Leveraging High-Throughput Sequencing Data to Understand Complex Systems: Effects of Heat Stress Along the Swine Gastrointestinal Microbiome, Inflammation in the Bovine Milk Microbiome, and the Genetics of Human Lactation with Respect to Human Milk Oligosaccharide Synthesis

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctorate of Philosophy

with a

Major in Bioinformatics and Computational Biology

in the

College of Graduate Studies

University of Idaho

by

Sarah L. Brooker

Major Professor: Mark A. McGuire, Ph.D.

Committee Members: Michelle K. McGuire, Ph.D.; James A. Foster, Ph.D.;

Christopher H. Remien, Ph.D.; Gordon K. Murdoch, Ph.D.

Department Administrator: Dave Tank, Ph.D.

May 2019

Authorization to Submit Dissertation

This dissertation of Sarah L. Brooker, submitted for the degree of Doctorate of Philosophy with a major in Bioinformatics and Computational Biology and titled "Leveraging High-Throughput Sequencing Data to Understand Complex Systems: Effects of Heat Stress Along the Swine Gastrointestinal Microbiome, Inflammation in the Bovine Milk Microbiome, and the Genetics of Human Lactation with Respect to Human Milk Oligosaccharide Synthesis," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____Date_____
Mark A. McGuire, Ph.D.

Committee Members: _____Date_____
Michelle K. McGuire, Ph.D.

_____Date_____
James A. Foster, Ph.D.

_____Date_____
Christopher H. Remien, Ph.D.

_____Date_____
Gordon K. Murdoch, Ph.D.

Department Administrator: _____Date_____
Dave Tank, Ph.D.

Abstract

Research has become entwined with sequencing of DNA and RNA to contribute to our understanding of life. The progression of sequencing technology from pattern finding; to targeted, quantitative sequencing; to high-throughput sequencing has provided deep surveys into the breadth of variation in life across all ecosystems. This compilation of chapters seeks to leverage the development of high-throughput sequencing to assess changes in the microbiomes of the gastrointestinal tracts (GIT) of heatstressed pigs, the milk microbiome of cows with varying levels of mammary inflammation, and the host genetic variation of milk carbohydrate-related genes in women around the globe.

In heatstressed pigs, microbial composition clustered distinctly by location in the GIT (stomach, ileum, colon/feces) and differed in select microbes and diversity metrics. The major finding was that fecal assessment of bacterial composition may not accurately evaluate membership or function the entire gastrointestinal microbiome. Milk microbial composition differed in cows on two dairies and varied with somatic cell counts (SCC; a proxy for inflammation). Bacterial richness and diversity of milk with low SCC (<200,000 cells/mL) was greater than in milk with greater SCC (>200,000 cells/mL). Multivariate analysis showed that quarters within a cow were more similar in their microbial composition than among all quarters analyzed. The genetic regulation and composition of milk carbohydrates examined variation on a global scale among women. Some of the genes (*FUT2*, *FUT3*, *ST6GalNAc5*), which are critical in the synthesis of human milk oligosaccharides (HMO), were found to be under selective pressure. Additionally, novel single nucleotide polymorphisms (SNP) were identified and known ones confirmed and related to the HMO and lactose in human milk.

These studies provide just a glimpse into the field of milk and lactation, addressing how the GIT may change with external input, how the milk microbiome may fluctuate due to inflammation, and how genetic regulation likely plays a key role in milk composition. The complexities in the field of lactation have only grown with the advancement of sequencing

technologies. However, our ability to leverage new information, particularly related to milk, may be the key to further benefitting mother and neonate.

Vita

Williams, JE, Carrothers, JM, Lackey, KA, Beatty, NF, **Brooker, SL**, Peterson, HK, Steinkamp, KM, York, MA, Shafii, B, Price, WJ, McGuire, MA, and McGuire, MK. Strong Multivariate Relationships Exist Among Milk, Oral, and Fecal Microbiomes in Mother-Infant Dyads During the First Six Months Postpartum. J Nutr 2018.

McGuire, MK, Meehan, CL, McGuire, MA, Williams, JE, Foster, J, Sellen, DW, Kamau-Mbuthia, EW, Kamundia, EW, Mbugua, S, Moore, SE, Kvist, LJ, Otoo, GE, **Brooker, SL**, Price, WJ, Shafii, B, Placek, C, Lackey, KA, Robertson, B, Manzano, S, Ruiz, L, Rodriguez, JM, Pareja, R, and Bode, L. What's normal? Oligosaccharide concentrations and profiles in milk produced by healthy women vary geographically. AJCN. 2017 May. Ridenhour, BJ, Brooker, SL, Williams, JE, Van Leuven, JT, Miller, AW, Dearing, D, and Remien, CH. Modeling time-series data from microbial communities. The ISME journal 11.11 (2017): 2526.

Lackey, KA; Williams, JE; Price, WJ; Carrothers, JM; **Brooker, SL**; Shafii, B; McGuire, MA; McGuire, MK. Comparison of Commercially-Available Preservatives for Maintaining the Integrity of Bacterial DNA in Human Milk. Journal of microbiological methods 141 (2017): 73-81.

Williams, JE, Carrothers, JM, Lackey, KA, Beatty, NF, York, MA, **Brooker, SL**, Shafii, B, Price, WJ, Settles, ML, McGuire, MA, and McGuire, MK. Human milk microbial community structure is relatively stable and related to variation in macronutrient and micronutrient intakes in healthy lactating women. J Nutr 147.9 (2017): 1739-1748.

McGuire, MK, Meehan, CL, **Brooker, SL**, Williams, JE, and McGuire, MA. An Evolutionary, Sociocultural Perspective on Variation in Human Milk Microbes and Oligosaccharides: An Example of Eco-Homeorhesis? In: McGuire, M.K., M.A. McGuire and L. Bode, Prebiotics and Probiotics in Human Milk. pp. 463-468. ISBN: 978-0-12-802725-7. Elsevier. San Diego CA.

Carrothers JM, York MA, **Brooker SL**, Lackey KA, Williams JE, Shafii B, Price WJ, Settles ML, McGuire MA, McGuire MK. Fecal Microbial Community Structure Is Stable over Time and Related to Variation in Macronutrient and Micronutrient Intakes in Lactating Women. J Nutr. 2015 Oct;145(10):2379-88.

Yahvah KM, **Brooker SL**, Williams JE, Settles M, McGuire MA, McGuire MK. Elevated dairy fat intake in lactating women alters milk lipid and fatty acids without detectable changes in expression of genes related to lipid uptake or synthesis. Nutr Res. 2015 Mar;35(3):221-8.

Acknowledgements

I would like to thank the following people and groups;

My major professor Dr. Mark McGuire for his endless patience, guidance, and support;

My committee members, Dr. Shelley McGuire, Dr. James Foster, Dr. Chris Remien, and Dr. Gordon Murdoch for their input, questions, and direction;

Dr. Janet Williams for being a guiding light in the lab and in life;

Dr. Bahman Shafii and Dr. Bill Price for their continued efforts in expanding statistics for highly multivariate data sets;

Dr. Brenda Murdoch, Kim Davenport, and their lab team for their expertise in genomics;

Dr. Ben Ridenhour for his time and effort which has made me a better coder;

Dr. Sam Hunter for he patience in teaching and helping me improve my computational abilities;

Dan New and the Genomics Resource Core Team for providing expert help and sequencing, without whom this dissertation would not be possible;

The co-authors of the research chapters for their efforts in all aspects of the research;

The faculty and staff of the Institute for Bioinformatics and Evolutionary Studies (IBEST);

The faculty and students of both the Bioinformatics and Computational Biology Program and the Department of Animal Science for their friendship and support;

Overall funding sources: BCB Fellowship and the Idaho Agricultural Experiment Station

And all the members of the McGuire research laboratory, past and present, for their friendship and making a great work environment.

Dedication

This dissertation is dedicated to the many woman around the world who volunteer to donate their milk to science so that we can better understand the ways in which milk impacts mother and infant.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction: How advances in sequencing technology advanced the field of human milk and lactation

Modern research has adopted the idea that high-throughput genetic sequencing will be the answer to most scientific inquiry regarding living systems; concisely summarized by Frederick Sanger: "knowledge of sequences could contribute much to our understanding of living matter" (Heather, 2016). Genetic sequencing spans all kingdoms of taxonomy including bacteria, archaea, and eukaryotes and impacts our understanding of biology from the utility of commensal microbiomes to the impact of specific point mutations on genome structures. The advanced technology in high-throughput sequencing allows scientists to explore the vastness of what nucleotide sequences can tell us due to advances in time, quality, and cost of sequencing. This difference is particularly noticeable when sequencing human versus bacterial genomes where the human genome is approximately 3300 Mbp (megabase pairs) and bacterial genomes range from 130 kilobase pairs (kbp) to 14 Mbp (Van Leuven, 2014; Han, 2013; Venter, 2001). Moore's Law states the complexity of microchips (measured by number of transistors per unit cost) doubles approximately every two years, while sequencing capabilities between 2004 and 2010 doubled every five months (Stein, 2010), pointing out the lighting fast advances in sequencing capacity supporting the depth of knowledge gained. From the initial Sanger sequencing and the subsequent advent of 454 pyrosequencing, the forward-thinking aspects indicated ability to increase speed and quality of sequencing while markedly decreasing cost of analysis (Heather, 2016). One of the major advances of pyrosequencing (a sequence by synthesis method where colored pyrophosphate is released for every nucleotide added by DNA polymerase) over Sanger sequencing (a chain termination sequencing technique where a terminal base gets added by polymerase and the resulting terminated amplicons can determine the DNA sequence) was the application of mass parallelization (the ability to sequence many strands of DNA simultaneously) of analysis (Shendure, 2008), marked by the improvement in cost and quality in sequencing the human genome (Wheeler,

2008). The next advancements allowed for bridge-amplification (a technique whereby a DNA strand is amplified in two directions, forming a 'bridge,' both strands are sequenced with phosphorescence associated when each nucleotide is added) developed into the current Illumina technology, which has further increased the length, quality, and parallelization capacity for high-throughput sequencing (Heather, 2016). The wealth of knowledge gained from high-throughput sequencing has important implications for any health-related field, if and only if we can manage to process and accurately interpret data of this magnitude.

Computational advances (though outside the true scope of this review) have allowed for the accurate use of increasing amounts of sequencing data. Particularly related to the following topical studies, utilizing DNA sequence analysis to identify bacterial taxa and single nucleotide polymorphisms (SNPs) will continue to be a field in need of advancement (Schloss, 2010; Liu, 2008; Lluch, 2015; Callahan, 2016; McKenna, 2010). The ability to leverage sequencing data comes equally from these advancements in analysis, wherein DNA sequences are aligned to a reference genome (e.g., BWA, Bowtie2, Dada2) or *de novo* assembled into longer reads (e.g., SPAdes, MEGA); once the DNA sequences are known, their similarity to known DNA or other samples can be assessed (Dalca, 2010). Once DNA can be compared, sequences can be classified to specific bacteria for microbiome work (e.g., RDP, SpinGo) or host DNA can be assessed for changes such as SNPs, insertions, deletions, or structural changes (e.g., GATK, Samtools). The field of milk and lactation has progressed due to the advancement of sequencing technology, as briefly summarized in Figure 1. Early studies focused on the genetic regulation of nutritional components in milk, as well as studying pathogens that caused mammary infection, relevant in both dairy cows and humans. Analyses of bacteria in milk were reliant on culture-dependent techniques to identify living bacteria thought to be environmental contamination from the skin, bedding, or milking practices surrounding lactation (Björkstén, 1980; Schalm, 1971). Culture-dependent methods were able to identify staphylococci, streptococci, enterococci, and enterobacteria as mastitis pathogens (Jimenez, 2017). The first lactic acid bacteria (LAB) were isolated

in studies from the early 2000s in human and bovine milk (Heikkila, 2003, Martin, 2003; Park, 2007) which altered the perception of milk only containing pathogenic bacteria when infected, but potentially harboring commensal bacteria. Culture-dependent techniques in bovine milk continue to be integral to mastitis pathogen identification (Bouchard, 2015; Braem, 2014, Klostermann, 2008). For many years, culture-dependent techniques allowed for the identification of live, definable microbes across the biosphere, not just limited to milk. Bacterial culture is still used to enumerate bacterial counts and identify pathogens in both human and animal models (Martin, 2003; Mediano, 2017; Broom, 2006; Hojberg, 2005; Katouli, 1999). Culture-dependent techniques also allow expanded evaluation of commensal microbiome by contributing in-depth characterization for virulence factors, antibiotic response, and potential probiotic traits (Jimenez, 2017).

The advancement of knowledge truly escalated with fast and inexpensive sequencing technology sufficient to examine the DNA profile of bacteria. Additionally, the change in philosophy regarding the presence of a non-disease causing bacterial community was supported by the introduction of the Human Microbiome Project in which site-specific bacterial communities were characterized in healthy humans (Turnbaugh, 2007). Initial characterization of DNA patterns from milk can be found with studies using denaturing or temperature gradient gel electrophoresis (DGGE/TGGE). These techniques were popular in the early 1990s (Muyzer, 1993) and helped jump start the field of microbial ecology. Differences in DNA sequences based on migration of bands in the gel due to nucleotide sequence differences make species classification tangible. Progress made by using DNA migration patterns was that cells need not be cultivable, nor the conditions to grow the bacteria known to identify bacteria. Advances in lactation research from DGGE/TGGE include describing diverse microbial communities in human milk and maternal feces (Perez, 2007), identifying potential mastitis pathogens in women (Delgado, 2008) and cows (Kuang, 2009; Quigley, 2011), as well as the diversity of the gastrointestinal tract (GIT) microbes in humans (Muyzer, 1993; Matto, 2005) and swine (Haenen, 2013; Kostantinov, 2006). The limitation of this type

of analytical approach is that only known patterns of bacteria can be identified, thus the hypotheses can only be driven by the current knowledge and isolation of bacterial DNA from the site of interest.

Further expansion of microbial ecology came with the ability to look at bacterial sequence directly, over the pattern detection available with DGGE/TGGE. The benefits of quantitative or real time PCR (qPCR) allowed genetic identification of specific bacteria via DNA-binding dye that fluoresces when bound to a specific target. Multiple sequences can be targeted and the DNA quantified by use of a calibration curve. The quantitative aspect is also beneficial for the enumeration of bacteria in a location, and can potentially be used as an estimation for the total bacterial load in a sample by amplifying the hypervariable regions of the 16S rRNA gene, which provides an avenue to treat bacteria as colony forming units instead of relative abundance (Li, 2015).This method is also timely, making it a commonly-used diagnostic tool for bovine mastitis (Koskinen, 2010; Shome, 2011, Mahmmod, 2013; Taponen, 2009; Zanardi, 2014), confirming the presence of specific bacterial taxa in milk such as Clostridium cluster IV and XIV, Prevotella, and Porphyormonas (Collado, 2009), determining the complexity of the human GIT microbiome (Rinttila, 2004), and identifying effects of feed additives on the GIT microbiome of pigs (Castillo, 2008; Costa, 2014; Heinritz, 2016). The time efficiency gain in identifying microbes and their quantity in myriad samples allowed for faster pathogen diagnosis, decreased time and cost to treat animals, as well as a benchmark for studying the etiology of mastitis (Mahmmod, 2013; Shome, 2011; Taponen, 2009; Zanardi, 2014) although this technology is still limited by the necessity of prior bacterial knowledge to prepare target specific primers.

The adaptation of sequencing to high-throughput, short-read technology (<1000 base pairs; as reviewed above and in Heather, 2016) increased the speed and accuracy to which taxa could be assigned with less stringent criteria for bacterial identification. The slow evolution rate of the 16S rRNA (small subunit bacterial ribosome) gene in bacteria enabled targeted sequencing as an ubiquitous way to identify bacteria with research growing to

encompass bacteria never cultured nor previously identified (Woese, 1987). The first study in human milk (Hunt et al., 2011) using 16S rRNA sequencing showed surprising results with known mastitis pathogens, staphylococci and streptococci comprising a significant percentage of the microbial community of milk from mothers self-identified as healthy. Further studies on the human milk microbiome have expanded into impacts of various environmental factors on community characteristics and have provided insight to the diversity of bacteria in a fluid that was once considered sterile (Jimenez, 2008; Flint, 2012; Jost, 2013; Patel, 2017; Williams, 2017). High-throughput sequencing also has provided deep surveys into the breadth of bacteria that inhabit most biological niches, such as the human GIT (Anderson, 2008, Wu, 2011), the bovine milk microbiome (Bhatt, 2011; Bonsaglia, 2017; Braem, 2012, Cressier, 2011, Doyle, 2017; Ganada, 2017; Falentin, 2016; Kuehn, 2013; Lima, 2017, Lima, 2018; Metzger, 2018; Oikonomou, 2012; Oikonomou, 2014; Rodrigues, 2017; Young, 2015), and the swine GIT (which may be a model species sufficient for examination of dysbiosis in the human GIT microbiome) (Buzoianu, 2012; Kim, 2011; Kim, 2012; Li, 2016; Looft, 2012; Mach, 2015; Mann, 2014; Niu, 2015; Pajarillo, 2014; Pajarillo, 2015; Ramavo-Caldas, 2016; Vahjen, 2010, Yu, 2017; Zhao, 2015). Analysis of 16S rRNA not only provides membership in the bacterial community as characteristics of the community are often summarized as diversity and/or richness estimators; where richness refers to the number of species identified in a sample, and diversity refers to the relative amounts of these species within the system (alpha diversity) or the relative amounts of bacteria across a system (beta diversity) (Hughes, 2001). These estimators allow for ecological exploration of bacterial communities as a complex 'holobiont,' versus looking at differences in individual species exemplified by findings such as decreased bacterial diversity in the case of human mastitis (Jimenez, 2017).

Advancements in sequencing technology not only helped the field with respect to the microbiome, but provided an avenue which to pursue questions of microbial function. Metagenomics is a "brute-force" method that looks at all DNA in a sample and uses sequence analysis to discover "who's there" (Mardis, 2008). Additionally, because the sequences iden-

tified are more than just marker genes (e.g., 16S rRNA) to identify bacterial taxa, metagenomics provides an avenue to discover the functional capacity in a sample based on the genes present from both the host and the microbiome (e.g., Lamendella, 2011) opening the door to exploration of host-microbe interactions. Metagenomics has been used to provide insight into the genetic functionality of various biological niches: human milk (Ward, 2013); human GIT (Turnbaugh, 2009; Gill, 2006); bovine milk (Ganada, 2016); and swine GIT (Looft, 2012; Looft, 2014; Poroyko, 2010, Willing, 2009; Yang, 2016). Metagenomics also provides insight into the expressed genes in a niche, which has the potential to expand the knowledge of milk composition through knowledge about the proteome and metabolome (or DNA associated with protein expression and the metabolic potential of a niche). Though metagenomics technology is still a fledgling process, discoveries of functional capacity of milk and GIT microbiomes may be beneficial to continued advances in the health of infants and other neonates.

The ability to detect the constituents of the microbiome, the potential functional genetics of the microbiome, as well as the abundant host DNA found in metagenomic sampling begs the question about microbiome-host interaction at the genetic level. While metagenomics does give some capacity to look at host gene function alongside bacterial function, high-throughput sequencing data also provide the capacity to look at whole host genomes and targeted sequencing which may also impact lactation. The instinct for genetic regulation with regard to lactation is largely drawn from the dairy industry, where breeding practices and genetics have long held a place in milk quality and production (Lemay, 2009a; Lemay, 2009b), yet little is known about how human milk is genetically regulated. Genetic testing helped advance the ability to select for traits beyond just the phenotypic features of a cow, but based on genetic heritability increase the rate at which famers can see genetic improvement in their herds (Wiggins, 2011). The same tools developed to scan and assess the bovine genome for lactation can be used to determine the genetic characteristics of human milk traits (Durbin, 2011).

The best human lactation example for an aspect of milk being driven by genetics is that of human milk oligosaccharides (HMO), a set of complex sugars that, despite not serving as nutrients per se for the infant, may act as prebiotics for the microbiota of the developing infant GIT and potentially the maternal milk microbiome (Bode, 2012). The genetic pathway for production of HMO is not well understood; however, some of the specific linkages of the various sugars are performed by well characterized genes from the ABO-Lewis blood group structures (Soejima, 2007; Soejima, 2009; Ferrer-Admetlla, 2009; Teppa, 2016). An example of such a gene is *FUT2*, known as the 'secretor' gene, which encodes α-(1,2)-fucosyltransferase whose function is to add a fucose sugar to a lactose backbone, creating both this moiety in HMO and one of the basic structures in the H antigen, the precursor for the AB blood group antigens (Guo, 2017). The general variation in this gene has focused on particular missense mutations that turn off this fucosyltransferase and create a 'non-secretor' phenotype (Heneghan, 2000). Literature has focused on the genetic variation in *FUT2* with occasional comparison to phenotype by blood typing (Soijima, 2007). However, this same linkage is a main driver in HMO profile in human milk, particularly driving the presence/absence of a particular HMO, 2-fucosyllactose (2'FL), which is a precursor for several other HMO moieties (Bode, 2012). Variation based on the predominance or presence of 2'FL has been demonstrated in comparisons of human milk from populations around the world (McGuire, 2017; Erney, 2000). Additionally, these genes may be under selective pressure from environmental, pathogen, or historical reasons that not only drive the diversity of ABO blood groupings, but subsequently drive the diversity seen in HMO profiles (Fumagalli, 2009).

Some of the next major questions to answer in the field of milk and lactation are how host genomics (particularly in humans) impact the nutrient composition and the microbial composition in milk, and how the interactions may also play an essential role in the development of the suckling neonate's GIT.

Though the field of lactation has become more complex due to the advancement of se-

quencing technology by discovering a complex microbial ecosystem and potential host genetic impact, the field is also integral to understanding infant nutrition, health, and welfare. Thus, lactation could be a pivotal point to leverage high-throughput sequencing data examining the interrelationship of the milk microbiome, the GIT microbiome, and host genetic factors that influence mammalian health. The following chapters will utilize high-throughput sequencing technology to address questions in the field of lactation.

1) Are fecal samples the most representative way to capture the complexity of the whole GIT microbiome? Using a swine model, it is hypothesized that the microbiome will vary across sampling sites (stomach, ileum, colon, and feces) along the GIT. In addition, supplementation with Zn and exposure to heat stress will alter the membership, diversity and richness of the bacterial community at each site.

2) How does the bovine milk microbiome vary across levels of inflammation? Using a cow model, it is hypothesized that the milk microbiome will be similar among quarters within a dairy cow, but differ based on somatic cell count (a proxy for inflammation).

3) How do SNPs in lactation-related genes relate to milk composition? In a diverse group of healthy, lactating women, it is hypothesized that variation will be found among these women with regards to oligosaccharide-related genes, and that this variation will be associated with known phenotypic differences with evidence of selective pressure on the variation detected.

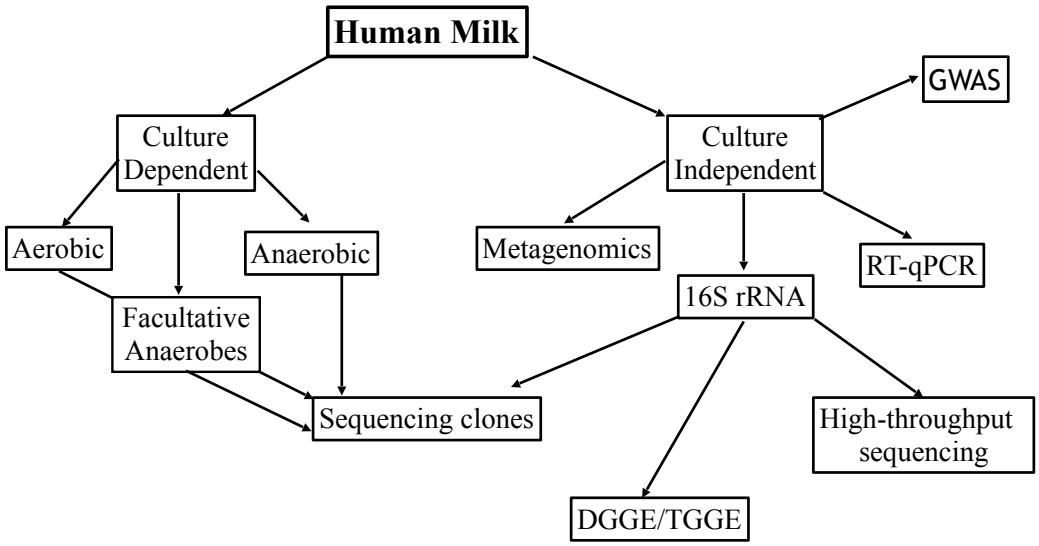


Figure 1.1: Summary of techniques used to leverage sequencing data regarding human milk.

CHAPTER 2

Impacts of heat stress and zinc supplementation on the microbiome along the porcine gastrointestinal tract

2.1 Abstract

Heat stress is known to lead to high morbidity and mortality in production animal practice. However, little is known about the impact of heat stress on the microbial composition in the gastrointestinal tract (GIT) of pigs. The study objectives were to evaluate the microbial composition along GIT segments under acute (1-day) and chronic (7-day) heat stress conditions (36 °C; ∼50% humidity; temperature-humidity index ≈ 85.5), and levels of Availa® Zn 100 (ZnAA). Crossbred gilts were fed *ad libitum* one of three diets: control (120 ppm zinc as ZnSO4, n = 15), Zn-220 (control + 100 ppm zinc as ZnAA, n = 7), or Zn-320 (control + 200 ppm zinc as ZnAA, n = 8) for 25 days prior to heat stress exposure. After heat exposure, pigs were euthanized and luminal contents were collected from three segments (stomach, ileum, and colon), and feces. The microbial community at each site was assessed by sequencing the 16S rRNA V1-V2 region. Sequences were filtered using Mothur and classified using Spingo, RDP, and the Silva database. ZnAA supplementation did not affect bacterial community membership at any sampling site. Microbial composition distinctly clustered by sampling site in PCoA analysis, but only *Dolichospermum* in the ileum and *Muribaculum* in feces were found to be affected by heat stress. This study demonstrates that segment of GIT is an important aspect of GIT microbiome studies as fecal samples are not representative of community composition in the upper GIT. Further work examining the impact of ZnAA or heat stress would require greater numbers of pigs.

2.2 Introduction

Production systems work to improve animal growth and health for animal welfare and economic benefit. Heat stress (HS) is well established as a detriment to animal production through reduced growth and feed efficiency, compromised intestinal barrier function, and increased animal health costs (Lee, 2016). Nutrient supplementation can be beneficial to animal production (Song, 2014; Sanz Fernandez, 2013), and there is limited evidence that supplemental zinc (Zn) can ameliorate HS effects on growth and health including intestinal integrity (Sales, 2013; Sanz-Fernandez et al., 2014). However, mechanisms by which HS negatively impacts animal production or acts as a vector for increased risk of diarrhea are poorly understood (Castillo, 2008; Costa, 2014; Janczyk, 2013).

Both HS and dietary supplementation have been hypothesized to lead to changes in the gastrointestinal (GI) microbiome (Starke, 2014; Buzoianu, 2012; Li, 2016), where clear evidence of modulating health and disease in the host has been demonstrated (Dethlefson, 2007; Flint, 2012; Mach, 2015). A vast diversity of microbes exists in the GI tract with estimates of over 500 species in humans (Eckburg, 2005) and 1000 species in pigs (Looft, 2012). Microbes may confer health to the host through participation in digestive processes, but dysbiosis of the GI microbiome may lead to a variety of disease states such as diarrhea (Alam, 1994; Allen, 2012), irritable bowel and Crohn's disease (Walker, 2011), enterocolitis in newborns (Bering, 2018), and with long lasting impacts such as obesity (Cani, 2008; Turnbaugh, 2006; Yang, 2016) or atopic diseases and allergy (Stefka, 2014; Stokholm, 2018). Animals, in particular, are impacted by changes in bacterial communities which impact growth rates (Ramayo-Caldas, 2016), state of inflammation (Cani, 2008; Walker, 2011), and diarrheal disease (Castillo, 2008; Costa, 2014).

The GI microbiome can be impacted by a number of factors including animal genetics and husbandry practices (Pajarillo, 2014; Doyle, 2017; Lamendella, 2011). In particular, supplementation can alter the microbial composition along the GI tract. Zinc oxide (ZnO)

has been used to ameliorate the impact of diarrhea in weaning piglets, but there is conflicting evidence as to the impact on the microbial composition of the GI tract (Katouli, 1999; Vahjen, 2010; Pieper, 2012; Yu, 2017). Other strategies to alter the GI microbiome include addition of fiber or starch (Haenen, 2013; Niu, 2015; Heinritz, 2016; Metzler-Zebeli, 2010, Tako, 2008), calcium phosphate (Trowman, 2006; Govers, 1993; Metzler-Zebeli, 2010b; Mann, 2014), and antibiotics (Looft, 2012; Kim, 2012). There is clear need to understand the effects of supplementation on animal health and production outcomes through alterations in the GIT microbial communities.

Health outcomes for pigs are also negatively impacted by HS, yet the response of the microbiome in the GI tract is not well studied. Evidence from poultry shows impacts of HS on the bacterial community of the small intestine of broiler chickens, particularly with increases in coliform bacteria, which include potential GIT pathogens (Song, 2014). Due to the potential impact of the GI microbiome to health and production outcomes in pigs, we evaluated the change in GI microbiome in response to HS conditions and supplementation with Zn. We hypothesized that HS would alter the relative abundance of some bacterial genera in segments of the pig GI tract; supplemental Zn would reduce genera associated with inflammation/pathogenesis in the small intestine; and supplemental Zn would mitigate changes in the GI microbiome due to HS.

2.3 Materials and Methods

Animals and Experimental Design

The Iowa State University Institutional Animal Care and Use Committee approved all procedures involving animals. Forty-three crossbred gilts were assigned to one of three diets, as described in Sanz-Fernandez, 2014. Zinc was provided above the control (ZnSO4) in a form complexed to amino acids (ZnAA; Availa®, Zinpro Corporation, Eden Prairie, MN, USA). Treatments were briefly: control diet (120 ppm ZnSO4) (pair-fed/Zn-control), control plus 100 ppm ZnAA (Zn-220), and control plus 200 ppm ZnAA (Zn-320). There were three

experimental phases during the study. Pigs were housed in groups according to their dietary treatment during the first step and fed experimental diets for 20 ± 1 days. During the second time step, pigs were housed individually and kept in thermoneutral conditions (19 °C; ∼61% humidity; temperature-humidity index ≈ 64) for 5 d. In the third step, pigs were exposed to constant heat stress for a period of 1 or 7 d (36 °C; ∼50% humidity; temperature-humidity index ≈ 85.5) except for a thermoneutral group of pigs kept at conditions of step 2 for 1 or 7 additional days (Table 1). At the designated day in step three, pigs were sacrificed using the captive bolt technique with exsanguination. After euthanasia, intestinal segments were collected and the luminal contents collected aseptically, frozen on ice, and stored at -80 °C. Samples of luminal contents were shipped frozen to the University of Idaho and kept at -80 °C until analysis.

Extraction and Amplification of Bacterial DNA

DNA was extracted from ∼200 mg of frozen luminal contents using the QIAamp DNA Stool Mini Kit (Qiagen Cat. 51504, Valencia, CA) following the manufacturer's protocol. DNA was eluted in AE buffer (10 mM Tris-Cl; 0.5 mM tris-acetate-ethylenediamine tetraacetic acid (EDTA, 0.5 M, pH 8, Sigma, St. Louis, MO); pH 9.0; Qiagen) and stored at -80 °C.

Polymerase chain reaction (PCR) was conducted to amplify the V1-V2 hypervariable region of the bacterial 16S rRNA gene. The PCR protocol was carried out in a dedicated PCR hood pre-cleaned with 70% ethanol and sterilized under ultraviolet light for 30 min. The PCR mixture (50 μL) contained the following: 0.5 μM forward primer (7F- GCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG), 0.5 μM barcoded reverse primer (338R-GCCTCCCTCGCGCCATCAGTGNNNNNNNNCATGCTGCCTCCCGTAGGAGT), 1 μL DNA extract from colon and fecal samples, 5 μL DNA extract from stomach and ileum samples, and PCR master mix containing a final concentration of 0.05 μM target specific primers with barcodes, 1x PCR buffer (Life Technologies, Carlsbad, CA), 3.12 mM MgCl2 (Life Technologies), 0.25% mM DMSO, 0.2 mM dNTP (Life Technologies), and 25 U/mL

AmpliTaq DNA 360 polymerase (Life Technologies) in 50 μL total volume. Thermal cycler settings included a 5 min denaturation step at 94 °C followed by 35 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 2 min. A final elongation step at 72 °C for 2 min was then performed to complete each reaction before storing PCR products at -20 °C until further use. Products from PCR were electrophoresed on 1% agarose gels made with EDTA (0.5 M, pH 8, Sigma, St. Louis, MO), TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) and containing ethidium bromide (0.0007 mg/mL). Gels were run for 30 min at 80 V before bands were viewed using the Bio-Rad Fluor-S Multimager (Bio-Rad Laboratories). Samples with high-quality product (relatively bright band of interest at 381 bp), low primer-dimers, and absence of unwanted bands or smears were considered acceptable to move on to sequencing. Due to low amplification, 7 samples failed to be sequenced. Ten μL of each PCR reaction were added to a pool sample that underwent emulsion PCR as previously described (Margulies, 2005). Pyrosequencing was performed at the University of Idaho, Genomics Resource Core on a 454 Life Sciences Genome Sequencer FLX machine (Roche, Bradford, CT).

Sequence Analysis

Mothur (v.1.24.0; Schloss, 2009) was utilized to bin sequences by sample and carry out quality control procedures. Sequences were removed from the data set if they had any ambiguous bases, contained homopolymer runs greater than 8 bases, greater than one difference from the barcode, or greater than 2 differences from the forward primer. Sequences were also removed if they did not maintain an average quality score of 35 over a sliding window of 50 bases. Sequences were then subjected to a pairwise alignment using Smith-Waterman global alignment (Vingron, 1994) against Mothur's Silva reference database and were trimmed to cover the same region. Sequences that did not align correctly were then removed from the dataset. A 2% single-linkage, precluster method was employed as it has been shown to remove sequences that may contain sequencing error (Huse, 2010). Potential chimeras were identified and removed using Mothur's implementation of the ChimeraSlayer algorithm (Haas, 2011). To determine taxonomy, SPINGO (Allard, 2015) was used to classify unique

sequences to the genus level against the Silva database (Quast, 2012). Further analysis was performed on sequences which SPINGO returned as "ambiguous". These sequences were taken through the Silva aligner against both the Silva and RDP databases (Wang, 2007). Where there was agreement on taxonomic classification, sequences were reassigned to the lowest possible taxonomic level. Sequences which remained ambiguous or were classified as chloroplast or mitochondria were removed. The taxonomy was then merged with the count data produced in Mothur, and relative abundances calculated for further analysis.

Data Analysis

After generation of relative abundances at the lowest taxonomic level, all exploratory and statistical analyses were performed in R (version 3.5.1). Thermoneutral samples collected on day 1 and day 7 were compared using Wilcoxon Rank Sum test and determined to not be different and were collapsed into a single group. One-way ANOVA was used to assess diversity indices in thermoneutral samples across GI segment. Non-parametric Kruskal-Wallis tests were used to calculate F statistic values to determine differences among environmental treatments. If the F statistic was significant ($P<0.05$), non-parametric paired Wilcoxon Rank Sum test, with Bonferroni adjustment to correct for multiple comparisons was used to determine differences within environmental treatments. Diversity indices were assessed at the lowest identified taxonomic level using richness, Pielou's evenness, Shannon diversity, and Simpson evenness and calculated using the vegan package in R as described in Williams, 2019. The Ordiplot function was used to create principal coordinates analysis (PCoA) plots of microbial community data using the Hellinger transformation and stacked bar charts were made using ggplot2 (version 3.0.0).

2.4 Results

Sequence Data Disposition

Over 230,000 sequence reads passed quality control measures, and clustered into 14,890 unique sequences based on 3% similarity. After removing unclassified sequences, an average

of 2036 ± 716 sequences per sample remained. By using a combination of the SPINGO and Silva databases to classify unique sequences, only 3% of sequences remained unclassified and were removed from analysis.

Microbial Composition by GI Segment

Overall, the bacterial community composition across the GI segments and feces under thermoneutral conditions differed in their most abundant taxa with the exception of *Prevotella* (18-32%) which was the most abundant genera identified (Tables 2-5).

The dominant bacteria in the luminal contents from the stomach from pigs under thermoneutral conditions were *Prevotella* and *Actinobacillus* (Table 2). The dominant bacteria in luminal contents from the ileum were *Prevotella* followed by *Escherichia/Shigella* and *Romboutsia* (Table 3). The dominant bacteria in luminal contents from the colon (Table 4) and in feces (Table 5) was *Prevotella*; no other genera exceeded a relative abundance of 10%.

Under thermoneutral conditions the bacterial community in the stomach generally had lower richness and diversity than the colon and feces, but similar in diversity to the ileum (Tables 2-5). The bacterial community in the ileum had the lowest richness of all GIT segments, and notably lower Shannon diversity, despite having the highest evenness scores across the GIT. The bacterial community in the colon had the highest richness and diversity across GIT segments while the fecal samples did not differ from any other segment in diversity or evenness.

The PCoA summary of the community structure of all samples after Hellinger transformation showed distinct clustering by sampling location (Figure 1). Bacterial communities of the stomach and ileum clustered independently and were mostly driven by the presence of *Lactobacillus*, *Escherichia/Shigella*, *Actinobacillus*, and *Clostridium*. The bacterial communities in the colon and feces clustered tightly together, separate from the stomach and ileum and were driven by the presence of *Prevotella*.

Zn Supplementation

There were no significant differences detected by non-parametric Kruskal-Wallis test, nor

clustering by PCoA (Figure 2) on the microbial community composition at any GIT segment related to supplementation with ZnAA. Therefore, Zn treatment groups were collapsed into HS-related categories only.

Heat Stress

Stomach No differences in the relative abundance of the top 9 bacterial genera were detected between acute and chronic HS in the stomach, but differences were found between thermoneutral and chronic HS conditions (Table 2 and Figure 3). The stomach, under HS, was still dominated by *Prevotella* (16%), but had a greater ($P < 0.008$) relative abundance of *Dolichospermum* (10%) in chronic HS than in pigs under thermoneutral conditions (1%; Table 2). *Actinobacillus* was affected by HS ($P < 0.05$), but no significance could be declared among environmental treatments. No significant differences in diversity, richness, or evenness in bacterial communities from the luminal contents of the stomach were found due to HS (Table 2).

Ileum The relative abundance of *Actinobacillus* ($P < 0.04$), *Romboutsia* ($P < 0.05$), and *Klebsiella* ($P < 0.03$) in ileal contents differed due to HS, but separation within environmental treatments was not found after Bonferroni correction (Table 3). *Escherichia/Shigella* (17%), *Actinobacillus* (16%), and *Clostridium* (11%) were the most abundant genera from the bacterial community of luminal contents from the ileum of pigs under HS (Table 3 and Figure 3). The diversity of the bacterial community in the ileum was impacted by HS. Pielou's evenness ($P < 0.04$) and Shannon diversity ($P < 0.04$) were lower under acute HS as compared to thermoneutral conditions (Table 3).

Colon and Feces No differences were found due to environmental treatment for either bacterial taxa or diversity indices from the community in the colon dominated by *Prevotella* (Table 4). The bacterial community of feces changed little due to HS exposure and were dominated in all treatments by Prevotella (Figure 3). Chronic HS decreased ($P < 0.004$) the abundance of *Muribaculum* when compared to the relative abundance during thermoneutral and acute HS conditions. Environmental treatment affected ($P < 0.04$) the relative abun-

dance of *Ruminococcus* but within treatment differences were undetected. No significant differences were found with respect to effect of environmental treatment on diversity indices of the bacterial community of feces (Table 5).

2.5 Discussion

Heat stress negatively impacts animal agriculture and contributes to challenges of animal health and economic success (Renaudeau, 2011). Research is needed to counter the effects of heat stress, particularly due to inconclusive results of supplementation studies (Sales, 2013). Pigs exposed to HS in this study had increased respiratory rate and rectal temperature and decreased feed intake as compared to thermoneutral conditions demonstrating the presence of heat stress (Sans Fernandez, 2013).

Zn Supplementation

No differences were detected in the microbiome due to ZnAA supplementation in any of the GI segments. Several other studies have been conducted looking at Zn supplementation on changes in the microbiome mostly in the ileum and using varying techniques to identify microbes. Using 16S sequencing of ileal samples and comparing low ZnO (∼150 ppm) to high Zn (∼2500 ppm) the following observations were reported: increase in ileum richness and diversity, but decrease in the colon (Yu, 2017), changes in the relative abundance of 5 bacterial genera and 9 bacterial species due to supplementation (Vahjen, 2010 and 2011, respectively). One of the early studies of Zn supplementation showed no changes in number of coliforms using culture dependent techniques when comparing dietary supplementation of 155 ppm ZnO to 2500 ppm Zn in feces. Hojberg, et. al. used culture-dependent methods and targeted sequencing to look at differences in microbes in the stomach, ileum, cecum, and colon fed 100 ppm ZnO or 2500 ppm Zn. They found lower anaerobic bacteria, lactic acid bacteria (LAB), and *Lactobacilli* in the stomach and ileum, but increased coliforms in the stomach and increased *Enterococci* in the stomach and ileum (Hojberg, 2005). The only studies to compare no or <100 ppm ZnO supplementation used culture-dependent techniques

with some targeted sequencing and found that there were no differences in bacterial counts due to Zn supplementation (Broom, 2006) with the exception of *Clostridium* group XIV and *Enterobacteriacea* which only differed between 50 mg/kg ZnO and 2500 mg/kg Zn, and distinct hierarchical clustering between those same Zn categories (Pieper, 2012). Based on these results, 100 ppm Zn may be a sufficient dose to see changes in the microbiome due to supplementation. Thus, pigs in the current study have been sufficiently supplemented with Zn in the control and no changes were seen due to Zn treatment.

Heat Stress

Heat stress exposure had limited impact on the microbial composition of the GI tract in pigs. We found elevated abundance of *Dolichospermum* due to HS in the stomach which are cyanobacteria and have known toxin-producing species found in algae blooms (Li, 2016b), and lowered levels of *Muribaculum* in feces due to chronic HS. Though diarrheal incidence tend to increase with heat stress (Castillo, 2008; Costa, 2014), there is only one study looking at the alteration in microbiome of the GI tract due to heat stress and it was in poultry. Song et. al. (2014) found lower viable counts of *Lactobacillus* and *Bifidobacterium* and higher counts of coliforms and *Clostridium* in contents of the small intestine from chronically heat stressed broilers. These results are not supported by our findings in pigs, but the bacterial community in the intestine of poultry (Pan and Yu, 2014) is very different from that of the pig questioning the appropriateness of the comparison. This leaves conflicting evidence as to whether heat stress induces an environment that promotes the growth of pathogenic bacteria or creates an opening for their proliferation. Thus, more work needs to be done to determine the disease pathway that creates harmful conditions under heat stress.

Konstantinov et. al. (2006) reported that newborn piglets, not undergoing heat stress, have identifiable *Escherichia/Shigella*, *Lactobacillus*, and *Streptococcus* in the small intestine as early as two days old, so there may be risk of increase in those bacteria due to heat stress, particularly at an early age. However, we saw treatment effects on *Actinobacillus*, *Romboutsia*, and *Klebsiella*, and significant changes in Pielou's evennes and Shannon diversity

which decreased due to acute HS. This demonstrates that certain bacteria may be prone to proliferate under heat stress conditions, driving down the evenness and richness of the community and causing dysbiosis.

Differences by GI segment

The bacterial communities detected in segments of the GI tract and feces of the growing pig shared some members but clear segregation of the communities was apparent even across treatments (Figure 1). There was overlap of community structure between that in the colon and feces suggesting that an evaluation of bacteria in feces may represent that of the colon. Zhao et al. (2015) reported correlations between the colon and fecal microbiomes; however, they also detected separation of bacterial community structure from the colon and feces compared to other segments of the GI tract by PCA, providing ambiguous evidence as to the differences (if any) between the colon and feces. In addition, the bacterial community of the stomach differed from that of the ileum in the composition of the most abundant bacteria, and had the presence of cyanobacteria (*Dolichospermum* and *Pseudanabaena*) possibly contributing to clustering away from the ileum. The segregation of colon/feces from the stomach and ileum clearly points to the inability to assess changes in the whole GI microbiome through analysis of feces.

