COMPARISON OF MICROBIAL DIVERSITY OF FIFTEEN AGED CHEDDAR CHEESES FROM DIFFERENT REGIONS USING NEXT GENERATION SEQUENCING

by

Sophie L. Overbeck

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Nutrition and Food Sciences

Approved:

_____	_____
Donald McMahon, Ph.D. Major Professor	Taylor Oberg, Ph.D. Co-Major Professor
_____	_____
Mike Lefevre, Ph.D. Committee Member	Craig J. Oberg, Ph.D. Committee Member

_____
D. Richard Cutler, Ph.D.
Interim Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2021

ABSTRACT

Comparison of Microbial Diversity of Fifteen Aged Cheddar Cheeses from Different Regions Using Next Generation Sequencing

by

Sophie Overbeck, Master of Science

Utah State University, 2021

Major Professor: Dr. Donald J. McMahon
Department: Nutrition, Dietetics, and Food Science

The aim of this project was to test a newly developed method and its efficacy in identifying bacterial community diversity within aged Cheddar cheese as well as identifying the relative abundance of the different bacteria. This project will provide further understanding of how using next generation sequencing can benefit the study of bacterial communities and their influence on the aging process of Cheddar cheese along with the possibility of identifying defect causing microorganisms. The hypothesis of this study is that using next generation sequencing with different Cheddar cheese samples from various regions will result in cheeses clustering by region and the cheese's microbiota will be dominated by non-starter lactic acid bacteria.

To test this hypothesis, 15 different Cheddar cheese sample's microbiota were determined by amplifying the hypervariable V4 region of the 16S rRNA gene and sequenced using the Illumina MiSeq system from the Center for Integrate Biosystems at Utah State University for DNA analysis. Sequencing data was evaluated using two different denoising pipelines, within Qiime 2 for comparison of variant identification and as a method of comparing starter and non-starter lactic acid bacteria in cheeses based on regions and manufacturers.

The samples clustered into two groups based on the dominant starter lactic acid bacteria (SLAB). The US cheeses tended to be dominated by *Lactococcus lactis* while the cheeses manufactured outside of the United States were dominated by *Streptococcus thermophilus*. The two denoising pipelines evaluated were Deblur and DADA2. Using Deblur, 40.5% of reads were retained resulting in 74 total amplicon sequence variants (ASVs) which represent a unique sequence within the samples. In contrast, DADA2 retained 76% of reads resulting in 247 unique ASVs. Amplicon sequence variants were categorized into 4 different types *Lc. lactis*, *St. thermophilus*, lactobacilli, and other species.

In both denoising methods the dominating ASVs were SLAB. Of the dominating *Lc. lactis* ASVs when using Deblur only 5 ASVs were identified compared to 34 ASVs identified by DADA2. Similarly, only 3 St. thermophilus ASVs were identified by Deblur and 35 ASVs were identified by DADA2. Typically, during Cheddar manufacture, multiple strains of SLAB are used as a way of preventing the failure of a cheese make because of phage infection. Therefore, the ASVs identified by DADA2 provide a more realistic representation of a true Cheddar cheese microbiome, leading to the conclusion that DADA2 is the better denoising pipeline to use when testing cheese samples because it provides a more realistic representation of the actual microbiota in the aging samples.

(91 Pages)

PUBLIC ABSTRACT

Comparison of Microbial Diversity of Fifteen Aged Cheddar Cheeses from Different Regions Using Next Generation Sequencing

Sophie Overbeck

This project was funded by BUILD Dairy to test the efficacy of using next generation sequencing to study Cheddar cheese microbiomes. The 15 different cheese samples used in this study were purchased in the retail market from various manufacturers and different origins of manufacture. Sequencing was done by the Center for Integrated Biosystems at Utah State University using Illumina MiSeq.

The aim of this project is to provide further understanding of how using next generation sequencing can benefit the study of bacterial communities and their influence during the aging process of Cheddar cheese and the possibility of identifying defect causing microorganisms. We tested Cheddar cheese samples microbiota by amplifying and sequencing the hypervariable V4 region of the 16S rRNA. Sequencing data was evaluated using two different denoising pipelines within Qiime 2: Deblur and DADA2. We used variant identification as a method of comparing starter and non-starter lactic acid bacteria in cheeses based on regions and manufacturers.

In the data denoised by both methods the samples clustered into two groups separated by the dominant starter lactic acid bacteria (SLAB). DADA2 identified over quadruple the number of amplicon sequence variants (ASVs), which represent a unique sequence within the samples, compared to Deblur. Using Deblur we identified 5 *Lactococcus lactis* ASVs and 3 *Streptococcus thermophilus* ASVs compared to DADA2 which identified 34 *L. lactis* ASVs and 35 *St. thermophilus*. Thus, it can be said that using DADA2 to denoise Cheddar cheese data provides a more realistic representation of the microbiome

ACKNOWLEDGMENTS

First and foremost, I want to express my deep gratitude to my advisor, Professor Donald J. McMahon for giving me the amazing opportunity to study and research with him. His expertise and encouragement have transformed my way of thinking about problems and challenges shaping me into a better scientist and writer. I would like to thank him for his support, understanding, and guidance throughout my time at Utah State University.

I would like to extend a special thank you to my committee member Prof. Craig J. Oberg for his moral support during my studies as well as for introducing me to BUILD Dairy and food science research during my undergraduate studies, I wouldn't be where I am if it weren't for you. I would also like to thank my committee members Prof. Michael Lefevre, and Prof. Taylor Oberg for their time, support, and expertise in microbiome research and bioinformatics. In addition, I would like to thank Deanna Larsen for her lab training and time.

I would like to express my gratitude to Dr. Eric Bastian for his work with BUILD Dairy and providing an environment where Dairy researchers feel encouraged and connected to the industry and farmers. Thank you to Kim Rasmussen and Tara B. Johnson for always being willing and able to help with questions and provide words of support and encouragement. I would also like to thank my fellow lab mates Rhitika Poudel and Brynli Tattersall for their help. Special thanks to Ireland Green my friend and lab mate who provided me with constant support both morally and physically.

Lastly, I would like to thank my family, my parents Cristen and Kirk Ottley and Chuck Overbeck, for teaching me to always prioritize the pursuit of knowledge and giving their unconditional love and support. I want to express my gratitude to my siblings Tom Overbeck, Chloe and Artie Casto, Emmett and Jenna Overbeck, and Phil Overbeck for their advice,

emotional support, and love throughout this process and helping me through the challenges I faced. I would like to thank Trae Kelley for his love, encouragement, understanding, and patience.

Sophie Overbeck

CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

LIST OF ABBREVIATIONS

LAB = Lactic acid bacteria

SLAB = Starter lactic acid bacteria

NSLAB = Non-starter lactic acid bacteria

ALAB = Adjunct lactic acid bacteria

BCAAs = Branched-chain amino acids

DGGE = Denaturing gradient gel electrophoresis

TGGE = Temperature gradient gel electrophoresis

SSCP = Single-strand conformation polymorphism

NGS = Next generation sequencing

HTS = High-throughput sequencing

HVR(s) = Hypervariable region(s)

V = Variable

OTUs = Operational taxonomic unit

ASVs = Amplicon sequence variants

CPU = Core processing unit

q2 = QIIME 2 2019.4

INTRODUCTION

In general, it is understood that over the cheese aging process there is a shift in the lactic acid bacteria (**LAB**) microbial community from starter lactic acid bacteria (**SLAB**) dominance to a non-starter lactic acid bacteria (**NSLAB**) dominated community. Every cheese has its own unique microbiome responsible for flavor development as well as producing defects. Recent advancements in next generation sequencing (**NGS**) facilitates a more accurate identification of microbial communities and thus implying their potential affect on functional properties.

Next generation sequencing has the potential to reveal microorganisms and their associated pathways responsible for favorable characteristics as well as unfavorable properties in aging cheese (Jonnala et al., 2018). Starter lactic acid bacteria, adjunct lactic acid bacteria (**ALAB**), and NSLAB collectively contribute to the flavor development in Cheddar cheese, while some NSLAB are also associated with causing defects. The development of flavor occurs through several basic biological mechanisms including lactose fermentation, conversion of milk proteins (primarily caseins) into peptides and free amino acids, and the breakdown of lipids, esters, citrate, and amino acids into volatile aroma compounds (Broadbent and Steele, 2005).

A challenge in understanding and controlling the microflora of cheese during the aging process is being able to identify and enumerate the different species of bacteria present, both SLAB and NSLAB. An ideal method should be relatively inexpensive and have the capability of tracking individual strains of SLAB, ALAB, and NSLAB over the storage time of cheese. The aim of this project was to use NGS to study the 16S rRNA V4 region of DNA extracted from cheese to determine the diversity of the bacterial community within aged Cheddar cheese as well as to identify the relative abundance of variants.

HYPOTHESIS AND OBJECTIVES

I hypothesize that using next generation sequencing with different Cheddar cheese samples from various regions will result in cheeses clustering by region.

To test this hypothesis, I will perform the following objectives:

1. Extract DNA from Cheddar cheese and analyze the microbiota using primers from the 16S rRNA V4 region obtain next generation sequencing of the DNA.
2. Analyze the sequence data using Qiime2 bioinformatics and compare two denoising platforms for comparison of variant identification.
3. Compare microbiota community differences between the cheeses.

LITERATURE REVIEW

Cheddar Manufacture

Modern Cheddar cheese manufacture is dependent on being able to control the drop in pH and the expulsion of moisture within a certain time frame. Availability of reliable starter cultures has made this possible. Despite the mechanical manufacturing advancements made to the cheese making process, Cheddar cheese remains a difficult cheese variety to manufacture with large variability due to the lengthy ripening period needed for flavor development (Lawrence et al., 2004).

Cheddar cheesemaking involves the removal of moisture from milk inoculated with starter cultures and coagulated using rennet. The rennet-induced coagulum is cut into small cubes to aid in moisture removal. The cubes are then cooked and stirred until the required amount of acid has been produced. The whey is then removed and the cubes of curd are fused into slabs by cheddaring with the cheddared curd then milled, salted, and pressed. After pressing the salted curd, the cheese is packaged and stored at 2-7 °C until maturation. Cheddar cheese is typically ripened for 3-12 months and over this time flavor formation occurs as well as changes in textural properties (Lawrence et al., 1984).

Lactic Acid Bacteria

The role of LAB is to produce lactic acid in fermented milk products. Lactic acid bacteria are a heterogeneous group of microorganisms that can be characterized as anaerobic or facultatively anaerobic, non-spore forming, non-motile, catalase-negative, gram positive rods or cocci that convert carbohydrates into lactic acid (Chen and Hang, 2019). Milk fermentation by LAB can occur spontaneously or through inoculation using starter cultures (Widyastuti et al., 2014). Acid production from LAB contributes to preservation of products as well as flavor

development and, therefore, has been widely utilized in both dairy and non-dairy food fermentations. The main LAB in dairy products are species of *Lactococcus, Lactobacillus*, *Leuconostoc*, *Pediococcus*, and *Streptococcus*. In cheddar Cheese manufacture, LAB can be divided into two categories: starter lactic acid bacteria (SLAB) defined as intentionally added LAB and non-starter lactic acid bacteria (NSLAB), which are LAB already present in the milk flora or introduced via manufacturing equipment.

Starter Cultures

The primary cultures of LAB used in cheese making are referred to as SLAB because they produce lactic acid from lactose in the cheese making process. The improvement of milk hygiene and the commercialized standardization and acceleration of ripening has led to blander-flavored cheeses. Blander-flavored cheeses prompted cheese makers to incorporate adjuncts or secondary starters to improve the organoleptic qualities of the cheese especially during accelerated ripening (Parente and Cogan, 2004). Secondary starters, also referred to as adjunct lactic acid bacteria (ALAB), contribute to the organoleptic properties of the finished cheese but do not contribute to acid production during cheesemaking.

Starter lactic acid bacteria available today are selected for specific properties being derived either from selective pressure by means of heat treatment, incubation temperature, or low pH, or from natural starters of undefined composition by means of backslopping and the use of natural whey cultures (Parente and Cogan, 2004). Lactic acid bacteria starters can be categorized as single-strain, multi-strain, and mixed strain. Single-strain starters only contain one strain, multi-strain starters contain more than one strain and the strains may belong to the same species or they can be a combination of species that have similar characteristics, while mixed starters contain many unknown strains (Chen and Hang, 2019). The starters are further characterized

according to their optimum fermentation temperature with the two main categories in the dairy industry being mesophilic starters with optimal growth temperatures between 20 and 30°C, or thermophilic starters with optimal growth temperatures between 40 and 45°C.

Non-starter Cultures

In Cheddar cheese made from pasteurized milk there are adventitious ("contaminants"), termed non-starter lactic acid bacteria (NSLAB), these LAB are always nonpathogenic (Broadbent, 2011). During the early aging stage, the young cheese has low levels of NSLAB, but after a few months of ripening the NSLAB population inevitably begins to grow eventually reaching high numbers. Cheese made using both high quality milk and good sanitary conditions will contain initial NSLAB populations below 10^2 cfu g^{-1} but can reach a plateau of cell density between 10^7 to 10^8 cfu g^{-1} after 6 mo. (Broadbent, 2011). Non-starter lactic acid bacteria in Cheddar most commonly found in aged cheese include *Lacticaseibacillus casei, Lacticaseibacillus paracasei, Latilactobacillus curvatus, Lacticaseibacillus rhamnosus Levilactobacillus brevis,* and *Lactobacillus plantarum* (Banks and Williams, 2004). These were reclassified in 2020 and were previously in the Lactobacillus genus (Zheng et al., 2020) although are still considered lactobacilli as described by McMahon et al. (2020).

The final numbers of NSLAB in Cheddar cheese appears to be stable and is not impacted by their initial levels within the fresh curd. Populations of NSLAB in the U.S. are dominated by facultatively heterofermentative species of lactobacilli such as *Lc. paracasei* and *Latl. curvatus*. Facultatively heterofermentative LAB are characterized by their ability to produce a variety of end products depending on which sugars are available, since they can ferment six carbon glucose sugars to lactic acid and can also convert five carbon pentose sugars to lactic acid, acetic acid, carbon dioxide, and other end products. Many of these end products contribute to flavor

development during cheese aging. Obligatory heterofermentative lactobacilli are usually found at lower numbers in Cheddar cheese but have often been associated with unwanted gas production (Oberg et al., 2016).

Although NSLAB are generally required for mature flavor development, they can also contribute to undesirable defects such as abnormal flavor, lactate crystal formation, and gas slits (Banks and Williams, 2004). Extent of growth as well as diversity of lactobacilli in Cheddar during ripening is dependent on processing factors, including milk quality, factory hygiene, the cooling rate of the cheese after manufacture, and the storage temperature for ripening (Fox et al., 1998). In addition, NSLAB can cause a positive, neutral, or negative influence on cheese flavor development mainly through catabolizing amino acids during cheese ripening (Broadbent, 2011).

Flavor Development

Starter lactic acid bacteria, adjunct or secondary cultures, and NSLAB collectively contribute to the flavor development in Cheddar. Flavor development occurs through several basic biological mechanisms including lactose fermentation, conversion of milk proteins (primarily caseins) into peptides and free amino acids, and breaking down lipids, esters, citrate, and amino acids into volatile aroma compounds (Broadbent and Steele, 2005). During ripening, microorganisms and enzymes within the cheese matrix act on available nutrients including carbohydrates, proteins, citrate, and lipids. The combination of both primary and secondary proteolysis of caseins influences cheese flavor in several ways. Breaking down the casein network softens the cheese texture causing flavor compounds to be released when consumed, as well as producing low-molecular-weight peptides directly affecting flavor (causing bitterness), and liberating free amino acids influencing flavor development.

Metabolism of branched-chain amino acids (**BCAAs**), aromatic amino acids, and methionine results in compounds that have a strong effect on cheese flavor (Broadbent, 2011). The metabolism of BCAAs contributes to a wide range of flavor compounds, including aldehydes, alcohols, and various acids. Branched-chain amino acids metabolism is extremely varied among NSLAB, which can contribute to 'cheesy' aroma, 'goaty' aroma, or imparting 'dark chocolate' flavor. Catabolism of aromatic amino acids by NSLABs can produce floral and pungent aromas commonly associated with off-flavors, in contrast, metabolism of methionine has a positive effect by producing methanethiol and is associated with the development of 'cheesy' flavors (Singh et al., 2003).

Studying Cheese Microbial Communities

In studying cheese microbial communities, the methods used can be grouped into two categories: culture-dependent methods and culture-independent methods. Culture-dependent methods use culture plating techniques for counting and isolation of viable microorganisms followed by the application of phenotypic and genotypic tools for evaluation (Ndoye and Lapointe, 2011). Phenotypic tools used in culture-dependent methods include microscopic examination, growth tests, determination of assimilation and fermentation patterns, and use of SDS-polyacrylamide gel separation of cellular proteins.

Genotypic tools used in culture-dependent techniques include DNA restriction mapping, ribotyping, DNA amplification, RNA or protein-coding gene analysis, using specific nucleic acid probes, and DNA arrays. One of the drawbacks of using culture-dependent methods is that only a fraction of microorganisms can be isolated using these techniques (Martin-Platero et al., 2009). Therefore, when using culture-dependent methods the real composition of the whole microbiome may not be reflected by the isolated strains. Another downside to using culture-based methods is

that some species are unable to grow because they are either outnumbered by more abundant microbial species or lack the capability to grow *in vitro*. Furthermore, some microbial species have long culture periods making propagation methods time consuming.

Unlike culture-dependent techniques, culture-independent methods do not require actually propagating microorganisms. Culture-independent techniques are based on direct extraction of total DNA or RNA from the cheese sample (Ndoye and Lapointe, 2011). Some of the common culture-independent methods used in cheese studies include denaturing gradient gel electrophoresis (**DGGE**), temperature gradient gel electrophoresis (**TGGE**), single-strand conformation polymorphism (**SSCP**), and next-generation sequencing (**NGS**). Both DGGE and TGGE are used as fingerprinting methods to study bacterial population dynamics and diversity during cheese manufacturing (Muyzer, 1999). Several drawbacks to these methods are that they are not well adapted to detect nondominant populations and band intensity does not always agree with plate counts, which means these techniques are not quantitative methods. Similar to T/DGGE, SSCP relies on electrophoretic separation of PCR products and has been used in studies to analyze dynamics between microbial populations in cheese (Delbès, 2007). The SSCP method is only able to provide a general overview of the relative abundance of dominant species within the cheese and is not able to precisely identify microbes at family or genus level. The newest technique currently used in researching microbial populations is NGS which will be further discussed in the next section.

