

**Characterization of uncultured, human oral microbiota using a
targeted, single-cell genomics approach**

**A Dissertation Presented for the
Doctor of Philosophy
Degree**

The University of Tennessee, Knoxville

Alisha Gail Campbell

May 2013

**Copyright © 2013 by Alisha Gail Campbell
All rights reserved.**

DEDICATION

I dedicate this dissertation to my husband Jim for his patience and unwavering support during my graduate career. In addition, I dedicate this work to my parents, Scott and Linda Jo and Jeff and Marie, who have encouraged me throughout the pursuit of my education.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Mircea Podar for giving me the opportunity to work on such an exciting project. I would also like to acknowledge my committee members: Drs. Alison Buchan, Robert Hettich, Chris Schadt and John Biggerstaff for the time, comments and suggestions they provided during each of our meetings.

I would like to thank several past and present members of the Biosciences Division at ORNL for their friendship, advice and assistance throughout my dissertation work including Migun Shakya, Scott Hamilton-Brehm, Jennifer Mosher, Donna Kridelbaugh, Sarah Kauffman, Zhiwu Wang, Taniya Roy Chowdhury, Steve Allman, Marilyn Kerley, Dawn Klingeman, Sue Carroll and Kim Drew.

In addition, I would like to thank the Genome Science and Technology program and its staff including Kay Gardner, Terrie Yeatts and Dr. Albrecht Von Arnim for all of their support during my graduate career. In addition, I would like to thank Dr. Cynthia Peterson for her guidance as I began my graduate experience. I would like to acknowledge Joe Mays at the UT sequencing facility for his great work and helpful discussions. I would also like to thank all of my friends within the GST program that commiserated with me about all the shared experiences of graduate school.

I would like to thank my many collaborators including Drs. Jeff Becker and Melinda Hauser at UT, as well as the Tanja Woyke lab at the Joint Genome Institute and collaborators at Ohio State University for the insight and expertise they were able to provide for many aspects of my project.

The research presented within this dissertation was supported by grant R01 HG004857 from the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) and by grant 1R56DE021567 from the National Institute for Dental and Cranial Research (NIDCR) of the NIH to Mircea Podar.

Lastly, I would like to thank my entire family for all of their love and support throughout my education, even when they thought I might continue to go to school forever. A special thank you goes to my husband Jim, who believed in me and continued to encourage me despite my occasional anxiety and bad temper.

ABSTRACT

Microbial communities associated with the human oral cavity are complex, and many oral microbes have yet to be cultured. These uncultured community members are of interest ecologically and phylogenetically, and a number of uncultured species have been positively correlated with oral diseases such as periodontitis. Thus, an approach was adapted to selectively separate single cells from mixed populations of oral bacteria and obtain genomic information for uncultured community members. A combination of fluorescent labeling, cell sorting with flow cytometry and multiple displacement amplification was used to obtain sufficient genomic material for whole-genome pyrosequencing.

The first targets were from uncultured oral lineages within *Deltaproteobacteria* groups *Desulfobulbus* and *Desulfovibrio*, and cells were selectively isolated with group-specific, fluorescent *in situ* hybridization probes. *Deltaproteobacteria* were targeted to better understand their increased abundance in periodontitis patients. The resulting oral *Deltaproteobacteria* genomes encoded several unique proteins that appear to function in survival in a host environment, including proteins similar to virulence factors of known human pathogens. Thus, it is likely that the oral *Deltaproteobacteria* are not just opportunistic community members of periodontitis sites but also have the potential to play a role in disease etiology.

A second target was a low abundance community member belonging to the class *Anaerolineae*, within *Chloroflexi* subphylum I. This diverse group of organisms has only a few cultured isolates, and limited genomic information is

available, making it difficult to discern their ecological significance. Genomic data suggest that oral *Anaerolineae* are anaerobic heterotrophs with an abundance of carbohydrate transport and metabolism genes. The presence of a unique phosphotransferase system and other genes necessary for *N*-acetylglucosamine metabolism suggests this organism grows by scavenging material from surrounding lysed bacterial cells.

These single-cell genomic studies have provided insight into the ecological niches of two uncultured groups within the oral environment, revealed differences between environmental and host-associated organisms within this groups and provided possible gene targets for future studies. Ultimately, methods employed in these studies could be readily applied to gain information on other uncultured microbes from any environment.

TABLE OF CONTENTS

Chapter 1. The importance of uncultured members of human-associated, oral microbial communities.....	1
Chapter 2. Methods for isolation and analysis of single amplified genomes of oral microbiota.....	15
Chapter 3. Multiple single-cell genomes provide insight into functions of uncultured <i>Deltaproteobacteria</i> in the human oral cavity.....	49
Chapter 4. Single-cell genomic analyses of an uncultured member of phylum <i>Chloroflexi</i> found in the human oral cavity.....	94
Chapter 5. Conclusions and future directions of the studies of uncultured microbiota.....	118
References.....	123
Vita.....	136

LIST OF TABLES

Table 2.1. Composition of lysis, neutralization and amplification buffers used for single-cell lysis and amplification procedures.....	37
Table 2.2. Reagents for the setup of one polymerase chain reaction (50 µL)....	39
Table 2.3. Commands used for genome normalization, assembly and gene prediction.....	43
Table 2.4. Optimization of resulting Velvet assemblies over a range of hash lengths for SAG Dsb3 discussed in Chapter 3.....	44
Table 3.1. Metadata and genome statistics for single and multi-cell oral amplicons.....	59
Table 3.2. Similarity matrix of <i>Desulfobulbus</i> genomes based upon average nucleotide identities.....	63
Table 3.3. Similarity matrix of <i>Desulfobulbus</i> genomes based upon tetranucleotide frequencies.....	64
Table 3.4. Categories of putative virulence factors found in Dsb1-5 and Dsv1.....	74-75
Table 3.5. Dsb1-5 contigs that contain large numbers of horizontally acquired polysaccharide metabolism genes.....	78-79
Table 3.6. COGs unique to either human-associated or environmental <i>Desulfovibrio</i> genomes.....	84-85
Table 4.1. Assembly statistics and metadata for combined Chl1-2 genome....	103

Table 4.2. COGs present in Chl1-2 and absent in other available <i>Chloroflexi</i> genomes.....	108-109
--	---------

LIST OF FIGURES

Figure 1.1. A comparison of a healthy tooth (left) versus the effects of periodontitis (right).....	8
Figure 2.1. An overview of single-cell fluorescent labeling and isolation and subsequent genome amplification, sequencing and data processing.....	16
Figure 2.2. Output from the flow cytometer showing the selection of a specific group of cells based upon forward scatter (FSC) and fluorescent emission (528-538).....	24
Figure 2.3. Overlapping scatter plots for four known cultures.....	26
Figure 2.4. Scatter plots for oral samples from (A) a healthy individual and (B) an individual with periodontitis.....	28
Figure 2.5. Organisms captured with each sorting gate in a healthy oral sample.....	31
Figure 2.6. Organisms captured with each sorting gate for a sample from an individual with periodontitis.....	32
Figure 2.7. Neighbor-joining phylogenetic tree of a subset of Synergistes based on SSU rRNA genes.....	33
Figure 2.8. Image depicting the process of multiple displacement amplification using ϕ 29 DNA polymerase and protected random primers.....	35
Figure 2.9. Illustration of MDA product PCR amplification and subsequent quality check using resulting chromatograms.....	39

Figure 2.10. Data reduction during the normalization process used for SAGs.....	43
Figure 2.11. The range of GC content seen in scaffolds originally assembled for SAG Dsb3 (Chapter 3).....	45
Figure 2.12. A tetranucleotide plot generated to detect possible contamination in a SAG Chl1 (Chapter 4).....	46
Figure 3.1. Optimization of sample hybridization times.....	58
Figure 3.2. Maximum likelihood phylogenetic tree based on the SSU rRNA gene of select <i>Deltaproteobacteria</i> sequences.....	60
Figure 3.3. Number of Dsv1 genes with homologs in gut species <i>D. sp.</i> 3_1_syn3, <i>D. piger</i> or environmental species <i>D. vulgaris</i> Hildenborough.....	61
Figure 3.4. Map of enzymes used in propionate metabolism and the methylmalonyl pathway.....	66
Figure 3.5. Comparison of COG categories (presence/absence) in (A) Dsb1-5 and <i>D. propionicus</i> and in (B) human-associated <i>Desulfovibrio</i> Dsv1, <i>D. sp.</i> 3_1_syn3 and <i>D. piger</i>	83
Figure 3.6. Maximum likelihood tree of putative <i>srfB</i> genes.....	87
Figure 3.7. Maximum likelihood tree of putative <i>srfC</i> genes.....	88
Figure 3.8. Gene neighborhood that includes two genes (<i>srfB</i> and <i>srfC</i>) unique to host-associated deltaproteobacterial genomes and with low homology to <i>Deltaproteobacteria</i> from other environments.....	89

Figure 3.9. Brief overview of proposed metabolic and transport capabilities of Dsb1-5.....	92
Figure 3.10. Brief overview of proposed metabolic and transport capabilities of Dsv1.....	93
Figure 4.1. Neighbor-joining phylogenetic tree based on <i>Chloroflexi</i> with the SSU rRNA Greengenes database.....	100
Figure 4.2. Maximum likelihood phylogenetic tree of human-associated, oral <i>Chloroflexi</i> based on partial SSU rRNA genes.....	101-102
Figure 4.3. Brief overview of proposed metabolic and transport capabilities of Chl1-2.....	106
Figure 4.4. Diagram of <i>N</i> -acetylglucosamine (GlcNAc) and sialic acid <i>N</i> -acetylneuraminate (Neu5Ac) metabolism.....	112

Chapter 1. The importance of uncultured members of human-associated, oral microbial communities

1.1. Uncultured microbes

Studies of microbes have been ongoing for centuries. A great advance in the field came in the 19th century when bacteria were first isolated and cultured, allowing for characterization of nutrient dependencies and, later, the use of genetic techniques for the characterization of specific gene content. Despite the advances made with culturable bacteria, there were also inaccessible, uncultured microbes. These uncultured microbes were first noticed microscopically. The number of microbes that could be observed under the microscope was always greater than the number of bacterial colonies that could be grown from serial dilutions in the laboratory, a phenomenon that came to be known as the “the great plate count anomaly” [1].

These uncultivated microorganisms were largely ignored until Woese first looked at bacterial taxonomy using ribosomal RNA (rRNA) genes to form a phylogenetic tree of life [2-4]. These studies revealed how little diversity was captured by studies of higher eukaryotes and the large amount of diversity found within the bacterial and archaeal domains [2, 5]. Conserved regions of the small subunit (SSU) rRNA gene were soon targeted with “universal” PCR primers, and the variable regions found within the SSU rRNA gene could be used for molecular identification of previously unknown microbes directly from environmental DNA samples [5-7]. This allowed for the study of microbial diversity without the need for cultivation. Ultimately, the use of SSU rRNA gene-

based phylogenetic tools led to surveys of numerous environments [5], and the power of these types of studies was amplified further with the use of shotgun metagenomics [8] and pyrosequencing [9, 10].

The number and diversity of microbes is enormous and has been estimated between 10^{30} - 10^{31} genomes [11]. Further, it has been estimated that >99% of the bacterial cells identified in most environments have yet to be cultured in a laboratory setting [12]. Many of these microbes are closely related to cultured organisms, but there are also many groups of organisms with no closely cultured relatives. These represent previously unknown groups of bacteria, even at the phylum level, such as TM7 [13]. Molecular studies have revealed the ubiquitous distributions of some of these uncultured bacteria, which could suggest important functional roles within their respective ecosystems [14, 15]. For example, uncultured organisms have been associated with nitrogen cycling and wastewater treatment processes [16], sulfur cycling [17] and health and disease states such as oral periodontitis [18]. Thus, it is of utmost importance that technology be developed and used to understand more about the ecological roles of these uncultured microbes.

1.2. The human microbiota and its importance

Uncultured organisms are of importance in all environments, including host-associated microbial communities. After the human genome was successfully sequenced, many researchers advocated for the sequencing of human-associated microbial communities [19]. The microbes present both in and on the

body are an important component of the human ecosystem and have been found to outnumber the human cells of the body by an order of magnitude [20]. These communities are known as the human microbiota, and the collective genomes of these organisms are known as the human microbiome [21, 22]. Early, culture-dependent studies of human-associated microbes often focused only on pathogens and revealed only a small subset of bacterial diversity within the human-associated microbial communities. Use of SSU rRNA gene studies and rapidly advancing sequencing platforms have allowed for the microbial ecology of the human body to be re-examined [22].

1.2.1. Human microbiome initiatives

To this end, several research institutes worldwide began large initiatives within the last ten years to describe and catalog members of the human microbiome to better understand their functional roles as well as their role in the health and disease states of humans. The ultimate goal of these studies is to use this knowledge in treatments for human disease. In the US, the National Institute of Health (NIH) funded the Human Microbiome Project (HMP) [19, 22], and, in Europe, the Metagenomics of the Human Intestinal Tract (MetaHIT) project [23] was initiated to better understand the microbes that reside in the human gut. These large-scale studies and many smaller projects provided a great deal of information about the human microbiome, and sequencing data from these efforts continues to grow exponentially [24].

One of the major successes of the HMP was an extensive SSU rRNA gene and whole-genome shotgun (WGS) metagenomic sequencing of 242 adults from 15-18 body sites [25]. These body sites included the gut, nasal passages and several distinct sites of the skin, vagina and oral cavity. In addition, hundreds of strains were sequenced to provide genomic reference data, and a number of tools and protocols were developed to deal with the data produced [25]. Although smaller than the HMP dataset, MetaHIT provided a large WGS metagenomic dataset for the gut microbes of healthy individuals, as well as overweight and obese individuals and subjects with inflammatory bowel disease [23].

1.2.2. Current knowledge of the human microbiome

One of the initial goals of the HMP was to determine if there is a “core” microbiome that is common to all humans. Finding a common core community shared among all individuals would make downstream studies of health and disease much easier, but the majority of body sites sampled have shown a large degree of individual variability [25]. At the species level, all humans appear to harbor a distinct set of microbes [25, 26]. Comparison of higher-order taxa, such as phyla, revealed larger overlap among individuals. However, even at this level, the abundance of phyla differed greatly among individuals [26-29]. Despite the distinct microbes found in individuals, it seems that there is a shared core of functional genes between microbial communities [26, 29, 30].

The gut microbiota is the largest and most intensely studied human-associated community. Thus, some of the greatest advances have come from

these studies. Research has addressed development [31] and evolution [32] of microbiota, functional contributions of microbiome constituents [26, 33] and contributions of host genetics [26] and diet [34] in shaping a microbiome. There have been links of certain bacterial community members to diseases ranging from obesity [35] and diabetes [36] to Crohn's disease [37]. Large numbers of reference genomes have been collected through cultivation and single-cell genomics [25, 38]. In addition, the large amount of data available from the gut microbiome allowed for the identification and confirmation of several novel bacterial taxa, including a number of sequences that are most closely related to the genus *Barnesiella* [39]. This enormous amount of data was retrieved with a broad spectrum of methods including SSU rRNA gene and WGS metagenomics, metaproteomics, metabolomics, genomics and the use of animal models including gnotobiotic mice.

Human-associated microbiota from other body sites have also been characterized, but a large number of these studies have used only SSU rRNA gene and small-scale WGS metagenomics. Although SSU rRNA gene pyrosequencing is important to initially describe the members of a microbial community, it is important to further characterize the functionality of these communities and address different states of health and disease.

1.3. The oral microbiota

Studies of microbes in the human oral cavity have been ongoing since the discovery of “animalcules” by Antony van Leeuwenhoek in 1676. Despite this

long history, it is estimated that approximately 30 - 60% of oral microbes remain uncultured [18, 40-42]. Over 500 species had been cultivated from the human oral cavity [41], and the most recent molecular studies predict a diversity of approximately 1000 species [18, 43]. Although a larger proportion of the organisms associated with the oral cavity have been cultured relative to other environments, information for the remaining, uncultured organisms is no less important. In fact, many of the uncultured organisms have been linked to oral or systemic diseases [18, 44].

Initial studies focused on only a few culturable microbial species that were found in association with caries, gingivitis or periodontitis, but more recent studies have begun to look at the overall ecology of the oral microbial community. Many of the global microbial studies of the oral cavity used clone libraries to characterize communities, [28, 40-42] resulting in a large repository of full-length, SSU rRNA gene sequences associated with the oral cavity [43]. Unfortunately, the number of sequences captured in most clone libraries does not reveal the full diversity associated with the oral microbial communities. However, this resource can be used as an important tool for classification of smaller amplicons used for pyrosequencing.

Such large-scale pyrosequencing efforts have only just begun for the oral microbiome, but have provided a number of interesting findings. These studies are providing insight into a number of previously unknown species in the oral environment [45] and have addressed inter and intra-individual variation [46]. Although all human-associated microbiomes sampled have shown unexpected

variation between individuals, the oral microbiome appears to be less variable than other body sites and may have a small “core” assemblage [46-48]. Several studies have also shown that different sites within the mouth harbor distinct microbiomes [42, 47, 49]. In addition to baseline studies of the oral microbiome of healthy individuals, several advances have been made in the study of microbes associated with oral diseases such as periodontitis.

1.3.1. Oral microbes associated with periodontitis

Periodontitis is an inflammatory gum disease that has been studied since the early 19th century (**Figure 1.1**). It is one of the most common diseases in humans and is the leading cause of tooth loss worldwide. Approximately 35% of adults in the US are estimated to have periodontitis [42]. Therefore, periodontitis is pervasive and results in significant treatment costs. Even more concerning, occurrence of periodontitis has been linked to a number of systemic diseases, including diabetes, cardiovascular disease, osteoporosis and preterm low birth weight [50-55]. Consequently, understanding roles of periodontitis-associated microbiome members is of utmost importance. Although it is known that bacteria are involved in the disease, the exact role of microbial communities associated with the disease remains unclear [56].

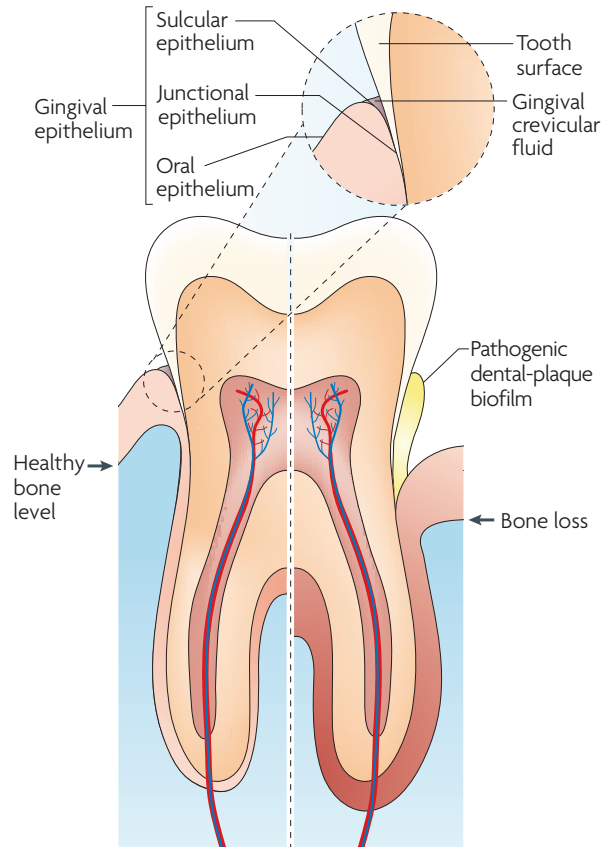


Figure 1.1. Comparison of a healthy tooth (left) versus the effects of periodontitis (right). Image reprinted from Darveau 2010 [56]. In periodontitis, inflammation occurs and plaque biofilm accumulates in a deepening subgingival crevice. The disease results in the loss of connective epithelium tissue and bone and can lead to tooth loss.

Periodontitis was originally thought to be caused by bacteria, but initial studies stalled when the disease could not be linked to a single, pathogenic species. Subsequent studies revealed that periodontitis is a polymicrobial disease in which the microbial community shifts from the community present in healthy individuals [56]. This shift was originally detected as a decrease in Gram-positive organisms and an increase in, anaerobic Gram-negative organisms during periodontitis [57]. DNA probe and DNA-DNA hybridization studies found that the presence of the “red complex”, *Porphyromonas gingivalis*, *Treponema*

denticola and *Tannerella forsythia*, was closely associated with several disease parameters [58].

More recent pyrosequencing efforts of periodontal samples have agreed with the basic tenets of older studies, although it was found that the microbiome associated with periodontitis is much more complex than originally thought [18]. Species of the “red complex” are present in both healthy individuals and those with periodontitis. *P. gingivalis*, *T. denticola* and *T. forsythia* were found to be more abundant in disease [18], but the presence and abundance of a large number of other, unrelated species were also associated with periodontitis. In one study, *Fusobacterium* and *Porphyromonas* were not found to be more abundant in periodontitis, although this was thought to be due to small sample size [59]. Further, there has been some support of a Gram-negative dominated community in the disease state versus a Gram-positive dominated healthy state [18, 59], but it was suggested that this view of the disease was too simplistic. In one case, the skew toward Gram-negative organisms in the disease state was mostly due to the decrease of two health-associated species of *Streptococcus* [18]. In addition, *Filifactor alocis*, a Gram-positive organism, is one of most abundant organisms correlated with periodontitis [18].

It has been shown that there is more bacterial diversity associated with periodontitis as opposed to the lower diversity associated with dental caries and the lowered pH [18, 43, 59]. It was also observed that members of the bacterial community associated with periodontitis are most often present in healthy individuals, but at low levels [59]. The use of metagenomics was able to suggest

enhanced functional traits of the periodontitis microbiome including fatty acid metabolism, ferredoxin oxidation and acetyl-CoA degradation. However, large amounts of human DNA contamination and a small sample size limited data from this study [59].

1.3.2. Future studies of the oral microbiota

Although information about the oral microbiome has increased a great deal over the past few years, there is still much information needed to explain disease etiology and suggest preventative measures. Large-scale metagenomic information is currently available for only healthy individuals [25]. The broad range of tools that has been used to expand knowledge of the gut microbiome could easily be applied to sites within the oral cavity. This includes more extensive metagenomic sequencing and addition of metaproteomic, metatranscriptomic and metabolomic information. Additional studies of different sites within the oral cavity and expanded studies of disease states would help explain initial differences seen with SSU rRNA gene pyrosequencing. Further, whole genome information would provide a way to assess functional potential, as well as serve as a reference for other fragmented sequencing efforts.

Unfortunately, a limited number of oral bacterial genomes are available, and information for periodontitis-associated bacteria is limited to a few, cultured bacteria [60-62]. One method used to obtain genomic information for uncultured organisms involves amplification of genomic material from a single, isolated cell. This technique was originally used in the oral environment to obtain single-cell-genomic information for the uncultured, candidate phylum TM7 [63]. Although the

small portion of the TM7 genome that was captured limited functional conclusions, the method of single-cell genomics has rapidly developed, making this tool an attractive option for obtaining genomic information and adding to the knowledge of uncultured organisms of the oral cavity.

1.4. Use of single-cell genomics

Successfully amplification of whole genome information from a single cell was first accomplished in 1992 [64], but the use of a more efficient enzyme [65, 66] with random primers protected from exonuclease activity was able to significantly increase quality and quantity of DNA obtained from the reaction [67]. Initial validation of multiple displacement amplifications (MDA) for amplification of single cells was first shown with *Escherichia coli* separated by flow cytometry [68] and has rapidly progressed to a number of environmental studies [38, 69].

Initial studies used single-cell genomics to obtain functional information for the uncultivated phylum TM7 [63, 70], and this method is still used for the study of uncultivated phyla [71]. Other early single-cell studies allowed for the functional characterization of uncultured organisms that were found to be abundant community members. Flow cytometry and MDA was used to retrieve genomic information for several uncultured marine bacteria, and revealed a proteorhodopsin gene in several flavobacterium cells, suggesting a role for this group in photometabolism [72].

As the process was refined and methods were developed to limit contamination [73, 74], the quality of MDA data increased. In 2009, two marine

flavobacterium single amplified genomes (SAGs) were estimated to be 78-91% complete and were compared to available metagenomic data to further describe the ecological niche of these organisms [75]. The first complete genome obtained using MDA was published in 2010. The polyploidy sharpshooter symbiont, *Candidatus* Sulcia muelleri DMIN, was sequenced from a single cell, and the genome was closed using Sanger sequencing. Metagenomics were used for comparison, allowing for the verification of assembly measures and comparisons of single nucleotide polymorphisms [76].

More recent studies have provided an abundance of functional information, including the discovery of chemolithoautotrophic pathways in uncultured marine proteobacteria [77] and insight into the role of ammonia-oxidizing archaea [78] and the probable role of Verrucomicrobia in polysaccharide degradation [79]. Single-cell genomics has also been used to resolve the phylogeny of the controversial *Leptothrix ochracea* [80]. Many of these studies have shown the benefits of combining information from single-cell genomes with other types of data, such as metagenomics, to gain a more complete understanding of the organism and environment of interest. This was illustrated in a recent study of oil-degrading bacteria associated with the *Deepwater Horizon* oil spill, in which a combination of single-cell genomics, metagenomics and metaproteomics were applied [81]. Combined use of metagenomics and single-cell genomics can provide a wealth of data, but specific benefits of single-cell genomics include targeted access of community

members and the ability to capture specific associations with symbionts, viruses and plasmids [38, 69].

1.5. Scope of the dissertation

This dissertation will discuss the approaches of flow cytometry and single-cell genomics for the use of obtaining uncultured oral microbes of interest. Chapter 2 details the streamlined methodology developed for the isolation of single cells of interest from complex community samples and downstream genomic sequencing and analyses of these cells. These methods were adapted to ensure proper sample processing and storage, limit contamination and provide accurate and complete genomic information for interpretation of genomic data. An additional method more thoroughly discussed in Chapter 2 was developed and used to rapidly access a larger number of uncultured organisms in human-associated samples. The use of defined sectors in flow cytometry, as opposed to random sorting, increases the likelihood of isolating cells of less abundant, uncultured targets and was successfully used to amplify genomes of more than 100 bacteria of interest.

Chapter 3 describes the use of single-cell genomic techniques in association with group-specific probes to selectively isolate oral *Deltaproteobacteria*. Both *Desulfobulbus* and *Desulfovibrio* species from the oral environment have been associated with oral and systemic disease, and the large amount of genomic data gathered from sequencing multiple single cells revealed

a suite of genes that potentially allow these bacteria to be successful in the oral environment and suggest a possible role in pathogenesis.

In Chapter 4, isolation and characterization of a rare, oral member of the *Chloroflexi* phylum is discussed. Comparative analyses revealed little overlap with genes found in photosynthetic members of the same phylum. The resulting genomic data revealed the presence of genes for *N*-acetylglucosamine metabolism and could indicate that this organism grows cryptically by scavenging material from lysed bacterial cells in the surrounding oral environment.

Data acquired throughout these studies has provided genomic evidence that host-associated *Deltaproteobacteria* have genes similar to other host-associated microbes (Chapter 3), provide insight into the ecological niche of a rare and phylogenetically interesting member of the oral community (Chapter 4) and will continue to provide a wealth of additional genomic data with the sequencing of numerous reference genomes (Chapter 2). The final chapter (Chapter 5) provides an outlook for the future of single-cell genomics and study of the human microbiome, particularly the study of oral-associated microbes.

Chapter 2. Methods for isolation and analysis of single amplified genomes of oral microbiota

This chapter details the methods used for all the studies performed throughout this dissertation. All methods performed by collaborators have been clearly noted. Unless stated, Alisha G. Campbell performed all methods discussed. Methods that were developed or amended during the studies in this dissertation are discussed in greater detail than previously published methods.

2.1. Introduction

Single-cell genomics is a relatively new combination of technologies [68] that allows scientists to access genomic material from a single, retrieved cell of an organism that cannot yet be cultivated. This technique has allowed for the sequencing of partial to essentially complete genomes for several bacteria [63, 70, 72, 75, 76, 79], archaea [78] and protists [82, 83].

Only one dataset generated using single-cell genomics from a human microbiome sample has been published [63], and this study used micromanipulation for cell retrieval. Therefore, established techniques for sample handling and processing, other methods of cell retrieval from human samples and efficient data analysis are limited. Also, despite the obvious benefits of single-cell genomics, several aspects of the technique are problematic. Aspects that have been addressed but continue to provide challenges to the field include low efficiency lysis procedures [72], contamination of samples and reagents [73, 74], formation of chimeric amplicons [84] and biased genomic information [75, 85]. Thus, it was important to establish an efficient and reliable workflow for retrieving selected oral bacteria and effectively analyzing the resulting genomic data. To this end, methods were optimized and developed for processing and

analyzing oral samples, and the overview of this protocol is shown in **Figure 2.1**.

This process included retrieval of labeled cells using a flow cytometer and subsequent amplification, sequencing and analysis of genomic data. Each step will be discussed in detail throughout this chapter. Steps that were developed during the course of this dissertation or amended from original protocols will be noted in the text.

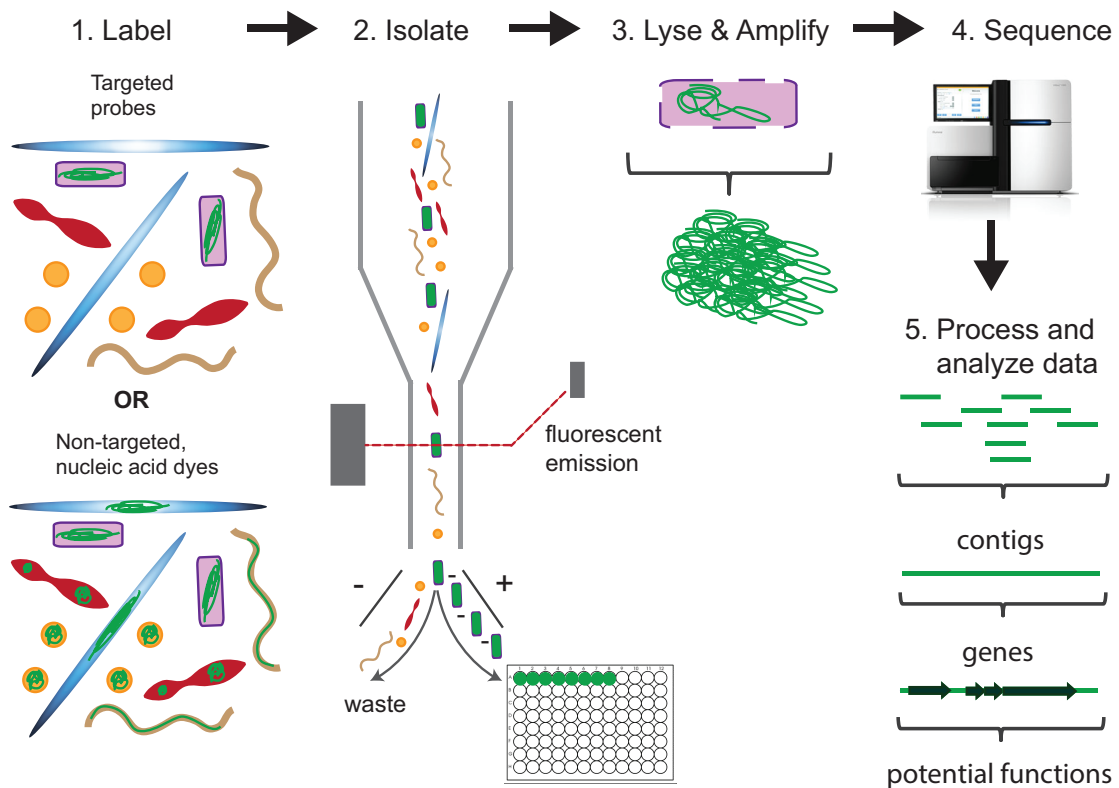


Figure 2.1. An overview of single-cell fluorescent labeling and isolation and subsequent genome amplification, sequencing and data processing. Cells were labeled with either a group-specific fluorescent *in situ* hybridization probe for targeted isolation or with a non-specific, nucleic acid dye for a more high-throughput method discussed in section 2.4.3. All cells were isolated based on fluorescent emission, size and shape using a flow cytometer. Amplification was accomplished using a multiple displacement amplification and sequencing was performed using 100 bp paired-end libraries on an Illumina Hi Seq.

