2015

Clinical and Genomic Characterization of Two Vaginal Megasphaera Species

Abigail L. Glascock
Virginia Commonwealth University

CLINICAL AND GENOMIC CHARACTERIZATION OF VAGINAL *MEGASPHAERA* SPECIES

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by

ABIGAIL LEIGH GLASCOCK
Bachelor of Science, James Madison University, 2010

Director: JENNIFER M. FETTWEIS, PH.D.
ASSISTANT PROFESSOR, CENTER FOR THE STUDY OF BIOLOGICAL COMPLEXITY,
DEPARTMENT OF OBSTETRICS AND GYNECOLOGY, SCHOOL OF MEDICINE

Virginia Commonwealth University
Richmond, Virginia
December 2015

Acknowledgment

First and foremost, I would like to extend my sincerest gratitude to my advisor, Dr. Jennifer M. Fettweis. Dr. Fettweis has served as a dedicated and caring mentor to me since I first arrived at VCU in 2010 as a laboratory technician. She has sacrificed countless hours of her time and spent many late nights in the lab discussing my research, reviewing my writing, scientific posters and presentations, and guiding me towards a successful and fulfilling science career. She has been supportive and proactive every step of the way in helping me to become a better scientist. Most importantly, Dr. Fettweis has served as an excellent role model to me as a young woman and scientist with aspirations of becoming an independent researcher. Her dedication to the quality of her work and science as a whole is truly inspiring. I again extend my thanks to her for accepting me as her graduate student and mentee.

I am also deeply thankful to my committee members Dr. Gregory A. Buck and Dr. Brian C. Verrelli. Dr. Buck first accepted me into his laboratory as an employee and has since fostered my development as a scientist and encouraged my pursuit of further education. He has been heavily involved in the review of my writings, presentations and posters and can always be counted on for meaningful and constructive advice. Dr. Buck has taught me how to think in

terms of the big picture and how to view our work within that perspective. Dr. Verrelli has been an invaluable student advocate for myself and others since arriving at VCU. I am grateful for his openness, kindness, valuable scientific and professional advice and support during the past few years. Dr. Verrelli has also helped spark my interest in evolutionary biology in all realms of life, for which I am especially grateful. As Dobzhanksy stated, "Nothing in biology makes sense, except in the light of evolution."

I would also like to extend my deepest gratitude to the members of the Vaginal Microbiome Consortium and members of the Buck and Jefferson Labs. I thank Dr. Kimberly K. Jefferson and Melissa Prestosa for their aid in culturing *Megasphaera* isolates. I also thank the VCU Nucleic Acids Research Facilities for their help in generating data for this study. I would like to recognize Dr. Vishal Koparde for his development of the Genome Annotation Pipeline used herein and Shaun Norris for his help in retrieving data for the study. Special thanks to Dr. Bernice Huang, Dr. Anita Marinelli, Dr. Jennifer I. Drake, Dr. Myrna G. Serrano, Dr. J. Paul Brooks and Dr. Sophonie Jean for their advice and expertise in helping me to navigate this project.

I would also like to recognize the excellent professors and mentors that I have had in my time as a student. Without their dedication and willingness to share their extensive knowledge the analyses performed herein would not have been possible. Thank you to Dr. Paul M. Fawcett, Dr. Andrew J. Eckert, Dr. Maria C. Rivera, Dr. William B. Eggleston, Dr. Peter H. Uetz, Dr. Herschell S. Emery and Dr. Allison A. Johnson. I would also like to thank my fellow students in the Buck and Fettweis Labs for their friendship and advice. Thank you to Niels C. Asmussen, Katie R. Bradwell, Dr. Preethi R. Iyengar, Sarah K. Rozycki, William K. Potts, Nicole R. Jimenez, Dylan N. David, Quy-Hien R. Dang and Suryanaren “Sunny” Kummarapurugu.

Lastly, I would like to thank my parents Mat and Jeanya Wylie, my sister Anna Jane Glascock, my husband Justain Murphy and my dogs Dakota and Sadie for keeping me sane, encouraging me to pursue my dreams and never sell myself short and providing me with continuous love and support throughout my time as a graduate student. I certainly never could have made it to this point in my life and career without your support and I am forever grateful and honored to have you in my life. Thank you.

Table of Contents

Page

List of Tables

List of Figures

List of Abbreviations

VaHMP	Vaginal Human Microbiome Project
DNA	Deoxyribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
rDNA	Ribosomal Deoxyribonucleic Acid
PCR	Polymerase Chain Reaction
VCU	Virginia Commonwealth University
RDP	Ribosomal Database Project
STIRRUPS	Species-level Taxon Identification using a USEARCH Pipeline Strategy
OTU	Operational Taxonomic Unit
mL	milliliter
sBHI	Supplemented Brain-Heart Infusion
NCBI	National Center for Biotechnology Information

SNP	Single Nucleotide Polymorphism
BV	Bacterial Vaginosis
RAST	Rapid Annotation using Subsystem Technology
NMPDR	National Microbial Pathogen Data Resource
NARF	Nucleic Acids Research Facilities
tRNA	Transfer Ribonucleic Acid
COG	Clusters of Orthologous Groups
EMBL	European Molecular Biology Laboratory
EBI	European Bioinformatics Institute
EMBOSS	European Molecular Biology Open Software Suite
ALTER	ALignment Transformation EnviRonment
MUSCLE	MUltiple Sequence Comparison Log-Expectation
RAxML- HPC	Randomized Axelerated Maximum Likelihood- High Performance Computing
WAG	Whelan And Goldman
bp	base pairs
TBE	Tris-Borate EDTA Buffer

g	grams
Mb	megabases

Abstract

CLINICAL AND GENOMIC CHARACTERIZATION OF VAGINAL *MEGAPHAERA*

SPECIES

by

Abigail Leigh Glascock

B.S. Biology, James Madison University, 2010

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2015

Major Director: Jennifer M. Fettweis, Ph.D.

Assistant Professor, Center for the Study of Biological Complexity, VCU Life Sciences, Department of Obstetrics and Gynecology, School of Medicine

Two vaginal phylotypes of the genus *Megasphaera* (phylotype 1 and phylotype 2) were recently associated with bacterial vaginosis (BV), an infection characterized by vaginal dysbiosis. Through an analysis of 16S rRNA profiles of 3,986 women enrolled in the Vaginal Human Microbiome Project, we confirmed that while both phylotypes were associated with BV, *Megaspheara* phylotype 1 had higher specificity for the condition. *Megasphaera* phylotype 2 was strongly associated with trichomoniasis. Previous studies have reported that BV-associated organisms are excluded in pregnancy. We observed that *Megasphaera* phylotype 1, which has been associated with adverse pregnancy outcomes, exhibited a trend of increased prevalence in the pregnant cohort. We sequenced the genomes of isolates of the two phylotypes and performed comparative analyses. We demonstrate that these two phylotypes have distinct genomic features and unique potential for metabolic processes that reveal niche specialization. These findings may provide insight into their differential associations with vaginal infections.

INTRODUCTION

The vaginal microbiome and women's health

Our bodies are home to trillions of bacteria, collectively known as the human microbiome. The bacterial cells that inhabit our bodies outnumber our human cells ten to one and thus play an important role in our health and well-being (Boleij & Tjalsma, 2012). Although research involving host-related bacteria, both commensal and pathogenic, has been ongoing for centuries, the characterization of entire bacterial communities in and on the body is a developing field, which holds great promise for furthering the understanding of bacterial contributions to human health and host-microbe interactions. The vaginal microbiome is of particular interest given its associations with the transmission and acquisition of sexually transmitted diseases, role in women's reproductive health and fertility, and implications in pregnancy and neonatal health. Over the past several years, many groups have performed studies to characterize the bacterial composition and structure of vaginal microbiome (Brown et al., 2007; Fettweis, Serrano, Girerd, Jefferson, & Buck, 2012a; Kiss et al., 2007; Ravel et al., 2011).

The functionality of the human microbiome has been found to differ by body site, and the model of optimal health in the vaginal microbiome is different from the paradigm for the gut microbiome (Human Microbiome Project Consortium, 2012). Gut microbiome research is rapidly developing and has already yielded new therapeutic strategies (Barbut, Collignon, Butel, & Bourlioux, 2015; Gibson, McCartney, & Rastall, 2005; Rao & Samak, 2013). Thousands of different bacterial species inhabit the gut and many likely play a role in keeping the human gut operating as a functional habitat. Loss of bacterial species often causes a loss of function in the gut and loss of diversity can result in dire health consequences such as the development of inflammatory bowel diseases and even death in some severe cases, such as infections with the opportunistic pathogen *Clostridium difficile* (Jeffery, Lynch, & O'Toole, 2015; Ohkusa & Koido, 2015; Pérez-Cobas et al., 2014; Segata, 2015). The vaginal microbiome structure is strikingly different. An individual's vaginal microbiome is composed of tens of different species, instead of hundreds, and is often completely dominated by a single bacterial species. Dominant vaginal bacterial species are often members of the genus *Lactobacillus*, a lactic-acid producing taxon that has frequently been associated with reproductive and vaginal health (Ravel et al., 2011).

Albert Döderlein, a German gynecologist, published a monograph in 1892, detailing a study of the vaginal secretions of 200 pregnant women. He discovered the presence of a vaginal bacillus that was capable of acidifying the vaginal environment and preventing the colonization of pathogenic bacteria. Döderlein's bacillus was later classified as *Lactobacillus acidophilus* in 1928, and has since been found to represent many different vaginal *Lactobacillus* species including *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus jensenii* and *Lactobacillus iners* (Cruickshank, 1931). Given these findings and several further publications supporting them, *Lactobacillus* species are thought to serve a protective function in the vaginal microbiome, preventing the colonization of anaerobic bacteria, pathogens and sexually transmitted organisms through a number of mechanisms including lactic acid production, hydrogen peroxide production, secretion of bacteriocins and adhesion to the vaginal epithelium (Balkus et al., 2012; Kaewsrichan, Peeyananjarassri, & Kongprasertkit, 2006; O'Hanlon, Moench, & Cone, 2013; Ortiz, Ruiz, Pascual, & Barberis, 2014; Wilks et al., 2004).

It is thought that the optimal vaginal microbiome is one dominated by *Lactobacillus* species with few other bacterial species present. Vaginal pH is often used as a measure of vaginal community health due to the fact that presence of *Lactobacillus* species results in a decrease in pH caused by the production of D- and L- isomers of lactic acid and hydrogen peroxide (Balkus et al., 2012; O'Hanlon et al., 2013; Wilks et al., 2004; Witkin et al., 2013). A "healthy" vaginal pH is characterized as being less than or equal to 4.5. However, this standard is based on previous literature, which has largely characterized women of European ancestry. Recent work characterizing the vaginal communities of women of different ethnic backgrounds has revealed that this estimate may be too low (Fettweis, Brooks, et al., 2014a). Clinically healthy women of African ancestry have higher vaginal pH on average than women of European ancestry. All species of *Lactobacillus* are not equivalent in their ability to protect against colonization of other organisms. *L. crispatus, L. gasseri, L. jensenii* and others are often the most dominant organisms in vaginal samples. *L. iners* is a common constituent in the vaginal microbiome that sometimes dominates the vaginal microbiome and can coexist with anaerobic bacteria and bacterial vaginosis-associated organisms (Ravel et al., 2011; Verstraelen et al., 2009). A mechanism clarifying why *L. iners* is less protective than other *Lactobacillus* species has yet to be elucidated.

Traditionally, the presence of anaerobic bacteria in the vaginal microbiome such as *Gardnerella vaginalis* has been thought of as a marker of vaginal dysbiosis and poor vaginal health (Africa, Nel, & Stemmet, 2014; Brotman, 2011; Donders et al., 2009). Anaerobic species are not able to flourish in the presence of protective *Lactobacillus* species and thus their presence also signals the lack of these protective species. Anaerobic species are also associated with increased vaginal pH, a common marker of bacterial vaginosis. Many of these taxa such as *Sneathia, Atopobium, Mobiluncus,* BVAB1, BVAB2, BVAB3 and *Megasphaera* have been strongly associated with clinical diagnosis of bacterial vaginosis, the most common vaginal infection across the world (Allsworth & Peipert, 2007; Fethers et al., 2012; Harwich et al., 2012; Malaguti, Bahls, Uchimura, Gimenes, & Consolaro, 2015; Zozaya-Hinchliffe, Martin, & Ferris, 2008). Bacterial vaginosis is characterized by a shift from a *Lactobacillus*-dominated community to one characterized by the presence of anaerobic species, a strong amine odor, vaginal discharge and vaginal itching. Bacterial vaginosis (BV) often goes undiagnosed as it is not an inflammatory condition and can be asymptomatic (Allsworth & Peipert, 2007). BV also has a high recurrence rate due to the standard of care treatment with metronidazole that rarely causes a permanent shift to a *Lactobacillus*-dominated state (Bradshaw et al., 2006). That said, BV is important to treat because it has been associated with increased risk of sexually transmitted infections including HIV, pelvic inflammatory disease, and reproductive and obstetric disorders including preterm birth (Cohen et al., 2012; Gallo et al., 2012; Leitich et al., 2003; Purwar, Ughade, Bhagat, Agarwal, & Kulkarni, 2001).

The cause of bacterial vaginosis has not been clarified due to its inability to fit Koch's postulates. In the mid-20th century the organism *G. vaginalis* was thought to serve as the causative agent, but it has since been shown that *G. vaginalis* can exist in the vaginal communities of healthy women (Fettweis, Brooks, et al., 2014a; Leppäluoto, 2011). It seems that BV is associated with a loss of the protective functionality of the vaginal microbiome, via a shift in the taxa present in the community. Further vaginal microbiome research has revealed many more bacterial associations with diseases, which are not causative, but may potentially increase risk for transmission or acquisition of the disease. Examples of these findings include the association of *Mycoplasma hominis* and "*Ca.* Mycoplasma girerdii" with trichomoniasis, and the association of various anaerobic genera including *Atopobium*, *Mobiluncus*, and *Megasphaera* with bacterial vaginosis (Fethers et al., 2012; Fettweis, Serrano, et al., 2014b; Rappelli et al., 2001; Zozaya-Hinchliffe et al., 2008). From these association studies, we can see how understanding the underlying structure of the vaginal microbiome and the role that each organism plays in the community may hold valuable information about which organisms determine overall reproductive health, which organisms increase the risk of acquiring or transmitting which diseases and which organisms confer a lessened or heightened risk of complications during pregnancy.

Megasphaera

Megasphaera was first described in 1959 by Gutierrez *et al*., who classified isolates from the bovine and ovine rumen as members of the genus *Peptostreptococcus* (GUTIERREZ, DAVIS, LINDAHL, & WARWICK, 1959). In 1971, Rogosa *et al.* published an article renaming the organisms *Megasphaera elsdenii* and creating the new genus name *Megasphaera* due to the morphology of the cells observed while viewing them under the microscope (Fig.1). *Megasphaera* are Gram-negative cocci, which often associate as diplococci. They occasionally form longer chains in stationary phase. These cocci are larger in size (2-10µm) than the average coccus, which measures roughly 0.5-1.0µm, hence the prefix "Mega-" in the genus name (Rogosa, 1971). They are non-spore forming, non-motile and obligately anaerobic, as expected given their isolation from rumen samples. Since their discovery, *Megasphaera* have been isolated from disparate environments including spoiled beer and the human gastrointestinal and reproductive tracts (Juvonen & Suihko, 2006; Lanjekar, Marathe, Ramana, Shouche, & Ranade, 2014; Padmanabhan et al., 2013; Zozaya-Hinchliffe et al., 2008).

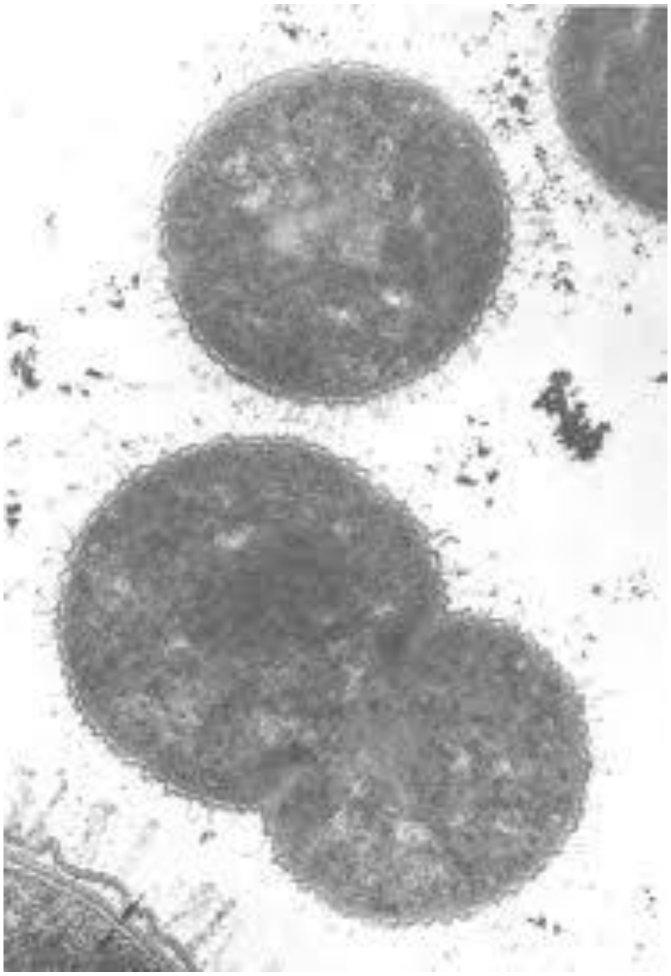
Two vaginal phylotypes, termed *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2, were first identified in 2008 by Zozaya-Hinchliffe *et al.* through cultivation-independent techniques examining the 16S rRNA gene (Zozaya-Hinchliffe et al., 2008). They reported the two phylotypes to be more prevalent among women with a current diagnosis of bacterial vaginosis. *Megasphaera* sp. (not sub-classified) have also been identified in male urethral and coronal sulcus samples, although it is still unclear whether the organisms can be sexually transmitted (Manhart et al., 2013; D. E. Nelson et al., 2012). *Megasphaera* phylotype 1, the more prevalent phylotype, has been associated with increased number of sexual partners, elevated viremia in HIV-positive patients and women who have sex with women (Dang et al., 2012; Fethers, 2001; Fethers et al., 2012). Due to its high specificity and sensitivity for the condition, detection of *Megasphaera* phylotype 1 has been used in combination with other BV-associated organisms for molecular diagnosis of bacterial vaginosis (Datcu et al., 2014).

It has also been shown to be capable of invading the upper genital tract in a recent study characterizing the UGT microbiome of women undergoing hysterectomies (Mitchell et al., 2015). Nelson *et al.* followed a cohort of pregnant women with a history of spontaneous preterm birth, collecting vaginal swabs at multiple time points throughout the pregnancy (D. B. Nelson et al., 2014). *Megasphaera* phylotype 1 was found to be associated with increased risk for spontaneous preterm delivery in this cohort, especially if the relative proportion of the organism increased during pregnancy. In a recent publication characterizing the metabolomics profiles of bacterial vaginosis, *Megasphaera* phylotype 1 was strongly correlated with 12-HETE,

Figure 1. ***Megasphaera elsdenii***

Electron micrograph of a section of *Megasphaera elsdenii* (ATCC 25940) at 46,000X magnification published by Rogosa *et al.* Note the coccal shape and diplococcus structure. Coccal structures shown range from 1.2-1.9µm. License ID: 3761521157733.

Figure 1.



an inflammatory metabolite observed at high levels in actively laboring women (Srinivasan et al., 2015). One suggested cause of spontaneous preterm delivery is the ascension of vaginal bacteria into the upper genital tract, which triggers an inflammatory response in the host resulting in early onset of labor. Given its association with negative health outcomes and correlation with spontaneous preterm delivery, *Megasphaera* phylotype 1 is a worthy of investigation as a potential biomarker for vaginal dysbiosis and high-risk pregnancies.

Fewer studies have focused on associations of *Megasphaera* phylotype 2, likely due to its lower overall prevalence. Most reported associations have either focused on *Megasphaera* phylotype 1 or not discriminate between the two phylotypes. One study noted an associated between *Megasphaera* phylotype 2 and trichomoniasis, the most common non-viral STI worldwide (Zozaya-Hinchliffe et al., 2008). Trichomoniasis is a major public health concern, affecting roughly 3.7 million women in the United States. However, it has largely been ignored and was recently named by the Centers for Disease Control and Prevention as a disease worthy of more scientific study (Meites et al., 2015). Trichomoniasis infection is characterized by itching, inflammation, dysuria, dyspareunia and malodorous discharge. However, roughly 30% of infections are asymptomatic, increasing the risk of transmission.

Trichomoniasis is also associated with increased risk of STI acquisition and pregnancy complications such as spontaneous preterm delivery and low birth weight (Cotch et al., 1997; Edwards, Burke, Smalley, & Hobbs, 2014; Silver, Guy, Kaldor, Jamil, & Rumbold, 2014). Intriguingly, trichomoniasis has previously been strongly associated with the presence of specific vaginal bacteria such as *Mycoplasma hominis* and "*Ca*. Mycoplasma girerdii" (Fettweis, Serrano, et al., 2014b; D. H. Martin et al., 2013; Rappelli et al., 2001; Rappelli, Addis, Carta, & Fiori, 1998) . This suggests that trichomoniasis may contribute to shifts in vaginal microbiome composition to favor presence of these organisms or vice versa.

The two *Megasphaera* phylotypes have frequently been grouped together in microbiome analyses at the genus level. Based on their distinct clinical associations in the literature, we hypothesized that while these two phylotypes apparently occupy the same niche and are often present together in vaginal communities, they likely contribute to the vaginal microbiome uniquely and exhibit niche specialization. In characterizing the differences between associations of the phylotypes with clinical infections and demographic data and analyzing their phylogeny and predicted function using comparative genomics, we aimed to develop an understanding of how each phylotype contributes to the microbiome and influences health outcomes.

Research Objectives

Two vaginal *Megasphaera* phylotypes (phylotype 1 and phylotype 2) have recently been identified in vaginal samples. Since their description in the literature by Martin *et al* in 2008, they have been associated with negative health outcomes in a number of vaginal microbiome studies, including bacterial vaginosis and trichomoniasis (see above). Although these two phylotypes are related and are sometimes grouped together in vaginal microbiome association analyses, they exhibit different clinical associations in the published literature. We sought to investigate the differential roles that these two phylotypes play in the vaginal community using a dataset of 16S rDNA and associated health and lifestyle information from 3,986 women generated by the Vaginal Human Microbiome Project (VaHMP) at Virginia Commonwealth University (VCU) (Fettweis, Serrano, Girerd, Jefferson, & Buck, 2012a).

Megasphaera phylotype 1 and *Megasphaera* phylotype 2 were grouped together in the initial species-level analysis of this dataset. Thus, we set out to develop a method to successfully distinguish between the two phylotypes using the existing 16S rDNA dataset. Our goal was to analyze the 16S rDNA dataset and the associated health history data to examine associations between the *Megasphaera* phylotypes and clinical diagnoses, demographics and lifestyle factors. We also sought to examine the correlation of these two phylotypes with other vaginal organisms present in the community.

While 16S rDNA association studies are useful for identifying biomarkers and targeting species worthy of further investigation, we must move towards understanding the mechanism of how these organisms interact in the vaginal environment and contribute to adverse health outcomes. We sought to cultivate and isolate *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 and perform whole genome sequencing. Since we began this project, additional

Megasphaera phylotype 1 and *Megasphaera* phylotype 2 genomes were sequenced by other groups and made available at NCBI. Thus, we used the publicly available genomes in conjunction with those sequenced in-house to conduct comparative genomic analyses. Our goal was to gain insight into why differences exist between the phylotypes at the level of clinical associations and vaginal microbiome composition. We sought to determine if the observed differences in clinical presentation and genomic composition could signal niche specialization within the vaginal community between these two closely related species.

MATERIALS AND METHODS

16S rDNA vaginal microbiome data

As a part of the Vaginal Human Microbiome Project (VaHMP) at Virginia Commonwealth University, mid-vaginal wall swab samples were obtained from 3,986 participants (Fettweis, Serrano, Girerd, Jefferson, & Buck, 2012a). DNA was isolated from the swabs using the MoBio Powersoil DNA Isolation Kit. DNA samples were randomized in an effort to avoid batch effects and the V1-V3 region of the 16S rRNA gene was amplified using polymerase chain reaction (PCR). The amplified 16S rDNA fragments were then sequenced on the Roche 454 GS FLX Titanium platform. Sequences were classified using both the Ribosomal Database Project (RDP) classifier and the in-house STIRRUPS (Species-level Taxon Identification of rDNA Reads using a USEARCH Pipeline Strategy) classifier to achieve species-level classification (Fettweis, Serrano, Sheth, Mayer, Glascock, Brooks, et al., 2012b). AbundantOTU analysis was also performed to detect prevalent OTUs (operational taxonomic units) for which no reference was present in the Vaginal 16S rDNA Reference Database (Fettweis, Serrano, Sheth, Mayer, Glascock, Brooks, et al., 2012b; Ye, 2011). The Institutional Review Boards for Human Subjects at Virginia Commonwealth University and The Virginia

Department of Health approved the VaHMP study and consent was obtained from all participants, under IRB protocol HM12169.

Subclassification

Reads terminally classified as *Megasphaera* cluster52 by STIRRUPS were aligned using MUSCLE, and visually inspected for single nucleotide polymorphisms (SNPs) using Jalview (Edgar, 2004; A. M. Waterhouse, Procter, Martin, Clamp, & Barton, 2009). To analyze intratype differences, sequences assigned to *Megasphaera* cluster52 that were closest to *Megasphaera* phylotype 1 were selected, aligned using MUSCLE and visualized using Jalview. For further resolution, that alignment was trimmed for informative regions using Gblocks, and used to create a phylogenetic tree using PhyML and TreeDyn (Castresana, 2000; Chevenet, Brun, Bañuls, Jacq, & Christen, 2006; Guindon et al., 2010). This process was repeated for *Megasphaera* phylotype 2. USEARCH v4.0 was used with a 97% cutoff to separate reads assigned to *Megasphaera* cluster52 into *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 (Edgar, 2010).

OTU analysis of *Megasphaera* and related reads

AbundantOTU clustering analysis was used to group reads terminally assigned to the genus *Megasphaera* by the RDP classifier into clusters based on sequence similarity using default parameters. Consensus sequences from each cluster were identified using BLAST and the non-redundant nucleotide database at NCBI (Wang, Garrity, Tiedje, & Cole, 2007; Ye, 2011).

Reclassification of *Megasphaera* in VaHMP dataset

All reads terminally classified at the genus level *Megasphaera,* family level Veillonellaceae, order level Clostridiales, class level Clostridia, phylum level Firmicutes or kingdom level Bacteria by the Ribosomal Database Project (RDP) classifier were reclassified using USEARCH v4.0 and an updated and comprehensive *Megasphaera* 16S rDNA V1-V3 region database. In-house scripts were utilized to integrate this updated classification data into an existing dataset containing 16S rDNA microbiome profile data, health history data and demographic data (Edgar, 2010; Wang et al., 2007).

Demographic and Clinical Associations

Megasphaera phylotypes were determined to be "present" if they comprised at least 0.1% of the microbiome of a given sample. Samples from participants enrolled in outpatient clinics were included; samples were excluded from the analysis if the participant was pregnant, was recruited in Labor & Delivery, was recruited in the twin cohort, if the samples were processed with a Qiagen DNA extraction kit instead of the MoBio DNA extraction kit or if the vaginal swab sample yielded less than 5,000 reads during 16S rDNA sequencing. Demographic and health history data was extracted from an extensive questionnaire, which was completed by the participants. Basic health statistics including height, weight, blood pressure, pulse, vaginal pH and diagnosis data on the date of visit were gathered by clinical coordinators under advisement of a physician. Associations were calculated based on the presence or absence of a taxon of interest in combination with given demographic or clinical data. Statistical significance was calculated using a two-tailed Student's T-test with an alpha level of 0.05.

Relative Risk

Relative risk values and their corresponding 95% confidence interval values were calculated based on the standard relative risk formula used in epidemiological studies. Relative Risk = (A/A+B) / (C/C+D) where A represents the number of samples where the taxon is present and the participant is diagnosed with the disease, B represents the number of samples where the taxon is present but the participant is not diagnosed with the disease, C represents the number of samples where the taxon is absent but the participant is diagnosed with the disease and D represents the number of samples where the taxon is not present and the participant is not diagnosed with the disease. Bacterial taxa were determined to be present if at least 0.1% of the reads from the sample were assigned to that taxon. The outpatient cohort used for this analysis (n=2633) was comprised of non-twin, non-pregnant participants (by diagnosis, self-report or clinic-ID). Samples met the threshold of at least 5,000 reads during sequencing and were processed using the MoBio Power Soil DNA extraction kit. Vaginal infection status was determined based on clinician diagnosis at time of visit.

Case Matching

An in-house perl script was developed to case match the participants in the pregnant cohort with non-pregnant controls. For each pregnant participant, a non-pregnant participant was identified as a case match if they were the same age, same ethnicity and fell into the same socioeconomic bracket. If a match was not found, the restrictions were loosened to be the same age plus or minus one year, the same race and the same socioeconomic status plus or minus one bracket. This case-matching methodology yielded 421 pregnant and 421 non-pregnant case matched pairs.

Pregnancy Analysis

The case-matched cohort was used for this analysis. Bacterial taxa were determined to be present if at least 0.1% of the reads from the sample were assigned to that taxon. For a number of BV-associated taxa, the proportion of the cohort containing that taxon was calculated for both pregnant and non-pregnant women. The difference in the prevalence of bacterial taxa between the pregnant and non-pregnant cohorts was analyzed for statistical significance using a two-tailed Student's T-test. To test if *Megasphaera* phylotype 1 was significantly higher in pregnant women than in non-pregnant women, a non-parametric Mann-Whitney U Test was performed (Whitney, 1997).

Microbial Co-occurrence

Prevalence of each taxon in the cohort as a whole was calculated as a proportion. The expected proportion of the cohort in which any two species would appear together simply by chance was also calculated for each pairwise combination of taxa by multiplying together their respective prevalence. The actual proportion of the cohort containing any two taxa was also calculated for each pairwise combination of taxa. The actual proportion to expected proportion ratio was calculated as a representative metric of how likely the actual co-occurrence proportion was based on a stochastic model.

Cultivation of *Megasphaera*

At the time of visit, a second mid-vaginal swab sample was collected from participants. This duplicate swab was used to inoculate 1.0mL of culture media with an added cryo-protectant, sBHI + 20% glycerol (Table 1). Vaginal samples were targeted for cultivation based on the presence of bacterial targets of interest in the 16S rDNA survey. A scraping of the frozen vaginal culture media from the selected targets was used to inoculate media plates for bacterial clone culture. Scrapings were plated on both ThermoScientific Remel Chocolate agar (lysed blood agar) and ThermoScientific Remel Brucella Blood agar (containing 5% sheep's blood) at four dilutions: 1:10, 1:100, 1:1000 and 1:10000. Plates were stored at 37°C for 24-48 hours. The plates were enclosed in three nested Ziploc bags along with a Mitsubishi Anaeropack-Anaero, which served to both absorb oxygen and release anaerobic gas, creating a mostly anaerobic environment, similar to the vagina, for the growth of fastidious anaerobic and microaerophilic organisms. Individual colonies were selected for growth and purification from the dilution plates based on colony morphology and differential growth characteristics. Isolates were re-streaked repeatedly until visibly pure. They were then re-streaked three more times to ensure purity. A single colony was used to inoculate 5mL of sBHI in a 15mL falcon tube. Tubes were loosely capped to allow gas exchange and stored in a rack at 37°C for 24-48 hours in three nested Ziploc bags containing an Anaeropack. The DNA was then harvested using the Qiagen Spin Miniprep Kit and quantified using the Nanodrop 2000 spectrophotometer.

Table 1. Supplemented Brain-Heart Infusion Recipe

Ingredient	Quantity
Brain-Heart Infusion Powder (Oxoid)	9.25g
Yeast Extract	2.50g
Gelatin	2.50g
Dextrose	0.25g
Sucrose	0.25g
Deionized Water	250mL

Identification of Isolates

Bacterial clone isolates were identified by colony PCR amplification of the full 16S rRNA gene using universal 16S primers (Table 2). Amplicons were purified using the Qiagen QIAquick PCR Purification Kit and sequenced using Sanger sequencing technology on the Applied Biosystems 3730 DNA Analyzer. Sequences were trimmed for quality and chromatograms were analyzed for purity. Sequences were classified based on the combined results of a blastn comparison against the in-house Vaginal 16S rDNA Reference Database and a BLAST search against the non-redundant nucleotide database at NCBI.

Sequencing and Assembly of Genomic DNA

Purified genomic DNA from the *Megasphaera* phylotype 1 isolate OTU70 was sequenced using the Roche 454 GS FLX Titanium platform. The OTU70 reads were assembled using Newbler v2.8. Purified genomic DNA from the two *Megasphaera* phylotype 2 isolates M2-4 and M2-8 were sequenced using the Illumina MiSeq platform and reads were assembled using Velvet. Quality trimming, low complexity filtering, and Poly-A, Poly-T and N filtering were implemented. All sequencing and assembly was performed at the Nucleic Acids Research Facilities at Virginia Commonwealth University.

