



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas
INSTITUTO DE BIOTECNOLOGÍA

“Análisis bioinformático del metagenoma fermentativo del pozol.”

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:
Rafael López Sánchez

TUTOR PRINCIPAL
Dr. Lorenzo P. Segovia Forcella
Instituto de Biotecnología UNAM

MIEMBROS DEL COMITÉ TUTOR

Dr. Agustín López Munguía Canales
Instituto de Biotecnología UNAM

Dr. Luis David Alcaraz
Facultad de Ciencias UNAM

Cuernavaca, Morelos, Diciembre 2018



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ÍNDICE

.....	1
Agradecimientos	2
Resumen.....	3
Introducción	5
Antecedentes	8
1. Metagenómica	8
2. Técnicas de secuenciación masiva para metagenómica	9
2.1 Métodos clásicos	9
2.2 Métodos de secuenciación de nueva generación	10
3. Estudios de Metagenómica en alimentos fermentados	12
4. El pozol	16
4.1 Preparación del pozol	18
4.2 Microbiota del pozol	20
5. CAZymes	25
Hipótesis	27
Objetivos	27
Metodología y material	28
1. Programas	28
2. Metodología	28
2.1 Muestreo	28
2.2 Extracción del DNA genómico	28
2.3 Amplificación por PCR	29
2.4 Taxonomía y α y β diversidad: Perfil taxonómico de amplicones de 16S rRNA	30
2.5 Taxonomía: Metagenoma Shotgun	32
2.6 Ensamblados	34
2.7 Anotación funcional: Metagenoma Shotgun	36
Resultados	38
1.1 Extracción del DNA genómico	38
1.2 Amplificación por PCR	39
1.3 Metagenoma de amplicones 16S	40
1.4 Esfuerzo de muestreo, α y β diversidad	41
2. Taxonomía	43
2.1 Amplicones 16S	43
2.2 Metagenoma tipo shotgun	50
2.3 Taxonomía de Amplicones vs Shotgun	61
2.4 Probióticos	69
3. Ensamblado	70
4. Anotación funcional	71
5. Reconstrucción Metabólica	77
Discusión	79
Conclusiones	85
Perspectivas	87
Referencias	88
Material suplementario	94

Agradecimientos

Al Conacyt por beca posgrado y a los donativos DGAPA PAPIIT IN223917 y IN206918, a la Dra. Carmen Wachter y al Dr. Agustín López Munguía Canales, al Dr. Lorenzo P. Segovia Forcella por permitirme formar parte del grupo de trabajo del Laboratorio 12 y del pozol.

A la Maestra en Ciencias Blanca Ramos Cerrillo por la parte técnica de este proyecto, así como a M.C Jerome Jean Verleyen por el apoyo bioinformático y el uso del cluster del Instituto de Biotecnología de la UNAM "Teopanzolco".

A la Unidad de Bioinformática del Instituto de Biotecnología de la UNAM, al Dr. Fidel Alejandro Sánchez Flores a las Dras. Grisel Alejandra Escobar Zepeda y Elizabeth Ernestina Godoy Lozano y la M.C. Verónica Jimenez Jacinto.

A la Unidad de Cómputo de la UNAM, conformada por Arturo Ocádiz Ramírez Juan Manuel Hurtado Ramírez, Servando Aguirre Cruz, Roberto Pablo Rodríguez Bahena y David Santiago Castañeda Carreón.

Resumen

El pozol es una bebida no alcohólica, prehispánica fermentada a base de masa de maíz nixtamalizado que se consume principalmente en el sureste de México y parte de Centroamérica. Es utilizado por los indígenas y la población en general como remedio medicinal, fuente de nutrientes ya que tiene aminoácidos esenciales y para usos ceremoniales en la cultura maya. Actualmente no se sabe mucho sobre la interacción que tienen los microorganismos presentes en esta fermentación, ni que sustratos son los que utilizan de la masa fermentada. Es por eso por lo que se realizó una aproximación metagenómica para obtener la anotación taxonómica y funcional de cuatro tiempos claves de la fermentación (0,9,24 y 48 horas) y así poder realizar una reconstrucción metabólica de la fermentación. Para la anotación taxonómica se utilizaron dos métodos, mediante el gen 16S ribosomal y el método de metagenómica shotgun. Para la anotación funcional se buscaron los genes que codificaban para enzimas activas de carbohidratos (CAZymes) en contra de las bases de datos de CAZY. La reconstrucción metabólica se realizó mediante la anotación funcional de KEGG y se resaltaron los módulos de aminoácidos esenciales que se encuentran en mayor proporción en el pozol que en la masa cruda. Se encontró que los géneros más abundantes en el pozol son *Streptococcus*, *Exiguobacterium*, *Anoxybacillus* y *Lactobacillus*. El género *Streptococcus* es abundante en todos los tiempos, pero empieza a tener más presencia a partir de las 9 horas de fermentación. *Exiguobacterium* permanece desde el inicio hasta el final de la fermentación de la misma manera, *Anoxybacillus* se encuentra al principio en gran proporción y a partir de las 9 horas empieza a decrecer en abundancia. *Lactobacillus* por su parte tiene mayor presencia a partir de las 24 horas de la fermentación. Se encontraron genes que codifican para enzimas que degradan almidón, hemicelulosa y sacarosa en los cuatro tiempos de la fermentación y se pudo realizar una reconstrucción metabólica del metagenoma en el cual se resaltaron de manera completa los módulos de treonina, lisina, isoleucina, triptofano, arginina y riboflavina en la fermentación del pozol.

Abstract

Pozol is a non-alcoholic, pre-Hispanic fermented beverage based on nixtamalized corn dough that is consumed mainly in southeastern Mexico and part of Central America. In the Maya culture, pozol is used as a medicinal remedy, source of nutrients and ceremonial uses. Currently, not much is known about the interaction of the microorganisms present in this fermentation, nor what substrates are those that use the fermented dough. That is why a metagenomic approach was used to obtain the taxonomic and functional annotation of four key times of the fermentation (0,9,24 and 48 hours) and thus be able to perform a metabolic reconstruction of the fermentation. For the taxonomic annotation, two methods were used, using the 16S ribosomal gene and the shotgun metagenomic method. For the functional annotation, the genes that codified for active carbohydrate enzymes (CAZymes) were searched against the CAZY databases. The metabolic reconstruction was carried out through the functional annotation of KEGG and the essential amino acid modules that are found in a greater proportion in the pozol than in the raw mass were highlighted. It was found that the most abundant genera in the pozol are *Streptococcus*, *Exiguobacterium*, *Anoxybacillus* and

Lactobacillus. *Streptococcus* is abundant at all times but begins to have more presence after 9 hours of fermentation. *Exiguobacterium* remains from the beginning until the end of fermentation in the same way, *Anoxybacillus* is found at first in large proportion and starting at 9 o'clock starts to decrease in abundance. *Lactobacillus* in turn has a greater presence after 24 hours of fermentation. We found genes that encode enzymes that degrade starch, hemicellulose and sucrose in all four times of fermentation and we were able to perform a metabolic reconstruction of the metagenome in which the modules of threonine, lysine, isoleucine, tryptophan, arginine and riboflavin were completely highlighted in the pozol fermentation.

Introducción.

Por miles de años los humanos han optimizado las condiciones para promover el crecimiento de cierto tipo de comunidades microbianas. La actividad metabólica de estas comunidades es la llave para la seguridad, el buen sabor, el aroma y la textura de los alimentos fermentados (Wolfe et al., 2015).

La fermentación es un proceso complejo, ya que los alimentos son transformados por medio de microorganismos del medio ambiente que utilizan los componentes de las materias primas de frutas, cereales, vegetales, carnes o lácteos, metabolizando principalmente los azúcares de estos alimentos y llevando a cabo la transformación de ácidos orgánicos como el acético, el láctico, el butírico y el propiónico, así como de algunos alcoholes como el etanol y la liberación de algunos aminoácidos (Wacher et al., 2014). Esto hace que se transforme su sabor, olor, color y textura, así como su valor nutrimental. Los alimentos fermentados hacen una contribución en la dieta humana debido a que la tecnología que requieren es muy barata y con ella se pueden preservar los alimentos además de mejorar sus propiedades sensoriales (Wolfe et al., 2015).

La manipulación del crecimiento microbiano varía de acuerdo con la temperatura, la salinidad y la humedad y da como resultado una amplia gama de alimentos fermentados como son los quesos, la cerveza, el vino, el chocolate, etc (Wolfe et al., 2015). Además, el bajo valor de pH y la presencia de fermentaciones ácidas significa que los productos fermentados son menos susceptibles a infección por envenenamiento de bacterias que los productos similares no fermentados (Cruz et al., 1973). También pueden contribuir a la destrucción y desintoxicación de factores indeseables presentes en los alimentos crudos como son los taninos y los polifenoles (Reddy et al., 2008). La generación de ácido láctico no sólo permite conservar los vegetales, sino que también les confiere un sabor agradable. Así, la fermentación vino a dar una muy importante variedad a la dieta, al tiempo que permitió la conservación de los alimentos (Wacher et al., 2014).

Para encontrar un rol productivo importante en las fermentaciones es esencial que las técnicas de producción estén libres de bacterias patógenas. Sin embargo, los métodos de preparación tradicional conducen a productos con altos niveles de contaminación a la hora de almacenarlos y procesarlos de manera artesanal (Cruz et al., 1973).

Por diversas circunstancias, no es sino hasta hace unas cuantas décadas que se analiza con profundidad el tipo de microorganismos responsables de lo que podríamos considerar los alimentos fermentados prehispánicos. Muchos de estos microorganismos, además de presentar actividades fisiológicas interesantes, podrían usarse como cultivos iniciadores para producir dichos alimentos en condiciones controladas (Wacher et al., 2014). La identificación debida de los microorganismos clave en las fermentaciones es un paso necesario para poder desarrollar una cultura mixta que permitiría estandarizar y manufacturar estos productos (Omar et al., 2000).

El análisis de microorganismos en los alimentos fermentados presenta una complejidad que requiere de estudios minuciosos para conocer las funciones y efectos que tienen sobre el

sustrato y, al mismo tiempo los cambios que producen y determinan otra fase de la sucesión de cada producto (Cruz et al., 1973). A pesar de eso, la experimentación utilizando comunidades microbianas de alimentos fermentados es relativamente fácil debido a que los ecosistemas bacterianos son muy accesibles. La producción de alimentos fermentados procede en intervalos regulares de tiempo y las comunidades microbianas se desarrollan en tiempos cortos (de días a meses), permitiendo al investigador la posibilidad de realizar muchas repeticiones de muestreo por un periodo de tiempo relativamente corto. Los alimentos fermentados son normalmente producidos en muchas regiones geográficas y también ha aumentado la accesibilidad a las muestras. Estas comunidades forman parte de entidades discretas lo que permite un muy bien definido muestreo tanto temporal como espacial. También permiten la oportunidad de cuantificar los roles relativos de estos procesos ecológicos porque la dinámica de las comunidades se puede medir fácilmente y se puede controlar (Wolfe et al., 2015).

Sin embargo, hay un gran número de barreras para entender los ecosistemas microbianos. La primera y más importante es la enorme diversidad y complejidad de la mayoría de las comunidades microbianas. La secuenciación basada en DNA ha sido actualmente aplicada a muchos hábitats y nos muestra una imagen de la diversidad microbiana de los distintos ambientes. A pesar de haber una variación sustancial entre los diferentes ecosistemas, el número de especies en la mayoría de los hábitats puede alcanzar un orden de cientos de millones (Wolfe et al., 2015). Muchas metodologías han sido empleadas para estudiar las fermentaciones (Cárdenas et al., 2014) La dificultad de aislar la gran mayoría de las especies de sistemas naturales mediante técnicas de cultivo es otra barrera que es significativa para caracterizar el rol de algunas especies en un ecosistema. Una restricción relacionada es la dificultad de recrear las condiciones experimentales en un procedimiento *in vitro* para recrear el proceso en condiciones controladas (Wolfe et al., 2015). Para ello se han utilizado tanto las técnicas de cultivo tradicionales y las independientes de cultivo (Cárdenas et al., 2014). Las técnicas independientes de cultivo han ayudado a cambiar la manera de estudiar la ecología microbiana, considerando a las poblaciones como un consorcio (De Filippis et al., 2017).

Debido a esto es que hoy en día las innovaciones en el campo de la genómica microbiana están haciendo posible explotar la riqueza genética y la variabilidad fenotípica existente entre los microorganismos de las fermentaciones de alimentos. Gracias a la reducción de los costos, la secuenciación de genomas completos de muchas cepas de las mismas especies es ahora una práctica común. Además, las herramientas bioinformáticas desarrolladas permiten recuperar genomas microbianos directamente de los metagenomas, permitiendo un proceso de monitoreo a nivel cepa durante el proceso y comparación genómico (De Filippis et al., 2017).

Los investigadores se están enriqueciendo con catálogos de diversidad bacteriana, fúngica y viral en muchos alimentos tradicionalmente fermentados. Algunos ejemplos de grupos bacterianos son las bacterias ácido-lácticas (LAB) y las bacterias acéticas ácido lácticas (AAB), así como algunas especies de hongos o levaduras como son *Saccharomyces cerevisiae*, quienes tienen roles bien establecidos en las fermentaciones (Wolfe et al., 2015).

A raíz de esto se han empezado a secuenciar muchas más muestras de fermentaciones, ya sea para mejorar su calidad, tener un mejor entendimiento de su hábitat natural, mejorar su seguridad o incluso encontrar las bacterias que le dan la denominación de origen. México no es

distinto al resto del mundo y ya ha empezado a secuenciar el DNA de sus fermentaciones. En casi todos los casos, los productos están fuertemente arraigados a la cultura de las regiones en las que fueron desarrollados o adaptados; en otros casos los alimentos se han beneficiado de los avances científicos y tecnológicos, como la cerveza, el vino o la producción de yogurt; en otros más, como los casos del pozol o del pulque en México, el proceso se mantiene con muy poca innovación (Wacher et al., 2014) (Giles-Gómez et al., 2016).

El pozol.

El pozol es una bebida prehispánica ácida fermentada de maíz que se produce en el sureste de México y algunas regiones de Centroamérica. Es rica en aminoácidos esenciales y su microbiota está formada principalmente por bacterias ácido-lácticas, hongos y levaduras. Es parte de la cultura maya y se le han atribuido muchas propiedades medicinales y también forma parte de la cultura y la dieta prehispánica ya que es utilizado en ceremonias de grupos indígenas (Cruz et al., 1973).

Poco se sabe de la expresión genética en fermentaciones complejas como la del pozol. La microbiología de esta fermentación espontánea ha sido extensivamente estudiada y ha sido caracterizada por la presencia de un significativo número de bacterias probióticas. Sin embargo, no sabemos mucho del rol de los microorganismos en la dinámica de esta fermentación ni cómo contienen a este ambiente en particular, así que la metagenómica puede darnos información para ayudarnos a entender este sistema complejo.

En este trabajo se pretende hacer un estudio de la metagenómica de la fermentación del pozol, obtener información sobre los genes relacionados a la degradación de carbohidratos y la reconstrucción metabólica de la fermentación.

El gran aumento en la población mundial no sólo demanda una mayor producción de carbohidratos, sino que, seguramente de manera más urgente, una mayor producción de proteínas (Cruz et al., 1973). En este sentido el pozol resulta interesante ya que presenta un alto contenido proteico, mayor que el de la materia prima que es el maíz con el que se elabora.

Es por medio de las técnicas de secuenciación masiva y el estudio bioinformático que se pretende complementar los innumerables estudios de cultivo que se han realizado para poder detectar microorganismos en ambientes tan complejos como el de la fermentación del pozol. Este análisis bioinformático será realizado a cuatro muestras en tiempos claves de la fermentación y se pretende obtener además del perfil metabólico, el perfil taxonómico y filogenético de las mismas por medio de las técnicas de análisis de secuencias de metagenómica del gen ribosomal 16S y metagenómica shotgun.

Antecedentes.

1. Metagenómica

La metagenómica es la aplicación de técnicas genómicas actuales para el estudio de las comunidades de microorganismos directamente de sus ambientes naturales, evitando la necesidad del aislamiento y el cultivo en laboratorio de especies individuales (Chen et al., 2005).

Debido al uso de métodos moleculares hoy en día no se requiere el cultivo de microorganismos, lo que ha mejorado nuestro conocimiento de la diversidad microbiana en comunidades complejas (Díaz-Ruiz et al., 2003). Los cuellos de botella que se generan por el cultivo de microorganismos pueden ser resueltos mediante el uso de métodos cultivo-independientes como la metagenómica, donde se extrae el DNA y se analiza de la muestra microbiana sin tener que previamente aislarla (Kunath et al., 2017).

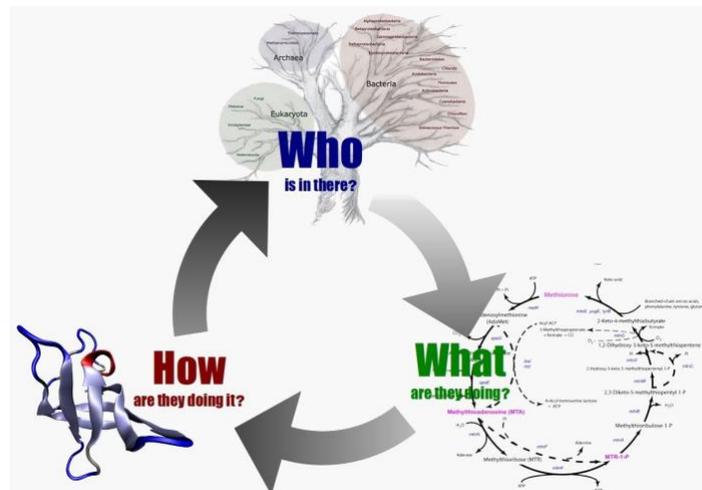


Figura 1. Esquema de la Metagenómica. Imagen obtenida de la página:

<https://www.slideshare.net/FAOoftheUN/presentacin-sobre-importancia-de-la-diversidad-biolgica-del-suelo>

En un segundo sentido la metagenómica busca el entendimiento de la biología en un nivel agregado, trascendiendo de un solo organismo al enfoque de los genes de una comunidad y su influencia en las actividades de los otros para la obtención de funciones colectivas (Handelsman et al., 2007).

El resultado del análisis metagenómico basado en la secuenciación de DNA es el potencial de las actividades de la comunidad microbiana, porque un gen en específico puede realmente no expresarse en esa condición debido a que el DNA puede surgir de células muertas o inactivas (De Filippis et al., 2017).

2. Técnicas de secuenciación masiva para metagenómica.

Se puede obtener mucha información con los métodos moleculares sobre las comunidades microbianas de los alimentos. Esta información incluye datos confiables sobre la diversidad y por lo tanto una identificación taxonómica de los microorganismos que se encuentran en cada etapa de la producción de alimentos, así como la cuantificación, la distribución y las correlaciones en crecimiento y actividad de estos (Díaz-Ruiz et al., 2003).

El coleccionar y procesar las muestras ambientales es la primera etapa para planear un proyecto metagenómico. Lo primero que se debe hacer es extraer y purificar las muestras de DNA obtenidas de los alimentos. (Díaz-Ruiz et al., 2003).

Además, se requieren otras consideraciones claves que incluyen el número de muestras necesario para ver adecuadamente la dinámica y la variabilidad en el ecosistema, así como los requerimientos del software bioinformático que necesita múltiples muestras (Kunath et al., 2017).

2.1 Métodos clásicos.

Comparados con métodos clásicos basados en el cultivo, los acercamientos metagenómicos están diseñados para dar una descripción rápida, más comprensible y complementaria de la diversidad microbiana asociada a los procesos de fermentación. La metagenómica se basa en los análisis de los ácidos nucleicos DNA y RNA asociados con la secuenciación masiva y las herramientas bioinformáticas. Sin embargo, otros métodos moleculares han sido utilizados para investigar la diversidad microbiana. Por ejemplo, el buscar regiones del DNA o el RNA y analizarlas para encontrar “huellas digitales” con métodos acoplados a la amplificación por PCR, como lo son Geles Desnaturalizados de Gradiente de Electroforesis (DGGE), Geles de Gradiente de Temperatura de Electroforesis (TGGE), entre otros. La técnica de Reacción en Cadena de la Polimerasa (PCR) consiste en amplificar algunas regiones del gen ribosomal procarionte 16S (16S rRNA) (Díaz-Ruiz et al., 2003). El gen 16S rRNA tiene una longitud aproximada de 1500 pb y contiene nueve regiones variables intercaladas entre las regiones conservadas. Las regiones variables de 16S rRNA son frecuentemente usadas para la clasificación filogenética de género o especie en diversas poblaciones microbianas (Cruz et al., 1973). Cómo elegir la región de interés para secuenciar es y ha sido un área de debate ya que ésta varía dependiendo de los objetivos experimentales, el diseño del experimento o el tipo de muestra (Illumina 2013). Cuando se realiza un estudio de la comunidad microbiana de la ecología de los alimentos, la amplificación de la PCR se lleva a cabo con cebadores cuyo objetivo son las regiones codificantes rRNA 16S para bacterias, 26S y 28S para levaduras y hongos (Bigot et al., 2015).

Para poder separar los fragmentos que fueron amplificados del DNA ribosomal (rDNA), se construyen bibliotecas de las clonas de rDNA 16S y es gracias a éstas que nos es posible determinar la diversidad microbiana de una muestra y así poder identificar o determinar cuáles

son los organismos conocidos más cercanos a los miembros de la comunidad mediante la comparación de los fragmentos de rDNA 16S con los de las bases de datos. El grupo de microorganismos nos dará una idea de qué cebadores son los adecuados (Díaz-Ruiz et al., 2003).

Para identificar a los organismos necesitamos tener un análisis de las cepas de referencia, o tener e identificar cepas representativas mediante la comparación de secuencias del 16S rRNA. Normalmente son utilizados los enfoques de taxonomía polifásica, además de otros métodos de clasificación y de identificación (Díaz-Ruiz et al., 2003).

Una de las limitaciones de esta técnica es que en la extracción de ácidos nucleicos de las muestras no está garantizado que todos los microorganismos estén presentes ni la recuperación de esos ácidos nucleicos intactos. También es de vital importancia purificar los ácidos nucleicos y así eliminar sustancias que puedan inhibir las reacciones de PCR o la acción de las enzimas de restricción (Díaz-Ruiz et al., 2003). Además, los métodos basados en 16S son vulnerables a la amplificación de los cebadores y esto representa un sesgo debido a que se puede dar una evaluación errónea de la abundancia de los miembros de la comunidad debido a la alta similitud correspondiente a las secuencias de 16S.

2.2 Métodos de secuenciación de nueva generación

Existen otros métodos utilizados en la Secuenciación de Nueva Generación (NGS), en los cuáles es posible identificar nuevas especies a partir de abundancia pequeña en datos de secuencias. Por ejemplo, el análisis metagenómico utilizando el método de shotgun, no involucra la clonación de los genes 16S rRNA y provee un acercamiento de cultivo independiente extremadamente importante debido a que supera los problemas relacionados a la identificación de las especies (Nalbantoglu et al., 2014).

La secuenciación shotgun está revolucionando nuestro entendimiento de los microbiomas asociados a humanos y otros ambientes, ya que junto con el análisis computacional están siendo usados para comparar perfiles funcionales y taxonómicos de comunidades microbianas (Nayfach et al., 2016).

La metagenómica tipo shotgun puede teóricamente generar secuencias llamadas lecturas de todos los genomas presentes en la muestra, lo que como colectivo ya hemos llamado metagenoma (Kunath et al., 2017). Estos conjuntos (sets) de datos consisten en millones de lecturas cortas (fragmentos genómicos) que idealmente representarían todos los microorganismos que habitan la muestra ambiental (Segata et al., 2014). A pesar de la complejidad y el reto de analizar estas bibliotecas metagenómicas se pueden utilizar para la cuantificación del taxa o de los genes para poder saber qué microorganismos están ahí y qué están haciendo (*Figura 1*) (Nayfach et al., 2016). La identificación de los organismos presentes y la abundancia de estos es usualmente el primer paso para conocer la biología de estas comunidades y se hace mediante anotación taxonómica. La anotación taxonómica es la operación computacional de inferir los clados taxonómicos que están viviendo en una comunidad microbiana y en qué proporciones (abundancia relativa) (Segata et al., 2014).

A continuación, y para obtener una visión completa del material genómico de una muestra ambiental, se necesitaría recuperar los genomas completos de cada individuo de la muestra. Claramente esto es idealista, pero mediante la utilización de técnicas de ensamblado *de novo*, las lecturas del metagenoma tipo shotgun pueden ser primeramente ensambladas y posteriormente analizadas (Segata et al., 2014). El ensamblado es un paso clave requerido para la generación de fragmentos contiguos de secuencia (contigs), que son necesarios para maximizar el número de marcos de lectura abiertos y operones río abajo del screening de las enzimas (Kunath et al., 2017).

Los acercamientos basados en ensamblado son particularmente buenos para microbiomas que incluyen una amplia proporción de microorganismos no secuenciados. (Segata et al., 2014).

Posteriormente se asignan los contigs a los genomas de organismos y esto nos sirve para poder interpretar sus interacciones. Las herramientas que se utilizan para la asignación de los contigs actúan como clasificadores y etiquetadores de éstos mediante una taxonomía existente, como lo puede ser la base taxonómica de NCBI (Kunath et al., 2017).

Un ejemplo de esto son los proyectos que se dedican al descubrimiento de CAZyme en microorganismos de alimentos. El objetivo es muy directo: la obtención de fragmentos largos y contiguos que son acoplados con la menor cantidad posible de malos ensamblados para que muchos genes completos y operones estén disponibles para screening (Kunath et al., 2017).

Es así como la metagenómica tipo shotgun es un método que ya ha mostrado un gran potencial para encontrar nuevas bacterias no cultivables y degradadoras de polisacáridos e interpretar las relaciones sinérgicas entre los filotipos no cultivables en una comunidad celulolítica (Kunath et al., 2017). Además, y a diferencia de los estudios del 16S, la secuenciación de tipo shotgun nos proporciona datos útiles no sólo para la perspectiva taxonómica, sino también para la diversidad metabólica y funcional de la población microbiana durante las etapas fermentativas (Sulaiman et al., 2014).

A pesar de los beneficios, el uso de la metagenómica tipo shotgun también tiene muchos inconvenientes (Kunath et al., 2017). En particular, la diversidad microbiana de las especies combinada con la inmensa cantidad de datos obtenidos con las plataformas de NGS. Estas plataformas producen gigabytes de datos que necesitan estar fuertemente anotados antes de ser interpretados de manera significativa. Aún para especies que tienen secuenciados sus genomas, las bases de datos no pueden capturar una colección completa de genes a través de las diferentes cepas (Nayfach et al., 2016).

Otro inconveniente es que la compatibilidad de los algoritmos es actualmente obstaculizada por el uso de estadísticas sobre la abundancia que no estiman un parámetro significativo de la comunidad microbiana. Detectar las diferencias en abundancia de los microorganismos individualmente o de sus genes requiere muestras muy grandes (Nayfach et al., 2016).

Típicamente la secuenciación tipo shotgun arroja resultados de una amplia cobertura de especies dominantes y menos lecturas de especies poco abundantes. Para acomodar esta discrepancia y tener acceso a miembros de la comunidad más raros, los esfuerzos de

secuenciación deben duplicarse; sin embargo, se tienen que considerar las limitaciones computacionales. Es por eso por lo que una evaluación de la estructura de la comunidad utilizando un análisis del gen amplicón 16S rRNA es recomendada para ayudar a determinar la complejidad, así como también la selección de la plataforma de NGS y la profundidad de la secuenciación para la obtención de los datasets necesarios para la búsqueda de genes. (Kunath et al., 2017).

3. Estudios de Metagenómica en alimentos fermentados.

Las técnicas de cultivo independiente han ayudado a cambiar la manera en la que estudiamos la ecología de los microorganismos que existen en los alimentos. El estudio de la ecología microbiana es relevante para la biotecnología y es la base no sólo del desarrollo de las fermentaciones, sino también de la comprensión de las interacciones microbianas que dan lugar a un proceso de alta calidad. Este entendimiento nos ofrece también la posibilidad de comprender las funciones potenciales que pueden ser cambiadas o moduladas, y por ende mejorar la calidad de los alimentos (De Filippis et al., 2017).

Los ejemplos de estudios metagenómicos en fermentaciones son variados y numerosos. Por mucho, son los lácteos los ambientes más estudiados (De Filippis et al., 2017). Un ejemplo de esto ha sido la caracterización taxonómica y funcional del microbioma del queso Cotija a un nivel altamente detallado que ayuda a explicar sus características sensoriales y de inocuidad (Escobar-Zepeda et al., 2016)

Otro ejemplo son los estudios de los tres quesos más importantes en Italia, que son el Mozzarella, el Grana Padano y el Parmigiano Reggiano (*Figura 2*). Aquí se analizaron las regiones V1 y V3 del gen 16 rRNA para poder identificar la microbiota involucrada en la acidificación de la cuajada. Gracias a esto tenemos el conocimiento de que los tres sueros de los quesos están caracterizados por una comunidad de bacterias ácido lácticas, en las cuales están incluidas tanto las mesofílicas, como las termofílicas y que comparten un núcleo similar de bacterias en su microbiota, el cual está compuesto de *Streptococcus thermophilus*, *Lactobacillus Delbrueckii*, *Lactobacillus Helveticus* y *Lactobacillus fermentum* (De Filippis et al., 2014).

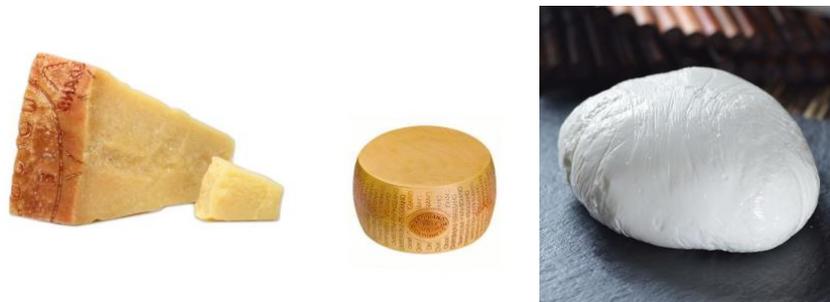


Figura 2. Quesos Grana Pandano, Parmigiano Reggiano y Mozzarella. Imágenes obtenidas de: <https://ronjamesitaliankitchen.blogspot.com/2016/05/italian-grana-padano-cheese-makers-sue.html>

En un estudio más general se secuenciaron 137 comunidades diferentes de la corteza de los quesos de 10 países distintos para llegar a la conclusión de que hay 24 géneros cultivables ampliamente distribuidos entre el reino bacteria y fungi, siendo los más abundantes *Staphylococcus* y *Debarymyces* respectivamente como miembros dominantes de la comunidad. También se pudo comprobar que la comunidad microbiana de la corteza del queso está extendida y es reproducible en condiciones de laboratorio (Wolfe et al., 2014).

Además, los ambientes relacionados con los alimentos albergan una microbiota beneficiosa para el ser humano involucrada en los lácteos (De Filippis et al., 2017). El kefir han sido objeto de estudio debido a ser un excelente probiótico. El kefir es una bebida tradicional obtenida de la fermentación de leche mediante la fermentación de “los granos de kefir” que son una mezcla compleja de bacterias, levaduras y polisacáridos producidos por esta microbiota y sus propiedades son propagadas a las siguientes generaciones para dar nuevos granos, lo cual convierte a esta bebida en un alimento funcional, debido a su gran valor nutricional. Mediante estudios de shotgun se ha podido identificar que es el género *Lactobacillus* el que domina esta fermentación y son tres las especies más abundantes: *Lactobacillus kefiranofaciens*, *Lactobacillus buchneri* y *Lactobacillus helveticus* (Nalbantoglu et al., 2014).

Como ya ha sido mencionado con anterioridad, mejorar la calidad de los productos alimenticios es algo que se pretende lograr con la metagenómica. En el caso de la soya tradicional china se realizaron estudios para determinar la comunidad microbiana de la salmuera tradicional y su rol metabólico en los tanques de fermentación mediante un estudio de shotgun durante seis meses de presentación. Esto con la finalidad de intentar entender cómo se desarrolla la fermentación sin involucrar un inóculo conocido y así evitar que la soya sea menos aromática. Esta microbiota fue dominada por el género bacteriano *Weisella* y posteriormente por el género fúngico *Candida* (Sulaiman et al., 2014).