2.2. Sample collection and preparation

To prepare oral samples for sorting in a flow cytometer and subsequent amplification of resulting single cells, it was essential to properly collect, fix and store samples to ensure that microbial cells remain intact. A comprehensive workflow was developed and followed at the time of sample collection, during transport and once the samples arrived at the laboratory. These sampling procedures were used for all samples discussed in the following chapters. Prior to all sampling, human subject enrollment and sample collection protocols were approved by the Ohio State University Institutional Review Board and by the Oak Ridge Site-Wide Institutional Review Board. Signed, informed written consent was obtained from all human subjects that provided samples for this study. Approved research safety summaries were also established at Oak Ridge National Laboratory (ORNL) to ensure proper handling and disposal of all human-related, biohazard samples.

2.2.1. Sample classification

Samples were collected from both periodontally healthy subjects and subjects with periodontitis. Subjects for this analysis were sampled locally (Knoxville, TN, USA) or by collaborators at The Ohio State University (Columbus, OH, USA). All samples analyzed in this study were taken from the subgingival crevice. The majority of samples were obtained using a paper point, and one sample was obtained using a curette instrument to retrieve a scraping of subgingival plaque. Patients were classified as healthy or diseased based upon previously published criteria [18]. Briefly, subjects were considered to have periodontitis if three non-

adjacent sites in at least two quadrants showed ≥ 4 mm attachment loss and 5 mm probing depth. Subjects were considered healthy if no subgingival pockets were identified with a probing depth > 4 mm.

2.2.2. Sample preparation and fixation

In all cases, samples were taken and immediately stored in 1× PBS at 4 °C and transported on ice to the lab within 24 h. When samples arrived at the lab, each was vortexed vigorously for one minute, and PBS from multiple samples was combined. Samples from multiple subjects were combined in order to increase the probability of obtaining our low-abundance targets that are not present in all subjects. Final samples came from a mix of either healthy subjects or subjects with periodontitis.

The volume of combined samples was reduced by centrifugation at $3000 \times g$ for 5 min. The resulting pellets were resuspended in 500 μ L of 1× PBS, and an equal volume of ethanol was added to each sample. Properly fixed bacterial cells are left permeable, yet intact. Although paraformaldehyde (PFA) is often used for efficient fixation of bacterial cells [86], this protocol was not appropriate for our analyses. PFA crosslinks DNA and could cause problems for the downstream amplification and sequencing process. Thus, cells were fixed in ethanol, a process initially used for gram positive organisms [87] and that has been used in at least one other single-cell study [70]. Optimization within our lab revealed a minimum of 24 h was needed for efficient permeabilization of oral samples.

Proper storage of human oral samples for subsequent use in single-cell analyses had not been previously addressed in the literature and was optimized

during the course of this work. Although a number of storage conditions were shown to be satisfactory for cell preservation when tested with bacterial cultures, effective storage of oral samples proved more difficult. When examined microscopically, fixed oral samples showed signs of aggregation and cell lysis after approximately one week at -20 C. Therefore, to ensure samples were useful for downstream cell sorting, samples used for these studies were stored for a maximum of 72 h at -20 C before cells were labeled and isolated.

2.3. Cell labeling

An important part of the cell retrieval process is labeling of cells. A fluorescent label allows cells to be distinguished from background debris and can also be used to distinguish a subset of bacteria that are of interest (as seen in Chapter 3). To selectively target one family of bacteria, FISH probes were used. These probes are small, fluorescently labeled oligonucleotides used to target specific regions of SSU ribosomal RNA that can effectively differentiate between the cells of interest and other bacteria present in the sample. Although original protocols were for visualization of environmental bacterial groups of interest on microscope slides, FISH has been adapted to work in-solution to label cells in a way that can be used for further analysis on a flow cytometer [86, 88]. Efficient FISH depends not only on an effective probe design but also on appropriate hybridization conditions. Therefore, optimization of hybridization times was necessary in these studies despite the use of previously published probes. To capture a larger, diverse set of uncultured organisms from the oral environment, fluorescently labeled nucleic acid dyes were used.

2.3.1. In-solution, fluorescent in situ hybridization

The following procedure was used for the targeted sorting discussed in 2.4.2 and successfully captured the cells discussed in Chapter 3. The protocol was a combination and adaptation of protocols provided on <http://www.arb-silva.de/fish-probes/> and in published literature [86, 88-90]. First, hybridization and wash buffers were made as described [90]. Both formamide concentration and hybridization times were optimized during these studies using a cultured isolate that was related to the uncultured target.

Ethanol-fixed samples were first removed from -20 C, and 500 μ L was centrifuged at 3000 $\times$ g for 5 min. Resulting pellets were resuspended in 50 μ L of a probe/hybridization buffer mixture. The probe/hybridization buffer mixture contained one volume of 50 ng/ μ L probe solution to 9 volumes of hybridization buffer. Samples were incubated at 46 C for 3-20 h, depending upon optimized hybridization times. Hybridization was performed in a light-protected, rotating incubator on a low speed setting. After hybridization is complete, 1 mL of pre-warmed (48 C) wash buffer was added to each sample, and samples were incubated at 46 C for an additional 10 min. Samples were then centrifuged (3000 $\times$ g, 5 min), and the pellet was resuspended in 1 mL of 1 $\times$ PBS. Prior to sorting, the entire sample was passed through a CellTrics 30- μ m, disposable filter (Partec, Görlitz, Germany) to eliminate particles large enough to obstruct the tubing of the cell sorter. At all times, samples were kept on ice, in the dark and used for cell sorting within 2 h.

2.3.2. Hybridization with nucleic acid dyes

Samples that were sorted using a non-targeted, sectoring method discussed in 2.4.3. were not labeled with organism-specific probes. Instead, 500 μ L of fixed sample was centrifuged ($3000 \times g$, 5 min) and the resulting pellet was resuspended in 50 μ L of $1\times$ PBS with 5 μ M of nucleic acid dyes SYTO 9 (green) and SYTO 62 (red) (Life Technologies, Grand Island, NY, USA) for 15 min. Prior to sorting, samples were diluted in 1 mL additional $1\times$ PBS and filtered as discussed above.

2.4. Cell sorting

All single cells sequenced and analyzed in the following chapters were sorted based on fluorescent emission using a Cytopeia INFLUX sorter (BD, Franklin Lakes, NJ, USA). In this process, a sample with labeled bacterial cells was attached to the sorter, and the sample was passed through the sorter in a flow of $1\times$ PBS sheath fluid at a rate that allowed cells to be interrogated individually. Cells were passed through the path of multiple lasers. Forward scatter of the light was used to separate cells based on differences in size and optical density. An additional laser was used for excitation of fluorescently labeled cells, creating a detectable fluorescent emission. Cells with the desired forward scatter and levels of fluorescence emission were then selected using charge deflection and binned into tubes or individual wells of a 96-well plate.

2.4.1. Cleaning procedures for flow cytometer

Because material for genome sequencing comes from only a single cell, it was important that the risk of contamination was minimized. Any contaminating bacterial or human cells could easily be amplified in downstream processes, making it difficult to glean any useful information from the intended target. Thus, cleaning procedures are an important part of the process and have been discussed in the literature. The influx cell sorter used for the methods discussed in this dissertation was cleaned prior to use in a similar manner to those described earlier [72, 91].

Fortunately, the design of the Influx sorter minimizes the potential for contamination. Sheath fluid is held in a UV-sterilized reservoir tank, flow rate is controlled by head pressure using filtered air, and the fluid flows through easily replaceable tubing and a disposable filter until it reaches the nozzle assembly. Similarly, the sample is injected into the center of the nozzle assembly through replaceable line. For initial cleansing, fluidic lines were rinsed with 10% bleach for 45 min and 0.2- μm filtered ddH₂O for 30 min, followed by rinsing with sheath fluid for 1 h prior to sorting. The final rinse with sheath fluid allowed the flow rate to become constant, ensuring accurate sorting of individual cells. The sheath fluid was UV-sterilized overnight. Between successive samples, fluidic lines were flushed with 10% bleach, followed by sheath fluid.

Periodically, the machine was checked to ensure fluidic lines were clean, and there was no cross-contamination of samples. This was done by running filtered 1 $\times$ PBS from a sample tube and comparing the event rate to the

background rate seen from only the sheath fluid reservoir. Normal background rates were between 1-10 events per second and were always seen outside the sorting gate. Sorting rates with actual samples were typically 2000-7000 events per second.

2.4.2 Sorting of oral samples labeled with targeted probes

After the sorter was cleaned, oral samples were placed on the sorter for interrogation. Cells of interest labeled with a specific probe were selected based on forward scatter position and either green (528-538 nm) or red (670-730 nm) fluorescent emission. This is done in the software associated with the flow cytometer by creating a sorting “gate” on the desired portion of the scatter pattern seen on the flow cytometer output (**Figure 2.2**). This can be used to separate labeled cells from other cells and debris in the sample that also appear as “events” in the scatter pattern. For example, in **Figure 2.2**, a *Deltaproteobacteria*-specific probe has been hybridized with the sample, and a sorting gate was used to select these cells, which have higher fluorescent emission.

Single cells of interest were sorted into 3 μ L UV-sterilized, 1 $\times$ TE (10 mM Tris, pH 8.0, 1 mM EDTA, pH 8.0) buffer in individual wells of a UV-sterilized, 96-well plate (BioRad MLL-9601). Plates were covered with adhesive seals and centrifuged briefly (1500 $\times$ g, 30 s) to ensure the sorted cells and TE were in the bottom of the wells. Plates were immediately used for lysis amplification or were stored at -80 C. Stored plates were placed in plastic and vacuum-sealed to ensure adhesive seals did not become loose and allow samples to become contaminated.

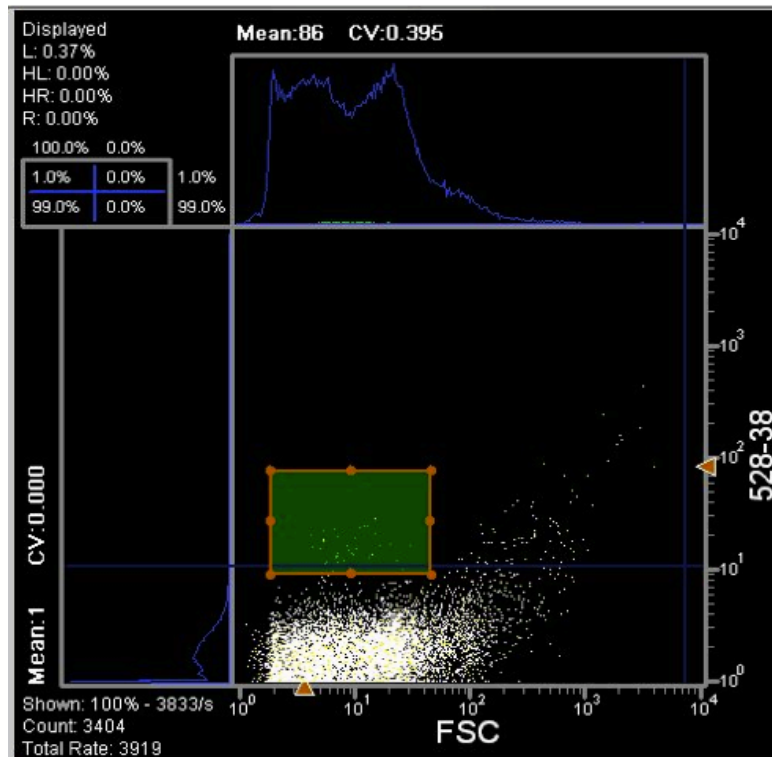


Figure 2.2. Output from the flow cytometer showing the selection of a specific group of cells based upon forward scatter (FSC) and fluorescent emission (528-538). Each white dot represents one event seen in the flow cytometer. Events that fall within the green selection box (also known as a “gate”) will be sorted into a selected tube or 96-well plate. The remaining events in the figure represent non-labeled bacterial cells as well as other debris.

2.4.3. Sectoring sorting of oral samples labeled with nucleic acid dyes

Although a group-specific FISH probe was used to target some bacteria as discussed above, the number of uncultured bacteria within the oral cavity was too large to specifically target each one individually. Therefore, a technique was developed during the course of this dissertation that could be used to quickly and efficiently retrieve multiple cells of uncultured organisms from an environment of interest.

To determine if different species of microbes could be preferentially enriched based on the forward scatter pattern from a flow cytometer, the patterns of four cultures were compared. This test was done using the following organisms: *Desulfovibrio piger*, *Staphylococcus aureus*, *Bacteroides vulgatus* and *Akkermansia muciniphila*. Each culture was processed on the flow cytometer using the same parameters, and the resulting scatter plots were color-coded and superimposed after sorting (**Figure 2.3**). Each species had a unique pattern, although some cultures had regions of overlap. There were discrete regions of the pattern for both *D. piger* and *S. aureus* that could be selected to easily isolate these cells from others in a mixed population. *B. vulgatus* and *A. muciniphila* overlapped to a greater extent, but there were positions in the scatter that sorting gates could be placed to greatly enhance the population of these species at the exclusion of other cells in a mixed sample. Although simplistic, this test verified that scatter plots from a flow cytometer did show distinct patterns for different bacterial cells that should be useful for isolation of different community members.

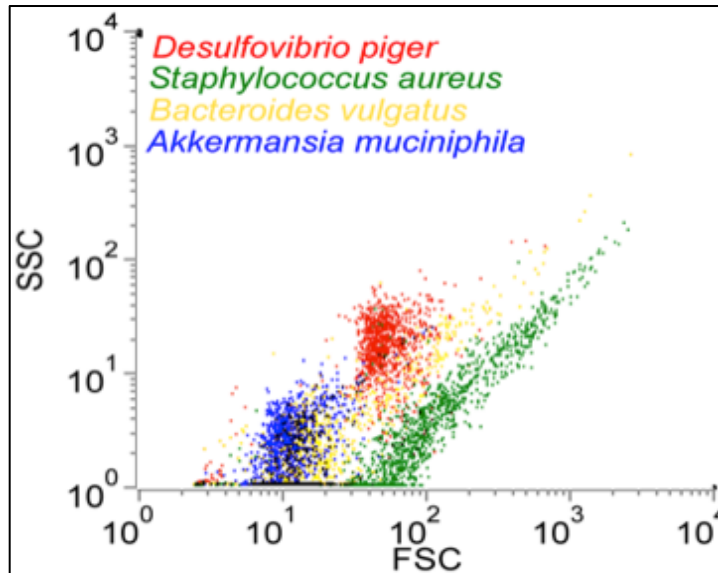


Figure 2.3. Overlapping scatter plots for four known cultures. The forward and side scatter pattern for each culture has been color-coded and corresponds to the culture name found in the top left corner. Samples were all sorted consecutively with identical settings.

The technique of sectoring a flow cytometry scatter pattern was next applied to oral samples from a healthy individual and one with periodontitis. The samples were stained with the nucleic acid dyes SYTO 9 and SYTO 62 (Life Technologies, Grand Island, NY USA), and the separation of the samples was observed using a scatter plot for forward scatter pattern (x-axis) and fluorescent emission at 528-538 nm (y-axis). Sorting gates were based on 7 distinct sectors of the scatter pattern labeled A – G (**Figure 2.4**). The method of sectorized sorting was able to further separate high-density areas on the scatter plot and isolate low density areas of the scatter plot, thus increasing the chance of retrieving novel cells for further processing.

All sectors were not identical in size, but were adjusted based upon the number of events captured within them. For example, sorting gate “G” was made

much larger because a relatively low number of events were captured in this area, increasing the amount of time and sample needed to capture these cells. The overall scatter patterns in the healthy and periodontitis samples were similar but not identical; therefore, sorting gates A-G were not identical but were found in similar positions relative to the overall sample pattern (**Figure 2.4**).

Approximately 20,000 cells from each sector were sorted into a single tube using green fluorescent emission (SYTO 9). This tube was then subjected to a second round of sorting based upon red fluorescent emission (SYTO 62) to further dilute any possible contaminating DNA found in the original sample [91]. Cells were sorted as single cells into 96-well plates in 3 μ L UV-sterilized, 1 $\times$ TE buffer. Two, 96-well plates of single-cells were sorted from each sector.

Sorted plates were lysed and amplified, and successful amplifications were identified using universal bacterial primers 27fm (5'-AGA GTT TGA TYM TGG CTC AG-3') and 1492r (5'-TAC GGY TAC CTT GTT ACG ACT T-3') [92]. The identified cells from each sector confirmed the usefulness of this technique for separating and capturing the diversity of microbes found in the oral samples (**Figures 2.5, 2.6**). Using two oral samples, 48 distinct taxa were sorted, and a large amount of diversity was captured in several of these groups.

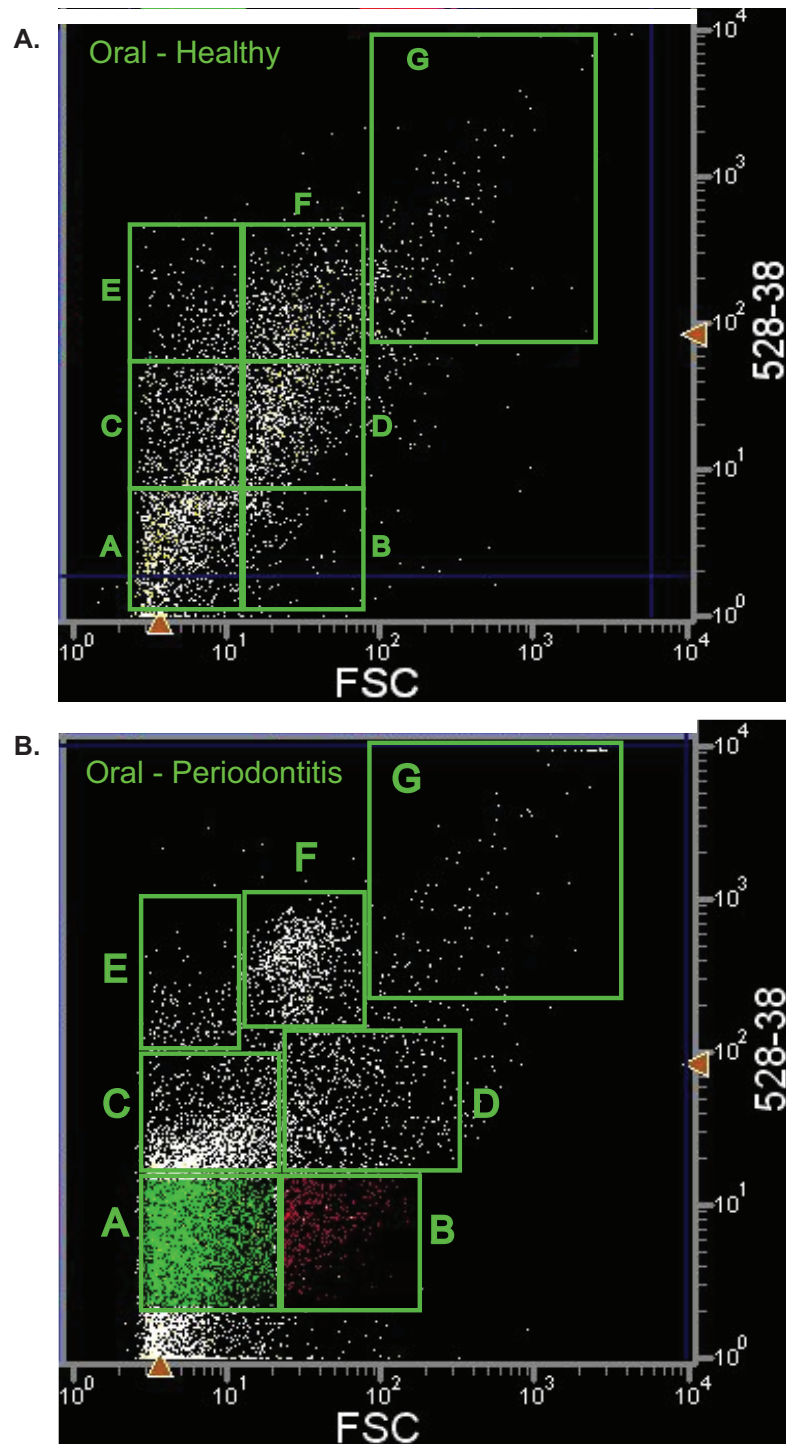


Figure 2.4. Scatter plots for oral samples from (A) a healthy individual and (B) an individual with periodontitis. Sectors, and the resulting sorting gates, are shown in green and labeled A-G.

Gates C and D from the healthy oral sample preferentially sorted *Selenomonas* species, *Prevotella* species were most abundant in gate A and Synergistes dominated in gate F (**Figure 2.5**). While several gates successfully captured one dominant member, other gates were able to capture a wide variety of organisms. For example, sector G in the healthy oral sample captured several unique microbes, including *Dialister*, *Pirellula*, *Propionivibrio* and *Shuttleworthia* species. A random sorting method without the use of sectors would likely miss these species because more abundant targets in the other sectors would be sorted with higher frequency.

Separation of distinct microbes within each sector was also observed in the periodontitis sample (**Figure 2.6**). Gate C preferentially captured *Tannerella* species, gate E isolates were mainly *Fusobacterium* species and gate F captured preferentially Synergistes. Several organisms were captured by similar gates in both the healthy and diseased oral sample. In both samples, gate E captured an abundance of *Fusobacterium*, gate F was dominated by Synergistes and gate B was dominated by *Streptococcus* species. However, many differences were observed between the samples. The healthy oral sample had an abundance of *Selenomonas* species in gates C and D not seen in the diseased sample. Likewise, the periodontitis sample captured many more *Porphyromonas* cells than were seen in the healthy sample, particularly in gates D and G. Unique microbes were captured from both samples.

Perhaps the most striking example of successful isolation was found in gate F, which was dominated by uncultured members of the phylum

Synergistetes in both samples. This phylum accounts for <0.5% of all the clones found in the human oral microbiome database [43] and was found at an average of <2% in the healthy and periodontitis samples studies in Griffen *et al.* [18]. Therefore, the method of sectored sorting was able to selectively isolate this low-abundance group of organisms from both healthy and periodontitis samples. There is a large diversity of Synergistetes associated with humans, and many have been linked to states of health and disease [18, 93-95], making this group a relevant target for genomic characterization. Sectored sorting of both healthy and diseased oral samples captured a wide range of diversity within the Synergistetes (**Figure 2.7**). Because only two individuals were sampled, it is unclear if these groups can be linked to health or disease.

To date, the sectored sorting technique has been used in our laboratory to sort and amplify over 2000 single amplified genomes (SAGs) from both oral and fecal samples, yielding over 75 bacterial genera. Approximately 150 single SAGs have been purified and sent to be sequenced at HMP sequencing centers (upon their request) in order to capture genomic data for several “most wanted” bacteria [96] and expand a reference dataset for the taxa associated with the human body. Thus, this method has proven to be an effective and high-throughput procedure that can be used to generate large amounts of genomic data for a given environment.

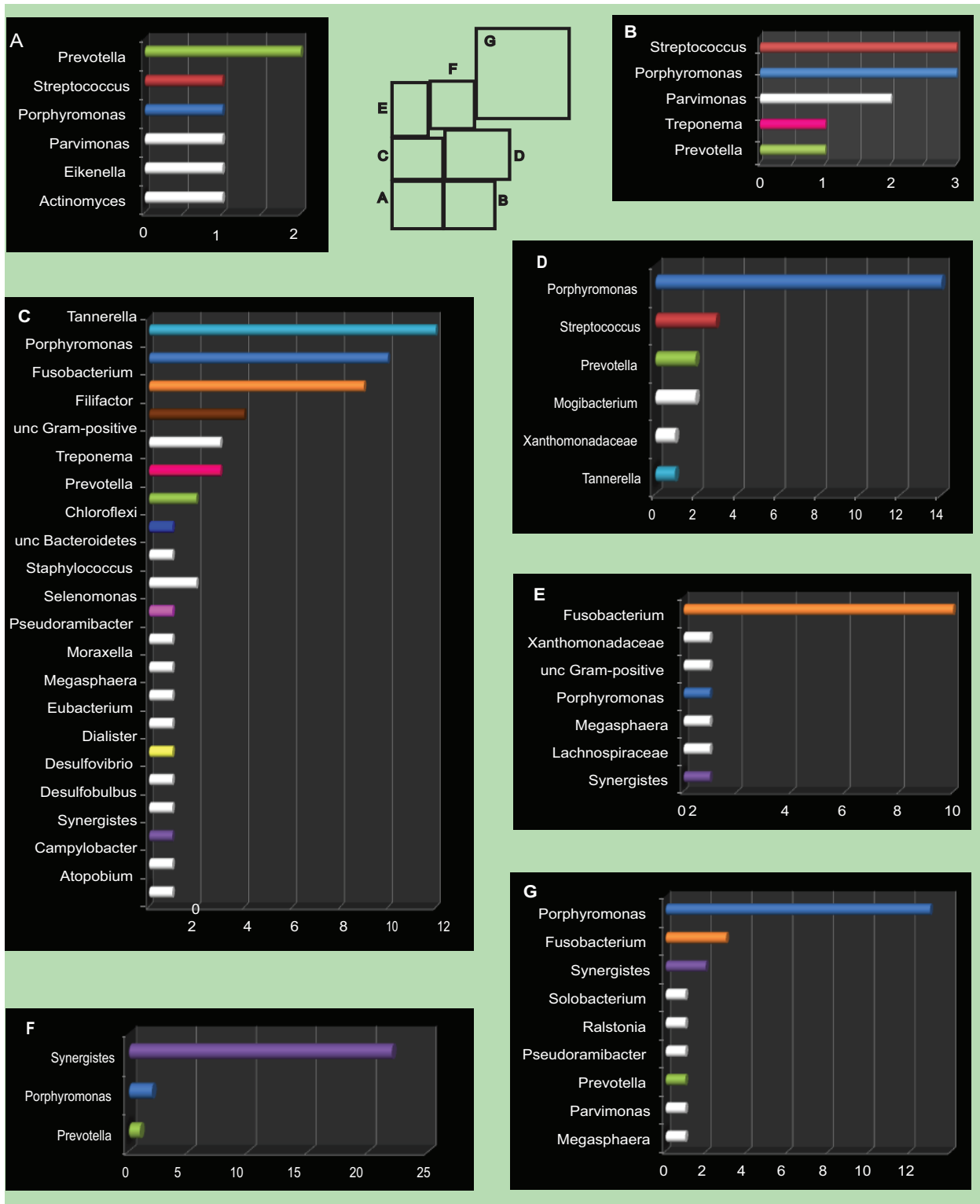


Figure 2.5. Organisms captured with each sorting gate in a healthy oral sample. The x-axis shows the number of cells collected, and groups of interest have been colored. Identifications for all successful amplifications in each gate are shown.

SYNERGISTES

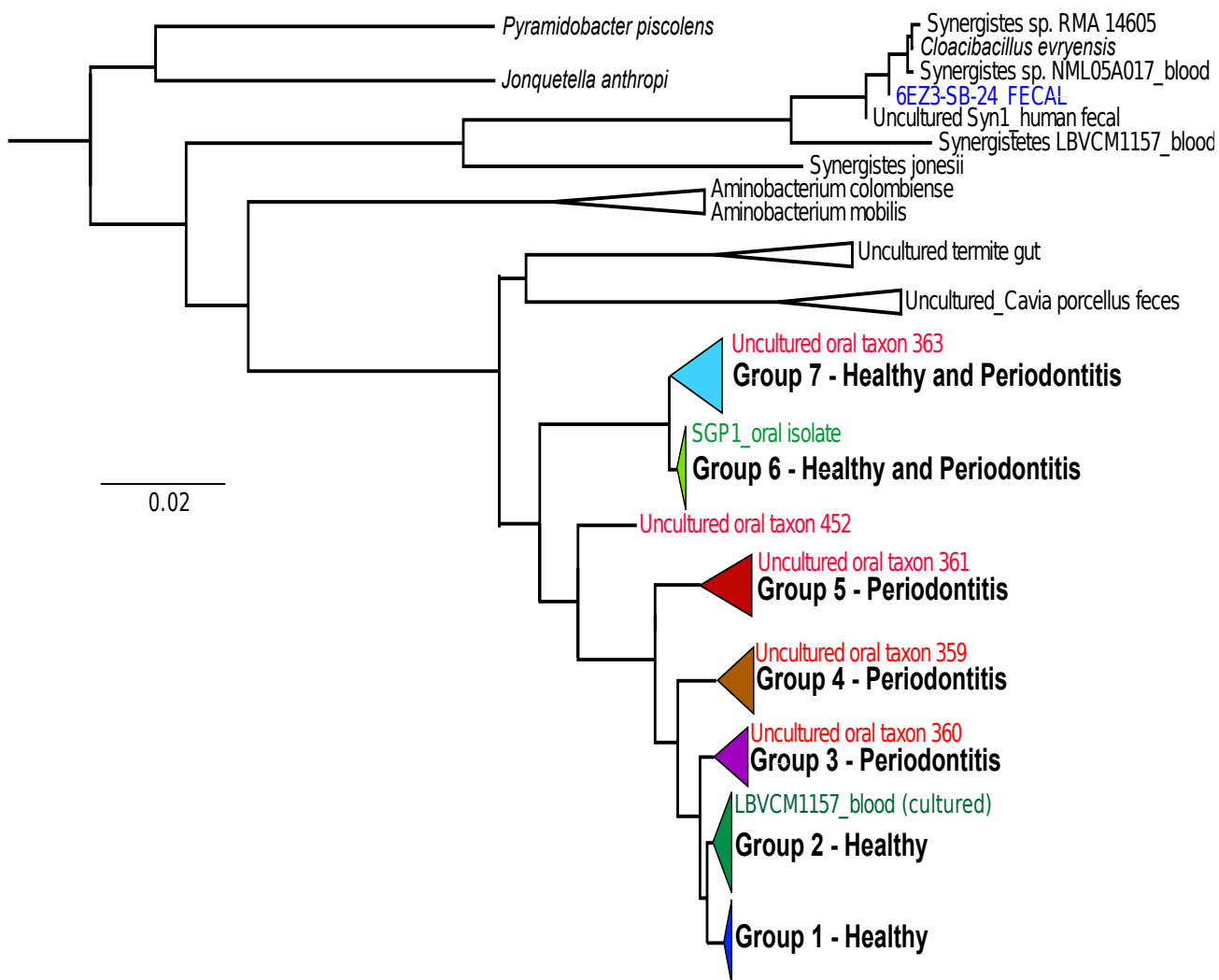


Figure 2.7. Neighbor-joining phylogenetic tree of a subset of *Synergistes* based on SSU rRNA genes. SAGs captured in this study are shown as groups 1-7 and have been labeled according to the sample type from which they were retrieved.