Next Generation Sequencing

Next generation sequencing also known as high-throughput sequencing (**HTS**) is a catch-all term describing modern sequencing techniques and refers to large-scale DNA sequencing that allows for the whole genome, whole exome, or exons for select genes to be queried. The earliest

environmental microbial DNA cloning was first proposed by Lane et al., (1985), and the term ‘metagenome’ was proposed by Handelsman et al. (1998) describing the entire collection of genomic information of all microorganisms found within a given environment, nowadays also described as the microbiome.

Microbiome studies require high-quality sequence data to generate valid results and, therefore, require having sufficient numbers of samples, isolating samples from biologically relevant sites, controlling for confounding factors, as well as using appropriate analytical tools (Bella et al., 2013). Furthermore, choosing the appropriate sequencing technology is of vital importance as each technology has its own maximum capacity controlling the number of sequencing reads they can undertake, which influences how many samples can be analyzed at once. Sequencing depth, meaning the number of reads that can be produced during sequencing, is very important in regard to being able to resolve rare species or detect differences between samples (Bella et al., 2013).

Utilization of NGS technology can provide valuable information related to manufacturing processes, seasonal variation, geography, climatic conditions, use of raw or pasteurized milk, and a variety of other factors which contribute to the cheese microbiota. There are three primary approaches for microbiota analysis using NGS: amplicon sequencing using a fragment of a highly conserved gene, ideally containing 16S hypervariable region(s) (**HVR(s)**) for sequencing using comparison to databases allowing taxonomic assignment, shotgun metagenomic sequencing involving non-targeted sequencing of DNA in a sample also using databases for taxonomic assignment, and metatranscriptomics RNA sequencing where the total mRNA in the sample is sequenced revealing the extent of different genes being expresses which demonstrates the relative activity of different components of the microbial community (Jonnala et al., 2018).

Hypervariable Region

When applying HTS to profile a microbial community, selecting a suitable HVR(s) of the 16S rRNA gene is a critical step which influences the achieved resolution of bacterial diversity (Rajeev et al., 2020). Over the past decade, there have been various studies researching different variable (**V**) regions or multiple regions of the 16S rRNA gene (Zhang et al., 2018). Deep 16S rRNA gene sequencing of V regions is currently the predominant tool for studying microbial ecology.

The 16S rRNA gene is comprised of approximately 1600 bp and contains nine hypervariable regions with varying conservation named V1-V9 (Bukin et al., 2019). There are regions within the 16S rRNA gene that are more conserved. These more conserved regions are used to determine higher-ranking taxa, in contrast to the more quickly evolving regions with hyper variability help identify genus or species. In a previous study for soft-cheese manufacture they tested two different hypervariable regions V1-V3 and V3-V4 for their abilities to reveal Lactococcus lactis diversity (Saltaji et al., 2020). The study showed that the V3-V4 produced more reads overall. Several studies have shown V4 as a promising region with results allowing higher species richness determination than other regions with higher coverage and a broader spectrum in the Bacteria domain (Zhang et al., 2018; Rajeev et al., 2020). Previous HTS studies on dairy specific products have used the HVR V4 region for analysis (Salazar et al., 2019; Duarte et al., 2020; Choi et al., 2020).

Bioinformatics

The term “bioinformatics” first began being used in the mid- 1980s as a way to describe the application of information science and technology in the life sciences. This definition was very general and included everything from robotics to artificial intelligence. Currently

bioinformatics is more appropriately recognized as the science of how information is generated, transmitted, received, stored, processed, and interpreted in biological systems, or more plainly defined as the application of information science to biology (Ramsden, 2015).

Next generation sequencing techniques produce massive amounts of data and, consequently, to draw any useful conclusions it is essential to computationally analyze the data using bioinformatic pipelines (Bella et al., 2013). Of the bioinformatics pipelines available, Qiime is the most popular (Caporaso, 2010). Qiime 2 is the newest version of Qiime and both are next-generation bioinformatics platforms that are extensible, free, open source, and community developed (Bolyen et al., 2019). It is strongly recommended by Qiime's website to use a computer with a minimum of 8 gigabytes of RAM. In addition to memory requirements, Qiime 2 must be in a conda environment to operate. Conda is an open-source environment management system that runs on Windows, macOS, and Linus (Anaconda Software Distribution, 2020).

Denoising

High-throughput sequencing of the 16S rRNA gene is a strategy commonly used to study microbial communities. Sequence reads are traditionally clustered into operational taxonomic units (**OTUs**) at a defined identity threshold in measure to avoid sequencing errors of false taxonomic units and sequencing generated artifacts. As means of eliminating sequencing errors to determine real biological sequences, numerous bioinformatic packages have been made available to denoise microbiome data. Denoising in bioinformatics is a crucial feature that reduces the amount of erroneous OTUs and, therefore, increases the accuracy of the entire analysis pipeline.

The different denoising pipelines determine real biological sequences using single nucleotide resolution by generating amplicon sequence variants (**ASVs**) also referred to as zero

noise OTUs and sub-OTUs. Two of the most commonly used denoising packages are Deblur and DADA2 which are both available through Qiime 2. The two pipelines differ in their algorithms used to denoise as well as in run times. The pipelines result in similar general community structure but there is considerable variation in the number of ASVs/OTUs and resulting alpha-diversity that should be taken into consideration when attempting to identify rare organisms (Nearing et al., 2018). Deblur and DADA2 differ from other denoising pipelines by reconstructing exact biological sequences within the sample creating ASVs as opposed to OTUs (Prodan et al., 2020). Previous studies comparing these two denoising pipelines used mock communities to evaluate the differences and effectiveness of the algorithms (Nearing et al., 2018; Prodan et al., 2020).

Deblur

Deblur is open source under the Berkley Software Distribution license making it easily and freely accessible through the Qiime 2 platform. The Deblur algorithm works independently on each sample. Deblur aligns sequences together into ASVs based upon an upper error rate with the mean error rate and a constant probability of indels removing predicted error-derived reads from adjacent sequences (Amir et al., 2017). Indel is a term used in molecular biology used to describe an insertion or deletion of bases in the genome which is then applied to sample-by-sample approach having an advantage of reducing memory requirements and computational demand (Nearing et al., 2018).

The algorithm compares sequence-to-sequence Hamming distances within a sample. Hamming distance is a metric used to compare two binary data strings of equal length with the distance being the number of bit positions in which the two strings are different. The implementation of the Deblur algorithm can be broken down into three steps. Firstly, sequences

are sorted by abundance. Secondly, the number of predicted error-derived reads are subtracted from the neighboring reads using the Hamming distance in an order from the most to least abundant sequence. Lastly, any sequence with an abundance that drops to 0 after a subtraction of the Hamming distance is removed from the list of valid sequences (Amir et al., 2017).

DADA2

DADA2 stands for Divisive Amplicon Denoising Algorithm 2 and is open source under the Lesser General Public License (LGPL-3.0). DADA2 uses a parametric model to determine true biologic sequences from reads. The algorithm relies on input read abundance and distances with the reasoning of true reads. Meaning true reads are likely to be more abundant and less abundant reads with a few base-differences from more abundant sequences being more likely to be error-derived (Prodan et al., 2020). There are some structural differences between DADA2 and most other denoising pipelines. One of the biggest differences is DADA2 performs merging of paired-end reads after denoising samples. The core denoising algorithm uses the empirical relationship between the quality score and error rates, and when reads are merged, the relationship will differ between the overlapping, forward-only, and reverse-only portions of the merged read creating a variation that interferes with the denoising algorithm (Callahan et al., 2016). Therefore, greater accuracy can be achieved using DADA2 by merging after samples are denoised.

DADA2's core denoising algorithm is slower than others because of the computational demands. The DADA2 pipeline includes filtering, dereplication, denoising, chimera detection, and merging paired ends. Filtering trims sequences to a specified length, removing any sequences shorter than that length, and filters based on a minimum quality score and the number of ambiguous bases (Callahan et al., 2016). Dereplication creates a list of unique sequences and

their abundances and creates consensus quality scores for the unique sequences by taking the mean of the positional qualities of the component reads. The denoising step implements the core denoising algorithm. In the chimera detection sequences that are exact bimeras, and then searching for combinations of right-parent and left-parent sequences all chimeric sequences are flagged and removed. DADA2 uses a method with sensitive chimera detection allowing this pipeline to differentiate closely related variants creating ASVs. The final step is merging the paired sequences if they match exactly.

METHODS

Cheese

Fifteen aged Cheddar cheeses manufactured in different locations were purchased in the retail market (Table 1). This included an aged Cheddar cheese made at Utah State University (Logan, UT) for which the manufacturing protocol was known including the SLAB and ALAB used. Information on the typical NSLAB present in the cheese was also available.

DNA Extraction

Bacterial DNA in the Cheddar was extracted using a pretreatment followed by a modified DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) extraction protocol. Cheese samples were stomached for 4 min at 260 rpm using 11 g of cheese into 99 mL of sterile 2% sodium citrate (Fisher Scientific, Fair Lawn, NJ). Then 10 mL of homogenate was transferred into a 15-mL sterile centrifuge tube and centrifuged at 8,000 x *g* for 15 min. Following centrifugation, the supernatant was removed, and the pellet suspended into 1 mL of sterile 0.2 M sodium phosphate buffer (pH 7) (Fisher Scientific). The suspended pellet was then transferred into a sterile 1.5-mL microfuge tube and centrifuged at 13,000 x *g* for 10 min and supernatant discarded. The pellet was suspended in warmed(60°C) bead solution from the PowerSoil kit and transferred back into the bead tube included in PowerSoil Kit. Reagent C1 was then added per PowerSoil kit protocol with the addition of 10 μL of sterile 5-mg/mL lysozyme (Fisher Scientific) in phosphate buffer (Fisher Scientific) and incubated at 37 °C for 1 h.

The PowerSoil kit's protocol was followed with the exceptions of adding 25 μL proteinase K (Qiagen, not included in the kit) during step 6 used to digest proteins and remove contamination. Additional modifications included taking all of the supernatant at steps 9 and 12 instead of the 600 μL and 750 μL listed in the PowerSoil protocol as a way of maintaining as

much DNA in the samples as possible. A final modification to the kits protocol was using a final volume of 50 μL of reagent C6 instead of 100 μL at the end of the protocol (step 19) as a means of not diluting the final DNA product (Appendix A.). Following extraction DNA samples were quality checked using NanoQuant Plate (Tecan, Zürich, Switzerland).

DNA Amplification, Barcoding, and Sequencing

The samples underwent two rounds of PCR using Thermo Fisher Scientific PCR reagents. During the first round of PCR the set of primers used were S-D-Bact-0341-b-S-17 (CCTACGGGNGGCWGCAG) and S-D-Bact-0785-a-A-21 (GACTACHVGGGTATCTAA TCC) obtained from Integrated DNA Technologies (IDT; Coralville, IA). These primers amplify 460 bp from the 5' end of the 16S rRNA gene targeting the hypervariable V3-V4 region. A second round of PCR added unique forward and reverse barcode primer combinations to each sample (see Appendix B) obtained by IDT allowing samples to be identified during downstream analysis. The prepared DNA samples were then sequenced by the Center for Integrated Biosystems (Utah State University, Logan, UT) using the Illumina MiSeq system with MiSeq Reagent Kit v3 (600 cycle).

Downstream Analysis

Microbiome bioinformatics were performed using QIIME 2 2019.4 (q2) software (Bolyen et al., 2019) with the data uploaded and accessed remotely at the Center for High Performance Computing (**CHPC**; University of Utah, Salt Lake City, UT) using FastX2, from StarNet. The CHPC remote server provided access to 80 core processing units (CPUs) and 376 GB of memory to use for analysis.

Raw sequence data was demultiplexed, joined, and quality filtered using the q2-demux plugin followed by denoising. The sequence data was denoised using two plugin programs:

Deblur (Amir et al., 2017; see Appendix B) and DADA2 (Callahan et al., 2016; see Appendix C). After the samples were denoised, ASVs were aligned with mafft (Katoh et al., 2002) using q2-allignment and then phylogenic trees created using fasttree2 (Price et al., 2010) using q2-phylogeny. Taxonomic analysis was performed using q2-feature-classifier to assign taxonomy to ASVs using Silva 132 99% OTUs aligned sequences (Quast et al 2012). All ASVs were then assigned an OTU identification using the National Center for Biotechnology Information (NCBI) database.

RESULTS AND DISCUSSION

Sample Identities

The fifteen different samples were manufactured in different regions, with 9 of the cheeses manufactured in the United States, 5 originating from Ireland, and one originating from Australia (Table 1). The types of Cheddar cheeses purchased for this study included medium, sharp, extra sharp, aged, and reserved. The term sharp refers to the flavor profile resulting from the aging process. There are no legal definitions or requirements for Cheddar cheese to be labeled as “sharp” or “aged” resulting in a retail market where products commercially labeled as “sharp” or “aged” may actually vary widely in actual ripening time (Drake et al., 2009).

Even though there is not an industry standard for labeling, Cheddar cheese types are assigned based on how long they have been aged with mild Cheddar being ripened the least amount of time typically 1 to 3 mo, medium Cheddar is ripened 3 to 6 mo, and sharp Cheddar is ripened over 6 mo or more as defined by the U.S. Dairy Export Council (2016), extra sharp Cheddar is ripened 15 mo or greater, and reserve Cheddar is often a Cheddar cheese that has been aged between 2 to 6 yr. Although there may be slight variation among manufacturers these ranges of categorizations of sharpness have been accepted for this study. The Cheddar cheeses purchased have a type listed which indicates what type they were at the time they left the packaging facility. However, there is no way of knowing the exact age of the cheeses at the time of DNA extraction for a number of reasons including not knowing how long it took to for the Cheddar cheeses to reach the retail outlet or how long the Cheddar cheeses were there before they were purchased.

Both the manufacturer and origin of manufacture are listed with the cheese sample identity (Table 1). In a previous study, Cheddar cheeses from Ireland, New Zealand and the

Table 1. Cheeses used for DNA extraction.

Sample	Manufacturer/Brand	Type	Origin
A	Tillamook	Medium Cheddar	Oregon, USA
B	Cracker Barrel	Vermont Sharp White	Vermont, USA
C	Cabot	NY Extra Sharp	New York, USA
D	Old Crock	Sharp Cheddar	Australia
E	Kerrygold	Skelling	Ireland
F	Kerrygold	Reserve Cheddar	Ireland
G	Kerrygold	Dubliner	Ireland
H	Oscar Wilde	Vintage Cheddar	Ireland
I	Utah State University	Aged Cheddar	Utah, USA
J	Murray's	Irish Cheddar	Ireland
K	Spring Hill Jersey Cheese	Yellow Cheddar	California, USA
L	Hilmar	Medium Cheddar	California, USA
M	Hilmar	Vintage Sharp Cheddar	California, USA
N	Carr Valley Cheese	Medium Cheddar	Wisconsin, USA
O	Babcock	White Aged Cheddar	Wisconsin, USA

United States were analyzed for differences in flavor (Drake et al., 2005). The results showed distinctive differences between Cheddar cheeses from each of the geographic locations as noted by the highly trained tasting panel, suggesting international differences in Cheddar cheese. It is known that SLAB, adjunct or secondary cultures, and NSLAB collectively contribute to the flavor development in Cheddar. Whether the milk comes from grass fed cows or cows fed a total mixed ratio diet is also important to final cheese flavor because milk from grass fed cows has a distinctive flavor even if the same cultures are used to make the cheese (Khanal et al., 2005; Crossiant et al., 2007). Therefore, the manufacturer and origin of manufacture are important categorizations for our cheese samples.

Sequencing Output

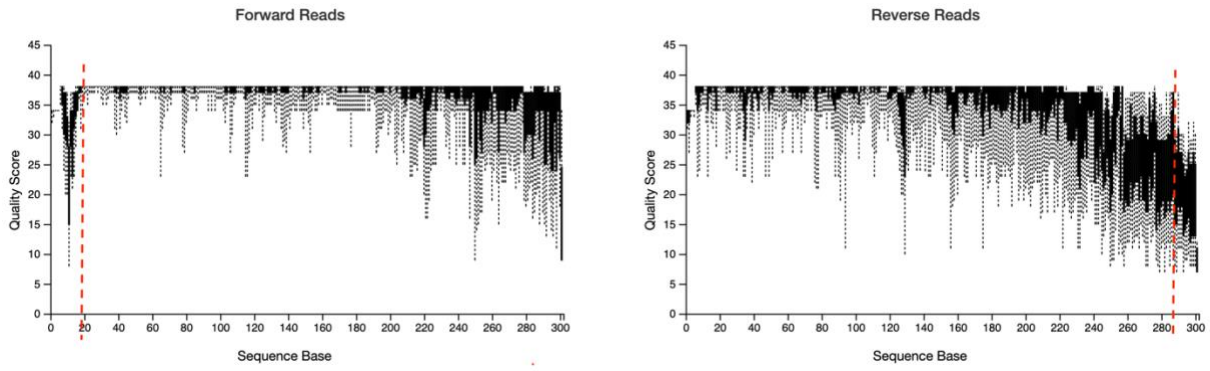
The total number of raw sequence reads before denoising the 15 Cheddar cheese samples was 15,193,666. Individual samples ranged between 854,340 and 1,197,119 reads with an

average of 1,012,911 reads per sample. Looking at the interactive quality plot generated in demux-paired-end.qzv (see Appendices C and D), I decided to trim left 17 base pairs (bp) and to trim right from 460 bp to 448 bp (Figure 1).

After sequencing of the DNA and initial processing of data in Qiime2 using q2-vsearch and applying a quality filter (see Appendix C), there were 13,669,044 reads going into the Deblur denoising algorithm. Using a denoising step is a crucial step in the bioinformatic analysis used to eliminate sequencing errors and determine real biological sequences within the samples. Denoising sequences allows for greater taxonomic resolution and is now the standard in bioinformatics. After using Deblur to denoise the reads, there were 6,152,601 filtered reads resulting in only 40.5% of the raw reads being retained. This is an average of 410,000 reads per cheese sample.

When using DADA2 there were 15,193,666 sequence reads going into the denoising algorithm, about one and a half million more sequences than the sequences that went into Deblur. This difference is because unlike Deblur, DADA2 joins paired ends and denoises sequence data. After using DADA2 to denoise the raw reads, there were 11,547,210 filtered reads (770,000 reads per sample) resulting in 76% of the reads being retained. DADA2 retained 1.9 times more reads compared to Deblur.

Figure 1 Interactive quality plot generated by Qiime 2 with areas to trim marked by red dashed lines.



Deblur

After denoising using Deblur, 74 unique ASVs across the 15 different samples were identified (see Appendices E and F). Each ASV has its own DNA sequence. Deblur had 30 DNA sequences that were only identified by Deblur and had 44 DNA sequences that were also found by DADA2. These 74 ASVs were then grouped as *Lc. lactis*, *St. thermophilus*, lactobacilli, or other species (Table 2). The category of other species does not contain bacteria that traditionally contribute to the cheese making process or considered as LAB typically present in cheese.

