



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Caracterización metagenómica de la fermentación del pulque

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:
Julián Torres Morales

TUTOR PRINCIPAL
Dr. José Adelfo Escalante Lozada
Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR
Dr. Agustín López Munguía Canales
Instituto de Biotecnología, UNAM
Dr. Adrián Ochoa Leyva
Instituto de Biotecnología, UNAM

Ciudad de México. Septiembre, 2022



Universidad Nacional
Autónoma de México



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.

Agradecimientos

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM <IN207917 e IN211420>

En el proyecto de investigación <Estudio de la diversidad microbiana, genética y metabólica del pulque por medio de un enfoque polifásico>

CONACYT

Beca de maestría

Apoyo técnico

Dr. Gamaliel López Leal, Departamento de Microbiología Molecular, IBT, UNAM.

M. en C. Georgina Teresa Hernández Chávez, Laboratorio de Ingeniería de Vías Metabólicas, Departamento de Ingeniería Celular y Biocatálisis, IBT, UNAM.

M. en T.I. Juan Manuel Hurtado Ramírez, Unidad de Cómputo, IBT, UNAM.

Dra. Shamayim Tabita Ramírez Puebla, MBL, University of Chicago.

M. en C. María Elena Rodríguez Alegría, Departamento de Ingeniería Celular y Biocatálisis, IBT, UNAM.

Índice

1. Resumen	1
2. Introducción	3
2.1. Alimentos fermentados	3
2.1.1. Clasificación de los alimentos fermentados	4
2.1.2. Importancia de los estudios metagenómicos.....	5
3. Antecedentes	6
3.1. Caracterización de comunidades microbianas	6
3.1.1. Microorganismos cultivables y no cultivables.....	6
3.1.2. Secuenciación masiva de comunidades microbianas	7
3.1.3. Microscopía de comunidades microbianas	8
3.2. Metagenómica en alimentos fermentados	9
3.3. El pulque	12
3.3.1. Aspectos generales	12
3.3.2. La elaboración de pulque	12
3.3.3. Aspectos históricos del pulque	14
3.3.4. Aporte nutricional del pulque	15
3.3.5. Microorganismos en el pulque	16
4. Justificación	22
5. Hipótesis	22
6. Objetivo General	23
7. Objetivos Específicos	23
8. Desarrollo metodológico	23
8.1. Recolección de aguamiel y pulque	26
8.2. La fermentación y el muestreo	26
8.3. Extracción de ADN metagenómico	27
8.4. Determinación de metabolitos	27
8.5. Secuenciación	28
8.6. Análisis bioinformático	28
8.6.1. Análisis de calidad, limpieza y <i>trimming</i>	30

8.6.2. Determinación de la diversidad y composición microbiana	30
8.6.3. Obtención de MAGs	32
8.6.4. Análisis funcional (pangenómico) de los MAGs	35
8.7. Microscopía	36
9. Resultados	40
9.1. Determinación de metabolitos	40
9.2. Extracción de ADN metagenómico y secuenciación	47
9.3. Análisis bioinformático	51
9.3.1. Análisis de calidad, limpieza y <i>trimming</i>	51
9.3.2. Determinación de la diversidad y composición microbiana	52
9.3.3. Obtención de MAGs	67
9.3.4. Análisis funcional (pangenómico) de los MAGs	76
9.3.5. Pangenoma de <i>Zymomonas mobilis</i>	79
9.4. Microscopía	81
10. Conclusiones	84
11. Perspectivas	87
12. Referencias	88

Índice de figuras

Figura 1. Aplicaciones de la metagenómica	6
Figura 2. Proceso de elaboración de pulque	13
Figura 3. Bacterias identificadas en pulque de 3 regiones de México.....	17
Figura 4. Diagrama de Venn de bacterias identificadas por un método dependiente de cultivo y un método independiente de cultivo	19
Figura 5. Mapa de calor de abundancias de los géneros microbianos en aguamiel, contrapunta y pulque	21
Figura 6. Resumen del desarrollo metodológico	25
Figura 7. Diagrama de flujo del análisis bioinformático	29
Figura 8. Ecuaciones para caracterizar la composición microbiana	31
Figura 9. Diagrama de flujo de trabajo para el análisis microscópico de aguamiel y pulque empleando CLASI-FISH	37
Figura 10. Perfil de metabolitos	41
Figura 11. Aportación de carbono de los metabolitos en distintas fermentaciones de pulque	44
Figura 12. Perfil de oligosacáridos extracelulares durante la fermentación de pulque	46
Figura 13. Integridad del ADN metagenómico de las muestras enviadas para secuenciación	49
Figura 14. Eficiencia de KRAKEN2 para los 5 metagenomas	55
Figura 15. Precisión de KRAKEN2 para 6 niveles taxonómicos	56
Figura 16. Composición genómica microbiana de bacterias y hongos	60
Figura 17. Composición microbiana de bacterias y hongos por clasificación de genes ribosomales	62
Figura 18. Composición genómica microbiana de bacterias y hongos a nivel de especie.....	64
Figura 19. Composición genómica de virus (viroma)	67
Figura 20. Gráfica de frecuencias del número de <i>contigs</i> por intervalos de tamaño	70
Figura 21. Microorganismos del AgM, PQ y de la fermentación identificados por distintos métodos	75

Figura 22. Representatividad funcional de los MAGs en el meta-ensamble	76
Figura 23. Pangenoma de los 23 MAGs	77
Figura 24. Distribución de los MAGs en los 4 principales procesos metabólicos, propuestos por Escalante et al. (2016).....	79
Figura 25. Pangenoma de <i>Zymomonas mobilis</i>	81
Figura 26. Organización espacial de bacterias y hongos en muestras de aguamiel y pulque utilizando CLASI-FISH	84
Figura 27. Resumen de los resultados más relevantes de cada una de las 5 etapas del trabajo.....	87

Índice de tablas

Tabla 1. Ventajas y desventajas de la secuenciación masiva dependiente e independiente de sonda.....	8
Tabla 2. Análisis de secuenciación masiva en alimentos fermentados de México.....	11
Tabla 3. Tratamiento de las muestras.....	26
Tabla 4. Combinación de ensambladores y tipo de ensamblados	33
Tabla 5. Descripción de las sondas para microscopía	39
Tabla 6. Concentración de metabolitos, pH y temperatura	42
Tabla 7. Pureza y cuantificación de ADN metagenómico para secuenciación.....	50
Tabla 8. Archivos de secuenciación.....	51
Tabla 9. Resumen de las lecturas y del <i>trimming</i>	52
Tabla 10. Comparación de los anotadores taxonómicos.....	53
Tabla 11. Índices de diversidad alfa.....	57
Tabla 12. Composición genómica microbiana; nivel taxonómico: dominio.....	58
Tabla 13. Microorganismos reportados previamente por el grupo y en este trabajo.....	63
Tabla 14. Comparación de los métodos para ensamblar genomas	68
Tabla 15. Comparación de los ensamblados	69
Tabla 16. Resumen del mapeo de las lecturas de cada metagenoma con el meta-ensamble	71

Tabla 17. Resumen de la obtención de MAGs	72
Tabla 18. Resumen de los MAGs	73

1. Resumen

Los alimentos y bebidas fermentadas son de gran importancia cultural, económica, nutricional y social en todo el mundo ¹⁻¹².

El pulque (PQ) es una bebida fermentada, tradicional mexicana producida a partir de la savia (aguamiel – AgM) de algunas especies de Agave. Diversos estudios han descrito la composición y diversidad microbiana del pulque y el aguamiel utilizando métodos dependientes de cultivo. Recientemente se realizó un perfil bacteriano por medio de secuenciación masiva del gen ribosomal 16S permitiendo identificar nuevos grupos bacterianos. Además, algunos sustratos (carbohidratos) y productos (ácidos orgánicos) de la fermentación se han asociado con las diversas poblaciones bacterianas¹³⁻¹⁷.

La metagenómica es el estudio de la totalidad del material genético de una muestra ambiental empleando secuenciación masiva. Esta herramienta ha permitido ensamblar genomas sin la necesidad de aislar a los microorganismos y explorar la capacidad funcional de la comunidad secuenciada ¹⁸⁻²². Nuevas técnicas de microscopía han evidenciado que los microorganismos pueden organizarse espacialmente añadiendo otro nivel de complejidad al estudio de las comunidades microbianas ^{23,24}.

En este contexto, el presente trabajo tuvo como objetivo caracterizar la comunidad microbiana a lo largo de la fermentación de pulque mediante metagenómica y caracterización de los perfiles metabólicos. La fermentación se realizó mezclando AgM y PQ (2:3) en un volumen total de 5 litros. Se muestreó el AgM, el PQ, al inicio de la fermentación (0 horas), a las 3 y 6 horas. Se secuenció una réplica de cada tiempo de muestreo y los controles. Para el perfil metabólico se analizaron tres réplicas de la fermentación por HPLC.

Este es uno de los primeros trabajos que reporta el ensamble simultáneo de bacterias y hongos a partir de una bebida fermentada. Se ensamblaron 23 genomas (2 levaduras y 21 bacterias) que corresponden a las especies más abundantes en aguamiel y pulque. Estos genomas aportan más del 94% de la capacidad funcional de la bebida. Durante la

fermentación se observó el consumo de sacarosa, la producción de etanol, fructosa y lactato, y la disminución del pH, característicos de la bebida. *Zymomonas mobilis* y *Saccharomyces* spp. se asociaron con la producción de etanol. *Z. mobilis* se enriqueció al final del proceso (26% - 36%), mientras que, *Saccharomyces* sp. fue inoculado del PQ (21%). Bacterias ácido-lácticas, como *Lactobacillus* spp., responsables de la acidificación del pulque incrementaron su abundancia en más 4 veces durante la fermentación. El género asociado a rizosfera, *Acinetobacter*, fue el más abundante en el aguamiel indicando que proviene del Agave. Por último, mediante microscopía se observó levaduras y bacterias, las cuales formaban una estructura en forma de malla.

En conclusión, el aguamiel y el pulque tienen una comunidad distinta de bacterias y levaduras. Durante la fermentación se enriquecieron bacterias etanológicas y ácido-lácticas que contribuyen a las propiedades sensoriales características de dicha bebida.

2. Introducción

2.1 Alimentos fermentados

Los alimentos fermentados tienen un gran impacto cultural, económico y social. Representan un reservorio importante microorganismos con actividades metabólicas que podrían ser explotadas con fines biotecnológicos ³.

La fermentación es el proceso metabólico por el cual las células utilizan una fuente de carbono en condiciones aeróbicas produciendo ácidos orgánicos; principalmente etanol, ácido acético, ácido láctico y ácido cítrico ²⁵. Por ejemplo, los egipcios utilizaron la fermentación desde hace más de 4000 años para producir cerveza. En la actualidad se conocen más de 5000 distintos alimentos y bebidas fermentados en el mundo siendo los lácteos y las bebidas alcohólicas los que dominan el mercado ^{1,10,26,27}. Esta gran variedad puede explicarse por la diversidad de microorganismos, sustratos y condiciones de producción.

Las bacterias ácido lácticas (BAL), ácido acéticas (BAA) y las levaduras son los microorganismos fermentadores por excelencia utilizados en la industria alimenticia. *Saccharomyces cerevisiae* (*S. cerevisiae*) es un microorganismo crítico en la producción de bebidas fermentadas alcohólicas, como el vino y la cerveza. *S. cerevisiae* requiere de concentraciones bajas de oxígeno (anaerobio facultativo), pH 4.5–6.5, una temperatura entre 20°–30° C y una fuente de carbono (sacarosa, fructosa, glucosa, maltosa, maltotriosa, etc.). Los principales productos son el etanol, CO₂ y energía (2 ATP). *S. cerevisiae* es sensible a concentraciones elevadas de etanol (> 10-15% w/v), por lo cual se han generado levaduras resistentes a dicho metabolito para mejorar su producción ¹⁰. Por otro lado, las BAL y BAA presentan una gran plasticidad metabólica para la síntesis de diversos compuestos (ácidos orgánicos), una tasa de división alta y pueden cultivarse y manipularse con facilidad en el laboratorio. Dichas propiedades se han explotado en la industria de alimentos fermentados para la producción de quesos, yogurt, algunos cárnicos, bebidas alcohólicas, etc. Además, dada su fácil manipulación genética e

inocuidad son utilizadas en la industria biotecnológica para la síntesis de antibióticos, biopolímeros y en el control de plagas ^{28,29}.

2.1.1 Clasificación de los alimentos fermentados

Propuesta de clasificación de los alimentos fermentados. Existen varias clasificaciones de los alimentos fermentados, estas son complejas y confusas ^{25,27,30-33}. La versión aquí presentada tiene como objetivo clasificar a los fermentados partiendo de una visión general hasta una particular y técnica. Los ejemplos tienen la intención de evidenciar que las clasificaciones no son excluyentes sino complementarias:

- i. Por el objetivo del productor: venta (cerveza), conservador (queso Camembert).
- ii. Por el uso del consumidor: recreativo (sake), complemento dietético (yogurt con búlgaros).
- iii. Por su forma de consumo: alimento (sourdough) y bebida (vino).
- iv. Por el grado de industrialización: industrializado (cerveza), semi-industrializado (sourdough) y no industrializado/tradicional (pulque).
- v. Por el origen de los microorganismos: fermentación espontánea (microorganismos silvestres – segunda fermentación en cervezas) y fermentación no espontánea (inóculo- quesos maduros).
- vi. Por la composición microbiana: mono-fermentación (un microorganismo – masa fermentada), poli-fermentación (consorcio de microorganismos – crème fraîche).
- vii. Por el origen del sustrato: animal (jamón serrano), vegetal (sauerkraut).
- viii. Por el tipo de sustrato: cereales (pan de masa madre), vegetales (kimchi), leguminosas (kinema), raíces-tubérculos (fu fu), lácteos (yogurt), cárnicos-embutidos (salami), pescados-mariscos (surströmming) y productos misceláneos (cacao).
- ix. Por el principal metabolito producido: fermentación alcohólica – etanol (sake), fermentación ácido-láctica - lactato (kéfir) y fermentación acética – acetato (vinagre).

Los métodos de secuenciación masiva podrían ayudar a clasificar los alimentos fermentados en función del tipo de microorganismos y su capacidad funcional. Por ejemplo, agrupándolos en oligotipos o filotipos de manera similar a los enterotipos ^{34,35}

2.1.2 Importancia de los estudios metagenómicos

Las comunidades microbianas están conformadas por los microorganismos que viven en un espacio definido y las interacciones que tienen entre ellos (microorganismo-microorganismo, microorganismo-huésped y microorganismo-factores abióticos) ³⁶. La gran mayoría de los microorganismos no se pueden crecer en el laboratorio, son no cultivables. La secuenciación masiva es una técnica independiente de cultivo siendo ideal para describir la composición microbiana en ambientes distintos o a través del tiempo ^{4,37}. Además, permite conocer el potencial funcional (funciones presentes en una comunidad) y ensamblar genomas que sirvan como referencia y ayudando en la elaboración de medios de cultivo adecuados para aislar microorganismos no cultivables.

La metagenómica es una herramienta muy poderosa ya que ha permitido establecer la diversidad y composición microbiana, definir el potencial funcional y metabólico, ensamblar genomas y hacer inferencias sobre posibles interacciones entre los microorganismos (Figura 2) ³⁸⁻⁴⁰. Por ejemplo, Parks et al. en 2018 lograron ensamblar cerca de 8000 genomas a partir de datos metagenómicos públicos, Delmont et al. en 2018 identificaron bacterias nitrificantes en mares que no habían sido reportadas y que podrían tener un impacto en procesos biogeoquímicos marinos, y en los últimos años se ha propuesto métodos para inferir interacciones metabólicas dentro de una comunidad microbiana ^{41,42}.

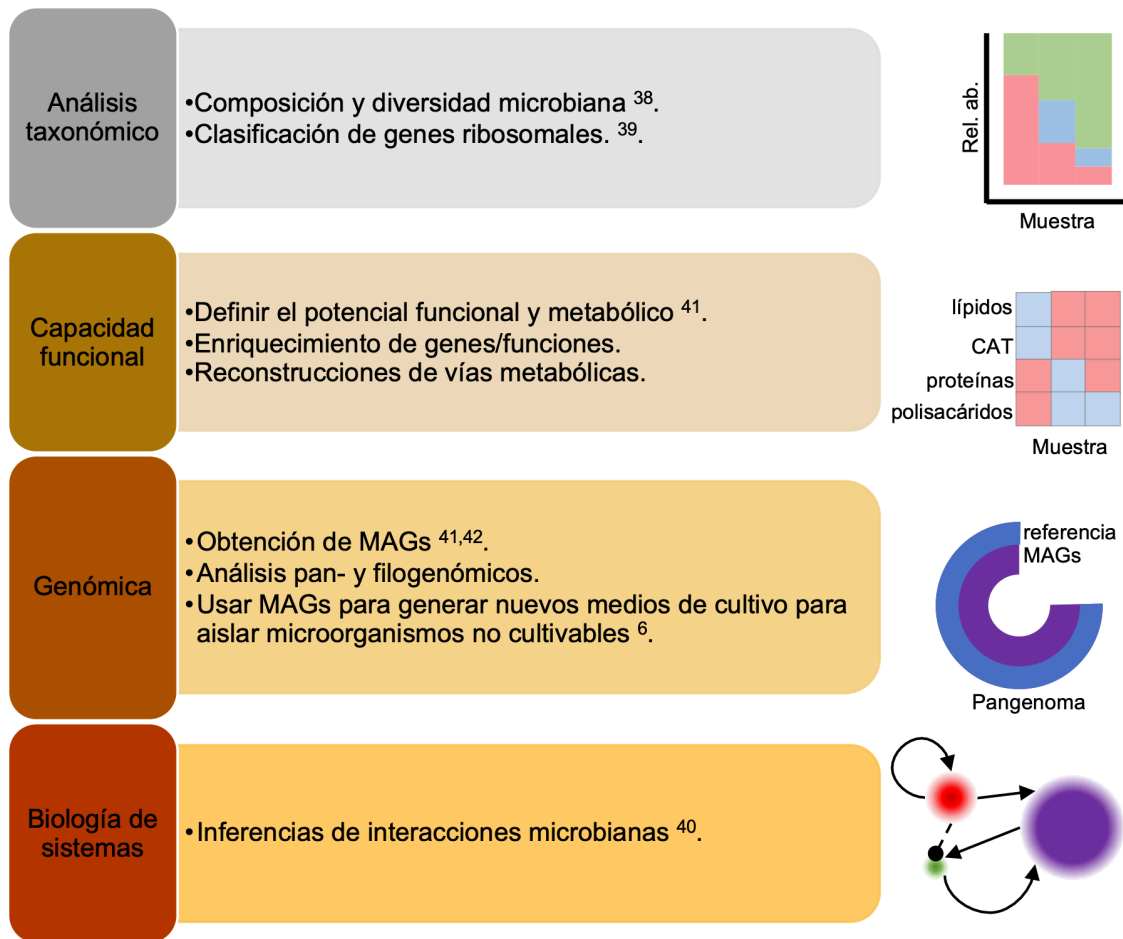


Figura 1. Aplicaciones de la metagenómica.

3 Antecedentes

3.1 Caracterización de comunidades microbianas

3.1.1 Microorganismos cultivables y no cultivables

Los microorganismos se pueden clasificar en cultivables y no cultivables en función de la capacidad que tenemos para crecerlos en el laboratorio. La gran mayoría de éstos pertenecen al segundo grupo y podrían tener un papel crítico en sus comunidades. Nuestra capacidad para aislar microorganismos se limita por la gran diversidad metabólica que tienen y porque la gran mayoría de la investigación se centra en los

patógenos. La incorporación de uno o algunos pocos nutrientes (sales), así como modificar las condiciones de crecimiento (concentración de oxígeno) pueden mejorar nuestro porcentaje de aislamiento. En otros casos se requieren medidas extremas como utilizar una temperatura de crecimiento muy elevada ^{43,44}. También se han identificado microorganismos (no cultivables) con genomas reducidos que viven dentro de otros organismos (endosimbiontes). Con ayuda de la secuenciación masiva se ha demostrado que algunos carecen de vías de síntesis de aminoácidos esenciales para ser aislados en el laboratorio. Al complementar el medio de cultivo dicho se han logrado aislar algunas bacterias ^{6,45,46}.

3.1.2 Secuenciación masiva de comunidades microbianas

La secuenciación masiva surge a partir del Proyecto del Genoma Humano con la intención de mejorar la profundidad de secuenciación no lograda por secuenciadores de la época. El inconveniente fue la generación de grandes cantidades de información que debía ser procesada promoviendo el rápido desarrollo de la bioinformática. Bases de datos de fácil y rápido acceso, plataformas y programas automatizados y semiautomatizados y programas para ensamblar genomas y anotar genes son resultado de la generación de masiva de información ⁴⁷.

Desde una perspectiva microbiológica la secuenciación masiva tiene dos variantes: i) a partir de los genes ribosomales 16S o las regiones espaciadoras intergénicas (ITS – por sus siglas en ingles) y ii) a partir del ADN total. El primer tipo de secuenciación es conocido como secuenciación de perfiles taxonómicos o secuenciación por 16S siendo ideal para identificar a los miembros de una comunidad. El segundo tipo se le llama metagenómica y permite conocer los genes presentes en el ambiente y ensamblar genomas ^{46,48}.

Cada tipo de secuenciación tiene ventajas y desventajas como la información que se puede obtener, los requerimientos computacionales y el costo (Tabla 1) ⁴⁹. En este

trabajo se utilizará el término metagenómica para hacer referencia a la secuenciación masiva independiente de sonda.

Tabla 1. Ventajas y desventajas de la secuenciación masiva dependiente e independiente de sonda.

	Tipo de secuenciación	
	Dependiente en sonda Perfil metagenómico (ADNr 16S ó ADNr 18S)	Independiente de sonda Metagenómica (<i>shotgun</i> / genoma completo)
Ventajas	<p>Se analizan microorganismos cultivables y no cultivables.</p> <p>Se obtiene información cualitativa y cuantitativa.</p> <p>Muy útil para explorar.</p> <p>Bases de datos de genes ribosomales actualizadas y curadas.</p>	<p>Se analizan microorganismos cultivables y no cultivables.</p> <p>Se obtiene información funcional.</p> <p>Se pueden ensamblar genomas.</p> <p>Los genomas se pueden clasificar hasta nivel de cepa.</p>
Desventajas	<p>No se obtiene información sobre las funciones.</p> <p>No todos los microorganismos se clasifican bien con los genes ribosomales.</p> <p>Es muy complicado asignar hasta nivel de especie.</p>	<p>Los parámetros de calidad para el ensamblaje son poco claros para metagenomas.</p> <p>Las bases de datos de funciones no están actualizadas o curadas.</p> <p>Muchos genes no se logran anotar funcionalmente.</p>

Mardis et al. (2008) y Ranjan et al. (2016)

3.1.3 Microscopia de comunidades microbianas

La distribución espacial de microorganismos es un fenómeno reconocido desde los inicios de la microbiología. El entendimiento de procesos biogeoquímicos como la fijación de nitrógeno en los océanos usando una columna de Winogradsky son ejemplos claros de una distribución selectiva y posible interacción entre microorganismos y sus metabolitos ⁵⁰. El quórum *sensing* es otro ejemplo de interacción espacial de microorganismos ⁵¹. A microescala se sabe que los microorganismos pueden agregarse y formar biofilms. Sin embargo, estudiar estas micro comunidades representa un gran reto. Amann et al. (1990) desarrollaron un método basado en citometría de flujo e hibridación de sondas dirigidas contra el ADNr 16S, para analizar comunidades bacterianas mixtas ⁵². Valm et al. (2012) visualizaron de manera selectiva diferentes taxones bacterianos en cultivo mixto usando microscopia de fluorescencia. En dicho trabajo diseñaron sondas marcadas con distintos fluoróforos dirigidas a distintos taxones bacterianos que hibridan con distintas regiones de los genes ribosomales Esta técnica lleva el nombre de CLASI-

FISH (Combinatorial Labelling and Spectral Imaging) y se empleó para evaluar la distribución espacial de las bacterias que conforman la placa dentobacteriana ^{23,24}. Esto incrementa la complejidad de entender a las comunidades microbianas; ya que no solo debemos tomar en cuenta: ¿quién está ahí?, ¿qué puede hacer?, ¿qué está haciendo?; sino también, ¿cómo están organizados espacialmente?

3.2 Metagenómica en alimentos fermentados

Los alimentos y bebidas fermentados han sido una fuente importante de enzimas y microorganismos de interés biotecnológico. Éstos pueden proveer de nutrientes y probióticos – “microorganismos vivos que, al administrarse en cantidades adecuadas, confieren beneficio a la salud del hospedero” – de acuerdo con la Organización de las Naciones Unidas para la Alimentación y la Agricultura (ONUAA) y la OMS. Además, con el incremento de microorganismos multi-drogo resistentes, se están buscando nuevos péptidos y enzimas con actividad antimicrobiana que pudieran sustituir a los antibióticos. Los alimentos fermentados, como el pulque, podrían ser una fuente de probióticos y péptidos antimicrobianos ^{3,53–57}. La búsqueda de estas enzimas, péptidos y microorganismos puede realizarse mediante el aislamiento de los microorganismos y/o por metagenómica. La mayor parte de la investigación en alimentos fermentados está enfocada en la identificación de patógenos y en el aislamiento y caracterización de microorganismos con potencial industrial. Esto reduce de manera importante la capacidad de encontrar nuevas funciones.

Pocos alimentos fermentados han sido estudiados empleando técnicas de secuenciación masiva, aunque se realiza con mayor frecuencia. Estos estudios han permitido conocer de manera más precisa la diversidad y abundancia de bacterias, hongos y virus que están presentes antes, durante y al final de la fermentación, identificar genes que codifican para potenciales bacteriocinas, defensinas y péptidos antimicrobianos, definir el potencial funcional, etc. ^{32,58,59}. Cao et al. (2017) y Rizo et al. (2018) hacen muy buenas revisiones sobre los resultados más relevantes en alimentos fermentados utilizando

metagenómica y otras ómicas (metatranscriptómica, proteómica y metabolómica), respectivamente ^{7,60}.

Los vegetales fermentados y los productos lácteos han sido ampliamente estudiados. Por ejemplo, uno de los primeros trabajos de metagenómica en alimentos fermentados fue el realizado por Jung et al. (2011). En dicho estudio, a partir de pirosecuenciación, comparan el perfil de bacterias (identificación del gen ADNr 16S), con el perfil metabólico (resonancia magnética nuclear - RMN) y el enriquecimiento de funciones a lo largo de la fermentación de kimchi, sugiriendo que 3 géneros (*Leuconostoc*, *Lactobacillus* y *Weisella*) son los principales responsables del proceso metabólico ⁶¹.

Wolfe et al. (2014) analizaron las comunidades microbianas presentes en quesos maduros. Ellos identifican que las condiciones ambientales (el tipo de costra y la humedad) correlaciona mejor que el origen geográfico del queso con la diversidad microbiana (secuenciación del gen ADNr 16S y el ITS para bacterias y hongos, respectivamente) y con el potencial funcional obtenido de secuenciación por *shotgun*, de tal forma que la comunidad microbiana y el enriquecimiento de algunas funciones son características del tipo de costra de los quesos ¹¹.

En México se han realizado muy pocos trabajos en alimentos fermentados; la mayoría se han enfocado en los quesos y en describir la composición microbiana (Tabla 2) ^{4,62-64}. El trabajo realizado por Escobar-Zepeda et al. (2016) es el único, hasta la fecha, que ha empleado metagenómica por *shotgun*. En dicho trabajo identifican a 3 especies (*Lactobacillus plantarum*, *Leuconostoc mesenteroides* y *Weisella paramesenteroides*) como las más abundantes en el queso Cotija (80% de abundancia relativa) y obtienen sus genomas. Además, identifican potenciales bacteriocinas que podrían ser analizadas *in vitro*. Por último, ellos sugieren que tienen una huella bacteriana y genética que podría ser utilizada como biomarcador del queso Cotija ⁴.

La escasez de estudios metagenómicos en alimentos fermentados de México abre la posibilidad de explorar y analizar nuevos productos.

Tabla 2. Análisis de secuenciación masiva en alimentos fermentados de México.

Alimento	Tipo de secuenciación (microorganismos analizados)	Relevancia
Queso de Poro ⁶²	Basado en sonda (bacteria)	Cambios en la diversidad de bacterias dependiendo de la temporada del año.
Queso "bola de Ocosingo" ⁶³	Basado en sonda (bacteria)	<i>Sterptococcus</i> spp. y <i>Lactobacillus</i> spp. son los microorganismos dominantes. Identificación de <i>E. coli</i> en casi todas las muestras.
Queso Cotija ⁴	Metagenómica (bacteria)	Identificación de 3 especies dominantes (80%): <i>Lactobacillus plantarum</i> , <i>Leuconostoc mesenteroides</i> y <i>Weissella paramesenteroides</i> . Identificación de clusters de bacteriocinas.
Varios quesos (adobera, asadero, canasta, cíncho, Chihuahua, Cotija, panela, joqcoque, rancharo, tipo-manchecho, Zacatecas) ⁶⁴	Basado en sonda (bacteria / hongos)	Análisis de bacterias y hongos, su co-presencia y exclusión. Sugieren un núcleo (<i>core</i>) de microorganismos.
Aguamiel, contrapunta y pulque ¹⁷	Basado en sonda (bacteria / hongos)	Primer trabajo de bacterias y hongos usando genes ribosomales.

3.3 El pulque

3.3.1 Aspectos generales

El pulque es una bebida tradicional mexicana que se caracteriza por ser ligeramente ácida, viscosa, alcohólica (4-7° GL), y de apariencia lechosa. En la actualidad, la mayor producción y consumo se concentra en el centro del altiplano mexicano donde colindan los estados de Hidalgo, Tlaxcala, el Estado de México, y Querétaro. La máxima producción y consumo se da en el valle del Mezquital en el estado de Hidalgo. Otros estados productores de pulque son: Morelos, Michoacán y Oaxaca; así como, algunas regiones rurales de la Ciudad de México ^{15,16}.

