

UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ingeniería



**Desarrollo de *Pipeline* informático para la  
evaluación e identificación de microbioma intestinal  
de pacientes esquizofrénicos activos y/o pasivos**

trabajo de graduación en modalidad de tesis presentado por  
Licenciada María Isabel Montoya Valladares  
para optar al grado académico de licenciado en Ingeniería en  
Bioinformática

Guatemala,  
2023



UNIVERSIDAD DEL VALLE DE GUATEMALA  
Facultad de Ingeniería



**Desarrollo de *Pipeline* informático para la  
evaluación e identificación de microbioma intestinal  
de pacientes esquizofrénicos activos y/o pasivos**

trabajo de graduación en modalidad de tesis presentado por  
Licenciada María Isabel Montoya Valladares  
para optar al grado académico de licenciado en Ingeniería en  
Bioinformática

Guatemala,  
2023


Vo.Bo.:



(f)

MSc. Luis Augusto Franco López

Tribunal Examinador:



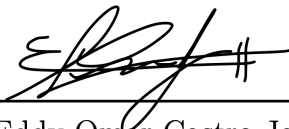
(f)

MSc. Luis Augusto Franco López



(f)

MSc. Jorge Eduardo Chang Estrada



(f)

Lic. Eddy Omar Castro Jauregui

Fecha de aprobación: Guatemala, 6 de diciembre 2023.

---

## Agradecimientos

---

Agradezco a mis padres, quienes me apoyaron en mis estudios desde el colegio hasta mi último año de universidad. A mi hermana por ser siempre el apoyo emocional.

Agradezco a Luis Pedro Cuéllar por apoyarme con sus conocimientos más específicos de computación y me orientó en la utilización de máquinas virtuales para la elaboración del producto del presente de trabajo de graduación.

Agradezco a mi asesor, Augusto, por el apoyo en los comandos bioinformáticos y en la revisión del presente trabajo.

Agradezco a Jorge Chang que me ayudó a orientar el tema de aplicación para el presente trabajo.

<b>Agradecimientos</b>	III
<b>Lista de figuras</b>	VI
<b>Lista de cuadros</b>	VII
<b>Resumen</b>	IX
<b>1. Introducción</b>	1
<b>2. Objetivos</b>	4
2.1. Objetivo general	4
2.2. Objetivos específicos	4
<b>3. Justificación</b>	5
<b>4. Marco teórico</b>	7
4.1. Trastornos neurodegenerativos	7
4.2. Metagenómica	8
4.3. Antecedentes de análisis de microbioma intestinal humana	10
4.4. Esquizofrenia	13
4.4.1. Diagnóstico	14
4.4.2. Esquizofrenia activa	15
4.4.3. Esquizofrenia pasiva	15
4.4.4. Vías neuronales en la esquizofrenia	15
4.5. Tratamientos existentes	16
4.6. Estado del arte de estudios genéticos enfocados en pacientes esquizofrénicos y la microbiota	18
4.6.1. Implicaciones para la investigación y el tratamiento futuros	20
4.7. Implicaciones éticas de la investigación de microbioma	21
4.8. Herramientas para el análisis de secuencias	23
4.8.1. SeqKit	23
4.8.2. FastQC	23
4.8.3. Fastp	24
4.8.4. Clasificación de taxonomía	25
4.9. Automatización de procesos computacionales	27
4.10. <i>Storytelling</i> : Visualización de resultados	28

4.10.1. R	29
4.10.2. Shiny	29
<b>5. Metodología</b>	<b>31</b>
5.1. Metodología	31
5.1.1. Control de calidad y filtrado de secuencias	33
5.1.2. Asignación de taxonomía	40
5.1.3. Análisis de secuencias	43
5.1.4. Visualización de resultados	47
<b>6. Resultados</b>	<b>49</b>
6.1. Alfa diversidad	49
6.2. Beta diversidad	49
6.3. Abundancia	51
6.3.1. Abundancia absoluta de filos	51
6.3.2. Abundancia relativa de filos	53
6.3.3. Abundancia relativa de géneros	53
<b>7. Discusión</b>	<b>61</b>
<b>8. Conclusiones</b>	<b>66</b>
<b>9. Recomendaciones</b>	<b>67</b>
<b>10. Bibliografía</b>	<b>68</b>
<b>11. Anexos</b>	<b>74</b>
11.1. Archivo pipeline.sh creado	74

5.1. Diagrama de flujo de pasos para el <i>pipeline</i> . . . . .	32
5.2. Estructura inicial de archivos en sistema Linux . . . . .	33
5.3. Resultado de "stats" de seqkit sobre la secuencias descargadas . . . . .	34
5.4. Esquema de directorio . . . . .	35
5.5. Reporte generado por Fastqc . . . . .	35
5.6. Gráficas generadas por Fastqc . . . . .	37
5.7. Reporte generado por fastp por secuencia acerca del estado anterior de las seuencias y luego de ser filtradas . . . . .	38
5.8. Archivos creados por bowtie2-build . . . . .	39
5.9. Ejemplo de archivo .kraken . . . . .	44
5.10. Reporte generado por clasificación Kraken2 . . . . .	44
5.11. Archivo HTML creado por Krona . . . . .	45
5.12. Ejemplo de la estructura final de archivos a obtener . . . . .	48
5.13. Layout principal de <i>Dashboard</i> generado por Shiny App . . . . .	48
6.1. Alfa diversidad para muestras control . . . . .	50
6.2. Alfa diversidad para muestras de pacientes esquizofrenicos . . . . .	50
6.3. Beta diversidad - DCA . . . . .	51
6.4. Beta diversidad - CA . . . . .	51
6.5. Beta diversidad - NMDS . . . . .	52
6.6. Abundancia absoluta de las muestras . . . . .	52
6.7. Abundancia relativa de filos 0.5% > . . . . .	53
6.8. Abundancia relativa de filos 0.9% . . . . .	54
6.9. Ejemplo de tabla en Shiny . . . . .	54
6.10. Abundancia relativa de géneros 0.5% . . . . .	55
6.11. Tabla resumen de los géneros con abundancia relativa 0.5% > . . . . .	56
6.12. Abundancia relativa de géneros 0.7% . . . . .	56
6.13. Tabla resumen de los géneros con abundancia relativa 0.7% > . . . . .	57
6.14. Abundancia relativa de géneros 2% . . . . .	57
6.15. Tabla resumen de los géneros con abundancia relativa 2% > . . . . .	58
6.16. Abundancia relativa de géneros 5% . . . . .	59
6.17. Tabla resumen de los géneros con abundancia relativa 5% > . . . . .	60

---

Lista de cuadros

---

4.1. Medicamentos recetados en Guatemala con sus efectos adversos . . . . . 17

5.1. Instalación de canales . . . . .	33
5.2. Creación de ambientes fastqc . . . . .	33
5.3. Activación de ambiente fastqc . . . . .	33
5.4. Instalación de Seqkit y uso . . . . .	34
5.5. Instalación de fastqc y uso . . . . .	34
5.6. Filtración de lecturas basadas en calidad de secuenciación a través de FASTP . . . . .	35
5.7. Instalación de bowtie2 y creación de index . . . . .	38
5.8. Filtrado de lecturas . . . . .	38
5.9. Conversión de formato SAM a FASTQ.GZ . . . . .	39
5.10. Uso de Kaiju . . . . .	40
5.11. Instalación de Kraken2 y construcción de base de datos . . . . .	42
5.12. Descarga de base de datos standard de Kraken2 . . . . .	42
5.13. Clasificación de bacterias con Kraken2 . . . . .	43
5.14. Visualización con krona . . . . .	45
5.15. Utilización de KrakenTools . . . . .	46
5.16. Utilización de Kraken-biom . . . . .	47
5.17. Ejecución del archivo que contiene el <i>pipeline</i> creado . . . . .	47
11.1. Contenido del archivo pipeline.sh . . . . .	74

La importancia del desarrollo de *pipelines* bioinformáticos de metagenómica radica en que permiten analizar de manera eficiente y precisa grandes cantidades de datos de secuenciación. Estos datos se generan a partir de muestras ambientales, como suelo, agua, o aire, que contienen una gran diversidad de microorganismos. Estos *pipelines* se encargan de procesar estos datos para identificar y caracterizar los microorganismos presentes en la muestra. Esto permite realizar un estudio detallado de la diversidad microbiana de un ecosistema, así como identificar microorganismos potencialmente patógenos o beneficiosos. En el presente trabajo se aborda el caso de aplicación del desarrollo de un *pipeline* bioinformático para el análisis de microbioma intestinal de pacientes con esquizofrenia. Las muestras fueron obtenidas de la base de datos pública de <https://www.ebi.ac.uk> bajo los códigos de acceso de PRJEB41217 y PRJEB41786. Estos contienen las muestras secuenciadas de pacientes diagnosticados con esquizofrenia y pacientes 'sanos' de Dinamarca.

Con el *pipeline* se lleva a cabo un proceso de automatización en donde se realiza las acciones de: control de calidad, filtrado de lecturas dentro de las secuencias, remoción de ADN del hospedero de las muestras (en este caso el humano), clasificación taxonómica de las muestras y la creación de archivos en formato BIOM que puedan ser procesados por la librería phyloseq de R. Dado la alta carga computacional del *pipeline* se hizo uso de una máquina virtual de Digital Ocean que lleve a cabo la ejecución de la automatización desarrollada para 53 secuencias provenientes del microbioma intestinal de pacientes diagnosticados con esquizofrenia y 53 de pacientes sanos. Finalmente, usando R Shiny, shinydashboard y phyloseq como librerías principales para la creación de un *dashboard*, se lleva a cabo tanto el análisis de alfa y beta diversidad como también de abundancias absolutas y relativas. Añadiendo que se toma en cuenta que la visualización creada es interactiva con el usuario. Esto significa que este puede realizar diferentes filtrados de datos o indicar con que métodos trabajar para la generación de las gráficas para analizar.

El *pipeline* fue exitoso en la clasificación del microbioma intestinal de las muestras utilizadas. Con este se encontró que en términos de alfa y beta diversidad no hay una diferencia significativa entre el conjunto de muestras de pacientes con esquizofrenia y los pacientes control. Sin embargo, al analizar las abundancias relativas por género de las muestras se encontró que hay una abundancia relativa alta de los géneros *Flavonifractor*, *Dysosmobacter*, *Escheria* y *Lactobacillus* en las muestras de pacientes con esquizofrenia.

La metagenómica es una rama de la genómica que se enfoca en el estudio de comunidades microbianas complejas presentes en diversos ambientes, como lo es el suelo, el agua, el intestino humano y muchos otros ([Metagenomics](#), s.f.). A diferencia de la genómica convencional, la cual se centra en el estudio de un solo organismo o especie a la vez; la metagenómica proporciona información sobre la diversidad y composición de las comunidades microbianas y aborda la biodiversidad presente en una muestra ambiental sin necesidad de aislar y cultivar individualmente a los microorganismos. Esto significaría un menor esfuerzo de laboratorio para el análisis de estos microorganismos ([Lu y cols.](#), 2022).

Los microbiomas, desempeñan un papel fundamental en numerosos procesos biológicos y ambientales, como la descomposición de materia orgánica, la síntesis de nutrientes y la influencia en la salud humana ([Aggarwal y cols.](#), 2022). La metagenómica nos permite analizar y comprender la diversidad genética y funcional de estas comunidades microbianas de una manera más completa y precisa ([Cryan y Dinan](#), 2012). Esto también se debe a que solamente el 2% de los microorganismos pueden ser cultivados exitosamente dentro de un laboratorio; la metagenómica ahorra esta pérdida de información ([Pipeline-Pitfalls](#), s.f.).

Para los estudios metagenómicos existen pasos fundamentales que incluyen la recolección de muestras ambientales, la extracción de ADN, la secuenciación masiva de genes o genomas, y el análisis bioinformático de los datos resultantes. Este análisis puede revelar información sobre la identidad de los microorganismos presentes, sus funciones metabólicas, interacciones microbianas y cómo responden a cambios en su entorno ([Taxonomic Analysis with R - Data Processing and Visualization for Metagenomics](#), s.f.).

Los *pipelines* de metagenómica son una herramienta esencial para los investigadores que estudian la diversidad microbiana. Estos *pipelines* pueden automatizar el proceso de análisis de datos de secuenciación de ADN. Lo que permite a los investigadores centrarse en la interpretación de los resultados ([Lu y cols.](#), 2022). El desarrollo de un *pipeline* de metagenómica puede ser una tarea compleja y desafiante. Sin embargo, los *pipelines* bien diseñados pueden ahorrar tiempo y esfuerzo a los investigadores. Además, esta automatización puede ayudar a garantizar la precisión de los resultados y su reproductibilidad. Cabe añadir que, para investigadores con pocos conocimientos bioinformáticos, presenta una gran ventaja ya que, el desarrollo y uso de comandos para llegar a producir el resultado para este análisis puede resultar tedioso.

Un *pipeline* de metagenómica consta de una serie de pasos que se utilizan para procesar y analizar

los datos de secuenciación de ADN de una muestra ambiental. Estos pasos típicos incluyen:

- **Preprocesamiento:** Este paso se utiliza para limpiar los datos de secuenciación y eliminar los artefactos. Para la metagenómica este paso suele incluir: filtrado de lecturas, corrección de errores, eliminación de adaptadores, corte de bases de mala calidad, etc.
- **Alineamiento:** Este paso se utiliza para comparar las secuencias de la muestra con un conjunto de referencia de secuencias conocidas.
- **Clasificación:** Este paso se utiliza para asignar las secuencias alineadas a taxones biológicos.
- **Análisis:** Este paso se utiliza para identificar las características de la comunidad microbiana, como la diversidad, la composición y la abundancia de los microorganismos.

*(Taxonomic Analysis with R – Data Processing and Visualization for Metagenomics, s.f.)*

El *pipeline* ayuda a este análisis al automatizar estos pasos y al proporcionar un flujo de trabajo reproducible. Al estar hablando de metagenómica también se comprende el uso de secuencias de tamaño considerable, por lo que el estar ejecutando los comandos manualmente puede conllevar a que el análisis de estas secuencias sea un proceso más tardado de lo que debe de ser.

La metagenómica en general tiene una amplia gama de aplicaciones, desde la investigación en ecología microbiana hasta la mejora de procesos industriales, la vigilancia epidemiológica y el estudio de la microbiota humana. En el campo de la salud, por ejemplo, se ha utilizado para comprender mejor la influencia de la microbiota intestinal en la digestión, el sistema inmunológico y enfermedades como la obesidad o enfermedades autoinmunitarias. Sin embargo, en el presente proyecto se enfocará en el análisis del microbioma intestinal de pacientes esquizofrénicos.

La esquizofrenia es una enfermedad mental severa que puede afectar la forma en la que piensa una persona, siente y se comporta. Es un desorden psiquiátrico caracterizado por síntomas activos o “positivos” como lo son las alucinaciones, flujo erróneo de pensamientos, diluciones; como también síntomas pasivos o “negativos” como lo son la apatía, paranoia y desapego. Este desorden afecta alrededor de 21 millones de personas en el mundo y puede llegar a generar una mórbida física y social significativa (*Schizophrenia, s.f.-a*) (*Weickert, Weickert, Pillai, y Buckley, 2013b*).

La etimología de este desorden puede variar dependiendo del individuo que está siendo tratado, sin embargo, la enfermedad se atribuye a las siguientes causas: genética ya que existe la posibilidad que la esquizofrenia pueda ser heredada de padres a hijos; neuroquímica, siendo el individuo deficiente en la regulación de neurotransmisores que controlan ciertas vías metabólicas o “circuitos” que afecta en la manera de pensar o de comportarse; estructura anormal del cerebro; y ambiente/crianza en la cual el individuo creció, ya que un ambiente en el cual la persona fue expuesta a infecciones virales, toxinas o exposición frecuente a situaciones altamente estresantes puede causar que la persona presente síntomas de esquizofrenia (*Contributors, s.f.*).