Of the 74 unique ASVs, 39% were identified as *Lc. lactis* or *St. thermophilus*, although within each cheese they accounted for 98.25% of the relative abundance. Unexpectedly, this makes SLAB the most dominant category of bacteria in the aged Cheddar cheese samples at almost 2 logs higher than NSLAB. Twenty-three percent of ASVs were identified as various lactobacilli which can include ALAB and NSLAB. Within individual cheeses, lactobacilli only accounted for an average of 1.73% relative abundance. There were 28 other ASVs identified as various species but at a very low relative abundance within a cheese of 0.01% or less.

It has been noted that during cheese maturation, the salt-in-moisture concentration and lack of sugars as nutrients causes some SLAB to lose their viability and at the same time NSLAB lactobacilli begin to multiply, plateauing after 3 to 9 mo at cell densities of 10^7 to 10^8 cfu/g (Broadbent et al., 2011). The cheese samples used in this study were all aged at least 6 mo and it was expected that ASV relative abundance would show a greater abundance of lactobacilli than *Lc. lactis* and *St. thermophilus*. However, the relative abundance observed showed SLAB to have populations 2 logs higher than NSLAB populations.

Starter LAB reach their peak cell density during the day of manufacture with counts up to 10^9 cfu/mL and within the first mo. of ripening the viable SLAB decline to approximately 1%

Table 2. Relative abundance of amplicon sequence variants after denoising using Deblur identified as starter culture (*Lactococcus lactis* or *Streptococcus thermophilus*), adjunct and nonstarter lactobacilli and other species in aged cheeses color coded from lowest abundance (yellow) to greatest abundance (red) with black representing none detected.

Cheese[1]	*Lactococcus lactis*	*Streptococcus thermophilus*	Lactobacilli	Other species.
	----------(%)----------			
A	99.8048	0.0388	0.1560	0.0002
B	46.8270	52.1020	0.0278	0.9502
C	99.4799	0.0757	0.4371	0.0071
D	18.1103	80.8374	0.5322	0.5201
E	32.4472	67.4674	0.0519	0.0335
F	91.1881	3.6333	5.0828	0.0957
G	0.1752	98.8798	0.9414	0.0037
H	14.3854	85.0878	0.4608	0.0660
I	88.5547	0.0919	11.3532	0.0002
J	22.6283	76.0801	0.9910	0.3006
K	36.7049	63.0468	0.0000	0.2482
L	94.4577	0.0749	5.4329	0.0346
M	99.6077	0.0378	0.3544	
N	99.8632	0.0466	0.0818	0.0080
O	99.9256	0.0580	0.0133	0.0031

[1]Cheeses as described in Table 1.

of maximum numbers (McSweeney et al., 1994; Banks and Williams, 2004). Many culture-dependent approaches have inherent limitations in terms of their abilities to recreate real conditions under which most bacteria are growing in their natural habitat leading to the cultivatable populations not being completely reflective of the communities and the actual microbial diversity (Ruggirella et al., 2014). One way to explain the SLAB dominance in relative abundance is the possibility that the SLAB populations are in a viable non-culturable (**VNC**) state.

During cheese storage after carbohydrate depletion, lactococci can enter a VNC state, meaning that although the cells have an inability to replicate, they do not necessarily die. During this stage, *Lc. lactis* cells have been observed using spectofluorometry and demonstrated their ability to express lytic and cell wall repair genes showing that they did not undergo lysis but rather maintained their cellular membrane aiding in preservation of nucleic acids (Ganesan et al., 2007). At the beginning of Cheddar cheese making, milk is inoculated with high levels of starter cultures and during the cheese ripening some starter cultures die or enter a VNC state where they preserve their nucleic acids. Therefore, the dominance of SLAB in the relative abundance observed reflects a population of initial starter cultures within the cheeses.

The lactobacilli group was observed ranging in ASV relative abundances between 0.0004% to 11.35%. Generally, the lactobacilli group is thought of as the NSLAB population within aged Cheddar cheese. However, it is known that the cheese sample from Utah State University (Cheese I) uses *Lactobacillus helveticus* as an ALAB when making this particular aged Cheddar cheese explaining the high relative abundance of 11.35% (Table 3) observed within the lactobacilli group. Cheese F also had a higher relative abundance of 4.5% for *Lb. helveticus* when compared to the other relative abundances observed for the identified OTUs.

Table 3. Relative abundance of best match of operational taxonomic unit of amplicon sequence variants after denoising using Deblur sequenced from aged cheese grouped as various lactobacilli species and color coded from lowest abundance (yellow) to greatest abundance (orange) with black representing none detected.

	Lactobacilli OTU Identifications											
Cheese[1]	*Lacticaseibacillus paracasei*	*Lacticaseibacillus rhamnosus*	*Lactiplantibacillus plantarum*	*Lactobacillus crispatus*	*Lactobacillus delbrueckii*	*Lactobacillus helveticus*	*Latilactobacillus curvatus*	*Lentilactobacillus buchneri*	*Lentilactobacillus kefiri*	*Loigolactobacillus coryniformis* subsp. *torquens*	*Paucilactobacillus wasatchensis*	*Secundilactobacillus malefermentans*
	----------%----------											
A	0.1362	0.0198										
B					0.0278							
C		0.1519						0.0144	0.1192		0.1517	
D				0.0029	0.0022	0.5265					0.0006	
E			0.0035	0.0003	0.0006	0.0469					0.0006	
F	0.0675					4.5141					0.0432	0.4580
G						0.9414						
H	0.3993					0.0579			0.0005		0.0031	
I						11.3532						
J				0.0030		0.9830						0.0050
K												
L	0.0615				4.0943	0.9502	0.3117		0.0147			0.0005
M			0.0006		0.2920	0.0278	0.0040		0.0301			
N							0.0806					0.0012
O							0.0004			0.0129		

[1]Cheeses as described in Table 1.

Cheese L contains *Lactobacillus delbrueckii* subsp. *lactis* with a relative abundance of 4%. All other lactobacilli observed had relative abundances <1%.

The use of ALAB during cheese making can be for a number of purposes including preventing bitterness, modifying the flavor profile, or for their ability to produce proteins that protect against the growth of unwanted bacteria (Fox et al., 1998; Orsi and Zambrini, 2017). Knowing that the highest level of abundance within the *lactobacillus* grouping is because of the use of *Lb. helveticus* as an ALAB means that the maximum relative abundance possible for a NSLAB in the lactobacilli group is 4.5%. Although, *Lb. helveticus* is used as an ALAB in Cheese I, it cannot be stated with certainty that this is the same for Cheese F because the SLAB and ALAB used to manufacture all cheeses other than Sample I are unknown in this study.

Deblur Filtered > 0.01% Relative Abundance

Using Deblur to denoise the sequence data resulted in 74 total ASVs, of which only 42% had a relative abundance greater than 0.01% (see Appendix G). The 31 ASVs with relative abundances greater than 0.01% make up 99.95 to 99.99 % total abundance across the samples (Table 4). After filtering out ASVs with relative abundances less than 0.01% there were only 5 unique ASVs for *Lc. lactis*. Of these ASVs, only one ASV had a relative abundance greater than 0.09% (Table 5). This single dominant ASV was identified as a strain of *Lc. lactis* subsp. *lactis*. There were only 3 unique ASVs identified as *St. thermophilus* identified. Two had abundances greater than 1% (Table 6).

Based on both the microbiological and biochemical characteristics of *Lc. lactis* is a preferred organism for manufacturing aged Cheddar cheese and is widely used in commercial starter cultures (Vedamuthu et al., 1966). A single strain of *Lc. lactis* dominating across all samples manufactured in various regions and by various manufactures is not a realistic

Table 4. Relative abundances greater than 0.01% of amplicon sequence variants after denoising using Deblur identified as starter culture (*Lactococcus lactis* or *Streptococcus thermophilus*), adjunct and nonstarter lactobacilli and other species in aged cheeses color coded from lowest abundance (yellow) to greatest abundance (red) with black representing abundances less than 0.01%

	Lactococcus lactis	*Streptococcus thermophilus*	Lactobacilli	Other species.
Cheese[1]	----------(%)----------			
A	99.78	0.04	0.16	0.00
B	46.81	52.08	0.03	0.95
C	99.47	0.08	0.44	0.01
D	18.11	80.83	0.53	0.50
E	32.44	67.46	0.05	0.02
F	91.18	3.63	5.08	0.07
G	0.18	98.87	0.94	0.00
H	14.38	85.08	0.46	0.06
I	88.55	0.09	11.35	
J	22.61	76.07	0.98	0.28
K	36.70	63.03		0.25
L	94.44	0.07	5.43	0.02
M	99.59	0.04	0.34	
N	99.85	0.05	0.08	0.01
O	99.91	0.06		

[1]Cheeses as described in Table 1.

Table 5. Relative abundances greater than 0.01% for the 5 unique amplicon sequence variants (ASVs) obtained after denoising using Deblur, all identified as being *Lactococcus lactis* subsp. *lactis*. Color coded as lowest abundance (yellow) to greatest abundance (red) and all cells less than 0.01% abundance are blacked out.

	***Lc. lactis* ASVs**				
Cheese[1]	**1**	**2**	**3**	**4**	**5**
	----------(%)----------				
A	99.78				
B	46.80		0.09	0.02	0.01
C	99.47				
D	18.11				
E	32.43			0.01	0.01
F	91.18				
G	0.18				
H	14.38				
I	88.55				
J	22.61			0.01	0.01
K	36.68			0.02	
L	94.43	0.01			
M	99.59	0.01			
N	99.85				
O	99.91				

[1]Cheeses as described in Table 1.

Table 6. Relative abundances greater than 0.01% for the 3 unique amplicon sequence variants (ASVs) obtained after denoising using Deblur, all identified as being *Streptococcus thermophilus*. Color coded as lowest abundance (yellow) to greatest abundance (red) and all cells less than 0.01% abundance are blacked out.

	Streptococcus thermophilus ASVs		
Cheese[1]	1	2	3
	----------%----------		
A			0.04
B			52.08
C			0.08
D			80.83
E			67.46
F		0.97	2.67
G			98.87
H			85.08
I			0.09
J		1.70	74.38
K	0.13	20.32	42.59
L			0.07
M			0.04
N			0.05
O			0.06

[1]Cheeses as described in Table 1.

representation of what the actual microbiomes of the cheese samples would be. This would mean that all of the cheeses used the same single strain of *Lc. lactis* in their starter culture. For Cheddar cheese manufacture, companies use defined or mixed starters composed of a blend of multiple *Lc. lactis* strains that are ideally phage-unrelated (Bissonnette et al., 1999). It is known that in Cheese I, the starter culture consisted of mixture of 6 different stains of *Lc. lactis* including both *lactis* and *cremoris* subspecies.

DADA2

Using DADA2, there were 247 unique ASVs identified across the 15 different cheeses (see Appendices H and I). DADA2 had 203 DNA sequences that were only identified by using its algorithm and shared 44 DNA sequences with Deblur. These 247 unique ASVs were then split into 5 different classifications; *Lc. lactis*, *St. thermophilus*, lactobacilli, and other species (Table 7). Of the 247 unique ASVs, 38% were identified as *Lc. lactis* and *St. thermophilus*. *Lactococcus lactis* ranging up to 99.25% abundance and *St. thermophilus* ranged up to 96%. Lactobacilli had 38 unique ASVs making up 15% of the total number ASVs present with an average of 0.04% relative abundance within the samples.

The lactobacilli group had ASV relative abundances ranging between 0.0003% to 10.029%. The highest relative abundance was observed in Cheese I for the OUT group of *Lb. helveticus* (Table 8). This difference in the range of observed relative abundance for lactobacilli when compared to the range observed when using Deblur comes from the difference in total ASVs identified.

DADA2 Filtered >0.01% Relative Abundance

Using DADA2 to denoise the sequence data resulted in 247 total ASVs. Of these, only 43% have a relative abundance greater than 0.01%. After filtering out any ASV with relative

Table 7. Relative abundance of amplicon sequence variants after denoising using DADA2 identified as starter culture (*Lactococcus lactis* or *Streptococcus thermophilus*), adjunct and nonstarter lactobacilli and other species in aged cheeses color coded from lowest abundance (yellow) to greatest abundance (red) with black representing none detected.

Cheese[1]	*Lactococcus lactis*	*Streptococcus thermophilus*	Lactobacilli	Other species
	----------%----------			
A	99.6056	0.1171	0.2768	0.0004
B	38.8660	60.9461	0.0251	0.1628
C	99.3002	0.1851	0.5063	0.0084
D	17.1963	82.2402	0.4906	0.0729
E	30.7423	69.1828	0.0523	0.0227
F	88.6470	5.9894	4.3430	0.4658
G	0.3474	98.7812	0.8698	0.0016
H	13.8005	85.5437	0.5628	0.0931
I	89.5246	0.1797	10.2522	0.0435
J	15.4018	83.9265	0.6356	0.0314
K	25.4110	74.5863		0.0027
L	94.0464	0.2289	5.6646	0.0609
M	99.4637	0.1513	0.3838	0.0012
N	99.7896	0.1247	0.0736	0.0114
O	99.8047	0.1640	0.0245	0.0067

[1]Cheeses as described in Table 1.

Table 8. Relative abundance of best match of operational taxonomic unit of amplicon sequence variants after denoising using DADA2 sequenced from aged cheese grouped as various lactobacilli species and color coded from lowest abundance (yellow) to greatest abundance (orange) with black representing none detected.

	Lactobacilli OTU Identifications														
Cheese[1]	Lacticaseibacillus casei	Lacticaseibacillus paracasei	Lacticaseibacillus rhamnosus	Lactiplantibacillus plantarum	Lactobacillus delbrueckii	Lactobacillus helveticus	Lactobacillus hokkaidonensis	Latilactobacillus curvatus	Lentilactobacillus buchneri	Lentilactobacillus farraginis	Lentilactobacillus kefiri	Lentilactobacillus kisonensis	Loigolactobacillus coryniformis	Paucilactobacillus wasatchensis	Secundilactobacillus malefermentans
	----------%----------														
A		0.1729	0.1033			0.0007									
B					0.0251										
C			0.1992						0.0398		0.1214	0.0013		0.1446	
D					0.0017	0.4880								0.0009	
E						0.0515								0.0005	
F		0.1281				4.1708								0.0441	0.5548
G						0.8698									
H		0.4769	0.0189			0.0623								0.0047	
I						10.2505	0.0008			0.0008					
J	0.0004					0.6349									0.0048
K															
L		0.1096			4.3214	0.8836		0.3331			0.0162				
M				0.0006	0.3174	0.0297		0.0050			0.0311				
N								0.0731							0.0007
O								0.0003					0.0242		

[1]Cheeses as described in Table 1.

abundances less than 0.01% only 107 ASVs remained (see Appendicies K and L) making up 99.42 to 99.99% total relative abundance across samples (Table 9). After filtering out ASVs with less than 0.01% there were 34 unique ASVs for *Lc. lactis*, four had abundances greater than 1% (Table 10). Similarly, there 35 unique ASVs for *St. thermophilus*, two had abundances greater than 1% (Table 11).

Both the greater number of ASVs and there spread across the different Cheddar chesses demonstrates a more realistic representation of what would be expected in a commercial manufacture. To manufacture Cheddar cheese, an inoculation of carefully selected SLAB is required. The SLAB drives the fermentation process. When phage sensitive cultures are present, the consequent lysis of sensitive cells can delay or halt milk fermentation leading to low-quality products or the need to discard the inoculated milk entirely (Garneau and Moineau, 2011). The risk of LAB phages has led the industry to select multiple SLAB strains for use in the starter inoculum that are phage-unrelated as a way of preventing a fermentation from failing. Therefore, DADA2 provides a better representation of the SLAB Cheddar cheese samples and will be used for further comparisons. Similar data obtained by Deblur will be given in the Appendix.

Regional Comparison of Cheeses

The Cheddar samples clustered into two groups (Figure 2). The first grouping included cheeses manufactured in Ireland, Australia, California, and Vermont. The second grouping included cheese from Ireland, Utah, California, Oregon, New York, and Wisconsin. Initially it was predicted that the cheeses were going to cluster according to their regions of manufacture. However, this was not entirely the case. Cheeses clustered by the most abundant SLAB with the first group dominated by *St. thermophilus* and the second group dominated by *Lc. lactis*

Table 9. Relative abundances greater than 0.01% of amplicon sequence variants after denoising using DADA2 identified as starter culture (*Lactococcus lactis* or *Streptococcus thermophilus*), adjunct and nonstarter lactobacilli and other species in aged cheeses color coded from lowest abundance (yellow) to greatest abundance (red) with black representing abundances less than 0.01%.

Cheese[1]	*Lactococcus lactis*	*Streptococcus thermophilus*	Lactobacilli	Other species
	----------%----------			
A	99.59	0.12	0.28	
B	38.86	60.94	0.03	0.16
C	99.29	0.19	0.50	0.01
D	17.19	82.24	0.49	0.05
E	30.74	69.18	0.05	0.00
F	88.63	5.99	4.34	0.46
G	0.33	98.77	0.87	
H	13.79	85.54	0.56	0.07
I	89.51	0.18	10.25	0.04
J	15.39	83.92	0.63	0.03
K	25.41	74.58		
L	94.04	0.23	5.66	0.02
M	99.45	0.15	0.38	
N	99.78	0.12	0.07	0.00
O	99.79	0.16	0.02	

[1]Cheeses as described in Table 1.

Table 10. Relative abundances greater than 0.01% for the 34 unique amplicon sequence variants (ASV) after denoising using DADA2 all identified as being *Lactococcus lactis*. Color coded as lowest abundance (yellow) to greatest abundance (orange) with ASV <0.01% or not detected as black (continued on page 36).

	Lactococcus lactis ASVs																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Cheese[1]	---------- % ----------																
A	0.11		99.25				0.20										
B			38.86														
C			98.29				0.24	0.17	0.13			0.15	0.15	0.13			
D			17.17														
E			30.71														
F	0.12		88.02					0.16		0.11						0.12	0.12
G			0.29														
H			13.74														
I			83.41		2.74	1.11	0.06	0.14	0.12	0.11	0.33	0.12	0.11	0.12		0.15	0.13
J			15.39														
K		0.01	25.37														
L			93.54				0.09	0.16		0.12							
M			98.92				0.19		0.11	0.11							
N			87.84	11.68											0.11		
O			99.46						0.11						0.14		

[1]Cheeses as described in Table 1.

	Lactococcus lactis ASVs																
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Cheese[1]	----------%----------																
A							0.02									0.11	
B																	
C							0.02										
D										0.01					0.01		
E									0.03								
F																0.12	
G																	
H													0.02				
I	0.15	0.14	0.13	0.12		0.10	0.02	0.04				0.07		0.08			
J																	
K															0.03		0.01
L											0.14						
M					0.13												
N											0.06		0.08				
O												0.08					

[1]Cheeses as described in Table 1.