3.3.2 La elaboración de pulque

La producción de pulque se lleva a cabo en cuatro etapas que prácticamente no han cambiado desde tiempos prehispánicos (Figura 2) ¹³⁻¹⁶:

- a) La maduración, castración del maguey y el reposo: los agaves se dejan crecer entre 7 y 15 años hasta que comienzan a surgir los pedúnculos embrionarios florales. Cuando esto sucede, los pedúnculos embrionarios florales son eliminados y se realiza una cavidad en el núcleo de la planta (corazón del maguey) y se cubre el orificio. El agave castrado se deja reposar por algún tiempo (semanas o meses) previo a la extracción de aguamiel (Figura 3a).
- b) El raspado: esta operación consiste en estimular los tejidos de la pared interna de la cavidad del maguey para permitir la salida del aguamiel. Por lo general, se realiza dos veces al día (en la mañana y por la noche) durante todo el tiempo de vida de producción de la planta (~1 mes). El aguamiel es una solución azucarada que funcionará como sustrato de la fermentación (Figura 3b).
- c) Generación de la semilla: cuando se quiere empezar una nueva fermentación (un nuevo lote), aguamiel alta calidad (de acuerdo con el productor) se deja fermentar (fermentación espontánea) hasta que alcance las propiedades que busca el productor. Este fermentado es llamado *semilla* y es utilizada como un inóculo (Figura 3c).

d) La fermentación: cuando se empieza un lote nuevo de fermentación, la semilla es colocada en el tinacal (función de biorreactor) y se le adiciona aguamiel. En caso de que se quiera continuar la fermentación de un lote, el aguamiel es adicionado al tinacal done se está llevando a cabo la fermentación. Este proceso simula una fermentación de tipo lote alimentado sin agitación. La duración de la fermentación varía entre productores (Figura 3d).

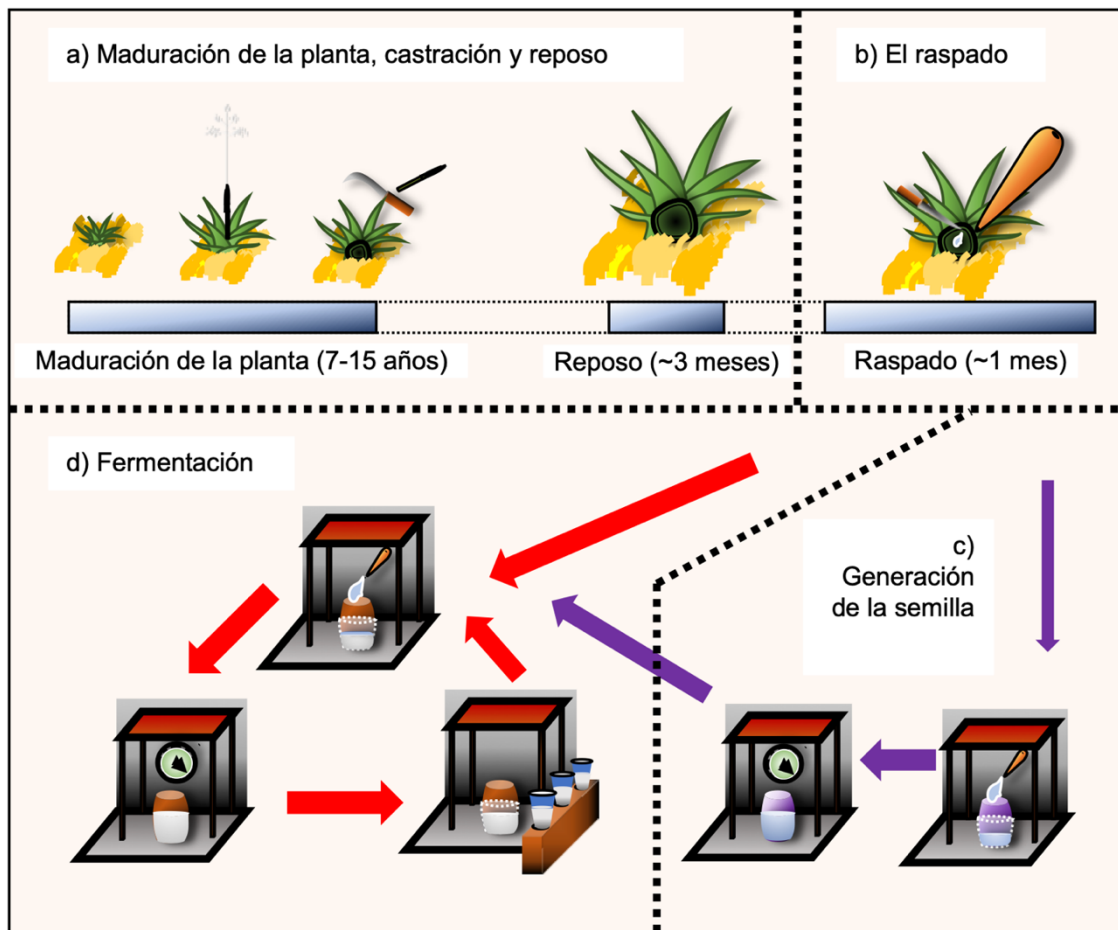


Figura 2. Proceso de elaboración de pulque. Representación gráfica de las 4 etapas involucradas en la producción de pulque: a) Maduración de la planta, castración y reposo – cuadro azul indica el tiempo aproximado de cada etapa, b) el raspado, c) generación de la semilla, y, d) fermentación. Las líneas rojas indican cuando a la fermentación se le adiciona aguamiel simulando una fermentación de tipo lote alimentado. Las líneas moradas indican cuando se genera una semilla y un nuevo lote de fermentación.

El tinacal es el recipiente donde se lleva a cabo la fermentación, y éste puede ser de distintos materiales y formas. Por lo general son de plástico y en algunas ocasiones de madera. Por lo general, el tinacal tiene forma de barril o de barril acostado. Además, el tinacal es resguardado de la intemperie en un lugar fresco y sin luz solar directa, por ejemplo, se deja dentro de un cuarto fresco. El tinacal no es tapado, y esto permite el intercambio de materia (incluyendo insectos), por lo cual el pulque se filtra antes de servirlo ¹⁶.

3.3.3 Aspectos históricos del pulque

Como se comentó anteriormente, la producción de pulque se da desde tiempos prehispánicos. Diversos códices y pinturas prehispánicas sugieren que el pulque era considerado una bebida sagrada en las culturas náhuatl y otomí. Correa-Asencio et al (2014) analizaron lípidos presentes en vasijas de la zona arqueológica de Teotihuacán mediante resonancia magnética nuclear (RMN) e identificaron hopanoides que son sintetizados por la α -PB *Zymomonas mobilis*. Esto sugiere que el pulque era producido y consumido desde los años 200-550 DC ⁶⁵.

San Bernardino de Sahagún describió en “Los usos y costumbres de la Nueva España” que el pulque se consumía como bebida religiosa en los asentamientos indígenas del Altiplano Mexicano y estaba restringido a personas y días específicos. Inclusive, se tenían penas por el incumplimiento de las normas de consumo, las cuales podían llegar a la pena de muerte. También existían usos medicinales para el pulque, y la planta de maguey, del cual se obtiene el aguamiel para la producción del pulque, era utilizada para la fabricación de mecate, utensilios y algunos alimentos como las hojas de la penca. Durante la época colonial, a pesar de las restricciones en el consumo de bebidas embriagantes por parte de los conquistadores, el pulque se popularizó, principalmente entre los indígenas. Además, debido a la explotación de la bebida y de la planta de maguey que generaban abundantes recursos, comenzaron a generarse grandes plantíos en las cercanías de la Ciudad de México. A finales del siglo XIX y principios del siglo XX,

durante el Porfiriato y hasta algunos años después de la Revolución Mexicana, el pulque se convirtió en la bebida nacional. La época de oro del pulque fue posible debido a la alta demanda de esta bebida, así como por la construcción del sistema ferroviario que facilitaba el transporte hacia la capital desde ciudades y provincias de los estados de Hidalgo, Puebla, Tlaxcala y del Estado de México. La gran derrama económica que generaba permitió el establecimiento de las grandes haciendas pulqueras en las cercanías de la capital, así como expendios de pulque llamados pulquerías. La prohibición del consumo de bebidas alcohólicas durante los años 30, la asociación del pulque con un estatus socioeconómico bajo, el largo tiempo de maduración de la planta, la rápida fermentación de la bebida, las reformas agrarias y, eventualmente, la llegada de la cerveza (bebida industrializada) generaron una debacle en la industria pulquera. Para mediados y finales del siglo XX la producción y el consumo de pulque se restringía a comunidades rurales e indígenas del Altiplano mexicano y las haciendas pulqueras quedaron abandonadas, destruidas y saqueadas ⁶⁶.

En los últimos años se ha observado un incremento en la demanda de pulque, principalmente en la capital del país. No obstante, existe una gran competencia en el mercado de bebidas alcohólicas en el país, y el largo tiempo de maduración de la planta, el conocimiento para la elaboración de la bebida (que era transmitido de padres a hijos), la corta vida de anaquel y su bajo costo, hacen al pulque un producto poco redituable.

3.3.4 Aporte nutricional del pulque

El consumo de pulque en comunidades rurales es alto y ha sido asociado con alcoholismo, un problema importante de salud pública en México ^{67,68}. En algunas comunidades, el pulque podría ser una fuente de nutrientes, por ejemplo, se ha cuantificado la concentración de ácido ascórbico en pulque. Éste se encuentra en niveles funcionales, sugiriendo que esta bebida puede ser una fuente de este metabolito. Además, se han identificado fitasas, (enzimas que fosforilan a los fitatos, como el mio-inositol hexaquisfosfato, presentes en semillas) que incrementan la biodisponibilidad de fierro y zinc. El frijol y el maíz son ricos en fitatos y son alimentos básicos de la dieta tradicional

mexicana. Además, algunas bacterias ácido-lácticas aisladas del pulque se han descrito como posibles probióticos (*Lactobacillus* spp.) y como prebióticos en el caso de los exopolisacáridos (EPS) sintetizados principalmente por *Leuconotoc* spp. La vida rural e indígena en México es compleja ya que existe una alta pobreza caracterizada por hambruna y un déficit de nutrientes esenciales en la alimentación. Mas allá del contenido alcohólico de la bebida y la asociación con cirrosis, el pulque pareciera ser una alternativa nutricional para dichas comunidades proporcionando vitaminas, carbohidratos simples y complejos, así como, incrementando la biodisponibilidad de micronutrientes esenciales a partir de frijol y tortillas ^{5,9,69,70}.

3.3.5 Microorganismos en el pulque

A mediados del siglo XX, Sánchez-Marroquín aisló, identificó y caracterizó a distintos microorganismos provenientes del pulque. Además, utilizó un consorcio para producir pulque en condiciones industriales ^{71,72}.

En un trabajo realizado por Escalante et al. (2004), analizaron la composición bacteriana de PQ de 3 distintas regiones geográficas (Aculco, estado de México; Huitzilac, estado de Morelos; y Tizayuca, estado de Hidalgo) a partir de análisis de restricción (ARDRA) y secuenciación del gen ADNr 16S proveniente de bancos de clones de cada región. Los resultados de dicho trabajo indican una diversidad microbiana mayor que la reportada previamente (cuando se utilizaron métodos dependientes de cultivo) en cualquiera de las muestras. Se destaca una elevada proporción de BAL, siendo alrededor del 80% de las clones identificadas en cada banco, siendo el grupo más abundante independientemente de la región. Se identificaron BAL homo- y hetero-fermentativas. Además, se reportó por primera vez a *Lactobacillus acidophilus* en el PQ. La región de Aculco y Tizayuca tienen un perfil (presencia / ausencia) de bacterias más parecido entre ellas que con Huitzilac. (Figura 3) Y, el pulque de Huitzilac contiene α -PB. Desde el punto de vista geográfico, Aculco y Tizayuca son regiones más cercanas entre ellas y ambas al norte de la Ciudad de México, mientras que Huitzilac se encuentra al sur de la Ciudad de México. Además, de manera interesante, *Lactobacillus cepa ASF*, *Lactobacillus acidophilus* y *Leuconostoc*

mesenteroides subsp. *mesenteroides* están presentes en las 3 regiones, pudiendo ser microorganismos claves en la fermentación (miembros del core bacteriano del pulque)

13.

	Aculco (Estado de México)	Tizayuca (Hidalgo)	Huitzilac (Morelos)
<i>Lactobacillus</i> strain ASF	azul	azul	azul
<i>Lactobacillus acidophilus</i>	azul	azul	azul
<i>Lactobacillus kefir</i>	azul	azul	rojo
<i>Lactobacillus acetotolerans</i>	azul	azul	rojo
<i>Lactobacillus hilgardii</i>	rojo	rojo	azul
<i>Lactobacillus plantarum</i>	rojo	azul	rojo
<i>Leuconostoc mesenteroides</i> spp. <i>mesenteroides</i>	azul	azul	azul
<i>Leuconostoc pseudomesenteroides</i>	rojo	rojo	azul
<i>Microbacterium arborescens</i>	rojo	azul	rojo
<i>Flavobacterium johnsoniae</i>	rojo	azul	rojo
<i>Acetobacter pomorium</i>	rojo	rojo	azul
<i>Gluconobacter oxydans</i>	rojo	rojo	azul
<i>Zymomonas mobilis</i>	rojo	azul	azul
<i>Hafnia alvei</i>	azul	azul	rojo

Figura 3. Bacterias identificadas en 3 regiones de México. Se muestra la presencia y ausencia de 14 especies bacterianas identificadas en pulque en 3 regiones distintas de México. Colores: azul - presencia y rojo – no presente. Datos tomados de Escalante et al. (2004).

En un estudio posterior, Escalante et al. (2008) compararon la diversidad microbiana en AgM y a lo largo de una fermentación, empleando una aproximación dependiente e independiente de cultivo, reológica y microscópica. La diversidad microbiana fue determinada por el patrón de restricción (ARDRA) del gen 16S rRNA; las levaduras fueron identificadas por su crecimiento en medio. Las bacterias identificadas a partir de aislados fueron distintas a las de ADNmet., La diversidad bacteriana estuvo compuesta

principalmente por BAL, α -PB y γ -PB. Destacan en el AgM varias especies del género *Leuconostoc* y enterobacterias, como *Enterobacter spp.*, *Erwinia rhapontici*, *Kluyvera cochleae* y *Serratia grimensii*. Durante la fermentación se detectaron BAL del género *Lactobacillus*, y *Leuconostoc*. Además, volvieron a identificar enterobacterias (*Enterobacter agglomerans*, *Kluyvera ascorbata* y *Providencia spp.*). De acuerdo con los autores, la asociación entre los cambios de la diversidad microbiana en la fermentación con los cambios físicos observados les permitió hacer inferencias sobre las características metabólicas de estos microorganismos y proponer su impacto sobre el desarrollo de las características sensoriales de la bebida ¹⁴.

Solamente 6/36 especies y 5/20 géneros fueron identificados por ambos métodos (dependiente e independiente de cultivo) (Figura 3). Esto puede deberse a distintas razones, como el haber descartado a todas las bacterias que no se asignaron a nivel de especie, al método de ARDRA, o a los *primers* que utilizaron. Lo que llama la atención es que ni a nivel de especie ni a nivel de género, utilizando el gen ADNr 16S, se observa una homología entre el MDC y el MIC. Sin embargo, este resultado no es sorprendente si consideramos que las BAL (por ejemplo del género *Lactobacillus*) tienen una identidad de hasta el 99% en el gen ADNr 16S, pero con actividades metabólicas distintas, haciéndolas difíciles de clasificar utilizando solo este gen ⁴³.

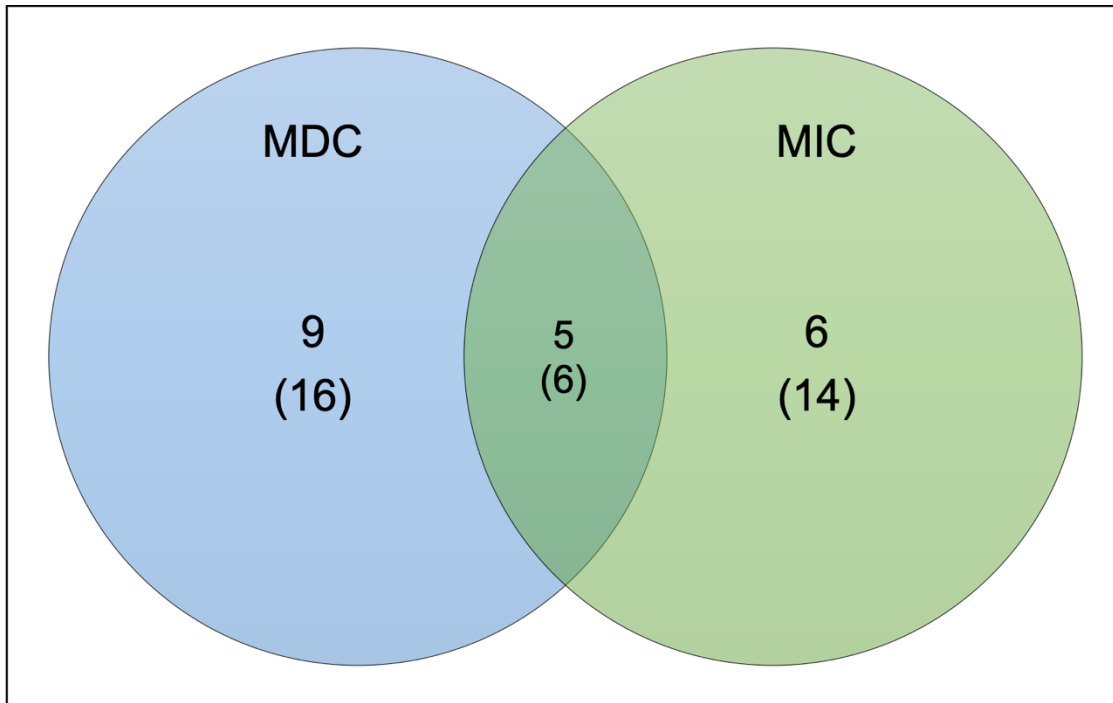


Figura 4. Diagrama de Venn de bacterias identificadas por un método dependiente de cultivo y un método independiente de cultivo. A partir de la Tabla 1 de Escalante et al. (2008); se comparó el número de géneros y especies únicas y compartidas identificadas por método dependiente (MDC) e independiente de cultivo (MIC). Color: azul – MDC; y, verde – MIC. Las especies se muestran entre paréntesis.

Escalante et al. (2016) proponen que durante la fermentación de pulque ocurren 4 procesos microbianos ¹⁵:

- i) Fermentación láctica: promovida por BAL homo- y hetero-fermentativas, principalmente, de los géneros *Enterococcus*, *Lactobacillus*, *Pediococcus*, *Streptococcus* y *Leuconostoc*.
- ii) Fermentación acética: dada por BAA de los géneros *Acetobacter* y *Gluconobacter*.
- iii) Fermentación alcohólica: característica de levaduras y algunas bacterias como *Saccharomyces* spp. y *Z. mobilis*, respectivamente.
- iv) Síntesis de polisacáridos: en este rubro *Leuconostoc* spp. y *Zymomonas mobilis* han sido descritas como productoras de polisacárido extracelular.

Recientemente, se realizó un trabajo donde caracterizan la composición microbiana del AgM, PQ y contrapunta (etapa media de la fermentación) empleando secuenciación masiva del gen ADNr 16S y el ITS. Ellos reportan que el AgM contenía la mayor cantidad de OTU's (unidades taxonómicas operacionales) para bacterias y hongos e identificaron a los filos Proteobacteria y Ascomycota como los más abundantes, respectivamente. Los géneros más abundantes que ellos identifican correlacionan con resultados previos, aunque es de llamar la atención que el género *Sphingomonas* es el más abundante en PQ, disminuye la abundancia de *Leuconostoc* y la de *Lactobacillus* es apenas observable. Este es el primer trabajo empleando el gen ADNr 16S y el ITS para caracterizar la dinámica microbiana en la fermentación de PQ, sin embargo, no se observa una dinámica clara (Figura 5) ¹⁴. Es probable que requieran más muestras y estudiar más tiempos de la fermentación y no debe perderse de vista que varios de estos géneros contienen a especies con propiedades genómicas y metabólicas muy distintas entre ellas como es el caso del género *Lactobacillus* ⁷³⁻⁷⁵. Este tipo de estudios debería contemplar el uso de estrategias de secuenciación y/o de análisis bioinformático que permita resolver los taxones hasta nivel de especie, como *oligotyping* ³⁵.

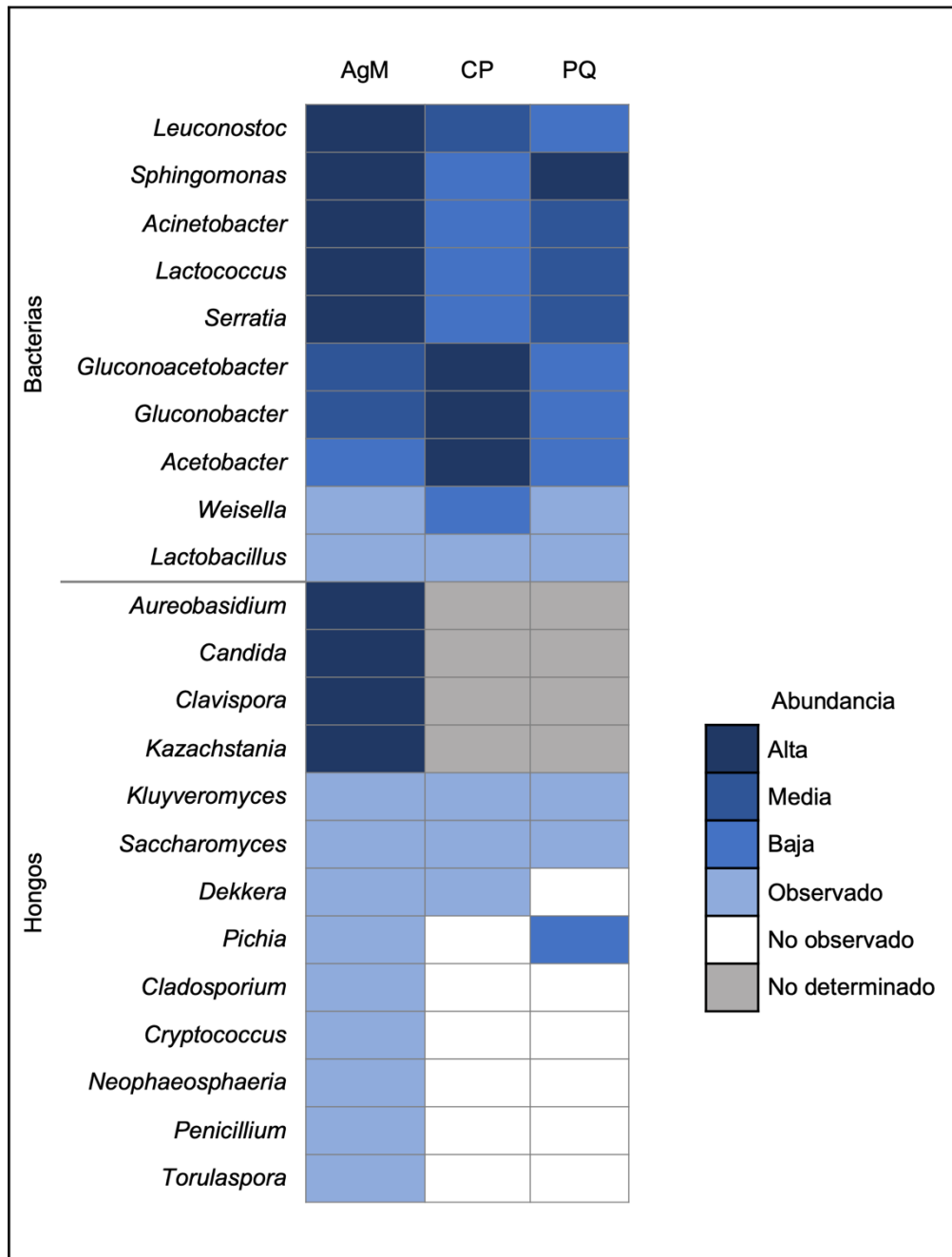


Figura 5. Mapa de calor de abundancias de los géneros microbianos en aguamiel, contrapunta y pulque. Los datos se re-organizaron dividiendo entre bacterias y hongos, de mayor a menor abundancia en aguamiel, contrapunta y pulque. Colores: nivel de abundancia – gradiente de azul; No observado – blanco; y, no determinado (inconcluso) – gris. Arriba – bacterias; y abajo – hongos. AgM – aguamiel, CP –contrapunta, PQ – pulque. Datos tomados y modificados de Rocha Arriaga (2017).

El PQ ha demostrado ser una fuente de microorganismos con potencial biotecnológico. *Z. mobilis*, aislada de agaves e identificada en el PQ, es utilizada para la producción de bioetanol ^{2,12}. Algunos microorganismos aislados del PQ han sido utilizados como consorcio para la producción de una bebida tipo pulque ^{71,72}. Además, se han obtenido y caracterizado cepas con potencial probiótico y/o con la capacidad de producir algún prebiótico ^{5,57,76,77}. Los géneros microbianos más abundantes en AgM y PQ de distintos estudios son similares entre sí, con algunas excepciones. Sin embargo, es probable que existan más funciones en el PQ de las que ya se han caracterizado. El poder analizar la diversidad microbiana y funcional en AgM, PQ y a lo largo de una fermentación sería de gran importancia en la búsqueda de nuevas y mejores funciones.

4. Justificación

El pulque es una bebida fermentada de importancia cultural, económica y social de México. Los trabajos realizados hasta la fecha para caracterizar a los microorganismos de esta bebida no han abordado el potencial funcional microbiano. Mediante el uso de secuenciación masiva por *shotgun*, la cuantificación de metabolitos y un acercamiento microscópico se pretende definir la diversidad, composición y organización espacial microbiana a lo largo de la fermentación de pulque. Además, se busca definir el potencial funcional microbiano y ensamblar genomas microbianos.

5. Hipótesis

A partir del análisis de metabolitos, del metagenómico y del microscópico, se observará la dinámica microbiana y metabólica a lo largo de la fermentación de pulque. Estos datos permitirán ensamblar genomas microbianos representativos de la bebida. A través de la microscopía se identificará la organización espacial de hongos y bacterias en aguamiel y pulque.

6. Objetivo general

Analizar el perfil de metabolitos, diversidad, composición y organización microbiana en aguamiel, pulque y a lo largo de la fermentación de pulque realizada en el laboratorio, y ensamblar los genomas de algunos microorganismos.

7. Objetivos específicos

1. Determinar el perfil de metabolitos (sacarosa, fructosa, glucosa, acetato, lactato, succinato, etanol y oligosacáridos), el pH y la temperatura en aguamiel, pulque y en 3 tiempos de fermentación (inicio, mitad y final).
2. A partir de datos de secuenciación por *shotgun*, establecer la diversidad, la composición genómica microbiana (archaeas, bacterias, hongos, protozoarios y virus) y la composición microbiana por genes ribosomales (archaeas, bacterias y hongos) para el aguamiel, pulque y en 3 tiempos de fermentación (inicio, mitad y final).
3. Ensamblar y anotar genomas microbianos tomando en consideración los dos primeros objetivos.
4. Analizar la distribución espacial de hongos y bacterias por microscopía confocal usando sondas dirigidas a los genes ribosomales.

8. Desarrollo metodológico

La Figura 6 muestra el resumen del desarrollo metodológico. Brevemente, en 5 ocasiones distintas se obtuvo AgM y PQ de un productor local de Huitzilac, Estado de Morelos, México. En el laboratorio, se realizó la fermentación de pulque mezclando el AgM y el PQ recién recolectados. La fermentación tuvo una duración de 6 horas. Se tomaron muestras de AgM, PQ, al inicio (T_0h), a la mitad (T_3h) y al final (T_6h) de 4 fermentaciones independientes. A partir de 3 fermentaciones se analizaron los metabolitos por HPLC (sacarosa, fructosa, glucosa, acetato, lactato, succinato, etanol y oligosacáridos), el pH, y la temperatura. De la cuarta fermentación se analizaron los

metabolitos por HPLC (sacarosa, fructosa, glucosa, acetato, lactato, succinato, etanol) y se extrajo el ADNmet para su secuenciación (NextSeq500, pareado 2X 150). Las lecturas (*reads*) de cada muestra (metagenoma) fueron sometidas a múltiples análisis bioinformáticos. Para la microscopía se recolectó AgM y PQ muestras de cada bebida fueron fijadas, procesadas y enviadas a EUA para su análisis.

