CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE GENÉTICA Y BIOLOGÍA MOLECULAR

"CARACTERIZACIÓN DE LA COMUNIDAD DE ARQUEAS EN MUESTRAS BINOMIALES DE CALOSTRO Y COPRO DE NEONATO"

TESIS

Que presenta
Biol. Maricarmen Salas López

Para obtener el grado de
Maestra en Ciencias
En la especialidad de Genética y Biología Molecular

Director de Tesis:
Jaime García Mena

Ciudad de México Agosto 2023.

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE GENÉTICA Y BIOLOGÍA MOLECULAR

"CHARACTERIZATION OF THE ARCHAEAL COMMUNITY OF HUMAN COLOSTRUM AND NEONATAL STOOL IN PAIRED SAMPLES"

A Thesis

Submitted in Partial Fulfillment of The

Requirements for The Degree of *Master of Sciences*

in the field of Genetics and Molecular Biology

Presented by
Biol. Maricarmen Salas López

Thesis Director
Jaime García Mena

Mexico City August 2023.

Advisory Research Committee

Dr. José Manuel Hernández Hernández
Departamento de Genética y Biología Molecular
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Dra. Claudia Pérez Cruz
Departamento de Farmacología
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Este trabajo se realizó en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional Unidad Zacatenco, Laboratorio de Referencia y Apoyo para la Caracterización de Genomas, Transcriptomas y Microbiomas del Departamento de Genética y Biología Molecular, bajo la dirección del Dr. Jaime García Mena. Parte de este trabajo se realizó en el Hospital General José María Rodríguez, Ecatepec-de-Morelos, Estado de México.

Se agradece el apoyo del Biol. Alberto Piña Escobedo por su asistencia técnica con los reactivos, al Sr. Rodrigo García Gutiérrez por su asistencia con el material del laboratorio, a la Lic. Viridiana Rosas Ocegueda por su apoyo administrativo y a la Dra. Carmen Josefina Juárez Castelán y la Dra. Karina Corona Cervantes por su asesoramiento técnico especializado durante el trabajo experimental.

Este proyecto fue financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT-163235) INFR-2011-01, por el Fondo SEP-Cinvestav, proyecto No. 174: "Caracterización de la microbiota del calostro y leche humana por los subtipos de IgA y su relación con episodios infecciosos durante el embarazo" y CONACyT FORDECYT-PRONACES/6669/2020 "Estudio de la vía Tracto-Digestivo-Materno►Leche-Materna►Tracto-digestivo-neonato. Una aproximación Ómica".

Durante el desarrollo de este trabajo Maricarmen Salas López fue becaria de Maestría del Consejo Nacional de Ciencia y Tecnología (CONACyT) beca CVU: 1140881 durante el periodo de Septiembre del año 2021 a Agosto del año 2023.

Academic Acknowledgements

First, I would like to express my gratitude to my advisor, Dr Jaime García Mena, for allowing me to work in his laboratory and for providing me guidance throughout this project, for his patience, encouragement, and knowledge. I would like to extend my sincere thanks to the members of my advisory research committee, Dr José Manuel Hernández Hernández and Dra. Claudia Pérez Cruz for their insightful comments and feedback.

I am grateful to Biol. Alberto Piña for his technical support and mentorship. Thanks should also go to Sr. Rodrigo García Gutiérrez for his assistance with the maintenance of the laboratory and the laboratory's material. I am also deeply thankful to all the research team members: Dra. Karina Corona, Mtra. Tizziani Benítez, Dra. Carmen Juárez, Mtro. Juan Manuel Vélez and Biol. Diana Laura Rojas, thank you for your willingness to help, for your reassurance, and for being examples of professionality and companionship. To Dra. Karina Corona, for being so generous with your knowledge and expertise. To Mtra. Tizziani and Dra. Carmen Juárez, for your patience, relentless work, and counselling both academic and personal. To Mtro. Juan Manuel Vélez, for your invaluable help and advice, your unyielding support, your teachings and, most of all, for your friendship. To Biol. Diana Laura Rojas, for being an excellent peer, for your kindness and your companionship. To all members of the laboratory, thank you for your warmth and openness, your closeness and availability, for making the laboratory such a welcoming place, and for being a source of inspiration. Finally, I would like to thank Gabriela Mora for her exceptional help with all administrative processes.

Personal Acknowledgements

Foremost, I am especially grateful for my parents. Thank you for all your efforts and your unceasing support, I will always be indebted to you. To my siblings, thank you for your love, for always being there for me regardless of the distance, for keeping me open to new perspectives, and for helping me grow in so many ways. To my extended family, thank you for your words of encouragement and your affection.

To my laboratory colleagues: Tizziani, Carmen, Juan, and Diana, I am most grateful for your teachings both on academic and personal levels. Thank you for being so generous with your knowledge and abilities. I have learned so much from each one of you.

To my companions throughout this journey: Samantha, Dariel, and Cris. I am so fortunate to have met you, thank you for being a support net, for being there in the good times and the bad, for caring for me, for your friendship and, overall, for being a second home.

To my good old friends: Ana, Raúl, Izandro. Thank you for the immeasurable gift of your friendship, for listening to me notwithstanding time or distance, for your precious advice, for the laughs and the love. I will always be grateful for you guys.

To all the people I have encountered along the way from courses to social gatherings, to everyone that has had a word of comfort, that has taught me something whether directly or indirectly, thank you.

Finally, I would like to express my gratitude to everyone that has come before me and has shared their knowledge. Thank you to scientists, video tutorials, people at forums, and everyone who, with their collaborative work, helps to improve science. I would like to also extend my deep appreciation to the musicians that have accompanied me with their art and to all artists that have inspired me.

Contents

Table index

Figure index

Abbreviations

ASV	Amplicon Sequence Variant
BLAST	Basic Local Alignment Search Tool
DESeq	Differential Abundance Analysis based on the Negative Binomial Distribution
DNA	Deoxyribonucleic Acid
PCR	Polymerase Chain Reaction
QIIME	Quantitative Insights Into Microbial Ecology
RNA	Ribonucleic Acid

Abstract

Lactation is a stage of great importance for the mother's and the newborn's health. Many components exist in the milk, among which are microorganisms, such as bacteria, viruses, and archaea. Archaea is a domain of prokaryotic microorganisms closely related to eukaryotes. Little is known about this group's presence in human milk and its role in the intestine of neonates since most studies have focused on bacteria. Given this, it is relevant to study archaeal diversity for a better understanding of its role in human health and disease. In that order of ideas, we sought to identify the archaeal community of colostrum and neonatal stool samples obtained from 42 Mexican mothers and neonates. We hypothesized that the diversity of the archaeal community in the colostrum would be similar to the one found in newborn neonatal stool samples. We developed a cross-sectional descriptive study in which amplicons of the V5-V6 regions of the 16S rRNA gene were obtained from DNA isolated from colostrum and neonatal stool samples and analyzed by next-generation massive sequencing. Intra- and inter-sample analyses were carried out to characterize the Archaean diversity in each type of sample. The relative abundance results show a dominance of the genera *Methanoculleus* and *Methanosarcina* in both meconium and colostrum. Our results indicate that alpha diversity differences between colostrum and stool are significant, possibly due to the difference in biomass between samples. On the other hand, the beta diversity analysis featured that the sample types are quite similar. More robust analyses revealed the absence of differential taxa between mother and neonatal samples, which is consistent with our hypothesis. This result is confirmed by the prediction of metabolic pathways, which indicates that there are no differential metabolic pathways between colostrum and neonatal stool.

Resumen

La lactancia materna es una etapa de mucha importancia para la salud de la madre y el recién nacido. Se han identificado gran cantidad de componentes en la leche materna, dentro de los que se encuentran componentes biológicos, tales como: bacterias, virus y arqueas. Las arqueas son un dominio de microorganismos procariontes relacionados cercanamente a las eucariotas. De este grupo, poco se conoce sobre su presencia en la leche materna y su papel en la colonización del intestino de los neonatos debido a que la mayoría de los estudios se han enfocado en bacterias. Ante esto, resulta importante estudiar la diversidad de arqueas en la población mexicana con el fin de entender mejor su papel dentro del cuerpo humano. Por lo anterior, en este estudio buscamos caracterizar la comunidad de arqueas presentes en 42 muestras de calostro y copro de neonato de una población mexicana. Hipotetizamos que la diversidad de la comunidad de arqueas de las muestras de leche materna sería semejante a la de las muestras de copro de neonato. Se realizó un estudio descriptivo transversal en el cual se obtuvieron amplicones de la región V5-V6 del gen 16S rRNA de muestras de calostro y copro de neonato y se analizaron por secuenciación masiva de última generación. Posteriormente, se realizaron análisis intra e inter-muestrales para caracterizar la diversidad arqueana de cada tipo de muestra. Los resultados de abundancia relativa muestran una dominancia de los géneros *Methanoculleus* y *Methanosarcina* tanto en meconio como en calostro. Aparte, nuestros resultados indican que las diferencias de diversidad alfa entre el calostro y el copro de neonato son significativas, posiblemente debido a la diferencia en la biomasa entre las muestras. Por otra parte, el análisis de diversidad beta mostró que los tipos de muestra son bastante similares. Análisis más robustos revelan la ausencia de taxa diferenciales entre ambos grupos, lo cual es consistente con nuestra hipótesis. Este resultado se confirma con la predicción de las rutas metabólicas, que indica que no existen rutas metabólicas diferenciales entre calostro y copro de neonato.

I. Introduction

1.1 Importance of breastfeeding

Breastfeeding plays a crucial role in the development and growth of newborns. Evidence suggests that an exclusive breastmilk diet for the first six months of life can reduce the risk of various conditions such as asthma, gastrointestinal infections, and obesity (Dogaru et al., 2014; Frank et al., 2019; Marseglia et al., 2015). Additionally, research has shown that children who are breastfed for longer periods have higher intelligence and lower morbidity and mortality rates than those who are not breastfed or breastfed for shorter periods (Victora et al., 2016). Lengthy lactation periods are associated with a 26% reduction in the likelihood of developing overweight or obesity (Horta et al., 2015). Moreover, breastfeeding offers numerous benefits for mothers, including a lower risk of breast and ovarian cancer, type 2 diabetes, metabolic syndrome, and obesity, among others (Binns et al., 2016; Chowdhury et al., 2015; Sattari et al., 2019).

Human milk is the ideal food for infants (Victora et al., 2016). It contains multiple nutrients and bioactive constituents, including proteins, carbohydrates, cytokines, oligosaccharides, fatty acids, immune factors, and microbiota (Gila-Diaz et al., 2019; Le Doare et al., 2018; Pace et al., 2021). Further, its composition changes over time to meet the evolving nutritional needs of the growing infant (Selma-Royo et al., 2022).

The first milk produced during the first few days after birth, typically around 2-4 days is called colostrum (Thapa, 2005). The exact duration varies among women and depends on factors such as delivery mode and parity (Dewey et al., 2003). Colostrum is a thick and yellowish fluid rich in protein, minerals, vitamins, growth factors and immunoglobulins, essential for providing passive immunity to newborns and protecting them against infections until their immune system develops (Uruakpa et al., 2002). Additionally, colostrum's growth factors help promote the maturation and growth of the infant's gastrointestinal tract (Ballard & Morrow, 2013). Then, colostrum's composition gradually changes giving place to transitional milk, which has a higher fat and lactose content (Yates, 2010). This usually occurs from 5 days to two weeks postpartum (Ballard & Morrow, 2013). From that moment on, milk is considered mature, and it is characterized by a high concentration of carbohydrates (Ballard & Morrow, 2013; Jenness, 1979).

1.2 Human milk microbiota

The human microbiota is the aggregate of microorganisms (bacteria, archaea, viruses, fungi) that inhabit the human body (Jandhyala et al., 2015; Sekirov et al., 2010). These are found in various parts, such as the skin, nose, mouth, lungs, vagina, urinary tract, and gastrointestinal tract (Dekaboruah et al., 2020). As for human milk, it is known that it has a bacterial concentration of

approximately 1 X 10^3 CFUs/mL and it is estimated that in each breastfeeding the newborn ingests 1 X 10^5 bacteria (Notarbartolo et al., 2022; Ogra et al., 2020). The bacterial community of human milk has already been characterized. Currently, the main phyla and genera that make up the human milk microbiota are known. The genera that constitute the central bacteriome are *Staphylococcus*, *Streptococcus*, *Serratia*, *Pseudomonas*, *Corynebacterium*, *Ralstonia*, *Propionibacterium*, *Sphingomonas* and *Bradyrhizobium* (Notarbartolo et al., 2022). Few is known about the presence of other members of the microbiota in human milk.

1.3 Entero-mammary pathway

The origin of the bacteria in human milk is mainly attributed to two sources: the entero-mammary route and retrograde flow (Fernández et al., 2013; Moossavi & Azad, 2020; Notarbartolo et al., 2022). The first consists of immune cell-mediated bacterial translocation from the maternal gastrointestinal tract to the mammary gland (Moossavi & Azad, 2020; Ogra et al., 2020). The second refers to external contamination or the transfer from the mouth of the newborn or the mother to the breast tissue (Moossavi & Azad, 2020).

1.4 What are archaea?

Archaea were defined as a new domain of life in 1977 by Carl Woese and George Fox after their phylogenetic analysis based on the 16S ribosomal RNA subunit (Cavicchioli, 2011; Thursby & Juge, 2017). These microorganisms are distinguished from the Bacteria and Eukarya domains by multiple characteristics. One of them is the cell membrane. Archaeal membranes are made up of isoprenoids and ether-linked glycerol-1-phosphate, while those of bacteria and eukaryotes consist of straight-chain fatty acids with ester-linked glycerol-3-phosphate (Cavicchioli, 2011; Eme & Doolittle, 2015; Kandler & König, 1998). In addition, archaea lack murein, a cell wall compound present in most bacterial taxa (Kandler & König, 1998). Instead, their cell walls are made of a pseudomurein layer and a second heteropolysaccharide layer. Due to pseudomurein, archaea are resistant to antibiotics that attack the cell wall, such as penicillin or vancomycin (Hilpert et al., 1981). Most archaea are gram-negative. Their cell envelopes are protein, glycoprotein (S-layers) or in some cases, a reinforcement of the cytoplasmic membrane similar to the glycocalyx of eukaryotic cells (Kandler & König, 1998). The S layers of the archaea have a stabilizing function since they are the only envelope outside the cytoplasm (Kandler & Konig, 1993).

Another relevant characteristic of archaea is their genomic architecture. On the one hand, they are similar to bacteria, their genomes being circular with a size of around 0.50 to 5.75 Mpb (DasSarma et al., 2009). However, on the other hand, their replication and transcription mechanisms are similar to those of eukaryotes (Cavicchioli, 2011; Offre et al., 2013). For example, they have replication and repair proteins homologous to eukaryotes, as well as one of their DNA polymerases (DasSarma et al., 2009).

Archaea have heterogeneous metabolisms, being able to metabolize a wide variety of organic compounds and even fix carbon from inorganic compounds. They are the only ones capable of carrying out the metabolic pathways of methanogenesis and the anaerobic oxidation of methane (carried out through the syntrophic association with reducing sulfate bacteria) (Eme & Doolittle, 2015; Offre et al., 2013). In addition, they also contribute significantly to the nitrogen cycle through the anaerobic oxidation of ammonia, being the main nitrifiers worldwide (Eme & Doolittle, 2015; Offre et al., 2013). In this way, archaea play a key role in biogeochemical cycles and greenhouse gas emissions.

The Archaea domain consists of five phyla: Crenarchaeota, Euryarchaeota, Korarchaeota, Nanoarchaeota and Thaumarchaeota (Horz, 2015). However, thanks to information from whole genome sequencing, Thaumarchaeota, Crenarchaeota, Konarchaeota and a recently discovered phylum: Aigarchaeota, have been grouped into a superphylum called TACK (by their initials) (Horz, 2015). Recent phylogenetic studies suggest that eukaryotes branched from this group (Eme et al., 2017).

They are ubiquitous. Archaea inhabit a wide variety of ecosystems, from extreme habitats such as hot springs and hydrothermal vents, to soils, oceans, and the intestines of animals, including humans (Chaban et al., 2006). These are part of the human intestinal microbiota.

II. Background

Methanogenic archaea are the most abundant in the digestive tract of adults, specifically sp. *Methanobrevibacter smithii*. This specie has a prevalence of 95.7% in adult faeces, followed by *Methanosphaera stadmanae* with a prevalence of 29.4% likewise, *Methanomassiliicoccus luminyensis* has been detected with a prevalence of 4% (Dridi et al., 2009). These archaea are responsible for producing methane through the absorption of H_2 and CO_2, which are waste products of the fermentation of polysaccharides by bacteria (van de Pol et al., 2017). However, they have different methanogenesis metabolism. *M. smithii* uses H_2 to reduce CO_2, while *M. stadtmanae* reduces methanol (CH_3OH) (Gaci et al., 2014).