The predominant phyla in the pig GI tract are *Firmicutes* and *Bacteroides*, which agrees with our findings (Costa, 2014; Zhou 2015). Additionally, more types of *γ-proteobacteria* (common pathogens) were found in the ileum and stomach, than in the colon, similar to other findings (Zhao et al. 2015; Isaacson et al. 2012). The most abundant genera detected are typical of those in the literature (Hojberg, 2005; Pajarillo, 2014, 2015; Leser, 2002; Mach 2015, Kim, 2011; Vahjen, 2010) with *Prevotella* being the most dominant among pig GI microbes, as well as *Lactobacillus*, *Clostridium*, *Streptococcus*, *Mitsuokella*, *Neisseria* and *Ruminococcus*. We found similar levels of *Escherichia/Shigella* in the ileum to previous findings, confirming that this is one of the most abundant taxa in that portion of the GI tract (Zhao, 2015; Konstantinov, 2006; Broom, 2006). The distinct genera detected but not pre-

viously reported in pigs were *Dolichospermum*, *Porphyromonas*, and *Pseudanabaena* in the ileum and stomach, and *Muribaculum*, *Treponema*, and *Selenomonas* in the colon and feces (Tables 2-5). These genera are low in abundance in comparison to the aforementioned, but the differences are likely due to a combination of factors including environment, diet, breed (genetics), and care practices, as described in other studies as impacting unique microbiota (Pajarillo, 2014; Doyle, 2017; Lamendella, 2011).

The presence of potential pathogens in the intestinal microbiome has been considered to have immune-modulating capability, preparing the small intestine and the immune system for environmental and food borne pathogens (Allen, 2012). This work is dependent on the advancement of detecting individual bacterial species level data with metagenomics to identify active immunological components. Continued research to identify species level bacteria or toxins present in the GI tract could help to determine whether pathogenic bacteria are contributing to poor outcomes under HS.

Limitations

Clearly, more animals are needed in microbial studies to identify differences in bacterial abundance across multiple targets (GI segment, diet treatment, environmental treatment) particularly if potential interaction of treatments are potentially present. Additionally, the limitations of 454-pyrosequencing in identifying bacteria only to taxonomic rank of genus mean there is no species or toxin potential for clarity to draw conclusions regarding pathogenic or inflammatory risk. Further research should be conducted to elucidate the potential pathogens that increase in presence or pathogenicity due to HS and/or common toxins that may contribute to HS declines in performance.

Conclusion

Overall our findings show that the supplementation of ZnAA to pigs exposed to heat stress did not alter the microbiome of the stomach, ileum, colon, or feces. However, heat stress did cause limited changes in the microbiome at certain locations in the pig GI tract. We found effects of HS along the GIT and in feces: *Dolichospermum* and *Actinobacillus* in the

stomach; *Actinobacillus*, *Romboutsia*, *Klebsiella*, Pielou's evenness, and Shannon diversity in the ileum; and *Ruminococcus* and *Muribaculum* in the feces. There are distinct differences in microbial composition of luminal contents along the pig GIT and feces demonstrating clearly that fecal assessment of bacterial composition may not be an accurate proxy for conclusions about the overall GI microbial composition.

2.6 Acknowledgements

I would like to acknowledge the assistance of the following individuals for their work as co-authors on this paper: Katherine M. Yahvah, William J. Price, Lance H. Baumgard, and Mark A. McGuire.

This work was supported by the Idaho Agricultural Experiment Station.

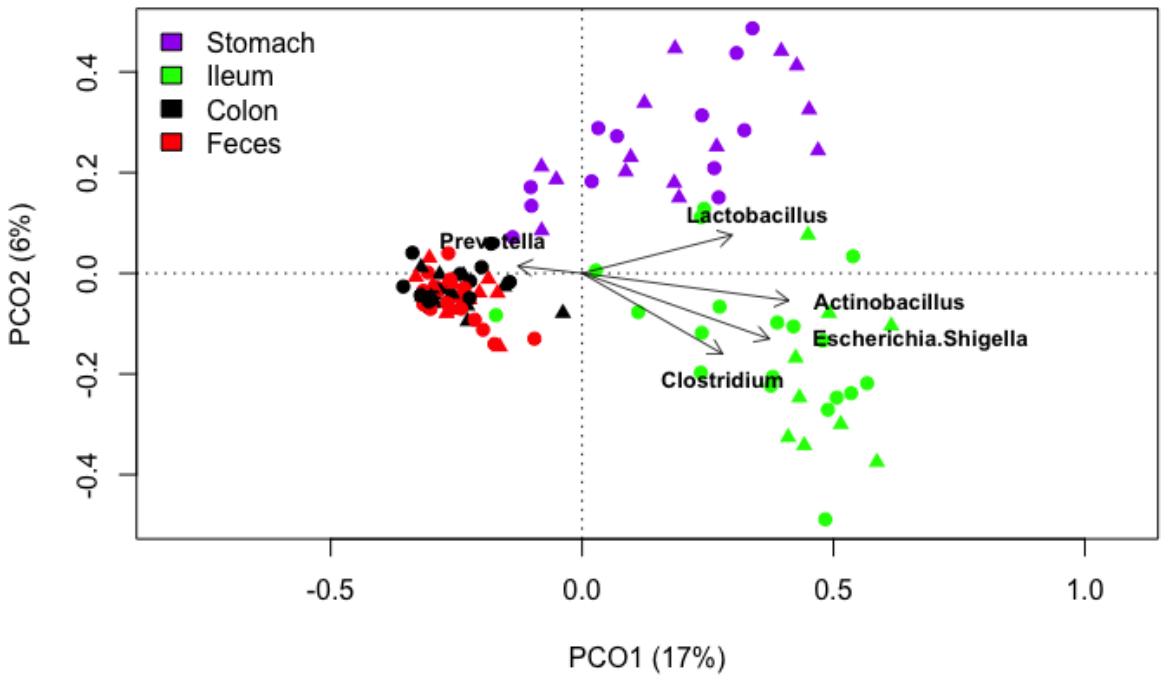


Figure 2.1: Principal coordinates analysis (PCoA) plot of Hellinger transformed genus-level relative abundance of bacteria composition across gastrointestinal segments and in feces of growing pigs. Points are colored by site of sampling. Treatment duration is indicated by shape - circles indicate acute treatment (1-day) and triangles indicate chronic treatment (7-days). Driving bacterial are displayed as vectors in the direction of contribution.

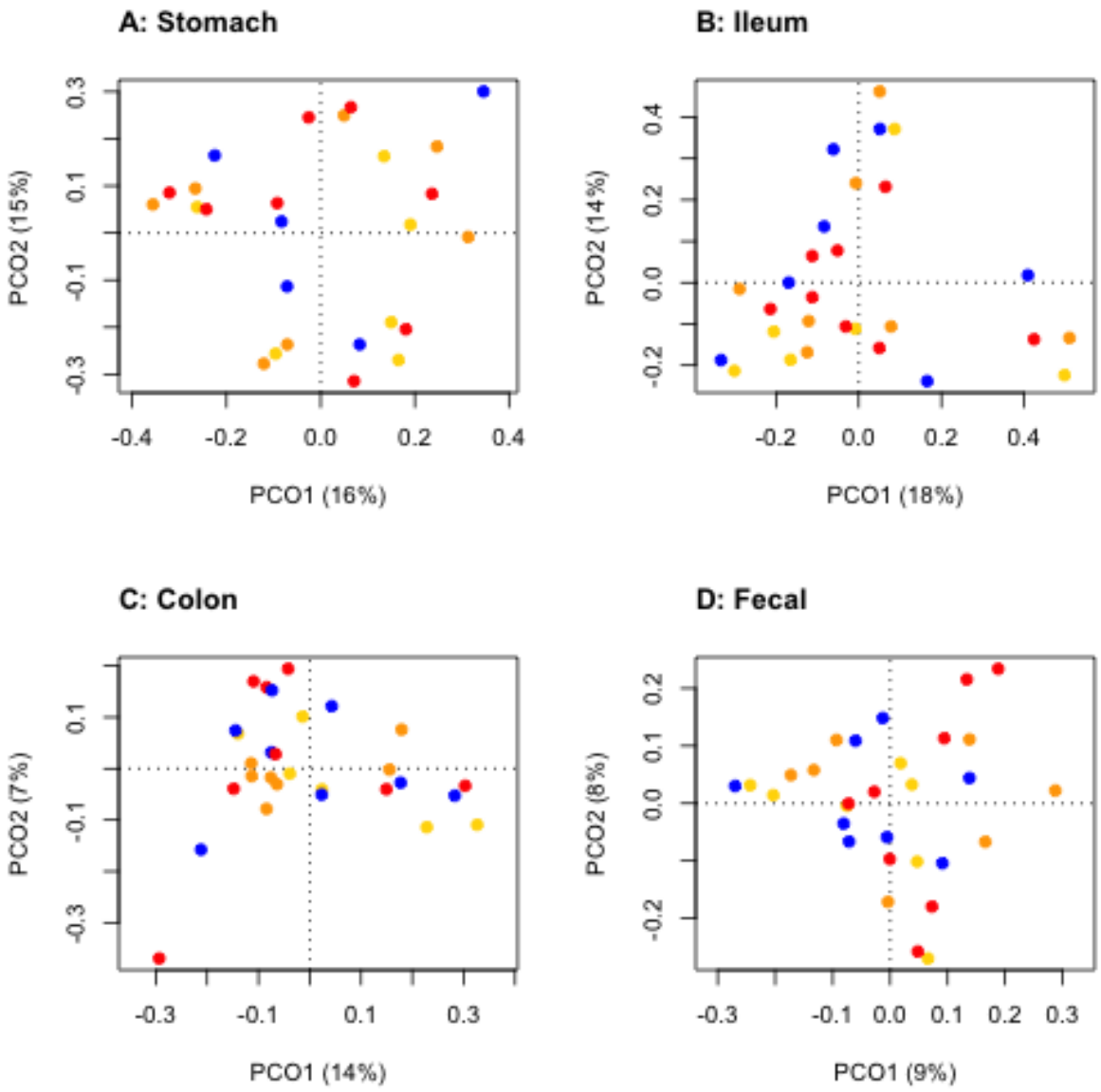


Figure 2.2: Principal coordinates analysis (PCoA) plots of Hellinger transformed genus-level relative abundance of bacteria composition across gastrointestinal segments and in feces of growing pigs: stomach (A), ileum (B), colon (C), and feces (D). Each point represents a single sample. Treatments are identified by color: blue - thermoneutral, Zn-control diet; yellow - Zn-control; orange - Zn-220; and red - Zn-320.

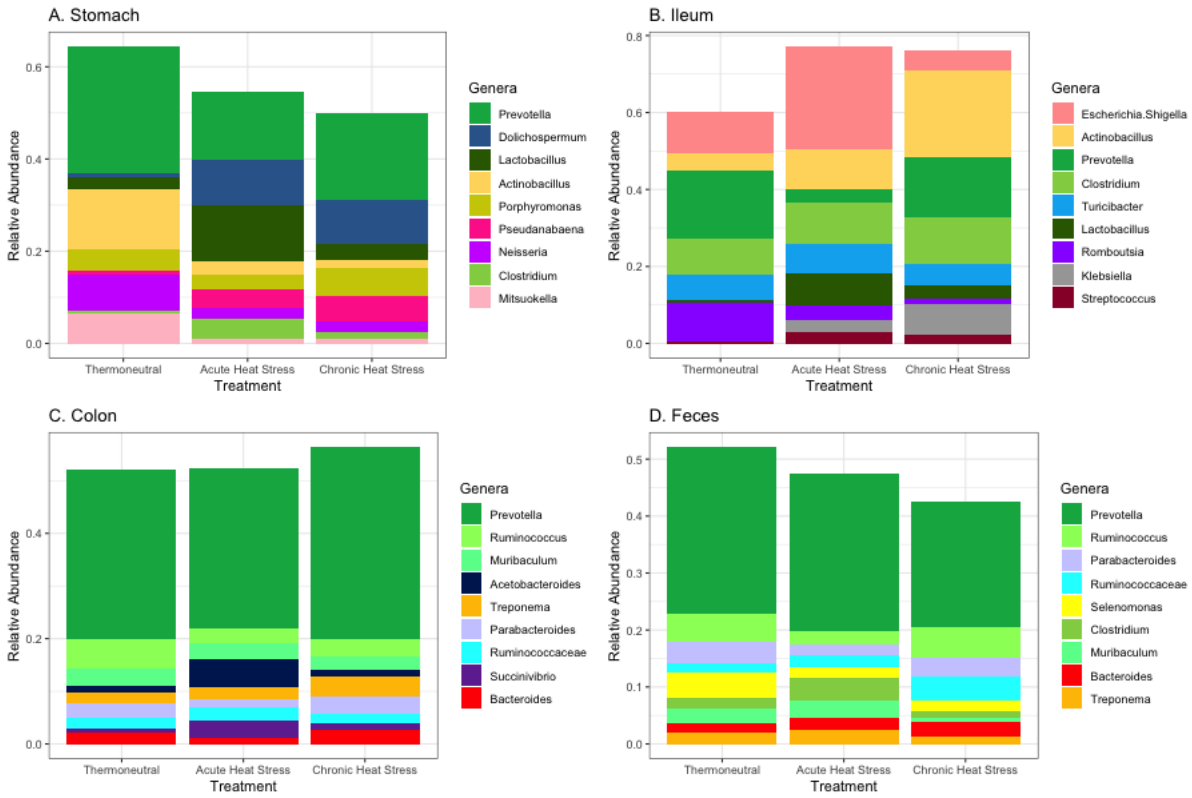


Figure 2.3: Average relative abundance of the 9 most abundant bacterial genera from luminal contents of the (A) stomach, (B) ileum, and (C) colon, and in (D) feces of growing pigs exposed to acute heat stress (1 day), chronic heat stress (7 days), and thermoneutral conditions.

Table 2.1: Treatment descriptions and animal distributions.

Treatment Classification	Dietary Treatment	Environmental Treatment	Day of Sacrifice	Number of pigs (n)
Zn-Control	120 ppm of zinc from ZnSO4	Heat Stress (HS): 36°C; 50% humidity	1	4
	No supplemental zinc	Ad libitum feeding	7	3
Zn-220	120 ppm of zinc from ZnSO4	Heat Stress (HS): 36°C; 50% humidity	1	4
	100 ppm of zinc from Availa Zn 100 ®	Ad libitum feeding	7	3
Zn-320	120 ppm of zinc from ZnSO4	Heat Stress (HS): 36°C; 50% humidity	1	4
	100 ppm of zinc from Availa Zn 100 ®	Ad libitum feeding	7	4
Pair-Fed[1]	120 ppm of zinc from ZnSO4	Thermoneutral (TN): 19°C; 61% humidity	1	4
	No supplemental zinc	Pair-feeding	7	4

[1]Pair-fed samples were matched to feed intake of Zn-control group under heat stress to mitigate differences due to altered feed intake.

Table 2.2: Relative abundance of the top 9 bacterial taxa at lowest identified classification and diversity indices from luminal contents in the stomach of growing pigs exposed to heat stress.[1]

	Thermoneutral (n = 6)	Acute HS (n = 11)	Chronic HS (n = 9)
Bacteria taxa			
Prevotella	0.27 ± 0.03	0.15 ± 0.03	0.19 ± 0.04
Dolichospermum	0.01 ± 0.00^{a}	0.10 ± 0.03^{ab}	0.10 ± 0.01^{b}
Lactobacillus	0.03 ± 0.01	0.12 ± 0.03	0.03 ± 0.01
Actinobacillus	0.13 ± 0.03	0.03 ± 0.01	0.02 ± 0.01
Porphyromonas	0.05 ± 0.02	0.03 ± 0.01	0.06 ± 0.04
Pseudanabaena	0.01 ± 0.00	0.04 ± 0.01	0.05 ± 0.02
Neisseria	0.08 ± 0.02	0.02 ± 0.01	0.02 ± 0.01
Clostridium	0.01 ± 0.00	0.04 ± 0.02	0.02 ± 0.01
Mitsuokella	0.06 ± 0.03	0.01 ± 0.01	0.01 ± 0.00
Diversity indices			
Pielou evenness	0.66 ± 0.02	0.68 ± 0.03	0.71 ± 0.02
Richness	42 ± 4.2	50.2 ± 6.0	56.1 ± 4.7
Shannon diversity	12.5 ± 1.3	17.6 ± 3.0	18.6 ± 2.0
Simpson evenness	0.18 ± 0.01	0.20 ± 0.02	0.17 ± 0.02

[1] *All values are means ± SEMs. Acute HS - acute heat stress for 1 day; Chronic HS - chronic heat stress for 7 days. Values in a row within a sampling location that have different superscripts are significantly different based on Wilcoxon Rank Sum Test. P <0.05 corrected for multiple comparisons.*

Table 2.3: Relative abundance of the top 9 bacterial taxa at lowest identified classification and diversity indices from luminal contents in the ileum of growing pigs exposed to heat stress.[1]

	Thermoneutral	Acute HS	Chronic HS
	(n = 6)	(n = 12)	(n = 10)
Bacterial taxa			
Prevotella	0.18 ± 0.04	0.04 ± 0.01	0.16 ± 0.05
Lactobacillus	0.01 ± 0.00	0.08 ± 0.03	0.03 ± 0.01
Actinobacillus	0.04 ± 0.01	0.10 ± 0.04	0.23 ± 0.04
Clostridium	0.09 ± 0.03	0.11 ± 0.04	0.12 ± 0.03
Escherichia/Shigella	0.11 ± 0.04	0.27 ± 0.08	0.05 ± 0.02
Turicibacter	0.07 ± 0.01	0.08 ± 0.04	0.05 ± 0.02
Romboutsia	0.10 ± 0.02	0.04 ± 0.02	0.02 ± 0.01
Klebsiella	0	0.03 ± 0.02	0.08 ± 0.04
Streptococcus	0.01 ± 0.00	0.03 ± 0.01	0.02 ± 0.01
Diversity indices			
Pielou evenness	0.74 ± 0.02^{a}	0.54 ± 0.04^{b}	0.65 ± 0.03^{ab}
Richness	24.8 ± 2.9	16.6 ± 1.7	22.7 ± 1.4
Shannon diversity	11.1 ± 1.2^{a}	4.9 ± 0.6^{b}	8.2 ± 0.9^{ab}
Simpson evenness	0.28 ± 0.02	0.22 ± 0.02	0.24 ± 0.03

[1] *All values are means ± SEMs. Acute HS - acute heat stress for 1 day; Chronic HS - chronic heat stress for 7 days. Values in a row within a sampling location that have different superscripts are significantly different based on Wilcoxon Rank Sum Test. P <0.05 corrected for multiple comparisons.*

Table 2.4: Relative abundance of the top 9 bacterial taxa at lowest identified classification and diversity indices from luminal contents in the colon of growing pigs exposed to heat stress.[1]

	Thermoneutral	Acute HS	Chronic HS
	(n = 8)	(n = 12)	(n = 9)
Bacterial taxa			
Prevotella	0.32 ± 0.03	0.30 ± 0.05	0.36 ± 0.04
Ruminococcus	0.06 ± 0.01	0.03 ± 0.01	0.03 ± 0.01
Muribaculum	0.03 ± 0.01	0.03 ± 0.01	0.03 ± 0.01
Acetobacteroides	0.01 ± 0.00	0.05 ± 0.02	0.01 ± 0.00
Treponema	0.02 ± 0.01	0.02 ± 0.01	0.04 ± 0.01
Parabacteroides	0.03 ± 0.00	0.01 ± 0.00	0.03 ± 0.00
Ruminococcaceae	0.02 ± 0.00	0.02 ± 0.01	0.02 ± 0.01
Succinivibrio	0.01 ± 0.00	0.03 ± 0.02	0.01 ± 0.00
Bacteroides	0.02 ± 0.01	0.01 ± 0.00	0.03 ± 0.01
Diversity indices			
Pielou evenness	0.73 ± 0.02	0.67 ± 0.03	0.68 ± 0.04
Richness	52.9 ± 2.2	68.2 ± 5.8	49.1 ± 2.2
Shannon diversity	19.5 ± 1.6	21.2 ± 3.5	16.7 ± 1.8
Simpson evenness	0.17 ± 0.02	0.14 ± 0.02	0.16 ± 0.02

[1] *All values are means ± SEMs. Acute HS - acute heat stress for 1 day; Chronic HS - chronic heat stress for 7 days. Values in a row within a sampling location that have different superscripts are significantly different based on Wilcoxon Rank Sum Test. P <0.05 corrected for multiple comparisons.*

Table 2.5: Relative abundance of the top 9 bacterial taxa at lowest identified classification and diversity indices from luminal contents in the feces of growing pigs exposed to heat stress.[1]

	Thermoneutral	Acute HS	Chronic HS
	(n = 8)	(n = 12)	(n = 10)
Bacterial taxa			
Prevotella	0.29 ± 0.02	0.28 ± 0.03	0.22 ± 0.03
Clostridium	0.02 ± 0.01	0.04 ± 0.01	0.01 ± 0.00
Ruminococcus	0.05 ± 0.01	0.02 ± 0.00	0.05 ± 0.01
Muribaculum	0.03 ± 0.00^{a}	0.03 ± 0.00^{a}	0.01 ± 0.00^{b}
Treponema	0.02 ± 0.01	0.03 ± 0.01	0.01 ± 0.00
Parabacteroides	0.04 ± 0.01	0.02 ± 0.01	0.03 ± 0.01
Ruminococcaceae	0.02 ± 0.00	0.02 ± 0.01	0.04 ± 0.01
Bacteroides	0.02 ± 0.00	0.02 ± 0.01	0.02 ± 0.01
Selenomonas	0.04 ± 0.01	0.02 ± 0.01	0.02 ± 0.00
Diversity indices			
Pielou evenness	0.74 ± 0.01	0.71 ± 0.02	0.76 ± 0.02
Richness	48.8 ± 3.1	57.8 ± 2.9	49.5 ± 1.9
Shannon diversity	18.2 ± 1.2	18.2 ± 1.4	19.7 ± 1.1
Simpson evenness	0.19 ± 0.01	0.17 ± 0.02	0.22 ± 0.02

[1] *All values are means ± SEMs. Acute HS - acute heat stress for 1 day; Chronic HS - chronic heat stress for 7 days. Values in a row within a sampling location that have different superscripts are significantly different based on Wilcoxon Rank Sum Test. P <0.05 corrected for multiple comparisons.*

CHAPTER 3

Characterizing the bovine milk microbiome among quarters with variable somatic cell counts

3.1 Abstract

Mastitis continues to be a major cost to the dairy industry with significant animal welfare concerns. Generally, mastitis is treated as a pathogen-driven disease due to a limited set of bacteria. The objective of this study was to survey the microbiome of milk from all four quarters of cows on two dairies. A total of 412 quarters were sampled, somatic cell count (SCC) determined, and DNA isolated and amplified for the V1-V3 region of the 16S rRNA gene to classify the microbial community. Samples were categorized as low SCC (<200,00 cells/mL), mid SCC (200,000-400,00 cells/mL) and high SCC (>400,000 cells/mL). Neither differences in relative abundance nor diversity metrics among quarters were detected, and multivariate analysis found quarters to be more similar within a cow than among cows. Dairy A had significantly less *Staphylococcus* ($P < 0.01$) and was more rich ($P < 0.01$) and even ($P < 0.01$) in community structure than Dairy B. Low-SCC quarters had significantly less *Staphylococcus* ($P < 0.01$) and were more even ($P < 0.01$) and diverse ($P < 0.01$) than mid- or high-SCC quarters. Multivariate analysis confirmed that community structure differed between the dairies, and analysis of similarity identified a difference in community structure between low-, mid-, and high-SCC quarters. These results suggest that milk bacterial community structure is different among low-, mid-, and high-SCC quarters, as well as between dairies.

3.2 Introduction

Mastitis causes production and economic loss among dairy herds (Heikkila, 2003; Contreras, 2011), and is a complex disease caused by multiple, often undiagnosed pathogens, that may

be in dynamic relation to the natural bacteria of the udder (Kuang, 2009; Patel, 2017; Koskinen, 2010; Shome, 2011). In addition to being complex, mastitis can be difficult to treat with the use of antibiotics in the dairy industry due to climbing rates of antibiotic resistance (Ganada, 2016; Bonsiglia, 2011). These complications demand further research to determine the best course of action for bovine health and productivity, particularly in regard to alternative treatments (Crispie, 2008; Jimenez, 2008; Derakhshani, 2018). Untangling the complex relation between intramammary infection and the natural microbiota of the mammary gland depends on learning how the microbiome impacts the presence or type of mastitis in a dairy herd. To further understand this relationship, we must address the following concerns; 1. What are the bacteria present in milk at multiple somatic cell count ranges? 2. Is milk from quarters within a cow similar in microbial composition? Do quarters have the same or increased risk of infection? 3. Does the microbial community structure provide support against pathogen infection, and if so, can we determine a bacterial profile that may prevent mastitis development?

Diagnosing and treating mastitis

The known complexity of the microbial community and pathogenicity has increased with the introduction of next generation sequencing. Mastitis has moved from a cultivable pathogen diagnosis (supplemented by SCC; with approximately 25% of cases uncultivable), to a complex disease which may be more than just driven by host-pathogen interaction (Oikonomou, 2014; Bhatt, 2011). For example, *Streptococcus* is a common commensal milk bacterium in both humans (Hunt, 2011; Jost, 2013, Williams, 2017) and cows (Falentin, 2016) but contains many pathogenic strains which cause mastitis (Cressier, 2011). Pathogen identification itself does not capture every case of mastitis, which indicates a need for deeper understanding of its etiology (Jadhav, 2018; Leitner, 2000; Malik, 2018; Taponen, 2009). This complicates treatment protocols, particularly in terms of the impact of non-antibiotic treatments, such as probiotics, on the commensal milk microbiome (Pellegrino, 2017; Klostermann, 2008; Yu, 2017).

Environmental influence

The Human Microbiome Project established that the presence of commensal bacteria plays an integral role at the interface of host and environment (Turnbaugh, 2007) and provides a variety of health benefits to the host (Tambourini, 2016; Flint, 2012; Stefka, 2014). There is also evidence that 'exposure' to new environments can influence the community composition of the microbiome (Meehan, 2018; Ruiz 2017; Doyle 2017). Particularly related to mastitis, infection may be caused by an environmental pathogen, likely dominating the community of bacteria naturally occurring in the mammary gland (Almeida, 1996; Ganada, 2017). The teat apex and milk have been shown to have distinct microbiomes. However, both locations contain *Staphylococcus*, *Corynebacterium*, and *Ruminococcus* (Zanardi, 2014; Mediano, 2017; Braem, 2014; Falentin, 2016) known to harbor potential pathogens. There are also bacteria found in milk that are known to be in the digestive tract and environment of ruminants such as *Ruminococcaceae* (Oikonomou, 2012; Young 2015; Rodrigues, 2017). Additionally, there is little evidence available to determine if quarters within a cow are independent with respect to the microbiome and inflammation (Kuehn, 2013), and whether quarters play an integral role in development of mastitis (Falentine, 2016). These types of environment-microbiome interactions need further assessment to determine the risk factors of developing mastitis and potential avenues for prevention and successful treatment.

In the present work, we used high-throughput sequencing to analyze the bacterial community structure in milk from mammary glands by quarter of dairy cows with varying levels of inflammation, as indicated by SCC. By determining the potential interplay of quarters and mastitis risk, as well as deploying new multivariate analyses, we hope to contribute to the better understanding of the risk involved in cows contracting mastitis.

3.3 Materials and Methods

Animal Treatment and Approval

All animal work was approved by the Institutional Animal Care and Use Committee at

the University of Idaho before initiation.

Sampling

Milk was collected following the methods of the National Mastitis Council (1987) from cows (n=103) located at two dairies as they were going through the normal milking procedure. Briefly, prior to milking, cows were treated with an iodine-based teat dip. Teat ends were further cleaned with sterile gauze soaked in 70% ethanol, and the first few milliliters of milk discarded. A subsequent 50 mL sample was collected from each quarter, before the continuation of standard milking. All samples were stored at 4 °C until transported back to the University of Idaho laboratory on ice. Somatic cell count was determined using an automatic cell counter (DeLaval, Tumba, Sweden), and aliquots of milk stored at -80 °C for further analysis. Samples were categorized as having low-SCC (<200,00 cells/mL), mid-SCC (200,000-400,00 cells/mL) and high-SCC (>400,000 cells/mL).

DNA Extractions

Milk was thawed at 37 °C in a benchtop hybridization oven until no ice remained in the sample. Samples were thoroughly vortexed and 1 mL of milk was centrifuged at 4000 x g at 4 °C to form a pellet of cellular components. The lipid layer was removed along with the supernatant, and the cell pellet was resuspended in 500 μL buffer (10 mM tris-HCl:50 mM tris-acetate-ethylenediamine tetraacetic acid (EDTA, 0.5 M, pH 8, Sigma, St. Louis, MO): TE50). The suspension was subjected to an enzymatic lysis wherein samples were lysed for 1 hr at 37 °C using an enzyme cocktail consisting of 50 μL lysozyme (10 mg/mL), 6 μL mutanolysin (25,000 U/mL), and 3 μL lysostaphin (4000 U/mL in sodium acetate) in 41 μL TE50. Samples were then subjected to a physical disruption step using 0.1 mm zirconia/silica beads (BioSpec Products, Bartlesville, OK) in a FastPrep FP120 (Qbiogene, Carlsbad, CA) set at 5 for 2 x 30 sec. DNA was then extracted from the cell lysate using the QIAamp® DNA Mini Kit (Qiagen catalog no. 51304, Valencia, CA) according to manufacturer's protocol. Samples were eluted in 50 μL nuclease free water and stored at -80 °C until further processing.

PCR Amplification

Microbial DNA was amplified using seven-fold degenerate universal primers targeting the V1-V3 region of the 16S rRNA gene (Williams et al, 2017). The initial PCR mixture contained 0.05 μM target-specific primers (Integrated DNA Technologies, Coralville, IA), 5 μL 10X 360 PCR buffer (Life Technologies, Carlsbad, CA), 3 mM MgCl2 (Life Technologies), 0.24 mg/mL bovine serum albumin (Sigma, St. Louis, MO), 0.2 mM dNTP (Life Technologies), 0.255% DMSO, 0.05 U/μL AmpliTaq® DNA 360 polymerase (Life Technologies), nuclease-free water (Ambion, City, State), and 10 μL of DNA template in a total volume of 50 μL Due to the low biomass nature of samples, some samples (n = 161) were unable to be amplified with this protocol.These samples were amplified using Q5 2X Hot Start High Fidelity Master Mix (New England Biolabs, Ipswich, MA), the same seven-fold degenerate primers targeting V1-V3, and 5 μL of DNA template in a 25 μL reaction. PCR was conducted using either an Applied Biosystems 2720, Veriti, or ProFlex model thermocycler under the following conditions: initial denaturation for 2 min at 95 °C followed by 20 cycles of 95 °C for 20 sec, 60 °C for 30 sec, and 72 °C for 1 min with a 0.5 °C step-down in the annealing temp each cycle, an additional 20 cycles at 95 °C for 20 sec, 50 °C for 30 sec, and 72 °C for 1 min, and a final extension step of 72 °C for 5 min. Samples were held at 4 °C in the thermocycler until being stored at -20 °C. Products from the PCR were electrophoresed on 1% agarose gels made with tris-acetate-ethylenediamine tetraacetic acid (TAE; 40 mM Tris, 20 mM acetic acid, 1 mM EDTA) buffer and containing ethidium bromide (0.0007 mg/mL). Gels were allowed to electrophorese for 30 min at 80 V, and PCR products viewed using the BioRad UltraCam Digital Imaging System (Hercules, CA). Samples with high-quality amplicons (relatively bright band of interest at 534 bp), low primer-dimers, and absence of unwanted bands or smears were deemed acceptable for a second PCR reaction. First PCR products were diluted 1:100 with nuclease-free water, and 4 μL of DNA template were added to a second round of PCR in a reaction mix containing 75 nM primers with dual-index barcodes and Illumina sequencing adapters obtained from the University of Idaho's Institute

for Bioinformatics and Evolutionary Studies Genomics Resources Core (IBEST GRC), Q5 2X Hot Start High Fidelity Master Mix (NewEngland Biolabs) and nuclease free water to a volume of 20 μL. PCR was conducted using an Applied Biosystems 2720 thermocycler under the following conditions: initial denaturation for 30 sec at 98 °C, followed by 15 cycles of 98 °C for 10 sec, 60 °C for 20 sec, 72 °C for 20 sec, and a final extension step of 72 °C for 2 min. Samples were held at 4 °C in the thermocycler until being stored at -20 °C. Quality of 2nd PCR amplicons was evaluated using a QIAxcel DNA screening cartridge (Qiagen), and DNA quantified using the Qubit double stranded DNA high sensitivity assay (Life Technologies).

Sequencing

An appropriate volume of each amplicon (containing 50 ng DNA) was pooled to create a composite sample for high-throughput multiplexed sequencing and submitted to the IBEST GRC for further processing and sequencing. Amplicon pools were size-selected using AM-Pure beads (Beckman Coulter, Indianapolis, IN) and processed on a DNA 1000 chip for the Agilent 2100 Bioanalyzer. The cleaned amplicon pool was quantified using the KAPA Illumina quantification kit (KAPA Biosciences, Wilmington, MA) and the Applied Biosystems StepOne Plus real-time PCR system. Sequences were obtained using an Illumina MiSeq (San Diego, CA) v3 paired-end 300-bp protocol for 600 cycles.

Data Processing and Statistics

Raw sequences were demultiplexed using a custom python application dbcAmplicons (https://github.com/msettles/dbcAmplicons) where reads were identified by expected primer sequence and barcode as described previously (Williams et al. 2017). Reads were subsequently split by sample for further filtering using a custom python script. Reads were then processed using the DADA2 package (version 1.6.0; Callahan, 2016) using a trim length of 270 bp, a max error rate of 30 (11%); parameters to filter reads using only the read 1 sequencing data (27-290 bp of 16S; V1-V2), and taxonomic assignment used the SILVA database (version 128). Note, read 1 was used exculsively due to an inability to merge forward and reverse reads after quality filtering. Unique sequences resulting from filtering using

DADA2 were further aggregated to the genus taxonomic level using phyloseq (version1.22.3) and processed to represent the most abundant (>0.001%) of the overall data, and all genera representing <0.001% of the data were condensed into an "other" category.

Aggregation and summary statistics were performed in phyloseq (version 1.22.3) in the R environment (version 3.4.2; R Core Team, 2016); one genus *Ruminococcacea_UCG-005* was identified in the Silva database, but has not been assigned a genus level name and will be referred to simply as *Ruminococcacea*. Table statistics were calculated in SAS (version 9.4) using PROC MIXED with dairy, SCC category, and quarter as dependent variables, and Bonferonni adjusted values used to declare significance at P <0.05. Principal componenet analysis (PCA) plots were made including all genera using ggplot2 package (version 2.2.1). NMF analysis was performed using the NMF package with factorization rank 20, 100 runs, under the Brunet algorithm (version 0.21.0). Additive main effects and multiplicative interaction (AMMi) models of genotype-by-environment (GxE) interactions of bacteria and environments were performed using both SAS (version 9.4) robustreg procedure and R packages rlm (version 1.2) and pcaMethods (version 1.70.0) (methods adapted from Rodrigues, 2015). The final biplots were made using SAS sgplot procedure. Chi square, ADONIS (PERMANOVA), and ANOSIM analyses were run in the R environment. Chi square was run on presence or absence of bacterial genera, ADONIS was run with dairy x quarter x SCC category interactions in the model, and ANOSIM was run with a Bray-Curtis dissimilarity on each variable.

3.4 Results

General disposition of the data

Initially, 14,479,995 sequence reads (average: 34,807 reads/sample; range: 60-510,748 reads/sample) were generated through dbcAmplicons. Further processing using DADA2 (Callahan, 2016) yielded 10,800,318 reads (average: 25,962 reads/sample; range: 23-427,149 reads/sample). Samples with fewer than 1000 reads were eliminated from the study providing

relative abundances in milk from 400 quarters for further analysis. Unique sequence variants which could not be classified to the kingdom "bacteria" and were "uncharacterized" at the phylum level were removed, and unique sequences were collapsed by their genus assignment using the SILVA database (version 128); 18,473 unique sequences were unidentifiable at the genera level and were removed from further analysis. The agglomeration step yielded 925 unique genera. At the genus level, bacterial units which represented less than 0.1% of the total data were removed from the analysis leaving 596 unique genera. These genera were further reduced to the top 59 most abundant (genera summed across all samples made up for >1% of the total), with the remaining genera collapsed into a composite category "other".

Relative Abundances and Diversity

Staphylococcus (12.5%), *Ruminococcaceae* (7.3%), *Corynebacterium* (6.1%), *Turicibacter* (4.7%), and *Intestinibacter* (4.4%) made up the top 5 most abundant taxa overall. Several differences were found between dairies and among SCC categories. Milk from cows in dairy A had less *Staphylococcus* ($P < 0.01$) and *Ruminococcaceae* ($P < 0.01$), and more *Tiricibacter* ($P < 0.01$), *Clostridium sensu stricto* ($P < 0.05$), *Pseudomonas* ($P < 0.01$), *Paeniclostridium* ($P < 0.01$), and "other" bacteria ($P < 0.01$) than milk from cows in dairy B (Table 1). Milk from cows in dairy A had a richer microbial composition ($P < 0.01$), being more even ($P < 0.01$) and more diverse ($P < 0.01$) than milk from cows in dairy B (Table 1). Milk with low-SCC had less *Staphylococcus* than milk produced by either mid- or high-SCC quarters ($P < 0.01$) (Table 2). Milk with low-SCC also had significantly higher abundance of *Ruminococcaceae* ($P < 0.01$) and *Atopostipes* ($P < 0.05$) than milk with high SCC, but milk with mid SCC was not different from milk with low or high SCC (Table 2). Milk with low SCC was also richer than milk with high SCC ($P < 0.01$), more even (Pielou's) ($P < 0.01$), and more diverse ($P < 0.01$) than milk with mid or high SCC quarters, with the exception of Simpson Evenness which did not different among categories (Table 2). No differences were found in the relative abundance of any bacteria taxa or any diversity index by quarter (Table 3).

Chi square, ANOSIM, and ADONIS

Contingency tables of bacterial presence/absence were used to look at differences via chi-square test across the categories of data. Based on presence/absence, bacterial communities in milk were different (P <0.001) in structure while bacterial communities among quarters and SCC categories were not different (P = 1 and P = 0.9864, respectively). Community differences between dairies (P <0.001), among SCC categories (P <0.001), the interaction between dairy and SCC category (P <0.002), and the interaction between quarter and SCC category (P <0.042) were declared when a relative abundance (as opposed to presence/absence) ADONIS (PERMANOVA) multivariate approach based on a Bray-Curtis dissimilarity was used. A pairwise post hoc test was applied to determine differences between the 3 SCC categories which revealed that the community structure of milk with low SCC was different from the mid and high SCC categories (P <0.015), but the composition of the bacterial community in milk classified as mid and high SCC were not different (Figure 1). Analysis of similarity (ANOSIM) also showed community structure differences by asking if the community was more similar within a given category than compared to the whole set of quarters. The ANOSIM approach determined that microbial communities in milk from a given quarter were most similar within a cow (R = 0.4538, p = 0.001), more similar within a dairy (R = 0.3931, p = 0.001), not different among quarters (R = -0.0048, p = 0.946), and more similar within a SCC category (R = 0.2335, p = 0.001). This provides further evidence that some community differences exist based on animal, dairy, and SCC category.

NMF

Non-negative matrix factorization (NMF) was employed to see if there were distinct differences driving the pattern of data (Brunnet, 2004; Gaujoux, 2010). This analysis looks for patterns in the data by reducing the matrix dimensions and determining the probability of contribution to the data structure, which allows a look beyond the relative abundances of the bacteria. A NMF score closer to one indicates higher probability it has contributed to the data structure. Based on the results from ADONIS and ANOSIM, NMF helped to

discover bacterial taxa that may differ in contribution to the categorical variables. The overall data structure was strongly influenced by *Streptococcus*, *Bifidobacterium*, *Serratia*, *Janibacter*, *Sphingomonas*, and *Clostridium sensu stricto*. The similarities across dairies showed that *Kocuria*, *Aerococcus*, *Staphylococcus*, and *Faecalibacterium* (score = 0.95 - 1) contributed strongly to each dairy, and to the overall community structure. The genera that contributed equally strongly (0.95-1) to the data differed in the following ways: dairy A had the least number of genera that contributed at this level to the data structure; which included *Sphingomonas* and *Corynebacterium*; dairy B included *Ornithinimicrobium*, *Enterococcus*, *Leucobacter*, *Aequorivita*, *Glutamicibacter*, *Streptococcus*, and *Serratia* (Table 4). The SCC category showed more variability than the dairy category, with low SCC having nine genera contributing at scores equal to one, the highest ranked genera contributing to the variability in milk at the mid SCC was *Enterococcus* (0.975), and the high SCC category had only 2 genera contributing strongly to the data (Table 5). The quarter data look very similar, with the top contributors having 6 genera in common across quarters. Small variations among quarter data structure did exist, particularly *Escherichia/Shigella* in the right front quarter which could be due to a single quarter outlier contained only detected *Escherichia/Shigella* in that sample and was categorized as high SCC (Table 6). Notably, many of these genera were not found in the most abundant taxa but play a role in the community composition of the data. Heat maps showing the differences in contribution across these 3 categorical variables are available in Figures 2-4.

PCA and AMMi

To further identify community differences between categorical variables, principal component analysis (PCA) was used to see if there were any groupings by categories. Figures 5 and 6 show that the only grouping in the data was by dairy, with a very strong dairy effect. Grouping by SCC category or quarter lack clear distinction (Figures 5 and 6). Further, SCC does not cluster by dairy. Because the relative abundance may not capture the important differences in the data, a robust genome by environment type test was employed, additive

main effects and multiplicative interaction (AMMi) models. These models look at the variation in the data directly and then plot the variation of samples along with the bacterial contribution to the axes. The AMMi results support a very strong dairy effect (clustering by dairy; Figure 7), and no clustering by SCC category or quarter (Figure B and C, respectively). Again, no clustering was noted based on SCC category when each dairy was plotted (Figure 8).

3.5 Discussion

The purpose of the current study was to survey the microbial community in milk from quarters of dairy cows with varying SCC across two dairies. These data provide a survey of the bacteria present in mostly healthy quarters and show that microbial composition is related to inflammation and environment.

Environment/Dairy

There were differences in the milk microbiome and diversity indices between dairy A and dairy B; likely due to differences in environment, milking practices, animal husbandry, and/or genetics (Figure 1). Dairies clustered separately when examining the microbiome of healthy quarters using discriminant analysis similar to what others have found (Oikonomous, 2014). The genera driving the differences between dairies were *Glutamicibacter*, (found in cheese cultures, Jannala, 2018) *Janibacter*, *Ornithinimicrobium* (both previously reported in the raw milk microbiome, Raats, 2011; Li, 2018), and *Bifidobacterium* (a common microbe used as a probiotic and found in milk; Gueimonde, 2007). Differences to this degree have been shown, most notably, between dairy breeds; Kankrej, Gir, and crossbred *Bos taurus* x *B. indicus* (Bhatt et al., 2012) where different breeds produced milk which only shared ∼12% of bacteria. Bhatt et al. (2012) evaluated their sequences at the family level and saw distinct differences between breeds in *Bacillaceae*, *Staphylococcaceae*, *Streptococcaceae*, and *Enterococcaceae*, which contain many of the common genera in the bovine milk microbiome. Only Holstein cows were sampled on the two dairies in the current study, eliminating the

effect of breed as a part of the differences in milk microbiome between dairies. Environmental aspects from milking procedures, to feed, housing, and animal care practices may explain some of the differences found between our locations (Doyle, 2017). In addition to the potential environmental differences that may distinguish predominant bacterial genera, the history of mastitis may determine the community structure of the milk microbiome. While historical data were not collected in the present study, Falentin et al. (2016) found distinct clustering of the teat microbiome by PCA of healthy (defined as never having mastitis) from infected quarters. This demonstrates that the environment (milking practices, environmental exposure, feeding practices, disease history etc.) may play a large role in the community structure of milk from dairy cows. Historical data about inflammation may play a larger role in diagnosing and treating mastitis in the future.