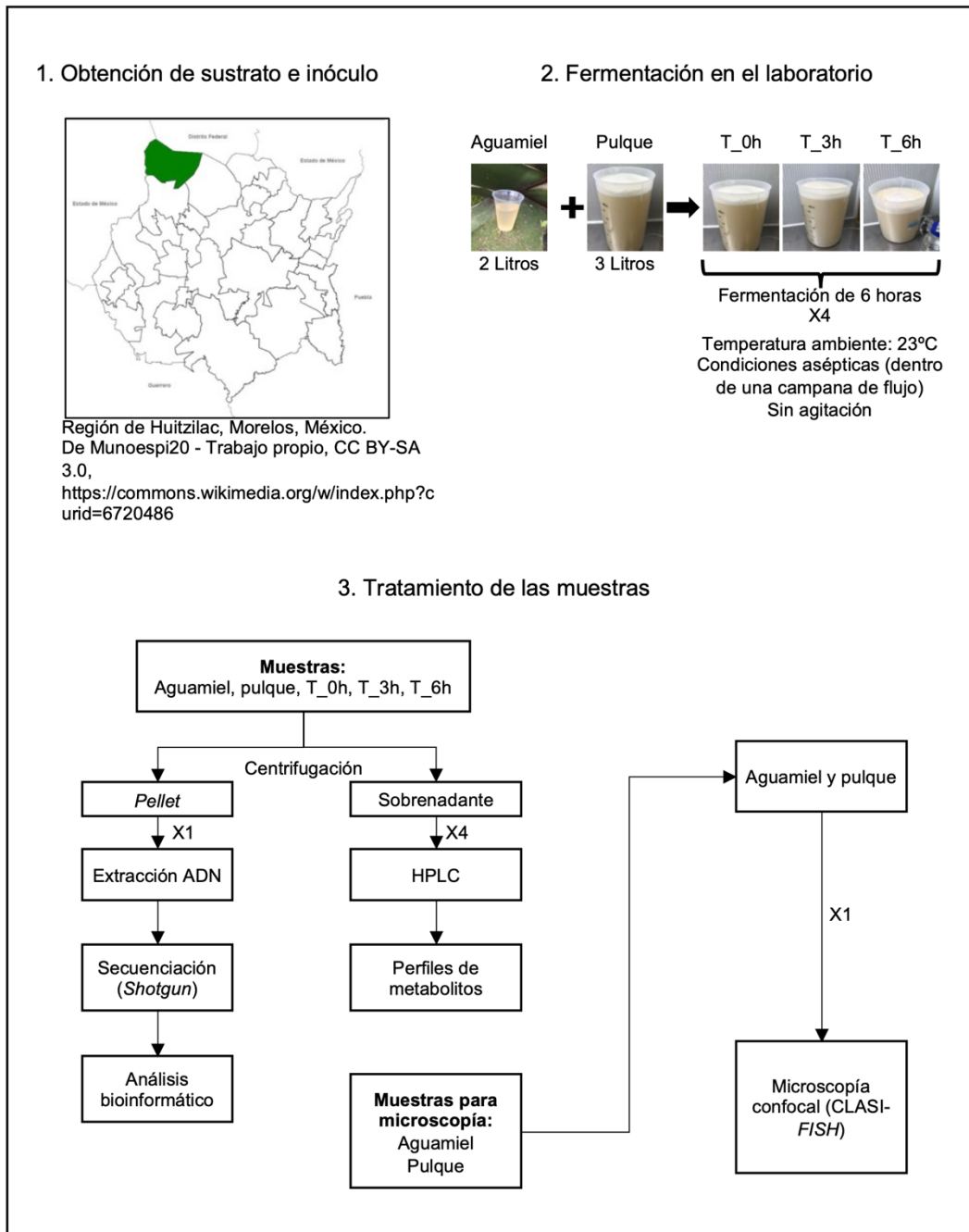


Figura 6. Resumen del desarrollo metodológico. La metodología consistió en 3 etapas: i) la obtención de la materia prima (aguamiel y pulque); ii) la realización de fermentaciones en el laboratorio; y iii) la preparación de las muestras para los 3 experimentos centrales: el análisis bioinformático de los metagenomas, la determinación de metabolitos en el sobrenadante y la adecuación de un protocolo de microscopía confocal para el análisis de microorganismos empleando sondas para genes ribosomales.

8.1 Recolección de aguamiel y pulque

En 5 ocasiones se obtuvieron muestras frescas de AgM y PQ de la localidad de Huitzilac, Estado de Morelos, México, de un productor local (Figura 6 “1. Obtención de sustrato e inóculo”). Las muestras fueron almacenadas en bolsas estériles, en hielo y transportadas de inmediato al laboratorio. Las 5 recolecciones fueron realizadas en distintos tiempos entre enero 2017 y agosto 2018 (Tabla 3).

Tabla 3. Tratamiento de las muestras.

No. Recolección	Fecha (dd/mm/aaa)	Fermentación	Extracción de ADN/ Secuenciación	Análisis de metabolitos				Fijación / Microscopía
				Carbohidratos, ac. orgánicos y etanol	Oligosacáridos	pH	Temperatura	
1	27/01/2017	Si	Si*/Si*	Si*	No	No	No	No/No
2	03/01/2018	No	No	No	No	No	No	Si/Si
3	19/04/2018	Si	No	Si*	Si*	Si*	Si*	Si*/Si**
4	26/04/2018	Si	No	Si*	Si*	Si*	Si*	Si*/Si**
5	30/08/2018	Si	No	Si*	Si*	Si*	Si*	Si*/Si**

* Indica que se analizaron las muestras de cada fermentación: AgM, PQ, T_0h, T_3h y T_6h.

** Indica que se eligió al azar una fermentación para analizar las 5 muestras: AgM, PQ, T_0h, T_3h y T_6h y no se muestran estos resultados porque están en proceso.

8.2 La fermentación y el muestreo

La fermentación se realizó en laboratorio con las muestras de AgM y PQ recién colectadas. Para ello, el AgM y PQ se mezclaron en un biorreactor en una relación 2:3 con un volumen final de 5 litros. El inicio de la fermentación (T_0h) se definió como el momento en el cual se realizó la mezcla y duró 6 horas (de acuerdo con el productor local). Las fermentaciones se llevaron a cabo el mismo día de la recolección y en el momento en que llegaron las muestras al laboratorio (Tabla 3). Todo fue realizado bajo condiciones asépticas (material estéril y el biorreactor se colocó dentro de una campana de flujo); además, la temperatura del cuarto se mantuvo a 23°C, no hubo agitación y el

biorreactor no se tapó, permitiendo la transferencia de gases. Esto fue para mantener las condiciones lo más parecidas a la fermentación en el tinacal del productor.

Las muestras (AgM, PQ, T_0h, T_3h y T_6h) se tomaron por triplicado (volumen de 10 mL) para cada una de las fermentaciones. Uno de los triplicados se utilizó para determinar pH y temperatura, otro para la determinación de metabolitos / extracción de ADNmet y el otro para la microscopía. No en todas las fermentaciones se utilizaron 3 muestras (Tabla 3).

8.3 Extracción de ADN metagenómico

Para separar la fracción no soluble (celular) de la soluble; cada muestra (volumen de 10 mL) fue centrifugada 40 min a 4000 X g a 4°C. La fracción soluble se almacenó en tubos cónicos a -70°C hasta su uso (determinación de metabolitos). La fracción no soluble fue resuspendida en 9.5 mL de buffer TE y se adicionó 0.5 mL de SDS 10% (m/v) y una punta de espátula de Lisozima -A y se incubó por 1 h a 40°C (lisis celular). Se agregó 1.8 mL de NaCl 5M, se mezcló vigorosamente, y se añadió 1.5 mL de CTAB/NaCl; se incubó por 20 min a 65°C (eliminación de proteínas y polisacáridos). Se adicionó un volumen de cloroformo/alcohol isoamílico y se centrifugó a 6000 X g a 4°C. A partir de este punto se usó el kit de MoBio (PowerSoil, DNA Isolation Kit, MOBIO Laboratories Inc.) para la limpieza del ADN. La fracción acuosa (arriba) se transfirió a un Spin Filter y se centrifugó de acuerdo con las especificaciones del kit (este proceso se repitió hasta que se filtró toda la fase acuosa). El ADN se lavó con la solución M4 y se concentró con una solución de Tris en H₂O grado milliQ. No se utilizó la solución M5 ya que contiene EDTA y éste puede afectar la actividad de la polimerasa durante la secuenciación ⁷⁸.

8.4 Determinación de metabolitos

Los sobrenadantes se descongelaron y se centrifugaron a 10,000 X g 40min a 4°C y se colocaron en hielo. Cada sobrenadante fue filtrado utilizando membranas con un poro de 0.45 µm hasta obtener un volumen entre 2-3 mL. Un mL del filtrado fue utilizado para

cuantificar por HPLC: sacarosa, fructosa, glucosa, acetato, lactato, succinato y etanol; mientras que el otro mL se empleó para conocer el perfil de oligosacáridos por HPLC.

Para la cuantificación de sacarosa, fructosa, glucosa, acetato, lactato, succinato y etanol se generaron 2 curvas estándar; una para los ácidos orgánicos que incluyó a etanol, acetato, lactato y succinato, y otra para los carbohidratos (glucosa, fructosa y sacarosa). En el caso de los oligosacáridos se utilizó un estándar de inulina.

8.5 Secuenciación

Antes de enviar el ADN a secuenciar se analizó su integridad en un gel de agarosa al 1%, se determinó la relación 260/280 y 260/230 y la concentración de ADN por Nanodrop. Además, se cuantificó el ADN de doble cadena por fluorescencia con el equipo Qubit en la Unidad de Bioinformática y Secuenciación del IBT.

Se envió 1 µg de ADN de doble cadena de cada muestra a la Unidad de Secuenciación del INEMGEN para su secuenciación. Para generar la librería, se utilizaron 50 ng de ADN de doble cadena cuantificado por Qubit HS para ADN de doble cadena. El ADN se etiquetó, se amplificó para ambos índices y se limpió con el kit Agencourt-AMPure XP (Beckman Coulter). Después, las librerías se normalizaron, y se juntaron para ser cargadas en el chip NextSeq 500. La configuración de la corrida fue de muestras-pareadas 2 X 150. Todo el proceso se realizó de acuerdo con las especificaciones de cada proveedor.

8.6 Análisis bioinformático

El análisis bioinformático se dividió en 4 etapas generales (Figura 7).

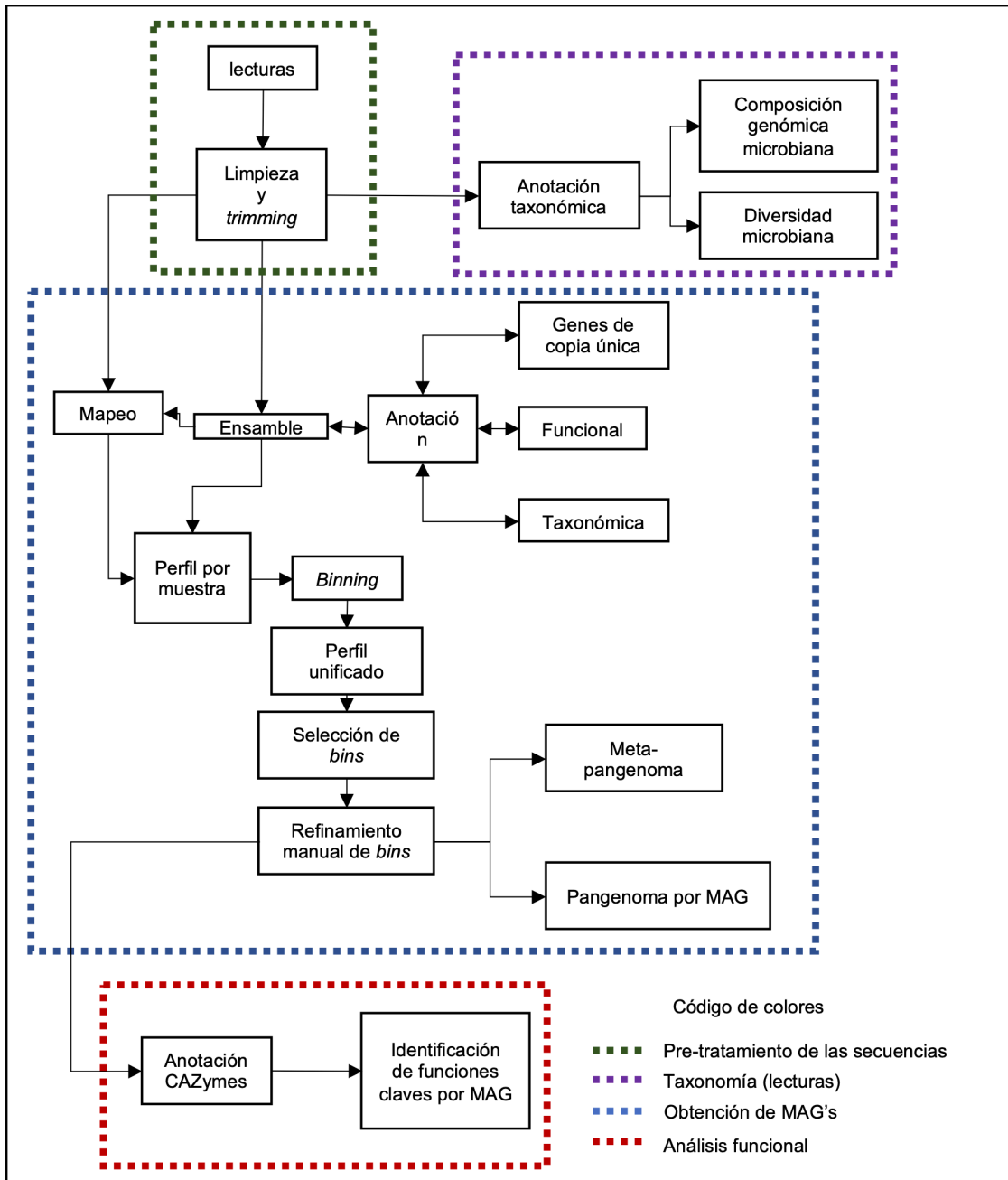


Figura 7. Diagrama de flujo del análisis bioinformático. El trabajo se dividió en 4 procesos generales: i) Análisis de calidad, limpieza y trimming de las lecturas – recuadro verde punteado; ii) Determinación de la composición y diversidad microbiana – recuadro morado punteado, iii) Obtención de MAGs – recuadro azul punteado; y, iv) Análisis funcional de los MAGs – recuadro color rojo punteado.

8.6.1 Análisis de calidad, limpieza y *trimming* (Figura 7, recuadro verde)

A las lecturas obtenidas de la secuenciación se les hizo un análisis de calidad (FastQC), se limpiaron (eliminando aquellas de mala calidad - PHRED SCORE < 20) y recortaron con Trim Galore, eliminando aquellas con una distribución elevada de k-meros repetidos y con adaptadores remanentes.

Programas utilizados: FastQC v0.11.5 (análisis de calidad), Cutadapt v1.13 (remoción de adaptadores), Trim Galore v0.4.4 (limpieza y *trimming*)⁷⁹⁻⁸¹.

8.6.2 Determinación de la diversidad y composición microbiana (Figura 7, recuadro morado)

Diversidad microbiana

Para la diversidad microbiana se definió un OTU como cualquier taxón asignado a cualquier nivel taxonómico. Se empleó Phyloseq (R) de acuerdo con la recomendación del programa y se estimaron los índices: *Observed*, Chao1, Shannon y Simpson.

Composición genómica microbiana

La composición genómica microbiana se definió como el número de lecturas asignadas por OTU dividido por el total de lecturas asignadas a nivel de dominio de la misma librería; empleando como referencia una base de datos (BD) de genomas completos (Figura 8 *Ec. 1*). Se utilizaron 5 anotadores distintos (Centrifuge, KAIJU, KRAKEN, KRAKEN2 y Metaphlan2) y para cada uno de ellos se generó una BD con genomas completos de la RefSeq para: archaeas, bacterias, protozoarios, virus y hongos (cuando fue posible). Metahplan2 utiliza marcadores no ribosomales para la anotación y se utilizó la BD de default. Los resultados se compararon y se eligió una sola anotación, para la cual se calculó la eficiencia (porcentaje de lecturas asignadas por muestra) y precisión de la asignación (Figura 8 *Ec. 2-4*).

$\text{Composición genómica microbiana}(l) = \sum \frac{\# \text{ lecturas } (t)}{\sum \text{lecturas asignadas } (D, l)}$	<i>Ec. 1</i>
$\text{Eficiencia}(D, l) = \frac{\sum \text{lecturas asignadas } (D, l)}{\text{total de lecturas de la librería } (l)}$	<i>Ec. 2</i>
$\text{Eficiencia media}(n) = \frac{\sum \text{Eficiencia}(n)}{m}$	<i>Ec. 3</i>
$\text{Precisión}(n) = \frac{\text{Eficiencia media } (n)}{\text{Eficiencia Max}(D, l)}$	<i>Ec. 4</i>
<p> <i>l = librería</i> <i>t = OTU (taxón)</i> <i>D = nivel taxonómico "dominio"</i> <i>Eficiencia Max (D,l) = Eficiencia (D,l) con el valor más alto</i> <i>m = # librerías</i> <i>n = nivel taxonómico</i> </p>	

Figura 8. Ecuaciones para caracterizar la composición microbiana. La composición genómica microbiana (*Ec. 1*) permitió comparar la proporción de los OTUs (taxones) entre las muestras, sin considerar a las lecturas no asignadas. La eficiencia de asignación (*Ec. 2*) permite comparar la proporción de lecturas asignadas entre librerías (librerías con un mismo origen y tratamiento deberían tener niveles similares de lecturas asignadas). Con la eficiencia media (*Ec. 3*), se compara la proporción de lecturas asignadas en las muestras para cada nivel taxonómico (se espera que la media disminuya al bajar en los niveles taxonómicos). La precisión de asignación (*Ec. 4*) indica que tan cercana o lejana es la eficiencia media respecto a la muestra con el valor más alto de eficiencia de asignación.

Composición de hongos y bacterias por genes ribosomales

Otra forma de conocer la composición microbiana (hongos y bacterias) es caracterizando a los genes ribosomales. Se usaron las lecturas de cada muestra, se identificaron los

OTUs con SortMeRNA (filtrado, mapeo y selección de OTUs) usando la BD de SILVA y luego se clasificaron con la BD de RDP (RDPTools).

Programas utilizados: Centrifuge v1.0.3-beta (anotador taxonómico), KAIJU v1.5.0 (anotador taxonómico), KRAKEN v1.1 (anotador taxonómico), KRAKEN2 v2.0.7-beta (anotador taxonómico), Metaphlan2 v2.6.0 (anotador taxonómico), Phyloseq v1.27.0 (análisis de diversidad), RDPTools v2.0.2 (clasificación de genes ribosomales), SortMeRNA v2.1 (identificación de genes ribosomales)⁸²⁻⁹⁰.

8.6.3 Obtención de MAGs (Figura 7, recuadro azul)

Elección de la plataforma

Lo primero que se hizo fue comparar distintos protocolos y paqueterías para ensamblar, anotar y obtener MAGs; Se compararon 4 atributos: el control (capacidad para modificar parámetros del programa principal y de los programas auxiliares), la transparencia (accesibilidad al código del programa, flexibilidad para revisar los archivos generados y, de manera muy importante, que tuvieran un manual y tutoriales comprensibles), la versatilidad (utilizar programas auxiliares distintos a los sugeridos) y el funcionamiento modular (la facilidad para pasar de una etapa a otra de manera sencilla).

Ensamble

Se probaron 3 tipos de ensamble: i) cada muestra independiente, ii) juntando las muestras en un solo archivo (meta-ensamble) y iii) todas las muestras juntos en archivos distintos (co-ensamble). Se evaluaron 3 ensambladores (IDBA-UD, MEGAHIT y metaSPADES). Se eliminaron los *contigs* con un tamaño menor a 1000 nt para la comparación de la calidad estadística (QUAST) y elegir el mejor ensamble (Tabla 4).

Tabla 4. Combinación de ensambladores y tipo de ensambles probados.

Ensamblador	Ensamble por muestra	Meta-ensamble	Co-ensamble
IDBA-UD*	Si	Si	No
metaSPADES*	Si	Si	No
MEGAHIT**	No	Si	Si

* Ensamblador para datos metagenómicos (1 muestra a la vez)

** Co-ensamblador (+1 librería a la vez)

Mapeo

Las lecturas de cada muestra se mapearon (samtools) contra los *contigs* del meta-ensamble. Primero se generó un índice con los *contigs*. Después, las lecturas de cada metagenoma fueron mapeadas a dicho índice bajo ciertas restricciones: i) el alineamiento fue muy sensible, ii) la lectura completa debía ser alineada (se prohibió que se recortara), iii) los 2 fragmentos (R1 y R2) de cada lectura debían estar cercanas entre ellas; y, iv) las lecturas que no mapearon no se incluyeron en el archivo de salida. Luego, se cuantificó el número de lecturas que mapearon (*total mapping reads*) y el número de lecturas que mapearon correctamente (*total CORRECT mapping reads*) para cada metagenoma. Finalmente, los archivos se transformaron de formato SAM a BAM, se ordenaron y se indexaron.

Generación de la CONTIGS.DB con Anvi'o y predicción de genes (*Gene calling*)

Los encabezados de los *contigs* fueron renombrados para eliminar caracteres extraños (“#,.!” , etc.) que impidieran ser leídos por Anvi'o (python3.6) y facilita la trazabilidad (se les asigna un nombre y un número - “Nombre_000#” a cada *contig*). Después, se generó una BD de *contigs* (CONTIGS.DB) incluyendo sólo los *contigs* \geq 1000 nt y se hizo la predicción de genes (Prodigal).

Anotación de genes de copia única y ribosomales

Anvi'o incluye cuatro BD curadas de genes de copia única. Éstos sirven como marcadores genómicos y taxonómicos. La CONTIGS.DB fue anotada con estas 4 BD (HMMER3).

Anotación funcional

Los genes hipotéticos fueron anotados funcionalmente con 17 BD, que da información para 18 clasificaciones de proteínas. Se anotaron (Blast+) con la BD de ortólogos de NCBI COG's (función / categoría), con GhostKoala de KEGG en su versión gratuita en línea y con la BD de firmas de proteínas de InterPro (incluye 15 BD: PIRSF, ProSitePatterns, SMART, SUPERFAMILY, ProDom, PRINTS, TIGRFAM, ProSiteProfiles, MobiDBLite, Coils, Gene3D, Hamap, CDD, Pfam, SFLD). Las anotaciones fueron incorporadas a la CONTIGS.DB.

Para identificar genes que pudieran codificar para enzimas de degradación o síntesis de carbohidratos en los MAGs, los genes hipotéticos se anotaron con la BD de CAZymes (HMMER3).

Anotación taxonómica de los *contigs*

Los genes hipotéticos fueron anotados taxonómicamente (KAIJU) con la BD no redundante de proteínas de NCBI.

Perfiles únicos, *binning* y perfil unificado

Utilizando la información del mapeo se generaron perfiles únicos para cada muestra para obtener información sobre la cobertura del meta-ensamble, cobertura de cada *contig*, cobertura por gen, %GC, #genes, # genes /kbp, etc. Además, identifica posibles variantes únicas de nucleótidos (SNV) y frecuencias de codones. Se estableció una cobertura mínima de 10X para el llamado de SNV.

Los perfiles únicos se juntaron generando un perfil unificado. Durante este proceso se realizó el *binning* no supervisado (CONCOCT). Éste consiste en agrupar a los *contigs* en *bins* ("canastas" - archivos) en función de su cobertura y del uso de codones. En teoría, estos *bins* ya contienen la información de un genoma. Sin embargo, cada *bin* debe ser revisado manualmente.

Refinamiento de los *bins* (curación de los MAGs)

Se eliminaron aquellos *bins* que no cumplieron con las siguientes condiciones > 1 Mb ó > 90% de genes de copia única. Luego, cada *bin* se revisó y limpió manualmente. La limpieza consistió en seleccionar, basados en un análisis jerárquico, aquellos *clusters* de genes con los cuales se obtuvieran *bins* > 90% de genes de copia única ó > 2 Mbp y < 10% de redundancia de genes de copia única. La anotación taxonómica y los genes de copia única se usaron para la anotación taxonómica. Estos *bins* revisados y limpios ya son MAGs. Por último, se generó un reporte que incluyó información de cobertura, secuencia de nucleótidos, secuencia de aminoácidos y los resultados de la anotación de cada gen para cada MAG.

Programas utilizados: Anvi'o (plataforma y visualizador), Blast+ v2.6.0 (alineamiento de COG's NCBI), Bowtie2 v2.3.4.3 (alineamiento), CONCOCT v1.0.0 (*binning*), GhostKOALA v2.0 (anotación de KO's – versión gratuita en línea), HMMR v3.1b2 (anotación de genes de copia única – archaea, bacteria, eukarya y ribosomales – y CAZymes), IDBA-UD v1.1.3 (ensamblador), InterProScan v5.28 (anotación de InterPro), KAIJU v1.5.0 (anotación taxonómica de *contigs*), MEGAHIT v1.1.3 (ensamblador), metaSPADES v3.12.0 (ensamblador), Prodigal v2.6.3 (*gene calling* – búsqueda de genes), QUAST v5.0.0 (estadísticas de calidad para los ensamblados), samtools v 1.9 (mapeo) ^{19,91–106}.

8.6.4 Análisis funcional (pangenómico) de los MAGs (Figura 7, recuadro rojo)

Análisis pangenómico de los 23 MAGs

Utilizando la información de los 23 MAGs (el reporte generado al final del paso anterior – “*Refinamiento de los bins*”) se hizo una comparación de sus genomas. Este proceso comenzó con el alineamiento de todos los genes en todos los MAGs (Blast+), luego se usó MCL (con un valor de 1; genomas poco cercanos filogenéticamente) para definir la sensibilidad con la cual se hacen los *clusters* de genes y, por último, con MUSCLE, se hizo el alineamiento múltiple de aminoácidos dentro de cada *cluster*.

Análisis pangenómico de *Zymomonas mobilis*

Se descargaron todos los genomas (n=14) de *Z. mobilis* que se encontraron en NCBI (RefSeq). Para cada genoma se generó una CONTIGS.DB y se anotó de la misma manera que los MAG's (Sección 8.6.3 "Anotación funcional"). El análisis pangenómico fue realizado de la misma manera y con los mismos parámetros que para los 23 MAGs (Sección 8.6.4 "Análisis pangenómico de los 23 MAGs"), con excepción del valor de MCL (mcl=1; genomas de la misma especie).

Programas utilizados: Anvi'o (plataforma y visualizador), Blast+ v2.6.0 (alineamiento de COG's NCBI y alineamientos entre genes), GhostKOALA v2.0 (anotación de KO's – versión gratuita en línea), InterProScan (anotación de InterPro), MUSCLE v3.8.1551 (alineamientos múltiples), MCL v14-137 (agrupamiento – *clustering*)^{19,20,83,90,91,94,101-110}.

8.7 Microscopía

Debido a que la muestra es líquida y no se tenía un protocolo estandarizado para analizar por microscopía este tipo de muestras; se probaron 4 combinaciones de fijación: dos tratamientos (fijador) y dos condiciones (centrifugando y sin centrifugar). Se eligieron distintas sondas dirigidas contra genes ribosomales de hongos y bacterias (1 de hongos y 4 de bacterias). Y se analizó la disposición espacial de los microorganismos mediante microscopía confocal siguiendo el protocolo de CLASI-FISH (Figura 9)^{23,24}.

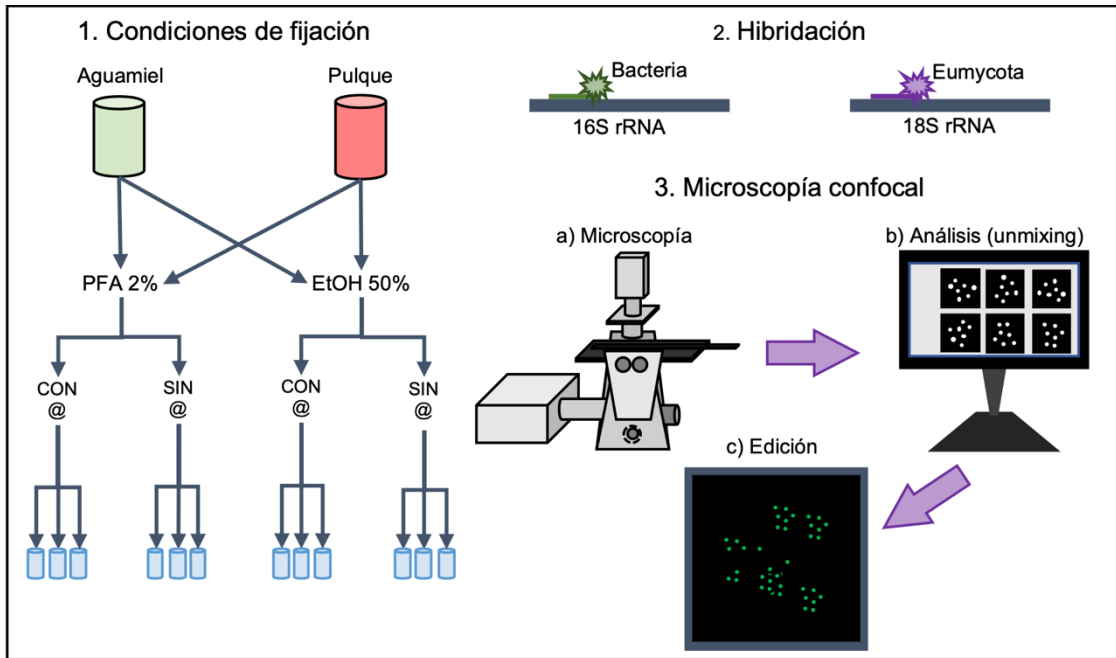


Figura 9. Diagrama de trabajo para el análisis microscópico de aguamiel y pulque empleando CLASI-FISH. 1. Organigrama para la determinar las condiciones de fijación. 2. Ilustración de la hibridación a genes ribosomales. 3. Pasos para la microscopía confocal.

Fijación

Se utilizaron dos métodos de fijación (etanol 50% y paraformaldehído 2%) y dos condiciones (con centrifugación y sin centrifugación) (Figura 9 – “1. Condiciones de fijación”):

- Muestras centrifugadas

Se centrifugaron 3 mL de cada una de las muestras a 1000 rpm por 5 minutos y fueron resuspendidas en 3mL TE (10 mM Tris, pH 7.5, 1 mM EDTA) en tres ocasiones.

Para la fijación en etanol 50%, se adicionaron 3 mL de etanol 100% (volumen final de 6 mL), se mezclaron lentamente por inversión y se colocaron en crioviales a -20°C hasta su envío. Para la fijación en PFA 2%, se adicionaron 3 mL de PFA 4% y se dejaron fijando toda la noche. Las muestras se lavaron 3 veces con 10 volúmenes de TE centrifugando a 1000 rpm por 5 minutos. Las muestras se re-suspendieron en 3

mL TE, se les adicionó 3 mL de etanol 100%, se mezclaron lentamente por inversión y se colocaron en crioviales a -20°C hasta su envío.

- Muestras no centrifugadas

Para la fijación en etanol 50%, se adicionaron 3 mL de etanol 100% (volumen final de 6 mL), se mezclaron lentamente por inversión y se colocaron en crioviales a -20°C hasta su envío.