La esquizofrenia suele tener un curso crónico, aunque con patrones fluctuantes, y discapacidad cognitiva. Su rasgo distintivo es la psicosis, caracterizada principalmente por síntomas positivos, alucinaciones y delirios, a menudo acompañados de síntomas negativos (déficit) síntomas negativos (deficitarios), como la disminución de las emociones, el habla y el interés, y la desorganización de la vida cotidiana, y el interés, y por la desorganización del habla y el comportamiento. Los síntomas del estado de ánimo están presentes en muchos casos, lo que dificulta el diagnóstico preciso. No existen pruebas de laboratorio para diagnosticar la esquizofrenia, sino que el diagnóstico se basa en la observación clínica y en la observación clínica y el autoinforme obtenido e interpretado por clínicos experimentados (*Gejman, Sanders, y Kendler, 2011*).

Actualmente el tratamiento para las personas que han sido diagnosticadas con el presente desorden consisten en ayudarlas en: manejar sus síntomas, mejorar su comportamiento en el día a día y

poder llevar lo que se identifica como “vida normal” en la que pueden desarrollarse como profesionales y mantener relaciones interpersonales sanas. Entre ellos se encuentran los tratamientos psicosociales en los cuales se apoya en la búsqueda de soluciones para obstáculos que una persona esquizofrénica se puede encontrar, apoyo y educación familiar, cuidado especial coordinado, tratamiento comunal asertivo, y medicación antipsicótica (Contributors, s.f.).

A pesar de que existen diferentes tipos de tratamientos, se encuentra que en los pacientes de este desorden mental son apoyados con ayuda psicológica en conjunto de una medicación. Siendo la solución de proveer antipsicóticos a la persona una constante en este tipo de diagnósticos. Estos medicamentos tienen el objetivo de hacer que los síntomas activos sean menos intensos y frecuentes. Sin embargo, cabe mencionar que estos medicamentos pueden causar efectos adversos en los pacientes. Estos siendo problemas hormonales, subida de peso, psicosis tardía, disquinesia, movimientos abruptos del cuerpo, ataques epilépticos, brotes de alergia en la piel, decoloración de piel, incremento de sensibilidad, puede llegar a producir pensamientos y comportamientos suicidas, taquicardia, temblores, entre otros (Side effects of antipsychotics, s.f.) (Stroup, 2018).

Siendo la esquizofrenia uno de los desórdenes mentales más severos en tema de su afección hacia la vida del individuo que ha sido diagnosticado. Por otro lado, en estudios para analizar la implicación de diversos factores que pueden estar relacionados con este tipo de patologías, se ha identificado al microbioma intestinal como una variable a tomar en cuenta. Como lo puede ser el ejemplo del caso de que en desórdenes relacionados con depresión, aquí se ha encontrado que la disbacteriosis, condición de desequilibrio en el número o tipo de colonias microbianas en el humano, y trasplantes fecales de personas diagnosticadas con depresión hacia ratas sin presencia de este microbioma induce comportamientos depresivos en estas ratas “sanas” (Kelly y cols., 2016). Dando así una nueva mirada en la búsqueda de soluciones o tratamientos para personas esquizofrénicas.

El microbioma intestinal del ser humano es una comunidad rica de microorganismos que son diversos y en ciertos casos “personalizados” por las condiciones únicas de cada persona. Este microbioma también es un ambiente dinámico. Este es influenciado por factores de estilo de vida, como lo puede ser la dieta, horas de sueño o cantidades grandes de estrés (Karl y cols., 2017). Por la naturaleza de esta dinámica es que la definición de una disbacteriosis es un concepto arduo de definir, además de que crea la limitación en la definición de cuales cambios cualitativos y cuantitativos establecen una desviación significativa de una cantidad estándar de bacterias dentro del intestino.

Esto teniendo consecuencias que se buscan a partir de una diferencia significativa a nivel macro de poblaciones control en comparación con pacientes diagnosticados con esquizofrenia. Teniendo así, la oportunidad de encontrar si estas bacterias desequilibradas tienen una función significativa en el proceso patológico del paciente. Buscando una relación entre la patogénesis de la esquizofrenia y el microbioma intestinal del paciente, para la identificación de bacterias clave del desorden, brindando así una perspectiva más grande acerca de la patogénesis de la esquizofrenia.

### 2.1. Objetivo general

Desarrollar un *pipeline* bioinformático que permita el análisis automatizado de secuencias de microbioma de pacientes esquizofrénicos de la población de Dinamarca.

### 2.2. Objetivos específicos

- Determinar el tipo de microbioma presente en poblaciones diagnosticadas con esquizofrenia pasiva o activa en Dinamarca.
- Generar una visualización que permita el análisis de la clasificación generada por el *pipeline* desarrollado.
- Clasificar las bacterias encontradas en el microbioma intestinal de pacientes diagnosticados con esquizofrenia.

El microbioma intestinal, la comunidad de microorganismos que habita en el intestino humano, ha sido objeto de estudios recientes que revelan diferencias significativas entre el microbioma de pacientes esquizofrénicos y el de personas sanas. Esta posible correlación cobra relevancia al considerar su asociación con diversos aspectos vinculados a la esquizofrenia, tales como la inflamación, la alteración de la inmunidad y la función cognitiva. En este contexto, resulta de gran interés observar cómo los cambios en la microbiota pueden desempeñar un papel clave en los problemas asociados con la esquizofrenia, estableciendo así un vínculo crucial entre la salud intestinal y la salud mental (Akhondzadeh, 2019).

El desarrollo de un *pipeline* bioinformático para el análisis de metagenoma puede facilitar la investigación sobre la relación entre el microbioma intestinal y diversas enfermedades, entre ellas la esquizofrenia. Este *pipeline* podría utilizarse para analizar los datos de secuenciación de ADN de microbiomas intestinales de pacientes esquizofrénicos, con el objetivo de identificar los cambios en la composición y la función de las comunidades microbianas que pueden estar asociados con la enfermedad. Además, la implementación de un *pipeline* bioinformático no solo automatiza el proceso de análisis de metagenoma, sino que también emerge como una herramienta vital para ahorrar tiempo y esfuerzo a los investigadores. Esta eficiencia no solo beneficia directamente a los científicos, sino que también tiene implicaciones positivas en el ámbito de la salud pública (Afshinnekoo y cols., 2017). La rapidez en el análisis metagenómico puede traducirse en diagnósticos más rápidos y precisos, lo que, a su vez, puede acelerar la implementación de tratamientos personalizados para los pacientes. Así, el uso de *pipelines* bioinformáticos no solo impulsa la investigación en el ámbito bioinformático de la enfermedad, sino que también se convierte en un componente esencial para el avance de soluciones más viables y accesibles en la salud.

Actualmente en Guatemala la investigación de enfermedades neurodegenerativas, en general, es un tema escaso dada la poca importancia que se le da a la salud mental dentro de la nación (Mangandid, s.f.). Dentro de esto, los estudios genéticos de la población guatemalteca diagnosticada con enfermedades mentales es aún más escaso; esto añadiendo la causa de tener un escaso presupuesto para las investigaciones de dicha índole y el tema de una secuenciación, la cual es necesaria para llevar a cabo estos estudios, tiene un precio muy elevado (McBride, 2022).

En la actualidad, el diagnóstico y tratamiento de pacientes psiquiátricos, especialmente aquellos con esquizofrenia, a menudo se lleva a cabo mediante métodos rudimentarios, como entrevistas y la observación de características cualitativas (Australia, 2022). Posteriormente, se receta un único medicamento, generalmente antipsicóticos típicos o atípicos, para abordar los síntomas activos, con

la posibilidad de administrarse a través de dosis inyectables o ingeribles en diferentes momentos del día. Sin embargo, este enfoque presenta posibles efectos adversos, como taquicardia, agitaciones, movimientos lentos, fiebre, mareos y malestares estomacales (Vásquez, s.f.).

La búsqueda de características diferenciadoras en el microbioma intestinal de pacientes esquizofrénicos representa un avance significativo. Al identificar biomarcadores específicos, podemos complementar los métodos tradicionales de diagnóstico, proporcionando una herramienta más precisa y objetiva para la identificación de la enfermedad (Weickert, Weickert, Pillai, y Buckley, 2013a). Además, este enfoque no solo contribuye a un diagnóstico más preciso, sino que también señala comunidades bacterianas que podrían convertirse en objetivos terapéuticos para el tratamiento de la esquizofrenia. La exploración de estas comunidades ofrece la posibilidad de desarrollar terapias más específicas y dirigidas, evitando así los efectos adversos negativos asociados con los antipsicóticos convencionales (Aggarwal y cols., 2022).

Este proyecto busca encontrar si existe una relación entre el microbioma intestinal y la presentación de síntomas activos del paciente a través del uso de herramientas bioinformáticas a través de un *pipeline* automatizado que realice desde el control de calidad hasta la clasificación de microbioma intestinal de pacientes esquizofrénicos de Dinamarca y una visualización que permita el análisis de estos resultados.

## 4.1. Trastornos neurodegenerativos

Los trastornos neurodegenerativos se caracterizan por la degeneración progresiva de las neuronas del sistema nervioso central (SNC), que incluye el cerebro y la médula espinal (Mawe y cols., 2022). Esta degeneración suele ser irreversible y provoca la pérdida gradual de la estructura y la función de las neuronas afectadas. Una característica común de muchos trastornos neurodegenerativos es la acumulación anormal de ciertas proteínas en las neuronas o en los tejidos circundantes (Mawe y cols., 2022). Por ejemplo, la enfermedad de Alzheimer se asocia a la acumulación de placas beta-amiloides y ovillos tau, mientras que la enfermedad de Parkinson implica la agregación de alfa-sinucleína. Estos agregados proteínicos interfieren con la función celular normal y contribuyen a la disfunción y muerte neuronal (Salahuddin y cols., 2021).

A medida que las enfermedades neurodegenerativas progresan, las neuronas sufren muerte celular, lo que conduce a una reducción del número de neuronas en funcionamiento (Mawe y cols., 2022). Esta pérdida neuronal puede provocar el encogimiento de las regiones cerebrales afectadas, lo que afecta a las funciones cognitivas y motoras. Además, suele haber una pérdida simultánea de sinapsis, las conexiones entre neuronas cruciales para la comunicación (Mawe y cols., 2022). La pérdida de sinapsis contribuye al declive de la actividad de la red neuronal y agrava aún más los trastornos cognitivos y motores.

Nos encontramos que estas enfermedades también se caracterizan por la presentación de una inflamación y estrés oxidativo en el SNC. La microglía, las células inmunitarias del cerebro, se activa y libera moléculas inflamatorias en respuesta al daño neuronal (Cummings, 2021). Aunque la inflamación es una respuesta protectora, la inflamación crónica o excesiva en el cerebro puede contribuir a un daño mayor. El estrés oxidativo, resultante de un desequilibrio entre la producción de especies reactivas del oxígeno y la capacidad del organismo para desintoxicarlas, también puede contribuir al daño celular en las enfermedades neurodegenerativas (Cummings, 2021) (Mawe y cols., 2022).

Los trastornos neurodegenerativos abarcan un grupo diverso de afecciones, cada una con su propio conjunto de síntomas, regiones cerebrales afectadas y patología subyacente. La enfermedad de Alzheimer se asocia principalmente con pérdida de memoria y deterioro cognitivo, la enfermedad de Parkinson con trastornos motores y la enfermedad de Huntington con síntomas motores y psiquiátricos (Salahuddin y cols., 2021). Esta diversidad refleja la complejidad de estos trastornos y plantea

retos a la hora de desarrollar tratamientos universales. A medida que nos adentramos en la compleja red que conecta los trastornos neurodegenerativos y la salud mental, se hace evidente que estas afecciones trascienden la mera disfunción neurológica, influyendo también en aspectos cognitivos y emocionales (Levenson, Sturm, y Haase, 2014).

La imbricación de los aspectos neurológicos y de salud mental en los trastornos neurodegenerativos es compleja. La acumulación anormal de proteínas, la pérdida sináptica y la inflamación en el cerebro contribuyen no sólo a los síntomas físicos, sino también a alteraciones en el estado de ánimo y el comportamiento. Comprender esta interacción es crucial para una atención y un apoyo holísticos (Levenson y cols., 2014).

La esquizofrenia es una enfermedad neuropsiquiátrica compleja que suele caracterizarse por alteraciones profundas del pensamiento, la percepción y la regulación emocional. Aunque tradicionalmente se ha clasificado como un trastorno mental funcional, las nuevas investigaciones han arrojado luz sobre las bases neurobiológicas que sitúan a la esquizofrenia dentro del espectro de los trastornos neurodegenerativos (Levenson y cols., 2014). Los estudios de neuroimagen revelan cambios estructurales en el cerebro de las personas con esquizofrenia, como alteraciones en el volumen de materia gris y en los patrones de conectividad (Keshavan y cols., 2020). Además, los estudios moleculares implican anomalías en los sistemas neurotransmisores, en particular la dopamina, el glutamato y la serotonina (Mandal y cols., 2022).

## 4.2. Metagenómica

La metagenómica es una técnica que permite estudiar la diversidad microbiana en su contexto natural. A diferencia de la genómica convencional, que se centra en el análisis de genomas individuales, la metagenómica se enfoca en la recolección y secuenciación de material genético a partir de comunidades microbianas complejas. En lugar de aislar y cultivar microorganismos en laboratorios, la metagenómica permite el acceso directo al ADN luego de su debida extracción en el laboratorio. Este material genético puede provenir de una variedad de fuentes, como suelos, océanos, aguas residuales, el tracto gastrointestinal humano y otros hábitats (Pereira, 2019).

La metagenómica se apoya en una serie de metodologías y tecnologías avanzadas que permiten la secuenciación masiva y el análisis de los datos genómicos recopilados de muestras ambientales. La utilización de tecnologías de secuenciación de próxima generación (NGS) ha sido el avance más notable en el campo de la metagenómica. Estas tecnologías han mejorado tanto en velocidad y en el alcance de la recopilación de datos metagenómicos de muestras (*Taxonomic Analysis with R – Data Processing and Visualization for Metagenomics*, s.f.).

Las tecnologías NGS permiten secuenciar grandes cantidades de ADN o ARN en una sola ejecución. Esto genera grandes conjuntos de datos que requieren un procesamiento masivo. Para ello, se han desarrollado herramientas bioinformáticas especializadas que pueden procesar y analizar estos datos de manera eficiente (*Metagenomics Next Generation Sequencing*, s.f.).

En el proceso de metagenómica, se extrae el material genético de una muestra ambiental. Luego, se preparan librerías necesarias para la secuenciación y se secuencian. A menudo, se secuencian múltiples muestras para obtener una imagen más completa de la comunidad microbiana presente en un entorno. Además, se pueden utilizar técnicas de meta-transcriptómica para estudiar la expresión génica en estas comunidades (Pereira, 2019). Esto proporciona información sobre las funciones que están activas en un momento dado.

El análisis de datos metagenómicos implica la identificación y anotación de genes, la construcción de perfiles de meta-genomas y la inferencia de funciones metabólicas. Para ello, se utilizan bases de datos de genes y proteínas, algoritmos de asignación taxonómica y funcional, y herramientas de minería de datos para extraer información relevante (*Taxonomic Analysis with R – Data Processing*

---

*and Visualization for Metagenomics*, s.f.).