Table 11. Relative abundances greater than 0.01% for the 35 unique amplicon sequence variants (ASV) after denoising using DADA2 all identified as being *Streptococcus thermophilus*. Color coded as lowest abundance (yellow) to greatest abundance (orange) with ASV <0.01% or not detected as black (continued on page 38).

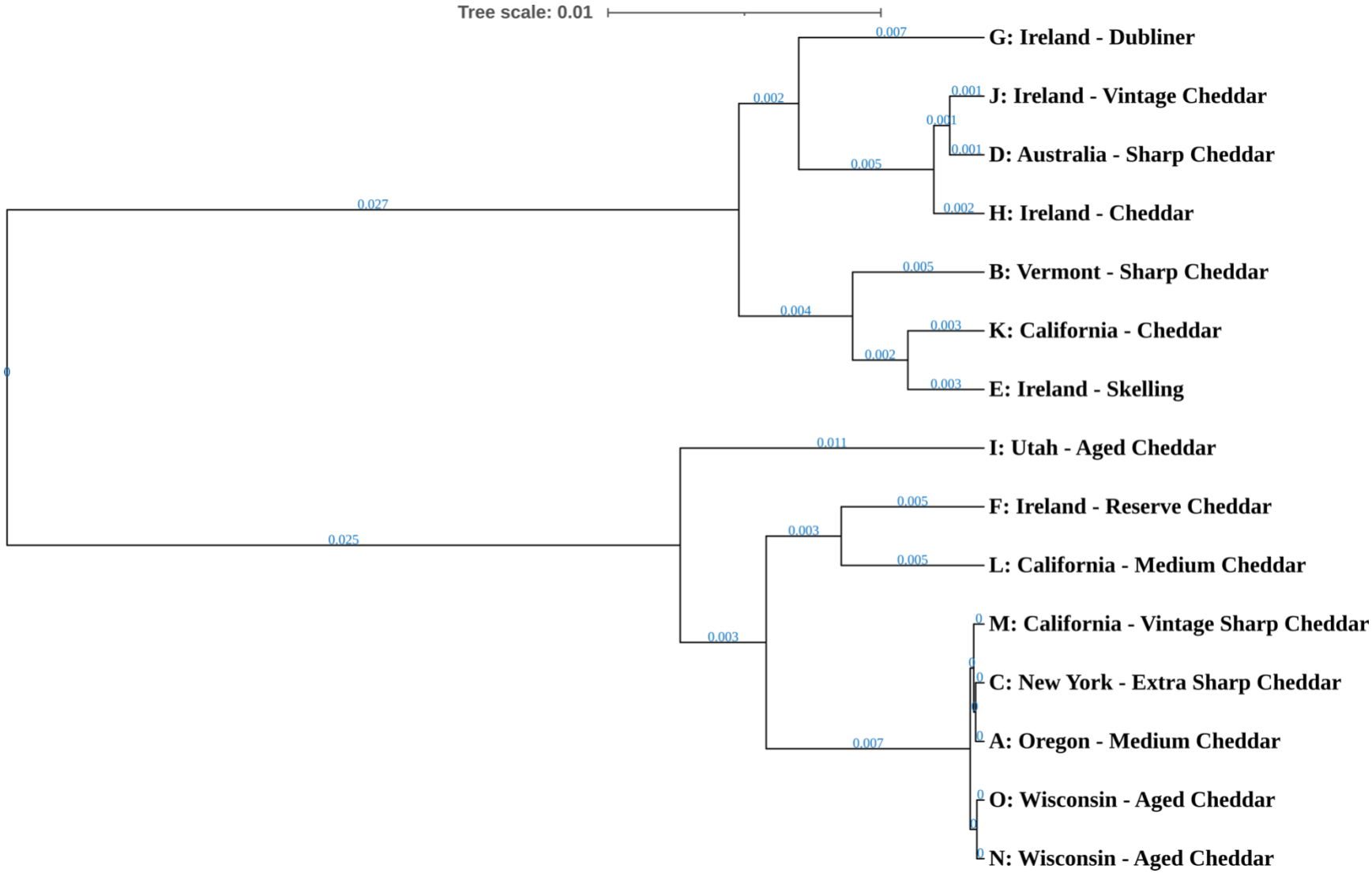
	Streptococcus thermophilus ASVs																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Cheese[1]	----------%----------																
A	0.12																
B	46.55	13.73	0.66														
C	0.19																
D	81.05		0.33	0.29	0.03		0.10	0.15	0.17		0.09				0.02		
E	67.78		0.34			0.10	0.14	0.15		0.11	0.07	0.08		0.09	0.23		
F	3.38	2.57															
G	96.29		0.38	0.19		0.14	0.08	0.13	0.12	0.14	0.08	0.11	0.17	0.13	0.02	0.15	
H	82.16		0.21	0.21	0.68	0.14	0.07		0.11	0.14	0.06	0.11	0.14			0.13	0.27
I	0.18																
J	55.30	27.69	0.24	0.20		0.07	0.10		0.07		0.06			0.06			
K	40.01	33.69	0.08			0.06	0.05	0.13				0.04		0.06	0.11		
L	0.23																
M	0.09		0.06														
N	0.12																
O	0.16																

[1]Cheeses as described in Table 1.

	Streptococcus thermophilus ASVs																	
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Cheese[1]	----------%----------																	
A																		
B																		
C																		
D																		
E													0.09					
F																0.04		
G	0.14	0.14		0.11	0.08			0.14						0.05				
H	0.12	0.12	0.12	0.11	0.05	0.11	0.13		0.12	0.11	0.10						0.04	
I																		
J			0.07		0.06													
K			0.06			0.04						0.13			0.06			0.04
L																		
M																		
N																		
O																		

[1]Cheeses as described in Table 1.

Figure 2. Phylogenetic clustering of samples based on their microbiomes displaying the branch lengths in blue.



Letunic and Bork, 2019 *Nucleic Acids Res* **doi: 10.1093/nar/gkz239**

(Table10). In the first cluster group, Cheese K and B, both from the United States, had the highest abundance of *Lc. lactis* of the group. In the second cluster group Cheese F from Ireland had the highest abundance of *St. thermophilus* of the group.

The biggest influence on how cheeses clustered in this study was the dominant SLAB used in manufacture. In a previous study, commercial Cheddar cheeses from various regions across the United States were evaluated for regional flavor differences (Drake et al., 2008). The evidence in that study suggested that region of manufacture and general make procedure did not have as large of an influence as the facility practices as a whole. Further, the starter culture had a bigger influence on the overall microbiome of aged Cheddar cheese not the region of manufacture.

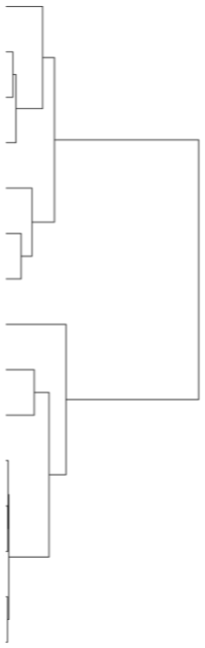
Environmental Bacteria

During the cheese making process environmental bacteria may be introduced through the milk, processing equipment, or by the cheese handlers. The environmental bacteria found in the category other species (see Appendices E and I) were not exclusionary strains of bacteria meaning that although some of the bacteria present were not ideal for cheese making, they were not the types of bacteria that pose a health risk and make the cheese unfit for sale. Environmental bacteria including both other species and NSLAB have the potential to cause cheese defects, poor flavor development, and affect shelf life. The amount and types of environmental bacteria found in cheeses may indicate how clean the cheese making process is for any given cheese making facility.

An example of an environmental contaminant introduced by milk during manufacture is *Streptococcus uberis*, which was found in two of the cheeses studied (see Appendices E and I).

Table 12. Relative abundances greater than 0.01% of amplicon sequence variants after denoising using DADA2 identified as starter culture (*Lactococcus lactis* or *Streptococcus thermophilus*), adjunct and nonstarter lactobacilli and other species in aged cheeses color coded from lowest abundance (yellow) to greatest abundance (red) with black representing abundances less than 0.01%.

Cheese[1]	Cluster	Region	*Lactococcus lactis*	*Streptococcus thermophilus*	Lactobacilli	Other species
			----------%----------			
G	**1**	Ireland	0.33	98.77	0.87	
J	**1**	Ireland	15.39	83.92	0.63	0.03
D	**1**	Australia	17.19	82.24	0.49	0.05
H	**1**	Ireland	13.79	85.54	0.56	0.07
B	**1**	Vermont	38.86	60.94	0.03	0.16
K	**1**	California	25.41	74.58		
E	**1**	Ireland	30.74	69.18	0.05	0.00
I	**2**	Utah	89.51	0.18	10.25	0.04
F	**2**	Ireland	88.63	5.99	4.34	0.46
L	**2**	California	94.04	0.23	5.66	0.02
M	**2**	California	99.45	0.15	0.38	
C	**2**	New York	99.29	0.19	0.50	0.01
A	**2**	Oregon	99.59	0.12	0.28	
O	**2**	Wisconsin	99.79	0.16	0.02	
N	**2**	Wisconsin	99.78	0.12	0.07	



[1]Cheeses as described in Table 1.

Streptococcus uberis is an environmental bacterium responsible for a high percentage of mastitis in dairy cattle (Di Domenico et al., 2015). Another example of an environmental contaminant is *Thermus thermophilus* (see Appendices E and I) a carotenoid-producing thermophile which can cause a pink discoloration defect in cheese (Quigely et al., 2016). The most probable source of *T. thermophilus* contamination is from hot water used during cheese making. *Staphylococcus aureus* was found in 2 cheese samples (see Appendix I) which is found on the skin indicating it was introduced by cheese handlers and possible poor hygiene practice in the manufacturing facilities.

CONCLUSIONS

Originally, it was hypothesized that when testing Cheddar cheeses from various regions of manufacture samples would cluster according to the region of manufacture when analyzed using NGS technology. However, when DNA was extracted from the 16S rRNA hypervariable V3-V4 region the most important factor impacting how samples cluster was not region of manufacture but rather the dominant SLAB population. Cheeses tested were either dominated by *Lc. lactis* or *St. thermophilus* which determined the clustering group.

When analyzing the NGS data using the Qiime 2 bioinformatics platform we used two different denoising algorithms both accessible through the platform. The two denoising methods were used for comparison of variant identification. After initial sequencing using Illumina MiSeq there were 15,193,666 raw sequence reads with an average of 1,012,911 raw reads per sample. When using Deblur to denoise sequence data 40.5% of reads were retained. When DADA2 was used to denoise the same sequence data 76% of reads were retained. Using DADA2 resulted in the retention of about 1.9 times more reads in comparison to Deblur. Greater number of reads allows less abundant ASVs to be detected. The ability to detect low-abundance bacteria is of particular importance especially in regard to low-abundance spoilage bacteria as demonstrated by Xue et al., 2021.

When comparing the two denoising methods it was important to compare the ASVs generated by each denoising algorithm. Deblur resulted in 74 unique ASVs and DADA2 resulted in 247 ASVs. The defining characteristic of the first cluster was dominance of *St. thermophilus* Deblur only identified 3 ASVs with greater than 0.01% relative abundance and DADA2 identified 35 ASVs. *Lactococcus lactis* dominance was the defining characteristic of the second cluster grouping and when using Deblur for analysis only 5 ASVs were identified for the OTU

classification of *Lc. lactis* compared to the 34 ASVs identified by DADA2. The starter composition for the cheese made by Utah State University known and uses 6 different *Lc. lactis* strains when making their cheese. This would mean for the results generated by using Deblur to be accurate in resolution of the data there was only 1 *Lc. lactis* starter used in all of the different cheeses tested. Leading to the conclusion that when doing a Cheddar microbiome study DADA2 should be used for denoising sequence data as a way of identifying more ASVs and as a means of creating a more accurate representation of the actual cheese composition.

It has been shown that SLAB and NSLAB dynamically change according to the modification of environmental conditions during cheese gaining (Stefanovic et al., 2018). In aging Cheddar cheese studies using culturing techniques it has been observed that over the aging process SLAB lose viability rapidly in the initial phases of ripening and NSLAB slowly increase in numbers to become the dominant microbiota in cheese (Gatti et al., 2014). The results of this study showed that SLAB remains the most abundant after being aged for over 6 months suggesting that SLAB remain present in a viable but non-culturable state (VBNC).

Microorganisms generally live in a feast or famine existence in their natural environments due to the variability of nutrient and energy availabilities. During cheese ripening conditions SLAB encounter famine type conditions leading to a severe reduction of the growth rate. Lactic acid bacteria are able to survive these periods of extremely slow growth, while still forming aroma compounds (Mastrigt et al., 2018). Using NGS as a way of detecting VBNC bacteria in food products would be a powerful tool to utilize in the dairy industry allowing for the detection of foodborne pathogenic bacteria. It has been suggested that 80% of foodborne illness can be attributed to VBNC microorganisms (Ünlu, 2021).

When comparting the microbiota community differences between the cheeses the differences observed were mainly differences in relative abundances of various SLAB. The first cluster grouping contained cheeses from Ireland, Australia, California, and Vermont with the OTU identifications of highest relative abundance being *St. thermophilus*. In the second cluster grouping there were cheeses from Ireland, Utah, California, Oregon, New York, and Wisconsin with OTU identifications of highest relative abundance being *Lc. lactis.* Although, of the cheeses in the first grouping both the cheeses from the United States had the highest relative abundance of *L. lactis* across the group and the Irish cheese from the second grouping had the highest relative abundance of *St. thermophilus* among the cheeses in its grouping. Demonstrating that there might be regional preferences of SLAB.

FUTURE RESEARCH

To build upon this project a follow up study could be done by choosing 6 different defined starter strains for which there is a genome sequence. Then using these 6 strains different cheese makes could be done with defined ratios to be used during manufacture sampling at different storage points. The previously outlined protocol could then be followed for DNA extraction and analysis. Then after bioinformatic analysis the method could be evaluated by checking if 6 unique ASVs were identified and what relative abundances were observed as well as evaluating any changes in the relative abundances observed over the Cheddar aging process.

Based on the results of this study either everyone is using the same SLAB strains to make cheese, or these are the only strains surviving. Another follow up study that could be done separately or in conjunction with the proposed study previously mentioned. Where cheeses are tested using next generation sequencing as well through use of traditional plating techniques for both SLAB and NSLAB.

REFERENCES

Amir, A., D. Mcdonald, J.A. Navas-Molina, E. Kopylova, J.T. Morton, Z.Z. Xu, E.P. Kightley, L.R. Thompson, E.R. Hyde, A. Gonzalez, and R. Knight. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2:e00191-16. https://doi:10.1128/msystems.00191-16.

Anaconda Software Distribution. 2020. Anaconda documentation. Anaconda Inc. Retrieved from https://docs.anaconda.com. Accessed Sept. 23, 2020.

Banks, J.M., and A.G. Williams. 2004. The role of the nonstarter lactic acid bacteria in Cheddar cheese ripening. Int J Dairy Technol. 57:145–152. doi:10.1111/j.1471-0307.2004.00150.x.

Bella, J.M.D., Y. Bao, G.B. Gloor, J.P. Burton, and G. Reid. 2013. High throughput sequencing methods and analysis for microbiome research. J Microbiol Methods 95:401–414. doi:10.1016/j.mimet.2013.08.011.

Bissonnette, F., S. Labrie, H. Deveau, M. Lamoureux, and S. Moineau. 2000. Characterization of mesophilic mixed starter cultures used for the manufacture of aged cheddar cheese. J Dairy Sci. 83:620–627. https://doi:10.3168/jds.s0022-0302(00)74921-6.

Bolyen, E., J.R. Rideout, M.R. Dillon, N.A. Bokulich, C. Abnet, G.A. Al-Ghalith, H. Alexander, E.J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J.E. Bisanz, K. Bittinger, A. Brejnrod, C.J. Brislawn, C.T. Brown, B.J. Callahan, A.M. Caraballo-Rodríguez, J. Chase, E. Cope, R.D. Silva, P.C. Dorrestein, G.M. Douglas, D.M. Durall, C. Duvallet, C.F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J.M. Gauglitz, D.L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. Huttley, S. Janssen, A.K. Jarmusch, L. Jiang, B. Kaehler, K.B. Kang, C.R. Keefe, P. Keim, S.T. Kelley, D.

Knights, I. Koester, T. Kosciolek, J. Kreps, M.G. Langille, J. Lee, R. Ley, Y.-X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B.D. Martin, D. Mcdonald, L.J. Mciver, A.V. Melnik, J.L. Metcalf, S.C. Morgan, J. Morton, A.T. Naimey, J.A. Navas-Molina, L.F. Nothias, S.B. Orchanian, T. Pearson, S.L. Peoples, D. Petras, M.L. Preuss, E. Pruesse, L.B. Rasmussen, A. Rivers, I.M.S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S.J. Song, J.R. Spear, A.D. Swafford, L.R. Thompson, P.J. Torres, P. Trinh, A. Tripathi, P.J. Turnbaugh, S. Ul-Hasan, J.J.V.D. Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M.V. Hippel, W. Walters, et al., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 37:852–857. doi: 10.1038/s41587-019-0209-9

Broadbent, J., M. Budinich, and J. Steele. 2016. Cheese: NSLAB. Reference Module in Food Sci. doi:10.1016/b978-0-08-100596-5.00676-4.

Broadbent, J., C. Brighton, D. McMahon, N. Farkye, M. Johnson, and J. Steele. 2013. Microbiology of cheddar cheese made with different fat contents using a *Lactococcus lactis* single-strain starter. J Dairy Sci. 96:4212–4222. doi:10.3168/jds.2012-6443.

Broadbent, J., M. Budinich, and J. Steele. 2011. Cheese | non-starter lactic acid bacteria. Ency. Dairy Sciences. 639–644. doi:10.1016/b978-0-12-374407-4.00077-7.

Broadbent, J., and Steele, J. 2005. Cheese flavor and the genomics of lactic acid bacteria. ASM News. 71. 121-128.

Bukin, Y.S., Y.P. Galachyants, I.V. Morozov, S.V. Bukin, A.S. Zakharenko, and T.I. Zemskaya. 2019. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci Data. 6. doi:10.1038/sdata.2019.7.

Callahan, B.J., P.J. Mcmurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, and S.P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 13:581–583. doi:10.1038/nmeth.3869.

Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons, C.A. Lozupone, P.J. Turnbaugh, N. Fierer, and R. Knight. 2010. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A. 108:4516–4522. doi:10.1073/pnas.1000080107.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of highthroughput community sequencing data. Nat. Methods 7, 335–336.

Chen, W., and F. Hang. 2019. Lactic acid bacteria starter. Lactic Acid Bacteria. 93–143. doi:10.1007/978-981-13-7283-4_5.