Para la fijación en PFA 2%, se adicionaron 3 mL de PFA 4% y se dejaron fijando toda la noche. Las muestras se lavaron 3 veces con 10 volúmenes de TE (10 mM Tris, pH 7.5, 1 mM EDTA) dejando que sedimentaran toda la noche. Las muestras se re-suspendieron en 3 mL TE (10 mM Tris, pH 7.5, 1 mM EDTA), se les adicionó 3 mL de etanol 100%, se mezclaron lentamente por inversión y se colocaron en crioviales a -20°C hasta su envío.

Hibridación

La hibridación y microscopía se realizó en el MBL por la Dra. Shamayim Tabita Ramirez Puebla y la Dra. Jessica Mark Welch.

Brevemente, se adicionaron 2 mL de la solución humidificadora (Formamida 20% v/v en H₂O) en 2 toallas (Kimwipe) y se colocó cada toalla en orillas opuestas de la cámara húmeda. Se colocaron 30 µL de las células en un portaobjetos y se dejaron secar por 5 minutos a 46° C en el horno de hibridación (previamente calentado a - 46° C). Los portaobjetos se sumergieron en etanol 70% v/v, etanol 80% v/v, etanol 95% v/v y etanol 100% v/v por 1 minuto cada uno y se dejaron secar al aire a temperatura ambiente. Los portaobjetos se colocaron dentro de la cámara húmeda y se aplicaron 20 µL de la solución para sondas (NaCl 900 mM, Tris pH 7.5 20 mM, SDS 0.01 % m/v, formamida desionizada 20% v/v, y cada sonda 2 pmol/µL; mezclado en H₂O) a las células adheridas. Inicialmente se probaron 5 sondas, sin embargo, solo se utilizaron 2 para los análisis (Eub338 y MY1574) (Tabla 5). Para evitar la evaporación, se colocaron cubreobjetos (22 x 30 mm) sobre las muestras. La cámara húmeda se selló con parafilm y se metió a la campana de extracción. Se dejó incubando por 10 minutos. Se metió la cámara húmeda

al horno de hibridación por 2 horas a 46°C evitando que la cámara estuviera en contacto directo con el piso del horno de hibridación. En la campana de extracción, se adicionaron de 100 - 200 µL de la solución de lavado (NaCl 215 mM, Tris pH 7.5 20 mM, EDTA 5 mM en agua estéril) previamente calentada a 46°C en una de las esquinas del cubreobjetos y se dejó que permeara hasta que el cubreobjetos flotó y se removió con unas pinzas. Los portaobjetos se lavaron sumergiéndolos por 15 minutos en tubos cónicos de 50 mL con solución de lavado precalentada a 48° C. Se adicionaron 100 µL de la solución de DAPI-WGA (Tris pH 7.5 10 mM, DAPI 1 µg/mL y WGA 20 µg/mL) y se incubó por 30 minutos a temperatura ambiente en oscuridad. Los portaobjetos se lavaron con agua helada y con etanol 100%. Se adicionó 1 gota del reactivo para montar (ProLong Gold Antifade Mountant) y se colocó el cubreobjetos (#1.5). Las muestras se almacenaron en oscuridad hasta que se realizó la microscopía (Figura 9 – “2. Hibridación”).

Tabla 5. Descripción de las sondas para microscopía.

Sonda	Organismo blanco	Fluoróforo	Gen blanco	Secuencia (5' - 3')	Referencia
Eub338 (universal)	Bacterias (casi todas)	Alexa 488	16S rRNA	GCTGCCTCCCGTAGGAGT	Amman et al. (1990)
MY1574 (eumycota)	Eumycota (varios grupos-ya probada con <i>Saccharomyces</i>)	Alexa 647	18S rRNA	TCCTCGTTGAAGAGC	Bastien et al. (2008)

Microscopía

Las imágenes se tomaron con un microscopio confocal de escaneo por láser, Zeiss 780 (Carl Zeiss) equipado con un detector de espectro de multiánodos con 32 canales y 3 o 5 líneas de laser. Las imágenes se adquirieron a 40X. El escaneo (Zeiss) de los espectros de las imágenes se adquirió para cada campo en orden descendente de la longitud de onda de excitación; 633 nm, 561 nm y 488 nm. Se adquirió una imagen de fluorescencia estándar a 405 nm de excitación para cada campo cuando se leía la señal de la sonda Eub388- Pacific Blue (Figura 9 – “3. Microscopía confocal”).

9. Resultados

9.1 Determinación de metabolitos

Los metabolitos del AgM y del PQ se han caracterizado en trabajos previos. En nuestro caso, el perfil metabólico se realizó con el objetivo de caracterizar la fermentación realizada en el laboratorio y comparar el final de nuestra fermentación con el pulque del productor local. Además, el perfil permitía monitorear el avance de las fermentaciones.

Se cuantificaron 7 metabolitos: sacarosa, fructosa, glucosa, acetato, lactato, succinato y etanol; así como la temperatura y el pH. La sacarosa, glucosa y fructosa son carbohidratos utilizados como fuente de carbono por muchos microorganismos, mientras que el acetato, lactato, succinato y etanol son considerados productos de fermentación. La temperatura es un indicador indirecto de la actividad microbiana y el pH es un indicador de la acidificación de la bebida por la producción de lactato y acetato (u otros ácidos orgánicos).

Los resultados se muestran en la Figura 10. Las mediciones para aguamiel y pulque se muestran del lado izquierdo de las gráficas (representando los controles) y las mediciones de la fermentación (T_0h, T_3h y T_6h), del lado derecho (representando la fermentación). El pH y la temperatura se monitorearon durante la toma de las muestras (Tabla 6).

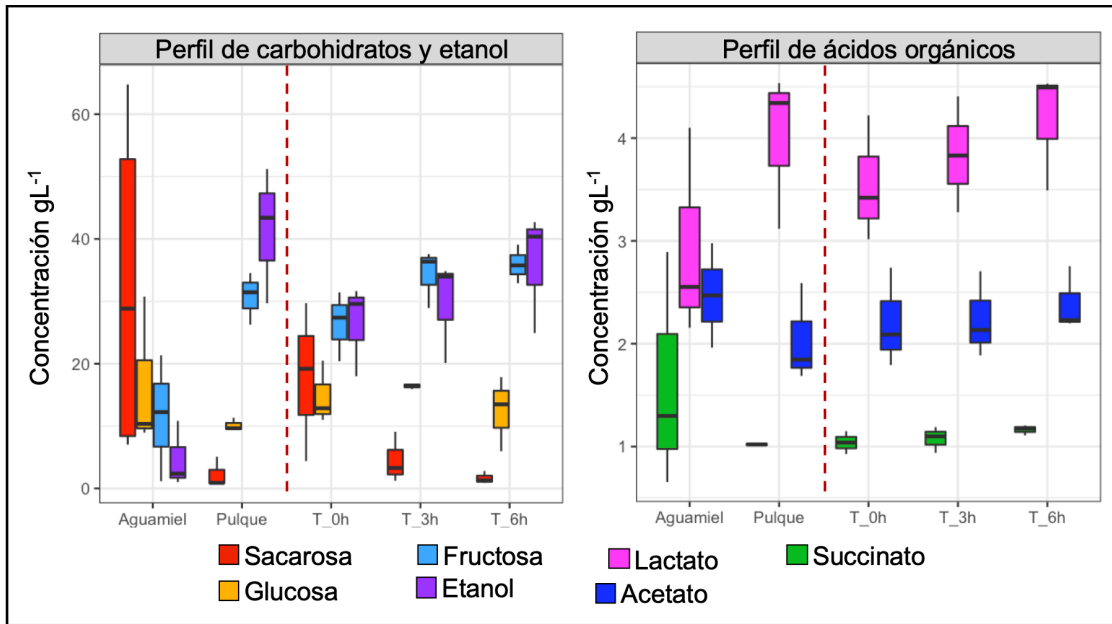


Figura 10. Perfil de metabolitos. Se determinó la concentración en gL⁻¹ mediante HPLC de algunos metabolitos clásicos de la fermentación de pulque. Se realizaron 4 fermentaciones independientes; distinto agave y distinta época del año. Los valores de los metabolitos para el AgM y el PQ se colocaron del lado izquierdo (línea roja punteada) ya que representan los controles de inicio y finalización de la fermentación. Por otro lado, los metabolitos de las muestras T_0h, T_3h y T_6h representan propiamente la fermentación. Izquierda: Gráfica de cajas y bigotes para los de carbohidratos (sacarosa, glucosa, fructosa) y el etanol. Colores: sacarosa – rojo; glucosa – amarillo; fructosa – azul claro; etanol – morado. Derecha: Gráfica de cajas y bigotes para los ácidos orgánicos (acetato, lactato y succinato). Colores: acetato – azul oscuro; lactato – magenta; y, succinato – verde.

Tabla 6. Concentración de metabolitos, pH y temperatura. Los metabolitos están en orden descendente respecto al AgM. Se muestra la concentración o valor medio y la desviación estándar (D.E.) de 3 fermentaciones independientes (No. de recolección 3, 4, 5; Tabla3 “Resumen del tratamiento al que fueron sometidas las muestras de cada fermentación”). La temperatura y el pH se muestran en la parte inferior.

Metabolito/ Temperatura y pH	Aguamiel		Pulque		T_0h		T_3h		T_6h	
	Conc. (g L ⁻¹)	D.E. (±)	Conc. (g L ⁻¹)	D.E. (±)	Conc. (g L ⁻¹)	D.E. (±)	Conc. (g L ⁻¹)	D.E. (±)	Conc. (g L ⁻¹)	D.E. (±)
Sacarosa	40.81	28.80	2.26	2.44	17.75	12.72	4.53	4.08	1.66	1.02
Glucosa	16.04	12.85	10.14	1.05	14.78	5.03	16.41	0.45	12.43	6.01
Fructosa	14.17	6.43	30.75	4.17	26.40	5.58	34.29	4.69	35.91	3.09
Etanol	4.80	5.28	41.43	10.90	26.40	7.36	29.64	8.26	35.99	9.67
Lactato	3.05	0.98	4.00	0.77	3.55	0.61	3.84	0.56	4.17	0.59
Succinato	2.50	0.51	2.04	0.48	2.21	0.48	2.24	0.42	2.39	0.31
Acetato	1.00	0.32	1.02	0.02	1.04	0.11	1.07	0.13	1.16	0.05
Temperatura (°C)	17.3	1.5	19.7	1.2	18.3	1.5	22.3	0.6	24.7	0.6
pH	4.90	0.21	4.24	0.08	4.45	0.14	4.29	0.10	4.15	0.18

El AgM está compuesto principalmente por sacarosa (40.81 g L⁻¹ ± 28.8), glucosa (16.04 g L⁻¹ ± 12.85) y fructosa (14.17 g L⁻¹ ± 6.43). El etanol (4.80 g L⁻¹ ± 5.28) presente en AgM puede deberse a la rápida fermentación de éste, que pudo suceder durante el transporte o desde la colecta. Mientras que el lactato fue el ácido orgánico más concentrado en esta muestra (3.05 g L⁻¹ ± 0.98), seguido por el succinato (2.50 g L⁻¹ ± 0.51) y el acetato (1.00 g L⁻¹ ± 0.32). (Figura 10 – Sección roja y Tabla 6). El pH fue de 4.90 ± 0.21 indicando que el AgM ya estaba ligeramente ácido, y la temperatura fue de 17.3 °C ± 1.5 (Tabla 6).

Por otro lado, el PQ tiene como metabolito más concentrado el etanol (41.43 g L⁻¹ ± 10.90), seguido por la fructosa (30.75 g L⁻¹ ± 4.17), la glucosa (10.14 g L⁻¹ ± 1.05), el lactato (4.00 g L⁻¹ ± 0.77), la sacarosa (2.26 g L⁻¹ ± 2.44), el succinato (2.04 g L⁻¹ ± 0.48) y el acetato (1.02 g L⁻¹ ± 0.02) (Figura 10 – Sección roja y Tabla 6). Su pH fue de 4.24 ± 0.01 y una temperatura de 19.7 °C ± 1.2 (Tabla 6).

Durante la fermentación (T_0h, T_3h y T_6h) se observa un incremento en la concentración de etanol ($26.40 \text{ g L}^{-1} \pm 7.36$ a $35.99 \text{ g L}^{-1} \pm 9.67$), fructosa ($26.40 \text{ g L}^{-1} \pm 5.58$ a $35.91 \text{ g L}^{-1} \pm 3.09$) y lactato ($3.55 \text{ g L}^{-1} \pm 0.61$ a $\text{g L}^{-1} \pm$); una disminución de la sacarosa ($17.75 \text{ g L}^{-1} \pm 12.72$ a $\text{g L}^{-1} \pm$); y la glucosa ($14.78 \text{ g L}^{-1} \pm 5.03$ a $4.17 \text{ g L}^{-1} \pm 0.59$). El acetato ($1.04 \text{ g L}^{-1} \pm 0.11$ a $1.16 \text{ g L}^{-1} \pm 0.05$) y el succinato ($2.21 \text{ g L}^{-1} \pm 0.48$ a $2.39 \text{ g L}^{-1} \pm 0.31$) no tuvieron grandes cambios (Figura 10 – Sección azul y Tabla 6). El pH disminuyó durante el proceso, pasando de 4.45 ± 0.14 al inicio (T_0h) a 4.15 ± 0.18 al final (T_6h); mientras que la temperatura aumentó de $18.3 \text{ }^\circ\text{C} \pm 1.5$ al inicio (T_0h), $22.3 \text{ }^\circ\text{C} \pm 0.6$ a la mitad (T_3h) y $24.7 \text{ }^\circ\text{C} \pm 0.6$ al final (T_6h) de la fermentación (Tabla 6).

La fermentación de PQ llevada a cabo en el laboratorio se caracteriza por la disminución en el pH y el incremento en la temperatura que son características de procesos fermentativos y de actividad metabólica, respectivamente. Además, se observa una fermentación, principalmente etanológica y láctica.

Al comparar el PQ con el final de la fermentación (T_6h) (Figura 10 y Tabla 6), se observa que la concentración de metabolitos y el pH son similares entre las dos muestras. Esto sugiere que la fermentación llevada a cabo en el laboratorio produce una bebida similar al PQ del proveedor.

La temperatura es $2.4 \text{ }^\circ\text{C}$ mayor en el PQ, respecto al AgM. Durante la fermentación, incrementa $5.4 \text{ }^\circ\text{C}$ del inicio (T_0h) al final (T_6h). La temperatura del cuarto de fermentación estuvo controlada a $23 \text{ }^\circ\text{C}$, y esto contribuyó al incremento en la temperatura durante las primeras 3 horas de la fermentación. A pesar de ello, la temperatura final fue de $24.7 \text{ }^\circ\text{C} \pm 0.6$, lo que significa que incrementó $1.4 \text{ }^\circ\text{C}$ respecto a la temperatura ambiente mostrando una actividad microbiana importante.

Las concentraciones de los metabolitos entre fermentaciones variaron mucho (Figura 10 y Tabla 6), esto puede deberse a haber usado AgM y PQ recolectado de distintas plantas y en diferentes épocas del año. Una forma de normalizar y poder comparar entre

muestras es utilizando la cantidad de carbono (moles) que aporta cada metabolito (Figura 11). En este análisis se incluyó la cuantificación de metabolitos de las 4 fermentaciones realizadas.

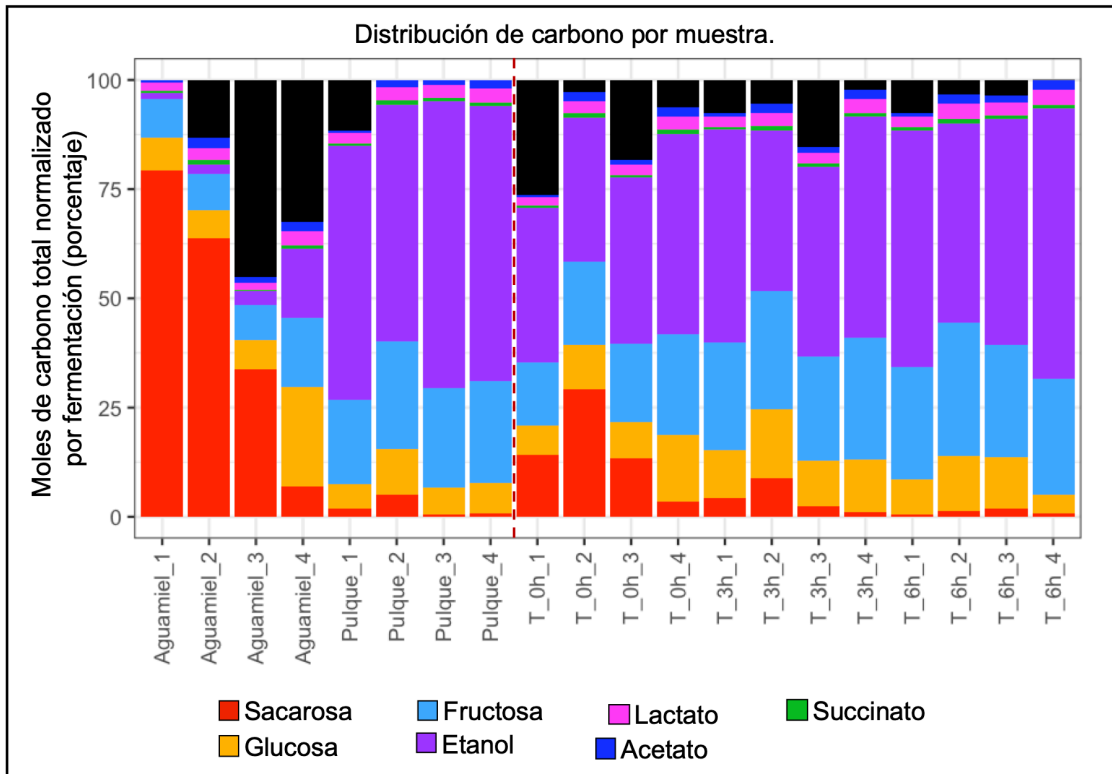


Figura 11. Aportación de carbono de los metabolitos en distintas fermentaciones de pulque. Para cada fermentación se calculó la cantidad de moles de carbono de cada metabolito y se sumó por muestra. La muestra con la mayor cantidad de carbono (moles) se utilizó para conocer el carbono faltante (no identificado) en las demás muestras (de la misma fermentación). Finalmente, se calculó el porcentaje de carbono por muestra para todas las fermentaciones. Separados por línea punteada roja; del lado izquierdo está el aguamiel y el pulque y, del lado derecho el T_0h, T_3h y T_6h. Colores: carbono no identificado – negro; acetato – azul oscuro; lactato – magenta; succinato –verde ; etanol + CO₂ – morado; glucosa – amarillo; fructosa – azul claro; y, sacarosa – rojo. NOTA: para el carbono (moles) aportado por el etanol se incluyeron los moles de CO₂, el cual es liberado durante la descarboxilación del piruvato a acetaldehído.

La sacarosa es el metabolito que aporta en mayor proporción carbono (moles) en el AgM (con excepción del AgM de la fermentación 4 – fecha 30/08/18); mientras que en el PQ, el etanol y la fructosa son los que se encuentran en mayor proporción en las 4 fermentaciones. Durante la fermentación, se observa una disminución en la proporción de sacarosa junto con un incremento del etanol y la fructosa. Esto sugiere que, a pesar de haber diferencias entre la concentración de los metabolitos (Tabla 6), la proporción de carbono (moles) que aporta cada metabolito es similar entre las muestras de cada fermentación.

El aguamiel y el pulque tienen perfiles metabólicos distintos. El AgM es rico en sacarosa y el pulque en etanol, CO₂ fructosa y lactato. En la fermentación se observa un cambio gradual de estos metabolitos (T_0h, T_3h y T6h), la sacarosa se consume casi en su totalidad a las 3 horas y a las 6 horas el etanol, la fructosa y el lactato alcanzan niveles similares al que tienen en el PQ. La concentración de glucosa se mantiene constante en todo momento, y es posible que este siendo utilizada como fuente de carbono de los microorganismos.

Debido a que se observó un incremento en la concentración de fructosa a lo largo de la fermentación, y esta podría estar siendo utilizada para la síntesis de oligosacáridos extracelulares (OSE), se decidió explorar si estaban presentes o no en nuestras muestras. Se utilizó HPLC con un método de detección basado en oxido-reducción y se empleó la inulina (polímero de fructosa) como estándar (Figura 12).

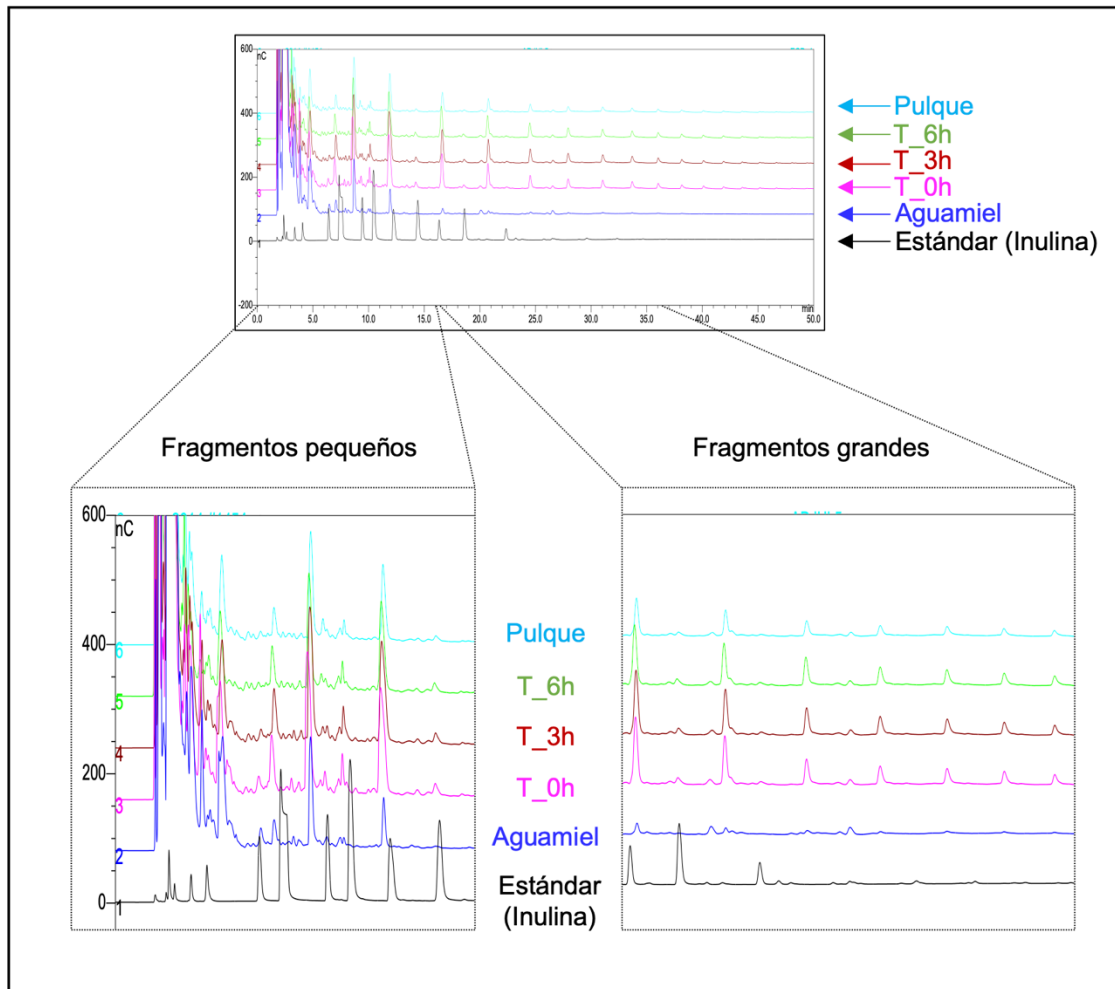


Figura 12. Perfil de oligosacáridos extracelulares durante la fermentación de pulque. Se determinó la presencia / ausencia de oligosacáridos extracelulares (OSE) mediante HPLC con un método de óxido-reducción. Primero, los OSE se separan por tamaño (los monómeros tienen el menor tiempo de retención y los más grandes o complejos son retenidos más en la columna). Después, ocurre una reacción de oxido-reducción y se determina la carga de las moléculas (nano Coulombs – nC). Las muestras fueron filtradas ($0.44\mu\text{m}$) y analizadas para las 3 fermentaciones (no se hidrolizaron). Se utilizó un estándar de inulina (polímero de fructosa) con fines comparativos. Cuadro agrandado izquierdo: fracción con un tiempo de retención 0 – 15 minutos. Cuadro agrandado derecho: fracción con un tiempo de retención 15 – 37 minutos. Colores: pulque – azul cielo, T_6h – verde; T_3h – rojo; T_0h – magenta; aguamiel – azul oscuro; y, estándar – negro.

En los primeros tiempos de retención (0 – 15 minutos), no se observan cambios en los perfiles de OSE (Figura 12 – “*Fragmentos pequeños*”). En tiempos tardíos de retención (15 – 37 minutos), varios picos se definen de manera clara, separando las muestras en dos grupos (Figura 12 – “*Fragmentos grandes*”). El primero conformado exclusivamente por AgM, y el segundo por el PQ, T_0h, T_3h y T_6h. De manera interesante, varios picos de las muestras no coinciden con el estándar sugiriendo que durante la fermentación se producen OSE de un tipo distinto a la inulina o distintos a los incluidos en el estándar de inulina utilizado. La abundancia y conformación estructural de estos OSE no se pueden determinar con este método. Pero, el AgM podría ser un utilizado como control negativo de OSE de la fermentación de PQ.

La fermentación de PQ realizada en el laboratorio es etanológica, láctica, contiene OSE, se acidifica (pH) y tiene actividad microbiana (temperatura). Utiliza la sacarosa como principal fuente de carbono la que es transformada en etanol y lactato y libera fructosa al medio. Además, a las 6 horas de fermentación, la bebida producida en el laboratorio es muy parecida al PQ del productor local.

El usar AgM y PQ de distintas plantas y en diferentes épocas del año pudieron ser los factores que más contribuyeron a la dispersión de las concentraciones de los metabolitos, aunque la proporción de carbono (moles) aportada por cada metabolito, el pH y la temperatura son similar entre las fermentaciones. Dos factores que no se midieron y que podrían haber sido de gran utilidad son la concentración de CO₂ y O₂, éstos son indicadores de la actividad metabólica y del crecimiento de los microorganismos. La biomasa no se determinó ni se calculó.

9.2 Extracción de ADN metagenómico y secuenciación

El ADN metagenómico (ADNmet) se extrajo de cada una de las muestras por triplicado y fue analizado para:

- i) Conocer la integridad del ADNmet usando un gel de agarosa al 1% (Figura 13). Las bandas ubicadas por encima del marcador de peso molecular corresponden al ADNmet y las bandas ubicadas en la parte inferior del gel corresponden a RNA.
- ii) Determinar la concentración de ADN, relación 260/230 y 260/280 por medio de Nanodrop (Tabla 7).
- iii) Determinar la concentración de ADN de doble cadena (Tabla 7).

Y se envió 1 µg de ADN de doble cadena por muestra para su secuenciación.

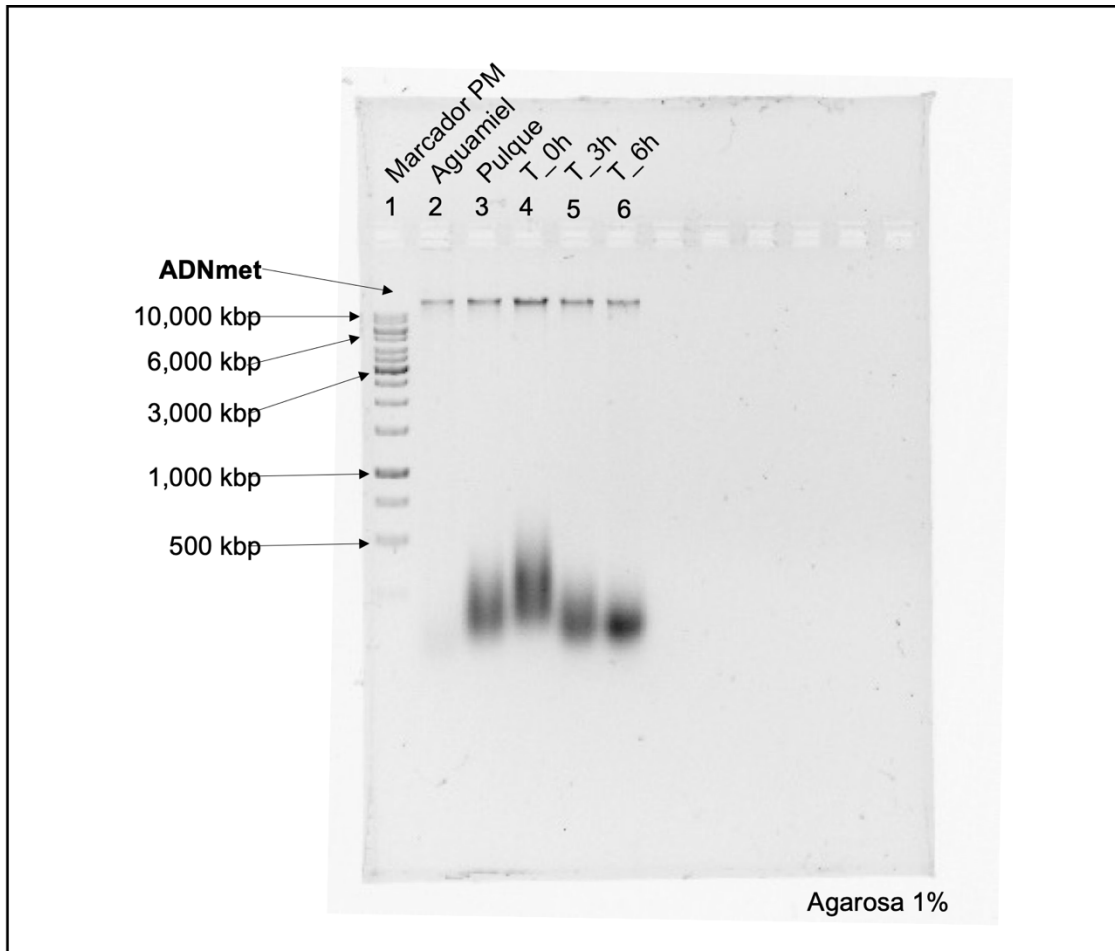


Figura 13. Integridad del ADN metagenómico de las muestras enviadas para secuenciación. Electroforesis del ADN metagenómico en gel de agarosa 1%. Carriles: Carril 1 – Marcador de peso molecular; Carril 2-6 – muestras (aguamiel, pulque, T_0h, T_3h y T_6h). Marcador de peso molecular: GeneRuler 1 kb DNA Ladder, Thermo scientific, #SM0311.