Synteny Analysis

Analysis of genomic synteny both within phylotype and between phylotypes was performed at the protein and nucleic acid level using PROmer and NUCmer respectively. Both

Table 2. Universal 16S rRNA primers

Primer Name	Sequence (5’ to 3’)[a]
16SF-YM	AGAGTTTGATYMTGGCTCAG
16SF-Bif	AGGGTTCGATTCTGGCTCAG
16SF-Bor	AGAGTTTGATCCTGGCTTAG
16SF-Chl	AGAATTTGATCTTGGCTTAG
1492R	TACCTTGTTACGACTT

[a] Degenerate bases are underlined. Forward primers were combined in a 4:1:1:1 ratio (16SF-YM : 16SF-Bif : 16SF-Bor : 16SF-Chl).

of these tools are part of the MUMmer 3.0 package. Synteny plots were created using gnuplot from the gnuplot 4.2 package and MUMmerplot, which is also a part of the MUMmer 3.0 package (Kurtz et al., 2004; Williams & Kelley, 2011).

Genome Annotation

Genomes were annotated using the in-house Genome Annotation Pipeline developed by Dr. Vishal N. Koparde of the VCU Nucleic Acids Research Facilities. This pipeline annotates genes using both Glimmer3 and GeneMarkS, identifies rRNA genes with rnammer, identifies tRNA genes using tRNAScan, calls orthologous genes with using rpsblast in conjunction with Pfam and COG databases, and assigns annotated genes a predicted function using blastx. Genome annotation was also performed using RAST, a web-based annotation tool provided by NMPDR (Bateman et al., 2002; Borodovsky & Lomsadze, 2011; Delcher, Harmon, Kasif, White, & Salzberg, 1999; Lagesen et al., 2007; Lowe & Eddy, 1997; Meyer et al., 2008; Tatusov, Galperin, Natale, & Koonin, 2000).

GC Composition Analysis

Protein-encoding genes were identified in all genomes using Glimmer3 and GC composition was calculated for each protein-encoding gene using in-house scripts. Average GC percentage of protein-encoding genes as well as whole genome GC composition were analyzed.

Codon Usage Analysis

Codon usage within the genomes was calculated using cusp, a program in the EMBOSS Tools package available through EMBL-EBI (Rice, Longden, & Bleasby, 2000).

Metabolic Reconstruction

Metabolic reconstruction was performed using ASGARD. Visual representations of phylotype differences within metabolic pathways were generated using color-maps (Alves & Buck, 2007). Metabolic reconstruction was also performed using RAST, which is available online through NMPDR.

Strain Resurrection

Scrapings from *Megasphaera* phylotype 1 and phylotype 2 isolates frozen in 1.0mL of sBHI media + 20% glycerol were used to inoculate ThermoScientifc Remel Chocolate agar plates, ThermoScientific Remel 5% Sheep Blood Brucella Blood agar plates and 5mL tubes of sBHI+s containing 10% human serum. Plates were cultivated anaerobically at 37°C for 24-96 hours in three nested Ziploc bags containing an Anaeropack. Tubes were loosely capped to allow gas exchange and stored in an anaerobic incubator at 37°C with 5% CO_2 for 24-96 hours.

Phylogenetic Analysis

OrthoDB, an online database for orthologous groups was used to determine which orthologous genes were conserved at the family level (Veillonellaceae) (R. M. Waterhouse, Tegenfeldt, Li, Zdobnov, & Kriventseva, 2013). These genes were verified using reciprocal blast and isolated from each of the six *Megasphaera* genomes used for our phylotype analysis as well as from all publicly available *Megasphaera* genomes and the single *Anaeroglobus* genome at NCBI. Different outgroup genomes were chosen from NCBI based on the analysis. Each gene was separately aligned using MUSCLE, a program within the EMBOSS Tools package available through EMBL-EBI (Edgar, 2004; Rice et al., 2000). Alignments were visually examined and those with large gaps or likely errors were discarded. For each genome, the remaining 321 orthologous genes were then concatenated together to create one large representative sequence. These sequences were then pruned using Gblocks to remove any uninformative stretches of sequence (Castresana, 2000). The resulting sequences were then converted from pir to phylip format using the online service, ALTER (Glez-Peña, Gómez-Blanco, Reboiro-Jato, Fdez-Riverola, & Posada, 2010). RAxML-HPC was used to perform a rapid bootstrap analysis using 100 bootstraps and search for the best scoring maximum likelihood tree using optimization of substitution rates, the gamma model of heterogeneity and the WAG amino acid substitution matrix. RAxML-HPC was also used to draw the bootstrap values on the best scoring maximum likelihood tree (Stamatakis, 2014). Aesthetic changes to the tree were made using TreeDyn (Chevenet et al., 2006).

RESULTS

Subclassification of *Megasphaera* reads into distinct phylotypes

The previous approach to classification utilized by the VaHMP project was the STIRRUPS (Species-level Taxon Identification using a USEARCH Pipeline Strategy) classifier, a pipeline developed to achieve high-quality species-level identification. For some analyses, the RDP classifier, which allows genus-level classification, was utilized (Fettweis, Serrano, Sheth, Mayer, Glascock, Brooks, et al., 2012b; Wang et al., 2007). The RDP methodology uses a large, carefully curated database of bacterial 16S ribosomal RNA sequences as a reference library. The classifier is a naïve Bayesian classifier meaning that it uses no "a priori" knowledge to classify reads and it calls classifications based on measures of likelihood. It calculates the probability that a given sequence was derived from a specific genus based on the presence of similar sequence portions in the read and reference sequences attributed to members of the genus in the reference library. Instead of using an alignment approach, it looks at smaller portions of the read by breaking it down into what it describes as "words", small portions of the sequence eight base pairs in length. It then uses a sliding window approach to calculate likelihood measures for each word in the read. Combining information from all words across the read yields the genus with the highest likelihood score. The RDP classifier is used to classify reads from the domain to the genus level in its native form (Wang et al., 2007).

Although the RDP classifier is useful for determining genus-level classifications, the original version cannot attain species-level resolution. This is pertinent for our study of the vaginal microbiome because members of the same genus may play very different roles in the vaginal community. For example, *L. crisptaus* is a vaginal taxon that has often been associated with low pH, a marker of vaginal health. It was been negatively associated with bacterial vaginosis as well. *L. iners*, a member of the same genus, is associated with higher pH and is often found in vaginal samples containing bacterial vaginosis associated bacteria such as *G. vaginalis* (Ravel et al., 2011). When trying to characterize a vaginal microbiome, knowing which vaginal *Lactobacillus* species are present is of great importance as it signals how protective that species is. For this reason, we developed the STIRRUPS classifier (Fettweis, Serrano, Sheth, Mayer, Glascock, Brooks, et al., 2012b). It utilizes an in-house reference library populated with the V1-V3 variable region of the 16S rRNA gene from vaginally relevant taxa. The top twenty-five most prevalent genera detected by RDP were selected and all species of those genera were added to the reference library if sequences were available. Other vaginally relevant bacteria were added as well including common pathogens such *Neisseria gonorrhoeae.* The STIRRUPS classifier uses the USEARCH v4.0 global alignment method to assign the read to the taxon to which it is most similar. The species-level taxon with the highest alignment score is selected as the best hit (Edgar, 2010).

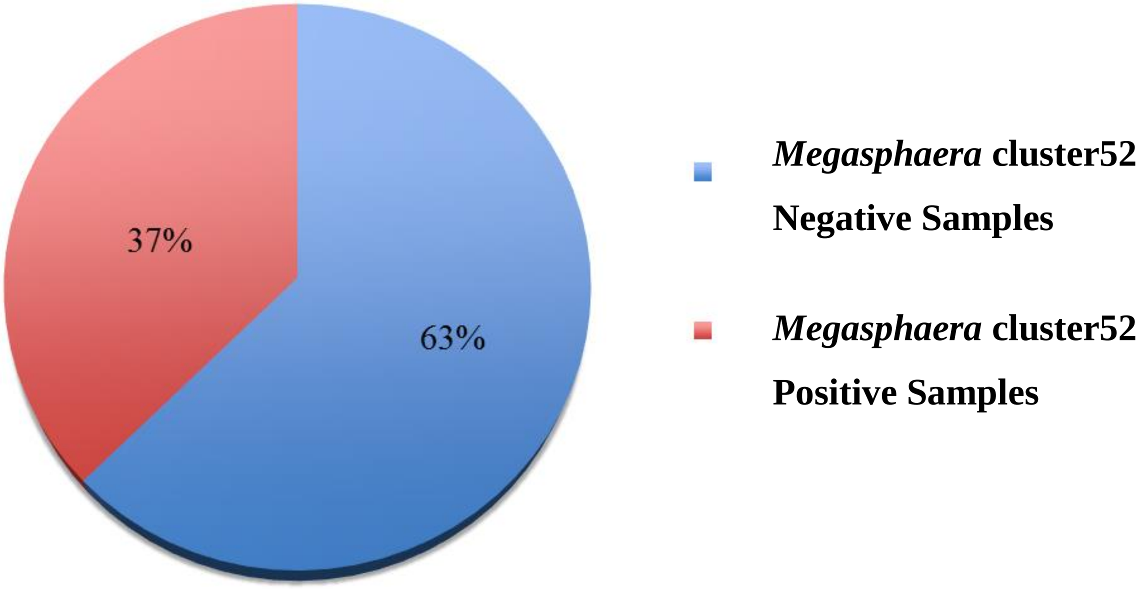
However, in some cases, the V1-V3 region of the 16S rRNA gene is not enough to confidently distinguish between two species. A common example of this is *Escherichia coli* and *Shigella* sp. Although they are separate species, they are nearly identical (99% similar) in the V1-V3 region of the 16S rRNA gene, and because of this they are grouped together to form a cluster in our analyses. Sequencing error can yield illegitimate single nucleotide polymorphisms, thus it is difficult to distinguish between closely related species, particularly when the true underlying references are unknown. Thus, we implemented a conservative approach to classification of different species in STIRRUPS, and clustered together closely related species. Read length is also an important factor to consider in the process of read assignment due to the nature of the V1-V3 region of the 16S rRNA gene. The VaHMP 16S rDNA dataset was generated using 454 GS FLX Titanium pyrosequencing, sequencing from the V1 side. Some closely related taxa may be nearly identical in the V1 region, but be differentiated in the V3 region. Thus, it would be impossible to discriminate with a short read of 200bp, but a longer read spanning the entire V1-V3 region would permit classification.

To address this issue, a collection of subsequences was created, in all possible one base incremented read lengths from 200bp (the minimum allowable read length) to the full length of the sequence for each sequence in the reference database. Each subsequence was then aligned to the Vaginal 16S rDNA Reference Database (our in-house database) using the USEARCH v4.0 global alignment algorithm. If any subsequence matched another taxon at ≥ 97% identity, those two taxa were clustered together. This was the case for the two *Megasphaera* phylotypes identified in the vaginal samples. Although they only share 96% similarity at the nucleotide level of the full length 16S rRNA gene, somewhere in the clustering analysis the subsequences of one phylotype aligned to the other phylotype at ≥ 97% identity causing these two taxa to be clustered together. This cluster is called "*Megasphaera* cluster52" in the VaHMP dataset. In order to be able to address the differences between these two phylotypes in terms of both microbiome composition and association with clinical and demographic data, it was necessary to be able to sub-classify the reads assigned to "*Megasphaera* cluster52" into either *Megasphaera* phylotype 1 or *Megasphaera* phylotype 2 (Fig.2).

Figure 2. ***Megasphaera*** **cluster52 prevalence**

A cohort of 4148 participants from the VaHMP study at VCU was included in this analysis. Representative 16S rDNA profiles for these participants were generated using RDP and STIRRUPS classifiers. *Megasphaera* cluster52 is defined as present if it comprises at least 0.1% of the microbiome profile.

Figure 2.


37%
63%
Megasphaera cluster52 Negative Samples
Megasphaera cluster52 Positive Samples

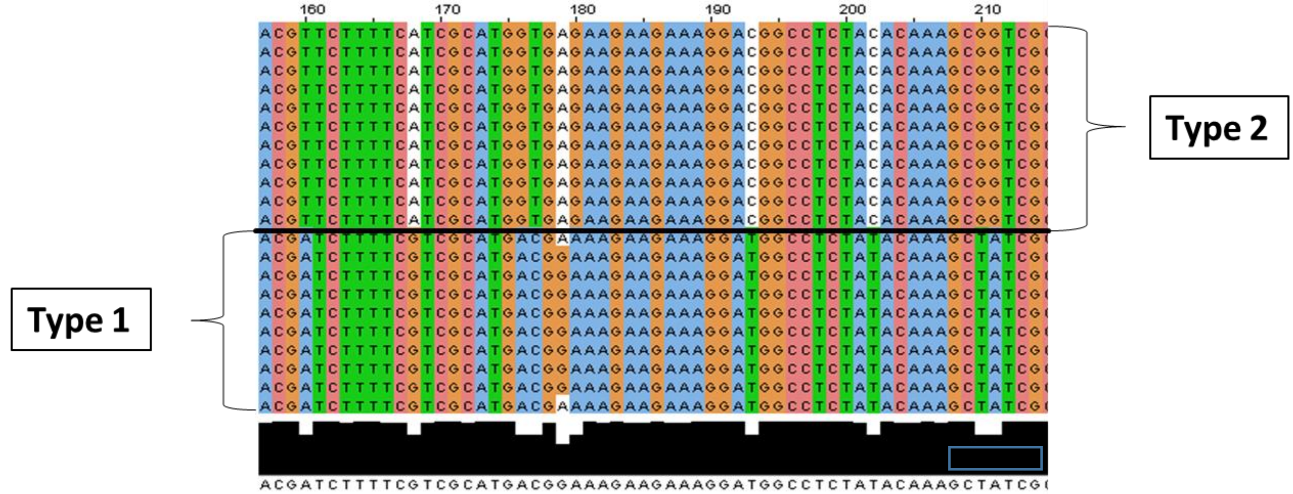
The full-length 16S rDNA sequences of the two phylotypes were 96% similar at the nucleotide level and 93% similar in the V1-V3 region of the gene. The two phylotypes were approaching the previous threshold for distinguishing separate species, so we determined whether it may be possible to separate the two phylotypes and sub-classify them from vaginal samples even using a basic strategy. The first step was examining the V1-V3 region of the 16S rRNA gene of both phylotypes visually to determine if there were specific sites in the sequence where they were different. These sites are referred to as SNPs. We also wanted to know if there were any apparent subtypes within the phylotypes that could be distinguished based on SNPs. The number of SNPs present and ability to distinguish between the phylotypes is largely contingent on the sequence length. In order to answer the above question, we created a dataset of 100 randomly selected vaginal reads assigned to *Megasphaera* by the RDP & STIRRUPS pipeline that could be of any length and could be closest to either phylotype based on the STIRRUPS alignment data. These reads were aligned using MUSCLE and visualized using Jalview (Edgar, 2004; A. M. Waterhouse et al., 2009). The alignments clearly clustered into two distinct groups, representing those most closely related to each phylotype. There were conserved SNPs that were segregated by phylotype, supporting our hypothesis that there might be enough differences between the two phylotypes to distinguish between them with our data (Fig.3).

We also performed this analysis using only long reads (>500bp) and only short reads (trimmed to 200bp) to determine if there were important SNPs only contained in long reads near the end of the sequence and if there was still separation between the phylotypes using short reads, which contained a smaller number of informative SNPs. We performed the analysis for each read length type (200bp, any read length and >500bp) using 250 randomly selected reads and 500

Figure 3. Visualization of alignment containing sequences of both phylotypes

Full-length 16S rDNA reads either most similar to *Megasphaera* phylotype 1 or *Megasphaera* phylotype 2 were combined. These reads were aligned using MUSCLE and visualized using Jalview. This figure shows a small portion of the alignment using 21 sequences. The portion shown is the 157bp-214bp region of the V1-V3 region. Single nucleotide polymorphisms and areas of incongruence in the alignment can be visualized using the solid black bar chart below the alignment. *Megasphaera* phylotype 1 reads are labeled as “Type 1” and *Megasphaera* phylotype 2 reads are labeled as “Type 2.”

Figure 3.


160
170
180
190
200
210
Type 2
Type 1
ACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTATACAAAGCTATCG

randomly selected reads to determine if adding more sequences had any effect on the clustering of the alignments, introduced any subgroups to the analysis, or introduced new informative SNPs. It was also important to determine if the SNPs we found in the smaller datasets were conserved across the larger datasets, which would determine if they would be useful in sub-classifying the *Megasphaera* reads (Table 3).

We also wanted to know if there were any obvious subtypes within the phylotypes that could be distinguished based on SNPs. We performed the same alignment and visual inspection analysis described above using only reads most similar to *Megasphaera* phylotype 1 by STIRRUPS analysis. At all read lengths and at all dataset sizes there were no obvious subgroups within the phylotype. There were smaller groups that clustered together but the SNPs were often seemingly random substitutions and did not seem to be conserved across a large number of reads (Fig.4). The same conclusion was drawn when performing the analysis using only reads most similar to *Megasphaera* phylotype 2 by STIRRUPS analysis (Fettweis, Serrano, Sheth, Mayer, Glascock, Brooks, et al., 2012b). Again, there were no obvious subtypes and SNPs were not conserved across large datasets (Fig.5). To further assess if there were any intratype differences, datasets with long reads which contained only reads most similar to *Megasphaera* phylotype 1 by STIRRUPS analysis were aligned using MUSCLE, trimmed to contain only informative portions using Gblocks and used to create a phylogenetic tree using PhyML and TreeDyn (Castresana, 2000; Chevenet et al., 2006; Edgar, 2004; Guindon et al., 2010). The trees were then analyzed for the presence of specific groupings or branching patterns in an effort to find subtypes. The same process was also performed for the datasets containing long reads most similar to *Megasphaera* phylotype 2. There were no obvious groupings in the trees. Although

Table 3 A. Informative SNPs to distinguish between *Megasphaera* phylotypes in the V1-V3 region

SNP Location	*Megasphaera* phylotype 1	*Megasphaera* phylotype 2
48	C	T
53	G	A
64	C	T
69	G	A
75	G	T
126	T	C
133	C	T
160	A	T
168	G	A
176	A	G
177	C	T
179	G	A
180	A	G
193	T	C
202	T	C
210	T	G
211	A	G
244	G	A
274	T	C
409	A	C
427	G	A
445-449	AAAAA	AAC
453	-	A
454	-	A
457	A	G/T
461	C	T

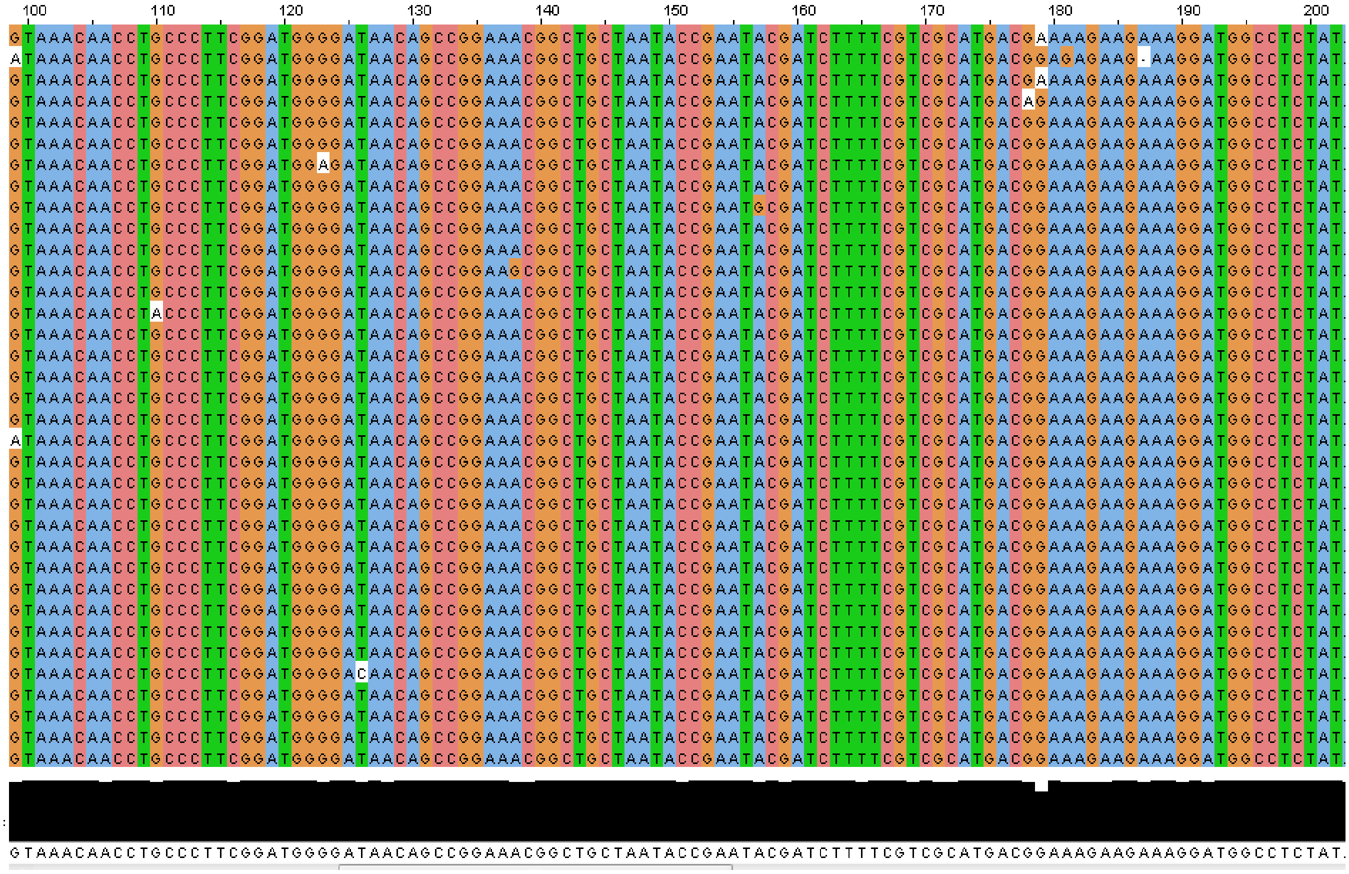
Table 3 B. Informative SNPs to distinguish between *Megasphaera* phylotypes in the V1-V3 region (cont'd)

SNP Location	*Megasphaera* phylotype 1	*Megasphaera* phylotype 2
462	C	A
465	C	T
468	T	C
469	-	C
472	C	-
474	G	A
476	C	G
477	C	A

Figure 4. Intratype alignment analysis of *Megasphaera* phylotype 1

16S rDNA reads in the V1-V3 region that were most closely related to *Megasphaera* phylotype 1 were combined, aligned using MUSCLE and visualized using Jalview. There were no obvious SNPs in the sequence that would suggest any subtypes might exist within the phylotype. The 100bp-200bp region is shown.

Figure 4.



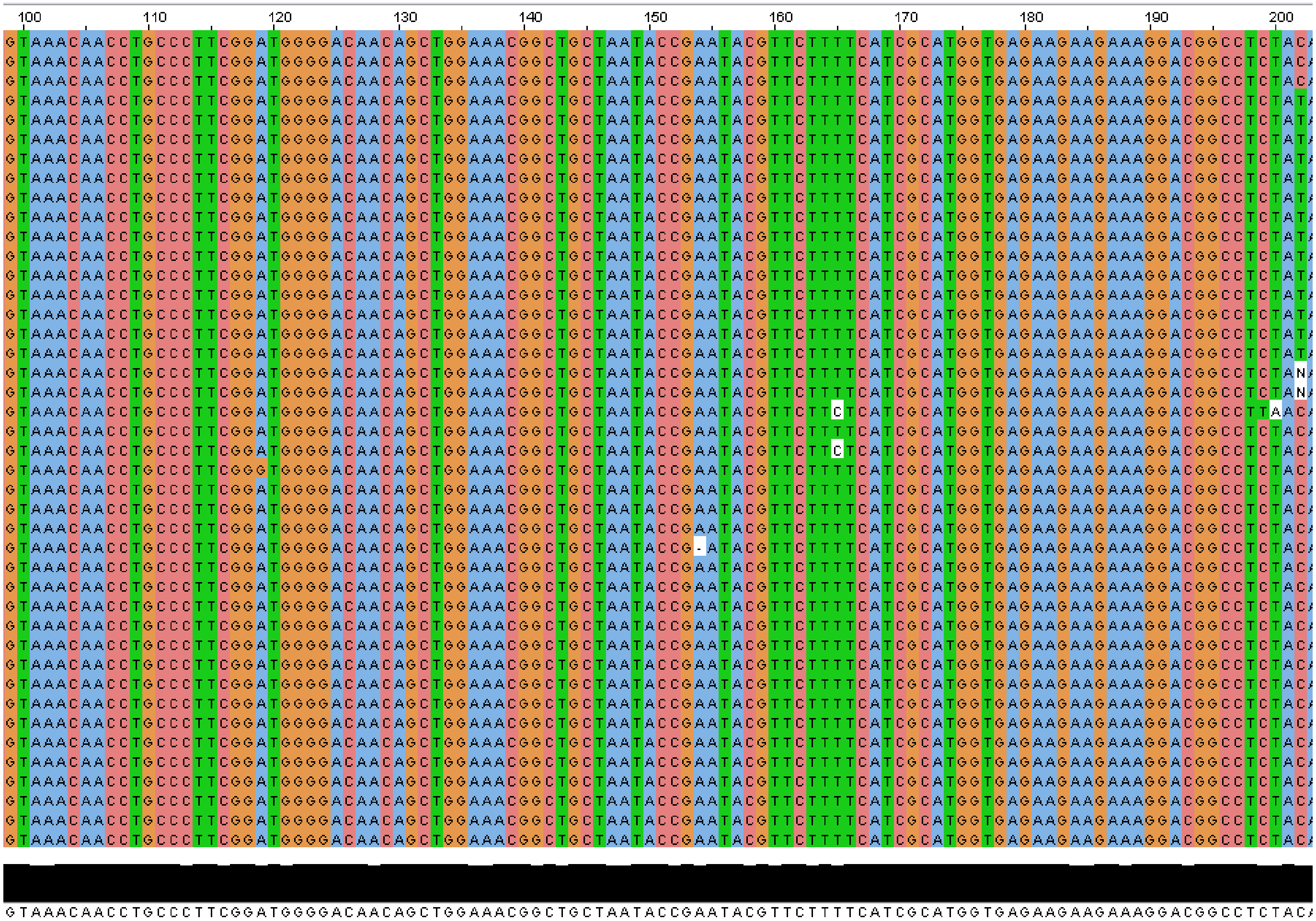
```
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGAAAAGAAGAAAGGATGGCCTCTAT.
ATAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAGAGAAG-AAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGAAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACAGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGAGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATGCGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAGCGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTACCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
ATAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.

GTAAACAACCTGCCCTTCGGATGGGGATAACAGCCGGAAACGGCTGCTAATACCGAATACGATCTTTTCGTCGCATGACGGAAAGAAGAAAGGATGGCCTCTAT.
```

Figure 5. Intratype alignment analysis of *Megasphaera* phylotype 2

16S rDNA reads in the V1-V3 region that were most closely related to *Megasphaera* phylotype 2 were combined, aligned using MUSCLE and visualized using Jalview. Like *Megasphaera* phylotype 1, the visual analysis yielded no obvious SNPs in the sequence that would suggest any subtypes might exist within the phylotype. The 100bp-200bp region is shown.

Figure 5.



```
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTATA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTANA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTANA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTCTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTTAACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTCTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGGTGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCG-ATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA

GTAAACAACCTGCCCTTCGGATGGGGACAACAGCTGGAAACGGCTGCTAATACCGAATACGTTCTTTTCATCGCATGGTGAGAAGAAGAAAGGACGGCCTCTACA
```

Figure 6. Intratype phylogenetic analysis of *Megasphaera* phylotype 1

A total of 25016S rDNA reads in the V1-V3 region that were most closely related to *Megasphaera* phylotype 1 were combined, aligned using MUSCLE and pruned for informative regions using Gblocks. A phylogenetic tree was created using PhyML. Although there were some distinct clades, they were not different enough based on sequence and SNPs to investigate the presence of any subtypes.

Figure 6.

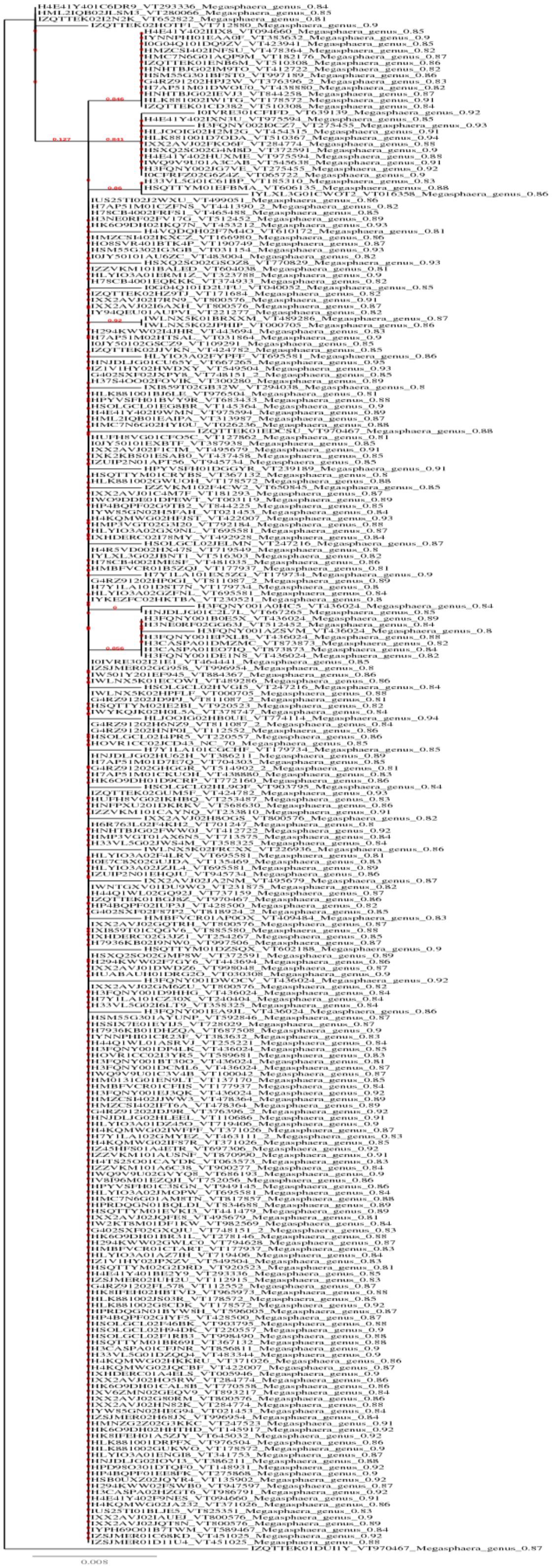
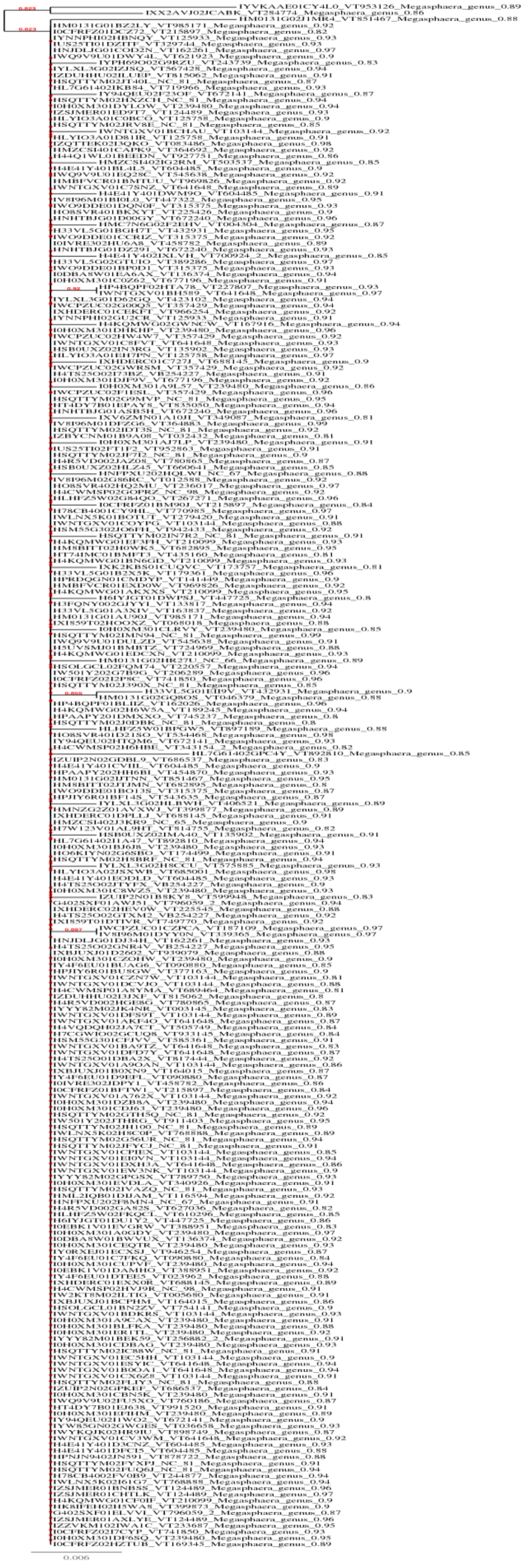