Choi, J., S.I. Lee, B. Rackerby, R. Frojen, L. Goddik, S.-D. Ha, and S.H. Park. 2020. Assessment of overall microbial community shift during Cheddar cheese production from raw milk to aging. Appl Microbiol and Biotechnol. 104:6249–6260. doi:10.1007/s00253-020-10651-7.

Choi, J., S.I. Lee, B. Rackerby, L. Goddik, R. Frojen, S.-D. Ha, J.H. Kim, and S.H. Park. 2020. Microbial communities of a variety of cheeses and comparison between core and rind region of cheeses. J Dairy Sci. 103:4026–4042. doi:10.3168/jds.2019-17455.

Croissant, A. E., S. P. Washburn, L. L. Dean, and M. A. Drake. 2007. Chemical properties and consumer perception of fluid milk from conventional and pasture-based production systems. J Dairy Sci. 90:4942–4953. https:// doi .org/ 10 .3168/ jds .2007 -0456.

Dawson, D.J., and J.T. Feagan. 1957. 673. Bacteriology of Cheddar cheese: a study of starter

organisms in manufacture and maturing. J Dairy Res. 24:210–224. doi:10.1017/s002202990000875x.

Delbès, C, L. Ali-Mandjee, and M.-C. Montel. 2007. Monitoring bacterial communities in raw milk and cheese by culture-dependent and -independent 16S rRNA gene-based analyses. Appl Environ Microbiol. 73:1882–1891. doi:10.1128/aem.01716-06.

Di Domenico, E.G., L. Toma, G. Prignano, L. Pelagalli, A. Police, C. Cavallotti, R. Torelli, M. Sanguinetti, and F. Ensoli. 2015. Misidentification of *Streptococcus uberis* as a human pathogen: a case report and literature review. Int J of Infect Dis. 33:79–81. doi:10.1016/j.ijid.2015.01.002.

Drake, M., M. Yates, and P. Gerard. 2008. Determination of Regional Flavor Differences in U.S. Cheddar Cheeses Aged for 6 Mo or Longer. J Food Sci. 73:199–207. doi:10.1111/j.1750-3841.2008.00750.x.

Drake, M.A., M.D. Yates, P.D. Gerard, C.M. Delahunty, E.M. Sheehan, R.P. Turnbull, and T.M. Dodds. 2005. Comparison of differences between lexicons for descriptive analysis of Cheddar cheese flavour in Ireland, New Zealand, and the United States of America. Int Dairy J. 473–483. doi:10.1016/j.idairyj.2004.09.006.

Drake, S. L., K. Lopetcharat, S. Clark, H.S. Kwak, S.Y. Lee, and M.A. Drake. 2009. Mapping differences in consumer perception of sharp Cheddar cheese in the United States. J of Food Sci. 74:276–285. doi:10.1111/j.1750-3841.2009.01219.x.

Duarte, V. D. S., M. Carlot, S. Pakroo, A. Tarrah, A. Lombardi, H. Santiago, V. Corich, and A. Giacomini. 2020. Comparative evaluation of cheese whey microbial composition from four Italian cheese factories by viable counts and 16S rRNA gene amplicon sequencing. Int Dairy J. 104:104656. doi:10.1016/j.idairyj.2020.104656.

Edgar, R. C. 2017. Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. PeerJ. 5:e3889. doi:10.7717/peerj.3889.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation. 61:1–10. doi:10.1016/0006-3207(92)91201-3.

Fitzsimons, N.A., T.M. Cogan, S. Condon, and T. Beresford. 1999. Phenotypic and genotypic characterization of non-starter lactic acid bacteria in mature Cheddar cheese. Appl and Environ Microbiol. 65:3418–3426. doi:10.1128/aem.65.8.3418-3426.1999.

Fox, P.F., P.L.H. McSweeney, C.M. and Lynch. 1998. Significance of non- starter lactic acid bacteria in Cheddar cheese. A J Dairy Technol. 53:83- 89.

Fox, P.F., T.M. Cogan, and P.L.H. McSweeney. 2017. Microbiology of cheese ripening. In Fundamentals of Cheese Sci. Springer, New York, NY. 206–235.

Ganesan, B., M.R. Stuart, and B.C. Weimer. 2007. Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis*. Appl Environ Microbiol. 73:2498–2512. doi:10.1128/AEM.01832-06.

Garneau, J.E., and S. Moineau. 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microb Cell Fact. 10. doi:10.1186/1475-2859-10-s1-s20.

Gatti, M., B. Bottari, C. Lazzi, E. Neviani, and G. Mucchetti. 2014. Invited review: Microbial evolution in raw-milk, long-ripened cheeses produced using undefined natural whey starters. J. Dairy Sci. 97:573–591. doi:10.3168/jds.2013-7187.

Jonnala, B.R.Y., P.L.H. Mcsweeney, J.J. Sheehan, and P.D. Cotter. 2018. Sequencing of the cheese microbiome and its relevance to industry. Front Microbiol. 9. doi:10.3389/fmicb.2018.01020.

Katoh, K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast

Fourier transform. Nucleic Acids Res. 30:3059–3066. doi:10.1093/nar/gkf436.

Khanal, R. C., T. R. Dhiman, A. L. Ure, C. P. Brennand, R. L. Boman, and D. J. McMahon. 2005. Consumer acceptability of conjugated linoleic acid-enriched milk and Cheddar cheese from cows grazing on pasture. J Dairy Sci. 88:1837–1847. https:// doi .org/ 10 .3168/ jds .S0022 -0302(05)72858 -7.

Lawrence, R., H. Heap, and J. Gilles. 1984. A controlled approach to cheese technology. J Dairy Science. 67:1632–1645. doi:10.3168/jds.s0022-0302(84)81486-1.

Lozupone, C.A., M. Hamady, S.T. Kelley, and R. Knight. 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol. 73:1576–1585. doi:10.1128/aem.01996-06.

Lozupone, C., and R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 71:8228–8235. doi:10.1128/aem.71.12.8228-8235.2005.

Martin-Platero, A.M., M. Maqueda, E. Valdiva, J. Purswani, and M. Martinez-Bueno. 2009. Polyphasic study of microbial communities of two Spanish farmhouse goats' milk cheeses from Sierra de Aracena. Food Microbiol. 26:294–304. doi:10.1016/j.fm.2008.12.004.

Mastrigt, O.V., T. Abee, S.K. Lillevang, and E.J. Smid. 2018. Quantitative physiology and aroma formation of a dairy *Lactococcus lactis* at near-zero growth rates. Food Microbiol. 73:216–226. doi:10.1016/j.fm.2018.01.027.

Mataragas, M., V. Alessandria, I. Ferrocino, K. Rantsiou, and L. Cocolin. 2018. A bioinformatics pipeline integrating predictive metagenomics profiling for the analysis of 16S rDNA/rRNA sequencing data originated from foods. Food Microbiol. 76:279–286. doi:10.1016/j.fm.2018.05.009.

McSweeney P. L. H., Walsh E. M., Fox P. F., Cogan T. M., Drinan F. D. and Castelo-Gopnzalez M. 1994 A procedure for the manufacture of Cheddar cheese under controlled bacteriological conditions and the effect of adjunct lactobacilli on cheese quality. Irish J Agri Food Res 33 183–192.

Muyzer, G. 1999. DGGE/TGGE a method for identifying genes from natural ecosystems. Curr Opin Microbiol. 2:317–322. doi:10.1016/s1369-5274(99)80055-1.

Nearing, J.T., G.M. Douglas, A.M. Comeau, and M.G.I. Langille. 2018. Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ. doi:10.7287/peerj.5364.

Orsi, C., and A.V. Zambrini. 2017. Starter and ancillary cultures. Advan Dairy Prod. 58–74. http://dx.doi.org/10.1002/9781118906460.ch1d.

Parente, E., and T.M. Cogan. 2004. Starter cultures: general aspects. In Cheese: chem, phys, microbiol. Elsevier Appl Sci, London u.a.

Price, M.N., P.S. Dehal, and A.P. Arkin. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE. 5. doi:10.1371/journal.pone.00 09490.

Prodan, A., V. Tremaroli, H. Brolin, A.H. Zwinderman, M. Nieuwdorp, and E. Levin. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. Plos One. 15. doi:10.1371/journal.pone.0227434.

Quigley, L., D.J. O'Sullivan, D. Daly, O. O'Sullivan, Z. Burdikova, R. Vana, T.P. Beresford, R.P. Ross, G.F. Fitzgerald, P.L. McSweeney, L. Giblin, J.J. Sheehan, and P.D. Cotter. 2016. *Thermus* and the pink discoloration defect in cheese. mSystems. 1. doi:10.1128/msystems.00023-16.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F.O. Glöckner.

2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41. doi:10.1093/nar/gks1219.

Rajeev, M., T. Sushmitha, S.R. Toleti, and S.K. Pandian. 2020. Sediment-associated bacterial community and predictive functionalities are influenced by choice of 16S ribosomal RNA hypervariable region(s): An amplicon-based diversity study. Genomics. 112:4968–4979. doi:10.1016/j.ygeno.2020.09.006.

Ramsden, J. 2015. Introduction. In Bioinformatics An Introduction. Springer-Verlag, London. 1–3.

Ruggirello, M., P. Dolci, and L. Cocolin. 2014. Detection and viability of Lactococcus lactis throughout cheese ripening. PLOS One. doi:10.1371/journal.pone.0114280.

Salazar, J.K., C.K. Carstens, P. Ramachandran, A.G. Shazer, S.S. Narula, E. Reed, A. Ottesen, and K.M. Schill. 2018. Metagenomics of pasteurized and unpasteurized gouda cheese using targeted 16S rDNA sequencing. BMC Microbiol. 18. doi:10.1186/s12866-018-1323-4.

Saltaji, S., O. Rué, V. Sopena, S. Sablé, F. Tambadou, S. Didelot, and R. Chevrot. 2020. *Lactococcus lactis* diversity revealed by targeted amplicon sequencing of Purr Gene, metabolic comparisons and antimicrobial properties in an undefined mixed starter culture used for soft-cheese manufacture. Foods. 9:622. doi:10.3390/foods9050622.

Singh, T., M. Drake, and K. Cadwallader. 2003. Flavor of Cheddar cheese: a chemical and sensory perspective. Comp Rev Food Sci Food Saf. 2:166–189. doi:10.1111/j.1541-4337.2003.tb00021.x.

Trépanier, G., R. Simard, and B. Lee. 1991. Lactic acid bacteria relation to accelerated maturation of Cheddar cheese. J Food Sci. 56:1238–1240. doi:10.1111/j.1365-2621.1991.tb04742.x.

Van Mastrigt, O., T. Abee, S.K. Lillevang, and E.J. Smid. 2018. Quantitative physiology and aroma formation of a dairy *Lactococcus lactis* at near-zero growth rates. Food Microbiol. 73:216–226. doi:10.1016/j.fm.2018.01.027.

Ünlü, G. 2021. Viable but nonculturable bacteria can threaten food safety. *IFT*. 75. doi:https://www.ift.org/news-and-publications/food-technology-magazine/issues/2021/march/columns/food-safety-and-quality-viable-but-nonculturable-bacteria-can-threaten-food-safety.

U.S. Dairy Export Council. 2016. 5.1 Cheddar and Colby. In Reference Manual for U.S. Cheese. Arlington, VA. 76–78.

Vedamuthu, E. R., W. E. Sandine, and P. R. Elliker. 1966. Flavor and texture in Cheddar cheese. Role of mixed strain lactic acid starter cultures. J Dairy Sci. 49:144–150.

Widyastuti, Y., R., and A. Febrisiantosa. 2014. The Role of Lactic Acid Bacteria in Milk Fermentation. Food Nutr Sci. 05:435–442. doi:10.4236/fns.2014.54051.

Xue, Z., J.T. Brooks, Z. Quart, E.T. Stevens, M.E. Kable, J. Heidenreich, J. McLeod, and M.L. Marco. 2021. Microbiota Assessments for the Identification and Confirmation of Slit Defect-Causing Bacteria in Milk and Cheddar Cheese. mSystems. 6. doi:10.1128/msystems.01114-20.

Zhang, J., X. Ding, R. Guan, C. Zhu, C. Xu, B. Zhu, H. Zhang, Z. Xiong, Y. Xue, J. Tu, and Z. Lu. 2018. Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. Sci Total Envr. 618:1254–1267. doi:10.1016/j.scitotenv.2017.09.228.

APPENDICES

APPENDIX A. MANUFACTURER (QIAGEN) INSTRUCTIONS USED FOR DNA EXTRACTION

1. Add 0.25 g of soil sample to the PowerBead Tube provided. Gently vortex to mix.
2. Add 60 µl of Solution C1 and invert several times or vortex briefly.

   Note: Solution C1 may be added to the PowerBead tube before adding soil sample
3. Secure PowerBead Tubes horizontally using a Vortex Adapter tube holder (cat. no. 13000–V1–24).
4. Vortex at maximum speed for 10 min.

   Note: If using the 24-place Vortex Adapter for more than 12 preps, increase the vortex time by 5–10 min.
5. Centrifuge tubes at 10,000 x *g* for 30 s.
6. Transfer the supernatant to a clean 2 ml collection tube.

   Note: Expect between 400–500 µl of supernatant. Supernatant may still contain some soil particles.
7. Add 250 µl of Solution C2 and vortex for 5 s. Incubate at 2–8°C for 5 min.

Sample to Insight

QIAGEN

Note: You can skip the 5 min incubation. However, if you have already validated the DNeasy PowerSoil extractions with this incubation we recommend you retain the step.

8. Centrifuge the tubes for 1 min at 10,000 x *g*.
9. Avoiding the pellet, transfer up to 600 µl of supernatant to a clean 2 ml collection tube.
10. Add 200 µl of Solution C3 and vortex briefly. Incubate at 2–8°C for 5 min.

    Note: You can skip the 5 min incubation. However, if you have already validated the PowerSoil extractions with this incubation we recommend you retain the step.
11. Centrifuge the tubes for 1 min at 10,000 x *g*.
12. Avoiding the pellet, transfer up to 750 µl of supernatant to a clean 2 ml collection tube.
13. Shake to mix Solution C4 and add 1200 µl to the supernatant. Vortex for 5 s.
14. Load 675 µl onto an MB Spin Column and centrifuge at 10,000 x *g* for 1 min. Discard flow through.

15. Repeat step 14 twice, until all of the sample has been processed.
16. Add 500 µl of Solution C5. Centrifuge for 30 s at 10,000 x *g*.
17. Discard the flow through. Centrifuge again for 1 min at 10,000 x *g*.
18. Carefully place the MB Spin Column into a clean 2 ml collection tube. Avoid splashing any Solution C5 onto the column.
19. Add 100 µl of Solution C6 to the center of the white filter membrane. Alternatively, you can use sterile DNA-Free PCR Grade Water for this step (cat. no. 17000–10).
20. Centrifuge at room temperature for 30 s at 10,000 x *g*. Discard the MB Spin Column. The DNA is now ready for downstream applications.
    Note: Solution C6 is 10 mM Tris-HCl, pH 8.5. We recommend storing DNA frozen (–20° to –80°C) as Solution C6 does not contain EDTA. To concentrate DNA see the Hints & Troubleshooting Guide.

APPENDIX B. UNIQUE BARCODE COMBINATIONS FOR EACH CHEESE SAMPLE

Cheese[1]	Forward Barcode	Reverse Barcode	Forward Sequence	Reverse Sequence
A	U701	U501	TCGCCTTA	TAGATCGC
B	U701	U502	TCGCCTTA	CTCTCTAT
C	U701	U503	TCGCCTTA	TATCCTCT
D	U701	U504	TCGCCTTA	AGAGTAGA
E	U701	U505	TCGCCTTA	GTAAGGAG
F	U702	U501	CTAGTACG	TAGATCGC
G	U702	U502	CTAGTACG	CTCTCTAT
H	U702	U503	CTAGTACG	TATCCTCT
I	U702	U504	CTAGTACG	AGAGTAGA
J	U702	U505	CTAGTACG	GTAAGGAG
K	U704	U501	GCTCAGGA	TAGATCGC
L	U704	U502	GCTCAGGA	CTCTCTAT
M	U704	U503	GCTCAGGA	TATCCTCT
N	U704	U504	GCTCAGGA	AGAGTAGA
O	U704	U505	GCTCAGGA	GTAAGGAG

[1]Cheeses as described in Table 1.

APPENDIX C: QIIME 2 DEBLUR DENOISING COMMAND PIPELINE

A. Data importing

1. Move sequence data folder (Sophie_Sequences) into the directory you will be using in Qiime 2. The sequence data are in individual compressed files (.gz) of forward and reverse sequences.
2. Rename files to include lane number: `A_S1_R1_001.fastq.gz > A_S1_L###_R1_001.fastq.gz` where "`L###`" is a fake lane number for each sample (can all be "`L001`").
3. Import data as type: "cassava-18-paired-end-demultiplexed".

```
qiime tools import \
  --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path Sophie_Sequences \
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path demux-paired-end.qza
```

4. Generate summary of the demultiplexed results.

```
qiime demux summarize \
  --i-data demux-paired-end.qza \
  --o-visualization demux-paired-end.qzv
```

```
#visualizations generated are .qzv files and you can view these
files in Qiime 2
```

5. View generated visualization. The first tab shows the distribution of sequences/samples. The second tab is an interactive graph of quality score as a function of sequence length.

```
qiime tools view demux-paired-end.qzv
```

```
#the interactive graph will show you if your sequence reads need to
be trimmed for analysis
```

B. Sequence quality control and feature table construction.

There are two options for quality control Deblur and DADA2. Deblur will only do denoising and clustering and thus joining paired ends must be done separately using q2-vsearch.

4. Join demultiplexed reads.

```
qiime vsearch join-pairs \
  --i-demultiplexed-seqs demux-paired-end.qza \
  --o-joined-sequences demux-joined.qza
```

5. Create a view of the summary of the joined data with read quality.

```
qiime demux summarize \
  --i-data demux-joined.qza \
  --o-visualization demux-joined.qzv
```

6. View results.

```
qiime tools view demux-joined.qzv
```

By hovering over the read quality graph, determine the maximum sequence length that is associated with 99% of the sequence data (e.g., less than 9,900 sequences). (In my sample, my max was 460.) Note the lower overall quality score at the beginning and end of the reads. By trimming the primers off both ends (REQUIRED), these areas of poor quality will be removed. The commands "`--p-left-trim-len 17 \`" and "`--p-trim-length 448 \`" trim the V4 primers used in this study.