Quarters and SCC

Overall, the most dominant bacteria across all milk samples were *Staphylococcus*, *Ruminococcaceae*, and *Corynebacterium*, all of which have some association with mastitis (Kuang, 2009; Rodrigues, 2017; Metzger, 2018). These genera have been shown to be key players in the milk microbiome of both healthy humans and cows (Park, 2007; Hunt, 2011; Oikonomou, 2014; Quigley, 2013). *Staphylococcus* became a larger percentage of the relative abundance of the bacterial community in milk as SCC increased, supporting significant knowledge about the role of *Staphylococcus* in mastitis (Delgado, 2009; Contreras, 2011; Almieda, 1996). In addition, the increase in *Staphylococcus* was associated with decreases in other bacterial abundance (Table 2). These changes could reflect simply the analysis of relative abundance or suggest that *Staphylococcus* can out-compete the other genera to establish a more biased community. Ma et al. (2015) have suggested a potential partnership of *Staphylococcus* and *Corynebacterium* in human milk that can lead to dysbiosis. These genera appear to have an inverse relationship relative to SCC category (higher *Staphylococcus* and lower *Corynebacterium* in mid and high SCC quarters) which is likely decreasing based on the increase in *Staphylococcus* and the nature of relative abundance data. Thus, the potential partnership

noted by Ma et al. (2015) is not supported. The presence of *Rumicococcaceae* which are commonly found in the digestive tract of ruminants (Lima, 2017), shows both the potential for environmental microbes to play a role in the milk microbiome, but may also be supportive of the enteromammary pathway, described in humans and mice (Young, 2015; Perez, 2007; Fernandez, 2013), though we found lower relative abundance at high SCC (Table 2). Contamination during milking and storage cannot be ruled out, though sampling milk by needle aspiration of the teat cistern rather than expression through the teat canal still shows a diverse microbial community in milk (McGuire, 2017a). Most studies to date (Ganda, 2017; Kuehn, 2013) have looked exclusively at healthy versus inflamed quarters, at most pairing these quarters in a cow, but have never looked at the potential independence of all four mammary glands (quarters) within a cow. Our direct analysis showed that milk across quarters were remarkably similar in bacterial composition with no significant taxa or diversity differences, even when one or multiple quarters had increased SCC (Table 3), and ANOSIM determined that quarters were more similar within a cow than among cows. This contrasts with previous findings where quarters were not necessarily more similar within a cow, especially in the case of inflamed quarters (Falentin, 2016). Previous findings also suggest that milk from mastitic quarters showed a predominance of a single bacterial pathogen, which accounts for the low diversity (Oikonomous, 2014), although subclinical mastitis may be the result of multiple pathogens which may limit changes in diversity or evenness (Bhatt, 2011). Our results generally support a limit with richness, diversity, and evenness consistently decreasing with elevated SCC indicating a less complex community associated with inflammation (Table 2). The highest diversity was in milk with low SCC in support of the general thought that greater diversity is associated with increased health or positive outcomes (Braem, 2012).

Analysis

Principal component analysis (PCA), principal coordinate analysis (PCoA), and discriminant analysis have all been leveraged to summarize the complexity of microbial composition

data by looking at the variation between samples. Specifically, these analyses have distinguished variation between healthy and mastitic quarters based on the microbial variation, and particularly that culture negative mastitic samples look distinct from others (Falentin, 2016; Khuen, 2013; Oikonomou, 2014; Lima, 2018). The data in the present study do not show clustering due to SCC level (as a proxy for mastitis) using PCA and AMMi analysis. The majority (360 out of 400 samples) of milk from quarters in our study were low in SCC, which may limit the ability to detect differences. A limitation of these clustering analyses are that they distill the data to an understandable 2-dimensional picture, based on variation which is often driven by the most abundant taxa. We explored two methods from the genetics field in order to determine if there were less abundant taxa playing key roles in the community structure of the milk microbiome, which can be highly important to function (e.g., oxalate degredation in woodrat populations (Ridenhour, 2017)). Non-negative matrix factorization (NMF) was employed which looks for patterns in high dimensional data but is independent of the relative abundance of the bacterial taxa. The purpose of this strategy is to decompose the data matrix into components and determine their potential probability of contribution to the underlying variability structure (Brunett, 2004; Shafii, 2017; McGuire, 2017b; Cai, 2017; Rodriguez, 2016). The results from this analysis show contributions from lower abundance bacteria, that create unique group distinctions, as well as show us which bacteria may contribute to milk bacteria structure overall. For example, *Kocuria* had high NMF scores in both dairies and at all SCC levels, but only makes up on average 1-2% of any of those communities, and though no known mastitis causing species are reported, certain species have been implicated in cholocystitis (Ma, 2005) and urinary tract infections (Chen, 2015). Similarly, *Serretia* makes up $\sim$ 1% of average milk microbial composition with low SCC but only 0.07% of milk microbial composition with mid or high SCC and has been reported as an opportunistic environmental pathogen (Zadoks, 2011). Though there is little known about the potential role for these microbes, they do appear to play a role in distinguishing milk microbial community structure. Additionally, the bacteria contributing with

the highest probability to the data structure across quarters is highly consistent, despite SCC variability, providing further evidence that quarters are very similar across cows (Table 6). Additionally, we adapted genotype by environment interaction models to look at the multivariate structure of the microbial communities. Additive main effects multiplicative interaction models (AMMi) decompose the environmental factors (such as sample type, location, etc.) and genetics (bacterial taxa) matrix into smaller contributing parts (Rodriguez, 2016). Again, this helps to eliminate the dependence on overly abundant data. We found that the strongest clustering was by dairy (Figure 7), further supporting environmental factors playing a key role in the microbial composition of milk from dairy cows based on their location and care.

Limitations

Despite the distinct differences by dairy in the bacterial community composition of milk, only sampling two locations limits the extent to which we can assume environment plays a key role in microbial community structure. We sampled all quarters within cows without discrimination of presence or absence of inflammation providing data skewed toward healthy milk (<200,000 cells/mL). This limits the conclusions we can make based on the lower number of inflamed quarters. Future work needs to sample more dairies, using the same techniques across samples, and look at the same statistical tests to determine true environmental/location differences. Other studies have also shown strong evidence of pathogenicity in mastitis by presenting the culture dependent results alongside the culture independent results, and this evidence could have been useful in determining a category of high SCC culture negative milk samples to corroborate previous studies (Taponen, 2009; Koskinen, 2009).

Conclusion

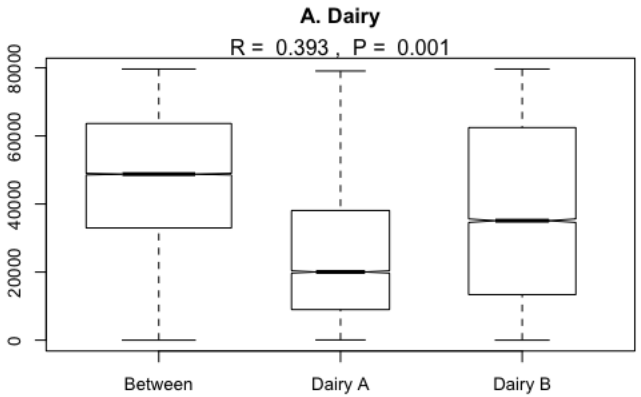
There continues to be a need for methods of treatment and prevention of mastitis through more knowledge. Many milk samples from cows with symptoms of mastitis fail to have an identifiable pathogen. Thus, more work needs to be done in the areas of multi-omics

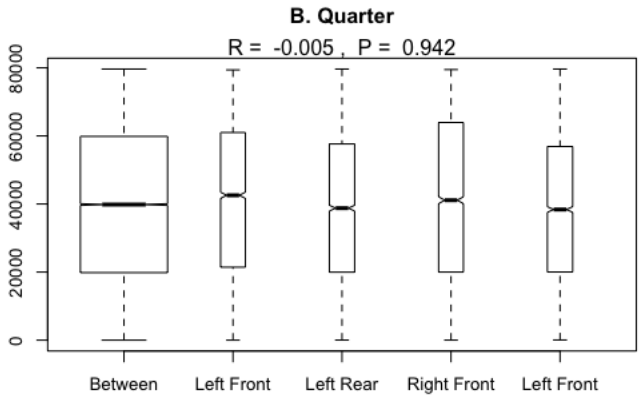
approaches (microbiome, metabolomics, proteomics, etc.) to determine the causative agents of intramammary inflammation. We determined that quarters are more similar within a cow than among cows and that environment and SCC may contribute to changes in diversity metrics and microbial constituents. Further research needs to be done in order to determine the active etiological agents that cause mastitis, and the microbiome may play an important role in both prevention and treatment of mastitis.

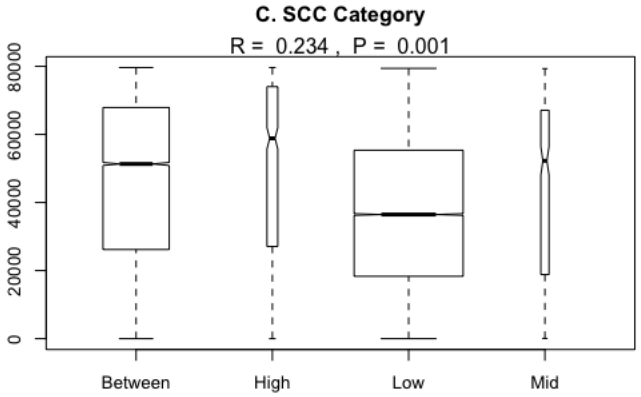
3.6 Acknowledgements

I would like to acknowledge the assistance of the following individuals for their work as co-authors on this paper: Janet E. Williams, Katherine M. Yahvah, Brittany A. Casperson, William J. Price, Juliette K. Tinker, Mark A. McGuire.

This work was supported by the Idaho State Board of Education HERC Incubation Fund (#IF13-006), USDA National Institute of Food and Agriculture (2013-01189), and the Idaho Agricultural Experiment Station.







D. Pairwise SCC

	High-SCC	Low-SCC
Low-SCC	0.0015	NA
Mid-SCC	0.575	0.0015

Figure 3.1: Analysis of similarity (ANOSIM) results showing the variation between groups ("Between" bar in each plot) and within group variation: (A) dairy, (B) quarter, (C) SCC category, and (D) the posthoc difference among the SCC Categories.

Figure 3.2: Heatmap of non-negative matrix factorization (NMF) contribution probabilities by dairy.

Figure 3.3: Heatmap of non-negative matrix factorization (NMF) contribution probabilities by SCC category.

Figure 3.4: . Heatmap of non-negative matrix factorization (NMF) contribution probabilities by quarter.

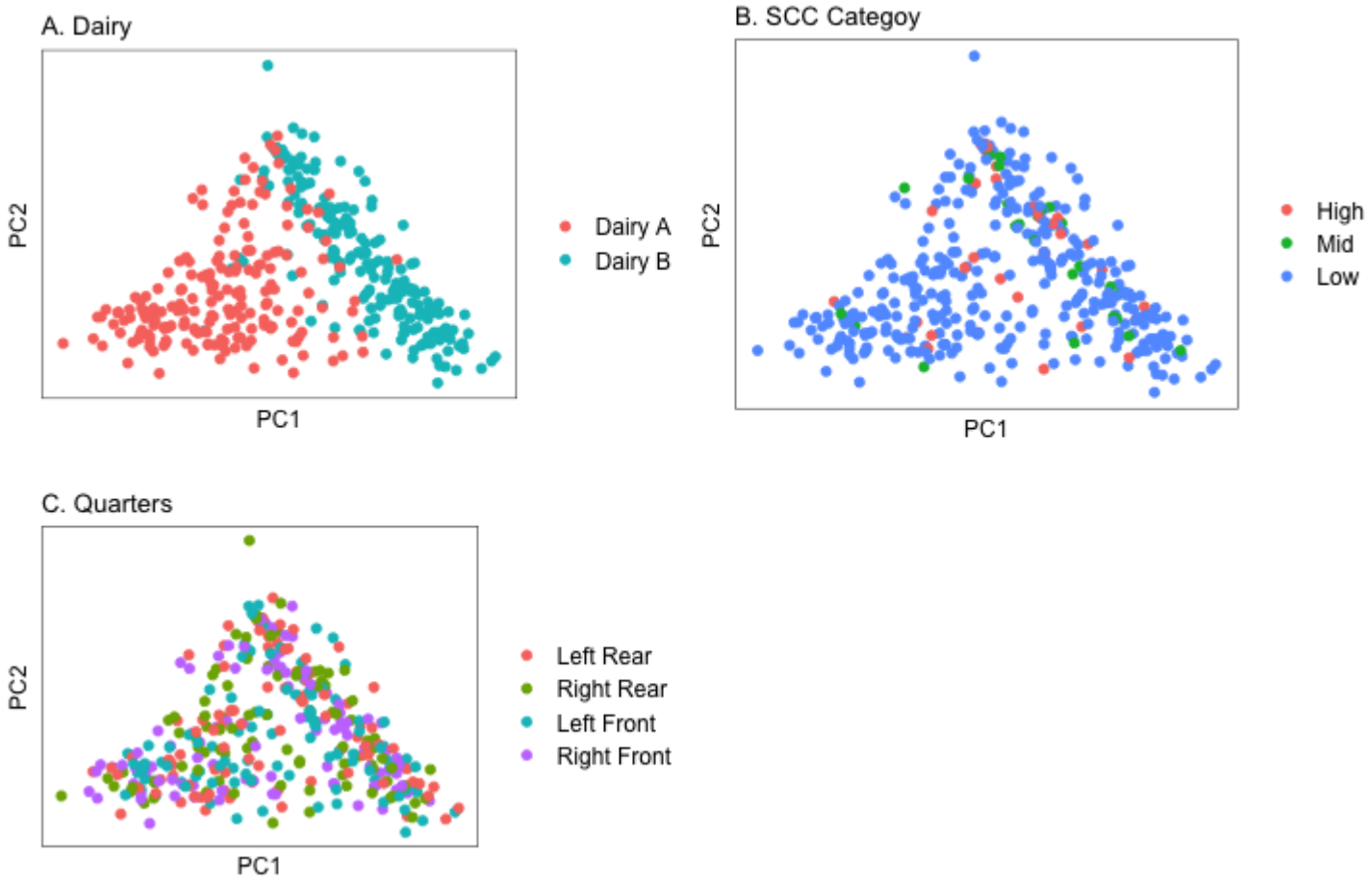


Figure 3.5: Principal component analysis (PCA) results of relative microbial abundance (PC1 and PC2), where data are colored by (A) dairy, (B) SCC Category, and (C) Quarter.

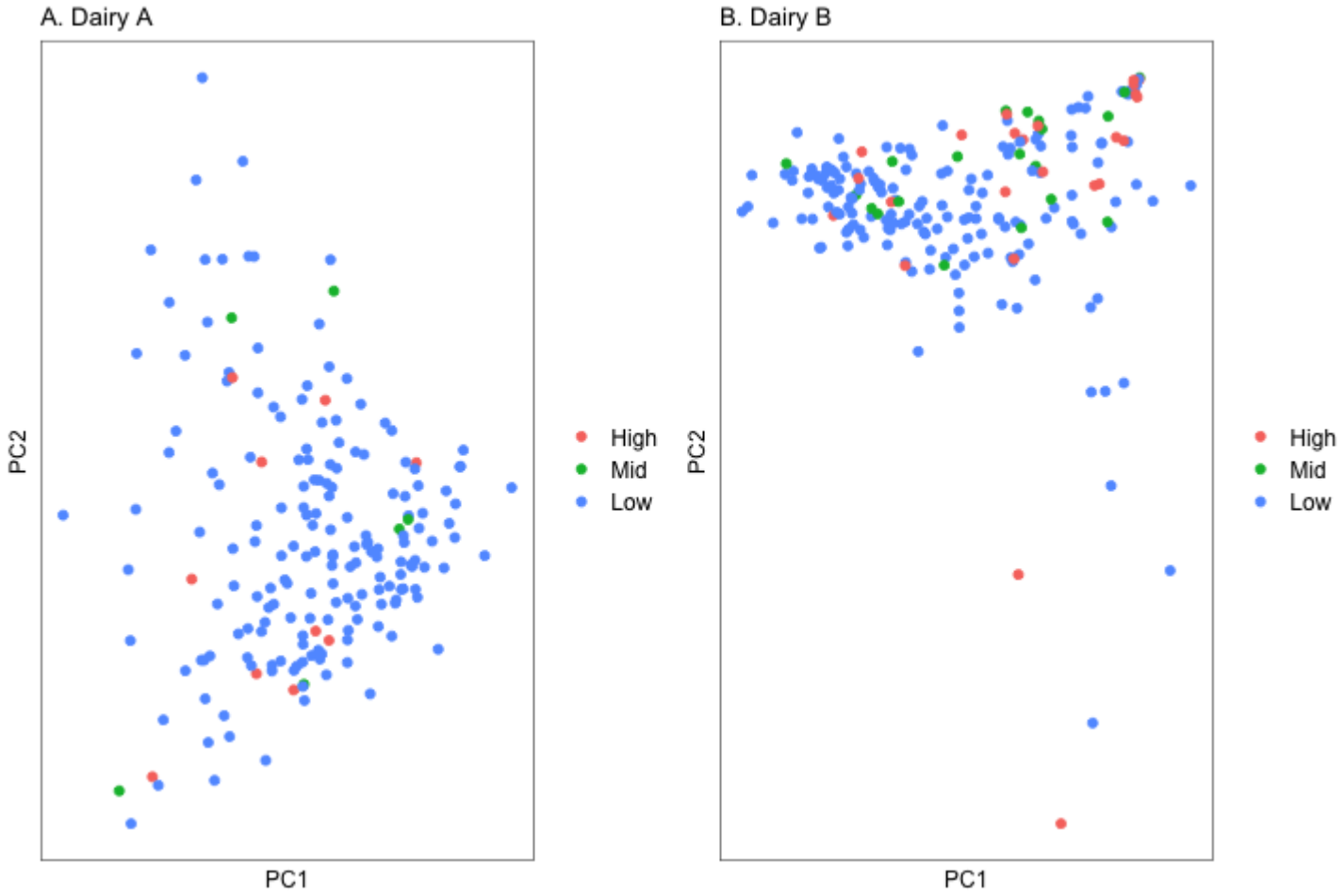


Figure 3.6: Principle component analysis (PCA) results of relative microbial abundance (PC1 and PC2) among individual dairies, where data are colored by SCC category: (A) dairy A and (B) dairy B.

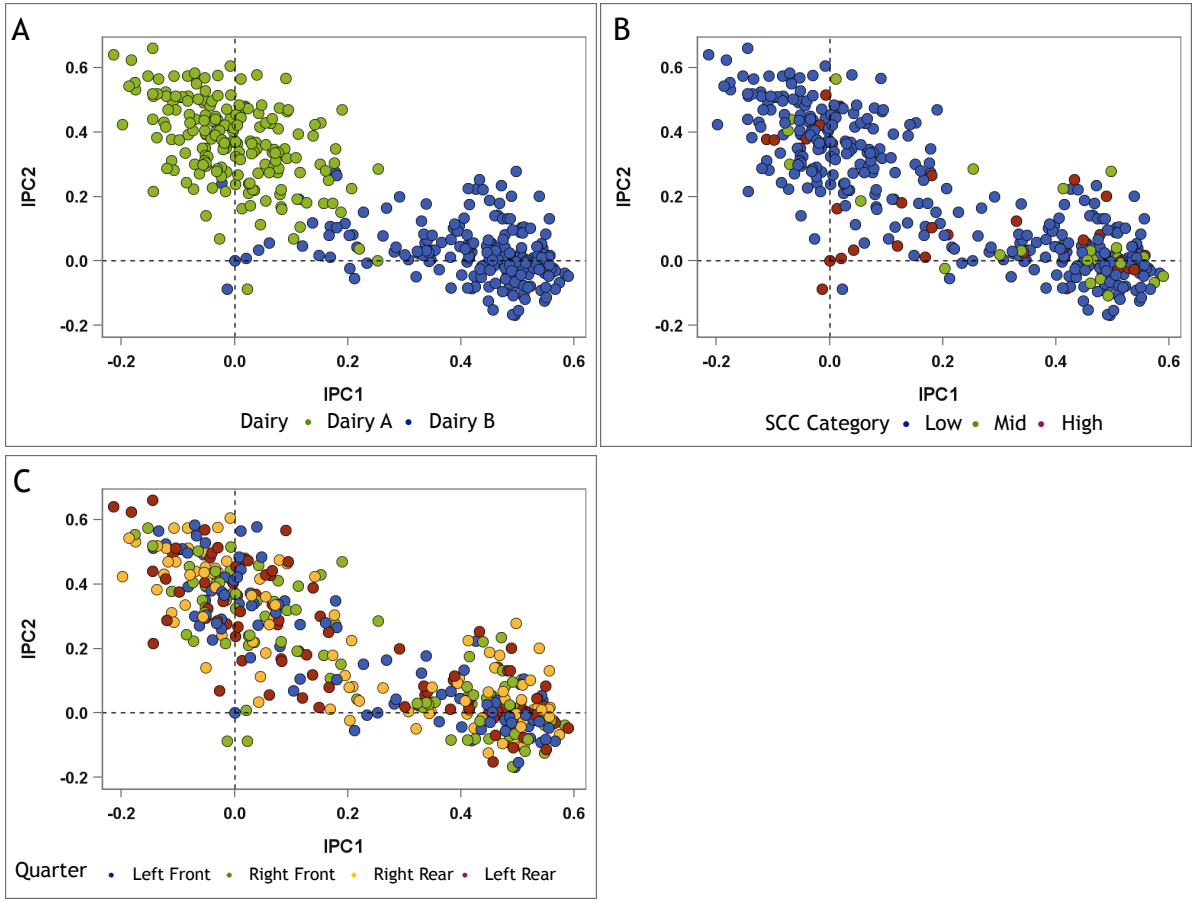


Figure 3.7: Additive main effects and multiplicative interaction (AMMi) model results of relative microbial abundance, where axes are equivalent to principal components 1 and 2, and data colored by (A) dairy, (B) SCC category, (C) quarter.

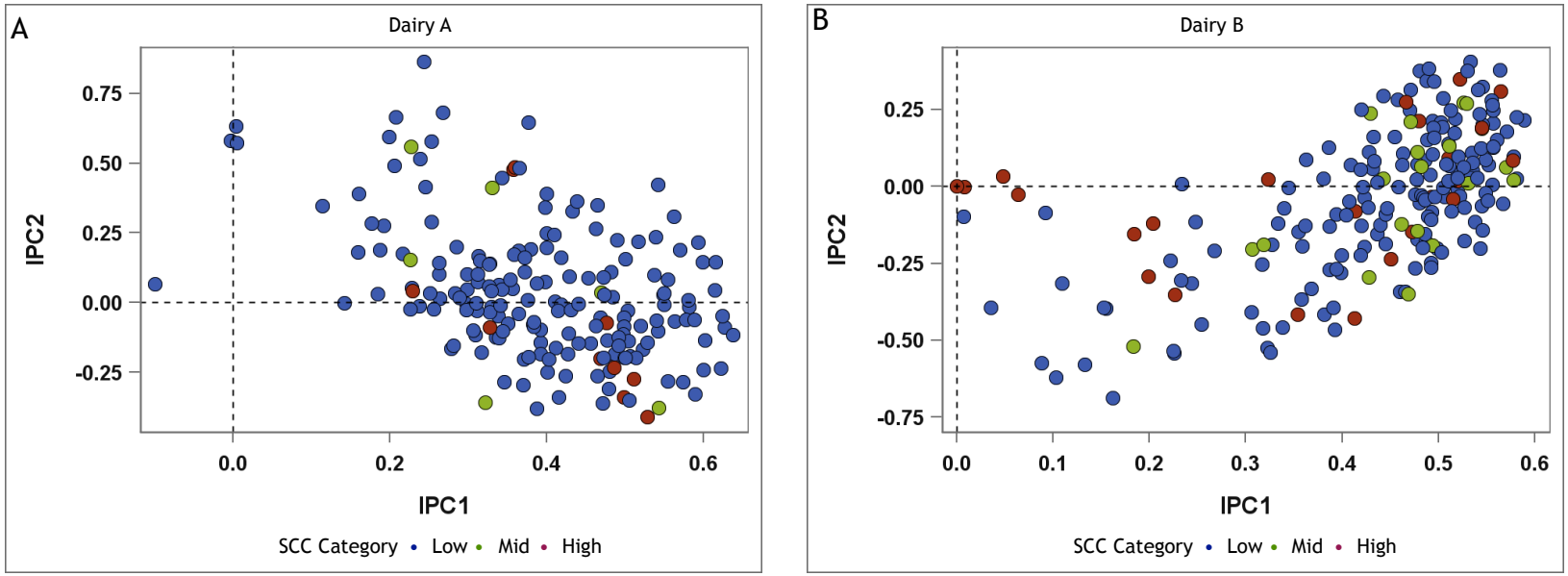


Figure 3.8: Additive main effects and multiplicative interaction (AMMi) model results of relative microbial abundance, among individual dairies, where axes are equivalent to principal components 1 and 2, and data colored by SCC category: (A) dairy A and (B) dairy B.

Table 3.1: Relative abundance of the top 11 bacterial genera and diversity metrics in bovine milk from two dairies.[1]

	Dairy A	Dairy B
	(n = 191)	(n = 209)
Bacterial taxa		
Staphylococcus	5.3 ± 1.0^a	19.0 ± 1.9^b
Ruminococcaceae	3.1 ± 0.2^a	11.1 ± 0.5^b
Corynebacterium	6.6 ± 0.4	5.6 ± 0.6
Turicibacter	5.3 ± 0.4^a	4.2 ± 0.3^b
Intestinibacter	4.2 ± 0.2	4.6 ± 0.2
Atopostipes	3.6 ± 0.3	3.1 ± 0.3
Clostridium sensu stricto	3.3 ± 0.3^a	2.6 ± 0.2^b
Pseudomonas	2.8 ± 0.3^a	1.4 ± 0.2^b
Paeniclostridium	2.3 ± 0.1^a	1.7 ± 0.1^b
Glutamicibacter	4.0 ± 0.3	0.1 ± 0.0
"Other"	18.9 ± 0.6^a	14.0 ± 0.7^b
Diversity indicies		
Richness	46.5 ± 0.65^a	38.4 ± 0.59^b
Pielou evenness	0.76 ± 0.01^a	0.67 ± 0.01^b
Shannon diversity	19.8 ± 0.4^a	13.7 ± 0.4^b
Shannon entropy	2.9 ± 0.05^a	2.4 ± 0.03^b
Simpson evenness	0.25 ± 0.01^a	0.22 ± 0.01^b
Inverse simpson	11.7 ± 0.3^a	8.2 ± 0.3^b

[1] *All values are means ± SEMs. Values in a row within a sampling location that have different superscripts are significantly different based on a mixed effect model. $P < 0.05$ corrected for multiple comparisons.*

Table 3.2: Relative abundance of the top 11 bacterial genera and diversity metrics in bovine from milk with varying SCC levels.[1]

	Low SCC	Mid SCC	High SCC
	(n = 340)	(n = 26)	(n = 34)
Bacterial taxa			
Staphylococcus	10.0 ± 1.1^{a}	29.6 ± 6.4^{b}	24.6 ± 5.5^{b}
Ruminococcaceae	7.5 ± 0.4^{a}	7.4 ± 1.2^{ab}	5.0 ± 0.9^{b}
Corynebacterium	6.4 ± 0.4	4.9 ± 1.1	4.6 ± 2.1
Turicibacter	4.7 ± 0.2	5.0 ± 1.0	5.0 ± 0.9
Intestinibacter	4.4 ± 0.2	4.2 ± 0.8	4.4 ± 0.8
Atopostipes	3.7 ± 0.2^{a}	2.0 ± 0.5^{ab}	1.6 ± 0.3^{b}
Clostridium sensu stricto	2.9 ± 0.2	3.3 ± 0.8	2.8 ± 0.6
Pseudomonas	2.3 ± 0.2	1.2 ± 0.6	1.0 ± 0.3
Paeniclostridium	2.0 ± 0.1	1.8 ± 0.3	1.6 ± 0.3
Glutamicibacter	2.1 ± 0.2	0.7 ± 0.3	1.1 ± 0.3
"Other"	16.8 ± 0.5	11.4 ± 1.8	14.7 ± 2.5
Diversity indicies			
Richness	43.4 ± 2.7^{a}	37.8 ± 0.5^{ab}	34.3 ± 2.0^{b}
Pielou evenness	0.73 ± 0.04^{a}	0.63 ± 0.04^{b}	0.61 ± 0.01^{b}
Shannon diversity	17.4 ± 1.4^{a}	12.0 ± 0.3^{b}	12.0 ± 1.3^{b}
Shannon entropy	2.7 ± 0.2^{a}	2.2 ± 0.0^{b}	2.1 ± 0.2^{b}
Simpson evenness	0.24 ± 0.01	0.20 ± 0.02	0.26 ± 0.04
Inverse simpson	10.4 ± 1.0^{a}	7.1 ± 0.2^{b}	7.1 ± 1.0^{b}

[1] *All values are means ± SEMs. Low SCC <200,000 cells; Mid SCC 200,000-400,000 cells; High SCC >400,000 cells Values in a row within a sampling location that have different superscripts are significantly different based on a mixed effects model. P >0.05 corrected for multiple comparisons.*

Table 3.3: Relative abundance of the top 11 bacterial genera and diversity metrics in bovine milk across quarters.[1]

	Right Front	Right Rear	Left Rear	Left Front
	(n = 101)	(n = 101)	(n = 99)	(n = 99)
Bacterial taxa				
Staphylococcus	10.0 ± 1.9	12.0 ± 2.3	13.7 ± 2.6	13.7 ± 2.6
Ruminococcaceae	7.9 ± 0.7	7.0 ± 0.7	6.9 ± 0.7	7.2 ± 0.8
Corynebacterium	6.3 ± 0.9	5.7 ± 0.6	6.3 ± 0.8	5.9 ± 0.6
Turicibacter	5.1 ± 0.5	4.7 ± 0.4	4.9 ± 0.5	4.3 ± 0.4
Intestinibacter	4.5 ± 0.3	4.3 ± 0.3	4.8 ± 0.4	4.0 ± 0.3
Atopostipes	4.3 ± 0.5	3.1 ± 0.3	3.1 ± 0.4	3.0 ± 0.4
Clostridium sensu stricto	2.8 ± 0.3	2.8 ± 0.3	3.5 ± 0.6	2.6 ± 0.3
Pseudomonas	2.1 ± 0.3	2.3 ± 0.4	2.0 ± 0.3	2.0 ± 0.3
Paeniclostridium	2.0 ± 0.2	1.9 ± 0.2	2.1 ± 0.2	1.9 ± 0.2
Glutamicibacter	1.7 ± 0.3	2.2 ± 0.4	1.8 ± 0.3	2.0 ± 0.3
"Other"	16.3 ± 0.9	16.7 ± 0.9	15.6 ± 1.0	16.6 ± 1.1
Diversity indicies				
Richness	42.8 ± 1.0	42.4 ± 0.9	42.0 ± 1.0	41.7 ± 0.9
Pielou evenness	0.73 ± 0.02	0.72 ± 0.01	0.70 ± 0.01	0.70 ± 0.02
Shannon diversity	17.0 ± 0.7	16.8 ± 0.6	16.1 ± 0.6	16.6 ± 0.7
Shannon entropy	2.7 ± 0.1	2.7 ± 0.1	2.6 ± 0.1	2.6 ± 0.1
Simpson evenness	0.25 ± 0.01	0.24 ± 0.01	0.22 ± 0.01	0.23 ± 0.01
Inverse simpson	10.2 ± 0.4	10.1 ± 0.4	9.5 ± 0.4	9.8 ± 0.5

[1] *All values are means ± SEMs. Values in a row within a sampling location that have different superscripts are significantly different based on a mixed effect model. $P < 0.05$ corrected for multiple comparisons.*

Table 3.4: Non-negative matrix factorization (NMF) scores by dairy.[1]

Dairy A		**Dairy B**		**Overall**	
Kocuria	1	*Ornithinimicrobium**	1	*Streptococcus**	1
Aerococcus	1	*Kocuria*	1	*Bifidobacterium**	1
Sphingomonas	1	*Faecalibacterium*	1	*Kocuria*	1
*Corynebacterium**	1	*Enterococcus**	1	*Aerococcus*	1
Staphylococcus	1	*Leucobacter**	1	*Serratia**	1
Faecalibacterium	0.95	*Escherichia/Shigella*	1	*Faecalibacterium*	1
*Propionibacterium**	0.94	*Aequorivita**	1	*Escherichia/Shigella*	1
*Clostridium sensu stricto**	0.92	*Glutamicibacter**	1	*Janibacter*	1
*Turicibacter**	0.92	*Aerococcus*	1	*Sphingomonas*	1
*Bacteroides**	0.85	*Staphylococcus*	1	*Staphylococcus*	1

[1]*are exclusively in the top10 of one category

Table 3.5: Non-negative matrix factorization (NMF) scores by SCC category. [1]

Low SCC		**Mid SCC**		**High SCC**	
Sphingomonas	1	*Enterococcus*	0.98	*Faecalibacterium*	1
Faecalibacterium	1	*Faecalibacterium*	0.77	*Escherichia/Shigella*	1
Serratia	1	*Sphingomonas*	0.76	*Streptococcus*	0.94
*Glutamicibacter**	1	*Tyzzerella**	0.63	*Aerococcus*	0.92
Streptococcus	1	*Kocuria*	0.63	*Serratia*	0.85
Bifidobacterium	1	*Escherichia/Shigella*	0.63	*Kocuria*	0.81
Aerococcus	1	*Leucobacter**	0.62	*Bifidobacterium*	0.77
Kocuria	1	*Serratia*	0.61	*Staphylococcus*	0.77
Staphylococcus	1	*Propionibacterium**	0.59	*Enterococcus*	0.67
*Clostridium_sensu_stricto**	0.97	*Dietzia**	0.58	*Aliicoccus**	0.66

[1] *are exclusively in the top10 of one category. *Low SCC <200,000 cells; Mid SCC 200,000-400,000 cells; High SCC >400,000 cells*

Table 3.6: Non-negative matrix factorization (NMF) scores by quarter.[1]

Left Rear		**Left Front**		**Right Front**		**Right Rear**	
Bifidobacterium	1	*Bifidobacterium*	1	*Escherichia/Shigella**	1	*Kocuria*	1
Serratia	1	*Kocuria*	1	*Aerococcus*	1	*Serratia*	1
Faecalibacterium	1	*Enterococcus*	1	*Sphingomonas*	1	*Sphingomonas*	1
Streptococcus	1	*Faecalibacterium*	1	*Streptococcus*	1	*Faecalibacterium*	1
Aerococcus	1	*Aerococcus*	1	*Kocuria*	0.98	*Bifidobacterium*	1
Staphylococcus	0.99	*Serratia*	0.99	*Staphylococcus*	0.94	*Glutamicibacter**	1
Kocuria	0.99	*Staphylococcus*	0.96	*Faecalibacterium*	0.92	*Aerococcus*	1
Janibacter	0.90	*Psychrobacter**	0.88	*Bifidobacterium*	0.90	*Streptococcus*	0.99
*Turicibacter**	0.82	*Streptococcus*	0.85	*Atopostipes**	0.89	*Staphylococcus*	0.98
Enterococcus	0.80	*Janibacter*	0.83	*Serratia*	0.88	*Janibacter*	0.88

[1]*are exclusively in the top10 of one category

CHAPTER 4

Global variation, evolutionary perspective, and associations of candidate genes with human milk oligosaccharides and lactose

4.1 Abstract

The evolution and genetic regulation of lactation has growing interest, particularly with regard to milk carbohydrates. This study sought to report on global variation in candidate genes associated with the synthesis of human milk carbohydrates, compare this variation to milk carbohydrate phenotypes, and assess the evidence for evolutionary selection on those genes. Primer pairs (n = 113) were designed to cover exonic regions of four genes associated with milk carbohydrate production and two pseudogenes for examination of signatures of selection. Sequences were mapped to the human reference and 230 single nucleotide polymorphisms (SNPs) were identified. These SNPs were then associated with 19 human milk oligosaccharide concentrations and lactose from paired samples. Previously unreported variants were discovered in each gene: one in *LALBA*, 49 in *FUT2*, two in *FUT3*, and 27 in *ST6GalNAc5*. Evolutionary statistics (F_{ST}, Tajima's D, and π) were calculated, and 25 SNPs were determined to be under selective pressure (elevated F_{ST}). These results show preliminary evidence that milk synthesis is likely under evolutionary pressure, genes involved in milk carbohydrate regulation are variable and impact milk carbohydrate concentrations, and more work needs to be conducted regarding the genetic regulation of milk constituents.

4.2 Introduction

The evolution of lactation has been debated for well over a century (Hayssen, 1985) and nutrient composition of milk has been of interest for decades as this food source provides the first nutrition to young (McGuire, 2017). Human milk has long been considered the gold standard of infant nutrition (WHO, 2017), and provides nutrients, immune-modulatory

factors, microbial constituents, and many other active factors to the developing infant (Hale and Hartmann, 2017). Composition has focused on macro- (e.g., fat, protein, carbohydrates) and micronutrients (e.g., vitamins and minerals) and the variation among mammals of the same species as well as variation among species (Capuco, 2009). There exists substantial evidence that milk nutrient composition is regulated by diet (Lonnerdal, 1986), but research on the role of genetics on milk nutrient composition is growing (Colodro-Conde, 2015; Fanos, 2017; Lee, 2016; Strucken, 2015). Even further than just genetic regulation, research has begun to address the evolutionary nature of lactation via genetics, and how nutrient composition has adjusted over time to develop into the ideal nutrient source for neonatal mammals (Lemay, 2009).

An initial factor that influences both genetic regulation as well as the evolution of lactation is the presence and nature of lactose. Lactose is a disaccharide made up of glucose and galactose linked together by lactose synthase, an enzyme complex composed of beta-galactosyltransferase 1 and alpha-lactalbumin. Considering lactose is an osmotic regulator in the mammary gland, driving milk production (Oftedal, 2009) and the genetic similarity to lysozyme, alpha-lactalbumin synthesis has been hypothesized to be at the center of the development of lactation (Capuco, 2009). Evidence has pointed to a gene duplication and subsequent adaptation of alpha-lactalbumin from lysozyme (Hayssen, 1985). The inception of alpha-lactalbumin is estimated at least 310 million years ago which is prior to the split between synapsids and sauropsids, whereas the earliest mammaliforms were identified approximately 100 million years later (Oftedal, 2009). More complex than lactose, oligosaccharides also make up a proportion of sugars in milk ranging from the most abundant in some monotremes and marsupials to low, but appreciable quantities in modern mammals (Oftedal, 2009). Oligosaccharides are all founded on a backbone of lactose with additional sugar units added in various branching structures (Bode, 2012). Variability in these components due to environmental factors and diet of the mother has been noted; however, there is likely a maternal genetic effect on milk oligosaccharides (Hale and Hartmann, 2017). Human milk

oligosaccharides (HMO) vary in type and quantity across human populations (Erney, 2000; Thurl, 2007; McGuire, 2017). Specifically, HMO are a set of complex carbohydrates found in human milk that, despite not serving as nutrients per se for the infant, can act as prebiotics for the microbiota of the developing infant gastrointestinal tract (GIT) (Bode, 2012; Kononova, 2017) and potentially the maternal milk microbiome (Triantis, 2018). Oligosaccharides also potentially act to modulate the infant immune system, as antiadhesive antimicrobials, and contribute to early brain development (Bode and Jantscher-Krenn, 2012). Morrow et al. (2005) determined that fucosylated HMO appear to be a major protector of the infant from enteric pathogen binding in both the GI and respiratory tract (Morrow, 2005).

Many of these HMO structures are also genetically regulated, much like lactose (Bode, 2012). Despite vast knowledge of how host genetics impact milk components in the dairy cow (Lemay, 2009), little is known about genetic variation among women and its impact on human milk. The genetic regulation of HMO is not well understood; however, some of the enzymes producing HMO are well characterized proteins of genes from the ABO-Lewis blood group structure pathways (Soejima, 2007). Specifically, *FUT2* and *FUT3* are genes which code for fucosyltransferases that add α-1-2-linked fucose to the terminal galactose of HMO and an α-4/3-linked fucose on the internal N-acetylglucosamine, respectively (Bode, 2012). Additionally, it is thought that there are specific sialic acid transferase genes (e.g., ST6GalNac5) which attach sialic acid to nascent oligosaccharides, including lactose.

Despite what we know about variation in *FUT2* and *FUT3* genes involved in synthesis of milk carbohydrates have yet to be truly investigated in the context of lactation. To leverage this perspective, genotypic data from the same women will be related to milk phenotypes (e.g., HMO and lactose). We hypothesize that genetic variation related to milk carbohydrate synthesis is associated with variation in milk HMO profiles around the world. Knowing that HMO isoforms are variable in milk and there exists extensive genetic variation in the regulatory genes of HMO synthesis, we aim to piece together the connection between maternal

genetic variation and HMO composition of human milk.

4.3 Materials and Methods

Subjects, Experimental Design, and Ethics Approvals

Phenotypic data and enrollment information about the subjects participating in this study have been previously described (McGuire, 2017). In brief, women were recruited from multiple international sites and were eligible for participation under the following criteria: women ($\geq$ 18 yr of age) had to be breastfeeding or pumping at least five times per day, self-reported as healthy with a healthy infant, and between two wk and five mo postpartum. Exclusion criteria included any indication of acute breast infection or pain that was 'abnormal' for lactation, maternal or infant use of antibiotics in the previous 30 d, or nursing a child with symptoms of illness in the previous seven d. The international locations included two European [Spain (n=40) and Sweden (n= 24)], one South American [Peru (n=43)], two North American [Eastern Washington/Northern Idaho (n=41) and Southern California, self-identified as Hispanic (n=19)], and six sub-Saharan African [rural and urban Ethiopia (n=40 and n=36, respectively), rural and urban Gambia (n=40 for both), Ghana (n=41), and Kenya (n=40)].

When enrolled, each woman completed several questionnaires including eligibility criteria, demographic information on general maternal and infant health, and anthropometry. Ethics approvals were obtained for all procedures from each participating institution, with overarching approval from the Washington State University Institutional Review Board (#13264). After being translated from English (as needed), informed, verbal or written consent (depending on locale and the subject's literacy level) was acquired from each participating woman.

Saliva Collection and Preservation

Using a gloved hand, research personnel opened a collection tube with collection funnel attached for the mother to collect approximately 2 mL of saliva in a SalivaGard DNA

kit (USA Biomatrica, San Diego, CA). Research personnel then added the preservative, SampleMatrix® (also known as QiaSafe, Lee, 2010) and capped the tube. All samples were shipped and stored at room temperature until analyzed.

DNA Extraction

DNA extractions were performed with the Gentra Pure Gene Blood Kit (Qiagen, Valencia, CA) using manufacturer's protocol. DNA was resuspended in 200 μL nuclease free water and stored at -20 °C until further analysis. DNA was quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher/Life Technologies, Grand Island, NY) and quality was assessed as 260/280 ratio using a Nanodrop 1000 (Thermo Scientific).

Primer design and validation

Targeted primers were designed using manufacturer instructions (Fluidigm, San Francisco, CA) for most exon specific regions of genes alpha lactalbumin (*LALBA*), fucosyltransferase 2 (*FUT2*), fucosyltransferase 3 (*FUT3*), and ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 5 (*ST6GALNAC5*). Additionally, regions of secretory blood group 1 (*SEC1P*) and gulonolactone (L-) oxidase (*GULOP*) pseudogenes were assessed for neutral genome comparisons. Primer pairs were developed using human reference build HG38 and Primer3 software ($http://frodo.wi.mit.edu$) with base parameters as follows; product size length <600 bp, Primer Tm min: 59.0, opt: 60.0 and max: 61.0, and Max Poly-X: 3. Primer pairs were *in silico* validated using BLAST (NCBI) to check for specificity of binding and alternate product yield (Altschul, 1990). Primer sequences were selected based on their specificity to the location (as few alternate products as possible) and the size of the product <600 bp. Additionally, a short consensus sequence (CS) tag was added to the 5' end of the target-specific primer. The inclusion of the CS tag on the target-specific primer and on the 3' end of the Illumina sequencing adaptor allows the design of multiple target-specific primers that can then be linked to primers containing a complementary CS tag, the Illumina sequencing adapter, and individual barcodes (University of Idaho IBEST Genomics Resources Core Facility).