2.5. Single-cell amplification

Single cells retrieved from the cell sorter need to be lysed and amplified in a multiple displacement amplification (MDA) reaction to produce enough material for further investigation. Once a single cell has been isolated from the complex oral samples, it is still not available for genomic sequencing. Cells must first be lysed to allow access to the genomic DNA. If the type of cell sorted is unknown, it is difficult to make a lysis protocol efficient for all cell types found in a sample. In order to increase the likelihood of cell lysis, both chemical and physical lysis procedures were used.

Although amplification from a single cell was achieved several years ago [64], initial amplifications were highly biased and resulted in short amplicons [38]. DNA polymerase ϕ 29, originally isolated from a bacteriophage of *Bacillus subtilis* and used for rolling circle amplifications [67], greatly enhanced amplification of single cells. This polymerase is a highly processive enzyme with high strand displacement properties [65, 66], 3' to 5' proofreading activity [97] and functionality at 30 C. Random hexamers are used as primers and have phosphothioate modifications at the 3' end to avoid the exonuclease activity of the polymerase [67]. When combined with random primers, ϕ 29 has the ability to randomly amplify genomic DNA to produce micrograms of DNA from a single cell (Figure 2.8).

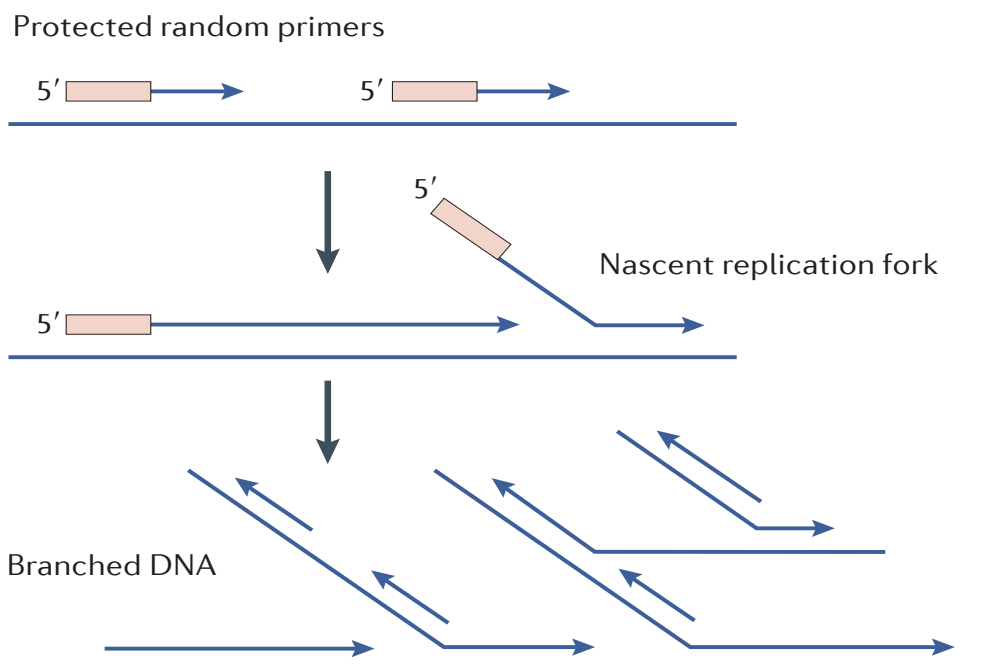


Figure 2.8. Image depicting the process of multiple displacement amplification using ϕ 29 DNA polymerase and protected random primers. Illustration reprinted from Lasken 2012 [38]. The high processivity and strand displacement activity of ϕ 29 DNA polymerase leads to the formation of branched DNA and highly efficient amplification of the genome.

2.5.1. Reagents used for cell lysis and amplification

The EDTA, Tris and water used for lysis and neutralization buffers were from Ambion (Austin, TX, USA) and DTT was purchased from Roche (Indianapolis, IN, USA). For the amplification mix, random hexamers with two protective, phosphothioate bonds on the 3' end were purchased from Integrated DNA Technologies (Coralville, IA, USA), dNTPs were purchased from Roche, ϕ 29 DNA polymerase buffer was purchase from New England BioLabs (Ipswich, MA, USA), and ϕ 29 DNA polymerase was made in-house with a construct generously provided by Paul Blainey using the method previously described [74].

All plasticware, water, lysis (without DTT) and neutralization buffers used for cell lysis and amplification were UV treated as described [98]. Briefly, a Stratalinker (Stratagene, Santa Clara, CA, USA) was used for UV treatment. The Stratalinker was modified to enhance the UV treatment. A stand was built to bring reagents closer to the light source during treatment, and the interior of the machine was covered with aluminum foil to reflect the UV source. Liquid reagents were placed in closed tubes on their side and treated for 30 min. Plasticware was treated for 10 min.

2.5.2. Procedure for single-cell lysis and amplification

Lysis, neutralization and MDA reactions were set up in a manner similar to that used by Rodrigue and colleagues [91]. Cells were lysed by addition of 3 μ L lysis buffer (**Table 2.1**), heated to 95 C for 30 s, and immediately placed on ice for 10 min. Lysis was halted by the addition of 3 μ L neutralization buffer (**Table 2.1**). Cells were initially sorted into 3 μ L of 1 $\times$ TE; therefore, lysates have a total volume of 9 μ L at this point. To the 9 μ L lysate, 11 μ L of an amplification master mix (**Table 2.1**) was added. Reactions were amplified for 10 h at 30 C and then inactivated at 80 C for 20 min. Desirable MDA products identified were used to generate secondary MDA products. Five secondary MDA reactions were set up for each single-cell amplicon, and each reaction used 1 μ L of a 1:10 dilution of the original MDA product. All buffers and amplification mixes for secondary reactions were constructed as described above for the primary MDA reactions. Secondary MDAs were amplified for 6 hrs before heat inactivation.

Table 2.1. Composition of lysis, neutralization and amplification buffers used for single-cell lysis and amplification procedures. Volumes are shown for the amount need to process one 96-well plate.

Lysis Buffer			
	Concentrations		
Reagent	Stock	Final	Volume (μL)
KOH	1.0 N	0.13 N	65.00
EDTA	500 mM	3.3 mM	3.30
DTT	1000 mM	27.7 mM	13.85
Water			417.85
Neutralization Buffer			
	Concentrations		
Reagent	Stock	Final	Volume (μL)
HCl	1.0 N	0.13 N	65.00
Tris pH 7.0	1 M	0.42 M	210.00
Tris pH 8.0	1 M	0.18 M	90.00
Water			135.00
Amplification Master Mix			
	Concentrations		
Reagent	Stock	Final	Volume (μL)
φ29 buffer	10 ×	1.82 ×	200.00
random primers	1000 μM	90.9 μM	99.99
dNTPs	25 mM	1.09 mM	47.96
DTT	1000 mM	7.27 mM	8.00
φ29 DNA polymerase	300 U/μL	272.73 U	75.00
water			669.05

2.6. Identification of target cells

Once the lysis and amplification process was complete for a 96-well plate, the identity of each single cell was still unknown. Further, it was possible that some wells of a 96-well plate did not receive a bacterial cell or that some cells did not lyse. Thus, a subset of each MDA amplicon was used for downstream PCR screening and identification.

For each MDA amplicon, 1 μ L was diluted (1:150) in PCR-grade water, and the remainder of each MDA was vacuum-sealed and stored at -80 C. Dilutions were used to set up PCR amplifications using universal bacterial primers 27fm (5'-AGA GTT TGA TYM TGG CTC AG-3') and 1492r (5'-TAC GGY TAC CTT GTT ACG ACT T-3') [92] for randomly sorted cells or group-specific primers when screening FISH-isolated cells. Fifty-microliter reactions were constructed as shown in **Table 2.2**. PCR reactions ran under the following thermal conditions: 94 C for 2 min, followed by 30 cycles of 94 C for 30 s, 55 C for 30 s, 72 C for 1.5 min and a final extension at 72 C for 5 min.

Amplicons were visualized by gel electrophoresis within 1% (w/v) agarose gels (**Figure 2.9**). All reactions that produced clean bands were purified with PCR cleanup filter plates (Millipore, Billerica, MA, USA) following the protocol provided. In this purification, samples were placed in a 96-well plate that can be attached to a vacuum manifold. Samples are retained in the plate on a size exclusion filter, rinsed twice with Tris (pH 8.0) buffer and removed from the purification plate for downstream sequencing.

Samples were sequenced directly from the 27fm primer with Sanger sequencing at the University of Tennessee Molecular Biology Resource Facility (Knoxville, TN, USA). Resulting chromatograms were manually edited in Geneious® Pro 5.6.5 [99] and used as an initial assessment of purity (**Figure 2.9**). Phylogenetic identifications were obtained using RDP Classifier [100].

Table 2.2. Reagents for the setup of one polymerase chain reaction (50 μ L).

PCR Master Mix			
	Concentrations		
Reagent	Stock	Final	Volume (μ L)
Polymerase buffer	10×	1×	5.00
dNTPs	25 mM	200 μ M	0.40
MgCl ₂	50 mM	2 mM	2.00
BSA	500 μ g	5 μ g	0.50
Forward Primer	10 mM	300 μ M	1.50
Reverse Primer	10 mM	300 μ M	1.50
<i>Taq/Pfu</i> polymerase	250 U	1 U	0.20
Water			37.90
1:10 diluted MDA product			1.00

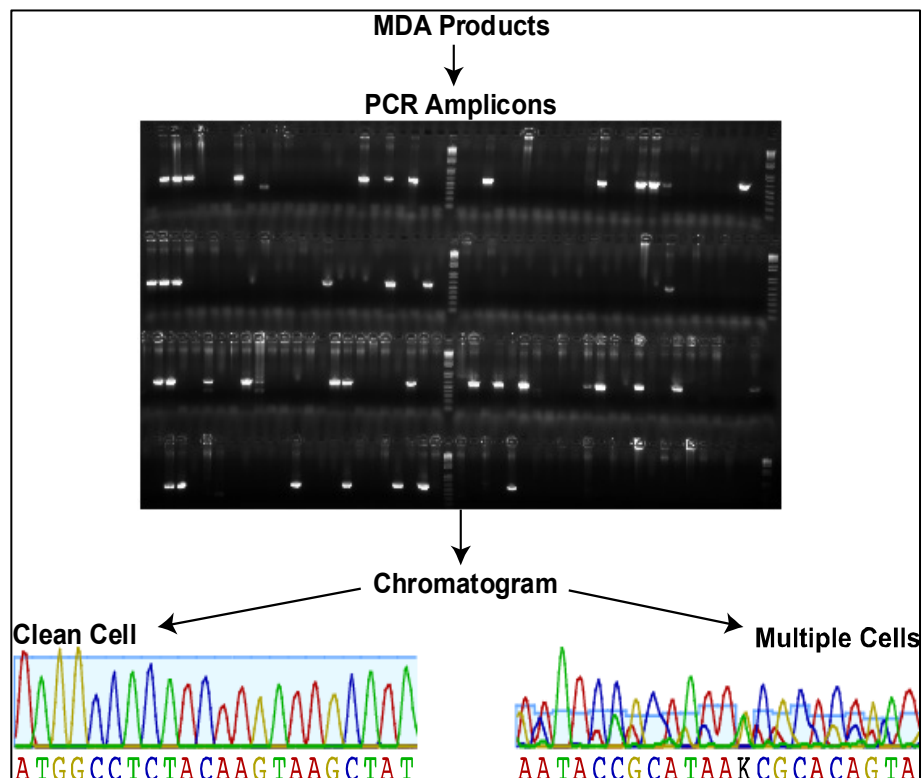


Figure 2.9. Illustration of MDA product PCR amplification and subsequent quality check using resulting chromatograms. As seen in the gel image, only a subset of sorted, single cells will result in positive amplifications with bacterial primers (two 96-well plates of MDAs are shown). An amplicon from a single cell will result in a clean chromatogram whereas an amplicon that originated from multiple cells will have background peaks.

2.7. Whole genome amplification

Both primary and secondary MDA products of interest were purified with a phenol:chloroform:isoamyl alcohol (PCIAA) extraction and alcohol precipitation method before being sent for sequencing. Illumina Hi-Seq 100-bp, paired-end libraries were constructed and sequenced at Hudson Alpha Institute for Biotechnology (Huntsville, AL, USA). In this process, the purified genomic DNA was first fragmented into small pieces and overhangs were removed to leave blunt-end fragments. The 3' end of the fragments was then prepped for the addition of adapters by attaching an 'A' base using the adenylation activity of Klenow fragment. Adapters were then attached to the 3' ends of fragments, and unligated adapters were removed using a gel purification procedure. PCR primers that match one end of the ligated adapters were used to amplify the genomic fragments.

Genome fragments with adapters were then added and attach to a flow cell that was coated with oligonucleotides. Here, a clonal amplification was performed, and the dsDNA was linearized into single strands that can be sequenced. To prevent nonspecific sequencing, the free 3'-OH ends of the dsDNA clusters must be blocked. Finally, the dsDNA is denatured to allow the hybridization of a sequencing primer, and the sample can be sequenced.

Samples discussed in this dissertation were indexed and sequenced with 2-5 samples in each lane. This resulted in a minimum of 30 Gb data for each sample.

2.8. Normalization, assembly and annotation of genomic information

Genomic data from SAGs is slightly different than that of genomes from cultured organisms. The major difference results from the random bias seen during MDAs. The area of genomic DNA where random primers first attach and amplification begins will generate a greater number of copies than areas of the genome that are not primed until later in the MDA reaction. This results in highly skewed sequencing data, with some sections of DNA represented by thousands of copies and other sections of the genome represented by only a few copies. Most genome-assembly software available is not equipped to deal with this highly variable coverage and does not provide an accurate assembly of SAG data. Therefore, several techniques have been developed in the field of single-cell genomics to normalize coverage across the genome.

Attempts to normalize coverage by single-cell researchers were first done in the laboratory, prior to genome sequencing [91]. After MDA amplification, a duplex-specific nuclease [101] that specifically degrades double-stranded DNA was added to denatured samples, and kinetics of re-annealing allowed copies of more abundant targets to be degraded [91]. However, as datasets from sequencing efforts became progressively larger and more inexpensive to produce, it became easier to normalize genome coverage computationally, after the sequencing process was complete. Several normalization software programs have been developed for both single-cell and metagenomic datasets [85, 102-104].

Some of the SAG data presented in this dissertation were processed using the Joint Genome Institute's (JGI) single-cell assembly and QC pipeline, similar to Illumina SAG libraries described previously [77]. However, the majority of data described in the following chapters was normalized, assembled, checked for contamination and annotated in-house using publicly available software. These processes were accomplished using digital normalization software khmer [85] and the assembly software package Velvet [105].

The digital normalization software compares the genomic data using a user determined fragment length and discards all data that exceeds a user defined maximum coverage. This results in an even coverage across the genome. The commands used for digital normalization can be found in **Table 2.3**, and the following parameters were used: (k) = 30, (C) = 30, (N) = 4 and (x) = 1e+09. The parameter (k) represents the length of k-mers that was used for the normalization process, and was set to 30 base pairs. (C) is the coverage threshold, and when a k-mer was represented by more than 30 reads, reads beyond 30 were removed from the dataset. (N) is the number of hashes used and (x) is the minimum hash size. These parameters are used to estimate the amount of memory that will be used during data processing, and 4 GB (hash number $\times$ minimum hash size) were used in the datasets discussed. A second step in this program was used to remove low abundance k-mers that were likely due to errors. For datasets analyzed using this software, less than 10% of original data remained after these normalization steps. One SAG from the *Deltaproteobacteria* dataset discussed in Chapter 3 can be seen in **Figure 2.10**.

Table 2.3. Commands used for genome normalization, assembly and gene prediction.

Program	Purpose	Command
Python	Combine both sequencing files from paired-end reads	<code>python /khmer/sandbox/interleave.py File1.fasta.qz File2.fasta.qz Files1and2.fastq &</code>
khmer	Ensure proper python is used	<code>export PYTHONPATH=/directory/khmer/python/</code>
khmer	Normalize data by removing coverage > 30×	<code>/khmer/scripts/normalize-by-median.py -C 30 -k 30 -N 4 -x 1e9 --savehash /directory/Data.kh /directory/Files1and2.fastq</code>
khmer	Remove low abundance k-mers/ output is Data.fastq.keep.abundfilt.pe	<code>/khmer/scripts/filter-abund.py Data.kh Data.fastq.keep</code>
Velvet	Assemble data using a range of overlaps (53-65 bp, in 2 step increments)	<code>./velveth Data 53,65,2 -shortPaired -fasta Data.fastq.keep.abundfilt.pe</code>
Velvet	Check for best assembly/run for each overlap used in the previous step	<code>./velvetg Data_59 -exp_cov auto -cov_cutoff auto</code>
Prodigal	Predict genes for best assembly	<code>./prodigal -i Data_59.fa -o Data -a Data_prot -f gff</code>

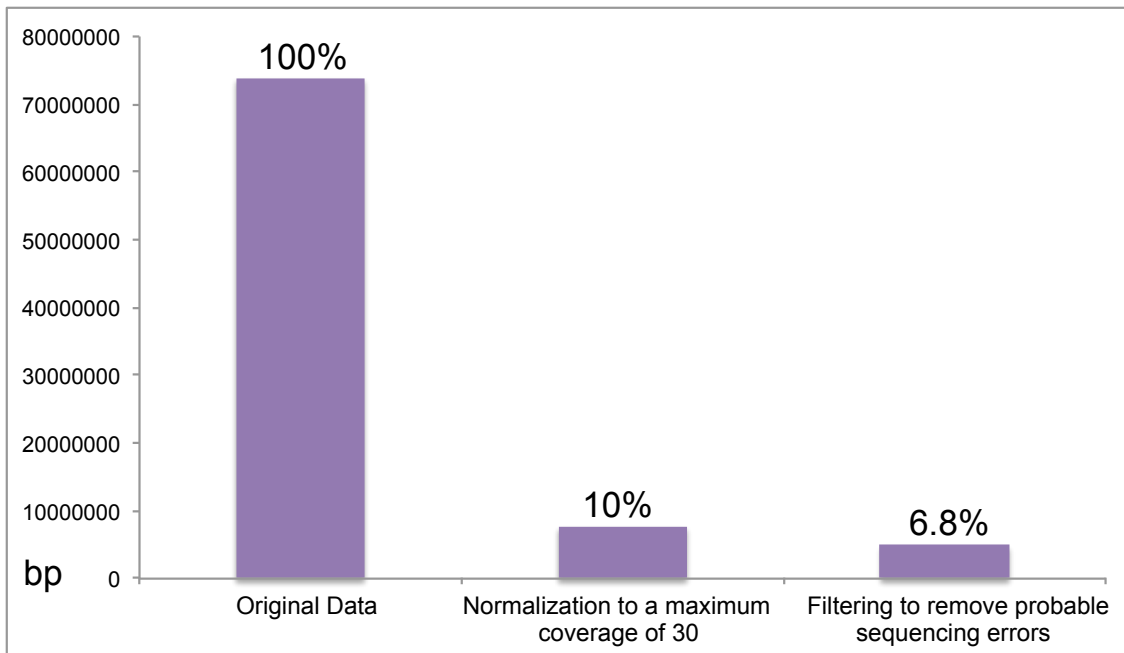


Figure 2.10. Data reduction during the normalization process used for SAGs. Data shown is for SAG Dsb3 discussed in Chapter 3.

Once normalized, each genome was assembled in Velvet using a range of hash lengths (53 to 65 with a step of 2). All other assembly parameters were left as default, and the commands are listed in **Table 2.3**. The hash length represents the number of overlapping basepairs that was required between two reads to allow assembly. For example, a length of 53 would assembly two reads only if at least 53 bp of each 100 bp read overlapped. The amount of overlap required between reads affects the specificity of the assembly and was optimized for each dataset processed. The number of nodes (scaffolds), n50, maximum number of bases in a scaffold and total number of bases for each hash length were compared to select the final assembled dataset. The optimization of one SAG from the dataset in Chapter 3 is shown in **Table 2.4**.

Assembled genomes were subjected to the JGI/ORNL automated annotation pipeline, which uses the gene calling program Prodigal [106]. Some datasets had gene predictions performed in-house, also using the program Prodigal. The codes used for gene prediction can be found in **Table 2.3**.

Table 2.4. Optimization of resulting Velvet assemblies over a range of hash lengths for SAG Dsb3 discussed in 3. The assembly that was used for downstream analyses has been highlighted in yellow.

Hash length	Nodes	n50	Maximum scaffold length	Total Assembly Size (bp)	Total Reads Used	Total Reads
53	298	9549	23703	1052803	624825	1532060
55	216	11982	23998	1049161	563368	1532060
57	157	13800	25758	1033939	508232	1532060
59	141	13800	30991	1004061	455941	1532060
61	132	14608	32539	994311	412123	1532060
63	144	14443	30828	1031995	382519	1532060
65	94585	36	1357	1838413	128300	1532060

2.9. Data analysis

The first step of analysis for each dataset was contamination removal. The main tool for contamination removal was BLASTP [107] for each predicted protein. This allowed for easy removal of human contaminants as well as other bacteria. Scaffolds that contained predicted proteins with top BLASTP hits (based on e-value) that were greater than 90% identical to an unrelated organism were removed. Secondary contamination checks were performed by analysis of GC content (**Figure 2.11**) and comparison of the nucleotide composition of scaffolds using tetramer patterns [108] (**Figure 2.12**). In each case, the data was plotted and extreme outliers were further analyzed and often removed.

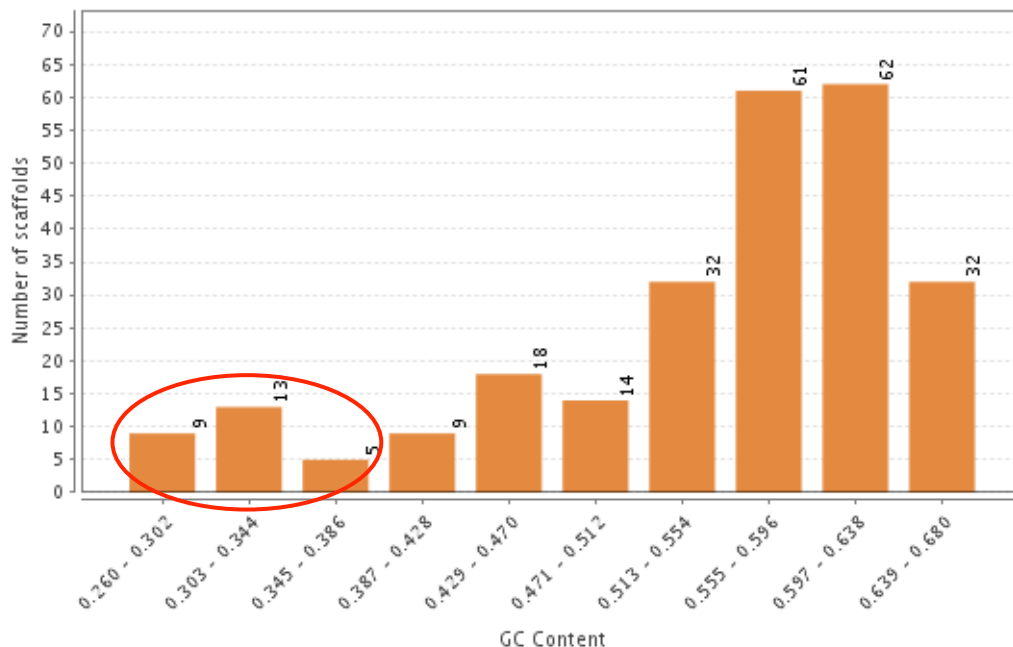


Figure 2.11. The range of GC content seen in scaffolds originally assembled for SAG Dsb3 (Chapter 3). Small scaffolds of human contamination circled in red were removed. Averaged GC content after contamination removal was 56.5%.

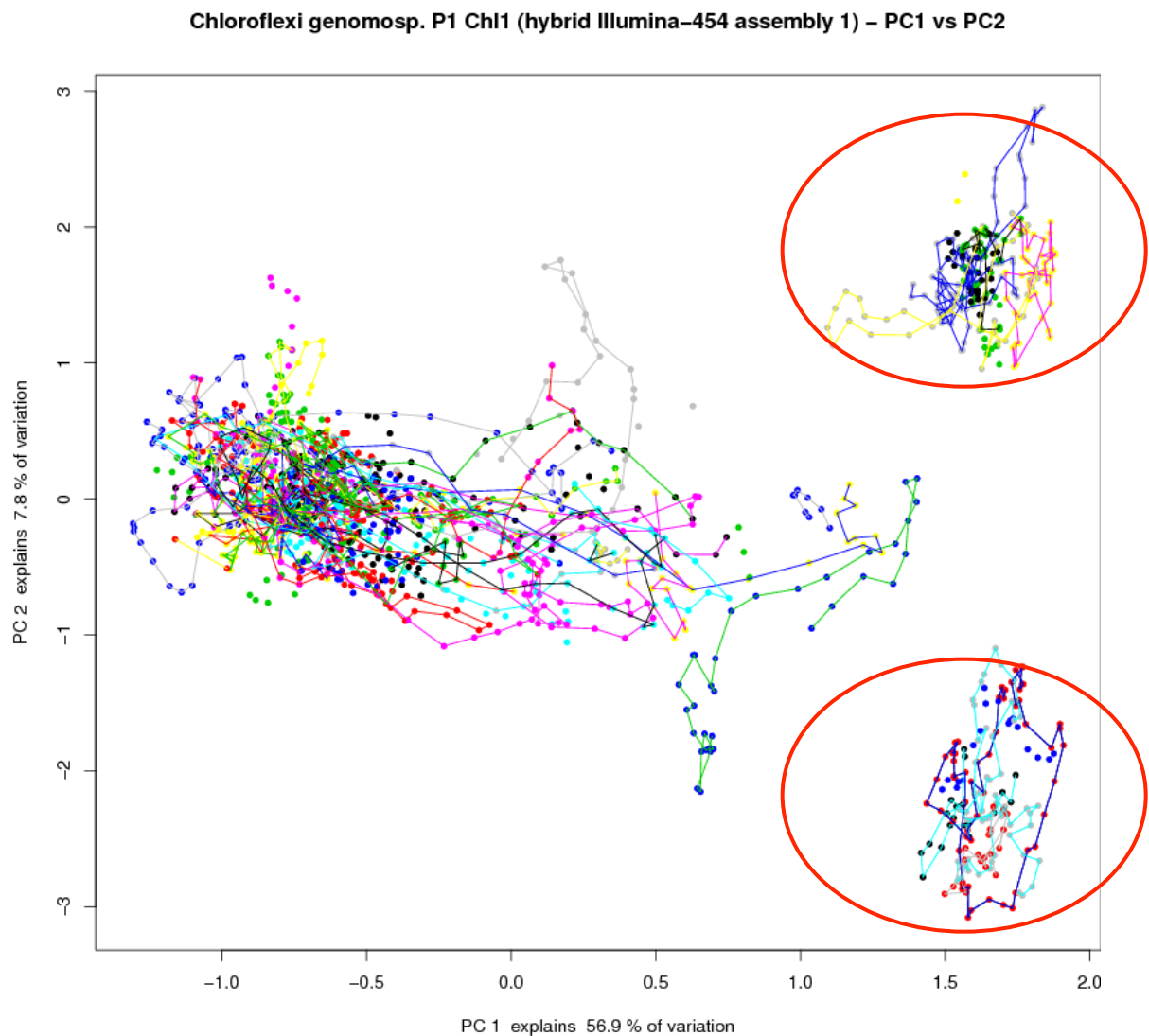


Figure 2.10. A tetranucleotide plot generated to detect possible contamination in a SAG Chl1 (Chapter 4). Scaffolds later found to be contaminants have been circled in red.

2.9.1. Metabolic potential and genomic properties

Most data analysis was done with tools available in Integrated Microbial Genomes (IMG) [109] and RAST [110]. Both websites allow the user to upload private genomic data and compare these data to publicly available genomes using a variety of tools. Initial metabolic capabilities of our genomes were assessed using KEGG maps [111]. Further steps of analysis used clusters of orthologous groups (COGs) of proteins [112] and their corresponding categories to break data into manageable subsets. COGs were also used for comparisons of genomes to closely related species. By comparing the presence and absence of COGs, differences between genomes were identified. Both protein families (pfam) database information [113] and BLASTP top hits were used to confirm the most likely function of genes of interest.

2.9.2. Combined non-redundant assembly

To assess the phylogenetic relationships between multiple related amplicons, collaborators at JGI performed comparisons using average nucleotide identity (ANI) and tetranucleotide frequency. ANI is used to directly compare the percent identity of overlapping sequence from two datasets while tetranucleotide frequency compares the usage of unique, 4-bp patterns of overlapping sequence. Both ANI and tetranucleotide frequencies were calculated using the software JSpecies [114]. When comparisons of ANI and tetranucleotide frequencies revealed closely related datasets (>97%), genomic information was combined to generate more complete datasets that were used to discuss the potential function of oral groups of bacteria as a whole. To obtain a combined

dataset purged of replicate genes, repetitive regions between amplicons were removed iteratively with the aid of custom perl scripts written and run by collaborators at JGI. A region was considered redundant if it exhibited more than 95% nucleotide identity over a region of at least 1000 bp. Remaining overlaps were assessed using BLASTCLUST [107] using parameters that formed clusters based on 95% identity over at least 70% of the length of both sequences ($S = 95$, $L = 0.7$). Overlaps were also identified using Mauve [115] and the *de novo* assembler within Geneious® Pro 5.6.5 [99] and removed manually.

2.9.3. Genome size estimation

To assess the impact of missing genes, it was important to get an accurate assessment of the amount of the genome that has been captured by our sequencing efforts. Thus, collaborators at JGI estimated genome size and coverage using a conserved single copy gene (CSCG) set that has been determined from all 1516 finished bacterial genome sequences in the IMG database [109]. The set consists of 138 CSCG that were found to occur only once in $\geq 90\%$ of all genomes by analysis of an abundance matrix based on hits to the pfam database [113]. Hidden Markov models of the identified pfams were used to search assemblies by means of the HMMER3 software [116]. Resulting best hits above pre-calculated cutoffs were counted, and the coverage was estimated as the ratio of found CSCG to total CSCGs in the set after normalization to 90%. Finally, complete genome size was estimated by multiplication of the predicted genome coverage by the total assembly size.

Chapter 3 - Multiple single-cell genomes provide insight into functions of uncultured *Deltaproteobacteria* in the human oral cavity

Alisha G. Campbell, James H. Campbell, Patrick Schwientek, Tanja Woyke, Alexander Sczyrba, Steve Allman, Cliff Beall, Ann Griffen, Eugene Leys and Mircea Podar

This chapter has been adapted from text that has been submitted for publication to the journal PLOS One. Alisha G. Campbell was responsible for the majority of the sample processing, data analysis and writing.

Acknowledgements

This research was supported by grant R01 HG004857 from the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) to M.P. and by grant 1R56DE021567 from the National Institute for Dental and Cranial Research (NIDCR) of the NIH to M.P., A.G. and E.L. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. PS, TW and AS were supported by the U.S. Department of Energy Joint Genome Institute and the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank the NIH Human Microbiome Project research community for making available sequence data used for comparative analyses. We also thank Dr. Paul Blainey for the phi29 polymerase overexpression clone, Sarah Kauffman for technical assistance with the project, Dr. Matt Heaton and Dr. Jeffrey Becker for advice and help with sample collection and processing.