7. Apply quality control filter.

```
qiime quality-filter q-score-joined \
  --i-demux demux-joined.qza \
  --o-filtered-sequences demux-joined-filtered.qza \
  --o-filter-stats demux-joined-filter-stats.qza
```

8. Create visualization of quality control filter outcome.

```
qiime metadata tabulate \
  --m-input-file demux-joined-filter-stats.qza \
  --o-visualization demux-joined-filter-stats.qzv
```

9. View results.

```
qiime tools view demux-joined-filter-stats.qzv
```

10. Denoise your sequences with Deblur

```
qiime deblur denoise-16S \
  --i-demultiplexed-seqs demux-joined-filtered.qza \
  --p-left-trim-len 17 \
  --p-trim-length 448 \
  --p-sample-stats \
  --p-jobs-to-start 80 \
  --o-representative-sequences rep-seqs.qza \
  --o-table table.qza \
  --o-stats deblur-stats.qza
  --verbose
#--verbose shows what the command is doing and where it is in the
denoising process
```

11. Create visualization summary of Deblur feature table and stats

```
qiime feature-table summarize \
  --i-table table.qza \
  --o-visualization table.qzv
```

```
qiime feature-table tabulate-seqs \
  --i-data rep-seqs.qza \
  --o-visualization rep-seqs.qzv
```

```
qiime deblur visualize-stats \
  --i-deblur-stats deblur-stats.qza \
  --o-visualization deblur-stats.qzv
```

12. Generate a phylogenetic tree based on the representative sequences

```
qiime phylogeny align-to-tree-mafft-fasttree \
  --i-sequences rep-seqs.qza \
```

```
  --o-alignment aligned-rep-seqs.qza \
  --o-masked-alignment masked-aligned-rep-seqs.qza \
  --o-tree unrooted-tree.qza \
  --o-rooted-tree rooted-tree.qza

qiime tools export \
  --input-path rooted-tree.qza \
  --output-path exported-rooted-tree
```

13. Alpha and beta diversity analysis.

Using the table.qzv look at the interactive sample detail to choose sampling depth with the highest level of retained sequences for all of the samples

```
qiime diversity core-metrics-phylogenetic \
  --i-phylogeny rooted-tree.qza \
  --i-table table.qza \
  --p-sampling-depth 220000 \
  --m-metadata-file sophie-metadata.tsv \
  --output-dir core-metrics-results
#at a sampling depth of 220000 53.64% of features are retained in
100% of samples

qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/faith_pd_vector.qza \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization core-metrics-results/faith-pd-group-
significance.qzv

qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/evenness_vector.qza \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization core-metrics-results/evenness-group-
significance.qzv

qiime diversity beta-group-significance \
  --i-distance-matrix core-metrics-
results/unweighted_unifrac_distance_matrix.qza \
  --m-metadata-file sophie-metadata.tsv \
  --m-metadata-column Country \
  --o-visualization core-metrics-results/unweighted-unifrac-country-
significance.qzv \
  --p-pairwise
```

14. Alpha rarefaction plotting.

```
qiime diversity alpha-rarefaction \
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-max-depth 200000 \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization alpha-rarefaction.qzv
```

15. Beta-diversity jackknife on 0.01% filtered table.

```
qiime diversity beta-rarefaction \
```

```
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-metric unweighted_unifrac \
  --p-clustering-method upgma \
  --m-metadata-file sophie-metadata.tsv \
  --p-sampling-depth 200000 \
  --o-visualization unweighted_unifrac_jackknife.qzv


qiime diversity beta-rarefaction \
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-metric weighted_unifrac \
  --p-clustering-method upgma \
  --m-metadata-file sophie-metadata.tsv \
  --p-sampling-depth 200000 \
  --o-visualization weighted_unifrac_jackknife.qzv
```

16. Train classifier and assign taxonomies

```
qiime feature-classifier classify-sklearn \
  --i-classifier silva_V3V4_classifier.qza \
  --i-reads rep-seqs.qza \
  --o-classification rep-seqs-taxonomy.qza

qiime metadata tabulate \
  --m-input-file rep-seqs-taxonomy.qza \
  --o-visualization rep-seqs-taxonomy.qzv
17.
```

17. Taxonomy bar chart

```
qiime taxa barplot \
  --i-table table.qza \
  --i-taxonomy rep-seqs-taxonomy.qza \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization taxa-bar-plots.qzv
```

18. Exporting feature table with taxonomy

```
qiime tools export \
  --input-path table.qza \
  --output-path exported table

qiime tools export \
  --input-path rep-seqs-taxonomy.qza \
  --output-path exported rep-seqs-taxonomy

#exported files will be exported into a file that can be
opened in excel
```

APPENDIX D: QIIME 2 DADA2 DENOISING COMMAND PIPELINE

A. Data importing

19. Move sequence data folder (Sophie_Sequences) into the directory you will be using in Qiime 2. The sequence data are in individual compressed files (.gz) of forward and reverse sequences.
20. Import data as type: “cassava-18-paired-end-demultiplexed”.

```
qiime tools import \
  --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path Sophie_Sequences \
  --input-format CasavaOneEightSingleLanePerSampleDirFmt \
  --output-path demux-paired-end.qza
```

21. Generate summary of the demultiplexed results.

```
qiime demux summarize \
  --i-data demux-paired-end.qza \
  --o-visualization demux-paired-end.qzv

#visualizations generated are .qzv files and you can view these
files in Qiime 2
```

22. View generated visualization. The first tab shows the distribution of sequences/samples. The second tab is an interactive graph of quality score as a function of sequence length.

```
qiime tools view demux-paired-end.qzv

#the interactive graph will show you if your sequence reads need to
be trimmed for analysis
```

B. Sequence quality control and feature table construction

Unlike Deblur, DADA2 will both join paired ends and do denoising resulting in the clustering of ASVs. Therefore, there is no need to join the paired ends using q2-vsearch.

23. Denoise sequence data with DADA2

```
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs demux-paired-end.qza \
  --p-n-threads 12 \
  --p-trunc-len-f 250 \
  --p-trim-left-f 17
  --p-trunc-len-r 250 \
  --p-trim-left-r 21 \
  --o-representative-sequences rep-seqs-dada2.qza \
  --o-table table-dada2.qza \
  --o-denoising-stats stats-dada2.qza
  --verbose

#--p-n-threads is not required to run the DADA2 command however by
specifying 12 threads to be used it will cut down the amount of time
needed to run the command. However, note that too many threads can
max out your memory and cause the command to fail.
```

```
#the truncate length and trim lengths were chosen based off the
interactive quality score graph and reflected the trimming used in
the Deblur denoising method
```

24. Rename output using the next two commands

```
mv rep-seqs-dada2.qza rep-seqs.qza
mv table-dada2.qza table.qza
```

25. Create visualization summary of DADA2 feature table and stats

```
qiime feature-table summarize \
  --i-table table.qza \
  --o-visualization table.qzv \
  --m-sample-metadata-file sophie-metadata.tsv

qiime feature-table tabulate-seqs \
  --i-data rep-seqs.qza \
  --o-visualization rep-seqs.qzv

qiime metadata tabulate \
  --m-input-file stats-dada2.qza \
  --o-visualization stats-dada2.qzv

#the metadata file created labeled the sample IDs with features like
type of cheddar, origin of manufacture, and manufacturer
```

26. Create Phylogenic Tree

```
qiime phylogeny align-to-tree-mafft-fasttree \
  --i-sequences rep-seqs.qza \
  --o-alignment aligned-rep-seqs.qza \
  --o-masked-alignment masked-aligned-rep-seqs.qza \
  --o-tree unrooted-tree.qza \
  --o-rooted-tree rooted-tree.qza
```

27. Alpha and Beta Diversity Analysis

Using the table.qzv look at the interactive sample detail to choose sampling depth with the highest level of retained sequences for all of the samples

```
qiime diversity core-metrics-phylogenetic \
  --i-phylogeny rooted-tree.qza \
  --i-table table.qza \
  --p-sampling-depth 451028 \
  --m-metadata-file sophie-metadata.tsv \
  --output-dir core-metrics-results
#sampling depth of 451028 was chosen because it retained 57.25% of
features in 100% of samples
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/faith_pd_vector.qza \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization core-metrics-results/faith-pd-group-
significance.qzv

qiime diversity alpha-group-significance \
```

```
  --i-alpha-diversity core-metrics-results/evenness_vector.qza \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization core-metrics-results/evenness-group-
significance.qzv

qiime diversity beta-group-significance \
  --i-distance-matrix core-metrics-
results/unweighted_unifrac_distance_matrix.qza \
  --m-metadata-file sophie-metadata.tsv \
  --m-metadata-column Country \
  --o-visualization core-metrics-results/unweighted-unifrac-country-
significance.qzv \
  --p-pairwise
```

28. Alpha rarefaction plotting.

```
qiime diversity alpha-rarefaction \
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-max-depth 220000 \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization alpha-rarefaction.qzv
```

29. Beta-diversity jackknife on 0.01% filtered table.

```
qiime diversity beta-rarefaction \
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-metric unweighted_unifrac \
  --p-clustering-method upgma \
  --m-metadata-file sophie-metadata.tsv \
  --p-sampling-depth 200000 \
  --o-visualization unweighted_unifrac_jackknife.qzv


qiime diversity beta-rarefaction \
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-metric weighted_unifrac \
  --p-clustering-method upgma \
  --m-metadata-file sophie-metadata.tsv \
  --p-sampling-depth 200000 \
  --o-visualization weighted_unifrac_jackknife.qzv
```

30. Train classifier and assign taxonomies

```
qiime feature-classifier classify-sklearn \
  --i-classifier silva_V3V4_classifier.qza \
  --i-reads rep-seqs.qza \
  --o-classification rep-seqs-taxonomy.qza

qiime metadata tabulate \
  --m-input-file rep-seqs-taxonomy.qza \
```

```
  --o-visualization rep-seqs-taxonomy.qzv
```

31. Taxonomy bar chart

```
qiime taxa barplot \
  --i-table table.qza \
  --i-taxonomy rep-seqs-taxonomy.qza \
  --m-metadata-file sophie-metadata.tsv \
  --o-visualization taxa-bar-plots.qzv
```

32. Exporting feature table with taxonomy

```
qiime tools export \
  --input-path table.qza \
  --output-path exported table

qiime tools export \
  --input-path rep-seqs-taxonomy.qza \
  --output-path exported rep-seqs-taxonomy

biom convert \
-i exported /feature-table.biom \
-o exported/feature-table.txt \
  --to-tsv

#exported files will be exported into a file that can be
opened in excel
```

APPENDIX E: DEBLUR OTU IDENTIFICATIONS OF NON-CHEESE AMPLICON SEQUENCE VARIANTS AND RELATIVE ABUNDANCE WITHIN EACH SAMPLE WITH SAMPLES ANONYMIZED

			Cheese Identity														
Best Match (OTU Identification)	**Percent Identity**	**E-value**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Acinetobacter baumannii	100	0.00E+00								0.0024					0.0427	0.0013	
Acinetobacter baumannii	100	0.00E+00		0.0009											0.0160		
Anoxybacillus flavithermus	100	0.00E+00		0.0026											0.0024	0.0032	
uncultured Bacteroides sp.	100	0.00E+00									0.0031						
uncultured Flavobacteriaceae	100	0.00E+00													0.0026		
Escherichia coli	100	0.00E+00								0.0005					0.0024		
Geobacillus stearothermophilus	100	0.00E+00		0.0026												0.0013	
Klebsiella pneumoniae	100	0.00E+00								0.0026							
Streptococcus anginosus	93.49	3.00E-179							0.0034								
Lactococcus sp.	94.2	0.00E+00				0.0050				0.0050							
Lactococcus taiwanensis	96.77	0.00E+00			0.0929							0.0002	0.0002	0.0004			
Lactococcus lactis	95.59	0.00E+00									0.0052						
Macrococcus caseolyticus	100	0.00E+00			0.0278												
Mycoplasma alkalescens	100	0.00E+00								0.0031							
Pediococcus parvulus	100	0.00E+00					0.0295					0.0050					
Pseudomonas lactis	100	0.00E+00		0.0009					0.0007					0.0028		0.0010	
Pseudomonas sp.	99.77	0.00E+00								0.0063							
Staphylococcus aureus	100	0.00E+00								0.0029							
Streptococcus dysgalactiae	100	0.00E+00								0.0037		0.0019					
Streptococcus lutetiensis	100	0.00E+00		0.0032												0.0035	
Streptococcus sp.	96.97	0.00E+00	0.0004	0.0029	0.0040				0.0023							0.0026	
Streptococcus anginosus	96.96	0.00E+00						0.0029									
uncultured Enterococcaceae	98.38	0.00E+00			0.5579				0.0973			0.0002				0.3358	
Streptococcus saliviloxodontae	98.38	0.00E+00	0.2478	0.0198	0.3586			0.0007	0.1812	0.0005			0.0002			0.1583	
Streptococcus anginosus	95.82	0.00E+00														0.0064	
Streptococcus uberis	100	0.00E+00		0.0006	0.0018		0.0080		0.0017	0.0126				0.0052			
Streptococcus anginosus	97.43	0.00E+00							0.0064								
Streptococcus anginosus	96.73	0.00E+00							0.0077							0.0067	
Thermus thermophilus	99.3	0.00E+00					0.0089										
Thermus thermophilus	99.53	0.00E+00					0.0048										
Thermus thermophilus	100	0.00E+00					0.0156										
Thermus thermophilus	99.77	0.00E+00					0.0184										
Thermus thermophilus	99.53	0.00E+00					0.0046										
Thermus thermophilus	100	0.00E+00					0.0033										
Thermus thermophilus	100	0.00E+00					0.0026										0.0002

APPENDIX F: DEBLUR OTU IDENTIFICATIONS OF CHEESE-RELATED AMPLICON SEQUENCE VARIANTS AND RELATIVE ABUNDANCE WITHIN EACH SAMPLE

			Cheese Identity														
Best Match (OTU Identification)	**Percent Identity**	**E-value**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**
Lacticaseibacillus paracasei	100	0.00E+00	0.1362					0.0486		0.3993				0.0225			
Lacticaseibacillus paracasei	100	0.00E+00						0.0189						0.0390			
Lacticaseibacillus rhamnosus	100	0.00E+00	0.0198		0.1519												
Lactiplantibacillus plantarum	100	0.00E+00					0.0035								0.0006		
Lactobacillus crispatus	96.75	0.00E+00				0.0029	0.0003					0.0030					
Lactobacillus delbrueckii	99.77	0.00E+00												0.0450	0.0067		
Lactobacillus delbrueckii	100	0.00E+00		0.0278		0.0022	0.0006							4.0493	0.2853		
Lactobacillus helveticus	100	0.00E+00				0.5265	0.0469			0.0579		0.9830					
Lactobacillus helveticus	100	0.00E+00						4.5141	0.9414		11.3532			0.9502	0.0278		
Lactococcus lactis	95.60	0.00E+00												0.0013		0.0048	
Lactococcus lactis	95.36	0.00E+00												0.0026			
Lactococcus lactis	95.59	0.00E+00															0.0052
Lactococcus lactis subsp. lactis	96.75	0.00E+00		0.0099			0.0111					0.0054					
Lactococcus lactis subsp. lactis	97.22	0.00E+00					0.0015						0.0036				
Lactococcus lactis subsp. lactis	97.68	0.00E+00		0.0166		0.0022	0.0073					0.0070	0.0188				
Lactococcus lactis subsp. lactis	100	0.00E+00	0.0004				0.0006								0.0004	0.0006	0.0002
Lactococcus lactis subsp. lactis	100	0.00E+00	0.0019		0.0028	0.0006	0.0009	0.0022			0.0018	0.0010	0.0013	0.0003	0.0019	0.0024	0.0015
Lactococcus lactis subsp. lactis	95.36	0.00E+00	0.0075											0.0008			
Lactococcus lactis subsp. lactis	97.22	0.00E+00	0.0045	0.0022	0.0052			0.0035		0.0017	0.0049	0.0010	0.0018	0.0024	0.0040	0.0038	0.0035
Lactococcus lactis subsp. lactis	99.08	0.00E+00	0.0011		0.0019						0.0014		0.0009	0.0021		0.0014	0.0011
Lactococcus lactis subsp. lactis	100	0.00E+00			0.0006						0.0006			0.0010			
Lactococcus lactis subsp. lactis	100	0.00E+00	99.7835	46.7983	99.4695	18.1074	32.4259	91.1825	0.1752	14.3838	88.5459	22.6139	36.6786	94.4302	99.5898	99.8503	99.9142
Lactococcus lactis subsp. lactis	94.43	0.00E+00	0.0060														
Lactococcus sp.	94.90	0.00E+00												0.0120	0.0065		
Lactococcus sp.	94.20	0.00E+00												0.0050	0.0050		
Latilactobacillus curvatus	100	0.00E+00												0.3117	0.0040	0.0806	0.0004
Lentilactobacillus buchneri	100	0.00E+00			0.0144												
Lentilactobacillus kefiri	100	0.00E+00			0.1192					0.0005				0.0147	0.0301		
Loigolactobacillus coryniformis subsp. torquens	99.77	0.00E+00															0.0042
Loigolactobacillus coryniformis subsp. torquens	100	0.00E+00															0.0087
Paucilactobacillus wasatchensis	100	0.00E+00			0.1517	0.0006	0.0006	0.0432		0.0031							
Secundilactobacillus malefermentans	100	0.00E+00						0.4369				0.0050		0.0005		0.0012	
Secundilactobacillus malefermentans	99.54	0.00E+00						0.0211									
Streptococcus thermophilus	96.29	0.00E+00											0.1299				
Streptococcus thermophilus	97.22	0.00E+00		0.0090									0.0067				
Streptococcus thermophilus	96.76	0.00E+00				0.0032							0.0031				
Streptococcus thermophilus	100	0.00E+00		0.0018		0.0026	0.0032		0.0015	0.0031		0.0017					
Streptococcus thermophilus	100	0.00E+00	0.0388	52.0809	0.0757	80.8252	67.4589	2.6679	98.8732	85.0823	0.0919	74.3755	42.5865	0.0749	0.0378	0.0466	0.0580
Streptococcus thermophilus	100	0.00E+00						0.9655				1.6979	20.3175				
Streptococcus thermophilus	97.45	0.00E+00		0.0031		0.0026	0.0020		0.0051	0.0024		0.0037	0.0022				
uncultured Streptococcus sp.	97.90	0.00E+00		0.0072		0.0038	0.0032					0.0013	0.0009				

APPENDIX G: DEBLUR FILTERED > 0.01% RELATIVE ABUNDANCE OPERATIONAL TAXONOMIC UNIT IDENTIFICATIONS FOR EACH NON-CHEESE AMPLICON SEQUENCE VARIANT WITH SAMPLES ANONYMIZED