Tabla 7. Pureza y cuantificación de ADN metagenómico para secuenciación. Se realizaron 3 extracciones (réplicas técnicas) de ADN metagenómico por muestra. Cada extracción fue evaluada por espectrofotometría de UV para estimar la pureza (260/280 y 260/230) así como la cantidad de material genético presente (Nanodrop). Se eligió la 2ª réplica para determinar la cantidad de ADN de cadena doble mediante fluorometría (ensayo de dsDNA HS). Se eliminó una de las extracciones de aguamiel, ya que se obtuvo un valor negativo en el índice 260/230 y muy poco ADN.

Muestra	Ensayo de pureza y cuantificación de ADN (Nanodrop)					Ensayo dsDNA HS (Quibit)**				
	Volumen extracción (mL)	Triplicado	Volumen elución (µL)	260/280	260/230	DNA (ng/µL)	DNA (µg)	C _{ensayo} (ng/mL)	C _{ADNcd} (ng/µL)	C _{ADNcd 2:10} (ng/µL)
Aguamiel	10	1	50.00	2.15	2.38	662.60	33.13			
	10	2	50.00	2.19	2.53	722.40	36.12	206	103	20.6
	10	3	50.00	2.17	-6.72*	53.40	2.67			
Pulque	10	1	50.00	2.04	1.75	1368.40	68.42			
	10	2	50.00	2.20	2.40	1028.50	51.43	233	116.5	23.3
	10	3	50.00	2.21	2.60	753.40	37.67			
T_0h	10	1	50.00	2.17	2.49	534.10	26.71			
	10	2	50.00	2.23	2.67	639.30	31.97	494	247	49.4
	10	3	50.00	1.93	1.37	1391.00	69.55			
T_3h	10	1	50.00	2.05	1.75	50.50	2.53			
	10	2	50.00	2.21	2.61	398.50	19.93	210	105	21
	10	3	50.00	2.24	2.53	767.10	38.36			
T_6h	10	1	50.00	2.02	1.69	751.50	37.58			
	10	2	50.00	2.17	2.28	781.10	39.06	252	126	25.2
	10	3	50.00	2.05	1.83	1987.50	99.38			

* Muestra eliminada.

** dsDNA = ADN de doble cadena
Para el ensayo de dsDNA HS (Quibit) se eligió 1 de las réplicas.

9.3 Análisis bioinformático

9.3.1 Análisis de calidad, limpieza y *trimming*

El número de nucleótidos totales obtenidos de la secuenciación fue ~30,000 millones y el tamaño de todos los archivos fue de ~70 gigabytes. A partir de este momento, a las muestras se les llamará metagenoma. Cada metagenoma fue secuenciado en 4 líneas para garantizar la homogeneidad de la secuenciación, que se realizó de manera pareada (R1 y R2), obteniendo 8 archivos de secuenciación para cada metagenoma (total archivos = 40). Se concatenó la información de las 4 líneas de cada metagenoma para R1 y R2. Con esto se obtuvieron 2 archivos por metagenoma, uno con las secuencias R1 y otro con las R2 (total archivos concatenados = 10) (Tabla 8).

Tabla 8. Archivos de secuenciación. En la primer columna está el número de la muestra (n=5). En la segunda columna se indica el nombre de los archivos finales (concatenados) con los cuales se trabajó. La tercer columna muestra el tipo de lectura (R1/R2), la cuarta indica las líneas de secuenciación utilizadas para generar los archivos concatenados. Y , la última columna indica si alguna línea fue eliminada. En nuestro caso no se tuvo que eliminar ningún archivo.

# Metagenoma	Archivos concatenados	Pareado (R1 / R2)	Líneas de la secuenciación	Línea eliminada (Si /No)
1	Aguamiel_R1.fq	R1	L1, L2, L3, L4	No
	Aguamiel_R2.fq	R2	L1, L2, L3, L4	No
2	T_0h_R1.fq	R1	L1, L2, L3, L4	No
	T_0h_R2.fq	R2	L1, L2, L3, L4	No
3	T_3h_R1.fq	R1	L1, L2, L3, L4	No
	T_3h_R2.fq	R2	L1, L2, L3, L4	No
4	T_6h_R1.fq	R1	L1, L2, L3, L4	No
	T_6h_R2.fq	R2	L1, L2, L3, L4	No
5	Pulque_R1.fq	R1	L1, L2, L3, L4	No
	Pulque_R2.fq	R2	L1, L2, L3, L4	No

Para conocer la calidad inicial de las lecturas se utilizó el programa FastQC y se observó una distribución elevada de k-meros repetidos (*k-mer content*). Dichos *k-meros* coincidían con las secuencias de los códigos de barras utilizadas durante la secuenciación, por lo cual se decidió recortar las lecturas con el programa Trim Galore.

La calidad de las lecturas recortadas se volvió a analizar con FastQC y no se observó la presencia de k-meros. Inicialmente, las 5 librerías contenían 253,173,126 lecturas pareadas y después de la limpieza y el *trimming* se conservaron 251,702,144 lecturas (se eliminaron menos del 0.6% del total de lecturas iniciales) (Tabla 9).

Las 5 librerías tuvieron un tamaño similar (65,682,782 – 49,501,142 lecturas) con excepción de la librería 3 (T_0h) donde se generaron 20,761,154 lecturas siendo considerablemente menor al resto. Pese a ello, se decidió utilizarla ya que en todos los análisis se normalizó con el tamaño de cada librería. El tamaño promedio de las lecturas de cada librería estuvo entre 117 y 121 nt con máximos de 127 y mínimos de 20 (establecido durante el *trimming*) (Tabla 9).

Tabla 9. Resumen de las lecturas y del *trimming*.

# Librería	Metagenoma	# lecturas totales	# lecturas limpias	Lecturas limpias (%)	Tamaño promedio (nt)	Max (nt)	Min (nt)
1	Aguamiel	49,501,142	49,260,394	99.51	121	127	20
2	T_0h	20,761,154	20,675,772	99.59	120	127	20
3	T_3h	62,411,428	62,007,472	99.35	117	127	20
4	T_6h	54,816,620	54,532,706	99.48	120	127	20
5	Pulque	65,682,782	65,243,800	99.33	118	127	20
-	Total	253,173,126	251,720,144	99.43	119	127	20

9.3.2 Determinación de la diversidad y composición microbiana

Elección del anotador taxonómico

Las lecturas de cada metagenoma (librería o muestra) se anotaron con distintos programas (Metaphlan2, Centrifuge, KAIJU, KRAKEN y KRAKEN2); para elegir la mejor anotación se consideraron los siguientes criterios:

1. La capacidad de anotar a nivel de género y especie.
2. La capacidad de anotar hongos.
3. Que el archivo de resultados permitiera la extracción de información de manera sencilla.

4. Que existiera documentación clara de cada programa

KRAKEN2 fue el mejor anotador taxonómico evaluado, ya que anota de buena forma a nivel de género y especie, la BD incluye hongos y es de genomas completos, es fácil extraer la información y clasifica de manera rápida (Tabla 10).

Tabla 10. Comparación de los anotadores taxonómicos.

Condición	Metahplan2	Centrifuge	KAIJU	KRAKEN	KRAKEN2
Base de datos	Default	Construida	Construida	Construida	Construida
Referencia	Marcadores no ribosomales	Genomas completos	Genomas completos (proteínas)	Genomas completos	Genomas completos
Capacidad de anotar a nivel de género y especie	Malo	Bueno	Muy bueno	Muy bueno	Excelente
Microorganismos que clasifica	Archaeas, bacterias, hongos y virus.	Archaeas, bacterias, virus y humano.	Archaeas, bacterias, hongos, protozoarios y virus.	Archaeas, bacterias y virus.	Archaeas, bacterias, hongos, protozoarios y virus.
Extracción sencilla de la información	Excelente	Malo (resultado en el nivel taxonómico más bajo que clasifica)	Regular	Muy bueno	Muy bueno
Velocidad para clasificar	Muy rápido	Rápido	Muy rápido	Rápido	Rápido
Resultados con información adicional	No	Si (p.e. tamaño del genoma).	Si	No	No
Notas	Grafica los resultados. Clasificó muy bien microorganismos modelo.	Clasificó > 90% de las lecturas en cada muestra.	Clasifica con proteínas. La construcción de la BD es muy tardada.	Actualizado con KRAKEN2 No clasifica hongos	Flexible para construir la BD

Para evaluar de manera más cuantitativa la capacidad de anotar de KRAKEN2 se estimó la eficiencia y precisión usando los 5 metagenomas (se emplearon las ecuaciones de la Figura 8).

La eficiencia de asignación de KRAKEN2 estuvo entre 0.3248 para el metagenoma del AgM y 0.4851 para el del T_3h. Los otros 3 metagenomas estuvieron por encima de 0.45 (Figura 14.). Esto significa que en 4 de los 5 metagenomas se lograron asignar casi la mitad de las lecturas de cada librería, al menos a nivel de dominio. En el caso del metagenoma de AgM, es probable que algunos de los microorganismos presentes en esta muestra no estén reportados en la RefSeq y, por lo tanto, podrían ensamblarse genomas no reportados a partir de esta muestra. Otra posibilidad es la presencia de organismos que no estuvieron considerados en la BD. Por ejemplo, no se consideraron los genomas de insectos o del humano como posible fuente de contaminación. Los insectos pudieron haber estado en contacto con el AgM antes de la recolección, y, por el otro, durante la extracción de AgM, donde se emplea el acocote y se succiona con la boca, se habría podido contaminar con células humanas. Sin embargo, cuando se utilizó el anotador “*Centrifuge*”, cuya BD contiene secuencias humanas, se observó una baja contaminación por genoma humano (<5%), descartando que haya afectado en la eficiencia de asignación de KRAKEN2 en el AgM. En el caso de los insectos no se analizó si había o no secuencias en nuestros metagenomas.

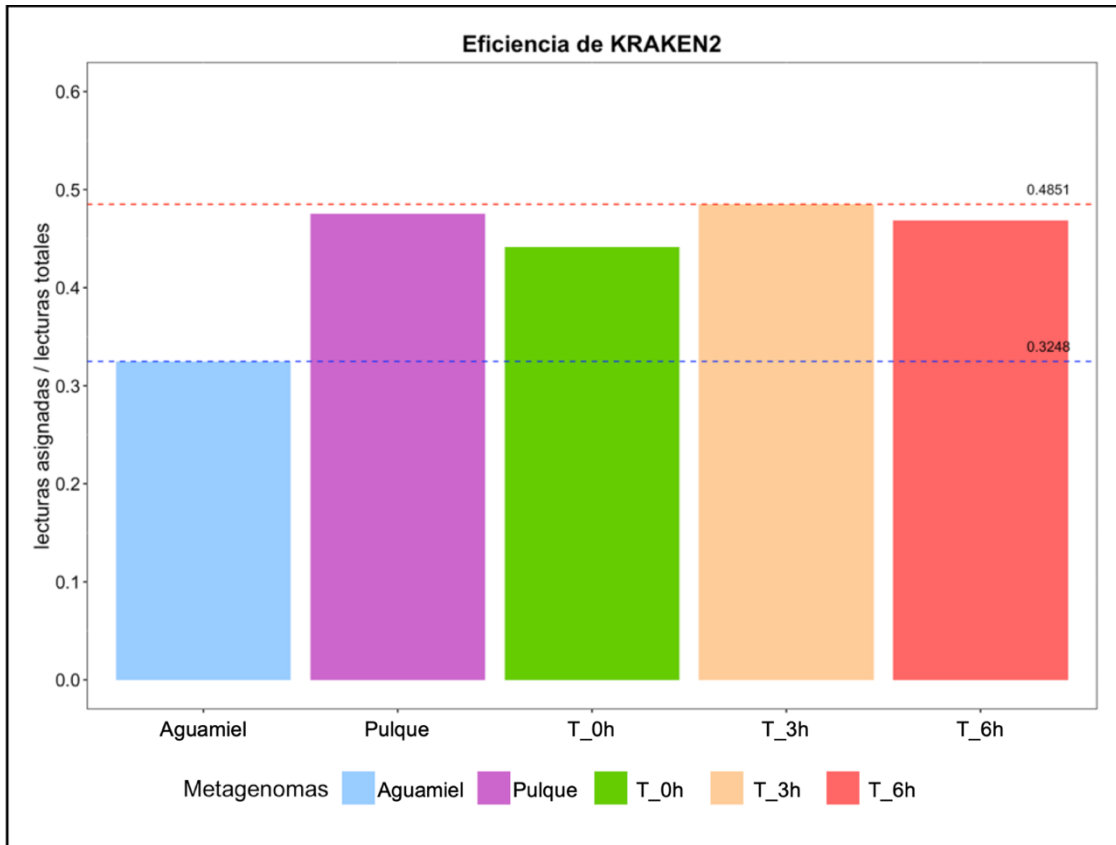


Figura 14. Eficiencia de KRAKEN2 para los 5 metagenomas. El metagenoma con la mayor eficiencia de asignación fue el T_3h (0.48) y el de menor eficiencia el AgM (0.32). Colores: Aguamiel – azul; Pulque – morado T_0h – verde; T_3h – naranja; y, T_6h – rojo. Línea punteada roja eficiencia más alta (0.4861) y línea punteada azul la eficiencia más baja (0.3248).

Una vez que sabíamos la eficiencia de asignación, se estimó la precisión de asignación en 6 niveles taxonómicos (filo, clase, orden, familia, género y especie). En los 6 niveles taxonómicos analizados, la precisión con la cual asignó KRAKEN2 fue excelente, ya que se obtuvo un valor mayor a 0.9 en cada uno de ellos. Esto significa que, en promedio, más del 90% de las lecturas de cada metagenoma que se asignaron a nivel de dominio, también se asignaron en los demás niveles taxonómicos (Figura 15.). Esto nos permitió analizar la composición microbiana hasta nivel de especie.

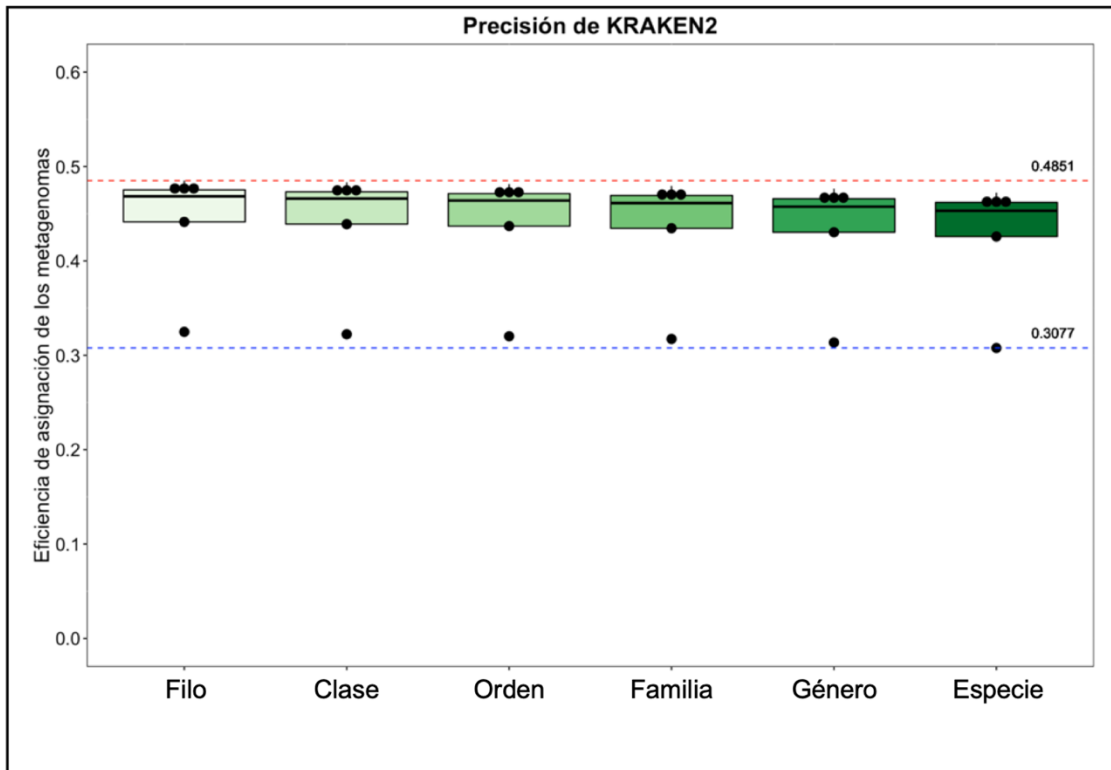


Figura 15. Precisión de KRAKEN2 para 6 niveles taxonómicos. Gráfica de cajas y bigotes de la precisión de KRAKEN2 entre los 5 metagenomas para cada nivel taxonómico. Los bigotes representan la proporción de asignación máxima y mínima; y las cajas el cuartil 3 (Q3), la mediana y el cuartil 1 (Q1). Color degradado: filo a especie – verde claro a verde oscuro. Eficiencia máxima - línea punteada roja (0.4851) y valor mínimo – línea punteada azul (0.3077). La precisión fue mayor al 90% en todos los niveles taxonómicos.

Diversidad microbiana

La diversidad de especies es una propiedad de las comunidades ecológicas, la cual se cree que es el principal factor descriptivo que determina la estructura, funcionalidad y dinámica de las comunidades ¹¹¹. Existen distintas maneras para analizar la diversidad microbiana, en los últimos años la más utilizada se basa en utilizar el gen ribosomal y estimar diversos índices ^{111–113}.

En nuestro caso, se decidió estimar la diversidad microbiana enfocada a la riqueza y similitud (*Observed*, *Chao1*, *Shannon* y *Simpson*) por dos razones: la primera como un

factor descriptivo y la segunda porque estos resultados se tomaron en cuenta, más adelante, en la elección del tipo de ensamble.

Con *Observed* y *Chao1*, se estimó la riqueza de OTUs únicos y ajustados, respectivamente. Con *Shannon* y *Simpson* se estimó la riqueza de OTUs similares y la probabilidad de que dos OTUs tomados al azar representen al mismo OTU, respectivamente. Se utilizaron únicamente los taxones asignados a nivel de especie.

La riqueza (*Observed* y *Chao1*) es similar en los 5 metagenomas (Tabla 11). El metagenoma del AgM fue el de mayor riqueza (*Observed* = 4285 y *Chao1* = 4658.02). El metagenoma T_0h tuvo los valores más dispersos y menores de riqueza (*Observed* = 3784 y *Chao1* = 4160.51); esto es de esperarse, ya que es la librería más pequeña. Los demás metagenomas tuvieron valores cercanos al del AgM.

Se calculó el exponente del índice de Shannon ($e^{Shannon}$) para poder comparar entre los metagenomas. El índice de *Shannon* ($e^{Shannon}$) y el de *Simpson* muestran que el metagenoma del AgM tiene la mayor cantidad de OTUs dominantes ~100 y está dominado por 1 OTU. Este OTU fue identificado como *Zymomonas mobilis* (abundancia relativa 11.4% en AgM). Los demás metagenomas tuvieron índices *Shannon* ($e^{Shannon}$) entre 14.26 y 29.69, y de *Simpson* entre 0.76 y 0.86, sugiriendo que hay menos especies dominantes en la fermentación y en el PQ respecto al AgM, pero sigue habiendo 1 especie que domina, la cual volvió a ser identificada como *Zymomonas mobilis*.

Tabla 11. Índices de diversidad alfa. Para la estimación de los índices se usaron los taxones anotados con KRAKEN2 a nivel de especie, cada taxón único fue considerado como un OTU.

Índice	Aguamiel	Pulque	T_0h	T_3h	T_6h	Media	D.E.
<i>Observed</i>	4285	4224	3784	4185	4222	4140	202.22
<i>Chao1</i>	4658.02	4521.98	4160.51	4419.41	4553.01	4462.58	189.09
<i>Shannon</i>	4.61	2.76	3.39	2.66	3.07	NA	NA
EXP- <i>Shannon</i> *	100.56	15.87	29.69	14.26	21.54	36.38	36.38
<i>Simpson</i>	0.96	0.77	0.86	0.76	0.76	0.82	0.09

*Se calculó el exponente del índice de Shannon ($e^{Shannon}$) para hacer comparables los resultados.

No se estimó la diversidad beta porque este considera la abundancia relativa de los OTUs. La anotación que se realizó fue con genomas completos, por lo cual, el número de lecturas asignadas a cada OTU está afectada por el tamaño del genoma al cual fue asignada. Por ejemplo, se esperaría que el número de lecturas asignadas y la abundancia relativa de una levadura fuera mayor que de una bacteria, porque su genoma es más grande.

Composición microbiana a nivel de género

La composición genómica microbiana (CGM) se definió como la sumatoria del número de lecturas asignadas a un taxón dividido por el total de lecturas de la librería (Figura 8. – Ec. 1) y se calculó la CGM para los siguientes niveles taxonómicos (dominio, filo, clase, orden, familia, género y especie).

Las bacterias y los eucariotas fueron los dominios más abundantes en los 5 metagenomas, con > 99.84 % de las lecturas asignadas (Tabla 12). Se decidió no explorar las archaeas debido a su baja proporción; sin embargo, a pesar de ser el grupo menos abundante, si se analizaron los virus debido a que sus genomas son de menor tamaño y podrían tener importancia en el control de la población microbiana. Los virus se analizaron por separado de las bacterias y eucariotas.

Tabla 12. Composición genómica microbiana; nivel taxonómico: dominio.

Dominio	Aguamiel	Pulque	T_0h	T_3h	T_6h
Archaea	0.07	0.04	0.04	0.04	0.04
Bacteria	99.16	78.61	80.84	71.09	90.27
Eucariota	0.69	21.32	19.08	28.86	9.65
Virus	0.08	0.03	0.03	0.02	0.03
Bacteria + Eucariota	99.84	99.93	99.92	99.94	99.92

Para analizar la abundancia relativa de bacterias y eucariotas, se incluyeron aquellos taxones $\geq 0.5\%$ en al menos 1 muestra y se decidió analizar los resultados en 3 niveles taxonómicos: clase, género y especie. Se identificaron 4 clases dominantes de bacterias

(Bacilli, Clostridia, Alfaproteobacteria y Gammaproteobacteria) y 1 de eucariotas (Saccharomycetes), que representan entre el 96.4% y 98.2% de las lecturas asignadas. Además, se identificaron 14 géneros, que pertenecieron a 4 de las 5 clases más abundantes (Bacilli, Alfaproteobacteria, Gammaproteobacteria y Saccharomycetes) y en conjunto contenían entre el 82.6% y el 91.8% de lecturas asignadas. Finalmente, se obtuvieron 38 especies distribuidas en 9 de los 14 géneros y en 4 de las 5 clases (Bacilli, Alfaproteobacteria, Gammaproteobacteria y Saccharomycetes) y todas las especies se compartieron en las 5 muestras. Varias de las especies clasificadas pertenecen a BAL, BAA y levaduras que han sido identificadas previamente en AgM y PQ ^{5,13-15,17,76}.

El AgM está compuesto principalmente por géneros de la clase Bacilli – 41.3% (*Leuconostoc* – 21.7%, *Lactococcus* – 9.9%, *Streptococcus* – 3.4%, *Lactobacillus* – 1.4%, *Bacillus* – 0.6%, *Enterococcus* – 0.6% y *Staphylococcus* – 0.5%), Gammaproteobacteria – 40.3% (*Acinetobacter* – 26.7%, *Hafnia* – 3.5%, *Rahnella* – 1.3% y *Obesumbacterium* – 0.7%), Alfaproteobacteria 13.8% (*Zymomonas* – 11.4% y *Acetobacter* – 1.1%) y Saccharomycetes – 0.5% (*Saccharomyces* – 0.1%).

Por su parte, el PQ tuvo una composición distinta. La clase más abundante fue la Alfaproteobacteria 40.3% (*Zymomonas* – 35.8% y *Acetobacter* – 0.8%), seguida por los Bacilli – 33.8% (*Leuconostoc* – 15.9%, *Lactobacillus* – 6.0%, *Lactococcus* – 5.8%, *Streptococcus* – 2.3%, *Enterococcus* – 0.4%, *Bacillus* – 0.4% y *Staphylococcus* – 0.4%), las Gammaproteobacteria – 13.8% (*Acinetobacter* 2.5%, *Hafnia* – 0.7%, *Rahnella* – 0.2% y *Obesumbacterium* – 0.1%), y los Saccharomycetes – 21.2% (*Saccharomyces* – 20.3%).

El cambio más notorio entre el AgM y el PQ y que es consistente durante la fermentación, fue el enriquecimiento de los géneros *Saccharomyces* (240 veces), *Zymomonas* (3.15 veces) y *Lactobacillus* (4.2 veces), junto con la disminución de todos los géneros de la clase Gammaproteobacteria en el PQ y al final de la fermentación (T_6h) (Figura 16).

De acuerdo con otros estudios (Figura 5), los géneros *Acinetobacter*, *Lactococcus* y *Leuconostoc* son abundantes en AgM. Anteriormente no se había a *Zymomonas* en el

PQ; estas diferencias pueden deberse a que las muestras son de distinto origen geográfico o por usar secuenciación masiva. *Z. mobilis* se encuentra en otras regiones geográficas y algunos genomas se han aislado de otras zonas. El encontrar a *Z. mobilis* pudo deberse a que la metagenómica no depende de un solo gen, permitiendo una mejor clasificación e identificación taxonómica.

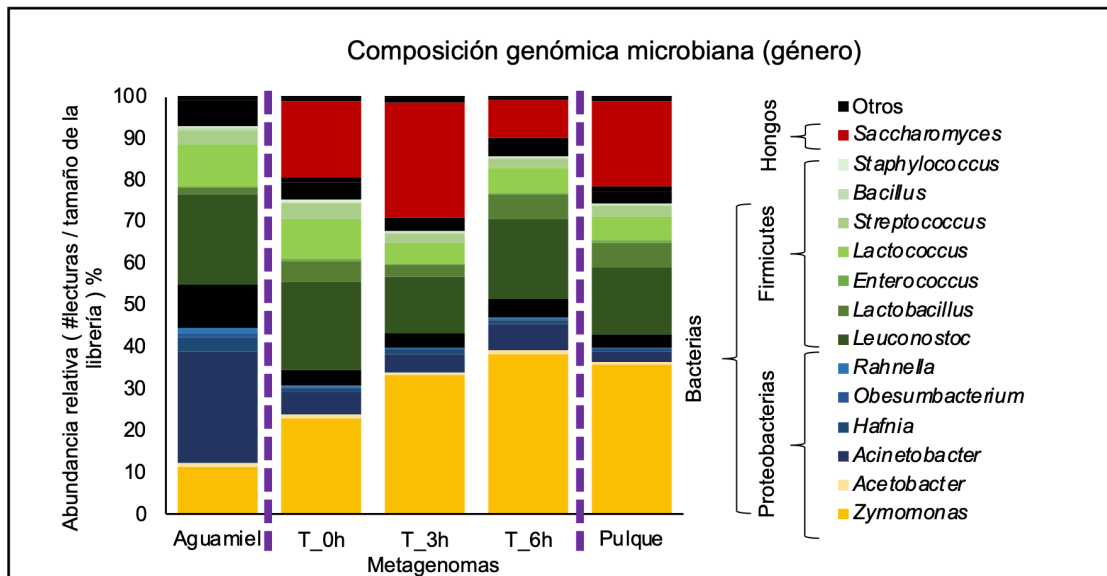


Figura 16. Composición genómica microbiana de bacterias y hongos. Gráfica de barras apiladas de la abundancia relativa de los bacterias y hongos a nivel de clase y género. Los géneros mostrados tienen una abundancia relativa $\geq 0.5\%$ en al menos uno de los metagenomas y representan a 4 clases (Alfaproteobacteria, Gammaproteobacteria, Bacilli y Saccharomycetes). Aquellos géneros cuya abundancia relativa fue $< 0.5\%$ en todas las muestras se agruparon como otros. Color: Otros – negro; géneros de la clase Saccharomycetes – tonos de rojo; géneros de la clase Bacilli – tonos de verde; géneros de la clase Gammaproteobacteria – tonos de azul; géneros de la clase Alfaproteobacteria – tonos de amarillo. Orden (abajo hacia arriba): Alfaproteobacteria, Gammaproteobacteria, Otras Proteobacterias, Bacilli, Otros Firmicutes, Otras bacterias, Saccharomycetes, Otros Ascomycota, Otros.

Composición microbiana usando genes ribosomales

Una manera de corroborar la anotación taxonómica es mediante la asignación de las lecturas a genes ribosomales³⁹. Por ello, se decidió recuperar (identificar u OTU *picking*) posibles genes ribosomales de bacterias (ADNr 16S) y hongos (ITS) con la BD de SILVA v121 (curada) y clasificarlos con la BD de RDP con un número de *bostraps* = 100. La abundancia relativa se definió como (#hits = # genes ribosomales; #hits / total de hits de cada metagenoma) y se graficaron los géneros con una abundancia relativa $\geq 0.5\%$ en al menos una muestra (Figura 17.).

Se clasificaron 6 géneros bacterianos: *Streptococcus*, *Lactobacillus*, *Lactococcus*; *Leuconostoc*, *Acinetobacter* y *Zymomonas*; y, 4 géneros de hongos: *Kluyveromyces*, *Naumovozyma*, *Saccharomyces* y *Torulaspota*. De estos, *Streptococcus*, *Lactobacillus*, *Lactococcus* y *Leuconostoc* pertenecen a los Bacilli y son la clase más abundante identificada en las 5 muestras. *Acinetobacter* y *Zymomnas* fueron los únicos géneros de las clases γ -PB y α -PB, respectivamente.