H4E41Y401C6DR9_VT293336_Megasphaera_genus_0.84
HML2IQB02JLSM3_VT280066_Megasphaera_genus_0.85
IZQTTEK02I2N2K_VT652822_Megasphaera_genus_0.81
IZQTTEK02HOTF1_VT712880_Megasphaera_genus_0.9
H4E41Y402IIX8_VT094660_Megasphaera_genus_0.85
IYNNPHI01EAA0F_VT383632_Megasphaera_genus_0.9
I0G04Q101DQ9ZV_VT423941_Megasphaera_genus_0.85
HMZCSI402INFSU_VT478364_Megasphaera_genus_0.82
HMC7N6G01AQP9M_VT182176_Megasphaera_genus_0.87
IZQTTEK01ENB6M_VT510308_Megasphaera_genus_0.86
HNHTBJG02IM9TO_VT412722_Megasphaera_genus_0.82
HSM55G301BFST0_VT997189_Megasphaera_genus_0.86
G4RZ91202HPJ2W_VT376396_2_Megasphaera_genus_0.83
H7AP51M01DWOU0_VT438880_Megasphaera_genus_0.82
HNHTBJG02IEVJ3_VT844258_Megasphaera_genus_0.87
HLK881002IW1TG_VT178572_Megasphaera_genus_0.91
IZQTTEK01CD382_VT510308_Megasphaera_genus_0.84
0.046
I0IVRE301CFIFD_VT639139_Megasphaera_genus_0.92
H4E41Y402IXN3U_VT975594_Megasphaera_genus_0.85
H3FQNY002I0CZ7_VT275455_Megasphaera_genus_0.93
HLJOOIG02H2M2G_VT454315_Megasphaera_genus_0.91
HLK881001D7ODA_VT510367_Megasphaera_genus_0.94
IXX2AVJ02FKO6F_VT284774_Megasphaera_genus_0.88
HSXQ2SO02G4M8D_VT372591_Megasphaera_genus_0.9
H4E41Y402HUXME_VT975594_Megasphaera_genus_0.88
IWQ9V9U01A3CAB_VT545638_Megasphaera_genus_0.91
H3FQNY002JG7VE_VT275455_Megasphaera_genus_0.92
I0CFRFZ02G6Z4Z_VT065722_Megasphaera_genus_0.9
0.127
0.841
H33VL5G01C61BP_VT185310_Megasphaera_genus_0.83
HSQTTYM01EFBMA_VT606135_Megasphaera_genus_0.88
IYLXL3G01CWOT2_VT016358_Megasphaera_genus_0.86
0.06
IUS25TI02J2WXU_VT499051_Megasphaera_genus_0.86
H7AP51M01CZFNS_VT441390_2_Megasphaera_genus_0.82
H78CB4002FRFS1_VT465488_Megasphaera_genus_0.85
H3NE0RF02FV17G_VT512452_Megasphaera_genus_0.89
HK6O9DH02IKQ7N_VT453212_Megasphaera_genus_0.93
H4VQDQH02F7M4O_VT610172_Megasphaera_genus_0.81
HMZCSI402IKXCZ_VT166980_Megasphaera_genus_0.86
HO8SVR401BTK4P_VT190749_Megasphaera_genus_0.87
HSM55G302IG3GB_VT031154_Megasphaera_genus_0.93
I0JY50101AU6ZC_VT483004_Megasphaera_genus_0.82
HSXQ2SO02GSOZ8_VT770829_Megasphaera_genus_0.93
IZZVKM101BALED_VT604038_Megasphaera_genus_0.81
HLYIO3A01ERM1Z_VT323788_Megasphaera_genus_0.9
H78CB4001EQKKK_VT374933_Megasphaera_genus_0.82
I0G04Q101D2UFU_VT040052_Megasphaera_genus_0.85
IZQTTEK02HZ9TJ_VT171684_Megasphaera_genus_0.82
IXX2AVJ02I7RN9_VT800576_Megasphaera_genus_0.91
IXX2AVJ02I6AXH_VT800576_Megasphaera_genus_0.87
IY94QEU01AUPVI_VT221277_Megasphaera_genus_0.82
IWLNX5K01BRXXM_VT489286_Megasphaera_genus_0.87
IWLNX5K02JPHIP_VT000705_Megasphaera_genus_0.86
0.92
H294KWW02I4JHR_VT443694_Megasphaera_genus_0.83
H7AP51M02HTSAL_VT031864_Megasphaera_genus_0.9
I0JY50102GSCZ9_VT109291_Megasphaera_genus_0.85
IZQTTEK02JJVKN_VT424782_Megasphaera_genus_0.85
HLYIO3A02FYPFF_VT695581_Megasphaera_genus_0.86
HNJDLJG01CU65Y_VT667265_Megasphaera_genus_0.95
IZ1V1HY02HWDXY_VT549504_Megasphaera_genus_0.93
G402SXF02JXPY8_VT748151_2_Megasphaera_genus_0.85
H37S4OO02FOVIK_VT300280_Megasphaera_genus_0.89
IXI859T02GB32W_VT294038_Megasphaera_genus_0.8
HLK881001BJ6LE_VT976504_Megasphaera_genus_0.81
HPYVSFH01BVY9R_VT683433_Megasphaera_genus_0.88
HSOLGCL01EG8BR_VT145364_Megasphaera_genus_0.9
H4E41Y402I9WMN_VT975594_Megasphaera_genus_0.89
HML2IQB01EAIPA_VT313987_Megasphaera_genus_0.87
HMC7N6G02HYI0U_VT026236_Megasphaera_genus_0.88
IZQTTEK01EDCSU_VT970467_Megasphaera_genus_0.88
HUFH8VG01CFO5C_VT127862_Megasphaera_genus_0.81
I0JY50101EXBTF_VT387938_Megasphaera_genus_0.85
IXX2AVJ02F1CIM_VT495679_Megasphaera_genus_0.91
IXK2KBS01ESAB0_VT437458_Megasphaera_genus_0.85
IZUIP2N01APT56_VT945734_Megasphaera_genus_0.85
HPYVSFH01DGGYR_VT239189_Megasphaera_genus_0.91
HSQTTYM01CRYBS_VT367132_Megasphaera_genus_0.8
HLK881002GWUOH_VT178572_Megasphaera_genus_0.88
IZZVKM102F4CW2_VT650845_Megasphaera_genus_0.85
IXX2AVJ01C4M7F_VT181293_Megasphaera_genus_0.87
IWO9DDE01DPEWT_VT003119_Megasphaera_genus_0.89
HP4BQPF02G9TB2_VT844225_Megasphaera_genus_0.85
IYW85GN02H5FAH_VT021453_Megasphaera_genus_0.84
H4KQMWG02HFJST_VT422007_Megasphaera_genus_0.93
HMP3VGT02G3I20_VT792184_Megasphaera_genus_0.88
HLYIO3A02GX9NL_VT695581_Megasphaera_genus_0.87
IXHDERC02I78MY_VT492928_Megasphaera_genus_0.84
HSOLGCL02JELMN_VT247216_Megasphaera_genus_0.87
H4R5VD002HX47S_VT719549_Megasphaera_genus_0.8
IYLXL3G02JBNT1_VT516303_Megasphaera_genus_0.82
H78CB4002IMESF_VT481035_Megasphaera_genus_0.86
HMBFVCR01B5ZQJ_VT177937_Megasphaera_genus_0.81
H7Y1LA101EX5ZG_VT179734_Megasphaera_genus_0.9
G4RZ91202HP0GI_VT811087_2_Megasphaera_genus_0.89
H7Y1LA101DST7N_VT179734_Megasphaera_genus_0.8
HLYIO3A02GZFNL_VT695581_Megasphaera_genus_0.84
IYKEZFC02HKTBA_VT230521_Megasphaera_genus_0.8
H3FQNY001A0HC5_VT436024_Megasphaera_genus_0.84
HNJDLJG01C2L7L_VT667265_Megasphaera_genus_0.85
0
H3FQNY001B0E5X_VT436024_Megasphaera_genus_0.89
H3NE0RF02GG63J_VT512452_Megasphaera_genus_0.84
H3FQNY001AZSVM_VT436024_Megasphaera_genus_0.8
H3FQNY001EPXLB_VT436024_Megasphaera_genus_0.88
H3CASPA01DMZMC_VT873873_Megasphaera_genus_0.82
H3CASPA01EO7IQ_VT873873_Megasphaera_genus_0.84
H3FQNY001DE1N8_VT436024_Megasphaera_genus_0.82
0.056
I0IVRE302I21E1_VT464441_Megasphaera_genus_0.85
IZSJMER02GG958_VT996954_Megasphaera_genus_0.8
IW501Y201EF945_VT884367_Megasphaera_genus_0.86
IWLNX5K01ECOWI_VT489286_Megasphaera_genus_0.86
HSOLGCL02HVGI5_VT247216_Megasphaera_genus_0.84
IWLNX5K02HPFLF_VT000705_Megasphaera_genus_0.88
G4RZ91202JD9PJ_VT811087_2_Megasphaera_genus_0.81
HSQTTYM02IE2BI_VT920523_Megasphaera_genus_0.82
IWYKQJK02H0L5A_VT378747_Megasphaera_genus_0.84
HLJOOIG02HB0UE_VT774114_Megasphaera_genus_0.94
G4RZ91202H6NZ9_VT811087_2_Megasphaera_genus_0.84
G4RZ91202HNP0I_VT112552_Megasphaera_genus_0.86
HSOLGCL02I4PR5_VT220557_Megasphaera_genus_0.86
HOVR1CC02JCD43_NC_70_Megasphaera_genus_0.85
H7Y1LA101CGCHF_VT179734_Megasphaera_genus_0.85
HNJDLJG02HU62H_VT386211_Megasphaera_genus_0.89
H7AP51M01D7E7Q_VT704303_Megasphaera_genus_0.85
G4RZ91202GHGGR_VT514902_2_Megasphaera_genus_0.81
H7AP51M01CKUOH_VT438880_Megasphaera_genus_0.83
HK6O9DH01D9CRP_VT772160_Megasphaera_genus_0.86
HSOLGCL02HL9OF_VT903795_Megasphaera_genus_0.84
IZQTTEK02GUM5F_VT424782_Megasphaera_genus_0.93
HUFH8VG02IKHBQ_VT253487_Megasphaera_genus_0.83
HNFPXU201DKRKV_VT568630_Megasphaera_genus_0.86
IZZVKM101CAYNQ_VT233810_Megasphaera_genus_0.91
IXX2AVJ02H8OGS_VT800576_Megasphaera_genus_0.82
H6R763L02F4KH2_VT701247_Megasphaera_genus_0.8
HNHTBJG02FWW0J_VT412722_Megasphaera_genus_0.92
HMP3VGT01AX6N5_VT713575_Megasphaera_genus_0.84
H33VL5G02JWS4M_VT358325_Megasphaera_genus_0.84
IWLNX5K02FRCXX_VT226936_Megasphaera_genus_0.86
HLYIO3A02F4LRV_VT695581_Megasphaera_genus_0.81
I0E7C8X02GUJDA_VT135469_Megasphaera_genus_0.83
HLYIO3A02JZJL4_VT695581_Megasphaera_genus_0.89
IZUIP2N01EHQJU_VT945734_Megasphaera_genus_0.86
IXX2AVJ02JA2NM_VT495679_Megasphaera_genus_0.87
IWNTGXV01DU9WO_VT231875_Megasphaera_genus_0.82
H44Q1WL02GQ92J_VT737159_Megasphaera_genus_0.87
IZQTTEK01BGJ8Z_VT970467_Megasphaera_genus_0.86
HP4BQPF02IUP3J_VT428500_Megasphaera_genus_0.82
G402SXF02F87P2_VT818924_2_Megasphaera_genus_0.85
HMBFVCR01AP0OX_VT409484_Megasphaera_genus_0.83
IXX2AVJ02GQTRH_VT800576_Megasphaera_genus_0.87
IXI859T01CQGV6_VT885580_Megasphaera_genus_0.88
IXHDERC02G3JZ1_VT254267_Megasphaera_genus_0.85
H7936KB02I9NW0_VT997506_Megasphaera_genus_0.87
HSQTTYM01DZSQX_VT602188_Megasphaera_genus_0.9
HSXQ2SO02GMP8W_VT372591_Megasphaera_genus_0.89
H294KWW02F7GY6_VT443694_Megasphaera_genus_0.86
IXX2AVJ01DWDZ6_VT998048_Megasphaera_genus_0.87
HUABAUH01DRG2O_VT030308_Megasphaera_genus_0.9
H3FQNY001DWOCV_VT436024_Megasphaera_genus_0.92
IXX2AVJ02GM6ZU_VT800576_Megasphaera_genus_0.82
H3FQNY001D9HHG_VT436024_Megasphaera_genus_0.84
H7Y1LA101CZ30X_VT240404_Megasphaera_genus_0.84
H33VL5G02I6LT9_VT358325_Megasphaera_genus_0.84
H3FQNY001EA9JL_VT436024_Megasphaera_genus_0.86
HSM55G301AYUNP_VT592846_Megasphaera_genus_0.87
HSSIX7E01EYIJ5_VT728029_Megasphaera_genus_0.87
H7936KB01DHZQA_VT687508_Megasphaera_genus_0.9
IYNNPHI01CR23F_VT383632_Megasphaera_genus_0.83
H44Q1WL01ASRVJ_VT255221_Megasphaera_genus_0.84
H3FQNY001DP4LK_VT436024_Megasphaera_genus_0.85
HOVR1CC02I3YR5_VT589681_Megasphaera_genus_0.83
H3FQNY001BT30O_VT436024_Megasphaera_genus_0.81
H3FQNY001DCML6_VT436024_Megasphaera_genus_0.87
IWQ9V9U01C3V4B_VT100042_Megasphaera_genus_0.87
HM0131G01EN9LT_VT137170_Megasphaera_genus_0.85
HMBFVCR01CFIIS_VT177937_Megasphaera_genus_0.84
H3FQNY001EJ3QK_VT436024_Megasphaera_genus_0.92
HMZCSI402JJWW3_VT478364_Megasphaera_genus_0.89
HMZCSI402IFT6A_VT478364_Megasphaera_genus_0.89
G4RZ91202JDJ9R_VT376396_2_Megasphaera_genus_0.92
HNJDLJG02HLEEL_VT110686_Megasphaera_genus_0.91
HLYIO3A01DZ45O_VT719406_Megasphaera_genus_0.9
H4KQMWG02IWFPF_VT371026_Megasphaera_genus_0.87
H7Y1LA102GMYEZ_VT463111_2_Megasphaera_genus_0.83
H4KQMWG02IF87R_VT371026_Megasphaera_genus_0.85
IZ45HFS01A4ETR_VT697306_Megasphaera_genus_0.92
IZZVKM101AUSNF_VT870990_Megasphaera_genus_0.91
H4TS25O01CAYDK_VT063573_Megasphaera_genus_0.83
IZZVKM101A6C38_VT900277_Megasphaera_genus_0.84
IWQ9V9U02GVYQ8_VT686193_Megasphaera_genus_0.9
IV8I96M01EZQJI_VT752056_Megasphaera_genus_0.86
HPYVSFH01C3SGN_VT949145_Megasphaera_genus_0.86
HLYIO3A02JMOPW_VT695581_Megasphaera_genus_0.84
HMC7N6G01AM8TN_VT817857_Megasphaera_genus_0.88
HPRDQGN01BQLDI_VT834688_Megasphaera_genus_0.89
HSQTTYM01EVKI3_VT441479_Megasphaera_genus_0.89
IXX2AVJ02JQFE8_VT495679_Megasphaera_genus_0.81
IW2KT8M01DF1KW_VT982569_Megasphaera_genus_0.84
G402SXF02GXQJU_VT748151_2_Megasphaera_genus_0.83
HK6O9DH01BR31L_VT278146_Megasphaera_genus_0.88
H294KWW02GWLC0_VT794628_Megasphaera_genus_0.87
HMBFVCR01CTART_VT177937_Megasphaera_genus_0.83
HLYIO3A01AZ7IH_VT719406_Megasphaera_genus_0.84
IZ1V1HY02JPXZV_VT549504_Megasphaera_genus_0.83
HSQTTYM02G2DRD_VT920523_Megasphaera_genus_0.81
H4E41Y401BE2Y9_VT293336_Megasphaera_genus_0.85
IZSJMER02IUH2U_VT112915_Megasphaera_genus_0.83
G4RZ91202FL578_VT112552_Megasphaera_genus_0.87
HK8IFEH02HBTVD_VT965973_Megasphaera_genus_0.88
HLK881002JS03R_VT178572_Megasphaera_genus_0.85
HLK881002G8CDK_VT178572_Megasphaera_genus_0.92
HPRDQGN01BYW8H_VT596005_Megasphaera_genus_0.87
HP4BQPF02GIYF5_VT428500_Megasphaera_genus_0.87
HSOLGCL02F46BK_VT903795_Megasphaera_genus_0.88
HSOLGCL02H94DK_VT220557_Megasphaera_genus_0.9
HSOLGCL02F1RB3_VT998490_Megasphaera_genus_0.88
HSQTTYM01BR69I_VT367132_Megasphaera_genus_0.88
H3CASPA01CFJNR_VT856811_Megasphaera_genus_0.9
H33VL5G01DZQQ4_VT483344_Megasphaera_genus_0.9
H4KQMWG02HKKRU_VT371026_Megasphaera_genus_0.86
H4KQMWG02JQCBF_VT422007_Megasphaera_genus_0.87
IXHDERC01A4ELS_VT005946_Megasphaera_genus_0.9
IXX2AVJ02HO5RW_VT284774_Megasphaera_genus_0.86
HK6O9DH01CAL8B_VT770558_Megasphaera_genus_0.86
IXV6ZMN02GEQV9_VT893217_Megasphaera_genus_0.84
IXX2AVJ02G80RM_VT800576_Megasphaera_genus_0.86
IXX2AVJ02HN82K_VT284774_Megasphaera_genus_0.88
IYW85GN02HEG94_VT021453_Megasphaera_genus_0.84
IZSJMER02H68JX_VT996954_Megasphaera_genus_0.84
HMNZG2Z02G3KKC_VT247523_Megasphaera_genus_0.91
HK6O9DH02HHTHD_VT145917_Megasphaera_genus_0.92
HK8IFEH01A5ZJY_VT645032_Megasphaera_genus_0.92
HLK881001DRPFX_VT976504_Megasphaera_genus_0.86
HLK881002GUKWO_VT178572_Megasphaera_genus_0.9
HLYIO3A01ENGI8_VT341753_Megasphaera_genus_0.87
HNJDLJG02IOVI3_VT386211_Megasphaera_genus_0.88
HPD98O301DTQF0_VT148931_Megasphaera_genus_0.92
HP4BQPF01EE8FK_VT275868_Megasphaera_genus_0.9
HSB0UXZ02JQYR4_VT135902_Megasphaera_genus_0.92
H294KWW02FSWB0_VT947597_Megasphaera_genus_0.87
H3CASPA02HZGT6_VT986791_Megasphaera_genus_0.92
H4E41Y402F9NES_VT094660_Megasphaera_genus_0.91
H4KQMWG02JA232_VT371026_Megasphaera_genus_0.86
IUS25TI01BLJE5_VT825351_Megasphaera_genus_0.83
IXX2AVJ02IAUEJ_VT800576_Megasphaera_genus_0.9
IXX2AVJ02JQT8N_VT800576_Megasphaera_genus_0.89
IYPH69O01B7TWM_VT589467_Megasphaera_genus_0.84
IZSJMER01C68KD_VT451025_Megasphaera_genus_0.92
IZSJMER01D11U4_VT451025_Megasphaera_genus_0.88
IZQTTEK01DUI1Y_VT970467_Megasphaera_genus_0.87
0.008

Figure 7. Intratype phylogenetic analysis of *Megasphaera* phylotype 2

A total of 25016S rDNA reads in the V1-V3 region that were most closely related to *Megasphaera* phylotype 2 were combined, aligned using MUSCLE and pruned for informative regions using Gblocks. A phylogenetic tree was created using PhyML. There were no distinct clades or groupings. These sequences were all remarkably similar, and as such no subtypes were discovered.

Figure 7.


0.823
IYVKAAE01CY4L0_VT953126_Megasphaera_genus_0.89
IXX2AVJ02JCABK_VT284774_Megasphaera_genus_0.86
HM0131G02J1MR4_VT851467_Megasphaera_genus_0.88
0.823
HM0131G01BZ2LY_VT985171_Megasphaera_genus_0.92
I0CFRFZ01DCZ72_VT215897_Megasphaera_genus_0.82
IYNNPHI02HBNQY_VT125933_Megasphaera_genus_0.93
IUS25TI01DZITF_VT329744_Megasphaera_genus_0.93
HNJDLJG01COD2N_VT162261_Megasphaera_genus_0.97
IWQ9V9U01DWY4L_VT621923_Megasphaera_genus_0.9
IYPH69O02G9RZU_VT243739_Megasphaera_genus_0.83
IYLXL3G02IZJSQ_VT567428_Megasphaera_genus_0.94
IZDUHHU02ILUEP_VT815062_Megasphaera_genus_0.91
HSQTTYM02JT40L_NC_81_Megasphaera_genus_0.87
HL7G61402IKB84_VT719966_Megasphaera_genus_0.93
IY94QEU02F23OF_VT672141_Megasphaera_genus_0.87
HSQTTYM02HXZCH_NC_81_Megasphaera_genus_0.94
I0H0XM301DYLOW_VT239480_Megasphaera_genus_0.94
IZSJMER01ED9T7_VT124489_Megasphaera_genus_0.93
HLYIO3A01C0BCO_VT125758_Megasphaera_genus_0.9
HSQTTYM02JRV8E_NC_81_Megasphaera_genus_0.85
IWNTGXV01BCHAU_VT103144_Megasphaera_genus_0.92
HLYIO3A01D81IR_VT125758_Megasphaera_genus_0.91
IZQTTEK02I3QKO_VT083486_Megasphaera_genus_0.98
HMZCSI401CAPK9_VT364692_Megasphaera_genus_0.92
H44Q1WL01BEEDN_VT927751_Megasphaera_genus_0.86
HMZCSI402IG2RM_VT503537_Megasphaera_genus_0.85
H4E41Y401BL4L5_VT604485_Megasphaera_genus_0.9
IWQ9V9U01EQ28C_VT545638_Megasphaera_genus_0.92
HMBFVCR01BMTUU_VT969826_Megasphaera_genus_0.92
IWNTGXV01C7SNZ_VT641648_Megasphaera_genus_0.89
H4E41Y401DWM9O_VT604485_Megasphaera_genus_0.91
IV8I96M01BI0L0_VT447322_Megasphaera_genus_0.95
IWO9DDE01DQN0F_VT315375_Megasphaera_genus_0.93
HO8SVR401BKXYT_VT225426_Megasphaera_genus_0.9
HNHTBJG01D00GY_VT672240_Megasphaera_genus_0.96
HMC7N6G02F2EHV_VT004304_Megasphaera_genus_0.87
H33VL5G01BGH7T_VT432931_Megasphaera_genus_0.95
IWO9DDE01CCRIZ_VT315375_Megasphaera_genus_0.92
I0IVRE302HU6A8_VT458782_Megasphaera_genus_0.89
HNHTBJG01DZ291_VT672240_Megasphaera_genus_0.93
H4E41Y402IXLVH_VT700924_2_Megasphaera_genus_0.85
H33VL5G02GTUIO_VT389286_Megasphaera_genus_0.97
IWO9DDE01BP0D1_VT315375_Megasphaera_genus_0.93
I0DBA8W01EA6AX_VT136374_Megasphaera_genus_0.94
I0H0XM301C0Z62_VT677196_Megasphaera_genus_0.91
0.82
HP4BQPF02HTA78_VT227807_Megasphaera_genus_0.93
IWNTGXV01BH589_VT641648_Megasphaera_genus_0.97
IYLXL3G01D62GQ_VT423102_Megasphaera_genus_0.94
IWCPZUC02G00Q5_VT357429_Megasphaera_genus_0.94
IXHDERC01CEKFT_VT966254_Megasphaera_genus_0.92
IYNNPHI02GU2CR_VT125933_Megasphaera_genus_0.91
H4KQMWG02GWNCW_VT167916_Megasphaera_genus_0.94
I0H0XM301DHKHP_VT239480_Megasphaera_genus_0.96
IWCPZUC02HW4W7_VT357429_Megasphaera_genus_0.92
IWNTGXV01C8FVT_VT641648_Megasphaera_genus_0.93
HSB0UXZ02IN3RG_VT135902_Megasphaera_genus_0.93
HLYIO3A01EH7PN_VT125758_Megasphaera_genus_0.97
IXHDERC01C727J_VT688145_Megasphaera_genus_0.9
IWCPZUC02GWRSM_VT357429_Megasphaera_genus_0.92
H4TS25O02I73BZ_VB254227_Megasphaera_genus_0.91
I0H0XM301DJF9V_VT677196_Megasphaera_genus_0.92
I0H0XM301A9L57_VT239480_Megasphaera_genus_0.86
IWCPZUC02F1ESL_VT357429_Megasphaera_genus_0.96
HSQTTYM02G9MVP_NC_81_Megasphaera_genus_0.95
HT4DY7B01EPAY8_VT835050_Megasphaera_genus_0.94
HNHTBJG01ASB5H_VT672240_Megasphaera_genus_0.96
IXV6ZMN01A10JI_VT349087_Megasphaera_genus_0.81
IV8I96M01DFZG6_VT364883_Megasphaera_genus_0.99
HSQTTYM02IDT3S_NC_81_Megasphaera_genus_0.92
IZBYCNM01B9A08_VT032432_Megasphaera_genus_0.81
I0H0XM301AJ7LP_VT239480_Megasphaera_genus_0.91
IUS25TI02FT1F2_VT952863_Megasphaera_genus_0.91
HSQTTYM02JF7I2_NC_81_Megasphaera_genus_0.9
H4R5VD002JAZ08_VT780865_Megasphaera_genus_0.87
HSB0UXZ02HLZ45_VT660641_Megasphaera_genus_0.85
HNFPXU202HQLWI_NC_67_Megasphaera_genus_0.88
IV8I96M02G86RC_VT012588_Megasphaera_genus_0.92
HO8SVR402HQ2MU_VT236017_Megasphaera_genus_0.97
H4CWMSP02GOPRZ_NC_98_Megasphaera_genus_0.92
HLHFZ5W02G84QO_VT267271_Megasphaera_genus_0.96
I0CFRFZ01BM90J_VT215897_Megasphaera_genus_0.84
H78CB4001CY9HL_VT770985_Megasphaera_genus_0.97
IWLNX5K01BOTUT_VT279420_Megasphaera_genus_0.91
IWNTGXV01COYPG_VT103144_Megasphaera_genus_0.88
HSM55G302JO6FH_VT942433_Megasphaera_genus_0.92
HSQTTYM02IN7R2_NC_81_Megasphaera_genus_0.91
H4KQMWG01EF3FH_VT210099_Megasphaera_genus_0.93
HM8BITT02H0WK5_VT682895_Megasphaera_genus_0.95
HT74IMC01BMFT3_VT435160_Megasphaera_genus_0.81
H4KQMWG01BN6GD_VT210099_Megasphaera_genus_0.93
IXK2KBS01CUQVC_VT173757_Megasphaera_genus_0.81
H33VL5G01B2X5K_VT179361_Megasphaera_genus_0.96
HPRDQGN01CMDYP_VT141449_Megasphaera_genus_0.9
HMBFVCR01EXD0W_VT969826_Megasphaera_genus_0.92
H4KQMWG01AKXXS_VT210099_Megasphaera_genus_0.95
H6IYJGT01DWPSJ_VT447725_Megasphaera_genus_0.8
H3FQNY002GJYYI_VT133817_Megasphaera_genus_0.94
H33VL5G01A3XIV_VT163837_Megasphaera_genus_0.92
HM0131G01AU90J_VT985171_Megasphaera_genus_0.94
IXI859T02HOOXZ_VT068018_Megasphaera_genus_0.88
I0H0XM301CLRVY_VT239480_Megasphaera_genus_0.85
HSQTTYM02IMN94_NC_81_Megasphaera_genus_0.99
IWQ9V9U01DULZD_VT545638_Megasphaera_genus_0.91
H5UVSMJ01BMBTZ_VT724969_Megasphaera_genus_0.88
H4KQMWG01EDCXN_VT210099_Megasphaera_genus_0.93
HM0131G02HR27U_NC_66_Megasphaera_genus_0.89
HSOLGCL02FQM74_VT220557_Megasphaera_genus_0.94
IW501Y202G7B9G_VT206289_Megasphaera_genus_0.96
I0CFRFZ02I2P8C_VT741850_Megasphaera_genus_0.96
HSQTTYM02J390X_NC_81_Megasphaera_genus_0.85
0.888
H33VL5G01EII9V_VT432931_Megasphaera_genus_0.9
HM0131G02GQ8OS_VT046379_Megasphaera_genus_0.88
HP4BQPF01BLIIZ_VT162026_Megasphaera_genus_0.96
H4KQMWG02H6W5A_VT189245_Megasphaera_genus_0.94
HPAAPY201DMXXO_VT745237_Megasphaera_genus_0.8
HSQTTYM02J0DBK_NC_81_Megasphaera_genus_0.8
HLHFZ5W01BPGW5_VT897189_Megasphaera_genus_0.88
HO8SVR401D21SO_VT534468_Megasphaera_genus_0.98
IY94QEU02HTQM6_VT672141_Megasphaera_genus_0.93
H4CWMSP02H6HBE_VT343154_2_Megasphaera_genus_0.82
HL7G61402GPC4Y_VT892810_Megasphaera_genus_0.85
IZUIP2N02GDBL9_VT686537_Megasphaera_genus_0.83
H4E41Y401CVIIL_VT604485_Megasphaera_genus_0.9
HPAAPY202HH6BI_VT454870_Megasphaera_genus_0.93
HM0131G02IJTNN_VT851467_Megasphaera_genus_0.95
HM8BITT02JTJMN_VT682895_Megasphaera_genus_0.8
IWO9DDE01BO13S_VT315375_Megasphaera_genus_0.87
HPJIY6R01BF14S_VT543635_Megasphaera_genus_0.87
IYLXL3G02HLBWH_VT406521_Megasphaera_genus_0.89
HMNZG2Z01AVXWJ_VT399877_Megasphaera_genus_0.89
IXHDERC01DPLLJ_VT688145_Megasphaera_genus_0.91
HMZCSI402J3KR9_NC_65_Megasphaera_genus_0.9
H7W123V01AL9HT_VT814755_Megasphaera_genus_0.82
HSB0UXZ02IMA40_VT135902_Megasphaera_genus_0.91
HL7G61402I1A47_VT892810_Megasphaera_genus_0.94
I0H0XM301BJ6J8_VT239480_Megasphaera_genus_0.93
HO6KIYN02G6SBO_VT174499_Megasphaera_genus_0.91
HSQTTYM02H8BRF_NC_81_Megasphaera_genus_0.94
IYLXL3G02H8CCU_VT575885_Megasphaera_genus_0.93
HLYIO3A02JSXWB_VT685001_Megasphaera_genus_0.98
H4E41Y401EODLD_VT604485_Megasphaera_genus_0.93
H4TS25O02JTYFX_VB254227_Megasphaera_genus_0.9
I0H0XM301C8WZ5_VT239480_Megasphaera_genus_0.93
IZUIP2N01B8KY8_VT599948_Megasphaera_genus_0.83
G402SXF01AWJ51_VT796059_2_Megasphaera_genus_0.94
IXHDERC02HEV0W_VT225545_Megasphaera_genus_0.88
H4TS25O02GTXM2_VB254227_Megasphaera_genus_0.92
IXI859T01DTIVR_VT749770_Megasphaera_genus_0.92
0.897
IWCPZUC01CZPCA_VT187109_Megasphaera_genus_0.97
IV8I96M01DYY0N_VT339365_Megasphaera_genus_0.97
HNJDLJG01DJ34H_VT162261_Megasphaera_genus_0.93
H4TS25O02GNR4V_VB254227_Megasphaera_genus_0.95
IXBJUXJ01D2602_VT939079_Megasphaera_genus_0.88
I0H0XM301CZOHW_VT239480_Megasphaera_genus_0.9
IY4F6EU01BUAG6_VT090880_Megasphaera_genus_0.85
HPJIY6R01BU8GW_VT377163_Megasphaera_genus_0.9
IWNTGXV01CZN7W_VT103144_Megasphaera_genus_0.81
IWNTGXV01DCVJO_VT103144_Megasphaera_genus_0.88
H4CWMSP01A8YMA_VT689464_Megasphaera_genus_0.81
IZDUHHU02I3JXF_VT815062_Megasphaera_genus_0.8
H4R5VD002HGE8G_VT780865_Megasphaera_genus_0.87
IYYY82M02JK4NR_VT003145_Megasphaera_genus_0.83
IWNTGXV01DFS9T_VT103144_Megasphaera_genus_0.89
IWNTGXV01AKF4O_VT641648_Megasphaera_genus_0.87
H4VQDQH02JA7CT_VT505749_Megasphaera_genus_0.84
H7CGWIO02GCUQ8_VT933145_Megasphaera_genus_0.84
HSM55G301CFJVV_VT585361_Megasphaera_genus_0.91
IWNTGXV01BA9TZ_VT641648_Megasphaera_genus_0.83
IWNTGXV01DFD7Y_VT641648_Megasphaera_genus_0.87
H4TS25O01DBA2X_VT817444_Megasphaera_genus_0.92
IWNTGXV01A9OAN_VT103144_Megasphaera_genus_0.86
IXBJUXJ01B0XN9_VT164015_Megasphaera_genus_0.87
IY4F6EU01D9EFL_VT090880_Megasphaera_genus_0.87
I0IVRE302JDPY1_VT458782_Megasphaera_genus_0.86
I0CFRFZ01BFTW1_VT215897_Megasphaera_genus_0.84
IWNTGXV01A762X_VT103144_Megasphaera_genus_0.92
I0H0XM301DZB8A_VT239480_Megasphaera_genus_0.94
I0H0XM301CDJ63_VT239480_Megasphaera_genus_0.96
HSQTTYM02GTH5Q_NC_81_Megasphaera_genus_0.92
IW501Y202JTHRG_VT911403_Megasphaera_genus_0.95
HSQTTYM02JH100_NC_81_Megasphaera_genus_0.89
IWLNX5K02H8C0P_VT768888_Megasphaera_genus_0.89
HSQTTYM02G56UR_NC_81_Megasphaera_genus_0.94
HSQTTYM02JFYCJ_NC_81_Megasphaera_genus_0.91
IWNTGXV01CPIEX_VT103144_Megasphaera_genus_0.85
IWNTGXV01EI0VN_VT103144_Megasphaera_genus_0.94
IWNTGXV01DXH3A_VT641648_Megasphaera_genus_0.86
IWNTGXV01EW3NK_VT103144_Megasphaera_genus_0.9
IYYY82M02GPGSX_VT789750_Megasphaera_genus_0.93
I0H0XM301EVDLA_VT340926_Megasphaera_genus_0.91
HSQTTYM02IVAZQ_NC_81_Megasphaera_genus_0.93
HML2IQB01DIJAM_VT116594_Megasphaera_genus_0.92
HNFPXU202F8MN4_NC_67_Megasphaera_genus_0.91
H4R5VD002GA82S_VT627036_Megasphaera_genus_0.82
HLHFZ5W02FKQCL_VT610296_Megasphaera_genus_0.85
H6IYJGT01DU1Y2_VT447725_Megasphaera_genus_0.86
I0EBK1V01EVGRW_VT388951_Megasphaera_genus_0.83
I0H0XM301A0GDY_VT239480_Megasphaera_genus_0.97
I0DBA8W01BWVUX_VT136374_Megasphaera_genus_0.92
I0H0XM301CEQTR_VT239480_Megasphaera_genus_0.93
IY0RXEJ01ECXSJ_VT946254_Megasphaera_genus_0.87
IY4F6EU01C7FKQ_VT090880_Megasphaera_genus_0.84
I0H0XM301CUPVF_VT239480_Megasphaera_genus_0.94
I0EBK1V01DAMHO_VT388951_Megasphaera_genus_0.92
IY4F6EU01DTEE5_VT023962_Megasphaera_genus_0.88
IXHDERC01EXX0R_VT688145_Megasphaera_genus_0.89
H4CWMSP02HVJ9R_NC_98_Megasphaera_genus_0.91
IW2KT8M02ILTIG_VT005680_Megasphaera_genus_0.91
IXBJUXJ01BCPHM_VT164015_Megasphaera_genus_0.86
HSOLGCL01BN2ZV_VT754141_Megasphaera_genus_0.9
IWNTGXV01BDKRS_VT103144_Megasphaera_genus_0.93
I0H0XM301A9CAX_VT239480_Megasphaera_genus_0.91
I0H0XM301BLFKA_VT239480_Megasphaera_genus_0.88
I0H0XM301ER1TL_VT239480_Megasphaera_genus_0.92
IYYY82M01BEK59_VT256882_2_Megasphaera_genus_0.91
I0H0XM301CDBAG_VT239480_Megasphaera_genus_0.93
HSQTTYM02IC88W_NC_81_Megasphaera_genus_0.91
IWNTGXV01EC5HH_VT103144_Megasphaera_genus_0.9
IWNTGXV01ESYIC_VT641648_Megasphaera_genus_0.94
IWNTGXV01B0OA1_VT641648_Megasphaera_genus_0.94
IWNTGXV01CX6Z8_VT103144_Megasphaera_genus_0.91
HSQTTYM02FLIY3_NC_81_Megasphaera_genus_0.88
IZUIP2N02GPKEF_VT686537_Megasphaera_genus_0.84
I0H0XM301CBN5K_VT239480_Megasphaera_genus_0.91
IWQ9V9U02FU5XO_VT760186_Megasphaera_genus_0.87
HT4DY7B01EJ638_VT991520_Megasphaera_genus_0.91
I0H0XM301EFIHM_VT239480_Megasphaera_genus_0.89
IY94QEU02I1WO2_VT672141_Megasphaera_genus_0.9
IYW85GN02GWGES_VT036658_Megasphaera_genus_0.93
IWYKQJK02HR9IU_VT898749_Megasphaera_genus_0.87
IWNTGXV01CV3WM_VT641648_Megasphaera_genus_0.92
H4E41Y401D3CNZ_VT604485_Megasphaera_genus_0.93
H4E41Y401DFCI5_VT604485_Megasphaera_genus_0.88
HPNJN9402JN591_VT878722_Megasphaera_genus_0.88
HSQTTYM02FYXPJ_NC_81_Megasphaera_genus_0.91
HSQTTYM02FUQ6J_NC_81_Megasphaera_genus_0.94
H78CB4002FV0B9_VT244877_Megasphaera_genus_0.94
IWLNX5K02I61G7_VT768888_Megasphaera_genus_0.94
IZSJMER01BNBSS_VT124489_Megasphaera_genus_0.96
IZSJMER01CHTLK_VT124489_Megasphaera_genus_0.97
H4KQMWG01CF0IF_VT210099_Megasphaera_genus_0.9
HK8IFEH02H5WA8_VT399873_Megasphaera_genus_0.9
G402SXF01ELVVI_VT796059_2_Megasphaera_genus_0.87
IZSJMER01AXLYE_VT124489_Megasphaera_genus_0.96
IZZVKM102IWA1C_VT233687_Megasphaera_genus_0.95
I0CFRFZ02I7CYP_VT741850_Megasphaera_genus_0.93
I0H0XM301DF6SQ_VT239480_Megasphaera_genus_0.95
I0CFRFZ02HZTUB_VT169345_Megasphaera_genus_0.89
0.006