			Cheese Identity														
Best Match (OTU Identification)	**Percent Identity**	**E-value**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Acinetobacter baumannii	100	0.00E+00													0.0427		
Acinetobacter baumannii	100	0.00E+00													0.0160		
Lactococcus taiwanensis	96.77	0.00E+00			0.0929												
Macrococcus caseolyticus	100	0.00E+00			0.0278												
Pediococcus parvulus	100	0.00E+00					0.0295										
uncultured Enterococcaceae	98.38	0.00E+00			0.5579				0.0973							0.3358	
Streptococcus saliviloxodontae	98.38	0.00E+00	0.2478	0.0198	0.3586				0.1812							0.1583	
Streptococcus uberis	100	0.00E+00								0.0126							
Thermus thermophilus	100	0.00E+00					0.0156										
Thermus thermophilus	99.77	0.00E+00					0.0184										

APPENDIX H: DEBLUR FILTERED > 0.01% RELATIVE ABUNDANCE OTU IDENTIFICATIONS FOR EACH CHEESES-RELATED ASV

			Cheese Identity														
Best Match (OTU Identification)	**Percent Identity**	**E-value**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**
Lacticaseibacillus paracasei	100	0.00E+00	0.1362					0.0486		0.3993				0.0225			
Lacticaseibacillus paracasei	100	0.00E+00						0.0189						0.0390			
Lacticaseibacillus rhamnosus	100	0.00E+00	0.0198		0.1519												
Lactobacillus delbrueckii	99.77	0.00E+00												0.0450			
Lactobacillus delbrueckii	100	0.00E+00		0.0278										4.0493	0.2853		
Lactobacillus helveticus	100	0.00E+00				0.5265	0.0469			0.0579		0.9830					
Lactobacillus helveticus	100	0.00E+00						4.5141	0.9414		11.3532			0.9502	0.0278		
Lactococcus lactis subsp. lactis	96.75	0.00E+00					0.0111										
Lactococcus lactis subsp. lactis	97.68	0.00E+00		0.0166									0.0188				
Lactococcus lactis subsp. lactis	100	0.00E+00	99.7835	46.7983	99.4695	18.1074	32.4259	91.1825	0.1752	14.3838	88.5459	22.6139	36.6786	94.4302	99.5898	99.8503	99.9142
Lactococcus sp.	94.90	0.00E+00												0.0120			
Latilactobacillus curvatus	100	0.00E+00												0.3117		0.0806	
Lentilactobacillus buchneri	100	0.00E+00			0.0144												
Lentilactobacillus kefiri	100	0.00E+00			0.1192									0.0147	0.0301		
Paucilactobacillus wasatchensis	100	0.00E+00			0.1517			0.0432									
Secundilactobacillus malefermentans	100	0.00E+00						0.4369									
Secundilactobacillus malefermentans	99.54	0.00E+00						0.0211									
Streptococcus thermophilus	96.29	0.00E+00											0.1299				
Streptococcus thermophilus	100	0.00E+00	0.0388	52.0809	0.0757	80.8252	67.4589	2.6679	98.8732	85.0823	0.0919	74.3755	42.5865	0.0749	0.0378	0.0466	0.0580
Streptococcus thermophilus	100	0.00E+00						0.9655				1.6979	20.3175				

APPENDIX I: DADA2 OTU IDENTIFICATIONS OF NON-CHEESE AMPLICON SEQUENCE VARIANTS AND RELATIVE ABUNDANCE WITHIN EACH SAMPLE WITH SAMPLES ANONYMIZED

			Cheese Identity														
Best Match (OTU Identification)	Percent Identity	E-value	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Achromobacter denitrificans	99.77	0.00E+00									0.00032438						
Acinetobacter baumannii	100	0.00E+00								0.00336568					0.04878109	0.0006924	
Acinetobacter baumannii	100	0.00E+00		0.00063929											0.02445261		
Acinetobacter lwoffii	100	0.00E+00					0.00101395										
Anoxybacillus flavithermus	100	0.00E+00		0.00239732											0.00186187	0.00484677	
Anoxybacillus flavithermus	99.77	0.00E+00								0.000439							
Bacteroides pyogenes	100	0.00E+00								0.000439							
Bacteroidetes bacterium	100	0.00E+00												0.00020498			
Bos taurus	100	2.00E-124		0.00047946					0.00054289	0.00263401	0.0004325						
Brevundimonas bullata	100	0.00E+00							0.00027145								
Carnobacterium maltaromaticum	99.77	0.00E+00												0.00051244			
Chlamydiales bacterium	96.49	0.00E+00					0.00050698										
Chryseobacterium haifense	100	0.00E+00					0.00190116										
Corynebacterium sp.	100	0.00E+00													0.00322725		
Deinococcus sp.	99.76	0.00E+00					0.00025349										
Enterococcus faecalis	100	0.00E+00					0.00038023										
Enterococcus faecalis	100	0.00E+00		0.00063929						0.00117067						0.0005193	
Enterococcus italicus	99.3	0.00E+00						0.00039827									
Escherichia coli	100	0.00E+00													0.00359962		
Flavobacteriaceae bacterium	100	0.00E+00												0.00030746			
Fundicoccus ignavus	99.77	0.00E+00								0.00058534							
Fusobacterium necrophum	99.75	0.00E+00								0.00117067							
Fusobacterium necrophum	100	0.00E+00									0.00075688						
Geobacillus sp.	100	0.00E+00					0.00088721										
Geobacillus sp.	98.6	0.00E+00															0.00051791
Geobacillus sp.	99.53	0.00E+00													0.00024825		
Geobacillus stearothermophilus	100	0.00E+00		0.00239732													
Geobacillus stearothermophilus	99.77	0.00E+00															0.00082865
Helcococcus ovis	100	0.00E+00									0.00064875						
Klebsiella pneumoniae	100	0.00E+00								0.00234134							
Klebsiella pneumoniae	100	0.00E+00								0.00160967							
Macrococcus caseolyticus	99.58	1.00E-116			0.03991024												
Macrococcus caseolyticus	100	0.00E+00			0.01961303												
Mycoplasma alkalescens	100	0.00E+00								0.00292668							
Novosphingobium sp.	98.76	0.00E+00					0.00076046										
Novosphingobium sp.	98.51	0.00E+00															0.00041433
Pasteurella multocida	100	0.00E+00												0.00020498			
Pediococcus parvulus	100	0.00E+00					0.02763023					0.00550904					
Planomicrobium flavidum	100	0.00E+00												0.00020498			
Prototheca zopfii	100	0.00E+00								0.00175601							
Pseudomonas fluorescens	100	0.00E+00												0.00040995			
Pseudomonas lactis	100	0.00E+00												0.00317713			
Pseudomonas lactis	100	0.00E+00		0.00079911				0.00026551	0.00067861							0.0008655	
Pseudomonas sp.	99.77	0.00E+00								0.00848736							
Pseudomonas sp.	100	0.00E+00					0.00139419									0.0008655	
Pseudomonas sp.	100	0.00E+00								0.001317							
Pseudomonas sp.	100	0.00E+00												0.0008199			
Pseudomonas sp.	100	0.00E+00												0.00051244			
Pseudomonas sp.	96.96	0.00E+00												0.00020498			
Ralstonia insidiosa	100	0.00E+00					0.00025349										
Rhodococcus sp.	100	0.00E+00															0.00062149
Romboutsia	100	0.00E+00					0.00088721		0.00108578	0.00190234		0.00030606			0.00062062		
Rothia marina	100	0.00E+00							0.02633022								
rumen bacterium	96.26	0.00E+00					0.00038023										
Sanguibacter sp.	100	0.00E+00												0.00020498			
Secundilactobacillus malefermentans	100	0.00E+00					0.45640571		0.0047503					0.00071742			
Secundilactobacillus malefermentans	100	0.00E+00					0.09835347										
Shewanella putrefaciens	99.53	0.00E+00															0.00031074
Staphylococcus aureus	100	0.00E+00								0.00234134							
Staphylococcus aureus	99.53	0.00E+00								0.000439							
Staphylococcus aureus	100	0.00E+00				0.00024268											
Staphylococcus chromogenes	100	0.00E+00															0.0011394
Staphylococcus uberis	100	0.00E+00					0.00811163			0.01697472				0.00389454			
Staphylococcus uberis	99.53	0.00E+00	0.00162217		0.00091223	0.00097071	0.00012674	0.00066378	0.00067861	0.000439							
Streptococcus anginosus	95.81	0.00E+00		0.00031964													
Streptococcus dysgalactiae	100	0.00E+00								0.00760936		0.00234644					
Streptococcus lutetiensis	100	0.00E+00		0.00543392												0.00501987	
Streptococcus parasuis	98.83	0.00E+00								0.00014633							
Streptococcus phocae	98.83	0.00E+00							0.00013572								
Streptococcus salivarius	95.58	0.00E+00			0.00114029												
Streptococcus uberis	99.77	0.00E+00	0.00054072		0.00182447				0.00095006								
Thermus thermophilus	100	0.00E+00					0.25500925								0.00111712		0.00196805
Thermus thermophilus	99.52	0.00E+00					0.16451392										
Trueperella pyogenes	100	0.00E+00															0.03770364
uncultured Actinomycetales bacterium	100	0.00E+00													0.0004965		
uncultured anaerobic bacterium	99.51	0.00E+00							0.00027145								
uncultured anaerobic bacterium	99.77	0.00E+00	0.00036048														
uncultured anaerobic bacterium	98.36	0.00E+00								0.00029267							
uncultured anaerobic bacterium	100	0.00E+00								0.00029267							
uncultured anaerobic bacterium	97.87	0.00E+00						0.00026551									
uncultured anaerobic bacterium	99.01	0.00E+00													0.00024825		
uncultured bacterium	100	0.00E+00					0.00139419										
uncultured bacterium	100	0.00E+00								0.00102434	0.0004325						
uncultured bacterium	100	0.00E+00														0.0006924	
uncultured bacterium	99.75	0.00E+00												0.00020498			
uncultured bacterium	100	0.00E+00													0.00211012		
uncultured bacterium	94.51	4.00E-173												0.00051244			
uncultured bacterium	100	0.00E+00									0.00054063						
uncultured bacterium	100	0.00E+00					0.00038023										
uncultured bacterium	99.75	0.00E+00		0.00271696												0.00103859	
uncultured bacterium	100	0.00E+00													0.00211012		
uncultured bacterium	98.25	0.00E+00							0.00040717								

uncultured bacterium	100	0.00E+00										0.00021927				
uncultured bacterium	100	0.00E+00												0.00024825		
uncultured Bacteroides sp.	100	0.00E+00									0.00335189					
uncultured Bacteroides sp.	99.53	0.00E+00			0.00068418											
uncultured Flavobacteriaceae bacterium	100	0.00E+00												0.003972		
uncultured Porphyromonadaceae	99.76	0.00E+00									0.00021625					
uncultured Porphyromonas sp.	99.76	0.00E+00								0.000439						
uncultured proteobacterium	99.75	0.00E+00													0.0003462	
uncultured rumen bacterium	93.28	9.00E-165										0.00021927				
uncultured rumen bacterium	98.51	0.00E+00													0.0003462	
uncultured Streptococcus sp.	97.89	0.00E+00			0.09874933											
uncultured Streptococcus sp.	97.89	0.00E+00													0.05071801	
uncultured Streptococcus sp.	99.53	0.00E+00		0.00655267											0.00605847	
uncultured Streptococcus sp.	94.39	0.00E+00													0.0008655	
uncultured Streptococcus sp.	94.88	0.00E+00		0.00031964												
uncultured Streptococcus sp.	100	4.00E-131	0.00018024													
uncultured Thermus bacterium	100	0.00E+00										0.00020404				

APPENDIX J. DADA2 OTU IDENTIFICATIONS OF CHEESE-RELATED AMPLICON SEQUENCE VARIANTS AND RELATIVE ABUNDANCE WITHIN EACH SAMPLE

			Cheese Identity														
Best Match (OTU Identification)	Percent Identity	E-value	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Lacticaseibacillus paracasei	100	0.00E+00	0.14099362					0.05868255		0.3820565				0.0493145			
Lacticaseibacillus paracasei	100	0.00E+00	0.03190447					0.0510779						0.06028954			
Lacticaseibacillus paracasei	100	0.00E+00								0.09483144							
Lacticaseibacillus paracasei	100	0.00E+00						0.0183779									
Lacticaseibacillus rhamnosus	100	0.00E+00	0.09012189		0.16608737					0.01886699							
Lacticaseibacillus rhamnosus	100	0.00E+00	0.01315648		0.03315626												
Lacticaseibacillus casei	95.09	0.00E+00										0.0040717					
Lactiplantibacillus plantarum	99.77	0.00E+00													0.0060669		
Lactobacillus delbrueckii	100	0.00E+00		0.02508643		0.0017309								4.3205062	0.31742064		
Lactobacillus delbrueckii	90.63	2.00E-157												0.00058534			
Lactobacillus delbrueckii	90.55	3.00E-145												0.00029267			
Lactobacillus helveticus	100	0.00E+00						4.17026303	0.8698111		10.0289614			0.88356364	0.02972785		
Lactobacillus helveticus	100	0.00E+00				0.48796616	0.05146244			0.06231071		0.63491116					
Lactobacillus helveticus	99.77	0.00E+00									0.02548103						
Lactobacillus helveticus	99.77	0.00E+00									0.02517029						
Lactobacillus helveticus	99.77	0.00E+00									0.02382373						
Lactobacillus helveticus	99.77	0.00E+00									0.02382373						
Lactobacillus helveticus	99.77	0.00E+00									0.02330582						
Lactobacillus helveticus	99.77	0.00E+00									0.01999122						
Lactobacillus helveticus	99.77	0.00E+00									0.01957689						
Lactobacillus helveticus	99.77	0.00E+00									0.01947331						
Lactobacillus helveticus	99.77	0.00E+00									0.01926615						
Lactobacillus helveticus	100	0.00E+00									0.01905898						
Lactobacillus helveticus	98.13	0.00E+00									0.00155372						
Lactobacillus helveticus	90.63	2.00E-157						0.00050698			0.00103581						
Lactobacillus helveticus	100	0.00E+00	0.00065782														
Lactobacillus hokkaidonensis	98.59	0.00E+00									0.00082865						
Lactococcus cremoris	92.41	3.00E-145	0.00603005	0.0027367	0.00479491		0.00159821	0.00633721		0.00086887	0.00642205	0.00027145	0.00234313	0.00439001	0.00424684	0.00502191	0.00616316
Lactococcus cremoris	91.6	3.00E-140				0.0008655											
Lactococcus cremoris	92.41	3.00E-145				0.0008655											
Lactococcus lactis	97.66	0.00E+00														0.00112737	
Lactococcus lactis	98.36	0.00E+00														0.00092239	
Lactococcus lactis	97.77	0.00E+00	0.10996625					0.11635114									
Lactococcus lactis	97.89	0.00E+00							0.00902734	0.00570975							
Lactococcus lactis	97.19	0.00E+00							0.00451367	0.00186187		0.00692186					
Lactococcus lactis	97.19	0.00E+00											0.01225636				
Lactococcus lactis	97.42	0.00E+00								0.00459262							
Lactococcus lactis	97.19	0.00E+00							0.00371714								
Lactococcus lactis sub. lactis	94.61	0.00E+00												0.000439			
Lactococcus lactis sub. lactis	100	0.00E+00	99.253589	38.8632196	98.2881167	17.1698655	30.7128507	88.0205224	0.2903353	13.7428628	83.4131946	15.3928632	25.3679612	93.5354106	98.9181489	87.8425661	99.4635887
Lactococcus lactis sub. lactis	99.77	0.00E+00														11.6804547	
Lactococcus lactis sub. lactis	100	0.00E+00									2.74387212						
Lactococcus lactis sub. lactis	99.77	0.00E+00									1.11205025						
Lactococcus lactis sub. lactis	98.36	0.00E+00	0.20162307		0.2423978						0.06173453			0.09189764	0.18613275		
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.17322872			0.15538834			0.14252805			0.15555286			
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.13456341						0.12212251				0.1096897		0.10931502
Lactococcus lactis sub. lactis	99.77	0.00E+00						0.10988719			0.10523873			0.1198474	0.10677759		
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.33456802						
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.15007034						0.12429772						
Lactococcus lactis sub. lactis	98.58	0.00E+00			0.15374304						0.10927841						
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.13089072						0.11901506						
Lactococcus lactis sub. lactis	98.59	0.00E+00	0.00526259													0.11027709	0.13807644
Lactococcus lactis sub. lactis	99.77	0.00E+00						0.11647788			0.14967517						
Lactococcus lactis sub. lactis	99.77	0.00E+00						0.11546393			0.12947679						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.14905368						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.14366745						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.13103051						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.12108669						
Lactococcus lactis sub. lactis	99.77	0.00E+00													0.13262262		
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.09964534						
Lactococcus lactis sub. lactis	100	0.00E+00	0.02225638		0.02152607						0.0172981						
Lactococcus lactis sub. lactis	97.19	0.00E+00	0.00657824		0.00091817		0.00063929	0.00405581		0.00136537	0.0055934	0.0017644			0.00533888	0.00584182	0.00724442
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.0362535						
Lactococcus lactis sub. lactis	98.13	0.00E+00					0.02716961										
Lactococcus lactis sub. lactis	98.36	0.00E+00				0.01384792											
Lactococcus lactis sub. lactis	96.24	0.00E+00						0.00152093									
Lactococcus lactis sub. lactis	97.19	0.00E+00														0.00092239	
Lactococcus lactis sub. lactis	97.42	0.00E+00													0.00060669		
Lactococcus lactis sub. lactis	97.89	0.00E+00						0.00012674					0.00018024			0.00010249	
Lactococcus lactis sub. lactis	100	2.00E-142	0.00021927														
Lactococcus lactis sub. lactis	100	2.00E-129	0.00010964														
Lactococcus lactis sub. lactis	97.66	0.00E+00													0.00012134		
Lactococcus lactis sub. lactis	90.02	5.00E-153														0.00020498	
Lactococcus lactis sub. lactis	97.89	0.00E+00												0.13740746		0.00425998	
Lactococcus lactis sub. lactis	97.89	0.00E+00									0.07219626						0.07839109
Lactococcus lactis sub. lactis	98.13	0.00E+00								0.02159774						0.07789088	
Lactococcus lactis sub. lactis	98.13	0.00E+00									0.0753037						
Lactococcus lactis sub. lactis	97.89	0.00E+00				0.01090524							0.02829777				
Lactococcus piscium	100	0.00E+00															0.00194626
Lactococcus sp.	97.42	0.00E+00												0.001317			
Lactococcus sp.	100	0.00E+00						0.00088721									
Lactococcus sp.	98.36	0.00E+00							0.03982652	0.02159774							
Latilactobacillus curvatus	100	0.00E+00												0.33305579	0.00497486	0.07307394	0.00032438
Lentilactobacillus buchneri	100	0.00E+00			0.03978751												
Lentilactobacillus farraginis	99.77	0.00E+00									0.00082865						
Lentilactobacillus kefiri	100	0.00E+00			0.12140293									0.01624305	0.03106257		
Lentilactobacillus kisonensis	100	0.00E+00			0.00132625												
Leptotrichiaceae bacterium	99.51	0.00E+00												0.00073167			
Leuconostoc lactis	96.96	0.00E+00					0.00031964										
Loigolactobacillus coryniformis	100	0.00E+00															0.01124507
Loigolactobacillus coryniformis subsp. torquens	100	0.00E+00															0.01297508
Paucilactobacillus wasatchensis	100	0.00E+00			0.1445613			0.04410697		0.00471675							
Paucilactobacillus wasatchensis	99.3	0.00E+00				0.0008655	0.00047946										
Streptococcus thermophilus	96.72	0.00E+00					0.00063929										
Streptococcus thermophilus	100	0.00E+00	0.11709268	46.5549484	0.18506295	81.0485993	67.7800348	3.37659111	96.2871067	82.1576635	0.17971378	55.2965814	40.0135901	0.2288661	0.09161032	0.12472789	0.1640266
Streptococcus thermophilus	100	0.00E+00		13.727981				2.56796339				27.6919256	33.6907504				
Streptococcus thermophilus	100	0.00E+00		0.65612428		0.33183083	0.34185767		0.37689159	0.20641974		0.24077221	0.08381187		0.05969838		
Streptococcus thermophilus	98.83	0.00E+00				0.29219116			0.18625534	0.21386724		0.20290555					
Streptococcus thermophilus	99.77	0.00E+00				0.03133092				0.6829353							
Streptococcus thermophilus	99.77	0.00E+00					0.10196596		0.13740148	0.13777866		0.06596127	0.06182252				
Streptococcus thermophilus	98.83	0.00E+00				0.09832025	0.14144181		0.07832548	0.06653096		0.09690606	0.04866496				
Streptococcus thermophilus	98.83	0.00E+00				0.15388503	0.15230966		0.13209128				0.12724985				
Streptococcus thermophilus	98.83	0.00E+00				0.17102184			0.11669169	0.10538206		0.06786139					
Streptococcus thermophilus	99.77	0.00E+00					0.10516239		0.1444375	0.14063353							