Al igual que cuando se asignó con genomas completos (KRARKEN2), el AgM tiene una composición mayoritaria de géneros pertenecientes a los Bacilli, seguidos por las Gammaproteobacterias, las Alfabroteobacterias y por último los Saccharomycetes. Estos dos últimos se incrementan en el PQ y durante la fermentación. Los 10 géneros identificados por genes ribosomales también fueron identificados cuando las lecturas se asignaron con los genomas completos sin restringir la abundancia. *Kluyveromyces*, *Torulaspota* y *Naumovozyma* habían sido identificados como poco abundantes ($< 0.5\%$ en todas las muestras). Otra diferencia muy evidente es la dominancia de BAL ($> 60\%$ abundancia relativa) y una baja cantidad de levaduras ($< 10\%$ abundancia relativa) en todas las muestras.

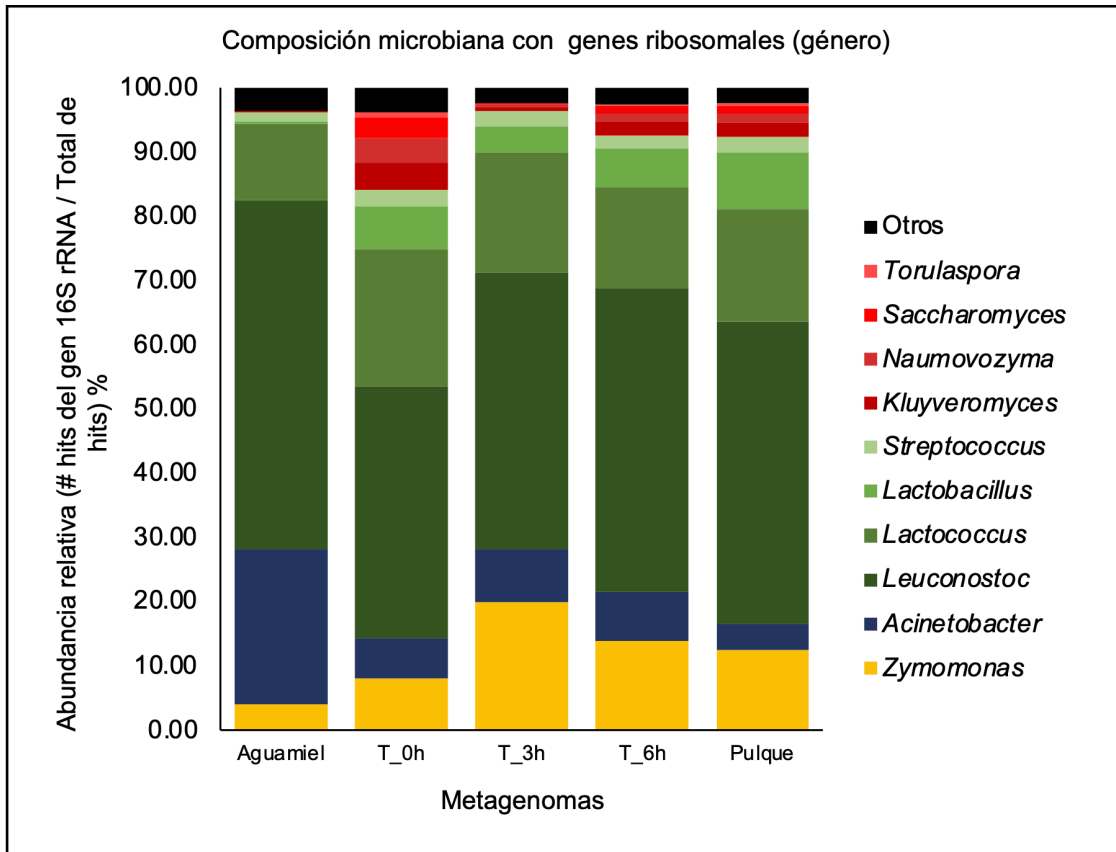


Figura 17. Composición microbiana de bacterias y hongos por clasificación de genes ribosomales. Las lecturas fueron alineadas con la BD de SILVA v121 usando SortMeRNA y un porcentaje de identidad del 97%. Los genes ribosomales fueron clasificados con la BD de RDP y un número de *bootstrap* = 100. Se muestran los géneros con una abundancia relativa de hits $\geq 0.5\%$. Color nivel taxonómico: Otros – Negro; géneros de la clase Saccharomycetes – tonos de rojo; géneros de la clase Bacilli – tonos de verde; géneros de la clase Gammaproteobacteria – tonos de azul; géneros de la clase Alfacaproteobacteria – tonos de amarillo. Orden (abajo hacia arriba): Alfacaproteobacteria, Gammaproteobacteria, Bacilli, Saccharomycetes, Otros.

El análisis de la composición microbiana recuperando genes ribosomales refleja un comportamiento similar al observado cuando se usan genomas completos. Este análisis es más estricto y restringido a un solo gen; y por eso no se pudieron identificar todos los géneros predichos cuando se usó una referencia con genomas completos. Este análisis

permitió ratificar la presencia de 10 de los 14 géneros (71.4%), es probable que éstos 10 géneros sean claves en la estructura funcional y dinámica de la fermentación de PQ. En la Tabla 13 se comparan los microorganismos, nivel de género, identificados en este trabajo con trabajos previos.

Composición microbiana a nivel de especie

Como se había mencionado con anterioridad; la precisión de KRAKEN2 (>90% a nivel de especie) permitió analizar la composición microbiana hasta nivel de especie. Para este análisis se incluyeron aquellas especies que tuvieran una abundancia relativa > 1% en al menos una de las muestras (Figura 18).

Tabla13. Microorganismos reportados previamente por el grupo y en este trabajo.

Clasificación	Aislamiento / Banco de clonas 16S rRNA ^{13,14}	Este estudio (genomas completos / recuperación genes ribosomales) (abs. rel.>=0.5%)
Hongos	<i>Candida</i> sp. <i>Kluyveromyces</i> sp. <i>Saccharomyces</i> sp.	<i>Kluyveromyces</i> sp. <i>Naumovozyma</i> sp. <i>Saccharomyces</i> sp. <i>Torulosporella</i> sp.
Bacteria	<i>Acetobacter</i> sp. <i>Acinetobacter</i> sp. <i>Bacillus</i> sp. <i>Enterobacter</i> sp. <i>Enterococcus</i> sp. <i>Gluconobacter</i> sp. <i>Lactobacillus</i> sp. <i>Leuconostoc</i> sp. <i>Pediococcus</i> sp. <i>Streptococcus</i> sp. <i>Zymomonas</i> sp.	<i>Acetobacter</i> sp. <i>Acinetobacter</i> sp. <i>Bacillus</i> sp. <i>Enterococcus</i> sp. <i>Hafnia</i> sp. <i>Lactobacillus</i> sp. <i>Lactococcus</i> sp. <i>Leuconostoc</i> sp. <i>Obesumbacterium</i> sp.* <i>Rahnella</i> sp.* <i>Staphylococcus</i> sp. <i>Streptococcus</i> sp. <i>Zymomonas</i> sp.

* Microorganismos no descritos en los trabajos previos

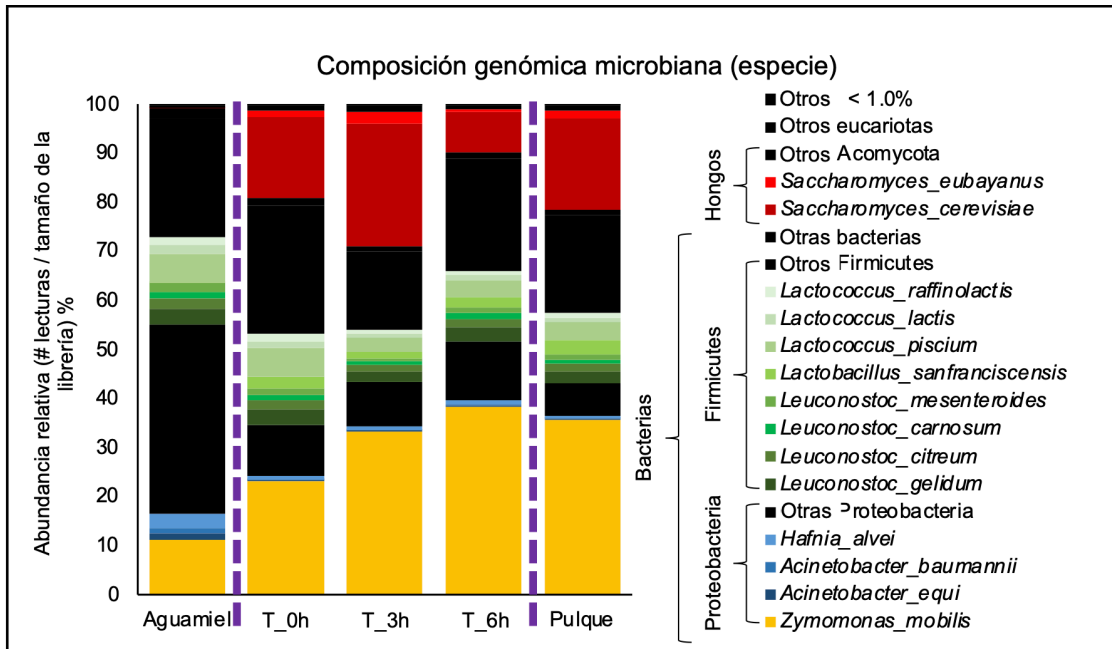


Figura 18. Composición genómica microbiana de bacterias y hongos a nivel de especie. Gráfica de barras apiladas de la abundancia relativa de los bacterias y hongos a nivel de clase y especie. Los géneros mostrados tienen una abundancia relativa $\geq 1.0\%$ en al menos uno de los metagenomas y representan a 4 clases (Alfaproteobacteria, Gammaproteobacteria, Bacilli y Saccharomycetes). Aquellas especies cuya abundancia relativa fue $< 1\%$ en todas las muestras se agruparon al final de cada clase. Color: Otros – negro; especies de la clase Saccharomycetes – tonos de rojo; especies de la clase Bacilli – tonos de verde; especies de la clase Gammaproteobacteria – tonos de azul; especies de la clase Alfaproteobacteria – tonos de amarillo. Orden (abajo hacia arriba): Alfaproteobacteria, Gammaproteobacteria, Otras Proteobacterias, Bacilli, Otros Firmicutes, Otras bacterias, Saccharomycetes, Otros Ascomycota, Otros.

Las BAL aportan la mayor cantidad de especies y géneros al AgM, PQ y durante la fermentación; es probable que cada una de ellas este teniendo un papel característico debido a su diversidad genómica⁷³⁻⁷⁵. Por otro lado, *Zymomonas mobilis* es una bacteria microaerofílica y la única especie del género *Zymomonas*. Esta bacteria fue aislada por primera vez del pulque y se tienen pocos genomas secuenciados completos hasta la

fecha (16), a pesar de su alto potencial en la producción de bioetanol ^{2,12}. La presencia y participación de esta bacteria en la fermentación de pulque no es clara, aunque todo indicaría que está involucrada en la síntesis de etanol y polisacáridos. Estudios de diversidad microbiana en pulque y aguamiel mostraron que su abundancia es baja con respecto a levaduras etanológicas clásicas (*S. cerevisiae*) y de bacterias productoras de EPS (*L. mesenteroides*) ¹⁵. Es probable que su representatividad haya sido subvalorada por ser una bacteria de difícil aislamiento y poco representada en bases de datos. En este trabajo estamos reportando por primera vez que *Zymomonas mobilis* es la especie más abundante en aguamiel, durante la fermentación y en el pulque.

Estos resultados sugieren que, a pesar de que *Z. mobilis* y *S. cerevisiae* son la bacteria y la levadura más abundante en las muestras, las BAL son las que tienen la mayor cantidad de especies y, probablemente, la mayor diversidad de funciones.

El viroma

Los virus son microorganismos que dirigen la evolución de otros microorganismos y promueven la transferencia horizontal de material genético; además, se cree que la mayoría de los virus son fagos que infectan a bacterias. El viroma, que se define como la colección de material genético de virus en un ecosistema, ha sido estudiado principalmente en ambientes marinos en donde han encontrado que cambios genéticos en los genomas de los virus explica de mejor manera la diferencia entre las regiones geográficas que la presencia o ausencia de los virus. Dicho en otras palabras, los virus son universales y los cambios genéticos se deben a factores ambientales los cuales ejercen presión de selección ^{114,115}. Aunque, en alimentos fermentados se ha encontrado que la diversidad de virus explica de mejor manera el origen geográfico del kimchi que la diversidad bacteriana ¹¹⁶.

En este marco, se decidió analizar la abundancia de los virus en las 5 muestras secuenciadas. La abundancia relativa de los virus no superó el 0.08% en ninguna de las muestras; el metagenoma del AgM fue el que tuvo la mayor proporción de virus (0.08 %) y los demás metagenomas estuvieron entre 0.3% (PQ, T_0h y T_6h) y 0.2% (T_3h)

(Tabla 12). La composición de virus se analizó a nivel de especie con la finalidad de identificar de manera más clara a los miembros de esta comunidad y se incluyeron los más abundantes (27 especies ≥ 10 lecturas en al menos una muestra) (Figura 19).

De los 780 virus identificados, se muestran las 27 especies más abundantes, la mayoría de ellos fagos de bacterias de la clase Bacilli (96.3%) con excepción del *Cellulophaga* phage phi39:1 el cual se encuentra en bacterias del género *Cellulophaga*. Este fago ha sido asociado con el viroma de océanos y la biósfera rara ¹¹⁷. La presencia de fagos de los océanos en nuestras muestras sugiere que éstos podrían tener diferentes nichos ecológicos. Los fagos de bacterias de la clase Bacilli pertenecen a los géneros *Enterococcus* (n=2), *Lactococcus* (n=9), *Leuconostoc* (n=7), *Staphylococcus* (n=2) y *Lactobacillus* (n=6). Una diferencia interesante que se encontró es que los fagos del género *Lactococcus* (0.0193%), que eran los más abundantes en AgM y los del género *Leuconostoc* (0.0074%), que eran los terceros más abundantes en AgM, pasaron a ser los segundos (0.0044%) y cuartos (0.0013%) más abundantes en PQ y al final de la fermentación, respectivamente, mientras que, *Enterococcus* y *Lactobacillus* pasaron del segundo (0.0178%) y cuarto (0.0009%) lugar en AgM, al primero (0.0050%) y tercero (0.0029%) en PQ y al final de la fermentación, respectivamente. Estos resultados coinciden con el incremento en la población bacteriana de *Lactobacillus* y *Enterococcus* y la disminución de *Leuconostoc* y *Lactococcus* del AgM al PQ y durante la fermentación (Figura 16).

la única plataforma que permitió hacer el *binning* híbrido y la curación manual de los *bins* 19,22,91

Tabla 14. Comparación de los métodos para ensamblar genomas.

Consideraciones	Método probados para obtener genomas a partir de datos metagenómicos		
	<i>Read-mapping</i> (dirigido)	Metagenoma (estándar)	<i>Binning</i> (híbrido)
1. Asignación taxonómica lecturas	Indispensable	No indispensable	Indispensable
2. Ensamble	Por taxón	Por muestra	Muestras juntas
3. Método de agrupación de los <i>contigs</i>	No aplica	Supervisado BD tax. BD func. %GC SGC	No supervisado Cobertura %GC Supervisado BD tax. BD func. %GC SGC
4. Desventaja	Lecturas poco complejas no se asignan. Las lecturas son forzadas a asignarse.	Genomas repetidos entre las muestras. Genomas muy fragmentados.	Pocos manuales y se requiere de varias muestras similares (composición y diversidad).
Método probado	Si	Si	Si
Método elegido	No	No	Si

Elección del tipo de ensamble y ensamblador

El siguiente paso consistió en elegir entre un meta-ensamble o un coensamble. Para ello, se realizaron 3 ensambles utilizando 3 distintos ensambladores. Para el meta-ensamble se juntaron las 5 muestras en 2 archivos (R1 y R2) y el ensamble se realizó con IDBA-UD y con metaSPADES. Para el coensamble, cada muestra se emplea como una réplica y se usó MEGAHIT. Los encabezados de los *contigs* obtenidos de cada ensamble fueron editados y se eliminaron aquellos *contigs* con un tamaño < 1000 nt de longitud y se compararon utilizando QUAST (Tabla 15 y Figura 20).

El mejor ensamble fue obtenido con metaSPADES. El número de *contigs* y el %GC no variaron entre los 3 ensambles; la longitud total del ensamble fue similar entre MEGAHIT

y metaSPADES pero con IDBA-UD fue mucho menor; el *contig* más largo fue obtenido con metaSPADES y tuvo un tamaño de 287,616 nt y con los otros dos ensambladores no superó los 20,000 nt. Cuando se comparó la N50 (la longitud del *contig* con el cual se alcanza el 50% de la longitud del ensamble cuando los *contigs* son ordenados de mayor a menor tamaño) y la L50 (la posición que ocupa el *contig* de la N50) el ensamble de metaSPADES fue claramente superior a los otros 2; al igual que cuando se analizó la distribución de los *contigs* por tamaño, donde metaSPADES fue el único que generó *contigs* > 25,000 y 50,000 nt (Figura 20).

Tabla 15. Comparación de los ensambles. Se utilizó QUAST para comparar los resultados del ensamble y *contigs* \geq 1000 nucleótidos.

Ensamble	# <i>contigs</i>	Longitud total (nt)	<i>Contig</i> más largo	GC (%)	N50	L50
IDA-UD	19,398	32,479,456	8,280	43.3	1,694	6,636
Megahit*	27,325	77,212,334	14,742	44.0	1,542	9,662
metaSPADES	22,089	97,073,248	287,616	42.9	8,567	1,974

Durante el ensamble no se permitió la incorporación de Ns.
*Coensamble

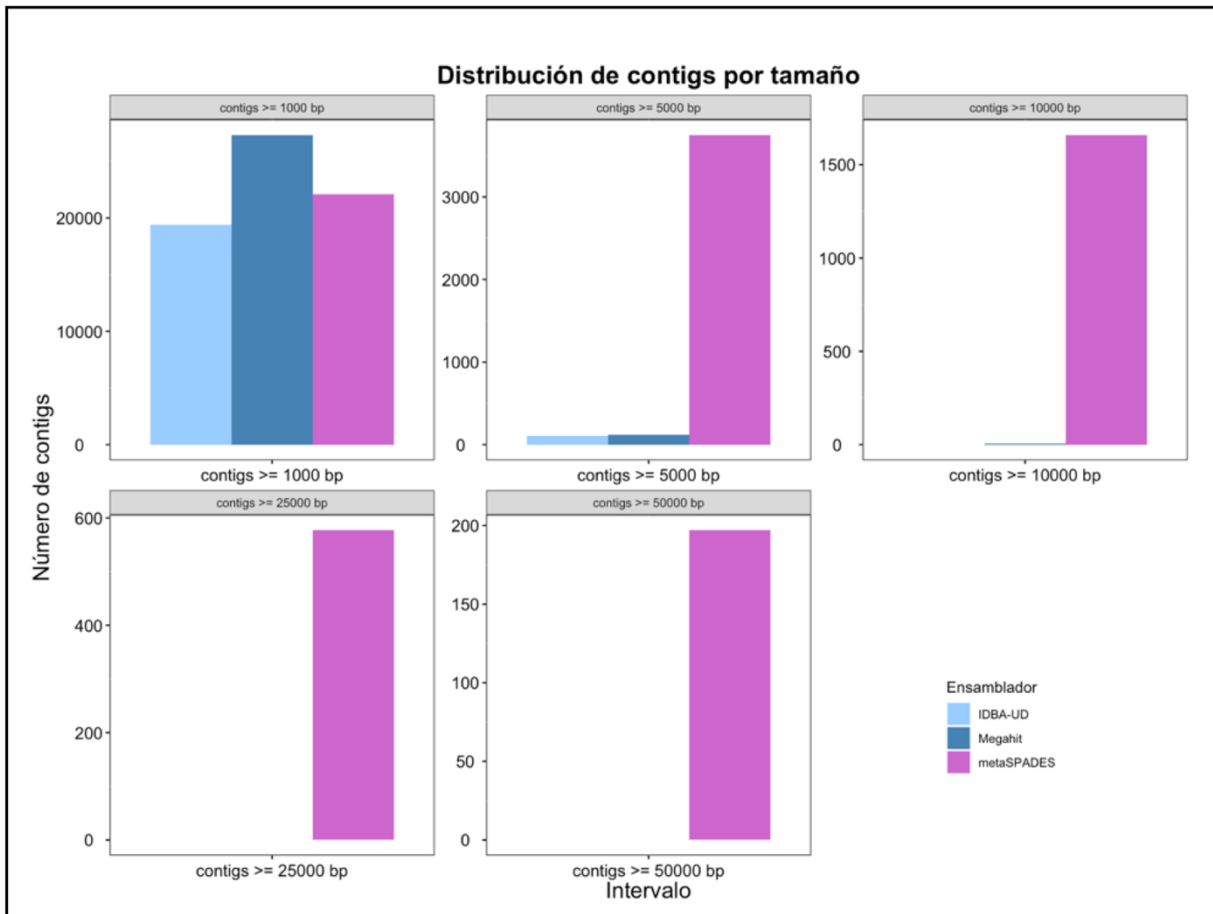


Figura 20. Gráfica de frecuencias del número de *contigs* por intervalos de tamaño. Colores: IDBA-UD – azul cielo; Megahit – azul; y, metaSPADES – morado.

En ocasiones, la asignación taxonómica con lecturas o con *contigs* es distinta. Para descartar que fuera nuestro caso, los *contigs* se asignaron con KRAKEN2 y con los mismos parámetros y la misma BD empleada para las lecturas. Se seleccionaron los 30 géneros más abundantes de los 5 metagenomas (asignados con las lecturas), obteniendo 55 géneros únicos. Estos 55 géneros fueron buscados en la anotación taxonómica del meta-ensamble y se encontraron 53 de los 55 géneros (96.4%); los géneros *Kazachstania* y *Zygosaccharomyces* fueron los 2 géneros ausentes en el meta-ensamble. Es probable que sus genomas tengan una alta similitud con otras levaduras o que no hayan podido ser ensambladas correctamente. Aun así, el resultado muestra

una muy buena correlación entre la clasificación taxonómica usando lecturas o ensamblajes para con los géneros más abundantes.

Mapeo

El *binning* híbrido requiere información de la cobertura y el uso de codones, para ello cada metagenoma de ser mapeado contra el meta-ensamble. Se utilizó Bowtie2 y Samtools para este paso; se espera que el porcentaje de lecturas mapeadas esté alrededor del 90% y las lecturas mapeadas correctamente sean más del 70% (Tabla 16). Estos valores van a variar dependiendo del número de *contigs* eliminados por su tamaño, por ejemplo, el ensamble generado con metaSPADES produjo un total de 105,458 *contigs* de los cuales se eliminaron 83,369 debido a que tenían un tamaño < 1000 nt.

Tabla 16: Resumen del mapeo de las lecturas de cada metagenoma con el meta-ensamble.

Mapeo	Aguamiel	Pulque	T_0h	T_3h	T_6h
Lecturas totales (#)	49,260,394	65,243,800	20,675,772	62,007,472	54,532,706
Lecturas mapeadas (#)	43,368,909	62,317,303	19,716,570	59,286,520	51,435,577
Lecturas mapeadas correctamente (#)	34,099,140	46,128,054	15,180,674	40,849,892	40,666,300
Lecturas mapeadas (%)	88.04	95.51	95.36	95.61	94.32
Lecturas mapeadas correctamente (%)	69.22	70.7	73.42	65.87	74.57

Generación de la CONTIGS.DB y de los MAGs

Se generó la CONTIGS.DB de acuerdo con el programa Anvi'o y se predijeron los genes hipotéticos con Prodigal. Además, utilizando HMMER3, se anotaron los genes hipotéticos con las 4 BD de genes de copia única. Con Blast+ (E-value < 1e⁻⁵), se anotaron con la BD de genes ortólogos COGs de NCBI. Luego, los genes hipotéticos fueron extraídos de la CONTIGS.DB y traducidos a proteínas para ser anotados con InterProScan con la BD de InterPro y con GhostKoala de KEGG en su versión gratuita en línea. Por último, se utilizó KAIJU para anotar taxonómicamente los genes hipotéticos con la BD no redundante de proteínas de NCBI. Las anotaciones se incorporaron a la CONTIGS.DB. Utilizando los resultados del mapeo y la CONTIGS.DB anotada, se generaron los perfiles para cada metagenoma empleando una herramienta de Anvi'o. Como se comentó en la metodología, durante esta etapa se generan muchos datos estadísticos de los *contigs* y

genes y luego se unen los perfiles para hacer el *binning*. Una vez que se tienen los *bins*, hay que revisar cada uno y manualmente limpiarlos, considerando el porcentaje de genes de copia única y el tamaño total de los *contigs* ($\geq 90\%$ ó >2 Mega bases por *bin*) y la redundancia de los genes de copia única ($<10\%$ por *bin*) (Tabla 17). Los *bins* curados ya son MAGs.

Es importante destacar que con los genes de copia única uno puede identificar genomas de una misma especie o de otra, ya que son marcadores filogenéticos ¹¹⁸.

Tabla 17. Resumen de la obtención de MAGs. Se utilizó el ensamble producido por *metaSPADES*, para la predicción de genes se utilizó *prodigal*, los COGs se anotaron con *blastp*, los KOs se anotaron con *GhostKoala* (en línea), la anotación con Interpro se hizo con *InterProScan*, para el *binning* se utilizó *CONCOCT* y el refinamiento se hizo en Anvi'o.

<i>Contigs</i>	# genes hipotéticos	# COGs	# KOs	# firmas InterPro*	# <i>bins</i>	# MAGs*
22,089	99,730	66,869	24,336	509,823	59	23

COGs= cluster de genes ortólogos
 KO= ortólogos de KEGG

* Interpro incluye 14 bases de datos para proteínas

**Algunos MAGs no requirieron limpieza

Se obtuvieron 23 MAGs curados de los cuales 7 se lograron asignar a nivel de género, 14 a nivel de especie y 2 no fueron identificados ni en especie ni en género (Tabla 18).

Tabla 18. Resumen de los MAGs. Se obtuvieron 23 MAGs (> 90% GCU o > 2 Mb y < 10% redundancia GCU), de los cuales 7 se resolvieron a nivel de género (1 – 7), 14 a nivel de especie (8 – 21) y 2 no pudieron resolverse a nivel de género (ND – 22 y 23).

#	Asignación taxonómica	Nivel taxonómico	Longitud			GCU		Redundancia GCU (%)
			(Mbp)	N50	# contigs	GC (%)	(%)	
1	<i>Acetobacter</i> sp.	género	2.45	4,989	638	52.8	97.1	2.9
2	<i>Acinetobacter</i> sp.	género	2.27	28,379	162	37.4	99.3	0.0
3	<i>Lactobacillus</i> sp.	género	2.62	43,500	123	40.0	97.1	0.7
4	<i>Lactobacillus</i> sp.	género	1.56	21,002	127	43.3	99.3	1.4
5	<i>Leuconostoc</i> sp.	género	1.36	39,406	64	37.1	98.6	0.7
6	<i>Rahnella</i> sp.	género	3.30	59,158	104	51.7	53.2	2.9
7	<i>Saccharomyces</i> sp.	género	11.36	63,159	318	38.4	85.5	7.2
8	<i>Lactobacillus sanfranciscensis</i>	especie	1.53	67,568	47	36.3	98.6	0.0
9	<i>Lactobacillus otakiensis</i>	especie	2.29	18,296	171	42.2	85.6	2.9
10	<i>Lactobacillus pentosiphilus</i>	especie	2.66	18,796	212	43.3	99.3	4.3
11	<i>Lactococcus plantarum</i>	especie	1.61	26,142	108	35.8	95.7	0.7
12	<i>Lactococcus plantarum</i>	especie	1.80	33,099	93	35.9	74.1	1.4
13	<i>Lactococcus lactis</i>	especie	1.92	4,893	515	35.6	92.1	5.8
14	<i>Leuconostoc mesenteroides</i>	especie	1.94	6,312	370	36.8	97.1	2.9
15	<i>Gluconobacter roseus</i>	especie	2.55	2,276	1,245	60.3	74.1	6.5
16	<i>Hafnia alvei</i>	especie	4.48	38,589	181	48.6	99.3	5.0
17	<i>Kluyvera intermedia</i>	especie	3.76	10,722	431	52.4	72.7	6.5
18	<i>Kluyveromyces marxianus</i>	especie	7.20	3,782	2,144	40.0	66.3	2.4
19	<i>Zymomonas mobilis</i>	especie	2.08	20,587	170	46.2	97.8	6.5
20	<i>Zymomonas mobilis</i>	especie	1.69	13,428	187	46.6	97.8	0.7
21	<i>Zymomonas mobilis</i>	especie	1.84	9,117	283	45.1	59.0	0.7
22	ND	ND	1.70	3,708	599	34.5	86.3	8.6
23	ND	ND	4.36	2,073	2,246	52.3	54.7	7.9

GCU= genes de copia única

ND= no determinado a nivel de género

De los MAGs que lograron ser asignados taxonómicamente (n=21), se obtuvieron 2 levaduras, una a nivel de género (*Saccharomyces* sp.) y otra a nivel de especie

(*Kluyveromyces marxianus*). Ambas levaduras han sido reportadas como microorganismos clave en la fermentación de pulque y de otros alimentos fermentados 5,7,9,13–15,17. Las bacterias de la clase Bacilli fueron las que aportaron la mayor cantidad de MAGs (n=10), los cuales pertenecen a *Leuconostoc mesenteroides*, *Leuconostoc sp.*, *Lactobacillus otakiensis*, *Lactobacillus sanfranciscensis*, *Lactobacillus pentosiphilus*, *Lactobacillus sp.* (n=2), *Lactococcus lactis*, *Lactococcus plantarum* (n=2). *Lactococcus plantarum* no había sido reportado en la fermentación de pulque. Además, se logró obtener MAGs de las γ -PB: *Hafnia alvei*, *Rahnella sp.*, *Acinetobacter sp.* y *Kluyvera intermedia*. Dentro de las α -PB se obtuvieron 5 MAGs: uno de *Acetobacter sp.*, uno de *Gluconobacter roseus* y 3 de *Zymomonas mobilis*, este último es la especie más abundante en todas nuestras muestras, por lo que no es un resultado extraño.