Primer validation was conducted using the manufacturer's protocol (Fluidigm). PCR reactions were carried out in 5 μL volume using FastStart^{TM} High Fidelity PCR System and dNTPack (Roche), specifically 1x FastStart Reaction Buffer, 4.5 mM MgCl2, 5% DMSO, 200 μM of each nucleotide base, 0.05 U FastStart High Fidelity Enzyme Blend (polymerase), 400 nM Illumina barcode primers (University of Idaho IBEST Genomics Resources Core Facility), 10 ng/μL genomic DNA, and 50 μM specific target primer pairs (Sigma, St. Louis, MO) following manufacturer's instructions (Fluidigm). Amplification was validated using a QIAxcel screening cartridge (Qiagen) and amplicons were deemed successful with a moderate amplification peak assessed by electropherogram.

Fluidigm Protocol and Sequencing

The Fluidigm Access Array and Juno technologies were used for amplicon production, using the manufacturer's protocol. Briefly, a 48x48 Access Array IFC (Fluidigm) was used to simultaneously run 48 primer pairs x 48 samples. The reagents used were the same as listed above and a 20x loading reagent (Fluidigm). The reaction was run using the same thermal cycling conditions as a primer validation step. Amplicons were post-processed following manufacturer guidelines; amplicons were harvested and pooled equi-volume, cleaned with AMPure XP beads (Beckman), and qualified with Advanced Analytical Fragment Analyzer (22 cm capillary array) (Ames, IA). Libraries were quantified using qPCR, Kappa kit for Illumina libraries on the ABI StepOnePlus qPCR machine (ThermoFisher/Life Technologies), normalized, pooled and sequenced on an Illumina MiSeq (Illumina, San Diego, CA) using v3 2x300 method.

Data Processing

Amplicon sequences were preprocessed from raw sequences and demultiplexed using default settings of the custom python pipeline, dbcAmplicons and sequences were split by sample using a custom python script, splitReadsBySample.py
(*https* : *//github.com/msettles/dbcAmplicons/*). Primer sequences were not trimmed to facilitate mapping to the reference. Raw, non-overlapped reads were mapped to HG38 using

Burrows Wheeler Aligner (BWA) mem version 0.7.12 (Li, 2009) adding read tag groups (for variant calling) and using default settings. Variant calls were made via the Genome Analysis Toolkit (GATK) version 3.5 Unified Genotyper (McKenna, 2010) across each subpopulation using GATK best practices (Van der Auwera, 2013). The raw calls were then filtered against a BED file of primer locations to prevent calls within the primer sequence and filtered with the following parameters using VCFtools version 0.1.14 (Danecek, 2011). Individual variant calls required >15x coverage, >0.05 minor allele frequency (MAF), data present in >70% of the samples, and Hardy-Weinberg Equilibrium conditions. Individuals with <70% SNP call rate were also removed from further analysis for a total number of 281 subjects.

HMO and Lactose Analysis

Milk collection and HMO analysis was as previously described (McGuire 2017). Analysis of lactose in milk was adapted from Polberger and Lonnerdal, 1993. First, protein was precipitated from 0.25 uL whole milk using 0.3 M barium hydroxide and 5% zinc sulfate. Samples, lactose standards, and glucose standards were digested by beta-galactosidase (Roche 105-031, 1500 U/1 mL, from Escherichia coli). and assessed via spectrophotometry at 450 nm. Total glucose was then back calculated to infer the lactose concentration in the samples.

Statistical Analysis

After filtering, allele frequency, F_{ST} (fixation index; Holsinger, 2009), Tajima's D, and π (nucleotide diversity) were calculated using VCFtools (Danecek, 2011). Allele frequencies were calculated per population, F_{ST} was calculated using the Weir and Cockerham calculation (1984) among all populations, and a sliding window of 5 bp was used to calculate Tajima's D and π. F_{ST} values were considered of interest at >0.1 and Tajima's D were considered of interest at >1.5. Basic ANOVA analysis of lactose and HMO concentrations by population was calculated in R version 3.5.1 (R core team) and significance was declared at P <0.01.

Associations

Association analysis was run using a linear model in Plink version 1.9 (*http* : *//pngu.mgh.harvard.edu/purcell/plink/*; Purcell, 2007) with HMO phenotypes and lactose. Significance was declared at P <0.01 with a Bonferroni correction to account for multiple comparisons.

4.4 Results

Targeted exonic regions of *LALBA*, *FUT2*, *FUT3*, *ST6GALNAC5*, and pseudogenes *GULOP* and *SEC1P* in 281 individuals covering a variety of human populations were sequenced and compared to concentrations of HMO and lactose in milk collected.

General disposition of the data

We sequenced products of 113 custom designed primer pairs (Table 1) spanning $\sim$ 500 bp each of the coding regions of four genes and two pseudogenes from 281 lactating women located in 11 populations, nine countries, and four continents. The women were generally similar in age, parity, time postpartum, weight, and height with some differences between populations (Table 2). Spanish women were older but had limited parities whereas women in Sweden and the United States were some of the tallest and heaviest. Time postpartum was not different among sites. Lactose concentration in milk was not different among populations (Figure 1) with a range of 71.6 to 85.6 g/dL for means by site. Oligosaccharides were variable by country, and described previously for the full dataset (McGuire, 2017). A summary of the HMO for the women by population included in the genetic evaluation is found in Figure 2.

Nucleotide Variation and Phenotype Association

Alpha Lactalbumin (*LALBA*) Using 5 primer pairs, 2,706 bp of the total 32,428 bp were sequenced. The initial three primer pairs covering exons 1-3, though, had very low read coverage having $\leq$ 6000 reads per sample, where successful calls were made in the final two primers whose average read count was >114,000 reads per sample (Table 1 and Figure 3).

Variation was detected in five locations (Table 3): one novel (referring to no rsID available via NCBI) at chr12:48569365, one nonsynonymous (change which confers an amino acid change to the protein) rs1261272249, one intergenic (very low frequency) rs113190598, and two upstream variants rs73104702 and rs923867024. No significant association was found between a SNP and any concentration of HMO or lactose (Figure 4).

Fucosyltransferase 2 (*FUT2*) encodes the enzyme with alpha-(1,2)-fucosyltransferase activity. Seventeen primer pairs were used to sequence the coding region of the gene, as well as some potential upstream promotor region sites, and coverage was approximately 8,000 bp of the 9,980 bp gene. Primers FUT2-6, FUT2-1, and FUT2-P3 had <42,000 average reads per sample and no SNPs were identified in these regions. The remaining primers were more successful in identifying variants with >118,000 average reads per sample (Table 1 and Figure 5). We found 119 SNPs: 42 upstream, two downstream, 12 nonsynonymous (including one stop gain), seven synonymous, and 56 in the 3'UTR (untranslated region) (Table 4). Among these loci, 49 were novel, ten very low frequency, and five appear to be population specific. Of the novel SNPs, most were found in intergenic regions or upstream of the coding region. Of those in the coding region, one SNP was found to be synonymous (chr19:48703974) and three SNPs were found to be nonsynonymous; chr19:48703384 changes a glutamic acid changed for a glycine, chr19:48703419 changes a threonine to proline, and chr19:48703912 changes an isoleucine to threonine. These novel variants had low allele frequencies yet several noncoding, upstream variants, such as chr19:48694285 have allele frequencies as high as 20-39% (Table 4).

Variants in *FUT2* had many significant associations with HMO moieties (Figures 6 and 7). A novel SNP (chr19: 48694463) was significantly associated with concentrations of LNnT and LNT; a large group of SNPs at the end of FUT2 (chr19: 48702888 - 48705969) were associated with several outcomes including, secretor status (2'FL >200 nmol/mL are secretors), and concentrations of 2'FL, DFLac, FDSLNH, LNFPI, LNFPII, LNnT (far fewer SNPs), LNT, LSTb, total HMO, total fucosylated HMO (sum of all fucose containing HMO) and

total sialylated HMO (sum of all sialic acid containing HMO; far fewer SNPs) Additionally rs603985 and rs485186 was associated with concentration of DSLNT and a novel SNP (chr19: 48705048) was associated with concentration of DFLNH.

Fucosyltransferase 3 (*FUT3*) encodes the enzyme with alpha-(1,3)-fucosyltransferase and alpha-(1,4)-fucosyltransferase activities and is the last step of Lewis antigen biosynthesis. Using 8 primer pairs, about 3,600 bp of the 9,980 bp gene was sequenced. Coverage was lowest at the beginning of the gene with <5000 reads per sample for FUT3-P and FUT3-1, subsequent primers had average reads per sample >139,000 (Table 1 and Figure 8). We identified 15 SNPs: nine nonsynonymous (one novel), and four synonymous (one novel), and two 3'UTR variants (Table 5). One of the identified variants appeared to be population specific, but none were very low frequency. The two novel SNPs were identified in the coding region; chr19:5843757 is a synonymous variant and chr19:5843782 is a nonsynonymous variant changing a valine to glycine. Importantly, two known SNPs that impact Lewis secretor status were confirmed at relatively high frequencies: rs3745635 and rs28362459.

Associations were found with concentrations of four HMO moieties (Figure 9): rs778986 and rs28362459 were associated with DFLNT concentration, and rs28362459 was also associated with concentrations of LNFPI, LNFPII, and LNH. No significant associations for 3FL or total fucosylated HMO were found with any *FUT3* SNP (Figure 9).

ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 5 (*ST6GALNAC5*) is one of a family of Golgi type II transmembrane glycosyltransferase proteins which add sialic acid residues to cell surface proteins and it is thought to perform this function on HMO. The sequencing results across this gene mostly captured data in intergenic regions, using 27 primers, spanning 12,866 bp of the 201,727 bp that make up the gene. Coverage across *ST6GALNAC5* was mostly good with primers (ST6-2 and ST6-20) having <41,000 average reads per sample, however most of the sequences mapped outside of the main exon (Table 1 and Figure 10). We found 58 SNPs: three nonsynonymous (all novel), six intergenic (one novel, one potential splice site), 16 downstream (eight novel), and 33 3'UTR (15 novel),

within these SNPs ten appear to be population specific (Table 6). Of the novel SNPs, three were in the coding region of the gene; chr1:77044448 is a nonsynonymous mutation changing an isoleucine to a serine, chr1:77044472 is also a nonsynonymous mutation changing a methionine to arginine, and chr1:77063069 which changes a glycine to a cystine. Allele frequencies for all SNPs can be found in Table 6.

Very few associations between ST6GalNAc5 and HMO concentrations were found (Figure 11): rs11162262 was associated with 2'FL, rs115329926 was associated with 6'SL, and a novel SNP (chr1: 77067914) was associated with LNFPIII. No significant associations were found with total sialylated HMO (Figure 11).

Gulonolactone (L-) oxidase, pseudogene (*GULOP*) was sequenced as a 'neutral' marker, due to non-functionality of the L-gulono-gamma-lactone gene that influences vitamin C synthesis (Yang, 2013). General gene coverage and read depth are reported in Figure 12. Sequencing was limited to four primer pairs, covering 2,373 bp of the 28,800 bp pseudogene. In this limited selection, we identified 26 SNPs, eight novel, two of which appear to be population specific (rs78422197 and rs146127661 in GN; Table 7). Although none of these mutations have expression effects in this pseudogene, the variation helps to determine the relative rate of mutations in populations (Table 7). Significant associations were found with DFLac and rs146127661 and rs78422197 (Figure 13).

Secretory blood group 1, pseudogene (*SEC1P*) is a pseudogene found nested in the family of fucosyltransferase genes on chromosome 19 with relatively unknown function; it was sequenced as a potential neutral marker. General gene coverage and read depth are reported in Figure 14. Three primer pairs were designed to cover regions of this gene which did not overlap with other open reading frames associated with alternate genes, covering 1,500 bp of the 44,207 bp pseudogene. We identified seven SNPs, four of which were novel (Table 8). Of the four novel SNPs, chr19:48679949 and chr19:48680161 were relatively low frequency (Table 8). There were no significant associations found between SNPs in SEC1P and HMO or lactose concentrations in milk (Figure 15).

Interpopulation Differentiation Analysis

To assess the interpopulation differences across these evaluated genes, F_{ST}, Tajima's D, and nucleotide diversity (π) were calculated across all populations. π was calculated on a 5 bp sliding window and used as a basis marker for the level of variation across populations and is generally associated with the overall allele frequencies (Tables 3-8). F_{ST} was calculated on a per site basis across all populations, and elevated F_{ST} was found (>0.1) in three out of four genes and one out of two pseudogenes sequenced: zero SNPs in *LALBA*, 14 SNPs in *FUT2*, two SNPs in *FUT3*, four SNPs in *ST6GalNAc5*, five SNPs in *GULOP* and zero SNPs in *SEC1P*. Tajima's D was calculated on a 5 bp sliding window and elevated D values (>1.5) were found in the same three out of four genes and one of the two pseudogenes had elevated F_{ST}, though the SNPs did not often overlap. Values of π, F_{ST}, and D are plotted as scatter plots along the length of each gene in Figures 16-21.

4.5 Discussion

Lactose and HMO

Lactose concentrations were not different among populations, and were consistent with concentrations reported at similar times postpartum (Gay, 2018; Spavecek, 2015; Zhang, 2013) Variation in HMO phenotypes has been described across populations (Erney, 2000; McGuire et al. 2017). Briefly, the complete dataset from which the current study was used found total HMO concentration of each individual HMO moiety except for LNFP I differed by population. These results are similar to the relative concentrations of HMO moieties from the 281 women paired to sequencing data in the current study (Figure 2). We found differences (P <0.05) among populations in the following HMO: 2'FL, 3FL, LNnT, DFLac, 6'SL, LNT, LNFPII, LNFPIII, LSTb, LSTc, DSLNT, FLNH, DFLNH, FDSLNH, and DSLNH. This differs from previous reporting where differences were also found in: 3'SL, DFLNT and LNH, likely due to the reduced sample representation. The variation in HMO phenotype provides a basis to look at the genetic regulation of HMO synthesis and evolutionary processes

potentially in play. This variability and unknown genetic influence may be indicative that milk is not a 'one-size-fits-all' substance and cannot be characterized the same on a global level.

Nucleotide Variation

The genetic variation of genes tied to milk carbohydrate synthesis was explored through custom, targeted sequencing. Beside genes involved in milk carbohydrate synthesis, two pseudogenes were included as a comparison of the baseline rate of variation and selective pressure (Fumagalli, 2009).

LALBA *LALBA* is the gene encoding the alpha-lactalbumin protein which is a key feature regulating lactose synthesis (Blackburn, 1989: Jabed, 2012) and has been reported as highly divergent across Mammalia (Lemay, 2009). Several studies have looked at the folding affinities for *LALBA* (Ramboarina, 2009; Mizuguchi, 2005) and human alpha-lactalbumin folding has been associated with tumor apoptosis (Svensson, 2003; Casbarra, 2004) suggesting protein sequence has biological impact. Further, absence of alpha-lactalbumin leads to failure of lactation (Stacey, 1995) demonstrating the critical importance of this gene. Due to limited sequencing reads, only one coding region SNP was identified as a nonsynonymous glycine to alanine amino acid change which occurred at low frequencies in urban Gambia and Sweden. This is consistent with findings listed on NCBI of a global MAF of >0.01 of a globally diverse study cohort (Sherry, 2001). Few SNPs found in *LALBA* is likely the result of poor sequencing, however reports in bovine genetics show a substantial divergence in the sequence of *LALBA* across species indicating that *LALBA* is not constrained of variation despite the necessity of lactose (Lemay, 2009).

FUT2 *FUT2* is involved in making the precursor of the H antigen in the A/B antigen synthesis pathway. This gene is also associated with the production of fucosylated HMO, particularly 2'FL (Kumazaki, 1984), denoting 'secretor status' when 2'FL is present in milk. Variations in *FUT2* included 18 cited variants and 3 clinical variants (NCBI) associated with presence/absence of 2'FL in human milk, and presence of the H antigen on red blood cells

and the mucosal lining of the gastrointestinal tract (Sherry, 2001). HMO (phenotypes) and their suspected regulatory genes (genotypes) are variable among women and populations, and the fucosyltransferase gene family, in particular, has been studied across subpopulations of humans (Koda, 2001; Ferrer-Admetlla, 2009). However, these lines of research have (to our knowledge) proceeded independently of one another. Soejima's group in Japan detected wide genetic variation in *FUT2* (Soejima, 2007, 2009). Likely due to sequencing limitations, we found a subset of SNPs in common with both Soejima (2007, 2009) and with Ferrer-Admetella (2009), covering varying populations across all 3 studies. The overlapping SNPs with their data are: rs1800021 (functional allele), rs492602, rs681343, rs28362836 (nonfunctional allele), rs281377 (functional allele), rs601338 (nonfunctional allele), rs1800027 (functional allele), rs1800025 (functional allele), rs602662, rs485186 (functional allele), rs485073 (functional allele), rs603985 (functional allele), which do not include 24 SNPs identified in previous studies, likely due to low sequencing coverage and/or population coverage, particularly considering evidence of several Asian specific polymorphisms (Soejima, 2007; Henry, 2014; Koda, 2001). Of the many reported SNPs in the dbSNP database, we found similar rates of variation in rs601338 49% reported in African populations where we found 43-59%; 44% in European populations where we found 41-50%; and 34% in the Americas where we found 10-40% (Sherry, 2001). rs1800459 and rs755843863 are reported at <1% across all populations and we found much higher frequencies in our data (8-48% and 0-11%, respectively; Sherry, 2001).

FUT3 FUT3 has also been studied broadly in the context of blood groups and functionally provides alpha-(1,3) and alpha-(1,4) linkages to many fucosylated HMO such as 3FL (Bode, 2012). Extensive variation has been discovered across *FUT3* including 10 cited variants, and 2 clinical variants (NCBI) have been associated with the Lewis antigen system (Corvelo, 2013; Soejima, 2009; Koda, 1993), miRNA and gastric cancer (Cai, 2016), ulcerative colitis and inflammatory bowel disease (Hu, 2016; Guo, 2015), cardiovascular disease (Silander, 2008), and schizophrenia (Yazawa, 1999). Our results differed from the findings

of Soejima et. al. (2009) identifying different polymorphisms across *FUT3* considering only 3 SNPs overlapped between our studies. Found in all populations in the current study and in three populations (Ghana, Caucasian, and Mongolian) from Soejima were rs3745635, rs812936, and rs28262459. Despite identifying different SNPs than those reported by Soejima et al. (yr), only 2 new variants (chr19:5843757 and chr19:5843782) were not previously reported. This could be due to the increased diversity of populations used in the current study in addition to sequencing past the only intron in this gene, which had not been done. Some SNPs (17) reported (Soejima) were not identified in the current study possibly due to low sequencing coverage, especially considering the two studies overlapped in both Caucasian and Ghanaian populations. As compared to reported allele frequencies to NCBI, we found similar rates of variation at the following loci: rs5844332 is reported at 30% in African populations where we found 9-38%; 2% in European populations where we found 2-3%; and 22% in the Americas where we found 0-45%; and rs28362459 is reported at 34% in African populations where we found 21-45%; 10% in European populations where we found 6-11%; and 27% in the Americas where we found 5-50% (Sherry, 2001).

ST6GALNAC5 Far less is known about the sialyltransferase genes that create sialylated HMO, but there is limited evidence that other genes similar to *ST6GalNAc5* are under evolutionary pressure (Teppa, 2016). There are many variants reported in NCBI including 6 cited variants associated with coronary artery disease (CAD) (Amini, 2014), colon cancer (Tsuchida, 2003), and nicotine response (Rose, 2010). The *ST6GalNAc5* gene was of particular interest because sialylated HMO appear to have many health benefits to the infant, including response to necrotizing enterocolitis (Jantscher-Krenn, 2011). The sialyltransferases are categorized into 4 families of genes: *ST6Gal*, *ST3Gal*, *ST6GalNAc*, and *ST8Sia* and there are at least 7 common HMO isoforms that contain sialic acid residues (Teppa, 2016; McGuire, 2017). We chose to sequence *ST6GalNAc5* due to evidence that it may contribute specifically to the structure of DSLNT found to be protective of very low birth weight neonates (Jantscher-Krenn, 2011; Autran, 2017). Despite significant differences in the quan-

tity (218-370 nmol/mL, P <0.01) of DSLNT among populations (McGuire, 2017), there were very little genotypic differences among populations in this sialyltransferase gene. Difficulty in detecting genetic differences could be because *ST6GalNAc5* is a very long gene, hard to tile unique primers for adequate coverage. The custom primer pairs (Table 1) covered the 5 exons across the length of *ST6GalNAc5*. Primers 1 and 2 had low sequence coverage and no variants were found and despite having sequence coverage, however Primer 4, 5, 7, and 8 were directly over coding portions of the gene, but only 3 coding SNPs were identified. This may be due to lack of variation in this gene, but further studies with more subjects will be needed to clarify. The many SNPs identified in the noncoding regions (Table 6) have 4 loci (rs11162262, rs199722, rs199724, and rs199663) which have elevated diversity statistics, particularly F_{ST} and Tajima's D. While this provides evidence that some selective pressure may be underlying the polymorphisms in this gene, little has been reported on the evolutionary history of this particular gene, although orthologs are found in most vertebrate species for the *ST6Gal* gene family (Teppa, 2016). The ability to connect functional *ST6GalNAc5* with sialylation of HMO is also limited based on associations as the only HMO associated with variants of *ST6GalNAc5* were 2'FL, LNFP III and 6SL, and only 6SL is sialylated. Little evidence that *ST6GalNAc5* provides the majority of structure to sialylated HMO was found, but further sequencing of *ST6* family genes may determine the genetic variation responsible for the sialylated HMO. Compared to allele frequencies reported to NCBI, we saw similar frequencies in our populations: rs200277344 reported at <1% in European populations where we found 6% in SW specifically; rs144737930 reported at 2% in African populations where we found 3% in GBR specifically; rs199724, rs199662, and re199663 all reported at 58% in African populations where we found 32-78%, 31% in European populations where we found 25-47%, and 50% in the America's where we found 23-60% (Sherry, 2001).

GULOP Five cited variants exist even in this pseudogene and have been primarily discovered in relation to L-gulono-gamma-lactone oxidase (*GULO*) and scurvy (Inai, 2003; Yang, 2013). Compared to reported allele frequencies in the dbSNP database, we found similar

results. rs78422197 and rs146127661 are only reported in African populations (3% and 2%, respectively) and we found this allele only in the Ghanaian population. rs17057419 was found at similar rates across our populations demographics: 36% in African populations where we found 28-62%; 15% in European populations where we found 17-23%; and 21% in the Americas where we found 0-19% (Sherry, 2001).

SEC1P Four cited variants have been identified in this region, mostly associated with serum carcinoembryonic antigen levels (Liang, 2014). Other studies have noted some potential interactions with *SEC1P* variants and *FUT2* variants as they are in close proximity to one another (Soejima, 2008; Koda 1997). One of the novel SNPs chr19:48680213 had higher allele frequency (0-24%) than the other identified SNPs.

Selection

The variation in HMO found in human milk suggests a potential for importance for the unique structures. To evaluate the value of these structures, one can assume evolutionary pressure could lead to greater reproductive success if certain HMO provide greater survival. Some statistics have been developed to test for signatures of evolutionary selection. These include F_{ST} or fixation index which measures interpopulation variation (Beaumont, 2005; Holsinger, 2009), Tajima's D compares the average number of pairwise differences to the number of segregating sites (Tajima, 1989), and nucleotide diversity (π) which measures the average nucleotide differences per site (Misawa, 1997). Elevated F_{ST} has often been reported as evidence for balancing selection between populations, however, there are few strong F_{ST} values across genes that would indicate strong population differentiation due to balancing selection in this case. There were no SNPs in *LALBA* that had elevated F_{ST} nor Tajima's D which is indicative that the alleles are not undergoing selection, likely due to the essential functionality of alpha-lactalbumin in lactation (Lemay, 2009). *FUT2* appears to be under non-neutral evolutionary selection, with evidence pointing toward balancing selection forces that appear to impact the expression of fucosylated antigens (Silva, 2009). Because the protein products of these genes are known to have immunologic roles in both mother and

infant, the variation and evolutionary pressure could be, at least in part, pathogen-driven. For instance, Fumagalli (2009) found evidence for selection and haplotype variation in *FUT2* with respect to pathogen density, indicating that host-pathogen pressure may contribute to the selective pressure on *FUT2*. This work concurs with conclusions from Bode that 'secretor status' (presence or absence of a functional FUT2 gene; presence or absence of 2'FL) can contribute to the risk of maternal and infant infections, including but not limited to caliciviruses, HIV, respiratory viruses, urinary tract infections, yeast infections, and some diarrheal infections (Bode, Jantscher-Krenn, 2012; Fumagalli, 2009). *FUT3* has two loci with elevated F_{ST} and two additional sites with elevated Tajima's D which would indicate some selective pressure, likely balancing the variation in this gene. *ST6GalNAc5* had 1 loci with elevated F_{ST} but is in a non-coding region of the gene but may be linked to variation not captured in this study. Comparatively, 'neutral' genes are likely to experience random variation due to genetic drift (Fay, 2002). We sequenced portions of *GULOP* and *SEC1P* as comparative non-coding regions of the genome, however we saw increased F_{ST} at 4 loci in *GULOP* which may mean that the among population differences are increased, or there may be a higher rate of mutation at this part of the genome (Holsinger, 2009). *LALBA* and *SEC1P* which saw no elevated evolutionary statistics are likely close to the level of background differentiation expected among human populations.

Values for Tajima's D were positive and elevated for several loci across these genes (Figures 16-21) suggesting some balancing selection. Of particular interest are the fucosyltransferase genes. The genes here have had polymorphisms identified that appear to be under some positive or balancing selection previously (Ferrer-Admetella, 2009). A key limit to estimating both FST and Tajima's D in this data is the lack of an Asian cohort which may directly contribute to the lower values than previously reported (Ferrer-Admetella, 2009; Fumagalli, 2009).

Limitations

Some limitations of this study are the low sample size, limited coverage across the genome

and targeted genes, and bias in global variation, specifically due to a lack of an Asian cohort. A targeted approach was attempted to determine polymorphisms across coding regions of selected genes thought to potentially alter milk carbohydrate composition. Had there been more sequence coverage of the genes, there would have been more potential to tie those genes directly to milk phenotypes. The limited sample size per population impaired the potential to detect variation in populations, particularly of very low frequency SNPs.

Conclusions

Variation in selected genes related to proteins involved in synthesis of carbohydrates (lactose and HMO) in milk was found confirming previously identified polymorphisms and novel ones: 5 SNPs in *LALBA*, 119 SNPs in *FUT2*, 15 SNPs in *FUT3*, 58 SNPs in *ST6GalNAc5*, 26 SNPs in *GULOP*, and 7 SNPs in *SEC1P*. Evidence for selective pressure was found in 46 SNPs across *FUT2*, *FUT3*, *ST6GalNAc5* and *GULOP* which have elevated F_{ST} (>0.1) or Tajima's D (>1.0) indicating that these genes may be under selective pressure. The long-term goals of this research are to determine the extent to which maternal genetics contributes to infant nutrition through variation in milk composition and identify selective pressures in genes associated with milk composition. This may alter how milk is viewed as a 'one-size-fits-all' source of nutrients.

4.6 Acknowledgements

I would like to acknowledge the assistance of the following individuals for their work as co-authors on this paper: Brenda Murdoch, Sam Hunter, Kelcey McBride, Kim Davenport, Hannah Jagger, Michelle K McGuire, Courtney L Meehan, Mark A McGuire, Janet E Williams, James Foster, Daniel W Sellen, Elizabeth W Kamau-Mbuthia, Egidioh W Kamundia, Samwel Mbugua, Sophie E Moore, Andrew M Prentice, Linda J Kvist, Gloria E Otoo, Caitlyn Placek, Kimberly A Lackey, Bianca Robertson, Susana Manzano, Lorena Ruiz, Juan M Rodriguez, Rossina G Pareja, and Lars Bode.

This work was supported by the National Science Foundation Science and Technology Center

BEACON DBI-0939454, the National Science Foundation (award 1344288), and the Idaho Agricultural Experiment Station.

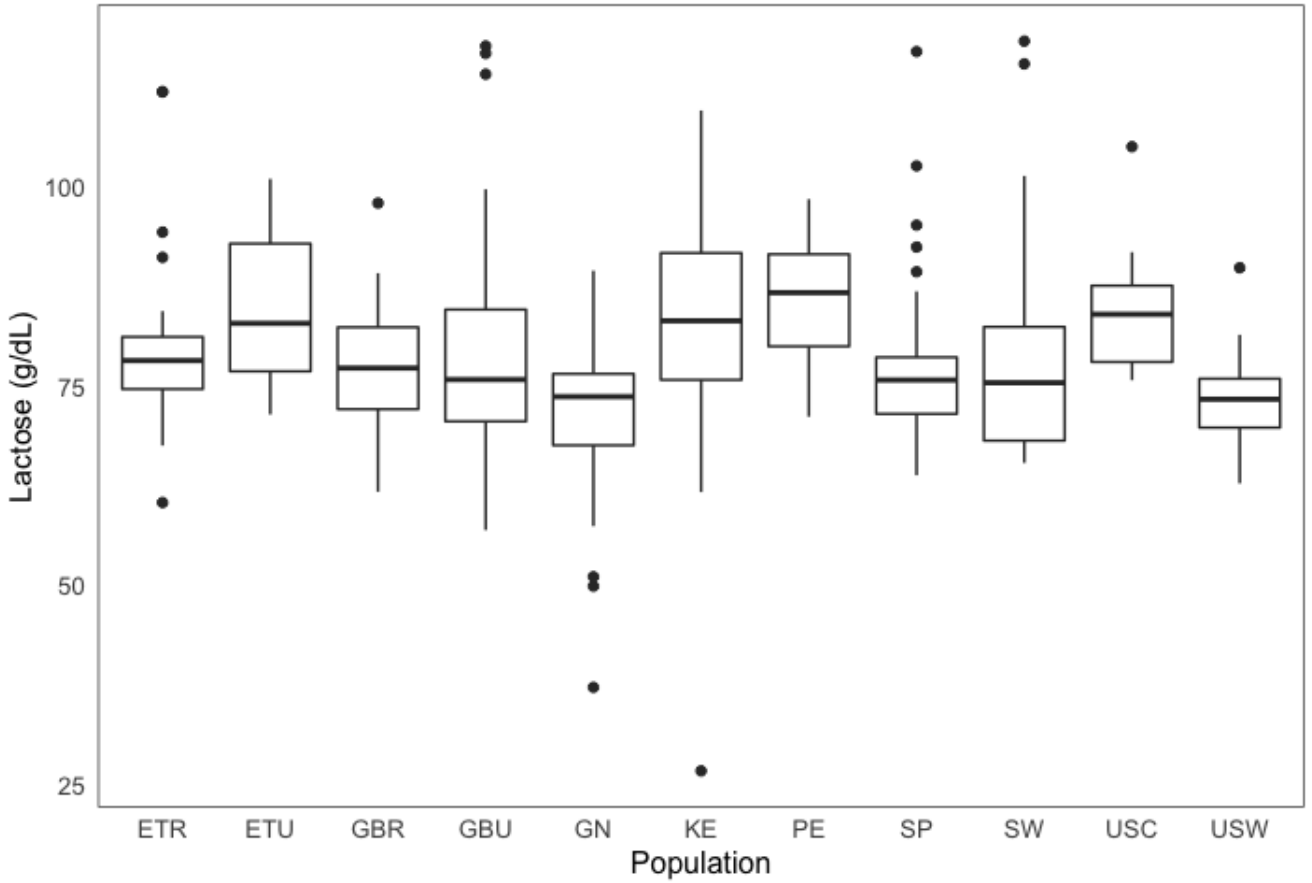


Figure 4.1: Boxplot of lactation concentration (g/dL) by population. Box indicates population median with first and third quartile. Outliers are indicated as dots. ETR - rural Ethiopia; ETU - urban Ethiopia; GBR - rural Gambia; GBU - urban Gambia; GN - Ghana; KE - Kenya; PE - Peru; SP - Spain; SW - Sweden; USC - United States, California; USW - United States, Washington/Idaho.

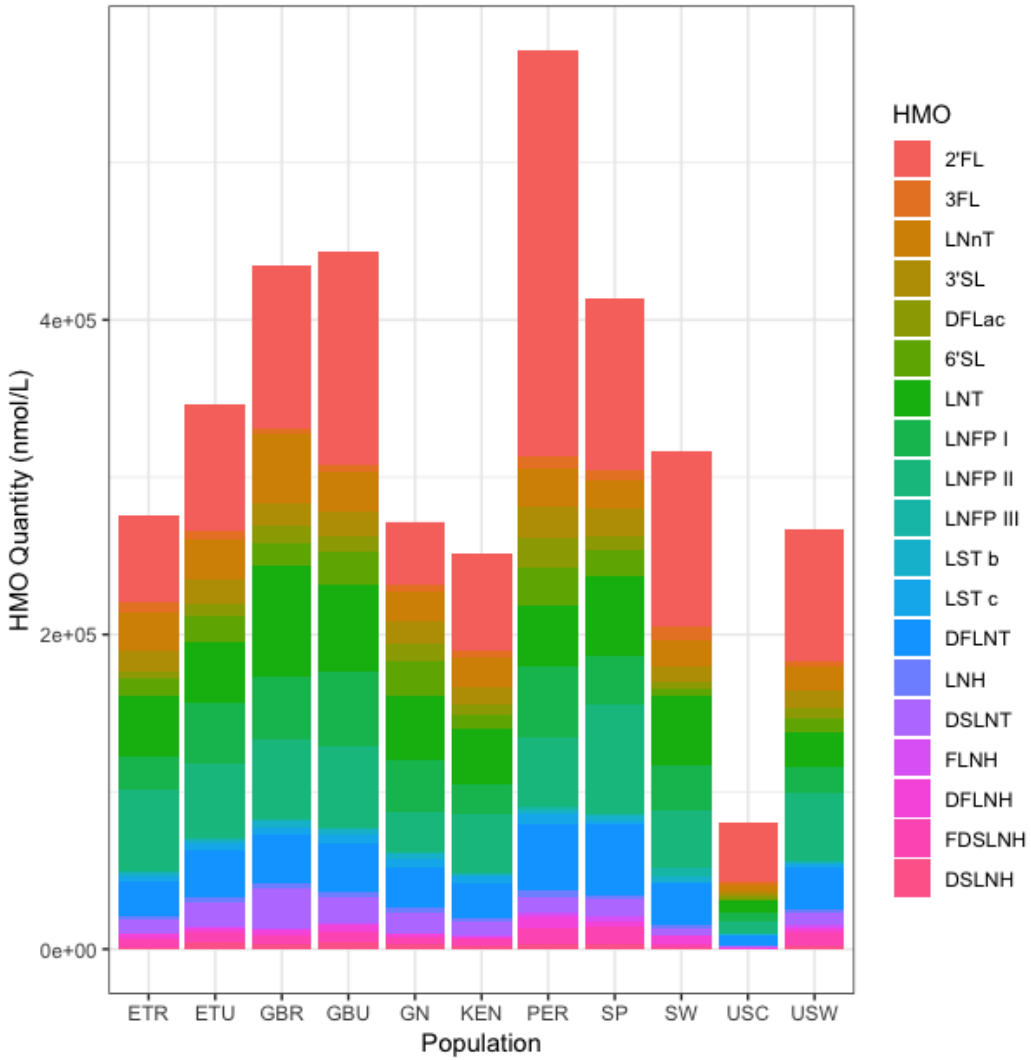


Figure 4.2: Stacked bar chart of HMO abundance (nmol/mL) of 19 HMO moieties averaged within populations with 281 total women represented.

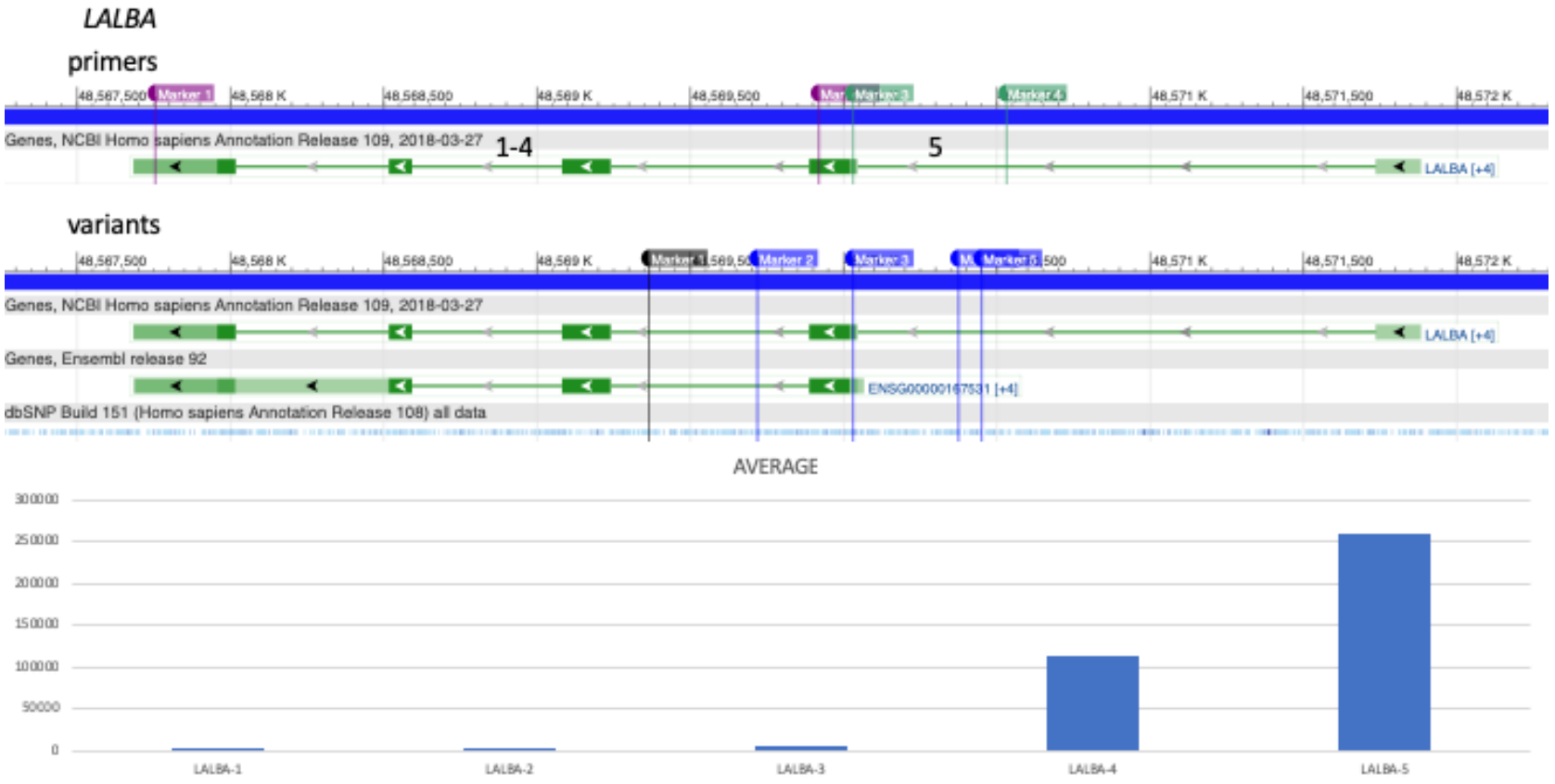


Figure 4.3: Graphic showing position of custom primer pairs for *LALBA* and histogram of total read counts sequenced for each primer pair.

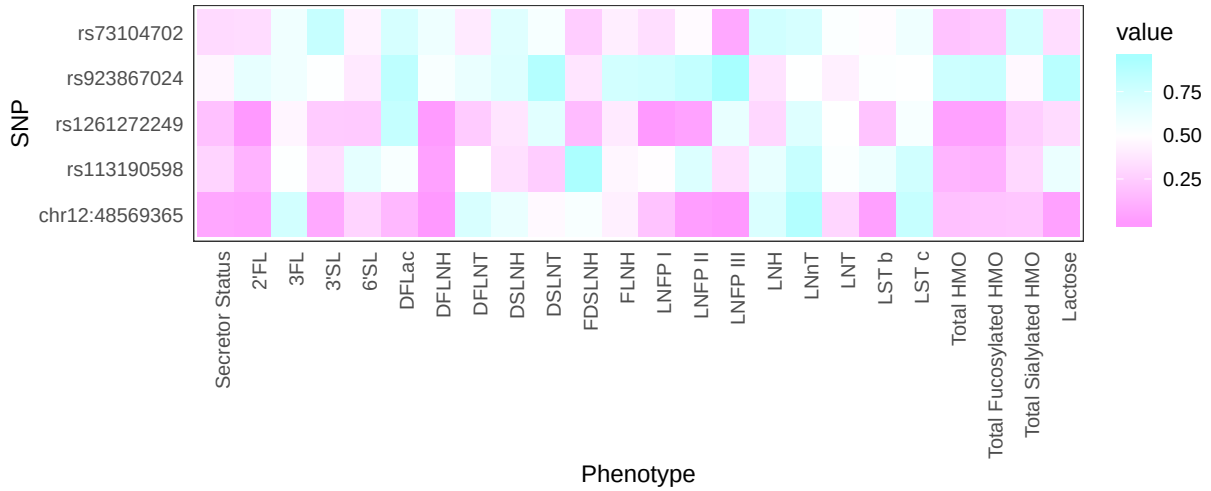


Figure 4.4: Heatmap of association test p-values for all SNPs identified in *LALBA*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

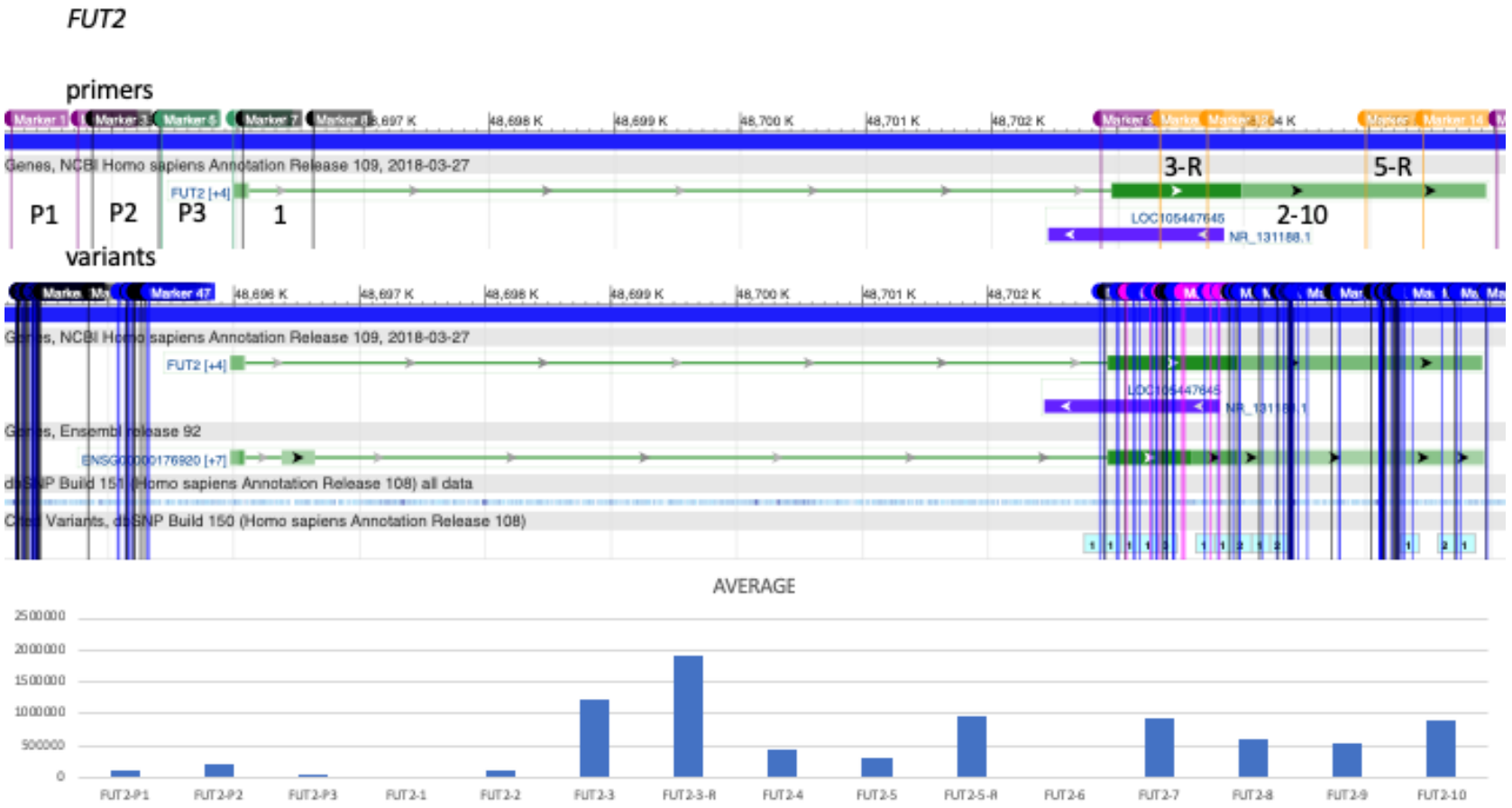


Figure 4.5: Graphic showing position of custom primer pairs for *FUT2* and histogram of total read counts sequenced for each primer pair.