3.1. Introduction

Recent efforts have profiled the large number of microbial communities associated with the human body and highlighted the importance of determining the composition and function of these communities and, ultimately, their effects on human health [22, 43]. Studies of microbes in the human oral cavity have

been ongoing since the discovery of “animalcules” by Antony van Leeuwenhoek in 1676, but despite this long history, over 50% of oral microbes remains uncultured. Likewise, periodontitis, an inflammatory gum disease, has been under investigation since the early 19th century, but the role of microbial communities associated with the disease remains unclear [56]. Periodontitis is the leading cause of tooth loss worldwide and has been linked to a number of systemic diseases, including diabetes, cardiovascular disease, osteoporosis and preterm low birth weight [50-55]. Thus, understanding the roles of periodontitis-associated microbial community members is of importance. Genomic information provides a way to assess functional potential; however, only a few, cultured bacteria associated with periodontitis have been sequenced and analyzed [60-62], leaving it challenging to fully comprehend this polymicrobial disease. Additional genomic information for associated microbes could ultimately result in more effective treatments and preventative measures for periodontitis.

Recent culture-independent studies have associated several uncultured organisms with periodontitis [18, 117]. One such group is sulfate-reducing bacteria (SRB). These organisms have been of long-standing interest because of their ability to produce hydrogen sulfide, a compound that can be toxic to human cells. Potential SRB within the oral microbial community include several members of *Deltaproteobacteria*. Both culture-based studies [118, 119] as well as quantitative PCR studies based upon dissimilatory sulfate reduction genes (*dsrAB*) [120] have linked SRB to periodontitis, particularly more clinically severe cases with deeper periodontal pockets.

Several culture-independent studies have linked the deltaproteobacterial genus *Desulfobulbus*, with progressive periodontitis [117, 121]. *Desulfobulbus* sp. R004 is present in low levels in most adults [122] but has been found in greater abundance in both healthy and diseased sites of subjects with periodontitis compared to healthy control subjects [18]. Despite these findings, little is known about oral *Desulfobulbus* species because there are no cultured representatives from this environment and only one genome published from this genus [123]. Isolates of *Desulfovibrio fairfieldensis* and closely related species have also been associated with periodontitis [124], and these organisms have been found as the causative agent of several cases of bacteremia [125, 126]. Although cultured isolates are available for *Desulfovibrio fairfieldensis*, no genomic information has been published. Obtaining genomic information for these oral *Deltaproteobacteria* would greatly enhance our knowledge and shed light on their potential function in the etiology of progressive periodontitis.

Thus, our aim was to selectively isolate single cells of oral *Deltaproteobacteria* and sequence their genomes. We sequenced both individuals and groups of single-cell amplicons most closely related to the uncultured *Desulfobulbus* sp. R004, as well as a group of single-cell amplicons most closely related to the uncultured *Desulfovibrio* sp. BB161 [43]. Several single-cell amplicons were sequenced to provide more complete genomic information for each genus. Genomic data were analyzed and compared to sequenced, environmental isolates of related organisms and other host-associated, *Deltaproteobacteria* genomes.

3.2. Materials and methods

3.2.1. Ethics Statement

Human subjects enrollment and sample collection protocols were approved by the Ohio State University Institutional Review Board and by the Oak Ridge Site-Wide Institutional Review Board. Signed informed written consent was obtained from all human subjects that provided samples for this study.

3.2.2. Sample collection

We acquired three single-cells (Dsb1, Dsb2, Dsb3) and two combined pools (Dsb4, Dsb5) of *Desulfobulbus* species, as well as one pool (Dsv1) of a *Desulfovibrio* species. Samples came from both periodontally healthy subjects and subjects with periodontitis. Patients were identified as having periodontitis based upon previously published criteria [18]. Dsb1 and Dsb2 were retrieved from a subgingival paper point sample taken from a healthy volunteer. Dsb3 was retrieved from a mixture of paper point samples taken from two subjects with periodontitis, and Dsb4 was retrieved from samples obtained by curette from a subject with periodontitis. Dsb5 and Dsv1 were retrieved from a mixture of paper point samples taken from four subjects with periodontitis. In all cases, samples were taken and immediately stored in 1× PBS at 4C and transported on ice to the lab within 24 hours. Samples were processed, fixed and stored as described in Chapter 2 before fluorescent *in situ* hybridization with group-specific probes and analysis on a Cytopeia Influx cell sorter.

3.2.3. Fluorescence *in situ* hybridization

For fluorescence *in situ* hybridization (FISH) we used the *Deltaproteobacteria*-specific probes DELTA495a [127] and SRB385 [86]. These are group-specific probes that have been shown to capture a large percentage of known *Desulfobulbus* and *Desulfovibrio* species [128]. Hybridization and wash buffers were prepared as previously described, and previous hybridization protocols were slightly adapted to work in-solution [90]. Hybridization buffer contained 35% formamide [127]. This mixture was hybridized at 46C for 3 hrs or overnight (~20 hrs). Based on control hybridizations using *D. propionicus* and *D. piger*, the signal was enhanced by overnight incubations and that condition was applied to the oral samples. Following centrifugation, the cells were washed using standard FISH solution and resuspended in PBS for cell sorting. Both Alexa488-labeled DELTA495a probe alone and Alexa488-labeled DELTA495a mixed with Cy5-dual-labeled SRB385 [129] probe combinations were used. When samples were hybridized with the combined probes, cells had both red and green fluorescence.

3.2.4. Cell sorting

Three Dsb cell isolates (Dsb1, Dsb2, Dsb3) were obtained by random sorting using a Cytopeia INFLUX sorter (BD, Franklin Lakes, NJ, USA). The influx cell sorter was cleaned prior to use in a similar manner to those described earlier [72, 91]. Ethanol-fixed oral samples from periodontitis patients were first passed through a CellTrics 30 µm disposable filter (Partec, Görlitz, Germany). Prior to sorting, samples were labeled by FISH or stained with 5 µM of nucleic acid-binding dyes SYTO 9 (green) and SYTO 62 (red) (Life Technologies, Grand

Island, NY USA) for 15 min. Cells labeled by FISH probes were sorted only once, as single cells. Samples labeled with nucleic acid dyes were sorted twice, using SYTO 9 emission for the first sort and SYTO 62 emission for the final sort, to further dilute any possible contaminating free DNA found in the original sample [91]. Cells sorted in the final round were deposited as single cells into 3 μ L UV-sterilized TE buffer.

3.2.5. Multiple displacement amplification

Single-cells retrieved by sorting were used for multiple displacement amplification (MDA). All plasticware, water and reagents used for MDA reactions (except the primers, dNTPs and enzyme) were UV treated as described [98]. Reactions were set up in a manner similar to [91] using the reagents and concentrations discussed in Chapter 2. Amplification was performed for 10 hrs at 30C followed by inactivation at 80C for 20 min and storage at -20C.

3.2.6. Target identification

A small aliquot of each single cell amplified genomic product (SAG) was diluted (1:150) in PCR-grade water (Ambion, Austin, TX, USA), and the remainder product was stored at -80C. Dilutions were used for SSU rRNA gene PCR amplifications using universal bacterial primer 27fm (5'-AGA GTT TGA TYM TGG CTC AG-3') [92] and *Deltaproteobacteria*-specific primer Delta495a (5'-AGT TAG CCG GTG CTT CCT-3') [127]. Fifty microliter reactions were setup and screened as described in Chapter 2. All confirmed *Deltaproteobacteria* products were used to generate secondary MDA products.

3.2.7. MDA purification and sequencing

Secondary *Deltaproteobacteria* SAGs were purified by phenol:chloroform: isoamyl alcohol extraction and alcohol precipitation before being sent for sequencing. Illumina Hi-Seq 100-bp paired-end libraries were constructed and sequenced at Hudson Alpha Institute for Biotechnology (Huntsville, AL, USA).

3.2.8. Genome assembly and annotation

SAG libraries Dsb2 and Dsb3 were submitted to the Joint Genome Institute's single-cell assembly and QC pipeline, similar to Illumina SAG libraries described previously [77]. Dsb1, Dsb4, Dsb5 and Dsv1 genomic data were normalized and assembled in-house using digital normalization [85] and Velvet assembly [105]. Assembled genomes were subjected to the JGI/ORNL automated annotation pipeline, which uses Prodigal for ORF calling [106], before submission into the Integrated Microbial Genomes (IMG) [109] for further analysis. Additionally, data was submitted into RAST [110] for analysis. The search for potentially contaminating DNA contigs (human and non-*Deltaproteobacteria*) was conducted using BLASTP [107] as well as by GC content and tetramer analysis [108] of each scaffold. Contigs that were aberrant based on all these analyses were removed from further analysis but represented a small fraction of the dataset.

3.2.9. Combined non-redundant assembly

To assess the phylogenetic relationships between the multiple Dsb SAGs, Dsb1-Dsb5 were compared using average nucleotide identity (ANI) and tetranucleotide frequency calculated using the software JSpecies [114]. In order to discuss the potential function of oral *Desulfobulbus* as a whole, all datasets were combined

(Dsb1-5). Repetitive regions between amplicons were removed iteratively with the aid of custom perl scripts. Repetitive regions were defined as described in Chapter 2.

3.2.10. Genome size estimation

Genome size and coverage was estimated using a set of 138 conserved single copy genes (CSCG) that has been determined from 1516 finished bacterial genome sequences in the IMG database [109]. Hidden Markov models of the identified Pfams [113] associated with each CSCG were used to search assemblies by means of the HMMER3 software [116]. The coverage was estimated as the ratio of found CSCG to total CSCGs, and the estimated complete genome size was calculated by division of the estimated genome coverage by the total assembly size.

3.3. Results and Discussion

3.3.1. Cell Isolation, sequencing and phylogenetic analysis

Single cells of both oral *Desulfobulbus* (n=7) and *Desulfovibrio* (n=5) were isolated from ethanol-fixed samples using a combination of fluorescence *in situ* hybridization (FISH) and flow cytometric cell sorting. FISH was performed using *Deltaproteobacteria*-specific probes [86, 127]. A culture of *Desulfobulbus propionicus* 1pr3 (DSM 2032) was used to test and optimize hybridization conditions (**Figure 3.1**). The genomic DNA of single cells was amplified using multiple displacement amplification (MDA) followed by taxonomic characterization of the single cell amplified genomes (SAGs) by SSU rRNA gene

amplification and sequencing. SAGs confirmed to represent target *Deltaproteobacteria* were sequenced individually (Dsb1, Dsb2, Dsb3) or in groups (Dsb4, Dsb5, Dsv1). The confirmed deltaproteobacterial SAGs originated from samples collected from both healthy individuals and individuals with periodontitis (**Table 3.1**).

Oral *Desulfobulbus* cells sequenced in this study were chosen for genomic sequencing based on SSU rRNA gene similarity to a previously known oral clone, *Desulfobulbus* sp. oral clone R004 [43], that has been associated with periodontitis [18]. Comparisons of full-length SSU rRNA genes revealed that amplicons Dsb1 – Dsb4 are identical, and these amplicons are over 99% identical to Dsb5 and *Desulfobulbus* sp. clone R004 (**Figure 3.2**). However, all the oral taxa form a distinct clade compared to *Desulfobulbus* from other environments.

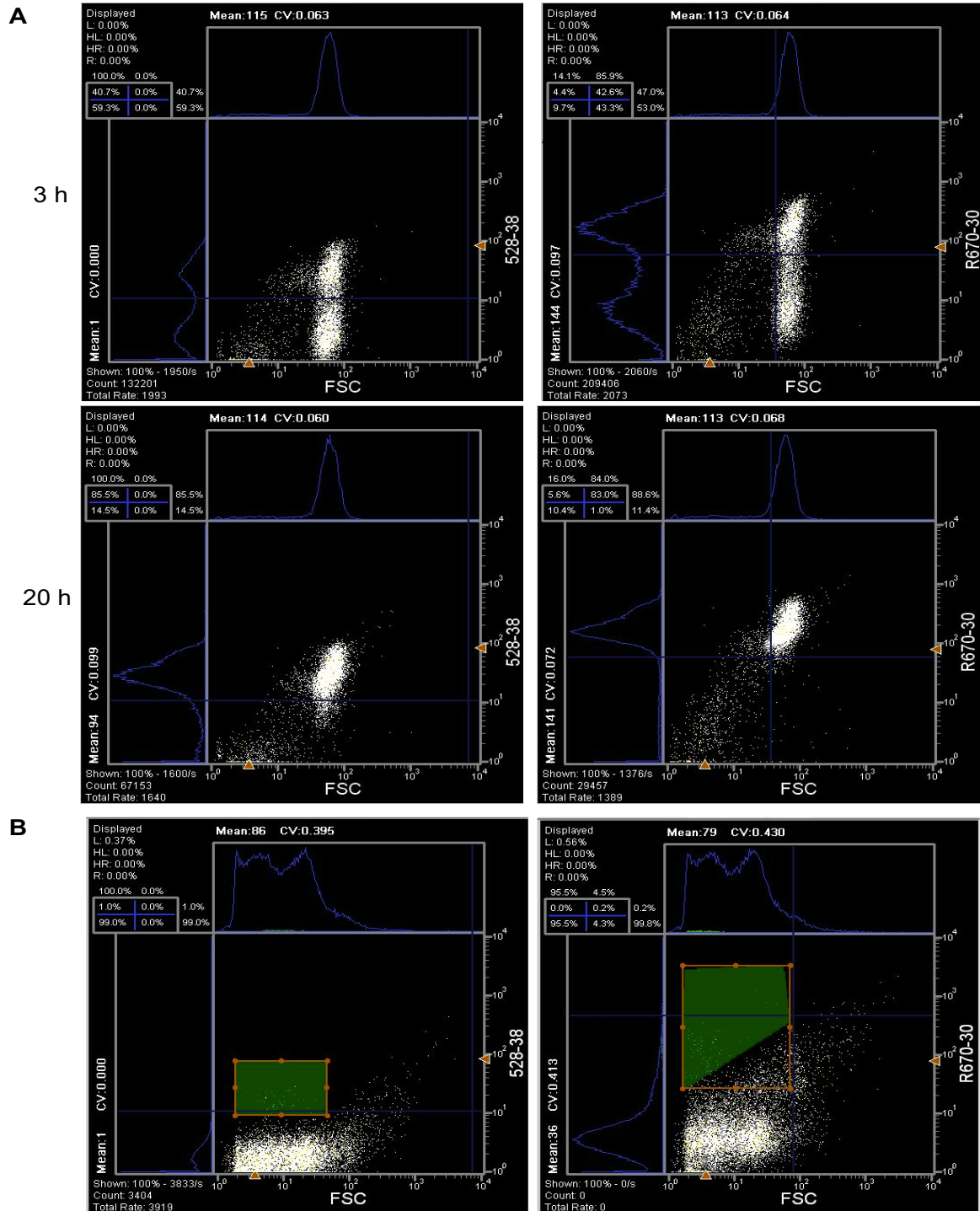


Figure 3.1. Optimization of sample hybridization times. Panel A shows the scatter pattern of a *D. propionicus* culture after 3 (top) or 20 (bottom) hours of hybridization. Panel B shows actual sorting gates for an oral sample after 20 hour hybridization. All scatter images of 528-38 emission (left) were samples hybridized with Alexa488-labeled DELTA495a. All scatter images of 670-30 emission (right) were samples hybridized with Cy5-dual labeled SRB385.

Table 3.1. Metadata and genome statistics for single and multi-cell oral amplicons.

	Genome Assembly						
	Dsb1	Dsb2	Dsb3	Dsb4	Dsb5	Dsb1-5	Dsv1
Number of cells sequenced	1	1	1	2	2	7	5
Donor health status	Healthy	Healthy	Periodontitis	Periodontitis	Periodontitis	Healthy+ Periodontitis	Periodontitis
Assembly size (bp)	455,123	768,341	798,161	748,829	959,378	1,883,075	2,603,557
DNA scaffolds	64	164	231	70	112	349	259
G+C (%)	56.6	57.6	57.2	59	59.7	58.6	59.9
Estimated genome size (Mbp)	1.39	2.09	2.12	3.75	2.26	2.48	2.63
Estimated genome recovery (%)	32.8	36.8	37.6	20.0	42.4	76.0	99.1
CRISPR	2	0	1	2	0	---	0
Genes total number	527	808	860	781	1007	1936	2890
5S rRNA	1	1	1	1	1	1	1
SSU rRNA	1	1	1	1	1	1	1
23S rRNA	1	1	1	1	0	1	1
tRNA genes	10	9	12	16	15	23	36
Other RNA genes	0	0	1	0	0	0	4
Protein coding genes	514	796	844	762	990	1910	2847

The SSU rRNA gene of Dsv1 is over 99% identical to *Desulfovibrio* sp. oral clone BB161 [43] and 98% identical to *Desulfovibrio fairfieldensis*, a cultured oral isolate [131]. In addition, the SSU rRNA gene of Dsv1 is also highly similar (>98%) to *Desulfovibrio* sp. 3_1_syn3 and *Desulfovibrio* sp. 6_1_46AFAA, two human-associated, sequenced species from the gut (**Figure 3.2**). When comparing the predicted proteins of Dsv1 to those of *D. sp. 3_1_syn3*, BLASTP [107] percent identity between the two organisms was 80% or greater for roughly half the proteins predicted in Dsv1. Only 13% of Dsv1 predicted proteins had the same level of similarity to human-associated *Desulfovibrio piger*, and 3% of genes were this closely related to environmental *Desulfovibrio vulgaris* Hildenborough (**Figure 3.3**).

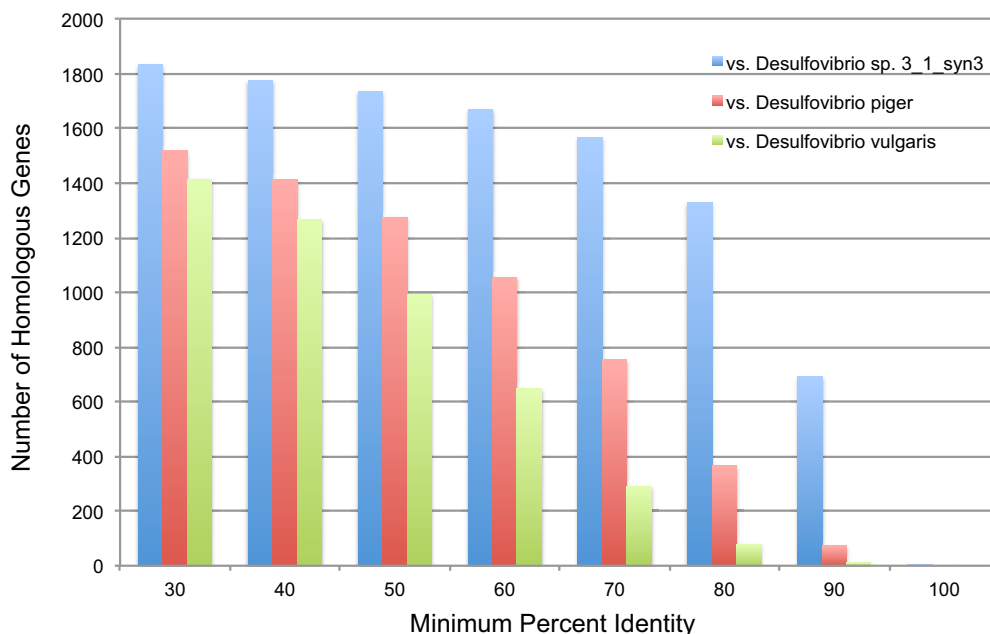


Figure 3.3. Number of Dsv1 genes with homologs in gut species *D. sp. 3_1_syn3*, *D. piger* or environmental species *D. vulgaris* Hildenborough. The search for homologous genes was performed in IMG [109] with increasing minimum percent identity requirements. A total of 2890 Dsv1 genes were tested.

3.3.2. Genome statistics

After normalization and assembly, the *Desulfobulbus* datasets had 0.45 – 0.96 Mbp each, with 527-1007 total genes (**Table 3.1**). Although Dsb4 and Dsb5 were a combination of two single amplified genomes (SAGs), only Dsb5 showed a noticeable increase in data (**Table 3.1**). Higher levels of contamination removal from Dsb4 most likely explain this discrepancy. Analyses comparing average nucleotide identity (ANI) of *Desulfobulbus* data showed that the amplicons were an average of 97.4% similar to one another (**Table 3.2**). Similarly, comparing tetranucleotide frequencies revealed a 98.7% average similarity between *Desulfobulbus* datasets Dsb1 through Dsb5 (**Table 3.3**). Overlapping regions between datasets were used for ANI and tetranucleotide analyses, and overlaps were between 14-25% for each comparison. Although the *Desulfobulbus* cells were isolated from both patients with periodontitis as well as healthy controls, there was no obvious separation between these groups. This similarity is expected since previous studies have shown the same *Desulfobulbus* species present in most adults but increasing in abundance during periodontitis [18, 122]. In light of these findings, we treated these cells as members of a single operational taxonomic unit (OTU), and all downstream metabolic analyses were performed on a combined assembly of *Desulfobulbus* data. This combined data set will be referred to as Dsb1-5 throughout. After removing redundant data, Dsb1-5 had a total of 1910 unique genes and 1.9 Mbp, with a GC content of 58.6%. Using a genome size estimation based on a conserved single copy gene

set, the genome of this oral *Desulfobulbus* OTU is approximately 2.48 Mbp. Thus, we have captured approximately 76% of the complete genome.

For the oral *Desulfovibrio*, five separate and taxonomically confirmed SAGs were combined before sequencing and will be referred to as Dsv1. The combination of genomic DNA from five SAGs provided a large increase of sequence data compared to the single and two-cell assemblies sequenced for the *Desulfobulbus* species. Dsv1 contained 2.6 Mbp of non-redundant data with a total of 2890 predicted genes and a GC content of 59.9%. A genome size estimate of 2.63 Mbp revealed that we were able to capture over 99% of the *Desulfovibrio* genome. Analysis with BLASTCLUST [107] revealed that the genomic dataset of these five cells, all sorted from the same sample, assembled well. Of the 2842 gene clusters formed at a similarity of 95% and an overlapping length of 0.7, 2839 were unique. The three overlapping clusters contained small, hypothetical genes, with at least one cluster having high similarity to integrase/transposase-like proteins.

Table 3.2. Similarity matrix of *Desulfobulbus* genomes based upon average nucleotide identities.

	Dsb1	Dsb2	Dsb3	Dsb4	Dsb5
Dsb1	---				
Dsb2	99.11	---			
Dsb3	98.00	97.79	---		
Dsb4	97.02	96.62	96.89	---	
Dsb5	95.86	96.03	97.60	97.72	---

Table 3.3. Similarity matrix of *Desulfobulbus* genomes based upon tetranucleotide frequencies.

	Dsb1	Dsb2	Dsb3	Dsb4	Dsb5
Dsb1	---				
Dsb2	0.990	---			
Dsb3	0.989	0.992	---		
Dsb4	0.983	0.989	0.987	---	
Dsb5	0.978	0.984	0.982	0.993	---

3.3.3. Metabolism

Although Dsb1-5 is human-associated, several aspects of its predicted metabolism appear to be similar to those found in the most closely related sequenced species, *D. propionicus*, an isolate from anaerobic mud. Forty four percent of Dsb1-5 predicted proteins had top BLASTP [107] hits to proteins in *D. propionicus*. *D. propionicus* has a complete pathway for glycolysis/gluconeogenesis, and most of these genes are also present in Dsb1-5. Similar genes are also found in the pentose phosphate pathway, although several genes were missing in our partial Dsb1-5 genome. The ability to respire sulfate appears to be present in Dsb1-5. The genes for dissimilatory sulfite reductase (*dsrAB*) are present, as well as dissimilatory adenylylsulfate reductase (*aprAB*). In addition, genes encoding QmoA and B and DsrJ, K, M, O and P [132] are present. A dissimilatory nitrate reductase gene is also present in *D. propionicus* and Dsb1-5, although Dsb1-5 had no evidence of the nitrogenase found in *D. propionicus*.

Despite the similarities, key differences can be seen in Dsb1-5 compared to some of the characteristics most associated with the *Desulfobulbus* genus. Several genes needed for propionate metabolism and the methylmalonyl

pathway are missing in Dsb1-5 (**Figure 3.4**). In particular, genes capable of converting propionyl phosphate to propionyl-CoA are present only in the *D. propionicus* genome. However, Dsb1-5 does code for formate C-acetyltransferase (E.C. 2.3.1.54), an enzyme not seen in *D. propionicus*, which would directly convert 2-oxobutyrate to propionyl-CoA.

Fermentation of pyruvate and alcohols has been found to proceed via the methylmalonyl-CoA pathway in *D. propionicus* [133]. Two of the three essential genes for the methylmalonyl-CoA fermentation pathway (EC 5.1.99.1 and EC 5.4.99.2) were not found in Dsb1-5, and pyruvate dehydrogenase and malate dehydrogenase, proteins that are active in *D. propionicus* [133], were also absent. Key enzymes of an alternative, acetyl-CoA pathway (formyltetrahydrofolate synthetase, carbon monoxide dehydrogenase and acetyl-CoA synthase) were also absent from Dsb1-5. Although this could be due to the incomplete nature of the Dsb1-5 assembly, it allows speculation that these capabilities may have been lost in the human-associated species. Additionally, these genes are not localized in *D. propionicus* but are found across the genome, making it even more unlikely that all such genes would be missed by our sequencing efforts. Dsb1-5 also lacks genes involved in butanoate metabolism and lactate utilization. In addition, *D. propionicus* encodes a partial TCA cycle that includes genes for 2-oxoglutarate:ferredoxin reductase, but Dsb1-5 has only three genes involved in the TCA cycle, leaving its use of this cycle unclear.

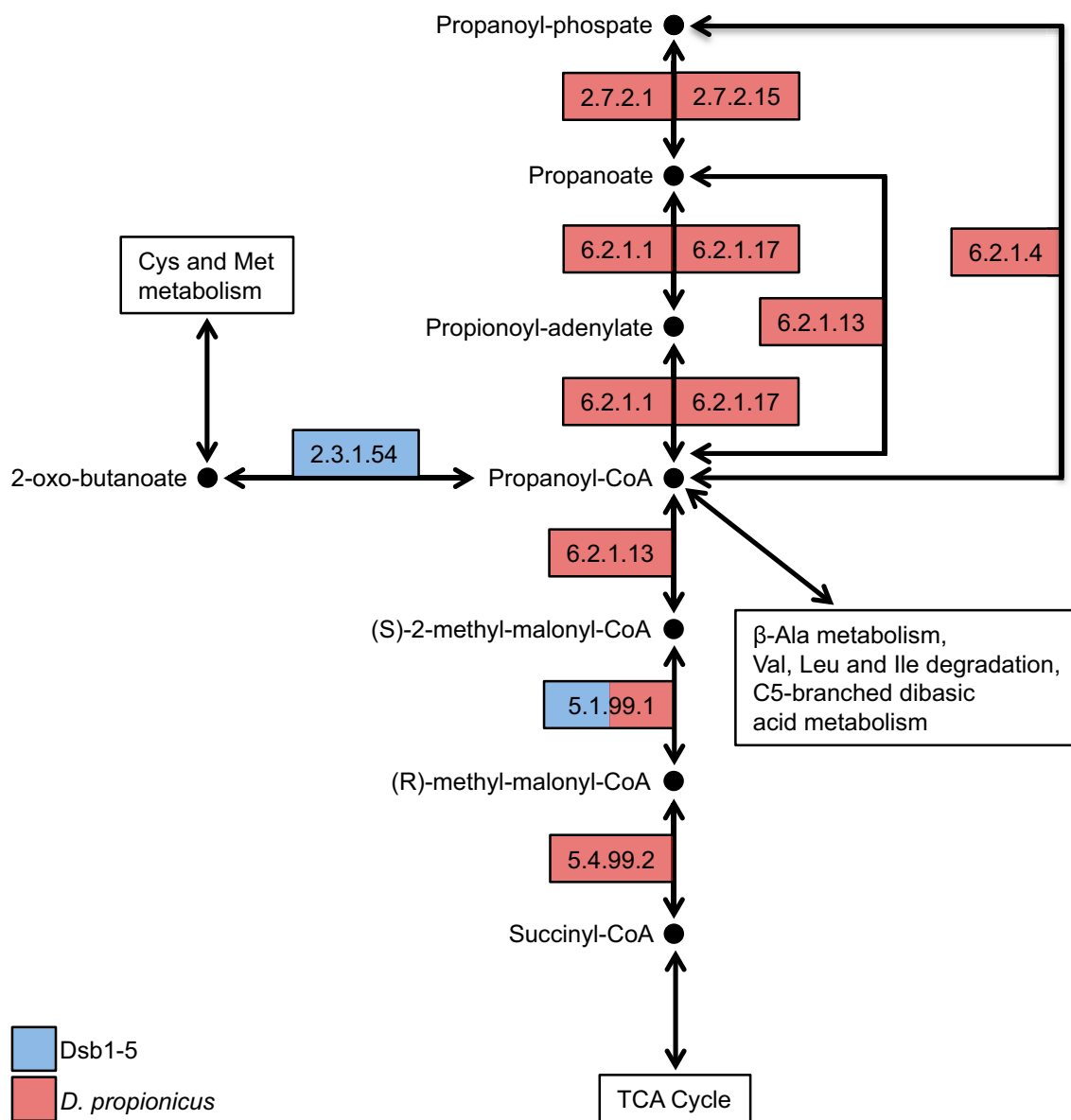


Figure 3.4. Map of enzymes used in propionate metabolism and the methylmalonyl pathway. Genes found in Dsb1-5 (based on E.C. number) are shown in blue. Genes in *D. propionicus* are shown in red.

Metabolic capabilities of Dsv1 seem to be similar to those found in other human-associated *Desulfovibrio* genomes (especially *Desulfovibrio* sp. 3_1_syn3), and these genomes also share many similarities to sequenced *Desulfovibrio* species from other environmental niches. Using BLASTP [107], 42% of Dsv1 predicted proteins had top hits to the *Desulfovibrio* sp. 3_1_syn3 genome. Approximately 32% of Dsv1 predicted proteins showed greatest similarity to other host-associated *Deltaproteobacteria* (*Desulfovibrio* sp. 6_1_46AFAA = 22%, *Desulfovibrio piger* = 2%, *Desulfovibrio desulfuricans* subsp. *desulfuricans* str. ATCC 27774 = 2%, *Bilophila wadsworthia* 3_1_6 = 4%, *Bilophila* sp. 4_1_30 = 2%).

All of the human-associated *Desulfovibrio* species, including Dsv1, contain genes for glycolysis/gluconeogenesis and key components of the pentose phosphate pathway, although the gene for glucose-6-phosphate isomerase is missing in Dsv1. Dsv1 also has the genes necessary for sulfate respiration, including *dsrAB*, *aprAB*, *dsrJKMOP* [132], *qmoABC* and *hmc* genes. All three human-associated *Desulfovibrio* genomes harbor the genes needed to reduce nitrate to ammonia that are also found in many environmental *Desulfovibrio* species. Human-associated *Desulfovibrio* all have an F-type ATPase, cytochrome *bd* complex and cytochrome *c* family III, much like environmental *Desulfovibrio* species. However, there is no evidence for a cytochrome *c* oxidase in the human-associated *Desulfovibrio* species. All host-associated *Desulfovibrio* species have a complete (*D. desulfuricans* ATCC 27774, *D. piger*, Dsv1) or

partial (*D. sp. 3_1_syn3*) transporter for methionine that is not seen in any sequenced, environmental *Desulfovibrio* species.