La Figura 21 muestra los distintos microorganismos, a nivel de género, que han sido identificados en el AgM, PQ y en la fermentación empleando distintos métodos. Se utilizaron todos los géneros reportados en 3 trabajos y los identificados en este estudio; los géneros únicos se agruparon en 5 categorías y se seleccionaron aquellos que hubieran sido identificados en al menos 2 de ellas. Se obtuvieron 73 géneros que cumplían con nuestras condiciones; la mayoría fueron identificados por secuenciación masiva tipo *shotgun* y asignadas con la BD de genomas completos (n=69) seguida de recuperación de genes ribosomales de la misma secuenciación (recuperación de genes ribosomales) (n=57). Destacaron 6 géneros que han sido identificados en todos los estudios, independientemente del método empleado: *Acinetobacter*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Kluyveromyces* y *Saccharomyces*. Estos géneros podrían ser considerados el *core* microbiano de la bebida (Figura 21a). Es probable que existan varias subpoblaciones de *Z. mobilis* en nuestras muestras porque se ensamblaron 3 genomas y fue la especie más abundante en todas nuestras muestras. La presencia de subpoblaciones podría explicar su alta abundancia en condiciones distintas, AgM y PQ.

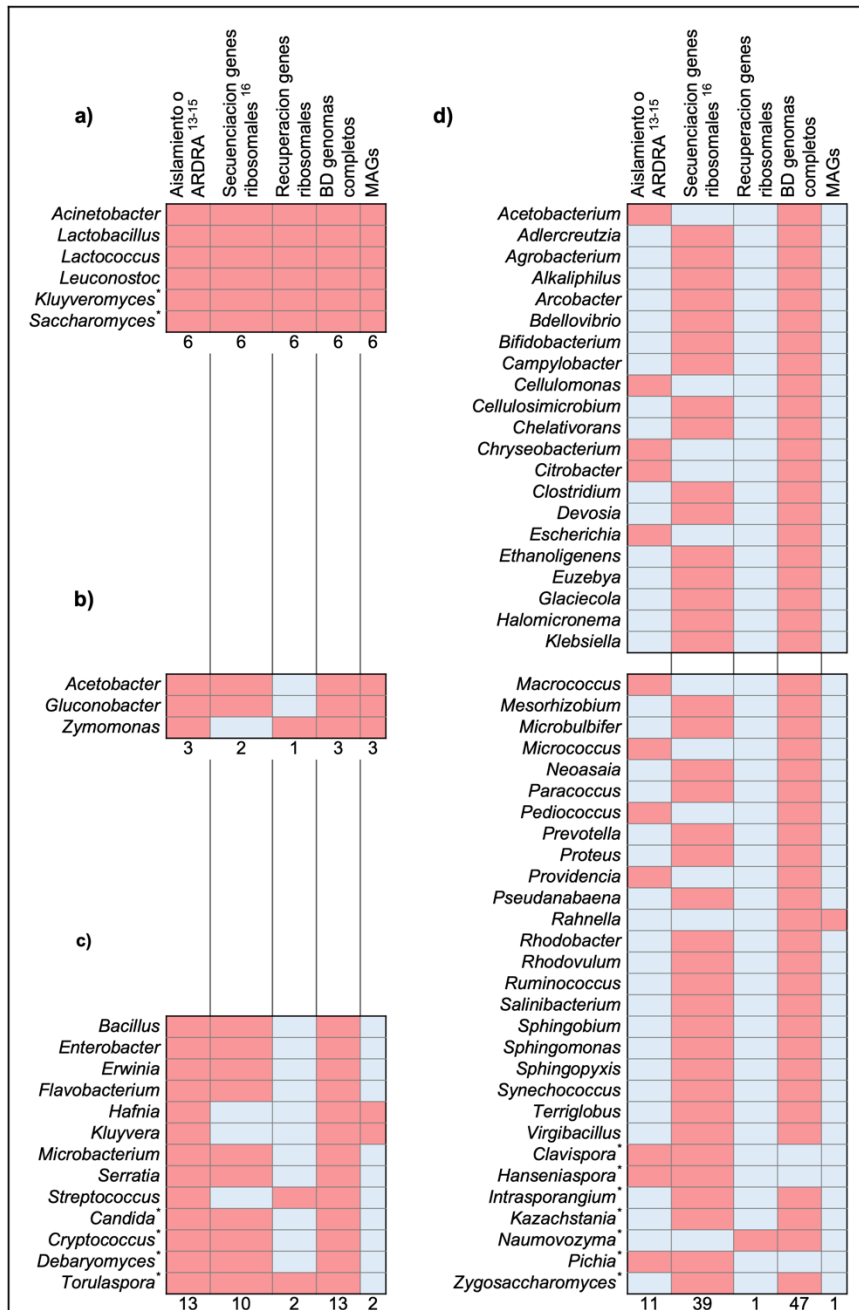


Figura 21. Microorganismos del aguamiel, pulque y de la fermentación identificados por distintos métodos. Se consideraron sólo las bacterias y hongos que se identificaron con al menos 2 de 5 métodos distintos : aislamiento ó ARDRA (Aculco, Edo. de México; Tizayuca, Hidalgo; y Huitzilac, Morelos), secuenciación de genes ribosomales (Tepeapulco, Hidalgo), recuperación de genes ribosomales de metagenomas (Huitzilac, Morelos), por identificación con BD de genomas completos (Huitzilac, Morelos), obtención de MAGs (Huitzilac, Morelos). a) microorganismos identificados por los 5 métodos ; b) microorganismos identificados con 4 métodos; y c) microorganismos identificados con 3 de los 4 métodos; y d) microorganismos identificados por 2 métodos. Color: presencia – rojo; y, ausencia – azul. * Indica hongos. Los números debajo de los mapas de calor indican el # de microorganismos aportados por cada método.

Para conocer el potencial funcional que podrían tener estos 23 MAGs; se comparó el número de genes hipotéticos y de funciones únicas de los MAGs con respecto a los del meta-ensamble (Figura 22.). El número total de genes hipotéticos para el meta-ensamble fue de 99,730 (100%) y el de los 23 MAGs de 64,548 (64.72%). Para identificar las funciones únicas, éstas se definieron como las combinaciones únicas de los anotadores funcionales para cada gen hipotético. Por ejemplo, la proteína-A que está anotada de la

siguiente manera (KO-0001-COG-001- PS00301) es distinta a la proteína B (KO-0001-COG-001) y tienen “funciones” distintas. El número total de funciones únicas en el meta-ensamble fue de 30,693 (100%) y 28,884 (94.10%) para los MAGs. Esto significa que, a pesar de tener el ~65% de los genes hipotéticos, los MAGs representan ~95% de la capacidad funcional de la fermentación de pulque.

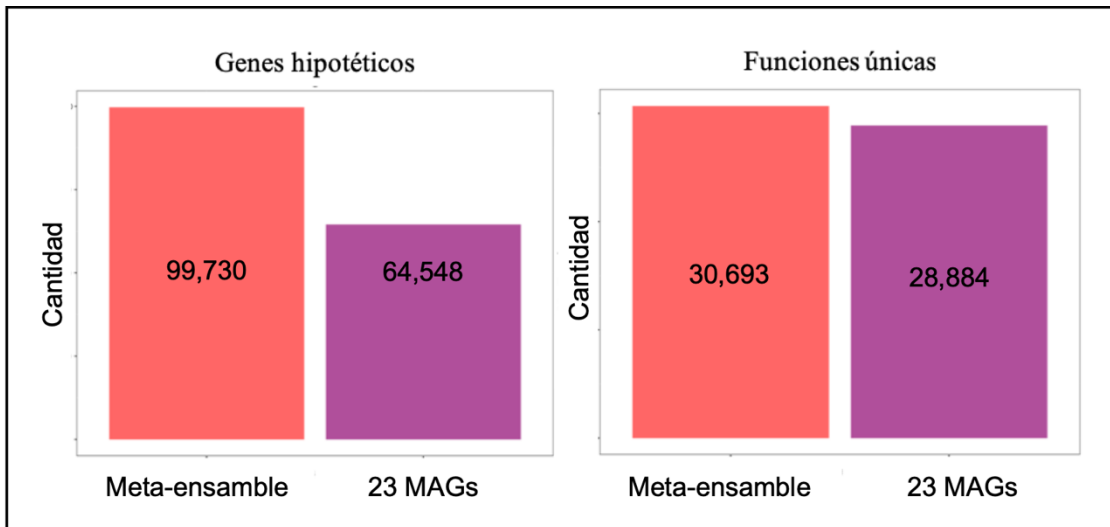


Figura 22. Representatividad funcional de los MAGs en el meta-ensamble. Izquierda: Genes hipotéticos totales en el meta-ensamble y en los 23 MAGs. Derecha: Funciones únicas en el meta-ensamble y en los MAGs. Colores: meta-ensamble – rojo y MAGs – morado.

9.3.4 Análisis funcional (pangenómico) de los MAGs

Una manera de comparar la diversidad de funciones es mediante el análisis de *clusters* de todos los genes dentro de varios genomas (pangenoma). Este tipo de análisis consiste en agrupar los genes en función de su similitud a nivel de secuencia (blastp), seguido de la separación en *clusters* (MCL) y, finalmente, realizando un alineamiento múltiple de los genes de cada *cluster* (MUSCLE). Se utilizaron los 23 MAGs para hacer el análisis pangenómico y de manera ilustrativa se muestra en la Figura 23.

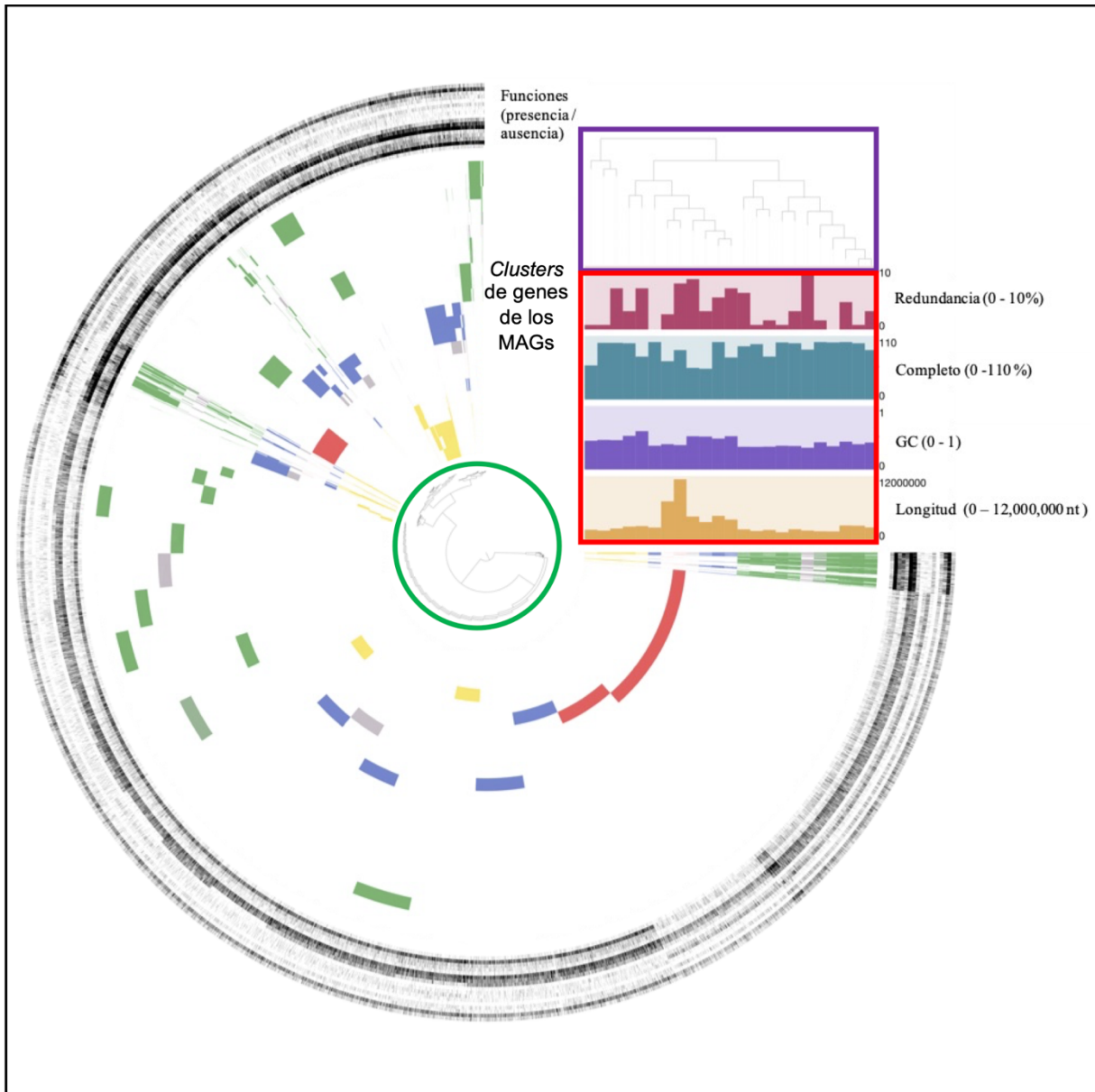


Figura 23. Pangenoma de los 23 MAGs. Recuadro morado: histograma de la frecuencia de genes en los *cluster* de genes los MAGs. Recuadro rojo: características de los MAGs (redundancia de genes de copia única, fracción de genes de copia única en cada MAG, porcentaje de GC, longitud del genoma). Círculo verde: histograma de presencia o ausencia de genes en los *clusters* de genes. *Clusters* de genes de los MAGs; Bacilli: verde; Gammaproteobacterias - azul; Saccharomycetes: rojo; Alphaproteobacterias - amarillos; Desconocidos - gris.

Se formaron 3 grandes *clusters* de funciones (circulograma en orden de las manecillas del reloj); el primero muestra funciones representadas en todos los grupos bacterianos; el segundo muestra funciones únicas entre los MAGs, el tercero son funciones compartidas entre algunos MAGs.

Por otro lado, se formaron 8 clusters taxonómicos (Figura 23 – recuadro morado): i) las 3 *Zymomonas mobilis* (α -PB); ii) *Acetobacter* sp. (α -PB) y *Gluconobacter roseus* (α -PB); iii) *Acinetobacter* sp. (γ -PB); iv) *Saccharomyces* sp. y *Kluyveromyces marxianus*; v) *Rahnella* (γ -PB), *Kluyvera intermedia* (γ -PB), *Hafnia alvei* (γ -PB) y Desconocido_2; vi) 2 *Lactococcus plantarum* (Bacilli) y *Lactococcus lactis* (Bacilli); vii) *Leuconostoc mesenteroides* (Bacilli) y *Leuconostoc* sp. (Bacilli); y, viii) Desconocido_1, *Lactobacillus* sp_1. (Bacilli), *Lactobacillus sanfranciscensis* (Bacilli), *Lactobacillus pentosiphilus* (Bacilli), *Lactobacillus* sp_2. y *Lactobacillus otakiensis* (Bacilli). Es probable que uno de los MAGs desconocidos pertenezca al género *Lactobacillus* y que el otro sea una γ -PB, sin embargo, se requieren otros análisis para corroborarlo.

Este análisis soporta la propuesta de que existen al menos 4 grupos de microorganismos con actividades metabólicas muy características; sin embargo, no se pudieron identificar las funciones metabólicas características de cada uno de los microorganismos.

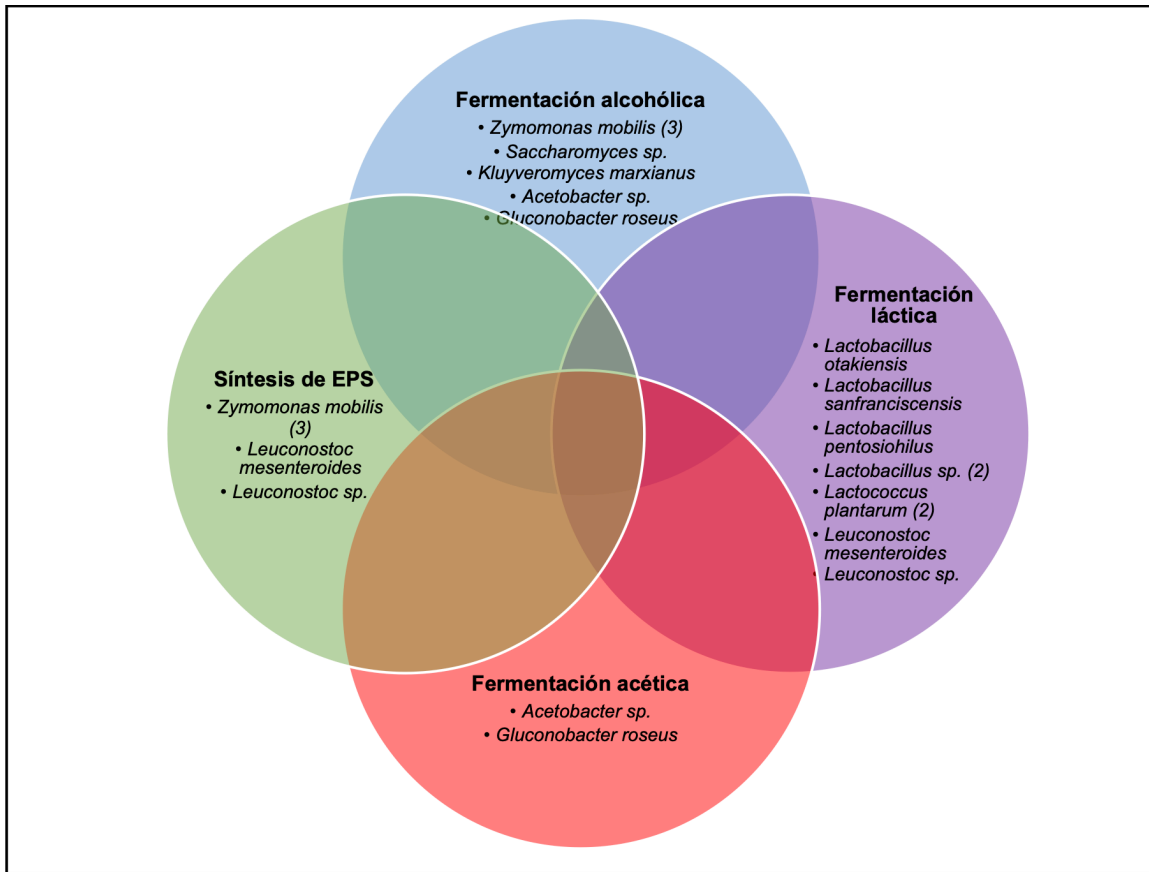


Figura 24. Distribución de los MAGs en los 4 principales procesos metabólicos, propuestos por Escalante et al. (2016). EPS= polisacáridos extracelulares. Entre paréntesis es el número de genomas ensamblados, cuando es mayor a 1.

9.3.5 Pangenoma de *Zymomonas mobilis*

Z. mobilis es la especie más abundante en las 5 muestras que se analizaron, están reportados pocos genomas (n=16), es de importancia biotecnológica y se obtuvieron 3 MAGs, por lo cual se decidió hacer un análisis pangenómico de la especie. Se descargaron 16 genomas de NCBI y se descartaron 2 que no estaban reportados en la RefSeq. Después, se generó una CONTIGS.DB y se anotó cada uno de ellos con las mismas BD y condiciones que los MAGs. El análisis pangenómico se realizó con los

mismos parámetros que en el pangenoma de los MAGs, el único parámetro que se modificó fue el valor de MCL (valor = 1).

Este análisis permitió identificar los genes que pertenecen al *core* de *Z. mobilis* el cual está compuesto por ~2000 genes (Figura 25). La redundancia de los genomas fue mínima, excepto por uno de los MAGs (CB_Zymomonas_mobilis_3 ~8% redundancia). CB_Zymomonas_mobilis_1 fue el MAG más pequeño de los 3. El %GC fue muy similar al de los genomas de referencia. La longitud total de los MAGs fue ligeramente menor que las referencias. CB_Zymomonas_mobilis_1 y CB_Zymomonas_mobilis_2 se separaron de los demás genomas, mientras que, CB_Zymomonas_mobilis_3 se agrupó entre los genomas ASM57616v1v1 y ASM305204v1. Es importante recalcar que no realizamos un análisis filogenómico y el agrupamiento se basa en la presencia/ausencia de genes. CB_Zymomonas_mobilis_1 y CB_Zymomonas_mobilis_2 tienen *clusters* faltantes que parecieran complementarse.

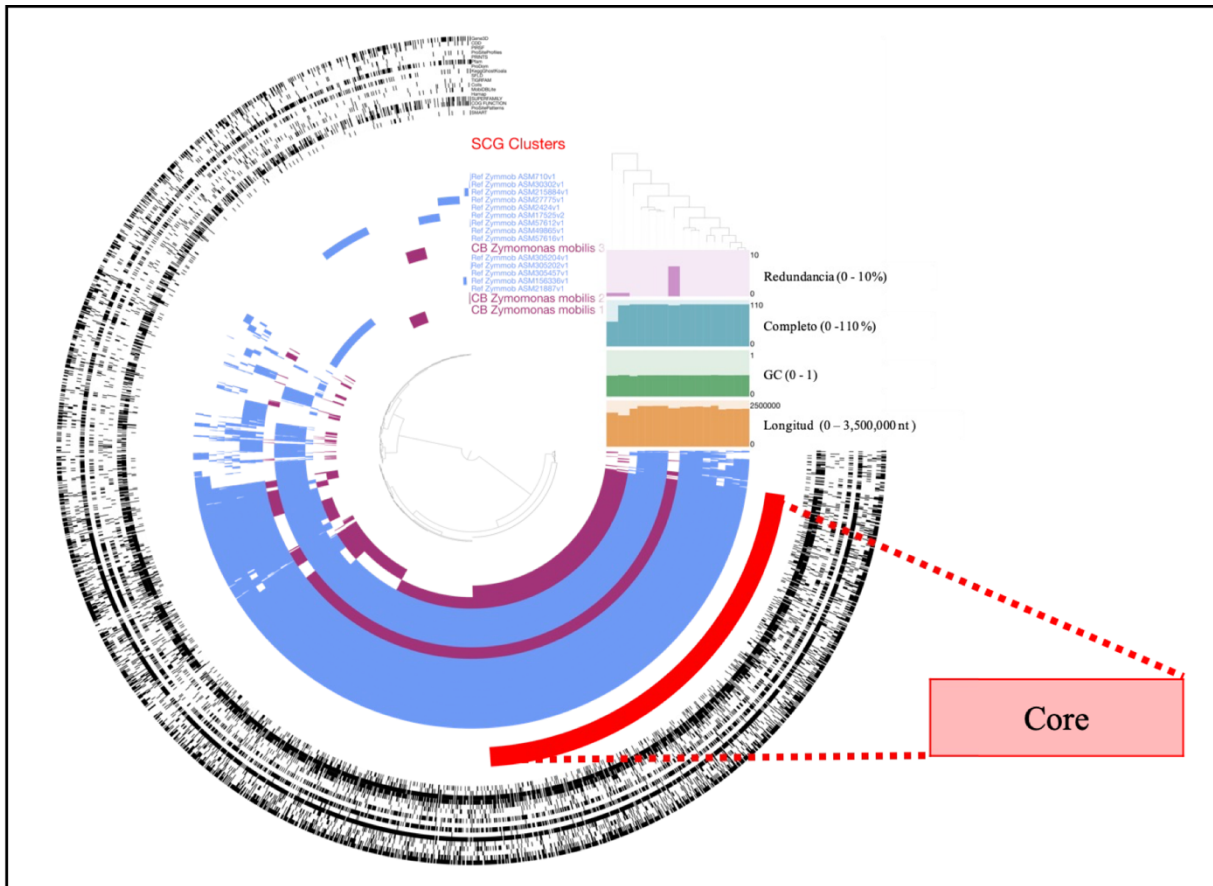


Figura 25. Pangenoma de *Zymomonas mobilis*. Se usaron los 3 MAGs y 14 genomas de referencias (RefSeq). Cluster de genes de copia única presente en los 17 genomas – línea roja; genomas de referencia – azul; y MAGs – morado.

9.4 Microscopía

La distribución y organización espacial de los microorganismos puede tener un impacto importante desde el punto de vista metabólico y de interacción entre ellos y el ambiente. En ratones gnotobióticos, se ha estudiado la organización espacial en el intestino a macroescala (centenas de micrómetros) donde se observan comunidades heterogéneas; mientras que a microescala (decenas de micrómetros), se observan microhábitats homogéneos similares a lo observado en biorreactores ^{23,119}.

En este contexto y en colaboración con la Dra. Jessica Mark Welch y la Dra. Shamayim Tabita Ramírez Puebla del MBL de la Universidad de Chicago, se analizaron muestras de AgM y PQ empleando el método de CLASI-FISH. Se eligieron 5 sondas que ya se habían utilizado, pero se decidió analizar las muestras sólo con 2 de manera exploratoria (Tabla 5). Las muestras se fijaron en el IBT y se enviaron a EUA para la hibridación y microscopía. La muestra de AgM fijada con PFA 2% y centrifugada no se analizó por problemas técnicos.

En AgM, independientemente del tratamiento, se observa una menor densidad de microorganismos representada en su mayoría por bacterias (verde) y los hongos fueron casi imperceptibles, en comparación con PQ, pero no se ven cambios entre las muestras de AgM o entre las del PQ (Figura 26). Tampoco se distinguen diferencias en la densidad de microorganismos entre las muestras centrifugadas y no centrifugadas.

Los microorganismos no se cuantificaron, sin embargo, a pesar del incremento en la cantidad de microorganismos, pareciera que la proporción bacteria-hongo se mantiene entre el AgM y el PQ. Interesantemente, se observan distintas formas bacterianas entre el AgM y el PQ; por ejemplo, en el AgM la mayoría de las bacterias forman racimos y en el PQ hay diplococos, cadenas de bacilos y bacilos.

La distribución de los microorganismos pareciera no se aleatoria. En las 3 muestras de AgM, las bacterias forman cúmulos dentro de los cuales se aprecian hongos, lo cual sugiere una interacción entre estos grupos. Los hongos podrían estar siendo protegidos de las altas concentraciones de azúcar en el AgM por las bacterias, y éstas últimas podrían estar inhibiendo el crecimiento de los hongos y alimentándose de ellos. En cambio, en el PQ, las bacterias y los hongos parecieran distribuirse a lo largo de todo el campo. En las muestras de PQ no centrifugadas se aprecia una organización no aleatoria de las bacterias, es decir, pareciera que se agregan en forma de una red, la cual podría deberse a la organización de microorganismos alrededor de los polímeros. En este momento, es difícil saber si esta estructura en forma de red es real en el PQ o un artefacto del protocolo. Una manera de mejorar el análisis microscópico sería fijar las muestras mediante criopreservación (usando nitrógeno líquido) o con alguna resina, ya que

preservarían de mejor manera la disposición espacial. Además, el poder teñir los polisacáridos contribuiría de manera importante a saber si éstos tienen un impacto en la organización microbiana.

En resumen, se observan diferencias en la densidad microbiana, el tipo de morfología bacteriana, y una organización no aleatoria de bacterias y hongos (se agregan) entre el AgM y el PQ. La formación de una estructura microbiana durante la fermentación de PQ no es evidente y se debería abordar de otra manera, sin embargo, parecieran no distribuirse de manera azarosa.

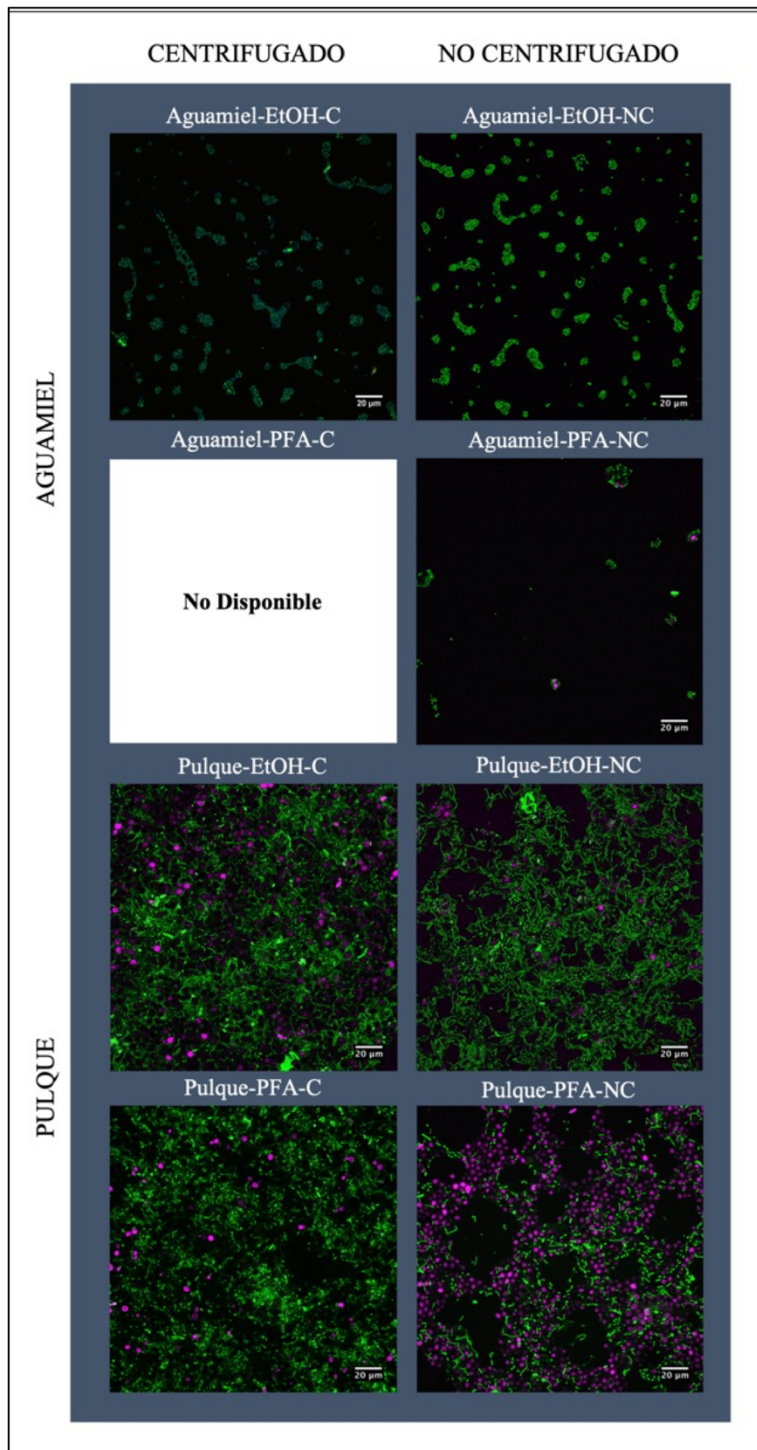


Figura 26. Organización espacial de bacterias y hongos en muestras de aguamiel y pulque utilizando CLASI-FISH. Condiciones: fijación con etanol – EtOH; fijación con paraformaldehído al 2% - PFA; muestra centrifugada – C; y, muestra no centrifugada – NC. Colores: bacteria – verde; ADN – azul; hongos – morado. Las imágenes son representativas de varios campos y de varias réplicas técnicas. NOTA: la muestra Aguamiel-PFA-C no fue analizada.