La metagenómica, con su capacidad única para explorar la diversidad genética y funcional de comunidades microbianas en su entorno natural, ha encontrado una amplia gama de aplicaciones en diversas áreas de la ciencia y la industria como lo son:

- Exploración de la biodiversidad microbiana: La metagenómica ha permitido la identificación de microorganismos previamente desconocidos, así como la caracterización de sus genes y funciones. Esto es esencial para comprender la biodiversidad microbiana en ecosistemas como los suelos, los océanos, los sedimentos y otros entornos.
- Salud humana y microbioma: La metagenómica ha sido fundamental en la caracterización de la microbiota humana, lo que ha llevado a un profundo entendimiento de su papel en la salud y las enfermedades. Se ha aplicado en la identificación de desequilibrios en la microbiota (disbiosis) y en el desarrollo de terapias basadas en la restauración de la microbiota, como los trasplantes fecales.
- Medio ambiente y conservación: La metagenómica se emplea para evaluar la salud de los ecosistemas naturales y afectados por la contaminación. Permite monitorear la respuesta de las comunidades microbianas a cambios ambientales, lo que es crítico para la conservación de la biodiversidad y la gestión sostenible de recursos naturales.
- Industria alimentaria y fermentación: En la producción de alimentos, la metagenómica es útil para controlar y optimizar procesos de fermentación, como la producción de queso, yogur, cerveza y vino. Permite entender las comunidades microbianas responsables de estos procesos y mejorar la calidad de los productos finales.
- Agricultura y suelos: La metagenómica se aplica en la agricultura para mejorar la calidad del suelo, promover la fertilidad y prevenir la erosión. Ayuda a identificar microorganismos beneficiosos y entender sus interacciones en el suelo.

(L. Zhang y cols., 2021)

En el caso del presente proyecto, se hará un análisis de microbioma humana y sus posibles implicaciones en la salud mental.

A pesar de los avances significativos en la metagenómica, esta disciplina no está exenta de desafíos y limitaciones importantes que deben ser abordados para aprovechar su pleno potencial, entre estas limitaciones se puede encontrar:

- Complejidad de los datos: Los conjuntos de datos metagenómicos son enormes y altamente complejos, lo que requiere una capacidad de procesamiento computacional significativa. El análisis de estos datos a menudo se convierte en un desafío técnico, y la interpretación de los resultados puede ser complicada debido a la variabilidad natural en las comunidades microbianas.
- Función genómica desconocida: Gran parte de los genes identificados en los meta-genomas no tienen una función conocida, lo que dificulta la asignación de funciones metabólicas específicas a los microorganismos. Esto limita nuestra comprensión de la ecología microbiana y su contribución a procesos biogeoquímicos.
- Contaminación cruzada de muestras: La contaminación cruzada de muestras durante la recolección, procesamiento y tratamiento de las muestras puede introducir errores humanos en los datos metagenómicos. Por lo que es importante que las personas involucradas en el procesamiento de laboratorio sean instruidos acerca de las medidas de higiene y técnicas propias para evitar estos desafíos.

- Limitaciones de secuenciación: Aunque las tecnologías de secuenciación de próxima generación han avanzado, todavía existen limitaciones en términos de longitud de las secuencias y errores de secuenciación. Esto puede afectar la precisión en la asignación de taxonomía y función.
- Bases de datos incompletas: Las bases de datos utilizadas para comparar y anotar secuencias genómicas a menudo no están actualizadas y pueden ser incompletas. Esto dificulta la identificación de microorganismos y funciones desconocidas.
- Ética y privacidad: En el contexto de la microbiota humana, surgen preocupaciones éticas y de privacidad en torno a la recopilación y el análisis de datos metagenómicos.
- Recursos financieros y tecnológicos: La metagenómica requiere recursos financieros considerables, tanto en equipos de secuenciación de última generación como en la capacidad de computación para el análisis de datos. Esto puede limitar el acceso a esta tecnología en ciertas áreas o instituciones.

(Y. Chen, Fan, Chai, y Xu, 2022)

Abordar estos desafíos es esencial para garantizar la precisión y la utilidad de los estudios metagenómicos.

La metagenómica es un campo de investigación prometedor que está teniendo un impacto cada vez mayor en la ciencia y la sociedad. El futuro de la metagenómica está lleno de promesas. Las tecnologías de secuenciación seguirán mejorando, lo que permitirá obtener datos más rápidamente y con mayor precisión. Esto hará que la metagenómica sea más accesible y permitirá realizar estudios a mayor escala en una variedad de entornos.

La colaboración será cada vez más importante para abordar los desafíos y oportunidades de la metagenómica. La construcción de bases de datos más completas y actualizadas, el desarrollo de algoritmos de análisis más sofisticados y la estandarización de prácticas metodológicas son áreas clave de desarrollo (Y. Chen y cols., 2022).

En resumen, la metagenómica es una herramienta poderosa y versátil que ha transformado nuestra comprensión de la diversidad microbiana y su impacto en el mundo que nos rodea. A medida que se superen los desafíos actuales, podemos esperar que la metagenómica continúe contribuyendo a importantes descubrimientos científicos, avances en la medicina y aplicaciones prácticas en una variedad de campos.

### 4.3. Antecedentes de análisis de microbioma intestinal humana

La interacción entre el intestino y el cerebro del cuerpo humano ha fascinado durante mucho tiempo a investigadores y profesionales de la salud. Aunque antes se consideraban sistemas distintos y separados. Ahora se reconoce que el intestino y el cerebro están interconectados con una vasta red de comunicación que los une. En el centro de esta conexión se encuentra el microbioma gastrointestinal (GI) (Cryan y Dinan, 2012) (Foster, Rinaman, y Cryan, 2017), un complejo ecosistema de billones de microorganismos, entre bacterias, virus y hongos, que reside en el tracto digestivo humano. En la última década, el microbioma gastrointestinal se ha convertido en una pieza clave en los campos de la neurociencia y la psiquiatría, ofreciendo profundos conocimientos sobre la relación entre la salud intestinal y el bienestar mental.

Este creciente campo de investigación, a menudo denominado 'eje-microbiota-intestino-cerebro', explora la dinámica comunicación bidireccional entre el intestino y el cerebro. Este sugiere que el microbioma intestinal puede ejercer una influencia sustancial en los resultados de la salud mental

y la funcionalidad del sistema nervioso central. Aunque en principio este concepto pueda parecer inusual, las pruebas que apoyan la conexión intestino-cerebro son convincentes y siguen ampliándose rápidamente. Además, está ofreciendo una perspectiva novedosa sobre la salud mental, destacando el papel potencial del microbioma en afecciones como la depresión, la ansiedad e incluso los trastornos del neurodesarrollo (Kelly y cols., 2016).

La influencia del microbioma gastrointestinal en el cerebro es objeto de intensa exploración. Al parecer, no sólo influye en los estados psicológicos, sino también en la función cognitiva e incluso en el comportamiento (Cryan y Dinan, 2012). Los mecanismos subyacentes a esta conexión son complejos y multifacéticos. Una posible vía es que los metabolitos microbianos, como los ácidos grasos de cadena corta, pueden atravesar la barrera hematoencefálica y afectar la actividad cerebral. Otra vía es que los compuestos neuroactivos producidos por las bacterias intestinales, como la serotonina y la dopamina, pueden actuar directamente sobre los receptores cerebrales (Carabotti, Scirocco, Maselli, y Severi, 2015). Finalmente, las respuestas inmunitarias desencadenadas por el microbioma intestinal también pueden afectar el cerebro. A medida que los investigadores profundizan en esta intrincada red de interacciones, pretenden desentrañar la contribución del microbioma a afecciones tan diversas como el estrés, los trastornos del estado de ánimo, los trastornos del espectro autista, etc (Carabotti y cols., 2015) (Mawe y cols., 2022).

Cryan y Dinan (2012) (Cryan y Dinan, 2012) realizaron un estudio que trata de la investigación emergente sobre la relación entre la microbiota intestinal y la función cerebral y el comportamiento. Los autores repasan acerca las distintas formas en que la microbiota intestinal puede comunicarse con el cerebro. Por ejemplo, a través del nervio vago, el nervio vago, también conocido como décimo par craneal, es un componente clave del sistema nervioso autónomo (de Theije y cols., 2011). Es el nervio craneal más largo y desempeña un papel crucial en la regulación de diversas funciones corporales. El mecanismo del nervio vago implica tanto funciones sensoriales como motoras. El nervio vago actúa como vía de comunicación bidireccional entre el cerebro y diversos órganos del cuerpo. Desempeña un papel crucial en el mantenimiento de la homeostasis y la regulación de las funciones corporales. Las disfunciones o anomalías del nervio vago pueden provocar diversos problemas de salud, como trastornos gastrointestinales, problemas cardiovasculares y problemas respiratorios (de Theije y cols., 2011).

También se menciona el sistema inmunitario como vía de comunicación. La activación inmunitaria en el intestino puede influir en la función cerebral y el comportamiento a través de varios mecanismos. Puede afectar a los sistemas neurotransmisores, las vías neuroendocrinas y los circuitos neuronales implicados en la regulación del estado de ánimo, la respuesta al estrés y la cognición (Gareau, Sherman, y Walker, 2010). Además, la activación inmunitaria puede conducir a la producción de especies reactivas del oxígeno y mediadores inflamatorios que pueden dañar directamente las células cerebrales y contribuir a la neuroinflamación (Gareau y cols., 2010). A la inversa, el estrés y otros factores que activan el sistema nervioso central (SNC) también pueden afectar a la función intestinal y alterar la composición de la microbiota intestinal. Esta comunicación bidireccional entre el intestino y el cerebro pone de relieve la importancia del eje intestino-cerebro en el mantenimiento de la salud general y el papel potencial de la activación inmunitaria en este eje (Gareau y cols., 2010) (Mawe y cols., 2022).

Otra vía de comunicación entre el intestino con el cerebro es la producción de neurotransmisores y otras moléculas de señalización. Un importante neurotransmisor que interviene en el eje intestino-cerebro es la serotonina (Cryan y Dinan, 2012) (Mandal y cols., 2022). La serotonina se produce principalmente en el intestino, donde regula diversas funciones gastrointestinales, como la motilidad y la secreción (Joo y cols., 2023). La microbiota intestinal puede influir en la producción de serotonina a través del metabolismo del triptófano, un aminoácido precursor de la serotonina (Mandal y cols., 2022). Ciertas bacterias del intestino pueden producir enzimas que modifican el metabolismo del triptófano, afectando a los niveles de serotonina. Otro neurotransmisor implicado en el eje intestino-cerebro es el ácido gamma-aminobutírico (GABA) (Cryan y Dinan, 2012). El GABA es un neurotransmisor inhibitor que interviene en la regulación de la ansiedad y las respuestas al

estrés. La microbiota intestinal puede producir GABA a través de la actividad de bacterias específicas. El GABA producido en el intestino puede atravesar la barrera hematoencefálica e influir en la señalización GABAérgica en el cerebro (Gareau y cols., 2010). Además de neurotransmisores, la microbiota intestinal también puede producir e interactuar con diversas moléculas de señalización, como los ácidos grasos de cadena corta (AGCC). Los AGCC son producidos por la fermentación de las fibras alimentarias por las bacterias intestinales. Pueden actuar como moléculas de señalización e influir en la función de las células inmunitarias, las neuronas entéricas y la barrera hematoencefálica. Los AGCC están implicados en la regulación de la motilidad intestinal, la inflamación y la función cerebral (Cryan y Dinan, 2012) (Mandal y cols., 2022).

En un estudio, realizado por Kelly J. et al. (2016), los investigadores recogen muestras fecales de ratas diagnosticadas de depresión y de ratas sanas (grupo de control) sin depresión. Estas muestras fecales contienen las comunidades microbianas del tracto gastrointestinal de los participantes (Kelly y cols., 2016). Los investigadores realizan un trasplante de microbiota fecal (TFM), en el que transfieren esta microbiota recogida al tracto gastrointestinal de ratas. Este proceso les permite estudiar cómo la microbiota intestinal de individuos con depresión podría afectar al comportamiento y la neurobiología de las ratas receptoras. Los resultados del estudio se resumen en donde, tras recibir microbiota intestinal de individuos con depresión, las ratas receptoras muestran comportamientos "similares a la depresión". Estos comportamientos pueden incluir cambios en los niveles de actividad, motivación y estado de ánimo, similares a los síntomas observados en la depresión humana (Kelly y cols., 2016). Este estudio subraya el papel potencial del microbioma intestinal en la salud mental, especialmente en la depresión.

El estudio titulado "El potencial neuroactivo de la microbiota intestinal humana en la calidad de vida y la depresión", realizado por Valles-Colomer et al. en 2019 (Valles-Colomer y cols., 2019), es una contribución significativa al campo de la investigación del microbioma. Este estudio se propuso explorar la posible influencia de la microbiota intestinal en la calidad de vida de un individuo y su asociación con la depresión. Su objetivo era descubrir vínculos entre el microbioma intestinal y la salud mental. En la metodología de investigación se analizaron los datos de una gran cohorte de individuos, que incluían datos metagenómicos sobre el contenido genético de la microbiota intestinal, junto con amplia información clínica de los individuos. Esta información incluía evaluaciones de la calidad de vida y el diagnóstico de depresión. De especial interés en este estudio fue la evaluación de los atributos funcionales del microbioma, demostrando que el microbioma intestinal produce y degrada diversos compuestos neuroactivos, entre ellos neurotransmisores como la serotonina y el ácido gamma-aminobutírico (GABA) (Cryan y Dinan, 2012) (Valles-Colomer y cols., 2019). La producción de estos neurotransmisores por las bacterias intestinales se ha observado en cepas pertenecientes a distintos géneros asociados al intestino. Por ejemplo, se ha observado experimentalmente la síntesis de serotonina en cepas pertenecientes a diez géneros diferentes asociados al intestino, como Akkermansia, Alistipes y Roseburia. El microbioma intestinal también tiene el potencial de producir el metabolito de la dopamina ácido 3,4-dihidroxifenilacético, que se ha descubierto que se correlaciona positivamente con la calidad de vida mental (Valles-Colomer y cols., 2019).

En las conclusiones del estudio, los investigadores identificaron taxones microbianos específicos que estaban asociados a la calidad de vida de un individuo y a la presencia de depresión. Las bacterias Faecalibacterium y Coprococcus productoras de butirato se asociaron constantemente con indicadores de mayor calidad de vida. Junto con Dialister, Coprococcus spp., se vieron altamente relacionadas con la microbiota de pacientes con depresión, incluso después de realizar un análisis posterior a la administración de antidepresivos a estos pacientes (Valles-Colomer y cols., 2019). También destacaron el papel potencial de ciertas bacterias intestinales en la producción de compuestos neuroactivos que podrían influir en la salud mental.

Se ha demostrado también acerca del papel importante del microbiota humano en el estrés. En un estudio realizado por Foster y et al., publicado en 2017. El cual tuvo como objetivo investigar cómo el microbioma intestinal influye y regula la respuesta del organismo al estrés, en particular en lo que respecta a la función cerebral (Foster y cols., 2017). Los investigadores adoptaron un enfoque

multidisciplinar que incorporaba diversas metodologías. Entre ellas, el uso de modelos animales, evaluaciones neurobiológicas y análisis moleculares. Estos métodos les permitieron profundizar en los complejos mecanismos que intervienen en la regulación por el microbioma del eje intestino-cerebro en situaciones de estrés. Un descubrimiento clave es cómo la microbiota intestinal puede influir en el eje hipotalámico-hipofisario-suprarrenal (HPA), un sistema central responsable de regular la respuesta del organismo al estrés (Foster y cols., 2017). La desregulación de este eje se ha vinculado a trastornos relacionados con el estrés, como la ansiedad y la depresión. El estudio también sugiere que el microbioma intestinal se comunica con el cerebro a través de vías neuronales, inmunitarias y endocrinas. Esta comunicación puede influir significativamente en el estado de ánimo, el comportamiento y la función cognitiva, sobre todo durante los periodos de mayor estrés.

Estos estudios subrayan la importancia de reconocer el microbioma intestinal como una entidad dinámica capaz de influir en el bienestar mental. Abre nuevas vías de investigación y el posible desarrollo de intervenciones dirigidas a la microbiota intestinal para mejorar la calidad de vida y controlar la depresión. A medida que el campo de la investigación sobre el microbioma sigue evolucionando, estudios como éste desempeñan un papel vital para descubrir las complejidades del eje microbiota-intestino-cerebro y el potencial de los enfoques basados en el microbioma para abordar los problemas de salud mental.