In addition, halophilic archaea have been found in stool samples using PCR fingerprinting in a population of Koreans. These sequences were associated with the species: *Halorubrum koreense, Halorubrum alimentarium* and *Halococcus morrhuae*. The presence of Euryarcheota has been related to dietary habits (Nam et al., 2008).

The presence of methanogens is dependent on age. It has been observed that colonization by these occurs gradually in the first years of life and reaches its maximum prevalence in adulthood (Grine et al., 2017; Palmer et al., 2007). *M. smithii* is an early colonizer, establishing in the gastric mucosa just after birth, with the possible source of colonization being the maternal intestinal microbiota (Grine et al., 2017). In this regard, it has been seen that there is a positive relationship between breastfeeding and the presence of *M. smithii* in the baby's gastric juice (Grine et al., 2017). This is consistent with the fact that *M. smithii* sequences have been found in human milk samples (Grine et al., 2017; Togo et al., 2019).

On the other hand, according to a recent study (Van de Pol et al. 2017), the consumption of yoghurt and milk is positively associated with the presence of *M. smithiii*. They found that dairy products are a possible source of colonization by this methanogen. Moreover, a higher abundance of *M. smithii* was found in raw milk than in processed milk, while for *M. stadtmanae* no detectable levels were found in any of the samples.

The relationship of methanogenic archaea with diseases is contradictory. On the one hand, it is suggested that *M. smithii* promotes the absorption of calories from the diet. In an animal model, it was observed that it influences the metabolism of the saccharolytic bacterium *Bacteroides thetaiotaomicron* and that this interaction generates an increase in energy production together with lipogenesis and fat accumulation in the host. On the other hand, the presence of *M.*

smithii has been associated to healthy humans and is decreased in obese populations (Togo et al., 2019). Apart, in severe acute malnutrition, there is an absence of *M. smithii*. For this reason, this species is proposed as a probiotic to restore intestinal symbiosis (Camara et al., 2021).

III. Justification

Most studies on gut microbiota have focused on bacteria without placing much relevance on the Archaea domain. This is partly because for several decades the knowledge of this group was restricted to cultivable organisms (Horz, 2015). For this reason, there is a lack of knowledge about the diversity of archaea in humans and its possible health implications. In the case of archaeal presence in human milk, publications are scarce. The characterization of archaea in samples of colostrum and neonatal stool in Mexican mothers and neonates would constitute an innovative work and a first step in evaluating the relationship of this group of organisms with health factors.

IV. Scope

This study sought to characterize the community of archaea present in 42 paired samples of colostrum and neonatal stool from patients of the General Hospital "Dr José María Rodríguez" located in Ecatepec de Morelos, State of Mexico. The latter was done through next-generation sequencing of the V5-V6 regions of the 16S rRNA gene.

V. Hypothesis

The diversity of the archaea community of colostrum samples is similar to that of neonatal stool samples.

VI. Aims

General aim

To characterize the archaeal community in paired samples of colostrum and neonatal stool obtained from patients at the General Hospital "Dr José María Rodríguez".

Specific aims

Specific aim 1. To characterize and compare the diversity of the archaea found in colostrum and neonatal stool.

Activity 1. DNA extraction from colostrum and neonatal stool samples.

Activity 1. Amplification of the V5 and V6 regions of the 16S rRNA gene.

Activity 3. Construction of libraries and high throughput DNA sequencing.

Activity 4. Quality control and taxonomic classification.

Activity 5. Alfa and beta diversity analysis.

Activity 6. Recognition of differential taxa by differential analysis.

Specific aim 2. To evaluate the possible source of neonatal stool microbiota.

Activity 1. To perform SourceTracker 2 Analysis

Specific aim 3. To determine the prediction metagenome.

Activity 1. To perform Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) analysis to predict functional metagenomes.

VII. Methodology

A cross-sectional descriptive study was carried out, which sought to characterize the archaeal community in binomial samples of colostrum and neonatal stool collected in Ecatepec de Morelos.

7.1 Experimental design

The cross-sectional descriptive study consisted of 42 mother-neonatal pairs of patients from the "Dr José María Rodríguez" General Hospital, located in Ecatepec de Morelos, State of Mexico (19°36′35″ N, 99°3′36"W). The samples were from healthy lactating women and exclusively breastfed newborns. Colostrum and neonatal stool samples were collected from each mother-infant pair 1 to 6 days after delivery. They were collected in a period from November 2017 to January 2018.

The inclusion criteria were as follows:

(1) Mexican origin with at least two generations of ancestry.

(2) gestational age between 37 and 41 weeks.

(3) vaginal delivery or non-elective cesarean section.

(4) birth weight between 2500g - 4500g.

(5) Apgar score greater than 7 at 5 min after birth.

Exclusion criteria:

(1) Probiotic and alcohol consumption.

(2) smoking

(3) diabetes, overweight and obesity before or during pregnancy.

(4) antibiotic use during the last trimester of pregnancy and before sampling.

The participants were given a survey where sociodemographic and clinical data was recorded (maternal age, gestational age at delivery, mode of delivery, sex, and age of the newborn). Written informed consent was obtained from all participants before the study, following the 2013 Declaration of Helsinki. The protocol was approved by the Ethics Committee of the General Hospital "Dr José María Rodríguez".

7.2 Sample collection

The samples were taken by a member of the research team wearing sterile gloves. The colostrum-neonatal stool sample pairs were collected on the same day. Colostrum was collected manually into a sterile polypropylene tube (about 5 to 10 mL). Breast sanitation was not employed to better represent the microorganisms. The neonatal stool was recovered from diapers into sterile containers by using sterile tongue depressors. The samples were promptly sent to the laboratory in a cold environment. They were distributed in aliquots of 1 mL (colostrum) or 200 mg (neonatal stool). They were stored at -20° C.

7.3 DNA extraction

First, 1 mL of milk was centrifuged at 10,000 g and 4° C for 15 min in a refrigerated centrifuge (Eppendorf 5415R) and the fat was removed with a roll of sterile dental cotton. The aqueous supernatant was removed by decantation, the sediment was resuspended in 1 ml of sterile PBS pH 7.4 and then recentrifuged at 10,000 g for 15 min. The sediment obtained was resuspended in 300 µL of PBS pH 7.4 and processed for DNA extraction using the FavorPrep Milk bacterial DNA extraction kit (Cat: FAMBD001, Favorgen, Biotech Corp, Taiwan) following the manufacturer's instructions. Faecal DNA was extracted from 200 mg stool samples using a QIAamp DNA Stool Mini Kit (Cat.: 12830-50, Qiagen, The Netherlands), following the manufacturer's instructions. In the two cases, 300 µL of PBS pH 7.4 was used as a negative control for DNA extraction. The DNA concentration in samples was measured at 260/280 absorbance using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA), no absorbance was detected for negative controls. DNA integrity was assessed by electrophoretic fractionation in 0.5% agarose gel stained with 0.80 µL of Midori Green advanced dye (1:15) using buffer TBE (Tris-Borato-EDTA). To charge, 1 µL of electrophoretic colour marker (bromophenol blue, 2X glycerol) with 5 µL of DNA were added, they ran for 50 min, applying 90 Volts. DNA was visualized with the MolecularImager® Gel DocTM XR System program (Bio-Rad Laboratories, Chicago, IL, USA).

7.4 Sample selection

For the present study, 42 binomial samples (colostrum and neonatal stool) were selected from the Laboratorio de Referencia y Apoyo para la Caracterización de Genomas, Transcriptomas y Microbiomas del Departamento de Genética y Biología Molecular of Cinvestav, Zacatenco based on the quality of spectrophotometry and the amplification by PCR. The samples were previously

obtained from the General Hospital "Dr José María Rodríguez" in Ecatepec de Morelos, State of Mexico (Corona-Cervantes et al., 2020).

7.5 Amplification of the V5-V6 region of the 16S rDNA.

The forward Arc787F (5′-ATT-AGA-TAC-CCG-BGT-AGT-CC-3′) and reverse primer Arc1059R (5′-GCC-ATG-CAC-CWC-CTC-T -3′) were used for the polymerase chain reactions (PCR) (Obtained from Yu et al., 2005). These had already been used by the working group to amplify the V5 and V6 regions of the archaeal ribosomal 16S gene (Gómez-Gallego et al., 2016).

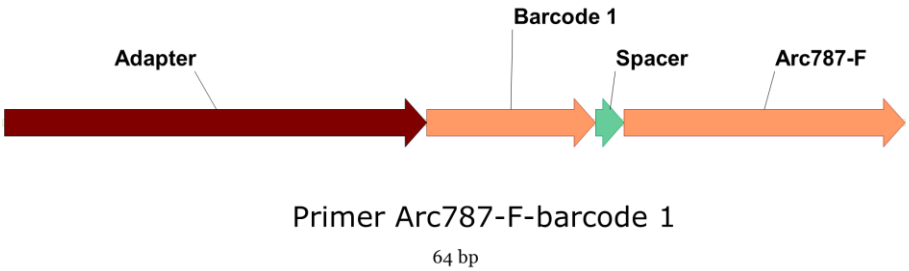

Figure 1. Structure of Forward Primer Arc787

In Figure 1, the forward primer structure can be seen, with the adapters needed for ion semiconductor sequencing and the identifier sequence, the barcode. The reverse primer (Figure 2) has the B and P1 adapters necessary for sequencing.

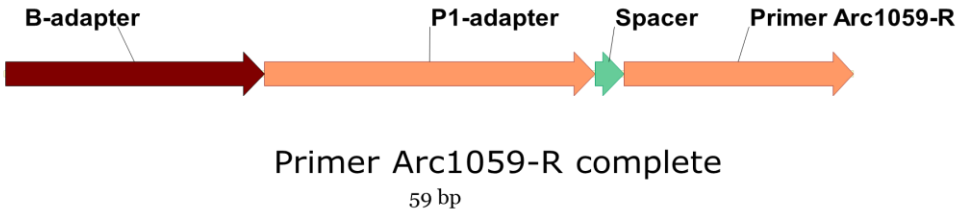

Figure 2. Structure of Reverse Primer Arc1059

The reactions were performed with the Phusion High-Fidelity PCR Kit (Cat F-530S), ThermoFisher Scientific, Waltham, MA, USA) following previously reported conditions (Gállego Bravo et al., 2019). The procedure was carried out in a PCR workstation, under aseptic conditions. The reaction mixture consisted of 4.0 uL of HF buffer (1X), 0.4 uL of dNTPs (200 uM), 0.2 uL of Phusion polymerase (0.02 U/uL), 1 uL of each primer (10 uM) and 0.2 uL of $MgCl_2$ (0.5 mM). The DNA template volume was adjusted to 13.2 uL with water for a final concentration of 4.0 ng in 20.0 uL. The reactions were programmed in 2729 Thermo Cycler (Applied Biosystems™, ThermoFisher Scientific, Waltham, MA, USA) with a 5 min hot start at 95° C, then initial denaturation at 95° C cycle for 5 min, 25 cycles of denaturation at 94° C for 15s, annealing at 56

° C for 15s, extension at 72° C for 15s and a final extension at 72° C for 7 min. Subsequently, the correct size of the amplicons (358 bp, Figure 3) was observed using a 1.5% agarose gel dyed with Midori Green (Nippon Genetics®, Dueren, Germany) in 0.5X TBE, with a weighted marker of 100 bp. Electrophoresis lasted 45 min at 80 V. The gel reading was performed with the Molecular Imager® Gel DocTM XR System program (Bio-Rad Laboratories, Chicago, IL, USA).

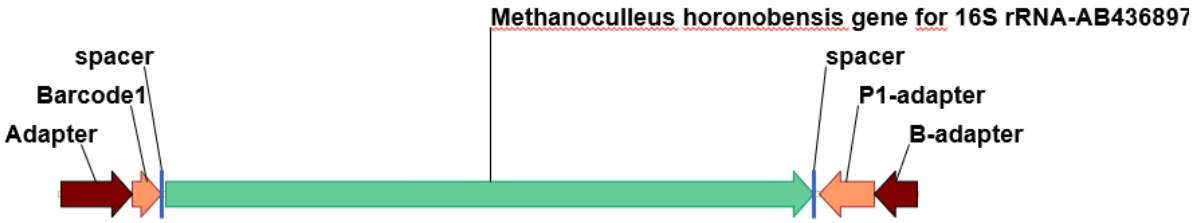

Figure 3. PCR product of *Methanoculleus horonobensis*, the amplicon has 358bp.

7.6 Library construction and Ion Torrent sequencing

The library preparation consisted in mixing the barcode samples, purifying them, and quantifying the DNA concentration. Then, the DNA quality was assessed with the 2100 Bioanalyzer equipment, verifying the amplicon's sizes. Subsequently, massive PCR reactions were carried out and the DNA's quality was evaluated with the Ion One Touch system and Qubit 2.0 fluorometer respectively. For the sequencing initialization, the Ion Torrent PGM was cleaned, the enzyme and primers in 5' were added and, finally, the sample mixture was loaded into the sequencing chip. The sequencing process lasts around 7h. After, the sequence data was obtained as well as the FASTQ files for the subsequent analysis.

7.7 Sequencing data analysis

The reading counts were processed with the Torrent Suite Software v4.0.2, assigning the reads to the samples with the barcodes. In addition, the low quality (quality score ≤ 20) and polyclonal sequences (homopolymers > 6) were filtered, and the adapters were trimmed. After, they were exported as FASTQ files. Then, they were further processed and analyzed with QIIME 2022.2. With this software, the ASVs were determined as well as the taxonomic assignations. The first was performed with the QIIME dada2 denoise-single plugin, with trimming at 238 nt. The latter was done with the feature-classifier classify-consensus-blast plugin, with a 97% percentage identity. The Silva 138 database was used for analysis with BLAST.

7.8 Archaeal Relative Abundance, Diversity, SourceTracker 2 and PICRUSt2 analyses

R 4.2.0 (R Core Team, 2021) in Rstudio (RStudio Team, 2022) was used for the relative abundance, diversity and DESeq2 analyses. For that, the following packages were used: to import the qiime artefacts, qiime2R (Bisanz, 2018), for alpha and beta diversity analyses, phyloseq 1.4.0 (McMurdie & Holmes, 2013), for the DESEQ2 analysis, DESEq2 1.3.6 (Love et al., 2023), for the heatmap elaboration, ComplexHeatmap 2.12.0 (Gu, 2022), for the ANOSIM, vegan 2.6-2 (Oksanen et al., 2022), and for figures, tidyverse 1.3.1 (Wickham et al., 2019), dplyr 1.09 (data frame manipulation), ggplot2 3.3.6 , scales 1.2.0, ggpubr 0.4.4 and, gridExtra 2.3 (Auguie, 2017). The SourceTracker analysis was carried out with the SourceTracker 2 giibs plugin (Knights et al., 2011). PICRUSt2 was executed following Douglas et al., 2020 pipeline tutorial.

7.9 Statistical Analyses

Archaeal diversity was estimated with alpha diversity (within samples) and beta diversity (between samples). The first one was calculated with the following indexes: Observed species, Shannon, Simpson, Inverse Simpson, and Fisher. The Shapiro-Wilk preliminary test was applied to know if the data followed a normal distribution. Then, U Mann-Whitney (for not parametrical data) or t-student (for parametrical data, only the Inverse Simpson) tests were applied. The differences in beta diversity were evaluated with an ANOSIM (Analyses of similarities). A differential abundance analysis (DESeq2) was performed to identify relevant taxa in the distinct sample types and was evaluated with a Wald test.

VIII. Results

8.1 Sample selection and characteristics of the sample studied.

Forty-two DNA-extracted samples of colostrum and neonatal stool were selected from the bank of the Laboratory of Reference and Support for the Characterization of Genomes, Transcriptomes and Microbiomes of the Department of Genetics and Molecular Biology of Cinvestav, Zacatenco. The DNA was assessed with spectrophotometry and with PCR amplification.

Colostrum and neonatal stool samples were collected on the first days after birth (1.53 ± 1.23). In Table 1 we can observe the sociodemographic and clinical data of the participants, most of them come from Estado de México (73.8%) and Ciudad de México (14.28%). The principal occupation of this cohort is being a housewife (95.23%) and almost half of them had superior education, meaning they reached high school (47.61%). The anthropometric (Table 2) and clinical data shows that most of them had normal weight (24.27 ± 4.21) and more than half of the deliveries were vaginal (69.04%). Table 3 shows the anthropometric data of the neonates, where 59.52% corresponded to males and 35.71% to females (for two of the samples we had no information).