Al igual que la salmuera de la soya china tradicional, el *nuruk* (Figura 3) es un iniciador natural. Se prepara cociendo los granos de cereal (principalmente el trigo) mediante una fermentación espontánea de una mezcla de microorganismos incorporados de manera natural. Este iniciador es extensamente utilizado en la fermentación del Makgeolli, una bebida muy popular coreana, a la cual se le pretende mejorar su calidad y en la que se pudieron detectar dos especies de hongos representativas en la fermentación que fueron los géneros *Aspergillus* y *mucorales*, mismos que no solamente influyen la fermentación, sino que potencian un sabor agradable en la bebida (Bal et al., 2016).



Figura 3. Pasteles de nuruk. Imagen obtenida de la página:
<https://en.wikipedia.org/wiki/Nuruk>

Así como en los casos de la soya y el trigo, el maíz es otro ejemplo de un cereal cuyas fermentaciones son de interés. En Africa la composición microbiana de alimentos tradicionales fermentados como son el potopoto de la República del Congo fue estudiada mediante la amplificación de la región V3 del gen 16S rRNA usando la técnica de Temperatura Temporal de Gel de Electroforesis (TTGE). El potopoto es una fermentación tradicional de maíz preparada en los hogares del Congo para su uso como alimento de destete. En esta fermentación se reveló la presencia de *Lactobacillus gasseri*, *Enterococcus sp.*, *Escherichia coli*, *Lactobacillus plantarum/paraplantarum*, *Lactobacillus acidophilus*, *Lactobacillus delbrueckii*, *Bacillus sp.*, *Lactobacillus reuteri* y *Lactobacillus casei*. (Abriouel et al., 2006).

Otro trabajo realizado en fermentaciones de maíz es la “chicha” (Figura 4), una bebida alcohólica producida en el noroeste de Argentina. En esta fermentación se identificaron 46 especies de bacterias ácido lácticas entre las cuales hubo predominancia de *Lactobacillus plantarum*, *Lactobacillus rossiae*, *Leuconostoc lactis* y *Weisella confusa* (Elizaquível et al., 2015).



Figura 4. Chicha Argentina. Imagen obtenida de la página:
<http://elnoti.com/tag/indigenas/>

Además de las ya mencionadas, otras preguntas se pueden resolver mediante la descripción de las comunidades microbianas durante las fermentaciones. Una de ellas es la identificación de biomarcadores para evitar el deterioro o contaminación de los alimentos (De Filippis et al., 2017). En el caso del poto-poto africano la detección de *E. coli* y *Bacillus cereus* es un indicador del riesgo asociado a la fermentación de este alimento (Abriouel et al., 2006).

En el caso del té Pu-erh (Figura 5) que es producido en Yunnan China, los estudios de secuenciación de amplicones rDNA no sólo identificaron las especies bacterianas y fúngicas, sino que encontró 25 metabolitos tóxicos procedentes principalmente de los hongos, lo cual sirve para la prevención en el consumo de esta bebida y así se aconseja a los consumidores hervir el agua con la cual harán el té (Zhang et al., 2016).



Figura 5. Te Pu-erh. Imagen obtenida de la página: <http://hojotea.com>

Otro ejemplo, la Ochratoxina A, es un metabolito secundario tóxico producido por el hongo del género *Aspergillus* y *Penicilium*, el cual ya ha sido demostrado en ratas que tiene propiedades carcinógenas e inmunotóxicas y que afecta a los granos de café de manera directa desde la cosecha y especialmente, en el proceso de almacenado. Por medio de un estudio de DGGE se mostró la evolución de la micoflora de los hongos del café en las distintas etapas de los tratamientos post-cosecha y se llegó a la conclusión de que los granos de café se suelen contaminar al caerse las bayas en el campo (Durand et al., 2013).

En todos estos ambientes de fermentaciones de alimentos los microorganismos están expuestos a una abundante mezcla compleja de carbohidratos como lo son la celulosa, xilano y fructano además de estar expuestos a diversos y abundantes sustratos incluyendo polisacáridos de plantas y glucosaminoglicanos animales encontrados en la comida y producidos por el huésped (Berlemont et al., 2016).

4. El pozol

En México existen numerosos grupos indígenas que utilizan diversos alimentos y bebidas fermentados con fines nutricionales, estimulantes, medicinales y rituales (Cruz et al., 1973).



Figura 6. Pozol listo para tomarse en una jícara. Imagen tomada de la siguiente página:
https://es.wikipedia.org/wiki/Pozol#/media/File:Tabasco_Pozol_en_j%C3%ADcara.jpg

Entre las bebidas y los alimentos fermentados autóctonos del país se encuentra el pozol (*Figura 6*) que es una bebida de maíz de origen maya, la cual forma parte de la alimentación básica de muchos grupos étnicos del sur y el sureste de México como son los chontales, los choles, los mayas, los lacandones, los tzotziles o chamulas, los tzetzales, los zoques, los mames y los zapotecos, así como la población mestiza de la región (Wacher et al., 2014).

Este alimento se prepara, tanto de forma doméstica para el consumo de la misma familia que lo hace, o en cantidades ligeramente mayores para ser vendido en el mercado (*Figura 7*), de una manera tradicional que ha sido pasada de generación en generación (Cruz et al., 1973).

El pozol es consumido por los campesinos en su jornada de trabajo varias veces durante el día, aumentando la frecuencia en las horas más calientes. Es un alimento fundamental principalmente para los campesinos de bajos recursos económicos. Un ejemplo de esto son los lacandones y chamulas, que portan consigo bolas de pozol como provisión en sus morrales, para sus jornadas por la selva. La cantidad de pozol consumido varía y va desde medio litro hasta dos litros o más por día. Algunas personas consumen hasta un kilo, o poco más de pozol (peso fresco) diluido en agua por día (Cruz et al., 1973).

Se han realizado estudios en los que se puede ver que el pozol contiene aminoácidos esenciales en mayor cantidad que la masa con la que fue fermentado, como los son el triptófano, la arginina, la lisina, la isoleucina, la treonina y la valina. También contiene más niacina y riboflavina (Cravioto et al., 1955).

Por otra parte, el pozol tiene asimismo usos medicinales y rituales (Wacher et al., 2014). Varios autores han publicado interesantes descripciones de tipo etnológico en relación con el pozol, indicando los usos alimenticios y ceremoniales (*Figura 8*) de este alimento por la cultura maya desde hace siglos, aún antes de la conquista española (Cruz et al., 1973).



Figura 7. Pozolería de Tabasco: Imagen tomada de la siguiente página:

https://es.wikipedia.org/wiki/Pozol#/media/File:Ilustraci%C3%B3n_sobre_pozol_tabasque%C3%B1o.jpg

Figura 8. El pozol en jícara, presente durante los rezos de un brujo lacandón. Fotografía del Instituto Nacional Indigenista.

En la actualidad todavía se efectúan ceremonias en las diferentes etapas de desarrollo en los cultivos de maíz, cuando está tierno, maduro y antes de la cosecha, siendo el pozol una de las ofrendas fundamentales en ellas (Cruz et al., 1973).

En ocasiones es utilizado para controlar la diarrea u otras afecciones intestinales, probablemente de manera semejante al uso que hacen otras personas de medicamentos o productos alimenticios que contienen levaduras o lactobacilos. En algunos registros históricos se indica el hecho de que las bolas de pozol enmohecidas eran utilizadas por los antiguos mayas como cataplasmas para prevenir o curar infecciones superficiales y heridas, lo cual se ha considerado una anticipación al uso de la penicilina, que se prescribió mucho tiempo después, con bases científicas (Herrera, 2006).

Otra propiedad curativa fue la que le encontraron los lacandones, ya que lo utilizan mezclado con miel para bajar la fiebre y controlar la diarrea (Wacher et al., 2014).

4.1 Preparación del pozol

La palabra pozol se deriva del náhuatl “pozolli”, que significa “espumoso”. Sin embargo, el producto resulta de una fermentación sólida, fundamentalmente láctica. Debido a eso no hay producción de gas ni de alcohol etílico como en el caso de las fermentaciones líquidas y por lo tanto el nombre se refiere de manera concreta a las bolas de masa fermentada y no a los efectos de la fermentación (Herrera, 2006).

El pozol se prepara de la siguiente manera:

Se le aplica una limpieza al maíz (que puede ser blanco, negro o amarillo como se ve en la *Figura 9*) en la cual se remueven los granos podridos y cualquier otro material ajeno al cereal. Se cubre con agua y se deja que repose unos minutos para posteriormente remover todo lo que este flotando. Una vez que se remueven los desechos sobrantes, se vacía el maíz en un canasto de manera tradicional. Posteriormente se remueven los granos malos que hayan quedado (Canas et al., 1993).



Figura 9. Distintos tipos de maíz. Fotografía obtenida de la Revista Biotecnología en Movimiento volumen 9 UNAM.

Una vez que se encuentra el maíz limpio se procede al proceso de nixtamalización; Este proceso consiste en hervir los granos de maíz durante más o menos una hora en unos dos litros de agua conteniendo cal (alrededor de 10% Ca(OH)_2 peso/volumen) que es posible adquirir del mercado local (Cruz et al., 1973) (Canas et al., 1993).



Figura 10. Proceso de nixtamalización del maíz. Fotografía obtenida de la Revista Biotecnología en Movimiento volumen 9 UNAM.

Esta cocción se lleva normalmente a cabo en hogueras y la cantidad de cal juega un papel muy importante, debido a que, si fuera insuficiente, daría dureza al maíz; Una vez que los granos de maíz se han hinchado y sus pericarpios se pueden quitar fácilmente con los dedos, los granos enfriados son frotados con las manos y enjuagados con agua varias veces para, una vez escurrido el exceso de agua, obtener el nixtamal (*Figura 10*) (Cruz et al., 1973) (Canas et al., 1993).

El nixtamal es lavado con agua potable y la finalidad de esto es la remoción de la cascarilla. Otra técnica consiste en agregar agua al nixtamal y hacer la transferencia de los granos a una cubeta con agua limpia.

Algunos componentes fibrosos no son completamente solubilizados por la nixtamalización y el sedimento está presente en la bebida cuando la masa es suspendida en agua. A la población mestiza no le agrada este sedimento y, probablemente por ello, modificaron el proceso indígena añadiendo una segunda cocción de los granos de nixtamal antes de molerlo (Cruz et al., 1973) (Canas et al., 1993) (Nuraida et al., 1995).

Posteriormente se procede a dejar remojando el producto durante toda la noche; en el caso del pozol indígena se hace en frío y se incluye el último lavado del nixtamal; en el caso del pozol mestizo se remoja en caliente con el agua del reventado. El remojo le permite a la masa el aumento y retención de la humedad para evitar que éste se seque durante el siguiente paso: la molienda (Canas et al., 1993).

Al día siguiente los granos son molidos en un molino comercial, de manera manual o con un metate o una piedra. Se considera que la calidad y las características sensoriales del pozol son capaces de variar de acuerdo con el tipo de molido que reciba la masa. Es sabido que durante este paso múltiples focos de contaminación microbiana se pueden observar y esto es debido a que los utensilios que se utilizan rara vez son desarmados y limpiados a fondo (Canas et al., 1993).

La masa es entonces moldeada en bolas que varían en forma y tamaño. Se dice que algunas de ellas llegan a pesar hasta un kilo. Este amasado se realiza de manera vigorosa en un tiempo

que varía entre 40 y 80 segundos y se hace con las dos manos sobre una mesa que es enjuagada previamente con agua (Cruz et al., 1973) (Canas et al., 1993).



Figura 11, Bolas de masa de pozol de maíz blanco, amarillo y azul. Imagen de Mari Carmen Wachter.

Las bolas amasadas (*Figura 11*) se envuelven tradicionalmente en hojas de plátano, de platanillo o de hoja blanca, con el objeto de reducir la desecación de la masa durante el tiempo que dure la fermentación, el cual varía de unas horas a cinco días o más; algunos grupos indígenas como los lacandones y los chamulas dejan fermentar el pozol dos semanas, pero el periodo puede incluso llegar a ser hasta de un mes y lo consumen enmohecido (Cruz et al., 1973) (Wachter et al., 2014).

Sin embargo, actualmente se prefiere la bolsa de plástico por la disminución en precio y volumen en la producción del pozol. También los consumidores indígenas han argumentado que la hoja retarda la acidificación del pozol (Cruz et al., 1973) (Canas et al., 1993).

La mayoría de los consumidores prefieren un pozol que haya sido preparado el mismo día. Este pozol no es muy ácido, pero el tiempo previo a su ingesta (aproximadamente 9 horas) es suficiente para que las bacterias ácido-lácticas que se encargan de la acidificación del producto tengan un máximo desarrollo (Canas et al., 1993).

Una vez fermentado, el pozol es diluido en agua en proporción de 1:2, aproximadamente, para obtener una bebida refrescante (Cruz et al., 1973).

4.2 Microbiota del pozol.

La dinámica de la comunidad microbiana responsable de la fermentación tradicional del maíz para la producción del pozol ha sido investigada por métodos cultivo dependientes principalmente (Omar et al., 2000) y de cultivo independientes mediante la secuenciación de los genes de DNA ribosomales 16S y mediante la extracción de proteínas totales para el estudio del metaproteoma (Cárdenas et al., 2014).

En los estudios realizados a esta bebida se nos indica que contiene en su mayoría microorganismos que confieren beneficios a la salud como son las ya mencionadas bacterias ácido-lácticas (LAB) y las bacterias amilolíticas ácido lácticas (ALAB), así como una comunidad de hongos y levaduras (Wachter et al., 1993).

Las LAB y las ALAB son un grupo de bacterias que están relacionadas en tener al ácido láctico como su mayor metabolito. Esa propiedad se les atribuye debido a que son capaces de producir ácido láctico a partir de los carbohidratos disponibles en una fermentación. En el caso de las ALAB una de sus fuentes de carbohidratos es el almidón (Reddy et al, 2008).

Estas bacterias se dividen en homofermentativas ya que sólo sintetizan ácido láctico a partir de los carbohidratos y heterofermentativas, si son capaces de producir ácido láctico y etanol a partir de los carbohidratos (Escobar-Zepeda et al., 2016).

Además, estas bacterias juegan un papel muy importante en la seguridad de la producción de compuestos antimicrobianos, que son muy importantes en la biopreservación del pozol (Reddy et al, 2008).

Tanto LAB como ALAB han sido reportadas en diferentes alimentos fermentados amiláceos preparados principalmente de cassava y cereales como lo son el sorgo y en este caso el maíz (Díaz-Ruiz et al., 2003). Su capacidad de mantener un gradiente de pH en la presencia de cantidades grandes de ácidos orgánicos les permite tener un papel importante en las fermentaciones en plantas (Ampe et al., 1999).

Las LAB han sido encontradas en todas las etapas del pozol. Los miembros del género *Leuconostoc*, *Streptococcus*, *Bacillus* y *Weisella*, fueron algunos de los organismos dominantes (Nuraida et al., 1995).

Una especie de *Streptococcus* dominó la fermentación y contabilizó del 25 al 75% de la microbiota total durante todo el proceso (Omar et al., 2000). Secuencias de clonas únicas fueron identificadas y atribuidas a *Streptococcus suis* por medio de estudios de 16S rRNA (Escalante et al. 2001) y *Streptococcus macedonicus* (Díaz-Ruiz et al., 2003). Otra cepa importante identificada fue *Streptococcus infantarius* 25124 (Cooper, en preparación) que, primeramente, se identificó como *Streptococcus bovis* 25124 (Díaz-Ruiz et al., 2003).

A continuación, se incluye la Tabla 1 con las cepas reportadas en los estudios de pozol:

Especie	Subgrupo	Especie	Subgrupo
<i>Acetobacter aceti</i>	Alphaproteobacteria	<i>Lactobacillus paracasei</i>	Firmicutes
<i>Achromobacter pozolis</i>	Betaproteobacteria	<i>Lactobacillus pentosus</i>	Firmicutes
<i>Agrobacterium azotophilum</i>	Alphaproteobacteria	<i>Lactobacillus plantarum</i>	Firmicutes
<i>Bacillus cereus</i>	Firmicutes	<i>Lactobacillus reuteri</i>	Firmicutes
<i>Bifidobacterium minimum</i>	Actinobacteria	<i>Lactococcus lactis</i>	Firmicutes
<i>Enterobacter aerogenes (aerobacter)</i>	Gammaproteobacteria	<i>Lactococcus raffinolactis</i>	Firmicutes
<i>Enterococcus saccharolyticus</i>	Firmicutes	<i>Lactobacillus zeae</i>	Firmicutes
<i>Enterococcus sulfureus</i>	Firmicutes	<i>Leuconostoc citreum</i>	Firmicutes
<i>Escherichia coli</i>	Gammaproteobacteria	<i>Leuconostoc mesenteroides</i>	Firmicutes
<i>Exiguobacterium acetylicum</i>	Firmicutes	<i>Streptococcus bovis</i>	Firmicutes
<i>Exiguobacterium aurantiacum</i>	Firmicutes	<i>Streptococcus infantarius</i>	Firmicutes
<i>Lactobacillus alimentarius</i>	Firmicutes	<i>Streptococcus intestinalis</i>	Firmicutes
<i>Lactobacillus casei</i>	Firmicutes	<i>Streptococcus macedonicus</i>	Firmicutes
<i>Lactobacillus confusus</i>	Firmicutes	<i>Streptococcus suis</i>	Firmicutes
<i>Lactobacillus delbrueckii</i>	Firmicutes	<i>Weissella confusa</i>	Firmicutes
<i>Lactobacillus fermentum</i>	Firmicutes	<i>Weissella paramesenteroides</i>	Firmicutes

Tabla 1. Bacterias cultivables reportadas en el pozol.

Los primeros estudios sobre la microbiota del pozol fueron reportados en 1958 por Salinas donde se registró una especie de bacterias aisladas en este alimento: *Bacillus cereus* (Cruz et al., 1973).

Luego en los 90s se pudo ver que en el remojo de los granos que recibieron un tratamiento alcalino durante la noche, prolifera el crecimiento de bacterias ácido-lácticas, las mesófilas aerobias y las *Enterobacteriaceae* y son éstas las que constituyen la primera microbiota de la masa de pozol. El molido en un molino comercial inocula la masa con las bacterias ácido-lácticas, las mesófilas aerobias, las *Enterobacteriaceae* y las levaduras (Wacher et al., 1993).

La presencia de *Bifidobacterium*, *Enterococcus* y *enterobacteria* sugiere el origen de algunos microorganismos importantes en el pozol (Omar et al., 2000).

A medida que la fermentación se va desarrollando se ha reportado que la mayor concentración de bacterias ácido-lácticas es a las 9 horas (Nuraida et al., 1995).

Posteriormente y mediante el análisis de los patrones de Electroforesis en gel con gradiente de desnaturalización (DGGE) obtenidos con los cebadores de bacterias y eucariotes orientados a genes 16S y 18S rDNA claramente se demostró que hubo un cambio mayor en la estructura de la comunidad después de las 24 horas y que una alta biodiversidad se mantuvo a través de todo el proceso (Omar et al., 2000).

Estos estudios mostraron que la microbiota aerobia fue progresivamente reemplazada de bacterias ácido-lácticas heterofermentativas por LAB homofermentativas como lo son *Lactobacillus plantarum*, *Lactobacillus casei* y *Lactobacillus delbrueckii* quienes fueron las encargadas de continuar con la acidificación de la masa y de acuerdo con el índice de Shannon-Weaver obtenido, una alta biodiversidad se mantuvo durante todo el proceso. Después se mantuvo la microbiota en las horas subsecuentes (Omar et al., 2000).

En el caso del maíz del pozol la definición de los sustratos utilizados por la microbiota presente es un problema aún sin resolver, debido a que la frecuencia con la que ciertas especies o géneros han sido reportados en el pozol podría estar relacionada con el tipo de sustrato y las condiciones ambientales, ya que gracias a la nixtamalización se remueve la gran mayoría de carbohidratos solubles y hace que se presenten altos niveles de almidón soluble que parece ser una buena fuente de carbono (Ampe et al., 1999).

En la primera etapa de la fermentación, sólo unos pocos microorganismos fueron capaces de utilizar el almidón. Sin embargo, el desarrollo de las ALAB soltó amilasas que degradaron el almidón en dextrinas y maltosas volviéndolo así disponible para un amplio número de cepas incluyendo las no amilolíticas.

En contraste, un número relativamente elevado de bacterias no amilolíticas ácido lácticas fueron observadas al final de la fermentación y eso podría explicar un uso eficiente de mono- y disacáridos presentes inicialmente y otros azúcares disponibles mediante la hidrólisis del almidón o la presencia de otros carbohidratos (Díaz-Ruiz et al., 2003).

Es por esa razón que mucha de la investigación que se ha hecho en la fermentación del pozol se ha centrado en el almidón como fuente de energía. Pero esto no explica del todo la presencia de muchas bacterias a través de todo el proceso de fermentado. La hemicelulosa, que es un compuesto soluble en soluciones alcalinas (en el caso del maíz arabinosilano) debería estar

más disponible debido al tratamiento de nixtamalización y podría ser otra fuente de carbono de las LAB presentes en la fermentación (Cooper, en preparación).

Sin embargo, la información relativa a la actividad de xilanasas por las LAB es escasa. Se ha detectado en las cepas de *S. infantarius* que logra crecer en el arabinoxilano del nejayote, que es el agua residual de la nixtamalización. Esto sugiere que *S. infantarius* 25124 posee la habilidad de adaptarse a un ambiente hemicelulolítico y puede ser la razón por la cual ha sido reportada su prevalencia durante toda la fermentación (Cooper, en preparación).

También la presencia de otra cepa, *Weissella confusa* L9, especie no amilolítica, ha sido reportada. Dicha especie crece muy bien en la masa del nixtamal y es capaz de utilizar el xilano, pero sobre todo los xilooligosacáridos (López-Hernández et al., en preparación).

En este contexto, se observó que tanto *S. Infantarius* 25124 como *W. Confusa* L9 son capaces de crecer en xilano como única fuente de carbono gracias a una actividad xilanolítica que ya fue demostrada. (Cooper et al., en preparación) (López-Hernández et al., en preparación).

De todo lo anterior se desprende que los microbios responsables de la degradación continúan sin ser completamente determinados. Es por eso que se concluye que la fermentación no es exclusivamente dependiente de azúcares libres y el almidón y los arabinoxilanos también contribuyen a la fermentación de este producto tradicional (Cooper et al., en preparación).

Una de las características benéficas de algunos alimentos es la presencia de microorganismos que se integran a la microbiota intestinal. A estos microorganismos, generalmente bacterias, se les denomina microorganismos probióticos. Se trata de sustancias químicas no metabolizables por el cuerpo humano y se ha determinado que si se consumen en concentraciones grandes tienen un efecto positivo en la microbiota intestinal del consumidor (Wacher et al., 2014). Se cree que la microbiota del pozol puede tener un gran potencial para la producción de prebióticos.

En el pozol la bacteria ácido-láctica *Leuconostoc citreum* constituye un buen ejemplo de probiótico. Esta bacteria se ha asociado a la producción de un polímero de glucosa denominado dextrana. También, se encontró que esta cepa es especial en el sentido de que produce igualmente fructanas, polímeros de fructosa que son consumidos por bacterias del intestino (Wacher et al., 2014). Una característica interesante de estas enzimas es su capacidad de catalizar la síntesis de oligosacáridos de bajo peso molecular a partir de la sacarosa, cuando moléculas aceptores eficientes, como es el caso de la maltosa, son agregadas al medio. Algunos de estos oligosacáridos son usados como prebióticos en aplicaciones cosméticas y alimentarias (Olivares-Illana et. al. 2002).

Las bacterias del género *Bifidobacterium* son bacterias Gram-positivas que habitan el tracto intestinal tanto de humanos como de animales y se usan comúnmente en la producción de productos prebióticos. Este tipo de bacteria además es capaz de competir contra cepas bacterianas que causan la diarrea (Handelsman et al., 2007). Esta cepa también ha sido detectada en el pozol (Ampe et. al 1999).

Otro ejemplo de esto son las ya mencionadas *Weisella confusa* L9 y *Streptococcus infantarius* 25214 capaces de crecer en la hemicelulosa del maíz (Cooper et al., en preparación) (López-Hernández et al., en preparación). La celulosa y la hemicelulosa son componentes importantes en cereales debido a su naturaleza fibrosa; el consumo de estos compuestos por los humanos se considera saludable debido a su carácter de fibra insoluble y a su actividad prebiótica. Estos compuestos (la fibra de la dieta) se encuentran principalmente en las capas externas de los cereales. Las xilanasas también han sido involucradas en la producción de xilooligosacáridos (XOs) y arabinooligosacáridos (AXOs) y tienen un potencial efecto benéfico en humanos como prebióticos y fibra soluble en la dieta (Cooper et al., en preparación) (López-Hernández et al., en preparación).

5. CAZymes

La degradación completa enzimática de polisacáridos involucra muchas enzimas activas (CAZymes) incluidas las GH, las polisacáridoliasas, las carbohidratoesterasas, entre otras. Las GH son importantes en las fermentaciones de alimentos ya que rompen los enlaces de los polisacáridos (ejemplo el almidón) y oligosacáridos y se obtienen productos metabolizables más pequeños (Berlemont et al., 2016).

Las enzimas de Carbohidratos-activos (CAZyme por sus siglas en inglés) son las responsables de la síntesis, degradación y modificación de todos los carbohidratos presentes en la Tierra y están clasificados como sigue:

- Glicosil hidrolasas (GHs por sus siglas en inglés). Son enzimas que catalizan la hidrólisis de las uniones glicosídicas de glicósidos, que dan la formación de azúcares hemicetales o hemiacetales. Estas pueden catalizar la hidrólisis de O-, N- y S-uniones glicosídicas. Se clasifican en *exo-* y *endo-* dependiendo de la hidrólisis glicosídica que tengan de romper un sustrato. (http://www.cazypedia.org/index.php/Glycoside_hydrolases).
- Glicosil transferasas (GTs por sus siglas en inglés). Son enzimas que catalizan la formación de uniones glicosídicas para formar un glucósido. Estas enzimas utilizan azúcares activados como donadores glicosídicos y catalizan la transferencia de grupos glicosílicos a un grupo nucleofílico, usualmente un alcohol. El producto de transferencia puede ser un O-, N-, S-, o C-glicosido y este puede ser monosacárido, oligosacárido o polisacárido. (<http://www.cazypedia.org/index.php/Glycosyltransferases>).
- Liasas de polisacáridos (PLs por sus siglas en inglés). Son un grupo de enzimas que hacen una escisión no hidrolítica en las cadenas de polisacáridos que contienen ácido urónico, vía los mecanismos de β -eliminación para generar un residuo de ácido hexenurónico insaturado y un nuevo extremo reductor.
- (<http://www.cazy.org/Polysaccharide-Lyases.html>).
- Esterasas de carbohidratos (CEs por sus siglas en inglés). Enzimas que catalizan la acilación de -O o de -N en sacáridos sustituidos. Tienen dos clases de sustratos; unos en los que el azúcar interpreta el papel de ácido como lo son los ésteres de pectina y

otros en los que el azúcar se comporta como alcohol como son los xilanos acetilados. (<http://www.cazy.org/Carbohydrate-Esterases.html>).

- Módulos no catalíticos de unión a carbohidratos (CBMs por sus siglas en inglés). Está definida como una secuencia contigua de aminoácidos con una enzima activa de carbohidratos con un doblez discreto que tiene actividad de unión a carbohidratos. (<http://www.cazy.org/Carbohydrate-Binding-Modules.html>).
- Actividades Auxiliares (AAs por sus siglas en inglés). Enzimas redox que actúan en conjunto con las CAZymes . (<http://www.cazy.org/Auxiliary-Activities.html>).

A pesar de no existir formalmente un estudio en alimentos fermentados, se realizó un enfoque metagenómico para estudiar las glicosilhidrolasas (GH) en 13 ecosistemas mundiales. La identificación de dominios específicos en GH en genomas y metagenomas secuenciados permite la predicción del potencial de descomposición de almidón, xilano, fructano, celulosa, quitina y dextrano siendo los dos primeros materias primas en muchas fermentaciones (Berlemont et al., 2016).

Hipótesis

A través de un análisis metagenómico de la dinámica poblacional a través del tiempo de los microorganismos involucrados en la fermentación del pozol se pueden determinar las vías metabólicas de carbohidratos.

Objetivos

General:

Realizar un estudio metagenómico de la dinámica poblacional de la fermentación del pozol.

Particulares:

- Analizar la composición taxonómica del consorcio y su dinámica en el transcurso de la fermentación.
- Identificar, a través de una aproximación metagenómica, el metabolismo de los carbohidratos disponibles en la masa del pozol, así como un modelo de su fermentación.
- Analizar la fermentación mediante el estudio bioinformático de las interacciones metabólicas que se establecen entre los microorganismos presentes en ella.

Metodología y material.

1. Programas

BLASTp (Johnson et al., 2008)
BLASTx (Johnson et al., 2008)
BWA-v0.7.16a (Li et al., 2013)
dbCAN (Yin et al., 2012)
FLASH-v1.2.11 (Magoč et al., 2011)
GeneMark-v3.26 (Zhu et al., 2010)
GhostKoala-v2.0 (Kanehisa et al., 2016)
HMMER-v2.3.2 (Jeffryes et al., 2011)
IDBA-UD -v1.1.3 (Peng et al., 2012)
Jellyfish-v1.1.11 (Marçais et al., 2011)
Kraken-v1.0 (Wood et al., 2014)
MEGAHIT-v1.1.1 (Li et al., 2015)
METAXA2-v2.1.3 (Bengtsson-Palme et al., 2015)
MegaBLAST (Morgulis et al., 2008)
MetaGeneMark-v1 (Zhu et al., 2010)
Parallel-Meta2.0-v3.3.2 (Su et al., 2012)
R-v3.4.1
Paquetes:

- Vegan (Dixon 2013)
- MetagenomeSeq (Paulson et al. 2013)
- Phyloseq (McMurdie et al. 2013)
- MASS (Tang et al., 2013)

SAMtools-v1.8 (Heng et al., 2009)
VSEARCH-v2.4.3 (Rognes et al., 2016)

2. Metodología

2.1 Muestreo.

La muestra fue obtenida del mercado de Villahermosa, Tabasco. La recolecta de la muestra fue el 15 de agosto del 2016. La muestra se extendió en bolsas de cierre hermético y se almacenó a -70°C. De ahí se pusieron a fermentar las bolas de pozol a distintos tiempos (0, 9, 24 y 48h) en hojas de plátano previamente esterilizado con fuego y etanol. La fermentación se desarrolla a 37°C en cámara húmeda hechiza.

2.2 Extracción del DNA genómico.

Se extrajo el DNA de las muestras de pozol a distintos tiempos de fermentación, los cuales fueron 0 horas, 9 horas, 24 horas y 48 horas de la muestra designada como “bola 2” (que estaba a -70 grados centígrados) por medio de los siguientes kits de extracción de DNA genómico: PowerSoil® DNA Isolation Kit catálogo No. 12888-50, PowerMax® Soil DNA Isolation Kit catálogo No. 12988-10 y, UltraClean® Microbial DNA Isolation Kit catálogo No. 12224-250 todos

obtenidos de la compañía Mo Bio Laboratories, Inc., de acuerdo con los protocolos del fabricante de cada kit . Cada muestra pesó aproximadamente 5 gramos.

Se obtuvieron las siguientes cuantificaciones utilizando el método de cuantificación de Qubit y se corrió también un gel al 1% de agarosa para comprobar la integridad del DNA. El DNA metagenómico se utilizó para generar una secuenciación de tipo whole metagenome shotgun. Para la preparación de la librería, se utilizó el protocolo TruSeq DNA PCR-Free HT Library Prep siguiendo las instrucciones del fabricante. La secuenciación se llevó a cabo en un secuenciador Next Seq500 de Illumina, con una configuración de lecturas de extremo pareado de XXX ciclos, en la Unidad Universitaria de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la UNAM

2.3 Amplificación por PCR.

Se purificó el DNA de una muestra de pozol y se amplificó una alícuota con cebadores específicos para la región V3-V4 del gen ribosomal 16S.

Se amplificaron las muestras de gDNA de los diferentes tiempos de fermentación por PCR utilizando la polimerasa Thermo Scientific Phusion Flash High-Fidelity PCR Máster Mix. El protocolo que se siguió es el adecuado para la secuenciación por medio del sistema Illumina MiSeq y consiste en lo siguiente:

Ordenar los cebadores de los amplicones. Este protocolo incluye los cebadores a utilizar que son los siguientes (Klindworth et al.,2013) :

16S Amplicon PCR Forward Primer = 5'
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG.
(Illumina, 2013).

