00081
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11030–11035. doi:10.1073/pnas.0404206101
- 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. *Mol. Biol. Evol.* 28, 2731–2739. doi:10.1093/molbev/msr121
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–9.
- Taras, D., Simmering, R., Collins, M.D., Lawson, P.A., Blaut, M., 2002. Reclassification of *Eubacterium formicigenerans* Holdeman and Moore 1974 as *Dorea formicigenerans* gen. nov., comb. nov., and description of *Dorea longicatena* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 52, 423–428.

- The enterococci: pathogenesis, molecular biology and antibiotic resistance and infection control., 2002. . ASM ; Blackwell, Washington, D.C.; Oxford.
- The Integrated Microbial Genomes (IMG) platform., n.d.
- Tissier, H., 1908. Recherches sur la flore intestinale normale des enfants âgés d'un à cinq ans. *Ann. Inst. Pasteur* 22, 189–208.
- Torsvik, V., Øvreås, L., Thingstad, T.F., 2002. Prokaryotic Diversity—Magnitude, Dynamics, and Controlling Factors. *Science* 296, 1064–1066.
- Touyama, M., Jin, J. s., Kibe, R., Hayashi, H., Benno, Y., 2015. Quantification of *Blautia wexlerae* and *Blautia luti* in human faeces by real-time PCR using specific primers. *Benef. Microbes* 6, 583–590. doi:10.3920/BM2014.0133
- Trevors, J.T., 1996. DNA in soil: adsorption, genetic transformation, molecular evolution and genetic microchip. *Antonie Van Leeuwenhoek* 70, 1–10. doi:10.1007/BF00393564
- Ueki, A., Ohtaki, Y., Kaku, N., Ueki, K., 2016. Descriptions of *Anaerotaenia torta* gen. nov., sp. nov. and *Anaerocolumna cellulositytica* gen. nov., sp. nov. isolated from a methanogenic reactor of cattle waste and reclassification of *Clostridium aminovalericum*, *Clostridium jejuense* and *Clostridium xylanovorans* as *Anaerocolumna* species. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/ijsem.0.001123
- Van den Abbeele, P., Belzer, C., Goossens, M., Kleerebezem, M., De Vos, W.M., Thas, O., De Weirtd, R., Kerckhof, F.-M., Van de Wiele, T., 2013. Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J.* 7, 949–961. doi:10.1038/ismej.2012.158
- Vanharanta, M., Voutilainen, S., Lakka, T.A., van der Lee, M., Adlercreutz, H., Salonen, J.T., 1999. Risk of acute coronary events according to serum concentrations of enterolactone: a prospective population-based case-control study. *Lancet* 354, 2112–2115.
- Varel, V.H., Tanner, R.S., Woese, C.R., 1995. *Clostridium herbivorans* sp. nov., a Cellulolytic Anaerobe from the Pig Intestine. *Int. J. Syst. Bacteriol.* 45, 490–494. doi:10.1099/00207713-45-3-490
- Vermeiren, J., Abbeele, P.V. den, Laukens, D., Vignæs, L.K., Vos, M.D., Boon, N., Wiele, T.V. de, 2012. Decreased colonization of fecal *Clostridium coccoides*/*Eubacterium rectale* species from ulcerative colitis patients in an in vitro dynamic gut model with mucin environment. *FEMS Microbiol. Ecol.* 79, 685–696. doi:10.1111/j.1574-6941.2011.01252.x