Dsv1 contains an RNF-type complex I similar to other *Desulfovibrio* species. In addition, Dsv1 appears to have a partial Nuo-type NADH dehydrogenase (subunits A-D, H-K) that is not present in either of the gut *Desulfovibrio* genomes, and all subunits have top matches to the NADH dehydrogenase genes from *D. propionicus*. The complex appears to lack the NADH binding and oxidizing subunits (NuoEFG) and may transfer electrons from ferredoxin to quinones of the respiratory chain. Similar complexes were found in Dsb1-5, *D. propionicus* and *D. desulfuricans desulfuricans* ATCC 27774, a bacterium isolated from a sheep rumen. Both Dsv1 and *D. sp. 3_1_syn3* have genes that encode fumarate reductase, an enzyme that functions in anaerobic respiration, but this appears to be absent in *D. piger*.

3.3.3.1. Chromosome partitioning and cell division

As expected, a gene encoding cell division protein FtsZ was found in Dsb1-5, along with other proteins involved in the cell division protein complex, including FtsA, FtsB, FtsE, FtsK, FtsQ and FtsX. Similar genes encoding the FtsZ system were also found in the Dsv1. In addition, Dsv1 also contained genes encoding the proteins MreB, MreC and RodA, proteins known to be involved in shape determination, particularly for elongated, rod-shaped bacteria.

3.3.3.2. Phage-related genes

Both Dsb1-5 and Dsv1 have large numbers of phage-related genes including Mu-like proteins. Both Dsv1 and Dsb1-5 harbor killer and antidote

proteins of a killer gene system. This system can be used for stress response, cell cycle arrest and maintenance of otherwise disposable genes, such as in an integron [134]. All of the toxin/antitoxin gene pairs found in Dsb1-5 and Dsv1 appear to be chromosomal and are found in close proximity to transposase, integrase and other phage-related genes.

3.3.3.3. Replication and Repair

Genes functioning in replication and repair were found in Dsb1-5 and Dsv1, including genes involved in DNA replication, homologous recombination and base excision, mismatch and nucleotide excision repair. The top BLASTP hits for Dsb1-5 were variable, with a wide range of *Deltaproteobacteria* and lower overall percent identity. Most proteins in Dsv1 had high BLASTP percent identity to proteins predicted in two HMP *Desulfovibrio* sp. that have been sequenced from the human gut (*D. sp. 3_1_syn3* and *D. sp. 6_1_46AFAA*). Thus, replication and repair machinery in the oral Dsb1-5 are further removed from any of the currently sequenced *Deltaproteobacteria* than that of Dsv1.

3.3.3.4. Motility

Cell mobility is an important factor of chemotaxis, biofilm formation and oral pathogenicity, and the oral *Deltaproteobacteria* characterized here have the potential for several types of mobility. Dsb1-5 contained proteins for chemotaxis, including methyl-accepting chemotaxis proteins for detection of chemical gradients and chemotactic proteins CheA, CheW and CheY. Dsv1 contained all the chemotaxis proteins found in Dsb, as well as CheB, CheV and CheZ. These

chemotactic proteins are involved in a signal transduction system that could control flagellar swimming motility and/or pilin twitching motility.

Evidence for flagellar biosynthetic and motor proteins was present in Dsb1-5 and Dsv1 and included FliG, FliL, FliM, FliN, FliP, FliQ, and motor proteins MotA and MotB. However, only Dsv1 had evidence of flagellar proteins necessary for the formation of the flagellar distal and proximal rods, hook and filament. Although *D. piger* is nonmotile, both Dsv1 and *D. sp. 3_1_syn3* have all the genes necessary for flagellar motility. *D. propionicus* strains 2pr4 and 3pr10 produce a single polar flagella and are motile while strain 1pr3 was nonmotile without a visible flagellum [135], despite the presence of several flagellar encoding genes. Based on the absence of several flagellar proteins, it is possible that Dsb1-5 has lost the ability to produce flagella, much like *D. propionicus* 1pr3.

In addition to an incomplete set of flagellar proteins, Dsb1-5 also includes a variety of genes for pilin proteins. These genes code for proteins PilA, PilB, PilF, PilO, PilT, PilV, PilW, PilX, and PilY1. The presence of PilT, a protein involved in pilus retraction and disassembly, suggests that Dsb1-5 forms type IV pili that can be used for twitching motility and biofilm formation [136, 137]. This is not surprising as twitching motility is found in organisms within Proteobacteria divisions Delta, Beta and Gamma [136]. *D. propionicus* strain 1pr3 also produces pili [135], and many of the pili genes in Dsb1-5 have top BLASTP hits to *D. propionicus*.

3.3.3.5. Secretion

Several of the genes found in Dsb1-5 for twitching motility also overlap with genes used in type II secretion systems. The presence of additional genes PulD, PulE, PulF and PulO suggest that Dsb1-5 also contain a type II secretion system. Genes that code for both the Sec-dependent protein export as well as twin arginine targeting protein export are present in both Dsb1-5 and Dsv1. A type IV secretion system is also present in both organisms. The type IV proteins match pfam categories that include both virulence and conjugal transfer proteins. The presence of a type IV secretion system potentially allows Dsb1-5 to adapt to environmental changes during. It is also possible that conjugative pili could be working to aid colonization and biofilm formation [138]. Both Dsb1-5 and Dsv1 contained a gene encoding a putative homolog of TadD, a Flp pilus assembly protein important for tight, nonspecific adherence of certain bacteria to surfaces [139]. The *tad* locus has been studied in *Aggregatibacter actinomycetemcomitans* and found to be essential for the ability to colonize tooth surfaces [139]. In addition, *tadD* was found to be an important virulence factor in the pathogens *Pasteurella multocida* and *Yersinia ruckeri* [140, 141]. The presence of a membrane fusion protein similar to HlyD points to type I secretion capabilities in both Dsb1-5 and Dsv1.

3.3.3.6. Defense Mechanisms

Several genes in Dsb1-5 are involved in self-defense. It appears that oral *Desulfobulbus* species have both type I and II restriction modification systems in place to protect against bacteriophage. Dsv1 shows evidence of type I and III

restriction modification systems. CRISPR regions were found in Dsb1-5, and CRISPR-associated genes were found in both Dsb1-5 and Dsv1. Dsb1-5 and Dsv1 also contain genes that provide a wide range of antibiotic and multi-drug resistance, including both primary and secondary active transporters. Both Multi-Antimicrobial Extrusion (MATE) and Resistance nodulation cell division (RND) families of secondary active transporters were found in Dsb1-5 and Dsv1. One multi-drug efflux pump found in both oral genera had high similarity (0.0e+00 E-value) with the AcrB/D/F family (pfam00873), which is known to have a wide range of substrate specificity and exports most antibiotics in use [142]. In addition, several multidrug and antimicrobial ATP-binding cassette (ABC) transporters are encoded in both Dsb1-5 and Dsv1. Dsb1-5 contains genes for bacteriocin/lantibiotic transporters. Also, oral Dsb1-5 carries class C beta-lactamase genes; thus, it is likely resistant to penicillin and other beta-lactam-containing antibiotics. Finally, Dsb1-5 included an undecaprenyl-diphosphatase gene, which confers resistance to bacitracin.

3.3.3.7. Oxygen-Tolerance

Although originally thought to be obligate anaerobes, it is now known that many sulfate-reducing bacteria are oxygen tolerant. In fact, both *Desulfobulbus* and *Desulfovibrio* species have been shown to perform aerobic respiration with several electron donors, including sulfur compounds. However, these bacteria were not able to grow with oxygen as the electron acceptor [143], but rather, they seem to use aerobic respiration as a defense mechanism to avoid production of toxic products such as hydrogen peroxide and superoxide radicals [144]. Genes

important for oxygen tolerance were found in both Dsb1-5 and Dsv1, suggesting that this trait is also important in the oral environment. A catalase gene was found in Dsb1-5, and this enzyme is able to detoxify certain oxygen species [144]. In addition, rubredoxin and desulfoferrodoxin genes were found in both Dsb1-5 and Dsv1, and a gene encoding rubrerythrin was found in Dsb1-5. Homologous genes in *Desulfarculus baarsii* were found to protect against damage caused by reactive oxygen species [145]. The ability to deal with oxygen stress efficiently would be an important trait for successful pathogens.

3.3.4. Pathogen-associated genes

Both Dsb1-5 and Dsv1 carry genes that are beneficial for a pathogen lifestyle. Some of these genes are in common with *Desulfovibrio* species found in the environment, but several genes seem to be uniquely associated with a host-associated niche. Pathogen-associated genes provide a microbe with a competitive advantage in a host environment. Putative virulence factors for oral pathogens include genes for adherence, defense, uptake of limited nutrients, stress adaptation and interactions with both host cells and other microbial cells found within a biofilm. Categories of putative virulence factors that were found in Dsb1-5 and Dsv1 are listed in **Table 3.4** and discussed below.

Table 3.4. Categories of putative virulence factors found in Dsb1-5 and Dsv1. The presence of a gene is noted by (+).

Category	Annotation	Dsb1-5	Dsv1
Acquisition of iron	FeoB ferrous iron uptake	+	+
	ABC-type Fe ³⁺ -siderophore transport system	+	+
	Hemolysin	+	+
	ferrous iron transport protein A	+	+
	Fur family ferric uptake regulator	+	+
	TonB-dependent receptor	+	+
Secretion	Sec system	+	+
	Type I secretion	+	+
	Type II secretion	+	
	Type IV secretion	+	+
Stress Response	Universal stress protein UspA	+	+
	Clp proteins	+	+
	RelA/SpoT family protein	+	+
	5'/3'-nucleotidase SurE	+	+
	Crp/Fnr family transcriptional regulator	+	+
	desulfoferrodoxin	+	+
	rubredoxin	+	+
	Ruberrythrin	+	
	Catalase	+	
	Peroxiredoxin		+
Evasion	2-dehydro-3-deoxyphosphooctonate aldolase	+	
	Acetyltransferase	+	+
	alginate o-acetyltransferase	+	
	Beta-glucosidase-related glycosidases		+
	CMP-2-keto-3-deoxyoctulosonic acid synthetase		+
	dTDP-4-dehydrorhamnose 3,5-epimerase		+
	dTDP-4-dehydrorhamnose reductase	+	
	dTDP-glucose 4,6-dehydratase		+
	glucose 1-phosphate thymidyltransferase	+	
	glycoside hydrolase family protein		
	Glycosyltransferase	+	+
	lytic transglycosylase	+	
	membrane-bound lytic murein transglycosylase		+
	N-acetylmuramoyl-L-alanine amidase	+	+
	NAD-dependent epimerase/dehydratase	+	
	nucleotide sugar dehydrogenase	+	+
	pantetheine-phosphate adenylyltransferase, bacterial		+
	peptidoglycan O-acetyltransferase PacA	+	
	GDP:alpha-D-mannose-1-phosphate guanylyltransferase	+	
	putative exopolysaccharide biosynthesis protein		+

Table 3.4. (continued)

Category	Annotation	Dsb1-5	Dsv1
	UDP-N-acetylglucosamine 2-epimerase	+	
	undecaprenyl-phosphate galactose phosphotransferase	+	
Defense mechanism	ABC-type bacteriocin/lantibiotic exporter	+	
	ABC-type multidrug transport system	+	+
	acriflavin resistance protein	+	+
	choline/carnitine/betaine transporter family	+	
	Cation/multidrug efflux pump	+	+
Protease/Peptidase	Clp protease	+	+
	La protease	+	+
	carboxyl-terminal protease	+	
	Collagenase and related proteases		+
	FtsH protease	+	
	Membrane proteins related to metalloendopeptidases		+
	O-sialoglycoprotein endopeptidase	+	
	signal peptide peptidase SppA	+	+
	Trypsin-like serine proteases, typically periplasmic		+
	Xaa-Pro aminopeptidase		+
	putative glycoprotease GCP		+
	periplasmic serine protease, Do/DeqQ family		+
	outer membrane protease	+	
Adhesion	Type IV pili	+	
	TPR repeat-containing protein	+	+
	YD repeat-containing protein	+	
	Surface antigens	+	+
	Lipoproteins	+	+

Iron is an essential element used in several metabolic processes; however, bio-available forms of iron are often limited in the environment [146]. The ability to acquire iron is a particularly important trait for human-associated bacteria, where iron can be even more limited by iron-binding proteins of host cells. In the oral environment, most iron is bound to the host protein lactoferrin [146]. Thus, successful human pathogens often have several genes that give them an advantage in acquiring iron. In addition to several iron transport systems, both Dsb1-5 and Dsv1 also contain hemolysin. Using IMG, these genes can also found in many environmental *Deltaproteobacteria* genomes and have

likely persisted in Dsb1-5 and Dsv1 because of their continued benefits in the host environment.

In order to be successful pathogens, bacteria must cope with environmental changes and harmful molecules. Both Dsb1-5 and Dsv1 encode a wide range of transporters and efflux systems to protect against drugs and antibiotics. In addition, Dsb1-5 and Dsv1 contain oxygen-tolerance genes discussed above, universal stress protein UspA and a suite of Clp proteins. The Clp proteins in *Porphyromonas gingivalis* have been shown to be important for stress response, as well as biofilm formation and entry into host epithelial cells [147]. Knockouts of ClpB resulted in reduced virulence in both *Leptospira interrogans* and *Enterococcus faecalis* and reduced general stress resistance in *L. interrogans* [148, 149]. Further, ClpAP mutants in *Helicobacter pylori* showed increased sensitivity to antibiotics and oxidative stress and disrupted colonization of macrophages [150].

Another important aspect of pathogenicity includes the ability to attach to other cells. The type IV pili found in Dsb1-5 are likely important for its adhesion [137, 151]. Additional surface proteins that may play a role in binding to cells include tetratrico peptide repeat (TPR) and YD dipeptide repeat-containing domains, surface antigens and lipoproteins [62]. Additionally, both Dsb1-5 and Dsv1 contain several proteases and peptidases that often serve as virulence factors that degrade host proteins for nutrients and may contribute to the cytotoxicity of periodontitis [62, 152].

Lastly, both genomes had several genes for polysaccharide metabolism, including an abundance of glycosyl transferases [62, 152]. Both genomes had a large number of genes involved in lipopolysaccharide biosynthesis, and there were also genes present in Dsb1-5 (UDP-N-acetylglucosamine 2-epimerase and polysaccharide biosynthesis protein CapD) that are putatively involved in extracellular polysaccharide (EPS) or capsular production. At least two scaffolds in Dsb1-5 include genes that may be involved in EPS or capsular production, and the top BLASTP hits for most genes on these scaffolds suggest these genes could have been acquired through horizontal gene transfer (**Table 3.5**). The presence of a capsule would be beneficial to a pathogen in the oral environment by providing adherence to oral surfaces and other biofilm members and resistance to both specific and nonspecific host immune systems [153].

Table 3.5. Two Dsb1-5 contigs that contain large numbers of polysaccharide metabolism genes. Information for the top BLASTP hit (based on e-value) for each predicted gene is shown.

Dsb1-5 Contig_Gene Number	Information for top BLASTP hit			
	E-value	Accession	Description	Organism
Dsb1-5_Contig22_1	1.25E-09	ZP_10104881	glycosyl transferase family 2	Thiothrix nivea DSM 5205
Dsb1-5_Contig22_2	2.94E-66	ZP_08821249	glycosyl transferase family 2	Thiorhodococcus drewsii AZ1
Dsb1-5_Contig22_3	0.00748	EGH05187	4-oxalocrotonate tautomerase	Pseudomonas syringae pv. aesculi str. 0893_23
Dsb1-5_Contig22_4	7.16E-44	ZP_07333891	conserved hypothetical protein	Desulfovibrio fructosovorans JJ
Dsb1-5_Contig22_5	3.55E-08	ZP_06298373	hypothetical protein pah_c004o242	Parachlamydia acanthamoebae str. Hall's coccus
Dsb1-5_Contig22_6	6.83E-20	YP_001230579	Ig family protein	Geobacter uraniireducens Rf4
Dsb1-5_Contig22_7	1.05E-53	YP_003809026	acetyltransferase	Desulfarculus baarsii DSM 2075
Dsb1-5_Contig22_8	9.88E-24	YP_003206524	membrane protein	Candidatus Methyloirabilis oxyfera
Dsb1-5_Contig22_9	1.31E-71	YP_003809023	hypothetical protein Deba_3076	Desulfarculus baarsii DSM 2075
Dsb1-5_Contig22_10	5.19E-140	YP_003806580	family 2 glycosyl transferase	Desulfarculus baarsii DSM 2075
Dsb1-5_Contig22_11	1.71E-69	YP_645837	glycosyl transferase family protein	Rubrobacter xylanophilus DSM 9941
Dsb1-5_Contig5_1	4.11E-97	YP_002354782	acetyl transferase protein	Thauera sp. MZ1T
Dsb1-5_Contig5_2	1.40E-95	ZP_10662483	glycosyl transferase (involved in LPS synthesis)	Pseudomonas sp. GM48
Dsb1-5_Contig5_3	2.84E-158	YP_728196	glycosyl transferase	Ralstonia eutropha H16
Dsb1-5_Contig5_4	1.85E-158	ZP_08931531	UDP-N-acetylglucosamine 2-epimerase	Thioalkalivibrio thiocyanoxidans ARh 4
Dsb1-5_Contig5_5	1.29E-113	YP_728198	hypothetical protein H16_B0027	Ralstonia eutropha H16
Dsb1-5_Contig5_6	1.75E-139	ZP_09507281	glycosyltransferase	Marinobacterium stanieri S30
Dsb1-5_Contig5_7	2.23E-168	YP_003506150	asparagine synthase	Meiothermus ruber DSM 1279
Dsb1-5_Contig5_8	4.25E-44	ZP_04414198	glycosyltransferase	Vibrio cholerae bv. albensis VL426
Dsb1-5_Contig5_9	2.15E-08	ZP_08271641	O-antigen polymerase precursor	gamma proteobacterium IMCC3088
Dsb1-5_Contig5_10	0	YP_983341	LPS biosynthesis protein WbpG	Polaromonas naphthalenivorans CJ2
Dsb1-5_Contig5_11	1.54E-82	YP_983344	polysaccharide biosynthesis protein	Polaromonas naphthalenivorans CJ2
Dsb1-5_Contig5_12	0	YP_001339692	DegT/DnrJ/EryC1/StrS aminotransferase	Marinomonas sp. MWYL1

Table 3.5. (continued)

Dsb1-5 Contig _Gene Number	Information for top BLASTP hit			
	E-value	Accession	Description	Organism
Dsb1-5_Contig5_13	2.56E-107	ZP_09911538	hexapaptide repeat-containing transferase	SAR324 cluster bacterium JCVI-SC AAA005
Dsb1-5_Contig5_14	0	ZP_08931518	oxidoreductase domain protein	Thioalkalivibrio thiocyanoxidans ARh 4
Dsb1-5_Contig5_15	0	YP_847464	UDP-glucose/GDP-mannose dehydrogenase	Syntrophobacter fumaroxidans MPOB
Dsb1-5_Contig5_16	0	YP_004194024	polysaccharide biosynthesis protein CapD	Desulfobulbus propionicus DSM 2032
Dsb1-5_Contig5_17	6.96E-14	CBL39994	hypothetical protein CK3_01200	butyrate-producing bacterium SS3/4
Dsb1-5_Contig5_18	0.0005242	ZP_08576528	Gp54 protein	Lactobacillus farciminis KCTC 3681

3.3.5. COG comparisons to other *Deltaproteobacteria* genomes

To further understand the unique properties of the oral *Deltaproteobacteria* characterized here, comparisons were done with other sequenced *Deltaproteobacteria* using the presence/absence of COGs. Overall percentages of genes in each COG category for Dsb1-5 and Dsv1 were similar to those seen in other sequenced *Deltaproteobacteria*. Comparisons of Dsb1-5 to *D. propionicus* closely matched initial comparisons made with BLASTP hits and revealed a large overlap of genes. Dsb1-5 shares 920 COGS with *D. propionicus*, whereas only 222 COGs (19%) were unique to the oral genomes (**Figure 3.5A**). A complete Dsb1-5 genome would likely show an even greater overlap with *D. propionicus*. Several of the unique COGs in Dsb1-5 were for related functions. Nine COGS included components of a type IV secretion system found in Dsb1-5, whereas only components of the type II secretion system were found in *D. propionicus*. Dsb1-5 also encodes an array of unique COGS for transporters, including those that transport dipeptides/oligopeptides, bacteriocins/lantibiotics, monosaccharides, betaine and cobalt. Dsb1-5 contained COGs for both Na⁺/alanine and Na⁺/H⁺-dicarboxylate symporters, while *D. propionicus* contains a unique multi-subunit, Na⁺/H⁺ antiporter. Many of the Dsb1-5 specific COGs contain proteins discussed above that may be used for host evasion and possible capsular production. Other differences include unique COGs in each genome for phage-related proteins and outer membrane proteins.

COG categories were also used to compare the oral Dsv1 genome to other available human-associated species, *D. sp. 3_1_syn3* and *D. piger* (**Figure**

3.5B). Again, the COGs were similar to BLASTP results and largely overlapped, with 947 COGs found in all three species. An additional 130 COGs overlapped in both Dsv1 and *D. sp. 3_1_syn3*. COGs that were unique to Dsv1 included six NADH dehydrogenase subunits, a toxin/antitoxin growth regulation system, an ABC-type sulfate/molybdate transporter and an organic solvent tolerance protein. The large number of overlapping COGs found in all species included many metabolic genes, genes for cell division, DNA replication and repair and a large number of ABC-type transporters and Na⁺ symporters and antiporters.

A broader comparison of genomes within the *Desulfovibrio* genus found COGs that are specific to either host-associated or environmental organisms (**Table 3.6**). Unfortunately, genes for four of the differential, human-associated COGs were predicted to encode uncharacterized proteins, leaving the function of these proteins unclear. Both pfam and top BLASTP hits for the uncharacterized COG proteins shed light on their potential function. The genes that fall into COG3735 also fall into pfam PF01963, the TraB family of proteins. In *Enterococcus faecalis*, TraB is an important regulator for the peptide pheromone cAD1 that is encoded on a hemolysin/bacteriocin plasmid and used to control both clumping and plasmid transfer between cells containing plasmids and those that do not [154]. COG2326-associated proteins are likely similar to polyphosphate kinase 2, an enzyme converts AMP to ADP. Another uncharacterized COG contained a protein that is found, in every host-associated genome analyzed, between a tripartite ATP-independent periplasmic (TRAP) transporter solute receptor protein and a TRAP fused permease component and

may be involved in this secondary transport system. The final human-specific COG was a putative amino acid racemase.

In addition to these differential genes, eight additional genes fell into COG categories that were present in both human-associated and sheep rumen-associated *Desulfovibrio*, but absent in environmental *Desulfovibrio*. Interesting genes in this group include hits to COG2011 and COG1464. Further investigation of pfam and KO modules revealed these two proteins are likely used as a D-methionine transport system, and BLASTP results show that the closest hits outside of the host-associated *Desulfovibrio* are to clostridia. An additional COG (1897) that includes homoserine O-succinyltransferase, the first step in the biosynthesis of methionine, was found in all host-associated *Deltaproteobacteria*, except the incomplete, Dsb1-5 genome. Both COG0214 and COG0311 contain proteins that are part of the pyridoxine (vitamin B₆) biosynthesis pathway.

Twenty-five COGs were detected only in the environmental *Desulfovibrio* species, eight of which are uncharacterized proteins. Two COGs encoded proteins similar to FtsE and FtsX, proteins involved in cell division and important under low osmotic conditions [155]. Additional differential COGs include a distinct endocuclease III and IV and Fe-S oxidoreductase.

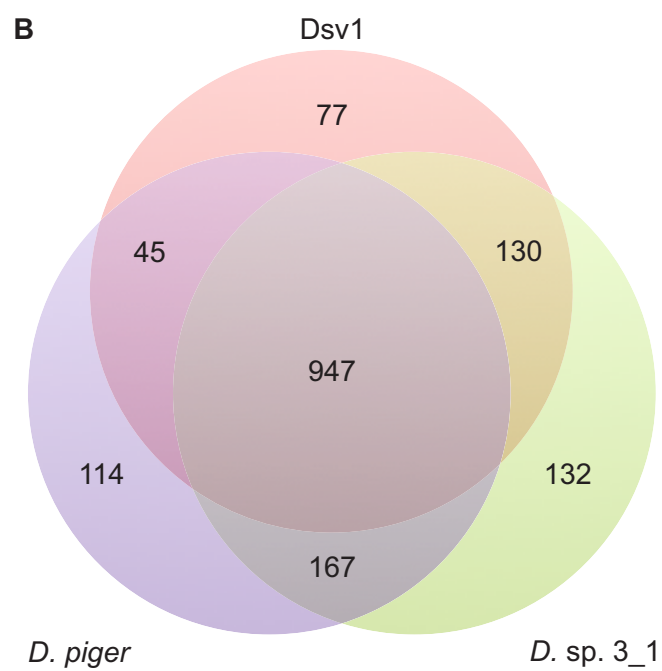
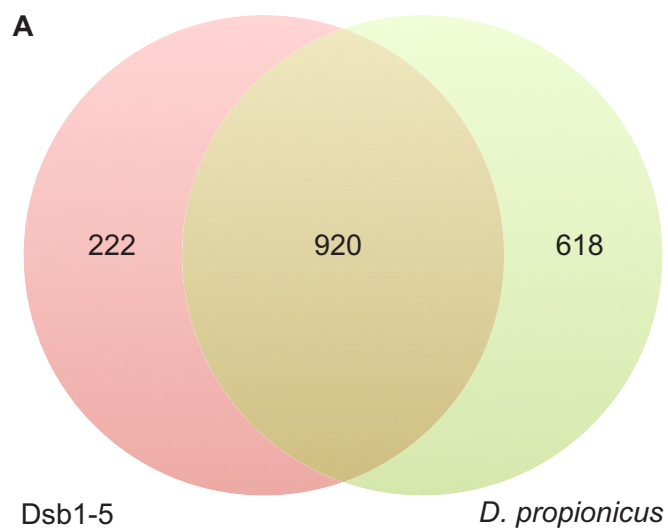


Figure 3.5. Comparison of COG categories (presence/absence) in (A) Dsb1-5 and *D. propionicus* and in (B) human-associated *Desulfovibrio* Dsv1, *D. sp. 3_1_syn3* and *D. piger*.

Table 3.6. COGs unique to either human-associated or environmental *Desulfovibrio* genomes.

COGS unique to host-associated <i>Desulfovibrio</i>	
COG0122	3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase
COG1464	ABC-type metal ion transport system, periplasmic component/surface antigen
COG2011	ABC-type metal ion transport system, permease component
COG2759	Formyltetrahydrofolate synthetase
COG3616	Predicted amino acid aldolase or racemase
COG0311	Predicted glutamine amidotransferase involved in pyridoxine biosynthesis
COG1058	Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA
COG5496	Predicted thioesterase
COG0214	Pyridoxine biosynthesis enzyme
COG2326	Uncharacterized conserved protein
COG4729	Uncharacterized conserved protein
COG2315	Uncharacterized protein conserved in bacteria
COG3735	Uncharacterized protein conserved in bacteria
COGS unique to environmental <i>Desulfovibrio</i>	
COG1042	Acyl-CoA synthetase (NDP forming)
COG1254	Acylphosphatases
COG2177	Cell division protein
COG0648	Endonuclease IV
COG0731	Fe-S oxidoreductases
COG0174	Glutamine synthetase
COG1443	Isopentenyl diphosphate isomerase
COG0346	Lactoylglutathione lyase and related lyases
COG1762	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
COG1075	Predicted acetyltransferases and hydrolases with the alpha/beta hydrolase fold
COG2884	Predicted ATPase involved in cell division
COG0702	Predicted nucleoside-diphosphate-sugar epimerases

Table 3.6. (continued)

COGS unique to environmental <i>Desulfovibrio</i>	
COG3635	Predicted phosphoglycerate mutase, AP superfamily
COG5001	Predicted signal transduction protein with membrane domain, an EAL and a GGDEF domain
COG1995	Pyridoxal phosphate biosynthesis protein
COG1663	Tetraacyldisaccharide-1-P 4'-kinase
COG0176	Transaldolase
COG1690	Uncharacterized conserved protein
COG1912	Uncharacterized conserved protein
COG0011	Uncharacterized conserved protein
COG0398	Uncharacterized conserved protein
COG1900	Uncharacterized conserved protein
COG2122	Uncharacterized conserved protein
COG2509	Uncharacterized FAD-dependent dehydrogenases
COG2231	Uncharacterized protein related to Endonuclease III

A comparison of all available *Deltaproteobacteria* genomes in IMG ($n = 53$) revealed two COG categories (COG4457, COG4458) that were found in all human-associated *Deltaproteobacteria* as well as *D. desulfuricans* ATCC 27774 but were not represented in any of the environmental *Deltaproteobacteria* genomes. These COGs included *srfB* and *srfC*, two genes of the SrfABC operon originally found in *Salmonella enterica* [156]. Although initially thought to be activated by SsrB, the transcriptional activator of a two-part regulatory system that controls gene activity of *Salmonella* pathogenicity island 2 (SPI2) [156], more recent work did not support this finding [157]. Additional studies in *Salmonella* have linked regulation of this operon to genes responsible for flagella production [158] and oxygen sensing [159]. Ultimately, the function of these genes is still unclear. BLASTP was used within IMG to find closely related predicted proteins in other genomes. When top hits were aligned and a phylogenetic tree was made, the most closely related proteins were found in a wide range of plant and animal-associated microbes, including both commensal and pathogenic bacteria (**Figures 3.6, 3.7**). Similar matches were found for a neighboring gene that encodes a putative *Haemophilus* adhesion and penetration (Hap)-like protein that is important for biofilm formation and adherence and entry into epithelial cell [160].