Una forma de actuar sobre la microbiota intestinal con fines terapéuticos según Cryan et al. (Cryan y Dinan, 2012), es mediante el uso de probióticos. Los probióticos son microorganismos vivos similares a las bacterias beneficiosas que viven de forma natural en el intestino. Se ha demostrado que los probióticos tienen una serie de beneficios para la salud, como la mejora de la salud intestinal, el refuerzo del sistema inmunitario y la reducción de la inflamación.

Otra forma de actuar sobre la microbiota intestinal con fines terapéuticos es mediante el uso de prebióticos. Los prebióticos son fibras no digeribles que sirven de alimento a las bacterias beneficiosas del intestino. Se ha demostrado que los prebióticos aumentan el número de bacterias beneficiosas en el intestino y favorecen la salud intestinal. Además de los probióticos y los prebióticos, existen otras posibles estrategias terapéuticas dirigidas a la microbiota intestinal:

- Trasplante de microbiota fecal (TFM): El TFM es un procedimiento en el que las heces de una persona sana se trasplantan al intestino de una persona con un microbioma poco saludable. El TFM ha demostrado su eficacia en el tratamiento de ciertos trastornos intestinales, como la infección por *Clostridium difficile*.
- Psicobióticos: Los psicobióticos son una nueva clase de probióticos diseñados específicamente para la función cerebral y el comportamiento. Se ha demostrado que los psicobióticos tienen una serie de efectos beneficiosos sobre la salud mental, como la reducción de la ansiedad y la depresión.

(Cryan y Dinan, 2012)

## 4.4. Esquizofrenia

La esquizofrenia es un trastorno mental complejo que afecta a millones de personas en todo el mundo. Se trata de una enfermedad crónica y grave caracterizada por una alteración de los procesos de pensamiento, las emociones y la percepción de la realidad. La esquizofrenia es un trastorno psiquiátrico relativamente frecuente, con un impacto sustancial en la salud pública. Según la Organización Mundial de la Salud (OMS), se estima que afecta a 20 millones de personas en todo el mundo, con importantes variaciones regionales en la prevalencia (*Schizophrenia*, s.f.-a). Según los informes, la prevalencia de la esquizofrenia a lo largo de la vida es de alrededor del 0,3% al 0,7% de la población general (Saha, 2005).

Además, el inicio de la esquizofrenia suele producirse durante los primeros años de la edad adulta, y la mayoría de los individuos experimentan su primer episodio entre los 15 y los 30 años (Owen, 2016). Los hombres suelen desarrollar el trastorno a una edad más temprana que las mujeres, con un pico de edad de inicio para los hombres en torno a los 18-25 años, mientras que las mujeres tienden a experimentar el inicio entre los 25-35 años (Abel, Drake, y Goldstein, 2010).

La esquizofrenia tiene profundas implicaciones para los individuos afectados por el trastorno, así como para sus familias y la sociedad en general. Se considera una de las principales causas de discapacidad en todo el mundo (Schizophrenia, s.f.-b). El empleo se ve particularmente afectado, con investigaciones que indican que aproximadamente el 80% de los individuos con esquizofrenia están desempleados (Marwaha, 2016). Los síntomas cognitivos asociados con el trastorno, como las dificultades en la atención, la memoria y la resolución de problemas, contribuyen a reducir la productividad laboral y los desafíos vocacionales.

La esquizofrenia también supone una carga considerable para los sistemas y recursos sanitarios. Las personas con esquizofrenia suelen requerir atención psiquiátrica a largo plazo, lo que incluye hospitalizaciones, tratamientos ambulatorios y control de la medicación. Según un estudio realizado en Estados Unidos, los costes sanitarios directos asociados a la esquizofrenia ascienden a miles de millones de dólares anuales (Cloutier y cols., 2016).

Además, el estigma social que rodea a la esquizofrenia puede conducir a la discriminación, el aislamiento y la reducción del acceso a recursos y oportunidades esenciales. Las personas con esquizofrenia pueden enfrentar desafíos para mantener relaciones sociales, asegurar la vivienda y perseguir esfuerzos educativos, lo que afecta su calidad de vida en general (Schizophrenia, s.f.-b).

Los síntomas y los trastornos funcionales asociados con la esquizofrenia pueden alterar significativamente varios aspectos de la vida de un individuo, incluida la educación, el empleo, las relaciones y el funcionamiento diario. La esquizofrenia suele asociarse a una amplia gama de síntomas que pueden clasificarse en categorías positivas, negativas y cognitivas (Kahn y cols., 2015). Los síntomas positivos incluyen alucinaciones (auditivas o visuales), delirios y pensamiento o habla desorganizados. Los síntomas negativos implican una disminución o pérdida del funcionamiento normal, como la disminución de la expresión emocional, la reducción de la motivación y el retraimiento social. Los síntomas cognitivos pueden incluir dificultades de atención, memoria y funciones ejecutivas.

Además, la esquizofrenia se clasifica en diferentes subtipos, como paranoide, desorganizada, catatónica, indiferenciada y residual, en función de la naturaleza y gravedad de los síntomas; los cuales, por el fin de la investigación no serán discutidos en el presente documento.

#### 4.4.1. Diagnóstico

El diagnóstico de la esquizofrenia implica una evaluación exhaustiva que considera una combinación de entrevistas clínicas, observación de los síntomas y, en algunos casos, pruebas complementarias para descartar otras afecciones médicas (Shabani y cols., 2021). Es importante tener en cuenta que el diagnóstico de la esquizofrenia suele ser realizado por profesionales de la salud mental cualificados, como psiquiatras o psicólogos clínicos. Los métodos comunes son los siguientes:

- Entrevista clínica estructurada para el DSM (SCID): Se trata de una herramienta ampliamente utilizada para diagnosticar trastornos psiquiátricos, incluida la esquizofrenia. Consiste en un conjunto sistemático de preguntas para evaluar la presencia y gravedad de los síntomas basándose en los criterios del *Manual Diagnóstico y Estadístico de los Trastornos Mentales* (DSM-5).
- Observación de síntomas positivos y negativos: La esquizofrenia se caracteriza por síntomas positivos (p. ej., alucinaciones, delirios, pensamiento desorganizado) y síntomas negativos (p.

ej., retraimiento social, expresión emocional reducida, déficits cognitivos). Un clínico observará y evaluará la presencia y gravedad de estos síntomas.

- Para ser diagnosticado de esquizofrenia, los síntomas deben persistir durante un período de tiempo significativo, normalmente seis meses o más. Esto ayuda a distinguirla de otros trastornos psicóticos transitorios.
- Pruebas médicas: Para descartar otras afecciones médicas que pueden presentarse con síntomas psicóticos, los médicos pueden realizar análisis de sangre, imágenes cerebrales (resonancia magnética o tomografía computarizada) y exámenes neurológicos.
- Exclusión del trastorno psicótico inducido por sustancias: Los clínicos explorarán si el consumo o la abstinencia de sustancias pueden estar contribuyendo a los síntomas.

(APA, 2013)

Es crucial tener en cuenta que el proceso de diagnóstico es complejo y requiere la experiencia de profesionales de la salud mental. Además, se tienen en cuenta factores culturales y contextuales para garantizar un diagnóstico preciso y culturalmente sensible.

#### 4.4.2. Esquizofrenia activa

La esquizofrenia activa se refiere a la fase aguda del trastorno, caracterizada por la presencia de síntomas psicóticos prominentes. La esquizofrenia activa se caracteriza por una serie de síntomas que alteran gravemente el pensamiento, la percepción y el comportamiento del individuo. Los síntomas positivos, como las alucinaciones (por ejemplo, oír voces), los delirios (por ejemplo, creencias falsas) y el habla o el comportamiento desorganizados, son prominentes durante la fase activa. Estos síntomas contribuyen a un sentido distorsionado de la realidad y pueden provocar una profunda confusión y angustia en las personas afectadas. Los síntomas negativos, como la reducción de la expresión emocional, el retraimiento social y la disminución de la motivación, también pueden estar presentes, deteriorando aún más el funcionamiento y las relaciones interpersonales (Kahn y cols., 2015).

#### 4.4.3. Esquizofrenia pasiva

La esquizofrenia pasiva, también conocida como esquizofrenia deficitaria o esquizofrenia negativa, se refiere a un subtipo del trastorno caracterizado por síntomas negativos predominantes y una expresión disminuida de las emociones. Este subtipo se distingue principalmente por la presencia de síntomas negativos, que implican la ausencia o reducción de comportamientos y funciones normales. Estos síntomas incluyen una expresión emocional disminuida, retraimiento social, motivación reducida y pobreza del habla. Los individuos con esquizofrenia pasiva pueden mostrar una falta general de interés o placer en las actividades cotidianas y tener dificultades para iniciar y mantener relaciones sociales. También pueden estar presentes alteraciones cognitivas, como disminución de la atención y dificultades de memoria (Kahn y cols., 2015).

#### 4.4.4. Vías neuronales en la esquizofrenia

La esquizofrenia implica disfunciones en varias vías neuronales que afectan la comunicación entre las células cerebrales. Entre estas, la dopaminérgica es particularmente relevante (Howes y Kapur, 2009). Se postula que un desequilibrio en la liberación de dopamina en ciertas áreas del cerebro contribuye a los síntomas positivos de la esquizofrenia, como las alucinaciones y delirios. Además, la glutamatérgica, serotoninérgica y colinérgica también juegan un papel crucial en la regulación del

humor, la cognición y la percepción, y sus alteraciones se asocian con distintos aspectos de la esquizofrenia (Javitt, 2007). Es importante destacar que la esquizofrenia es una enfermedad heterogénea, y la variabilidad en los síntomas puede deberse a la complejidad de las interacciones en estas vías.

La hiperactividad en la vía dopaminérgica, especialmente en la vía mesolímbica, se asocia con síntomas positivos de la esquizofrenia, como alucinaciones y delirios. Se cree que la liberación excesiva de dopamina en ciertas áreas del cerebro contribuye a la aparición de estos síntomas (Javitt, 2007). La glutamina es un neurotransmisor clave en la vía glutamatérgica, que desempeña un papel crucial en la regulación de la cognición y la percepción. Alteraciones en esta vía se asocian con síntomas negativos y cognitivos de la esquizofrenia (Javitt, 2007). La serotonina está involucrada en la regulación del estado de ánimo y la respuesta al estrés. Alteraciones en la vía serotoninérgica se asocian con síntomas afectivos y cognitivos en la esquizofrenia. La acetilcolina es un neurotransmisor que juega un papel en la función cognitiva y la atención. Alteraciones en la vía colinérgica se asocian con déficits cognitivos en la esquizofrenia (Javitt, 2007).

La microbiota intestinal desempeña un papel en el metabolismo y puede afectar la producción de neurotransmisores y la inflamación en el cuerpo (Cryan y Dinan, 2012). Se ha observado que ciertos metabolitos microbianos pueden cruzar la barrera hematoencefálica y afectar la función cerebral. A través de la modulación del metabolismo, la microbiota podría influir en las vías neuronales asociadas con la esquizofrenia (Foster y cols., 2017). Por ejemplo, cambios en la composición bacteriana podrían afectar la producción de compuestos que impactan la actividad neuronal, inflamación y la liberación de neurotransmisores clave. Como lo es la vía de comunicación que afecta al GABA, la modulación de neurotransmisores, anteriormente mencionados, y como lo es que algunos metabolitos microbianos pueden atravesar la barrera hematoencefálica, la cual protege al cerebro de sustancias potencialmente dañinas. Una vez en el cerebro, estos metabolitos pueden influir en la función neuronal (Sherwin, Sandhu, Dinan, y Cryan, 2016).

Es importante destacar que la relación entre la microbiota y la esquizofrenia es un campo de investigación en evolución, y se necesitan más estudios para comprender completamente cómo estos dos sistemas interactúan.

## 4.5. Tratamientos existentes

La esquizofrenia (SCZ) es un trastorno mental crónico que afecta a la percepción, el pensamiento y las emociones de las personas. Los tratamientos para la esquizofrenia suelen combinar enfoques farmacológicos y terapéuticos, y pueden variar según la gravedad y las necesidades individuales del paciente ((APA), 2019).

Dentro de los enfoques terapéuticos, para el tratamiento de SCZ, se encuentran las terapias psicosociales. Estas terapias pueden ayudar a los pacientes a lidiar con el estrés, mejorar sus habilidades sociales y fomentar la adherencia al tratamiento. Algunos enfoques terapéuticos comunes incluyen la terapia cognitivo-conductual (TCC), la es una forma de terapia psicológica que se enfoca en abordar los patrones de pensamiento negativos y las conductas problemáticas que pueden contribuir a los trastornos mentales y emocionales. Fue desarrollada en la década de 1960 por el psiquiatra Aaron T. Beck y el psicólogo Albert Ellis (Wykes, Steel, Everitt, y Tarrier, 2008); la terapia de apoyo cuyo enfoque terapéutico se centra en proporcionar apoyo emocional, comprensión y empatía a las personas que enfrentan dificultades emocionales o situaciones estresantes (Zachariae y cols., 2003); terapia de habilidades sociales, la cual es una forma de intervención psicológica que se enfoca en mejorar las habilidades de comunicación, interacción y comportamiento social de una persona (Bellack y Hersen, 2014); y terapias complementarias que puedan beneficiar al paciente y brindarle apoyo emocional a través de los intereses individuales como lo puede ser la música, el arte y las manualidades.

En complementario a este tipo de terapias, a los pacientes con SCZ se recomienda seguir con

programas de rehabilitación psicosocial, el cual puede ayudar a los pacientes con esquizofrenia a mejorar su capacidad para funcionar en la comunidad y en el trabajo. Estos programas pueden ofrecer capacitación laboral, servicios de vivienda supervisada y apoyo para la reintegración social (of Mental Health (NIMH), 2020).

La intervención farmacológica en el tratamiento de la esquizofrenia se centra en antipsicóticos, que buscan controlar los síntomas al influir en la actividad de diversos neurotransmisores cerebrales. Si bien la dopamina ha sido identificada como un objetivo crucial, no es el único neurotransmisor afectado. Los antipsicóticos, tanto típicos como atípicos, están diseñados para modular la transmisión sináptica de neurotransmisores clave, como la serotonina, la glutamato y la noradrenalina (of Mental Health (NIMH), 2020). La compleja interacción entre estos neurotransmisores desempeña un papel significativo en la regulación del estado de ánimo, la percepción y la cognición, aspectos que se ven afectados en la esquizofrenia. Hay dos categorías principales de antipsicóticos: los antipsicóticos típicos (clásicos) y los antipsicóticos atípicos (segunda generación) (of Mental Health (NIMH), 2020).

Entre los antipsicóticos típicos nos encontramos con Haloperidol (Haldol), Clorpromazina (Thorazine), Flufenazina (Prolixin), Perfenazina (Trilafon) y Trifluoperazina (Stelazine). Estos medicamentos funcionan, a gran escala, a través del bloqueo de receptores de dopamina y pueden ser utilizados para tratar otros trastornos psicóticos, como lo puede ser el trastorno bipolar. Estos antipsicóticos pueden tener efectos secundarios significativos, como movimientos involuntarios (discinesia tardía), rigidez muscular y síntomas extrapiramidales, lo que ha llevado a la preferencia por los antipsicóticos atípicos en muchas situaciones clínicas (Goodman y Brunton, 2018).

Por otro lado, los antipsicóticos atípicos nos encontramos con Risperidona (Risperdal), Olanzapina (Zyprexa), Quetiapina (Seroquel), Ziprasidona (Geodon), Aripiprazol (Abilify), Clozapina (Clozaril) y Lurasidona (Latuda). A diferencia de los antipsicóticos típicos, esta clase de fármacos en su mecanismo de acción bloquea no solo los receptores de dopamina, sino que también hace un bloqueo a los receptores de serotonina. A pesar que los antipsicóticos atípicos generalmente tienen un perfil de efectos secundarios más favorable que los antipsicóticos típicos, todavía pueden tener efectos secundarios significativos. Cada persona puede responder de manera diferente a estos medicamentos (Goodman y Brunton, 2018).