Megasphaera phylotype 1 seemed to have some smaller groups, with very few informative SNP differences, *Megasphaera* phylotype 2 was nearly completely uniform (Fig.6, Fig.7).

Finally, in order to determine experimentally if the USEARCH algorithm alone could distinguish correctly between the two phylotypes two large datasets were curated. The first contained all of the reads classified to "*Megasphaera* cluster52" generated through the VaHMP project totaling 439,530 reads at their full length. A second dataset was curated containing all reads classified to "*Megasphaera* cluster52" trimmed to the minimum allowable read length of 200 bp, in order to assess if short length reads could be accurately distinguished between the two phylotypes as well, The USEARCHv4.0 global alignment algorithm was used on both the trimmed dataset (200bp) as well as full length dataset, aligning reads to the Vaginal 16S rDNA Reference Database (Edgar, 2010; Fettweis, Serrano, Sheth, Mayer, Glascock, Brooks, et al., 2012b). A cutoff of 97% was used. Regardless of read length, the USEARCH analysis yielded similar results. The USEARCH method alone was capable of subclassifying *Megasphaera* cluster52 reads into *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 reads even using the minimum allowable read length of 200bp. There was some very minimal cross-classification between the two phylotypes between the full-length and short read analyses, likely attributable to lack of informative SNPs or sequencing error. After accounting for these reads, USEARCH was determined to call short reads (200bp) and full-length reads at the phylotype level concordantly 99.4% of the time. We decided to use this approach for sub-classification of the reads in further analyses (Table 4).

Table 4. USEARCH correctly classifies *Megasphaera* with concordance of 99.4%

Megasphaera sp.	Full Length Reads	Trimmed Reads	Net Change
Megasphaera phylotype 1	285,300	292,930	+7,630
Megasphaera phylotype 2	137,212	139,181	+1,969
Megasphaera BV3-C16	2,862	3,022	+160
Megasphaera elsdenii	17	18	+1
Megasphaera micronuciformis	1,170	1,330	+160
Megasphaera massiliensis	614	759	+145
Megasphaera sueciensis	0	0	0
Megasphaera paucivorans	0	0	0
All	427,175	437,240	+10,065

Read clustering analysis reveals a third vaginal *Megasphaera* taxon

After determining that we could sub-classify our *Megasphaera* cluster52 reads into the two vaginal phylotypes using the approach initially employed by the STIRRUPS classifier, we updated the *Megasphaera* reference library. We included all of the *Megasphaera* species for which the V1-V3 region of the 16S rRNA gene was available at NCBI: *M. elsdenii, M. cerevisiae, M. indica, M. massiliensis, M. micronuciformis, M. paucivorans, M. sueciensis, Megasphaera* phylotype 1 and *Megasphaera* phylotype 2. We also examined the raw reads that were classified as *Megasphaera* for the presence of other vaginal *Megasphaera* species. We performed an AbundantOTU clustering analysis on all of our reads that were assigned to *Megasphaera* by the RDP classification (Ye, 2011). We discovered three new clusters. One matched 100% to *Anaeroglobus geminatus*, a taxon originally isolated from a human oral sample, which likely represents a misclassified *Megasphaera* species based on its phylogenetic placement in 16S rRNA and multi-gene trees (see Figure 24). The second cluster matched 96% to the species *Megasphaera massiliensis,* an organism isolated from human fecal samples (Padmanabhan et al., 2013). Given that the identity is 96% it may represent a new species of bacteria. However, prevalence and proportion were both low, and it did not meet the criteria outlined for inclusion in the reference database for STIRRUPS classification.

Finally the third cluster mapped to an organism isolated from a human vaginal swab, whose genome has been sequenced. This organism was named BV3C16-1 (GCA_000478965.1), suggesting to us that it may have been a bacterial vaginosis isolate, although this is not mentioned in the online documentation of the two biosamples cited (SAMN00829149, SAMN02436562). Interestingly, when we added this new sequence to the reference library and re-classified our *Megasphaera* reads we found that it was present at low abundance in four women. We examined the health history data from the women who showed vaginal carriage of BV3C16-1, and found that they all had a current diagnosis of BV. Given the low prevalence of carriage, we did not pursue additional association studies with this organism.

Reclassification of all *Megasphaera* reads

Sometimes, the RDP classifier is unable to call a genus level classification with a bootstrap score of 0.8 or greater and may only be able to classify to the family, order, class, phylum or even kingdom level with confidence. To accurately reclassify *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 reads, we reexamined all reads terminally classified (*e.g.,* with a bootstrap score of 0.8 or greater) at any taxonomic level above the genus level of *Megasphaera.* Thus, we retrieved all reads with terminal classification of Veillonellaceae, Clostridiales, Clostridia, Firmicutes and Bacteria. *Megasphaera* was recently moved from the Clostridia class to the class Negativicutes and from the order Clostridiales to the order Selenomonadales. However, recent publications suggest that they should be placed back into the Clostidia class (Campbell, Adeolu, & Gupta, 2015; Yutin & Galperin, 2013). Here we used the earlier taxonomy for our work, as it was that used in the initial VaHMP analysis. Notably, a number of reads that had only been accurately classified Bacteria were assigned to both phylotypes using the STIRRUPS approach with the updated *Megasphaera* library (Table 5).

After completing this analysis, we developed a suite of scripts, which calculated the *Megasphaera* read profile for each sample in the dataset. This calculated the total number of reads assigned to each *Megasphaera* species or phylotype for a given sample. Another script extracted the previous *Megasphaera* columns from a large data file containing the sample ID,

Table 5. *Megasphaera* reads classified to higher taxonomic levels by RDP[a]

	Megasphaera	Veillonella	Clostridiales	Clostridia	Firmicutes	Bacteria
M. phylotype 1	287,054	458,182	10,411	254	154	122
M. phylotype 2	155,077	54,162	554	27	17	66
M. BV3-C16	13	0	0	0	0	0
M. elsdenii	91	0	0	0	0	0
M. cerevisiae	0	0	0	0	0	0
M. indica	20	0	0	0	0	0
M. massiliensis	2,291	0	0	0	0	0
M. micronuciformis	2,269	0	0	0	0	0
M. paucivorans	0	0	0	0	0	0
M. sueciensis	0	0	0	0	0	0

[a] Columns indicate the RDP classification results of the reads in question. Rows indicate the new USEARCH and updated reference database methodology results for those reads. Most reads are appropriately assigned to the genus *Megasphaera* by the RDP classifier but many of the *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 reads were classified at higher taxonomic levels.

16S rDNA data and health history data. It then added ten columns in the same space containing the updated *Megasphaera* 16S rDNA read numbers by matching up read IDs with sample IDs and finding the sample ID row in the data file. After this process was complete, we had sub-classified all of the *Megasphaera* reads terminal at any branch above the species level by RDP for every read in the 2013 Data Drop and replaced the old *Megasphaera* classification data with an updated, more informative and comprehensive *Megasphaera* dataset. The dataset was integrated with other 16S rDNA data and participant health history and demographic data to permit phylotype-specific association studies.

***Megasphaera* co-occur with other BV-associated bacteria**

We wanted to identify bacterial species that co-occur with one or both of the two vaginal *Megasphaera* phyotypes. Given that both phylotypes are obligately anaerobic organisms and associated with a clinical diagnosis of bacterial vaginosis, we hypothesized that the two phylotypes would often co-occur with other anaerobic organisms and be associated with BV (Zozaya-Hinchliffe et al., 2008). We also hypothesized that they would not often co-exist in samples with high levels of *Lactobacillus* species. In order to examine microbial co-occurrence, we developed a script that would calculate prevalence of each organism across the entire cohort. The script then performed a pairwise calculation to determine the proportion of the cohort in which two given species would co-exist simply by chance. This can be calculated by simply multiplying together the proportions at which each taxon is found in the total cohort. The script then calculated the actual prevalence of each pairwise co-occurrence and compared the two measures. The script generated a ratio value, which describes the relationship between the actual prevalence and the prevalence expected by chance of pairwise co-occurrence.

Based on this measure, if the value is high, the two organisms co-occur more likely than simply by chance. If the value is one, the two organisms co-occur as often as one would expect by chance, and if the value is below one then the two organisms occur less often than expected by chance suggesting a possible antagonistic relationship. The highest values calculated for each phylotype were with species of the genus *Prevotella*, a relatively common vaginal and oral taxon. *Prevotella* species have been associated with bacterial vaginosis but also exist in "healthy" vaginal communities (Datcu et al., 2014; Vitali et al., 2015). As expected, *Megasphaera* phylotype 1 often co-occurred with other BV-associated organisms such as *Parvimonas*, BVAB2, BVAB3, *Fusobacterium* and *Megasphaera* phylotype 2. *Megasphaera* phylotype 2 often co-occurred with other BV-associated organisms including *Sneathia, Peptoniphilus, Mobiluncus,* BVAB1, BVAB2, BVAB3, *Gemella, Anaerococcus, Dialister* and *Megasphaera* phylotype 1. *Megasphaera* phylotype 2 also co-occured with *Fusobacterium,* an uncultivated *Sneathia* species and other organisms associated with disease states including *Neisseria gonorrhoeae*, the causative agent of gonorrhea and *Mycoplasma hominis* and *Ca.* Mycoplasma girerdii, both previously strongly associated with trichomoniasis (Table 6) (Fettweis, Serrano, et al., 2014b; Rappelli et al., 1998).

Although these two organisms were both associated with BV-associated taxa and shared a few common co-associations including to each other, they seem to have somewhat unique co-occurrence profiles. We hypothesize that there may be different community types associated with BV, a disorder that is largely characterized by the presence of diverse anaerobic organisms instead of the presence of certain specific organisms, and that these subtypes may be unique. While these two *Megasphaera* phylotypes can co-occur in samples, they may represent members of two different community subtypes of BV. Further research using more statistically rigorous

Table 6. Microbial actual:expected co-occurrence ratio reveals synergistic relationships

	Megasphaera phylotype 1	*Megasphaera* phylotype 2
Megasphaera phylotype 1	N/A	3.20
Megasphaera phylotype 2	3.20	N/A
Lachnospiraceae BVAB1	2.02	2.56
Clostridiales BVAB2	2.72	2.93
Clostridiales BVAB3	2.85	4.31
Prevotella cluster2	1.95	2.35
Parvimonas OTU142	2.58	3.17
Fusobacterium gonidiaformans/equinum	1.06	4.38
Sneathia OTU65	1.14	6.40
Peptoniphilus lacrimalis	2.72	4.71
Mobiluncus mulieris	2.31	4.61
Gemella OTU86	3.62	3.51
Anaerococcus tetradius	1.66	2.93
Dialister cluster51	2.92	3.45
Mycoplasma hominis	2.08	3.78
Ca. Mycoplasma girerdii	1.32	3.07
Neisseria gonorrhoeae	1.90	5.74

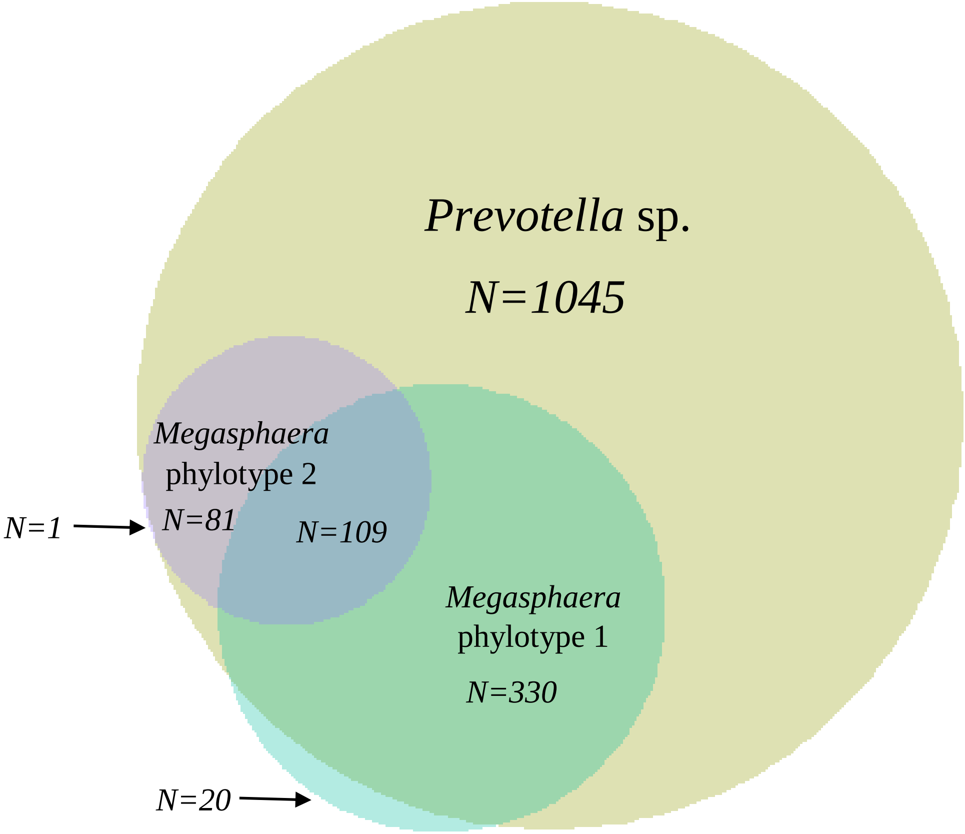
methodology would be necessary to confirm these associations and additional studies are needed to begin to understand how these organisms interact in the context of their host-microbiome communities.

Interestingly, when examining the strongest co-occurrence results between *Megasphaera* phyotopes and *Prevotella* sp., we found that most samples with either vaginal phylotype or both also have *Prevotella* in the sample. Using 0.1% as a definition of presence, we found that 99.5% of *Megasphaera* phylotype 2 positive samples also have *Prevotella* and 95.6% of *Megasphaera* phylotype 1 positive samples also have *Prevotella* (Fig.8). The most common taxon of *Prevotella* in these samples from the VaHMP STIRRUPS analysis is *Prevotella* cluster2, made up of the four species *Prevotella timonensis, Prevotella buccalis, Prevotella* OTU46 and *Prevotella* OTU47. *Prevotella timonensis* is the most prevalent among these four organisms. Determining which species within the cluster is actually present would require developing a sub-classification method for those species, which is beyond the scope of this work. Due to the strong correlation between *Prevotella* and the two *Megasphaera* phylotypes and the stringent growth requirements of vaginal *Megasphaera,* we hypothesize that the two vaginal phylotypes may possibly have a dependence on *Prevotella* species, or one of their metabolites or other products, for growth or survival. Alternatively, the *Prevotella* species may be earlier colonizers in vaginal biofilm formation that is associated with BV, and *Megasphaera* phylotypes may directly or indirectly require *Prevotella* species for co-aggregation. Additional studies are required to understand the relationship between these species in the vaginal microbiome community.

Figure 8. *Megasphaera* phylotypes largely co-occur with *Prevotella* species

Bacterial taxa were considered present if they comprised at least 0.1% of the microbiome. Venn diagram was created using the ‘venneuler’ package for R. *Megasphaera* phylotype 1 co-occurs with *Prevotella* in the microbiome 95.6% of the time, while *Megasphaera* phylotype 2 co-occurs with *Prevotella* in the microbiome 99.5% of the time.

Figure 8.


Prevotella sp.
N=1045
Megasphaera
phylotype 2
N=81
N=109
N=1
Megasphaera
phylotype 1
N=330
N=20

Vaginal *Megasphaera* demographic associations

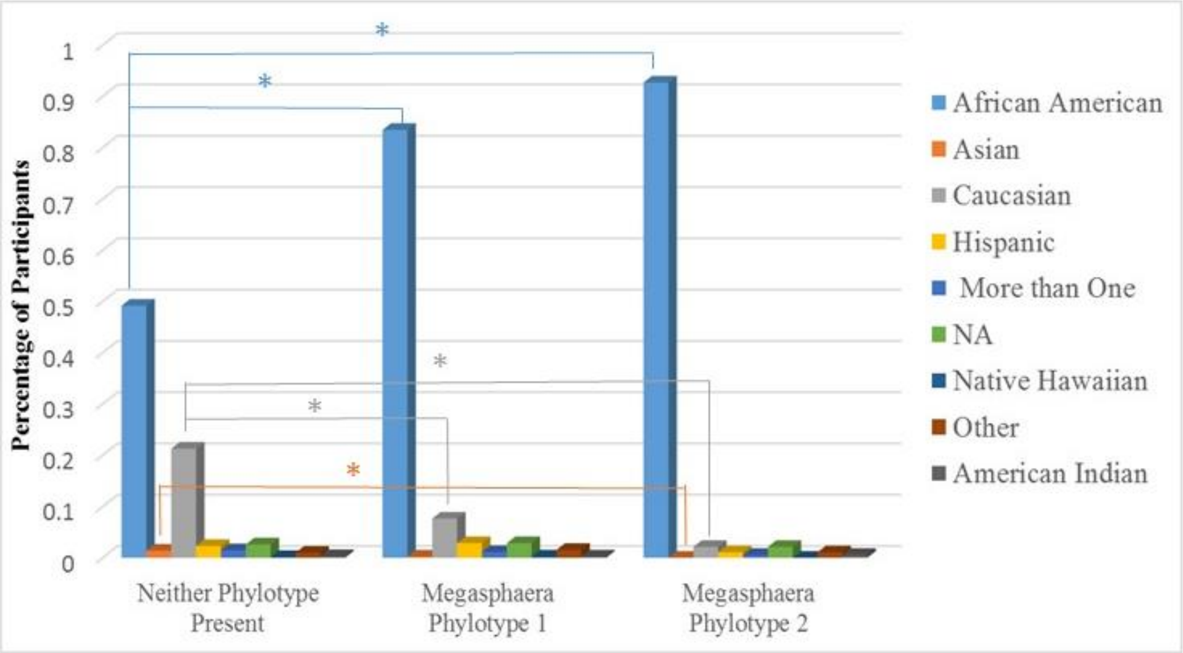
After integrating the updated *Megasphaera* columns into the VaHMP 2013 Data Drop dataset including all health history data, demographic and clinical associations with each phylotype were analyzed by comparing the prevalence of each phylotype between different demographic measures, diagnoses, behavioral measures, *etc.* The cohort for this analysis was made up of 2,633 participants. The size of this cohort is smaller than the previous cohort used for the sub-classification validation because these samples had not yet been fully bioinformatically processed and connected with their medical and health history data at the outset of this study. Twins, pregnant women (by self-report, diagnosis, or clinic ID), samples processed using the Qiagen DNA extraction kit instead of the MoBio DNA extraction kit and those whose samples did not pass the threshold of 5,000 reads during 16S rDNA sequencing were excluded. Twins were excluded to avoid genetic factors skewing the results of the analysis. Pregnant women were excluded because pregnancy has been shown to have an effect on the vaginal microbiome composition. The Qiagen-processed samples were excluded to avoid introducing DNA extraction-related bias in a small number of samples. Samples with less than 5,000 reads were excluded to avoid low sequencing quality. The overall prevalence of *Megasphaera* phylotype 1 was 17.4% and the overall prevalence of *Megasphaera* phylotype 2 was 7.3%. 4.1% of profiles contained both *Megasphaera* phylotype 1 *and Megasphaera* phylotype 2.

Both phylotypes were significantly associated with African ancestry and negatively associated with European ancestry. *Megasphaera* phylotype 1 was also negatively associated with Asian ancestry (Fig.9). Both phylotypes were also strongly associated with lower socioeconomic status with the highest prevalence of each phylotype existing in the "Under $1 5,000/year" income bracket. Both phylotypes were also significantly negatively associated with

Figure 9. ***Megasphaera*** **phylotypes strongly associated with ethnicity**

This analysis was performed on a cohort of 2,633 women passing the following inclusion criteria: non-pregnant (by clinic ID, self-report or diagnosis), non-twin, sample yielded at least 5,000 reads, process using MoBio Power Soil DNA extraction kit. Presence of a bacterial taxon is defined as comprising at least 0.1% of the microbiome. Statistical significance was assessed using a two-tailed Student's T-test with an alpha level of 0.05. Significant associations between ethnicity and taxa are denoted with an asterisk.

Figure 9.


Percentage of Participants
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
*
*
*
*
*
Neither Phylotype Present
Megasphaera Phylotype 1
Megasphaera Phylotype 2
African American
Asian
Caucasian
Hispanic
More than One
NA
Native Hawaiian
Other
American Indian

the higher income brackets of "$40-59,000/year", "$60-79,000/year" and "$80,000+/year" (Fig.10). Both phylotypes were also strongly associated with lower levels of education.

Although only 65 women in the cohort reported having a current female sexual partner, we found a significant association between women who have sex with women and *Megasphaera* phylotype 1 (p=0.006). This finding was previously reported in another publication (Fethers, Marks, Mindel, & Estcourt, 2000). An immediate cause for this correlation is not clear. We also found a strong association between douching and *Megasphaera* phylotype 1 (p=0.004). It has been previously described in the literature that douching is a risk factor for developing bacterial vaginosis as it removes the normal flora from the vagina (Cottrell, 2010). Both phylotypes were also associated with increased vaginal pH, a marker of vaginal dysbiosis and bacterial vaginosis and an increased number of lifetime sexual partners. In those who have had more than 20 partners, the prevalence of both phylotypes is nearly double the total prevalence in the cohort. A few of the positive associations with these phylotypes are often comorbid with poverty including the strong association with lower levels of education. This makes it difficult to determine what is a true association and what is a covariate considering that many of these parameters are not independent.

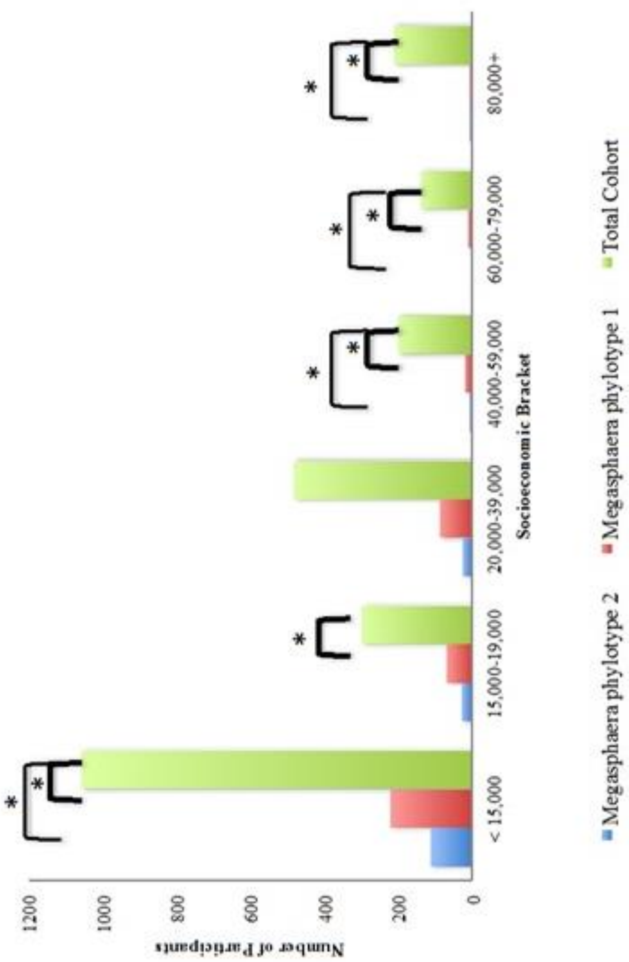
Vaginal *Megasphaera* exhibit differential clinical associations

Many previous publications had noted that *Megasphaera* phylotype 1 was associated with bacterial vaginosis (Datcu et al., 2014; Fethers et al., 2012; Zozaya-Hinchliffe et al., 2008).It has also been suggested to be used a diagnostic marker in first void urine samples due to

Figure 10. *Megasphaera* phylotypes associated with lower socioeconomic status

This analysis was performed on a cohort of 2,633 women passing the following inclusion criteria: non-pregnant (by clinic ID, self-report or diagnosis), non-twin, sample yielded at least 5,000 reads, processed using MoBio Power Soil DNA extraction kit. Presence of a bacterial taxon is defined as comprising at least 0.1% of the microbiome. Statistical significance was assessed using a two-tailed Student's T-test with an alpha level of 0.05. Significant associations between socioeconomic status and taxa are denoted with an asterisk.

Figure 10.


Number of Participants
0
200
400
600
800
1000
1200
< 15,000
15,000-19,000
20,000-39,000
40,000-59,000
60,000-79,000
80,000+
Socioeconomic Bracket
*
Megasphaera phylotype 2
Megasphaera phylotype 1
Total Cohort

its strong associations with the infection (Datcu et al., 2014). *Megasphaera* phylotype 2 has also been associated with BV, but to a lesser degree. One paper suggested that *Megasphaera* phylotype 1 was associated with elevated HIV viremia in the lingual microbiome (Dang et al., 2012). *Megasphaera* phylotype 2 was described as potentially correlated to trichomoniasis in a single publication (Zozaya-Hinchliffe et al., 2008). We wanted to see if we could confirm and extend these observations of differential associations with vaginal infection in the VaHMP cohort.

Vaginal itching, discharge and odor can be a symptom of common vaginal infections or sexually transmitted disease. We found that *Megasphaera* phylotype 2 was significantly correlated with self-report of vaginal itching, odor and discharge while *Megasphaera* phylotype 1 was significantly correlated with vaginal odor and discharge. We then analyzed the prevalence of vaginal infections and sexually transmitted diseases among those who had either one or both of the two phylotypes. Both phylotypes had a positive association with all of the following vaginal infections and diseases: genital herpes, chlamydia, gonorrhea, syphilis, trichomoniasis, and bacterial vaginosis. However, they were not positively correlated with viral diseases, urinary tract infections or yeast infections. We performed a relative risk analysis, a common epidemiological tool for analyzing the risk of disease associated with a given taxa. We analyzed three common vaginal infections of interest: bacterial vaginosis, yeast infections and trichomoniasis. Yeast infections are caused by an overgrowth of the eukaryotic *Candida albicans* and not typically correlated with anaerobic organisms common among women with bacterial vaginosis and trichomoniasis (Svobodová, Lysková, & Hamal, 2015). We selected a panel of taxa for comparison that were strongly associated with either BV or trichomoniasis. The relative risk analysis revealed that while both organisms conferred an increased risk of bacterial vaginosis, the risk conferred by *Megasphaera* phylotype 1was much greater than *Megasphaera* phylotype 2 (Table 7). In fact the risk associated with *Megasphaera* phylotype 1 for bacterial vaginosis was higher than that of *Gardnerella vaginalis,* on organism previously thought of as a marker of BV (Leppäluoto, 2011). Intriguingly, we also observed a strong association between *Megasphaera* phylotype 2 and trichomoniasis. This association, while not as strong as the near 100% association of "*Ca.* Mycoplasma girerdii", a likely symbiont, was stronger than the association of trichomoniasis with *Mycoplasma hominis. M. hominis* has repeatedly been associated with trichomoniasis infections and has been shown to live out part of its life inside of the trichomonad (Rappelli et al., 1998; 2001). These findings support the previous report that that *Megasphaera* phylotype 2 was associated with trichomoniasis. These findings also support the hypothesis that these two organisms play distinct roles in the vaginal microbiome.