16S Amplicon PCR Reverse Primer = 5'
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC.
(Illumina, 2013).

También incluye las secuencias cohesivas de los adaptadores que tienen que ser compatibles con el método de Illumina.

Posteriormente se preparan las bibliotecas mediante la amplificación de las regiones V3-V4 utilizando 30 ciclos de PCR. Se añaden los adaptadores y los códigos de barra de los objetivos de los amplicones.

Se pueden obtener hasta 96 bibliotecas para secuenciación utilizando los índices Nextera XT. El PCR amplificó un fragmento mayoritario de aproximadamente 500 pb en todas las horas de fermentación, el cual es muy cercano al reportado en la literatura (460 pb) (Illumina, 2013).

Las bibliotecas son secuenciadas por medio de MiSeq-Using® y se obtienen lecturas de 300 pbs pareadas.

2.4 Taxonomía y α y β diversidad: Perfil taxonómico de amplicones de 16S rRNA.

Una vez obtenidas las bibliotecas, los amplicones se reconstruyeron por medio del par de secuencias pareadas por el método de secuenciación de amplicones de cada uno de los 4 tiempos de fermentación. Se realizó un análisis taxonómico para determinar el perfil y composición filogenéticos de esta población por secuenciación masiva. Se realizó la asignación taxonómica, así como la α y β diversidad. El primer paso fue hacer una revisión de las secuencias obtenidas por el método de 16S previamente mencionado.

Una vez revisada la calidad de secuenciación se procedió a hacer la reconstrucción de los amplicones originales. Debido a que la secuenciación está en un formato pareado de 2X300 pares de bases para obtener los amplicones de aproximadamente 460 pares de bases se realiza un paso de la extensión de los fragmentos utilizando un programa llamado Fast Length Adjustment of Short (FLASH reads por sus siglas en inglés). (Magóc et al. 2011).

El uso de bibliotecas pareadas con un tamaño de fragmento menor al doble de la longitud de la secuencia brinda la oportunidad de generar lecturas mucho más largas mediante el solapamiento y la fusión de los pares de lecturas antes de ensamblar un genoma (Magóc et al. 2011).

FLASH es una herramienta computacional que extiende la longitud de las lecturas cortas mediante el solapamiento de lecturas pareadas de los fragmentos de las bibliotecas que son lo suficientemente cortos (Magóc et al. 2011).

FLASH requiere como un archivo fastq de lecturas pareadas en las que al menos algunas de las lecturas solapen la lectura generada por el extremo opuesto del mismo fragmento de DNA (Magóc et al. 2011). Se procesa cada par de lecturas por separado y se busca el solapamiento correcto entre las lecturas pareadas.

Cuando el solapamiento correcto es encontrado, las dos lecturas se fusionan, produciendo una lectura extendida que coincide con la longitud correcta del fragmento original de DNA del cual las lecturas pareadas fueron generadas.

Para encontrar el solapamiento correcto FLASH considera todas las posibilidades legales de solapamiento entre las lecturas pareadas, donde un solapamiento legal es definido como cualquier alineamiento sin gaps entre las dos lecturas tal que por lo menos un min-olap de bases solapen una sobre otra (Magóc et al. 2011).

Decidimos permitir solo alineamientos sin gaps debido a que con la secuenciación por las plataformas de Illumina, las inserciones y deleciones son bastante raras. FLASH puntúa todo solapamiento legal posible entre las dos lecturas pareadas. Por default los parámetros de min-olap son 10 pares de bases (pbs) y el max-olap es de 70 pbs.

El programa FLASH obtiene como archivo de salida un archivo fastq que contiene lecturas extendidas, donde cada base en la región del solapamiento se le asigna un valor de calidad en la base con la más alta calidad donde las lecturas coinciden, y un score de 0 a 255.

Los parámetros utilizados por el programa fueron los siguientes: un tamaño de lectura de 300 pbs, tamaño del fragmento de 450 pbs y una desviación estándar de ± 30 .

Las lecturas que no se logran extender se eliminan de los pasos subsecuentes como un método de control de calidad.

Una vez obtenidas las secuencias extendidas se procede a utilizar los archivos que obtuvimos de salida como input para hacer la desreplicación de las secuencias. Los archivos extendidos se cambian del formato fastq a fasta para poder ser procesados con el programa VSEARCH (Rognes et al., 2016).

VSEARCH incluye la búsqueda y clusterización por similitud utilizando un preordenado por longitud y abundancia. También hace detección de quimeras *de novo* o en base a referencias, desreplicación completa o de prefijos, alineamientos pareados, complementación reversa, ordenamiento y submuestra. (Rognes et al., 2016). A los archivos obtenidos libres de réplicas se les eliminan las quimeras y posteriormente concatenamos el resultado para la generación de OTUs con un porcentaje de identidad del 97%.

Una vez obtenidos los clusters se filtran los singletons (secuencias únicas) de la matriz resultante y se obtienen los parámetros de α y β diversidad.

Para esto utilizamos los paquetes estadísticos de R: VEGAN (Dixon, 2003), metagenomeSeq (Paulson et al. 2013), Phyloseq (McMurdie et al. 2013) y MASS (Tang et al., 2013), que nos proporcionan la tabla con los índices de Chao1, Shannon-Weaver y Simpson, la gráfica con el esfuerzo de muestreo para la α diversidad y posteriormente se realizó una gráfica de NMDS (Plot de Non-metric multidimensional scaling por sus siglas en inglés) para hacer una representación de la beta diversidad.

Para realizar la anotación taxonómica se parte de las secuencias reconstruidas por el programa FLASH y se corre una versión modificada del programa Parallel-meta 2.0 donde se obtuvieron las abundancias relativas utilizando Parallel-Meta 2.0 con la base de datos de METAXA2 (Escobar-Zepeda et al. 2018).

El pipeline de parallel-meta 2.0 incluye 4 pasos: la extracción de 16S rRNA para la predicción de los fragmentos de rRNA de las secuencias metagenómicas por HMM (Hidden Markov Model) en la búsqueda de HMMER y luego la extracción de esos fragmentos basados en los resultados de predicción. La parte del mapeo de 16S rRNA que consiste en mapear los fragmentos extraídos en el último paso en la base de datos de METAXA2 utilizando megaBLAST como la herramienta de alineamiento para la identificación (Su et al., 2012). El objetivo de HMMER es usar métodos avanzados probabilísticos para la detección de homología de secuencias disponibles con una amplia gama de útiles herramientas.

La base de datos de METAXA2 se construye con secuencias de 16S rRNA curadas manualmente de las bases de datos SILVA y MITOZOA verificadas por varios modos con datos de Greengenes, CRW y GenBank. De todas estas bases de datos todas las secuencias son recuperadas y su información taxonómica es evaluada. Para la base de datos curada, las secuencias del gen 16S rRNA sin una completa afiliación taxonómica a nivel de especie son excluidas junto con las secuencias que pueden ser quiméricas. Esto también significa que las secuencias de organismos no cultivables o de proyectos metagenómicos no son incluidas en esta clasificación de la base de datos. Finalmente, las secuencias que tienen información contradictoria en la base de datos curada también son excluidas. (Bengtsson-Palme et al., 2015)

La parte de clasificación que consiste en clasificar los fragmentos de 16S rRNA basándose en los términos de asignación taxonómica y mapeando cada muestra a su árbol filogenético.

Para ello utilizamos MegaBLAST que permite la búsqueda de alineamiento de secuencias de nucleótidos. Lo hace mediante la búsqueda de una estructura de índice de k-meros derivada del preprocesamiento de una base de datos. El módulo de BLAST MegaBLAST hace inicialmente la fase de encontrar pequeñas secuencias semillas para coincidencias mediante la búsqueda en la base de datos de índices. (Morgulis et al., 2008).

La comparación de muestras múltiples consiste en la comparación de la estructura taxonómica de todas las muestras en diferentes niveles biológicos. Después de todos estos pasos Parallel-META 2.0 reporta la clasificación taxonómica, la longitud de distribución, el resumen de las asignaciones de los fragmentos de secuencias del gen 16S rRNA y la estructura de los diferentes niveles filogenéticos en todas las muestras. (Su et al., 2012). Se utiliza un E-value de $1e-30$ en todos los tiempos.

2.5 Taxonomía: Metagenoma Shotgun.

Este mismo DNA también se usó para realizar un metagenoma "Shotgun". El metagenoma tipo Shotgun fue realizado utilizando el procedimiento de TruSeq® DNA PCR-Free Library Prep para la construcción de las bibliotecas. Dicho procedimiento consiste en la fragmentación del DNA en 75 pb: el cizallamiento genera fragmentos de dsDNA 3' o 5' con extremos cohesivos. Estos fragmentos son reparados por medio de una polimerasa cuya actividad de exonucleasa remueve esos extremos romos 3' y la actividad de polimerasa 5' 3' rellena los extremos romos 5' (Illumina, 2015).

En cuanto los extremos han sido ya reparados se selecciona el tamaño apropiado de las bibliotecas. Posteriormente los extremos cohesivos 3' son adenilados por un solo nucleótido de adenina para evitar que se ligan entre ellos durante la reacción de ligación (Illumina, 2015).

Un solo nucleótido correspondiente a timina en el extremo 3' del adaptador provee un solapamiento complementario para el fragmento del adaptador que es ligado. Esta estrategia nos sirve para tener un bajo grado de formación de quimeras (templados concatenados)

(Illumina, 2015). Se obtuvieron 20 millones de secuencias con una longitud de 75 pbs y una calidad aceptable.

En el caso de nuestras muestras de pozol 5 amplificaciones se tuvieron que hacer a las librerías shotgun (Illumina, 2015).

Para realizar la anotación taxonómica de nuestras secuencias de shotgun utilizamos el programa Kraken. Éste es un sistema para la asignación de etiquetas taxonómicas de secuencias cortas de DNA, usualmente obtenidas a través de estudios metagenómicos. La estimación de la abundancia se realiza mediante la creación de una base de datos de k-meros, que es mucho más pequeña que la colección de todos los genomas, lo que permite la clasificación mucho más rápida que los métodos que intentan identificar cada lectura en un set de datos. Las bases de datos que se construyen con Kraken contienen “marcadores de genes” (copias únicas de genes presentes en casi todos los microorganismos), o genes que son específicos de ciertos clados (Wood et al., 2014).

Esta base de datos consiste en un k-mero y del ancestro común en un nivel taxonómico mas bajo (LCA por sus siglas en inglés) de todos los organismos cuyo genoma contiene ese k-mero. Esto permite un vistazo rápido a los nodos más específicos del árbol asociado a esos k-meros. Su precisión se hace posible debido al gran número creciente de genomas microbianos secuenciados, actualmente unos 8500, lo que hace probable que secuencias muy similares de una especie en particular hayan sido vistas anteriormente (Wood et al., 2014).

Para clasificar una secuencia de DNA se colectan todos los k-meros en esa secuencia en un set. Posteriormente se mapea cada k-mero al LCA taxon de todos los genomas que contienen ese k-mero. Ese LCA taxon y sus acentros en el árbol taxonómico forman “el árbol de clasificación” y el peso de la clasificación para cada nodo depende del número de k-meros que mapean al taxon asociado a ese nodo. El método de clasificación del árbol traza un camino que va de la raíz a la hoja mediante el cálculo de todos los nodos a través del camino. (Wood et al., 2014).

Las secuencias son clasificadas mediante la consulta en la base de datos de cada k-mero en una secuencia y posteriormente utilizando un set de LCA taxa para determinar la etiqueta apropiada para cada secuencia. Las secuencias que no tienen k-meros en la base de datos permanecen no clasificadas por Kraken. (Wood et al., 2014).

Kraken incluye una biblioteca por default, basada en genomas microbianos completos de la base de datos de NCBI RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>), pero la biblioteca se puede personalizar dependiendo de las necesidades del usuario (Wood et al., 2014). Por default Kraken crea una base de datos de k-meros con $k=31$. Para nuestro análisis creamos una base de datos que contiene la información del NCBI para bacterias y el genoma del maíz con un $k=31$.

Una vez que la biblioteca ha sido escogida, se utiliza Jellyfish que es un algoritmo de conteo de k-meros y está basado en una tabla multithreaded, lock-free optimizada por conteo de k-

meros de hasta 31 bases y es por eso que Kraken lo utiliza para crear la base de datos que contiene cada 31-mero distinto en la biblioteca (Marçais et al., 2011).

Los clasificadores generalmente adoptan una de dos estrategias; o clasifican todas las secuencias tan precisamente como sea posible o, como Kraken y Megablast, dejan afuera secuencias sin clasificar si existe evidencia insuficiente. Sin embargo el uso exacto de $k=31$ da una mayor precisión para Kraken. (Wood et al., 2014).

2.6 Ensamblés.

La masa utilizada en la preparación del pozol está altamente enriquecida en maíz, por lo cual fue necesario eliminar las secuencias provenientes de su material genético/DNA.

Se removieron las secuencias que pertenecían al genoma del maíz mediante el alineamiento de nuestras lecturas crudas contra el genoma de referencia del maíz *Zea mays* (assembly B73 RefGen_v4) (<https://www.ncbi.nlm.nih.gov/genome/?term=zea+mays>).

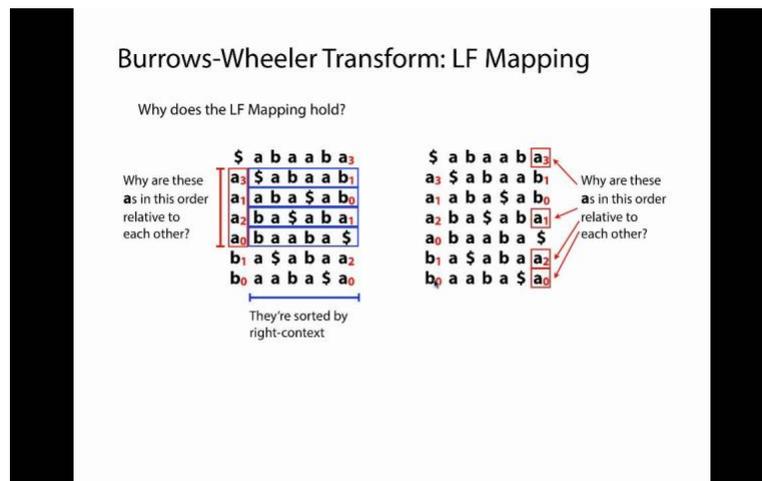


Figura 12. Transformada de Burrows-Wheeler.

Imagen obtenida en la página: <https://i.ytimg.com/vi/4n7NPK5Iwbl/maxresdefault.jpg>

El programa que utilizamos para hacer nuestro alineamiento es Burrows-Wheeler Alignment (BWA). BWA es una herramienta de alineamiento basada en la búsqueda de la transformadora de Burrows-Wheeler (BWT) (Figura 12), para alinear de manera eficiente las lecturas de secuencias cortas en contra de una secuencia de referencia larga, como en nuestro caso es el genoma de maíz, permitiendo los gaps y los mismatches (Li et al., 2013).

Para ello el algoritmo de BWA que utilizamos es BWA-MEM, que es el más nuevo y es generalmente recomendado para consultas de alta calidad ya que es más rápido y más preciso. BWA-MEM también funciona mejor para lecturas de Illumina de entre 70 y 100 pares de bases (Li et al., 2013).

Por medio de este programa se obtienen unos archivos de alineamiento de un formato conocido como SAM. El formato SAM (The Sequence Alignment Map) es un alineamiento genérico para

guardar alineados de lecturas en contra de secuencias de referencia, que soporta lecturas cortas y largas producidas por diferentes plataformas de secuenciación.

Este formato soporta lecturas pareadas o sencillas y lecturas combinadas de distintos tipos. La equivalente representación binaria, BAM, es compacta en tamaño y soporta la rápida recuperación de los alineamientos en regiones específicas. De nuestro alineamiento y con fines de ahorrar espacio en el disco utilizamos los archivos en formato BAM (Li et al., 2013).

Una vez obtenidos los archivos BAM, utilizamos SAMtools para la manipulación de estos. SAMtools es una suite de programas para interactuar con datos de alta resolución de secuenciación.

Puede manipular alineamientos en los formatos SAM/BAM: leer, escribir, editar e indexar, viendo y convirtiendo formatos SAM/BAM. SAMtools también puede indexar secuencias de referencia en el formato FASTA o extraer subsecuencias de secuencias de referencia indexadas (Li et al., 2009).

SAMtools implementa varias utilidades para post procesamiento de los alineamientos en el formato SAM, como lo son indexar y visualización de alineamientos y también provee de herramientas universales para procesar las lecturas alineadas (Li et al., 2009).

Utilizando el ordenamiento posicional y el indexado, las aplicaciones pueden realizar un procesamiento en regiones específicas genómicas sin cargar todo el archivo a la memoria. El formato SAM/BAM en conjunto con SAMtools, separa el paso de alineamiento para posterior análisis, permitiendo un acercamiento genérico y modular al análisis de los datos de secuenciación genómica (Li et al., 2009). En nuestro caso es la herramienta que se utiliza para quitar todas las secuencias que pertenecen al genoma del maíz.

Posteriormente una vez que ya tenemos las secuencias libres de DNA de maíz se procede a hacer el ensamblado. Al no existir un metagenoma de referencia del pozol se hizo un ensamblado *de novo* por medio de dos ensambladores para poder hacer una comparación de resultados y así proceder a realizar la anotación funcional. En este caso los dos ensambladores que utilizamos fueron IDBA-UD y Megahit.

IDBA-UD es un ensamblador *de novo* iterativo del gráfico de de Bruijn para datos de secuenciación de lecturas cortas con una alta y desigual profundidad de secuenciación. En lugar de utilizar un límite sencillo utilizamos límites relativos a profundidad para remover k-meros erróneos en regiones de baja y alta profundidad. (Peng et al., 2012).

El ensamblador IDBA provee una solución en la cual en lugar de utilizar un solo k, itera de $k=k_{\min}$ a $k=k_{\max}$. En cada iteración los contigs construidos son utilizados como lecturas para la siguiente iteración. Estos contigs acarrean los k-meros a la siguiente iteración y es así como solucionan el problema de los gaps. Es entonces que se depende de k largos para resolver las ramificaciones de las regiones repetidas (Peng et al., 2012).

Además de iterar k de pequeños a grandes, IDBA-UD reconstruye los k -meros faltantes mediante el ensamblado local y remueve errores mediante la remoción iterativa de contigs de baja profundidad (Peng et al., 2012).

MEGAHIT es un ensamblador *de novo* de NGS para ensamblar datos largos y complejos metagenómicos de forma que se ahorre tiempo y el proceso sea eficiente. MEGAHIT hace uso de un gráfico de de Bruijn sucinto, que es una representación comprimida del gráfico de Bruijn. Tiene raíz en un algoritmo paralelo y explota ese paralelismo de las unidades de procesamiento de gráficos (GPU por sus siglas en inglés), mediante la adaptación del algoritmo de construcción, que toma como ventaja del GPU el poder ordenar los sufijos de un set de lecturas de manera eficiente. Se adopta una estrategia de colocar particiones de k -meros en bloques de acuerdo a la longitud de los prefijos ($l=8$ por default). Los k -meros en particiones consecutivas que quedan sin la memoria de GPU se organizan juntos (Li et al., 2015).

MEGAHIT remueve potencialmente fillos erróneos mediante la remoción de tapas, juntando burbujas y removiendo fillos de baja cobertura local. Este acercamiento es útil en metagenómica que sufre de profundidad de secuencia no uniforme (Li et al., 2015).

A continuación y una vez obtenidos los contigs con cada ensamblador, se obtienen las estadísticas de cada uno. Se mapean los ensamblados con BWA y se calcula la cobertura de dichos ensamblados para posteriormente normalizarla.

Esto se hace con el fin de revisar nuestros ensamblados y supervisar que no existan las quimeras (malos ensamblados). BWA-MEM también identifica lecturas quiméricas, una característica crucial para el alineamiento de contigs y que no existe en la gran mayoría de los alineadores de NGS (Li et al., 2013).

El archivo de coberturas nos sirve para establecer la abundancia de cada gene en el metagenoma y la longitud del mismo que pudo ser cubierta con el mapeo.

Una vez obtenidas las coberturas se genera una matriz de abundancias para hacer una comparación de nuestras muestras.

2.7 Anotación funcional: Metagenoma Shotgun.

En base a los contigs obtenidos por ambos ensambladores se busca hacer la anotación funcional con el que haya dado como resultado el menor número de contigs posible y el mayor número en N50.

A partir de estos contigs se realiza la anotación funcional de las secuencias empezando con la predicción de genes. Esto se consigue mediante la predicción del potencial metabólico de las secuencias predichas como codificantes. Primero se utiliza el programa MetaGeneMark para obtener las secuencias en aminoácidos y la secuencia de genes mediante la predicción de Marcos Abiertos de Lectura (ORFs por sus siglas en inglés).

Para ello, los péptidos predichos serán anotados por búsquedas en las bases de datos. Para esto se buscan los dominios conservados con hmmscan de HMMER dentro de la base de datos formateada de dbCAN contra el archivo de proteínas y se ordenan en una tabla de salida de los dominios que coinciden con un criterio de selección de e-value 1×10^{-5} y 80% de cobertura. La base de datos dbCAN está definida por los dominios firma de cada familia de CAZyme derivados de la base de datos de dominios conservados (CDD por sus siglas en inglés) buscados en la literatura y posteriormente curados.

Para construir estos modelos específicos primeramente identificamos las regiones de los dominios con el componente de las proteínas de GenBank para cada familia de CAZyme, basándonos en cada modelo de CDD seleccionado utilizando la herramienta de HMMER llamada *hmmsearch*. Estos HMMs específicos de cada familia de CAZyme son los fundamentos de la anotación automatizada de dbCAN (*Yin et al. 2012*).

Posteriormente y para finalizar la anotación se hará una reconstrucción metabólica de las vías de degradación de los principales carbohidratos encontrados en la fermentación. La reconstrucción utiliza el programa GhostKOALA.

GhostKOALA permite la anotación de metagenomas: GhosKOALA es un servidor de anotación automático que utiliza el programa GHOSTX para la búsqueda en la base de datos de GENES no redundante. Este software es recomendado para anotar una gran cantidad de secuencias metagenómicas que realizan asignaciones KO (KEGG Orthology) para caracterizar funciones de genes individuales y la reconstrucción de rutas KEGG, jerarquías BRITE y módulos KEGG para inferir funciones de alto nivel de organismos o ecosistemas. El programa de búsqueda GHOSTX que es adecuado para la anotación de metagenoma y los archivos resultantes pueden ser descargados y manipulados para más análisis con KEGG Mapper.

La anotación de las secuencias metagenómicas usualmente involucra la asignación de información tanto funcional como taxonómica. Las bases de datos no redundantes de secuencias de pangenoma son suplementadas por la similitud de grupos de genes sin la asignación de números K para la mejor caracterización de la composición taxonómica de los metagenomas. En GhostKOALA, la asignación del número K es simplemente basada en la suma de todas las puntuaciones de GHOSTX normalizadas sin considerar ningún factor de peso. Esto se logra mediante la creación de Cd-hit clusters con el 50% de identidad para cada especie/género/familia/grupo y las secuencias son seleccionadas de cada cluster cuando ningún miembro es asignado a un número K y ningún miembro del genoma representativo está incluido.

Resultados

1.1 Extracción del DNA genómico.

Se realizó la extracción del DNA de las muestras de pozol a los tiempos de 0 horas, 9 horas, 24 horas y 48 horas. En la Tabla 2 se presentan las concentraciones de DNA cuantificadas por Qubit:

Muestra	Concentración (ng/ μ L)	Volúmen (μ L)	Cantidad (ng)
0 horas	1.47	100	147
9 horas	4.08	100	408
24 horas	6.51	100	651
48 horas	12.2	100	1220

Tabla 2. Cuantificación de DNA por Qubit.

En base a los resultados obtenidos en la extracción de DNA de la bola 2 de pozol, obtuvimos poca cantidad, lo cual representaba una limitación ya que normalmente se estima que para las muestras de shotgun es necesario tener como mínimo 1 μ g total. Al no cumplir con esta condición en tres de los tiempos muestreados (0hrs, 9hrs y 24hrs), se modificó el protocolo realizando 5 ciclos de amplificación a las muestras que fueron destinadas a las bibliotecas de shotgun.

Las bibliotecas de 16S rRNA requieren de nanogramos de DNA, por lo que las cantidades que extrajimos alcanzan perfectamente para su realización. (Figura 13)

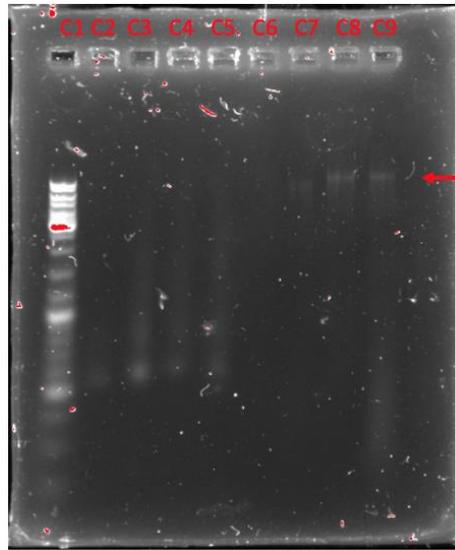


Fig 13. Gel de agarosa al 1% con 10 μ L de cada muestra + 2 μ L de LB. Carril 1(C1): Marcador de peso Molecular, Carril 6(C6): DNAg 0 horas Carril 7 (C7): DNAg 9 horas, carril 8(C8): DNAg 24 horas carril 9 (C9): DNAg 48 horas.

1.2 Amplificación por PCR

Para hacer la amplificación de las regiones V3-V4 se realiza la PCR ya descrita y mediante el gel de agarosa al 1% podemos ver los productos de PCR de aproximadamente 460 pbs.

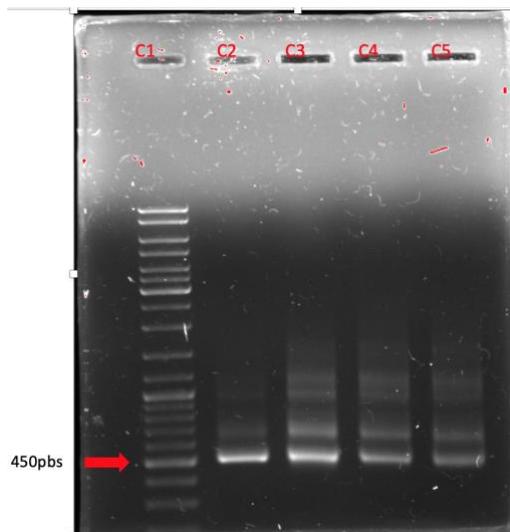


Fig 14. Gel de agarosa al 1% con 10 μ L de cada muestra +8 μ L agua + μ L 2 μ L de LB. Carril 1(C1): Marcador de peso Molecular, Carril 2 (C2): Amplicón 0 horas Carril 3 (C3): Amplicón 9 horas, Carril 4 (C4): Amplicón 24 horas Carril 5 (C5): Amplicón 48 horas.

1.3 Metagenoma de amplicones 16S

Se realizó un control de calidad de nuestras secuencias el cual se presenta en la Tabla 3:

Muestra	Lecturas totales	Bases totales	Longitud promedio	GC (%)	Calidad promedio	Lecturas Ns (%)	Q20(%)	Q30(%)
0 horas	927,980	278,394,000	300.00	56.55	30.78	1.34	96.14	62.03
9 horas	3,739,076	1,121,722,800	300.00	54.68	31.11	1.34	96.41	63.47
24 horas	634,154	190,246,200	300.00	54.34	31.15	1.33	96.35	65.33
48 horas	1,570,390	471,117,000	300.00	54.54	31.07	1.35	96.16	64.01

Tabla 3. Calidad secuencias de amplicones.

En la tabla se puede ver que la calidad de todas nuestras secuencias fue “aceptable” y por lo tanto no es necesario remover secuencias de baja calidad. Para llegar a la conclusión de esto tomamos como parámetro de “aceptable” a todas las lecturas que como mínimo de calidad en la escala de Qphred están en Q20. En el caso de nuestras secuencias más del 90% en todos los tiempos cumple con este requisito y es por eso y con la finalidad de no perder información valiosa que decidimos no someter las secuencias a ningún programa de corte inicialmente.

Posteriormente a la hora de hacer la reconstrucción de los amplicones presentamos la Tabla 4 del porcentaje de lecturas que sí produjeron lecturas combinadas:

Horas	Lecturas totales pareadas	Lecturas combinadas pareadas	Lecturas no combinadas pareadas	Porcentaje de combinación
0	463,990	392,883	71,107	84.67%
9	785,195	666,824	118,371	84.92%
24	1,869,538	1,597,598	271,940	85.45%
48	317,077	269,838	47,239	85.10%

Tabla 4. Porcentaje de lecturas combinadas.

Como podemos observar en la tabla todos los tiempos tienen más del 80% de sus lecturas extendidas lo cual es un buen porcentaje de las secuencias para poder realizar un análisis de calidad. Las lecturas que no se pudieron extender se descartaron del análisis.

1.4 Esfuerzo de muestreo, α y β diversidad.

A continuación se presentan los resultados de α y β diversidad de los distintos tiempos de la fermentación del pozol; Se realizaron 200 rarefacciones para hacer el esfuerzo de muestreo y en estos resultados se presentan los índices de Chao1, Shannon-Weaver y Simpson que son parámetros estadísticos de la α diversidad y un gráfico descriptivo del tipo NMDS para la β diversidad.

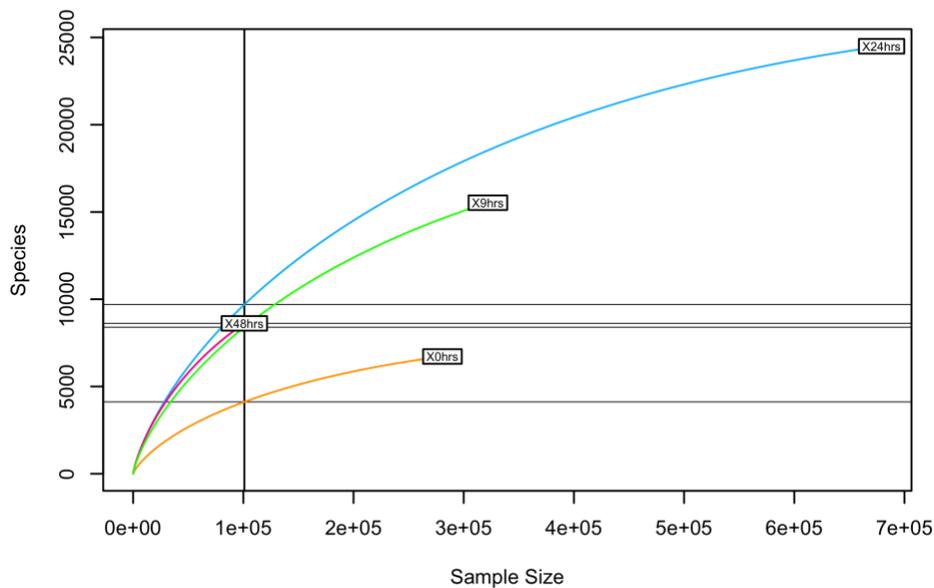


Figura 15. Esfuerzo de muestreo Número de rarefacciones total = 200.