- Vieira, F.C.S., Nahas, E., 2005. Comparison of microbial numbers in soils by using various culture media and temperatures. *Microbiol. Res.* 160, 197–202. doi:10.1016/j.micres.2005.01.004
- Vimr, E.R., Kalivoda, K.A., Deszo, E.L., Steenbergen, S.M., 2004. Diversity of Microbial Sialic Acid Metabolism. *Microbiol. Mol. Biol. Rev.* 68, 132–153. doi:10.1128/MMBR.68.1.132-153.2004
- Vital, M., Howe, A.C., Tiedje, J.M., 2014. Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data. *mBio* 5, e00889-14. doi:10.1128/mBio.00889-14
- Wachsman, J.T., Barker, H.A., 1954. Characterization of an Orotic Acid Fermenting Bacterium, *Zymobacterium Oroticum*, *Nov. Gen., Nov. Spec. J. Bacteriol.* 68, 400–404.
- Walther, R., Hippe, H., Gottschalk, G., 1977. Citrate, a specific substrate for the isolation of *Clostridium sphenoides*. *Appl. Environ. Microbiol.* 33, 955–962.
- Wang, J.-T., Zheng, Y.-M., Hu, H.-W., Li, J., Zhang, L.-M., Chen, B.-D., Chen, W.-P., He, J.-Z., 2016. Coupling of soil prokaryotic diversity and plant diversity across latitudinal forest ecosystems. *Sci. Rep.* 6, 19561. doi:10.1038/srep19561
- Wang, L.-Q., 2002. Mammalian phytoestrogens: enterodiol and enterolactone. *J. Chromatogr. B, Analytical and Biomedical Aspects of Natural Compounds with Estrogenic Activity* 777, 289–309. doi:10.1016/S1570-0232(02)00281-7
- Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi:10.1128/AEM.00062-07
- Warnick, T.A., Methé, B.A., Leschine, S.B., 2002. *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* 52, 1155–1160. doi:10.1099/ijs.0.02125-0
- Warren, Y.A., Tyrrell, K.L., Citron, D.M., Goldstein, E.J.C., 2006. *Clostridium aldenense* sp. nov. and *Clostridium citroniae* sp. nov. Isolated from Human Clinical Infections. *J. Clin. Microbiol.* 44, 2416–2422. doi:10.1128/JCM.00116-06
- Whitehead, T.R., Cotta, M.A., Collins, M.D., Lawson, P.A., 2004. *Hespellia stercorisuis* gen. nov., sp. nov. and *Hespellia porcina* sp. nov., isolated from swine manure storage pits. *Int. J. Syst. Evol. Microbiol.* 54, 241–245. doi:10.1099/ijs.0.02719-0

- Whitford, M.F., Yanke, L.J., Forster, R.J., Teather, R.M., 2001. *Lachnobacterium bovis* gen. nov., sp. nov., a novel bacterium isolated from the rumen and faeces of cattle. *Int. J. Syst. Evol. Microbiol.* 51, 1977–1981. doi:10.1099/00207713-51-6-1977
- Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci.* 95, 6578–6583.
- Wielen, P.W.J.J. van der, Rovers, G.M.L.L., Scheepens, J.M.A., Biesterveld, S., 2002. *Clostridium lactatifermentans* sp. nov., a lactate-fermenting anaerobe isolated from the caeca of a chicken. *Int. J. Syst. Evol. Microbiol.* 52, 921–925. doi:10.1099/ijs.0.02048-0
- Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* 74, 5088–5090. doi:10.1073/pnas.74.11.5088
- Wolin, M.J., Miller, T.L., Collins, M.D., Lawson, P.A., 2003. Formate-Dependent Growth and Homoacetogenic Fermentation by a Bacterium from Human Feces: Description of *Bryantella formatexigens* gen. nov., sp. nov. *Appl. Environ. Microbiol.* 69, 6321–6326. doi:10.1128/AEM.69.10.6321-6326.2003
- Yuli, S., Chengxu, L., Sydney, F., 2005. Multiplex PCR for rapid differentiation of three species in the “*Clostridium clostridioforme* group.” *FEMS Microbiol. Lett.* 244, 391–395. doi:10.1016/j.femsle.2005.02.017
- Yutin, N., Galperin, M.Y., 2013a. A genomic update on clostridial phylogeny: Gram-negative spore-formers and other misplaced clostridia. *Environ. Microbiol.* 15, 2631–2641. doi:10.1111/1462-2920.12173
- Yutin, N., Galperin, M.Y., 2013b. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* 15, 2631–2641. doi:10.1111/1462-2920.12173
- Zajec, N., Stres, B., Avguštin, G., 2012. Distinct approaches for the detection and removal of chimeric 16S rRNA sequences can significantly affect the outcome of between-site comparisons.
- Ze, X., Duncan, S.H., Louis, P., Flint, H.J., 2012. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* 6, 1535–1543. doi:10.1038/ismej.2012.4
- Zhang, Y., Zhou, L., Bao, Y.L., Wu, Y., Yu, C.L., Huang, Y.X., Sun, Y., Zheng, L.H., Li, Y.X., 2010. Butyrate induces cell apoptosis through activation of JNK MAP kinase pathway in human colon cancer RKO cells. *Chem. Biol. Interact.* 185, 174–181. doi:10.1016/j.cbi.2010.03.035

- Zhao, J., Binns, A.N., 2011. Characterization of the mmsAB-araD1 (gguABC) Genes of *Agrobacterium tumefaciens*. *J. Bacteriol.* 193, 6586–6596. doi:10.1128/JB.05790-11
- Zhulin, I.B., 2001. The superfamily of chemotaxis transducers: From physiology to genomics and back, in: *Physiology, B.-A. in M. (Ed.)*, . Academic Press, pp. 157–198.