Megasphaera **phylotype 1 not excluded in pregnancy**

Recent studies in the field of vaginal microbiome research have shown that pregnant women typically have a higher prevalence of protective *Lactobacillus* species than non-pregnant women (MacIntyre et al., 2015; Romero et al., 2014). Coincidentally, it has also been demonstrated that many bacterial vaginosis associated organisms such as *Gardnerella vaginalis* are less prevalent in pregnant women (MacIntyre et al., 2015). We case-matched a cohort of 421 pregnant women based on age, race and socioeconomic status. All matches had to be within plus or minus one year of age and plus or minus one socioeconomic bracket. The ethnicity match was exact. For each one of these samples, the relative proportions of a number of bacterial vaginosis associated organisms and other disease-associated organisms were calculated. Student's two-tailed T-tests were performed at an alpha level of 0.05 to determine if there was a statistically

Table 7. Relative Risk Analysis

Relative risk values and 95% confidence interval values are shown. Bacterial taxa were determined to be present if at least 0.1% of the reads from the sample were assigned to that taxon. The outpatient cohort used for this analysis (n=2633) is comprised of non-twin, non-pregnant participants (by diagnosis, self-report or clinic-ID). Samples met the threshold of at least 5,000 reads and were processed using the MoBio Power Soil DNA extraction kit. Vaginal infection status was determined based on clinician diagnosis at time of visit.

Table 7.

	Bacterial Vaginosis	Trichomoniasis	Yeast Infection
Megasphaera phylotype 1	6.01 (5.15-7.02)	1.87 (1.11-3.16)	1.01 (0.71-1.43)
Megasphaera phylotype 2	3.55 (3.00-4.19)	4.70 (2.79-7.90)	1.31 (0.83-2.05)
Gardnerella vaginalis	5.45 (3.68-8.07)	2.33 (1.12-4.85)	0.60 (0.45-0.78)
Prevotella cluster2	4.13 (3.34-5.12)	1.69 (1.04-2.75)	0.40 (0.30-0.54)
BVAB2	3.41 (2.89-4.03)	1.00 (0.60-1.68)	0.43 (0.30-0.63)
Sneathia amnii	3.00 (2.51-3.60)	2.07 (1.28-3.36)	0.53 (0.39-0.71)
Mycoplasma hominis	2.36 (2.01-2.78)	3.66 (2.29-5.86)	0.94 (0.68-1.30)
"Ca. Mycoplasma girerdii"	0.93 (0.58-1.49)	18.44 (11.85-28.69)	0.68 (0.29-1.62)

significant difference in the prevalence of any of these organisms between the pregnant and non-pregnant cohorts. The major trend, which has been observed by other groups, was that BV and disease associated organisms tend to be lower in prevalence in the pregnant cohort. The decreased prevalence of the following organisms was found to be significant: *Gardnerella vaginalis, Atopobium vaginae, Prevotella* cluster2, *Sneathia amnii, Dialister* cluster51, *Dialister micraerophilus, Sneathia sanguinegens, Mobiluncus mulieris* and Clostridiales BVAB3. Other organisms where the trend was less prevalence in pregnancy but the results were not significant included *Ureaplasma urealyticum/parvum, Prevotella amnii, Mycoplasma hominis, Megasphaera* phylotype 2 and *Mobiluncus curtisii* (Fig.11).

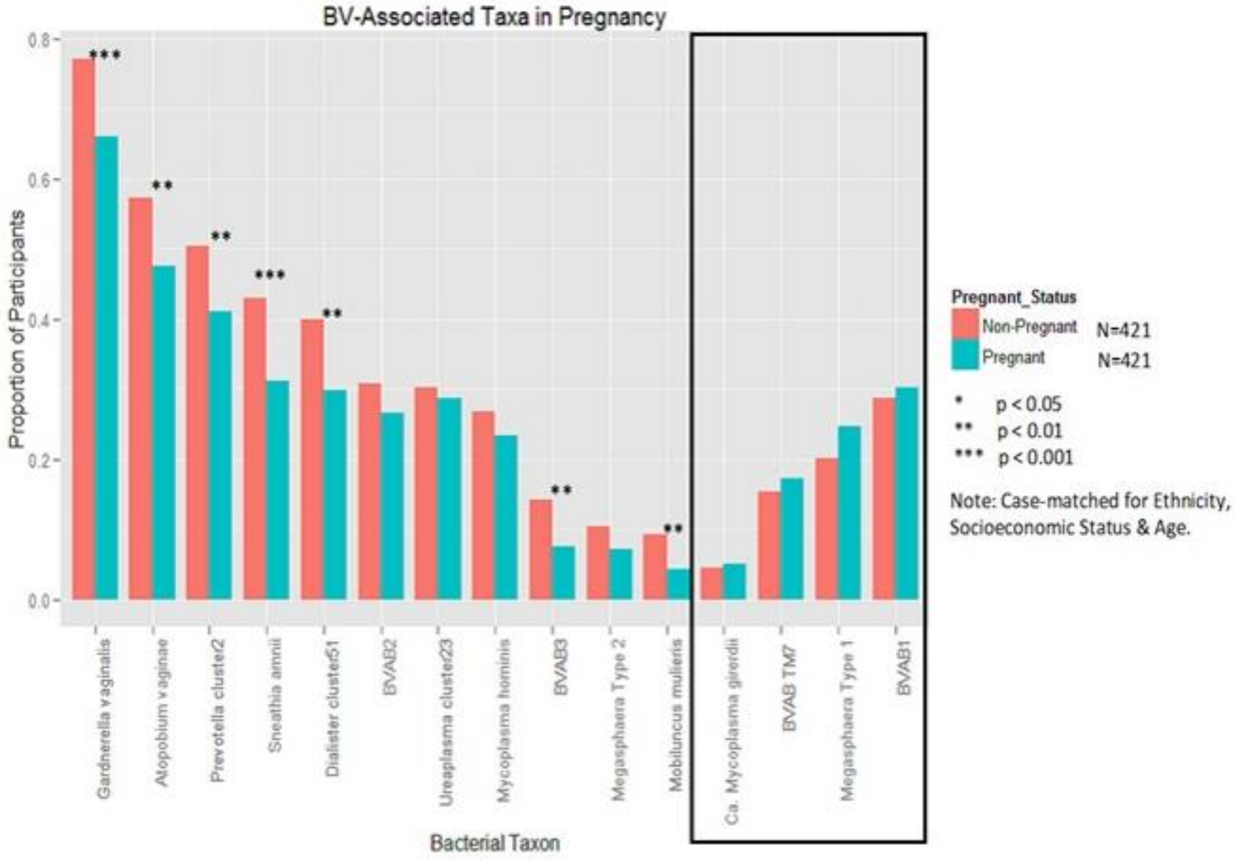
Surprisingly, some other BV and disease associated organisms were actually more prevalent in the pregnant cohort. None of these findings were significant at a 0.05 alpha level. The following organisms followed the trend of being more prevalent in pregnancy: Lachnospiraceae BVAB1, "*Ca.* Mycoplasma girerdii", Phylum TM7 bacterium and *Megasphaera* phylotype 1. The p-value for the *Megasphaera* phylotype 1 prevalence was 0.12. Because we expected that due to its bacterial vaginosis-associated status that *Megasphaera* phylotype 1 would be less prevalent in pregnancy and it is not, we decided to perform a statistical test to determine if it can be stated with confidence that it is not less prevalent in pregnant women. In order to test this, we performed a non-parametric Mann-Whitney U Test. We accepted the null hypothesis and found that the measurements came from the same population (p=0.06).

Megasphaera phylotype 1 is of special importance because it has also recently been shown to be capable of invading the upper genital tract in one publication examining the microbiota present in the UGT of women having hysterectomies (Mitchell et al., 2015). It has

Figure 11. BV and disease associated organisms in pregnancy

This analysis was performed on a case-matched group of 421 pregnant and 421 non-pregnant women recruited through the VaHMP project at VCU. Individuals were case-matched based on age, socioeconomic status and ethnicity. Presence was defined as comprising at least 0.1% of the microbiome. Statistical analysis was performed using a two-tailed Student’s T-test with an alpha level of 0.05. Organisms were selected for study based on their association with bacterial vaginosis or another reproductive disease state, such as trichomoniasis.

Figure 11.


BV-Associated Taxa in Pregnancy
Proportion of Participants
0.8
0.6
0.4
0.2
0.0
***
**
**
***
**
**
**
Gardnerella vaginalis
Atopobium vaginae
Prevotella cluster2
Sneathia amnii
Dialister cluster51
BVAB2
Ureaplasma cluster23
Mycoplasma hominis
BVAB3
Megasphaera Type 2
Mobiluncus mulieris
Ca. Mycoplasma girerdii
BVAB TM7
Megasphaera Type 1
BVAB1
Bacterial Taxon
Pregnant_Status
Non-Pregnant N=421
Pregnant N=421
* p < 0.05
** p < 0.01
*** p < 0.001
Note: Case-matched for Ethnicity, Socioeconomic Status & Age.

also been associated with the pro-inflammatory and labor-related lipid 12-HETE in a single paper addressing the metabolomics of BV (Srinivasan et al., 2015). These associations by no means solidify its role as an antagonistic agent in pregnancy but it does highlight this organism as a potential target for future study in pregnant cohorts, especially in women at high risk for preterm labor and/or delivery.

Identification, cultivation, isolation and sequencing of vaginal *Megasphaera*

To better understand what underlying genomic features may contribute to the differential associations with vaginal infections exhibited by the two *Megasphaera* phylotypes, we attempted to culture and sequence their genomes. No genomes representative of these two phylotypes were publicly available at the beginning of this study. Vaginal samples were targeted for cultivation based on the 16S rRNA profiles and presence of previously uncultivated taxa including the two vaginal *Megasphaera* phylotypes. The 16S rRNA gene of colonies was used to identify *Megasphaera* isolates by PCR amplification using a PCR protocol developed specifically for this screening project and universal 16S primers (Table 8). We were able to successfully cultivate and isolate one *Megasphaera* phylotype 1 clone on ThermoScientific Remel Brucella blood agar medium and two *Megaspahera* phylotype 2 clones both on ThermoScientific Remel chocolate agar medium. The genomic DNA prepared for the single *Megasphaera* phylotype 1 clone was sequenced using the Roche 454 GS FLX Titanium Pyrosequencing platform according to standard protocols as described by the manufacturer (F.Hoffman-La Roche AG). The raw reads were assembled using Newbler v2.8. Genomic DNA from the two *Megaspahera* phylotype 2 clones was sequenced on the Illumnia MiSeq platform according to standard protocols as described by the manufacturer (Illumina, Inc.). Reads were assembled using Velvet. Quality

Table 8. PCR protocol for 16S gene amplification

Step	Temperature	Time (min)
1	94°C	0:30
2	94°C	0:30
3	48°C	0:30
4	72°C	1:30
5	Repeat Steps 1-4 29 times	
6	72°C	10:00
7	4°C	∞

Each PCR reaction tube contained 45µl of Invitrogen Platinum SuperMix, 2µl of deionized sterile water, 1µl of 16S forward primer mix and 1µl of reverse primer thoroughly homogenized on ice.

trimming, low complexity filtering, and Poly-A, Poly-T and N filtering were also implemented to increase the quality of the assemblies.

The single *Megaspahera* phylotype 1 genome, OTU70, was assembled into 129 contigs and had an estimated genome size of 1.8Mb. The largest contig was 281,917bp in length and the assembly coverage was 244X. The average read lengths for the two *Megaspahera* phylotype 2 clones, identified as strains M2-4 and M2-8, were 107bp and 106bp respectively. The M2-4 genome was assembled into 311 contigs with the largest contig being 214,182bp in size. The M2-4 genome was estimated to be 1.74Mb in size and the assembly had a coverage level of 675X. The M2-8 genome was assembled into 328 contigs, comparable to M2-4, and its largest contig was 283,583bp in length. The projected size of the M2-8 genome was 1.71Mb and the coverage level of the assembly was 2203X.

Single colonies from each isolate were used to inoculate 1.0mL of sBHI + 20% glycerol for future culture work. However, after multiple attempts at resurrecting the clones for confirmatory biochemical analyses of genomic observations, they were unable to be recovered. Scrapings from the single *Megasphaera* phylotype 1 and both *Megasphaera* phylotype 2 isolates were used to inoculate ThermoScientific Remel chocolate agar plates pre-heated to 37°C, ThermoScientific Remel 5% Sheep Blood Brucella blood agar plates pre-heated to 37°C and 5mL tubes of sBHI+s containing 10% human serum. Plates were stored anaerobically at 37°C for 24-96 hours in three nested Ziploc bags containing an Anaeropack. Tubes were loosely capped to allow gas exchange and stored in an anaerobic incubator at 37°C with 5% CO_2 for 24-96 hours. Multiple rounds of this approach were attempted with scrapings. Finally, the frozen cultures were thawed and resuspended in 5mL of sBHI+s. This yielded no growth. A second duplicate tube created when storing frozen cultures was also used for scrapings with the same resurrection methodology. Unfortunately, none of these efforts yielded growth of either phylotype.

Genome analysis and annotation of six vaginal *Megasphaera*

Six genomes were selected for comparative genomic analyses including three cultivated at VCU and three genomes publicly available at NCBI including two *Megasphaera* phylotype 1 genomes: strains 28L and UPII 199-6 (GCA_000177555.1, GCA_000214495.2) and one *Megasphaera* phylotype 2 genome: strain UPII 135-E (GCA_000221545.2). The three genomes sequenced at VCU included one *Megasphaera* phylotype 1 genome: strain OTU70 and two *Megasphaera* phylotype 2 genomes: strains M2-4 and M2-8. Thus, the comparative genomic analysis contained an equal number of *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 genomes.

Basic genome comparisons were performed including calculation of genome size, number of protein coding genes and syntenic conservation (Table 9). *Megasphaera* phylotype 1 genomes were slightly larger and encoded more genes on average than the *Megasphaera* phylotype 2 genomes. The average genome size of the three *Megasphaera* phylotype 1 genomes was 1.72Mb while the average genome size of the *Megasphaera* phylotype 2 genomes was 1.70Mb. The average number of protein-coding genes in the *Megasphaera* phylotype 1 genomes was 1606 while the average number of protein-coding genes in the *Megasphaera* phylotype 2 was 1536. Later metabolic and functional analyses would reveal that these small differences could have been attributed to a loss of protein-coding genes involved in amino acid biosynthesis and nucleotide salvage. Discrepancies in genome size could also be attributed to various sizes of

Table 9. Genomic characteristics of six vaginal *Megasphaera* genomes

	Megasphaera phylotype 1			*Megasphaera* phylotype 2		
	OTU70	28L	UPII 199-6	M2-4	M2-8	UPII 135-E
Genome Size (Mb)	1.78	1.73	1.64	1.74	1.71	1.65
GC Percentage	46.33	46.05	46.37	38.94	39.09	38.88
No. of Contigs	129	34	45	311	328	49
N50 Length	179993	156177	100595	102411	131070	64000
No. Contigs @ N50	4	5	7	6	5	8
Transcriptome Size (Mb)	1.55	1.55	1.46	1.46	1.41	1.44
No. Predicted Genes	1647	1715	1457	1591	1508	1510

phage genetic material, existing embedded in the genomes.

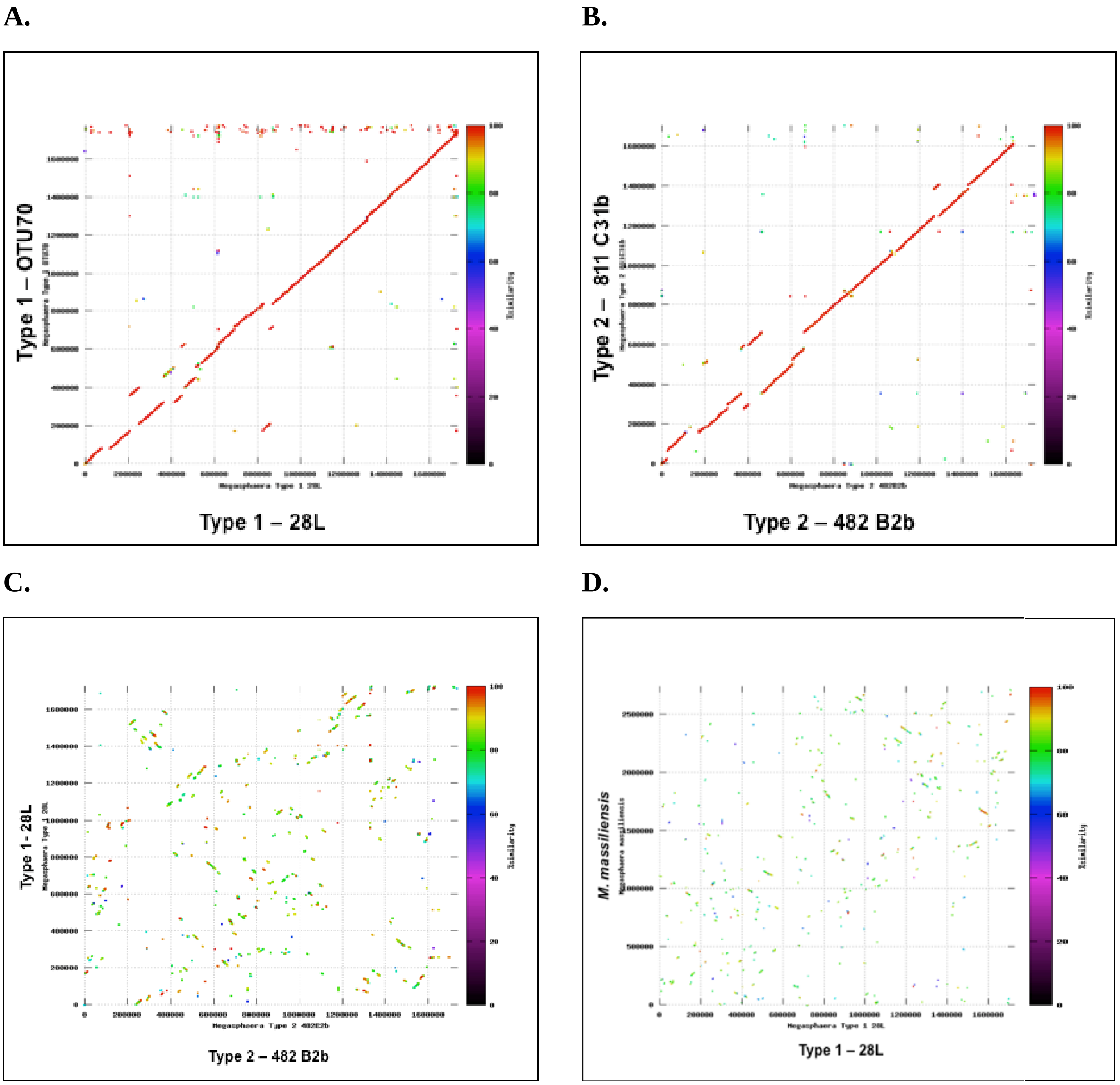
Syntenic analysis was also performed at both the protein and nucleotide level to examine if the synteny was conserved both between genomes of the same phylotype and between genomes of different phylotypes. Synteny describes the arrangement of the genes and other genetic information in the genome. You would expect two very closely related organisms to have more conserved synteny than distantly related organisms. We first analyzed the syntenic conservation among *Megasphaera* phylotype 1 genomes (Fig.12). The synteny was strongly conserved among all of the genomes and the sequence similarity at the protein level was nearly identical across the entire length of the genome. We then examined *Megasphaera* phylotype 2 genomic synteny, which likewise was highly conserved. Finally we examined the conservation of synteny between the two distinct phylotypes. There was massive genome rearrangement with no large stretches of sequence being conserved. Even at the protein level, the sequence similarity was found to average approximately 80 percent. We compared syntenic conservation between *Megasphaera* phylotype 1 and another *Megasphaera* species, *Megasphaera massiliensis.* The level of syntenic conservation was similar to that found between *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2 genomes, suggesting that although these two organisms are closely related at the 16S rDNA level, they likely represent distinct species.

Protein-encoding genes were identified in all genomes and GC composition was calculated for each protein-encoding gene. We analyzed average GC percentage of protein-encoding genes as well as whole genome GC composition. The average GC composition of genes was conserved among the phylotypes but was starkly different between the two groups. The average GC percentage of a gene in a *Megasphaera* phylotype 1 genome was 46.25 percent

Figure 12. Syntenic conservation between phylotypes resembles that of different species

Panel of protein level syntenic analyses created using PROmer, gnuplot and MUMmerplot. Panel A shows strong syntenic conservation and sequence similarity between genomes of *Megasphaera* phylotype 1. Panel B shows a similarly string syntenic conservation and sequence similarity between *Megasphaera* phylotype 2 genomes. Panel C shows the complete loss of synteny between the two phylotypes and decreased sequence similarity. Panel D is used for comparison of syentenic conservation between two distinct species within the same genus. This panel shows the syntenic conservation between a *Megasphaera* phylotype 1 genome and a *Megasphaera massiliensis* genome.

Figure 12.


A.
Type 1 – OTU70
Megasphaera Type 1 28L
%similarity
Type 1 – 28L
B.
Type 2 – 811 C31b
Megasphaera Type 2 482B2b
%similarity
Type 2 – 482 B2b
C.
Type 1- 28L
Megasphaera Type 2 482B2b
%similarity
Type 2 – 482 B2b
D.
M. massiliensis
Megasphaera massiliensis
Megasphaera Type 1 28L
%similarity
Type 1 – 28L

while the average GC percentage of a gene in a *Megasphaera* phylotype 2 genome was 38.9 percent. It has yet to be fully elucidated what causes divergent GC evolution, but the current belief is that effects of extreme temperatures, extreme acidic or basic environments, loss of DNA repair genes and genome size can impact the GC composition (Nishida, 2012). Host-associated bacteria often have smaller genomes and lower GC content (Bohlin, Skjerve, & Ussery, 2008). This may hold true for *Megasphaera* phylotype 2 given that it has lost a few essential metabolic genes, seems to have a potentially dependent relationship on other bacteria and has been associated with other small host-related organisms including *Mycoplasma hominis, Neisseria gonorrhoeae,* and "*Ca.* Mycoplasma girerdii". Another hypothesis is the effect of horizontally transferred genes. The literature suggests that bacteria cannot take up DNA from the environment with a GC content higher than their own chromosome although it may take up more AT-rich DNA. This may cause the genome to become more and more GC-poor over time (Nishida, 2012). It is less likely that any environmental factors such as temperature or pH play a role in the divergent GC composition between these two phylotypes given that they inhabit the same niche (Fig. 13).

Given the divergent GC composition between the two phylotypes, we decided to analyze codon usage as well to see if this was also different between phylotypes (Fig.14). Across all six genomes, the first letter of the codon is often GC rich, even more so than the genome as a whole with most first letter GC percentage hovering above 50%. At the second letter of the codon, both phylotypes experience a sharp dip in GC percentage, which ranges between about 37-40% GC depending on the phylotype. The position of interest is the third letter in the codon because it can often be changed and not result in any functional effect on the protein due to codon degeneracy. The *Megasphaera* phylotype 1 genomes exhibit a GC percentage of approximately 47%, similar

Figure 13. GC composition of protein coding genes across phylotypes

Protein-encoding genes were identified in all genomes using Glimmer3. GC composition was calculated for each gene and plotted accordingly. All blue-colored dots represent genes from *Megasphaera* phylotype 1 genomes. All green-colored dots represent genes from *Megasphaera* phylotype 2 genomes. The average GC percentage of protein-coding genes for each phylotype is represented by a black bold line. The average GC percentage of a protein-coding gene in a *Megasphaera* phylotype 1 genome was 46.25 percent while the average GC percentage of a protein-coding gene in a *Megasphaera* phylotype 2 genome was 38.9 percent. The gene number is random and not indicative of placement in the genome.

Figure 13.

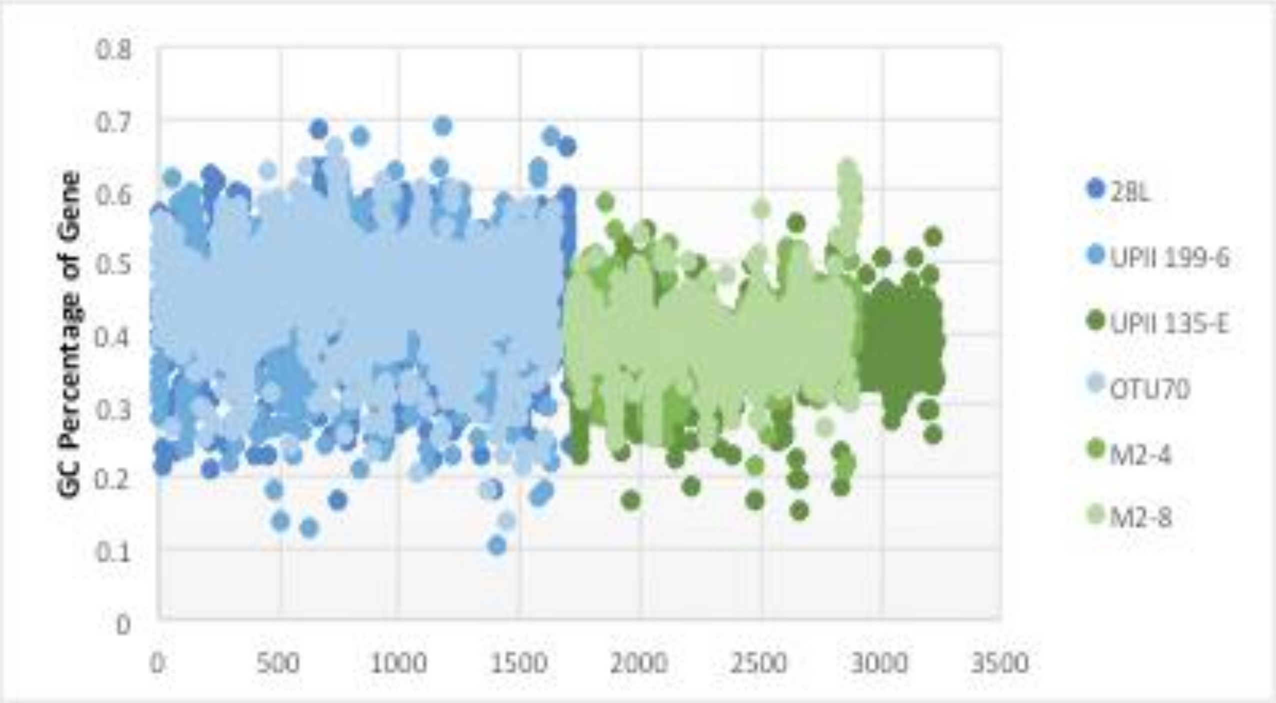

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
GC Percentage of Gene
0
500
1000
1500
2000
2500
3000
3500
● 28L
● UPII 199-6
● UPII 135-E
● OTU70
● M2-4
● M2-8

Figure 14. Phylotypes show distinct codon preference

Codon usage was calculated using cusp, a tool in the EMBOSS Tools package. GC percentages at each letter position in the codon for all six genomes are shown. *Megasphaera* phylotype 1 genomes are on the top panel *and Megasphaera* phylotype 2 genomes are on the bottom panel.

Figure 14.

Megasphaera phylotype 1 28L Coding GC 46.63% 1st letter GC 53.53% 2nd letter GC 39.67% 3rd letter GC 46.70%	*Megasphaera* phylotype 1 OTU70 Coding GC 46.92% 1st letter GC 53.72% 2nd letter GC 39.90% 3rd letter GC 47.14%	*Megasphaera* phylotype 1 UPII 199-6 Coding GC 47.05% 1st letter GC 53.86% 2nd letter GC 40.00% 3rd letter GC 47.27%
Megasphaera phylotype 2 UPII 135-E Coding GC 39.57% 1st letter GC 50.49% 2nd letter GC 37.06% 3rd letter GC 31.17%	*Megasphaera* phylotype 2 M2-4 Coding GC 39.59% 1st letter GC 50.43% 2nd letter GC 37.06% 3rd letter GC 31.29%	*Megasphaera* phylotype 2 M2-8 Coding GC 39.42% 1st letter GC 50.43% 2nd letter GC 37.03% 3rd letter GC 30.80%

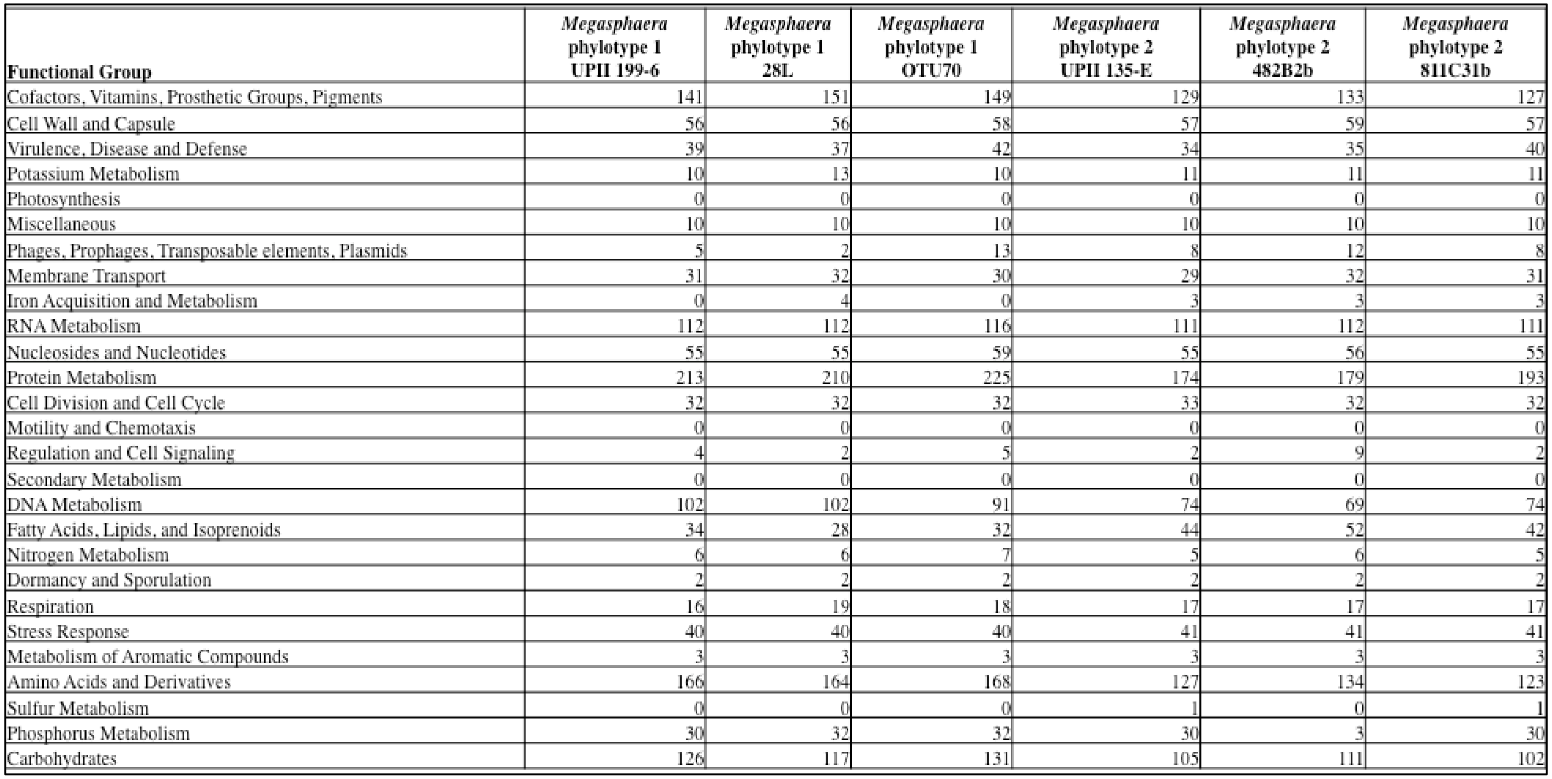
to their overall genome GC content. In contrast, the *Megasphaera* phylotype 2 genomes also exhibit an AT-rich trend with GC percentages at the third letter position between 30-31%. The mechanism for this selection causing the third letter shift and distinction in codon preference between the phylotypes has yet to be determined. We hypothesize *Megasphaera* phylotype 2 organisms may have a stronger dependence on the host, which may be related to this selection.

Functional analysis was also performed using RAST and the GAP pipeline, which uses COG and PFAM databases in coordination with rpsblast (Table 10) (Bateman et al., 2002; Marchler-Bauer et al., 2015; Meyer et al., 2008; Tatusov et al., 2000). The largest differences observed were in protein metabolism, DNA metabolism and carbohydrate metabolism. We observed phylotype-specific differences in entire amino acid biosynthetic pathways and nucleotide salvage pathways missing important components. The carbohydrate differences were largely related to the use of glucose as a carbon source. *Megasphaera* phylotype 2 was predicted to lack the enzyme hexokinase, which functions to convert glucose into glucose 6-phosphate.