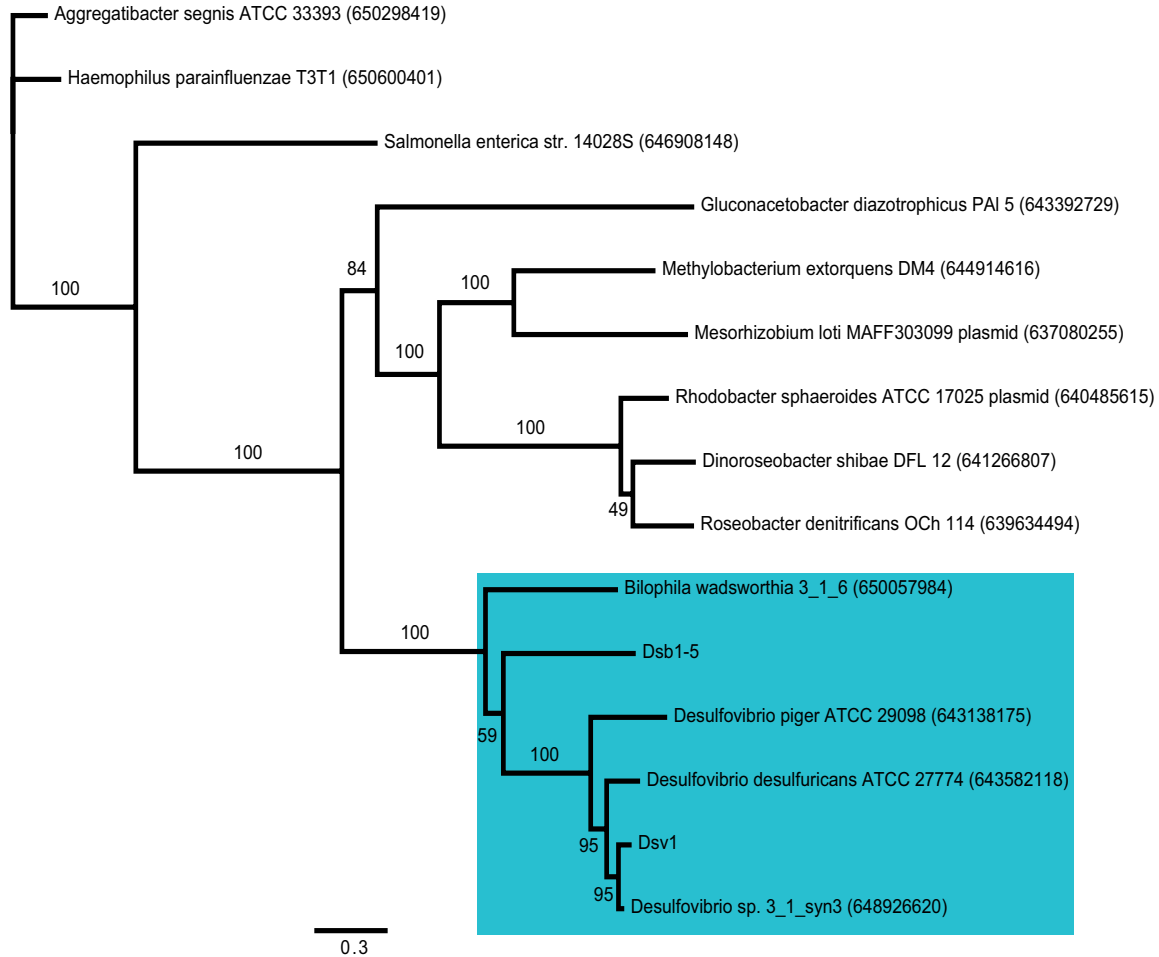


Figure 3.6. Maximum likelihood tree of putative *srfB* genes. The tree was constructed using PHYML [130] in the program Geneious® Pro 5.6.5 with a JTT (+ gamma + invariant sites) substitution model. Predicted proteins from host-associated *Deltaproteobacteria* are denoted by a blue box. The scale bar indicates 0.3 substitutions per nucleotide position. Numbers given at the nodes represent bootstrap percentages calculated on 100 replicates.

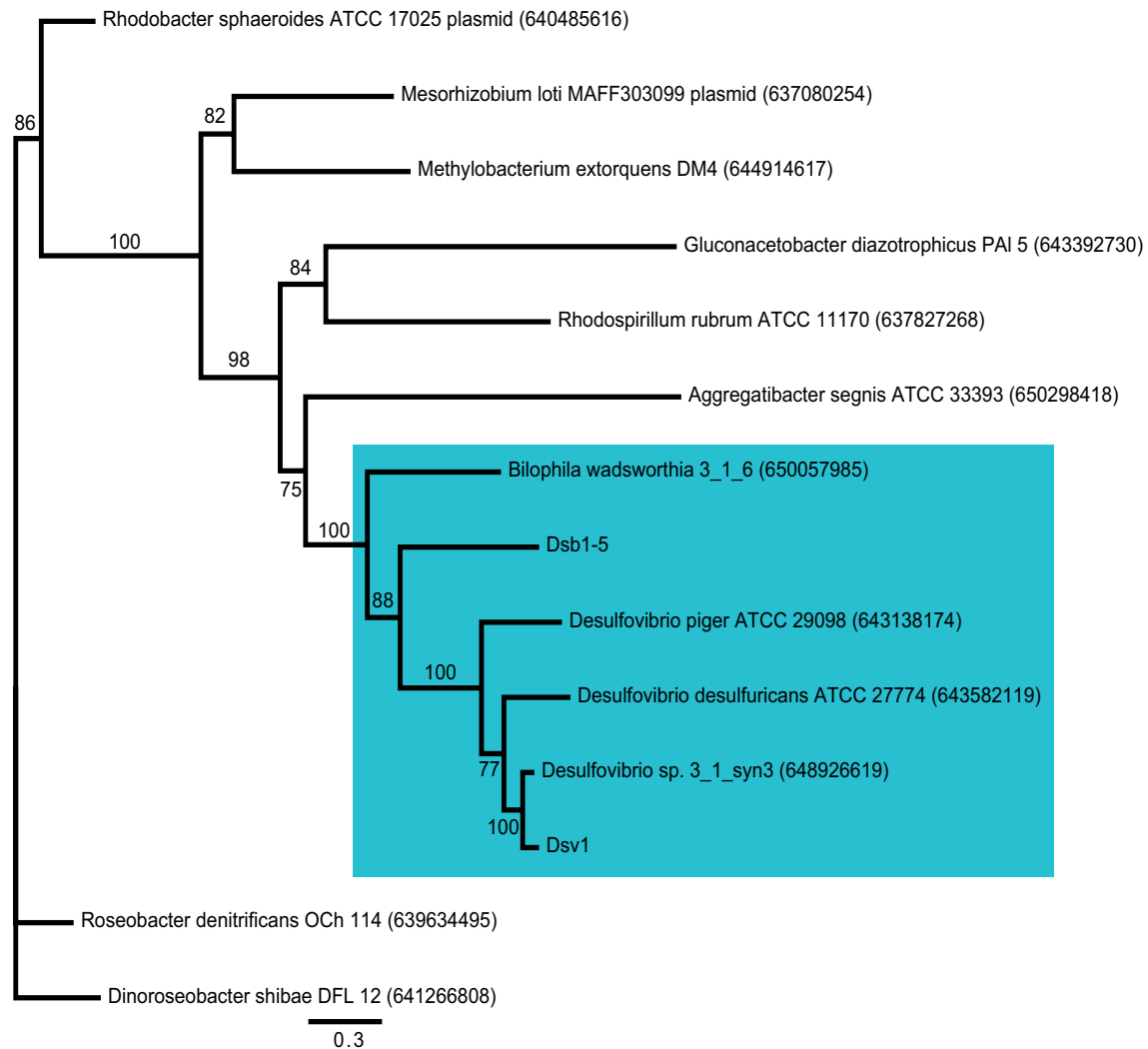


Figure 3.7. Maximum likelihood tree of putative *srfC* genes. The tree was constructed using PHYML [130] in the program Geneious® Pro 5.6.5 with a JTT (+ gamma + invariant sites) substitution model. Predicted proteins from host-associated *Deltaproteobacteria* are denoted by a blue box. The scale bar indicates 0.8 substitutions per nucleotide position. Numbers given at the nodes represent bootstrap percentages calculated on 100 replicates.

In addition to putative SrfABC and Hap encoding genes, all host-associated *Deltaproteobacteria* genomes share at least 3 other genes (*tagQ*, *ppkA* and *ligA*) in the same neighborhood. The majority of host-associated *Deltaproteobacteria* contained 3 additional homologous genes (*tagR*, *tagS*, *tagT*), resulting in a neighborhood of 10 genes with low homology to deltaproteobacterial genomes from other environments (**Figure 3.8**). Five of these genes (TagQ, TagR, TagS, TagT and PpkA) have been linked to type VI secretion (T6SS) post-translational control proteins in the threonine phosphorylation pathway (TPP) [161]. These genes act as part of a transmembrane signaling pathway that promotes T6SS activity under optimal conditions. Despite this, only *Desulfovibrio* sp. 3_1 contains genes necessary for type VI secretion machinery.

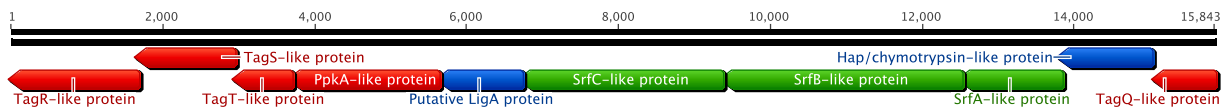


Figure 3.8. Gene neighborhood that includes two genes (*srfB* and *srfC*) unique to host-associated deltaproteobacterial genomes and with low homology to *Deltaproteobacteria* from other environments. Genes are color-coded based on an association with the *srf* operon (green), type VI secretion (T6SS) post-translational control (red) or unknown association (blue). Genes encoding TagR, TagT and TagS were not found in Dsv1 or *B. wadsworthia*.

PpkA is a threonine protein kinase responsible for the phosphorylation of T6SS component Fha1, an event that triggers the export of an effector protein [161]. Thus, it is possible that PpkA could trigger the export of SrfC, which was also predicted to be an effector protein [156]. TagQ is an outer membrane lipoprotein and is suspected to be involved in signal detection, due to its similarity to a 17kDa *Rickettsia* surface antigen [162]. TagR is proposed to be a co-receptor for the environmental signal that activates T6SS (Hsu, Schwarz et al. 2009). TagT and TagS form an ABC transporter that falls into a family of transporters that interacts with membrane-integrated histidine kinases and seem to play a role in regulated responses to external environment (Casabona, Silverman et al. 2012). All of the host-associated deltaproteobacterial genomes have TPP genes that are most closely related to those found in *Pseudomonas* species. This system in *Pseudomonas aeruginosa* has been shown to be an important virulence component [161].

One additional gene in this gene neighborhood encodes for a protein of unknown function. However, it was annotated as a LigA protein in at least one genome, and top BLASTP hits included *Taylorella asinigenitalis*, *Xenorhabdus bovienii* SS-2004, and *Proteus penneri* ATCC 35198, all host-associated organisms. Although it remains unclear how these genes function, they are likely important to survival in the host and interactions with both host and other bacterial cells.

3.4. Conclusion

The genomes of Dsb1-5 and Dsv1 enabled the first insight into the potential functions of these *Deltaproteobacteria* within the oral environment, and the metabolic and transport properties of each group have been outlined in **Figure 3.9** and **3.10**. Even with an extreme environmental change, both groups still had considerable metabolic overlap with related environmental organisms. Although no other host-associated *Desulfobulbus* species have been sequenced, comparisons of Dsv1 with gut-associated *Desulfovibrio* revealed that there is good congruence in these genomes. However, it is clear that some oral and gut-associated species are more closely related than others (Dsv1 and *D. sp_3_1_syn3* versus Dsv1 and *D. piger*). Further, analyzing distinct genes of Dsb1-5 and Dsv1 revealed a suite of genes that is essential for a host-adapted lifestyle. These unique genes are associated with adhesion, stress resistance, defense mechanisms and possible host-cell interactions and degradation. Thus, it is possible that associations of these organisms with disease are due to these virulent properties. This insight into the potential importance of uncultured, low-abundance members of microbial communities to disease will hopefully encourage similar genomic studies of other putative pathogens.

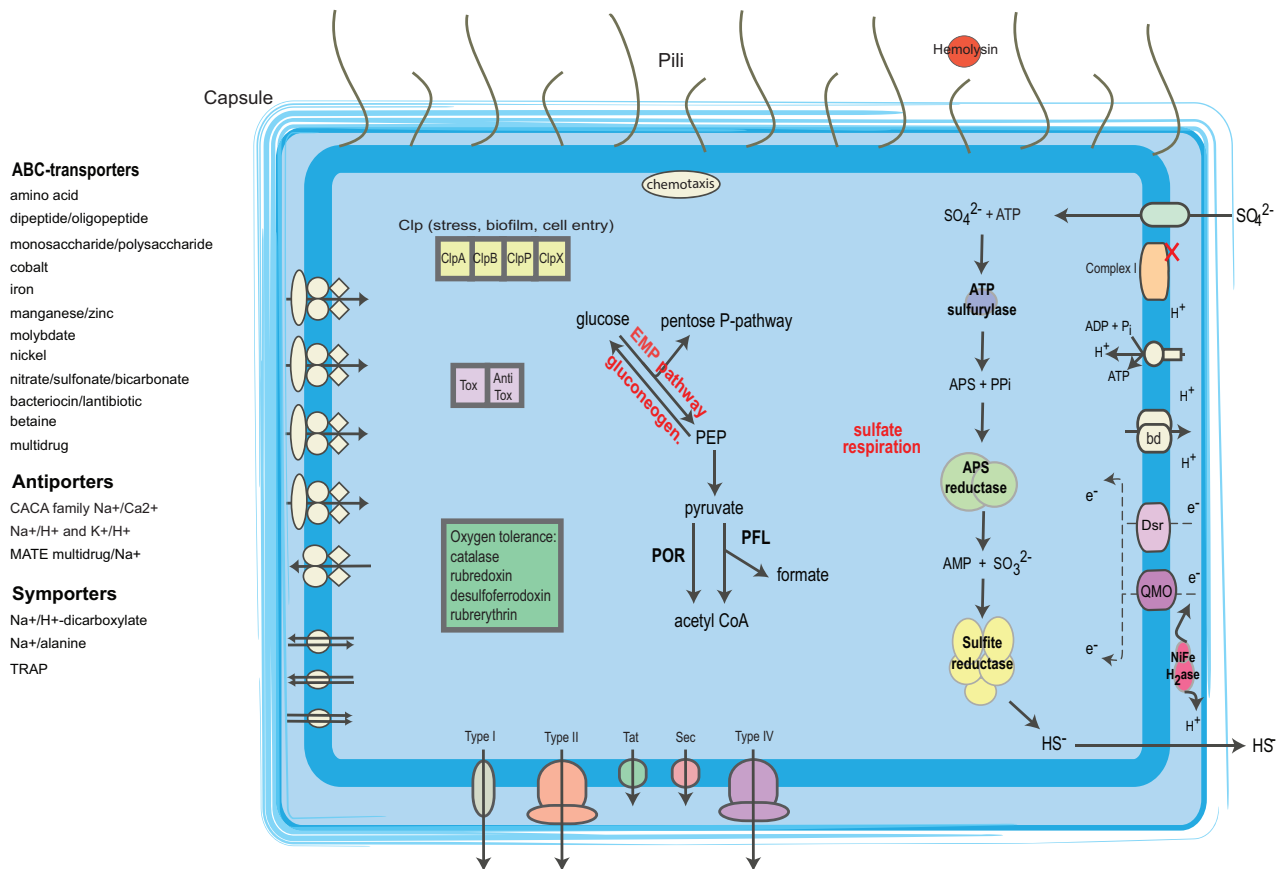


Figure 3.9. Brief overview of proposed metabolic and transport capabilities of Dsb1-5. EMP, Embden-Meyerhof pathway. APS, adenosine 5'-phosphosulfate. bd, cytochrome bd. Dsr, complex DsrMKJOP. QMO, QmoABC complex. Complex I, Nuo-type complex I (missing periplasmic components are denoted by a red X). H₂ase, hydrogenase. PEP, phosphoenolpyruvate. PFL, pyruvate formate lyase. POR, pyruvate ferredoxin oxidoreductase. TRAP, tripartite ATP-independent periplasmic. MATE, multidrug and toxic compound extrusion. Tox/AntiTox, killer gene system.

Chapter 4. Single-cell genomic analyses of an uncultured member of phylum *Chloroflexi* found in the human oral cavity

4.1. Introduction

The phylum *Chloroflexi* is a diverse and ubiquitous group of organisms that has been found to play important roles in various environments, from wastewater treatment systems [163, 164] to sponges [165]. Hugenholtz *et al.* first phylogenetically grouped these organisms (formerly green non-sulfur bacteria) into four subdivisions [15]. Most cultured representatives of the *Chloroflexi* phylum are within subdivision III. *Chloroflexus* species in subdivision III are the earliest branching bacteria capable of photosynthesis and are therefore, of great interest evolutionarily [166]. However, the majority of environmental clones and other deposited *Chloroflexi* sequences fall within subdivision I, leading to speculation that this subdivision has important environmental roles [15, 167]. Recent strides have resulted in several cultured isolates in subdivision I classes *Anaerolineae* and *Caldilineae*, including recently sequenced *Anaerolinea thermophila*. Although normally thought of as environmental organisms, several pyrosequencing and clone studies have shown a distinct clade of *Anaerolineae* as minority members of host-associated communities, including the oral cavity and fecal samples from a variety of mammals [32, 43].

Because there are no cultured representatives of host-associated *Chloroflexi*, there is nothing known about the biological role of *Chloroflexi* within this environment. Further, it is not known how these organisms and their

genomic capabilities compare to those found in the environment. Therefore, we used flow cytometry and multiple displacement amplification to obtain genomic DNA from two single cells of *Anaerolineae* retrieved from human subgingival samples. These single amplified genomes (SAGs) were used to further investigate the functional potential of this organism and compare it to other environmental isolates.

4.2. Materials and methods

4.2.1. Sample collection

Two cells (Chl1 and Chl2) belonging to the phylum *Chloroflexi* were isolated in two distinct sorting events. Both Chl1 and Chl2 were retrieved from subgingival paper point samples taken from diseased sites of periodontitis patients. Patients were identified as having periodontitis based upon established criteria [18]. Institutional Review Board approval and informed consent was obtained for all samples. In all cases, samples were taken, processed, fixed and stored as described in Chapter 2.

4.2.2. Cell retrieval

Chl1 and Chl2 were randomly sorted with a Cytopeia INFLUX sorter (BD, Franklin Lakes, NJ, USA). The influx cell sorter was cleaned prior to use in a similar manner to those described earlier [72, 91] as discussed in Chapter 2. Ethanol-fixed oral samples from periodontitis patients were passed through a CellTrics 30 µm disposable filter (Partec, Görlitz, Germany) prior to sorting to eliminate particles large enough to obstruct the tubing of the cell sorter. Samples

were then incubated with 5 μ M of nucleic acid dyes SYTO 9 (green) and SYTO 62 (red) (Life Technologies, Grand Island, NY, USA) for fifteen minutes. Approximately 20,000 cells were sorted into a single tube based upon 528-538 fluorescent emissions. This tube was then subjected to a second round of sorting based upon 670-730 fluorescent emissions to further dilute any possible contaminating DNA found in the original sample [91]. In the final round of sorting, single cells were sorted into 3 μ L UV-sterilized, 1 $\times$ TE buffer in individual wells of a 96-well plate.

4.2.3. Multiple displacement amplification

Multiple displacement amplification (MDA) reactions were used to produce enough genomic DNA from individual cells for further investigation. Reactions were set up in a manner similar to [91] as discussed in Chapter 2. All identified MDA products of interest were used to generate secondary MDA products.

4.2.4. Target identification

One microliter of each MDA amplicon was used to make dilutions (1:150) in PCR-grade water (Ambion, Austin, TX, USA), and the remainder of each amplicon was vacuum-sealed and stored at -80C. Dilutions were used to set up PCR amplifications using universal bacterial primers 27fm and 1492r [92] using the conditions discussed in Chapter 2. Phylogenetic identifications were obtained using RDP Classifier [100].

4.2.5. MDA purification and sequencing

Secondary MDA products of interest were purified with a phenol:chloroform:isoamyl alcohol extraction and alcohol precipitation method

before being sent for sequencing. Illumina Hi-Seq 100-bp paired-end libraries were constructed and sequenced at Hudson Alpha Institute for Biotechnology (Huntsville, AL, USA).

4.2.6. Genome Assembly and Annotation

Single amplified genome (SAG) library Chl1 were subjected to the JGI single-cell assembly and QC pipeline, similar to Illumina SAG libraries described previously [77]. In addition, Chl2 and a combined genomic dataset was normalized and assembled in-house using digital normalization [85] and Velvet [105] for assembly. Assembled genomes were subjected to the JGI/ORNL automated annotation pipeline, which uses the gene calling program Prodigal [106], before submission into Integrated Microbial Genomes (IMG) [109] for further analysis. Additionally, data was submitted into RAST [110] for analysis. Contamination was found and removed based on BLASTP [107] results for each translated protein, as well as GC content and tetramer analysis [108] of each scaffold.

4.2.7. Combined non-redundant assembly

To assess the phylogenetic relationships between our SAGs Chl1 and Chl2, comparisons of average nucleotide identity (ANI) were calculated using the software JSpecies [114]. In order to discuss the potential function of oral *Chloroflexi* as a whole, both datasets were combined (Chl1-2).

4.2.8. Genome size estimation

Genome size and coverage was estimated using a conserved single copy gene (CSCG) set that has been determined from 1516 finished bacterial genome

sequences in the IMG database [109]. The set consists of 138 CSCG that were found to occur only once in at least 90% of all genomes by analysis of an abundance matrix based on hits to the protein family (Pfam) database [113]. Hidden Markov models of the identified Pfams were used to search assemblies by means of the HMMER3 software [116]. Resulting best hits above pre-calculated cutoffs were counted and the coverage was estimated as the ratio of found CSCG to total CSCGs in the set after normalization to 90%. The estimated complete genome size was then calculated by division of the estimated genome coverage by the total assembly size.

4.3. Results and discussion

4.3.1. Cell isolation and phylogeny

Single cells of oral *Chloroflexi* were isolated from ethanol-fixed subgingival samples using nucleic acid dyes and an in-flux cell sorter. The genomic DNA of single cells was amplified using MDA, and genomic material was sequenced for two discreet cells (referred to as Chl1 and Chl2). Both Chl1 and Chl2 came from individuals with periodontitis.

Based on SSU rRNA genes, Chl1 and Chl2 fall into the class *Anaerolineae* within subdivision I of the *Chloroflexi* phylum and form part of a clade that is distinct from its closest sequenced relative, *Anaerolinea thermophila* (**Figure 4.1**). The sequences from this study are 99.7% identical to a sequence seen in other human oral clone studies [43, 168], and a similar sequence was also found in the oral cavity of dogs [169]. A more distantly related clade contains

sequences from mammalian rumen and fecal samples [32, 170]. Comparisons of Chl1 and Chl2 with recent pyrosequencing data of partial SSU rRNA genes [18, 49] reveal that the majority of *Chloroflexi* identified are at least 99% identical to Chl1 and Chl2 (**Figure 4.2**). The number of oral sequences identified in classes other than *Anaerolineae* is low and often found in only one individual, suggesting that these may be transient community members or environmental contaminants.

4.3.2. Genome statistics

After normalization, assembly and contamination removal, Chl1 contained 1.1 Mb of data, and Chl2 contained 1.2 Mb of data (**Table 4.1**). The GC content for these partial genomes ranged from 53-54%. Analysis of average nucleotide identity (ANI) revealed an identity of 98.29% between Chl1 and Chl2, based on 37.24% overlap of the genomes. Due to this high similarity and the low *Chloroflexi* diversity seen in past oral studies, we treated these SAGs as members of a single operational taxonomic unit (OTU). All downstream metabolic analyses were performed on a combined assembly of the *Chloroflexi* data referred to as Chl1-2 throughout the paper. After removing redundant data, Chl1-2 had a total of 1774 unique genes and 1.8 Mb of data, with a GC content of 53%. Estimation of the genome size based on a conserved single copy gene set is approximately 2.7 Mb. Thus, we were able to capture approximately 67% of this novel oral genome.

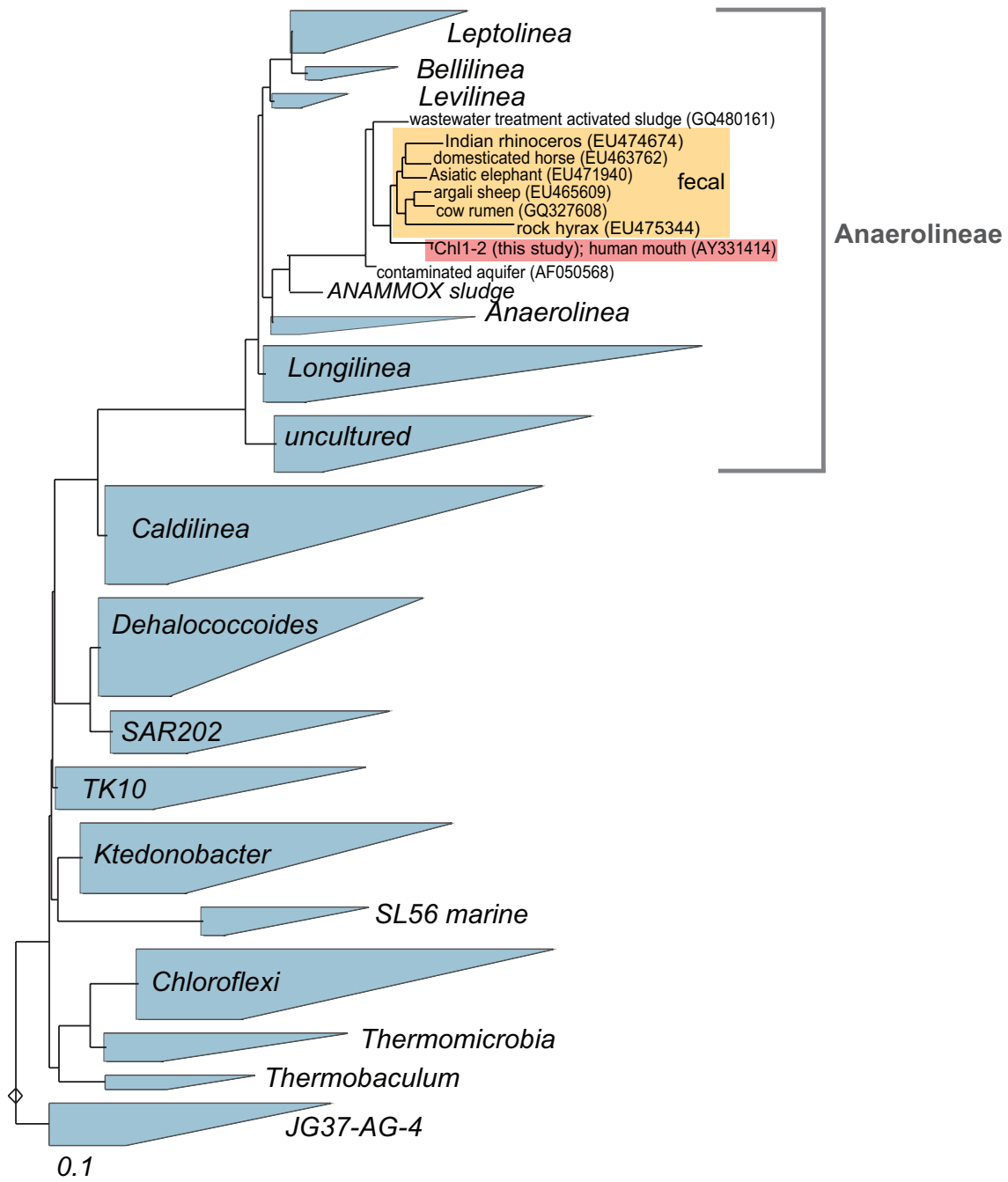


Figure 4.1. Neighbor-joining phylogenetic tree based on *Chloroflexi* with the SSU rRNA Greengenes database. The tree was constructed using ARB. Oral isolates are highlighted in red. Fecal and rumen isolates are highlighted in yellow. Accession numbers for sequences are given in parentheses.

Figure 4.2. Maximum likelihood phylogenetic tree of human-associated, oral *Chloroflexi* based on partial SSU rRNA genes. The tree was reconstructed using PHYML [130] in the program Geneious® Pro 5.6.5 with a JC69 substitution model. Sequences from this study are shown in red. Sequences from [49] are highlighted in blue and labeled as “HMP” followed by an abbreviation of the body site where the sequence was collected. BM – buccal mucosa, HP – hard palate, KG – keratinized gingiva, SB – subgingival, SP – supragingival, TD – tongue dorsum. Sequences from [18] are all subgingival, highlighted in green and are labeled as “OSU” followed by “H” and “P” to denote healthy and periodontitis samples, respectively. The number of sequences is shown in parentheses. A * indicates that sequences came from only one individual. The scale bar indicates 0.07 substitutions per nucleotide position. Numbers given at the nodes represent bootstrap percentages calculated on 100 replicates.

Figure 4.2.

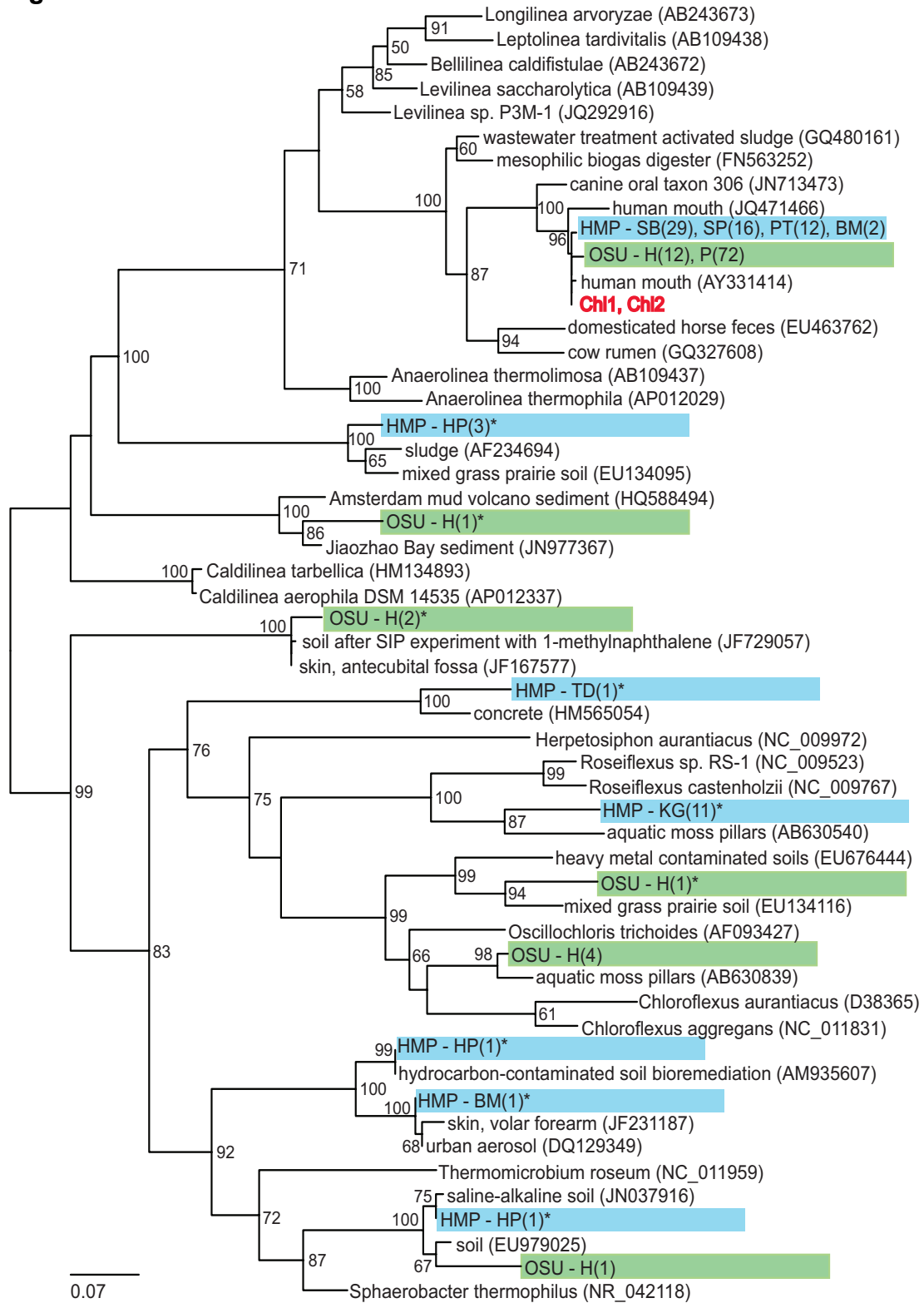


Table 4.1. Assembly statistics and metadata for combined Chl1-2 genome.