Muestra	Observadas	Chao1	se. chao1	Índice Shannon-Weaver	Índice de Simpson.
0 horas	6720	8,086.66	74.14	2.72	0.76
9 horas	24,452	26,652.2	77.78	5.25	0.95
24 horas	8,595	13,125.25	182.88	5.27	0.96
48 horas	15,491	20,837.15	173.55	4.95	0.94

Tabla 5. Parámetros de α diversidad. Están reportados los índices de Chao1, Shannon-Weaver y Simpson

Chao1 conceptualiza la diversidad como riqueza total de especies. Debido a que es sensible al tamaño de la muestra se subestima la riqueza real cuando el muestreo es pobre como lo es en este caso. Aquí se ve reportado el número máximo de especies esperado en una muestra dadas las observaciones y a partir de la rarefacción de ésta. Se puede ver como la diversidad varía de un tiempo a otro siendo la mayor diversidad a las 9 horas, seguida de 48 horas, 24 horas y por último 0 horas. El esfuerzo de muestreo es bueno debido a que no se observa mucha diferencia entre las especies observadas vs Chao1, a excepción de la hora 24 donde sí hay una diferencia significativa. (Figura 15 Tabla 5)

El índice de Shannon-Weaver es la medida de entropía donde el sistema es la comunidad. Este calcula la probabilidad de predecir la especie del siguiente individuo tomado al azar de la comunidad. El índice vale 0 para comunidades con una sola especie hasta valores altos para comunidades con muchas especies, cada una con pocos individuos. Por lo general toma valores mayores a 5 para muestras muy diversas (Brown et al., 2001). En nuestras muestras se observa que a 0 horas la muestra no es muy diversa mientras, que en 9, 24 y 48 horas sí lo es. (Tabla 5)

El índice de Simpson es una medida de la heterogeneidad de abundancia en la muestra. Se basa en la probabilidad de que, si se toman dos individuos, éstos sean de la misma especie. Toma valores desde 0-1, índices cercanos a 1 indican que sólo unas pocas especies predominan en la muestra (Brown et al., 2001). En el caso de nuestros tiempos esto indica que en la muestra de 0 horas hay más especies que predominan en comparación con los otros tres tiempos cuyo índice es muy cercano a 1. (Tabla 5)

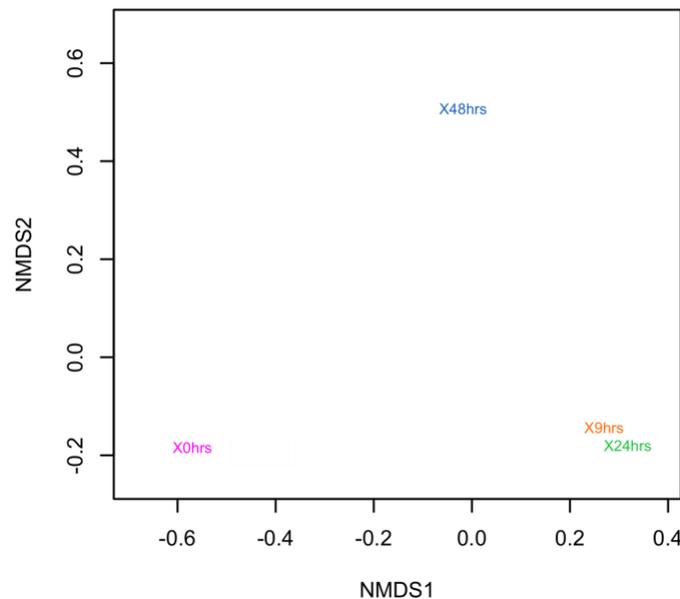


Figura 16. Gráfica NMDS de las muestras de pozol. Las muestras de 9 y 24 horas se correlacionan, las de 0 y 48 horas no.

La β diversidad nos muestra qué tan diversa es una muestra en comparación con otra. Para obtener la beta diversidad se realizó una normalización de la matriz por el método metagenomeSeq y el índice de disimilitud de Bray-Curtis para la obtención de la gráfica NMDS (Non-metric multidimensional scaling). Para el índice de disimilitud de Bray-Curtis hay una escala de 0 a 1; 0 si ambos sitios tienen la misma composición y 1 si no comparten ninguna especie (Bloom, 1981).

En nuestro metagenoma se puede ver mediante este gráfico que la muestra de 0 horas es muy distinta, así como la de 48 horas y que tanto la de 9 como la de 24 horas comparten una composición microbiana parecida.(Figura 16)

2. Taxonomía

2.1 Amplicones 16S

Como se puede ver en la Tabla 6, con Parallel-Meta se anotaron prácticamente todas las lecturas totales combinadas en todos los casos.

Tiempo	Lecturas combinadas rRNA	totales de	Lecturas mapearon rRNA	que	Porcentaje (%)
0hrs	392,883		392,881		99.99
9hrs	666,824		666,824		100
24hrs	1,597,598		1,597,598		100
48hrs	269,838		269,836		99.99

Tabla 6. Lecturas combinadas que se utilizaron para el análisis de 16S.

La asignación taxonómica de nuestras lecturas combinadas arrojó una matriz integrada con todos los niveles taxonómicos, misma que dividimos en subniveles (dominio, filo, clase, orden, familia, género) para la obtención de las gráficas de calceta correspondientes a los mencionados subniveles.

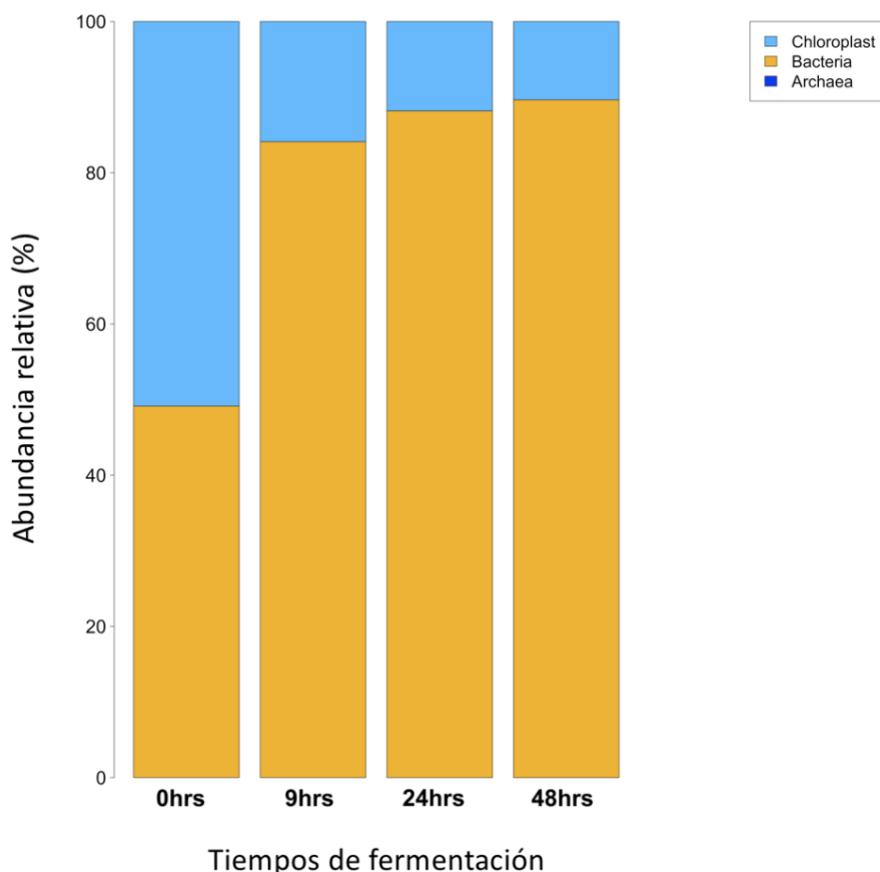


Figura 17. Abundancia relativa de los principales dominios en la fermentación del pozol por el método de amplicones 16S. La leyenda se lee de abajo hacia arriba.

Las muestras de 16S en su gran mayoría constituyen en sus cuatro tiempos al dominio bacteriano (49.13%, 84.11%, 88.18%, 89.63% a 0,9,24 y 48hrs respectivamente), aunque poseen un porcentaje significativo de cloroplasto – correspondiente al cloroplasto del Maíz principalmente - (50.86%,15.89%,11.82%,10.36% a 0,9,24 y 48hrs respectivamente), sobre todo al tiempo de 0hrs. (Figura 17)

También y en una proporción mucho menos significativa nuestras muestras tenían una cantidad pequeña del dominio Archaea (menor al 0.5% en todos los tiempos). (Figura 15)

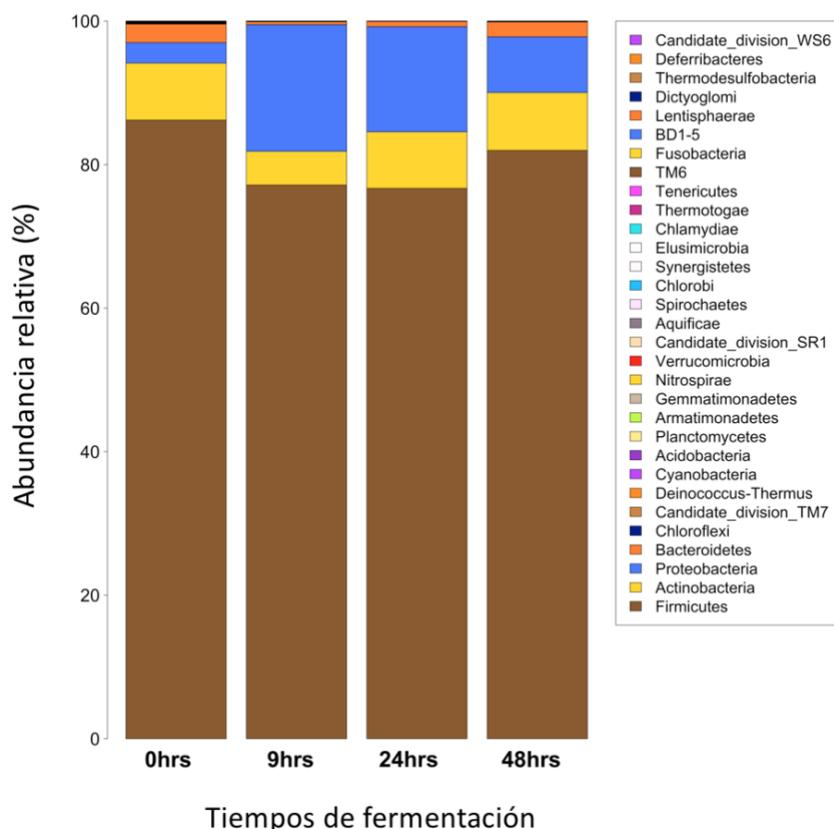


Figura 18. Abundancia relativa de los principales filos microbianas en la fermentación del pozol por el método de amplicones 16S. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 31 filos (*Figura 18*) de los cuales se ve que Firmicutes es el más abundante en los cuatro tiempos de fermentación (86.20%, 77.16%, 76.69%, 81.99% a 0, 9, 24 y 48 horas respectivamente), seguido en mucha menor proporción por Actinobacteria, Proteobacteria y Bacteroidetes.

Actinobacteria tiene su mayor proporción a las 0hrs (7.93%) para posteriormente bajar un poco en abundancia a las 9 horas (4.70%) y posteriormente tener un aumento a las 24 horas (7.87%) y a las 48hrs (8.05%).

En el caso de Proteobacteria a 0 horas se encuentra en poca cantidad (2.86%), tiene un aumento considerable a las 9 horas (17.65%), para luego bajar un poco en porcentaje a las 24 horas (14.66%) y una disminución mayor a las 48 horas (7.78%).

Bacteroidetes presenta su mayor representación en los tiempos de 0 horas (2.58%) y 48 horas (2.07%) y de menor proporción a 9 horas (0.40%) y 24 horas (0.72%).

La contribución de los otros filos es tan pequeña que no es representativa ni apreciable en la figura.

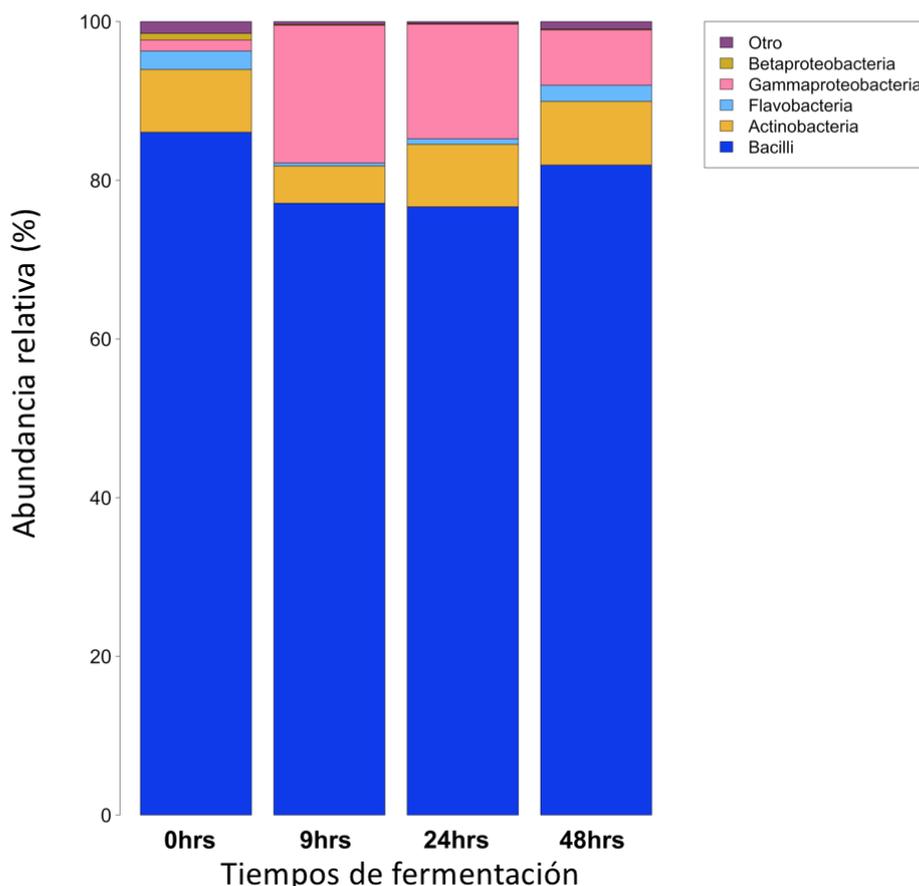


Figura 19. Abundancia relativa de las principales clases microbianas en la fermentación del pozol por el método de amplicones 16S. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 75 clases (*Figura 19*). De esas clases, hay cuatro grupos predominantes; Bacilli seguido de Actinobacteria, Flavobacteria y Gammaproteobacteria. El más predominante de los cuatro es Bacilli (86.06%, 77.11%, 76.66%, 81.93% a 0, 9, 24 y 48 horas respectivamente).

Actinobacteria tiene sus mayores porcentajes a las 0, 24 y 48 (7.90%, 7.87%, 8.02% respectivamente) horas teniendo un descenso a las 9 horas (4.69%). Flavobacteria se presenta en mayor proporción a las 0 horas (2.33%) y a las 48 horas (2.04%) y en menor proporción a las 9 horas (0.38%) y a las 24 horas (0.70%).

Gammaproteobacteria se encuentra en poca proporción a las 0 horas (1.39%) aumenta de manera considerable a las 9 horas (17.35%) disminuye ligeramente a las 24 horas (14.44%) y posteriormente disminuye más a las 48 horas (6.97%).

Como Otro nos referimos a todas las clases cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes 1.49%, 0.36%, 0.24%, 0.95% a 0, 9, 24 y 48 horas respectivamente.

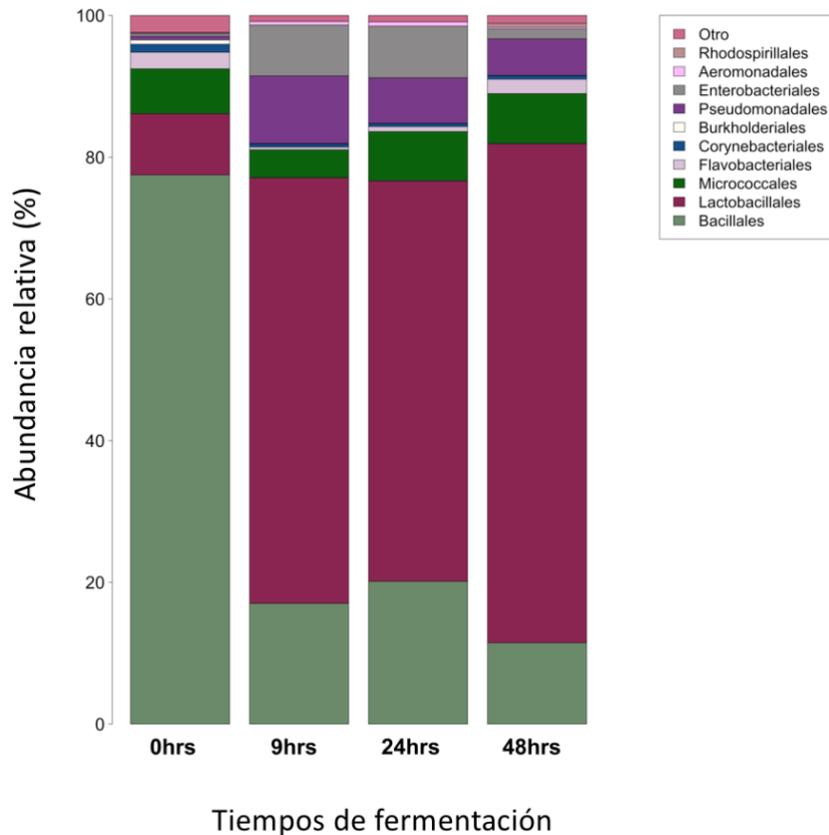


Figura 20. Abundancia relativa de los principales órdenes microbianos en la fermentación del pozol por el método de amplicones 16S. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 155 órdenes (*Figura 20*).

Los principales órdenes son Bacillales, Lactobacillales, Micrococcales, Flavobacteriales, Pseudomonadales, Enterobacteriales.

A las 0 horas de fermentación el orden que predomina es Bacillales (77.50%), seguido de Lactobacillales (8.64%), Micrococcales (6.34%), Flavobacteriales (2.31%).

A las 9 horas de fermentación el orden que predomina es Lactobacillales (60.08%), seguido de Bacillales (17.03%), seguido de Pseudomonadales (9.55%), Enterobacteriales (7.18%) y Micrococcales (3.94%).

A las 24 horas Lactobacillales continúa siendo el orden más abundante (56.52%), seguido de Bacillales (20.14%), Enterobacteriales (7.29%), Micrococcales (6.95%) y Pseudomonadales (6.44%).

A las 48 horas Lactobacillales continúa siendo el orden más abundante (70.46%), seguido de Bacillales (11.47%), Micrococcales (7.05%), Pseudomonadales (5.16%) y Flavobacteriales (2.03%).

Como Otro nos referimos a todos los órdenes cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes 2.40%, 0.84%, 0.90%, 1.11% a 0, 9, 24 y 48 horas respectivamente.

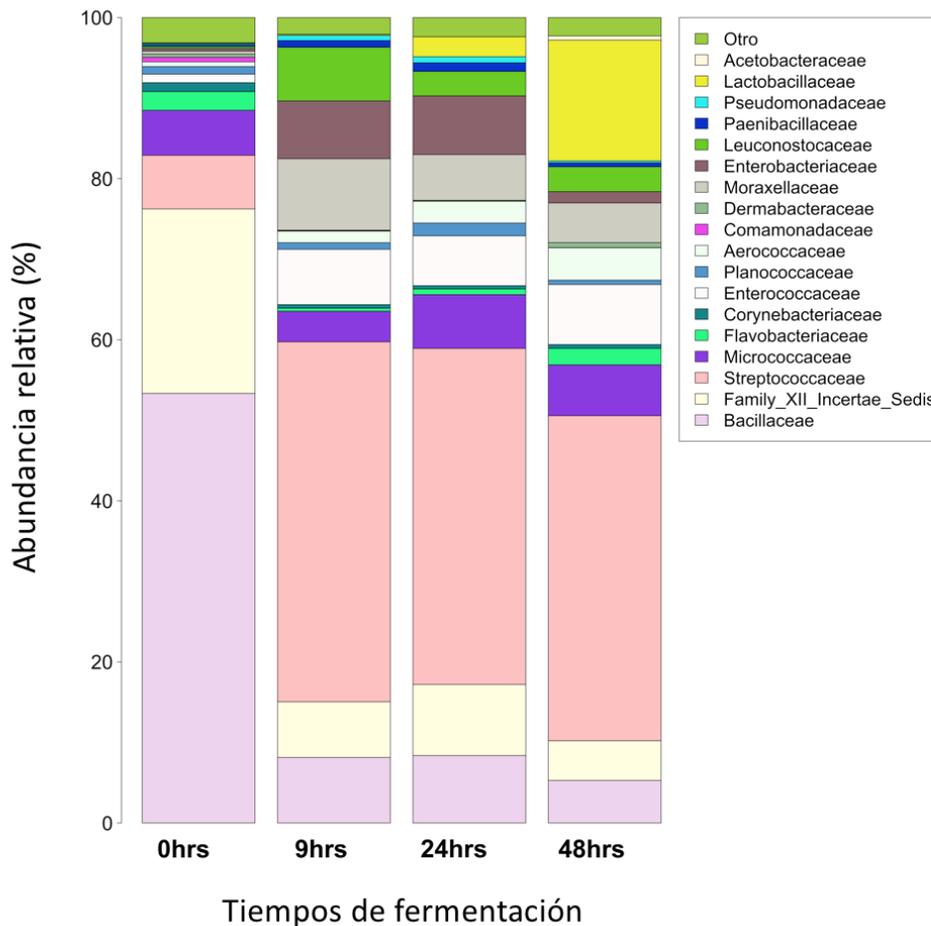


Figura 21. Abundancia relativa de las principales familias microbianas en la fermentación del pozol por el método de amplicones 16S. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 342 familias (Figura 21).

A las 0 horas la familia más abundante es Bacillaceae (53.34%), seguida de XII Incertae Sedis,(22.90%), Streptococcaceae (6.67%), Micrococccaceae (5.58%), Flavobacteriaceae (2.33%) y Corynebacteriaceae (1.12%).

A las 9 horas la familia más abundante es Streptococcaceae (44.68%), seguido de Moraxellaceae (8.88%), Bacillaceae (8.14%), Enterobacteriaceae (7.18%), XII Incertae Sedis (6.92%), Enterococcaceae (6.86%), Leuconostocaceae (6.66%) y Micrococccaceae (3.80%).

A las 24 horas la familia que continúa siendo la más abundante es Streptococcaceae (41.72%), seguida de XII Incertae Sedis (8.83%), Bacillaceae (8.37%), Enterobacteriaceae (7.29%), Micrococcaceae (6.68%), Enterococcaceae (6.18%), Moraxellaceae (5.68%), Leuconostocaceae (3.03%), Aerococcaceae (2.70%), Lactobacillaceae (2.47%) y Planococcaceae (1.59%).

A las 48 horas la familia que continúa siendo la más abundante es Streptococcaceae (40.35%), Lactobacillaceae (14.99%), Enterococcaceae (7.45%), Micrococcaceae (6.32%), Bacillaceae (5.29%), Moraxellaceae (4.93%), XII Incertae Sedis (4.92%), Aerococcaceae (4.02%), Leuconostocaceae (3.09%), Flavobacteriaceae (2.04%) y Enterobacteriaceae (1.41%).

Todas las familias cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación son referidas como "Otro" y se agrupan dando los siguientes porcentajes 3.14%, 2.08%, 2.38%, 2.28% a 0, 9, 24 y 48 horas respectivamente.

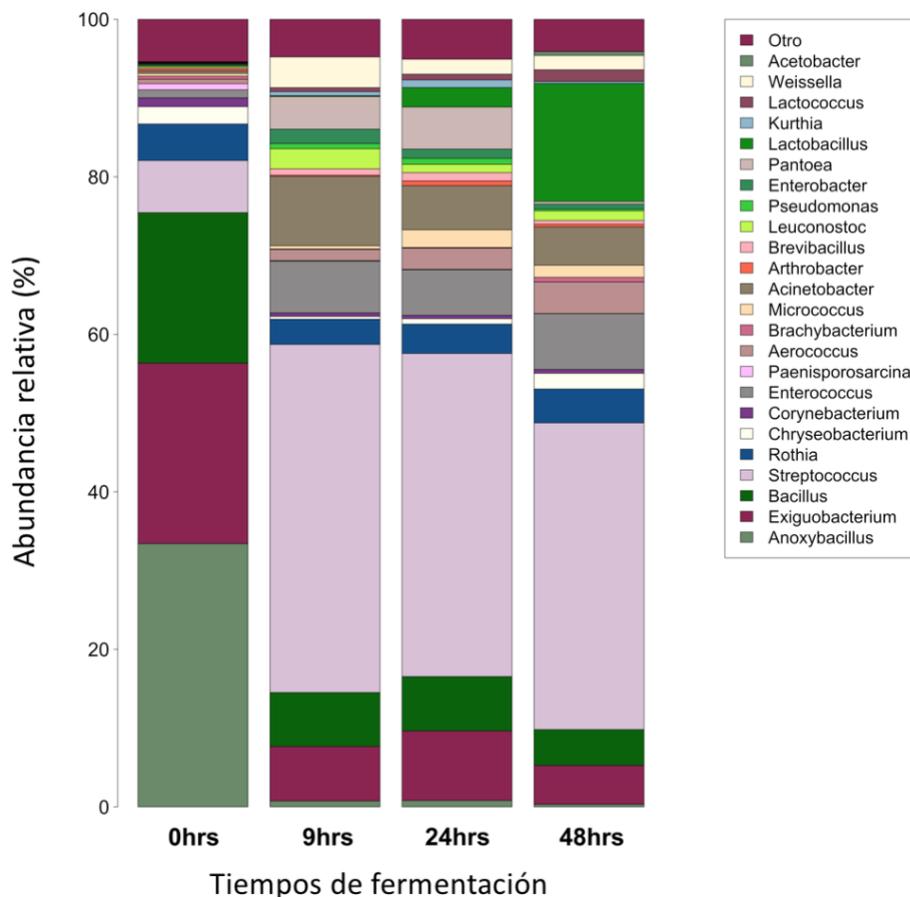


Figura 22. Abundancia relativa de los principales géneros microbianos en la fermentación del pozol por el método de amplicones 16S. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 1042 géneros (Figura 22).

A las 0 horas de la fermentación *Anoxybacillus*, *Exiguobacterium* y *Bacillus* representan más de la mitad de la población microbiana (33.42%, 22.92% y 19.11% respectivamente). En menor proporción encontramos a *Streptococcus* (6.62%).

A las 9 horas *Streptococcus* aumenta para convertirse en el género con más abundancia representando casi la mitad de la población microbiana (44.20%), seguido por *Acinetobacter* que también aumentó en proporción (8.83%), *Exiguobacterium* que disminuyó en abundancia (6.93%) así como *Bacillus* (6.85%).

A las 24 horas de la fermentación *Streptococcus* continúa siendo el género más abundante (41.02%) seguido de *Exiguobacterium* (8.84%), *Bacillus* (6.91%), *Enterococcus* (5.80%), *Lactobacillus* que aumentó de manera considerable en comparación con su proporción en los dos tiempos anteriores (2.48%).

A las 48 horas *Streptococcus* continúa siendo el género más abundante, aunque en ligera menor proporción (38.94%), seguido de *Lactobacillus* que tuvo un aumento considerable en relación con los tiempos anteriores, alcanzando un máximo en abundancia en este punto de la fermentación (15.00%), seguido de *Enterococcus* (7.09%), *Exiguobacterium* (4.93%), *Acinetobacter*(4.88%) y *Bacillus* (4.55%).

Como Otro nos referimos a todos los géneros cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes 5.39%, 4.77%, 5.06% y 4.08% a 0, 9, 24 y 48 horas respectivamente.

Las tablas con las matrices de cada nivel taxonómico se encuentran anexadas en el material suplementario de esta tesis.

2.2 Metagenoma tipo shotgun.

Se revisó la calidad de las secuencias tipo shotgun y se hizo el filtrado de las lecturas para obtener el porcentaje de ellas sin maíz (Tabla 7).

Muestra	Lecturas totales	Bases totales	Longitud promedio	GC (%)	Calidad promedio	Lecturas Ns (%)	Q20(%)	Q30(%)	Porcentaje de secuencias sin maíz (%)
0 horas	39,039,810	2,927,985,750	75.00	47.12	33.32	0.05	99.39	88.26	5.72
9 horas	35,370,260	2,652,769,500	75.00	46.13	33.44	0.05	99.23	89.30	48.28
24 horas	46,044,100	3,453,307,500	75.00	46.80	33.47	0.05	98.83	89.17	55.92
48 horas	41,007,846	3,075,588,450	75.00	45.46	33.42	0.05	99.49	89.12	31.54

Tabla 7. Calidad de las secuencias tipo shotgun.

Al igual que en las bibliotecas creadas para 16S, podemos ver que la calidad es aceptable debido a que casi el 100% de las secuencias tienen calidad de Q20, la cual es aceptable debido a que nuestras secuencias tienen un 99% de certeza. Debido a que se tuvo que filtrar de la muestra sólo las secuencias que no contenían DNA de maíz nos encontramos con muy poco DNA bacteriano a las 0 horas.

Kraken nos arroja los siguientes resultados de la anotación taxonómica (Tabla 8):

Tiempo de fermentación	Secuencias pareadas	Clasificadas	No clasificadas
0hrs	19,519,905 (2947.51 Mbp)	526,431 (2.70%)	18,993,474 (97.30%).
9hrs	20,503,923 (3096.09 Mbp)	4,159,207 (20.28%)	16,344,716 (79.72%)
24hrs	17,685,130 (2670.45 Mbp)	4,946,374 (27.97%)	12,738,756 (72.03%)
48hrs	23,022,050 (3476.33 Mbp)	2,994,026 (13.01%)	20,028,024 (86.99%)

Tabla 8. Secuencias anotadas taxonómicamente con Kraken.

Al igual que en los amplicones de 16S, la asignación taxonómica de nuestras lecturas pareadas arrojó una matriz integrada con todos los niveles taxonómicos, misma que dividimos en subniveles (dominio, filo, clase, orden, familia, género) para la obtención de las gráficas de calceta correspondientes a los mencionados subniveles.

En el caso de la anotación con Kraken se sigue la condición de “root-to-leaf” lo que significa que en cada nivel taxonómico el anotador verá si se puede o no anotar. Por lo tanto, hay secuencias que a pesar de cambiar de nivel taxonómico no podrán ser anotadas y se denominarán con el último nivel que se logró anotar seguido de “; no_rank”.

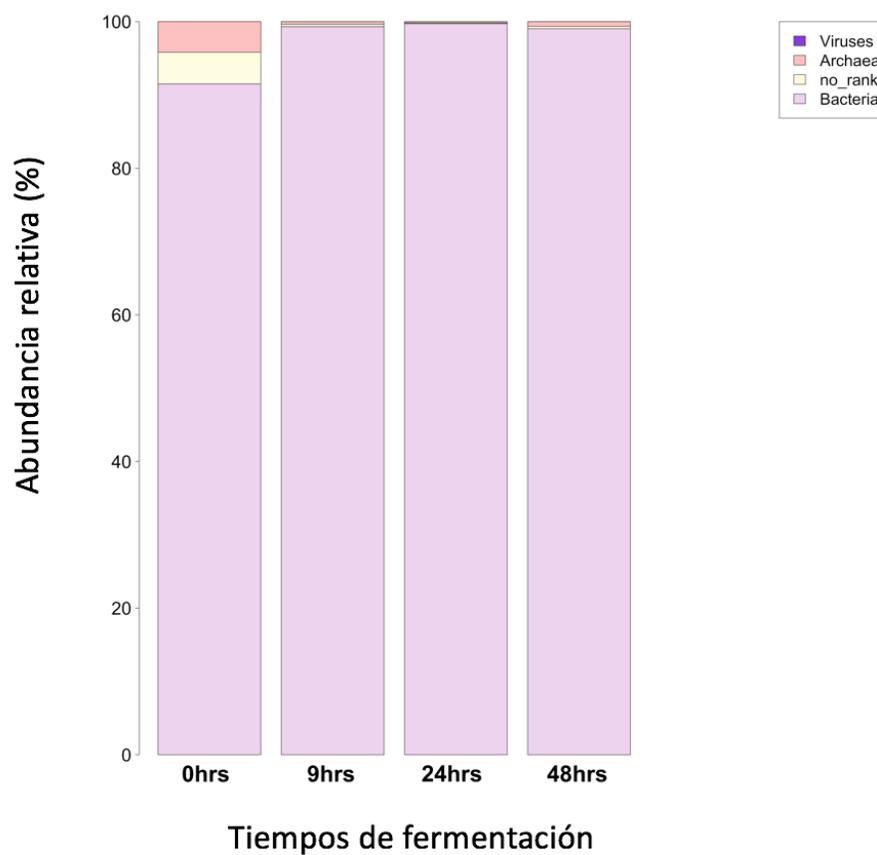


Figura 23. Abundancia relativa de los principales dominios en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

La base de datos de Kraken está diseñada para tres dominios, que son: Bacteria, Archaea y Virus. En nuestro caso Bacteria es el dominio más importante en todas las muestras siendo un porcentaje arriba del 90% en todos los tiempos (91.50%, 99.30%, 99.72% y 99.02% en 0,9,24 y 48 horas respectivamente). Archaea (4.32%, 0.36%, 0.17% y 0.64% en 0,9,24 y 48 horas respectivamente) y virus (menos del 0.01% en los cuatro tiempos) se encuentran en proporciones muy bajas en todas las muestras de la fermentación. También hubo un porcentaje que no se pudo asignar a nada (no rank) y es 4.32%, 0.33%, 0.11% y 0.34% en 0,9,24 y 48 horas respectivamente (Figura 23).