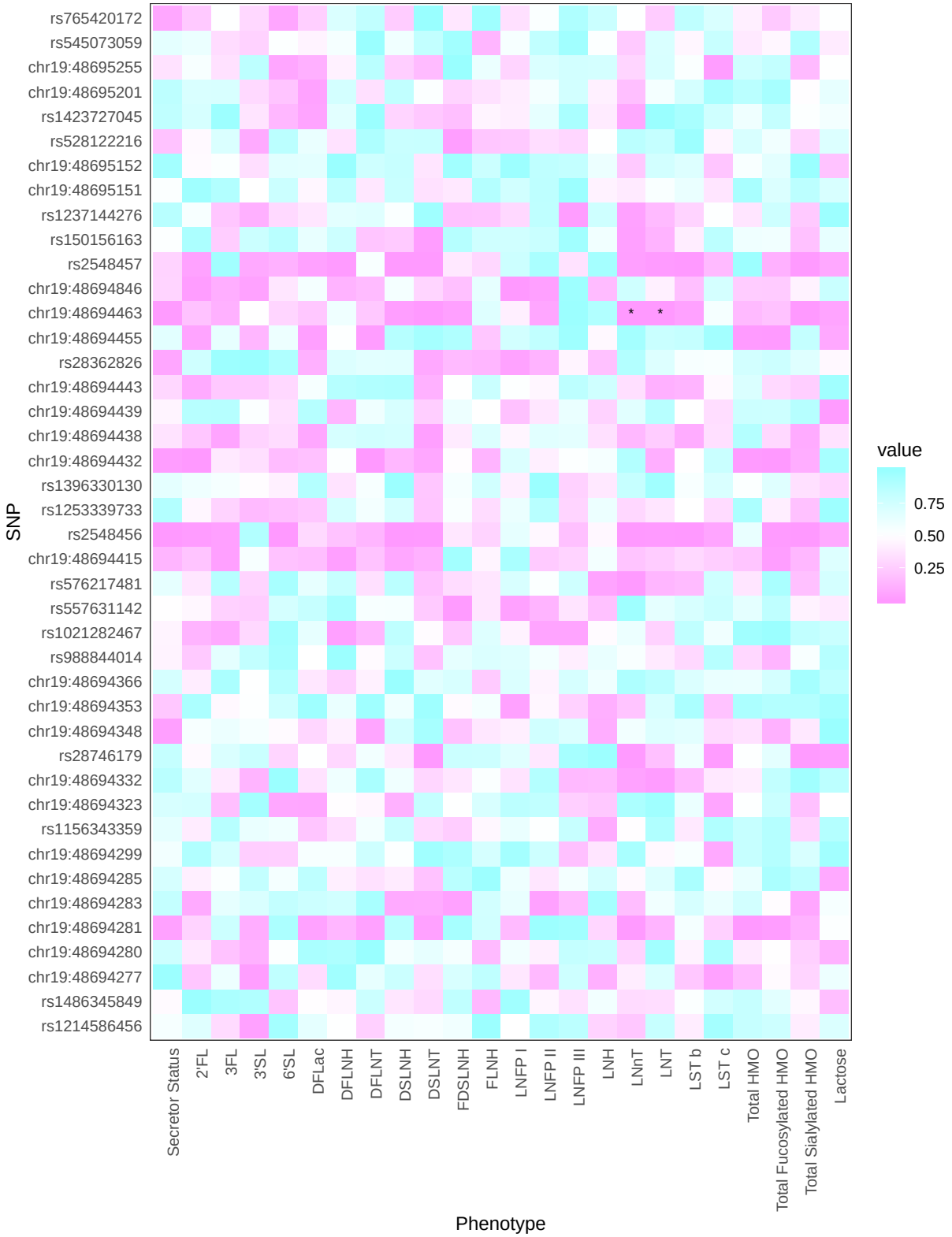


Figure 4.6: Heatmap of association test p-values for all SNPs identified upstream in *FUT2*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

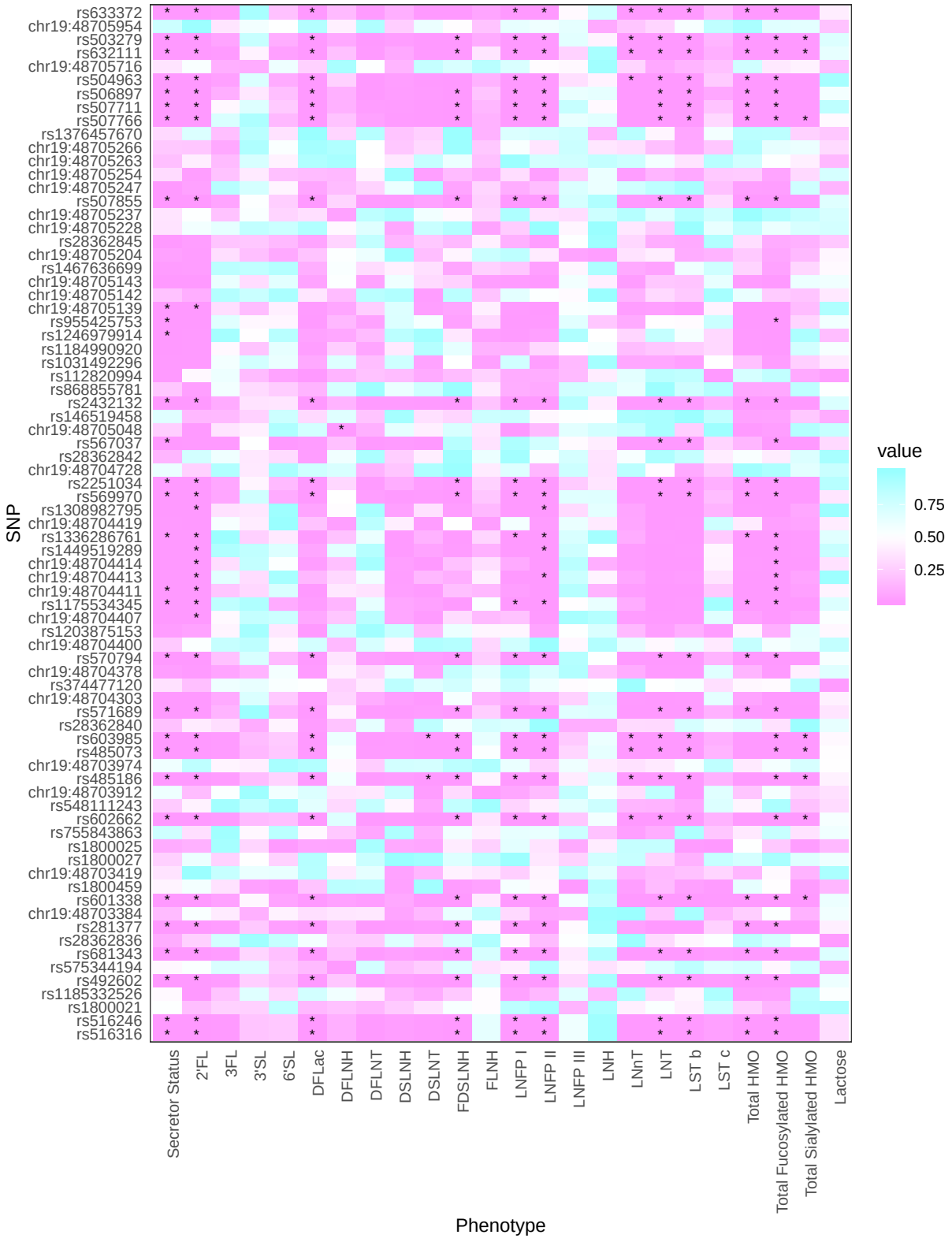


Figure 4.7: Heatmap of association test p-values for all SNPs identified near the coding region in *FUT2*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

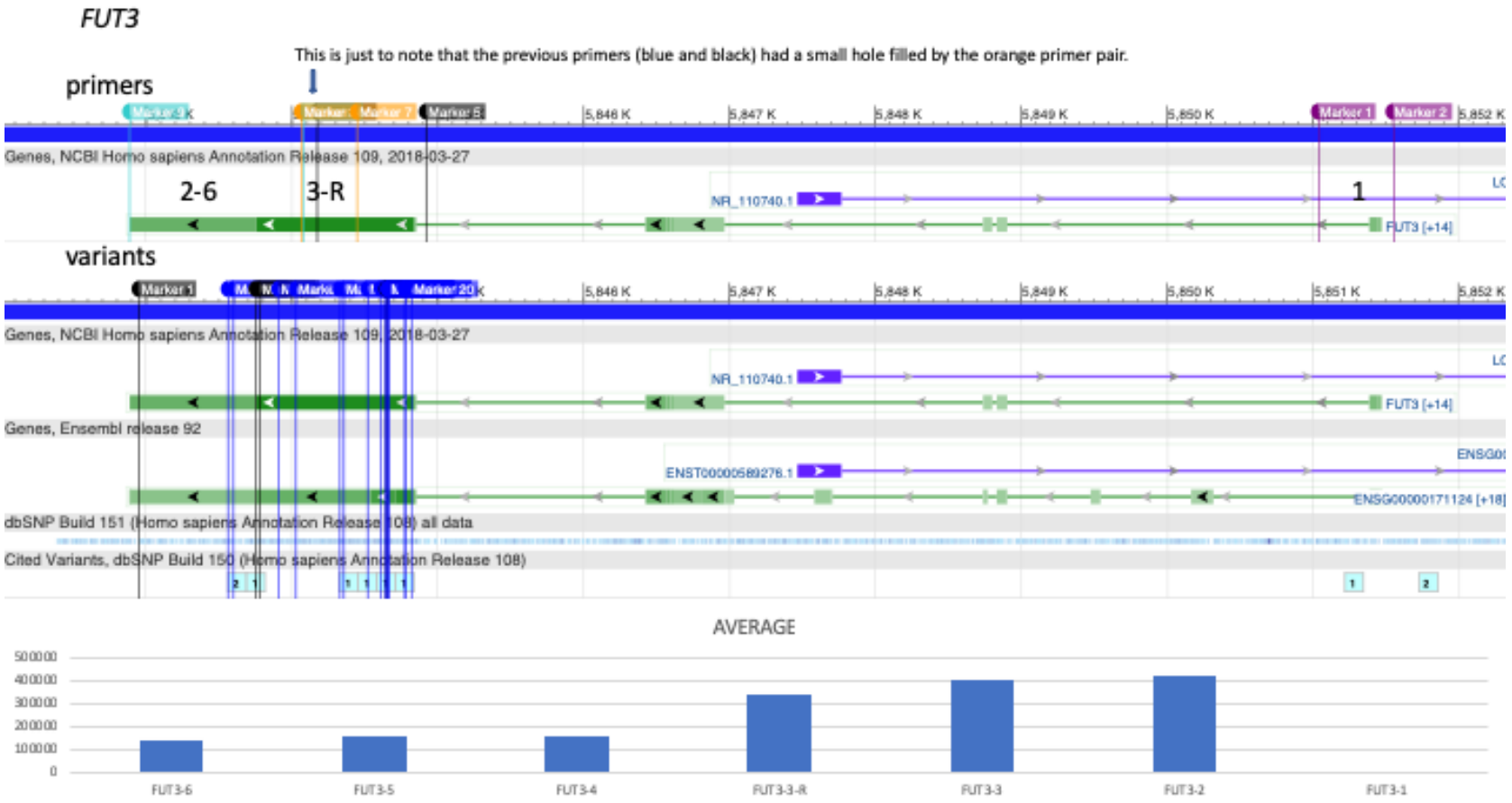


Figure 4.8: Graphic showing position of custom primer pairs for *FUT3* and histogram of total read counts sequenced for each primer pair.

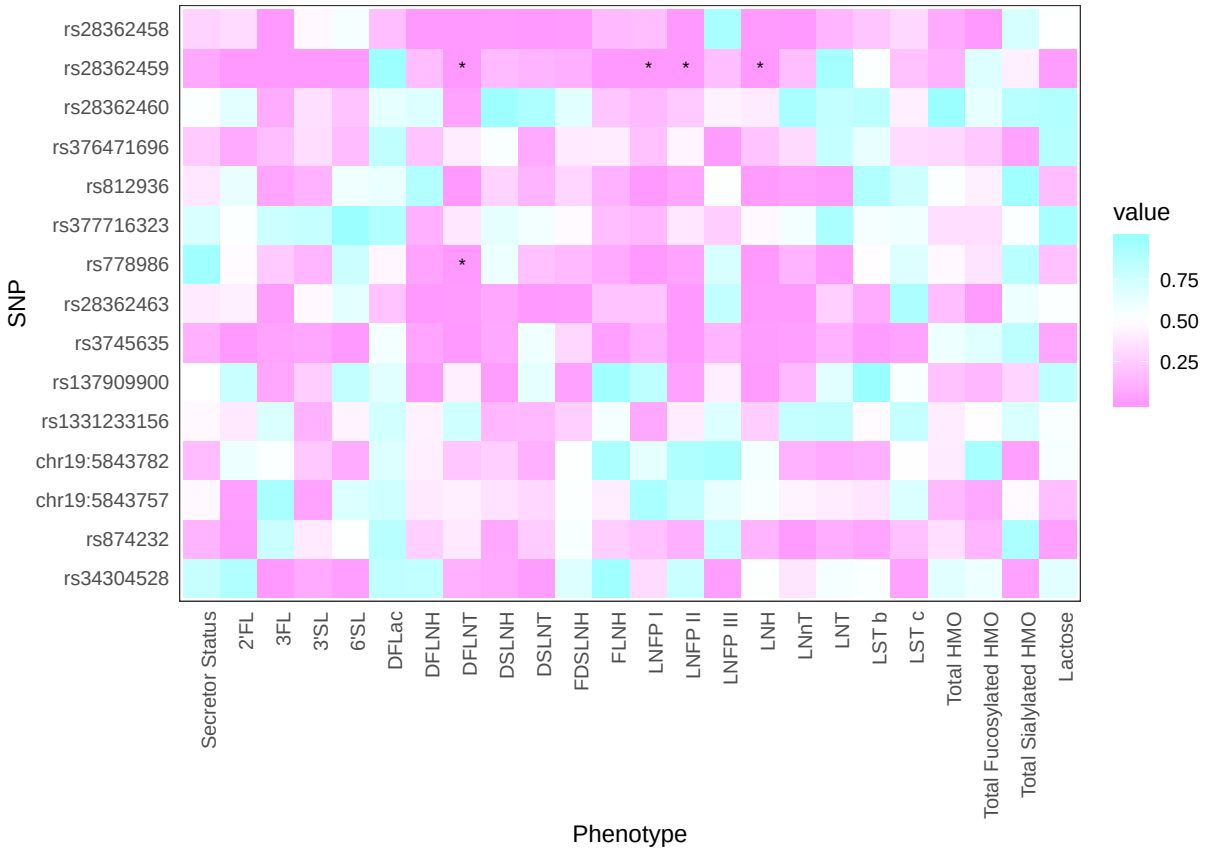


Figure 4.9: Heatmap of association test p-values for all SNPs identified in *FUT3*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

Figure 4.10: Graphic showing position of custom primer pairs for *ST6GalNAc5* and histogram of total read counts sequenced for each primer pair.

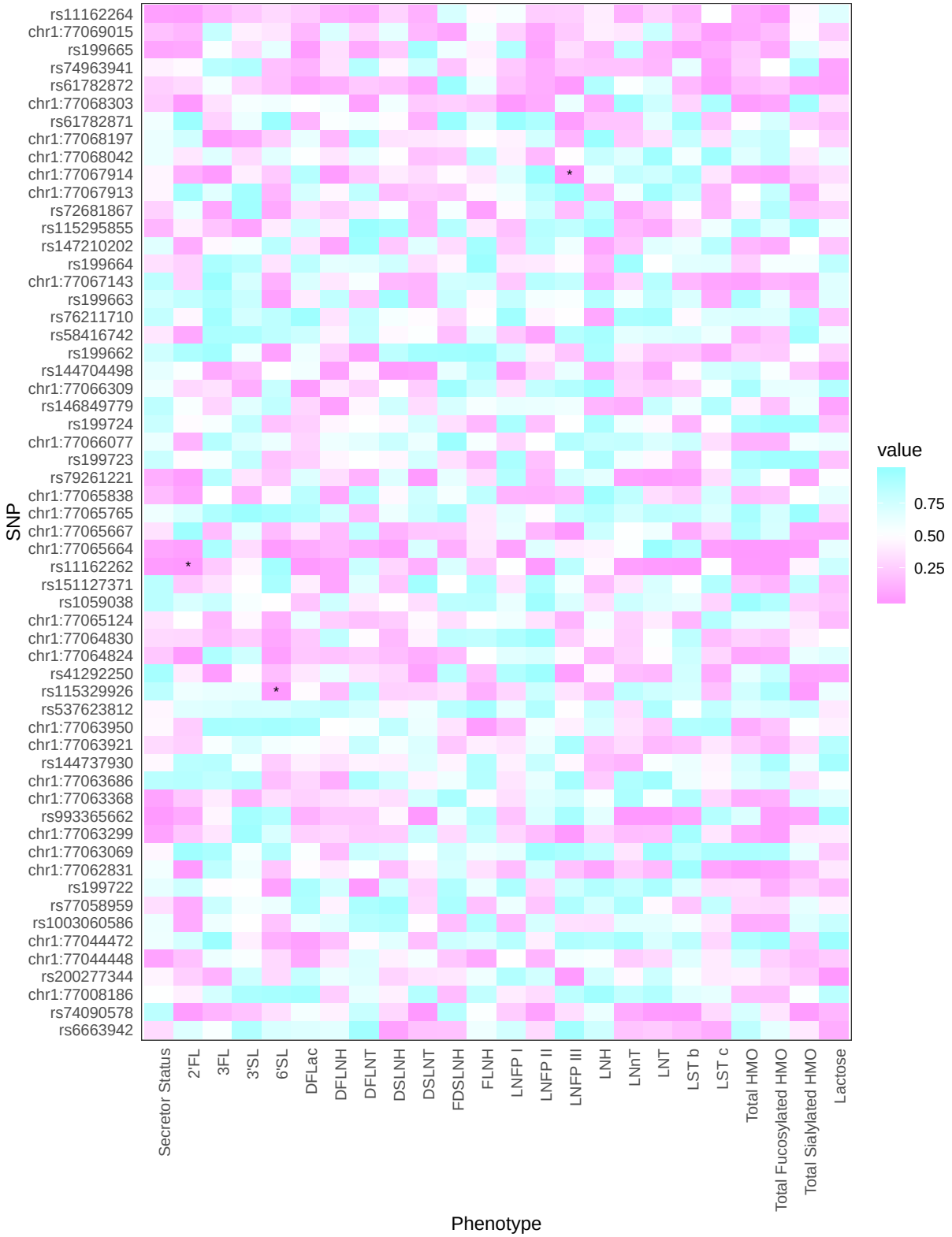


Figure 4.11: Heatmap of association test p-values for all SNPs identified in *ST6GalNAc5*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

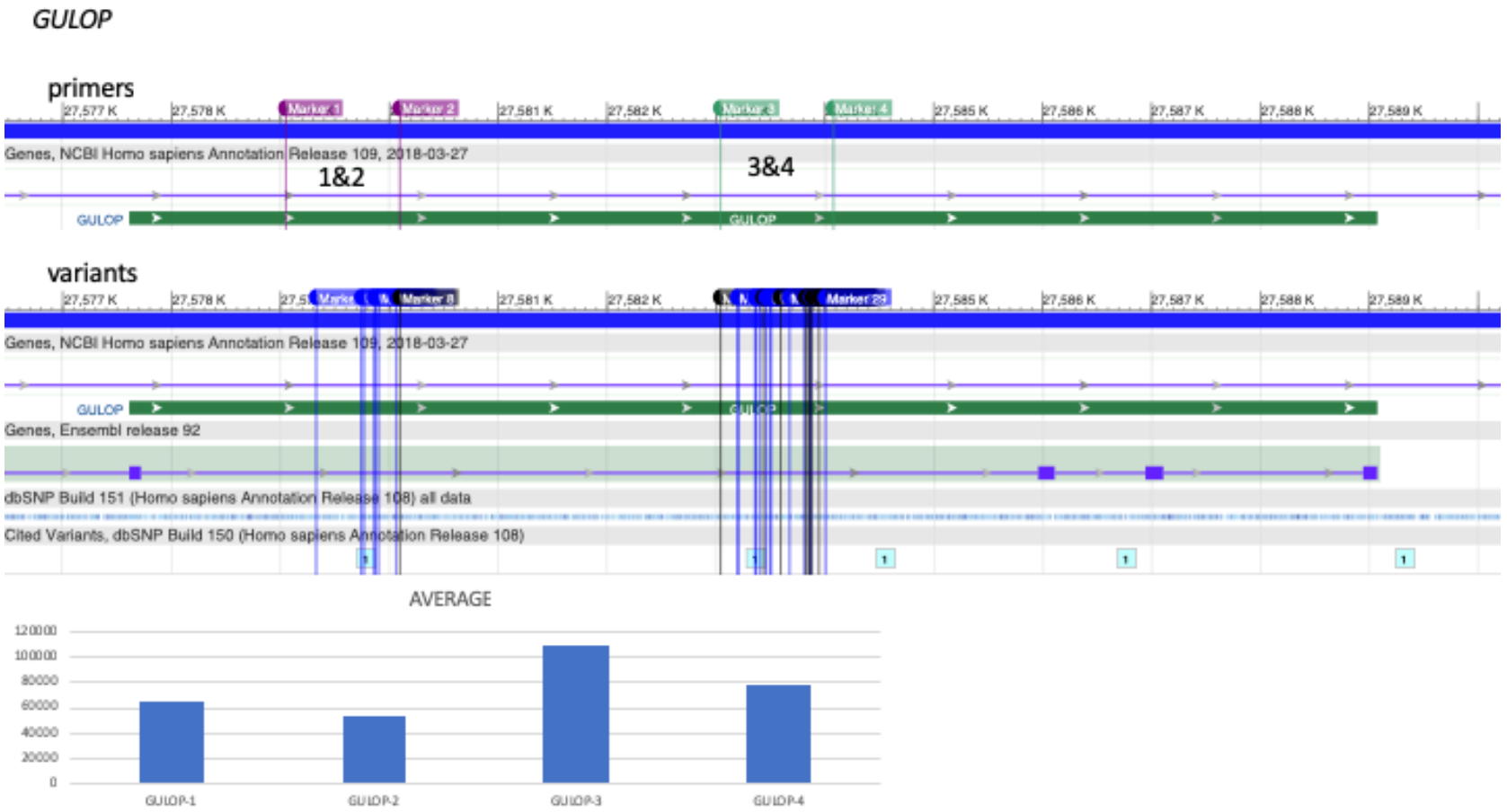


Figure 4.12: Graphic showing position of custom primer pairs for *GULOP* and histogram of total read counts sequenced for each primer pair.

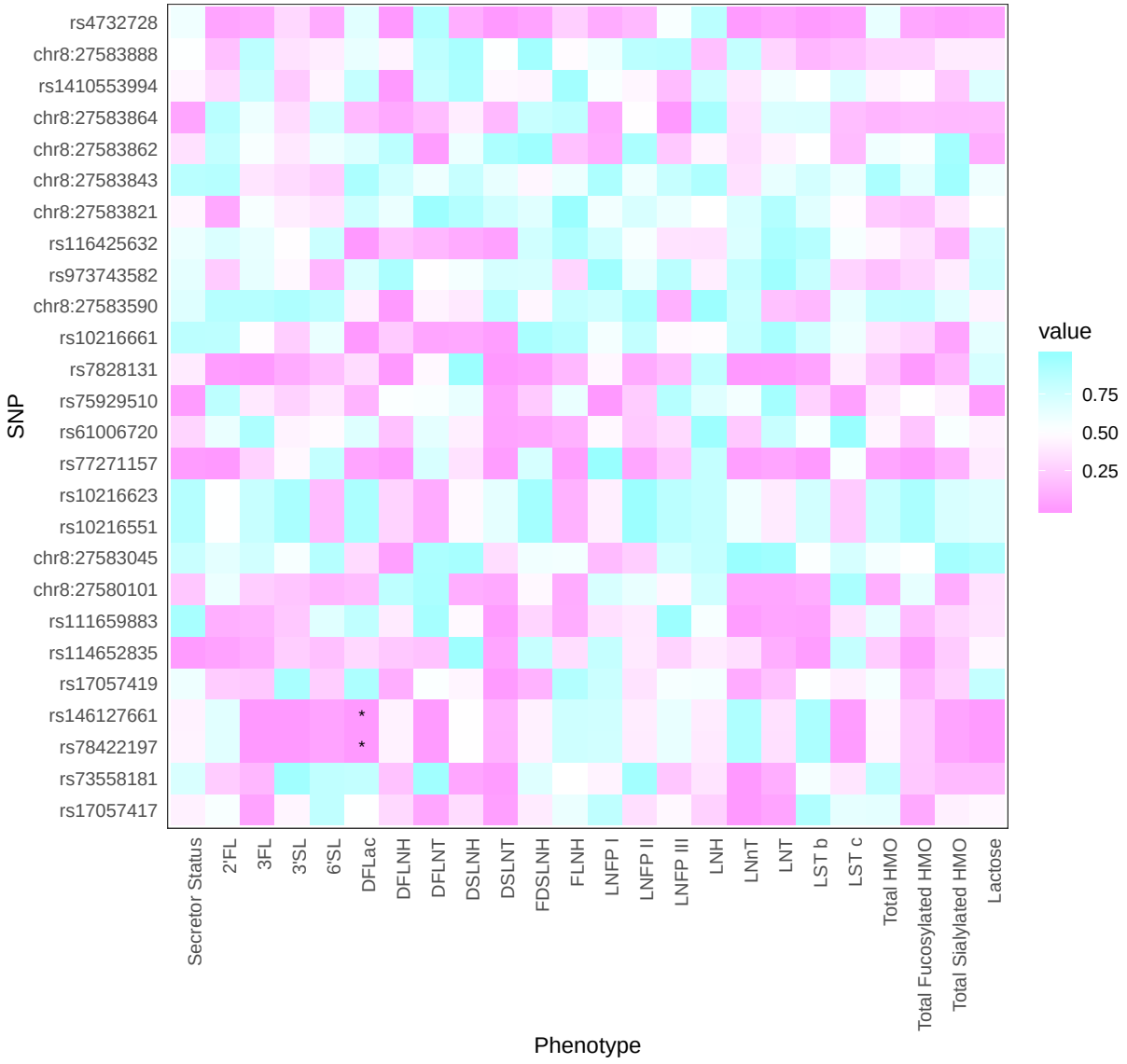


Figure 4.13: Heatmap of association test p-values for all SNPs identified in *GULOP*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

Figure 4.14: Graphic showing position of custom primer pairs for *SEC1P* and histogram of total read counts sequenced for each primer pair.

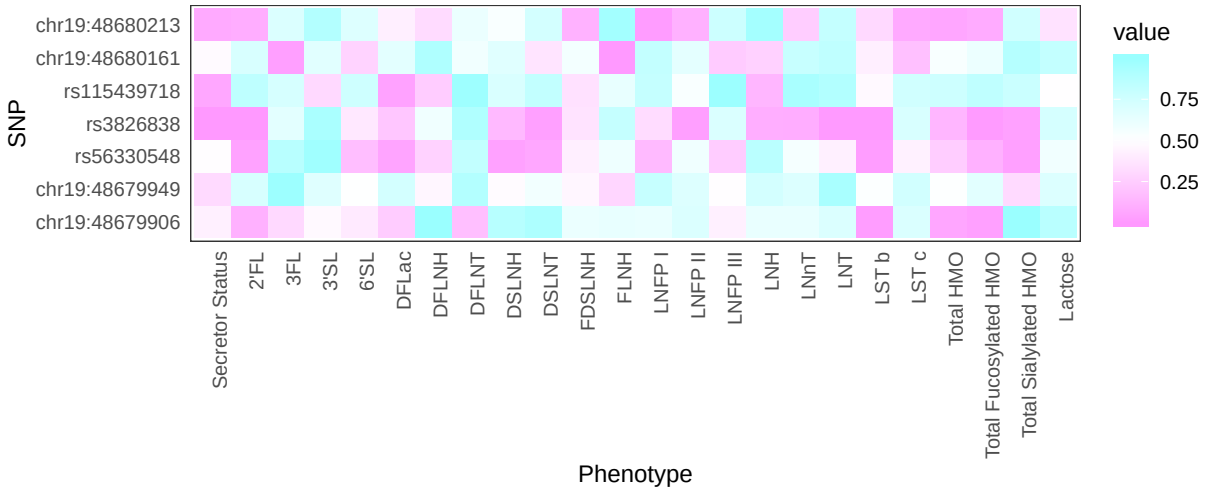


Figure 4.15: Heatmap of association test p-values for all SNPs identified in *SEC1P*. * indicates significance (P <0.01) with a Bonferroni correction for multiple comparisons.

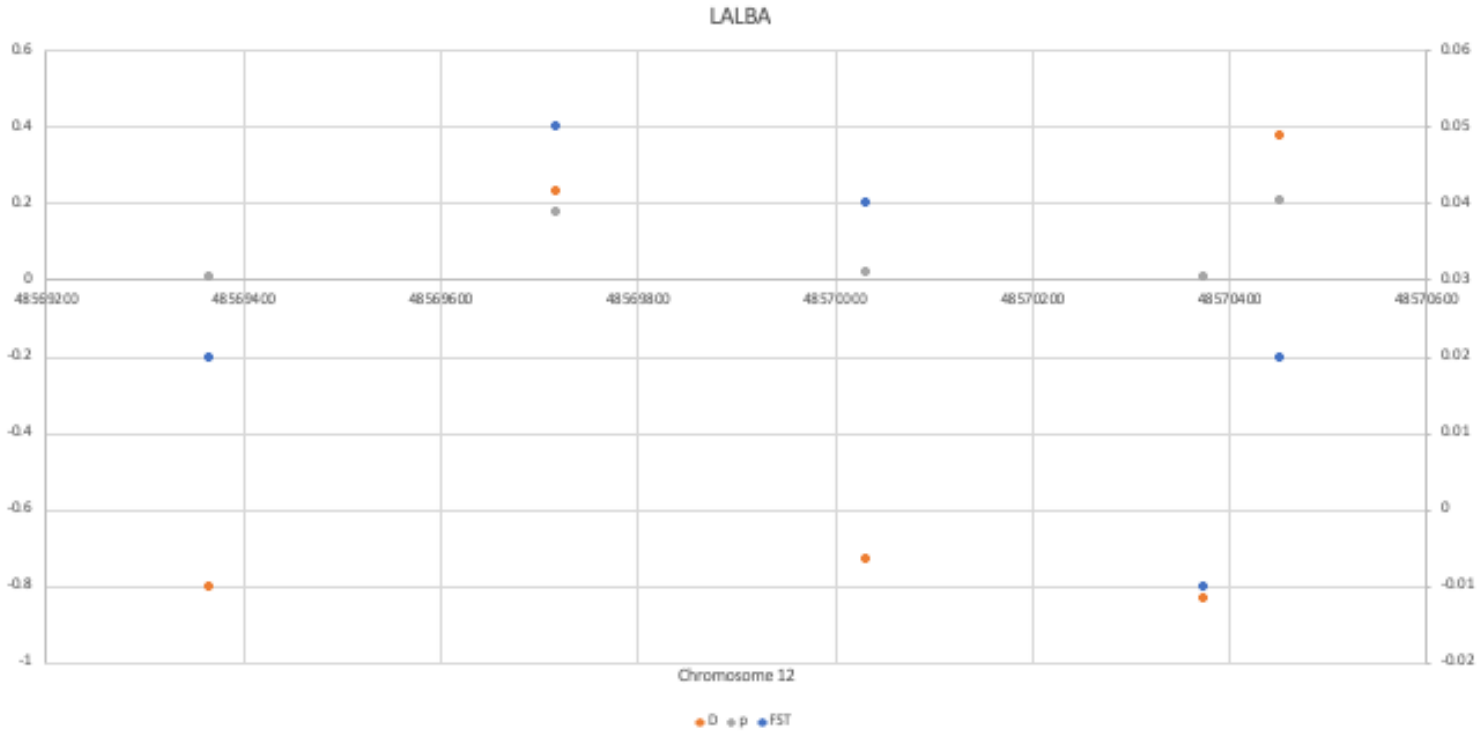


Figure 4.16: Scatterplot of evolutionary statistics by SNP displayed across chromosomal position for *LALBA*. Blue represents F_{ST}, or fixation index (y-axis right), orange represents Tajima's D (D; y-axis left), and grey represents π (p) or nucleotide diversity.

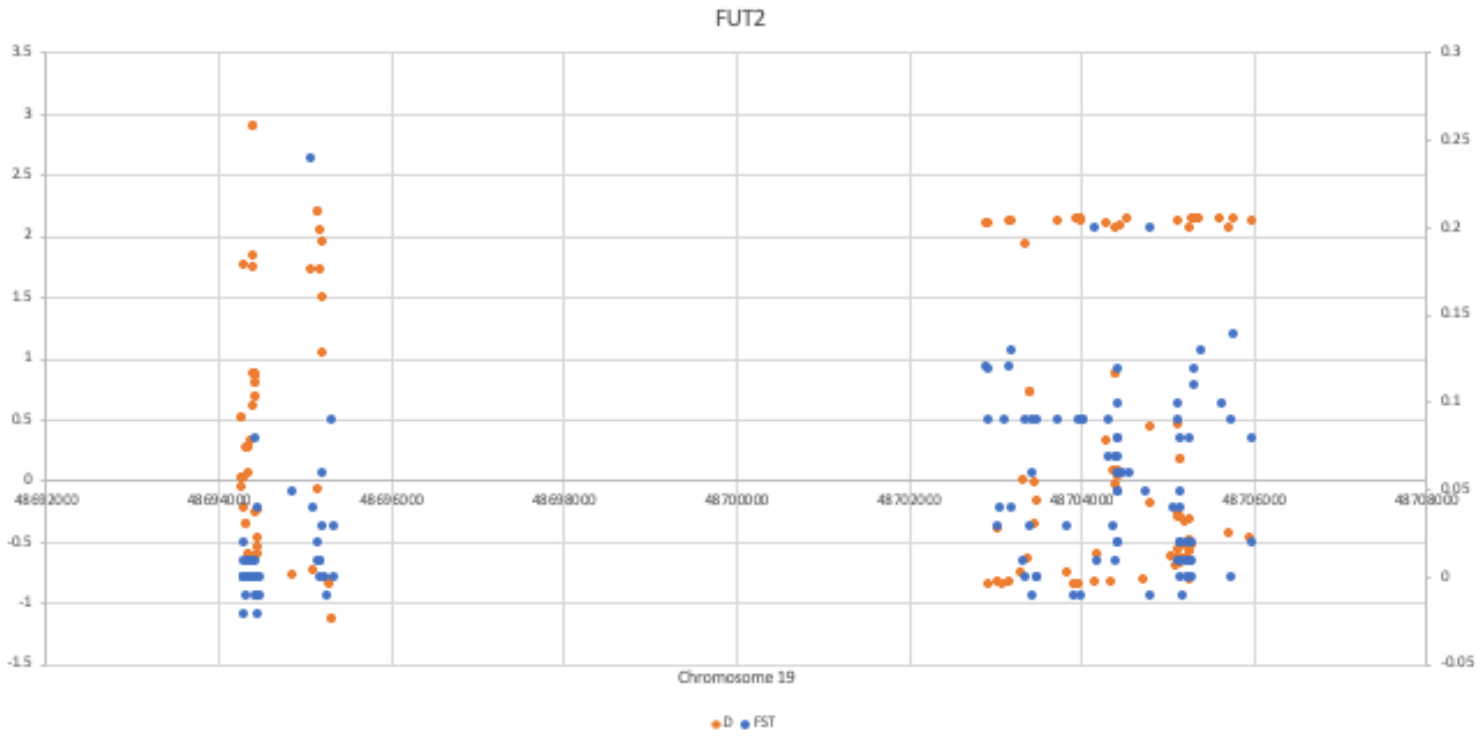


Figure 4.17: Scatterplot of evolutionary statistics by SNP displayed across chromosomal position for *FUT2*. Blue represents F_{ST}, or fixation index (y-axis right), orange represents Tajima's D (D; y-axis left), and grey represents π (p) or nucleotide diversity.

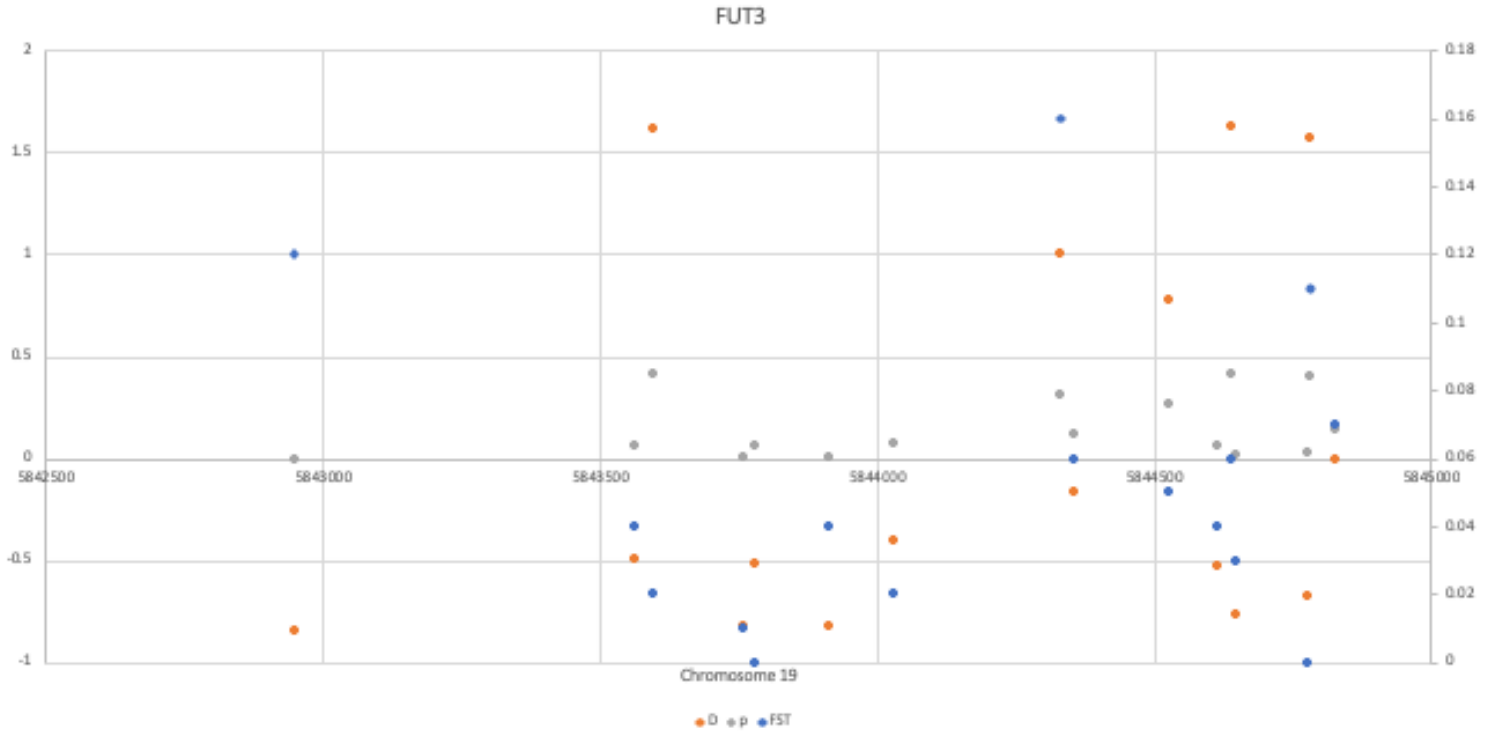


Figure 4.18: Scatterplot of evolutionary statistics by SNP displayed across chromosomal position for *FUT3*. Blue represents F_{ST}, or fixation index (y-axis right), orange represents Tajima's D (D; y-axis left), and grey represents π (p) or nucleotide diversity.

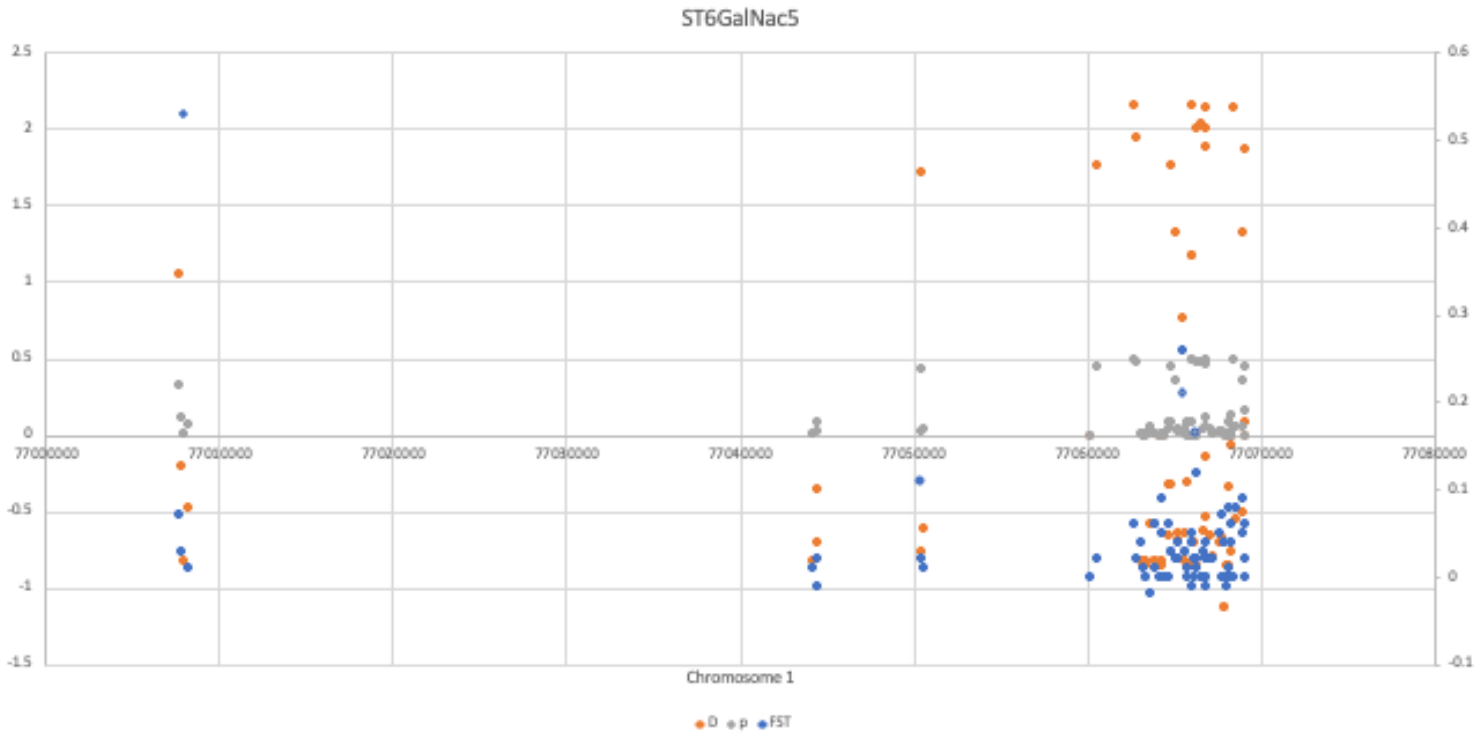


Figure 4.19: Scatterplot of evolutionary statistics by SNP displayed across chromosomal position for *ST6GalNAc5*. Blue represents F_{ST}, or fixation index (y-axis right), orange represents Tajima's D (D; y-axis left), and grey represents π (p) or nucleotide diversity.