***Megasphaera* phylotypes display unique metabolic strategies**

The two vaginal phylotypes exhibited distinct metabolic pathways in a number of functional categories as evidenced above. Firstly, they were unique in their carbohydrate utilization. While both phylotypes had the majority of the genes required for glycolysis encoded in the genome, *Megasphaera* phylotype 2 genomes lacked hexokinase, an essential gene required for catalyzing the reaction converting glucose to glucose-6-phosphate. Because of this loss of an essential gene, *Megasphaera* phylotype 2 organisms likely cannot use glucose as a carbon source. All of the genomes were lacking phosphoglucomutase, another enzyme in the pathway

Table 10. Functional group analysis of *Megasphaera* phylotypes using RAST



Functional Group	*Megasphaera* phylotype 1 UPII 199-6	*Megasphaera* phylotype 1 28L	*Megasphaera* phylotype 1 OTU70	*Megasphaera* phylotype 2 UPII 135-E	*Megasphaera* phylotype 2 482B2b	*Megasphaera* phylotype 2 811C31b
Cofactors, Vitamins, Prosthetic Groups, Pigments	141	151	149	129	133	127
Cell Wall and Capsule	56	56	58	57	59	57
Virulence, Disease and Defense	39	37	42	34	35	40
Potassium Metabolism	10	13	10	11	11	11
Photosynthesis	0	0	0	0	0	0
Miscellaneous	10	10	10	10	10	10
Phages, Prophages, Transposable elements, Plasmids	5	2	13	8	12	8
Membrane Transport	31	32	30	29	32	31
Iron Acquisition and Metabolism	0	4	0	3	3	3
RNA Metabolism	112	112	116	111	112	111
Nucleosides and Nucleotides	55	55	59	55	56	55
Protein Metabolism	213	210	225	174	179	193
Cell Division and Cell Cycle	32	32	32	33	32	32
Motility and Chemotaxis	0	0	0	0	0	0
Regulation and Cell Signaling	4	2	5	2	9	2
Secondary Metabolism	0	0	0	0	0	0
DNA Metabolism	102	102	91	74	69	74
Fatty Acids, Lipids, and Isoprenoids	34	28	32	44	52	42
Nitrogen Metabolism	6	6	7	5	6	5
Dormancy and Sporulation	2	2	2	2	2	2
Respiration	16	19	18	17	17	17
Stress Response	40	40	40	41	41	41
Metabolism of Aromatic Compounds	3	3	3	3	3	3
Amino Acids and Derivatives	166	164	168	127	134	123
Sulfur Metabolism	0	0	0	1	0	1
Phosphorus Metabolism	30	32	32	30	3	30
Carbohydrates	126	117	131	105	111	102

involved in the isomerization reaction of glucose-6-phosphate to fructose-6-phosphate. This switching between glucose and fructose has a relatively low activation energy requirement and may occur spontaneously in the cell. Based on the lack of this intermediate gene in both genomes, it is unlikely that either of them utilize glucose as their many energy source. *Megasphaera* phylotype 1 genomes contain a gene that catalyzes the breakdown of glycogen, often present on the vaginal epithelium. However, this pathway is also incomplete and thus it is unclear whether or not *Megasphaera* phylotype 1 organisms are able to use glycogen as an energy source.

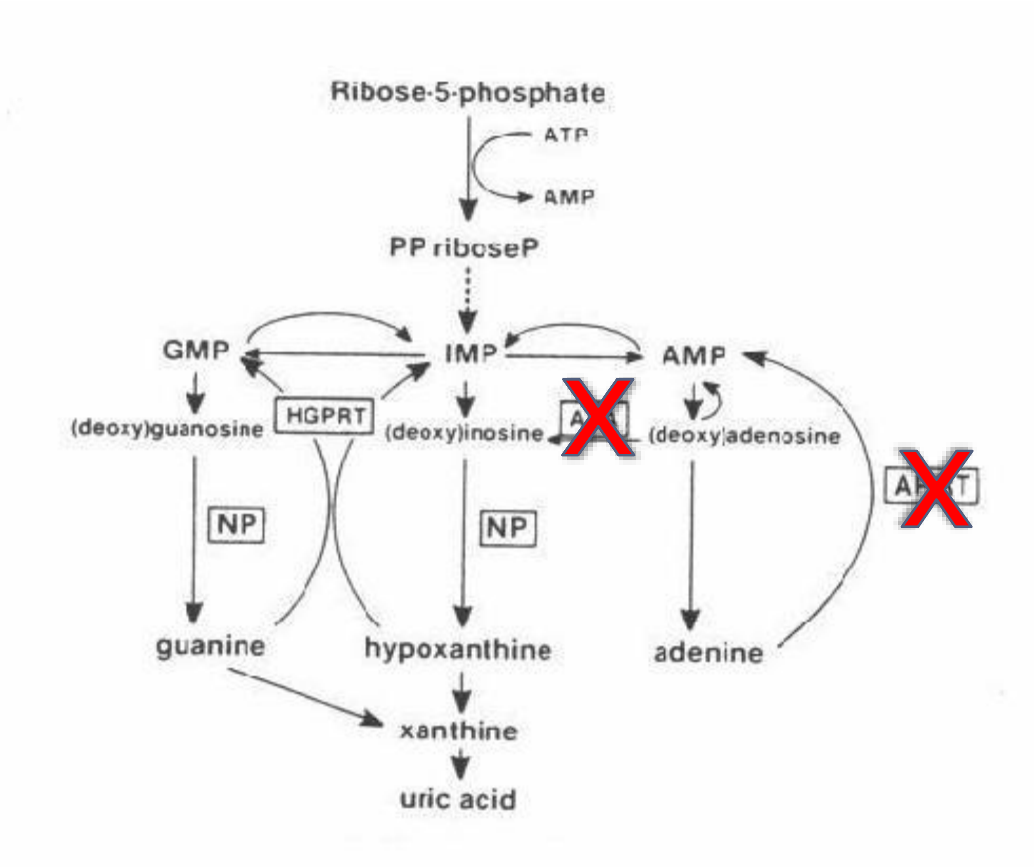
Another observed metabolic divergence between the two phylotypes was in the process of nucleotide salvage. Nucleotide salvage is pertinent to the survival of bacteria because they are constantly replicating their genetic material. This must be done efficiently and swiftly as those organisms that replicate faster will likely evolve faster and could gain a competitive advantage. This explains the push for more compact and utilitarian genomes as well as the use of nucleotide salvage pathways. Nucleotide salvage pathways utilize fragments of degraded DNA to rebuild nucleotides for new DNA strands. This saves the organism energy and replication time (Fasullo & Endres, 2015). *Megasphaera* phylotype 1 genomes are lacking both adenine phosphoribosyltransferase (APRT) and adenosine deaminase (ADA), both enzymes involved in the adenine salvage pathway (Fig. 15). Because multiple genes are knocked down, it seems as though these organisms have completely lost their ability to salvage adenine bases. They retain the other genes involved in the nucleotide salvage pathways.

Megasphaera phylotype 2 genomes encode all of the genes necessary for adenine nucleotide salvage as well as other nitrogenous bases. However, they have lost the gene cytidine deaminase, which functions in the breakdown of cytosine to uridine a precursor molecule to both

Figure 15. Adenine base nucleotide salvage pathways lost in *Megasphaera* phylotype 1

This representation shows a visualization of the various nucleotide salvage pathways of adenine nitrogenous bases. All three genomes of *Megasphaera* phylotype 1 have lost the genes adenine phosphoribosyltransferase (APRT) and adenosine deaminase (ADA), marked by large red Xs in the diagram.

Figure 15.


Ribose-5-phosphate
ATP
AMP
PP riboseP
GMP
IMP
AMP
(deoxy)guanosine
HGPRT
(deoxy)inosine
(deoxy)adenosine
NP
NP
guanine
hypoxanthine
adenine
xanthine
uric acid

cytosine and uracil. This loss of cytidine deaminase inhibits the organism's ability to efficiently recycle cytosine nitrogenous bases. This pattern is intriguing because the phylotype with the high GC composition has lost the ability to salvage adenine and the phylotype with the exceptionally low GC composition has lost the ability to salvage cytosine bases. This begs the question of whether these organisms lost the genes at some point in the past and have since had a strain on their replication speed leading to strong selection towards extreme GC divergence, or if, perhaps more likely, they have less of those bases to replicate in the genome and thus the salvage genes were lost over time due to weak selection and pressure for an ever smaller genome. This question is intriguing but could not be examined experimentally during the time course of this study.

The phylotypes also differ in their ability to synthesize amino acids. *Megasphaera* phylotype 2 genomes differ from *Megasphaera* phylotype 1 genomes in their ability to biosynthesize three important amino acids: leucine, tryptophan and cysteine. *Megasphaera* phylotype 2 organisms are capable of converting between serine and cysteine using a pathway of four conserved enzymes. *Megasphaera* phylotype 1 genomes have lost this pathway and must make cysteine bases the more time and energy consuming way. *Megasphaera* phylotype 2 genomes save energy by interconverting between amino acids based on what is freely available in the environment. *Megasphaera* phylotype 2 genomes have also lost a couple of very important pathways. They are incapable of making leucine due to the loss of five genes catalyzing the conversion of pyruvate to leucine. *Megasphaera* phylotype 1 genomes have retained this functionality. Leucine is one of three branched chain amino acids and is essential to the development of proper protein structure. *Megasphaera* phylotype 2 must be obtaining leucine from its environment or host.

Finally, *Megasphaera* phylotype 2 has lost the ability to biosynthesize the amino acid tryptophan after losing a six-gene enzymatic pathway converting chorismate to tryptophan (Fig.16, Fig. 17). Tryptophan is an interesting amino acid because it has been recently studied in terms of bacterial pathogenesis. It plays a role in chlamydial persistence as well as tissue tropism, an intriguing relationship to our study model. It is hypothesized that tryptophan may help bacteria evade IFN-γ related killing (Akers & Tan, 2006; Bhutia, Babu, & Ganapathy, 2015; Bonner, Byrne, & Jensen, 2014). This mode of bacterial killing usually acts by depleting the tryptophan and causes the cells to die off. If the organism is capable of biosynthesizing its own tryptophan, it may persist for longer periods of time.

Phylogenetic analyses reveal vaginal *Megasphaera* species to be an outgroup

In order to determine the placement of the vaginal *Megasphaera* phylotypes within the tree of life and within the family Veillonellaceae, we first performed a phylogenetic analysis of the full-length 16S rRNA gene. This gene is often used for phylogenetic analyses because it is very highly conserved across all bacterial species. It is important functionally for the translation of messenger RNA into proteins. Because a loss of function in this gene would be lethal, there is strong stabilizing pressure on the gene, preventing the fixing of substitutions along the length of the gene. Importantly, the gene contains interspersed conserved regions and nine variable regions. Because of the strength of conservation, it is easy to create universal primers to amplify

Figure 16. Tryptophan biosynthesis in *Megasphaera* phylotype 1

This metabolic map highlights which genes are present in the *Megasphaera* phylotype 1 genomes. Each color (red, yellow, blue) represents a different genome. Note the intact tryptophan biosynthesis pathway.

Figure 16.

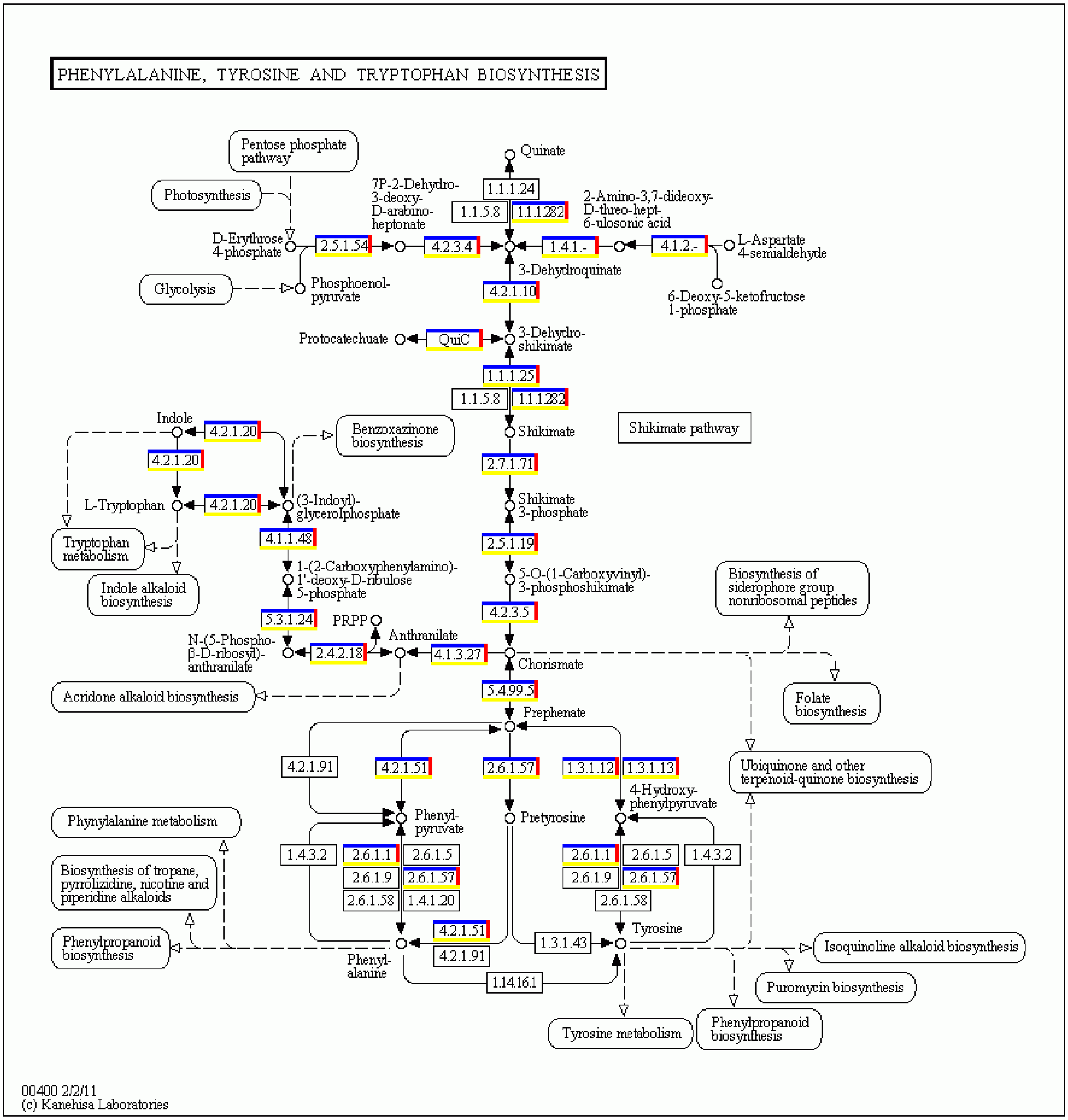
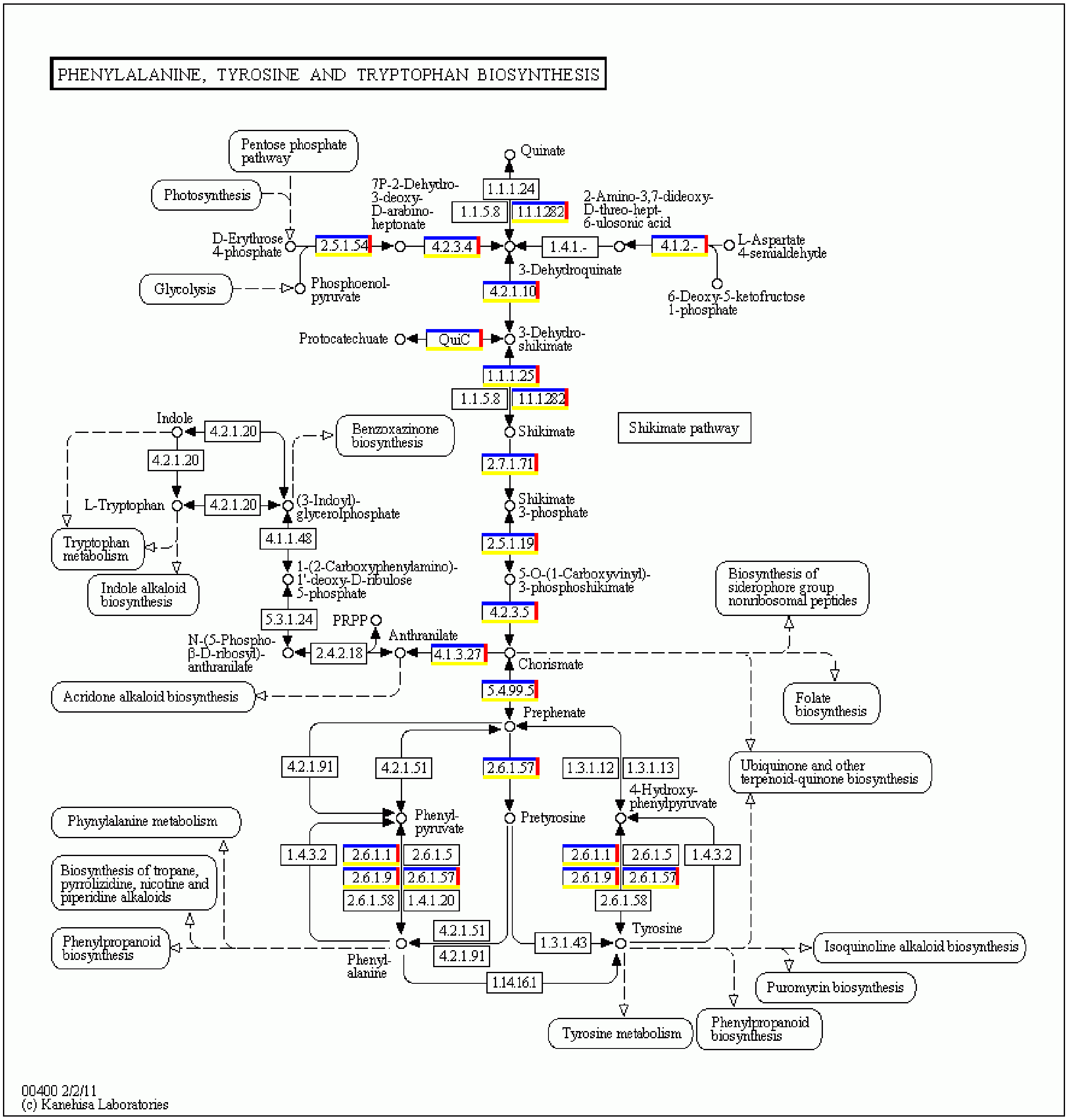

PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS
Pentose phosphate pathway
Photosynthesis
Quinate
1.1.1.24
1.1.5.8
1.1.1.282
7P-2-Dehydro-3-deoxy-D-arabino-heptonate
2-Amino-3,7-dideoxy-D-threo-hept-6-ulosonic acid
D-Erythrose 4-phosphate
2.5.1.54
4.2.3.4
1.4.1.-
4.1.2.-
L-Aspartate 4-semialdehyde
3-Dehydroquinate
Glycolysis
Phosphoenol-pyruvate
4.2.1.10
6-Deoxy-5-ketofructose 1-phosphate
Protocatechuate
QuiC
3-Dehydro-shikimate
1.1.1.25
1.1.5.8
1.1.1.282
Indole
4.2.1.20
Benzoxazinone biosynthesis
Shikimate
Shikimate pathway
4.2.1.20
2.7.1.71
L-Tryptophan
4.2.1.20
(3-Indoyl)-glycerolphosphate
Shikimate 3-phosphate
Tryptophan metabolism
4.1.1.48
2.5.1.19
1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5-phosphate
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
Indole alkaloid biosynthesis
Biosynthesis of siderophore group nonribosomal peptides
5.3.1.24
PRPP
4.2.3.5
N-(5-Phospho-β-D-ribosyl)-anthranilate
2.4.2.18
Anthranilate
4.1.3.27
Chorismate
Acridone alkaloid biosynthesis
5.4.99.5
Folate biosynthesis
Prephenate
4.2.1.91
4.2.1.51
2.6.1.57
1.3.1.12
1.3.1.13
Ubiquinone and other terpenoid-quinone biosynthesis
4-Hydroxy-phenylpyruvate
Phynylalanine metabolism
Phenyl-pyruvate
Pretyrosine
1.4.3.2
2.6.1.1
2.6.1.5
2.6.1.1
2.6.1.5
1.4.3.2
Biosynthesis of tropane, pyrrolizidine, nicotine and piperidine alkaloids
2.6.1.9
2.6.1.57
2.6.1.9
2.6.1.57
2.6.1.58
1.4.1.20
2.6.1.58
4.2.1.51
Tyrosine
Phenylpropanoid biosynthesis
1.3.1.43
Isoquinoline alkaloid biosynthesis
Phenyl-alanine
4.2.1.91
Puromycin biosynthesis
1.14.16.1
Tyrosine metabolism
Phenylpropanoid biosynthesis
00400 2/2/11
(c) Kanehisa Laboratories

Figure 17. Tryptophan biosynthesis in *Megasphaera* phylotype 2

This metabolic map highlights which genes are present in the *Megasphaera* phylotype 2 genomes. Each color (red, yellow, blue) represents a different genome. Note the lost six-gene pathway converting chorismate to tryptophan via the tryptophan biosynthesis pathway.

Figure 17.


PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS
Pentose phosphate pathway
Photosynthesis
Glycolysis
D-Erythrose 4-phosphate
Phosphoenol-pyruvate
2.5.1.54
7P-2-Dehydro-3-deoxy-D-arabino-heptonate
4.2.3.4
Quinate
1.1.1.24
1.1.5.8
1.1.1.282
2-Amino-3,7-dideoxy-D-threo-hept-6-ulosonic acid
1.4.1.-
4.1.2.-
L-Aspartate 4-semialdehyde
6-Deoxy-5-ketofructose 1-phosphate
3-Dehydroquinate
4.2.1.10
Protocatechuate
QuiC
3-Dehydro-shikimate
1.1.1.25
1.1.5.8
1.1.1.282
Shikimate
Shikimate pathway
2.7.1.71
Shikimate 3-phosphate
2.5.1.19
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
4.2.3.5
Chorismate
5.4.99.5
Prephenate
Indole
4.2.1.20
4.2.1.20
L-Tryptophan
4.2.1.20
(3-Indoyl)-glycerolphosphate
Benzoxazinone biosynthesis
Tryptophan metabolism
Indole alkaloid biosynthesis
4.1.1.48
1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5-phosphate
5.3.1.24
PRPP
N-(5-Phospho-β-D-ribosyl)-anthranilate
2.4.2.18
Anthranilate
4.1.3.27
Acridone alkaloid biosynthesis
Biosynthesis of siderophore group nonribosomal peptides
Folate biosynthesis
Ubiquinone and other terpenoid-quinone biosynthesis
4.2.1.91
4.2.1.51
2.6.1.57
1.3.1.12
1.3.1.13
Phenyl-pyruvate
Pretyrosine
4-Hydroxy-phenylpyruvate
Phynylalanine metabolism
1.4.3.2
2.6.1.1
2.6.1.5
2.6.1.9
2.6.1.57
2.6.1.58
1.4.1.20
2.6.1.1
2.6.1.5
1.4.3.2
2.6.1.9
2.6.1.57
2.6.1.58
Biosynthesis of tropane, pyrrolizidine, nicotine and piperidine alkaloids
Phenylpropanoid biosynthesis
4.2.1.51
4.2.1.91
Phenyl-alanine
1.3.1.43
Tyrosine
1.14.16.1
Isoquinoline alkaloid biosynthesis
Puromycin biosynthesis
Tyrosine metabolism
Phenylpropanoid biosynthesis
00400 2/2/11
(c) Kanehisa Laboratories

the 16S rRNA gene from the genome. Due to the presence of the variable regions, we are able to determine the relatedness of species within the tree of life.

The 16S rRNA gene is a good option for a convenient preliminary phylogeny placement, but it is not a perfect model. Due to the high level of conservation in the gene, elucidation of the tree structure out at the branches of a tree can be difficult. To permit accurate prediction of the true phylogeny, we would ideally utilize all genetic material that has been passed down directly from a common ancestor, known as orthologous genes. Inclusion of paralogs, which may have unique phylogenetic history, and any laterally transferred genes have the potential to convolute the tree. We utilized single copy orthologs conserved at the family level for this analysis. We used the online database OrthoDB to find 351 orthologs conserved across the family Veillonellaceae (R. M. Waterhouse et al., 2013). We used reciprocal BLAST to find and validate the orthologs within our genomes. We developed many in-house perl, python and bash scripts to automate the alignment, formatting and extraction of orthologs. Those orthologs that had multiple hits across the genome were excluded for clarity.

Once the orthologs were extracted, each gene was individually aligned using MUSCLE and visually examined for any large gaps or misalignments (Edgar, 2004). Genes with poor alignments were also removed yielding a total of 321 orthologs to be used in the analysis. For each genome, all orthologous genes were concatenated together to form one large sequence. A 100-bootsrap maximum-likelihood tree was created using these concatenated sequences (Stamatakis, 2014). The phylogenetic analysis grouped the oral *Megasphaera* together with a vaginal isolate and the lone *Anaeroglobus* species; the gut *Megasphaera* were grouped together as would be expected. Surprisingly, the *Dialister* species that was assumed to be an outgroup was placed inside of the two vaginal *Megasphaera* phylotypes, which grouped closely to each other and grouped cleanly into the two phylotypes (Fig.18). The bootstrap score was 100 on the branch placing *Dialister* inside the two vaginal *Megasphaera* clades. In order to determine if this was an error caused by not declaring the outgroup, we ran the analysis again, this time deliberately naming the *Dialister* as the outgroup in the RAxML-HPC arguments. This analysis returned a tree in which the groupings remained largely the same, with the oral and gut isolates internal. However it placed the two vaginal *Megasphaera* clades together and just inside the *Dialister*. The bootstrap score for the branch placing the vaginal clades inside of our forced outgroup was zero (Fig.19)

To rule out the possibility of artifacts that would cause such an error, we manually examined the alignments for each ortholog and created phylogenetic trees for each individual ortholog. We observed that the trees varied widely for each gene and no taxon was placed as an outgroup more often than the others. While variability is expected when examining gene trees even for organisms with a clearly defined phylogeny, we observed especially high variability. This result may be expected given the ambiguity of the phylogeny of these organisms. We added another genome to the full 321 ortholog analysis, by selecting *Veillonella parvula,* from the closely related genus *Veillonella*, an organism that is farther out in the Veillonellaceae family tree. We repeated the analysis and again it placed the two vaginal *Megasphaera* phylotype clades outside of all of the other taxa, grouping the *Veillonella parvula* and *Dialister micraerophilus* together, with the remainder of the species internal. Again, there was strong bootstrap support for these branch placements (Fig.20). We forced *Veillonella* to be the outgroup and the maximum likelihood tree placed *Dialister* just inside the *Veillonella* supporting previous research assigning these as closely related genera. It placed the two vaginal phylotypes inside of the *Dialister* and the other species internal. Support for the outgroup branching was zero (Fig. 21).

Figure 18. Phylogenetic analysis of 321 orthologs with *Dialister* as an outgroup

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. *Dialister micraerophilus* was selected as an outgroup. Here we see both *Megasphaera* phylotypes grouping outside of *Dialister* and the other *Megasphaera* genomes. Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, and GCA_000183445.2.

Figure 18.

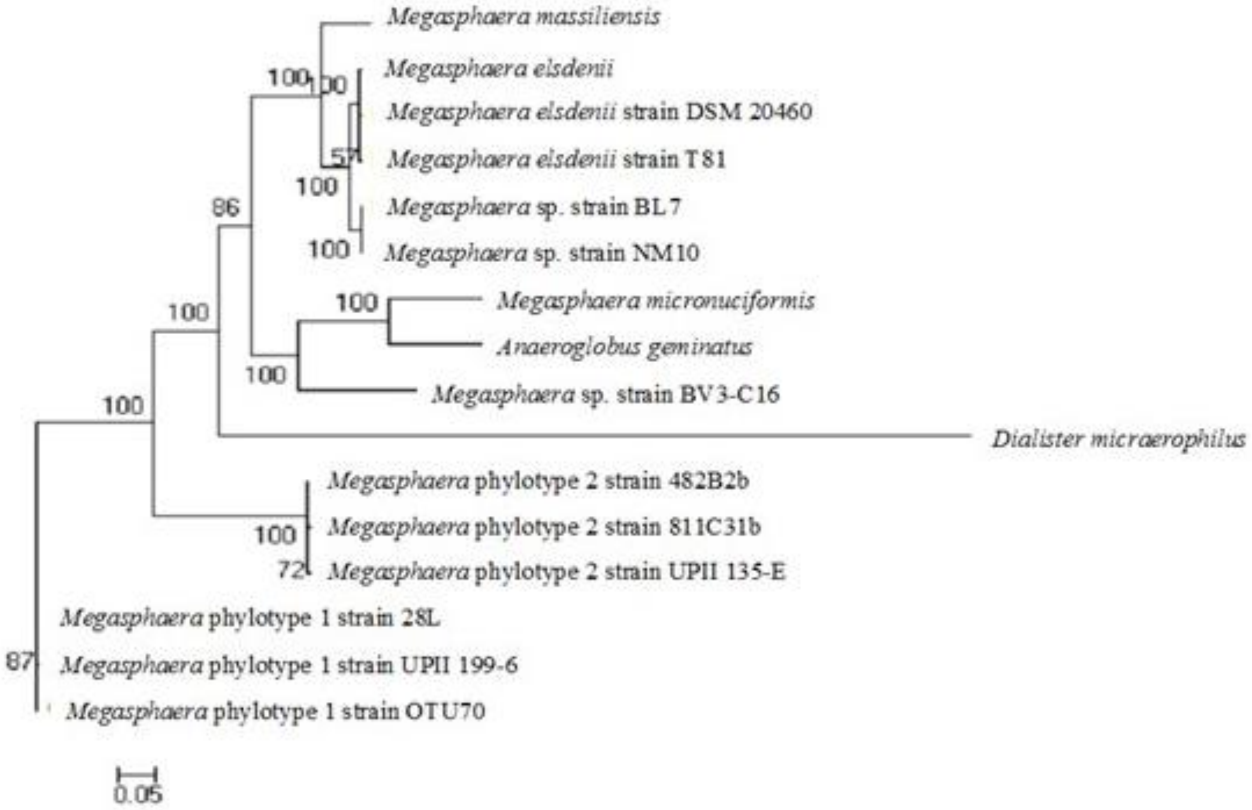

Megasphaera massiliensis
100
100
Megasphaera elsdenii
Megasphaera elsdenii strain DSM 20460
57
Megasphaera elsdenii strain T81
100
86
Megasphaera sp. strain BL7
100
Megasphaera sp. strain NM10
100
100
Megasphaera micronuciformis
Anaeroglobus geminatus
100
Megasphaera sp. strain BV3-C16
100
Dialister micraerophilus
Megasphaera phylotype 2 strain 482B2b
100
Megasphaera phylotype 2 strain 811C31b
72
Megasphaera phylotype 2 strain UPII 135-E
Megasphaera phylotype 1 strain 28L
87
Megasphaera phylotype 1 strain UPII 199-6
Megasphaera phylotype 1 strain OTU70
0.05

Figure 19. Phylogenetic analysis of 321 orthologs with *Dialister* forced to be an outgroup

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. *Dialister micraerophilus* was selected as an outgroup and explicitly stated as an outgroup when running the analysis. Note the bootstrap score of zero placing the *Dialister* at the root. Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, and GCA_000183445.2.

Figure 19.

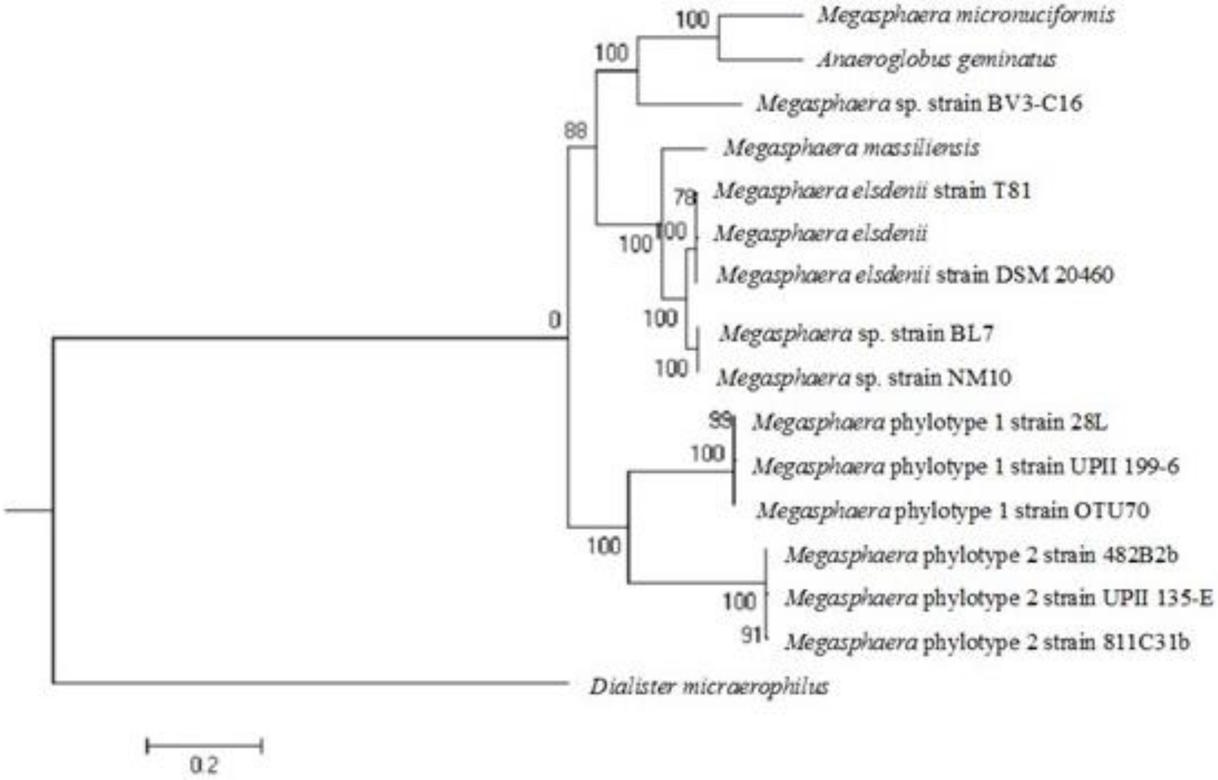

100
100
88
0
78
100
100
100
100
99
100
100
100
91
Megasphaera micronuciformis
Anaeroglobus geminatus
Megasphaera sp. strain BV3-C16
Megasphaera massiliensis
Megasphaera elsdenii strain T81
Megasphaera elsdenii
Megasphaera elsdenii strain DSM 20460
Megasphaera sp. strain BL7
Megasphaera sp. strain NM10
Megasphaera phylotype 1 strain 28L
Megasphaera phylotype 1 strain UPII 199-6
Megasphaera phylotype 1 strain OTU70
Megasphaera phylotype 2 strain 482B2b
Megasphaera phylotype 2 strain UPII 135-E
Megasphaera phylotype 2 strain 811C31b
Dialister micraerophilus
0.2

Figure 20. Phylogenetic analysis of 321 orthologs with *Veillonella* as an outgroup

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. *Veillonella parvula* was selected as an outgroup. Note the consistent outgrouping of the two vaginal *Megasphaera* species and the close relationship between *Veillonella* and *Dialister.* Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, GCA_000183445.2 and GCA_000024945.1.