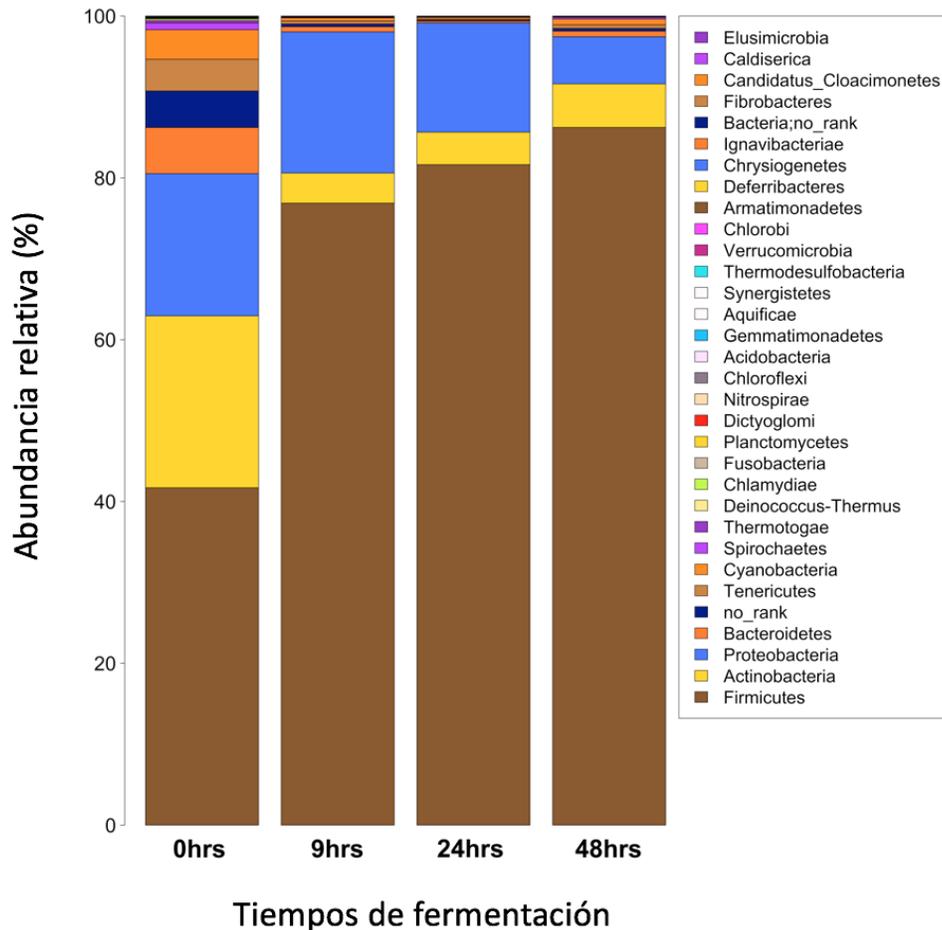


Figura 24. Abundancia relativa de los principales filios microbianos en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 32 filios (*Figura 24*).

Es el filo de Firmicutes el que predomina en toda la fermentación como el más abundante (41.70 %, 76.88 %, 81.64 %, 86.24 % a 0,9,24 y 48 horas respectivamente). A 0 horas el grupo siguiente de importancia es Actinobacteria y tiene una gran abundancia al principio de la fermentación (21.24 %).

Sin embargo, a partir de las 9 horas se vuelve menos abundante de manera significativa hasta el final de la fermentación a las 48 horas (3.72 %, 4.01 %, 5.38 % a 9,24 y 48 horas respectivamente). Le siguen los siguientes filios: Proteobacteria (17.58%, 17.45%,13.51% y 5.83% a 0,9,24 y 48 horas respectivamente), Bacteroidetes (5.70%, 0.64%, 0.22% y 0.69% a 0,9,24 y 48 horas respectivamente), Tenericutes (3.91%, 0.37%,0.13% y 0.45% a 0,9,24 y 48 horas respectivamente) y Cyanobacteria (3.66%, 0.38%, 0.20% y 0.71% a 0,9,24 y 48 horas respectivamente).

Los demás filos tienen menor representación del 1% y son imperceptibles en la figura.

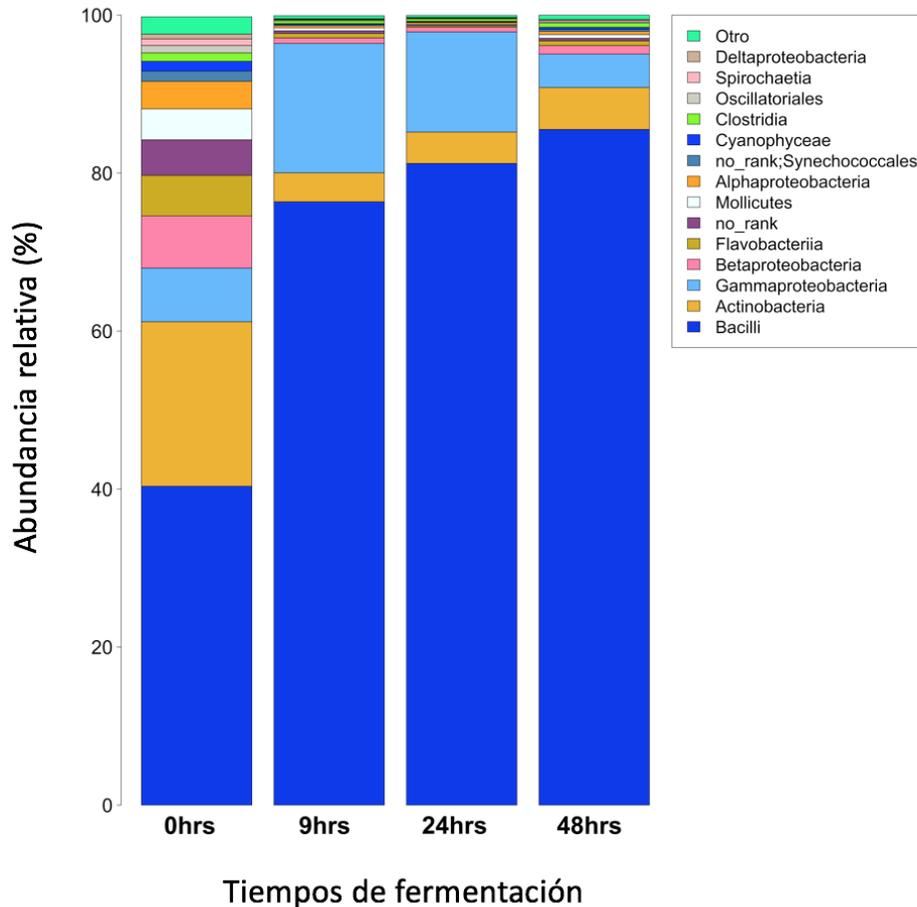


Figura 25. Abundancia relativa de las principales clases microbianas en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 75 clases (Figura 25).

A las 0 horas las clases predominantes son Bacilli (40.36%) que representa más de un cuarto de presencia en la fermentación, posteriormente aparece Actinobacteria (20.83%), seguido de Gammaproteobacteria (6.79%), Betaproteobacteria (6.62%), Flavobacteriia (5.12%), Mollicutes(3.91%), Alphaproteobacteria (3.49%), Synechococcales (1.30%), Cyanophyceae (1.22%) Clostridia (1.08%), Oscillatoriales (0.93%), Spirochaetia(0.83%) y Deltaproteobacteria (0.62%).

A las 9 horas Bacilli incrementa de manera importante para convertirse en la principal clase de la fermentación (76.37%) seguida de Gammaproteobacteria (16.38%), Actinobacteria (3.67%), Betaproteobacteria (0.70%) y Flavobacteriia (0.56%).

Las demás clases constituyen menos del 1% de abundancia en este tiempo: Mollicutes (0.37%), Clostridia (0.36%), Alphaproteobacteria (0.28%), Synechococcales (0.14%), Cyanophyceae (0.13%), Spirochaetia (0.10%), Oscillatoriales (0.09%) y Deltaproteobacteria (0.07%).

A las 24 y 48 horas Bacilli continúa siendo la clase más representativa (81.22% y 85.53% respectivamente), seguida de Gammaproteobacteria (12.68% y 4.23% respectivamente) y Actinobacteria (3.98% y 5.30% respectivamente, Betaproteobacteria (1.08% y 0.59% respectivamente) y en 48 horas le sigue Flavobacteriia (0.59%) y Clostridia (0.55%).

Las demás clases constituyen menos del 1% de abundancia en 24 horas: Clostridia (0.31%), Alphaproteobacteria (0.20%), Flavobacteriia (0.17%), Mollicutes (0.13%), Spirochaetia (0.09%), Synechococcales (0.07%), Cyanophyceae (0.06%), Oscillatoriales (0.04%) y Deltaproteobacteria (0.04%)

Las demás clases constituyen menos del 1% de abundancia en 48 horas: Mollicutes (0.45%), Alphaproteobacteria (0.42%), Synechococcales (0.26%), Cyanophyceae (0.25%), Spirochaetia (0.20%), Oscillatoriales (0.15%) y Deltaproteobacteria (0.08%).

Como Otro nos referimos a todas las clases cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes 2.19%, 0.38%, 0.26% y 0.55% a 0, 9, 24 y 48 horas respectivamente.

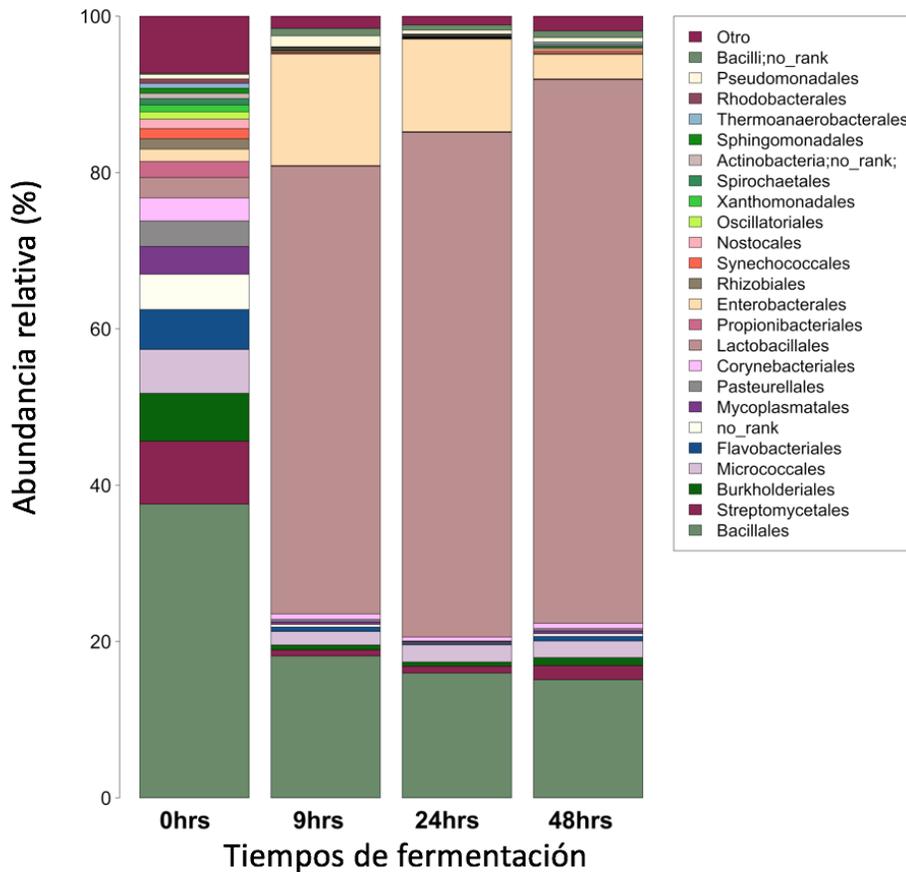


Figura 26. Abundancia relativa de los principales órdenes microbianos en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 157 órdenes (Figura 26).

A las 0 horas los órdenes más abundantes son Bacillales (37.59%) seguido de Streptomycetales (8.04%), Burkholderiales (6.13%), Micrococcales (5.60%) Flavobacteriales(5.12%), Mycoplasmatales(3.55%), Pasteurellales (3.28%),Corynebacterales (2.94%), Lactobacillales (2.62%), Propionibacterales (2.06%), Enterobacterales (1.57%), Rhizobiales(1.32%), Synechococcales(1.31%) y Nostocales (1.22%).

A las 24 horas Lactobacillales (64.58%) continúa siendo el orden dominante seguido de Bacillales (15.97%), Enterobacterales (11.84%) y Micrococcales (2.20%).

A las 48 horas Lactobacillales (69.55%) sigue predominando como el orden más abundante en la fermentación en este punto y le siguen los órdenes Bacillales (15.12%), Enterobacterales (3.12%) Micrococcales (2.12%), Streptomycetales (1.81%) y Burkholderiales (1.01%).

Como Otro nos referimos a todos los órdenes cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes a 7.29%, 1.56%, 1.11% y 1.87% a 0, 9, 24 y 48 horas respectivamente.

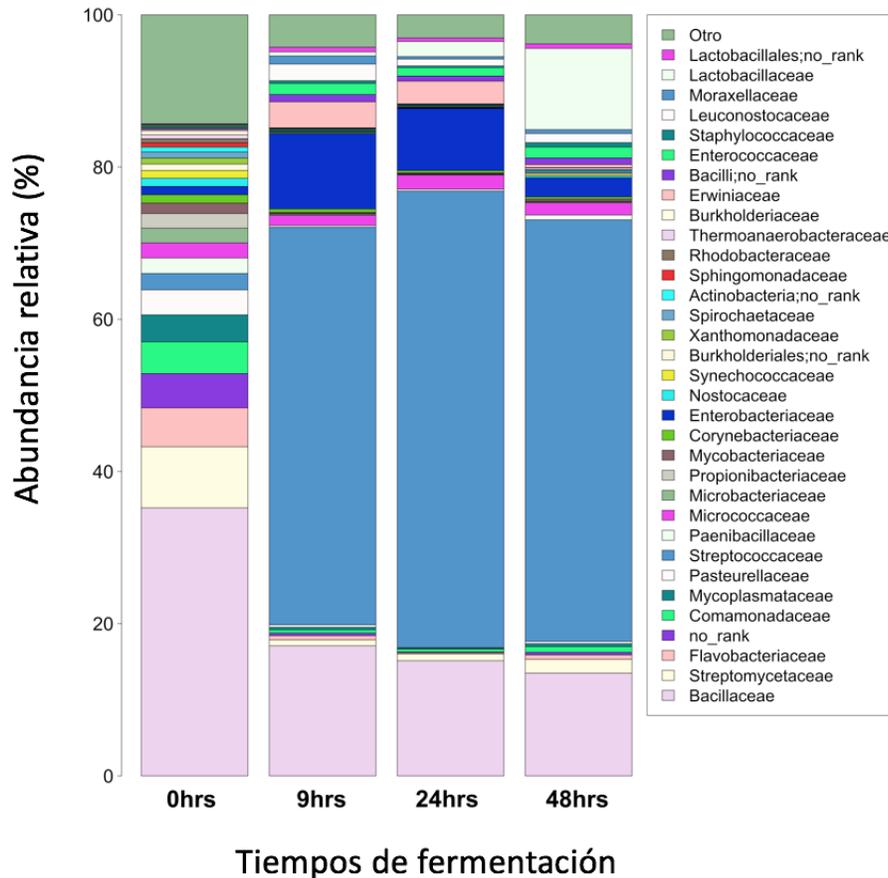


Figura 27. Abundancia relativa de las principales familias microbianas en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 356 familias (Figura 27).

A las 0 horas podemos ver que las familias más predominantes son Bacillaceae (35.21%), Streptomyetaceae (8.04%), Flavobacteriaceae (5.10%), Comamonadaceae (4.17%), Mycoplasmataceae (3.55%), Pasteurellaceae(3.28%), Streptococcaceae (2.16%), Paenibacillaceae (2.04%), Micrococcaceae (1.97%), Microbacteriaceae (1.94%), Propionibacteriaceae (1.93%), Mycobacteriaceae (1.37%), Corynebacteriaceae (1.10%), Enterobacteriaceae(1.09%), Nostocaceae (1.08%) y Synechococcaceae (1.01%).

A las 9 horas la familia que se encuentra en más abundancia es Streptococcaceae (52.26%), seguida de Bacillaceae (17.09%), Enterobacteriaceae (9.88%), Erwiniaceae (3.43%), Leuconostocaceae (2.18%), Enterococcaceae (1.47%), Micrococcaceae (1.35%) y Moraxellaceae (1.07%).

A las 24 horas Streptococcaceae continúa siendo en abundancia más de la mitad presente en este tiempo (59.98%) seguido de Bacillaceae (15.13%), Enterobacteriaceae (8.15%), Erwiniaceae (2.99%), Lactobacillaceae (1.98%), Micrococcaceae (1.87%) y Enterococcaceae (1.11%).

Y finalmente, a las 48 horas disminuye un poco la familia Streptococcaceae (55.40%); sin embargo sigue siendo la más abundante, seguida de Bacillaceae (13.50%), Lactobacillaceae (10.67%), Enterobacteriaceae (2.51%), Streptomyetaceae (1.81%), Micrococcaceae (1.60%), Enterococcaceae (1.44%) y Leuconostocaceae (1.19%).

Como Otro nos referimos a todas las familias cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes a 0, 9, 24 y 48 horas respectivamente.

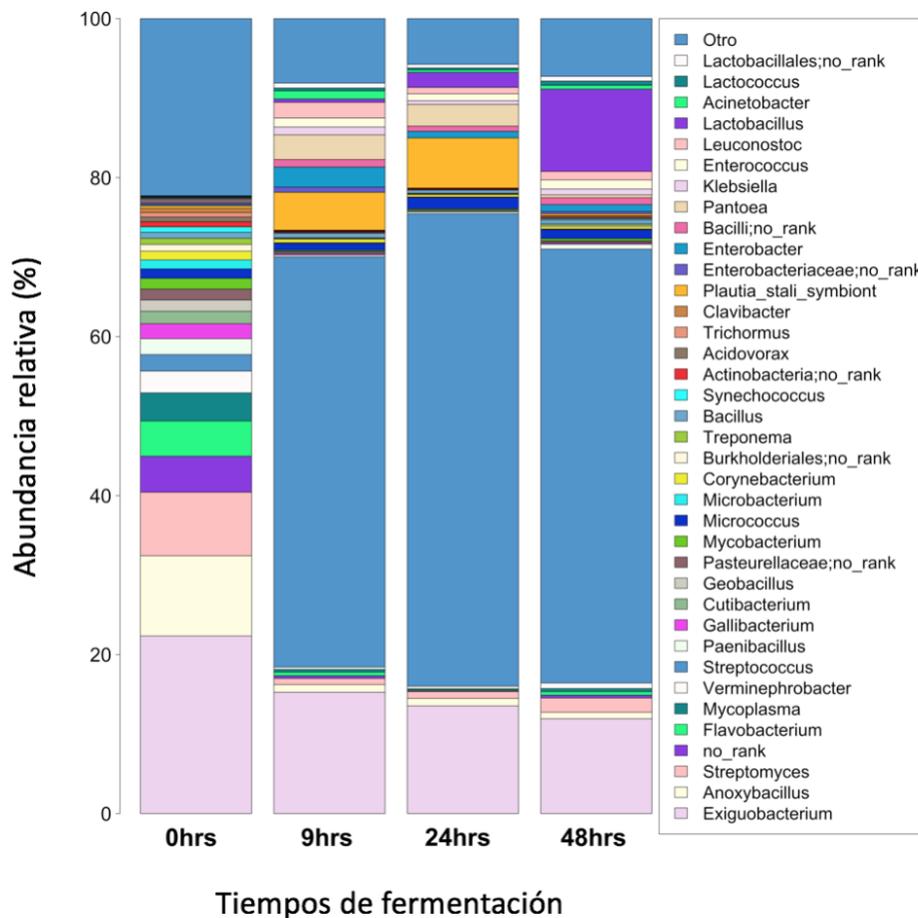


Figura 28. Abundancia relativa de los principales géneros microbianos en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 782 género (*Figura 28*).

A las 0 horas el género más abundante es *Exiguobacterium* (22.35%), seguido de *Anoxybacillus* (10.10%) *Streptomyces* (7.99%), *Flavobacterium* (4.44%), *Mycoplasma* (3.54%), *Verminephrobacter* (2.74%), *Streptococcus* (2.07%), *Paenibacillus* (2.00%), *Gallibacterium* (1.87%), *Cutibacterium* (1.57%), *Geobacillus* (1.42%), *Mycobacterium* (1.36%), *Micrococcus* (1.17%), *Microbacterium* (1.14%) y *Corynebacterium* (1.10%).

A 9 horas *Streptococcus* aumenta a más de la mitad en abundancia de los géneros presentes (51.59%), seguido de *Exiguobacterium* (15.26%), *Plautia stali symbiont* (4.77%), *Pantoea* (3.10%), *Enterobacter* (2.53%), *Leuconostoc* (1.97%), *Enterococcus* (1.14%), *Acinetobacter* (1.02%).

A las 24 horas *Streptococcus* sigue siendo el género más abundante (59.43%), seguido de *Exiguobacterium* (13.55%), *Plautia stali symbiont* (6.31%), *Pantoea* (2.71%), *Lactobacillus* (1.87%) *Micrococcus* (1.45%).

A las 48 horas *Streptococcus* (54.58%), seguido de *Exiguobacterium* (11.92%), *Lactobacillus* (10.35%) , *Streptomyces* (1.80%), *Enterococcus* (1.15%), *Micrococcus* (1.13%), *Leuconostoc* (1.07%).

Como Otro nos referimos a todos los géneros cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes a 22.28%, 8.11%, 5.74% y 7.26% a 0, 9, 24 y 48 horas respectivamente.

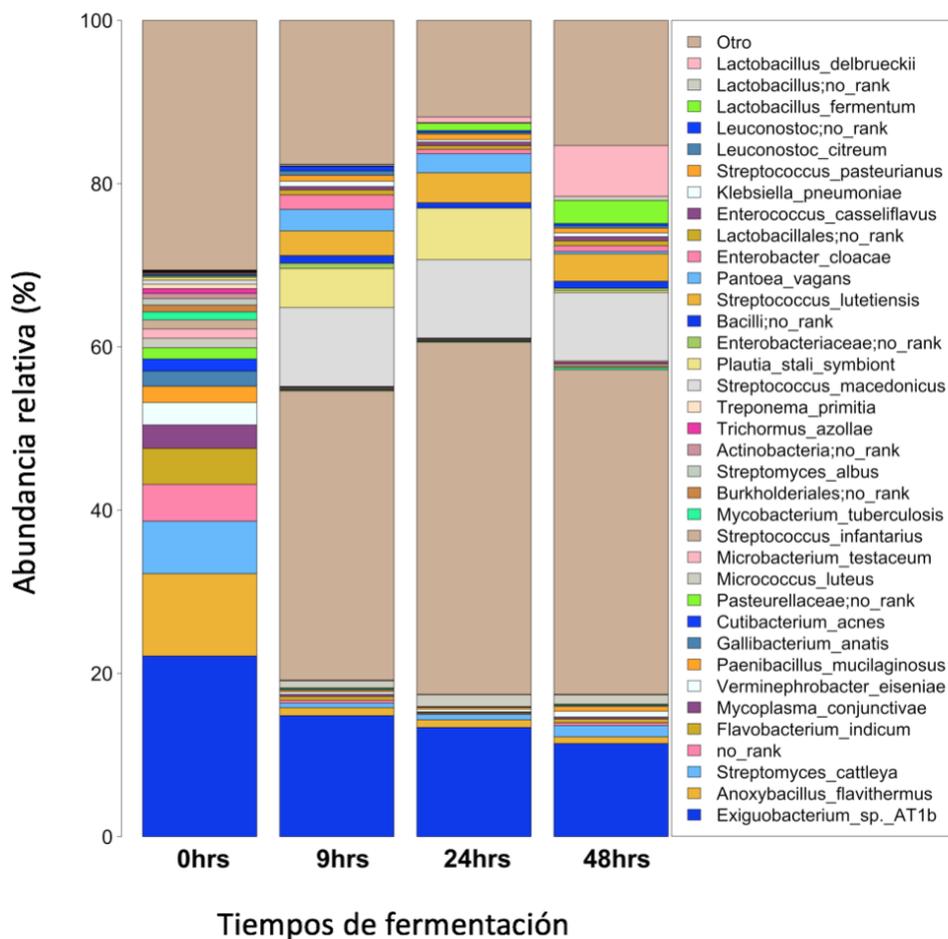


Figura 29. Abundancia relativa de las principales especies microbianas en la fermentación del pozol por el método de shotgun. La leyenda se lee de abajo hacia arriba.

Se obtuvieron un total de 1603 especies (Figura 29).

A las 0 horas *Exiguobacterium sp. AT1b* es la especie con mayor abundancia (22.13%), seguida de *Anoxybacillus flavithermus* (10.10%), *Streptomyces cattleya* (6.42%), *Flavobacterium indicum* (4.41%) *Mycoplasma conjunctivae* (2.87%), *Verminephrobacter eiseniae* (2.74%), *Paenibacillus mucilaginosus* (1.99%), *Gallibacterium anatis* (1.87%), *Cutibacterium acnes* (1.47%), *Micrococcus luteus* (1.17%), *Microbacterium testaceum* (1.13%) *Streptococcus infantarius* (1.11%).

A las 9 horas *Streptococcus infantarius* aumenta y se convierte en la especie más abundante (35.43%), seguida de *Exiguobacterium sp. AT1b* (14.82%), *Streptococcus macedonicus* (9.68%), *Plautia stali symbiont* (4.77%), *Streptococcus lutetiensis* (3.02%), *Pantoea vagans* (2.64%) y *Enterobacter cloacae* (1.77%).

A las 24 horas *Streptococcus infantarius* continua siendo la especie más abundante (43.15%), seguida de *Exiguobacterium sp. AT1b* (13.37%), *Streptococcus macedonicus* (9.61%), *Plautia stali symbiont* (6.31%), *Streptococcus lutetiensis* (3.68%) *Pantoea vagans* (2.32%) y *Micrococcus luteus* (1.45%).

A las 48 horas *Streptococcus infantarius* continúa siendo la especie más abundante (39.79%), seguida de *Exiguobacterium sp. AT1b* (11.41%), *Streptococcus macedonicus* (8.35%), *Lactobacillus delbrueckii* (6.23%), *Streptococcus lutetiensis* (3.35%), *Lactobacillus fermentum* (2.81%), *Streptomyces catleya* (1.39%) y *Micrococcus luteus* (1.13%).

Como Otro nos referimos a todas las especies cuya abundancia relativa es menor al 0.5% de abundancia en todos los tiempos de la fermentación y se agrupan dando los siguientes porcentajes: 30.59%, 17.63%, 11.83% y 15.33% a 0, 9, 24 y 48 horas respectivamente.

Las matrices con los porcentajes de abundancia relativa completos están en el material adicional de este escrito.

2.3 Taxonomía de Amplicones vs Shotgun

De los principales generos encontrados en cada uno de los experimentos (mayores al 0.5% en abundancia relativa) se obtuvieron 23 géneros en la taxonomía obtenida por medio del estudio de 16S y 30 géneros de la taxonomía obtenida por shotgun. 13 géneros son compartidos por ambos experimentos (*Figura 30*).

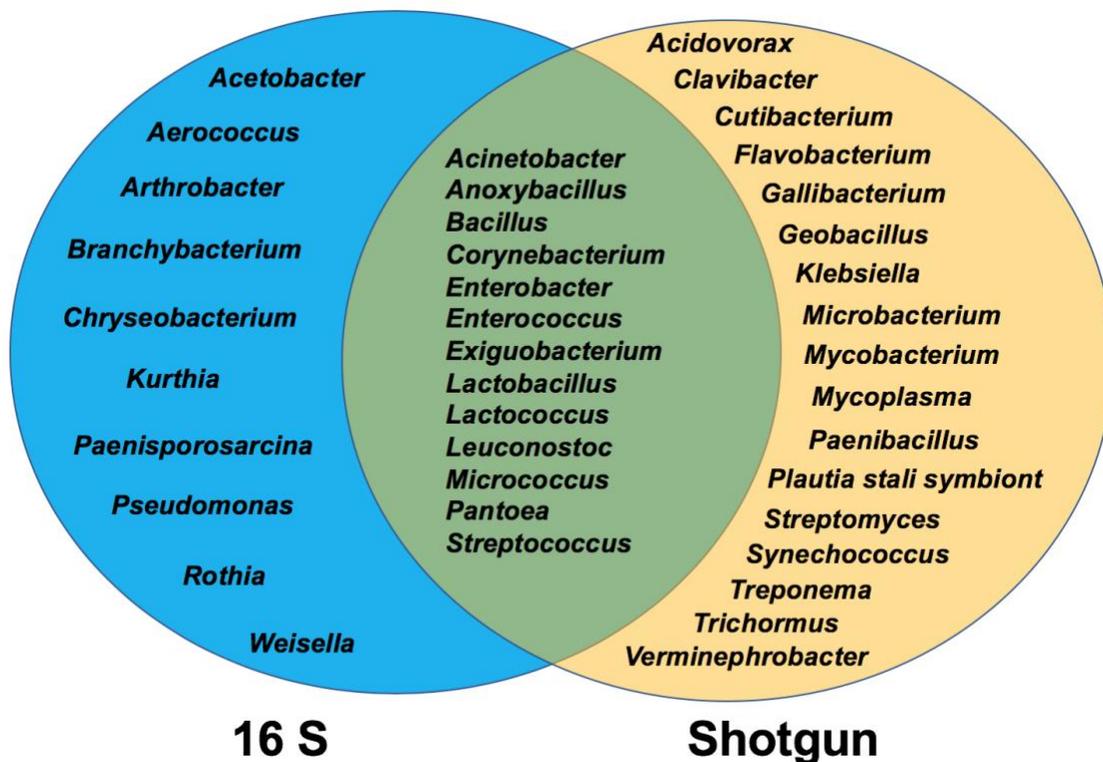


Figura 30. Diagrama de Venn con los principales géneros del pozol en 16S y shotgun.

A continuación, se mostrarán unas gráficas de los géneros que se comparten en ambos experimentos y su dinámica a través del tiempo en los experimentos de 16S y shotgun.

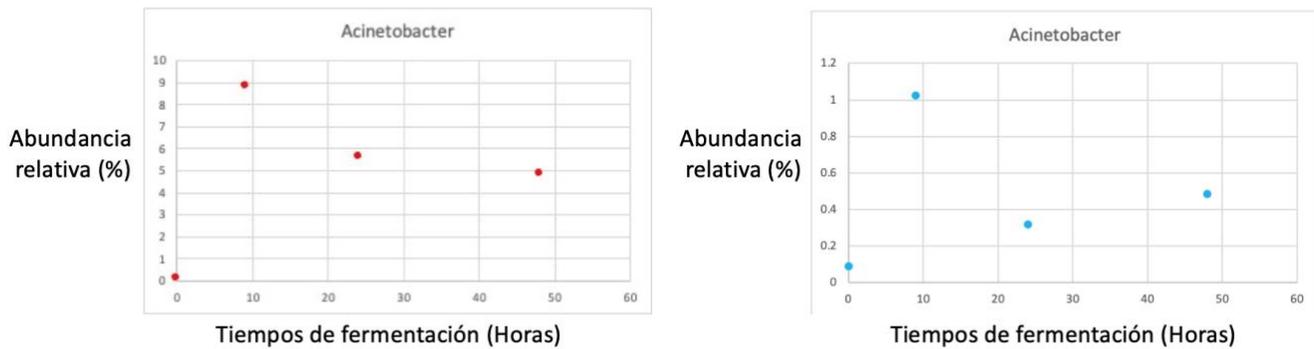


Figura 31. Abundancia relativa de *Acinetobacter* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 31* en ambos casos la abundancia del género *Acinetobacter* tiene su máximo a las 9 horas de la fermentación y su mínimo a las 0 horas.