Es importante destacar que los tratamientos para la esquizofrenia en Guatemala, al igual que en otros países, deben ser prescritos y supervisados por profesionales de la salud mental, como psiquiatras y psicólogos, y se basan en la evaluación individual del paciente y la gravedad de los síntomas (García Muñoz, 2014). Los tratamientos populares en Guatemala para la esquizofrenia pueden incluir los encontrados en la Tabla 1 encontrada a continuación.

**Tabla 4.1:** Medicamentos recetados en Guatemala con sus efectos adversos

Medicamento	Efectos adversos
Haloperidol	Síndrome de déficit inducido por neuroléptico, acatisia, síntomas extrapiramidales, parkinsonismo, discinesia tardía, galactorrea, amenorrea, mareos, sedación, boca seca, constipación, retención urinaria, visión borrosa, sudoración disminuida, hipotensión, taquicardia, hipertensión, aumento de peso.
Clorpromazina	Acatisia, priapismo, parkinsonismo, discinesias tardías, sedación, galactorrea, amenorrea, síndrome neuroléptico maligno, agranulocitosis y especial atención a pacientes de la tercera edad ya que incrementa el riesgo de muerte y eventos cerebrovasculares.
Tioridazina	Acatisia, priapismo, parkinsonismo, discinesias tardías, sedación, galactorrea, amenorrea, agranulocitosis, síndrome neuroléptico maligno e incluso se han reportado casos de retinitis pigmentaria con dosis altas de tratamiento.

Medicamento	Efectos adversos
Pimozida	Entre los principales efectos adversos están, acatisia, efectos extrapiramidales, parkinsonismo, discinesia tardías, hipotensión y ganancia de peso.
Palmitato de pipotiazina	Excitación, insomnio, inquietud, acatisia, parkinsonismo, discinesias tardías, sedación, galactorrea, amenorrea, agranulocitosis, síndrome neuroléptico maligno y distonías tardías.
Risperidona	Aumenta significativamente el riesgo de diabetes y dislipidemias. A pesar de ser un antipsicótico atípico pueden presentarse signos extrapiramidales e hiperprolactinemia.
Ziprasidona	Mareos, síntomas extrapiramidales, sedación, náusea, boca seca, astenia, erupciones en la piel, discinesia tardía e hipotensión ortostática. Entre sus efectos secundarios más severos se encuentra el síndrome neuroléptico maligno, convulsiones y aumenta la posibilidad de eventos cerebrovasculares.
Aripiprazol	Entre los principales efectos adversos están el insomnio, agitación, ansiedad, cefalea, vértigo, acatisia, somnolencia/sedación, temblor, síntomas extrapiramidales, visión borrosa; náuseas, vómitos, dispepsia, estreñimiento, hipersecreción salivar, fatiga, taquicardia e hipotensión ortostática.
Paliperidona	Infección de vías respiratorias superiores, infección del tracto urinario, insomnio, agitación, depresión, ansiedad, aumento de peso, disminución de peso, cefalea, parkinsonismo, acatisia, sedación o somnolencia, disfonía, mareos, discinesia, temblor, bradicardia, taquicardia, trastornos de la conducción, prolongación del intervalo QT en el ECG, hipertensión arterial, tos, congestión nasal, dolor abdominal, vómitos, náuseas, estreñimiento, diarrea.
Clozapina	Aumenta el riesgo de diabetes y dislipidemia, pueden presentar mareos, náuseas, cefalea, taquicardia, sudoración e hipotensión. La salivación puede ser un efecto muy severo, al igual que el aumento de peso y la sedación.

(García Muñoz, 2014)

#### 4.6. Estado del arte de estudios genéticos enfocados en pacientes esquizofrénicos y la microbiota

La intersección de estos dos campos, la epigenética y la microbiota, ha dado lugar a una creciente exploración de cómo las modificaciones epigenéticas pueden estar relacionadas con la microbiota y, en particular, cómo esta interacción podría estar implicada en la patogénesis de la esquizofrenia (Dinan y Cryan, 2017).

Cada vez hay más pruebas de que el eje microbioma-intestino-cerebro puede desempeñar un papel en el desarrollo o la progresión de la esquizofrenia.

Dinan y Cryan (2017) revisan las pruebas de que la microbiota intestinal está alterada en las personas con esquizofrenia. Descubrieron que las personas con esquizofrenia tienen niveles más bajos de ciertas bacterias beneficiosas, como *Bifidobacterium* y *Akkermansia* (Dinan y Cryan, 2017). También descubrieron, en concordancia por estudio realizado por Nguyen et al., que las personas con esquizofrenia tienen niveles más altos de ciertas bacterias dañinas, como *Proteobacteria* (Nguyen y cols., 2019).

En un estudio realizado por Severance et al. (2015) se analizan los posibles mecanismos por los que el eje microbioma-intestino-cerebro puede contribuir a la esquizofrenia. Sugieren que la microbiota intestinal puede influir en el desarrollo y la función cerebral a través de diversas vías, incluida la producción de neurotransmisores, la inflamación y el sistema inmunitario (Severance, Prandovszky, Castiglione, y Yolken, 2015).

Khavari et al. (2020) descubrieron que existían diferencias significativas en los perfiles epigenéticos del microbioma intestinal de personas con esquizofrenia y controles sanos. También descubrieron que estos cambios epigenéticos estaban asociados a cambios en la composición del microbioma intestinal (Khavari y Cairns, 2020).

En el estudio realizado por los investigadores se descubrió que:

- Las personas con esquizofrenia tenían niveles más altos de metilación del ADN en ciertos genes que intervienen en la regulación del microbioma intestinal.
- Las personas con esquizofrenia tenían niveles diferentes de ciertos microARN, que son pequeñas moléculas que pueden regular la expresión génica.
- Los cambios epigenéticos en el microbioma intestinal estaban asociados a cambios en la composición del microbioma intestinal. Por ejemplo, las personas con esquizofrenia tenían niveles más bajos de ciertas bacterias beneficiosas, como Bifidobacterium y Akkermansia.

(Khavari y Cairns, 2020)

Estos hallazgos fueron encontrados también por Munawar et al., en el estudio realizado por estos investigadores se reclutaron a 100 participantes con esquizofrenia y 100 controles sanos. Recogieron muestras de heces de todos los participantes y extrajeron ADN de la microbiota intestinal. Utilizaron una técnica llamada secuenciación de bisulfito del genoma completo para medir los patrones de metilación del ADN de la microbiota intestinal (Munawar, Ahsan, Muhammad, y et al., 2021).

Los investigadores descubrieron que había diferencias significativas en los patrones de metilación del ADN de la microbiota intestinal entre las personas con esquizofrenia y los controles sanos. Identificaron 51 regiones metiladas diferencialmente (DMR) en la microbiota intestinal de las personas con esquizofrenia. Estos DMR de la microbiota intestinal se vieron asociados a cambios en la composición del microbioma intestinal. Por ejemplo, los DMR se asociaron con niveles más bajos de ciertas bacterias beneficiosas, como Bifidobacterium y Akkermansia (Munawar y cols., 2021).

Los investigadores concluyeron que la epigenética puede desempeñar un papel en la regulación del microbioma intestinal en personas con esquizofrenia. Este hallazgo tiene implicaciones para el desarrollo de nuevas herramientas de diagnóstico y tratamientos del trastorno.

Estos biomarcadores intestinales pueden ser usados para diagnosticar la esquizofrenia o predecir el riesgo de desarrollar el trastorno. Además, los investigadores sugieren que podrían utilizarse terapias epigenéticas dirigidas al microbioma intestinal para tratar la esquizofrenia (Munawar y cols., 2021) (Dinan y Cryan, 2017).

Los investigadores también descubrieron que las DMR de la microbiota intestinal estaban asociadas a cambios en la composición del microbioma intestinal. Por ejemplo, los DMR se asociaron con niveles más bajos de ciertas bacterias beneficiosas, como Bifidobacterium y Akkermansia (Munawar y cols., 2021).

Estos estudios tienen varias implicaciones importantes. En primer lugar, sugiere que la microbiota intestinal puede desempeñar un papel en la esquizofrenia. En segundo lugar, sugiere que los patrones de metilación del ADN de la microbiota intestinal podrían utilizarse como biomarcadores para diagnosticar la esquizofrenia o predecir el riesgo de desarrollar el trastorno. En tercer lugar,

sugiere que podrían utilizarse terapias epigenéticas dirigidas a la microbiota intestinal para tratar la esquizofrenia.

Por otro lado Nemani et al. (2015) revisan las pruebas de que la desregulación epigenética del eje microbiota intestinal-intestino-cerebro puede desempeñar un papel en el desarrollo o la progresión de la esquizofrenia. Analizan los siguientes mecanismos:

- Cambios en la modificación de histonas: La modificación de las histonas es otro tipo de modificación epigenética que puede alterar la expresión génica. Nemani et al. (2015) discuten la evidencia de que los patrones de modificación de histonas de la microbiota intestinal también están alterados en personas con esquizofrenia. Estas alteraciones también pueden estar implicadas en el desarrollo o la progresión del trastorno.
- Cambios en la expresión de microARN: Los microARN son pequeñas moléculas que pueden regular la expresión génica. Nemani et al. (2015) discuten la evidencia de que los patrones de expresión de microARN de la microbiota intestinal también están alterados en personas con esquizofrenia. Estas alteraciones también pueden estar implicadas en el desarrollo o la progresión del trastorno.

(Nemani, Hosseini Ghomi, McCormick, y Fan, 2015)

#### 4.6.1. Implicaciones para la investigación y el tratamiento futuros

La investigación sobre el eje microbioma-intestino-cerebro en la esquizofrenia está aún en sus primeras fases, pero tiene el potencial de conducir a nuevas herramientas de diagnóstico y tratamientos para el trastorno.

Por ejemplo, estudios recientes, como el llevado a cabo por Hoffman Lee, et al. (2020) (Hoffman y cols., 2020), sugieren que los cambios epigenéticos en la microbiota podrían ser utilizados como biomarcadores para diagnosticar la esquizofrenia o incluso prever el riesgo de desarrollarla. Esta conexión entre la microbiota y la epigenética abre nuevas posibilidades para la identificación temprana y estrategias de intervención personalizadas. Adicionalmente, la exploración de terapias epigenéticas dirigidas a la microbiota emerge como un campo prometedor en el tratamiento de la esquizofrenia. La manipulación de la microbiota a través de intervenciones específicas podría no solo abordar los síntomas, sino también influir positivamente en los patrones epigenéticos, proporcionando un enfoque integral y novedoso en la gestión de esta compleja enfermedad mental (Hoffman y cols., 2020).

Se necesita más investigación para confirmar estos hallazgos y comprender los mecanismos exactos por los que el eje microbioma-intestino-cerebro contribuye a la esquizofrenia. Sin embargo, esta investigación tiene el potencial de conducir a formas nuevas e innovadoras de diagnosticar y tratar este complejo trastorno.

Se están estudiando una serie de posibles intervenciones terapéuticas dirigidas al eje microbioma-intestino-cerebro en la esquizofrenia. Entre ellas se incluyen

- Probióticos y prebióticos
- Psicobióticos
- Trasplante de microbiota fecal

(Hoffman y cols., 2020)

Además de las posibles intervenciones terapéuticas mencionadas anteriormente, las personas con esquizofrenia pueden hacer otras cosas para mejorar su salud intestinal y su bienestar general. Entre ellas se incluyen

- **Seguir una dieta sana:** Una dieta sana es esencial para la salud intestinal. Debe incluir una variedad de alimentos integrales y no procesados, como frutas, verduras, cereales integrales y proteínas magras. También es importante beber mucho líquido, como agua y té sin azúcar.
- **Controlar el estrés:** El estrés puede alterar la microbiota intestinal. Encontrar formas saludables de controlar el estrés, como el ejercicio, las técnicas de relajación y pasar tiempo en la naturaleza, puede ayudar a mejorar la salud intestinal.
- **Dormir lo suficiente:** El sueño es esencial para la salud y el bienestar general. También es importante para la salud intestinal. Duerma entre 7 y 8 horas cada noche.

(Hoffman y cols., 2020)

Se necesitan más investigaciones para determinar la seguridad y eficacia de estas intervenciones para el tratamiento de la esquizofrenia. Sin embargo, los resultados de los primeros estudios son prometedores y sugieren que actuar sobre el eje microbioma-intestino-cerebro puede ser una forma nueva y eficaz de tratar este trastorno.

## 4.7. Implicaciones éticas de la investigación de microbioma

El estudio del microbioma humano, que se refiere a la diversa comunidad de microorganismos que viven dentro y sobre nuestro cuerpo, plantea una serie de implicaciones éticas; tomando en cuenta que cual sea el objetivo de la investigación, estas implicaciones éticas siempre seguirán existiendo. Estas implicaciones giran en torno a cuestiones de consentimiento, privacidad, intercambio de datos, posible uso indebido de los hallazgos y acceso equitativo a los beneficios, entre otras.

En la ética de la investigación se busca siempre cumplir con los 4 principios base los cuales son:

1. **Autonomía:** Este principio se refiere al respeto por la capacidad de las personas para tomar decisiones informadas y voluntarias sobre su participación en la investigación. Implica obtener el consentimiento informado de los participantes y asegurarse de que comprendan plenamente los riesgos y beneficios del estudio.
2. **Beneficencia:** La beneficencia implica la obligación de maximizar los beneficios y minimizar los riesgos para los participantes de la investigación. Los investigadores deben diseñar estudios de manera que los posibles beneficios superen cualquier daño potencial y deben asegurarse de que los participantes reciban un trato ético y justo.
3. **Justicia:** La justicia en la investigación se refiere a la distribución equitativa de los beneficios y las cargas de la investigación. Esto significa que la selección de participantes debe ser justa y que los grupos vulnerables no deben ser explotados o excluidos de manera injusta. También se relaciona con el acceso equitativo a los resultados de la investigación.
4. **No maleficencia:** Este principio implica la obligación de no causar daño intencional a los participantes de la investigación. Los investigadores deben tomar medidas para prevenir o minimizar cualquier daño físico, emocional o psicológico que pueda resultar de la investigación. También se relaciona con la identificación y mitigación de riesgos potenciales.

(Schrag, 2011)

Dados estos 4 principios es importante asegurarse que en la realización de estudios de esta naturaleza que exista un consentimiento informado en donde los participantes del estudio deben de estar plenamente informados de la finalidad, los riesgos y los posibles beneficios de la investigación. Debido a la naturaleza íntima de la investigación, es crucial que los procesos de consentimiento sean claros y comprensibles. Los participantes deben comprender que sus datos microbianos podrían utilizarse para diversos análisis y deben tener derecho a retirar sus datos en cualquier momento si así desean (Touitou, Portaluppi, Smolensky, y Rensing, 2004).

Es de importancia añadir que la investigación del microbioma implica la recogida y el análisis de datos muy personales. Se deben de tomar las medidas necesarias para proteger la privacidad de los participantes y tener cuidado en desidentificar los datos, eliminando o encriptando cualquier identificador personal (Touitou y cols., 2004). Sin embargo, debido a la singularidad de los perfiles microbiómicos, siempre existe el riesgo de la reidentificación, de lo cuál debe de ser parte del consentimiento informado que deben de firmar los participantes. Este anonimato es esencial ya que puede existir una estigmatización de los individuos. La investigación del microbioma podría revelar información sensible sobre el estado de salud de un individuo, incluida la predisposición a ciertas enfermedades o afecciones. Esta información podría dar lugar a discriminación o estigmatización de los participantes. Las directrices éticas establecidas deben abordar el tratamiento responsable de dicha información (Schrag, 2011).

Sin embargo, se debe de tomar en cuenta que se necesita buscar la manera para compartir los datos de la investigación sin afectar la privacidad de los participantes. Compartir datos sobre el microbioma es esencial para el progreso científico, pero también conlleva riesgos para la privacidad (Touitou y cols., 2004). Los investigadores deben encontrar un equilibrio entre la promoción del intercambio abierto de datos y la protección de la privacidad de los participantes. Esto puede llegarse a hacer a través del establecimiento de directrices claras sobre el intercambio de datos, las técnicas de anonimización y el alcance del uso de los datos (Touitou y cols., 2004).