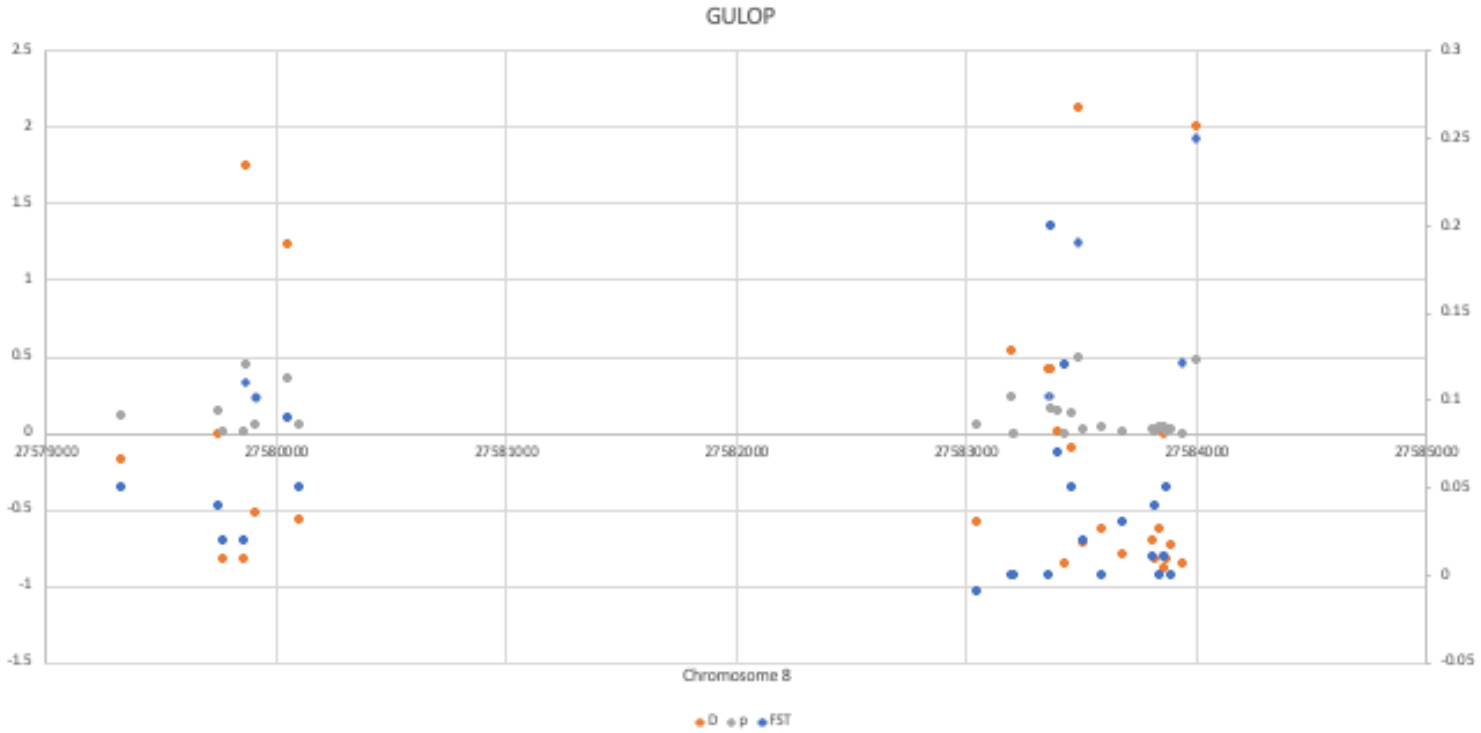


Figure 4.20: Scatterplot of evolutionary statistics by SNP displayed across chromosomal position for *GULOP*. Blue represents F_{ST}, or fixation index (y-axis right), orange represents Tajima's D (D; y-axis left), and grey represents π (p) or nucleotide diversity.

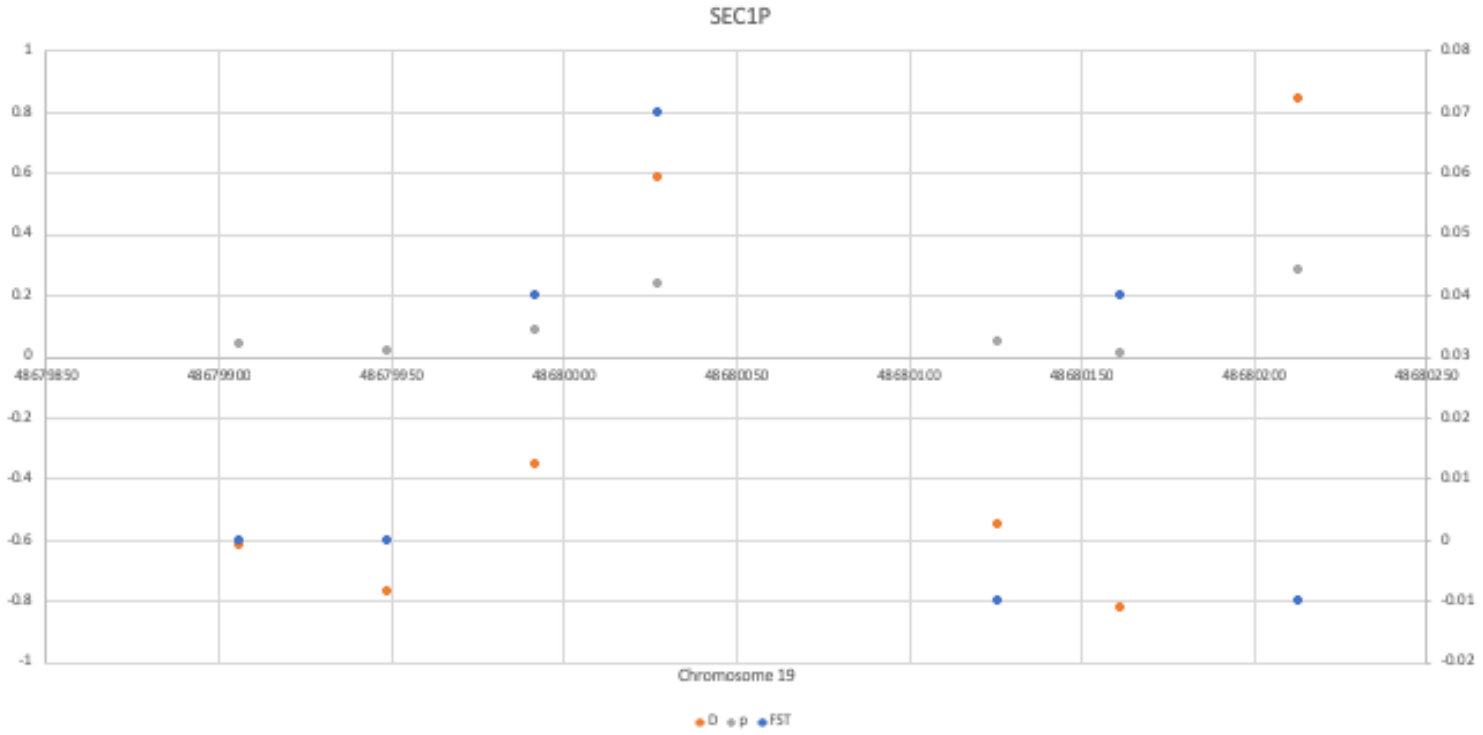


Figure 4.21: Scatterplot of evolutionary statistics by SNP displayed across chromosomal position for *SEC1P*. Blue represents F_{ST}, or fixation index (y-axis right), orange represents Tajima's D (D; y-axis left), and grey represents π (p) or nucleotide diversity.

Table 4.1: Primer sets used for targeted sequencing of human genes involved in milk lactose and oligosaccharide synthesis, and pseudogenes.[1]

Primer Pairs		
Name	"F" Sequence	"R" Sequence
LALBA-1	ACACTGACGACATGGTTCTACATGAGTAAGGGTCTAGAGCTCAGT	TACGGTAGCAGAGACTTGGTCTCGCTCCGCTCCTTGGTAG
LALBA-2	ACACTGACGACATGGTTCTACATTGCTACCAAGGAGCGGAG	TACGGTAGCAGAGACTTGGTCTCGGAGCTGTCTGAGAGTCTTG
LALBA-3	ACACTGACGACATGGTTCTACAAGACTCTCAGACAGCTCCGA	TACGGTAGCAGAGACTTGGTCTACTCTGAGCTTTGCCATCTTTG
LALBA-4	ACACTGACGACATGGTTCTACACTCCAAAGATGGCAAAGCTC	TACGGTAGCAGAGACTTGGTCTAGATGGTTATGGAGGCATCG
LALBA-5	ACACTGACGACATGGTTCTACACCCAAGAACCTGAAATGGAA	TACGGTAGCAGAGACTTGGTCTGCTGGATTGGTTGGACAAGT
FUT2-P1	ACACTGACGACATGGTTCTACATGGAGTTGGAAAGTCTAAGCCT	TACGGTAGCAGAGACTTGGTCTGGAGGACAAAGTTCCCAGTGA
FUT2-P2	ACACTGACGACATGGTTCTACAGAGTGGCTGCTGGTTTCTCT	TACGGTAGCAGAGACTTGGTCTAATACAAATGGGCCCGGTGT
FUT2-P3	ACACTGACGACATGGTTCTACATTGAAATAGGCAGGGCGAGG	TACGGTAGCAGAGACTTGGTCTAGCATCGGTGACTGGCAAA
FUT2-1	ACACTGACGACATGGTTCTACAAACCACTCTGTCCCGGTTTC	TACGGTAGCAGAGACTTGGTCTTCAGAGCGGGATTAGGGAGT
FUT2-2	ACACTGACGACATGGTTCTACACACTATGCCTGCACACCACCG	TACGGTAGCAGAGACTTGGTCTGCGGTATTCCTCCTCCATCCAGTC
FUT2-3	ACACTGACGACATGGTTCTACATGCAGATACCAGTGCTAGCC	TACGGTAGCAGAGACTTGGTCTGGTGACCACGAAGATGAGGG
FUT2-3R	ACACTGACGACATGGTTCTACAACCTGAACGACTGGATGGAG	TACGGTAGCAGAGACTTGGTCTGTCAATGTTCTCCCGACACC
FUT2-3b	ACACTGACGACATGGTTCTACACAGCTCCCTCATCTTCGTGG	TACGGTAGCAGAGACTTGGTCTCTGTTACTTGCAGCCCAACG
FUT2-4	ACACTGACGACATGGTTCTACACCGACCGGCGATACCTAC	TACGGTAGCAGAGACTTGGTCTCACAGAAGAGAGATGGGTCCT
FUT2-5	ACACTGACGACATGGTTCTACATGGGCTGCAAGTAACAG	TACGGTAGCAGAGACTTGGTCTGGATGATACAGCTAAGAAATGG
FUT2-6	ACACTGACGACATGGTTCTACAGTCACCTGAGCTCCATCCAT	TACGGTAGCAGAGACTTGGTCTGCTCTCTGGGTGGACACAAT
FUT2-6b	TGTCCTTGGCATTGTGTCCAACACTGACGACATGGTTCTACA	CCCTGAAGTTCACTGAGCCATACGGTAGCAGAGACTTGGTCT
FUT2-7	ACACTGACGACATGGTTCTACAAACCAGGTGTCCTTGGCATT	TACGGTAGCAGAGACTTGGTCTCATCCCAGGCCCTAGAAAGC
FUT2-7R	ACACTGACGACATGGTTCTACAGGTTGTTCACTGCAGGAAGT	TACGGTAGCAGAGACTTGGTCTAAACAGTGAGCTCTAGGGCC
FUT2-8	ACACTGACGACATGGTTCTACATGGCATGCATCCAAGTCCAT	TACGGTAGCAGAGACTTGGTCTTTCTCCCATCCGCAAAGTCA
FUT2-9	ACACTGACGACATGGTTCTACACTGCCTCTTCAATCCTGGCTTTCTA	TACGGTAGCAGAGACTTGGTCTCTGGCCTAAGATGTATTTTGGAGTT
FUT2-10	ACACTGACGACATGGTTCTACAGCATTGTGTCCACCCAGAGA	TACGGTAGCAGAGACTTGGTCTACGCCCTGAAGTTCACTGAG
FUT3-P2	CTGTCCTCATCCACTGCTCGACACTGACGACATGGTTCTACA	ATGCCACACCCAGAAAGACCTACGGTAGCAGAGACTTGGTCT
FUT3-1	ACACTGACGACATGGTTCTACACCCTGCTGCTGGGGAGAACA	TACGGTAGCAGAGACTTGGTCTGGGGAGGCTGTTGATGGGGT

FUT3-2	ACACTGACGACATGGTTCTACAGGACTCATGGCCCGGAGCTT	TACGGTAGCAGAGACTTGGTCTGGCTTCCAGGTGCTGGCAGTTA
FUT3-3	ACACTGACGACATGGTTCTACACCTGATCCTGCTACGGACAT	TACGGTAGCAGAGACTTGGTCTGTCCACCTTGAGATGAGCCT
FUT3-3R	ACACTGACGACATGGTTCTACATTCTCGGTGATGTAGTCGGG	TACGGTAGCAGAGACTTGGTCTCGCTGGATCTGGTTCAACTT
FUT3-4	ACACTGACGACATGGTTCTACACCCGACTACATCACCGAGAA	TACGGTAGCAGAGACTTGGTCTTAGCAGGCAAGTCTTCTGGA
FUT3-5	ACACTGACGACATGGTTCTACAGGCCAACCCTCTCTCTTACC	TACGGTAGCAGAGACTTGGTCTGGACCAACCCCTCTAGAGTG
FUT3-6	ACACTGACGACATGGTTCTACAGGCCCTGGGGAACTGGCTTA	TACGGTAGCAGAGACTTGGTCTGTGCCTGGCCGGCCTATTATTT
ST6-1	ACACTGACGACATGGTTCTACACGGACCAAGAAGTGGGTACA	TACGGTAGCAGAGACTTGGTCTCCTTTGTAACAGGCGACCTC
ST6-2	ACACTGACGACATGGTTCTACACGCCCCAAATCTCCCCCACT	TACGGTAGCAGAGACTTGGTCTTACTCACAGCGCCTCCCAGC
ST6-3	ACACTGACGACATGGTTCTACAAGGCAGATGCAAACCTGTGGGAG	TACGGTAGCAGAGACTTGGTCTATGCAGAGAGTCCCAGCCCC
ST6-4	ACACTGACGACATGGTTCTACATGCTGAGTGAGGGTTAAGCC	TACGGTAGCAGAGACTTGGTCTGCATCTTCTTCCCTGTGCCT
ST6-5	ACACTGACGACATGGTTCTACAGTTACTTTACCAGCTTGCAGACT	TACGGTAGCAGAGACTTGGTCTAAAGCCCTCACTGTCAAAGG
ST6-6	ACACTGACGACATGGTTCTACATCCCCACCTGCTTCCTCTAA	TACGGTAGCAGAGACTTGGTCTTGTCCAAGCAGCAGAAAAGTG
ST6-7	ACACTGACGACATGGTTCTACACTCTGGAAGCTTAGTGGAGTGG	TACGGTAGCAGAGACTTGGTCTATTGAATGTCCGTGCCCAGT
ST6-8	ACACTGACGACATGGTTCTACAGGGCAGTCATCACCGCTTTA	TACGGTAGCAGAGACTTGGTCTGCTTATGTTGCAGTTTCAAAGGC
ST6-9	ACACTGACGACATGGTTCTACAAAGTACCATGGACAGACGCC	TACGGTAGCAGAGACTTGGTCTGGCCATTTCCAACCATCATTCA
ST6-9b	TACCATGGACAGACGCCTACACACTGACGACATGGTTCTACA	TTCTAGGCCATTTCCAACCATTACGGTAGCAGAGACTTGGTCT
ST6-10	ACACTGACGACATGGTTCTACAAGCACATCTCCACTGACTTTCA	TACGGTAGCAGAGACTTGGTCTGGGTCAAGAAGAGGCCACAA
ST6-11	ACACTGACGACATGGTTCTACAGGCAGAATGATGTCACTGTACC	TACGGTAGCAGAGACTTGGTCTGAGTTGAATGATAGGGCATGTCC
ST6-12	ACACTGACGACATGGTTCTACATCTCGAAGCATCACCATCCG	TACGGTAGCAGAGACTTGGTCTTCATAAGCCAGGTCATGTCTGT
ST6-13	ACACTGACGACATGGTTCTACATGGTCTGAAGGAGTATAAAGGAC	TACGGTAGCAGAGACTTGGTCTTGTGTTCAGAGTCCATGCTGT
ST6-14	ACACTGACGACATGGTTCTACATGCTCAAACATAAGTGGCTCCT	TACGGTAGCAGAGACTTGGTCTACAGTCATGAAATGGTAGTGGGA
ST6-15	ACACTGACGACATGGTTCTACATCGTTTGTAACTTCTCCCTCTCA	TACGGTAGCAGAGACTTGGTCTAGGAGGGAGCTGATGACACA
ST6-15b	CCATCTCGTTTGTAACTTCTCCCACACTGACGACATGGTTCTACA	CCTCTGCCAAATGTTCTTGTGTTTACGGTAGCAGAGACTTGGTCT
ST6-16	ACACTGACGACATGGTTCTACATCCCACTACCATTTCATGACTGT	TACGGTAGCAGAGACTTGGTCTGGAGTACACACCAGGGTTAGC
ST6-16b	CTGTGTCATCAGCTCCCTCCACACTGACGACATGGTTCTACA	ACCTAAGGATTGTGAGAGACCATACGGTAGCAGAGACTTGGTCT
ST6-17	ACACTGACGACATGGTTCTACAGGCAACTAAGTCCCTGTGCT	TACGGTAGCAGAGACTTGGTCTCAAGGGCAAATATTTAGGCAGGT
ST6-17b	TGCTTTGCTAACCCTGGTGTACACTGACGACATGGTTCTACA	AGGAACATGGACCCTTCTTTCATACGGTAGCAGAGACTTGGTCT
ST6-18a	ACACTGACGACATGGTTCTACATGAATGACATGAGAATGGAGGGA	TACGGTAGCAGAGACTTGGTCTCCTTAACAGCACCAAAGACTGC

ST6-18b	ACACTGACGACATGGTTCTACATCTGAAAGAAGGGTCCATGTTCC	TACGGTAGCAGAGACTTGGTCTAGTTCTCTTTGACGTGGCCT
ST6-19	ACACTGACGACATGGTTCTACATCTCTCTCTGCAGTGTTGGG	TACGGTAGCAGAGACTTGGTCTCACTCTCACCTCTTTCCATTCCA
ST6-20	ACACTGACGACATGGTTCTACATCAGAAGGCCACGTCAAAGA	TACGGTAGCAGAGACTTGGTCTCCAGTGGGTCAACATTAAGAGGT
ST6-21	ACACTGACGACATGGTTCTACATGTCTCTGCAGCCTATACCTAGT	TACGGTAGCAGAGACTTGGTCTAGCTGTCAAGTTTAAAGCAGGTT
ST6-21b	CTCCTTTGTCTCTGCAGCCTATACACTGACGACATGGTTCTACA	TGTCGTTGAACCACAGAAGTCATACGGTAGCAGAGACTTGGTCT
ST6-22	ACACTGACGACATGGTTCTACAGCAGAAAGGCGGAAGAACAG	TACGGTAGCAGAGACTTGGTCTGTCCACATGCTCAAGAAGGGA
ST6-23	ACACTGACGACATGGTTCTACAGCATGTGGACACTCTAGAATTCC	TACGGTAGCAGAGACTTGGTCTTCTGTTCTCATCTTCATCACCCA
ST6-24	ACACTGACGACATGGTTCTACAGTCCTGTACCATGTGACTAGCT	TACGGTAGCAGAGACTTGGTCTGCTTCTAACCTGCCTTTCCCA
ST6-25	ACACTGACGACATGGTTCTACACAGCATGTCTTGTACCGAGC	TACGGTAGCAGAGACTTGGTCTTCAGGTAGCTTCAAAAGTAGTGA
ST6-26	ACACTGACGACATGGTTCTACAGTCCTGTACCATGTGACTAGCT	TACGGTAGCAGAGACTTGGTCTACACTCCAGAGACAGACCCT
GULOP-1	ACACTGACGACATGGTTCTACACTAGATGACTCAGGGTGCCA	TACGGTAGCAGAGACTTGGTCTGGAGTTGGACTGGCTGGAAA
GULOP-2	ACACTGACGACATGGTTCTACACCCTCTCTCATACCCAGCAG	TACGGTAGCAGAGACTTGGTCTATGATTACCTTGGCTGCTCG
GULOP-3	ACACTGACGACATGGTTCTACAAGACGTGAGCTACTGAACCC	TACGGTAGCAGAGACTTGGTCTTCCCTGCCCAGATCCAATAC
GULOP-4	ACACTGACGACATGGTTCTACATGAAAGGTGCTGGGAAGTGA	TACGGTAGCAGAGACTTGGTCTGGGCATCAGGTCTGGGTATA
SEC1P-2	ACACTGACGACATGGTTCTACAGTCAGGATTTCGAGACTGGC	TACGGTAGCAGAGACTTGGTCTCAGAACCACCCAATGAAGCC
SEC1P-3	ACACTGACGACATGGTTCTACATCAGGACAGAGGCTTGGATG	TACGGTAGCAGAGACTTGGTCTCCAGTCGTTCAGGTGGTAGT
SEC1P-4	ACACTGACGACATGGTTCTACACCCATCTTCAGAATCACCCTG	TACGGTAGCAGAGACTTGGTCTCTGTGTGAGCAGTGCGAAG

[1] *LALBA* - alpha-lactalbumin, *FUT2* - fucosyltransferase 2, *FUT3* - fucosyltransferase 3, *ST6GalNAc5* - ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase, *GULOP* - Gulonolactone (L-) oxidase, pseudogene, *SEC1P* - Secretory blood group 1, pseudogene

Table 4.2: Characteristics of the women (n = 281) participating in this study and effects of population. Values represent means ± SEM. Values in a row not sharing a common superscript differ (P < 0.05) using Bonferroni correction procedures for multiple comparisons.[1]

		Age (y)	Parity (#)	Time postpartum (d)	Weight (kg)	Height (cm)	BMI (kg/m^2)[2]
rural Ethiopia	(n = 30)	25.6 ± 1.0^{bc}	4.0 ± 0.4^{a}	67.4 ± 5.8	51.2 ± 1.6^{c}	154.4 ± 1.0^{d}	21.4 ± 0.5^{ab}
urban Ethiopia	(n = 28)	21.8 ± 0.6^{c}	1.7 ± 0.2^{b}	59.0 ± 2.4	55.6 ± 1.3^{bc}	158.3 ± 1.2^{cd}	22.2 ± 0.5^{ab}
rural Gambia	(n = 31)	26.3 ± 1.3^{bc}	3.9 ± 0.6^{a}	63.5 ± 3.1	55.9 ± 1.5^{bc}	162.6 ± 1.2^{abcd}	21.1 ± 0.5^{b}
urban Gambia	(n = 33)	26.6 ± 0.9^{bc}	3.1 ± 0.4^{ab}	61.8 ± 3.4	65.8 ± 1.8^{ab}	167.6 ± 1.2^{a}	23.4 ± 0.6^{ab}
Ghana	(n = 24)	30.5 ± 1.0^{ab}	2.5 ± 0.2^{ab}	59.0 ± 4.0	66.0 ± 2.4^{ab}	159.7 ± 1.3^{bcd}	25.9 ± 0.8^{ab}
Kenya	(n = 20)	25.5 ± 1.0^{bc}	2.3 ± 0.2^{ab}	73.1 ± 5.2	58.7 ± 2.2^{abc}	159.3 ± 1.4^{bcd}	23.1 ± 0.9^{ab}
Peru	(n = 39)	27.3 ± 1.0^{bc}	2.1 ± 0.2^{b}	61.3 ± 2.7	66.0 ± 2.1^{ab}	152.6 ± 0.8^{d}	28.3 ± 0.8^{a}
Spain	(n = 31)	34.1 ± 0.7^{a}	1.3 ± 0.1^{b}	66.8 ± 4.4	64.7 ± 1.8^{ab}	165.3 ± 1.3^{abc}	23.7 ± 0.7^{ab}
Sweden	(n = 20)	30.8 ± 1.2^{ab}	1.7 ± 0.2^{b}	49.8 ± 4.4	73.3 ± 2.7^{a}	168.5 ± 1.3^{a}	25.9 ± 1.1^{ab}
US Washington	(n = 20)	29.7 ± 1.1^{ab}	2.0 ± 0.2^{b}	64.9 ± 4.0	79.0 ± 3.3^{a}	167.3 ± 1.3^{ab}	28.3 ± 1.3^{a}
US California	(n = 5)	27.6 ± 1.5^{abc}	1.2 ± 0.2^{b}	82.5 ± 7.3	80.0 ± 5.4^{a}	164.9 ± 2.4^{abcd}	29.5 ± 2.2^{a}

[1]Values in a row not sharing a common superscript differ (P <0.05) using Bonferroni correction procedures for multiple comparisons.
[2]BMI - body mass index

Table 4.3: Identified genetic variants, allele frequency, and evolutionary statistics for alpha-lactalbumin (*LALBA*).

Position [a]		**rsID** [b]	**Change** [c]	**AA** [d]	**MAF** [e]												F_{ST} [f]	**D** [f]	π [f]
					ETR	ETU	GBR	GBU	GN	KE	PE	SP	SW	USC	USW	All			
					30	28	31	33	24	20	39	31	20	5	20	282			
chr12	48569365	.	intergenic		0	0.02	0	0	0	0	0	0	0.06	0	0	0.01	0.02	-0.8	0.01
chr12	48569718	rs113190598	intergenic		0.15	0.13	0.19	0.15	0.23	0.2	0	0	0	0	0.06	0.1	0.05	0.23	0.18
chr12	48570031	rs1261272249	nonsyn	G → A	0	0	0	0.08	0	0	0	0	0.03	0	0	0.01	0.04	-0.73	0.02
chr12	48570375	rs923867024	upstream		0.02	0.02	0	0	0	0	0	0	0	0	0	0	-0.01	-0.83	0.01
chr12	48570451	rs73104702	upstream		0.23	0.18	0.18	0.09	0.1	0.13	0.01	0.11	0.05	0	0.13	0.12	0.02	0.38	0.21

[a]Chromosomal location of variant.

[b]rsID corresponds to entry in NCBI dbSNP database, '.' indicates no entry available.

[c]Change based on snpEff analysis of location function and potential change of amino acid in protein sequence; syn - synonymous mutation; nonsyn - nonsynonymous mutation.

[d]Ammino acid change due to SNP

[e]dMinor Allele Frequency - ETR: Ethiopia, rural; ETU: Ethiopia, urban; GBR: The Gambia, rural; GBU: The Gambia, urban, GN: Ghana; KE: Kenya; PE: Peru; SP: Spain; SW: Sweden; USC: United States, California; USW: United States, Washington/Idaho.

[f]Evolutionary statistics to assess interpopulation differences: F_{ST}: fixation index; D: Tajima's D; π: nucleotide diversity

Table 4.4: Identified genetic variants, allele frequency, and evolutionary statistics for fucosyltransferase 2 *(FUT2)*.

Position [1]		**rsID** [2]	**Change** [3]	**AA** [4]	**MAF** [5]												F_{ST} [6]	**D** [6]	π [6]
					ETR	ETU	GBR	GBU	GN	KE	PE	SP	SW	USC	USW	All			
					30	28	31	33	24	20	39	31	20	5	20	282			
chr19	48694270	rs1214586456	upstream		0.06	0.07	0.05	0.02	0.05	0	0.06	0.02	0.03	0.2	0.05	0.04	0	0.51	0.09
chr19	48694274	rs1486345849	upstream		0.17	0.23	0.22	0.21	0.23	0.13	0.16	0.29	0.19	0.3	0.21	0.21	0	0.51	0.33
chr19	48694277	.	upstream		0.08	0.05	0.07	0.06	0	0.09	0.06	0.21	0.06	0	0.08	0.07	0.02	-0.05	0.14
chr19	48694280	.	upstream		0.27	0.23	0.21	0.18	0.18	0.28	0.16	0.1	0.08	0.3	0.13	0.18	0.01	0.02	0.3
chr19	48694281	.	upstream		0.02	0.02	0.02	0.03	0.05	0.06	0.03	0.04	0.06	0	0.03	0.03	-0.02	0.02	0.06
chr19	48694283	.	upstream		0.02	0.05	0.03	0.05	0.05	0.09	0.03	0.02	0	0.1	0.08	0.04	0	0.02	0.08
chr19	48694285	.	upstream		0.31	0.32	0.34	0.34	0.25	0.28	0.39	0.33	0.28	0.2	0.32	0.32	0	1.76	0.44
chr19	48694299	.	upstream		0.04	0.05	0.07	0.05	0.08	0.09	0.03	0.08	0.03	0	0.11	0.06	-0.01	-0.22	0.11
chr19	48694321	rs1156343359	upstream		0	0	0.05	0.02	0	0.03	0.01	0.02	0.03	0.1	0.08	0.02	0.01	-0.36	0.05
chr19	48694323	.	upstream		0.1	0.09	0.07	0.13	0.05	0.09	0.01	0.1	0.06	0.1	0.16	0.08	0	-0.36	0.15
chr19	48694332	.	upstream		0.06	0.16	0.16	0.06	0.08	0.13	0.1	0.15	0.14	0	0.08	0.11	0	0.27	0.19
chr19	48694335	rs28746179	upstream		0	0.05	0.07	0.02	0.05	0.09	0	0	0	0	0	0.02	0.01	-0.59	0.05
chr19	48694348	.	upstream		0.1	0.05	0.07	0.06	0.13	0.16	0.13	0.04	0.03	0.1	0.11	0.09	0	0.07	0.16
chr19	48694353	.	upstream		0.1	0.05	0.1	0.16	0.15	0.25	0.09	0.1	0.06	0	0.05	0.11	0.01	0.27	0.19
chr19	48694366	.	upstream		0.06	0.18	0.07	0.13	0.18	0.13	0.04	0.15	0.11	0.2	0.11	0.11	0.01	0.33	0.2
chr19	48694394	rs988844014	upstream		0.35	0.36	0.36	0.34	0.32	0.34	0.29	0.23	0.36	0.1	0.29	0.32	0	1.75	0.43
chr19	48694398	rs1021282467	upstream		0.29	0.32	0.29	0.24	0.38	0.44	0.37	0.4	0.44	0.4	0.32	0.34	0.01	1.84	0.45
chr19	48694406	rs557631142	upstream		0.15	0.07	0.16	0.08	0.15	0.25	0.17	0.15	0.14	0.1	0.18	0.14	0	0.62	0.25
chr19	48694414	rs576217481	upstream		0.19	0.16	0.18	0.18	0.18	0.19	0.16	0.13	0.14	0.4	0.21	0.17	-0.01	0.87	0.29
chr19	48694415	.	upstream		0.04	0.05	0.11	0.13	0.13	0.13	0.06	0.08	0.03	0	0.05	0.08	0.01	0.87	0.18
chr19	48694424	rs2548456	upstream		0	0	0.18	0.13	0.13	0.06	0	0.02	0	0	0	0.05	0.08	-0.26	0.1
chr19	48694428	rs1253339733	upstream		0.17	0.23	0.11	0.13	0.16	0.16	0.19	0.23	0.28	0	0.16	0.17	0	0.86	0.29
chr19	48694430	rs1396330130	upstream		0.06	0.07	0.1	0.11	0.1	0.06	0.07	0.08	0.06	0.1	0.08	0.08	-0.02	0.69	0.15
chr19	48694432	.	upstream		0.12	0.14	0.17	0.18	0.15	0.19	0.24	0.23	0.22	0.3	0.24	0.19	0	0.69	0.31

chr19	48694438	.	upstream		0.21	0.09	0.12	0.15	0.13	0.22	0.16	0.25	0.22	0.2	0.18	0.17	0	0.81	0.28
chr19	48694439	.	upstream		0.12	0.11	0.16	0.12	0.1	0.13	0.07	0.13	0.08	0.2	0.13	0.11	-0.01	0.81	0.2
chr19	48694443	.	upstream		0.08	0	0.02	0.05	0.03	0.03	0.03	0.02	0.06	0.1	0.03	0.03	-0.01	-0.47	0.07
chr19	48694447	rs28362826	upstream		0.04	0.11	0	0.06	0.03	0	0	0	0	0	0	0.02	0.04	-0.59	0.05
chr19	48694455	.	upstream		0	0.02	0.07	0.08	0.03	0.06	0.07	0.08	0.03	0.1	0.08	0.06	0	-0.24	0.1
chr19	48694463	.	upstream		0.04	0.02	0.07	0.03	0	0.03	0.01	0.02	0	0	0.05	0.03	-0.01	-0.54	0.06
chr19	48694846	.	upstream		0	0	0	0.07	0	0	0	0	0	0	0	0.01	0.05	-0.77	0.02
chr19	48695079	rs2548457	upstream		0.19	0.25	0.61	0.53	0.68	0.5	0	0.1	0.16	0.1	0.21	0.31	0.24	1.72	0.43
chr19	48695088	rs150156163	upstream		0	0	0.09	0.02	0	0	0	0	0	0	0	0.01	0.04	-0.73	0.02
chr19	48695142	rs1237144276	upstream		0.1	0.06	0.11	0.07	0.03	0.09	0.03	0.02	0.13	0.3	0.07	0.07	0.02	-0.07	0.13
chr19	48695151	.	upstream		0.26	0.42	0.34	0.28	0.21	0.18	0.22	0.23	0.28	0.4	0.29	0.27	0.01	2.20	0.4
chr19	48695152	.	upstream		0.33	0.4	0.36	0.33	0.24	0.21	0.26	0.29	0.28	0.4	0.32	0.31	0	2.20	0.43
chr19	48695167	rs528122216	upstream		0.28	0.44	0.3	0.31	0.34	0.21	0.28	0.23	0.38	0.5	0.32	0.31	0.01	1.72	0.43
chr19	48695199	rs1423727045	upstream		0.29	0.44	0.3	0.24	0.2	0.21	0.21	0.15	0.31	0.5	0.29	0.27	0.03	1.51	0.39
chr19	48695201	.	upstream		0.21	0.38	0.29	0.16	0.1	0.18	0.1	0.06	0.25	0.5	0.18	0.2	0.06	1.04	0.32
chr19	48695255	.	upstream		0.02	0	0.02	0	0.03	0	0	0	0	0	0	0.01	-0.01	-0.8	0.01
chr19	48695320	rs545073059	upstream		0	0	0	0	0	0	0	0.06	0	0	0	0.01	0.03	-1.12	0.01
chr19	48695323	rs765420172	upstream		0	0	0	0	0	0	0	0.02	0	0	0	0	0	-1.12	0
chr19	48702888	rs516316	downstream		0.63	0.41	0.57	0.47	0.43	0.64	0.09	0.62	0.35	0.1	0.47	0.44	0.12	2.11	0.49
chr19	48702915	rs516246	downstream		0.63	0.41	0.57	0.47	0.43	0.64	0.09	0.62	0.35	0.1	0.47	0.44	0.12	2.11	0.49
chr19	48703029	rs1800021	nonsyn	I → V	0	0.04	0.11	0.1	0.11	0.04	0	0	0	0	0	0.04	0.03	-0.39	0.08
chr19	48703036	rs1185332526	nonsyn	H → R	0	0	0	0.02	0	0	0	0	0	0.1	0	0	0.04	-0.83	0.01
chr19	48703160	rs492602	syn	A → A	0.65	0.45	0.57	0.47	0.46	0.63	0.08	0.65	0.42	0.2	0.5	0.47	0.12	2.13	0.5
chr19	48703167	rs575344194	nonsyn	R → C	0	0	0	0	0	0	0.01	0	0	0.1	0	0	0.04	-0.83	0.01
chr19	48703205	rs681343	syn	Y → Y	0.65	0.45	0.57	0.47	0.46	0.63	0.07	0.65	0.42	0.2	0.5	0.46	0.13	2.13	0.5
chr19	48703304	rs28362836	syn	S → S	0	0	0.02	0.02	0.04	0.07	0	0	0	0	0	0.01	0.01	-0.74	0.02
chr19	48703346	rs281377	syn	N → N	0.35	0.5	0.14	0.34	0.2	0.17	0.57	0.27	0.53	0.7	0.5	0.37	0.09	1.94	0.47
chr19	48703384	.	nonsyn	E → G	0.1	0.04	0	0	0	0	0	0	0.06	0	0.05	0.02	0.03	-0.63	0.04

chr19	48703417	rs601338	Stop gained	W → *	0.59	0.46	0.5	0.5	0.44	0.43	0.1	0.5	0.41	0.3	0.4	0.41	0.09	0.72	0.53
chr19	48703418	rs1800459	nonsyn	W → C	0.23	0.25	0.26	0.28	0.48	0.45	0.44	0.48	0.08	0.3	0.4	0.35	0.06	0.72	0.45
chr19	48703419	.	nonsyn	T → P	0	0	0	0	0.02	0	0.01	0.02	0	0	0	0.01	-0.01	0.72	0.01
chr19	48703469	rs1800027	syn	H → H	0.06	0.09	0.09	0.05	0.02	0.03	0.01	0.03	0.03	0	0.05	0.04	0	-0.36	0.08
chr19	48703470	rs1800025	nonsyn	D → N	0.06	0.02	0.17	0.2	0.21	0.13	0	0.02	0	0	0	0.08	0.09	-0.01	0.14
chr19	48703477	rs755843863	nonsyn	V → G	0.1	0.07	0.02	0.03	0.04	0	0.1	0.08	0.11	0	0.08	0.06	0	-0.16	0.12
chr19	48703728	rs602662	nonsyn	G → S	0.58	0.46	0.59	0.38	0.4	0.56	0.12	0.65	0.44	0.4	0.53	0.45	0.09	2.12	0.5
chr19	48703830	rs548111243	nonsyn	G → R	0	0	0	0	0	0	0.07	0	0	0	0	0.01	0.03	-0.75	0.02
chr19	48703912	.	nonsyn	I → T	0	0	0.02	0	0	0	0	0	0	0	0	0	-0.01	-0.85	0
chr19	48703949	rs485186	syn	T → T	0.62	0.48	0.6	0.43	0.4	0.61	0.12	0.62	0.44	0.4	0.53	0.47	0.09	2.14	0.5
chr19	48703974	.	syn	L → L	0	0	0	0.02	0	0	0	0	0	0	0	0	-0.01	-0.85	0
chr19	48703998	rs485073	3'UTR		0.62	0.48	0.6	0.47	0.4	0.56	0.12	0.65	0.47	0.4	0.53	0.47	0.09	2.14	0.5
chr19	48704000	rs603985	3'UTR		0.6	0.48	0.6	0.43	0.4	0.56	0.12	0.65	0.47	0.4	0.53	0.46	0.09	2.13	0.5
chr19	48704175	rs28362840	3'UTR		0.02	0	0.07	0.05	0.02	0.08	0	0	0	0	0.03	0.02	0.01	-0.59	0.05
chr19	48704297	rs571689	3'UTR		0.69	0.54	0.62	0.59	0.66	0.75	0.25	0.6	0.48	0.3	0.6	0.56	0.07	2.11	0.49
chr19	48704303	.	3'UTR		0.17	0.11	0.27	0.21	0.15	0.18	0	0.02	0	0	0	0.11	0.09	0.33	0.2
chr19	48704353	rs374477120	3'UTR		0	0	0	0	0	0.05	0	0	0	0	0	0	0.03	-0.83	0.01
chr19	48704378	.	3'UTR		0.18	0.11	0.07	0.09	0.1	0.05	0.03	0.11	0.08	0	0.05	0.09	0.01	0.09	0.16
chr19	48704394	rs570794	3'UTR		0.72	0.57	0.63	0.62	0.65	0.75	0.29	0.63	0.48	0.4	0.63	0.58	0.07	2.07	0.49
chr19	48704400	.	3'UTR		0.02	0.02	0	0	0	0	0	0.03	0	0.1	0	0.01	0.02	-0.04	0.02
chr19	48704403	rs1203875153	3'UTR		0.28	0.2	0.18	0.09	0.1	0.08	0.01	0.29	0.23	0.2	0.13	0.16	0.06	-0.04	0.26
chr19	48704407	.	3'UTR		0.04	0.02	0.05	0.05	0.02	0	0.24	0.05	0.08	0	0.05	0.06	0.07	0.87	0.11
chr19	48704408	rs1175534345	3'UTR		0.2	0.3	0.18	0.21	0.2	0.15	0.41	0.26	0.35	0.4	0.32	0.26	0.02	0.87	0.39
chr19	48704411	.	3'UTR		0.04	0.09	0.07	0.06	0.05	0	0.36	0.06	0.13	0.1	0.13	0.1	0.1	0.08	0.18
chr19	48704413	.	3'UTR		0.04	0.06	0.05	0.06	0.02	0	0.26	0.05	0.1	0.1	0.08	0.07	0.06	0.08	0.14
chr19	48704414	.	3'UTR		0.05	0.04	0.03	0.05	0.02	0	0.33	0.03	0.1	0.1	0.08	0.08	0.12	0.08	0.14
chr19	48704416	rs1449519289	3'UTR		0.02	0.06	0.03	0.08	0.05	0.03	0.26	0.06	0.1	0.1	0.11	0.08	0.05	0.04	0.15
chr19	48704418	rs1336286761	3'UTR		0.05	0.11	0.08	0.1	0.05	0.08	0.36	0.06	0.13	0.3	0.16	0.12	0.08	0.04	0.21

chr19	48704419	.	3'UTR	0	0.07	0.02	0.02	0.02	0	0.19	0	0.03	0	0.11	0.04	0.08	0.04	0.08
chr19	48704420	rs1308982795	3'UTR	0.04	0.09	0.07	0.05	0.05	0.03	0.26	0.05	0.05	0.1	0.16	0.08	0.05	0.05	0.15
chr19	48704461	rs569970	3'UTR	0.69	0.59	0.62	0.59	0.66	0.75	0.27	0.6	0.48	0.3	0.6	0.57	0.06	2.09	0.49
chr19	48704535	rs2251034	3'UTR	0.69	0.52	0.53	0.52	0.55	0.63	0.22	0.61	0.48	0.3	0.6	0.52	0.06	2.14	0.5
chr19	48704728	.	3'UTR	0	0	0	0	0.03	0	0	0.02	0	0.17	0	0.01	0.05	-0.8	0.01
chr19	48704791	rs28362842	3'UTR	0	0.02	0.11	0.08	0.03	0.06	0.08	0.08	0.05	0	0.06	0.06	-0.01	-0.18	0.11
chr19	48704795	rs567037	3'UTR	0	0	0.29	0.27	0.37	0.34	0	0.02	0	0	0	0.13	0.2	0.45	0.22
chr19	48705048	.	3'UTR	0	0	0	0.11	0	0.03	0.03	0.02	0.03	0	0	0.02	0.04	-0.62	0.04
chr19	48705106	rs146519458	3'UTR	0	0	0.03	0.02	0.06	0.03	0	0	0	0	0	0.01	0.01	-0.7	0.03
chr19	48705111	rs2432132	3'UTR	0.62	0.43	0.52	0.48	0.48	0.5	0.07	0.58	0.48	0.3	0.5	0.45	0.1	2.12	0.5
chr19	48705116	rs868855781	3'UTR	0.1	0.14	0	0	0	0.03	0	0	0	0	0	0.03	0.09	-0.56	0.05
chr19	48705123	rs112820994	3'UTR	0.02	0.02	0.12	0.09	0.21	0.1	0	0	0	0	0	0.05	0.08	-0.28	0.1
chr19	48705129	rs1031492296	3'UTR	0.2	0.27	0.2	0.11	0.06	0.18	0.03	0.03	0.15	0.1	0.1	0.13	0.05	0.46	0.22
chr19	48705130	rs1184990920	3'UTR	0.05	0.05	0.02	0.03	0.04	0.03	0	0	0.03	0	0.03	0.02	0	-0.61	0.05
chr19	48705132	rs1246979914	3'UTR	0.1	0.11	0.05	0.02	0.06	0.08	0.01	0	0.05	0	0.05	0.05	0.01	-0.61	0.09
chr19	48705136	rs955425753	3'UTR	0.08	0.05	0.07	0.03	0.06	0.05	0	0	0.03	0	0.08	0.04	0.01	-0.3	0.08
chr19	48705139	.	3'UTR	0.15	0.13	0.11	0.06	0.08	0.13	0	0.02	0.05	0	0.08	0.07	0.02	-0.3	0.14
chr19	48705142	.	3'UTR	0.22	0.27	0.13	0.15	0.06	0.18	0.06	0	0.13	0	0.15	0.13	0.04	0.17	0.22
chr19	48705143	.	3'UTR	0.1	0.11	0.11	0.05	0.04	0.1	0	0	0.05	0	0.05	0.06	0.02	0.17	0.11
chr19	48705161	rs1467636699	3'UTR	0.03	0.04	0.02	0	0.02	0.03	0	0	0.03	0	0.03	0.02	-0.01	-0.68	0.03
chr19	48705204	.	3'UTR	0.08	0.11	0.07	0.05	0.04	0.03	0.01	0	0.03	0	0.08	0.05	0.01	-0.34	0.09
chr19	48705214	rs28362845	3'UTR	0	0	0.02	0.06	0.06	0.08	0	0	0	0	0	0.02	0.02	-0.64	0.04
chr19	48705228	.	3'UTR	0.03	0.05	0.02	0.05	0.02	0	0.01	0	0.08	0	0.05	0.03	0	-0.54	0.06
chr19	48705237	.	3'UTR	0.02	0.04	0.05	0.03	0	0.05	0	0	0	0	0.03	0.02	0	-0.64	0.04
chr19	48705244	rs507855	3'UTR	0.6	0.39	0.47	0.44	0.38	0.48	0.1	0.55	0.43	0.3	0.5	0.42	0.08	2.07	0.49
chr19	48705247	.	3'UTR	0.02	0.04	0	0	0	0	0	0	0	0	0	0.01	0.01	-0.81	0.01
chr19	48705254	.	3'UTR	0.1	0.07	0.08	0.08	0.04	0.03	0.01	0	0.03	0.1	0.03	0.05	0.01	-0.32	0.09
chr19	48705263	.	3'UTR	0.02	0.02	0.02	0.05	0	0.08	0.04	0	0	0	0.05	0.02	0	-0.58	0.05

chr19	48705266	.	3'UTR	0.02	0.02	0.03	0.08	0	0.1	0.04	0	0	0	0.08	0.03	0.02	-0.48	0.07
chr19	48705272	rs1376457670	3'UTR	0.02	0.04	0.03	0.03	0	0.1	0.04	0	0	0	0.08	0.03	0.01	-0.52	0.06
chr19	48705286	rs507766	3'UTR	0.67	0.48	0.6	0.52	0.42	0.55	0.08	0.58	0.53	0.3	0.55	0.48	0.12	2.14	0.5
chr19	48705307	rs507711	3'UTR	0.67	0.52	0.59	0.52	0.44	0.55	0.09	0.6	0.5	0.3	0.53	0.48	0.11	2.15	0.5
chr19	48705372	rs506897	3'UTR	0.7	0.48	0.57	0.5	0.48	0.55	0.07	0.63	0.48	0.3	0.5	0.48	0.13	2.14	0.5
chr19	48705608	rs504963	3'UTR	0.65	0.52	0.52	0.5	0.48	0.68	0.09	0.58	0.43	0.4	0.48	0.47	0.1	2.14	0.5
chr19	48705716	.	3'UTR	0	0.09	0.05	0.05	0.02	0	0.06	0.03	0.03	0	0.03	0.04	0	-0.42	0.08
chr19	48705721	rs632111	3'UTR	0.48	0.3	0.53	0.45	0.48	0.55	0.08	0.6	0.4	0.3	0.48	0.42	0.09	2.07	0.49
chr19	48705753	rs503279	3'UTR	0.72	0.48	0.58	0.55	0.52	0.68	0.08	0.6	0.43	0.3	0.48	0.49	0.14	2.15	0.5
chr19	48705954	.	3'UTR	0.08	0	0.03	0	0	0.08	0.03	0.08	0.03	0	0.05	0.04	0.02	-0.46	0.07
chr19	48705969	rs633372	3'UTR	0.6	0.54	0.53	0.47	0.44	0.6	0.12	0.65	0.34	0.4	0.47	0.46	0.08	2.13	0.5

[1]Chromosomal location of variant.