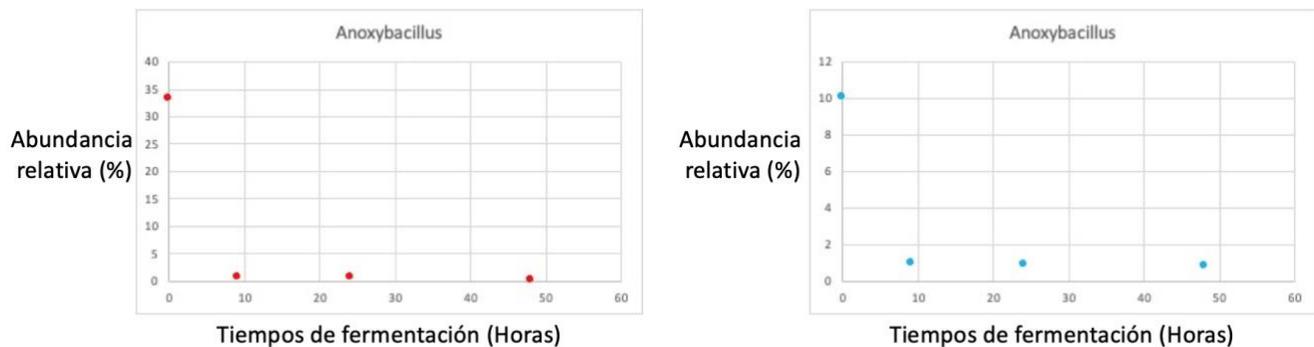


Figura 32. Abundancia relativa de *Anoxybacillus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 32* en ambos casos la abundancia del género *Anoxybacillus* tiene su máximo a las 0 horas de la fermentación y su mínimo a las 48 horas.

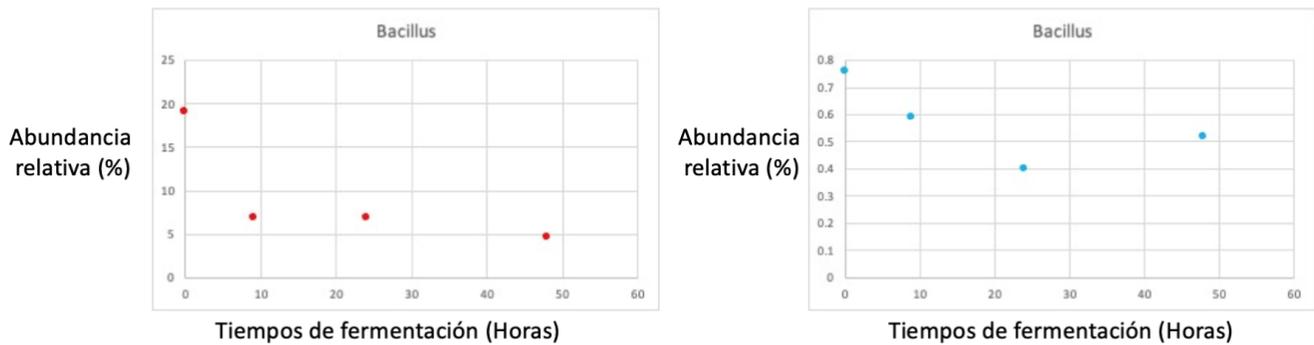


Figura 33. Abundancia relativa de *Bacillus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 33*, la abundancia del género *Bacillus* varía de manera significativa, sin embargo, en ambos casos tiene su máximo a las 0 horas de la fermentación y su mínimo a las en los experimentos de shotgun es a las 24 horas, mientras que en 16S es a las 48 horas.

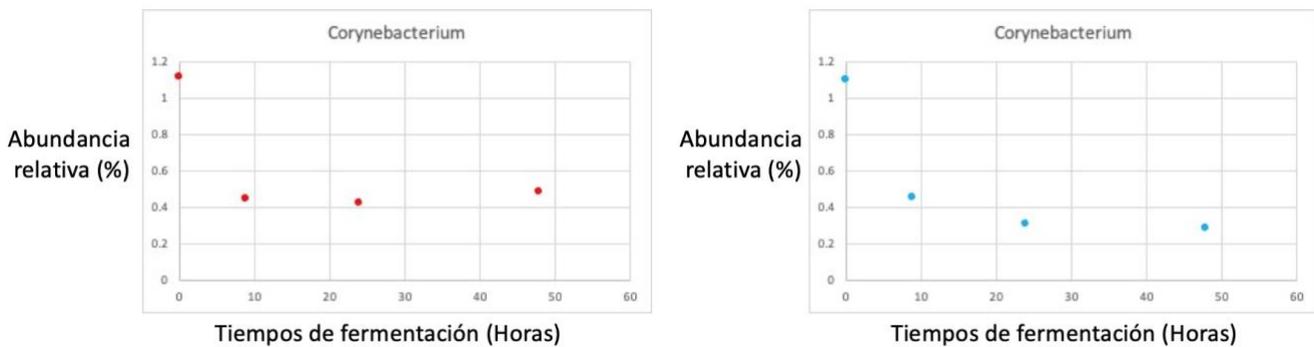


Figura 34. Abundancia relativa de *Corynebacterium* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 34*, la abundancia del género *Corynebacterium* tiene su máximo a las 0 horas de la fermentación y su mínimo a las en los experimentos de shotgun es a las 48 horas, mientras que en 16S es a las 24 horas.

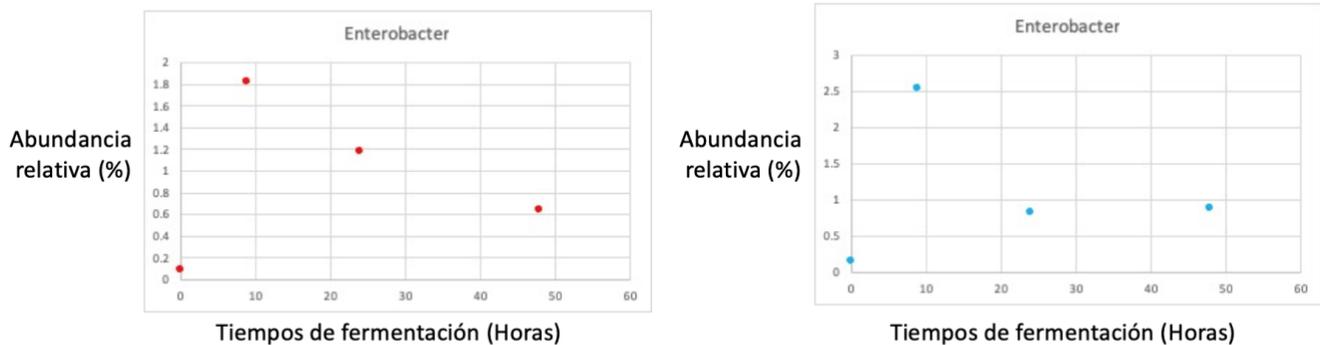


Figura 35. Abundancia relativa de *Enterobacter* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 35* en ambos casos la abundancia del género *Enterobacter* tiene su máximo a las 9 horas de la fermentación y su mínimo a las 0 horas.

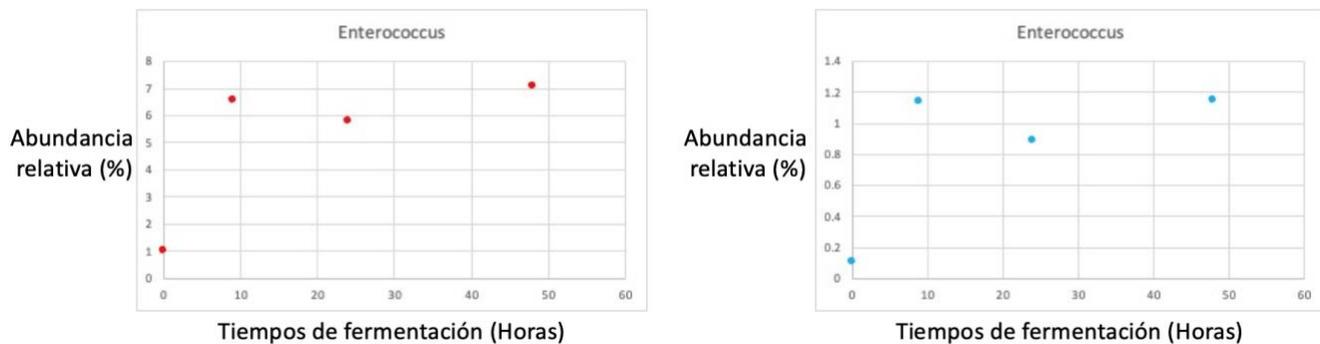


Figura 36. Abundancia relativa de *Enterococcus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 36*, en ambos casos la abundancia del género *Enterococcus* tiene su máximo a las 9 horas de la fermentación y su mínimo a las 0 horas.

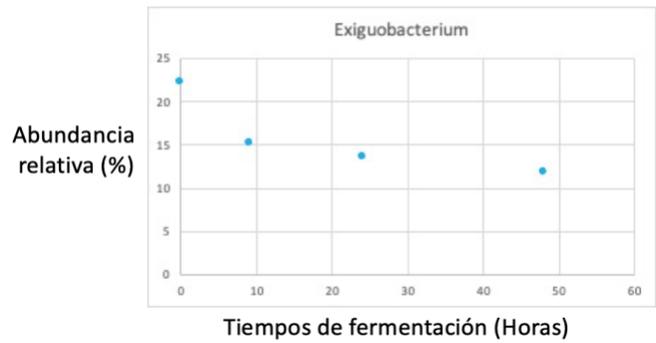
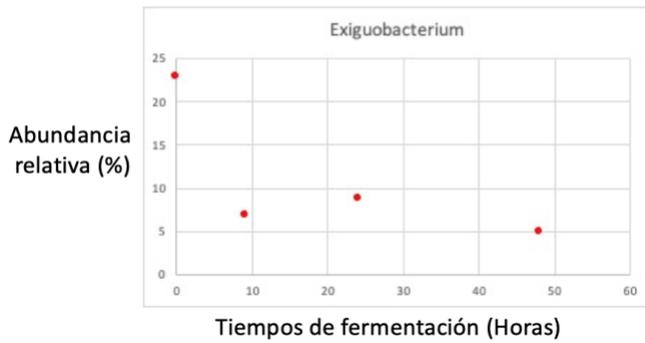


Figura 37. Abundancia relativa de *Exiguobacterium* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 37*, en ambos casos la abundancia del género *Exiguobacterium* tiene su máximo a las 0 horas de la fermentación y su mínimo a las 48 horas.

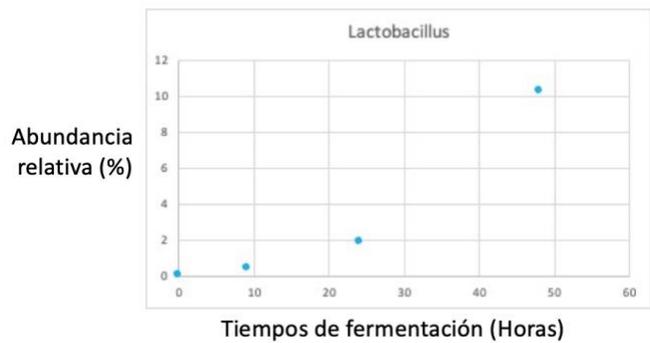
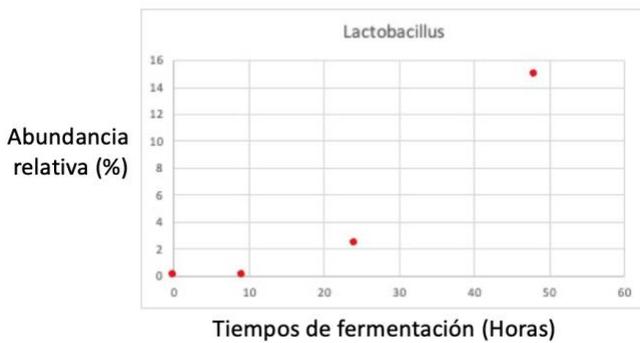


Figura 38. Abundancia relativa de *Lactobacillus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 38*, en ambos casos la abundancia del género *Lactobacillus* tiene su máximo a las 48 horas de la fermentación y su mínimo a las 0 horas.

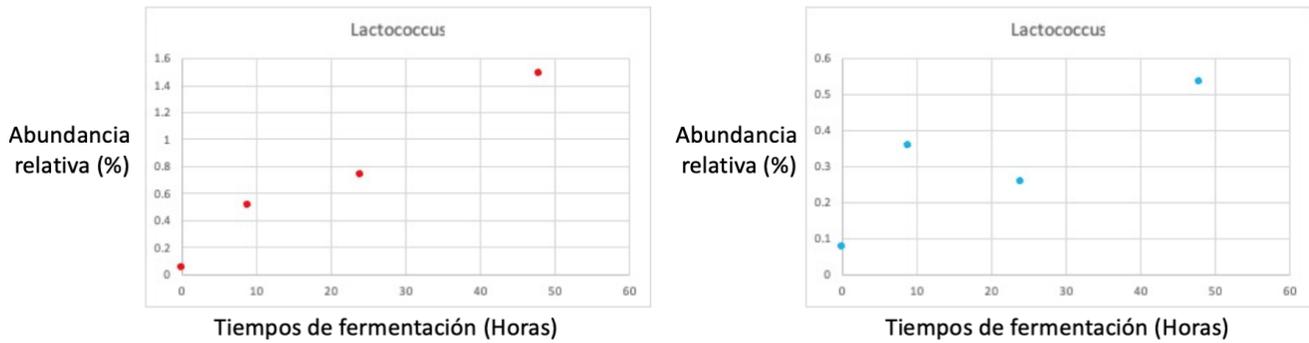


Figura 39. Abundancia relativa de *Lactococcus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 39*, en ambos casos la abundancia del género *Lactococcus* tiene su máximo a las 48 horas de la fermentación y su mínimo a las 0 horas.

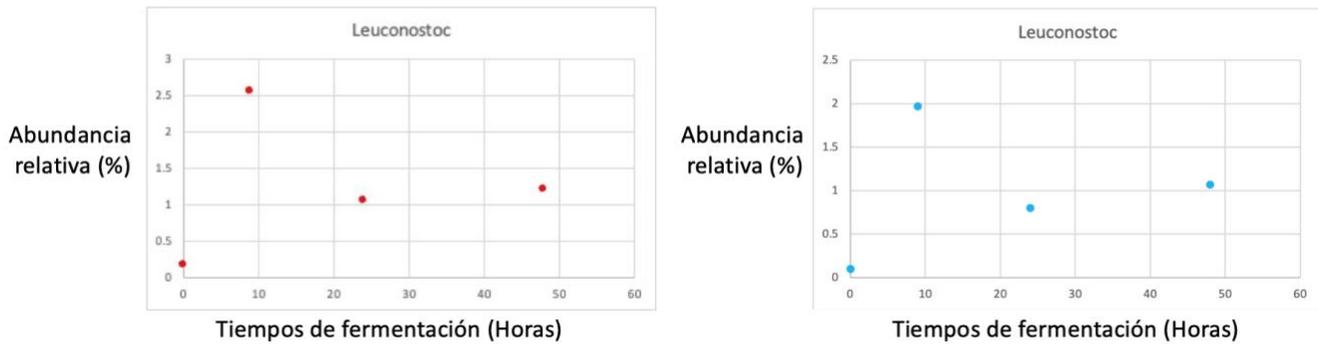


Figura 40. Abundancia relativa de *Leuconostoc* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 40*, en ambos casos la abundancia del género *Leuconostoc* tiene su máximo a las 9 horas de la fermentación y su mínimo a las 0 horas.

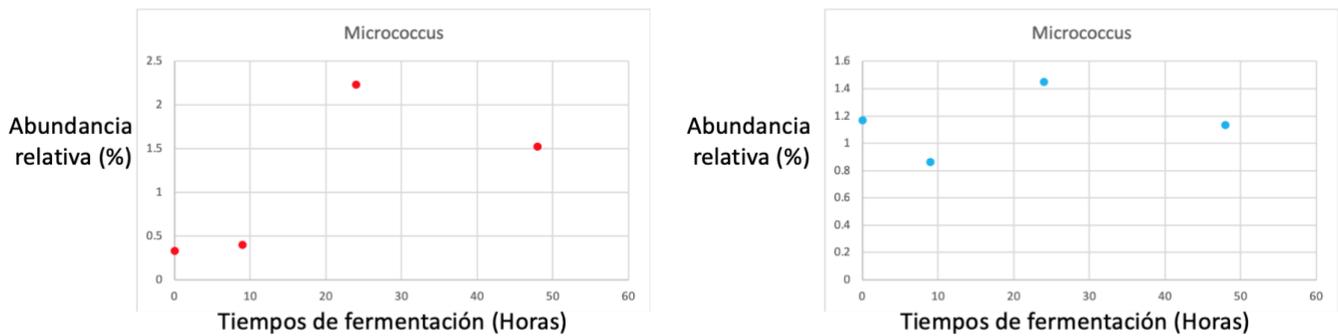


Figura 41. Abundancia relativa de *Micrococcus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 41*, en ambos casos la abundancia del género *Micrococcus* tiene su máximo a las 24 horas de la fermentación y su mínimo a las 0 horas en 16S y a las 9 horas en shotgun.

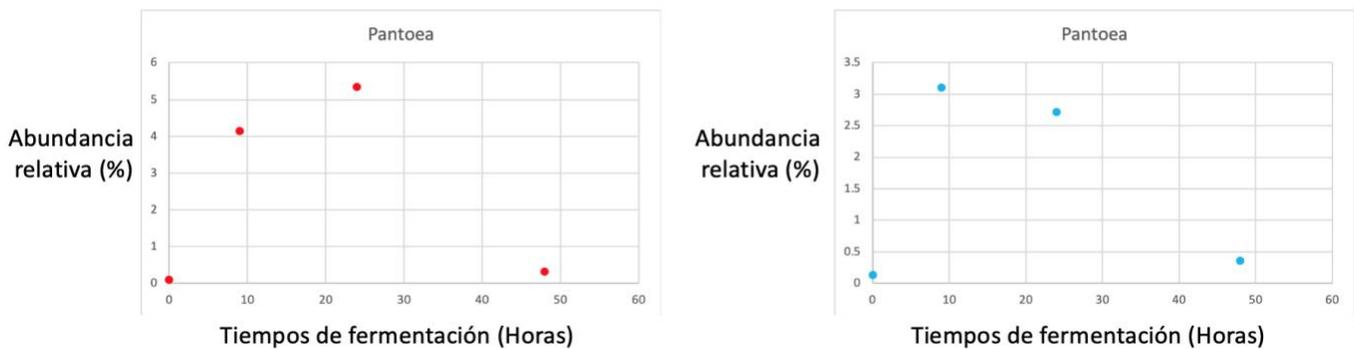


Figura 42. Abundancia relativa de *Pantoea* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 42*, la abundancia del género *Pantoea* tiene su máximo a las 24 horas en 16S y a las 9 horas en shotgun de la fermentación y su mínimo en ambos casos es a las 0 horas.

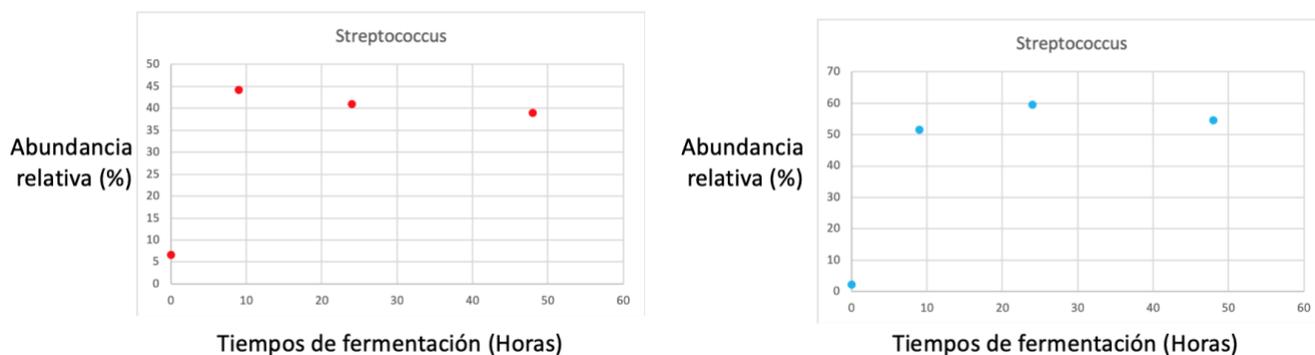


Figura 43. Abundancia relativa de *Streptococcus* a los distintos tiempos de fermentación. De rojo la abundancia de 16S. De azul la abundancia de shotgun.

Como se muestra en la *Figura 43*, la abundancia del género *Streptococcus* tiene su máximo a las 9 horas en 16S y a las 24 horas en shotgun de la fermentación y su mínimo en ambos casos es a las 0 horas.

2.4 Probióticos.

A continuación, presentamos un gráfico con las especies bacterianas que encontramos de manera abundante en los distintos tiempos de fermentación y que podrían tener potencial acción como probióticos.

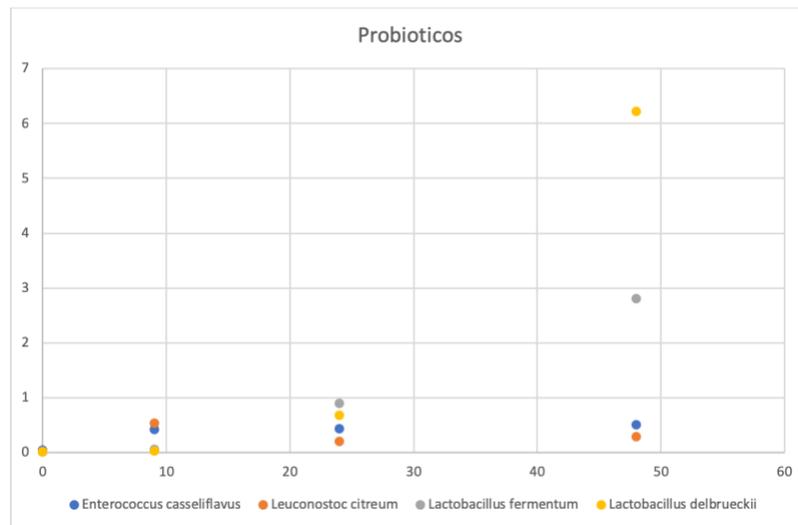


Figura 44. Abundancia relativa de especies con potencial probiotico en el pozo a los distintos tiempos de fermentación. De azul *Enterococcus casseliflavus*, de naranja *Leuconostoc citreum*, de gris *Lactobacillus fermentum* y de amarillo *Lactobacillus delbrueckii*.

Como se presenta en la *Figura 44*, *Enterococcus casseliflavus* tiene un máximo a las 24 horas y un mínimo al principio de la fermentación, *Leuconostoc citreum* tiene un máximo a las 9 horas y un mínimo a las 0 horas, *Lactobacillus fermentum* y *Lactobacillus delbrueckii* tienen un máximo a las 48 horas y un mínimo a las 0 horas.

3. Ensamblado

IDBA-UD	0hrs	9hrs	24hrs	48hrs
Número total de contigs	8,851	39,205	41,402	56,381
N50	996	3,092	2,504	1,374
Contig más largo	45,885	103,395	98,371	125,224
Contig más corto	200	200	200	200

Tabla 9. Estadísticas del ensamblado con el programa IDBA-UD.

Megahit	0hrs	9hrs	24hrs	48hrs
Número total de contigs	9,241	44,977	46,199	61,142
N50	900	2,518	2,176	1,215
Contig más largo	45,879	117,937	111,437	114,233
Contig más corto	201	200	200	200

Tabla 10. Estadísticas del ensamblado con el programa Megahit.

Tiempo (hrs)	Lecturas ensambladas (%)
0	28.77
9	83.06
24	87.05
48	78.92

Tabla 11. Porcentaje de lecturas crudas ensambladas con el programa IDBA-UD.

Se hizo una comparación entre los dos ensambladores que utilizamos para ver cual nos daba mejores resultados y llegamos a la conclusión de que IDBA-UD (Tabla 9 y Tabla 10) es el que da mejor resultados ya que se obtiene un menor número de contigs y un N50 más grande en todos los tiempos de fermentación; por lo tanto, utilizamos este ensamble para mapearlo contra nuestras secuencias sin maíz y obtener el porcentaje de lecturas que sí ensamblaron.

Como podemos observar en la tabla 0 horas tiene un porcentaje bajo de lecturas ensambladas, mientras que los otros tres tiempos mapean más del 75% de las lecturas (Tabla 11).

4. Anotación funcional

Se realizó la predicción de genes con los ORFs mediante el programa MetaGeneMark para posteriormente correr los resultados en la base de datos de dbCAN y así ver cuales de esos genes predichos tenían una firma única para CAZymes con la finalidad de poder anotar la función correspondiente. Los resultados se reportan en la Tabla 12:

Tiempo (horas)	Genes predichos	Genes anotados con dbCAN	Porcentaje (%)
0	13,419	2,491	18.56
9	76,735	15,338	19.98
24	83,917	17,328	20.64
48	76,162	13,528	17.76

Tabla 12. Porcentaje de genes anotados con dbCAN en base a los genes predichos.

Las CAZymes de todas las clases constituyeron 18.56, 19.98, 20.64 y 17.76 por ciento de todos los genes predichos a 0,9,24 y 48 hrs respectivamente (Tabla 11).

Tiempo (horas)	GH	CE	CBM	AA	PL
0	1,688	2137	150	207	23
9	9146	1040	700	1,251	291
24	10,324	1,224	743	1,601	312
48	8,456	828	582	1,153	221

Tabla 13. Genes anotados para CAZymes a los distintos tiempos de fermentación.

De todas esas CAZymes que están en la Tabla 13, se hizo una selección de las familias que eran de interés para nuestra investigación, basándonos en los carbohidratos que podrían fungir como sustrato en la fermentación del pozol, los cuales son almidón, celulosa, hemicelulosa y sacarosa. La lista de familias se redujo a 86 siendo 47 GHs, 10 CEs, 18 CBMs, 7 AAs y 4 PLs.

Haciendo una clasificación por el carbohidrato en el cual actúan estas familias obtuvimos lo siguiente:

Para Almidón tenemos 12 familias y son las siguientes: CBM20,CBM26,CBM34, CBM41,CBM68, GH126, GH13, GH14, GH15, GH57, GH76 y GH77.

Para hemicelulosa y celulosa tenemos 50 familias y son las siguientes: CBM1, CBM11, CBM22, CBM35, CBM37, CBM4, CBM56, CBM76, CE1, CE10, CE12, CE16, CE3, CE4, CE5, CE7, CE8, GH1, GH10, GH11, GH120, GH125, GH128, GH130, GH131, GH16, GH17, GH2, GH28, GH3, GH30, GH31, GH36, GH4, GH43, GH5, GH53, GH55, GH6, GH67, GH7, GH71, GH74, GH79, GH8, GH94, PL12, PL22, PL7 y PL9.

Para sacarosa tenemos 8 familias que son las siguientes: CBM24, CBM61, CBM66, GH32, GH66, GH68, GH70, GH72.

Para otros carbohidratos tenemos 9 familias que son: CBM32, CBM67, CE14, GH114, GH135, GH35, GH37, GH42 y GH65.

Y por último tenemos las enzimas auxiliares que ayudan a la degradación de varios carbohidratos y son las siguientes: AA1, AA12, AA2, AA3, AA4, AA7 y AA9.

Con estas enzimas realizamos un mapa de calor para ver su abundancia relativa en los distintos tiempos de nuestra fermentación.

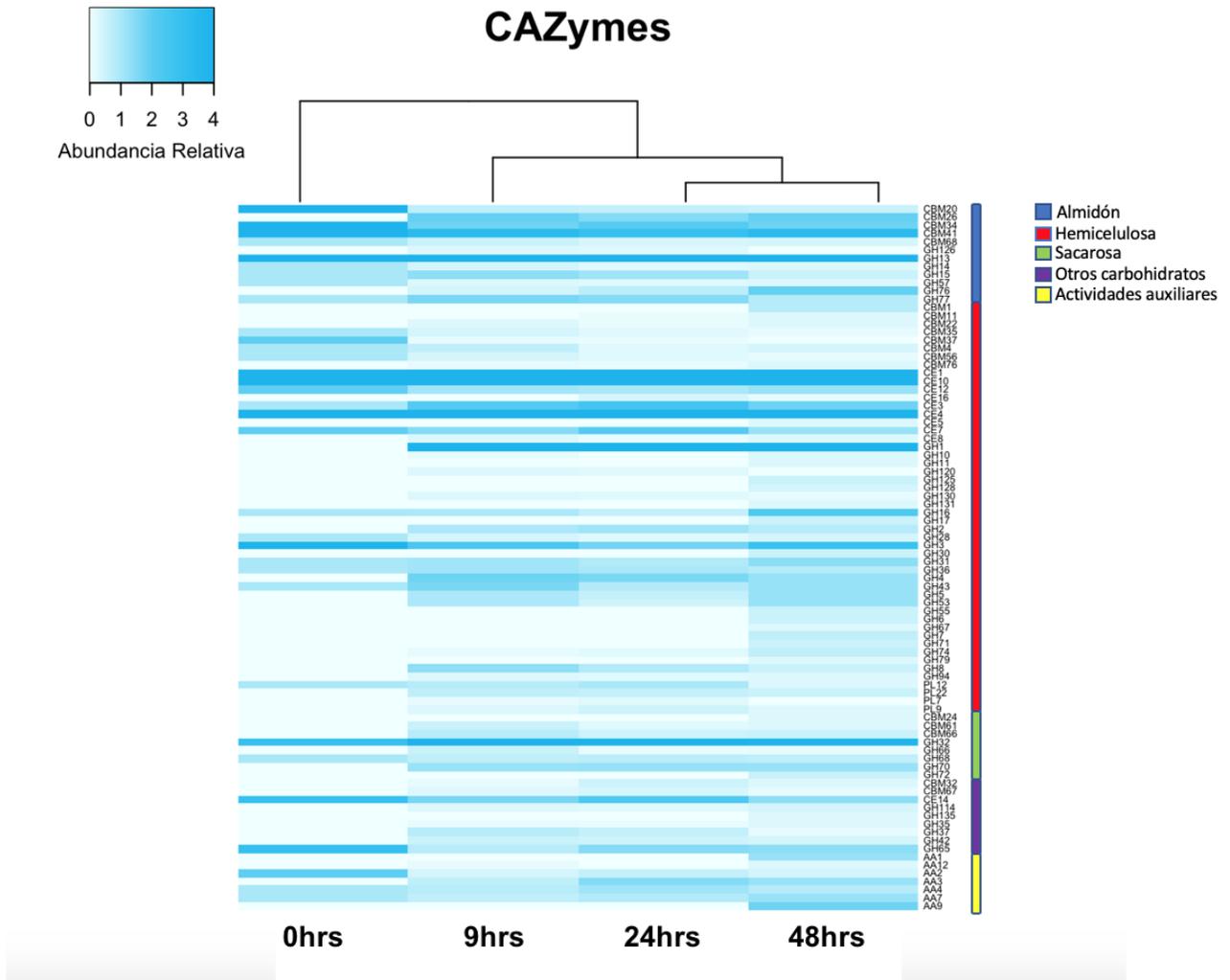


Figura 45. Mapa de calor de la abundancia relativa en porcentaje de las CAZymes en los distintos tiempos de la fermentación del pozol.

En el mapa de calor (*Figura 45*) podemos ver lo siguiente:

A las 0 horas la familia con más abundancia es GH13 (27.72%), seguida de CE4 (8.91%), CE1 (7.92%), CBM20 (4.95%), GH3 (3.96%), CE10 (3.96%), CBM41 (3.96%), CBM34 (3.96%), GH65 (2.97%), GH32 (2.97%) y CE14 (2.97%).

A las 9 horas la familia con más abundancia es GH13 (18.43%), seguido de CE1 (10.98%), GH1 (7.18%), CE10 (7.18%), CE4 (5.01%) GH32 (3.93%), CBM41 (3.39%), GH3 (2.44%) y CE3 (2.17%).