10. Conclusiones

En este trabajo se definió la composición microbiana en el aguamiel, el pulque y en distintos tiempos de fermentación, se evaluaron propiedades fisicoquímicas

(temperatura, pH, azúcares y ácidos orgánicos), se ensamblaron 23 genomas de microbianos y se observó, por microscopia de fluorescencia, a las levaduras y bacterias en aguamiel y pulque.

La Figura 27 muestra el resumen de cada etapa del trabajo.

La composición microbiana mostró que el aguamiel, el pulque y los 3 tiempos de fermentación comparte la gran mayoría de microorganismos. Se identificaron 38 especies (abundancia relativa. > 0.5% en al menos 1 muestra) que representan 14 géneros y 4 clases de bacterias y hongos cuando se clasificaron las lecturas. Como método de comprobación, se buscaron genes ribosomales (16S) y se clasificaron 7 de los 14 géneros previamente descritos. Además, al analizar el viroma se mostró que está compuesto principalmente por fagos de enterobacterias, BAL y BAA. Dichos fagos mostraron el mismo patrón de abundancias que sus hospederos y podrían ser utilizados como biomarcadores.

La abundancia de microorganismos varía en cada punto muestreado, por ejemplo, en el aguamiel *Acinetobacter* es abundante, pero disminuye a las 6 horas de fermentación. En cambio, *Z. mobilis* y *S. cerevisiae* se enriquecen conforme se desarrolla la bebida. Esto concuerda con los resultados metabólicos, en donde se observó el incremento en la concentración de etanol. La abundancia de las BAL (*Leuconostoc* y *Lactobacillus*) se mantuvo constante en todo momento a pesar del incremento del pH en la bebida. Es probable que se requiera de análisis de transcriptómica para identificar a los productores de ácido láctico.

El ensamble *de novo* permite recuperar secuencias consenso de los microorganismos presentes en una muestra. Se ensamblaron 23 MAGs de bacterias (n=21) y levaduras (n=2), de los cuales, 14 fueron asignados a nivel de especie, 7 a nivel de género y 2 no se asignaron a ninguno de estos dos niveles taxonómicos. Es probable que los genomas no clasificados sean artefactos del proceso de ensamble o representen microorganismos poco conocidos y no representados las bases de datos usadas. La curación manual de

MAGs es un proceso complejo y requiere validación con otros métodos como PCR y secuenciación de los extremos de los *contigs*. Este último paso no estaba contemplado en el trabajo, pero es necesario si se quiere tener más conocimiento de las poblaciones presentes en el pulque.

Comparación de MAGs con genomas de referencia permite validar de manera bioinformática la calidad de los genomas y compararlos con secuencias de referencia. El análisis pangénomico permite comparar todos los genes de un grupo de genomas. Se realizó un pangénoma de *Z. mobilis* con los 3 MAGs obtenidos en este trabajo y con 14 genomas de referencia obtenidos de NCBI, todos los disponibles hasta ese momento. Uno de los MAGs compartía más del 80% de sus genes con genomas previamente secuenciados, los otros 2 MAGs compartían alrededor del 50%. Esta diferencia podría indicar otros genotipos de *Z. mobilis*.

El uso de microscopía permite observar la distribución espacial de los microorganismos. Se observaron distintas agrupaciones microbianas entre el AgM, racimos de bacterias englobando hongos, y el PQ, diplococos, cadenas de bacilos y cadenas de hongos. Este tipo de organización supone un tipo de interacción complejo entre bacterias y levaduras. Por ejemplo, podría existir competencia por el sustrato, dada la rapidez en el consumo de sacarosa. También podría haber mutualismo o depredación, en donde la producción de polisacáridos podría permitir la supervivencia de algunos organismos y/o acaparar nutrientes.

Una gran parte del trabajo consistió en explorar y establecer el flujo de trabajo y los algoritmos para el análisis bioinformático. A la fecha no existen métodos estandarizados complicando la implementación e interpretación de los resultados. Pese a ello, los resultados son contundentes, existe una población muy definida en el pulque. El pulque es un sistema único en donde las poblaciones microbianas se renuevan en un ciclo en cada lote cuando se añade aguamiel y donde forman agregados microbianos. A pesar de ser un abierto (expuesto a la vida rural), no se encontraron patógenos. Es posible que la naturaleza ácida y la competencia por dicho hábitat prevenga la colonización por

microorganismos no deseados. La comunidad microbiana del pulque puede ser un excelente modelo de interacción microbiana y de producción de compuestos de interés biotecnológico como el bioetanol o compuestos nutricionales.

1. Perfil metabólico	2. Composición microbiana	3. Obtención de MAGs	4. Capacidad funcional de los MAGs	5. Microscopía
<ul style="list-style-type: none"> • Obtención de una bebida similar al PQ • Fermentación caracterizada por el incremento del etanol, fructosa y lactato y la disminución de la sacarosa • Disminución en el pH 	<ul style="list-style-type: none"> • Descripción basada en genomas completos, clasificación de genes ribosomales y viroma • Gammaproteobacterias disminuyen su abundancia en la fermentación y PQ. • <i>Z. Mobilis</i> y <i>S. cerevisiae</i> abundantes en la fermentación y PQ • <i>Lactobacillus</i> spp. se enriquecen durante el proceso • Los fagos reflejan la dinámica bacteriana 	<ul style="list-style-type: none"> • 23 genomas ensamblados (2 de levaduras y 21 de bacterias) • 7 clasificados a nivel de género • 14 clasificados a nivel de especie • 2 clasificados sólo como bacterias 	<ul style="list-style-type: none"> • Los 23 MAGs representan ~65% genes hipotéticos y ~95% de las funciones únicas 	<ul style="list-style-type: none"> • Densidad de hongos y bacterias mayor en PQ que en AgM • Organización no azarosa de bacterias y hongos y distinta entre AgM y PQ

Figura 27. Resumen de los resultados más relevantes de cada una de las 5 etapas del trabajo.

11. Perspectivas

- Desarrollar un modelo de fermentación de pulque para estudiar interacciones microbianas. Por ejemplo, seleccionando representantes de los 6-8 géneros *core*.
- Comparar los MAGs con microorganismos de otros alimentos fermentados para estimar que tan únicos o generales son.

- Aislar o secuenciar usando lecturas largas (PacBio) para obtener información de mejor calidad de los microorganismos presentes en la bebida. Esto permitiría comparar genes entre regiones geográficas
- Definir si los MAGs representan a microorganismos cultivables o no cultivables (p.e. vías de síntesis de aminoácidos) e intentar cultivarlos y aislarlos.

12. Referencias

1. Alba-Lois L, Segal-Kischinevsky C. Yeast Fermentation and the Making of Beer and Wine. *Nat Educ.* 2010;3(9):1-6.
2. Baratti JC, Bu'lock JD. *Zymomonas mobilis*: A bacterium for ethanol production. *Biotechnol Adv.* 1986;4(1):95-115. doi:10.1016/0734-9750(86)90006-6
3. Cook PE. Fermented foods as biotechnological resources. *Food Res Int.* 1994;27:309-316.
4. Escobar-Zepeda A, Sanchez-Flores A, Quirasco M. Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota. *Food Microbiol.* 2016;57:116-127. doi:10.1016/j.fm.2016.02.004
5. González-Vázquez R, Azaola-Espinosa A, Mayorga-Reyes L, Reyes-Nava LA, Shah NP, Rivera-Espinoza Y. Isolation, Identification and Partial Characterization of a *Lactobacillus casei* Strain with Bile Salt Hydrolase Activity from Pulque. *Probiotics Antimicrob Proteins.* 2015;7(4):242-248. doi:10.1007/s12602-015-9202-x
6. Lagier JC, Dubourg G, Million M, et al. Culturing the human microbiota and culturomics. *Nat Rev Microbiol.* 2018;16(9):540-550. doi:10.1038/s41579-018-0041-0
7. Rizo J, Guillén D, Farrés A, et al. Omics in traditional vegetable fermented foods and beverages. *Crit Rev Food Sci Nutr.* 2018;0(0):1-19. doi:10.1080/10408398.2018.1551189
8. Romero-Luna HE, Hernández-Sánchez H, Dávila-Ortiz G. Traditional fermented beverages from Mexico as a potential probiotic source. *Ann Microbiol.* 2017;67(9):577-586. doi:10.1007/s13213-017-1290-2
9. Torres-Maravilla E, Lenoir M, Mayorga-Reyes L, et al. Identification of novel anti-inflammatory probiotic strains isolated from pulque. *Appl Microbiol Biotechnol.* 2016;100(1):385-396. doi:10.1007/s00253-015-7049-4

10. Walker G, Stewart G. *Saccharomyces cerevisiae* in the production of fermented beverages. *Beverages*. 2016;2(4):30. doi:10.3390/beverages2040030
11. Wolfe BE, Button JE, Santarelli M, Dutton RJ. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell*. 2014;158(2):422-433. doi:10.1016/j.cell.2014.05.041
12. Yang S, Fei Q, Zhang Y, et al. *Zymomonas mobilis* as a model system for production of biofuels and biochemicals. *Microb Biotechnol*. 2016;9(6):699-717. doi:10.1111/1751-7915.12408
13. Escalante A, Elena Rodríguez M, Martínez A, López-Munguía A, Bolívar F, Gosset G. Characterization of bacterial diversity in Pulque, a traditional Mexican alcoholic fermented beverage, as determined by 16S rDNA analysis. *FEMS Microbiol Lett*. 2004;235(2):273-279. doi:10.1016/j.femsle.2004.04.045
14. Escalante A, Giles-Gómez M, Hernández G, et al. Analysis of bacterial community during the fermentation of pulque, a traditional Mexican alcoholic beverage, using a polyphasic approach. *Int J Food Microbiol*. 2008;124(2):126-134. doi:10.1016/j.ijfoodmicro.2008.03.003
15. Escalante A, López Soto DR, Velázquez Gutiérrez JE, Giles-Gómez M, Bolívar F, López-Munguía A. Pulque, a traditional Mexican alcoholic fermented beverage: Historical, microbiological, and technical aspects. *Front Microbiol*. 2016;7(JUN):1-18. doi:10.3389/fmicb.2016.01026
16. Valadez-Blanco R, Bravo-Villa G, Santos-Sánchez NF, Velasco-Almendarez SI, Montville TJ. The Artisanal Production of Pulque, a Traditional Beverage of the Mexican Highlands. *Probiotics Antimicrob Proteins*. 2012;4(2):140-144. doi:10.1007/s12602-012-9096-9
17. Rocha Arriaga C, Shimada Beltrán H, Cruz Ramírez LF, Alcaráz Peraza LD. *Denominación de Origen y Estudio Metagenómico Del Pulque Del Estado de Hidalgo , México*. Universidad Nacional Autónoma de México; 2017.
18. Park EJ, Kim KH, Abell GCJ, Kim MS, Roh SW, Bae JW. Metagenomic analysis of the viral communities in fermented foods. *Appl Environ Microbiol*. 2011;77(4):1284-1291. doi:10.1128/AEM.01859-10
19. Eren AM, Esen ÖC, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319. doi:10.7717/peerj.1319
20. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome . *PeerJ*. 2018;6:e4320. doi:10.7717/peerj.4320
21. Caputo A, Fournier PE, Raoult D. Genome and pan-genome analysis to classify emerging bacteria. *Biol Direct*. 2019;14(1):1-9. doi:10.1186/s13062-019-0234-0

22. Tamames J, Puente-Sánchez F. SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. *Front Microbiol.* 2019;9(January):1-10. doi:10.3389/fmicb.2018.03349
23. Valm AM, Mark Welch JL, Borisy GG. CLASI-FISH: Principles of combinatorial labeling and spectral imaging. *Syst Appl Microbiol.* 2012;35(8):496-502. doi:10.1016/j.syapm.2012.03.004
24. Mark Welch JL, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. Biogeography of a human oral microbiome at the micron scale. *Proc Natl Acad Sci.* 2016;113(6):E791-E800. doi:10.1073/pnas.1522149113
25. Muller M, Mentel M, van Hellemond JJ, et al. Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. *Microbiol Mol Biol Rev.* 2012;76(2):444-495. doi:10.1128/MMBR.05024-11
26. Research and Markets. Global Fermented Food and Ingredients Market to Reach \$ 888 . 76 Billion by 2023. *PRNewswire.* 2017:1-4.
27. Caplice E, Fitzgerald GF. Food fermentations: role of microorganisms in food production and preservation. *Int J Food Microbiol.* 1999;50:131-149. doi:10.1016/S0168-1605(99)00082-3
28. Vitorino LC, Bessa LA. Technological Microbiology: Development and Applications. *Front Microbiol.* 2017;8:827. doi:10.3389/fmicb.2017.00827
29. Peng Q, Jiang S, Chen J, et al. Unique Microbial Diversity and Metabolic Pathway Features of Fermented Vegetables From Hainan, China. *Front Microbiol.* 2018;9:399. doi:10.3389/fmicb.2018.00399
30. Sieuwerts S, de Bok FAM, Hugenholtz J, van Hylckama Vlieg JET. Unraveling Microbial Interactions in Food Fermentations: from Classical to Genomics Approaches. *Appl Environ Microbiol.* 2008;74(16):4997-5007. doi:10.1128/AEM.00113-08
31. Steinkraus KH. Fermentations in World Food Processing. *Compr Rev Food Sci Food Saf.* 2002;1(1):23-32. doi:10.1111/j.1541-4337.2002.tb00004.x
32. Tamang JP, Watanabe K, Holzapfel WH. Review: Diversity of microorganisms in global fermented foods and beverages. *Front Microbiol.* 2016;7(MAR). doi:10.3389/fmicb.2016.00377
33. Steinkraus KH. Classification of fermented foods: worldwide review of household fermentation techniques. *Fermented Food Saf.* 1997;8(5):311-317. doi:10.1016/S0956-7135(97)00050-9
34. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature.* 2011;473(7346):174-180. doi:10.1038/nature09944

35. Eren AM, Borisy GG, Huse SM, Mark Welch JL. Oligotyping analysis of the human oral microbiome. *Proc Natl Acad Sci*. 2014;111(28):E2875-E2884. doi:10.1073/pnas.1409644111
36. Konopka A. What is microbial community ecology? *ISME J*. 2009;3(11):1223-1230. doi:10.1038/ismej.2009.88
37. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: Genomic Analysis of Microbial Communities. *Annu Rev Genet*. 2004;38(1):525-552. doi:10.1146/annurev.genet.38.072902.091216
38. Franzosa EA, Hsu T, Sirota-Madi A, et al. Sequencing and beyond: Integrating molecular “omics” for microbial community profiling. *Nat Rev Microbiol*. 2015;13(6):360-372. doi:10.1038/nrmicro3451
39. Thompson LR, Sanders JG, McDonald D, et al. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature*. 2017;551(7681):457-463. doi:10.1038/nature24621
40. K. F, J. R. Microbial interactions: From networks to models. *Nat Rev Microbiol*. 2012;10(8):538-550. doi:10.1038/nrmicro2832
41. Parks DH, Chuvochina M, Waite DW, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018;36(10):996-1004. doi:10.1038/nbt.4229
42. Delmont TO, Quince C, Shaiber A, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol*. 2018;3(7):804-813. doi:10.1038/s41564-018-0176-9
43. Wade W. Unculturable bacteria—the uncharacterized organisms that cause oral infections. *J R Soc Med*. 2002;95:81-83.
44. Schloss PD, Handelsman J. Metagenomics for studying unculturable microorganisms: Cutting the Gordian knot. *Genome Biol*. 2005;6(8):6-9. doi:10.1186/gb-2005-6-8-229
45. Duncan SH, Louis P, Flint HJ. Cultivable bacterial diversity from the human colon. *Lett Appl Microbiol*. 2007;44(4):343-350. doi:10.1111/j.1472-765X.2007.02129.x
46. Pollock J, Laura G, Trong W, Mick W. The Madness of Microbiome : Attempting To Find Consensus. *Appl Environ Microbiol*. 2018;84(7):1-12.
47. Mardis ER. Next-Generation DNA Sequencing Methods. *Annu Rev Genomics Hum Genet*. 2008;9(1):387-402. doi:10.1146/annurev.genom.9.081307.164359

48. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front Genet.* 2015;6. doi:10.3389/fgene.2015.00348
49. Ranjan R, Rani A, Metwally A, McGee H, Perkins D. Analysis of the microbiome-Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun.* 469(4):967-977. doi:10.1007/s11065-015-9294-9.Functional
50. Zavarzin GA. Winogradsky and modern microbiology. *Microbiology.* 2006;75(5):501-511. doi:10.1134/s0026261706050018
51. B.L.Bassler MBM. Quorum Sensing in Bacteria. *Annu Rev Microbiol.* 2001;55:165-199.
52. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol.* 1990;56(6):1919-1925.
53. Sanders ME. Probiotics: Definition, Sources, Selection, and Uses. *Clin Infect Dis.* 2008;46(s2):S58-S61. doi:10.1086/523341
54. FAO. *Probiotics in Animal Nutrition - Production, Impact and Regulation* by Yadav S. Bajagai, Athol V. Klieve, Peter J. Dart and Wayne L. Bryden. Editor Harinder P.S. Makkar.; 2016.
55. Tamang JP, Shin DH, Jung SJ, Chae SW. Functional properties of microorganisms in fermented foods. *Front Microbiol.* 2016;7(APR):1-13. doi:10.3389/fmicb.2016.00578
56. Park SC, Park Y, Hahm KS. The role of antimicrobial peptides in preventing multidrug-resistant bacterial infections and biofilm formation. *Int J Mol Sci.* 2011;12(9):5971-5992. doi:10.3390/ijms12095971
57. Giles-Gómez M, Sandoval García JG, Matus V, Campos Quintana I, Bolívar F, Escalante A. In vitro and in vivo probiotic assessment of *Leuconostoc mesenteroides* P45 isolated from pulque, a Mexican traditional alcoholic beverage. *SpringerPlus.* 2016;5(1). doi:10.1186/s40064-016-2370-7
58. van Hijum SAFT, Vaughan EE, Vogel RF. Application of state-of-art sequencing technologies to indigenous food fermentations. *Curr Opin Biotechnol.* 2013;24(2):178-186. doi:10.1016/j.copbio.2012.08.004
59. Walsh AM, Crispie F, Claesson MJ, Cotter PD. Translating Omics to Food Microbiology. *Annu Rev Food Sci Technol.* 2017;8(1):113-134. doi:10.1146/annurev-food-030216-025729

60. Cao Y, Fanning S, Proos S, Jordan K, Srikumar S. A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Front Microbiol.* 2017;8(SEP):1-16. doi:10.3389/fmicb.2017.01829
61. Jung JY, Lee SH, Kim JM, et al. Metagenomic Analysis of Kimchi, a Traditional Korean Fermented Food. *Appl Environ Microbiol.* 2011;77(7):2264-2274. doi:10.1128/AEM.02157-10
62. Aldrete-Tapia A, Escobar-Ramírez MC, Tamplin ML, Hernández-Iturriaga M. High-throughput sequencing of microbial communities in Poro cheese, an artisanal Mexican cheese. *Food Microbiol.* 2014;44:136-141. doi:10.1016/j.fm.2014.05.022
63. Aldrete-Tapia A, Escobar-Ramírez CM, Tamplin ML, Hernández-Iturriaga M. Characterization of Bacterial Communities in Mexican Artisanal Raw Milk “Bola de Ocosingo” Cheese by High-Throughput Sequencing. *Front Microbiol.* 2018;9(October):1-7. doi:10.3389/fmicb.2018.02598
64. Murugesan S, Reyes-Mata MP, Nirmalkar K, et al. Profiling of bacterial and fungal communities of Mexican cheeses by high throughput DNA sequencing. *Food Res Int.* 2018;113(July):371-381. doi:10.1016/j.foodres.2018.07.023
65. Correa-Ascencio M, Robertson IG, Cabrera-Cortés O, Cabrera-Castro R, Evershed RP. Pulque production from fermented agave sap as a dietary supplement in Prehispanic Mesoamerica. *Proc Natl Acad Sci.* 2014;111(39):14223-14228. doi:10.1073/pnas.1408339111
66. Lorenzo MA. Las haciendas pulqueras de México. *Univ Nac Autónoma México.* 2007;(1):258.
67. Narro-Robles J, Gutiérrez-Ávila JH, López-Cervantes M, Phil M, Borges G, Rosovsky H. La mortalidad por cirrosis hepática en México II. Exceso de Mortalidad y consumo de pulque. *Salud Publica Mex.* 1992;34(388-405).
68. Narro-Robles J, Gutierrez-Avila JH. Correlacion ecologica entre consumo de bebidas alcoholicas y mortalidad por cirrosis hepatica en Mexico. *Salud Publica Mex.* 1997;39(3):217-220.
69. Backstrand JR, Goodman AH, Allen LH, Pelto GH. Pulque intake during pregnancy and lactation in rural Mexico: Alcohol and child growth from 1 to 57 months. *Eur J Clin Nutr.* 2004;58(12):1626-1634. doi:10.1038/sj.ejcn.1602019
70. Tovar LR, Olivos M, Gutierrez ME. Pulque, an alcoholic drink from rural Mexico, contains phytase. Its in vitro effects on corn tortilla. *Plant Foods Hum Nutr.* 2008;63(4):189-194. doi:10.1007/s11130-008-0089-5
71. Sánchez-Marroquín A, Larios C, Vierna L. Estudios sobre la microbiología del pulque. XIX. Elaboración de la bebida mediante cultivos puros, en planta piloto. *Rev Latinoam Microbiol Parasitol Mex.* 1967;9(2):83-85.

72. Sánchez-Marroquín A, Larios C, Vierna L. Estudios sobre la microbiología del pulque. XIX. Elaboración de la bebida mediante cultivos puros, en planta piloto. *Rev Latinoam Microbiol Parasitol Mex.* 1967;9(2):87-90.
73. Makarova KS, Koonin EV. Evolutionary Genomics of Lactic Acid Bacteria. *J Bacteriol.* 2007;189(4):1199-1208. doi:10.1128/JB.01351-06
74. Claesson MJ, van Sinderen D, O'Toole PW. The genus *Lactobacillus* – a genomic basis for understanding its diversity. *FEMS Microbiol Lett.* 2007;269(1):22-28. doi:10.1111/j.1574-6968.2006.00596.x
75. Wittouck S, Wuyts S, Meehan CJ, van Noort V, Lebeer S. A Genome-Based Species Taxonomy of the *Lactobacillus* Genus Complex. Gibbons SM, ed. *mSystems.* 2019;4(5):e00264-19. doi:10.1128/mSystems.00264-19
76. Torres-Maravilla E, Lenoir M, Mayorga-Reyes L, et al. Identification of novel anti-inflammatory probiotic strains isolated from pulque. *Appl Microbiol Biotechnol.* 2016;100(1):385-396. doi:10.1007/s00253-015-7049-4
77. Riveros-Mckay F, Campos I, Giles-Gómez M, Bolívar F, Escalante A. Draft Genome Sequence of *Leuconostoc mesenteroides* P45 Isolated from Pulque, a Traditional Mexican Alcoholic Fermented Beverage. *Genome Announc.* 2014;2(6):e01130-14. doi:10.1128/genomeA.01130-14
78. Troubleshooting your sequencing results, School of Life Sciences. University of Nottingham. Troubleshooting your sequencing results. Accessed January 10, 2017. <https://www.nottingham.ac.uk/life-sciences/facilities/dna-sequencing/troubleshooting.aspx>
79. Andrew S. *FASTQC. A Quality Control Tool for High Throughput Sequence Data.* Babraham Bioinformatics; 2010. <https://github.com/s-andrews/FastQC>
80. Kueger F. *Trim Galore. A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files, with Some Extra Functionality for MspI-Digested RRBS-Type (Reduced Representation Bisulfite-Seq) Libraries.* Babraham Bioinformatics; 2012. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
81. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 17(1):10-12.
82. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 2016;26(12):1721-1729. doi:10.1101/gr.210641.116
83. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun.* 2016;7(1):11257. doi:10.1038/ncomms11257

84. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15(3):R46. doi:10.1186/gb-2014-15-3-r46
85. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1):257. doi:10.1186/s13059-019-1891-0
86. Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods.* 2015;12(10):902-903. doi:10.1038/nmeth.3589
87. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, ed. *PLoS ONE.* 2013;8(4):e61217. doi:10.1371/journal.pone.0061217
88. Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 2014;42(D1):D633-D642. doi:10.1093/nar/gkt1244
89. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2012;41(D1):D590-D596. doi:10.1093/nar/gks1219
90. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics.* 2012;28(24):3211-3217. doi:10.1093/bioinformatics/bts611
91. Eren AM, Kiefl E, Shaiber A, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol.* 2021;6(1):3-6. doi:10.1038/s41564-020-00834-3
92. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2
93. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-359. doi:10.1038/nmeth.1923
94. Alneberg J, Bjarnason BS, de Bruijn I, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods.* 2014;11(11):1144-1146. doi:10.1038/nmeth.3103
95. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol.* 2016;428(4):726-731. doi:10.1016/j.jmb.2015.11.006
96. Eddy SR. Profile hidden Markov models. *Bioinformatics.* 1998;14(9):755-763. doi:10.1093/bioinformatics/14.9.755

97. Rinke C, Schwientek P, Sczyrba A, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature*. 2013;499(7459):431-437. doi:10.1038/nature12352
98. Campbell JH, O'Donoghue P, Campbell AG, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci*. 2013;110(14):5540-5545. doi:10.1073/pnas.1303090110
99. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-3212. doi:10.1093/bioinformatics/btv351
100. Seeman T. *Barnap: BAsic Rapid Ribosomal RNA Predictor*. <https://github.com/tseemann/barnap>
101. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28(11):1420-1428. doi:10.1093/bioinformatics/bts174
102. Zdobnov EM, Apweiler R. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001;17(9):847-848. doi:10.1093/bioinformatics/17.9.847
103. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674-1676. doi:10.1093/bioinformatics/btv033
104. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27(5):824-834. doi:10.1101/gr.213959.116
105. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119. doi:10.1186/1471-2105-11-119
106. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
107. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. doi:10.1093/bioinformatics/btt086
108. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5(1):113. doi:10.1186/1471-2105-5-113

109. van Dongen S, Abreu-Goodger C. Using MCL to Extract Clusters from Networks. In: van Helden J, Toussaint A, Thieffry D, eds. *Bacterial Molecular Networks: Methods and Protocols*. Springer New York; 2012:281-295. doi:10.1007/978-1-61779-361-5_15
110. Enright AJ. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002;30(7):1575-1584. doi:10.1093/nar/30.7.1575
111. Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS. Robust estimation of microbial diversity in theory and in practice. *ISME J*. 2013;7(6):1092-1101. doi:10.1038/ismej.2013.10
112. Štursa P, Uhlík O, Kurzawová V, et al. Approaches for diversity analysis of cultivable and non-cultivable bacteria in real soil. *Plant Soil Environ*. 2009;55(No. 9):389-396. doi:10.17221/136/2009-PSE
113. Bohannan BJM, Ricketts TH, Hellmann JJ, Hughes JB. Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. *APPL ENV MICROBIOL*. 2001;67:8.
114. Rohwer F, Prangishvili D, Lindell D. Roles of viruses in the environment. *Environ Microbiol*. 2009;11(11):2771-2774. doi:10.1111/j.1462-2920.2009.02101.x
115. Angly FE, Felts B, Breitbart M, et al. The Marine Viromes of Four Oceanic Regions. Moran NA, ed. *PLoS Biol*. 2006;4(11):e368. doi:10.1371/journal.pbio.0040368
116. Jung MJ, Kim MS, Yun JH, et al. Viral community predicts the geographical origin of fermented vegetable foods more precisely than bacterial community. *Food Microbiol*. 2018;76:319-327. doi:10.1016/j.fm.2018.06.010
117. Holmfeldt K, Solonenko N, Shah M, et al. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci*. 2013;110(31):12798-12803. doi:10.1073/pnas.1305956110
118. Aguilera G, Marthey S, Chiapello H, et al. Assessing the Performance of Single-Copy Genes for Recovering Robust Phylogenies. Ané C, ed. *Syst Biol*. 2008;57(4):613-627. doi:10.1080/10635150802306527
119. Mark Welch JL, Hasegawa Y, McNulty NP, Gordon JI, Borisy GG. Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. *Proc Natl Acad Sci*. 2017;114(43). doi:10.1073/pnas.1711596114