	Chl1-2
Cells sequenced	2
Assembly size (bp)	1,834,740
G+C %	53.2
DNA scaffolds	259
N50 (bp)	10,120
Largest scaffold (bp)	32,647
Est. genome size (Mb)	2.7
Est. genome recovery	67%
Genes total number	1820
RNA genes	47
rRNA genes	7
5S rRNA	3
16S rRNA	1
23S rRNA	3
tRNA genes	36
Protein coding genes	1773

4.3.3. Metabolism

Chl1-2 appears to be an anaerobic heterotroph and encodes a large percentage of genes for a wide variety of carbohydrate metabolism. In general, this finding is not surprising, as other, cultured *Anaerolineae* have been shown to be fermentative heterotrophs, and use of carbohydrates has been shown in the laboratory and observed in wastewater sludge granules [167]. Oral Chl1-2 shares many aspects of metabolism with its closest sequenced relative, *Anaerolinea thermophila*. *A. thermophila* was the top BLASTP hit for 23% of the predicted proteins. However, approximately 38% of predicted proteins within Chl1-2 match to unknown or hypothetical proteins.

Several central metabolism capabilities overlap with those found in other organisms within the *Chloroflexi* phylum. A complete glycolysis/gluconeogenesis pathway can be found in *A. thermophila*. It appears Chl1-2 also encodes a glycolysis/gluconeogenesis pathway, although genes for the enzymes needed to convert glucose to fructose-6-P are missing (**Figure 4.3**). It is unclear if the absent genes are the result of an incomplete genome. However, the additional archaeal type glyceraldehyde-3-phosphate ferredoxin oxidoreductase and fructose-1,6-bisphosphatase present in *A. thermophila* were not found in Chl1-2. Also, several glycolysis and gluconeogenesis genes in Chl1-2 seem to have been horizontally transferred, including glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase that show similarities to archaeal type enzymes. Chl1-2 also encoded a gene for an ADP-dependent phosphofructokinase, although it did not appear to be closely related to the enzyme seen in archaea. Several genes of the pentose phosphate pathway are also present in Chl1-2 (**Figure 4.3**).

A. thermophila appears to encode only a partial TCA cycle including 2-oxoglutarate ferredoxin oxidoreductase. Cultures of *Chloroflexus aurantiacus* indicate a complete TCA cycle that includes enzymes for the glyoxylate shunt [171, 172]. Few genes of the TCA cycle were present in the Chl1-2 dataset, and there was no evidence of 2-oxoglutarate ferredoxin oxidoreductase or glyoxylate shunt enzymes, making it difficult to speculate how the TCA cycle functions in this organism.

Evidence of energy production in Chl1-2 includes several subunits encoding ATP synthase (**Figure 4.3**). No evidence was found in Chl1-2 for cytochromes. All sequenced *Chloroflexi* available in IMG, including *A. thermophila*, encode NADH:quinone oxidoreductase (complex I) for electron transport. Although Chl1-2 does contain two genes similar to complex I subunits, these genes match more closely to *hymA* and *hymB*, genes that encode subunits responsible for electron transport in an iron-only hydrogenase [173]. The *hymAB*-like genes are adjacent to an iron-only hydrogenase catalytic unit gene (*hymC*) (**Figure 4.4**). Although these genes were originally found in association with a formate dehydrogenase, no evidence of this complex was found in Chl1-2.

A search for distinctive genes found in photosynthetic *Chloroflexus* or *Roseiflexus* species revealed little overlap with genes in Chl1-2. Genes characteristic for the 3-hydroxypropionate carbon fixation bi-cycle that has been described in *C. aurantiacus* [174] were not found in Chl1-2 or *A. thermophila*. Likewise, neither Chl1-2 or *A. thermophila* contain homologous genes for photosynthetic antenna and reaction center genes found in other *Chloroflexi* [166]. Although *A. thermophila* does encode some genes similar to those found in the alternative complex III in *C. aurantiacus*, there is no evidence for this complex in Chl1-2.

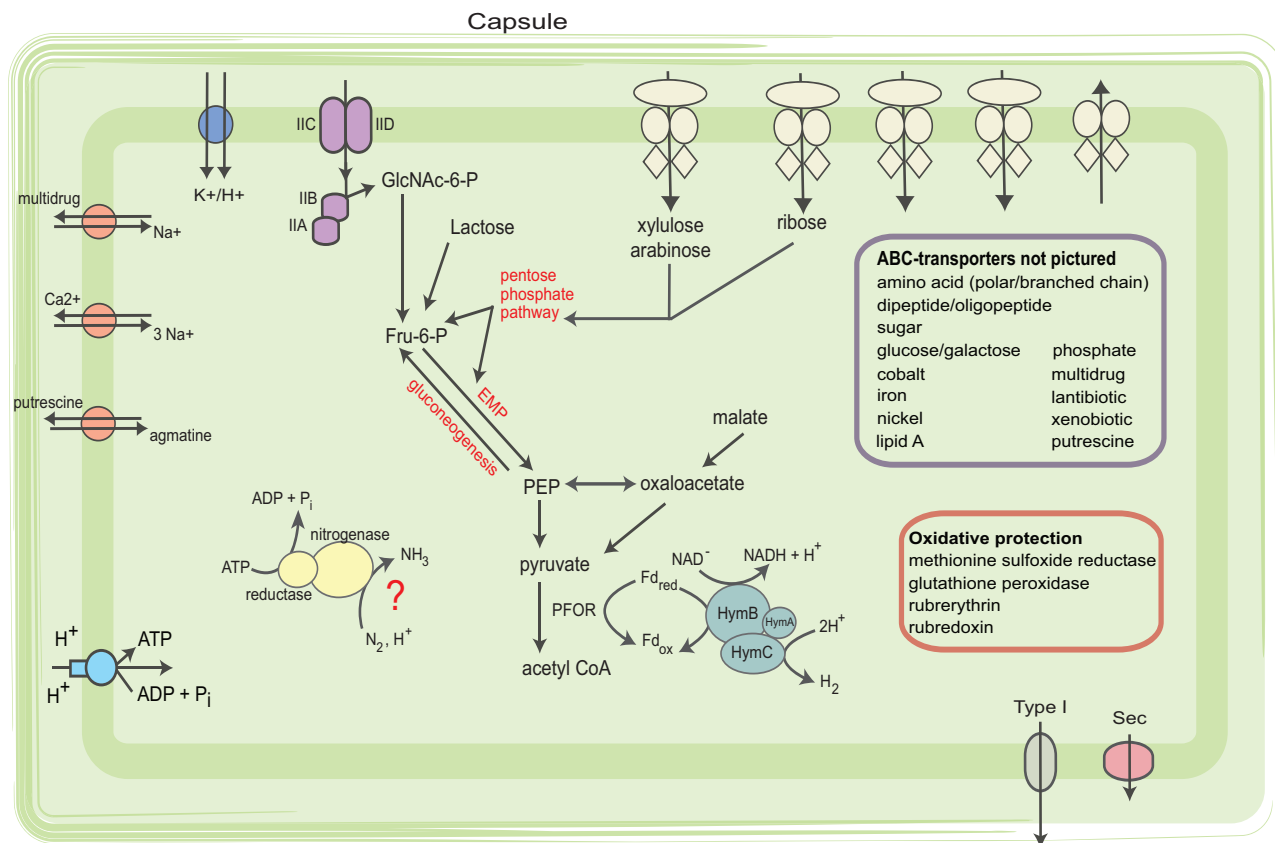


Figure 4.3. Brief overview of proposed metabolic and transport capabilities of Chl1-2. EMP, Embden-Meyerhof-Parnas glycolytic pathway. PEP, phosphoenolpyruvate. MATE, multidrug and toxic compound extrusion. GlcNAc-6-P, *N*-acetylglucosamine-6-phosphate. Fru-6-P, fructose-6-phosphate. PFOR, pyruvate: ferredoxin oxidoreductase. Fd, ferredoxin. HymA, electron transport subunit HymA. HymB, electron transport subunit HymB. HymC, iron-only hydrogenase HymC.

Other similarities to available *Chloroflexi* genomes include a lack of any genes that would encode pili or flagellar machinery for cell motility. In addition, all isolates of *Anaerolineae* have lacked gliding motility that has been seen in *Chloroflexus* species [175, 176], making it unlikely that Chl1-2 has gliding motility. It is not clear if Chl1-2 is filamentous, but all other isolates from subphylum I have been multicellular filaments. Also, similar to *A. thermophila* and other, photosynthetic *Chloroflexi*, Chl1-2 appears to have the genes necessary for a functioning mevalonate pathway. Like *A. thermophila*, Chl1-2 also encodes at least a partial urea cycle, and genes were found that could convert ammonia and aspartate to fumarate.

4.3.4. COG comparisons

To understand the unique functional role of Chl1-2 in a host-associated, oral environment, the presence/absence of clusters of orthologous groups (COGs) was used to compare Chl1-2 with its closest relative, *A. thermophila*, and other available *Chloroflexi* genomes. Comparisons to *A. thermophila* revealed 606 COGs found in both organisms and 182 unique to Chl1-2. *A. thermophila* contained a large number of distinct COGs (795), although some of these may be seen in Chl1-2 with a complete genome. A large number of differential COGs included uncharacterized or predicted proteins (38%), and several of these were membrane-associated proteins. Differential COGs included several ABC transporters, genes for a distinct toxin-antitoxin system, and genes for protection from oxidative and ultraviolet damage including genes similar to methionine sulfoxide reductase, photolyase and glutathione peroxidase.

Among the most interesting differential COGs, 38 were found exclusively in Chl1-2 and not in the 18 *Chloroflexi* genomes available in IMG (**Table 4.2**). Unfortunately, fifteen differential COGs had no functional prediction or only a general function prediction. However, several unique systems were observed including a number of genes involved in N-acetylglucosamine transport and lipopolysaccharide biosynthesis that will be discussed in detail below.

Table 4.2. COGs present in Chl1-2 and absent in other available *Chloroflexi* genomes.

COG	Description
Energy production and conversion	
1145	Ferredoxin
Amino acid transport and metabolism	
1687	Predicted branched-chain amino acid permeases (azaleucine resistance)
Nucleotide transport and metabolism	
3613	Nucleoside 2-deoxyribosyltransferase
0813	Purine-nucleoside phosphorylase
Carbohydrate transport and metabolism	
4809	Archaeal ADP-dependent phosphofructokinase/glucokinase
2731	Beta-galactosidase, beta subunit
3444	Phosphotransferase system, mannose/fructose/N-acetylgalactosamine-specific IIB
3715	Phosphotransferase system, mannose/fructose/N-acetylgalactosamine-specific IIC
3716	Phosphotransferase system, mannose/fructose/N-acetylgalactosamine-specific IID
3855	Uncharacterized protein conserved in bacteria
Transcription	
1996	DNA-directed RNA polymerase, subunit RPC10 (contains C4-type Zn-finger)
3279	Response regulator of the LytR/AlgR family
DNA replication, recombination and repair	
3723	Recombinational DNA repair protein (RecE pathway)
Cell envelope biogenesis, outer membrane	
3754	Lipopolysaccharide biosynthesis protein
1442	Lipopolysaccharide biosynthesis proteins, LPS:glycosyltransferases
4953	Membrane carboxypeptidase/penicillin-binding protein PbpC
0562	UDP-galactopyranose mutase
Posttranslational modification, protein turnover, chaperones	
4545	Glutaredoxin-related protein
Secondary metabolites biosynthesis, transport and catabolism	
4869	Propanediol utilization protein
Signal transduction mechanisms	
3480	Predicted secreted protein containing a PDZ domain

Table 4.2. (continued)

COG	Description
Signal transduction mechanisms	
3290	Signal transduction histidine kinase regulating citrate/malate metabolism
4725	Transcriptional activator, adenine-specific DNA methyltransferase
General function prediction only	
3740	Phage head maturation protease
218	Predicted GTPase
2194	Predicted membrane-associated, metal-dependent hydrolase
5618	Predicted periplasmic lipoprotein
Function unknown	
5283	Phage-related tail protein
3152	Predicted membrane protein
5346	Predicted membrane protein
1683	Uncharacterized conserved protein
2966	Uncharacterized conserved protein
3542	Uncharacterized conserved protein
3610	Uncharacterized conserved protein
5015	Uncharacterized conserved protein
5416	Uncharacterized integral membrane protein
3016	Uncharacterized iron-regulated protein
4496	Uncharacterized protein conserved in bacteria

4.3.5. Carbohydrate metabolism

Several parts of carbohydrate metabolism in Chl1-2 appear similar to those found in *A. thermophila* and the majority of *Anaerolineae* isolates [167]. Chl1-2 contains the necessary genes for the transport and degradation of fructose, xylose, ribose and arabinose (**Figure 4.3**). Additional partial pathways were found in Chl1-2 for the degradation of glucose, lactose, galactose and mannose. Chl1-2 also has genes for pyruvate ferredoxin oxidoreductase, an enzyme used in anaerobic fermentation of carbohydrates to convert pyruvate to acetyl CoA. No enzymes were found in Chl1-2 for conversion of acetyl CoA to acetate, but this could be

due to the incomplete genome. Common end products for all *Anaerolineae* isolates include acetate and hydrogen [175, 176].

4.3.6. *N*-acetylglucosamine transport and metabolism

In addition to several shared genes, Chl1-2 seems to have acquired a unique set for carbohydrate metabolism genes not seen in any other currently sequenced *Chloroflexi* genomes, including a phosphotransferase system (PTS) that has an ability to transfer a wide range of sugars. Studies in *E. coli* have shown that this PTS system can transport mannose, glucose, fructose, *N*-acetylglucosamine (GlcNAc), glucosamine and *N*-acetylmannosamine (ManNAc) [177]. Chl1-2 also contains homologs of GlcNAc-6-phosphate (GlcNAc-6-P) deacetylase and GlcNAc-6-P deaminase, enzymes involved in the catabolism pathway of GlcNAc-6-phosphate to fructose-6-phosphate [177]. The resulting fructose-6-phosphate could then feed into glycolysis. Together, these genes could encode a pathway for the metabolism of GlcNAc, a major component of bacterial cell walls (Figure 3). The presence of this pathway would support previous *in situ* reports that subphylum I *Chloroflexi* members are capable of utilizing GlcNAc [164, 178]. *Chloroflexi* members appear to retrieve GlcNAc from other lysed cells in the environment, as seen with several microautoradiography (MAR)-FISH studies [179-181].

In addition to the previously discussed GlcNAc genes, Chl1-2 has a gene homologous to *N*-acetylneuraminase (Neu5Ac) lyase. This protein converts the sialic acid Neu5Ac into ManNAc and pyruvate (**Figure 4.4**). Although no homologs of known sialic acid transporters were found in our partial genome, the

presence of Neu5Ac lyase suggests that Chl1-2 might also have the ability to transport and catabolize sialic acid from the environment. This ability would be useful in the oral environment, as most mammalian cell surfaces have an abundance of sialic acid residues within their glycoproteins and glycolipids [177]. Sialic acid appears to provide growth factors for some pathogenic oral bacteria, such as *Tannerella forsythia* [182]. Furthermore, Chl1-2 encodes Neu5Ac synthase and Neu5Ac cytidyltransferase, proteins necessary for the production of the sialic acid (**Figure 4.4**). Sialic acid biosynthesis genes are also present in *A. thermophila* but are not found in other available *Chloroflexi* data. Although it is unclear what the purpose of this pathway may be in *A. thermophila*, it is likely that the production of sialic acid in the oral environment helps Chl1-2 evade immune response by mimicking the exterior of host cells [182].

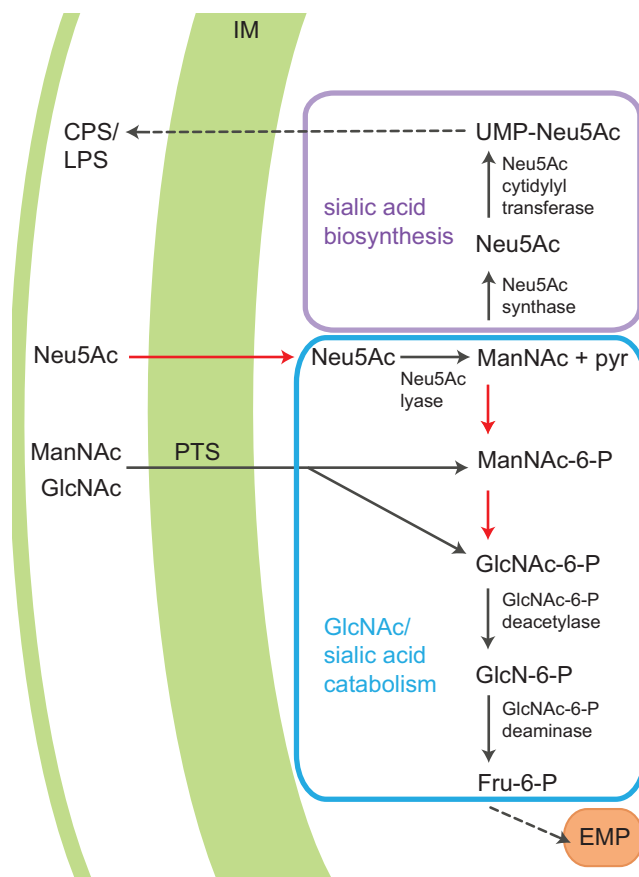


Figure 4.4. Diagram of *N*-acetylglucosamine (GlcNAc) and sialic acid *N*-acetylneuraminate (Neu5Ac) metabolism. Enzymes not found in Chl1-2 are shown with red arrows. IM, inner membrane. CPS, capsular polysaccharide. LPS, lipopolysaccharide. pyr, pyruvate. PTS, mannose-type phosphotransferase system. ManNAc, *N*-acetylmannosamine. GlcN-6-P, glucosamine-6-phosphate. Fru-6-P, fructose-6-phosphate. EMP, Embden-Meyerhof-Parnas glycolytic pathway.

4.3.7. Inositol metabolism

Another interesting gene neighborhood encodes several genes necessary for inositol catabolism. Chl1-2 contains a neighborhood of four genes that could convert 1D-*myo*-inositol (MI) to 2-deoxy-5-keto-D-gluconic acid-6-phosphate. It is likely that this compound is then converted to either acetyl-CoA or glyceraldehyde-3P for entry into glycolysis. These genes are only present in the distantly related *Chloroflexi* genera *Oscillochloris*, *Ktenobacter* and *Roseiflexus*,

and the genes in Chl1-2 seem more closely related to those found in *Firmicutes* based on BLASTP hits. Interestingly, although MI is abundant in soil [183], catabolic genes for MI have been found in many host-associated bacteria [183-185]. It has been suggested that the presence of the MI operon in gram-negative bacteria is due to horizontal gene transfer, as it is often seen in only a subset of related strains [183, 185].

4.3.8. Cellular envelope and outer membrane proteins

COG analysis revealed the presence of lipopolysaccharide (LPS) biosynthesis proteins that were unique to Chl1-2. These included proteins necessary for the production of galactose and rhamnose type O-antigens. As previously mentioned, Chl1-2 also carries genes necessary for the production of sialic acid that can be incorporated in LPS and capsular polysaccharides to mimic the surface of epithelial cells [182].

Chl1-2 also contains several genes that may function in capsule transport and production. These genes encode proteins involved in export of LPS to the cell surface [186], a protein similar to CapD [187] and poly-gamma-glutamate biosynthesis protein CapA. The presence of a capsule would benefit Chl1-2 in the oral environment by providing adherence and resistance to both specific and nonspecific host immune systems [153].

4.3.9. Nitrogen metabolism

Chl1-2 encodes both the nitrogenase molybdenum iron protein and the reductase protein for a putative nitrogenase. Thus, if this enzyme functions, Chl1-2 would be capable of nitrogen fixation. Other distantly related *Chloroflexi* genera

Dehalococcoides, *Roseiflexus* and *Oscillochloris* also contain nitrogenase genes. However, the proteins encoded in Chl1-2 have a broad phylogenetic range of BLASTP hits. The nitrogenase reductase protein NifH has also been seen in host-associated spirochetes. However, nitrogenase activity detected in treponemes from the termite gut was low, and no activity was detected in the oral spirochete *Treponema denticola* [188]. It is unclear what other functions these genes may encode.

4.3.10. Transport and secretion

Chl1-2 contains a large number of transporters for a wide range of substrates. Genes that were closely related to those found in *A. thermophila* encoded for ABC transporters that carried amino acids, branched amino acids and unknown substrates as well as a magnesium transport protein. Based on COG analysis, Chl1-2 contains transporters for sugar, glucose/galactose, maltose and oligopeptides that are not found in *A. thermophila*. Other uptake systems found in Chl1-2 include proteins specific for iron, nickel, ribose/xylose/arabinose/galactoside, dipeptides, polar amino acids, nucleosides, phosphate and potassium and the mannose-type PTS system discussed above.

Systems for export and efflux transport drugs, lantibiotics, xenobiotics, bacitracin, chromate, and lipid A. Export machinery in common with *A. thermophila* includes an HlyD-like secretion protein, a putative cation efflux protein and a MATE (multidrug and toxic compound extrusion) efflux transporter. Chl1-2 encodes the HlyD-like proteins and major facilitator superfamily (MFS)

proteins necessary for type I secretion and Sec proteins from the general secretion pathway. No evidence for Type II – VI secretion was found.

4.3.11. Stress adaptation and defense mechanisms

There are many proteins encoded by Chl1-2 that would help it effectively deal with environmental changes and stressors. These include heat shock proteins, cold shock proteins and a heat-inducible transcriptional repressor. Chl1-2 genes that could be involved in protection from oxidative stress include methionine sulfoxide reductase, glutathione peroxidase, rubrerythrin and rubredoxin (**Figure 4.3**). Several of these genes are not seen in *A. thermophila*, suggesting that Chl1-2 may be more sensitive to low levels of oxygen. These genes might prove useful for particularly oxygen sensitive enzymes in Chl1-2, including the putative nitrogenase. In addition to the multidrug and antibiotic transporters discussed above, Chl1-2 also carries a nitroimidazole antibiotic resistance protein. Chl1-2 encodes a Clp protease that has been shown to be important for stress response in a number of host-associated bacteria [147, 148, 150].

Competence proteins similar to ComEA and ComEC were found in Chl1-2, along with a regulator of competence-specific genes. This suggests that Chl1-2 might have natural competence capabilities that would enable horizontal acquisition of beneficial genes, although the pili-related genes that are often associated with competence are not present. In addition, Chl1-2 harbors killer and antidote proteins of a killer gene system. This system can be used for stress response, cell cycle arrest and maintenance of otherwise disposable genes

[134]. Chl1-2 also contains genes necessary for a type I restriction modification system.

In the oral environment, most iron is bound to the host protein lactoferrin and iron availability is limited [146]. Thus, successful acquisition of iron is important for survival. In addition to the iron transport systems discussed above, Chl1-2 also contains hemolysin III.

4.4. Conclusion

Through the use of flow cytometry and multiple single-cell genomes, we were able to capture approximately 70% of an uncultured member of the class *Anaerolineae* found in low abundance within the oral community. Genome analysis revealed Chl1-2 has metabolic properties that are similar to other isolated and sequenced *Anaerolineae*. The Chl1-2 genome encodes a large array of proteins used for carbohydrate transport and metabolism, including several substrates that have been used in environmental *Anaerolineae* isolate studies [175, 176]. Thus, it is likely that these capabilities are important for all *Anaerolineae*, regardless of the environment in which they reside. In addition, Chl1-2 encodes several proteins for carbohydrate metabolism that are not seen in its closest relative, *A. thermophila*. Chl1-2 encodes all the proteins necessary for uptake and metabolism of environmental GlcNAc. Based on previous MAR-FISH studies, this pathway in Chl1-2 is likely shared with uncultured *Anaerolineae* within wastewater treatment systems [179-181]. Thus, an important component of metabolism in Chl1-2 involves the uptake of cell wall components

from surrounding lysed bacterial cells. It is interesting to speculate that the larger number of Chl1-2 sequences seen in diseased individuals [18] may be due to increased cell turnover in the more abundant bacterial populations associated with periodontitis. In addition to uptake of lysed bacterial cells, it is possible that Chl1-2 might be able to use sialic acid from surrounding host, epithelial cells. Sialic biosynthesis genes found in Chl1-2 could allow this cell to mimic the exterior of host cells and enable this organism to escape immune response. Overall, the genomic information for Chl1-2 portrays an organism that is similar to other *Anaerolineae*, yet displays unique abilities that are advantageous for survival in a host environment.

Chapter 5. Conclusions and future directions of the studies of uncultured microbiota

5.1. Summary of presented work

Research presented in this dissertation provided an efficient method for the analysis of uncultured microbes through the use of flow cytometry and single-cell genomics and revealed insights into two interesting, uncultured groups found within the human oral cavity. Further, a high-throughput process was outlined which provided a way to isolate numerous uncultured microbes, including those in low abundance, from an environment of interest. This method could be used to add genomic information from a wide range of taxa to large-scale databases and provide reference information for many other studies. Thus, methods for non-specific, high-throughput uncultured studies and smaller, targeted approaches of single-cell isolation and analysis were examined. The combination of both methods should allow in-depth analyses of any environment. Such efforts would be augmented by complementary technologies, such as metagenomics and metaproteomics.

In addition to the methods discussed, specific genome studies detailed in Chapters 3 and 4 described many techniques that can be used to interpret resulting data and highlighted functional insights that can be gained with such studies. The study on *Desulfobulbus* and *Desulfovibrio* was able to pinpoint several genes that could give these organisms an advantage in the oral environment and contribute to pathogenesis. Many of the resulting data could be used in further studies to confirm the exact role of these organisms. For example,

using genome comparisons, a gene neighborhood was found that was unique to host-associated *Deltaproteobacteria* and was not found in any available genomes of environmental isolates. These genes would be ideal candidates for biochemical studies, which could be performed in other cultured, host-associated *Deltaproteobacteria*. Studies of *Chloroflexi* discussed in Chapter 4 were able to provide genomic information for a subphylum that is found in many environments but is largely uncharacterized. The carbohydrate metabolism and transport genes seen in the two available genomes and the fermentation properties of cultured isolates provided the first clue that members of the class *Anaerolineae* may play a similar role in a broad range of environments. Additionally, information from previous MAR-FISH studies [179] was able to complement the discovery of a *N*-acetylglucosamine metabolic pathway in our oral genome and suggested that these organisms scavenge and use material from neighboring, lysed cells.

Hopefully, the information provided in these studies will initiate further investigations into both *Deltaproteobacteria* and *Anaerolineae*. Specifically, genomic information from these studies could be used to design primers or antibodies for further studies on proteins of interest. In addition, a protein of interest could be studied in a related, cultured isolate or overexpressed in *E. coli* for further description. More broadly, the genomic data from these studies will be available as reference material for future, whole community studies with techniques such as metagenomics, metatranscriptomics or metaproteomics.

5.2. Future directions

The use of cutting-edge, molecular methods has provided a wealth of knowledge for previously uncharacterized microorganisms. This is an exciting step in the understanding of diverse microbes that were completely unknown prior to these molecular breakthroughs. However, our knowledge of yet-to-be cultured microbes is fragmented. The field of microbial ecology has already begun to address some of the missing information with progressive “omics” technologies. Pyrosequencing of the SSU rRNA gene has been groundbreaking but can only tell scientists “who is there”. Further data that addresses the physiological and ecological roles of uncultured microbes can be obtained through genomics, single-cell genomics, metagenomics, metatranscriptomics, metaproteomics and metabolomics.

The benefits of combining single-cell genomics and metagenomics have been discussed [38, 69], and many studies in the immediate future will likely take advantage of these complementary techniques. Metagenomics can be used to survey an entire community, but often at a lesser sequencing depth than can be obtained when sequencing a single target. Although sequencing coverage, depth and binning for metagenomics continue to improve, use of complete and nearly complete reference genomes is still important. In addition, genomic data from a single cell can provide information currently not accessible with metagenomic tools, such as information on associated symbionts, viruses and plasmids [38, 69]. Combining both types of data would allow researchers to work with larger, more complete datasets. These data would be further augmented with the use of

metatranscriptomics to provide information on gene expression or metaproteomics to determine which proteins are actively translated in a given environment.

It is of utmost importance that one future focus will be interpreting the ever-expanding amount of “omics” data available for uncultured organisms to guide the process of cultivation. Phenotypic changes that can be seen in cultivated organisms during genetic studies provide information that cannot be obtained from molecular techniques alone. Information from single-cell genomic studies of uncultured oral bacteria could inform initial cultivation attempts. Use of genomic information has already proven useful for cultivation of microbes from low diversity environments [189, 190].

To further cultivation efforts, isolation of cells will be a key step [191]. Therefore, the flow cytometry methods used to obtain single cells for genomic sequencing could also be used to isolate live single cells on growth medium. Genomic information from single-cell studies could aid in the production of antibodies for specific sorting of live cells. This allows for the cultivation of slow growing cells that can be overtaken by faster growing bacteria when using traditional cultivation techniques. The use of flow cytometry to cultivate cells of interests has been proven in the past [192, 193], and successful cultivation aided by single-cell flow cytometry has already been shown from the oral environment [44]. Additional studies, using samples from a broad range of individuals and sampling locations, would only add to this success.

The combined use of single-cell genomics approaches with whole-community data will be important across all fields of microbial ecology, but the benefits of these technologies will be particularly evident in studies of the human microbiome in the coming years due to the many initiatives aimed at describing these communities. However, understanding of human-associated microbiota provides additional difficulties because both bacterial cells and human cells influence the system. Information for microbial communities and host immunity will need to be combined for a more complete understanding of the system. This will be particularly important in studies of the oral cavity, where interactions between the host immune system and oral microbiota can cause inflammation and disease. Further, the oral epithelium is not protected by a mucosal layer and is highly porous [194]; thus, there is more direct contact between human cells and the microbial community in the oral cavity versus other human body sites. Studies that simultaneously investigate microbial composition and concentrations of key components of the immune system, such as neutrophils and cytokines, would reveal information about the interactions between these systems. Additionally, animal models have been important for previous studies of the immune system in the oral cavity [56], and further studies in germ-free mice could be used to deduce the effects of the oral microbiota. The knowledge gained from combining the single-cell genomic techniques used in this dissertation with additional whole-community methods and host information could ultimately be used to devise treatments and preventative measures for a wide array of oral diseases.

References

1. Staley, J.T. and A. Konopka, *Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats*. Annual Review of Microbiology, 1985. **39**.
2. Woese, C.R., *Bacterial evolution*. Microbiological Reviews, 1987. **51**(2).
3. Woese, C.R. and G.E. Fox, *Phylogentic structure of prokaryotic domain - primary kingdoms*. Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(11).
4. Woese, C.R., et al., *Conservation of primary structure in 16S ribosomal RNA*. Nature, 1975. **254**(5495).
5. Pace, N.R., *A molecular view of microbial diversity and the biosphere*. Science, 1997. **276**(5313).
6. Hugenholtz, P. and N.R. Pace, *Identifying microbial diversity in the natural environment: A molecular phylogenetic approach*. Trends in Biotechnology, 1996. **14**(6).
7. Olsen, G.J., et al., *Microbial ecology and evolution - a ribosomal RNA approach*. Annual Review of Microbiology, 1986. **40**.
8. Handelsman, J., *Metagenomics: Application of genomics to uncultured microorganisms*. Microbiology and Molecular Biology Reviews, 2004. **68**(4).
9. Liu, Z., et al., *Short pyrosequencing reads suffice for accurate microbial community analysis*. Nucleic Acids Research, 2007. **35**(18).
10. Sogin, M.L., et al., *Microbial diversity in the deep sea and the underexplored "rare biosphere"*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(32).
11. Whitman, W.B., D.C. Coleman, and W.J. Wiebe, *Prokaryotes: The unseen majority*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(12).
12. Amann, R.L., W. Ludwig, and K.H. Schleifer, *Phylogenetic identification and in situ detection of individual microbial cells without cultivation*. Microbiological Reviews, 1995. **59**(1).
13. Hugenholtz, P., et al., *Investigation of Candidate Division TM7, a recently recognized major lineage of the Domain Bacteria with no known pure-culture representatives*. Applied and Environmental Microbiology, 2001. **67**(1): p. 411-419.
14. Rappe, M.S. and S.J. Giovannoni, *The uncultured microbial majority*. Annual Review of Microbiology, 2003. **57**.
15. Hugenholtz, P., B.M. Goebel, and N.R. Pace, *Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity*. Journal of Bacteriology, 1998. **180**(18): p. 4765-4774.
16. Lucker, S., et al., *A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(30): p. 13479-13484.
17. Elshahed, M.S., et al., *Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring*. Applied and Environmental Microbiology, 2003. **69**(9): p. 5609-5621.