[2]rsID corresponds to entry in NCBI dbSNP database, '.' indicates no entry available.

[3]Change based on snpEff analysis of location function and potential change of amino acid in protein sequence; syn - synonymous mutation; nonsyn - nonsynonymous mutation.

[4]Ammino acid change due to SNP

[5]dMinor Allele Frequency - ETR: Ethiopia, rural; ETU: Ethiopia, urban; GBR: The Gambia, rural; GBU: The Gambia, urban, GN: Ghana; KE: Kenya; PE: Peru; SP: Spain; SW: Sweden; USC: United States, California; USW: United States, Washington/Idaho.

[6]Evolutionary statistics to assess interpopulation differences: F_{ST}: fixation index; D: Tajima's D; π: nucleotide diversity

Table 4.5: Identified genetic variants, allele frequency, and evolutionary statistics fucosyltransferase 3 *(FUT3)*.

Position [1]		**rsID** [2]	**Change** [3]	**AA** [4]	**MAF** [5]											F_{ST} [6]	**D** [6]	π [6]
					ETR	ETU	GBR	GBU	GN	KE	PE	SP	SW	USC	USW	All			
					30	28	31	33	24	20	39	31	20	5	20	282			
chr19	5843565	rs34304528	3'UTR		0.11	0	0	0	0	0	0.03	0.05	0.14	0.13	0.03	0.03	0.04	-0.49	0.06
chr19	5843598	rs874232	3'UTR		0.21	0.34	0.24	0.32	0.25	0.29	0.45	0.28	0.26	0.3	0.13	0.29	0.02	1.62	0.41
chr19	5843757	.	syn	T → T	0	0	0	0	0	0	0	0	0	0	0.06	0	0.01	-0.82	0.01
chr19	5843782	.	nonsyn	V → G	0.02	0.05	0.03	0.08	0	0.03	0.01	0	0.03	0	0	0.03	0	-0.51	0.06
chr19	5843913	rs1331233156	syn	L → L	0	0	0	0.02	0	0	0	0	0	0.1	0	0	0.04	-0.82	0.01
chr19	5844032	rs137909900	nonsyn	V → M	0.06	0.04	0.13	0.03	0.05	0.1	0	0	0	0	0	0.04	0.02	-0.4	0.08
chr19	5844332	rs3745635	nonsyn	G → S	0.09	0.14	0.12	0.38	0.32	0.18	0.45	0.02	0.03	0.25	0	0.19	0.16	1	0.31
chr19	5844356	rs28362463	nonsyn	D → N	0.07	0.06	0.2	0.1	0.1	0.17	0	0	0	0	0	0.06	0.06	-0.16	0.12
chr19	5844526	rs778986	nonsyn	M → T	0.65	0.7	0.81	0.88	0.87	0.98	0.95	0.84	0.76	1	0.85	0.84	0.05	0.78	0.27
chr19	5844615	rs377716323	syn	P → P	0.13	0	0.03	0.05	0	0	0.03	0	0.03	0	0.03	0.03	0.04	-0.53	0.06
chr19	5844638	rs812936	nonsyn	R → W	0.58	0.46	0.63	0.74	0.73	0.88	0.78	0.81	0.68	1	0.7	0.71	0.06	1.63	0.42
chr19	5844649	rs376471696	nonsyn	L → P	0	0	0	0	0.02	0	0.01	0	0.08	0	0	0.01	0.03	-0.76	0.02
chr19	5844652	.	nonsyn	L → P	0.38	0.38	0.29	0.3	0.37	0.25	0.4	0.29	0.25	0.4	0.38	0.33	0	1.81	0.44
chr19	5844656	rs1259810674	nonsyn	T → P	0.46	0.46	0.45	0.39	0.37	0.38	0.41	0.31	0.42	0.2	0.4	0.4	0.01	2.03	0.48
chr19	5844661	rs148881389	nonsyn	R → P	0.44	0.42	0.5	0.41	0.39	0.35	0.45	0.37	0.39	0.3	0.38	0.41	0	2.06	0.49
chr19	5844665	rs756103678	nonsyn	T → P	0.38	0.44	0.44	0.3	0.3	0.25	0.31	0.19	0.28	0.1	0.25	0.31	0.03	1.73	0.43
chr19	5844779	rs28362460	syn	L → L	0.02	0.02	0.02	0.02	0.07	0.03	0.01	0	0	0	0	0.02	0	-0.68	0.03
chr19	5844781	rs28362459	nonsyn	L → R	0.21	0.25	0.26	0.44	0.45	0.25	0.49	0.11	0.06	0.5	0.05	0.28	0.11	1.57	0.4

chr19	5844827	rs28362458	nonsyn	G → S	0.1	0.06	0.18	0.11	0.16	0.23	0	0	0	0	0	0.08	0.07	-0.01	0.14

[1]Chromosomal location of variant.

[2]rsID corresponds to entry in NCBI dbSNP database, '.' indicates no entry available.

[3]Change based on snpEff analysis of location function and potential change of amino acid in protein sequence; syn - synonymous mutation; nonsyn - nonsynonymous mutation.

[4]Ammino acid change due to SNP

[5]dMinor Allele Frequency - ETR: Ethiopia, rural; ETU: Ethiopia, urban; GBR: The Gambia, rural; GBU: The Gambia, urban, GN: Ghana; KE: Kenya; PE: Peru; SP: Spain; SW: Sweden; USC: United States, California; USW: United States, Washington/Idaho.

[6]Evolutionary statistics to assess interpopulation differences: F_{ST}: fixation index; D: Tajima's D; π: nucleotide diversity

Table 4.6: Identified genetic variants, allele frequency, and evolutionary statistics for ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 5 *(ST6GalNAc5)*.

Position [1]		**rsID** [2]	**Change** [3]	**AA** [4]	**MAF** [5]												F_{ST} [6]	**D** [6]	π [6]
					ETR	ETU	GBR	GBU	GN	KE	PE	SP	SW	USC	USW	All			
					30	28	31	33	24	20	39	31	20	5	20	282			
chr1	77007735	rs6663942	intergenic		0.29	0.22	0.25	0.3	0.26	0.46	0.04	0.02	0.21	0	0.11	0.2	0.07	1.05	0.32
chr1	77007808	rs74090578	intergenic		0	0.1	0.11	0.12	0.07	0.19	0	0	0.06	0	0.03	0.06	0.03	-0.2	0.11
chr1	77008186	.	intergenic		0.1	0	0.05	0.03	0.04	0.08	0.03	0	0.03	0	0	0.03	0.01	-0.47	0.07
chr1	77044197	rs200277344	intergenic		0	0	0	0	0	0	0	0	0.06	0	0	0	0.01	-0.82	0.01
chr1	77044448	.	nonsyn	I $\rightarrow$ S	0.02	0.02	0.02	0.11	0.06	0.06	0.04	0	0.12	0	0	0.04	0.02	-0.36	0.08
chr1	77044472	.	nonsyn	M $\rightarrow$ R	0	0.02	0.02	0	0.03	0.03	0.03	0	0	0	0.04	0.01	-0.01	-0.7	0.03
chr1	77050454	rs1003060586	intergenic		0.26	0.32	0.3	0.37	0.23	0.14	0.42	0.3	0.36	0	0.32	0.31	0.02	1.72	0.43
chr1	77050655	rs77058959	intergenic		0.09	0	0.02	0.03	0.08	0.04	0	0	0	0	0	0.02	0.01	-0.61	0.04
chr1	77062741	rs199722	downstream		0.6	0.52	0.66	0.45	0.52	0.73	0.55	0.33	0.39	0.33	0.18	0.5	0.06	2.15	0.5
chr1	77062831	.	downstream		0.27	0.33	0.45	0.45	0.32	0.37	0.43	0.28	0.42	0	0.35	0.37	0.02	1.94	0.47
chr1	77063069	.	nonsyn	G $\rightarrow$ C	0	0.02	0	0	0	0	0	0	0	0.1	0	0	0.04	-0.83	0.01
chr1	77063299	.	3'UTR		0	0	0	0	0	0.03	0	0	0	0	0	0	0.01	-0.85	0
chr1	77063362	rs993365662	3'UTR		0	0	0.02	0	0	0.03	0	0	0	0	0	0	0	-0.83	0.01
chr1	77063368	.	3'UTR		0.02	0	0	0	0	0	0	0	0	0	0	0	0	-0.85	0
chr1	77063686	.	3'UTR		0.02	0.04	0.03	0.03	0.02	0.03	0.03	0.02	0.03	0	0	0.03	-0.02	-0.58	0.05
chr1	77063880	rs144737930	3'UTR		0	0	0.03	0	0	0	0	0	0	0	0	0	0.01	-0.83	0.01
chr1	77063921	.	3'UTR		0	0	0	0	0.03	0	0	0	0	0.13	0	0	0.06	-0.83	0.01
chr1	77063950	.	3'UTR		0	0	0	0	0	0	0	0.02	0	0.13	0	0	0.06	-0.83	0.01
chr1	77064364	rs537623812	3'UTR		0	0	0	0	0	0	0	0	0	0.1	0.03	0	0.05	-0.83	0.01
chr1	77064710	rs115329926	3'UTR		0.02	0.04	0.02	0.03	0.07	0.03	0	0	0	0	0	0.02	0	-0.65	0.04
chr1	77064778	rs41292250	3'UTR		0.07	0.09	0	0	0	0	0	0.1	0.2	0.17	0.05	0.05	0.06	-0.32	0.09
chr1	77064824	.	3'UTR		0.17	0.38	0.3	0.4	0.27	0.42	0.44	0.24	0.33	0	0.26	0.32	0.03	1.77	0.44
chr1	77064830	.	3'UTR		0.13	0	0.1	0.02	0.02	0	0.03	0.1	0.03	0	0.05	0.05	0.03	-0.32	0.09

chr1	77065124	.	3'UTR	0.36	0.26	0.25	0.18	0.17	0.31	0.31	0.17	0.13	0	0.24	0.24	0.02	1.33	0.36
chr1	77065281	rs1059038	3'UTR	0.03	0.04	0	0	0	0.03	0	0.08	0	0.1	0	0.02	0.02	-0.64	0.04
chr1	77065303	rs151127371	3'UTR	0.02	0.09	0	0	0	0.06	0	0	0	0	0	0.02	0.04	-0.69	0.03
chr1	77065538	rs11162262	3'UTR	0	0.04	0.15	0.06	0.08	0.1	0.6	0.08	0.18	0.3	0.15	0.16	0.26	0.76	0.27
chr1	77065664	.	3'UTR	0.1	0.02	0.03	0.02	0	0	0	0.02	0	0	0	0.02	0.03	-0.64	0.04
chr1	77065667	.	3'UTR	0	0	0	0	0	0	0	0	0.05	0	0	0	0.03	-0.83	0.01
chr1	77065765	.	3'UTR	0	0	0	0	0	0	0	0.02	0	0	0	0	0	-0.85	0
chr1	77065838	.	3'UTR	0.08	0	0.1	0.02	0.02	0.03	0.04	0.1	0.08	0	0.03	0.05	0.01	-0.31	0.09
chr1	77066028	rs79261221	3'UTR	0.15	0	0	0	0	0	0.1	0.04	0.03	0.17	0.03	0.04	0.05	1.18	0.08
chr1	77066029	rs199723	3'UTR	0.48	0.5	0.65	0.45	0.55	0.76	0.6	0.35	0.47	0.5	0.25	0.51	0.04	1.18	0.5
chr1	77066077	.	3'UTR	0	0	0	0	0	0	0.01	0	0	0	0	0	-0.01	-0.85	0
chr1	77066095	rs199724	3'UTR	0.48	0.5	0.65	0.45	0.55	0.76	0.6	0.35	0.47	0.5	0.25	0.51	0.04	2.15	0.5
chr1	77066152	rs146849779	3'UTR	0.02	0.07	0.02	0	0	0.05	0	0	0	0	0	0.01	0.02	-0.7	0.03
chr1	77066309	.	3'UTR	0	0	0	0	0	0	0	0	0	0	0.03	0	0.01	-0.85	0
chr1	77066733	rs144704498	3'UTR	0	0	0.06	0.09	0.04	0	0	0	0	0	0	0.02	0.03	-0.62	0.04
chr1	77066868	rs199662	3'UTR	0.35	0.38	0.45	0.32	0.4	0.5	0.38	0.26	0.25	0.3	0.23	0.35	0.02	1.88	0.46
chr1	77066895	rs58416742	3'UTR	0.1	0.11	0.03	0.08	0.1	0.11	0	0.05	0.03	0.1	0.08	0.06	0	-0.14	0.12
chr1	77066910	rs76211710	3'UTR	0.03	0.02	0.06	0.05	0.02	0.06	0	0.03	0.03	0	0	0.03	-0.01	-0.53	0.06
chr1	77066917	rs199663	3'UTR	0.57	0.52	0.66	0.5	0.56	0.78	0.58	0.37	0.4	0.5	0.33	0.53	0.04	2.14	0.5
chr1	77067143	.	3'UTR	0.07	0.02	0	0.02	0	0	0	0.06	0	0	0	0.02	0.02	-0.66	0.04
chr1	77067212	rs199664	3'UTR	0	0	0	0	0.02	0.08	0	0	0	0	0	0.01	0.02	-0.8	0.01
chr1	77067670	rs147210202	downstream	0.02	0.09	0	0	0	0.04	0	0	0	0	0	0.01	0.05	-0.7	0.03
chr1	77067799	rs115295855	downstream	0	0	0.07	0.03	0.05	0	0	0	0	0	0	0.02	0	-0.67	0.03
chr1	77067800	rs72681867	downstream	0	0	0	0	0	0	0	0.05	0.16	0	0.03	0.02	0.07	-0.67	0.03
chr1	77067913	.	downstream	0	0	0	0	0	0	0	0.02	0	0.1	0	0	0.04	-1.13	0.01
chr1	77067914	.	downstream	0	0	0	0	0	0	0	0	0.06	0	0	0	0.04	-1.13	0.01
chr1	77068042	.	downstream	0	0	0.02	0	0	0	0	0	0	0	0	0	-0.01	-0.85	0
chr1	77068197	.	downstream	0	0	0	0	0	0	0	0	0.03	0	0	0	0.01	-0.85	0

chr1	77068224	rs61782871	downstream	0.02	0.09	0	0	0	0	0	0.17	0.15	0.2	0.06	0.05	0.08	-0.34	0.09
chr1	77068303	.	downstream	0	0	0	0.08	0	0	0	0	0	0	0	0.01	0.06	-0.76	0.02
chr1	77068366	rs61782872	downstream	0.08	0.08	0.02	0.03	0	0.09	0	0.17	0.18	0.2	0.09	0.07	0.04	-0.07	0.13
chr1	77068620	rs74963941	downstream	0	0.02	0	0.12	0.13	0	0	0	0	0	0	0.03	0.08	-0.55	0.05
chr1	77069007	rs199665	downstream	0.14	0.25	0.34	0.25	0.27	0.25	0.41	0.05	0.15	0.3	0.16	0.24	0.05	1.32	0.36
chr1	77069015	.	downstream	0.1	0	0	0	0	0.06	0	0.15	0	0	0	0.03	0.09	-0.51	0.06
chr1	77069103	rs11162264	downstream	0.23	0.05	0.19	0.1	0.07	0.17	0	0.07	0	0	0.03	0.09	0.06	0.09	0.16

[1]Chromosomal location of variant.

[2]rsID corresponds to entry in NCBI dbSNP database, '.' indicates no entry available.

[3]Change based on snpEff analysis of location function and potential change of amino acid in protein sequence; syn - synonymous mutation; nonsyn - nonsynonymous mutation.

[4]Ammino acid change due to SNP

[5]dMinor Allele Frequency - ETR: Ethiopia, rural; ETU: Ethiopia, urban; GBR: The Gambia, rural; GBU: The Gambia, urban, GN: Ghana; KE: Kenya; PE: Peru; SP: Spain; SW: Sweden; USC: United States, California; USW: United States, Washington/Idaho.

[6]Evolutionary statistics to assess interpopulation differences: F_{ST}: fixation index; D: Tajima's D; π: nucleotide diversity

Table 4.7: Identified genetic variants, allele frequency, and evolutionary statistics for Gulonolactone (L-) oxidase, pseudogene *(GULOP)*.

Position [1]		**rsID** [2]	**Change** [3]	**AA** [4]	**MAF** [5]												F_{ST} [6]	**D** [6]	π [6]
					ETR	ETU	GBR	GBU	GN	KE	PE	SP	SW	USC	USW	All			
					30	28	31	33	24	20	39	31	20	5	20	282			
chr8	27579326	rs17057417	intergenic		0.1	0.06	0.14	0.14	0	0.14	0	0	0	0	0.03	0.06	0.05	-0.17	0.12
chr8	27579750	rs73558181	intergenic		0.08	0.1	0.19	0.13	0.03	0.13	0	0	0	0	0.13	0.08	0.04	0	0.14
chr8	27579771	rs78422197	intergenic		0	0	0	0	0.06	0	0	0	0	0	0	0	0.02	-0.82	0.01
chr8	27579865	rs146127661	intergenic		0	0	0	0	0.07	0	0	0	0	0	0	0	0.02	-0.82	0.01
chr8	27579871	rs17057419	intergenic		0.32	0.28	0.37	0.58	0.37	0.63	0.17	0.23	0.06	0	0.19	0.32	0.11	1.75	0.44
chr8	27579906	rs114652835	intergenic		0	0.02	0.11	0	0.19	0	0	0	0	0	0	0.03	0.1	-0.52	0.06
chr8	27580056	rs111659883	intergenic		0.24	0.28	0.46	0.23	0.44	0.25	0.06	0.12	0.22	0	0.03	0.22	0.09	1.23	0.35
chr8	27580101	.	intergenic		0.13	0.02	0	0	0	0.09	0.04	0	0	0	0	0.03	0.05	-0.56	0.05
chr8	27583045	.	intergenic		0.02	0.04	0.04	0.03	0	0	0.03	0.05	0.03	0	0	0.03	-0.01	-0.58	0.05
chr8	27583197	rs10216551	intergenic		0.27	0.04	0.2	0.12	0.1	0.16	0.15	0.16	0.06	0.13	0.11	0.14	0	0.54	0.23
chr8	27583362	rs10216623	intergenic		0.27	0.04	0.2	0.12	0.1	0.16	0.15	0.16	0.06	0.13	0.11	0.14	0	0.42	0.23
chr8	27583364	rs77271157	intergenic		0	0	0	0	0	0	0.35	0.08	0.18	0.13	0.14	0.09	0.2	0.42	0.16
chr8	27583400	rs61006720	intergenic		0.05	0.07	0.14	0.26	0.02	0.03	0.05	0.03	0	0.13	0	0.08	0.07	0.01	0.15
chr8	27583457	rs75929510	intergenic		0	0.09	0.12	0.05	0.24	0.13	0	0	0.09	0	0.06	0.07	0.05	-0.1	0.13
chr8	27583493	rs7828131	intergenic		0.55	0.54	0.82	0.82	0.79	0.72	0.26	0.32	0.32	0.38	0.36	0.55	0.19	2.12	0.5
chr8	27583506	rs10216661	intergenic		0	0	0.08	0.02	0.02	0	0	0	0	0	0	0.01	0.02	-0.72	0.02
chr8	27583590	.	intergenic		0.02	0	0	0	0.04	0.05	0.03	0.02	0.05	0	0.03	0.02	0	-0.63	0.04
chr8	27583676	rs973743582	intergenic		0	0.02	0	0	0.02	0	0	0	0	0.13	0	0.01	0.03	-0.8	0.01
chr8	27583810	rs116425632	intergenic		0	0	0.07	0.03	0.02	0	0	0	0	0	0	0.01	0.01	-0.7	0.03
chr8	27583821	.	intergenic		0	0	0	0.02	0	0	0	0	0	0.1	0	0	0.04	-0.82	0.01
chr8	27583843	.	intergenic		0	0.02	0	0.02	0.02	0	0.01	0.07	0.03	0	0.03	0.02	0	-0.63	0.04
chr8	27583862	.	intergenic		0	0.06	0.02	0.05	0.05	0.03	0	0	0	0	0	0.02	0.01	-0.89	0.04
chr8	27583864	.	intergenic		0	0	0	0	0	0.06	0.03	0.04	0.06	0	0	0.02	0.01	N/A	0.03

chr8	27583870	rs1410553994	intergenic	0	0	0	0.02	0	0	0	0	0	0.13	0	0	0.05	-0.82	0.01
chr8	27583888	.	intergenic	0	0	0.02	0	0.05	0	0	0.04	0.03	0	0	0.01	0	-0.73	0.02
chr8	27584004	rs4732728	intergenic	0.7	0.65	0.93	0.81	0.86	0.84	0.24	0.44	0.44	0.25	0.33	0.61	0.25	2	0.48

[1]Chromosomal location of variant.

[2]rsID corresponds to entry in NCBI dbSNP database, '.' indicates no entry available.

[3]Change based on snpEff analysis of location function and potential change of amino acid in protein sequence; syn - synonymous mutation; nonsyn - nonsynonymous mutation.

[4]Ammino acid change due to SNP

[5]dMinor Allele Frequency - ETR: Ethiopia, rural; ETU: Ethiopia, urban; GBR: The Gambia, rural; GBU: The Gambia, urban, GN: Ghana; KE: Kenya; PE: Peru; SP: Spain; SW: Sweden; USC: United States, California; USW: United States, Washington/Idaho.

[6]Evolutionary statistics to assess interpopulation differences: F_{ST}: fixation index; D: Tajima's D; π: nucleotide diversity

Table 4.8: Identified genetic variants, allele frequency, and evolutionary statistics for Secretory blood group 1, pseudogene *(SEC1P)*.

Position [1]		**rsID** [2]	**Change** [3]	**AA** [4]	**MAF** [5]											F_{ST} [6]	**D** [6]	π [6]
				ETR	ETU	GBR	GBU	GN	KE	PE	SP	SW	USC	USW	All			
				30	28	31	33	24	20	39	31	20	5	20	282			
chr19	48679906	.	upstream	0.03	0	0.02	0	0	0.07	0.04	0.03	0.06	0	0	0.02	0	-0.62	0.04
chr19	48679949	.	upstream	0	0	0.02	0	0	0.07	0.01	0	0	0	0	0.01	0	-0.77	0.02
chr19	48679992	rs56330548	upstream	0	0.04	0.09	0.1	0.16	0	0.01	0	0	0	0	0.05	0.04	-0.35	0.09
chr19	48680027	rs3826838	Splice site	0.18	0.15	0.09	0.07	0.09	0	0.34	0.03	0.14	0.4	0.11	0.14	0.07	0.59	0.24
chr19	48680126	rs115439718	upstream	0	0.02	0.05	0.05	0.06	0.07	0	0	0	0	0.06	0.03	-0.01	-0.55	0.05
chr19	48680161	.	upstream	0	0	0	0	0	0	0.01	0	0	0.1	0	0	0.04	-0.82	0.01
chr19	48680213	.	upstream	0.18	0.24	0.16	0.15	0.13	0.14	0.19	0.13	0.19	0	0.19	0.17	-0.01	0.84	0.28

[1]Chromosomal location of variant.

[2]rsID corresponds to entry in NCBI dbSNP database, '.' indicates no entry available.

[3]Change based on snpEff analysis of location function and potential change of amino acid in protein sequence; syn - synonymous mutation; nonsyn - nonsynonymous mutation.

[4]Ammino acid change due to SNP

[5]dMinor Allele Frequency - ETR: Ethiopia, rural; ETU: Ethiopia, urban; GBR: The Gambia, rural; GBU: The Gambia, urban, GN: Ghana; KE: Kenya; PE: Peru; SP: Spain; SW: Sweden; USC: United States, California; USW: United States, Washington/Idaho.

[6]Evolutionary statistics to assess interpopulation differences: F_{ST}: fixation index; D: Tajima's D; π: nucleotide diversity

References

Alam, AN, et al. "Enteric Protein Loss and Intestinal Permeability Changes in Children during Acute Shigellosis and after Recovery: Effect of Zinc Supplementation." Vol. 35, no. 12, 1994, pp. 1707-11, doi:10.1136/gut.35.12.1707.

Allard, Guy, et al. "SPINGO: A Rapid Species-Classifier for Microbial Amplicon Sequences." BMC Bioinformatics, vol. 16, no. 1, Dec. 2015. Crossref, doi:10.1186/s12859-015-0747-1.

Allen, Heather. "Swine microbiota: What's changing." (2012).

Almeida, Raul A., et al. "Staphylococcus Aureus Invasion of Bovine Mammary Epithelial Cells." Journal of Dairy Science, vol. 79, no. 6, June 1996, pp. 1021-26. Crossref, doi:10.3168/jds.S0022-0302(96)76454-8.

Altschul, Stephen F., et al. "Basic Local Alignment Search Tool." Journal of Molecular Biology, vol. 215, no. 3, Oct. 1990, pp. 403-10. Crossref, doi:10.1016/S0022-2836(05)80360-2.

Andersson, Anders F., et al. "Comparative Analysis of Human Gut Microbiota by Barcoded Pyrosequencing." PLoS ONE, edited by Niyaz Ahmed, vol. 3, no. 7, July 2008, p. e2836. Crossref, doi:10.1371/journal.pone.0002836.

Autran, Chloe A., et al. "Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants." Gut 67.6 (2018): 1064-1070.

Beaumont, Mark A. "Adaptation and Speciation: What Can Fst Tell Us?" Trends Ecol Evol, vol. 20, no. 8, 2005, pp. 435-40, doi:10.1016/j.tree.2005.05.017.

Bering, Stine. "Human Milk Oligosaccharides to Prevent Gut Dysfunction and Necrotizing Enterocolitis in Preterm Neonates." Nutrients, vol. 10, no. 10, Oct. 2018, p. 1461. Crossref, doi:10.3390/nu10101461.

Bhatt, Vaibhav D., et al. "Identification and Antibiogram of Microbes Associated with Bovine Mastitis." Vol. 22, no. 3, 2011, pp. 163-69, doi:10.1080/10495398.2011.570132.

Bjorksten, B., et al. "Collecting and Banking Human Milk: To Heat or Not to Heat?" BMJ, vol. 281, no. 6243, Sept. 1980, pp. 765-69. Crossref, doi:10.1136/bmj.281.6243.765.

Blackburn, Daniel G., et al. "The Origins of Lactation and the Evolution of Milk: A Review with New Hypotheses." Mammal Review, vol. 19, no. 1, Mar. 1989, pp. 1-26. Crossref, doi:10.1111/j.1365-2907.1989.tb00398.x.

Bode, L. "Human Milk Oligosaccharides: Every Baby Needs a Sugar Mama." Glycobiology, vol. 22, no. 9, Sept. 2012, pp. 1147-62. Crossref, doi:10.1093/glycob/cws074.

Bode, L., and Jantscher-Krenn E. "Structure-Function Relationships of Human Milk Oligosaccharides." Vol. 3, no. 3, 2012, p. 383S391S, doi:10.3945/an.111.001404.

Bonsaglia, Erika C. R., et al. "Milk Microbiome and Bacterial Load Following Dry Cow Therapy without Antibiotics in Dairy Cows with Healthy Mammary Gland." Scientific Reports, vol. 7, no. 1, Dec. 2017. Crossref, doi:10.1038/s41598-017-08790-5.

Bouchard, Damien S., et al. "Lactic Acid Bacteria Isolated from Bovine Mammary Microbiota: Potential Allies against Bovine Mastitis." PLOS ONE, edited by Marie-Joelle Virolle, vol. 10, no. 12, Dec. 2015, p. e0144831. Crossref, doi:10.1371/journal.pone.0144831.

Braem, G., et al. "Antibacterial Activities of Coagulase-Negative Staphylococci from Bovine Teat Apex Skin and Their Inhibitory Effect on Mastitis-Related Pathogens." Journal of Applied Microbiology, vol. 116, no. 5, May 2014, pp. 1084-93. Crossref, doi:10.1111/jam.12447.

Braem, Gorik, et al. "Culture-Independent Exploration of the Teat Apex Microbiota of Dairy Cows Reveals a Wide Bacterial Species Diversity." Veterinary Microbiology, vol. 157, no. 3-4, June 2012, pp. 383-90. Crossref, doi:10.1016/j.vetmic.2011.12.031.

Broom, L. J., et al. "Effects of Zinc Oxide and Enterococcus Faecium SF68 Dietary Supplementation on the Performance, Intestinal Microbiota and Immune Status of Weaned Piglets." Research in Veterinary Science, vol. 80, no. 1, Feb. 2006, pp. 45-54. Crossref, doi:10.1016/j.rvsc.2005.04.004.

Brunet, J. P., et al. "Metagenes and Molecular Pattern Discovery Using Matrix Factorization." Proceedings of the National Academy of Sciences, vol. 101, no. 12, Mar. 2004, pp. 4164-69. Crossref, doi:10.1073/pnas.0308531101.

Buzoianu, Stefan G., et al. "High-Throughput Sequence-Based Analysis of the Intestinal Microbiota of Weanling Pigs Fed Genetically Modified MON810 Maize Expressing Bacillus Thuringiensis Cry1Ab (Bt Maize) for 31 Days." Applied and Environmental Microbiology, vol. 78, no. 12, June 2012, pp. 4217-24. Crossref, doi:10.1128/AEM.00307-12.

Cai, Yun, et al. "Learning Microbial Community Structures with Supervised and Unsupervised Non-Negative Matrix Factorization." Microbiome, vol. 5, no. 1, Dec. 2017. Crossref, doi:10.1186/s40168-017-0323-1.

Callahan, Benjamin J., et al. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." Nature Methods, vol. 13, no. 7, July 2016, pp. 581-83. Crossref, doi:10.1038/nmeth.3869.

Cani, Patrice D., et al. "Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet-Induced Obesity and Diabetes in Mice." Vol. 57, no. 6, 2008, pp. 1470-81, doi:10.2337/db07-1403.

Castillo, M., et al. "Use of Mannanoligosaccharides and Zinc Chelate as Growth Promoters and Diarrhea Preventative in Weaning Pigs: Effects on Microbiota and Gut Function1." Journal of Animal Science, vol. 86, no. 1, Jan. 2008, pp. 94-101. Crossref, doi:10.2527/jas.2005-686.

Chen, Hsin-Mao, et al. "Kocuria Kristinae: A True Pathogen in Pediatric Patients." Journal of Microbiology, Immunology and Infection, vol. 48, no. 1, Feb. 2015, pp. 80-84. Crossref, doi:10.1016/j.jmii.2013.07.001.

Collado, M. C., et al. "Assessment of the Bacterial Diversity of Breast Milk of Healthy Women by Quantitative Real-time PCR." Vol. 48, no. 5, 2009, pp. 523-28, doi:10.1111/j.1472-765X.2009.02567.x.

Colodro-Conde, Lucia, et al. "A Twin Study of Breastfeeding With a Preliminary Genome-Wide Association Scan." Twin Research and Human Genetics, vol. 18, no. 01, Feb. 2015, pp. 61-72. Crossref, doi:10.1017/thg.2014.74.

Contreras, Andres G., and Juan Rodriguez. "Mastitis: Comparative Etiology and Epidemiology." Vol. 16, no. 4, 2011, pp. 339-56, doi:10.1007/s10911-011-9234-0.

Costa, Matheus O., et al. "Characterization of the Fecal Microbiota of Pigs before and after Inoculation with 'Brachyspira Hampsonii."' PLoS ONE, edited by Gourapura J. Renukaradhya, vol. 9, no. 8, Aug. 2014, p. e106399. Crossref, doi:10.1371/journal.pone.0106399.

Cressier, B., and N. Bissonnette. "Assessment of an Extraction Protocol to Detect the Major Mastitis-Causing Pathogens in Bovine Milk." Journal of Dairy Science, vol. 94, no. 5, May 2011, pp. 2171-84. Crossref, doi:10.3168/jds.2010-3669.

Crispie, Fiona, et al. "Intramammary Infusion of a Live Culture for Treatment of Bovine Mastitis: Effect of Live Lactococci on the Mammary Immune Response." Journal of Dairy Research, vol. 75, no. 03, Aug. 2008. Crossref, doi:10.1017/S0022029908003385.

Dalca, A. V., and M. Brudno. "Genome Variation Discovery with High-Throughput Sequencing Data." Briefings in Bioinformatics, vol. 11, no. 1, Jan. 2010, pp. 3-14. Crossref, doi:10.1093/bib/bbp058.

Danecek, P., et al. "The Variant Call Format and VCFtools." Bioinformatics, vol. 27, no. 15, Aug. 2011, pp. 2156-58. Crossref, doi:10.1093/bioinformatics/btr330.

Delgado, Susana, et al. "PCR-DGGE Assessment of the Bacterial Diversity of Breast Milk in Women with Lactational Infectious Mastitis." Vol. 8, no. 1, 2008, p. 51, doi:10.1186/1471-2334-8-51.

Derakhshani, Hooman, et al. "Invited Review: Microbiota of the Bovine Udder: Contributing Factors and Potential Implications for Udder Health and Mastitis Susceptibility." Journal of Dairy Science, Oct. 2018. Crossref, doi:10.3168/jds.2018-14860.

Doyle, CJ, et al. "Impacts of Seasonal Housing and Teat Preparation on Raw Milk Microbiota: A High-Throughput Sequencing Study." Applied and environmental microbiology, 2017, doi:10.1128/AEM.02694-16.

Durbin, Richard M., et al. "A Map of Human Genome Variation from Population-Scale Sequencing." Nature, vol. 467, no. 7319, 2011, pp. 1061-73, doi:10.1038/nature09534.

Erney, RM, et al. "Variability of Human Milk Neutral Oligosaccharides in a Diverse Population." Journal of Pediatrics, 2000.

Espejo, Mariano Ruiz. "WHO Child Growth Standards: Methods and Development." Journal of the Royal Statistical Society: Series A (Statistics in Society), vol. 170, no. 2, Mar. 2007, pp. 512-512. Crossref, $doi : 10.1111/j.1467 - 985X.2007.00473_18.x$.

Falentin, Helene, et al. "Bovine Teat Microbiome Analysis Revealed Reduced Alpha Diversity and Significant Changes in Taxonomic Profiles in Quarters with a History of Mastitis." Front Microbiol, vol. 7, 2016, doi:10.3389/fmicb.2016.00480.

Fanos, Vassilios, et al. "Miracles and Mysteries of Breast Milk: From Egyptians to the 3 M's (Metabolomics, Microbiomics, Multipotent Stem Cells)." Journal of Pediatric and Neonatal Individualized Medicine, no. 2, Oct. 2017, p. e060204. mEDRA, doi:10.7363/060204.

Fernandez, Leonides, et al. "The Human Milk Microbiota: Origin and Potential Roles in Health and Disease." Vol. 69, no. 1, 2013, p. 110, doi:10.1016/j.phrs.2012.09.001.

Ferrer-Admetlla, A., et al. "A Natural History of FUT2 Polymorphism in Humans." Molecular Biology and Evolution, vol. 26, no. 9, Sept. 2009, pp. 1993-2003. Crossref, doi:10.1093/molbev/msp108.

Flint, Harry J., et al. "The Role of the Gut Microbiota in Nutrition and Health." Nature Reviews Gastroenterology & Hepatology, vol. 9, no. 10, Oct. 2012, pp. 577-89. Crossref, doi:10.1038/nrgastro.2012.156.

Fumagalli, Matteo, et al. "Widespread Balancing Selection and Pathogen-Driven Selection at Blood Group Antigen Genes." Biotechfor, vol. 19, no. 2, 2009, pp. 199-212, doi:10.1101/gr.082768.108.

Ganda, EK, et al. "Longitudinal Metagenomic Profiling of Bovine Milk to Assess the Impact of Intramammary Treatment Using a Third-Generation Cephalosporin." Scientific reports 6 (2016): 37565.

Ganda, Erika K., et al. "Normal Milk Microbiome Is Reestablished Following Experimental Infection with Escherichia Coli Independent of Intramammary Antibiotic Treatment with a Third-Generation Cephalosporin in Bovines." Microbiome, vol. 5, no. 1, Dec. 2017. Crossref, doi:10.1186/s40168-017-0291-5.

Gaujoux, Renaud, and Cathal Seoighe. "A Flexible R Package for Nonnegative Matrix Factorization." Bmc Bioinformatics, vol. 11, no. 1, 2010, pp. 1-9, doi:10.1186/1471-2105-11-367.

Gill, Steven R., et al. "Metagenomic Analysis of the Human Distal Gut Microbiome." Vol. 312, no. 5778, 2006, pp. 1355-59, doi:10.1126/science.1124234.

Govers, M. J., and R. Van der Meet. "Effects of Dietary Calcium and Phosphate on the Intestinal Interactions between Calcium, Phosphate, Fatty Acids, and Bile Acids." Gut, vol. 34, no. 3, Mar. 1993, pp. 365-70.

Guo, Meng, et al. "Distribution of Lewis and Secretor Polymorphisms and Corresponding CA19-9 Antigen Expression in a Chinese Population." FEBS Open Bio, vol. 7, no. 11, Nov. 2017, pp. 1660-71. Crossref, doi:10.1002/2211-5463.12278.

Haas, B. J., et al. "Chimeric 16S RRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons." Genome Research, vol. 21, no. 3, Mar. 2011, pp. 494-504. Crossref, doi:10.1101/gr.112730.110.

Haenen, Danielle, et al. "A Diet High in Resistant Starch Modulates Microbiota Composition, SCFA Concentrations, and Gene Expression in Pig Intestine." The Journal of Nutrition, vol. 143, no. 3, Mar. 2013, pp. 274-83. Crossref, doi:10.3945/jn.112.169672.

Hale, and Hartmann. Textbook of Human Lactation. 2007.

Han, Kui, et al. "Extraordinary Expansion of a Sorangium Cellulosum Genome from an Alkaline Milieu." Scientific Reports, vol. 3, no. 1, Dec. 2013. Crossref, doi:10.1038/srep02101.

Hayssen, V, and DG Blackburn. "α-Lactalbumin and the Origins of Lactation." Evolution, 1985, pp. 1147-49.

Heather, James M., and Benjamin Chain. "The Sequence of Sequencers: The History of Sequencing DNA." Genomics, vol. 107, no. 1, Jan. 2016, pp. 1-8. Crossref, doi:10.1016/j.ygeno.2015.11.003.

Heikkila, M. P., and P. E. J. Saris. "Inhibition of Staphylococcus Aureus by the Commensal Bacteria of Human Milk." Journal of Applied Microbiology, vol. 95, no. 3, 2003, pp. 471-78, doi:10.1046/j.1365-2672.2003.02002.x.

Heinritz, Sonja N., et al. "Intestinal Microbiota and Microbial Metabolites Are Changed in a Pig Model Fed a High-Fat/Low-Fiber or a Low-Fat/High-Fiber Diet." PLOS ONE, edited by Erwin G Zoetendal, vol. 11, no. 4, Apr. 2016, p. e0154329. Crossref, doi:10.1371/journal.pone.0154329.

Heneghan, MA, et al. "Relationship of Blood Group Determinants on Helicobacter Pylori Lipopolysaccharide with Host Lewis Phenotype and Inflammatory Response." Infection and Immunity, 2000, doi:10.1128/IAI.68.2.937-941.2000.

Henry, Stephen. "The Se w FUT2 Mutation A385T Does Not Result in a Nonsecretor Allele." Transfusion, vol. 54, no. 12, Dec. 2014, pp. 3255-3255. Crossref, doi:10.1111/trf.12909.

Hojberg, O., et al. "Influence of Dietary Zinc Oxide and Copper Sulfate on the Gastrointestinal Ecosystem in Newly Weaned Piglets." Applied and Environmental Microbiology, vol. 71, no. 5, May 2005, pp. 2267-77. Crossref, doi:10.1128/AEM.71.5.2267-2277.2005.

Holsinger, Kent E., and Bruce S. Weir. "Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting F_{ST}." Nat Rev Genet, vol. 10, no. 9, 2009, pp. 639-50, doi:10.1038/nrg2611.

Hughes, J. B., et al. "Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity." Applied and Environmental Microbiology, vol. 67, no. 10, Oct. 2001, pp. 4399-406. Crossref, doi:10.1128/AEM.67.10.4399-4406.2001.

Hunt, Katherine M., James A. Foster, et al. "Characterization of the Diversity and Temporal Stability of Bacterial Communities in Human Milk." Plos One, vol. 6, no. 6, 2011, p. e21313, doi:10.1371/journal.pone.0021313.

Huse, Susan M., et al. "Ironing out the Wrinkles in the Rare Biosphere through Improved OTU Clustering." Vol. 12, no. 7, 2010, pp. 1889-98, doi:10.1111/j.1462-2920.2010.02193.x.