Dadas las diferentes enfermedades que se atribuyen al microbioma intestinal, los estudios realizados de esta naturaleza puede llevar a una comercialización y desarrollo de patentes ya sea para la aplicación de nuevas terapias o productos. Creando así un interés en las empresas privadas para promover este tipo de estudios (Touitou y cols., 2004). En la investigación de debe de lograr un equilibrio entre el avance de la ciencia y los intereses comerciales para no crear una imparcialidad o maleficiencia por parte de los investigadores.

Esta comercialización o creación de productos lleva al punto de la importancia de una distribución equitativa de los beneficios a una alta gama de personas, incluyendo poblaciones vulnerables (Schrag, 2011). Destacando así que se debe de evitar el acceso exclusivo a los beneficios que puede desarrollar la investigación para prevenir disparidades sanitarias.

Finalmente, es esencial que se mantenga una integridad en la investigación e informar con precisión sobre los métodos, los resultados y los posibles conflictos de intereses. La tergiversación o la información selectiva pueden tener importantes implicaciones éticas y repercusiones en estudios posteriores acerca del mismo tema que se esta investigando (Touitou y cols., 2004). En los estudios de microbiomas es importante tomar en cuenta que existe una diversidad global, tanto de poblaciones como de dietas; y por ende, diversidad en microbiomas. Los estudios realizados principalmente en una población pueden no ser representativos de otras. Los investigadores deben aspirar a la inclusión y tener en cuenta las implicaciones culturales de su trabajo y dejar en sus resultados la transparencia del origen de las muestras utilizadas (Ma, Chen, Lan, y Ren, 2018).

Para abordar estas consideraciones éticas es necesaria la colaboración entre investigadores, comités de ética, responsables políticos y el público en general. A medida que avanza la investigación sobre el microbioma (Ma y cols., 2018), las directrices y normativas éticas deben adaptarse para garantizar que los beneficios de este campo se materialicen sin sacrificar los derechos, la dignidad y el bienestar de las personas implicadas (Touitou y cols., 2004).

## 4.8. Herramientas para el análisis de secuencias

La metagenómica aplicada y el análisis del microbioma han revolucionado nuestra comprensión de las complejas comunidades microbianas que habitan en diversos entornos, desde el cuerpo humano hasta los ecosistemas. Estas técnicas permiten estudiar la composición genética, el potencial funcional y las interacciones de las comunidades microbianas, arrojando luz sobre su papel en la salud, la enfermedad, la ecología y la biotecnología. Para aprovechar la gran cantidad de información que contienen estas comunidades, ha surgido un conjunto de potentes herramientas y tecnologías.

En el presente proyecto, dado que se tiene el objetivo de generar un programa que automatice la identificación de microbioma de secuencias, en específico de secuencias de pacientes esquizófrénicos, se toman en cuenta las siguientes herramientas:

### 4.8.1. SeqKit

Seqkit es una herramienta de línea de comandos y una biblioteca en lenguaje Go diseñada para manipular secuencias biológicas de forma eficiente. Se utiliza para realizar diversas tareas relacionadas con el procesamiento y análisis de datos de secuenciación, incluyendo la extracción de estadísticas, la conversión de formatos de archivo, la filtración y el muestreo de secuencias, entre otras cosas (Shen, Le, Li, y Hu, 2016).

Algunas de las funciones que seqkit puede realizar incluyen:

- Extracción de estadísticas: seqkit puede generar estadísticas sobre secuencias, como longitud promedio, longitud máxima, longitud mínima, GC contenido, etc.
- Conversión de formatos: Puede convertir secuencias entre varios formatos de archivo comunes, como FASTA, FASTQ y más.
- Filtración y muestreo: seqkit permite filtrar secuencias por longitud, calidad, contenido de bases, y también puede realizar muestreo aleatorio de secuencias.
- Manipulación de secuencias: Puede realizar diversas operaciones en secuencias, como la extracción de subsecuencias, la concatenación de secuencias y la inversión de secuencias.
- Manipulación de archivos multi-FASTA: Puede dividir y fusionar archivos multi-FASTA, facilitando la gestión de grandes conjuntos de datos.

(Shen y cols., 2016)

Dado que seqkit es una herramienta de línea de comandos, es muy útil para automatizar tareas de procesamiento de datos en análisis de secuenciación. Por lo que en el presente proyecto es utilizado como herramienta preliminar para conocer las características generales de las secuencias a tratar.

### 4.8.2. FastQC

FastQC es una herramienta de análisis de calidad para datos de secuenciación de ADN de alta velocidad, como los datos generados por secuenciadores de próxima generación (NGS) como Illumina. Su objetivo principal es proporcionar una evaluación rápida y visual de la calidad de los datos de secuenciación, lo que es esencial para tomar decisiones informadas durante el análisis posterior.

FastQC se utiliza comúnmente para realizar las siguientes tareas:

- Evaluación de calidad de lecturas: FastQC examina la calidad de las lecturas de secuencia a lo largo de sus longitudes. Esto incluye la detección de posibles problemas como bases de baja calidad, adaptadores residuales, regiones con alta variabilidad de calidad, contenido de bases, entre otros.
- Detección de contaminantes: La herramienta puede ayudar a identificar posibles contaminantes en los datos de secuenciación. Esto puede incluir secuencias provenientes de fuentes no deseadas o de contaminación cruzada.
- Análisis de contenido de bases y GC: FastQC proporciona gráficos que muestran la distribución de las bases nucleotídicas a lo largo de las lecturas y el contenido de GC. Esto es útil para detectar problemas potenciales en la preparación de la muestra.
- Identificación de adapters: Puede detectar la presencia de secuencias adaptadoras o secuencias no deseadas en las lecturas. Esto es especialmente importante después de procesar los datos de secuenciación.
- Generación de informes: FastQC genera informes interactivos y gráficos que pueden ayudar a los investigadores a comprender la calidad de los datos y a tomar decisiones sobre el procesamiento posterior.

(*Babraham Bioinformatics—FastQC A Quality Control tool for High Throughput Sequence Data*, s.f.)

FastQC se ejecuta desde la línea de comandos y es ampliamente utilizado en la comunidad de análisis de secuenciación para realizar una evaluación inicial de la calidad de los datos antes de continuar con pasos posteriores, como el recorte de adaptadores, la eliminación de bases de baja calidad y el ensamblaje de genomas.

### 4.8.3. Fastp

Fastp es otra herramienta de análisis y procesamiento de datos de secuenciación de ADN de alta velocidad, específicamente diseñada para datos generados por secuenciadores de próxima generación (NGS) como Illumina. Al igual que FastQC, fastp se utiliza para realizar tareas de control de calidad y preprocesamiento de datos antes de análisis más avanzados.

Sin embargo, esto no significa que sean independientes unos con otros. Fastqc genera los reportes más detallados acerca del control de calidad de las lecturas. Mientras que con fastp se realizan las tareas complementarias de:

- Recorte de adaptadores: La herramienta es capaz de detectar y eliminar secuencias adaptadoras y regiones de baja calidad al final de las lecturas. Esto es importante para garantizar que las secuencias de interés se conserven.
- Filtración de lecturas: Fastp permite filtrar las lecturas según criterios como la longitud mínima y la calidad mínima. Esto ayuda a eliminar lecturas de baja calidad o artefactos.
- Generación de reportes: Al igual que FastQC, fastp produce informes y gráficos interactivos que muestran estadísticas sobre la calidad y el procesamiento de las lecturas, también como generar carpetas por aparte acerca de las lecturas que fueron descartadas dado que no cumplía con el estándar de calidad.

(S. Chen y cols., 2018)

Además, algo que diferencia a Fastp, es su beneficios de tener una capacidad para realizar varias etapas de procesamiento en un solo paso y su enfoque en el rendimiento lo convierten en una opción valiosa para garantizar que los datos brutos sean de alta calidad antes de llevar a cabo análisis más avanzados, como el ensamblaje, la alineación y la anotación. Permitiendo trabajos en diferentes hilos y dando lugar al paralelismo para acortar tiempos de trabajo en el análisis de filtrado en genomas.

#### 4.8.4. Clasificación de taxonomía

Las herramientas de clasificación taxonómica en metagenómica son herramientas informáticas diseñadas para asignar secuencias de ADN o ARN a taxones específicos, como especies, géneros y familias, en el contexto de comunidades microbianas complejas.

Estas herramientas son esenciales para comprender la composición y la diversidad de los microbiomas presentes en diferentes entornos. Utilizando bases de datos de referencia y algoritmos de comparación, estas herramientas identifican las similitudes entre las secuencias desconocidas y las secuencias ya clasificadas, permitiendo a los investigadores inferir la presencia y abundancia de microorganismos en una muestra. Ejemplos de estas herramientas incluyen Kaiju, Kraken, y MetaPhlAn, que ofrecen enfoques diversos para abordar la clasificación taxonómica y facilitar la interpretación de los datos metagenómicos.

#### Kaiju

Kaiju es una herramienta de análisis taxonómico basada en secuencias que se utiliza en genómica comparativa y metagenómica para asignar secuencias de ADN o ARN a taxones específicos, como especies, géneros y familias. Su objetivo principal es la identificación rápida y precisa de la procedencia taxonómica de secuencias, especialmente en el contexto de grandes conjuntos de datos de secuenciación.

Algunas de las características y funciones clave de Kaiju son las siguientes:

- **Análisis de metagenómica y genómica comparativa:** Kaiju es especialmente útil en el análisis de datos metagenómicos, donde se busca identificar las especies o grupos taxonómicos presentes en una muestra. También se puede utilizar en el análisis de datos genómicos comparativos para evaluar la relación entre genomas.
- **Referencia de datos taxonómicos:** Kaiju utiliza una base de datos de referencia que contiene secuencias de referencia de diversos taxones, incluidos genomas completos y genes marcadores. Estos datos de referencia se utilizan para comparar y asignar las secuencias de consulta.
- **Mapeo rápido y eficiente:** Kaiju utiliza un enfoque de mapeo directo para asignar secuencias a taxones. Esto permite una velocidad de análisis rápida y es especialmente útil para analizar grandes conjuntos de datos.
- **Identificación sensible:** Kaiju es capaz de identificar secuencias incluso a nivel de especies o cepas cercanas, lo que puede ser valioso en el análisis de microbiomas y comunidades de microorganismos.
- **Interfaz de línea de comandos:** Kaiju se ejecuta desde la línea de comandos y admite varios parámetros para ajustar los criterios de asignación y la sensibilidad.
- **Resultados visuales y estadísticos:** Kaiju produce resultados en forma de archivos tabulares que pueden ser procesados posteriormente. También puede generar informes gráficos y estadísticas para resumir la composición taxonómica de las muestras.

- Personalización de la base de datos: Kaiju permite a los usuarios crear y utilizar bases de datos de referencia personalizadas para adaptarse a necesidades específicas de análisis.

(Menzel, Ng, y Krogh, 2016)

## Kraken

Kraken es una herramienta de clasificación taxonómica diseñada para analizar datos de secuenciación de ADN o ARN, como los generados por secuenciadores de próxima generación (NGS), y asignar secuencias a taxones específicos en comunidades microbianas o genomas (Wood, Lu, y Langmead, 2019). Su objetivo principal es identificar la procedencia taxonómica de secuencias en muestras complejas de microbiomas.

Entre las características clave de kraken se encuentran:

- Mapeo de secuencias a taxones: Kraken utiliza un enfoque de mapeo de k-mers (fragmentos cortos de secuencias) para asignar secuencias a taxones en una base de datos de referencia. Esto permite una clasificación rápida y eficiente.
- Base de datos de referencia completa: Kraken utiliza una base de datos de referencia que contiene secuencias de genomas completos, genes marcadores y otras secuencias taxonómicas. Esta base de datos abarca una amplia gama de taxones, lo que permite la identificación de microorganismos en diversos entornos.
- Identificación rápida: Debido a su enfoque basado en k-mers y su base de datos eficiente, Kraken es capaz de realizar análisis de clasificación taxonómica en muestras grandes en un tiempo relativamente corto.
- Niveles de taxonomía: Kraken es capaz de asignar secuencias a diferentes niveles taxonómicos, desde especies hasta categorías más amplias como géneros, familias y más.
- Informe detallado: Kraken produce informes que resumen la composición taxonómica de una muestra, indicando la abundancia relativa de diferentes taxones.
- Flexibilidad y personalización: Los usuarios pueden utilizar bases de datos de referencia personalizadas y ajustar los parámetros para adaptarse a sus necesidades de análisis.
- Uso complementario: Kraken se utiliza a menudo en combinación con otras herramientas para un análisis más completo de los datos metagenómicos, como herramientas de ensamblaje, anotación y análisis funcional.

(Wood y cols., 2019)

## MetaPhlAn

MetaPhlAn es una herramienta de análisis metagenómico que se utiliza para realizar la clasificación taxonómica y la cuantificación de microorganismos presentes en muestras metagenómicas. A diferencia de otras herramientas que utilizan enfoques de mapeo o alineación de secuencias, MetaPhlAn utiliza genes marcadores específicos para identificar y cuantificar la composición taxonómica de comunidades microbianas (Blanco-Míguez y cols., 2023).

Características clave de MetaPhlAn:

- **Marcadores genéticos específicos:** MetaPhlAn utiliza una base de datos de genes marcadores específicos para diferentes taxones microbianos. Estos marcadores son secuencias conservadas en la evolución que permiten la identificación de grupos taxonómicos.
- **Perfilado taxonómico:** La herramienta proporciona un perfil taxonómico que muestra la abundancia relativa de diferentes grupos microbianos en una muestra metagenómica.
- **Identificación específica:** Debido a su enfoque de genes marcadores, MetaPhlAn puede proporcionar información taxonómica hasta el nivel de especies y cepas, lo que es especialmente útil en el análisis de microbiomas complejos.
- **Rápido y eficiente:** MetaPhlAn es capaz de realizar análisis rápidos de clasificación taxonómica en grandes conjuntos de datos, lo que es beneficioso para el análisis de datos metagenómicos de alta velocidad.
- **Reportes visuales y datos de salida:** Proporciona informes gráficos y tabulares que resumen la composición taxonómica de las muestras. Los resultados también pueden ser exportados para análisis posteriores.
- **Uso complementario:** MetaPhlAn se puede usar junto con otras herramientas de análisis metagenómico y genómico para obtener una visión más completa de la comunidad microbiana y su funcionalidad.

(Blanco-Míguez y cols., 2023)

## 4.9. Automatización de procesos computacionales

La automatización se ha convertido en un componente esencial de los flujos de trabajo informáticos modernos, ya que agiliza los procesos y mejora la eficiencia en diversos ámbitos. Un potente método para automatizar tareas en los entornos Linux y Unix es el uso de “*pipelines*” o canalizaciones Bash. Estas cadenas consisten en una serie de herramientas de línea de comandos conectadas para procesar datos de forma secuencial, en la que cada herramienta recibe información de la anterior y pasa los resultados a la siguiente. Este enfoque no sólo mejora la productividad, sino que también permite la gestión eficiente de tareas complejas y repetitivas (*Bash-scripts*, s.f.).

Los *pipelines* de Bash son especialmente ventajosos para automatizar el procesamiento y el análisis de datos. Permiten a los usuarios combinar varios comandos, a menudo utilidades integradas o scripts personalizados, para crear un flujo continuo de manipulación de datos. Por ejemplo, en bioinformática, un investigador puede automatizar el análisis de datos de secuenciación de ADN encadenando comandos de control de calidad, alineación de lecturas, llamada de variantes y anotación. Cada paso del proceso se ejecuta automáticamente, lo que ahorra un tiempo valioso y reduce el riesgo de error humano (*Pipeline-Pitfalls*, s.f.).