Table 1. Sociodemographic and clinical characteristics of the mothers (N = 42).

	n	*%*
Origin		
Estado de México	31	73.8
Ciudad de México	6	14.28
Michoacán	1	2.38
Puebla	1	2.38
Veracruz	1	2.38
Unknown	2	4.76
Occupation		
Housewife	40	95.23
Student	1	2.38
Seller	1	2.38
Education		
Primary	10	23.8
Secondary	10	23.8
Superior	20	47.61
University	2	4.76
Gestational Age (weeks)		
< 37	3	7.14
37-40	36	85.71

> 40	2	4.76
Parity		
Primiparous	20	47.61
Multiparous	22	52.38
Mode of delivery		
Vaginal delivery	29	69.04
Emergency cesarean section	8	19.04
Elective cesarean section	5	11.9

Table 2. Anthropometric data of the mothers.

	Mean ± SD	*n*
Age (years)	22.52 ± 6.6	42
Weight (kg)	59.75 ± 10.64	33
Height (cm)	1.57 ± 0.05	33
IMC	24.27 ± 4.21	30

Table 3. Anthropometric data of the neonates.

Sex	
Female	59.52%
Male	35.71%
Size at birth (cm)	49.51 ± 1.53
Weight (g)	
<2500	7.14%
2500-4500	88.09%

8.2 Experimental assessment of archaeal primers

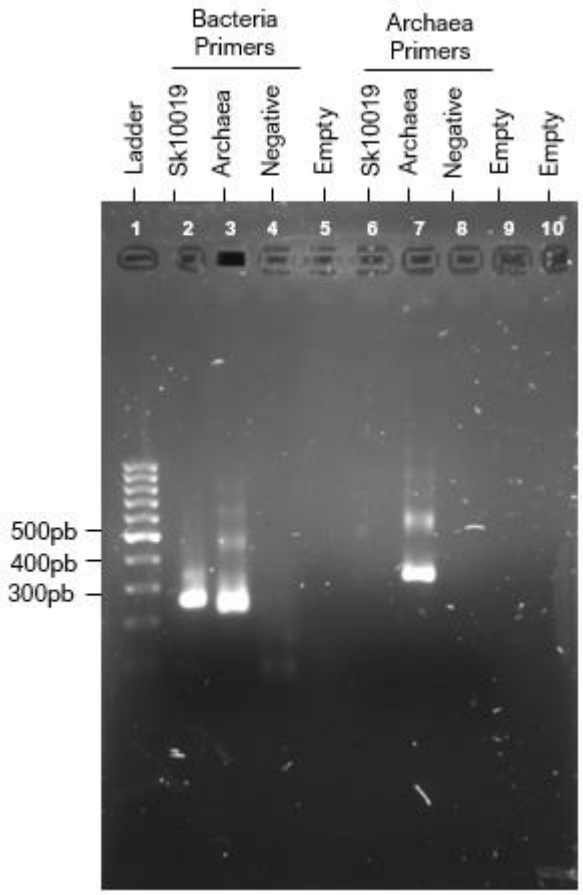

Figure 4. 1.5% agarose gel stained with Midori Green for PCR products of the V3 region (columns 2-4, they have CGO-465 bacterial primers) and V5-V6 region (columns 6-8, amplicon of 358 bp, they have archaeal primers) of the 16S ribosomal gene. (1: 100 pb ladder, 2: SK10019 *E.coli* strain, 3: Archaea culture, 4: Negative control; 6: SK10019 *E.coli* strain, 7: Archaea culture, 8: Negative control).

In **Figure 4**, we can see that the bacterial primers (CGO-465 Forward/Reverse) amplified for the *E. coli* sample (SK10019 strain) and the archaea sample. The latter is explained because the sample of archaea consisted of a reactor sample that favours the growth of archaea, however, that does not imply that there are no bacteria in the mixture. On the other hand, archaea primers only amplified the archaeal culture, generating a ~358bp amplicon. This result assured us that the archaeal primers were suitable to perform the PCRs.

8.3 Construction of Libraries for Ion Torrent Sequencing

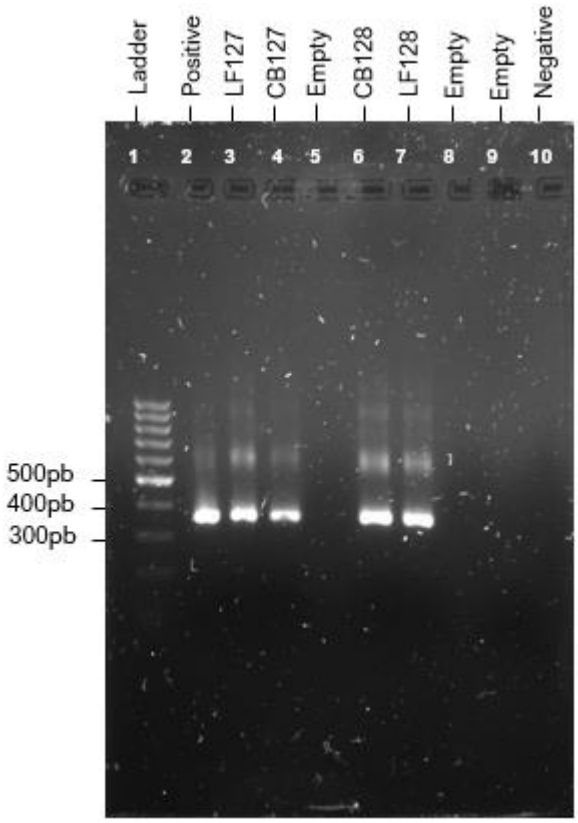

Figure 5. 1.5% agarose gel stained with Midori Green for V5-V6 PCR products of the 16S ribosomal gene (358 bp) in colostrum and neonatal stool (1: 100 pb ladder, 2: Positive control, 3: Colostrum sample LF127, 4: Neonatal stool sample CB127, 5: Empty, 6: Neonatal stool sample CB128, 7: Colostrum sample LF128, 8: Empty, 9: Empty, 10: Negative control).

In Figure 5, we can observe the PCR products of colostrum and neonatal stool samples (around 358 bp) that were obtained for the preparation of the DNA libraries.

8.4 Sequencing Results

Table 4. Sequencing summary.

Parameter	Colostrum (n=42)	Neonatal stool (n=42)
Number of reads	1’220,403	1’282,047
Mean of reads	29,057	30,524
Standard deviation	16,604	19,791
Standard error mean	2,562	3,053
Median of reads	24,871	24,702
Lowest sample reads	6,603	2,171
Highest sample reads	89,717	89,598

8.5 Results of assignations with BLAST

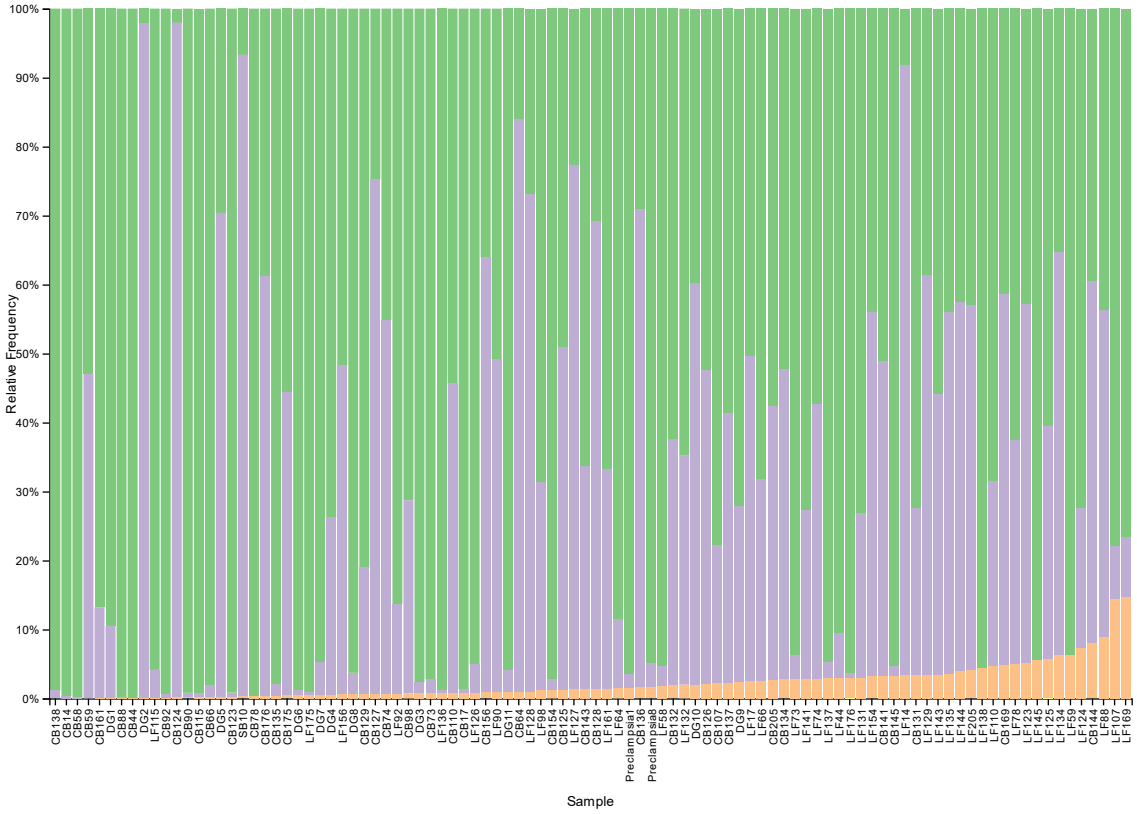

Figure 6. Barplot obtained from qiime2view showing the assignations for each of the samples using a BLAST classifier (green: bacteria assignations, purple: archaea assignations, orange: unknown assignations).

In Table 4, the Ion Torrent sequencing results are summarized. The number of reads, mean of reads and medians are comparable between the sample types. The results of taxonomic assignation at the domain level for each sample after performing BLAST with the qiime pipeline are shown in Figure 6. A comparison between the colostrum, neonatal stool and the positive control assignations can be seen in Appendices, Figure

8.6 Microbiota composition of colostrum and neonatal stool based on relative abundance.

The relative abundance of microbiota was determined at the domain, phylum, and genus levels in the two sample types.

In the relative domain abundance, Bacteria is the most predominant both in colostrum and neonatal stool (66% and 68%, respectively). Meanwhile, Archaea represent around 30% of the composition of the neonatal stool and colostrum samples. The rest of the composition was unassigned with a negligible quantity of Eukaryota reads (0.001%) in colostrum (Figure 7, Table 5).

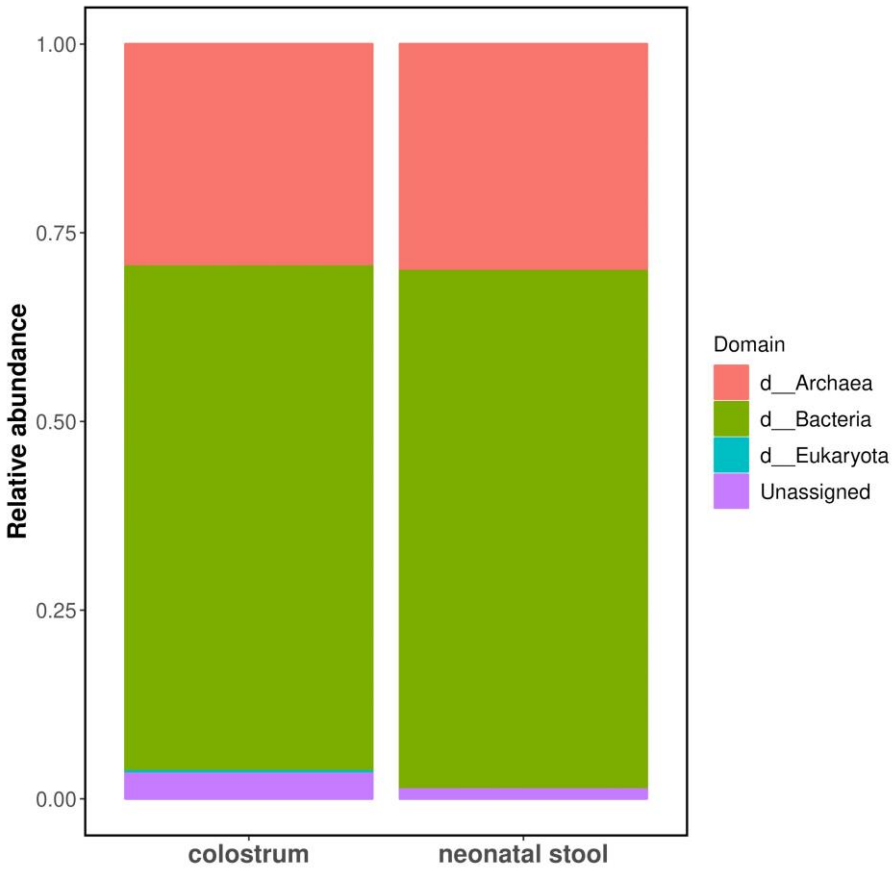

Figure 7. Relative Domain abundance for each sample type.

Tabla 5. Relative Domain abundance for each sample type.

Sample type	*Domain*	*Abundance*
colostrum	Unassigned	3.646
	Archaea	29.459
	Bacteria	66.894
	Eukaryota	0.001
neonatal stool	Unassigned	1.250
	Archaea	30.059
	Bacteria	68.691
	Eukaryota	0.000

After filtering the Bacteria and Eukarya domains from the samples, we observed the distribution of Archaea and unassigned classifications between the sample types. We can see that for colostrum 75.25% correspond to Archaea, whereas 24.74% were unassigned. As for neonatal stool, 85.45% are from Archaea and the lasting 14.54% were unassigned (Figure 8, Appendices: Table 2).

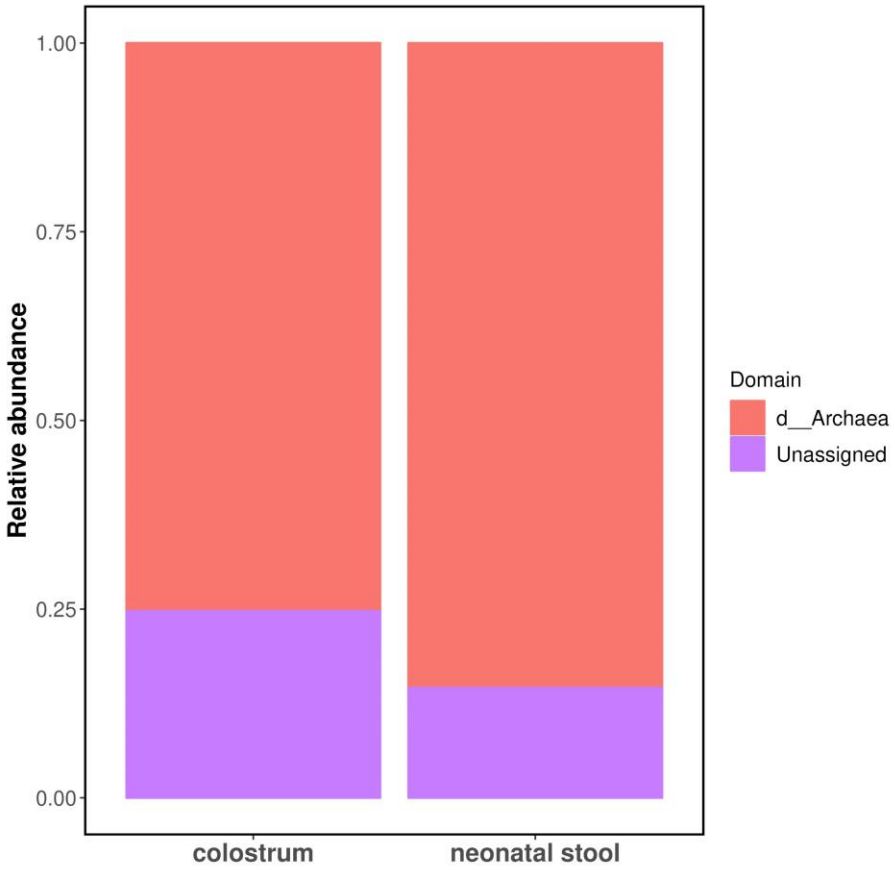

Figure 8. Relative Domain abundance filtered (only Archaeal and Unassigned classifications).

In the relative phylum abundance, four phyla were observed, all of which had similar relative abundances in colostrum and neonatal stool. The phyla were: Crenarchaeota, Euryarchaeota, Halobacterota and Thermoplasmota. The most abundant phylum was Halobacterota with 66.13% in colostrum and 75.36% in neonatal stool. Following was Euryarchaeota with 8.84% in colostrum and 8.82% in neonatal stool (Figure 9, Table 3).