A las 24 horas la familia con más abundancia es GH13 (19.78%) CE1 (10.75%), GH1 (8.03%), CE10 (7.05%), GH32 (4.45%), CE4 (4.45%), CBM41 (2.97%), CE3 (2.60%), CE14 (2.35%), CE7 (2.10%) y CBM34 (2.10%).

A las 48 horas la familia con más abundancia es GH13 (16.51%), CE1 (7.37%), CE10 (6.28%), GH1 (5.73%), CE4 (3.96%), GH32 (3.82%), CBM41, (3.41%) GH3 (2.86%) y GH16 (2.32%).

Con esta información decidimos buscar las especies a las cuales pertenecían las secuencias anotadas a los genes que predecían a esas familias. El criterio que utilizamos es que fueran mayor al 2% de identidad en por lo menos uno de los cuatro tiempos. Bajo este criterio acotamos las tablas 15, 16 y 17:

CAZyme	0 horas	9 horas	24 horas	48 horas
CBM20	<i>Anoxybacillus flavithermus</i>	-----	-----	<i>Lactobacillus fermentum</i> <i>Streptococcus infantarius</i>
CBM34	<i>Anoxybacillus flavithermus</i>	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i>
CBM41	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i>	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>

Tabla 15. Especies asociadas a las familias de CBMs.

CAZyme	0 horas	9 horas	24 horas	48 horas
CE1	<i>Anoxybacillus flavithermus</i> <i>Exiguobacterium</i> sp. AT1b	<i>Enterobacter cloacae</i> <i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Streptococcus infantarius</i>
CE3	<i>Anoxybacillus flavithermus</i> <i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Streptococcus infantarius</i>
CE4	<i>Anoxybacillus flavithermus</i> <i>Exiguobacterium</i> sp. AT1b	<i>Enterobacter cloacae</i> <i>Exiguobacterium</i> sp. AT1b <i>Streptococcus lutetiensis</i> <i>Streptococcus infantarius</i>	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Streptococcus infantarius</i>
CE7	<i>Anoxybacillus flavithermus</i> <i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus lutetiensis</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Streptococcus lutetiensis</i> <i>Streptococcus macedonicus</i>
CE10	<i>Anoxybacillus flavithermus</i> <i>Exiguobacterium</i> sp. AT1b	<i>Enterobacter cloacae</i> <i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus lutetiensis</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i>
CE14	<i>Exiguobacterium</i> sp. AT1b <i>Anoxybacillus flavithermus</i>	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i>

Tabla 16. Especies asociadas a las familias de CEs.

CAZyme	0 horas	9 horas	24 horas	48 horas
GH1	<i>Anoxybacillus flavithermus</i> <i>Exiguobacterium</i> sp. AT1b	<i>Enterobacter cloacae</i> <i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus lutetiensis</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Streptococcus infantarius</i> <i>Streptococcus lutetiensis</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Lactobacillus delbrueckii</i> <i>Lactobacillus plantarum</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>
GH3	<i>Exiguobacterium</i> sp. AT1b	<i>Enterobacter cloacae</i> <i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i>
GH13	<i>Anoxybacillus flavithermus</i> <i>Cutibacterium acnes</i> <i>Exiguobacterium</i> sp. AT1b	<i>Enterobacter cloacae</i> <i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i> <i>Pantoea vagans</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Micrococcus luteus</i> <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>
GH16	-----	<i>Streptococcus infantarius</i>	<i>Streptococcus infantarius</i>	<i>Streptococcus infantarius</i>
GH32	<i>Anoxybacillus flavithermus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Streptococcus infantarius</i> <i>Streptococcus macedonicus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Streptococcus infantarius</i>
GH65	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b	<i>Exiguobacterium</i> sp. AT1b <i>Micrococcus luteus</i>	<i>Exiguobacterium</i> sp. AT1b <i>Lactobacillus delbrueckii</i> <i>Lactobacillus fermentum</i> <i>Micrococcus luteus</i>

Tabla 17. Especies asociadas a las familias de GHs.

La matriz completa con la abundancia relativa de las CAZymes se encuentra en el material suplementario de este escrito.

5. Reconstrucción Metabólica

Se asignaron números de identificación de KEGG a más del 30% de todos los ORFs predichos en nuestros cuatro tiempos y se pusieron en distintos mapas de rutas metabólicas mediante el GhostKOALA (<https://www.kegg.jp/ghostkoala/>). Se pudo hacer la reconstrucción de prácticamente todo el metabolismo de Carbono Central, el de Aminoácidos, el de síntesis de Lípidos, cofactores y vitaminas. En resumen se pudo reconstruir con nuestro metagenoma completo prácticamente todo el mapa metabólico. (*Figura 28*)

Además se buscaron los módulos de síntesis de los aminoácidos esenciales, así como la vitamina riboflavina, que tienen aumento en el pozol en comparación con los de la masa cruda y que le confieren propiedades nutricionales importantes.

Se pudo ver que existen los módulos completos de biosíntesis de treonina (M00018 de KEGG Modules), lisina (M00016 y M00525 de KEGG Modules) Isoleucina (M00019, M00535 y M00570 de KEGG Modules) triptofano (M00023 de KEGG Modules), Arginina (M00844 y M00845 de KEGG Modules) y riboflavina (M00125 de KEGG Modules) en la fermentación del pozol. (*Figura 45*)

En base a todo esto se hizo un mapa donde se reconstuyó el modelo de fermentación y se resaltaron las rutas metabólicas en las que están involucrados esos aminoácidos vitamina.

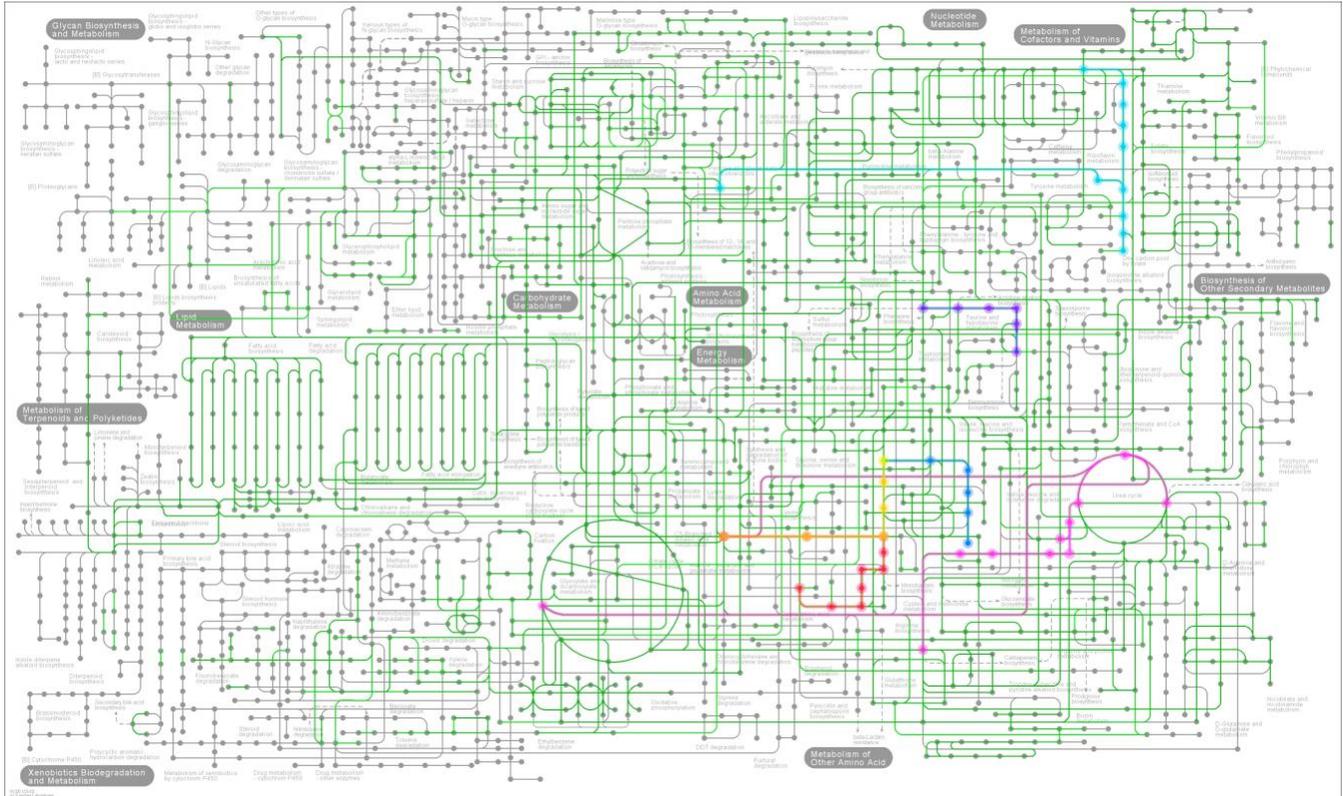


Figura 45. Atlas de Kegg que representa las rutas metabólicas encontradas en el metagenoma del pozol. Se resaltaron las rutas que son de biosíntesis de aminoácidos esenciales y la vitamina riboflavina. De azul claro la biosíntesis de riboflavina, de morado la biosíntesis de triptofano, de rosa la biosíntesis de arginina, de rojo la biosíntesis de lisina, de azul marino la biosíntesis de isoleucina, de amarillo la biosíntesis de treonina. De naranja el translope entre las síntesis de treonina y lisina.

Discusión

La fermentación del pozol como ya se ha dicho con anterioridad es un modelo de estudio complejo. Por los parámetros obtenidos de alfa y beta diversidad sabemos que nuestra fermentación empieza siendo no muy diversa y muy heterogénea, para ir volviéndose más homogénea y diversa con el transcurso del tiempo. También sabemos por medio de la beta diversidad que la fermentación tiene una microbiota más parecida en el transcurso de las 9 a las 24 horas y que no es tan parecida en el principio de la fermentación y a las 48 horas de la misma.

Se hicieron estudios comparativos a nivel taxonómico de las muestras mediante secuenciación por el gen ribosomal 16S y el método de shotgun.

En el caso de 16S los generos más representativos en la fermentación *Streptococcus*, *Lactobacillus*, *Exiguobacterium*, *Anoxybacillus*, *Streptomyces*, *Flavobacterium*, *Mycoplasma*, *Verminephrobacter*, *Paenibacillus*, *Gallibacterium*, *Cutibacterium*, *Geobacillus*, *Mycobacterium*, *Leuconostoc*, *Enterococcus* y *Micrococcus*.

En el caso de Shotgun los géneros más representativos en la fermentación son *Streptococcus*, *Lactobacillus*, *Exiguobacterium*, *Anoxybacillus*, *Streptomyces*, *Flavobacterium*, *Mycoplasma*, *Enterococcus*, *Leuconostoc* y *Pantoea*.

Mediante estos métodos pudimos ver qué géneros bacterianos eran los mismos en ambos experimentos *Acinetobacter*, *Anoxybacillus*, *Bacillus*, *Corynebacterium*, *Enterobacter*, *Enterococcus*, *Exiguobacterium*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Micrococcus*, *Pantoea* y *Streptococcus*.

Al principio de la fermentación podemos ver que los géneros que predominan son *Exiguobacterium*, *Anoxybacillus*, *Streptococcus* y *Micrococcus*, siendo los dos primeros los más abundantes.

Conforme pasa el tiempo, a las 9 horas podemos ver que el género *Streptococcus* tiene un incremento sustancial en la fermentación mismo que continúa prevaleciendo, convirtiéndose en el género más importante hasta el último tiempo de la fermentación. El género *Exiguobacterium* disminuye, pero prevalece el resto de la fermentación, mientras que *Anoxybacillus* deja de ser significativo de las 9 horas en adelante. *Pantoea*, *Enterobacter*, *Leuconostoc*, *Enterococcus* y *Acinetobacter* son otros géneros importantes en este tiempo.

A las 24 horas *Streptococcus* como ya se ha dicho con anterioridad sigue siendo el género con más abundancia, seguido de *Exiguobacterium*. *Pantoea*, *Lactobacillus* y *Micrococcus* son otros géneros con buena presencia en este tiempo. *Streptococcus* y *Exiguobacterium* siguen siendo los géneros más abundantes. Otros géneros con más presencia en este tiempo son *Leuconostoc*, *Enterococcus* y *Micrococcus*.

El género *Lactobacillus*, empieza a tener una presencia significativa a partir del tiempo de 24 horas y alcanza un máximo a las 48 horas de fermentación.

De dichos géneros pudimos hacer la asignación de especies con la anotación taxonómica de shotgun.

Del género *Exiguobacterium* la principal es *Exiguobacterium sp. AT1b*. De *Anoxybacillus* tenemos a *Anoxybacillus flavithermus*. Los *Streptococcus* que encontramos son *Streptococcus infantarius* en mayor proporción seguido de *Streptococcus macedonicus* y *Streptococcus lutetiensis*. En el género *Lactobacillus* las especies principales son *Lactobacillus Delbruecki* y *Lactobacillus fermentum*. Y por último de otros géneros con relevancia de *Micrococcus* *Micrococcus luteus*, de *Enterobacter*, *Enterobacter cloacae*, de *Pantoea*, *Pantoea vagans* y de *Leuconostoc*, *Leuconostoc citreum*.

De estas especies sabemos que *Lactobacillus delbruecki*, *Lactobacillus fermentum*, *Leuconostoc citreum* y *Enterococcus casseliflavus* tienen potencial probiótico y a excepción de *Leuconostoc citreum*, las otras tres especies predominan a tiempos tardíos de la fermentación, es decir de 24 horas en adelante.

En lo que respecta a las CAZymes encontradas mediante los genes predichos para la degradación de los carbohidratos de interés en la fermentación del pozol se hizo un análisis con las familias que degradan almidón, hemicelulosa y sacarosa en los distintos tiempos de fermentación.

En el caso del almidón se sabe que es prácticamente el 70% de la masa del pozol (Ampe et al., 1999).

Aquí encontramos que al principio de la fermentación, a las 0 horas la familia de GH que tiene mayor abundancia es GH13 donde se encuentran las enzimas α -amilasas (EC 3.2.1.1) que realizan la endohidrólisis of (1→4)- α -D-de uniones glicosídicas, las pululanases (EC 3.2.1.41) que actúan en la hidrólisis de (1→6)- α -D- uniones en pululano, amilopectina y glucógeno, y en los α - y β -límites de dextrinas en amilopectinas y glucógeno, la ciclomaltodextrinasa (EC 3.2.1.54) encargada de hidrolizar maltodextrina de forma linear y la amilasa maltogénica (EC 3.2.1.133) que actúa en almidón y cuyo producto es Acta α -maltosa. (<http://www.cazy.org>).

Se encontraron 557 genes de esta familia en nuestro metagenoma. Esta familia de enzimas tiene una gran abundancia en los 4 tiempos de la fermentación y sus genes se encuentran asociados a los contigs anotados a las especies de *Anoxybacillus flavithermus*, *Exiguobacterium sp. AT1b* a las 0 horas, a *Enterobacter cloacae*, *Exiguobacterium sp. AT1b*, *Micrococcus luteus*, *Streptococcus infantarius*, *Streptococcus macedonicus* y *Pantoea vagans* a las 9 horas, a *Exiguobacterium sp. AT1b*, *Micrococcus luteus*, *Streptococcus infantarius*, *Streptococcus macedonicus* y *Pantoea vagans* a las 24 horas y a *Exiguobacterium sp. AT1b*, *Lactobacillus delbrueckii*, *Lactobacillus fermentum*, *Micrococcus luteus*, *Streptococcus infantarius*, *Streptococcus macedonicus* a las 48 horas.

También están las familias de CBMs 20, 34 y 41. Las CBM20 y La CBM34 tienen la función de unir almidón. (<http://www.cazy.org>). Hay 36 genes en el metagenoma de la familia CBM20 y ésta es abundante al principio de la fermentación y a las 48 horas de la misma y las especies que se encontraron con los contigs fueron *Anoxybacillus flavithermus* a las 0 horas, *Lactobacillus fermentum* y *Streptococcus infantarius* a las 48 horas.

De la CBM34 se encontraron 51 genes en el metagenoma y es abundante en las 0 horas y a las 24 horas y las especies que se pudieron asociar a estos genes son *Exiguobacterium sp. AT1b* a las 0 horas, *Exiguobacterium sp. AT1b*, *Streptococcus infantarius* y *Streptococcus macedonicus* a las 24 horas.

La CMB41 une α -glucanos de amilosa, amilopectina, pululano y otros fragmentos de oligosacáridos derivados de estos polisacáridos. (<http://www.cazy.org>). Hay 42 genes en todo el metagenoma y está representada en todos los tiempos de la fermentación. Las especies asociadas a estos genes son *Exiguobacterium sp. AT1b* a 0 horas, *Exiguobacterium sp. AT1b* y *Streptococcus infantarius* a 9 horas, *Exiguobacterium sp. AT1b*, *Streptococcus infantarius* y *Streptococcus macedonicus* a 24 y 48 horas respectivamente.

Por otro lado, mediante la nixtamalización se remueven la gran mayoría de los carbohidratos solubles del maíz y se hacen disponibles algunos insolubles de la pared celular, como la hemicelulosa.

Se sabe que los polisacáridos no amiláceos del maíz son principalmente los arabinoxilanos ferulados y que éstos son un grupo heterogéneo de polisacáridos entre los cuales se encuentra el xilano, y cuyo monosacárido principal es la xilosa que está unida por enlaces glicosídicos β -(1 \rightarrow 4) en los que se unen residuos de arabinosa mediante α -(1 \rightarrow 3), α -(1 \rightarrow 2) o ambos. (Takao et al. 2002).

La xilosa presenta grados de sustitución monosustituídos o disustituídos y por lo tanto es común que los arabinoxilanos tengan residuos de celulosa, lignina, ácido glucorónico y galactosa. (Morales-Ortega et al. 2013).

Los genes asociados a las familias de GHs que encontramos en abundancia fueron de las familias GH1, GH3 y GH16; se encontraron 277,59,35 genes, respectivamente, en todo el metagenoma.

La familia GH1 tiene enzimas como la β -xilosidasa (EC 3.2.1.37) que realiza la hidrólisis de (1 \rightarrow 4)- β -D-xilanos, para remover los residuos sucesivos de D-xilosa de los extremos no reductores terminales. (<http://www.cazy.org>).

Esta familia tiene mucha abundancia a partir de las 9 horas de la fermentación hasta las 48 horas y las especies bacterianas que están asociadas a estos genes son *Anoxybacillus flavithermus* y *Exiguobacterium sp. AT1b* a las 0 horas, *Enterobacter cloacae*, *Exiguobacterium sp. AT1b*, *Streptococcus infantarius*, *Streptococcus lutetiensis*, *Streptococcus macedonicus* y *Pantoea vagans* a las 9 horas, *Streptococcus infantarius*, *Streptococcus lutetiensis*,

Streptococcus macedonicus y *Pantoea vagans* a las 24 horas, *Lactobacillus delbrueckii*, *Lactobacillus plantarum*, *Streptococcus infantarius* y *Streptococcus macedonicus* a las 48 horas.

La familia GH3 tiene enzimas como glucan 1,3- β -glucosidasa (EC 3.2.1.58) que realizan la hidrólisis sucesiva de unidades de β -D-glucosa de los extremos no reductores de (1 \rightarrow 3)- β -D-glucanos, liberando α -glucosa, α -L-arabinofuranosidase (EC 3.2.1.55) hidrólisis de residuos no reductores terminales de α -L-arabinofuranosidasa en α -L-arabinosidas. (<http://www.cazy.org>).

La familia GH3 tiene una abundancia importante al principio a las 0 horas y a las 48 horas y las especies asociadas a sus genes son *Exiguobacterium sp. AT1b* en ambos tiempos y *Lactobacillus delbrueckii* a las 48 horas.

La familia GH16 que tiene enzimas como la licheninasa (EC 3.2.1.73) que realiza la hidrólisis de uniones (1 \rightarrow 4)- β -D- glucosídicas en β -D-glucanos que contienen enlaces (1 \rightarrow 3)- y (1 \rightarrow 4), la endo-1,3- β -glucanasa (EC 3.2.1.39) que realiza la hidrólisis de enlaces (1 \rightarrow 3)- β -D-glucosídicos en (1 \rightarrow 3)- β -D-glucanos. (<http://www.cazy.org>).

Esta familia tiene una abundancia importante a las 48 horas y tiene genes asociados a *Streptococcus infantarius* en este tiempo.

También encontramos genes con abundancia para las familias de CEs, que son la CE1, CE3, CE4, CE7 y CE10. Se encontraron 619 genes para CE1, 152 para CE3, 140 para CE4, 402 para CE7 y 430 para CE10 en todo el metagenoma.

Las familias CE1, CE3, CE4 y CE7 tienen en común a la acetil xilan esterasa (EC 3.1.1.72) cuya función es la deacetilación mediante la hidrólisis de los xilanos y los xilo-oligosacáridos.

La familia CE1 que está conformada también por enzimas como la feruloil esterasa (EC 3.1.1.73) una enzima accesorio de la hemicelulosa (<http://www.cazy.org>).

Esta familia se encuentra en los 4 tiempos de fermentación y posee genes asociados a las siguientes especies: *Anoxybacillus flavithermus* a las 0 horas *Exiguobacterium sp. AT1b* en los cuatro tiempos de fermentación, *Enterobacter cloacae* a las 9 horas, *Streptococcus infantarius* a las 9, 24 y 48 horas, *Streptococcus macedonicus* a las 9 y 24 horas, *Pantoea vagans* a las 9 horas, *Micrococcus luteus* a las 24 horas, *Lactobacillus delbrueckii* y *Lactobacillus fermentum* a las 48 horas.

En la CE3 se encontraron genes en abundancia en los tiempos 9 y 24 horas, los cuales están asociados a las siguientes especies: *Exiguobacterium sp. AT1b*, *Streptococcus infantarius* y *Streptococcus macedonicus* en ambos tiempos, *Pantoea vagans* en 9 horas y *Micrococcus luteus* en 24 horas.

La familia CE4 está en todos los tiempos de la fermentación y las especies que están asociadas a sus géneros son las siguientes: *Anoxybacillus flavithermus* a las 0 horas *Exiguobacterium sp. AT1b* en todos los tiempos de la fermentación, *Enterobacter cloacae* a las 9 horas, *Streptococcus infantarius* a las 9, 24 y 48 horas, *Streptococcus lutetiensis*

a las 9 horas, *Micrococcus luteus* y *Streptococcus macedonicus* a las 24 horas y *Lactobacillus delbrueckii* a las 48 horas.

La familia CE7 se encuentra en abundancia a las 24 horas y tiene genes asociados a las especies *Exiguobacterium sp. AT1b*, *Micrococcus luteus*, *Streptococcus infantarius*, *Streptococcus lutetiensis*, *Streptococcus macedonicus* y *Pantoea vagans*.

La familia CE10 tiene enzimas como la acetil colin esterasa (EC 3.1.1.7) que actúa en una variedad de ésteres acéticos y además cataliza transacetilaciones (<http://www.cazy.org>).

Esta familia se encuentra en los 4 tiempos de fermentación y tiene genes asociados a las siguientes especies: *Anoxybacillus flavithermus* a las 0 horas, *Exiguobacterium sp. AT1b* en todos los tiempos, *Enterobacter cloacae*, a las 9 horas, *Streptococcus infantarius* a las 9, 24 y 48 horas, *Streptococcus macedonicus*, a las 9 y 24 horas, *Micrococcus luteus* a las 24 y 48 horas, *Pantoea vagans* 9 y 24 horas, *Streptococcus lutetiensis* a las 24 horas. *Lactobacillus delbrueckii*, *Lactobacillus fermentum* a las 48 horas.

Por último está la sacarosa que a pesar de ser un sustrato que se encuentra en menor proporción debido a la nixtamalización (Ampe et al., 1999), también tiene familias de enzimas que degradan este azúcar presentes en la fermentación. Entre ellas está la GH32 en abundancia.

Se encontraron 128 genes de la familia GH32 en nuestro metagenoma y ésta tiene enzimas como la invertasa (EC 3.2.1.26) cuya función es la hidrólisis de residuos no reductores terminales de β -D-fructofuranosidos en β -D-fructofuranosidos. Cataliza reacciones de sacarosa y fructotransferasas (<http://www.cazy.org>).

Esta familia se presenta de manera abundante en los 4 tiempos de fermentación y las especies que son asignadas a estos genes son las siguientes: *Anoxybacillus flavithermus* a las 0 horas, *Exiguobacterium sp. AT1b* y *Streptococcus infantarius* a 9, 24 y 48 horas, *Streptococcus macedonicus* a 9 y 24 horas, *Lactobacillus delbrueckii* y *Lactobacillus fermentum* a las 48 horas.

Y por último las familias que actúan sobre otros carbohidratos (monosacáridos, disacáridos) y que tienen abundancia en el metagenoma son la GH65 y CE14. Hay 50 genes para la GH65 y 56 para la CE14 en el metagenoma.

La GH65 tiene enzimas como α,α -trehalasa (EC 3.2.1.28) que tiene la función de catalizar la hidrólisis de O-uniones α -glucosídicas en α,α -trehalosa, liberando inicialmente cantidades equimolares de α - y β -D-glucosa. (<http://www.cazy.org>). Esta se encuentra en abundancia en el tiempo 0 horas y tiene genes asociados a *Exiguobacterium sp. AT1b*.

La CE14 tiene enzimas como N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy- α -D-glucopyranoside deacetylase (EC 3.5.1.89) cuya función es acilación de carbohidratos (<http://www.cazy.org>). Está presente en las 0 horas y en las 24 horas y tiene genes asociados a las especies *Exiguobacterium sp. AT1b* en ambos tiempos y *Anoxybacillus flavithermus* a las 0 horas.

Y por último podemos observar por medio de la reconstrucción metabólica de GhostKOALA, en la cual se puede ver que se tiene prácticamente el metabolismo completo y se pudieron resaltar las rutas metabólicas de interés de aminoácidos que son esenciales, así como de la vitamina riboflavina y que además tienen un aumento considerable en la masa fermentada del pozol en comparación de la masa con la que se prepara.

Conclusiones

Se puede concluir que el pozol es una fermentación compleja, llevada a cabo principalmente por bacterias del filo Firmicutes, más en específico por las bacterias ácido-lácticas (LAB).

También se ha comprobado que su microbiota es variada y que cambia conforme pasa el tiempo, que se parece mucho del transcurso de las 9 a las 24 horas y que es totalmente distinta a las 0 y a las 48 horas de fermentación.

Al principio de la fermentación la microbiota está representada por los géneros *Anoxybacillus*, *Exiguobacterium*, y *Streptococcus*. *Exiguobacterium* y *Streptococcus* han sido previamente reportados y se encuentran presentes en todos los tiempos de la fermentación siendo los géneros de mayor relevancia.

El género más abundante es *Streptococcus* y es a partir de las 9 horas de fermentación hasta el final de la misma que representa más de la mitad en abundancia de las bacterias presentes en el pozol. La especie que tiene más representación de ese género es *Streptococcus infantarius* y corroboramos que es la especie más importante en toda la fermentación (Díaz-Ruiz et al 2003).

El género *Exiguobacterium* ha sido aislado previamente de plantas de procesamiento de alimentos (Vishnivetskaya et. al 2009) y la especie previamente reportadas en pozol es *Exiguobacterium aurantiacum* (Ampe et al. 1999 Omar et al. 2000). Sin embargo en este trabajo podemos ver que la de mayor abundancia es *Exiguobacterium sp. AT1b*. Es probable que este género haya sido capaz de colonizar el pozol debido a su amplio rango de temperatura (-12°C—55°C) y de pH (5-11) (Vishnivetskaya et. al 2009) lo cual hace posible que creciera en las condiciones iniciales de la nixtamalización y que pudiera prevalecer a lo largo de la fermentación.

Anoxybacillus es un género que no ha sido reportado en pozol, pero sí en plantas de procesamiento de leche y en estiércol de vaca. Este género cuya especie más encontrada fue *Anoxybacillus flavithermus* es también, al igual que *Exiguobacterium*, tolerante a altas temperaturas y altos pHs (Goh et al, 2014) y es posible que sea por eso por lo que pudo colonizar la masa nixtamalizada, para luego disminuir en abundancia una vez que la acidificación de la masa hiciera que otras bacterias tuvieran ventaja sobre ella.

Se pudieron identificar otras especies en este estudio ya previamente reportadas como lo son *Leuconostoc citreum* (Olivares-Illana et al. 2002) previamente aislada del pozol, *Lactobacillus Delbruecki* y *Lactobacillus fermentum* (Ampe et al. 1999 Omar et al. 2000, Escalante et al. 2000) y *Streptococcus macedonicus* (Díaz-Ruiz et al 2003), así como bacterias nuevas de relevancia que no se habían reportado para el pozol, a saber, *Streptococcus lutetiensis*, *Micrococcus luteus*, *Enterobacter cloacae* y *Pantoea vagans*.

También por medio de este estudio corroboramos que géneros *Leuconostoc* tienen mayor proporción al principio de la fermentación (Ampe et al. 1999), específicamente a las 9 horas.

Otra cosa que se ha podido corroborar es que el género *Lactobacillus* tiene un aumento a partir de las 24 horas de la fermentación, donde podemos ver un reemplazo de las bacterias heterofermentativas a bacterias homofermentativas para continuar con la acidificación de la masa (Omar et al. 2000). Ejemplo de esto es *Lactobacillus fermentum* bacteria heterofermentativa que a las 24 horas representa la mayor parte de este género y a partir de las 48 horas *Lactobacillus delbrueckii* (bacteria homofermentativa) es la especie que adquiere mayor abundancia.

En lo que respecta al sustrato utilizado se puede ver que la microbiota principal del pozol en todos sus tiempos tiene la batería metabólica para degradar almidón, hemicelulosa y sacarosa en todos los tiempos de fermentación. En el caso de *Streptococcus Infantarius* se sabe que mostró un rendimiento muy bajo en amilasa y un grado específico de producción de amilasa. Esto le debe conferir una ventaja competitiva y puede ser una razón posible para su dominancia (Diaz-Ruiz et al 2003). También se ha visto en este trabajo que posee genes que codifican para familias de enzimas que degradan la hemicelulosa. Lo mismo ocurre para *Exiguobacterium sp. AT1b* que posee la batería metabólica para degradar ambos carbohidratos. Es quizás por esta razón que ambas especies prevalecen siendo las más abundantes en toda la fermentación.

Es por eso por lo que se concluye que los genes predichos que codifican para las familias de enzimas en todos los tiempos en buena abundancia son capaces de degradar cada uno de los tres polisacáridos en cuestión, así como para remover los “adornos” que estos poseen liberando monosacáridos y disacáridos que permitan que organismos que no son amilolíticos o xilanolíticos puedan desarrollarse en la masa.

Anteriormente se proponía que el almidón era el primero en degradarse por amilasas y que los azúcares resultantes, y el lactato producido permitieron a la segunda microbiota desarrollarse en la presencia de oxígeno. (Ampe Omar et al 1999). Ahora gracias a este estudio se contempla la utilización de la hemicelulosa por parte de muchas de las bacterias presentes en la masa.

Refiriendonos al potencial probiótico de la fermentación podemos pensar que mientras más pase el tiempo de fermentación, más bacterias con potencial probiótico aparecen, en especial del género *Lactobacillus*, que aparecen en las etapas tardías de los experimentos realizados.

Y por último mediante la reconstrucción metabólica de nuestro metagenoma pudimos corroborar que existen las rutas completas de biosíntesis de lisina, triptófano, riboflavina, isoleucina, arginina y treonina, ya que, como se ha mencionado en anterioridad, éstos aumentan en el pozol con respecto a la masa de maíz con la que se prepara.

También se puede ver mediante la anotación funcional que muchas de las rutas metabólicas que requiere la célula para su existencia, como las de todo el metabolismo de Carbono Central, el de Aminoácidos, la síntesis de lípidos, cofactores y vitaminas están presentes en nuestro metagenoma.

Perspectivas

- Complementar la taxonomía bacteriana con la de Levaduras, junto con estudios de ITS para obtener información de los Hongos presentes en el pozol.
- Realizar la anotación funcional de los metabolitos secundarios que se podrían encontrar en el pozol como lo son las bacteriocinas.
- Realizar estudios transcriptómicos para ver similitudes y poder predecir función y no solo el potencial metabólico.

Referencias.