Las ventajas de automatizar los programas informáticos mediante cadenas de Bash son múltiples. En primer lugar, la automatización elimina la necesidad de intervención manual en cada paso, reduciendo el potencial de error humano (*Pipeline-Pitfalls*, s.f.). Esto es crucial en campos como la investigación científica, el análisis de datos y la administración de sistemas, donde la precisión es primordial. Además, la automatización fomenta la reproducibilidad, ya que el mismo proceso puede ejecutarse de forma coherente en diferentes conjuntos de datos, lo que garantiza la fiabilidad de los resultados.

Los *pipelines* Bash son altamente escalables y personalizables. Los usuarios pueden añadir o eliminar pasos del proceso para adaptarlo a sus necesidades específicas. Esta flexibilidad las hace adecuadas para una amplia gama de aplicaciones, desde el preprocesamiento de datos hasta tareas

complejas de minería y transformación de datos. Además, la sencillez de las secuencias de comandos de Bash permite a los usuarios adaptar las canalizaciones a sus necesidades específicas, a menudo sin necesidad de conocimientos avanzados de programación ([Pipeline-Pitfalls](#), s.f.).

Aunque las ventajas de la automatización mediante cadenas de Bash son evidentes, es importante ser consciente de los posibles retos. Hay que tener cuidado al construir las canalizaciones para garantizar una gestión de errores y una validación de datos adecuadas. Además, el mantenimiento regular es crucial para mantener las canalizaciones actualizadas con los cambios en los formatos de datos o las dependencias de software. Las mejores prácticas incluyen una documentación exhaustiva, el control de versiones y la colaboración dentro de las comunidades científicas y computacionales para compartir y mejorar los *pipelines* existentes ([Pipeline-Pitfalls](#), s.f.).

La automatización de programas computacionales a través de los *pipelines* Bash es un potente enfoque para agilizar el procesamiento de datos, el análisis y otras tareas. Permite a los usuarios crear flujos de trabajo eficientes, reproducibles y personalizables, reduciendo el margen de error y mejorando la productividad. Aunque existen dificultades, los beneficios de la automatización compensan con creces el esfuerzo inicial necesario para crear y mantener estos canales. A medida que la tecnología y las contribuciones de código abierto sigan evolucionando, el futuro de la automatización a través de los procesos Bash promete un potencial aún mayor para el avance de la ciencia y otros ámbitos.

## 4.10. *Storytelling*: Visualización de resultados

En el ámbito de la toma de decisiones basada en datos y la comunicación científica, la narración o *storytelling* es un arte de suma importancia. Una narración convincente permite que los datos complejos trasciendan los números y las estadísticas, haciéndolos accesibles y atractivos para un público más amplio. Una de las herramientas más potentes del arsenal del narrador es la visualización de datos, el arte de transformar los datos en representaciones gráficas que hablen a la mente humana con claridad e impacto ([Storr, 2020](#)).

Una de las principales ventajas de la visualización de datos en la narración es su capacidad para mejorar la comprensión. Las visualizaciones simplifican datos complejos, haciéndolos más accesibles a un público amplio. Las visualizaciones bien diseñadas ofrecen una representación clara y concisa de los datos, lo que facilita que el público capte los puntos principales de una narración. Además, el contenido visual es más memorable que el texto, lo que garantiza que los puntos clave de la historia dejen una impresión duradera. También ayuda a detectar patrones y tendencias en los datos. Tanto si se trata de identificar correlaciones, valores atípicos o cambios temporales, las visualizaciones facilitan a los narradores y a su público el discernimiento de los aspectos significativos de los datos. Puede ayudar a revelar ideas que podrían ser menos evidentes al examinar los datos en bruto ([Storr, 2020](#)).

Asimismo, estas herramientas puede trascender las barreras lingüísticas y culturales, sirviendo como medio universal de comunicación. Esta característica permite que las narrativas basadas en datos lleguen a una audiencia global. Es una poderosa herramienta para simplificar conceptos complejos. Las historias a menudo implican datos o conceptos intrincados, y la visualización de datos destila estas complejidades en imágenes intuitivas que pueden ser entendidas tanto por expertos como por no expertos ([Storr, 2020](#)). Las visualizaciones de datos eficaces también pueden evocar emociones. Pueden añadir profundidad e impacto a una historia provocando empatía, sorpresa u otras respuestas emocionales de la audiencia. En la práctica, la visualización de datos contribuye a los procesos de toma de decisiones. Ya sea para influir en los responsables de la toma de decisiones, orientar las políticas o ayudar en la resolución de problemas, las visualizaciones de datos presentan pruebas y conocimientos de forma convincente y persuasiva. Ofrecen una representación visual de datos complejos que ayuda a tomar decisiones con conocimiento de causa.

En última instancia, la visualización de datos transforma los datos en un lenguaje narrativo que no sólo es más informativo, sino también más persuasivo. Cuando se utiliza con eficacia, puede convertir estadísticas áridas en una historia convincente, haciendo que los datos complejos sean más atractivos, comprensibles y procesables (Storr, 2020).

La investigación del microbioma es un campo plagado de datos complejos, y la visualización de los mismos desempeña un papel fundamental a la hora de entender esta complejidad. El microbioma es increíblemente diverso, con miles de microorganismos, cada uno con sus características genéticas y funcionales únicas. Caracterizar estas comunidades implica recopilar datos sobre las especies, su abundancia y sus interacciones. Visualizar estos datos multidimensionales es un reto formidable. A menudo se utilizan herramientas de visualización de datos como mapas térmicos, gráficos de barras apiladas y gráficos de redes para transmitir la abundancia relativa de los distintos microorganismos y sus interacciones (Chong, Liu, Zhou, y Xia, 2020).

Los datos del microbioma pueden ser muy dinámicos, ya que las comunidades microbianas cambian con el tiempo en respuesta a factores como la dieta, la medicación o los cambios ambientales. Los gráficos de series temporales y las visualizaciones animadas son valiosos para representar los cambios temporales en la composición del microbioma, ayudando a los investigadores a seguir la evolución y adaptación de las comunidades microbianas (Chong y cols., 2020).

La investigación del microbioma incluye datos tanto taxonómicos (composición) como funcionales (potencial genético y vías metabólicas). Las técnicas de visualización de datos, como los gráficos circulares y los diagramas de enriquecimiento de vías, representan la composición taxonómica y los aspectos funcionales, destacando el papel de determinadas vías metabólicas.

La visualización de datos es fundamental para simplificar esta complejidad, ayudando a los investigadores a extraer ideas significativas, descubrir patrones y comunicar sus hallazgos de manera eficaz. El campo sigue evolucionando con nuevas e innovadoras técnicas de visualización de datos que abordan los retos únicos que plantea la investigación del microbioma (Chong y cols., 2020).

Para abordar con el desafío se ha popularizado el uso de lenguaje R y aplicaciones Shiny los cuales son un aspecto dinámico e indispensable de la investigación del microbioma. R, con sus paquetes especializados como "phyloseq", sirve de base para el análisis y la manipulación de datos. "Phyloseq" está diseñado para estudios sobre el microbioma y permite a los investigadores importar y trabajar con datos específicos de este campo. El paquete "ggplot2" de R mejora aún más la visualización de datos al ofrecer una amplia gama de opciones de visualización. Es especialmente valioso para crear gráficos de alta calidad, como diagramas de barras, mapas de calor y gráficos de ordenación, que facilitan la representación de composiciones microbianas y estructuras comunitarias (Galloway-Peña y Hanson, 2020).

#### 4.10.1. R

R va más allá de los gráficos estáticos al incorporar interactividad con paquetes como "plotly". Los gráficos interactivos son cruciales para explorar datos microbianos complejos, ya que permiten a los usuarios interactuar con los datos, fomentando así una comprensión más profunda. Además, R admite varios modelos estadísticos, lo que lo convierte en una plataforma versátil para analizar la influencia de distintas variables en las comunidades microbianas (Galloway-Peña y Hanson, 2020).

#### 4.10.2. Shiny

Las aplicaciones Shiny ofrecen un enfoque dinámico y fácil de usar para la visualización de datos del microbioma. Los investigadores pueden crear paneles interactivos personalizados que permiten explorar los datos, interactuar con ellos y actualizarlos en tiempo real. Estos paneles pueden incor-

porar varios tipos de gráficos, tablas y componentes interactivos para una comprensión completa de los datos del microbioma. Una característica destacada es el filtrado dinámico, que permite a los usuarios seleccionar grupos taxonómicos específicos, puntos temporales u otras variables relevantes para afinar sus visualizaciones (Galloway-Peña y Hanson, 2020) (ProQuest Platform, s.f.).

Uno de los puntos fuertes de las aplicaciones Shiny es su capacidad para integrar modelos estadísticos, lo que ofrece a los usuarios la oportunidad de investigar cómo afectan los distintos factores a las comunidades microbianas (ProQuest Platform, s.f.). Esto es crucial para la investigación del microbioma, ya que ayuda a identificar el impacto de variables como la dieta, el estado de salud o el medio ambiente en las composiciones microbianas.

Además, las aplicaciones Shiny se han diseñado pensando en la facilidad de uso. Ofrecen interfaces intuitivas que no requieren grandes conocimientos de programación, lo que las hace accesibles a un público amplio. Las aplicaciones Shiny pueden utilizarse para crear cuadros de mando interactivos que facilitan la actualización de los datos en tiempo real, garantizando así que el público tenga acceso a los datos más recientes sobre el microbioma (ProQuest Platform, s.f.).

La combinación de R y aplicaciones Shiny en la investigación del microbioma proporciona un enfoque versátil y potente para la visualización de datos. Permite a los investigadores crear soluciones personalizadas que se adaptan a las necesidades específicas de sus estudios, ya se trate de visualizar composiciones taxonómicas, explorar dinámicas temporales o permitir la exploración interactiva de datos. Estas herramientas son esenciales para hacer más accesibles los datos complejos del microbioma y para comunicar eficazmente los resultados de la investigación a un público más amplio (Galloway-Peña y Hanson, 2020).

El presente proyecto identifica como los sujetos de estudio los siguientes:

- Pacientes con esquizofrenia de Dinamarca (n = 53)
- Secuencias control de sujetos de Dinamarca (n=53)

El proyecto parte desde el paso de la identificación de las secuencias del microbioma intestinal de los sujetos de estudio presentados anteriormente. Para esta sección preliminar se realizó una búsqueda exhaustiva en donde se encontraron la muestra fecales de individuos participantes de un estudio realizado previamente en Dinamarca bajo el título de '*Alteration of gut microbiome in patients with schizophrenia indicates links between bacterial tyrosine biosynthesis and cognitive dysfunction*' escrito por Thirion F, et al. La secuencias metagenómicas de las muestras utilizadas por Thirion pueden ser encontradas en la base de datos del Instituto Europeo de Bioinformática (<https://www.ebi.ac.uk>) bajo el código de acceso de PRJEB41217 de las muestras de pacientes con esquizofrenia y PRJEB41786 para las secuencias control.

El presente estudio tiene un diseño descriptivo, ya que se tiene como objetivo el de poder identificar bacterias en el microbioma intestinal de los sujetos de estudio, además de ser una investigación aplicada tecnológica ya que se busca esta identificación a través de un *pipeline* que permita la automatización de esta identificación bacteriana; añadiendo el poder proveer una visualización “user friendly” para su siguiente análisis.

La muestra a ser utilizada, como ya se determinó anteriormente, consiste en secuencias de microbioma de pacientes esquizofrénicos, teniendo así una muestra consistente de meta data de las secuencias.

## 5.1. Metodología

En esta metodología, se expone los pasos involucrados en el desarrollo de un *pipeline* para el análisis de microbioma con una aplicación específica para el análisis microbional de pacientes con

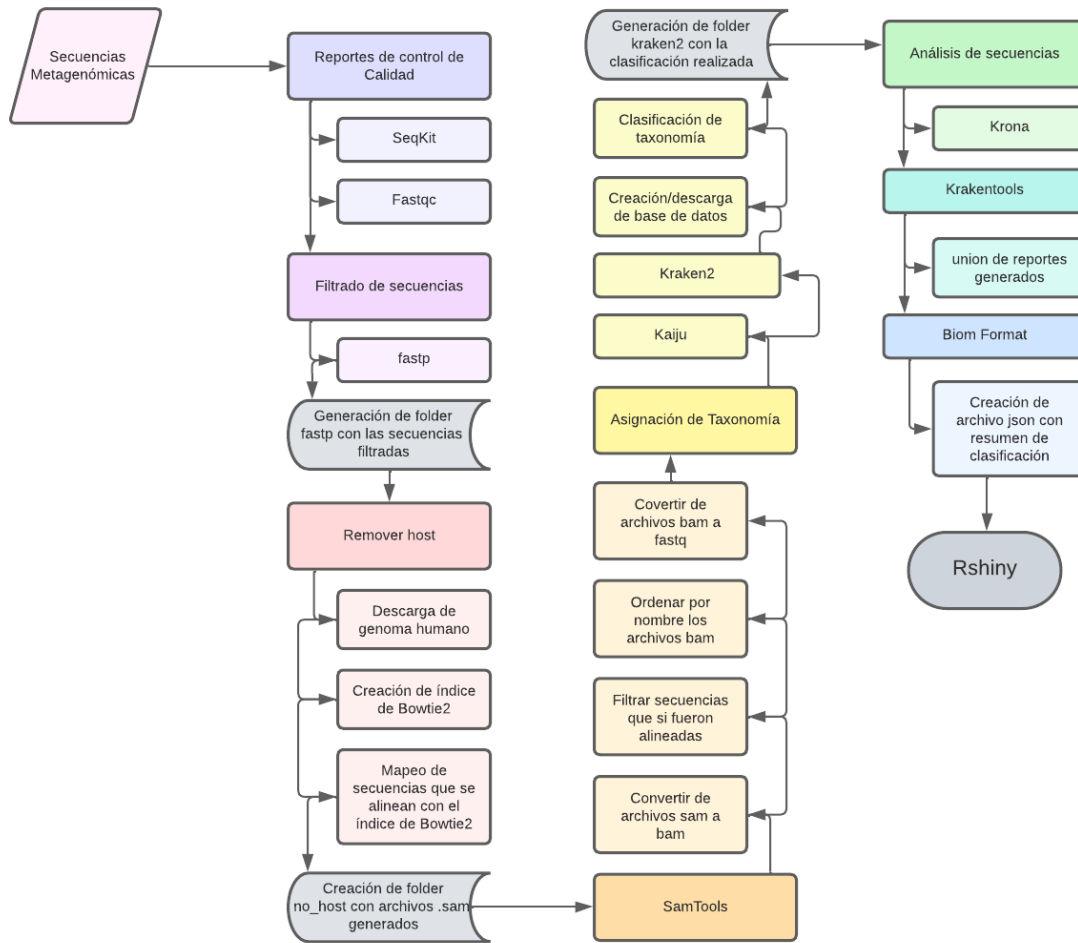


Figura 5.1: Diagrama de flujo de pasos para el *pipeline*

esquizofrenia vs. pacientes control. Las muestras a utilizar son 53 secuencias de personas control 'sanos' de Dinamarca y 53 pacientes con esquizofrenia.

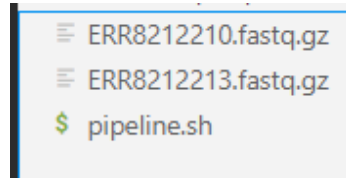
Este desarrollo incluye la descarga de la secuencias encontrada en la base de datos, el control de calidad de secuencias, filtrado de secuencias que no cumplan con el control de calidad estándar, descarga de base de datos de referencia para la clasificación bacteriana, la clasificación de bacterias dentro de la muestra, el curado de los resultados y finalmente el protocolo paso a paso para utilizar Shiny con la finalidad de tener una mejor visualización de los resultados obtenidos. Dentro de la metodología también se destacó las consideraciones clave como la selección de herramientas y algoritmos, la optimización de parámetros y la interpretación de resultados.

Esta metodología proporcionará una guía paso a paso para diseñar y construir un *pipeline* de análisis de microbioma eficaz y personalizado, que permitirá a los investigadores explorar y comprender mejor el mundo microbiano que nos rodea. Los pasos a seguir son los presentados en la Figura [5.1](#).