Figure 20.

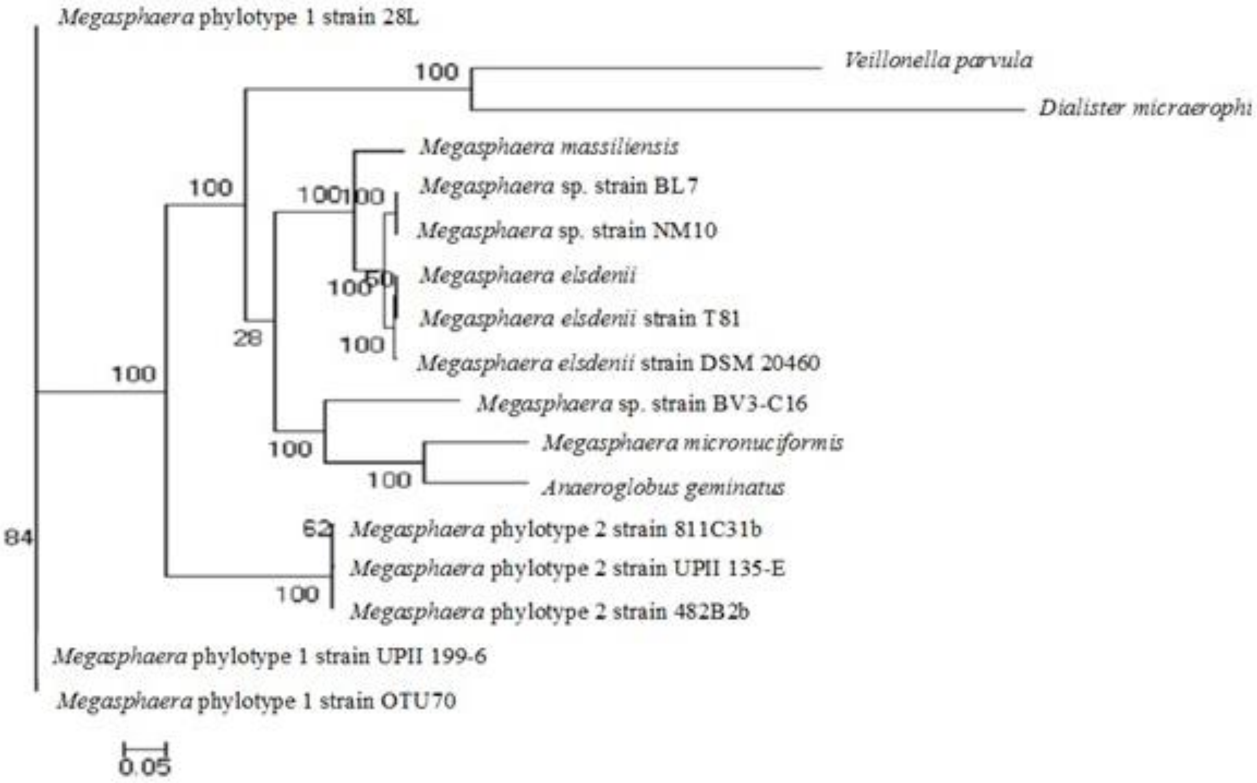
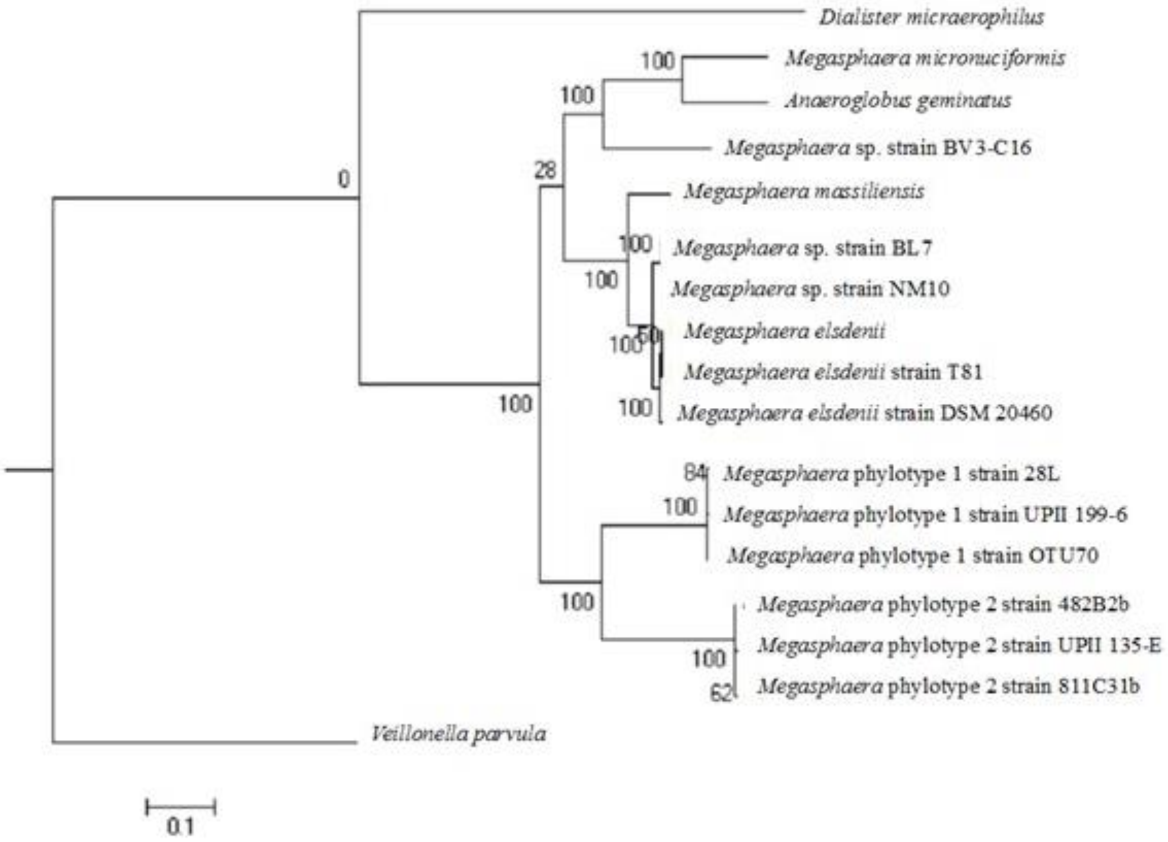

Megasphaera phylotype 1 strain 28L
100
Veillonella parvula
Dialister micraerophi
Megasphaera massiliensis
100
100100
Megasphaera sp. strain BL7
Megasphaera sp. strain NM10
10050
Megasphaera elsdenii
Megasphaera elsdenii strain T81
28
100
Megasphaera elsdenii strain DSM 20460
100
Megasphaera sp. strain BV3-C16
100
Megasphaera micronuciformis
100
Anaeroglobus geminatus
84
62
Megasphaera phylotype 2 strain 811C31b
Megasphaera phylotype 2 strain UPII 135-E
100
Megasphaera phylotype 2 strain 482B2b
Megasphaera phylotype 1 strain UPII 199-6
Megasphaera phylotype 1 strain OTU70
0.05

Figure 21. Phylogenetic analysis of 321 orthologs with *Veillonella* forced to be an outgroup

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. *Veillonella parvula* was selected as an outgroup and explicitly stated as an outgroup when running the analysis. Note the bootstrap score of zero placing the *Veillonella* at the root. Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, GCA_000183445.2 and GCA_000024945.1.

Figure 21.



Dialister micraerophilus
100
Megasphaera micronuciformis
100
Anaeroglobus geminatus
Megasphaera sp. strain BV3-C16
0
28
Megasphaera massiliensis
100
Megasphaera sp. strain BL7
100
Megasphaera sp. strain NM10
50
Megasphaera elsdenii
100
Megasphaera elsdenii strain T81
100
100
Megasphaera elsdenii strain DSM 20460
84
Megasphaera phylotype 1 strain 28L
100
Megasphaera phylotype 1 strain UPII 199-6
Megasphaera phylotype 1 strain OTU70
100
Megasphaera phylotype 2 strain 482B2b
100
Megasphaera phylotype 2 strain UPII 135-E
62
Megasphaera phylotype 2 strain 811C31b
Veillonella parvula
0.1

We hypothesized that our vaginal *Megasphaera* may possibly represent a new clade entirely outside of the *Megasphaeara, Veillonella* and the *Dialister.* We looked to the literature for a more comprehensive representation of the phylogeny of the family Veillonellaceae. We found that based on a 16S rRNA gene analysis the family groups into two clades. The first clade contains the genera *Megasphaera, Dialister, Veillonella,* and *Anaeroglobus* while the second clade was represented by a three way polytomy (Shetty, Marathe, Lanjekar, Ranade, & Shouche, 2013). We selected an organism from each branch of the polytomy, hypothesizing that the vaginal phylotypes would sit outside of the first clade but still inside of the second clade. We selected *Sporomusa ovata* (NZ_AUIL00000000.1), *Anaeromusa acidaminophila* (NZ_ARGA00000000.1) and *Selenomonas ruminantium* (GCA_000284095.1) and regenerated the orthologs. Due to poor assemblies of the genomes available at NCBI, many of the orthologs were not identified in the assemblies, leaving us with 225 conserved single copy orthologs. After running the analysis with no outgroup specification, the vaginal phylotypes were again placed outside of all of the other taxa (Fig.22).

We took another approach looking at conserved genes that other groups have used for phylogeny with the *Veillonella* outgrouped dataset (Arif et al., 2008; Aujoulat, Bouvet, Jumas-Bilak, Jean-Pierre, & Marchandin, 2014; Mashima, Kamaguchi, Miyakawa, & Nakazawa, 2013). When analyzing the DnaJ tree, *Veillonella* did outgroup on its own with a branch support of 81. When analyzing the RpoB tree, the two vaginal phylotypes outgrouped again with branch support of 100 (Fig.23, Fig. 24). Several groups have reported a great deal of heterogeneity within the Veillonellaceae family at the level of the 16S rRNA gene, particularly within the *Veillonella* (Marchandin et al., 2003; Michon et al., 2010). This may explain the difference between the 16S rRNA gene tree and the ortholog tree given that most of our genomes are not

Figure 22. Phylogenetic analysis of 225 orthologs with *Selenomonas, Sporomusa* & *Anaeromusa*

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. The three taxa combination of *Aneromusa, Sporomusa* and *Selenomonas* were selected to form an outgroup since they fell outside of the *Veillonella, Dialister, Megasphaera* clade but remained in the Veillonellaceae family. Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, GCA_000183445.2, GCA_000024945.1, GCA_000284095.1, GCA_000423685.1, and GCA_000374545.1.

Figure 22.

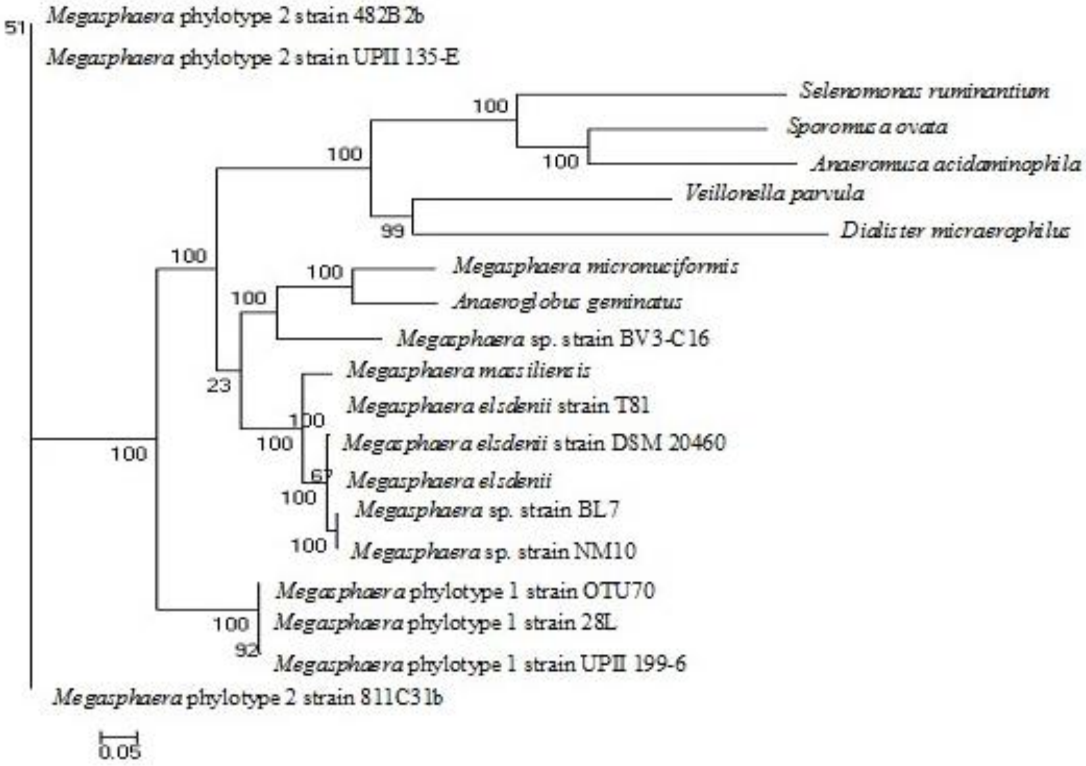
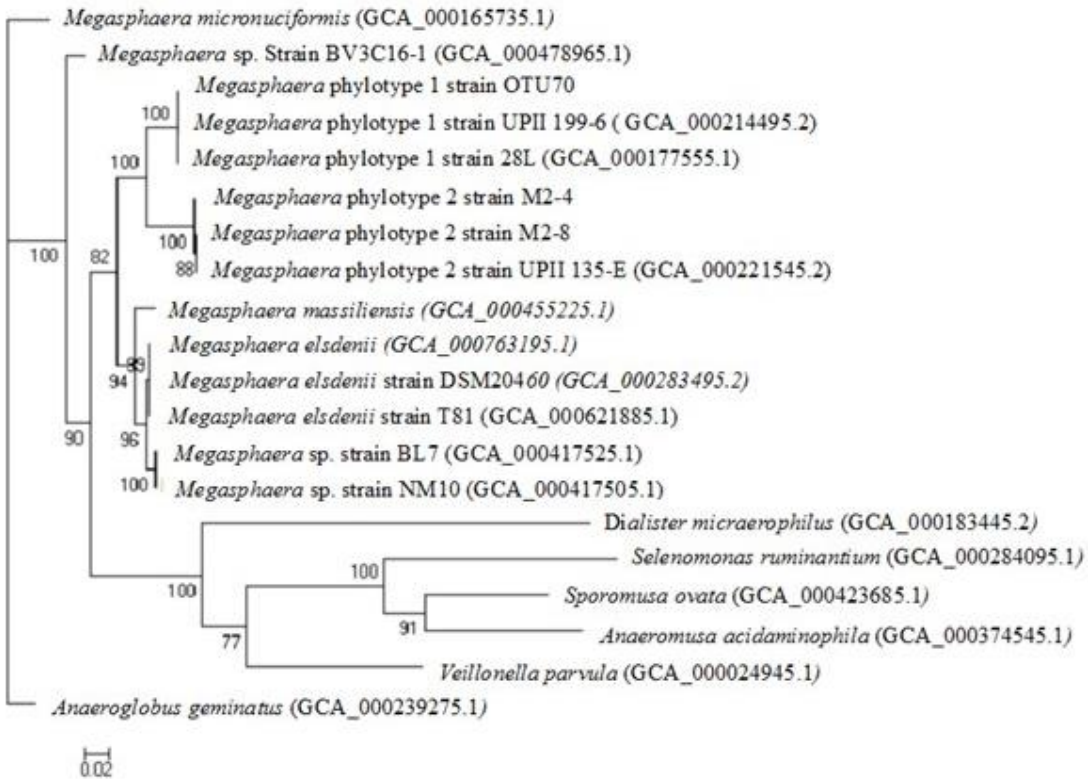

51
Megasphaera phylotype 2 strain 482B2b
Megasphaera phylotype 2 strain UPII 135-E
100
Selenomonas ruminantium
Sporomusa ovata
100
100
Anaeromusa acidaminophila
Veillonella parvula
99
Dialister micraerophilus
100
100
Megasphaera micronuciformis
100
Anaeroglobus geminatus
Megasphaera sp. strain BV3-C16
Megasphaera massiliensis
23
Megasphaera elsdenii strain T81
100
100
Megasphaera elsdenii strain DSM 20460
100
67
Megasphaera elsdenii
100
Megasphaera sp. strain BL7
100
Megasphaera sp. strain NM10
Megasphaera phylotype 1 strain OTU70
100
Megasphaera phylotype 1 strain 28L
92
Megasphaera phylotype 1 strain UPII 199-6
Megasphaera phylotype 2 strain 811C31b
0.05

Figure 23. Single copy orthologous gene DnaJ phylogenetic analysis

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. This analysis was performed using one single copy gene from each of the 20 genomes. Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, GCA_000183445.2, GCA_000024945.1, GCA_000284095.1, GCA_000423685.1, and GCA_000374545.1.

Figure 23.

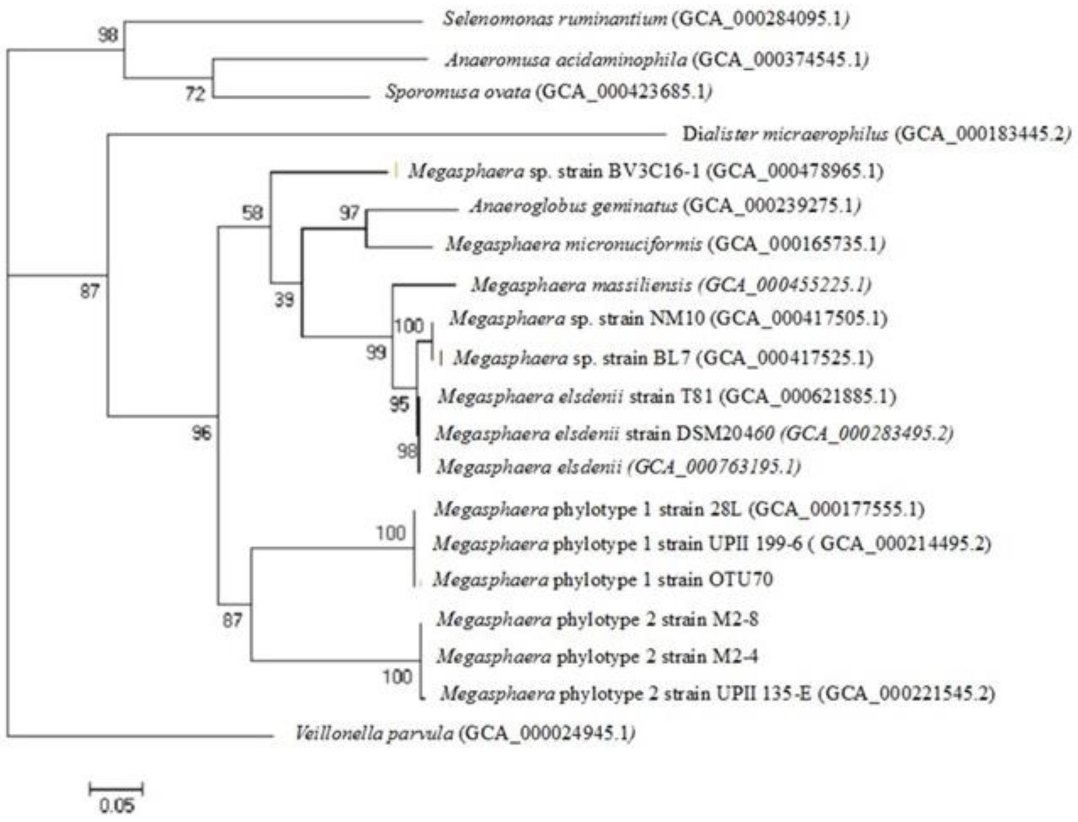


Megasphaera micronuciformis (GCA_000165735.1)
Megasphaera sp. Strain BV3C16-1 (GCA_000478965.1)
Megasphaera phylotype 1 strain OTU70
Megasphaera phylotype 1 strain UPII 199-6 (GCA_000214495.2)
Megasphaera phylotype 1 strain 28L (GCA_000177555.1)
Megasphaera phylotype 2 strain M2-4
Megasphaera phylotype 2 strain M2-8
Megasphaera phylotype 2 strain UPII 135-E (GCA_000221545.2)
Megasphaera massiliensis (GCA_000455225.1)
Megasphaera elsdenii (GCA_000763195.1)
Megasphaera elsdenii strain DSM20460 *(GCA_000283495.2)*
Megasphaera elsdenii strain T81 (GCA_000621885.1)
Megasphaera sp. strain BL7 (GCA_000417525.1)
Megasphaera sp. strain NM10 (GCA_000417505.1)
Di*alister micraerophilus* (GCA_000183445.2)
Selenomonas ruminantium (GCA_000284095.1)
Sporomusa ovata (GCA_000423685.1)
Anaeromusa acidaminophila (GCA_000374545.1)
Veillonella parvula (GCA_000024945.1)
Anaeroglobus geminatus (GCA_000239275.1)
100
100
100
100
88
82
83
94
96
90
100
100
100
91
77
0.02

Figure 24. Single copy orthologous gene RpoB phylogenetic analysis

This maximum likelihood tree was created using RAxML-HPC with 100 bootstrap replications. We employed the gamma model of heterogeneity, optimization of substitution rates and the WAG model of amino acid substitution. This analysis was performed using one single copy gene from each of the 20 genomes. Accession numbers of genomes used for this analysis are: GCA_000177555.1, GCA_000214495.2, GCA_000221545.2, GCA_000763195.1, GCA_000621885.1, GCA_000283495.1, GCA_000165735.1, GCA_000455225.1, GCA_000417505.1, GCA_000417525.1, GCA_000478965.1, GC_000239275.1, GCA_000183445.2, GCA_000024945.1, GCA_000284095.1, GCA_000423685.1, and GCA_000374545.1.

Figure 24.


Selenomonas ruminantium (GCA_000284095.1)
Anaeromusa acidaminophila (GCA_000374545.1)
Sporomusa ovata (GCA_000423685.1)
Dialister micraerophilus (GCA_000183445.2)
Megasphaera sp. strain BV3C16-1 (GCA_000478965.1)
Anaeroglobus geminatus (GCA_000239275.1)
Megasphaera micronuciformis (GCA_000165735.1)
Megasphaera massiliensis (GCA_000455225.1)
Megasphaera sp. strain NM10 (GCA_000417505.1)
Megasphaera sp. strain BL7 (GCA_000417525.1)
Megasphaera elsdenii strain T81 (GCA_000621885.1)
Megasphaera elsdenii strain DSM20460 (GCA_000283495.2)
Megasphaera elsdenii (GCA_000763195.1)
Megasphaera phylotype 1 strain 28L (GCA_000177555.1)
Megasphaera phylotype 1 strain UPII 199-6 (GCA_000214495.2)
Megasphaera phylotype 1 strain OTU70
Megasphaera phylotype 2 strain M2-8
Megasphaera phylotype 2 strain M2-4
Megasphaera phylotype 2 strain UPII 135-E (GCA_000221545.2)
Veillonella parvula (GCA_000024945.1)
98
72
87
58
97
39
100
99
95
98
96
100
87
100
0.05

circular and thus the multiple copies of the gene are not accurately represented. When a genome is assembled from contigs, if a genome is not circularized, 16S rRNA gene sequences often collapse together due to their similarity, yielding a consensus sequence representative of all of the 16S rRNA genes. Laterally transferred genes may also be contributing to the difference observed between the 16S rRNA and ortholog trees. Some publications have discussed the need for validation of the phylogeny of the family *Megasphaera* (Yutin & Galperin, 2013). The genus was once a member of the class Clostridia but was moved to the class Negativicutes. There is evidence that it should be placed back into Clostridia (Yutin & Galperin, 2013).

This is an intriguing result but this level of phylogenetic analysis is unfortunately beyond the scope of this current project. There may be a confounding evolutionary factor such as horizontal transfer or transduction at play. Future work will include analysis at the nucleotide level instead of at the protein level, recreating phylogeny of the entire Veillonellaceae family using single copy orthologs and/or conserved ribosomal proteins, an approach which was been utilized by other groups analyzing this family of bacteria (Yutin & Galperin, 2013).

DISCUSSION

Since they were first identified by 16S rRNA sequencing in 2008, the two known vaginal phylotypes of the genus *Megasphaera,* known in the vaginal microbiome research community as *Megasphaera* phylotype 1 and *Megasphaera* phylotype 2, have been repeatedly associated with negative reproductive health outcomes (Datcu, 2014; Datcu et al., 2014; Fethers et al., 2012; Marconi, Donders, Parada, Giraldo, & da Silva, 2013; D. B. Nelson et al., 2014; D. E. Nelson et al., 2012; Srinivasan et al., 2015; Zozaya-Hinchliffe et al., 2008). *Megasphaera* phylotype 1 has been observed to be associated with bacterial vaginosis in several publications. This more prevalent phylotype's association with the common vaginal infection is so consistent that it has been suggested as a biomarker for the condition using first-void urine samples (Datcu et al., 2014). *Megasphaera* phylotype 1 has been associated with increased HIV viremia in HIV-positive women in a single study (Dang et al., 2012). It was also associated with an increased risk of spontaneous preterm delivery in women who had previously given birth prematurely, demonstrated to be able to invade the upper genital tract in a single study of women having hysterectomies and found to be correlated with the inflammatory lipid marker 12-HETE, which is present in higher concentrations during active labor (Mitchell et al., 2015; D. B. Nelson et al., 2014; Srinivasan et al., 2015).

Megasphaera phylotype 2 has also been associated with bacterial vaginosis in a smaller number of studies and exhibits a weaker association with the condition compared to

Megasphaera phylotype 1. *Megasphaera* phylotype 2 has also been associated with trichomoniasis, the most common non-viral sexually transmitted disease worldwide (Zozaya-Hinchliffe et al., 2008). Given the strong associations found to exist between these two closely related organisms and highly prevalent vaginal infections as well as *Megasphaera* phylotype 1's association with preterm labor and delivery, these organisms were of interest to our group. Most of the information currently available about these organisms is limited to 16S rRNA association studies and at the time that this study began, there were no genomes available in any public databases. Our aims for this study were threefold. First, we set out to develop a method to sub-classify these organisms in our dataset of 16S rDNA V1-V3 reads. After achieving sub-classification, we sought to associate these microbial community data with clinical and demographic data. The third aim was to cultivate, isolate and sequence genomes representative of each of these phylotypes, characterize the genomes and examine the differences existing between the two phylotypes with the goal of informing our clinical findings using genomic insights.

We were able to develop a method for sub-classifcation using a USEARCH based approach in combination with an updated and comprehensive *Megasphaera* reference database (Edgar, 2010). This method was tested for precision and was found to concordantly assign reads to the phylotype level at the minimum read length of 200bp and at the full length greater than ninety-nine percent of the time. Reads assigned to the genus *Megasphaera* by the RDP classifier were used for a cluster analysis with AbundantOTU to search for new *Megasphaera* organisms detected in our samples that were not contained in the reference database (Wang et al., 2007; Ye, 2011). Three unique clusters were discovered including one representative of *Anaeroglobus geminatus*, an oral bacterium likely misclassified as a separate genus, one cluster ninety-six percent similar to *Megasphaera massiliensis*, which was only present in a few samples, and one cluster for which a full genome had been submitted to a public database. This genome was labeled BV3C16-1, suggesting that it may be associated with bacterial vaginosis. After further analysis, it was determined to be present only in BV-positive samples, representing a strongly associated taxon. However, it was only present in four individuals and thus was excluded from further analyses.

After sub-classifying the *Megasphaera* reads, reads terminal at higher taxonomic levels above *Megasphaera* including Veillonellaceae, Clostridiales, Clostridia, Firmicutes and Bacteria by the RDP classifier were also screened for potential *Megasphaera* hits using the new USEARCH approach. Thousands more *Megasphaera* reads were detected and added to further analyses. In-house scripts were utilized to add the updated *Megasphaera* data to an existing database of 16S rRNA gene microbiome data and clinical and demographic data. The 16S rRNA survey was used to analyze microbial co-occurrence. We sought to determine which organisms often co-occurred with each of the two phylotypes. As expected, the two phylotypes often co-occurred with other bacterial vaginosis associated bacteria. Interestingly, *Megasphaera* phylotype 2 was also associated with organisms known to be associated with trichomniasis including *Mycoplasma hominis* and "*Ca.* Mycoplasma girerdii". This finding suggested that this organism could be correlated with trichomoniasis infection and the associated microbial community. Both phylotypes were also strongly associated with *Prevotella* species. More than ninety-five percent of *Megasphaera* phylotype 1 and ninety-nine percent of *Megasphaera* phylotype 2 organisms were identified in samples that also contained *Prevotella* species, suggesting a potentially dependent relationship between the taxa.

Both phylotypes were positively associated with African descent and negatively associated with European descent. *Megasphaera* phylotype 1 was also negatively associated with Asian descent. The strength of these associations suggests that there may be a genetic factor predisposing certain ethnic groups to the acquisition of these phylotypes. Previous research has shown that African descent is positively associated with other BV-associated bacteria and the less protective *Lactobacillus iners* species (Fettweis, Brooks, et al., 2014a). Protective *Lactobacillus* species including *Lactobacillus crispatus* have been associated with European ancestry, potentially explaining the negative association with these BV-associated phylotypes (Ravel et al., 2011). Both phylotypes were also strongly associated with markers of lower socioeconomic status including yearly income and level of education and negatively associated with markers of higher socioeconomic status including yearly income, level of education, use of probiotics and vitamins. This suggests that there may be a socioeconomic associated risk to carriage of BV-associated organisms like these two phylotypes.

Both phylotypes were also associated with a number of common sexually transmitted infections including chlamydia, syphilis and gonorrhea. We hypothesize that these associations may be partly explained by the association of both phylotypes with an increased number of sexual partners. Interestingly, *Megasphaera* phylotype 1 was more strongly associated with bacterial vaginosis than *Gardnerella vaginalis* in a relative risk analysis. We hypothesize that this effect may be explained by the smaller prevalence of *Megasphaera* phylotype 1 in the total cohort in comparison to *Gardnerella. Gardnerella* has been prototypically described as a marker of bacterial vaginosis but can be found in microbial communities of women without any diagnosis (Fredricks, Fiedler, Thomas, Oakley, & Marrazzo, 2007). This subclinical presence of *Gardnerella vaginalis* is especially prevalent among women of African ancestry who have been largely underrepresented in previous vaginal microbiome work (Datcu, 2014; Ravel et al., 2011; Romero et al., 2014). The VaHMP dataset addresses this issue: more than two-thirds of the participants were women of African ancestry (Fettweis, Serrano, Girerd, Jefferson, & Buck, 2012a).

Megasphaera phylotype 2 was strongly associated with trichomoniasis infection by a relative risk analysis. This finding supported one previously published association of this phylotype with trichomoniasis (Zozaya-Hinchliffe et al., 2008). This phylotype was also associated with other trichomoniasis-associated organisms in the vaginal microbiome including *Mycoplasma hominis* and "*Ca*. Mycoplasma girerdii". *Megasphaera* phylotype 2 was also associated with self-reported vaginal itching, a common symptom of trichomoniasis, while *Megasphaera* phylotype 1 was not. These findings together suggest that *Megasphaera* phylotype 2 may be most prevalent among women with a current trichomoniasis infection.

After analyzing the prevalence of both organisms in a case-matched cohort of 421 pregnant and 421 non-pregnant women, *Megasphaera* phylotype 1 was found not to be excluded in pregnancy. Previous research suggests that pregnancy is characterized by decreasing prevalence of BV-associated organisms coupled with increasing prevalence of *Lactobacillus* species (Kiss et al., 2007; MacIntyre et al., 2015; Romero et al., 2014; Verstraelen et al., 2009). This organism defies that trend, although the finding was not statistically significant (p=0.12). Given this specific organism's lack of exclusion with pregnancy, association with spontaneous preterm birth, suggested ability to invade the upper genital tract and association with the inflammatory labor-associated lipid marker 12-HETE, it should be a target of future research examining the microbial contribution to spontaneous preterm labor and preterm birth.