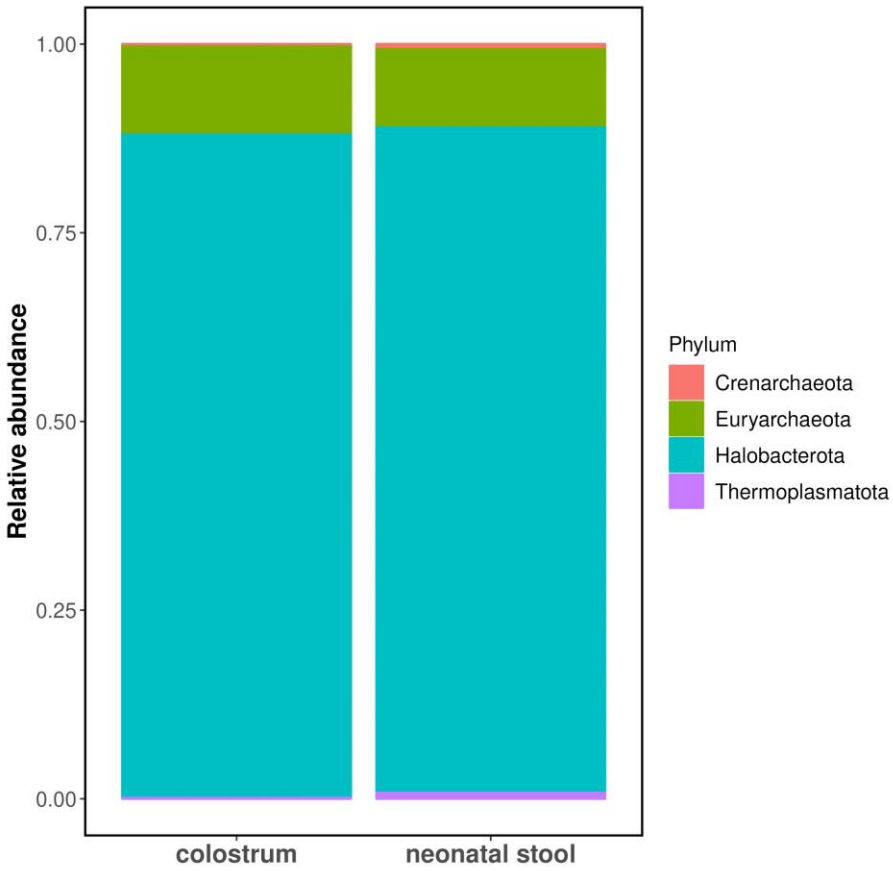

Figure 9. Relative abundance of phyla for each sample type.

Table 6. Relative abundance of phyla for each sample type.

Sample type	*Phylum*	*Abundance*
colostrum	Crenarchaeota	0.18
neonatal stool		0.60
colostrum	Euryarchaeota	8.84
neonatal stool		8.83
colostrum	Halobacterota	66.13
neonatal stool		75.36
colostrum	Thermoplasmatota	0.10
neonatal stool		0.66

As for the relative genus abundance, in both colostrum and neonatal stool, the main genus was *Methanoculleus* (38.44% colostrum, 45.75% neonatal stool), followed by *Methanosarcina* (27.58% colostrum, 29.45% neonatal stool) and then *Methanothermobacter*. The genus *Methanobrevibacter* was found in both sample types (3.36% colostrum, 2.15% neonatal stool), while *Methanosphaera* was only found in neonatal stool (2.19%) (Figure 10, Table 7).

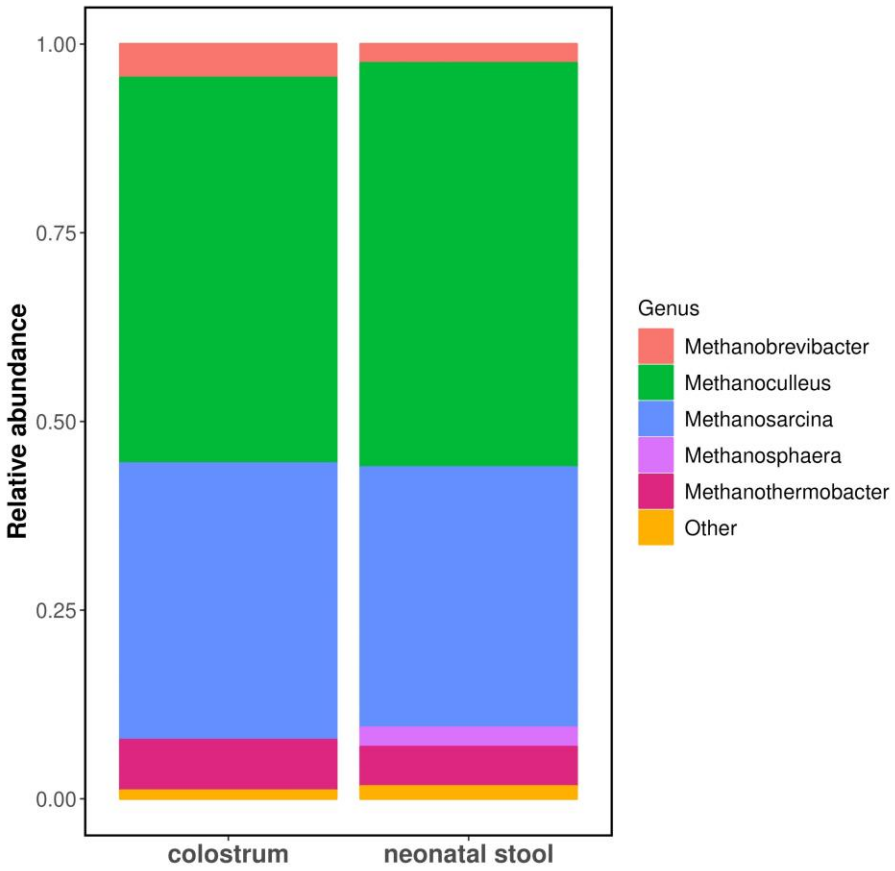

Figure 10. Relative abundance of genus for each sample.

Table 7. Relative abundance of genus for each sample.

Sample type	*Genus*	*Abundance*
colostrum	Methanobrevibacter	3.36
neonatal stool		2.15
colostrum	Methanoculleus	38.45
neonatal stool		45.76
colostrum	Methanosarcina	27.58
neonatal stool		29.46
neonatal stool	Methanosphaera	2.19
colostrum	Methanothermobacter	5.06
neonatal stool		4.46
colostrum	Other	0.05
neonatal stool		0.10

8.7 Visualization of the archaeal composition of colostrum and neonatal stool with the core microbiome

To further compare the archaeal communities of colostrum and neonatal stool, we graphed the core microbiome heatmap. Our criteria were that the genera had a 10% of prevalence and detection of 1%, meaning that the genera had to be present in at least 10% of samples with an abundance equal or superior to 1%. In panel A, the scale indicates the distance from the mean according to the max-min distance (mean normalization). In panel B, the scale indicates the percentages of relative abundance. As we can see in Figure 11, in both scales, the main archaea are members of the Euryarchaetoa, Halobacterota and Thermoplasmatota phyla. Moreover, the main archaea have almost identical core microbiome abundances. This goes to show that there is not much difference in the core archaeal community of colostrum and neonatal stool.

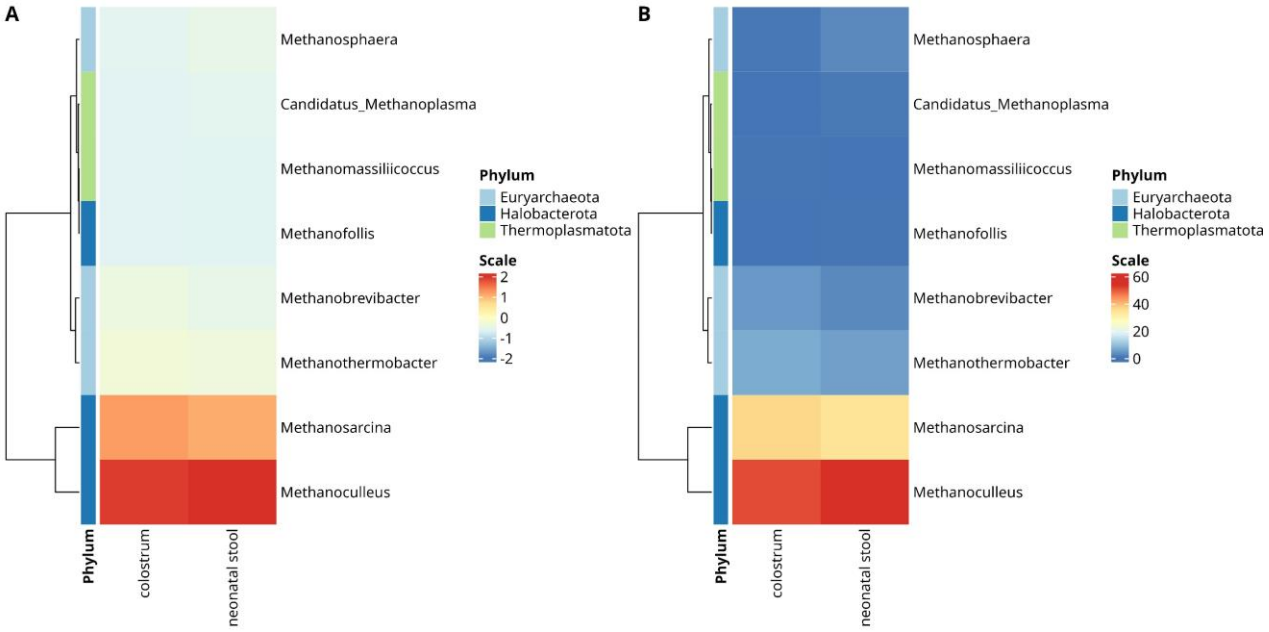

Figure 11. Core archaeal microbiota heatmap of colostrum and neonatal stool samples. A) Scale based on column, B) Scale based on relative abundance percentages.

8.8 Alpha diversity of Archaea in colostrum and neonatal stool

The alpha diversity analysis was carried out to characterize the number of species, richness, and homogeneity within samples and to compare the sample types (Figure 12, Table 8). The observed species index measures the richness of a sample, that is, the number of unique ASVs in a sample. For colostrum, we found that there were 32.8 ± 17.9 observed species while for neonatal stool there were 23.5 ± 15.9. The high standard deviation shows that the sample types are quite variable

from within. When comparing the richness of sample types with each other we obtained a p-value of 0.008, which means that colostrum and neonatal stool are significantly different. This result is expected as the biomass and environments of colostrum and neonatal stool are highly different. The Simpson index value for both sample types was high (colostrum = 0.9 ± 0.07, neonatal stool = 0.81 ± 0.19), which indicates that they possess communities dominated by few species, in other words, they are not diverse. The latter is further confirmed by the Inverse Simpson index. The Shannon index shows that of the two sample types, the most diverse one is colostrum and according to the Wilcoxon test applied, this difference is significant (p = 0.0013). Similarly, Fisher's index indicates that colostrum has more taxa and that this difference is significant (p = 0.002) (To see Shapiro test results and the statistics values go to Appendices: Table 3).

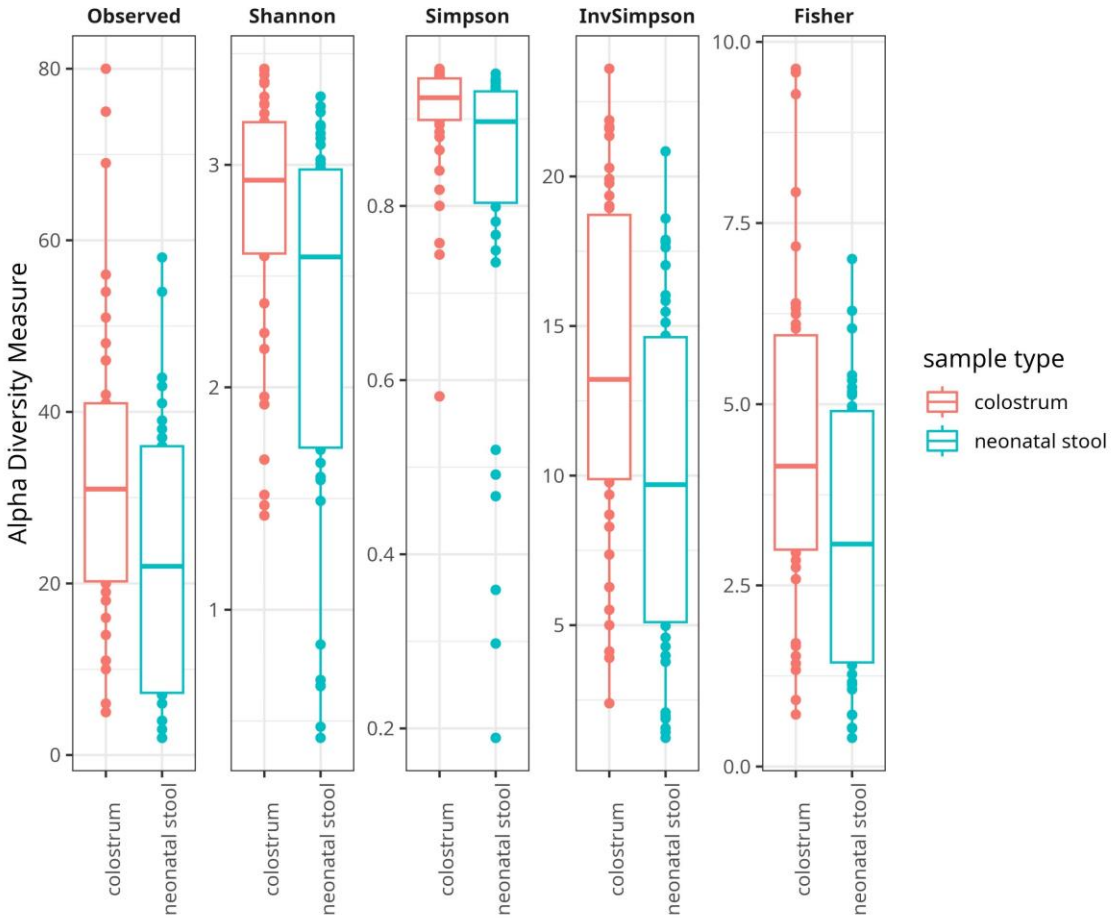

Figure 12. Boxplot of alpha diversity indices (Observed species, Shannon, Simpson, Inverse Simpson, and Fisher).

Table 8. Mean, standard deviation, and p-values for each index according to sample type.

Indices	*Colostrum*	*Neonatal stool*	*pvalues*	*Cohen's d*

Observed	32.83 ± 18	23.5 ± 15.92	0.009	0.79
Shannon	2.78 ± 0.56	2.31 ± 0.86	0.001	0.91
Simpson	0.9 ± 0.07	0.82 ± 0.19	0.001	0.92
InvSimpson	13.59 ± 5.62	9.95 ± 5.64	0.002	0.78
Fisher	4.44 ± 2.26	3.17 ± 1.91	0.002	0.86

8.9 Beta diversity of Archaea in colostrum and neonatal stool

To compare the community differences between the sample groups, we did a non-metrical multidimensional scaling based on the weighted UniFrac distance metric. The UniFrac utilizes a phylogenetic tree to evaluate the closeness of the communities, being weighted means that it uses the shared counts between samples instead of only presence and absence. In Figure 13, it can be noted that colostrum samples are more diverse than neonatal stool samples. However, there is also an overlapping between neonatal stool and colostrum. The R statistic of the ANOSIM test shows the sample types are quite similar since it's far from 1 (R = 0.06198, p= 0.002) and the stress plot (see Appendices: Figure 21) shows that the model is a good fit.