18. Griffen, A.L., et al., *Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing*. The ISME Journal, 2011.
19. Turnbaugh, P.J., et al., *The human microbiome project*. Nature, 2007. **449**(7164).
20. Savage, D.C., *Microbial ecology of gastrointestinal tract*. Annual Review of Microbiology, 1977. **31**.
21. Lederberg, J. and A.T. McCray, 'Ome sweet 'omics - A genealogical treasury of words. Scientist, 2001. **15**(7): p. 8-8.
22. Peterson, J., et al., *The NIH human microbiome project*. Genome Research, 2009. **19**(12): p. 2317-2323.
23. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing*. Nature, 2010. **464**(7285).
24. Gevers, D., et al., *The human microbiome project: a community resource for the healthy human microbiome*. Plos Biology, 2012. **10**(8).
25. Human Microbiome Project, C., *A framework for human microbiome research*. Nature, 2012. **486**(7402).
26. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins*. Nature, 2009. **457**(7228).
27. Ley, R.E., et al., *Worlds within worlds: evolution of the vertebrate gut microbiota*. Nature Reviews Microbiology, 2008. **6**(10).
28. Nasidze, I., et al., *Global diversity in the human salivary microbiome*. Genome Research, 2009. **19**(4).
29. Huttenhower, C., et al., *Structure, function and diversity of the healthy human microbiome*. Nature, 2012. **486**(7402).
30. Hamady, M. and R. Knight, *Microbial community profiling for human microbiome projects: Tools, techniques, and challenges*. Genome Research, 2009. **19**(7).
31. Palmer, C., et al., *Development of the human infant intestinal microbiota*. Plos Biology, 2007. **5**(7).
32. Ley, R.E., et al., *Evolution of mammals and their gut microbes*. Science, 2008. **320**(5883): p. 1647-1651.
33. Gill, S.R., et al., *Metagenomic analysis of the human distal gut microbiome*. Science, 2006. **312**(5778).
34. Turnbaugh, P.J., et al., *The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice*. Science Translational Medicine, 2009. **1**(6).
35. Turnbaugh, P.J., et al., *An obesity-associated gut microbiome with increased capacity for energy harvest*. Nature, 2006. **444**(7122).
36. Wen, L., et al., *Innate immunity and intestinal microbiota in the development of Type 1 diabetes*. Nature, 2008. **455**(7216): p. 1109-U10.
37. Dicksved, J., et al., *Molecular analysis of the gut microbiota of identical twins with Crohn's disease*. Isme Journal, 2008. **2**(7): p. 716-727.
38. Lasken, R.S., *Genomic sequencing of uncultured microorganisms from single cells*. Nature Reviews Microbiology, 2012. **10**(9): p. 631-640.

39. Wylie, K.M., et al., *Novel bacterial taxa in the human microbiome*. Plos One, 2012. **7**(6).
40. Paster, B.J., et al., *Bacterial diversity in human subgingival plaque*. Journal of Bacteriology, 2001. **183**(12): p. 3770-3783.
41. Kroes, I., P.W. Lepp, and D.A. Relman, *Bacterial diversity within the human subgingival crevice*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(25).
42. Aas, J.A., et al., *Defining the normal bacterial flora of the oral cavity*. Journal of Clinical Microbiology, 2005. **43**(11).
43. Dewhirst, F.E., et al., *The human oral microbiome*. Journal of Bacteriology, 2010. **192**(19): p. 5002-5017.
44. Sizova, M.V., et al., *New Approaches for Isolation of Previously Uncultivated Oral Bacteria*. Applied and Environmental Microbiology, 2012. **78**(1).
45. Keijser, B.J.F., et al., *Pyrosequencing analysis of the oral microflora of healthy adults*. Journal of Dental Research, 2008. **87**(11).
46. Lazarevic, V., et al., *Study of inter- and intra-individual variations in the salivary microbiota*. BMC Genomics, 2010. **11**.
47. Zaura, E., et al., *Defining the healthy "core microbiome" of oral microbial communities*. BMC Microbiology, 2009. **9**.
48. Costello, E.K., et al., *Bacterial community variation in human body habitats across space and time*. Science, 2009. **326**(5960).
49. Methe, B.A., et al., *A framework for human microbiome research*. Nature, 2012. **486**(7402).
50. Humphrey, L.L., et al., *Periodontal disease and coronary heart disease incidence: a systematic review and meta-analysis*. Journal of General Internal Medicine, 2008. **23**(12): p. 2079-2086.
51. Kuo, L.C., A.M. Poison, and T. Kang, *Associations between periodontal diseases and systemic diseases: A review of the inter-relationships and interactions with diabetes, respiratory diseases, cardiovascular diseases and osteoporosis*. Public Health, 2008. **122**(4): p. 417-433.
52. Preshaw, P.M., et al., *Periodontitis and diabetes: a two-way relationship*. Diabetologia, 2012. **55**(1): p. 21-31.
53. Ramirez-Tortosa, M.C., et al., *Periodontitis is associated with altered plasma fatty acids and cardiovascular risk markers*. Nutrition Metabolism and Cardiovascular Diseases, 2010. **20**(2): p. 133-139.
54. Scannapieco FA, D.A., Chhun N, *Does periodontal therapy reduce the risk for systemic diseases?* Dent Clin North Am., 2010. **54**(1): p. 18.
55. Kinane, D., P. Bouchard, and P. European Workshop, *Periodontal diseases and health: consensus report of the sixth European workshop on periodontology*. Journal of Clinical Periodontology, 2008. **35**: p. 333-337.
56. Darveau, R.P., *Periodontitis: a polymicrobial disruption of host homeostasis*. Nature Reviews Microbiology, 2010. **8**(7): p. 481-490.
57. Marsh, P.D., *Microbial ecology of dental plaque and its significance in health and disease*. Advances in Dental Research, 1994. **8**(2): p. 263-271.

58. Socransky, S.S., et al., *Microbial complexes in subgingival plaque*. Journal of Clinical Periodontology, 1998. **25**(2).
59. Liu, B., et al., *Deep sequencing of the oral microbiome reveals signatures of periodontal disease*. Plos One, 2012. **7**(6).
60. Kapatral, V., et al., *Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586*. Journal of Bacteriology, 2002. **184**(7): p. 2005-2018.
61. Nelson, K.E., et al., *Complete genome sequence of the oral pathogenic bacterium *Porphyromonas gingivalis* strain W83*. Journal of Bacteriology, 2003. **185**(18): p. 5591-5601.
62. Seshadri, R., et al., *Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(15): p. 5646-5651.
63. Marcy, Y., et al., *Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(29).
64. Zhang, L., et al., *Whole genome amplification from a single cell - implications for genetic analysis*. Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(13).
65. Blanco, L., et al., *Highly efficient DNA synthesis by the phage phi29 DNA polymerase - symmetrical mode of DNA replication*. Journal of Biological Chemistry, 1989. **264**(15).
66. Blanco, L. and M. Salas, *Characterization and purification of a phage phi29-encoded DNA polymerase required for the initiation of replication*. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1984. **81**(17).
67. Dean, F.B., et al., *Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification*. Genome Research, 2001. **11**(6).
68. Raghunathan, A., et al., *Genomic DNA amplification from a single bacterium*. Applied and Environmental Microbiology, 2005. **71**(6).
69. Stepanauskas, R., *Single cell genomics: an individual look at microbes*. Current opinion in microbiology, 2012. **15**(5).
70. Podar, M., et al., *Targeted access to the genomes of low-abundance organisms in complex microbial communities*. Applied and Environmental Microbiology, 2007. **73**(10).
71. Youssef, N.H., et al., *Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma)*. Applied and Environmental Microbiology, 2011. **77**(21): p. 7804-7814.
72. Stepanauskas, R. and M.E. Sieracki, *Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(21): p. 9052-9057.

73. Woyke, T., et al., *Decontamination of MDA Reagents for Single Cell Whole Genome Amplification*. Plos One, 2011. **6**(10).
74. Blainey, P.C. and S.R. Quake, *Digital MDA for enumeration of total nucleic acid contamination*. Nucleic Acids Research, 2011. **39**(4).
75. Woyke, T., et al., *Assembling the Marine Metagenome, One Cell at a Time*. Plos One, 2009. **4**(4).
76. Woyke, T., et al., *One bacterial cell, one complete genome*. Plos One, 2010. **5**(4).
77. Swan, B.K., et al., *Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean*. Science, 2011. **333**(6047): p. 1296-1300.
78. Blainey, P.C., et al., *Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis*. Plos One, 2011. **6**(2).
79. Martinez-Garcia, M., et al., *Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia*. PloS one, 2012. **7**(4).
80. Fleming, E.J., et al., *What's new Is old: resolving the identity of Leptothrix ochracea using single cell genomics, pyrosequencing and FISH*. Plos One, 2011. **6**(3).
81. Mason, O.U., et al., *Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill*. Isme Journal, 2012. **6**(9).
82. Heywood, J.L., et al., *Capturing diversity of marine heterotrophic protists: one cell at a time*. Isme Journal, 2011. **5**(4).
83. Martinez-Garcia, M., et al., *Unveiling in situ interactions between marine protists and bacteria through single cell sequencing*. Isme Journal, 2012. **6**(3).
84. Lasken, R.S. and T.B. Stockwell, *Mechanism of chimera formation during the Multiple Displacement Amplification reaction*. BMC biotechnology, 2007. **7**.
85. Brown, T., Howe, Adina, Zhang, Qingpeng, Pyrkosz, Alexis B. and Timothy H. Brom, *A reference-free algorithm for computational normalization of shotgun sequencing data*. 2012.
86. Amann, R.L., et al., *Combination of 16S ribosomal RNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations*. Applied and Environmental Microbiology, 1990. **56**(6): p. 1919-1925.
87. Roller, C., et al., *In situ probing of gram positive bacteria with high G+C content using 23S ribosomal RNA targeted oligonucleotides*. Microbiology-Uk, 1994. **140**: p. 2849-2858.
88. Wallner, G., R. Amann, and W. Beisker, *Optimizing fluorescent in situ hybridization with ribosomal RNA targeted oligonucleotide probes for flow cytometric identification of microorganisms*. Cytometry, 1993. **14**(2).

89. Amann, R., B.M. Fuchs, and S. Behrens, *The identification of microorganisms by fluorescence in situ hybridisation*. Current Opinion in Biotechnology, 2001. **12**(3).
90. Fuchs, B.M., J. Pernthaler, and R. Amann, *Single cell identification by fluorescence in situ hybridization*, in *Methods for General and Molecular Microbiology*, C.A. Reddy, et al., Editors. 2007, ASM Press: Washington, D.C. p. 886-896.
91. Rodrigue, S., et al., *Whole genome amplification and de novo assembly of single bacterial cells*. Plos One, 2009. **4**(9).
92. Weisburg, W.G., et al., *16S ribosomal DNA amplification for phylogenetic study*. Journal of Bacteriology, 1991. **173**(2): p. 697-703.
93. Horz, H.P., et al., *Synergistes group organisms of human origin*. Journal of Clinical Microbiology, 2006. **44**(8): p. 2914-2920.
94. Rocas, I.N. and J.F. Siqueira, *Detection of novel oral species and phylotypes in symptomatic endodontic infections including abscesses*. Fems Microbiology Letters, 2005. **250**(2): p. 279-285.
95. Vianna, M.E., et al., *Quantification and characterization of Synergistes in endodontic infections*. Oral Microbiology and Immunology, 2007. **22**(4): p. 260-265.
96. Fodor, A.A., et al., *The "most wanted" taxa from the human microbiome for whole genome sequencing*. Plos One, 2012. **7**(7).
97. Garmendia, C., et al., *The bacteriophage phi29 DNA polymerase, a proofreading enzyme*. Journal of Biological Chemistry, 1992. **267**(4).
98. Tamariz, J., et al., *The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA*. Journal of Forensic Sciences, 2006. **51**(4): p. 790-794.
99. Drummond, A.J., et al., *Geneious Pro V.5.6.3*, 2011.
100. Wang, Q., et al., *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy*. Applied and Environmental Microbiology, 2007. **73**(16): p. 5261-5267.
101. Shagin, D.A., et al., *A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas*. Genome Research, 2002. **12**(12): p. 1935-1942.
102. Bankevich, A., et al., *SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing*. Journal of Computational Biology, 2012. **19**(5).
103. Chitsaz, H., et al., *Efficient de novo assembly of single-cell bacterial genomes from short-read data sets*. Nature Biotechnology, 2011. **29**(10).
104. Peng, Y., et al., *IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth*. Bioinformatics, 2012. **28**(11): p. 1420-1428.
105. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs*. Genome Research, 2008. **18**(5): p. 821-829.
106. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification*. BMC Bioinformatics, 2010. **11**.

107. Altschul, S.F., et al., *Basic Local Alignment Search Tool*. Journal of Molecular Biology, 1990. **215**(3): p. 403-410.
108. Woyke, T., et al., *Assembling the marine metagenome, one cell at a time*. Plos One, 2009. **4**(4).
109. Markowitz, V.M., et al., *IMG: the integrated microbial genomes database and comparative analysis system*. Nucleic Acids Research, 2012. **40**(D1): p. D115-D122.
110. Aziz, R.K., et al., *The RAST server: Rapid annotations using subsystems technology*. BMC Genomics, 2008. **9**.
111. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Research, 2000. **28**(1): p. 27-30.
112. Tatusov, R.L., et al., *The COG database: an updated version includes eukaryotes*. BMC Bioinformatics, 2003. **4**.
113. Punta, M., et al., *The Pfam protein families database*. Nucleic Acids Research, 2012. **40**(D1): p. D290-D301.
114. Richter, M. and R. Rossello-Mora, *Shifting the genomic gold standard for the prokaryotic species definition*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(45): p. 19126-19131.
115. Darling, A.C.E., et al., *Mauve: Multiple alignment of conserved genomic sequence with rearrangements*. Genome Research, 2004. **14**(7): p. 1394-1403.
116. Eddy, S.R., *Accelerated profile HMM searches*. Plos Computational Biology, 2011. **7**(10).
117. Kumar, P.S., et al., *Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis*. Journal of Clinical Microbiology, 2005. **43**(8): p. 3944-3955.
118. Boopathy, R., et al., *Activity of sulfate-reducing bacteria in human periodontal pocket*. Canadian Journal of Microbiology, 2002. **48**(12): p. 1099-1103.
119. Langendijk, P.S., J.T.J. Hanssen, and J.S. Van der Hoeven, *Sulfate-reducing bacteria in association with human periodontitis*. Journal of Clinical Periodontology, 2000. **27**(12): p. 943-950.
120. Vianna, M.E., et al., *Quantitative analysis of three hydrogenotrophic microbial groups, methanogenic archaea, sulfate-reducing bacteria, and acetogenic bacteria, within plaque biofilms associated with human periodontal disease*. Journal of Bacteriology, 2008. **190**(10): p. 3779-3785.
121. Wade, W.G., et al., *Specificity of the oral microflora in dentinal caries, endodontic infections and periodontitis*. International Congress Series, 2005. **1284**(0): p. 150-157.
122. Kumar, P.S., et al., *New bacterial species associated with chronic periodontitis*. Journal of Dental Research, 2003. **82**(5): p. 338-344.
123. Pagani, I., et al., *Complete genome sequence of Desulfobulbus propionicus type strain (1pr3(T))*. Standards in Genomic Sciences, 2011. **4**(1): p. 100-110.
124. Langendijk, P.S., et al., *Isolation of Desulfomicrobium orale sp nov and Desulfovibrio strain NY682, oral sulfate-reducing bacteria involved in*

- human periodontal disease*. International Journal of Systematic and Evolutionary Microbiology, 2001. **51**: p. 1035-1044.
125. Loubinoux, J., et al., *Bacteremia caused by a strain of Desulfovibrio related to the provisionally named Desulfovibrio fairfieldensis* (vol 38, pg 931, 2000). Journal of Clinical Microbiology, 2000. **38**(4): p. 1707-1707.
 126. Pimentel, J.D. and R.C. Chan, *Desulfovibrio fairfieldensis Bacteremia associated with choledocholithiasis and endoscopic retrograde Cholangiopancreatography*. Journal of Clinical Microbiology, 2007. **45**(8): p. 2747-2750.
 127. Loy, A., et al., *Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment*. Applied and Environmental Microbiology, 2002. **68**(10): p. 5064-5081.
 128. Lucker, S., et al., *Improved 16S rRNA-targeted probe set for analysis of sulfate-reducing bacteria by fluorescence in situ hybridization*. Journal of Microbiological Methods, 2007. **69**(3): p. 523-528.
 129. Stoecker, K., et al., *Double labeling of oligonucleotide probes for fluorescence in situ hybridization (DOPE-FISH) improves signal intensity and increases rRNA accessibility*. Applied and Environmental Microbiology, 2010. **76**(3): p. 922-926.
 130. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*. Systematic Biology, 2010. **59**(3): p. 307-321.
 131. Loubinoux, J., et al., *Isolation of the provisionally named Desulfovibrio fairfieldensis from human periodontal pockets*. Oral Microbiology and Immunology, 2002. **17**(5): p. 321-323.
 132. Pires, R.H., et al., *Characterization of the Desulfovibrio desulfuricans ATCC 27774 DsrMKJOP complex - A membrane-bound redox complex involved in the sulfate respiratory pathway*. Biochemistry, 2006. **45**(1): p. 249-262.
 133. Tasaki, M., et al., *Acetogenesis from pyruvate by Desulfotomaculum thermobenzoicum and differences in pyruvate metabolism among 3 sulfate-reducing bacteria in the absence of sulfate*. FEMS Microbiology Letters, 1993. **106**(3): p. 259-264.
 134. Van Melderren, L. and M. Saavedra De Bast, *Bacterial toxin antitoxin systems: more than selfish entities?* PLoS Genet, 2009. **5**(3): p. e1000437.
 135. Widdel, F. and N. Pfennig, *Studies on dissimilatory sulfate-reducing bacteria that decompose fatty-acids II. Incomplete oxidation of propionate by Desulfobulbus propionicus gen. nov., sp. nov.* Archives of Microbiology, 1982. **131**(4): p. 360-365.
 136. Mattick, J.S., *Type IV pili and twitching motility*. Annual Review of Microbiology, 2002. **56**: p. 289-314.
 137. Li, Y.X., et al., *Type I and type IV pili of Xylella fastidiosa affect twitching motility, biofilm formation and cell-cell aggregation*. Microbiology-Sgm, 2007. **153**: p. 719-726.

138. Cascales, E. and P.J. Christie, *The versatile bacterial type IV secretion systems*. Nature Reviews Microbiology, 2003. **1**(2): p. 137-149.
139. Clock, S.A., et al., *Outer membrane components of the tad (tight adherence) secreton of Aggregatibacter actinomycetemcomitans*. Journal of Bacteriology, 2008. **190**(3): p. 980-990.
140. Fernandez, L., I. Marquez, and J.A. Guijarro, *Identification of specific in vivo-induced (ivi) genes in Yersinia ruckeri and analysis of ruckerbactin, a catecholate siderophore iron acquisition system*. Applied and Environmental Microbiology, 2004. **70**(9): p. 5199-5207.
141. Fuller, T.E., M.J. Kennedy, and D.E. Lowery, *Identification of Pasteurella multocida virulence genes in a septicemic mouse model using signature-tagged mutagenesis*. Microbial Pathogenesis, 2000. **29**(1): p. 25-38.
142. Nikaido, H., *Multidrug efflux pumps of gram-negative bacteria*. Journal of Bacteriology, 1996. **178**(20): p. 5853-5859.
143. Dannenberg, S., et al., *Oxidation of H₂, organic compounds and inorganic sulfur compounds coupled to reduction of O₂ or nitrate by sulfate-reducing bacteria*. Archives of Microbiology, 1992. **158**(2): p. 93-99.
144. Cypionka, H., *Oxygen respiration by Desulfovibrio species*. Annual Review of Microbiology, 2000. **54**: p. 827-848.
145. Sun, H., et al., *Complete genome sequence of Desulfarculus baarsii type strain (2st14(T))*. Standards in Genomic Sciences, 2010. **3**(3): p. 276-284.
146. Wang, R.K., et al., *The influence of iron availability on human salivary microbial community composition*. Microbial Ecology, 2012. **64**(1): p. 152-161.
147. Capestany, C.A., et al., *Role of the Clp system in stress tolerance, biofilm formation, and intracellular invasion in Porphyromonas gingivalis*. Journal of Bacteriology, 2008. **190**(4): p. 1436-1446.
148. Lourdault, K., et al., *Inactivation of clpB in the pathogen Leptospira interrogans reduces virulence and resistance to stress conditions*. Infection and Immunity, 2011. **79**(9): p. 3711-3717.
149. de Oliveira, N.E.M., et al., *clpB, a class III heat-shock gene regulated by CtsR, is involved in thermotolerance and virulence of Enterococcus faecalis*. Microbiology-Sgm, 2011. **157**: p. 656-665.
150. Loughlin, M.F., et al., *Helicobacter pylori mutants defective in the clpP ATP-dependant protease and the chaperone clpA display reduced macrophage and murine survival*. Microbial Pathogenesis, 2009. **46**(1): p. 53-57.
151. Winther-Larsen, H.C., et al., *Neisseria gonorrhoeae PilV, a type IV pilus-associated protein essential to human epithelial cell adherence*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(26).
152. Ajdic, D., et al., *Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(22): p. 14434-14439.

153. Roberts, I.S., *The biochemistry and genetics of capsular polysaccharide production in bacteria*. Annual Review of Microbiology, 1996. **50**: p. 285-315.
154. An, F.Y. and D.B. Clewell, *Characterization of the determinant (traB) encoding sex pheromone shutdown by the hemolysin bacteriocin plasmid pAD1 in Enterococcus faecalis*. Plasmid, 1994. **31**(2): p. 215-221.
155. Reddy, M., *Role of FtsEX in cell division of Escherichia coli: viability of ftsEX mutants is dependent on functional Sufl or high osmotic strength*. Journal of Bacteriology, 2007. **189**(1): p. 98-108.
156. Worley, M.J., K.H.L. Ching, and F. Heffron, *Salmonella SsrB activates a global regulon of horizontally acquired genes*. Molecular Microbiology, 2000. **36**(3): p. 749-761.
157. Garcia-Calderon, C.B., J. Casadesus, and F. Ramos-Morales, *Rcs and PhoPQ regulatory overlap in the control of Salmonella enterica virulence*. Journal of Bacteriology, 2007. **189**(18): p. 6635-6644.
158. Frye, J., et al., *Identification of new flagellar genes of Salmonella enterica serovar typhimurium*. Journal of Bacteriology, 2006. **188**(6): p. 2233-2243.
159. Fink, R.C., et al., *FNR is a global regulator of virulence and anaerobic metabolism in Salmonella enterica serovar typhimurium (ATCC 14028s)*. Journal of Bacteriology, 2007. **189**(6): p. 2262-2273.
160. Meng, G.Y., et al., *Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation*. Embo Journal, 2011. **30**(18): p. 3864-3874.
161. Mougous, J.D., et al., *Threonine phosphorylation post-translationally regulates protein secretion in Pseudomonas aeruginosa*. Nature Cell Biology, 2007. **9**(7): p. 797-U121.
162. Casabona, M.G., et al., *An ABC transporter and an outer membrane lipoprotein participate in posttranslational activation of type VI secretion in Pseudomonas aeruginosa*. Environmental microbiology, 2012: p. no-no.
163. Bjornsson, L., et al., *Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal*. Microbiology-Sgm, 2002. **148**: p. 2309-2318.
164. Miura, Y., Y. Watanabe, and S. Okabe, *Significance of Chloroflexi in performance of submerged membrane Bioreactors (MBR) treating municipal wastewater*. Environmental Science & Technology, 2007. **41**(22): p. 7787-7794.
165. Schmitt, S., et al., *Chloroflexi bacteria are more diverse, abundant, and similar in high than in low microbial abundance sponges*. Fems Microbiology Ecology, 2011. **78**(3): p. 497-510.
166. Tang, K.H., et al., *Complete genome sequence of the filamentous anoxygenic phototrophic bacterium Chloroflexus aurantiacus*. BMC Genomics, 2011. **12**.
167. Yamada, T. and Y. Sekiguchi, *Cultivation of uncultured Chloroflexi subphyla: significance and ecophysiology of formerly uncultured Chloroflexi 'subphylum I' with natural and biotechnological relevance*. Microbes and Environments, 2009. **24**(3): p. 205-216.

168. de Lillo, A., et al., *Novel subgingival bacterial phylotypes detected using multiple universal polymerase chain reaction primer sets*. Oral Microbiology and Immunology, 2006. **21**(1): p. 61-68.
169. Dewhirst, F.E., et al., *The canine oral microbiome*. Plos One, 2012. **7**(4).
170. Kong, Y.H., R. Teather, and R. Forster, *Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages*. Fems Microbiology Ecology, 2010. **74**(3): p. 612-622.
171. Krasilnikova, E.N. and E.N. Kondrateva, *Growth of Chloroflexus aurantiacus under anaerobic conditions in the dark and the metabolism of organic substrates*. Microbiology, 1987. **56**(3): p. 281-285.
172. Sirevag, R. and R. Castenholz, *Aspects of carbon metabolism in Chloroflexus*. Archives of Microbiology, 1979. **120**(2): p. 151-153.
173. Graentzdoerffer, A., et al., *Molecular and biochemical characterization of two tungsten- and selenium-containing formate dehydrogenases from Eubacterium acidaminophilum that are associated with components of an iron-only hydrogenase*. Archives of Microbiology, 2003. **179**(2): p. 116-130.
174. Zarzycki, J. and G. Fuchs, *Coassimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in Chloroflexus aurantiacus*. Applied and Environmental Microbiology, 2011. **77**(17): p. 6181-6188.
175. Sekiguchi, Y., et al., *Anaerolinea thermophila gen. nov., sp nov and Caldilinea aerophila gen. nov., sp nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level*. International Journal of Systematic and Evolutionary Microbiology, 2003. **53**: p. 1843-1851.
176. Yamada, T., et al., *Anaerolinea thermolimosa sp nov., Levilinea saccharolytica gen. nov., sp nov and Leptolinea tardivitalis gen. nov., so. nov., novel filamentous anaerobes, and description of the new classes anaerolineae classis nov and Caldilineae classis nov in the bacterial phylum Chloroflexi*. International Journal of Systematic and Evolutionary Microbiology, 2006. **56**: p. 1331-1340.
177. Plumbridge, J. and E. Vimr, *Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli*. Journal of Bacteriology, 1999. **181**(1): p. 47-54.
178. Kindaichi, T., T. Ito, and S. Okabe, *Ecophysiological interaction between nitrifying bacteria and heterotrophic bacteria in autotrophic nitrifying biofilms as determined by microautoradiography-fluorescence in situ hybridization*. Applied and Environmental Microbiology, 2004. **70**(3): p. 1641-1650.
179. Miura, Y. and S. Okabe, *Quantification of cell specific uptake activity of microbial products by uncultured Chloroflexi by microautoradiography combined with fluorescence in situ hybridization*. Environmental Science & Technology, 2008. **42**(19): p. 7380-7386.

180. Okabe, S., T. Kindaichi, and T. Ito, *Fate of C-14-labeled microbial products derived from nitrifying bacteria in autotrophic nitrifying biofilms*. Applied and Environmental Microbiology, 2005. **71**(7): p. 3987-3994.
181. Zang, K., et al., *Analysis of the phylogenetic diversity of estrone-degrading bacteria in activated sewage sludge using microautoradiography-fluorescence in situ hybridization*. Systematic and Applied Microbiology, 2008. **31**(3): p. 206-214.
182. Stafford, G., et al., *Sialic acid, periodontal pathogens and Tannerella forsythia: stick around and enjoy the feast!* Molecular Oral Microbiology, 2012. **27**(1): p. 11-22.
183. Zhang, W.Y., et al., *Comparative analysis of iol clusters in Lactobacillus casei strains*. World Journal of Microbiology & Biotechnology, 2010. **26**(11): p. 1949-1955.
184. Kawsar, H.I., et al., *Organization and transcriptional regulation of myo-inositol operon in Clostridium perfringens*. Fems Microbiology Letters, 2004. **235**(2): p. 289-295.
185. Kroger, C., J. Stolz, and T.M. Fuchs, *myo-Inositol transport by Salmonella enterica serovar Typhimurium*. Microbiology-Sgm, 2010. **156**: p. 128-138.
186. Frosch, M. and A. Muller, *Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in Neisseria meningitidis*. Molecular Microbiology, 1993. **8**(3): p. 483-493.
187. Lin, W.S., T. Cunneen, and C.Y. Lee, *Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in Staphylococcus aureus*. Journal of Bacteriology, 1994. **176**(22): p. 7005-7016.
188. Lilburn, T.C., et al., *Nitrogen fixation by symbiotic and free-living spirochetes*. Science, 2001. **292**(5526): p. 2495-2498.
189. Tyson, G.W., et al., *Genome-directed isolation of the key nitrogen fixer Leptospirillum ferroplasma sp nov from an acidophilic microbial community*. Applied and Environmental Microbiology, 2005. **71**(10).
190. Bomar, L., et al., *Directed culturing of microorganisms using metatranscriptomics*. Mbio, 2011. **2**(2): p. 8.
191. Zengler, K., *Central role of the cell in microbial ecology*. Microbiology and Molecular Biology Reviews, 2009. **73**(4).
192. Porter, J., et al., *Rapid, automated separation of specific bacteria from lake water and sewage by flow cytometry and cell sorting*. Applied and Environmental Microbiology, 1993. **59**(10).
193. Ferrari, B.C., G. Oregaard, and S.J. Sorensen, *Recovery of GFP-labeled bacteria for culturing and molecular analysis after cell sorting using a benchtop flow cytometer*. Microbial Ecology, 2004. **48**(2).
194. Bosshardt, D.D. and N.P. Lang, *The junctional epithelium: from health to disease*. Journal of Dental Research, 2005. **84**(1): p. 9-20.

Vita

Alisha Gail Campbell earned a Bachelor of Science in Biochemistry, Cellular and Molecular Biology from the University of Tennessee, Knoxville in May 2006 under the supervision of honor's thesis advisor Sundar Venkatachalam. She worked as a laboratory analyst for Clean Air Laboratories until August 2008 when she entered the Genome Science and Technology program at the University of Tennessee, Knoxville. She completed her dissertation work under the supervision of Mircea Podar and received her Ph.D. in May 2013.