Streptococcus thermophilus	98.83	0.00E+00				0.09001149	0.06744457		0.08363568	0.05548384		0.06175386						
Streptococcus thermophilus	100	0.00E+00					0.07511599		0.11324006	0.11071943			0.04415894					
Streptococcus thermophilus	99.77	0.00E+00							0.16501453	0.14324016								
Streptococcus thermophilus	100	0.00E+00					0.0887008		0.12903791			0.06392543	0.05983987					
Streptococcus thermophilus	98.59	0.00E+00				0.0180023	0.22998279		0.01500132				0.10850483					
Streptococcus thermophilus	99.06	0.00E+00							0.15094249	0.13455141								
Streptococcus thermophilus	99.77	0.00E+00								0.26537908								
Streptococcus thermophilus	99.77	0.00E+00							0.13766699	0.12114592								
Streptococcus thermophilus	100	0.00E+00							0.13979107	0.11866342								
Streptococcus thermophilus	99.77	0.00E+00								0.1205253		0.06826856	0.05983987					
Streptococcus thermophilus	99.77	0.00E+00							0.11204526	0.10935405								
Streptococcus thermophilus	98.83	0.00E+00							0.0773962	0.04555385		0.05890369						
Streptococcus thermophilus	100	0.00E+00								0.10575443			0.04253678					
Streptococcus thermophilus	99.77	0.00E+00								0.12921404								
Streptococcus thermophilus	99.77	0.00E+00							0.13700321									
Streptococcus thermophilus	99.77	0.00E+00								0.11878755								
Streptococcus thermophilus	99.77	0.00E+00								0.10786455								
Streptococcus thermophilus	99.77	0.00E+00								0.09917581								
Streptococcus thermophilus	97.89	0.00E+00											0.13103491					
Streptococcus thermophilus	98.83	0.00E+00					0.09381508											
Streptococcus thermophilus	98.83	0.00E+00							0.05164172									
Streptococcus thermophilus	99.77	0.00E+00											0.06416565					
Streptococcus thermophilus	98.13	0.00E+00						0.04461394										
Streptococcus thermophilus	98.59	0.00E+00								0.03971997								
Streptococcus thermophilus	99.77	0.00E+00											0.04253678					
Streptococcus thermophilus	92.95	1.00E-148		0.00364894		0.00484677	0.00367589		0.00331888	0.003972		0.00325735	0.00468626					
Streptococcus thermophilus	97.42	0.00E+00		0.00342088					0.00584122	0.00335137		0.00502174	0.00306409					
Streptococcus thermophilus	92.68	6.00E-147						0.00025349				0.00230729						
Streptococcus thermophilus	96.49	0.00E+00					0.00063929											
Streptococcus thermophilus	98.36	0.00E+00							0.00039827									
Streptococcus thermophilus	98.13	0.00E+00				0.0001731						0.00013572						
uncultured Lactobacillales bacterium	97.9	0.00E+00										0.00027145						
uncultured Lactobacillales bacterium	99.53	0.00E+00														0.00051244		
uncultured Lactococcus sp.	98.36	0.00E+00												0.0001463				

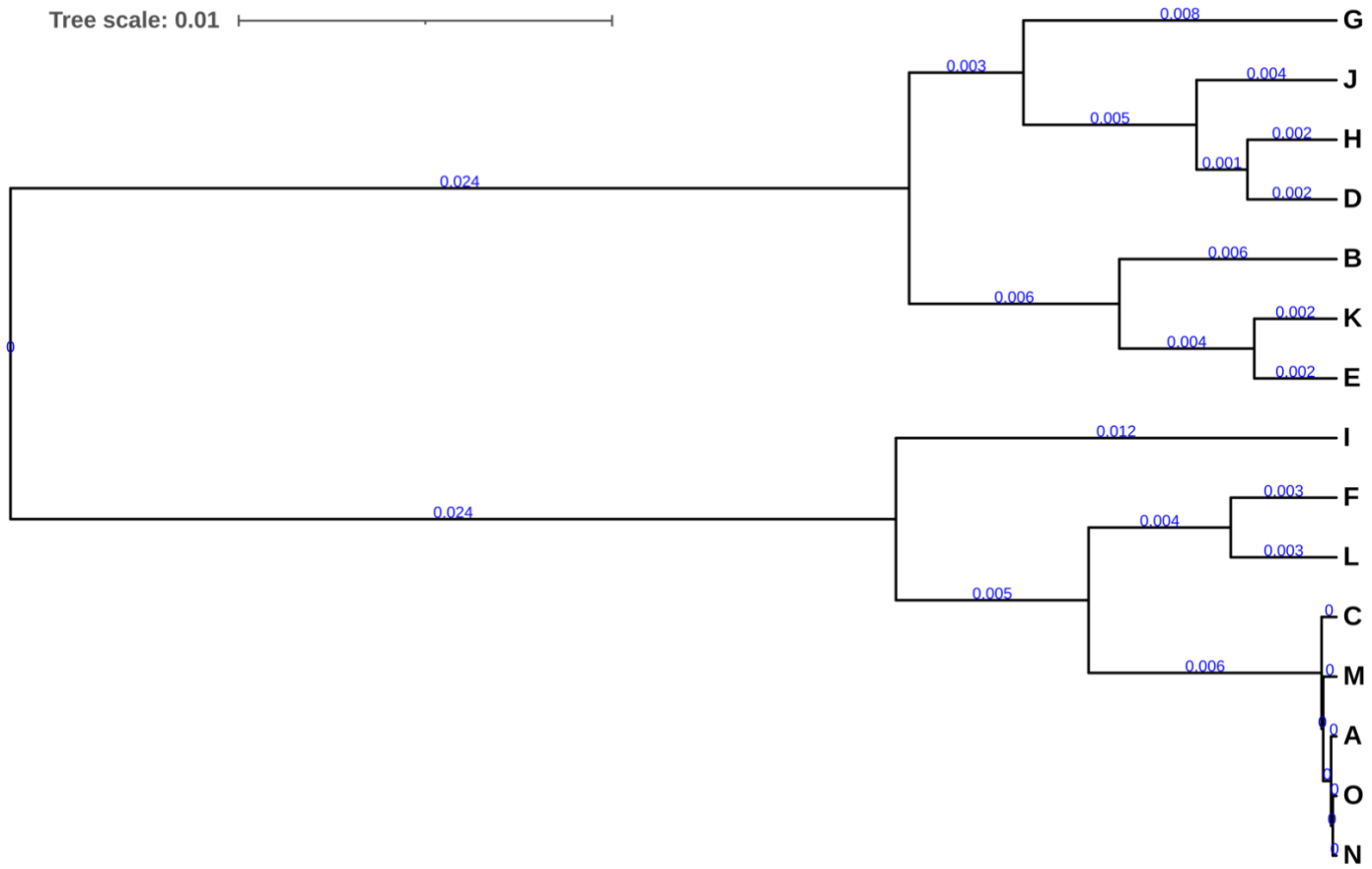
APPENDIX K: DADA2 FILTERED > 0.01% RELATIVE ABUNDANCE OPERATION TAXONOMIS UNIT IDENTIFICATIONS FOR EACH NON-CHEESE-AMPLICON SEQUENCE VARIANT WITH SAMPLES ANONYMIZED

			Cheese Identity														
Best Match (OTU Identification)	**Percent Identity**	**E-value**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Acinetobacter baumannii	100	0.00E+00													0.05		
Acinetobacter baumannii	100	0.00E+00													0.02		
Macrococcus caseolyticus	99.58	1.00E-116			0.04												
Macrococcus caseolyticus	100	0.00E+00			0.02												
Pediococcus parvulus	100	0.00E+00					0.03										
Rothia marina	100	0.00E+00							0.03								
Secundilactobacillus malefermentans	100	0.00E+00					0.46										
Secundilactobacillus malefermentans	100	0.00E+00					0.10										
Staphylococcus uberis	100	0.00E+00								0.02							
Thermus thermophilus	100	0.00E+00					0.26										
Thermus thermophilus	99.52	0.00E+00					0.16										
Trueperella pyogenes	100	0.00E+00															0.04
uncultured Streptococcus sp.	97.89	0.00E+00			0.10												
uncultured Streptococcus sp.	97.89	0.00E+00														0.05	

APPENDIX L: DADA2 FILTERED > 0.01% RELATIVE ABUNDANCE OPERATION TAXONOMIS UNIT IDENTIFICATIONS FOR EACH CHEESE-RELATED-AMPLICON SEQUENCE VARIANT WITH SAMPLES ANONYMIZED

			Cheese Identity														
Best Match (OTU Identification)	Percent Identity	E-value	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Lacticaseibacillus paracasei	100	0.00E+00	0.14					0.06		0.38				0.05			
Lacticaseibacillus paracasei	100	0.00E+00	0.03					0.05						0.06			
Lacticaseibacillus paracasei	100	0.00E+00								0.09							
Lacticaseibacillus paracasei	100	0.00E+00						0.02									
Lacticaseibacillus rhamnosus	100	0.00E+00	0.09		0.17					0.02							
Lacticaseibacillus rhamnosus	100	0.00E+00	0.01		0.03												
Lactobacillus delbrueckii	100	0.00E+00		0.03										4.32	0.32		
Lactobacillus helveticus	100	0.00E+00						4.17	0.87		10.03			0.88	0.03		
Lactobacillus helveticus	100	0.00E+00				0.49	0.05			0.06		0.63					
Lactobacillus helveticus	99.77	0.00E+00									0.03						
Lactobacillus helveticus	99.77	0.00E+00									0.03						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	99.77	0.00E+00									0.02						
Lactobacillus helveticus	100	0.00E+00									0.02						
Lactococcus lactis	97.77	0.00E+00	0.11					0.12									
Lactococcus lactis	97.19	0.00E+00											0.01				
Lactococcus lactis sub. lactis	100	0.00E+00	99.25	38.86	98.29	17.17	30.71	88.02	0.29	13.74	83.41	15.39	25.37	93.54	98.92	87.84	99.46
Lactococcus lactis sub. lactis	99.77	0.00E+00														11.68	
Lactococcus lactis sub. lactis	100	0.00E+00									2.74						
Lactococcus lactis sub. lactis	99.77	0.00E+00									1.11						
Lactococcus lactis sub. lactis	98.36	0.00E+00	0.20		0.24						0.06			0.09	0.19		
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.17			0.16			0.14			0.16			
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.13						0.12				0.11		0.11
Lactococcus lactis sub. lactis	99.77	0.00E+00						0.11			0.11			0.12	0.11		
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.33						
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.15						0.12						
Lactococcus lactis sub. lactis	98.58	0.00E+00			0.15						0.11						
Lactococcus lactis sub. lactis	99.77	0.00E+00			0.13						0.12						
Lactococcus lactis sub. lactis	98.59	0.00E+00														0.11	0.14
Lactococcus lactis sub. lactis	99.77	0.00E+00						0.12			0.15						
Lactococcus lactis sub. lactis	99.77	0.00E+00						0.12			0.13						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.15						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.14						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.13						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.12						
Lactococcus lactis sub. lactis	99.77	0.00E+00													0.13		
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.10						
Lactococcus lactis sub. lactis	100	0.00E+00	0.02		0.02						0.02						
Lactococcus lactis sub. lactis	99.77	0.00E+00									0.04						
Lactococcus lactis sub. lactis	98.13	0.00E+00					0.03										
Lactococcus lactis sub. lactis	98.36	0.00E+00				0.01											
Lactococcus lactis sub. lactis	97.89	0.00E+00												0.14		0.06	
Lactococcus lactis sub. lactis	97.89	0.00E+00									0.07						0.08
Lactococcus lactis sub. lactis	98.13	0.00E+00								0.02						0.08	
Lactococcus lactis sub. lactis	98.13	0.00E+00									0.08						
Lactococcus lactis sub. lactis	97.89	0.00E+00				0.01							0.03				
Lactococcus sp.	98.36	0.00E+00							0.04	0.02							
Latilactobacillus curvatus	100	0.00E+00												0.33		0.07	
Lentilactobacillus buchneri	100	0.00E+00			0.04												
Lentilactobacillus kefiri	100	0.00E+00			0.12									0.02	0.03		
Loigolactobacillus coryniformis	100	0.00E+00															0.01
Loigolactobacillus coryniformis subsp. torquens	100	0.00E+00															0.01
Paucilactobacillus wasatchensis	100	0.00E+00			0.14			0.04									
Streptococcus thermophilus	100	0.00E+00	0.12	46.55	0.19	81.05	67.78	3.38	96.29	82.16	0.18	55.30	40.01	0.23	0.09	0.12	0.16
Streptococcus thermophilus	100	0.00E+00		13.73				2.57				27.69	33.69				
Streptococcus thermophilus	100	0.00E+00		0.66		0.33	0.34		0.38	0.21		0.24	0.08		0.06		
Streptococcus thermophilus	98.83	0.00E+00				0.29			0.19	0.21		0.20					
Streptococcus thermophilus	99.77	0.00E+00				0.03				0.68							
Streptococcus thermophilus	99.77	0.00E+00					0.10		0.14	0.14		0.07	0.06				
Streptococcus thermophilus	98.83	0.00E+00				0.10	0.14		0.08	0.07		0.10	0.05				
Streptococcus thermophilus	98.83	0.00E+00				0.15	0.15		0.13				0.13				
Streptococcus thermophilus	98.83	0.00E+00				0.17			0.12	0.11		0.07					
Streptococcus thermophilus	99.77	0.00E+00					0.11		0.14	0.14							
Streptococcus thermophilus	98.83	0.00E+00				0.09	0.07		0.08	0.06		0.06					
Streptococcus thermophilus	100	0.00E+00					0.08		0.11	0.11			0.04				
Streptococcus thermophilus	99.77	0.00E+00							0.17	0.14							
Streptococcus thermophilus	100	0.00E+00					0.09		0.13			0.06	0.06				
Streptococcus thermophilus	98.59	0.00E+00				0.02	0.23		0.02				0.11				
Streptococcus thermophilus	99.06	0.00E+00							0.15	0.13							
Streptococcus thermophilus	99.77	0.00E+00								0.27							
Streptococcus thermophilus	99.77	0.00E+00							0.14	0.12							
Streptococcus thermophilus	100	0.00E+00							0.14	0.12							
Streptococcus thermophilus	99.77	0.00E+00								0.12		0.07	0.06				
Streptococcus thermophilus	99.77	0.00E+00							0.11	0.11							
Streptococcus thermophilus	98.83	0.00E+00							0.08	0.05		0.06					
Streptococcus thermophilus	100	0.00E+00								0.11			0.04				
Streptococcus thermophilus	99.77	0.00E+00								0.13							
Streptococcus thermophilus	99.77	0.00E+00							0.14								
Streptococcus thermophilus	99.77	0.00E+00								0.12							
Streptococcus thermophilus	99.77	0.00E+00								0.11							
Streptococcus thermophilus	99.77	0.00E+00								0.10							
Streptococcus thermophilus	97.89	0.00E+00											0.13				
Streptococcus thermophilus	98.83	0.00E+00					0.09										
Streptococcus thermophilus	98.83	0.00E+00							0.05								
Streptococcus thermophilus	99.77	0.00E+00											0.06				
Streptococcus thermophilus	98.13	0.00E+00						0.04									
Streptococcus thermophilus	98.59	0.00E+00								0.04							
Streptococcus thermophilus	99.77	0.00E+00											0.04				

APPENDIX M: DEBLUR SAMPLE CLUSTERING



Letunic and Bork (2019) *Nucleic Acids Res* **doi: 10.1093/nar/gkz239**

APPENDIX N: DEBLUR RELATIVE ABUNDANCE IN CLUSTER ORDER WITH REGION OF MANUFACTURE

			Relative Abundance			
			Lactococcus lactis	*Streptococcus thermophilus*	Lactobacilli	Other Species
Cheese[1]	Cluster	Region	----------%----------			
G	1	Ireland	0.18	98.87	0.94	
J	1	Ireland	22.61	76.53	0.98	
H	1	Ireland	14.38	85.08	0.46	
D	1	Australia	18.11	81.48	0.53	
B	1	Vermont	46.91	53.36	0.03	0.03
K	1	California	36.70	63.53		
E	1	Ireland	32.44	67.50	0.05	
I	2	Utah	88.55	0.09	11.35	
F	2	Ireland	91.18	3.63	5.08	0.06
L	2	California	94.44	0.09	5.43	
C	2	New York	99.47	0.08	0.44	
M	2	California	99.59	0.04	0.34	
A	2	Oregon	99.78	0.04	0.16	
O	2	Wisconsin	99.91	0.06		
N	2	Wisconsin	99.85	0.05	0.08	

[1]Cheeses as described in Table 1.