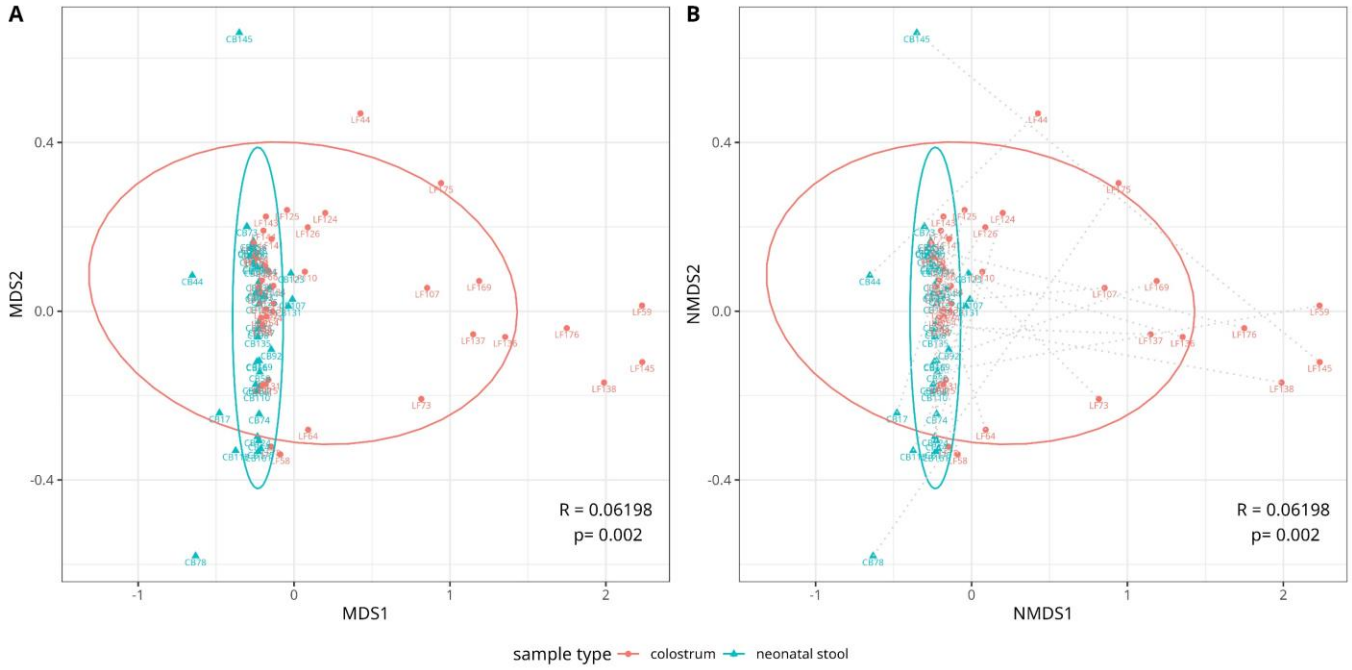

Figure 13. Non-metrical multidimensional scaling (NMDS) plot of weighted UniFrac distance. Each point in the figure represents a sample, the colour denotes the sample type, and the lines in panel B relate to each binomial sample.

Observing that some samples did not cluster, we evaluated them with partitioning around medoids (PAM) to identify the possible outliers (Figure 14, To see the k-medoids model, Appendices:Figure 22). This method identified 10 outlier samples. Then, another sample was removed on the basis that it was collected on day 6 and consisted almost only of the

Methanosphaera genus (see Appendices Figure 23). Performing the weighted Unifrac without these outlier samples resulted in an R statistic closer to 0 (R = 0.03) and a higher p-value (p = 0.06) as would be expected (Figure 15).

.

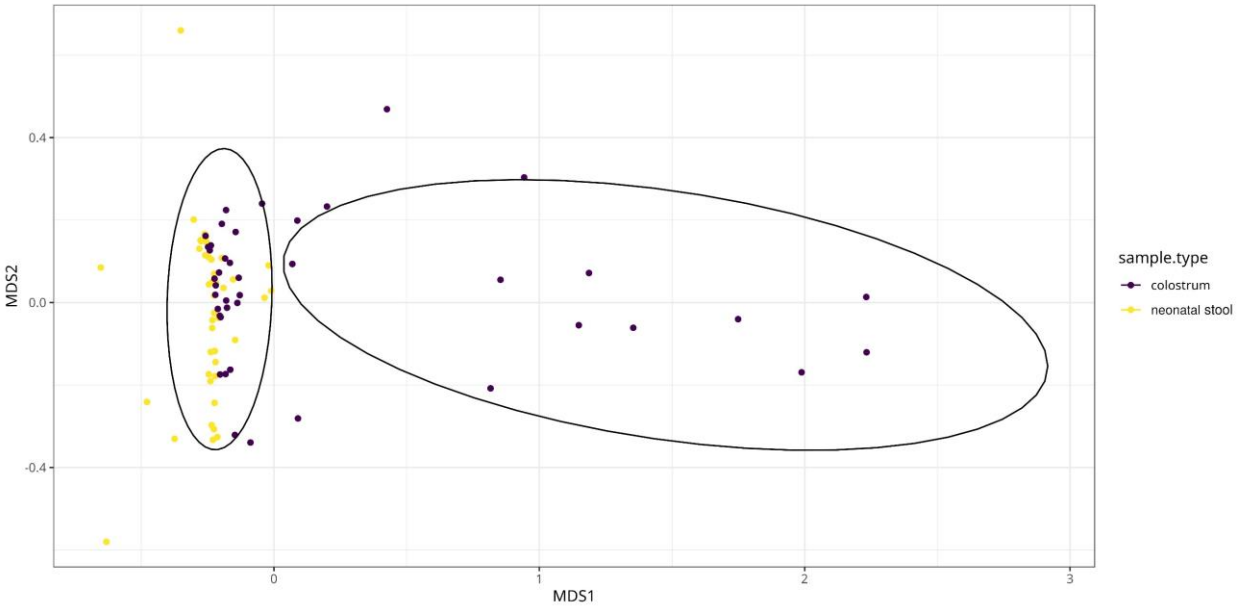

Figure 14. Clusters found with PAM.

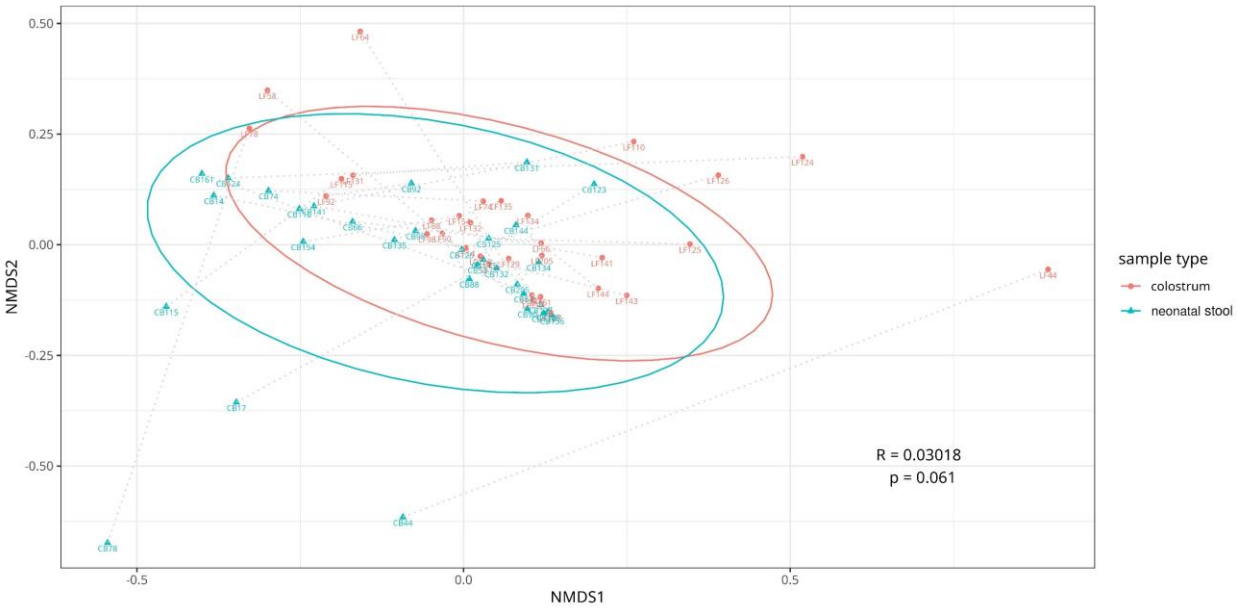

Figure 15. Non-metrical multidimensional scaling (NMDS) plot of weighted UniFrac distance without the outlier samples.

8.10 DESeq 2

A differential abundance analysis based on the negative binomial distribution was carried out to see what sequences were differentially abundant in each sample. In Figure 16, we can observe that there were four ASVs differentially expressed in colostrum with the greater p-value being 3.46×10^{-2}, meaning that all of them were significative (Table 9). After performing a Blast search, we found that these sequences are from Eukarya, corresponding to a fragment of chromosome 17 (Appendices: Figure 24). The alignment of the chromosome 17 sequence amplicon with Archaea amplicons showed that they were somewhat similar at the start of the sequence, approximately 20 nt equal (Figure 25).

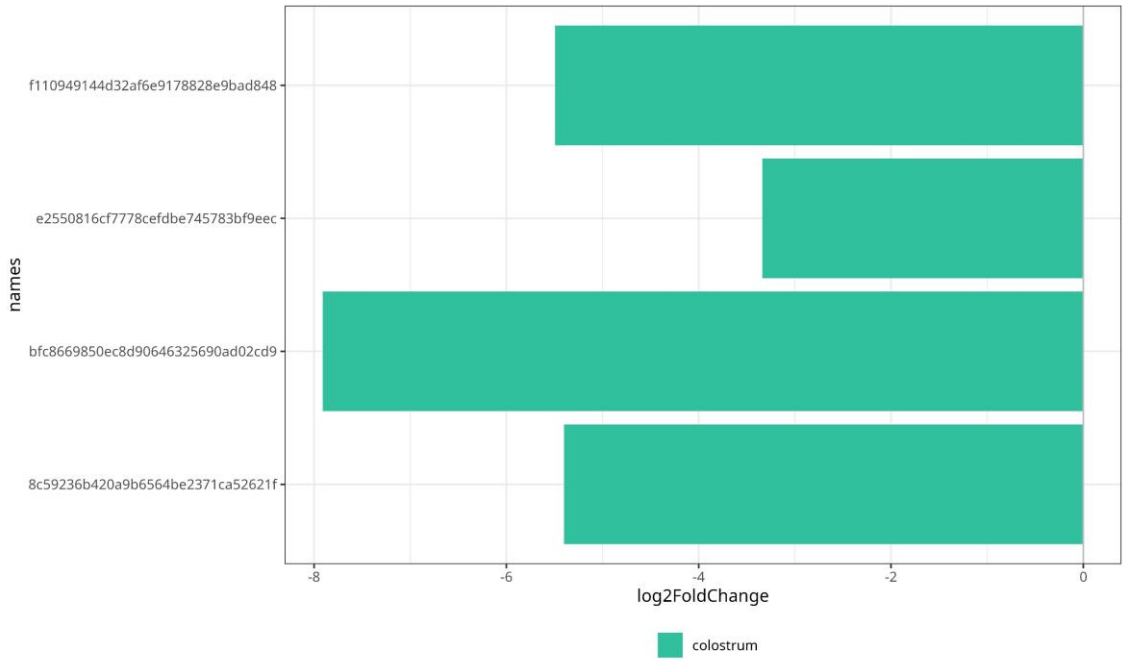

Figure 16. Differential abundance plot.

Table 9. Base means, log2foldchange (effect size estimate), standard errors, statistics, p-values and p-adjusted of the differentially abundant sequences.

Amplicon sequence variant	*Base mean*	*log2FoldChange*	*lfcSE*	*stat*	*p-value*	*p-adj*
8c59236b420a9b6564be2371ca52621f	83.544	-5.401	0.935	-5.776	$7.64X10^{-9}$	$2.17X10^{-6}$
bfc8669850ec8d90646325690ad02cd9	33.387	-7.910	1.481	-5.341	$9.22X10^{-8}$	$1.30X10^{-5}$
f110949144d32af6e9178828e9bad848	47.086	-5.494	1.129	-4.866	$1.13x10^{-6}$	$1.07X10^{-4}$
e2550816cf7778cefdbe745783bf9eec	45.322	-3.339	0.957	-3.487	$4.88X10^{-4}$	$3.46X10^{-2}$

8.11 SourceTracker 2

The SourceTracker 2 analysis (Figure 17) showed that approximately 70% of the archaea in neonatal stool had a potential origin in the colostrum, while a 30% came from unknown sources. In panel B, we can see the relative abundance of each source. *Methanoculleus*, *Methanosarcina* and *Methanothermobacter* were the most abundant genus having colostrum as their source, the same is true for the unknown source with the addition of *Candidatus_Methanoplasma* and *Nitrososphaeraceae*.

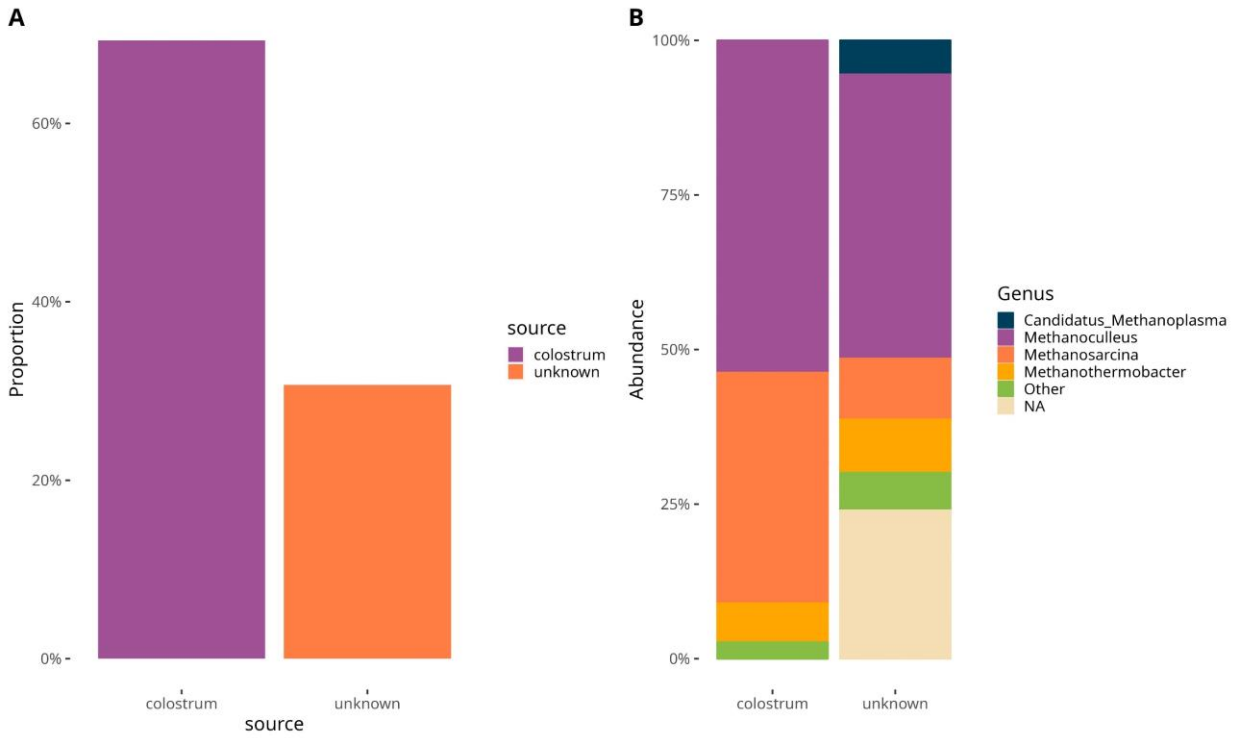

Figure 17. SourceTracker analysis of the possible origin of archaea in neonatal stool. A) The proportion of archaea that are attributed to colostrum, B) The archaeal genus and their possible provenance.

8.12 Association of microbiota diversity with metabolic pathways in colostrum and neonatal stool samples by PICRUSt

Finally, to better understand the role of the archaea present in the samples a predictive functional metagenomic analysis was performed. The PICRUSt analysis identified around 250 functional metabolic pathways, 10 of which are shown in Figure 18, which include Alpha-galactosidase, Glutamate synthase (NADPH) and DNA lyase, among others. However, when a multiple test correction is applied (Benjamin-Hochberg, Bonferroni, Sidak, FDR) none of the metabolic routes is significant. The latter indicates that the pathways found correspond to false positives, that is,

they were detected due to chance. Observing the difference between means of the metabolic pathways of colostrum and neonatal stool it can be noted that they're highly similar and none of them passes the *q-value* threshold (Figure 19).