1. Abriouel, H. *et al.* Culture-independent analysis of the microbial composition of the African traditional fermented foods poto poto and dégué by using three different DNA extraction methods. *Int. J. Food Microbiol.* **111**, 228–233 (2006).
2. Agbor, V. B., Cicek, N., Sparling, R., Berlin, A. & Levin, D. B. Biomass pretreatment: Fundamentals toward application. *Biotechnol. Adv.* **29**, 675–685 (2011).
3. Ampe, F., Ben Omar, N., Moizan, C., Wachter, C. & Guyot, J. P. Polyphasic study of the spatial distribution of microorganisms in Mexican pozol, a fermented maize dough, demonstrates the need for cultivation-independent methods to investigate traditional fermentations. *Appl. Environ. Microbiol.* **65**, 5464–5473 (1999).
4. Bal, J., Yun, S. H., Yeo, S. H., Kim, J. M. & Kim, D. H. Metagenomic analysis of fungal diversity in Korean traditional wheat-based fermentation starter nuruk. *Food Microbiol.* **60**, 73–83 (2016).
5. Bengtsson-Palme, J. *et al.* metaxa2: Improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.* **15**, 1403–1414 (2015).
6. Berlemont, R. & Martiny, A. C. Glycoside Hydrolases across Environmental Microbial Communities. *PLoS Comput. Biol.* **12**, 1–16 (2016).
7. Bigot, C., Meile, J.-C., Remize, F. & Strub, C. Applications of Metagenomics to Fermented Foods. *Fermented Foods, Part I (Biochemistry Biotechnol.* 333–346 (2015). doi:10.1201/b19872-19
8. Canas, A., Bárzana, E., David Owens, J. & Wachter, C. *La elaboración de pozol en los altos de Chiapas. Ciencia* **44**, (1993).
9. Cárdenas, C., Barkla, B. J., Wachter, C., Delgado-Olivares, L. & Rodríguez-Sanoja, R. Protein extraction method for the proteomic study of a Mexican traditional fermented starchy food. *J. Proteomics* **111**, 139–147 (2014).
10. Chen, K. & Pachter, L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput. Biol.* **1**, 0106–0112 (2005).
11. Cruz Ulloa, S. & Ulloa, M. Alimentos Fermentados De Maiz Consumidos En Mexico Y Otros Paises Latinoamericanos. 423–458 (1973).

12. De Filippis, F., La Storia, A., Stellato, G., Gatti, M. & Ercolini, D. A selected core microbiome drives the early stages of three popular Italian cheese manufactures. *PLoS One* **9**, (2014).
13. De Filippis, F., Parente, E. & Ercolini, D. Metagenomics insights into food fermentations. *Microb. Biotechnol.* **10**, 91–102 (2017).
14. Díaz Ruiz, G. & Wachter Rodarte, C. Métodos para el estudio de comunidades microbianas en alimentos fermentados. *Rev. Latinoam. Microbiol.* **45**, 30–40 (2003).
15. Díaz-Ruiz, G., Guyot, J. P., Ruiz-Teran, F., Morlon-Guyot, J. & Wachter, C. Microbial and physiological characterization of weakly amyolytic but fast-growing lactic acid bacteria: A functional role in supporting microbial diversity in pozol, a Mexican fermented maize beverage. *Appl. Environ. Microbiol.* **69**, 4367–4374 (2003).
16. Durand, N. *et al.* Application of PCR-DGGE to the study of dynamics and biodiversity of yeasts and potentially OTA producing fungi during coffee processing. *Food Control* **34**, 466–471 (2013).
17. Elizaquivel, P. *et al.* Pyrosequencing vs. culture-dependent approaches to analyze lactic acid bacteria associated to chicha, a traditional maize-based fermented beverage from Northwestern Argentina. *Int. J. Food Microbiol.* **198**, 9–18 (2015).
18. Escalante, A., Wachter, C. & Farrés, A. Lactic acid bacterial diversity in the traditional Mexican fermented dough pozol as determined by 16S rDNA sequence analysis. *Int. J. Food Microbiol.* **64**, 21–31 (2001).
19. Escobar-Zepeda, A., Sanchez-Flores, A. & Quirasco Baruch, M. Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota. *Food Microbiol.* **57**, 116–127 (2016).
20. Escobar-Zepeda A., Godoy-Lozano E., Raggi L., Segovia L., Merino E., Gutiérrez-Rios R., Juárez-Lopez K., Licea-Navarro A., Pardo-Lopez, L., Sanchez-Flores A., Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics. *submitted*. 2018.
21. Giles-Gómez, M. *et al.* In vitro and in vivo probiotic assessment of *Leuconostoc mesenteroides* P45 isolated from pulque, a Mexican traditional alcoholic beverage. *Springerplus* **5**, 708 (2016).
22. Handelsman, J., Tiedje, J., National Research Council (US) Committee on Metagenomics: Challenges and Functional & Applications. *THE NEW SCIENCE OF METAGENOMICS: Revealing the Secrets of Our Microbial Planet*. National Academies Press (2007). doi:NBK54006 [bookaccession]

23. Illumina. TruSeq ® DNA PCR-Free Library Prep Reference Guide. *Illumina* (2015). doi:FC-121-9006DOC
24. Illumina. 16S Metagenomic Sequencing Library Preparation. *Illumina.com* 1–28 (2013).
25. Jeffryes, M., Liakata, M. & Bateman, A. Crowdsourcing protein family database curation. *CEUR Workshop Proc.* **1747**, 29–37 (2016).
26. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, 5–9 (2008).
27. Kanehisa, M. *et al.* Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* **42**, 199–205 (2014).
28. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
29. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
30. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, 1–11 (2013).
31. Kunath, B., Bremges, A., Weimann, A., C. McHardy, A. & Pope, P. *Metagenomics and CAZyme Discovery. Methods in Molecular Biology* **1588**, (2017).
32. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
33. Li, H. Improving SNP discovery by base alignment quality. *Bioinformatics* **27**, 1157–1158 (2011).
34. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3 (2013).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

37. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
38. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
39. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
40. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
41. Morgulis, A. *et al.* Database indexing for production MegaBLAST searches. *Bioinformatics* **24**, 1757–1764 (2008).
42. Nalbantoglu, U. *et al.* Metagenomic analysis of the microbial community in kefir grains. *Food Microbiol.* **41**, 42–51 (2014).
43. Nayfach, S. & Pollard, K. S. Toward Accurate and Quantitative Comparative Metagenomics. *Cell* **166**, 1103–1116 (2016).
44. Nuraida, L., Wachter, M. C. & Owens, J. D. Microbiology of *pozol*, a Mexican maize dough. *World J. Microbiol. Biotechnol.* **11**, 567–571 (1995).
45. Olivares-Illana, V., Wachter-Odarte, C., Le Borgne, S. & López-Munguía, a. Characterization of a cell-associated inulosucrase from a novel source: a *Leuconostoc citreum* strain isolated from Pozol, a fermented corn beverage of Mayan origin. *J. Ind. Microbiol. Biotechnol.* **28**, 112–7 (2002).
46. Omar, N. & Ampe, F. Microbial Community Dynamics during Production of the Mexican Fermented Maize Dough Pozol. *Appl. Environ. Microbiol.* **66**, 3664–3673 (2000).
47. Pathways, M. Minoru Kanehisa. *Sci. Technol. Japan* **59**, 34–38 (1996).
48. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
49. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
50. Reddy, G., Altaf, M., Naveena, B. J., Venkateshwar, M. & Kumar, E. V. Amylolytic bacterial lactic acid fermentation - A review. *Biotechnol. Adv.* **26**, 22–34 (2008).

51. Rodarte, C. W. La Biotecnología Alimentaria Antigua: Los Alimentos Fermentados. **15**, 1607–6079 (2014).
52. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
53. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
54. Segata, N. *et al.* Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* **9**, 666–666 (2014).
55. Su, X., Pan, W., Song, B., Xu, J. & Ning, K. Parallel-META 2.0: Enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS One* **9**, 1–13 (2014).
56. Su, X., Xu, J. & Ning, K. Parallel-META: Efficient metagenomic data analysis based on high-performance computation. *BMC Syst. Biol.* **6**, S16 (2012).
57. Su, X. *et al.* Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Syst. Biol.* **6**, S16 (2012).
58. Sulaiman, J., Gan, H. M., Yin, W. F. & Chan, K. G. Microbial succession and the functional potential during the fermentation of Chinese soy sauce brine. *Front. Microbiol.* **5**, 1–9 (2014).
59. Universitaria, C. & Federal, D. Los Hongos En La Cultura Mexicana : Bebidas Y Alimentos Tradicionales Fermentados , Hongos Alucinógenos. **116**, 108–116 (2007).
60. Wachter, C., Cañas, A., Cook, P. E., Barzana, E. & Owens, J. D. Sources of microorganisms in pozol, a traditional Mexican fermented maize dough. *World J. Microbiol. Biotechnol.* **9**, 269–274 (1993).
61. Wachter, C. *et al.* Microbiology of Indian and Mestizo pozol fermentations. *Food Microbiol.* **17**, 251–256 (2000).
62. Wolfe, B. E., Button, J. E., Santarelli, M. & Dutton, R. J. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell* **158**, 422–433 (2014).
63. Wolfe, B. E. & Dutton, R. J. Fermented foods as experimentally tractable microbial ecosystems. *Cell* **161**, 49–55 (2015).
64. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).

65. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, (2014).
66. Zhang, Y. *et al.* The microbiome and metabolites in fermented Pu-erh tea as revealed by high-throughput sequencing and quantitative multiplex metabolite analysis. *PLoS One* **11**, 1–18 (2016).
67. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, 1–15 (2010).
68. Arabinoxylan, an alternative carbon source in *Streptococcus infantarius* 25124 lactic acid fermentation of pozol. Cooper, B. 1–26
69. <https://www.slideshare.net/FAOoftheUN/presentacin-sobre-importancia-de-la-diversidad-biolgica-del-suelo>
70. <https://en.wikipedia.org/wiki/Nuruk>
71. 70. https://es.wikipedia.org/wiki/Pozol#/media/File:Ilustraci%C3%B3n_sobre_pozol_tabasque%C3%B1o.jpg
72. www.cazy.org
73. <https://www.kegg.jp/ghostkoala/>
74. (<https://www.ncbi.nlm.nih.gov/refseq/>)

Material suplementario.

TAXONOMY	0Hrs	9Hrs	24Hrs	48Hrs
Archaea	0.001026	0.000619	0.000514	0.003445
Bacteria	49.134879	84.1056	88.177996	89.634268
Chloroplast	50.864095	15.893781	11.82149	10.362288

Tabla S1. Matriz a nivel dominio 16S.

TAXONOMY	0Hrs	9Hrs	24Hrs	48Hrs
Firmicutes	86.206008	77.1602	76.692665	81.985056
Actinobacteria	7.931702	4.696386	7.873359	8.047822
Proteobacteria	2.865714	17.646085	14.660504	7.77626
Bacteroidetes	2.580326	0.402022	0.718333	2.069599
Chloroflexi	0.066823	0.006624	0.003574	0.01281
Candidate_division_TM7	0.060558	0.00092	0.001969	0.004697
Deinococcus-Thermus	0.057078	0.004048	0.006127	0.019641
Cyanobacteria	0.057078	0.048758	0.031727	0.032878
Acidobacteria	0.054293	0.012143	0.004449	0.023484
Planctomycetes	0.031323	0.004232	0.001386	0.008967
Armatimonadetes	0.018098	0.000368	0.000146	0.000427
Gemmatimonadetes	0.016706	0.00184	0.000438	0.001708
Nitrospirae	0.011833	0.005704	0.001823	0.003843
Verrucomicrobia	0.009049	0.003496	0.001021	0.005978
Candidate_division_SR1	0.006961	0.000736	0.000146	0.001281
Aquificae	0.004872	0.001104	0.000802	0.003416
Spirochaetes	0.00348	0.00092	0.000219	0
Chlorobi	0.00348	0.000368	0.000073	0
Synergistetes	0.002088	0.00184	0.000146	0.001281
Elusimicrobia	0.002088	0.000368	0	0.000427
Chlamydiae	0.002088	0.000552	0.000219	0
Thermotogae	0.001392	0.000552	0.000365	0
Tenericutes	0.001392	0	0.000219	0
TM6	0.001392	0	0	0
Fusobacteria	0.001392	0	0.000073	0
BD1-5	0.001392	0	0	0
Lentisphaerae	0.000696	0.000184	0	0
Dictyoglomi	0.000696	0.000184	0	0.000427
Thermodesulfobacteria	0	0.000184	0.000073	0
Deferribacteres	0	0.000184	0.000073	0
Candidate_division_WS6	0	0	0.000073	0

Tabla S2. Matriz a nivel filo 16S.

TAXONOMY	0Hrs	9Hrs	24Hrs	48Hrs
Bacilli	86.058442	77.108682	76.659116	81.922716
Actinobacteria	7.903859	4.687555	7.870004	8.020922
Flavobacteria	2.327653	0.383255	0.702506	2.038002
Gammaproteobacteria	1.385177	17.354642	14.441846	6.973527
Betaproteobacteria	0.835282	0.101195	0.082854	0.09778
Otro	1.48958	0.364674	0.24368	0.947056

Tabla S3. Matriz a nivel clase 16S.

TAXONOMY	0Hrs	9Hrs	24Hrs	48Hrs
Bacillaceae	53.335888	8.139777	8.372696	5.292286
Family_XII_Incertae_Sedis	22.897955	6.920751	8.828631	4.922044
Streptococcaceae	6.665505	44.679806	41.716653	40.349147
Micrococcaceae	5.582692	3.798207	6.679003	6.321449
Flavobacteriaceae	2.325889	0.382728	0.702247	2.0374
Corynebacteriaceae	1.115563	0.443265	0.421581	0.486397
Enterococcaceae	1.047277	6.858926	6.181128	7.450965
Planococcaceae	0.957391	0.832065	1.593768	0.541912
Aerococcaceae	0.572066	1.430078	2.694839	4.016296
Comamonadaceae	0.567188	0.043977	0.022538	0.024768
Dermabacteraceae	0.400655	0.065689	0.104593	0.611092
Moraxellaceae	0.361635	8.882048	5.675085	4.930158
Enterobacteriaceae	0.345608	7.178173	7.290077	1.40581
Leuconostocaceae	0.220883	6.662962	3.034365	3.089196
Paenibacillaceae	0.208341	0.821209	1.061612	0.493229
Pseudomonadaceae	0.144933	0.666278	0.762712	0.231455
Lactobacillaceae	0.07595	0.116658	2.474348	14.992463
Acetobacteraceae	0.029962	0.000736	0.003355	0.525684
Otro	3.144638	2.076659	2.380795	2.278228

Tabla S4. Matriz a nivel familia 16S.

TAXONOMY	0Hrs	9Hrs	24Hrs	48Hrs
Bacillales	77.49709 1	17.02984 2	20.139294	11.46695 4
Lactobacillales	8.636797	60.07989 1	56.52147	70.46065 9
Micrococcales	6.339132	3.935101	6.953399	7.046536
Flavobacteriales	2.329713	0.383262	0.702524	2.038124
Corynebacteriales	1.140472	0.448948	0.430106	0.497041
Burkholderiales	0.598452	0.048759	0.025747	0.031599
Pseudomonadales	0.50649	9.547869	6.437666	5.161282
Enterobacteriales	0.345556	7.177829	7.289928	1.40572
Aeromonadales	0.128887	0.492739	0.59538	0.223754
Rhodospirillales	0.073152	0.012328	0.007658	0.553833
Otro	2.404256	0.843441	0.896829	1.114493

Tabla S5. Matriz a nivel orden 16S.

TAXONOMY	0Hrs	Hrs	24Hrs	48Hrs
Anoxybacillus	33.419828	0.74993	0.803323	0.32735
Exiguobacterium	22.922552	6.925211	8.837281	4.93207
Bacillus	19.113287	6.847143	6.909978	4.55551
Streptococcus	6.619652	44.198027	41.020935	38.941355
Rothia	4.640732	3.208333	3.737572	4.312886
Chryseobacterium	2.19725	0.350017	0.683003	1.984638
Corynebacterium	1.116761	0.443551	0.421995	0.487387
Enterococcus	1.023988	6.577772	5.800015	7.085303
Paenisporosarcina	0.769386	0.059287	0.046653	0.020112
Aerococcus	0.56989	1.420688	2.679957	4.001369
Brachy bacterium	0.398993	0.062417	0.096445	0.594792
Micrococcus	0.333424	0.401571	2.234162	1.522497
Acinetobacter	0.322961	8.830326	5.620412	4.876014
Arthrobacter	0.227398	0.135146	0.59123	0.352168
Brevibacillus	0.182755	0.808666	1.053526	0.480541
Leuconostoc	0.166712	2.552489	1.05411	1.210552
Pseudomonas	0.145088	0.666523	0.76346	0.231926
Enterobacter	0.09347	1.819497	1.178518	0.633304
Pantoea	0.088588	4.138887	5.338012	0.317508
Lactobacillus	0.076032	0.116734	2.472757	14.995614
Kurthia	0.073939	0.506705	0.963871	0.248615
Lactococcus	0.053013	0.510571	0.736446	1.489548
Weissella	0.051618	3.898607	1.892989	1.8062
Acetobacter	0.000698	0.000184	0.00219	0.509211
Otro	5.392086	4.771644	5.061133	4.083586

Tabla S6. Matriz a nivel género 16S.

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Bacteria	99.069292	99.911779	99.946403	99.791312
Archaea	0.930708	0.088191	0.053443	0.208602
Viruses	0	0.000031	0.000154	0.000086

Tabla S7. Matriz a nivel dominio Shotgun.

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Firmicutes	40.272811	77.795542	82.617715	85.489177
Bacteria;no_rank	25.866302	2.898477	3.213748	4.626931
Proteobacteria	13.653995	17.947507	13.258899	8.467652
Actinobacteria	12.499019	0.723326	0.630242	0.623386
Tenericutes	3.690261	0.273471	0.087174	0.2655
Cyanobacteria	3.158317	0.262479	0.134499	0.405345
Bacteroidetes	0.426845	0.063285	0.035674	0.071423
Thermotogae	0.258685	0.018768	0.011089	0.028423
Chlamydiae	0.079596	0.005511	0.001972	0.008103
Planctomycetes	0.034753	0.002786	0.001434	0.004587
Spirochaetes	0.016816	0.002725	0.001818	0.003387
Deinococcus-Thermus	0.016255	0.002143	0.001818	0.001629
Aquificae	0.009809	0.001531	0.001613	0.002444
Chloroflexi	0.006446	0.000857	0.000819	0.000643
Nitrospirae	0.004204	0.000459	0.000435	0.000214
Fusobacteria	0.002803	0.000429	0.000256	0.000343
Verrucomicrobia	0.001121	0.000031	0.000102	0.000086
Dictyoglomi	0.000561	0.000245	0.000077	0.000086
Chlorobi	0.000561	0.000092	0.000077	0.000086
Acidobacteria	0.000561	0.000245	0.000282	0.000472
Synergistetes	0.00028	0.000061	0.000128	0.000086
Thermodesulfobacteria	0	0.000031	0.000102	0
Ignavibacteriae	0	0	0.000026	0

Tabla S8. Matriz a nivel filo Shotgun

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Bacilli	39.87091	77.368865	82.231298	85.103853
Bacteria;no_rank	25.866302	2.898477	3.213748	4.626931
Actinobacteria	12.293864	0.70419	0.617719	0.596292
Proteobacteria;no_rank	5.357563	0.412686	0.183259	0.385067
Gammaproteobacteria	3.806291	17.156916	12.770276	7.654862
Mollicutes	3.690261	0.273471	0.087174	0.2655
Betaproteobacteria	2.829284	0.258101	0.217627	0.276517
Cyanobacteria;no_rank	1.561361	0.122315	0.054189	0.187432
Alphaproteobacteria	1.540061	0.105384	0.077749	0.131357
Synechococcales	1.094158	0.093259	0.053113	0.154421
Otro	2.089943	0.606337	0.493845	0.617771

Tabla S9. Matriz a nivel clase Shotgun.

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Exiguobacterium	30.23789	16.46736	15.203328	11.33383
Bacteria;no_rank	25.866302	2.898477	3.213748	4.626931
Propionibacteriaceae	9.537169	0.098127	0.098032	0.087714
Proteobacteria;no_rank	5.357563	0.412686	0.183259	0.385067
Bacillaceae	4.122431	1.08828	0.7714	0.78591
Mycoplasmataceae	3.41448	0.252376	0.075726	0.238148
Streptococcaceae	3.334604	53.639529	54.293316	45.769175
Cyanobacteria;no_rank	1.561361	0.122315	0.054189	0.187432
Enterobacteriaceae	1.468033	9.174677	8.864056	2.097889
Comamonadaceae	1.010639	0.075685	0.035494	0.028766
Synechococcaceae	0.892086	0.071552	0.038644	0.119396
Pseudomonadaceae	0.869665	0.30617	0.166665	0.05033
Burkholderiales;no_rank	0.840517	0.081227	0.095804	0.142717
Listeriaceae	0.819498	0.715886	0.486549	0.492716
Actinobacteria;no_rank	0.732335	0.151952	0.14039	0.14216
Burkholderiaceae	0.548761	0.05061	0.048939	0.066493
Moraxellaceae	0.392092	2.620689	2.129404	5.037891
Bacilli;no_rank	0.351173	1.491138	1.131721	1.195799
Lactobacillaceae	0.2954	0.804369	8.058084	23.136283
Enterococcaceae	0.139853	1.107477	0.940907	0.918038
Erwiniaceae	0.107062	3.85991	0.77793	0.138816
Lactobacillales;no_rank	0.105941	0.9915	0.798085	0.887685
Enterobacteriales;no_rank	0.061659	0.560229	0.432181	0.094573
Leuconostocaceae	0.033071	0.762821	0.305031	0.344211
Otro	7.90042	2.194961	1.657124	1.692034

Tabla S10. Matriz a nivel familia Shotgun.

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Bacillales	35.597135	18.520289	16.664921	12.802588
Bacteria;no_rank	25.866302	2.898477	3.213748	4.626931
Propionibacteriales	9.614242	0.110252	0.105561	0.095859
Proteobacteria;no_rank	5.357563	0.412686	0.183259	0.385067
Lactobacillales	3.922602	57.357438	64.434655	71.105465
Mycoplasmatales	3.41448	0.252376	0.075726	0.238148
Burkholderiales	2.522393	0.220901	0.188765	0.244536
Enterobacteriales	1.795944	13.83421	10.24626	2.38148
Cyanobacteria;no_rank	1.561361	0.122315	0.054189	0.187432
Pseudomonadales	1.261757	2.930808	2.297759	5.089851
Synechococcales	1.094158	0.093259	0.053113	0.154421
Rhizobiales	0.746068	0.052692	0.038414	0.055046
Actinobacteria;no_rank	0.732335	0.151952	0.14039	0.14216
Micrococcales	0.725889	0.157402	0.140569	0.140102
Corynebacteriales	0.638446	0.208655	0.163156	0.140317
Xanthomonadales	0.515689	0.028443	0.019207	0.020492
Otro	4.282464	1.96789	0.848586	0.994309

Tabla S11. Matriz a nivel orden Shotgun.

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Exiguobacterium	30.23789	16.46736	15.203328	11.33383
Bacteria;no_rank	25.866302	2.898477	3.213748	4.626931
Cutibacterium	9.518671	0.097393	0.097238	0.086857
Proteobacteria;no_rank	5.357563	0.412686	0.183259	0.385067
Mycoplasma	3.410556	0.251978	0.075573	0.237891
Streptococcus	3.196993	52.010126	50.33726	41.688334
Geobacillus	1.944485	0.163739	0.158982	0.120081
Cyanobacteria;no_rank	1.561361	0.122315	0.054189	0.187432
Bacillus	1.3492	0.837435	0.532978	0.599421
Burkholderiales;no_rank	0.840517	0.081227	0.095804	0.142717
Bacillaceae;no_rank	0.828746	0.087105	0.07944	0.066407
Listeria	0.819498	0.715886	0.486549	0.492716
Pseudomonas	0.807726	0.292637	0.158162	0.041113
Actinobacteria;no_rank	0.732335	0.151952	0.14039	0.14216
Synechococcus	0.659746	0.05609	0.029604	0.090972
Plautia_stali_symbiont	0.456273	5.273679	7.313753	0.23429
Acinetobacter	0.383684	2.578591	2.113782	5.020186
Enterobacteriaceae;no_rank	0.353415	0.939389	0.281855	0.397242
Bacilli;no_rank	0.351173	1.491138	1.131721	1.195799
Lactobacillus	0.283629	0.735389	7.966684	22.805062
Klebsiella	0.211321	1.317295	0.619793	0.854718
Enterococcus	0.11603	0.909201	0.800518	0.763274
Lactococcus	0.109023	1.103343	3.426048	3.632797
Lactobacillales;no_rank	0.105941	0.9915	0.798085	0.887685
Enterobacter	0.103979	1.039966	0.360859	0.306527
Pantoea	0.089125	3.678168	0.626938	0.113522
Enterobacteriales;no_rank	0.061659	0.560229	0.432181	0.094573
Leuconostoc	0.032511	0.723326	0.288283	0.325133
Streptococcaceae;no_rank	0.028587	0.526061	0.530008	0.448044
Otro	10.18206	3.486328	2.462996	2.679224

Tabla S12. Matriz a nivel género Shotgun.

TAXONOMY	0hrs	9hrs	24hrs	48hrs
Exiguobacterium_sp._AT1b	30.002747	16.035325	15.038379	10.89809
Bacteria;no_rank	25.866302	2.898477	3.213748	4.626931
Cutibacterium_acnes	9.441318	0.09629	0.09647	0.085828
Proteobacteria;no_rank	5.357563	0.412686	0.183259	0.385067
Mycoplasma_conjunctivae	3.195592	0.231311	0.06474	0.199821
Streptococcus_suis	1.571731	26.327368	22.882421	19.842124
Cyanobacteria;no_rank	1.561361	0.122315	0.054189	0.187432
Streptococcus;no_rank	1.133676	20.325097	22.654141	17.665095
Geobacillus;no_rank	1.035583	0.084687	0.083409	0.060148
Bacillus_megaterium	0.960191	0.562005	0.385854	0.420478
Burkholderiales;no_rank	0.840517	0.081227	0.095804	0.142717
Bacillaceae;no_rank	0.828746	0.087105	0.07944	0.066407
Listeria_monocytogenes	0.786987	0.384028	0.269946	0.257097
Actinobacteria;no_rank	0.732335	0.151952	0.14039	0.14216
Plautia_stali_symbiont	0.456273	5.273679	7.313753	0.23429
Enterobacteriaceae;no_rank	0.353415	0.939389	0.281855	0.397242
Bacilli;no_rank	0.351173	1.491138	1.131721	1.195799
Acinetobacter_baumannii	0.292037	1.712284	1.78094	4.572056
Streptococcus_thermophilus	0.255322	3.74874	3.277848	2.916852
Lactobacillus_fermentum	0.204874	0.421289	7.180892	18.933774
Lactobacillales;no_rank	0.105941	0.9915	0.798085	0.887685
Lactococcus_garvieae	0.085481	0.754279	3.176538	3.240871
Pantoea;Pantoea_ananatis	0.063901	3.228956	0.21883	0.068379
Enterobacteriales;no_rank	0.061659	0.560229	0.432181	0.094573
Acinetobacter;no_rank	0.056053	0.528449	0.224362	0.33165
Klebsiella_pneumoniae	0.05269	0.710436	0.312944	0.416062
Enterobacter_cloacae	0.045123	0.665888	0.206538	0.222029
Streptococcaceae;no_rank	0.028587	0.526061	0.530008	0.448044
Leuconostoc;no_rank	0.027466	0.653764	0.260036	0.291994
Lactobacillus_plantarum	0.017657	0.038914	0.091502	1.471373
Lactobacillus;no_rank	0.017376	0.127581	0.534182	1.84345
Otro	14.210323	9.82756	7.005592	7.454489

Tabla S13. Matriz a nivel especie Shotgun.

CAZymes	0 hrs	9 hrs	24 hrs	48 hrs
CBM20	4.950495	0.677507	0.618047	0.545703
CBM26	0	1.897019	1.483313	1.909959
CBM34	3.960396	1.761518	2.10136	1.773533
CBM41	3.960396	3.387534	2.966625	3.410641
CBM68	0.990099	0.542005	0.370828	0.409277
GH126	0	0.271003	0.247219	0
GH13	27.722772	18.428184	19.777503	16.507503
GH14	0.990099	0.406504	0.247219	0.272851
GH15	0.990099	1.355014	1.112485	0.545703
GH57	0.990099	0.271003	0.247219	0.272851
GH76	0	0.406504	0.741656	1.909959
GH77	0.990099	1.626016	1.483313	0.818554
CBM1	0	0	0	0.818554
CBM11	0	0	0.123609	0.272851
CBM22	0	0.271003	0.123609	0.272851
CBM35	0.990099	0.406504	0.247219	0.136426
CBM37	1.980198	0.135501	0.123609	0
CBM4	0.990099	0.677507	0.247219	0.409277
CBM56	0.990099	0.406504	0.247219	0.136426
CBM76	0	0.135501	0.123609	0.272851
CE1	7.920792	10.97561	10.754017	7.366985
CE10	3.960396	7.181572	7.045735	6.27558
CE12	1.980198	1.084011	0.988875	1.227831
CE16	0	0	0.370828	0.136426
CE3	0.990099	2.168022	2.595797	1.773533
CE4	8.910891	5.01355	4.449938	3.956344
CE5	0	0	0	0.272851
CE7	1.980198	1.490515	2.10136	1.227831
CE8	0	0.271003	0.123609	0.272851
GH1	0	7.181572	8.034611	5.729877
GH10	0	0.135501	0	0.272851
GH11	0	0	0	0.272851
GH120	0	0.271003	0.247219	0
GH125	0	0	0	0.545703
GH128	0	0	0	0.409277
GH130	0	0.271003	0.247219	0.136426
GH131	0	0	0	0.272851
GH16	0.990099	0.948509	0.618047	2.319236

GH17	0	0	0	0.545703
GH2	0	0.948509	1.112485	0.818554
GH28	0.990099	0.406504	0.247219	0.409277
GH3	3.960396	2.439024	1.730532	2.864939
GH30	0	0	0	0.545703
GH31	0.990099	1.084011	0.865266	1.364256
GH36	0.990099	1.084011	0.988875	0.818554
GH4	0	1.761518	1.606922	1.227831
GH43	0.990099	1.626016	0.865266	1.227831
GH5	0	0.948509	0.618047	1.227831
GH53	0	0.948509	0.494438	1.227831
GH55	0	0	0	0.545703
GH6	0	0	0	0.545703
GH67	0	0	0	0.272851
GH7	0	0	0	0.682128
GH71	0	0	0	0.545703
GH74	0	0.135501	0.247219	0.682128
GH79	0	0	0	0.272851
GH8	0	1.355014	0.865266	0.545703
GH94	0	0.271003	0.247219	0.272851
PL12	0.990099	0.813008	0.988875	0.272851
PL22	0	0.677507	0.618047	0.545703
PL7	0	0.135501	0.247219	0
PL9	0	0.271003	0.494438	0.272851
CBM24	0	0	0	0.272851
CBM61	0	0.542005	0.247219	0.272851
CBM66	0	0.813008	0.494438	0.545703
GH32	2.970297	3.929539	4.449938	3.819918
GH66	0	0.542005	0.123609	0.136426
GH68	0.990099	0.677507	0.741656	0.682128
GH70	0	1.219512	1.236094	1.227831
GH72	0	0	0	0.545703
CBM32	0	0.135501	0.494438	0.272851
CBM67	0	0.271003	0.370828	0.136426
CE14	2.970297	1.626016	2.348578	1.364256
GH114	0	0.271003	0.247219	0.409277
GH135	0	0	0	0.272851
GH35	0	0.135501	0.123609	0.272851
GH37	0	0.813008	0.618047	0.136426

GH42	0	0.542005	0.494438	0.409277
GH65	2.970297	0.813008	1.483313	1.364256
AA1	0	0	0	1.227831
AA12	0	0.135501	0	0.272851
AA2	1.980198	0.406504	0.618047	0.409277
AA3	0	0.677507	1.483313	1.227831
AA4	0.990099	0.813008	1.112485	0.818554
AA7	0.990099	0.677507	0.865266	1.227831
AA9	0	0	0	1.773533

Tabla S14. Matriz de abundancia relativa de las CAZymes encontradas en el metagenoma del pozol.