Ibeagha-Awemu, Eveline M., et al. "High Density Genome Wide Genotyping-by-Sequencing and Association Identifies Common and Low Frequency SNPs, and Novel Candidate Genes Influencing Cow Milk Traits." Scientific Reports, vol. 6, no. 1, Nov. 2016. Crossref, doi:10.1038/srep31109.

Isaacson, Richard, and Hyeun Bum Kim. "The Intestinal Microbiome of the Pig." Animal Health Research Reviews, vol. 13, no. 01, June 2012, pp. 100-09. Crossref, doi:10.1017/S1466252312000084.

Jabed, A., et al. "Targeted MicroRNA Expression in Dairy Cattle Directs Production of -Lactoglobulin-Free, High-Casein Milk." Proceedings of the National Academy of Sciences, vol. 109, no. 42, Oct. 2012, pp. 16811-16. Crossref, doi:10.1073/pnas.1210057109.

Jadhav, P. V., et al. "Threshold Somatic Cell Count for Delineation of Subclinical Mastitis Cases." Veterinary World, vol. 11, no. 6, June 2018, pp. 789-93. Crossref, doi:10.14202/vetworld.2018.789-793.

Janczyk, Pawel, et al. "No Protective Effects of High-Dosage Dietary Zinc Oxide on Weaned Pigs Infected with Salmonella Enterica Serovar Typhimurium DT104." Applied and Environmental Microbiology, vol. 79, no. 9, May 2013, pp. 2914-21. Crossref, doi:10.1128/AEM.03577-12.

Jantscher-Krenn, Evelyn, et al. "The human milk oligosaccharide disialyllacto-N-tetraose prevents necrotising enterocolitis in neonatal rats." Gut 61.10 (2012): 1417-1425.

Jimenez, E., et al. "Oral Administration of Lactobacillus Strains Isolated from Breast Milk as an Alternative for the Treatment of Infectious Mastitis during Lactation." Vol. 74, no. 15, 2008, pp. 4650-55, doi:10.1128/AEM.02599-07.

Jimenez, Esther. "From the Human Milk Microbiota to the Human Milk Metagenome." In: McGuire, M.K., M.A. McGuire and L. Bode, Prebiotics and Probiotics in Human Milk. pp. 463-468. ISBN: 978-0-12-802725-7. Elsevier. San Diego CA.

Jonnala, Bhagya R. Yeluri, et al. "Sequencing of the Cheese Microbiome and its relevance to Industry." Frontiers in microbiology 9 (2018).

Jost, Ted, et al. "Assessment of Bacterial Diversity in Breast Milk Using Culture-Dependent and Culture-Independent Approaches." 2013, p. 110, doi:10.1017/S0007114513000597.

Katouli, M., et al. "The Effect of Zinc Oxide Supplementation on the Stability of the Intestinal Flora with Special Reference to Composition of Coliforms in Weaned Pigs." Journal of Applied Microbiology, vol. 87, no. 4, Oct. 1999, pp. 564-73.

Kim, Hyeun, et al. "Longitudinal Investigation of the Age-Related Bacterial Diversity in the Feces of Commercial Pigs." Vol. 153, no. 1-2, 2011, p. 124133, doi:10.1016/j.vetmic.2011.05.021.

Kim, Hyeun, et al. "Microbial Shifts in the Swine Distal Gut in Response to the Treatment with Antimicrobial Growth Promoter, Tylosin." Vol. 109, no. 38, 2012, pp. 15485-90, doi:10.1073/pnas.1205147109.

Klostermann, Katja, et al. "Intramammary Infusion of a Live Culture of Lactococcus Lactis for Treatment of Bovine Mastitis: Comparison with Antibiotic Treatment in Field Trials." Journal of Dairy Research, vol. 75, no. 03, Aug. 2008. Crossref, doi:10.1017/S0022029908003373.

Koda, Y., H. Tachida, H. Pang, et al. "Contrasting Patterns of Polymorphisms at the ABO-Secretor Gene (FUT2) and Plasma α (1, 3) Fucosyltransferase Gene (FUT6) in Human Populations." Genetics 158.2, 2001, 747-756.

Koda, Y., H. Tachida, M. Soejima, et al. "Ancient Origin of the Null Allele Se 428 of the Human ABO-Secretor Locus (FUT2)." Journal of Molecular, 2000.

Koda, Yoshiro, et al. "Structure and Expression of the Gene Encoding Secretor-Type Galactoside 2-α-l-fucosyltransferase (FUT2)." Eur J Biochem, vol. 246, no. 3, 1997, pp. 750-55, doi:10.1111/j.1432-1033.1997.t01-1-00750.x.

Kononova, S. V. "How Fucose of Blood Group Glycotypes Programs Human Gut Microbiota." Biochemistry (Moscow), vol. 82, no. 9, Sept. 2017, pp. 973-89. doi:10.1134/S0006297917090012.

Konstantinov, Sergey R., et al. "Post-Natal Development of the Porcine Microbiota Composition and Activities." Environmental Microbiology, vol. 8, no. 7, July 2006, pp. 1191-99. Crossref, doi:10.1111/j.1462-2920.2006.01009.x.

Koskinen, MT, et al. "Field Comparison of Real-Time Polymerase Chain Reaction and Bacterial Culture for Identification of Bovine Mastitis Bacteria." Vol. 93, no. 12, 2010, pp. 5707-15, doi:10.3168/jds.2010-3167.

Kuang, Ying, et al. "Characterization of Bacterial Population of Raw Milk from Bovine Mastitis by Culture-Independent PCR-DGGE Method." Biochemical Engineering Journal, vol. 45, no. 1, June 2009, pp. 76-81. Crossref, doi:10.1016/j.bej.2009.02.010.

Kuehn, Joanna S., et al. "Bacterial Community Profiling of Milk Samples as a Means to Understand Culture-Negative Bovine Clinical Mastitis." Plos One, vol. 8, no. 4, 2013, p. e61959, doi:10.1371/journal.pone.0061959.

Kumazaki, T., and A. Yoshida. "Biochemical Evidence That Secretor Gene, Se, Is a Structural Gene Encoding a Specific Fucosyltransferase." Proceedings of the National Academy of Sciences 81.13 (1984): 4193-4197.

Lamendella, Regina, et al. "Comparative Fecal Metagenomics Unveils Unique Functional Capacity of the Swine Gut." BMC Microbiology, vol. 11, no. 1, 2011, p. 103. Crossref, doi:10.1186/1471-2180-11-103.

Lee, In Kyu, et al. "Stress, Nutrition, and Intestinal Immune Responses in Pigs - A Review." Asian-Australasian Journal of Animal Sciences, vol. 29, no. 8, May 2016, pp. 1075-82. Crossref, doi:10.5713/ajas.16.0118.

Lee, Sooyeon, and Shannon L. Kelleher. "Biological Underpinnings of Breastfeeding Challenges: The Role of Genetics, Diet, and Environment on Lactation Physiology." American Journal of Physiology-Endocrinology and Metabolism, vol. 311, no. 2, Aug. 2016, pp. E405-22. Crossref, doi:10.1152/ajpendo.00495.2015.

Leitner, G., et al. "Milk Leucocyte Population Patterns in Bovine Udder Infection of Different Aetiology." Vol. 47, no. 8, 2000, pp. 581-89, doi:10.1046/j.1439-0450.2000.00388.x.

Lemay, Danielle G., David J. Lynn, et al. "The Bovine Lactation Genome: Insights into the Evolution of Mammalian Milk." Vol. 10, no. 4, 2009, pp. 1-18, doi:10.1186/gb-2009-10-4-r43.

Lemay, Danielle G., Monique Rijnkels, et al. "Lessons from the Bovine Genome: Implications for Human Nutrition and Research." J Nutrition, vol. 139, no. 7, 2009, pp. 1271-72, doi:10.3945/jn.109.107656.

Leser, T. D., et al. "Culture-Independent Analysis of Gut Bacteria: The Pig Gastrointestinal Tract Microbiota Revisited." Applied and Environmental Microbiology, vol. 68, no. 2, Feb. 2002, pp. 673-90. Crossref, doi:10.1128/AEM.68.2.673-690.2002.

Li, Hongzhe. "Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis." Vol. 2, no. 1, 2015, pp. 73-94, doi:10.1146/annurev-statistics-010814-020351.

Li, Kaifeng, et al. "Microbial Composition in Different Gut Locations of Weaning Piglets Receiving Antibiotics." Asian-Australasian Journal of Animal Sciences, vol. 30, no. 1, June 2016, pp. 78-84. Crossref, doi:10.5713/ajas.16.0285.

Li, Nan, et al. "Variation in raw milk microbiota throughout 12 months and the impact of weather conditions." Scientific reports 8.1 (2018): 2371.

Li, Xiaochuang, et al. "An Overview of Diversity, Occurrence, Genetics and Toxin Production of Bloom-Forming Dolichospermum (Anabaena) Species." Harmful Algae, vol. 54, Apr. 2016, pp. 54-68. Crossref, doi:10.1016/j.hal.2015.10.015.

Lima, Svetlana F., et al. "The Bovine Colostrum Microbiome and Its Association with Clinical Mastitis." Journal of Dairy Science, vol. 100, no. 4, Apr. 2017, pp. 3031-42. Crossref, doi:10.3168/jds.2016-11604.

Lima, Svetlana Ferreira, et al. "Evaluation of Milk Sample Fractions for Characterization of Milk Microbiota from Healthy and Clinical Mastitis Cows." PLOS ONE, edited by Juan J. Loor, vol. 13, no. 3, Mar. 2018, p. e0193671. Crossref, doi:10.1371/journal.pone.0193671.

Liu, Zongzhi, et al. "Accurate Taxonomy Assignments from 16S RRNA Sequences Produced by Highly Parallel Pyrosequencers." Vol. 36, no. 18, 2008, pp. e120-e120, doi:10.1093/nar/gkn491.

Lluch, Jerome, et al. "The Characterization of Novel Tissue Microbiota Using an Optimized 16S Metagenomic Sequencing Pipeline." PLoS ONE, vol. 10, no. 11, 2015, p. e0142334, doi:10.1371/journal.pone.0142334.

Lonnerdal, Bo. "Effects of Maternal Dietary Intake on Human Milk Composition." The Journal of Nutrition, vol. 116, no. 4, Apr. 1986, pp. 499-513. Crossref, doi:10.1093/jn/116.4.499.

Looft, Torey, Heather K. Allen, et al. "Bacteria, Phages and Pigs: The Effects of in-Feed Antibiotics on the Microbiome at Different Gut Locations." The ISME Journal, vol. 8, no. 8, Aug. 2014, pp. 1566-76. Crossref, doi:10.1038/ismej.2014.12.

Looft, Torey, Timothy A. Johnson, et al. "In-Feed Antibiotic Effects on the Swine Intestinal Microbiome." Vol. 109, no. 5, 2012, pp. 1691-96, doi:10.1073/pnas.1120238109.

Ma, Edmond SK, et al. "Kocuria Kristinae Infection Associated with Acute Cholecystitis." BMC Infectious Diseases, vol. 5, no. 1, Dec. 2005. Crossref, doi:10.1186/1471-2334-5-60.

(Sam) Ma, Zhanshan, et al. "Network Analysis Suggests a Potentially 'Evil' Alliance of Opportunistic Pathogens Inhibited by a Cooperative Network in Human Milk Bacterial Communities." Scientific Reports, vol. 5, no. 1, July 2015. Crossref, doi:10.1038/srep08275.

Mach, Nuria, et al. "Early-Life Establishment of the Swine Gut Microbiome and Impact on Host Phenotypes: Role of Early-Life Gut Microbiome on Pigs' Health." Environmental Microbiology Reports, vol. 7, no. 3, June 2015, pp. 554-69. Crossref, doi:10.1111/1758-2229.12285.

Mahmmod, YS, et al. "Effect of Presampling Procedures on Real-Time PCR Used for Diagnosis of Intramammary Infections with Staphylococcus Aureus in Dairy Cows at Routine Milk Recordings." Vol. 96, no. 4, 2013, p. 22262233, doi:10.3168/jds.2012-6059.

Malik, Tariq Ahmad, et al. "Somatic Cells in Relation to Udder Health and Milk Quality-A Review." Journal of Animal Health and Production, vol. 6, no. 1, 2018. Crossref, doi:10.17582/journal.jahp/2018/6.1.18.26.

Mann, Evelyne, et al. "Mucosa-Associated Bacterial Microbiome of the Gastrointestinal Tract of Weaned Pigs and Dynamics Linked to Dietary Calcium-Phosphorus." PLoS ONE, edited by Francois Blachier, vol. 9, no. 1, Jan. 2014, p. e86950. Crossref, doi:10.1371/journal.pone.0086950.

Mardis, Elaine R. "The Impact of Next-Generation Sequencing Technology on Genetics." Trends in Genetics, vol. 24, no. 3, Mar. 2008, pp. 133-41. Crossref, doi:10.1016/j.tig.2007.12.007.

Margulies, Marcel, et al. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." Vol. 437, no. 7057, 2005, pp. 376-80, doi:10.1038/nature03959.

MartÃn, RocÃo, et al. "Human Milk Is a Source of Lactic Acid Bacteria for the Infant Gut." Vol. 143, no. 6, 2003, doi:10.1016/j.jpeds.2003.09.028.

Matto, Jaana, et al. "Composition and Temporal Stability of Gastrointestinal Microbiota in Irritable Bowel Syndrome - a Longitudinal Study in IBS and Control Subjects." FEMS Immunology & Medical Microbiology, vol. 43, no. 2, Feb. 2005, pp. 213-22. Crossref, doi:10.1016/j.femsim.2004.08.009.

McGuire, M.A. and M.K. McGuire. 2017. "Human milk microbes - Summary and research gaps." In: McGuire, M.K., M.A. McGuire and L. Bode, Prebiotics and Probiotics in Human Milk. pp. 463-468. ISBN: 978-0-12-802725-7. Elsevier. San Diego CA.

McGuire, M.A. and M.K. McGuire. 2017. "Isn't Milk Sterile? A Historical Perspective on Microbes in Milk." In: McGuire, M.K., M.A. McGuire and L. Bode, Prebiotics and Probiotics in Human Milk. pp. 463-468. ISBN: 978-0-12-802725-7. Elsevier. San Diego CA.

McGuire, Michelle K., et al. "What's Normal? Oligosaccharide Concentrations and Profiles in Milk Produced by Healthy Women Vary Geographically." The American Journal of Clinical Nutrition, vol. 105, no. 5, May 2017, pp. 1086-100. Crossref, doi:10.3945/ajcn.116.139980.

McKenna, A., et al. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." Genome Research, vol. 20, no. 9, Sept. 2010, pp. 1297-303. Crossref, doi:10.1101/gr.107524.110.

Mediano, Pilar, et al. "Microbial Diversity in Milk of Women With Mastitis: Potential Role of Coagulase-Negative Staphylococci, Viridans Group Streptococci, and Corynebacteria." Journal of Human Lactation, vol. 33, no. 2, May 2017, pp. 309-18. Crossref, doi:10.1177/0890334417692968.

Meehan, Courtney L., et al. "Social Networks, Cooperative Breeding, and the Human Milk Microbiome." American Journal of Human Biology, vol. 30, no. 4, July 2018, p. e23131. Crossref, doi:10.1002/ajhb.23131.

Metzger, Stephanie A., et al. "A Cohort Study of the Milk Microbiota of Healthy and Inflamed Bovine Mammary Glands From Dryoff Through 150 Days in Milk." Frontiers in Veterinary Science, vol. 5, Oct. 2018. Crossref, doi:10.3389/fvets.2018.00247.

Metzler-Zebeli, B. U., S. Hooda, et al. "Nonstarch Polysaccharides Modulate Bacterial Microbiota, Pathways for Butyrate Production, and Abundance of Pathogenic Escherichia Coli in the Pig Gastrointestinal Tract." Applied and Environmental Microbiology, vol. 76, no. 11, June 2010, pp. 3692-701. Crossref, doi:10.1128/AEM.00257-10.

Metzler-Zebeli, B. U., W. Vahjen, et al. "Ileal Microbiota of Growing Pigs Fed Different Dietary Calcium Phosphate Levels and Phytase Content and Subjected to Ileal Pectin Infusion1." Journal of Animal Science, vol. 88, no. 1, Jan. 2010, pp. 147-58. Crossref, doi:10.2527/jas.2008-1560.

Misawa, K., and F. Tajima. "Estimation of the Amount of DNA Polymorphism When the Neutral Mutation Rate Varies among Sites." Genetics, vol. 147, no. 4, Dec. 1997, pp. 1959-64.

Morrow, Ardythe L., et al. "Human-Milk Glycans That Inhibit Pathogen Binding Protect Breast-Feeding Infants against Infectious Diarrhea." The Journal of Nutrition, vol. 135, no. 5, May 2005, pp. 1304-07. Crossref, doi:10.1093/jn/135.5.1304.

Muyzer, G., et al. "Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S RRNA." Applied and Environmental Microbiology, vol. 59, no. 3, Mar. 1993, pp. 695-700.

Niu, Qing, et al. "Dynamic Distribution of the Gut Microbiota and the Relationship with Apparent Crude Fiber Digestibility and Growth Stages in Pigs." Scientific Reports, vol. 5, no. 1, Sept. 2015. Crossref, doi:10.1038/srep09938.

Oftedal, O. T. "The Evolution of Milk Secretion and Its Ancient Origins." Animal, vol. 6, no. 03, Mar. 2012, pp. 355-68. Crossref, doi:10.1017/S1751731111001935.

Oikonomou, Georgios, Marcela Bicalho, et al. "Microbiota of Cow's Milk; Distinguishing Healthy, Sub-Clinically and Clinically Diseased Quarters." Plos One, vol. 9, no. 1, 2014, p. e85904, doi:10.1371/journal.pone.0085904.

Oikonomou, Georgios, Vinicius Machado, et al. "Microbial Diversity of Bovine Mastitic Milk as Described by Pyrosequencing of Metagenomic 16s RDNA." Plos One, vol. 7, no. 10, 2012, p. e47671, doi:10.1371/journal.pone.0047671.

Pajarillo, Edward Alain B., Jong Pyo Chae, Hyeun Bum Kim, et al. "Barcoded Pyrosequencing-Based Metagenomic Analysis of the Faecal Microbiome of Three Purebred Pig Lines after Cohabitation." Applied Microbiology and Biotechnology, vol. 99, no. 13, July 2015, pp. 5647-56. Crossref, doi:10.1007/s00253-015-6408-5.

Pajarillo, Edward Alain B., Jong Pyo Chae, Marilen P. Balolong, et al. "Pyrosequencing-Based Analysis of Fecal Microbial Communities in Three Purebred Pig Lines." Journal of Microbiology, vol. 52, no. 8, Aug. 2014, pp. 646-51. Crossref, doi:10.1007/s12275-014-4270-2.

Pan, D. and Z. Yu. 2014. "Intestinal microbiome of poultry and its interaction with host and diet." Gut Microbes 5:1, 108-119, DOI: 10.4161/gmic.26945.

Park, Y. K., et al. "The Analysis of Milk Components and Pathogenic Bacteria Isolated from Bovine Raw Milk in Korea." Journal of Dairy Science, vol. 90, no. 12, Dec. 2007, pp. 5405-14. Crossref, doi:10.3168/jds.2007-0282.

Patel, Shriram H., et al. "Culture Independent Assessment of Human Milk Microbial Community in Lactational Mastitis." Scientific Reports, vol. 7, no. 1, Dec. 2017. Crossref, doi:10.1038/s41598-017-08451-7.

Pegolo, Sara, et al. "Integration of GWAS, Pathway and Network Analyses Reveals Novel Mechanistic Insights into the Synthesis of Milk Proteins in Dairy Cows." Scientific Reports, vol. 8, no. 1, Dec. 2018. Crossref, doi:10.1038/s41598-017-18916-4.

Pellegrino, M., et al. "Bovine Mastitis Prevention: Humoral and Cellular Response of Dairy Cows Inoculated with Lactic Acid Bacteria at the Dry-off Period." Beneficial Microbes, vol. 8, no. 4, Aug. 2017, pp. 589-96. Crossref, doi:10.3920/BM2016.0194.

Perez, Pablo F., et al. "Bacterial Imprinting of the Neonatal Immune System: Lessons From Maternal Cells?" Vol. 119, no. 3, 2007, pp. e724-32, doi:10.1542/peds.2006-1649.

Pieper, R., et al. "Dose-Dependent Effects of Dietary Zinc Oxide on Bacterial Communities and Metabolic Profiles in the Ileum of Weaned Pigs: Dose-Dependent Effects of Dietary Zinc Oxide." Journal of Animal Physiology and Animal Nutrition, vol. 96, no. 5, Oct. 2012, pp. 825-33. Crossref, doi:10.1111/j.1439-0396.2011.01231.x.

Poroyko, Valeriy, et al. "Gut Microbial Gene Expression in Mother-Fed and Formula-Fed Piglets." Vol. 5, no. 8, 2010, doi:10.1371/journal.pone.0012459.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

Quast, Christian, et al. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." Nucleic Acids Research, vol. 41, no. D1, Nov. 2012, pp. D590-96. Crossref, doi:10.1093/nar/gks1219.

Quigley, Lisa, et al. "Molecular Approaches to Analysing the Microbial Composition of Raw Milk and Raw Milk Cheese." Int J Food Microbiol, vol. 150, no. 2-3, 2011, pp. 81-94, doi:10.1016/j.ijfoodmicro.2011.08.001.

Raats, Dina, et al. "Molecular analysis of bacterial communities in raw cow milk and the impact of refrigeration on its structure and dynamics." Food Microbiology 28.3 (2011): 465-471.

R Development Core Team. "R: a language and environment for statistical computing." Vienna (Austria): R Foundation for Statistical Computing; 2016.

Ramayo-Caldas, Yuliaxis, et al. "Phylogenetic Network Analysis Applied to Pig Gut Microbiota Identifies an Ecosystem Structure Linked with Growth Traits." The ISME Journal, vol. 10, no. 12, Dec. 2016, pp. 2973-77. Crossref, doi:10.1038/ismej.2016.77.

Renaudeau, D., et al. "A Meta-Analysis of the Effects of High Ambient Temperature on Growth Performance of Growing-Finishing Pigs." Journal of Animal Science, vol. 89, no. 7, July 2011, pp. 2220-30. Crossref, doi:10.2527/jas.2010-3329.

Ridenhour, Benjamin J., et al. "Modeling Time-Series Data from Microbial Communities." The ISME Journal, vol. 11, no. 11, Nov. 2017, pp. 2526-37. Crossref, doi:10.1038/ismej.2017.107.

Rinttila, T., et al. "Development of an Extensive Set of 16S RDNA-Targeted Primers for Quantification of Pathogenic and Indigenous Bacteria in Faecal Samples by Real-Time PCR." Journal of Applied Microbiology, vol. 97, no. 6, Dec. 2004, pp. 1166-77. Crossref, doi:10.1111/j.1365-2672.2004.02409.x.

Rodrigues, M. X., et al. "The Microbiome of Bulk Tank Milk: Characterization and Associations with Somatic Cell Count and Bacterial Count." Journal of Dairy Science, vol. 100, no. 4, Apr. 2017, pp. 2536-52. Crossref, doi:10.3168/jds.2016-11540.

Rodrigues, Paulo C., et al. "A Robust AMMI Model for the Analysis of Genotype-by-Environment Data." Bioinformatics, Sept. 2015, p. btv533. Crossref, doi:10.1093/bioinformatics/btv533.

Ruiz, Lorena, et al. "What's Normal? Immune Profiling of Human Milk from Healthy Women Living in Different Geographical and Socioeconomic Settings." Frontiers in Immunology, vol. 8, June 2017. Crossref, doi:10.3389/fimmu.2017.00696.

Sales, James. "Effects of Pharmacological Concentrations of Dietary Zinc Oxide on Growth of Post-Weaning Pigs: A Meta-Analysis." Biological Trace Element Research, vol. 152, no. 3, June 2013, pp. 343-49. Crossref, doi:10.1007/s12011-013-9638-3.

Sanz Fernandez, MV, et al. "Effects of Supplemental Zinc Amino Acid Complex on Gut Integrity in Heat-Stressed Growing Pigs." Animal, vol. 8, no. 1, 2013, pp. 43-50, doi:10.1017/S1751731113001961.

Schalm, O. W., et al. "Bovine Mastitis." Lea & Febiger, 1971.

Schloss, Patrick D. "The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S RRNA Gene-Based Studies." Vol. 6, no. 7, 2010, doi:10.1371/journal.pcbi.1000844.

Schloss, Patrick D., et al. "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities." Vol. 75, no. 23, 2009, pp. 7537-41, doi:10.1128/AEM.01541-09.

Shafii, Bahman, Price, W.J., Holderman, Charlie, Minshall, Wayne, Anders, Paul. (2017). "Characterizing the Kootenai River Macroinvertebrate Community Structure Using Nonnegative Matrix Factorization." 10.15344/2456-8155/2017/120.

Shendure, Jay, and Hanlee Ji. "Next-Generation DNA Sequencing." Nature Biotechnology, vol. 26, no. 10, Oct. 2008, pp. 1135-45. Crossref, doi:10.1038/nbt1486.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308-11.

Shome, B. R., et al. "Multiplex PCR Assay for Species Identification of Bovine Mastitis Pathogens." Vol. 111, no. 6, 2011, pp. 1349-56, doi:10.1111/j.1365-2672.2011.05169.x.

Soejima, Mikiko, and Koda, Yoshiro. "Distinct Single Nucleotide Polymorphism Pattern at the FUT2 Promoter among Human Populations." Ann Hematol, vol. 87, no. 1, 2008, pp. 19-25, doi:10.1007/s00277-007-0362-y.

Soejima, Mikiko, Hao Pang, et al. "Genetic Variation of FUT2 in a Ghanaian Population: Identification of Four Novel Mutations and Inference of Balancing Selection." Ann Hematol, vol. 86, no. 3, 2007, pp. 199-204, doi:10.1007/s00277-006-0203-4.

Soejima, Mikiko, Lkhagvasuren Munkhtulga, et al. "Genetic Variation of FUT3 in Ghanaians, Caucasians, and Mongolians." Transfusion, vol. 49, no. 5, 2009, pp. 959-66, doi:10.1111/j.1537-2995.2008.02069.x.

Song, J., et al. "Effect of a Probiotic Mixture on Intestinal Microflora, Morphology, and Barrier Integrity of Broilers Subjected to Heat Stress." Poultry Science, vol. 93, no. 3, Mar. 2014, pp. 581-88. Crossref, doi:10.3382/ps.2013-03455.

Stacey, A., et al. "Lactation Is Disrupted by Alpha-Lactalbumin Deficiency and Can Be Restored by Human Alpha-Lactalbumin Gene Replacement in Mice." Proceedings of the National Academy of Sciences, vol. 92, no. 7, Mar. 1995, pp. 2835-39. Crossref, doi:10.1073/pnas.92.7.2835.

Starke, Ingo C., et al. "The Impact of High Dietary Zinc Oxide on the Development of the Intestinal Microbiota in Weaned Piglets." FEMS Microbiology Ecology, vol. 87, no. 2, Feb. 2014, pp. 416-27. Crossref, doi:10.1111/1574-6941.12233.

Stefka, A. T., et al. "Commensal Bacteria Protect against Food Allergen Sensitization." Proceedings of the National Academy of Sciences, vol. 111, no. 36, Sept. 2014, pp. 13145-50. Crossref, doi:10.1073/pnas.1412008111.

Stein, Lincoln D. "The Case for Cloud Computing in Genome Informatics." Genome Biology, vol. 11, no. 5, 2010, p. 207. Crossref, doi:10.1186/gb-2010-11-5-207.

Strucken, Eva M., et al. "Go with the Flow - Biology and Genetics of the Lactation Cycle." Front Genetics, vol. 6, 2015, p. 118, doi:10.3389/fgene.2015.00118.

Svensson, M., et al. "Conversion of α-Lactalbumin to a Protein Inducing Apoptosis." Vol. 97, no. 8, 2000, pp. 4221-26, doi:10.1073/pnas.97.8.4221.

Tajima, F. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." Genetics, vol. 123, no. 3, Nov. 1989, pp. 585-95.

Tako, E., et al. "Dietary Inulin Affects the Expression of Intestinal Enterocyte Iron Transporters, Receptors and Storage Protein and Alters the Microbiota in the Pig Intestine." British Journal of Nutrition, vol. 99, no. 03, Mar. 2008. Crossref, doi:10.1017/S0007114507825128.

Taponen, S., et al. "Real-Time Polymerase Chain Reaction-Based Identification of Bacteria in Milk Samples from Bovine Clinical Mastitis with No Growth in Conventional Culturing." Vol. 92, no. 6, 2009, pp. 2610-17, doi:10.3168/jds.2008-1729.

Teppa, Roxana E., et al. "Phylogenetic-Derived Insights into the Evolution of Sialylation in Eukaryotes: Comprehensive Analysis of Vertebrate β-Galactoside α2,3/6-Sialyltransferases (ST3Gal and ST6Gal)." Int J Mol Sci, vol. 17, no. 8, 2016, p. 1286, doi:10.3390/ijms17081286.

Thurl, Stephan, Manfred Munzert, et al. "Systematic Review of the Concentrations of Oligosaccharides in Human Milk." Nutrition Reviews, vol. 75, no. 11, Nov. 2017, pp. 920-33. Crossref, doi:10.1093/nutrit/nux044.

Triantis, Vassilis, et al. "Immunological Effects of Human Milk Oligosaccharides." Frontiers in Pediatrics, vol. 6, July 2018. Crossref, doi:10.3389/fped.2018.00190.

Trowman, Rebecca, et al. "A Systematic Review of the Effects of Calcium Supplementation on Body Weight." Vol. 95, no. 6, 2006, pp. 1033-38, doi:10.1079/BJN20051727.

Turnbaugh, Peter J., Ruth E. Ley, et al. "An Obesity-Associated Gut Microbiome with Increased Capacity for Energy Harvest." Nature, vol. 444, no. 7122, 2006, pp. 1027-31.

Turnbaugh, Peter J., Ruth E. Ley, et al. "The Human Microbiome Project." Vol. 449, no. 7164, 2007, pp. 804-10, doi:10.1038/nature06244.

Turnbaugh, Peter J., Vanessa K. Ridaura, et al. "The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice." Vol. 1, no. 6, 2009, pp. 6ra14-6ra14, doi:10.1126/scitranslmed.3000322.

Vahjen, W, et al. "Increased Dietary Zinc Oxide Changes the Bacterial Core and Enterobacterial Composition in the Ileum of Piglets." Vol. 89, no. 8, 2011, p. 24302439, doi:10.2527/jas.2010-3270.

Vahjen, W., et al. "Bar-Coded Pyrosequencing of 16S RRNA Gene Amplicons Reveals Changes in Ileal Porcine Bacterial Communities Due to High Dietary Zinc Intake." Applied and Environmental Microbiology, vol. 76, no. 19, Oct. 2010, pp. 6689-91. Crossref, doi:10.1128/AEM.03075-09.

Van Leuven, James T., et al. "Sympatric Speciation in a Bacterial Endosymbiont Results in Two Genomes with the Functionality of One." Cell, vol. 158, no. 6, Sept. 2014, pp. 1270-80. Crossref, doi:10.1016/j.cell.2014.07.047.

Venter, J. Craig, et al. "The Sequence of the Human Genome." Science, vol. 291, no. 5507, Feb. 2001, pp. 1304-51. Crossref, doi:10.1126/science.1058040.

Vingron, Martin, and Michael S. Waterman. "Sequence Alignment and Penalty Choice." Journal of Molecular Biology, vol. 235, no. 1, Jan. 1994, pp. 1-12. Crossref, doi:10.1016/S0022-2836(05)80006-3.

Walker, Alan W., et al. "High-Throughput Clone Library Analysis of the Mucosa-Associated Microbiota Reveals Dysbiosis and Differences between Inflamed and Non-Inflamed Regions of the Intestine in Inflammatory Bowel Disease." Vol. 11, no. 1, 2011, p. 7, doi:10.1186/1471-2180-11-7.

Wang, Qiong, et al. "NaÃ¯ve Bayesian Classifier for Rapid Assignment of RRNA Sequences into the New Bacterial Taxonomy." Vol. 73, no. 16, 2007, pp. 5261-67, doi:10.1128/AEM.00062-07.

Ward, Tonya L., et al. "Human Milk Metagenome: A Functional Capacity Analysis." BMC Microbiology, vol. 13, no. 1, 2013, p. 116. Crossref, doi:10.1186/1471-2180-13-116.

Wheeler, David A., et al. "The Complete Genome of an Individual by Massively Parallel DNA Sequencing." Nature, vol. 452, no. 7189, Apr. 2008, pp. 872-76. Crossref, doi:10.1038/nature06884.

WHO, Weltgesundheitsorganisation, editor. "Global Strategy for Infant and Young Child Feeding." 2003.

Wiggans, G. R., et al. "The Genomic Evaluation System in the United States: Past, Present, Future." Journal of Dairy Science, vol. 94, no. 6, June 2011, pp. 3202-11. Crossref, doi:10.3168/jds.2010-3866.

Williams, Janet E., et al. "Human Milk Microbial Community Structure Is Relatively Stable and Related to Variations in Macronutrient and Micronutrient Intakes in Healthy Lactating Women." The Journal of Nutrition, vol. 147, no. 9, 2017, pp. 1739-48. PubMed, doi:10.3945/jn.117.248864.

Williams, Janet E., et al. "Strong Multivariate Relationships Exist Among Milk, Oral, and Fecal Microbiomes in Mother-Infant Dyads During the First Six Months Postpartum." J Nutr 2018;148:1-13. https://doi.org/10.1093/jn/nxy299.

Willing, B. P., and A. G. Van Kessel. "Intestinal Microbiota Differentially Affect Brush Border Enzyme Activity and Gene Expression in the Neonatal Gnotobiotic Pig." Journal of Animal Physiology and Animal Nutrition, vol. 93, no. 5, Oct. 2009, pp. 586-95. Crossref, doi:10.1111/j.1439-0396.2008.00841.x.

Woese, C. R. "Bacterial Evolution." Microbiological Reviews, vol. 51, no. 2, June 1987, pp. 221-71.

Wu, Gary D., et al. "Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes." Vol. 334, no. 6052, 2011, pp. 105-08, doi:10.1126/science.1208344.

Yang, Hongwen. "Conserved or Lost: Molecular Evolution of the Key Gene GULO in Vertebrate Vitamin C Biosynthesis." Biochemical Genetics, vol. 51, no. 5-6, June 2013, pp. 413-25. Crossref, doi:10.1007/s10528-013-9574-0.

Yang, Hui, et al. "Uncovering the Composition of Microbial Community Structure and Metagenomics among Three Gut Locations in Pigs with Distinct Fatness." Scientific Reports, vol. 6, no. 1, July 2016. Crossref, doi:10.1038/srep27427.

Young, Wayne, et al. "Transfer of Intestinal Bacterial Components to Mammary Secretions in the Cow." Peerj, vol. 3, 2015, p. e888, doi:10.7717/peerj.888.

Yu, Jie, et al. "A Novel Lactobacilli-Based Teat Disinfectant for Improving Bacterial Communities in the Milks of Cow Teats with Subclinical Mastitis." Frontiers in Microbiology, vol. 8, Sept. 2017. Crossref, doi:10.3389/fmicb.2017.01782.

Yu, Ting, et al. "Dietary High Zinc Oxide Modulates the Microbiome of Ileum and Colon in Weaned Piglets." Frontiers in Microbiology, vol. 8, May 2017. Crossref, doi:10.3389/fmicb.2017.00825.

Zadoks, Ruth N., et al. "Molecular Epidemiology of Mastitis Pathogens of Dairy Cattle and Comparative Relevance to Humans." Journal of Mammary Gland Biology and Neoplasia, vol. 16, no. 4, Dec. 2011, pp. 357-72. Crossref, doi:10.1007/s10911-011-9236-y.

Zanardi, G., et al. "Short Communication: Comparing Real-Time PCR and Bacteriological Cultures for Streptococcus Agalactiae and Staphylococcus Aureus in Bulk-Tank Milk Samples."

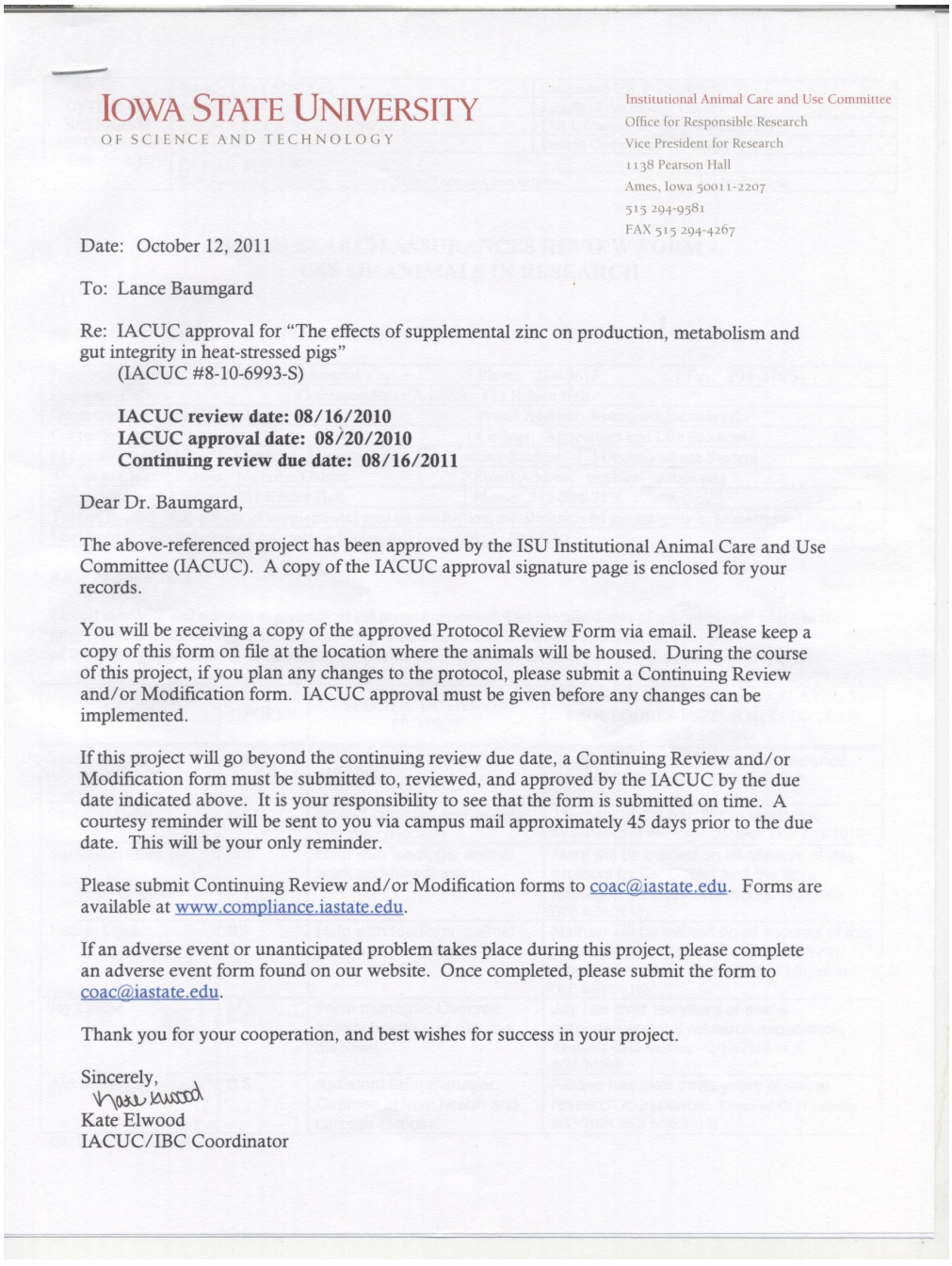
Journal of Dairy Science, vol. 97, no. 9, Sept. 2014, pp. 5592-98. Crossref, doi:10.3168/jds.2014-7947.

Zhang, Yinzhi, et al. "GLUT1 and Lactose Synthetase Are Critical Genes for Lactose Synthesis in Lactating Sows." Nutrition & Metabolism, vol. 15, no. 1, Dec. 2018. Crossref, doi:10.1186/s12986-018-0276-9.

Zhang, Zhiying, et al. "Amino Acid Profiles in Term and Preterm Human Milk through Lactation: A Systematic Review." Nutrients, vol. 5, no. 12, Nov. 2013, pp. 4800-21. Crossref, doi:10.3390/nu5124800.

Zhao, Wenjing, et al. "The Dynamic Distribution of Porcine Microbiota across Different Ages and Gastrointestinal Tract Segments." PLOS ONE, edited by Xiangzhen Li, vol. 10, no. 2, Feb. 2015, p. e0117441. Crossref, doi:10.1371/journal.pone.0117441.

Appendix A: Chapter 2 Project Approval



IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Animal Care and Use Committee
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-9581
FAX 515 294-4267

Date: October 12, 2011

To: Lance Baumgard

Re: IACUC approval for "The effects of supplemental zinc on production, metabolism and gut integrity in heat-stressed pigs"
(IACUC #8-10-6993-S)

IACUC review date: 08/16/2010
IACUC approval date: 08/20/2010
Continuing review due date: 08/16/2011

Dear Dr. Baumgard,

The above-referenced project has been approved by the ISU Institutional Animal Care and Use Committee (IACUC). A copy of the IACUC approval signature page is enclosed for your records.

You will be receiving a copy of the approved Protocol Review Form via email. Please keep a copy of this form on file at the location where the animals will be housed. During the course of this project, if you plan any changes to the protocol, please submit a Continuing Review and/or Modification form. IACUC approval must be given before any changes can be implemented.

If this project will go beyond the continuing review due date, a Continuing Review and/or Modification form must be submitted to, reviewed, and approved by the IACUC by the due date indicated above. It is your responsibility to see that the form is submitted on time. A courtesy reminder will be sent to you via campus mail approximately 45 days prior to the due date. This will be your only reminder.

Please submit Continuing Review and/or Modification forms to coac@iastate.edu. Forms are available at www.compliance.iastate.edu.

If an adverse event or unanticipated problem takes place during this project, please complete an adverse event form found on our website. Once completed, please submit the form to coac@iastate.edu.

Thank you for your cooperation, and best wishes for success in your project.

Sincerely,
Kate Elwood
Kate Elwood
IACUC/IBC Coordinator

Appendix B: Chapter 3 Project Approval

University of Idaho
Animal Care and Use Committee

Date: Friday, December 21, 2012
To: Mark McGuire
From: University of Idaho
Re: Protocol 2012-140
Characterization of Antigens for a Staphylococcal Bovine Mastitis Vaccine

Your animal care and use protocol for the project shown above was reviewed and approved by the University of Idaho on Friday, December 21, 2012.

This protocol was originally submitted for review on: Tuesday, November 27, 2012
The original approval date for this protocol is: Friday, December 21, 2012
This approval will remain in affect until: Saturday, December 21, 2013
The protocol may be continued by annual updates until: Monday, December 21, 2015

Federal laws and guidelines require that institutional animal care and use committees review ongoing projects annually. For the first two years after initial approval of the protocol you will be asked to submit an annual update form describing any changes in procedures or personnel. The committee may, at its discretion, extend approval for the project in yearly increments until the third anniversary of the original approval of the project. At that time, the protocol must be replaced by an entirely new submission.

Brad Williams, DVM
Campus Veterinarian
University of Idaho
208-885-8958

Appendix C: Chapter 4 Project Approval

University of Idaho

June 15, 2015

Office of Research Assurances
Institutional Review Board
875 Perimeter Drive, MS 3010
Moscow ID 83844-3010

Phone: 208-885-6162
Fax: 208-885-5752
irb@uidaho.edu

To: Mark McGuire

From: Jennifer Walker
Chair, University of Idaho Institutional Review Board
University Research Office
Moscow, ID 83844-3010

Title: ': INSPIRE Track 1: What is Normal Milk? Sociocultural, Evolutionary, Environmental, and Microbial Aspects of Human Milk Composition'
Project: 14-308
Approved: 06/27/15
Expires: 06/25/16

On behalf of the Institutional Review Board at the University of Idaho, I am pleased to inform you that the **first-year extension** of your proposal is approved as offering no significant risk to human subjects as no changes in protocol have been made on this project.

This extension of approval is valid until the date stated above at which time a second extension will need to be requested if you are still working on this project. If not, please advise the IRB committee when the project is completed.

Thank you for submitting your extension request.

Jennifer Walker

Jennifer Walker