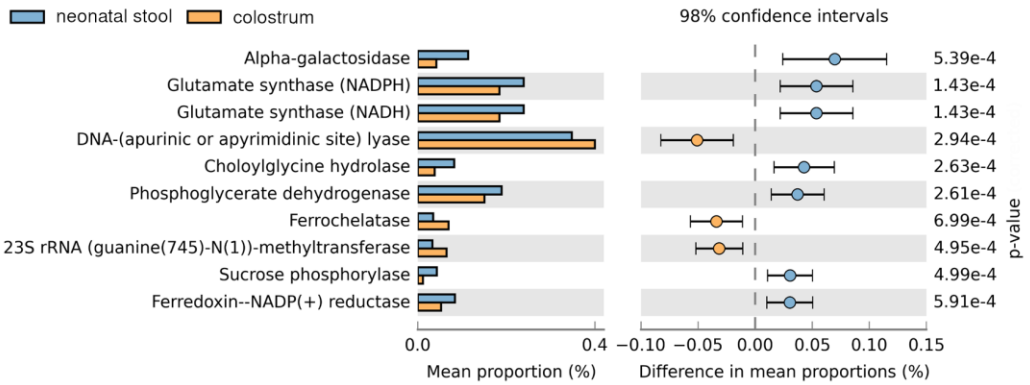

Figure 18. Functional variation of predictive metabolic pathways in colostrum and neonatal stool samples with a p-value of 0.001 (not adjusted).

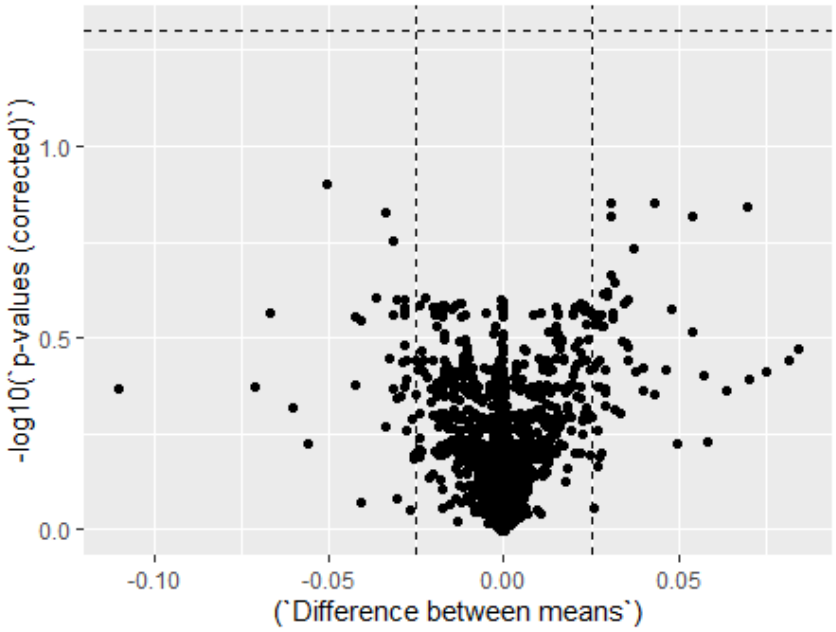

Figure 19. Volcano plot of the difference between means of metabolic pathways found with PICRUSt.

IX. Discussion

In this study, the archaeal composition of colostrum and neonatal stool from Mexican mother-neonate pairs was characterized, being the first report of such type in a Mexican sample. In our work, we found high levels of Halobacterota in both colostrum and neonatal stool (66% and 75%, respectively). Accordingly, the main genera are members of this phylum (*Methanoculleus* and *Methanosarcina*). The genus *Methanoculleus* had already been reported to be found in the human intestinal mucosa, however, this detection was indirect, by sequence analysis of *mcrA* gene clones present in colonic biopsies (Nava et al., 2012). On the other hand, for the genus *Methanosarcina*, there are not many reports, the literature suggests that the prevalence of this group ranges from 10% to 40% depending on age (Mihajlovski et al., 2010). Nonetheless, this evidence comes from a phylotype named Mx-01 (Mihajlovski et al., 2008) which is presumptively from the order Methanosarcinale (Scanlan et al., 2008). Further studies are needed to confirm its identity.

The genus *Methanobrevibacter* was present in the two sample types but in a low abundance (around 3%). This is interesting since *Methanobrevibacter smithii* is thought to be the most abundant methanogenic archaea in the gut, constituting 10% of the anaerobic microorganisms (Bang et al., 2014). Another highly reported methanogen is *Methanosphaera stadmanae*, which has been identified in low abundance in human faeces (Dridi et al., 2009). Here, the genus *Methanosphaera* was found in abundance less than 1% but was dominant in one outlier sample that will be discussed further below, the consensual low abundance of this genus agrees with the undetectable levels of *M. stadmanae* in a recent report in human milk samples (van de Pol et al., 2017).

The differences in genera proportions reported in the literature against this study could be attributed to lifestyle differences since there are no previous reports of archaeal composition in Mexican women. First, it is important to notice that hormonal changes during pregnancy affect the microbiota, therefore the microbial population of a pregnant woman will be distinct. Secondly, it has been observed that diet is one of the main factors influencing the human gut microbiome (David et al., 2014), this could also be the case with human milk considering the entero-mammary pathway (Fernández et al., 2013). Furthermore, the microbiota can also be influenced by geographical location and urban or rural lifestyle (Hasan & Yang, 2019). Considering the foregoing, we hypothesized that the differences found in the proportions of archaeal genera might be due to a combination of lifestyle and the recent pregnancy stage. We consider that future studies

should focus on Mexican populations without any health condition, in this manner we would get insight into the baseline conditions of the archaeal composition.

Moreover, in this study, we detected bacterial sequences in a high proportion. Considering the homology between the 16S rDNA of bacteria and archaea and the greater abundance of bacteria in neonatal stool and colostrum samples, this result is not quite unexpected. Besides, it was reported that the primers could detect the *Chloroflexaceae* bacterial genus (Yu et al., 2005). It is worth mentioning that these primers were thought for bioreactor conditions where archaea growth is favoured, thus in samples that have high concentrations of bacteria it is possible that bacteria monopolize the primers. This can be seen when we compare the positive control (Archaeal DNA from cultures of a bioreactor) assignations with those of colostrum and neonatal stool (Appendices, Figure 20). We think that a possible way of better selecting for archaea would be to add mechanical cell lyses in the DNA extraction (Dridi et al., 2009) and, it might also be useful to add antibiotics that damage bacteria but not archaea during the DNA extraction step.

Comparing the sample types, colostrum and neonatal stool were found to have a highly similar core microbiome, consisting of members of Euryarchaeota, Halobacterota and Thermoplasmota phyla. In opposition, the alpha diversity analysis showed that the sample types were significantly different, colostrum being the most diverse. This is to be expected, considering they are different sample types composing distinct ecosystems and also, that they have different biomass. Rarefaction could be used to reduce this difference; however, these approaches have been reported to introduce more bias (McMurdie & Holmes, 2014). Interestingly, some authors have concluded that alpha diversity measures might underestimate microbiota and more robust statistical methods might be necessary to assess the differences (Kers & Saccenti, 2022; Willis, 2019). We think that a good approach could be to use the DESEq normalization method to adjust the microbiota data to the zero-inflated distribution, which is more suitable for high-throughput data.

The ANOSIM further explained the similarity between the colostrum and neonatal stool samples, the NMDS beta diversity ordination showed that the colostrum and neonatal stool clusters were tightly gathered, suggesting that archaea are vertically transmitted through breastfeeding. Observing some of the outliers, we opted to use partitioning around medoids (PAM) to evaluate if we could discard them, as this method has been proven successful in detecting outliers (Kumar et al., 2013). Here, we saw that when filtrating these outlier samples, the ANOSIM showed higher similarity than before. Further examination of these outlier samples

showed that they had a high proportion of unassigned taxa and in one case, there was an unusual abundance of the genus *Methanosphaera* (Appendices, Figure 23). We hypothesized that this high proportion of *Methanosphaera* in this one neonatal stool sample is due to the day of collection (day 6), suggesting that at this time there is a replacement of archaea genera and a tendency towards dominance of a specie. We think that in future studies it would be interesting to explore the archaeal composition of the neonatal gut longitudinally.

The DESeq analysis did not show any differentially abundant archaea, further cementing the idea of similarity between the sample types. Interestingly, the DESeq showed a Eukarya sequence as differentially expressed in colostrum, we consider that this is due to the high presence of human cells in colostrum. Additionally, it seems that the presence of Eukarya sequences is a common problem in 16S rDNA sequencing studies. In fact, it has been reported that sequencing of breast tumour samples was highly affected by this issue, and recommended to sequence the V1-V2 regions because that reduces off-target amplification (Walker et al., 2020). However, this was focused on bacteria, so it remains to be known whether this applies to archaea. It is also important to mention that when filtering the outliers and performing the DESEQ,we did not see any differentially abundant sequences. The latter implies that possibly many of the unassigned sequences found in the outlier samples belonged to Eukarya. Parallelly, the PICRUSt2 analysis did not find any significatively different metabolic pathway between the sample types and the very low difference between means suggests a similarity between the colostrum and neonatal stool samples.

Finally, the SourceTracker 2 analysis revealed that around 69% of the archaea in neonatal stool probably originated from colostrum, this further support the hypothesis that there is a transmission of archaea from the mammary gland to the neonatal gut during lactation. It is noteworthy to mention that SourceTracker was designed to track the origin of microbial contaminants by a Bayesian probabilistic approach, one of its limitations is that the higher the similarity between communities, the higher the square error (Knights et al., 2011). In our case, beta diversity and core microbiota demonstrate high similarities between colostrum and neonatal stool communities. Additionally, the specificity of SourceTracker for contaminant detection implies that this method might not be optimal as archaea are thought to be vertically transmitted to the infant but also interact with the host gut and perform biological functions. Here, we cannot discard the fact that is possible that some of the archaea come from the oral and/or vaginal tract. We consider that in future studies it would be useful to sample the oral and vaginal microbiota of

the mother as well as the environment where the sample was taken and then apply the SourceTracker analysis to better infer what is the possible source of each genus.

X. Conclusions

In summary, in this study, we were able to characterize the archaeal composition of binomial samples mother-neonate and the similarity between the sample types implies that there is a vertical transmission of archaea during breastfeeding.

1. The archaeal community of colostrum and neonatal stool of Mexican mother-neonate pairs are mainly composed of *Methanoculleus* and *Methanosarcina* genus.
2. The archaeal community found in colostrum samples are more diverse than neonatal stool.
3. Colostrum and neonatal stool present a high dominance of archaea genera.
4. The archaeal communities of colostrum and neonatal stool are highly similar with no statistically significant difference in species or metabolic pathway.
5. The results of SourceTracker analysis along with the similarity between colostrum and neonatal stool suggests that there is a vertical transmission of archaea.

Though some of the limitations of this study were referred to in the discussion, it is important to mention that this study only applies to a population from Estado de México. Also, it only suggests vertical transmission (mother-newborn) but does not consider the entero-mammary pathway. In addition, it must be considered that stool samples are representative and that 16S rDNA sequencing is presumptive. All in all, this study represents a first step in understanding the origin of archaea in the gut from the beginning of life and remarks the importance to continue studying this often overlooked microorganisms.

XI. Future perspectives

Futures studies should search the contribution of archaea to the newborn from all the possible sources (skin, vagina, oral, human milk, and the environment). In this manner, we will better understand which source is the main contributor of archaea to the neonate. Apart, it would be interesting to confirm the origin of archaea in human milk, that is, evaluate if it’s coming from the gastrointestinal tract as proposed with the entero-mammary pathway. Moreover, studies of Mexican people without any health condition along with more diverse populations will improve our understanding of these communities and their possible function. In this order of ideas, further case-control studies will allow us to better comprehend the relationship of these microorganisms

with health. Finally, we consider that the interactions of archaea present in the neonate gut should be examined, as well as their change in composition through time.

XII. Bibliography

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for «Grid» Graphics*. https://CRAN.R-project.org/package=gridExtra

Ballard, O., & Morrow, A. L. (2013). Human Milk Composition: Nutrients and Bioactive Factors. *Pediatric Clinics*, *60*(1), 49-74. https://doi.org/10.1016/j.pcl.2012.10.002

Bang, C., Weidenbach, K., Gutsmann, T., Heine, H., & Schmitz, R. A. (2014). The Intestinal Archaea Methanosphaera stadtmanae and Methanobrevibacter smithii Activate Human Dendritic Cells. *PLOS ONE*, *9*(6), e99411. https://doi.org/10.1371/journal.pone.0099411

Binns, C., Lee, M., & Low, W. Y. (2016). The Long-Term Public Health Benefits of Breastfeeding. *Asia Pacific Journal of Public Health*, *28*(1), 7-14. https://doi.org/10.1177/1010539515624964

Bisanz, J. E. (2018). *Qiime2R: Importing QIIME2 artifacts and associated data into R sessions*. https://github.com/jbisanz/qiime2R

Camara, A., Konate, S., Tidjani Alou, M., Kodio, A., Togo, A. H., Cortaredona, S., Henrissat, B., Thera, M. A., Doumbo, O. K., Raoult, D., & Million, M. (2021). Clinical evidence of the role of Methanobrevibacter smithii in severe acute malnutrition. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-84641-8

Cavicchioli, R. (2011). Archaea—Timeline of the third domain. *Nature Reviews Microbiology*, *9*(1), Article 1. https://doi.org/10.1038/nrmicro2482

Chaban, B., Ng, S. Y. M., & Jarrell, K. F. (2006). Archaeal habitats—From the extreme to the ordinary. *Canadian Journal of Microbiology*, *52*(2), 73-116. https://doi.org/10.1139/w05-147

Chowdhury, R., Sinha, B., Sankar, M. J., Taneja, S., Bhandari, N., Rollins, N., Bahl, R., & Martines, J. (2015). Breastfeeding and maternal health outcomes: A systematic review and meta-analysis. *Acta Paediatrica*, *104*(S467), 96-113. https://doi.org/10.1111/apa.13102

Corona-Cervantes, K., García-González, I., Villalobos-Flores, L. E., Hernández-Quiroz, F., Piña-Escobedo, A., Hoyo-Vadillo, C., Rangel-Calvillo, M. N., & García-Mena, J. (2020). Human milk microbiota associated with early colonization of the neonatal gut in Mexican newborns. *PeerJ*, *8*, e9205. https://doi.org/10.7717/peerj.9205

DasSarma, S., Coker, J. A., & DasSarma, P. (2009). Archaea (overview). En M. Schaechter (Ed.), *Encyclopedia of Microbiology (Third Edition)* (pp. 1-23). Academic Press. https://doi.org/10.1016/B978-012373944-5.00108-5

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), Article 7484. https://doi.org/10.1038/nature12820

Dekaboruah, E., Suryavanshi, M. V., Chettri, D., & Verma, A. K. (2020). Human microbiome: An academic update on human body site specific surveillance and its possible role. *Archives of Microbiology*, *202*(8), 2147-2167. https://doi.org/10.1007/s00203-020-01931-x

Dewey, K. G., Nommsen-Rivers, L. A., Heinig, M. J., & Cohen, R. J. (2003). Risk Factors for Suboptimal Infant Breastfeeding Behavior, Delayed Onset of Lactation, and Excess Neonatal Weight Loss. *Pediatrics*, *112*(3), 607-619. https://doi.org/10.1542/peds.112.3.607

Dogaru, C. M., Nyffenegger, D., Pescatore, A. M., Spycher, B. D., & Kuehni, C. E. (2014). Breastfeeding and Childhood Asthma: Systematic Review and Meta-Analysis. *American Journal of Epidemiology*, *179*(10), 1153-1167. https://doi.org/10.1093/aje/kwu072

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), Article 6. https://doi.org/10.1038/s41587-020-0548-6

Dridi, B., Henry, M., Khéchine, A. E., Raoult, D., & Drancourt, M. (2009). High Prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae Detected in the Human Gut Using an Improved DNA Detection Protocol. *PLOS ONE*, *4*(9), e7063. https://doi.org/10.1371/journal.pone.0007063

Eme, L., & Doolittle, W. F. (2015). Archaea. *Current Biology*, *25*(19), R851-R855. https://doi.org/10.1016/j.cub.2015.05.025

Eme, L., Spang, A., Lombard, J., Stairs, C. W., & Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, *15*(12), Article 12. https://doi.org/10.1038/nrmicro.2017.133

Fernández, L., Langa, S., Martín, V., Maldonado, A., Jiménez, E., Martín, R., & Rodríguez, J. M. (2013). The human milk microbiota: Origin and potential roles in health and disease. *Pharmacological Research*, *69*(1), 1-10. https://doi.org/10.1016/j.phrs.2012.09.001

Frank, N. M., Lynch, K. F., Uusitalo, U., Yang, J., Lönnrot, M., Virtanen, S. M., Hyöty, H., Norris, J. M., Rewers, M., Bautista, K., Baxter, J., Felipe-Morales, D., Driscoll, K., Frohnert, B. I., Gallant, M., Gesualdo, P., Hoffman, M., Karban, R., Liu, E., … for the TEDDY Study Group. (2019). The relationship between breastfeeding and reported respiratory and gastrointestinal infection rates in young children. *BMC Pediatrics*, *19*(1), 339. https://doi.org/10.1186/s12887-019-1693-2