Three vaginal *Megasphaera* strains, including one *Megasphaera* phylotype 1 organism and two *Megasphaera* phylotype 2 organisms, were cultivated from mixed vaginal swab samples, isolated and sequenced using next-generation sequencing technologies. Three publicly available genomes including two *Megasphaera* phylotype 1 genomes and one *Megasphaera* phylotype 2 genome selected to be used in a comparative genomic analysis. These six genomes, three of each phylotype, were annotated and analyzed for gene content, metabolic potential and similarity. Overall these genomes were similar in genic content and size. *Megasphaera* phylotype 1 genomes were slightly larger and contained more genes on average than *Megasphaera* phylotype 2 genomes. Synteny between the two phylotypes was not conserved and massive genome rearrangement was apparent. The syntenic conservation between the two phylotypes mimicked the syntenic conservation present between two different species of *Megasphaeara*. This lack of synteny may be a characteristic of this genus. The GC composition of genes was also starkly different between the two phylotypes with the average GC composition of genes in *Megasphaera* phylotype 1 genomes being 46.25% and the average GC composition of genes in *Megasphaera* phylotype 2 genomes being 38.90%.

Although most of the metabolic pathways were conserved between these phylotypes, they differed in important ways. Most notably, they exhibited differential metabolic strategies for carbohydrate metabolism, amino acid biosynthesis and nucleotide salvage. *Megasphaera* phylotype 2 genomes had lost the enzyme hexokinase, potentially rendering them incapable of using glucose as a carbon source via the process of glycolysis. This finding suggests that these two phylotypes may use different carbon sources for energy in the vaginal microbiome. *Megasphaera* phylotype 1 genomes had lost adenosine deaminase and adenine phosphoribosyltransferase, two essential genes involved in the salvage of adenine bases.

Conversely, *Megasphaera* phylotype 2 genomes had lost cytidine deaminase, an essential gene in the salvage of cytosine bases. This was especially interesting given the drastic differences in GC composition between the two phylotypes. It is unclear whether these gene losses are the result of relaxed selection for maintaining these nucleotide salvage pathways due to the GC skew in the genomes or if these gene losses occurred first and resulted in a shift in GC composition over time.

Megasphaera phylotype 2 had lost the ability to biosynthesize the amino acid cysteine, although it maintained the ability to convert serine amino acids into cysteine. This organism had also lost the ability to biosynthesize leucine, an amino acid important for protein structure and tryptophan. Tryptophan biosynthesis has been suggested as a potential virulence factor as it allows the evasion of IFN-γ mediated killing (Bhutia et al., 2015). Collectively, these genomic findings suggest that *Megasphaera* phylotype 2 genomes have lost essential metabolic functions and reduced their genomes over time. This in combination with the AT-rich quality of the genomes suggests that these organisms may be more host-associated than *Megasphaera* phylotype 1. Intriguingly, *Megasphaera* phylotype 2 shares the AT rich quality, reduced genome and evidence of genome plasticity with vaginal *Mycoplasma*, which have the capability of living out part of their life cycle inside of the trichomonad (Rappelli et al., 2001). It is intriguing to hypothesize that *Megasphaera* phylotype 2 may also be capable of this sort of parasitic relationship, although this would require further experimentation.

We performed a phylogenetic analysis using 321 single-copy orthologous genes from 16 genomes including publicly available *Megasphaera, Anaeroglobus* and *Dialister* genomes. *Dialister micraerophilus* was selected as an outgroup for this analysis. Unexpectedly, this analysis placed the six vaginal *Megasphaera* pyhlotypes outside of the single *Dialister* species.

This branching had 100 percent bootstrap support. After forcing the *Dialister* species to be an outgroup in the analysis, its placement at the root had zero bootstrap support. We attempted the same analyses with further related organsism including *Veillonella parvula, Selenomonas ruminantium, Sporomusa ovata* and *Anaeromusa acidaminophila.* These analyses also placed the two phylotypes outside of the other organisms. There are two explanations for this finding. The first is that our selection of genes is somehow biased to select for genes placing the *Megasphaera* phylotypes in the incorrect place in the tree. The second is that these organisms may not in fact be *Megasphaera* and may represent two Veillonellaceae family species more dissimilar to other *Megasphaera* than previously described using 16S rRNA phylogenetic analyses.

In conclusion, these two phylotypes, although closely related at the 16S rRNA level, are distinct at the genome level both structurally and functionally. They exhibit unique metabolic strategies, genome structure and GC composition. The reduced genome and metabolic potential of *Megasphaera* phylotype 2 suggests that this organism may be more host-associated than *Megasphaera* phylotype 1. These organisms also exhibit distinct clinical features. While both phylotypes are associated with negative reproductive health outcomes and a number of risk factors for bacterial vaginosis, *Megasphaera* phylotype 1 exhibits a stronger association with BV. *Megasphaera* phylotype 2 is strongly associated with trichomoniasis, an infection often co-diagnosed with BV. In combination, these distinct clinical and genomic characteristics suggest that these organisms are less similar than anticipated. Although they have been grouped together at the genus level for some vaginal microbiome association analyses, our findings suggest that this is inadvisable. Based on the syntenic and phylogenetic analyses, these organisms likely represent separate species and are potentially incorrectly classified as *Megasphaera.* These two distinct vaginal *Megasphaera* phylotypes likely play unique roles in the vaginal microbial community and their genomic composition in combination with microbial and clinical associations suggest phylotype-specific niche specialization.

Literature Cited

Africa, C. W. J., Nel, J., & Stemmet, M. (2014). Anaerobes and bacterial vaginosis in pregnancy: virulence factors contributing to vaginal colonisation. *International Journal of Environmental Research and Public Health*, *11*(7), 6979–7000. http://doi.org/10.3390/ijerph110706979

Akers, J. C., & Tan, M. (2006). Molecular mechanism of tryptophan-dependent transcriptional regulation in Chlamydia trachomatis. *Journal of Bacteriology*, *188*(12), 4236–4243. http://doi.org/10.1128/JB.01660-05

Allsworth, J. E., & Peipert, J. F. (2007). Prevalence of bacterial vaginosis: 2001-2004 National Health and Nutrition Examination Survey data. *Obstetrics and Gynecology*, *109*(1), 114–120. http://doi.org/10.1097/01.AOG.0000247627.84791.91

Alves, J. M. P., & Buck, G. A. (2007). Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. *Chemistry & Biodiversity*, *4*(11), 2593–2602. http://doi.org/10.1002/cbdv.200790212

Arif, N., Do, T., Byun, R., Sheehy, E., Clark, D., Gilbert, S. C., & Beighton, D. (2008). Veillonella rogosae sp. nov., an anaerobic, Gram-negative coccus isolated from dental plaque. *International Journal of Systematic and Evolutionary Microbiology*, *58*(Pt 3), 581–584. http://doi.org/10.1099/ijs.0.65093-0

Aujoulat, F., Bouvet, P., Jumas-Bilak, E., Jean-Pierre, H., & Marchandin, H. (2014). Veillonella seminalis sp. nov., a novel anaerobic Gram-stain-negative coccus from human clinical samples, and emended description of the genus Veillonella. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt 10), 3526–3531. http://doi.org/10.1099/ijs.0.064451-0

Balkus, J. E., Mitchell, C., Agnew, K., Liu, C., Fiedler, T., Cohn, S. E., et al. (2012). Detection of hydrogen peroxide-producing Lactobacillus species in the vagina: a comparison of culture and quantitative PCR among HIV-1 seropositive women. *BMC Infectious Diseases*, *12*(1), 188. http://doi.org/10.1186/1471-2334-12-188

Barbut, F., Collignon, A., Butel, M.-J., & Bourlioux, P. (2015). [Fecal microbiota transplantation: review]. *Annales Pharmaceutiques Françaises*, *73*(1), 13–21. http://doi.org/10.1016/j.pharma.2014.05.004

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., et al. (2002). The Pfam protein families database. *Nucleic Acids Research*, *30*(1), 276–280.

Bhutia, Y. D., Babu, E., & Ganapathy, V. (2015). Interferon-γ induces a tryptophan-selective amino acid transporter in human colonic epithelial cells and mouse dendritic cells.

Biochimica Et Biophysica Acta, *1848*(2), 453–462. http://doi.org/10.1016/j.bbamem.2014.10.021

Bohlin, J., Skjerve, E., & Ussery, D. W. (2008). Investigations of oligonucleotide usage variance within and between prokaryotes. *PLoS Computational Biology*, *4*(4), e1000057. http://doi.org/10.1371/journal.pcbi.1000057

Boleij, A., & Tjalsma, H. (2012). Gut bacteria in health and disease: a survey on the interface between intestinal microbiology and colorectal cancer. *Biological Reviews of the Cambridge Philosophical Society*, *87*(3), 701–730. http://doi.org/10.1111/j.1469-185X.2012.00218.x

Bonner, C. A., Byrne, G. I., & Jensen, R. A. (2014). Chlamydia exploit the mammalian tryptophan-depletion defense strategy as a counter-defensive cue to trigger a survival state of persistence. *Frontiers in Cellular and Infection Microbiology*, *4*, 17. http://doi.org/10.3389/fcimb.2014.00017

Borodovsky, M., & Lomsadze, A. (2011). Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [Et Al.]*, *Chapter 4*, Unit 4.5.1–17. http://doi.org/10.1002/0471250953.bi0405s35

Bradshaw, C. S., Morton, A. N., Hocking, J., Garland, S. M., Morris, M. B., Moss, L. M., et al. (2006). High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *The Journal of Infectious Diseases*, *193*(11), 1478–1486. http://doi.org/10.1086/503780

Brotman, R. M. (2011). Vaginal microbiome and sexually transmitted infections: an epidemiologic perspective. *The Journal of Clinical Investigation*, *121*(12), 4610–4617. http://doi.org/10.1172/JCI57172

Brown, C. J., Wong, M., Davis, C. C., Kanti, A., Zhou, X., & Forney, L. J. (2007). Preliminary characterization of the normal microbiota of the human vulva using cultivation-independent methods. *Journal of Medical Microbiology*, *56*(Pt 2), 271–276. http://doi.org/10.1099/jmm.0.46607-0

Campbell, C., Adeolu, M., & Gupta, R. S. (2015). Genome-based taxonomic framework for the class Negativicutes: division of the class Negativicutes into the orders Selenomonadales emend., Acidaminococcales ord. nov. and Veillonellales ord. nov. *International Journal of Systematic and Evolutionary Microbiology*, *65*(9), 3203–3215. http://doi.org/10.1099/ijs.0.000347

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552.

Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., & Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, *7*(1), 439. http://doi.org/10.1186/1471-2105-7-439

Cohen, C. R., Lingappa, J. R., Baeten, J. M., Ngayo, M. O., Spiegel, C. A., Hong, T., et al. (2012). Bacterial vaginosis associated with increased risk of female-to-male HIV-1 transmission: a prospective cohort analysis among African couples. *PLoS Medicine*, *9*(6), e1001251. http://doi.org/10.1371/journal.pmed.1001251

Cotch, M. F., Pastorek, J. G., Nugent, R. P., Hillier, S. L., Gibbs, R. S., Martin, D. H., et al. (1997). Trichomonas vaginalis associated with low birth weight and preterm delivery. The Vaginal Infections and Prematurity Study Group. *Sexually Transmitted Diseases*, *24*(6), 353–360.

Cottrell, B. H. (2010). An updated review of of evidence to discourage douching. *MCN. the*

American Journal of Maternal Child Nursing, *35*(2), 102–7– quiz 108–9. http://doi.org/10.1097/NMC.0b013e3181cae9da

Cruickshank, R. (1931). Döderlein's Vaginal Bacillus: A Contribution to the Study of the <i>Lacto-Bacilli</i>. *Journal of Hygiene*, *31*(03), 375–381. http://doi.org/10.1017/S0022172400010901

Dang, A. T., Cotton, S., Sankaran-Walters, S., Li, C.-S., Lee, C.-Y. M., Dandekar, S., et al. (2012). Evidence of an increased pathogenic footprint in the lingual microbiome of untreated HIV infected patients. *BMC Microbiology*, *12*(1), 153. http://doi.org/10.1186/1471-2180-12-153

Datcu, R. (2014). Characterization of the vaginal microflora in health and disease. *Danish Medical Journal*, *61*(4), B4830.

Datcu, R., Gesink, D., Mulvad, G., Montgomery-Andersen, R., Rink, E., Koch, A., et al. (2014). Bacterial vaginosis diagnosed by analysis of first-void-urine specimens. *Journal of Clinical Microbiology*, *52*(1), 218–225. http://doi.org/10.1128/JCM.02347-13

Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, *27*(23), 4636–4641.

Donders, G. G., Van Calsteren, K., Bellen, G., Reybrouck, R., Van den Bosch, T., Riphagen, I., & Van Lierde, S. (2009). Predictive value for preterm birth of abnormal vaginal flora, bacterial vaginosis and aerobic vaginitis during the first trimester of pregnancy. *BJOG : an International Journal of Obstetrics and Gynaecology*, *116*(10), 1315–1324. http://doi.org/10.1111/j.1471-0528.2009.02237.x

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. http://doi.org/10.1093/nar/gkh340

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, *26*(19), 2460–2461. http://doi.org/10.1093/bioinformatics/btq461

Edwards, T., Burke, P., Smalley, H., & Hobbs, G. (2014). Trichomonas vaginalis: Clinical relevance, pathogenicity and diagnosis. *Critical Reviews in Microbiology*, 1–12. http://doi.org/10.3109/1040841X.2014.958050

Fasullo, M., & Endres, L. (2015). Nucleotide salvage deficiencies, DNA damage and neurodegeneration. *International Journal of Molecular Sciences*, *16*(5), 9431–9449. http://doi.org/10.3390/ijms16059431

Fethers, K. (2001). Is bacterial vaginosis a sexually transmitted infection. *Sexually Transmitted Infections*, *77*(5), 390–390. http://doi.org/10.1136/sti.77.5.390

Fethers, K., Marks, C., Mindel, A., & Estcourt, C. S. (2000). Sexually transmitted infections and risk behaviours in women who have sex with women. *Sexually Transmitted Infections*, *76*(5), 345–349. http://doi.org/10.1136/sti.76.5.345

Fethers, K., Twin, J., Fairley, C. K., Fowkes, F. J. I., Garland, S. M., Fehler, G., et al. (2012). Bacterial vaginosis (BV) candidate bacteria: associations with BV and behavioural practices in sexually-experienced and inexperienced women. *PloS One*, *7*(2), e30633. http://doi.org/10.1371/journal.pone.0030633

Fettweis, J. M., Brooks, J. P., Serrano, M. G., Sheth, N. U., Girerd, P. H., Edwards, D. J., et al. (2014a). Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology (Reading, England)*, *160*(Pt 10), 2272–2282. http://doi.org/10.1099/mic.0.081034-0

Fettweis, J. M., Serrano, M. G., Girerd, P. H., Jefferson, K. K., & Buck, G. A. (2012a). A new era of the vaginal microbiome: advances using next-generation sequencing. *Chemistry & Biodiversity*, *9*(5), 965–976. http://doi.org/10.1002/cbdv.201100359

Fettweis, J. M., Serrano, M. G., Huang, B., Brooks, J. P., Glascock, A. L., Sheth, N. U., et al. (2014b). An emerging mycoplasma associated with trichomoniasis, vaginal infection and disease. *PloS One*, *9*(10), e110943. http://doi.org/10.1371/journal.pone.0110943

Fettweis, J. M., Serrano, M. G., Sheth, N. U., Mayer, C. M., Glascock, A. L., Brooks, J. P., et al. (2012b). Species-level classification of the vaginal microbiome. *BMC Genomics*, *13 Suppl 8*, S17. http://doi.org/10.1186/1471-2164-13-S8-S17

Fredricks, D. N., Fiedler, T. L., Thomas, K. K., Oakley, B. B., & Marrazzo, J. M. (2007). Targeted PCR for detection of vaginal bacteria associated with bacterial vaginosis. *Journal of Clinical Microbiology*, *45*(10), 3270–3276. http://doi.org/10.1128/JCM.01272-07

Gallo, M. F., Macaluso, M., Warner, L., Fleenor, M. E., Hook, E. W., Brill, I., & Weaver, M. A. (2012). Bacterial vaginosis, gonorrhea, and chlamydial infection among women attending a sexually transmitted disease clinic: a longitudinal analysis of possible causal links. *Annals of Epidemiology*, *22*(3), 213–220. http://doi.org/10.1016/j.annepidem.2011.11.005

Gibson, G. R., McCartney, A. L., & Rastall, R. A. (2005). Prebiotics and resistance to gastrointestinal infections. *The British Journal of Nutrition*, *93 Suppl 1*, S31–4.

Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., & Posada, D. (2010). ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, *38*(Web Server issue), W14–8. http://doi.org/10.1093/nar/gkq321

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. http://doi.org/10.1093/sysbio/syq010

GUTIERREZ, J., DAVIS, R. E., LINDAHL, I. L., & WARWICK, E. J. (1959). Bacterial changes in the rumen during the onset of feed-lot bloat of cattle and characteristics of Peptostreptococcus elsdenii n. sp. *Applied Microbiology*, *7*(1), 16–22.

Harwich, M. D., Serrano, M. G., Fettweis, J. M., Alves, J. M. P., Reimers, M. A., Vaginal Microbiome Consortium (additional members), et al. (2012). Genomic sequence analysis and characterization of Sneathia amnii sp. nov. *BMC Genomics*, *13 Suppl 8*(Suppl 8), S4. http://doi.org/10.1186/1471-2164-13-S8-S4

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. http://doi.org/10.1038/nature11234

Jeffery, I. B., Lynch, D. B., & O'Toole, P. W. (2015). Composition and temporal stability of the gut microbiota in older persons. *The ISME Journal*. http://doi.org/10.1038/ismej.2015.88

Juvonen, R., & Suihko, M.-L. (2006). Megasphaera paucivorans sp. nov., Megasphaera sueciensis sp. nov. and Pectinatus haikarae sp. nov., isolated from brewery samples, and emended description of the genus Pectinatus. *International Journal of Systematic and Evolutionary Microbiology*, *56*(Pt 4), 695–702. http://doi.org/10.1099/ijs.0.63699-0

Kaewsrichan, J., Peeyananjarassri, K., & Kongprasertkit, J. (2006). Selection and identification of anaerobic lactobacilli producing inhibitory compounds against vaginal pathogens. *FEMS Immunology and Medical Microbiology*, *48*(1), 75–83. http://doi.org/10.1111/j.1574-695X.2006.00124.x

Kiss, H., Kögler, B., Petricevic, L., Sauerzapf, I., Klayraung, S., Domig, K., et al. (2007). Vaginal Lactobacillus microbiota of healthy women in the late first trimester of pregnancy. *BJOG : an International Journal of Obstetrics and Gynaecology*, *114*(11), 1402–1407.

http://doi.org/10.1111/j.1471-0528.2007.01412.x

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), R12. http://doi.org/10.1186/gb-2004-5-2-r12

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108. http://doi.org/10.1093/nar/gkm160

Lanjekar, V. B., Marathe, N. P., Ramana, V. V., Shouche, Y. S., & Ranade, D. R. (2014). Megasphaera indica sp. nov., an obligate anaerobic bacteria isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt 7), 2250–2256. http://doi.org/10.1099/ijs.0.059816-0

Leitich, H., Bodner-Adler, B., Brunbauer, M., Kaider, A., Egarter, C., & Husslein, P. (2003). Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. *American Journal of Obstetrics and Gynecology*, *189*(1), 139–147.

Leppäluoto, P. A. (2011). Bacterial vaginosis: what is physiological in vaginal bacteriology? An update and opinion. *Acta Obstetricia Et Gynecologica Scandinavica*, *90*(12), 1302–1306. http://doi.org/10.1111/j.1600-0412.2011.01279.x

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964.

MacIntyre, D. A., Chandiramani, M., Lee, Y. S., Kindinger, L., Smith, A., Angelopoulos, N., et al. (2015). The vaginal microbiome during pregnancy and the postpartum period in a European population. *Scientific Reports*, *5*, 8988. http://doi.org/10.1038/srep08988

Malaguti, N., Bahls, L. D., Uchimura, N. S., Gimenes, F., & Consolaro, M. E. L. (2015). Sensitive Detection of Thirteen Bacterial Vaginosis-Associated Agents Using Multiplex Polymerase Chain Reaction. *BioMed Research International*, *2015*(2), 645853–10. http://doi.org/10.1155/2015/645853

Manhart, L. E., Khosropour, C. M., Liu, C., Gillespie, C. W., Depner, K., Fiedler, T., et al. (2013). Bacterial vaginosis-associated bacteria in men: association of Leptotrichia/Sneathia spp. with nongonococcal urethritis. *Sexually Transmitted Diseases*, *40*(12), 944–949. http://doi.org/10.1097/OLQ.0000000000000054

Marchandin, H., Teyssier, C., Siméon De Buochberg, M., Jean-Pierre, H., Carrière, C., & Jumas-Bilak, E. (2003). Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus Veillonella: implications for phylogeny and taxonomy. *Microbiology (Reading, England)*, *149*(Pt 6), 1493–1501. http://doi.org/10.1099/mic.0.26132-0

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(Database issue), D222–6. http://doi.org/10.1093/nar/gku1221

Marconi, C., Donders, G. G., Parada, C. M. G. L., Giraldo, P. C., & da Silva, M. G. (2013). Do Atopobium vaginae, Megasphaera sp. and Leptotrichia sp. change the local innate immune response and sialidase activity in bacterial vaginosis? *Sexually Transmitted Infections*, *89*(2), 167–173. http://doi.org/10.1136/sextrans-2012-050616

Martin, D. H., Zozaya, M., Lillis, R. A., Myers, L., Nsuami, M. J., & Ferris, M. J. (2013). Unique vaginal microbiota that includes an unknown Mycoplasma-like organism is associated with Trichomonas vaginalis infection. *The Journal of Infectious Diseases*, *207*(12), 1922–1931. http://doi.org/10.1093/infdis/jit100

Mashima, I., Kamaguchi, A., Miyakawa, H., & Nakazawa, F. (2013). Veillonella tobetsuensis sp.

nov., an anaerobic, gram-negative coccus isolated from human tongue biofilms. *International Journal of Systematic and Evolutionary Microbiology*, *63*(Pt 4), 1443–1449. http://doi.org/10.1099/ijs.0.042515-0

Meites, E., Gaydos, C. A., Hobbs, M. M., Kissinger, P., Nyirjesy, P., Schwebke, J. R., et al. (2015). A Review of Evidence-Based Care of Symptomatic Trichomoniasis and Asymptomatic Trichomonas vaginalis Infections. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 8*(suppl 8), S837–48. http://doi.org/10.1093/cid/civ738

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*(1), 386. http://doi.org/10.1186/1471-2105-9-386

Michon, A.-L., Aujoulat, F., Roudière, L., Soulier, O., Zorgniotti, I., Jumas-Bilak, E., & Marchandin, H. (2010). Intragenomic and intraspecific heterogeneity in rrs may surpass interspecific variability in a natural population of Veillonella. *Microbiology (Reading, England)*, *156*(Pt 7), 2080–2091. http://doi.org/10.1099/mic.0.038224-0

Mitchell, C. M., Haick, A., Nkwopara, E., Garcia, R., Rendi, M., Agnew, K., et al. (2015). Colonization of the upper genital tract by vaginal bacterial species in nonpregnant women. *American Journal of Obstetrics and Gynecology*, *212*(5), 611.e1–9. http://doi.org/10.1016/j.ajog.2014.11.043

Nelson, D. B., Hanlon, A., Nachamkin, I., Haggerty, C., Mastrogiannis, D. S., Liu, C., & Fredricks, D. N. (2014). Early pregnancy changes in bacterial vaginosis-associated bacteria and preterm delivery. *Paediatric and Perinatal Epidemiology*, *28*(2), 88–96. http://doi.org/10.1111/ppe.12106

Nelson, D. E., Dong, Q., Van der Pol, B., Toh, E., Fan, B., Katz, B. P., et al. (2012). Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PloS One*, *7*(5), e36298. http://doi.org/10.1371/journal.pone.0036298

Nishida, H. (2012). Evolution of genome base composition and genome size in bacteria. *Frontiers in Microbiology*, *3*, 420. http://doi.org/10.3389/fmicb.2012.00420

O'Hanlon, D. E., Moench, T. R., & Cone, R. A. (2013). Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PloS One*, *8*(11), e80074. http://doi.org/10.1371/journal.pone.0080074

Ohkusa, T., & Koido, S. (2015). Intestinal microbiota and ulcerative colitis. *Journal of Infection and Chemotherapy : Official Journal of the Japan Society of Chemotherapy*, *21*(11), 761–768. http://doi.org/10.1016/j.jiac.2015.07.010

Ortiz, L., Ruiz, F., Pascual, L., & Barberis, L. (2014). Effect of two probiotic strains of Lactobacillus on in vitro adherence of Listeria monocytogenes, Streptococcus agalactiae, and Staphylococcus aureus to vaginal epithelial cells. *Current Microbiology*, *68*(6), 679–684. http://doi.org/10.1007/s00284-014-0524-9

Padmanabhan, R., Lagier, J.-C., Dangui, N. P. M., Michelle, C., Couderc, C., Raoult, D., & Fournier, P.-E. (2013). Non-contiguous finished genome sequence and description of Megasphaera massiliensis sp. nov. *Standards in Genomic Sciences*, *8*(3), 525–538. http://doi.org/10.4056/sigs.4077819

Pérez-Cobas, A. E., Artacho, A., Ott, S. J., Moya, A., Gosalbes, M. J., & Latorre, A. (2014). Structural and functional changes in the gut microbiota associated to Clostridium difficile infection. *Frontiers in Microbiology*, *5*, 335. http://doi.org/10.3389/fmicb.2014.00335

Purwar, M., Ughade, S., Bhagat, B., Agarwal, V., & Kulkarni, H. (2001). Bacterial vaginosis in early pregnancy and adverse pregnancy outcome. *The Journal of Obstetrics and Gynaecology Research*, *27*(4), 175–181.

Rao, R. K., & Samak, G. (2013). Protection and Restitution of Gut Barrier by Probiotics: Nutritional and Clinical Implications. *Current Nutrition and Food Science*, *9*(2), 99–107.

Rappelli, P., Addis, M. F., Carta, F., & Fiori, P. L. (1998). Mycoplasma hominis parasitism of Trichomonas vaginalis. *Lancet (London, England)*, *352*(9136), 1286. http://doi.org/10.1016/S0140-6736(98)00041-5

Rappelli, P., Carta, F., Delogu, G., Addis, M. F., Dessì, D., Cappuccinelli, P., & Fiori, P. L. (2001). Mycoplasma hominis and Trichomonas vaginalis symbiosis: multiplicity of infection and transmissibility of M. hominis to human cells. *Archives of Microbiology*, *175*(1), 70–74.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, S. L., et al. (2011). Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl 1*(Supplement_1), 4680–4687. http://doi.org/10.1073/pnas.1002611107

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics : TIG*, *16*(6), 276–277.

Rogosa, M. (1971). Transfer of Peptostreptococcus elsdenii Gutierrez et al. to a New Genus, Megasphaera [M. elsdenii (Gutierrez et al.) comb. nov.]. *International Journal of Systematic and Evolutionary Microbiology*, *21*(2), 187–189. http://doi.org/10.1099/00207713-21-2-187

Romero, R., Hassan, S. S., Gajer, P., Tarca, A. L., Fadrosh, D. W., Nikita, L., et al. (2014). The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome*, *2*(1), 4. http://doi.org/10.1186/2049-2618-2-4

Segata, N. (2015). Gut Microbiome: Westernization and the Disappearance of Intestinal Diversity. *Current Biology : CB*, *25*(14), R611–3. http://doi.org/10.1016/j.cub.2015.05.040

Shetty, S. A., Marathe, N. P., Lanjekar, V., Ranade, D., & Shouche, Y. S. (2013). Comparative genome analysis of Megasphaera sp. reveals niche specialization and its potential role in the human gut. *PloS One*, *8*(11), e79353. http://doi.org/10.1371/journal.pone.0079353

Silver, B. J., Guy, R. J., Kaldor, J. M., Jamil, M. S., & Rumbold, A. R. (2014). Trichomonas vaginalis as a cause of perinatal morbidity: a systematic review and meta-analysis. *Sexually Transmitted Diseases*, *41*(6), 369–376. http://doi.org/10.1097/OLQ.0000000000000134

Srinivasan, S., Morgan, M. T., Fiedler, T. L., Djukovic, D., Hoffman, N. G., Raftery, D., et al. (2015). Metabolic signatures of bacterial vaginosis. *mBio*, *6*(2), e00204–15. http://doi.org/10.1128/mBio.00204-15

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, *30*(9), 1312–1313. http://doi.org/10.1093/bioinformatics/btu033

Svobodová, L., Lysková, P., & Hamal, P. (2015). [Vulvovaginal candidiasis]. *Klinicka Mikrobiologie a Infekcni Lekarstvi*, *21*(2), 74–81.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33–36.

Verstraelen, H., Verhelst, R., Claeys, G., De Backer, E., Temmerman, M., & Vaneechoutte, M. (2009). Longitudinal analysis of the vaginal microflora in pregnancy suggests that L. crispatus promotes the stability of the normal vaginal microflora and that L. gasseri and/or L.

iners are more conducive to the occurrence of abnormal vaginal microflora. *BMC Microbiology*, *9*(1), 116. http://doi.org/10.1186/1471-2180-9-116

Vitali, B., Cruciani, F., Picone, G., Parolin, C., Donders, G., & Laghi, L. (2015). Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology*, *34*(12), 2367–2376. http://doi.org/10.1007/s10096-015-2490-y

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. http://doi.org/10.1128/AEM.00062-07

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, *25*(9), 1189–1191. http://doi.org/10.1093/bioinformatics/btp033

Waterhouse, R. M., Tegenfeldt, F., Li, J., Zdobnov, E. M., & Kriventseva, E. V. (2013). OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Research*, *41*(Database issue), D358–65. http://doi.org/10.1093/nar/gks1116

Whitney, J. (1997). Testing for differences with the nonparametric Mann-Whitney U test. *Journal of Wound, Ostomy, and Continence Nursing : Official Publication of the Wound, Ostomy and Continence Nurses Society / WOCN*, *24*(1), 12.

Wilks, M., Wiggins, R., Whiley, A., Hennessy, E., Warwick, S., Porter, H., et al. (2004). Identification and H(2)O(2) production of vaginal lactobacilli from pregnant women at high risk of preterm birth and relation with outcome. *Journal of Clinical Microbiology*, *42*(2), 713–717. http://doi.org/10.1128/JCM.42.2.713-717.2004

Williams, T., & Kelley, C. (2011). Gnuplot 4.4: an interactive plotting program. URL: http://www. gnuplot. info.

Witkin, S. S., Mendes-Soares, H., Linhares, I. M., Jayaram, A., Ledger, W. J., & Forney, L. J. (2013). Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *mBio*, *4*(4), e00460–13–e00460–13. http://doi.org/10.1128/mBio.00460-13

Ye, Y. (2011). Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. *Proceedings. IEEE International Conference on Bioinformatics and Biomedicine*, *2010*, 153–157. http://doi.org/10.1109/BIBM.2010.5706555

Yutin, N., & Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology*, *15*(10), 2631–2641. http://doi.org/10.1111/1462-2920.12173

Zozaya-Hinchliffe, M., Martin, D. H., & Ferris, M. J. (2008). Prevalence and abundance of uncultivated Megasphaera-like bacteria in the human vaginal environment. *Applied and Environmental Microbiology*, *74*(5), 1656–1659. http://doi.org/10.1128/AEM.02127-07

VITA

Abigail Leigh Glascock was born October 5, 1988 in Halifax, VA and is an American citizen. She graduated with an International Baccalaureate Diploma from Lee-Davis High School, Mechanicsville, VA in 2007. Abigail received her Bachelor of Science in Biology from James Madison University in Harrisonburg, VA in the summer of 2010. After graduating, she worked as a laboratory technician in the Department of Microbiology and Immunology at the Virginia Commonwealth University School of Medicine from the fall of 2010 through the summer of 2013. In the fall of 2013, Abigail entered both the Integrative Life Science Ph.D. and Bioinformatics MS programs as a member of the Vaginal Microbiome Consortium working with Dr. Gregory Buck and Dr. Jennifer Fettweis. She joined Dr. Fettweis' laboratory in the fall of 2014. Notable accomplishments are listed below.

AWARDS AND GRANTS:

2015…………….Golden Key International Honor Society

2015…………….Phi Kappa Phi Honor Society

SOCIETIES:

2013-Present…...American Association for the Advancement of Science

PUBLICATIONS:

Fettweis, J.M., Serrano, M.G., Huang, B., Brooks, J.P., Glascock, A.L., Sheth, N.U., Vaginal Microbiome Consortium, Strauss, J.F., Jefferson, K.K., and Buck, G.A. (2014). An emerging mycoplasma associated with trichomoniasis, vaginal infection and disease. PloS One *9*, e110943.

Fettweis, J.M., Serrano, M.G., Sheth, N.U., Mayer, C.M., Glascock, A.L., Brooks, J.P., Jefferon, K.K.,Vaginal Microbiome Consortium and Buck, G.A. (2012). Species-level classification of the vaginal microbiome. BMC Genomics 13, Suppl 8:S17.

Integrative HMP (iHMP) Research Network Consortium. (2014) The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe. 16 (3) 276-289.

Abdelmaksoud, A.A., Koparde, V.N., Sheth, N.U., Serrano, M.G., Glascock, A.L., Fettweis, J.M., Strauss, J.F. 3rd, Buck, G.A. and Jefferson, K.K. (In review) Comparison of *Lactobacillus crispatus* isolated from women with and without bacterial vaginosis.