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., & Brugère, J.-F. (2014). Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology : WJG*, *20*(43), 16062-16078. https://doi.org/10.3748/wjg.v20.i43.16062

Gállego Bravo, A. K., Salcedo Serrano, D. A., López Jiménez, G., Nirmalkar, K., Murugesan, S., García-Mena, J., Gutiérrez Castillo, M. E., & Tovar Gálvez, L. R. (2019). Microbial Profile of the Leachate from Mexico City's Bordo Poniente Composting Plant: An Inoculum to Digest Organic Waste. *Energies*, *12*(12), Article 12. https://doi.org/10.3390/en12122343

Gila-Diaz, A., Arribas, S. M., Algara, A., Martín-Cabrejas, M. A., López de Pablo, Á. L., Sáenz de Pipaón, M., & Ramiro-Cortijo, D. (2019). A Review of Bioactive Factors in Human Breastmilk: A Focus on Prematurity. *Nutrients*, *11*(6), Article 6. https://doi.org/10.3390/nu11061307

Gómez-Gallego, C., Pohl, S., Salminen, S., De Vos, W. m., & Kneifel, W. (2016). Akkermansia muciniphila: A novel functional microbe with probiotic properties. *Beneficial Microbes*, *7*(4), 571-584. https://doi.org/10.3920/BM2016.0009

Grine, G., Boualam, M. A., & Drancourt, M. (2017). Methanobrevibacter smithii, a methanogen consistently colonising the newborn stomach. *European Journal of Clinical Microbiology & Infectious Diseases*, *36*(12), 2449-2455. https://doi.org/10.1007/s10096-017-3084-7

Gu, Z. (2022). Complex Heatmap Visualization. *iMeta*. https://doi.org/10.1002/imt2.43

Hasan, N., & Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ*, *7*, e7502. https://doi.org/10.7717/peerj.7502

Hilpert, R., Winter, J., Hammes, W., & Kandler, O. (1981). The sensitivity of archaebacteria to antibiotics. *Zentralblatt Für Bakteriologie Mikrobiologie Und Hygiene: I. Abt. Originale C: Allgemeine, Angewandte Und Ökologische Mikrobiologie*, *2*(1), 11-20. https://doi.org/10.1016/S0721-9571(81)80014-2

Horta, B. L., Loret de Mola, C., & Victoria, C. G. (2015). Long-term consequences of breastfeeding on cholesterol, obesity, systolic blood pressure and type 2 diabetes: A systematic review and meta-analysis. *Acta paediatrica*, *104*, 30-37. https://doi.org/10.1111/apa.13133

Horz, H.-P. (2015). Archaeal Lineages within the Human Microbiome: Absent, Rare or Elusive? *Life*, *5*(2), Article 2. https://doi.org/10.3390/life5021333

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., & Reddy, D. N. (2015). Role of the normal gut microbiota. *World Journal of Gastroenterology : WJG*, *21*(29), 8787-8803. https://doi.org/10.3748/wjg.v21.i29.8787

Jenness, R. (1979). The composition of human milk. *Seminars in Perinatology*, *3*(3), 225-239.

Kandler, O., & Konig, H. (1993). Chapter 8 Cell envelopes of archaea: Structure and chemistry. En M. Kates, D. J. Kushner, & A. T. Matheson (Eds.), *New Comprehensive Biochemistry* (Vol. 26, pp. 223-259). Elsevier. https://doi.org/10.1016/S0167-7306(08)60257-4

Kandler, O., & König, H. (1998). Cell wall polymers in Archaea (Archaebacteria). *Cellular and Molecular Life Sciences CMLS*, *54*(4), 305-308. https://doi.org/10.1007/s000180050156

Kers, J. G., & Saccenti, E. (2022). The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.796025

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R., & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, *8*(9), Article 9. https://doi.org/10.1038/nmeth.1650

Kumar, V., Kumar, S., & Singh, A. K. (2013). *Outlier Detection: A Clustering-Based Approach*. *1*(7).

Le Doare, K., Holder, B., Bassett, A., & Pannaraj, P. S. (2018). Mother's Milk: A Purposeful Contribution to the Development of the Infant Microbiota and Immunity. *Frontiers in Immunology*, *9*. https://www.frontiersin.org/articles/10.3389/fimmu.2018.00361

Love, M., Ahlmann-Eltze, C., Forbes, K., Anders, S., Huber, W., FP7, R. E., NHGRI, N., & CZI. (2023). *DESeq2: Differential gene expression analysis based on the negative binomial distribution* (1.40.2) [Software]. Bioconductor version: Release (3.17). https://doi.org/10.18129/B9.bioc.DESeq2

Marseglia, L., Manti, S., D'Angelo, G., Cuppari, C., Salpietro, V., Filippelli, M., Trovato, A., Gitto, E., Salpietro, C., & Arrigo, T. (2015). Obesity and breastfeeding: The strength of association. *Women and Birth, 28*(2), 81-86. https://doi.org/10.1016/j.wombi.2014.12.007

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*(4), e61217.

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology, 10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Mihajlovski, A., Alric, M., & Brugère, J.-F. (2008). A putative new order of methanogenic Archaea inhabiting the human gut, as revealed by molecular analyses of the mcrA gene. *Research in Microbiology, 159*(7), 516-521. https://doi.org/10.1016/j.resmic.2008.06.007

Mihajlovski, A., Doré, J., Levenez, F., Alric, M., & Brugère, J.-F. (2010). Molecular evaluation of the human gut methanogenic archaeal microbiota reveals an age-associated increase of the diversity. *Environmental Microbiology Reports, 2*(2), 272-280. https://doi.org/10.1111/j.1758-2229.2009.00116.x

Moossavi, S., & Azad, M. B. (2020). Origins of human milk microbiota: New evidence and arising questions. *Gut Microbes, 12*(1), 1667722. https://doi.org/10.1080/19490976.2019.1667722

Nam, Y.-D., Chang, H.-W., Kim, K.-H., Roh, S. W., Kim, M.-S., Jung, M.-J., Lee, S.-W., Kim, J.-Y., Yoon, J.-H., & Bae, J.-W. (2008). Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people. *The Journal of Microbiology, 46*(5), 491-501. https://doi.org/10.1007/s12275-008-0199-7

Notarbartolo, V., Giuffrè, M., Montante, C., Corsello, G., & Carta, M. (2022). Composition of Human Breast Milk Microbiota and Its Role in Children's Health. *Pediatric Gastroenterology, Hepatology & Nutrition, 25*(3), 194-210. https://doi.org/10.5223/pghn.2022.25.3.194

Offre, P., Spang, A., & Schleper, C. (2013). Archaea in Biogeochemical Cycles. *Annual Review of Microbiology*, *67*(1), 437-457. https://doi.org/10.1146/annurev-micro-092412-155614

Ogra, P., Walker, W., & Lönnerdal, B. (2020). *Milk, Mucosal Immunity and the Microbiome: Impact on the Neonate*. *94*, 75-85. https://doi.org/10.1159/000505031

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Weedon, J. (2022). *vegan: Community Ecology Package*. https://CRAN.R-project.org/package=vegan

Pace, R. M., Williams, J. E., Robertson, B., Lackey, K. A., Meehan, C. L., Price, W. J., Foster, J. A., Sellen, D. W., Kamau-Mbuthia, E. W., Kamundia, E. W., Mbugua, S., Moore, S. E., Prentice, A. M., Kita, D. G., Kvist, L. J., Otoo, G. E., Ruiz, L., Rodríguez, J. M., Pareja, R. G., … McGuire, M. K. (2021). Variation in Human Milk Composition Is Related to Differences in Milk and Infant Fecal Microbial Communities. *Microorganisms*, *9*(6), Article 6. https://doi.org/10.3390/microorganisms9061153

Palmer, C., Bik, E. M., Giulio, D. D., Realman, D. A., & Brown, P. O. (2007, junio 26). *Development of the Human Infant Intestinal Microbiota | PLOS Biology*. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0050177

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team. (2022). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

Sattari, M., Serwint, J. R., & Levine, D. M. (2019). Maternal Implications of Breastfeeding: A Review for the Internist. *The American Journal of Medicine*, *132*(8), 912-920. https://doi.org/10.1016/j.amjmed.2019.02.021

Scanlan, P. D., Shanahan, F., & Marchesi, J. R. (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiology*, *8*(1), 79. https://doi.org/10.1186/1471-2180-8-79

Sekirov, I., Russell, S. L., Antunes, L. C. M., & Finlay, B. B. (2010). Gut Microbiota in Health and Disease. *Physiological Reviews*, *90*(3), 859-904. https://doi.org/10.1152/physrev.00045.2009

Selma-Royo, M., Calvo-Lerma, J., Bäuerl, C., Esteban-Torres, M., Cabrera-Rubio, R., & Collado, M. C. (2022). Human milk microbiota: What did we learn in the last 20 years? *Microbiome Research Reports*, *1*(3), 19. https://doi.org/10.20517/mrr.2022.05

Thapa, B. R. (2005). Health factors in colostrum. *The Indian Journal of Pediatrics*, *72*(7), 579-581. https://doi.org/10.1007/BF02724182

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *The Biochemical Journal*, *474*(11), 1823-1836. https://doi.org/10.1042/BCJ20160510

Togo, A. H., Grine, G., Khelaifia, S., des Robert, C., Brevaut, V., Caputo, A., Baptiste, E., Bonnet, M., Levasseur, A., Drancourt, M., Million, M., & Raoult, D. (2019). Culture of Methanogenic Archaea from Human Colostrum and Milk. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-54759-x

Uruakpa, F. O., Ismond, M. A. H., & Akobundu, E. N. T. (2002). Colostrum and its benefits: A review. *Nutrition Research*, *22*(6), 755-767. https://doi.org/10.1016/S0271-5317(02)00373-1

van de Pol, J. A. A., van Best, N., Mbakwa, C. A., Thijs, C., Savelkoul, P. H., Arts, I. C. W., Hornef, M. W., Mommers, M., & Penders, J. (2017). Gut Colonization by Methanogenic Archaea Is Associated with Organic Dairy Consumption in Children. *Frontiers in Microbiology*, *8*, 355. https://doi.org/10.3389/fmicb.2017.00355

Victora, C. G., Bahl, R., Barros, A. J. D., França, G. V. A., Horton, S., Krasevec, J., Murch, S., Sankar, M. J., Walker, N., & Rollins, N. C. (2016). Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effect. *The Lancet*, *387*(10017), 475-490. https://doi.org/10.1016/S0140-6736(15)01024-7

Walker, S. P., Barrett, M., Hogan, G., Flores Bueso, Y., Claesson, M. J., & Tangney, M. (2020). Non-specific amplification of human DNA is a major challenge for 16S rRNA gene sequence analysis. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-73403-7

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/articles/10.3389/fmicb.2019.02407

Yates, S. (2010). CHAPTER 3—Western approach to the postpartum. En *Pregnancy and Childbirth* (pp. 99-120). Churchill Livingstone. https://doi.org/10.1016/B978-0-7020-3055-0.00003-0

Yu, Y., Lee, C., Kim, J., & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering*, *89*(6), 670-679. https://doi.org/10.1002/bit.20347

XIII. Appendices

Appendix A. List of barcodes adapters sequences

Table 1: List of 100 barcode adapters sequences for each sequencing primer

Name	Adaptador	Barcode	Spacer	16S rRNA Primer (5→'3')
ArcF1	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GATCTGCGATCC	GT	ATTAGATACCCSBGTAGTCC
ArcF2	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAGCTCATCAGC	GT	ATTAGATACCCSBGTAGTCC
ArcF3	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAAACAACAGCT	GT	ATTAGATACCCSBGTAGTCC
ArcF4	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCAACACCATCC	GT	ATTAGATACCCSBGTAGTCC
ArcF5	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCGATATATCGC	GT	ATTAGATACCCSBGTAGTCC
ArcF6	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAGCAATCCTA	GT	ATTAGATACCCSBGTAGTCC
ArcF7	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTCGTGCACAT	GT	ATTAGATACCCSBGTAGTCC
ArcF8	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GTATCTGCGCGT	GT	ATTAGATACCCSBGTAGTCC
ArcF9	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAGGGAAAGTC	GT	ATTAGATACCCSBGTAGTCC
ArcF10	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAAATTCGGGAT	GT	ATTAGATACCCSBGTAGTCC
ArcF11	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGATTGACCAAC	GT	ATTAGATACCCSBGTAGTCC
ArcF12	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTTACGAGCTA	GT	ATTAGATACCCSBGTAGTCC
ArcF13	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCATATGCACTG	GT	ATTAGATACCCSBGTAGTCC
ArcF14	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAACTCCCGTGA	GT	ATTAGATACCCSBGTAGTCC
ArcF15	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTGCGTTAGCAG	GT	ATTAGATACCCSBGTAGTCC
ArcF16	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TACGAGCCCTAA	GT	ATTAGATACCCSBGTAGTCC
ArcF17	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CACTACGCTAGA	GT	ATTAGATACCCSBGTAGTCC
ArcF18	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGCAGTCCTCGA	GT	ATTAGATACCCSBGTAGTCC
ArcF19	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACCATAGCTCCG	GT	ATTAGATACCCSBGTAGTCC
ArcF20	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGGACATCTCTT	GT	ATTAGATACCCSBGTAGTCC
ArcF21	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GAACACTTTGGA	GT	ATTAGATACCCSBGTAGTCC
ArcF22	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GAGCCATCTGTA	GT	ATTAGATACCCSBGTAGTCC
ArcF23	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTGGGTACACGT	GT	ATTAGATACCCSBGTAGTCC
ArcF24	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AAGGCGCTCCTT	GT	ATTAGATACCCSBGTAGTCC
ArcF25	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TAATACGGATCG	GT	ATTAGATACCCSBGTAGTCC
ArcF26	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGGAATTAGAC	GT	ATTAGATACCCSBGTAGTCC
ArcF27	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGTGAATTCGGA	GT	ATTAGATACCCSBGTAGTCC
ArcF28	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CATTCGTGGCGT	GT	ATTAGATACCCSBGTAGTCC
ArcF29	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACGCACGCTAG	GT	ATTAGATACCCSBGTAGTCC
ArcF30	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACACTGTTCATG	GT	ATTAGATACCCSBGTAGTCC
ArcF31	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACCAGACGATGC	GT	ATTAGATACCCSBGTAGTCC
ArcF32	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCTCATGGAT	GT	ATTAGATACCCSBGTAGTCC
ArcF33	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTCACGGTATG	GT	ATTAGATACCCSBGTAGTCC
ArcF34	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGACCGTCAGAC	GT	ATTAGATACCCSBGTAGTCC
ArcF35	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCACGAGCCTA	GT	ATTAGATACCCSBGTAGTCC
ArcF36	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACAGACCACTCA	GT	ATTAGATACCCSBGTAGTCC

ArcF37	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACCAGCGACTAG	GT	ATTAGATACCCSBGTAGTCC
ArcF38	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGGATCGTCAG	GT	ATTAGATACCCSBGTAGTCC
ArcF39	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCTTGACAGCT	GT	ATTAGATACCCSBGTAGTCC
ArcF40	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACTGTGCGTAC	GT	ATTAGATACCCSBGTAGTCC
ArcF41	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACCGCAGAGTCA	GT	ATTAGATACCCSBGTAGTCC
ArcF42	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGGTGAGTGTC	GT	ATTAGATACCCSBGTAGTCC
ArcF43	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTCGATTCGAT	GT	ATTAGATACCCSBGTAGTCC
ArcF44	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGACTGCGTACT	GT	ATTAGATACCCSBGTAGTCC
ArcF45	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCAGTCGCGAT	GT	ATTAGATACCCSBGTAGTCC
ArcF46	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGGACGCACTGT	GT	ATTAGATACCCSBGTAGTCC
ArcF47	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AAGAGATGTCGA	GT	ATTAGATACCCSBGTAGTCC
ArcF48	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACAGCAGTGGTC	GT	ATTAGATACCCSBGTAGTCC
ArcF49	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGTACTCAGTG	GT	ATTAGATACCCSBGTAGTCC
ArcF50	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTCGCACAGGA	GT	ATTAGATACCCSBGTAGTCC
ArcF51	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGAGAGCAAGTG	GT	ATTAGATACCCSBGTAGTCC
ArcF52	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCATATGAGAG	GT	ATTAGATACCCSBGTAGTCC
ArcF53	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGGCTACACGAC	GT	ATTAGATACCCSBGTAGTCC
ArcF54	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AAGCTGCAGTCG	GT	ATTAGATACCCSBGTAGTCC
ArcF55	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACAGCTAGCTTG	GT	ATTAGATACCCSBGTAGTCC
ArcF56	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACCTGTCTCTCT	GT	ATTAGATACCCSBGTAGTCC
ArcF57	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGTCTGTAGCA	GT	ATTAGATACCCSBGTAGTCC
ArcF58	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGAGCAAGAGCA	GT	ATTAGATACCCSBGTAGTCC
ArcF59	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCCATACTGAC	GT	ATTAGATACCCSBGTAGTCC
ArcF60	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGGTGTGATCGC	GT	ATTAGATACCCSBGTAGTCC
ArcF61	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AATCAGTCTCGT	GT	ATTAGATACCCSBGTAGTCC
ArcF62	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGACGTCTTAG	GT	ATTAGATACCCSBGTAGTCC
ArcF63	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGTGAGAGAAT	GT	ATTAGATACCCSBGTAGTCC
ArcF64	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTGACAGCCAT	GT	ATTAGATACCCSBGTAGTCC
ArcF65	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCGACTGTGCA	GT	ATTAGATACCCSBGTAGTCC
ArcF66	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTACGCTCGAG	GT	ATTAGATACCCSBGTAGTCC
ArcF67	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AATCGTGACTCG	GT	ATTAGATACCCSBGTAGTCC
ArcF68	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGAGTGCTATC	GT	ATTAGATACCCSBGTAGTCC
ArcF69	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTGATCCTAGT	GT	ATTAGATACCCSBGTAGTCC
ArcF70	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGAGTCCTGAGC	GT	ATTAGATACCCSBGTAGTCC
ArcF71	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCGAGCTATCT	GT	ATTAGATACCCSBGTAGTCC
ArcF72	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTACTGCAGGC	GT	ATTAGATACCCSBGTAGTCC
ArcF73	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACACACTATGGC	GT	ATTAGATACCCSBGTAGTCC
ArcF74	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGATGCGACCA	GT	ATTAGATACCCSBGTAGTCC
ArcF75	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGTTAGCACAC	GT	ATTAGATACCCSBGTAGTCC
ArcF76	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTGTACGCGTA	GT	ATTAGATACCCSBGTAGTCC

ArcF77	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGATACACGCGC	GT	ATTAGATACCCSBGTAGTCC
ArcF78	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCGCTGATGTG	GT	ATTAGATACCCSBGTAGTCC
ArcF79	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACACATGTCTAC	GT	ATTAGATACCCSBGTAGTCC
ArcF80	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACATGATCGTTC	GT	ATTAGATACCCSBGTAGTCC
ArcF81	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCAACTGCTA	GT	ATTAGATACCCSBGTAGTCC
ArcF82	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTGTCGAAGCT	GT	ATTAGATACCCSBGTAGTCC
ArcF83	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCGTAGGTCGT	GT	ATTAGATACCCSBGTAGTCC
ArcF84	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACACGAGCCACA	GT	ATTAGATACCCSBGTAGTCC
ArcF85	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACATGTCACGTG	GT	ATTAGATACCCSBGTAGTCC
ArcF86	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCGATACTGG	GT	ATTAGATACCCSBGTAGTCC
ArcF87	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTACGTGTGGT	GT	ATTAGATACCCSBGTAGTCC
ArcF88	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTGTGACTTCA	GT	ATTAGATACCCSBGTAGTCC
ArcF89	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGATCTCTGCAT	GT	ATTAGATACCCSBGTAGTCC
ArcF90	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCTATCCACGA	GT	ATTAGATACCCSBGTAGTCC
ArcF91	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTCCATAGCTG	GT	ATTAGATACCCSBGTAGTCC
ArcF92	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACACGGTGTCTA	GT	ATTAGATACCCSBGTAGTCC
ArcF93	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACATTCAGCGCA	GT	ATTAGATACCCSBGTAGTCC
ArcF94	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTTGTAGCAGC	GT	ATTAGATACCCSBGTAGTCC
ArcF95	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGATGTTCTGCT	GT	ATTAGATACCCSBGTAGTCC
ArcF96	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCTCCATACAG	GT	ATTAGATACCCSBGTAGTCC
ArcF97	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCTATCTGGA	GT	ATTAGATACCCSBGTAGTCC
ArcF98	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACTATTGTCACG	GT	ATTAGATACCCSBGTAGTCC
ArcF99	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGAACACGTCTC	GT	ATTAGATACCCSBGTAGTCC
ArcF100	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTGAGAGAAGC	GT	ATTAGATACCCSBGTAGTCC
Reverse	CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT	NOT APPLICABLE	CC	GCCATGCACCWCCTCT

Appendix B. Relative Domain abundance for the sample types and the positive control.

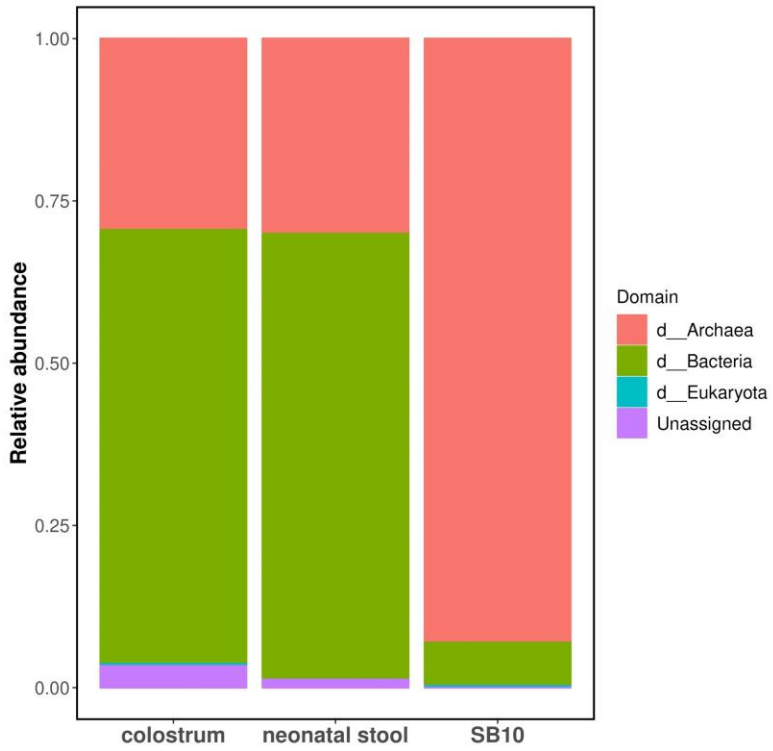

Figure 20. Relative Domain abundance for colostrum, neonatal stool, and the positive control (SB10).

Table 2. Relative Domain abundance for the positive control, colostrum and neonatal stool.

Sample type	*Domain*	*Abundance*
Positive control	Unassigned	0.287
	Archaea	93.043
	Bacteria	6.670
	Eukaryota	0.000
Colostrum	Unassigned	3.646
	Archaea	29.459
	Bacteria	66.894
	Eukaryota	0.001
Neonatal stool	Unassigned	1.250
	Archaea	30.059
	Bacteria	68.691
	Eukaryota	0.000

Appendix C. Relative abundance of archaea and unassigned

Table 3. Relative abundance of archaea and unassigned in each sample type.

Sample type	*Domain*	*Abundance*
colostrum	Unassigned	24.741
colostrum	Archaea	75.259
neonatal stool	Unassigned	14.544
neonatal stool	Archaea	85.456

Appendix D. Alpha diversity of Archaea

Table 4. Alpha diversity indices and their respective p and statistic values.

Indices	*pshap*	*pvals*	*stats*
Observed	0.0061	0.0088	661.5000
Shannon	0.0000	0.0013	702.0000
Simpson	0.0000	0.0013	702.0000
InvSimpson	0.0736	0.0020	3.2992
Fisher	0.0112	0.0023	690.0000

Table 5. Alpha diversity indices mean for each sample type.

Indices	*colostrum*	*neonatal stool*
Simpson	0.9 ± 0.07	0.82 ± 0.19
InvSimpson	13.59 ± 5.62	9.95 ± 5.64
Shannon	2.78 ± 0.56	2.31 ± 0.86
Observed	32.83 ± 18	23.5 ± 15.92
Fisher	4.44 ± 2.26	3.17 ± 1.91

Appendix E. Beta diversity of Archaea

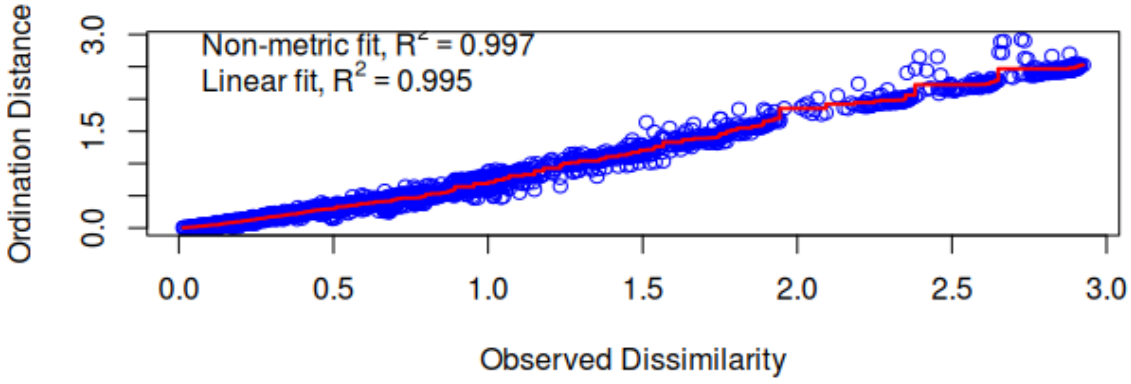

Figure 21. Stress plot of the NMDS analysis.

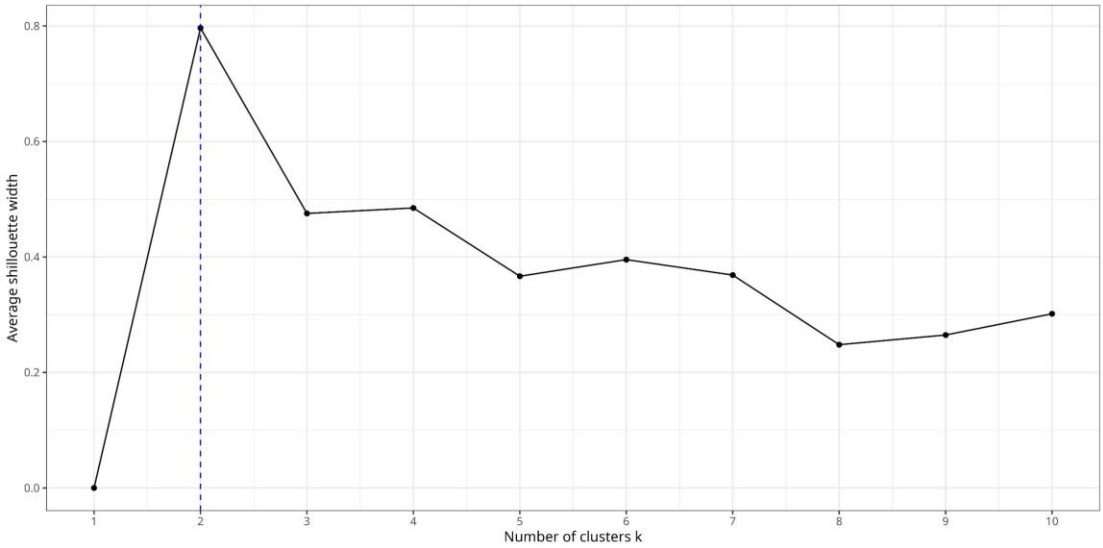

Figure 22. Graph showing the optimal number of clusters according to the average silhouette width.

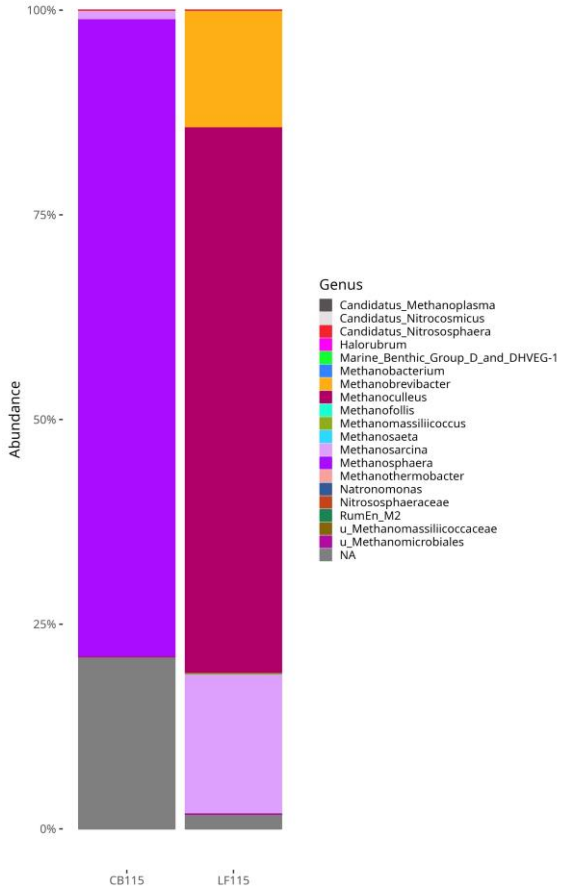

Figure 23. Relative abundance of binomial CB115-LF115. The high proportion of *Methanosphaera* can be seen in the neonatal stool sample (CB115).

Appendix F. DESeq

Eukaryotic synthetic construct chromosome 17

Sequence ID: CP034495.1 Length: 88299790 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 61742966 to 61743195 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
412 bits(223)	1e-110	228/230(99%)	1/230(0%)	Plus/Minus

```
Query  10        ACC-CCGGTAGTCCCCATGCCAGACTTTATCACATGCAGGTTGTCTTGAAAAACTGTCTT  68
                 ||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61743195  ACCTCCTGTAGTCCCCATGCCAGACTTTATCACATGCAGGTTGTCTTGAAAAACTGTCTT  61743136

Query  69        GCTCTAACGGGAAGTCAGAACCAAGGGGGAAGTGAATGTTCTTGCTGAAGCCTCCGATGT  128
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61743135  GCTCTAACGGGAAGTCAGAACCAAGGGGGAAGTGAATGTTCTTGCTGAAGCCTCCGATGT  61743076

Query  129       GCAATAGCTCTCTGCTAAGGGTCAAAAAGGAAACAACAGTCAAAGCCACATACTGGGGAG  188
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61743075  GCAATAGCTCTCTGCTAAGGGTCAAAAAGGAAACAACAGTCAAAGCCACATACTGGGGAG  61743016

Query  189       GCAGGCAAGTGAAGGGCAAGGGAGGGTCATAGTCTTGCACTCACAAAGAG  238
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61743015  GCAGGCAAGTGAAGGGCAAGGGAGGGTCATAGTCTTGCACTCACAAAGAG  61742966
```

Figure 24. Blastn results for the sequences of the differential expression analysis.

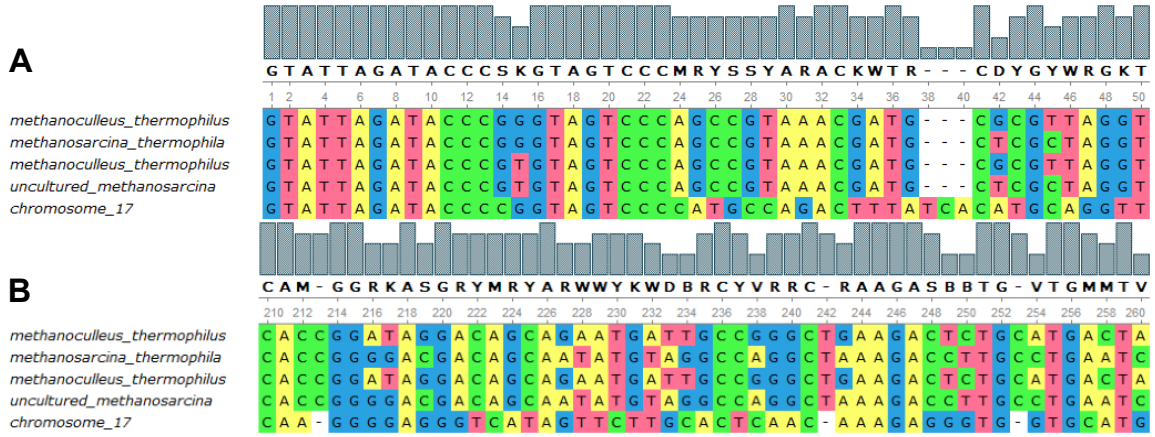

Figure 25. Alignment of Archaea amplicons against chromosome 17 amplicon. In A) Sequence segment from 1 to 50 nt., B) Sequence segment from 210-260 nt.