Dado a que se trata con secuencias de metagenómica, los archivos comprimidos que contienen el fastq son de un peso aproximado de 2.5 GB, por lo que, dada la cantidad a tratar dentro del proyecto, para la realización preliminar del *pipeline* se tratará con 2 secuencias iniciales a descargar en una computadora con procesador AMD RYZEN 5 3500U 4 Core(s) 8 GB RAM y 250 GB de

almacenamiento.

Se hace una instalación de conda en Ubuntu 20.04.06 LTS. Se descargan el archivo comprimido de las dos secuencias ERR8212210.fastq.gz y ERR8212213.fastq.gz y se crea el programa bash bajo el nombre pipeline.sh. Teniendo una estructura inicial como vista en la Figura [5.2](#)



**Figura 5.2:** Estructura inicial de archivos en sistema Linux

A continuación se comienza con la edición del archivo "pipeline.sh" las primeras líneas de código es para la instalación de los canales defaults, bioconda y conda-forge. Al instalar estos canales junto con Anaconda, se obtiene acceso a una amplia gama de paquetes que cubren diferentes áreas, desde las bibliotecas básicas de Python hasta herramientas especializadas en bioinformática y otros campos. Esto permite crear entornos de conda que incluyen fácilmente los paquetes que se necesita para el proyecto, sin tener que preocuparse por la compatibilidad de versiones o la disponibilidad de los paquetes.

```
1 conda config --add channels defaults
2 conda config --add channels bioconda
3 conda config --add channels conda-forge
4 conda init
```

**Listing 5.1:** Instalación de canales

### 5.1.1. Control de calidad y filtrado de secuencias

Para el control de calidad y filtrado a realizar dentro de las secuencias hay que tomar en cuenta la necesidad de la creación de diferentes ambientes dentro de conda para evitar problemas de compatibilidad de dependencias entre los paquetes a utilizar. Por lo que se crean los siguientes ambientes de fastqc, fastp, bowtie2 y samtools. Dado a que se busca crear un *pipeline* que no involucre la interacción con el usuario, se utiliza las banderas de '-y' para la creación de ambientes.

```
1 conda create -n fastqc -y
2 conda create -n fastp -y
3 conda create -n bowtie2 -y
4 conda create -n samtools -y
```

**Listing 5.2:** Creación de ambientes fastqc

Al momento de comenzar con el control de calidad es necesaria la activación del ambiente de fastqc, el cual se hace a través del siguiente comando:

```
1 source activate fastqc
```

**Listing 5.3:** Activación de ambiente fastqc

## Seqkit

Se toma en cuenta que el primer paso sea la elaboración de un .txt que nos pueda tener un resumen de la información de las secuencias que se están tratando en el análisis. Por lo que se ve necesaria la incorporación de la librería de Seqkit para esto. Se utiliza el comando de *stats* dentro de esta librería para cumplir con este objetivo.

```

1 conda install -c bioconda seqkit -y
2 seqkit stats *fastq.gz > stats.txt

```

**Listing 5.4:** Instalación de Seqkit y uso

Una vez se haya corrido el código especificado anteriormente se tiene como resultado un archivo llamado "stats.txt" que contenga la información que se desea. En el caso de las secuencias ejemplo el archivo "stats.txt" se presenta como se puede ver en la Figura 5.3.

```

≡ stats.txt
1 file                format type    num_seqs    sum_len  min_len  avg_len  max_len
2 ERR8212210.fastq.gz FASTQ  DNA    19,096,934  2,843,679,873    45    148.9    375
3 ERR8212213.fastq.gz FASTQ  DNA    3,064,886   424,327,576     45    138.4    317

```

**Figura 5.3:** Resultado de "stats" de seqkit sobre las secuencias descargadas

Como puede ser observado, este resultado proporciona información básica sobre el contenido de los archivos \*fastq.gz de secuencias, como la cantidad total de secuencias, la longitud promedio, la longitud máxima y mínima. Lo cual ayuda a comprender la naturaleza de los datos al momento de realizar análisis más avanzados.

## Fastqc

Luego se hace la instalación y uso de la librería de fastqc, la cuál nos brindará con los reportes acerca de la calidad de de las secuencias que se están tratando del análisis. Para ejecutar este análisis se necesitan los siguientes comandos:

```

1 conda install -c bioconda fastqc -y
2 mkdir fastqc
3 fastqc --noextract -o fastqc *.fastq.gz
4 conda deactivate

```

**Listing 5.5:** Instalación de fastqc y uso

Por motivos de orden se crea la carpeta de fastqc dentro de nuestro directorio de trabajo para que con el uso de la bandera o"fastqc deposite sus reportes dentro de esta carpeta y puedan ser encontrados de manera ordenada. Al momento de ejecutar estas tres líneas de código nos encontramos con la creación de los archivos encontrados en el esquema de directorio de la Figura 5.4.

Como puede observarse, la librería de fastqc nos crea dos archivos iniciales por secuencia dentro de la carpeta llamada "fastqc". Estos archivos que nos encontramos es el mismo resultado en diferentes formatos. Primero nos encontramos con un reporte interactivo y resumido en formato .html, cuyo ejemplo se puede ver en la Figura 5.5 y 5.6. Por otro lado genera una carpeta comprimida en donde se nos brindan los elementos que se utilizaron para construir el html visto.

Dado a que se busca automatizar el proceso, estos archivos quedan para uso del investigador posteriormente se haya puesto en ejecución el programa. Sin embargo, se hace necesaria el uso de una librería para el filtrado de secuencias que no cumplan con las condiciones estándar de calidad para no afectar en el resultado.

## Fastp

Con Fastp se realiza la tarea de filtración de secuencias a través de la línea de comandos. Para la implementación de la librería se debe de ejecutar las siguientes líneas de código:

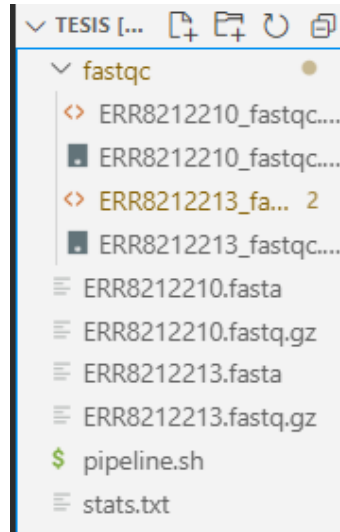


Figura 5.4: Esquema de directorio

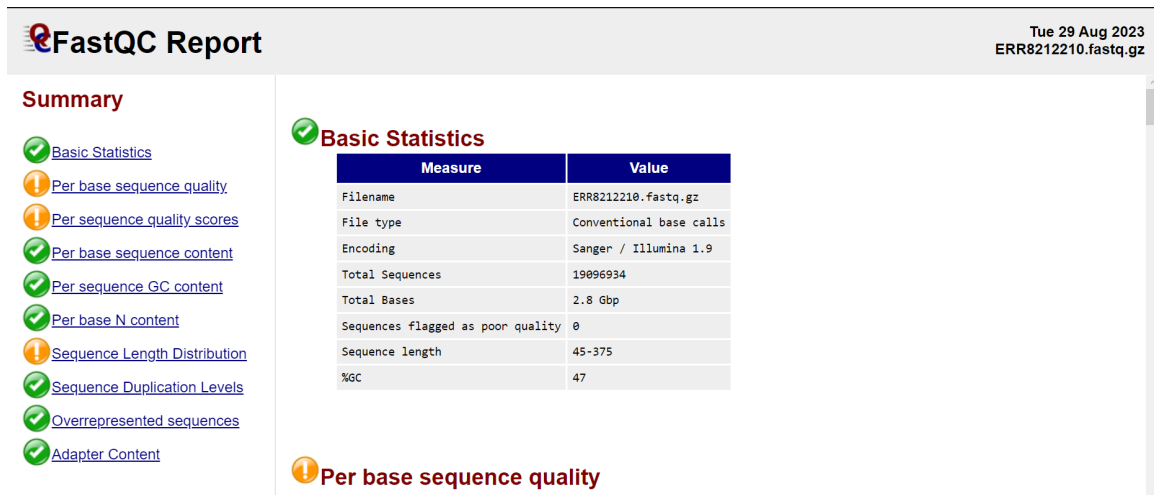


Figura 5.5: Reporte generado por Fastqc

```

1  source activate fastp
2  conda install -c bioconda fastp -y
3  mkdir filtered_reads
4  for infile in *.fastq.gz
5  do
6      base=$(basename ${infile})
7      #Extrae el nombre del archivo sin la extension ".fastq.gz"
8      newname="${base%.*}"
9      fastp -i $infile -o filtered_reads/${newname}.filt.fastq.gz
10 done
11 conda deactivate
12 rm *.fastq.gz
13 cd filtered_reads

```

Listing 5.6: Filtración de lecturas basadas en calidad de secuenciación a través de FASTP

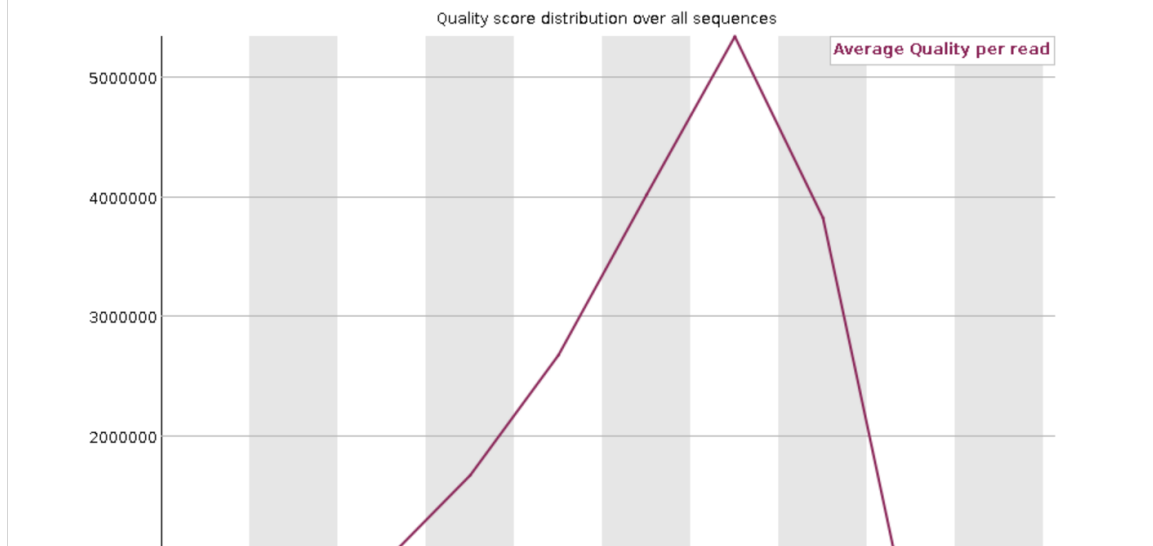
Este script primero activa un entorno conda llamado fastp. Este entorno contiene el paquete de software fastp, así como cualquier otra dependencia que fastp necesite para ejecutarse. A continuación, el script crea un directorio llamado filtered\_reads. Este directorio almacenará los archivos

### ! Per base sequence quality

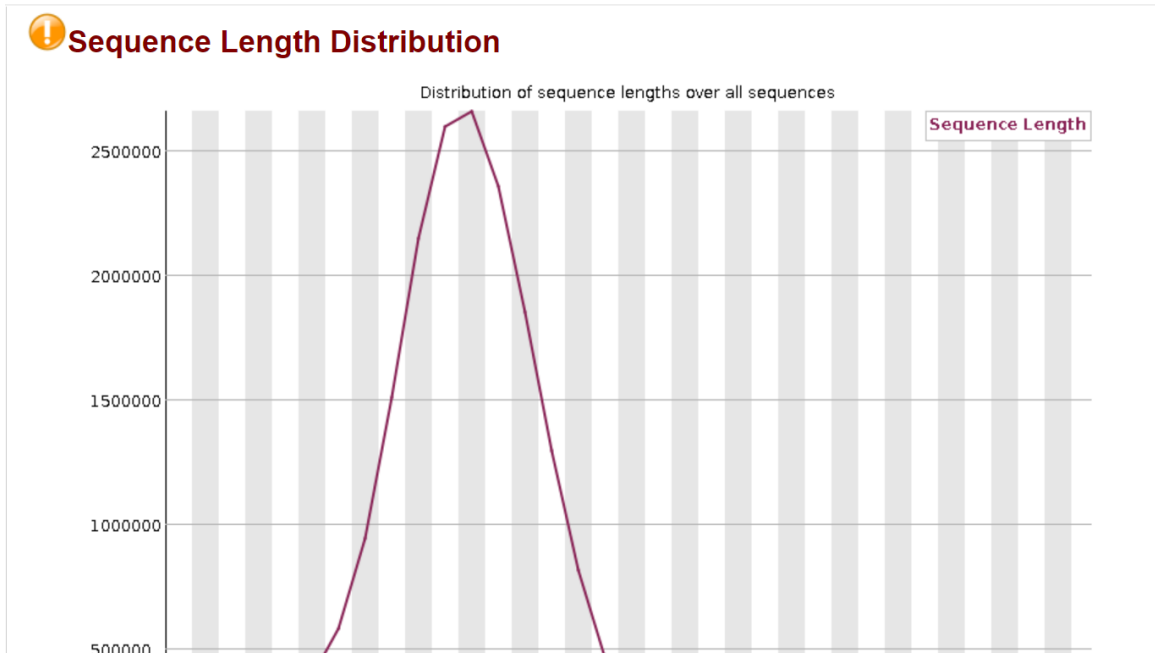


(a) Gráfica de Calidad de Secuencia por Base. En el informe de FastQC, el "Per Base Sequence Quality" muestra la calidad de cada base en todas las secuencias del archivo de secuencias. Aquí se puede observar que las secuencias se encuentran en el área amarilla, en donde se encuentran en un umbral de calidad aceptable, por lo general, se establece un umbral de calidad (por ejemplo, Phred score de 20) como mínimo aceptable para la mayoría de aplicaciones de análisis. Si la calidad de las bases cae por debajo de este umbral, podría ser necesario realizar recortes o filtrar las secuencias.

### ! Per sequence quality scores



(b) Puntuaciones de Calidad por Secuencia. Estas gráficas dan una idea de la calidad general de las secuencias en el conjunto de datos. Si la mayoría de las secuencias tienen puntuaciones de calidad altas y uniformes, es una señal de que los datos son de alta calidad.



(c) Distribución de Longitud de Secuencia. La "Sequence Length Distribution" es una métrica importante para evaluar la calidad de los datos de secuenciación de ADN. Esta métrica proporciona información sobre la distribución de las longitudes de las secuencias en el conjunto de datos.

**Figura 5.6:** Gráficas generadas por Fastqc

FASTQ filtrados generados por el script.

Por cada archivo de secuencias metagenómicas que se tengan en el directorio actual, se recorre y se realizan los siguientes pasos:

1. Extrae el nombre base del archivo, que es el nombre del archivo sin la extensión .fastq.gz.
2. Crea un nuevo nombre para el archivo eliminando la extensión .fastq.gz y añadiendo el sufijo .filt.
3. Ejecuta el comando fastp para filtrar el archivo FASTQ.GZ y guarda el resultado en el directorio filtered\_reads.

Una vez que el script ha terminado de procesar todos los archivos FASTQ.GZ, desactiva el entorno fastp conda. También elimina todos los archivos FASTQ.GZ originales del directorio actual y cambia el directorio actual al directorio filtered\_reads.

El resultado de la ejecución del presente código termina en tanto la creación de reportes de calidad de secuencia en formatos .html y .json (cuyo ejemplo puede visualizarse en la Figura 5.7) como en la creación del directorio 'filtered\_reads' con las secuencias filtradas por la librería.

## Bowtie2

En el proceso de Control de calidad y filtrado de secuencias también es necesaria la filtración de lecturas de secuencias del hospedero, el cual en el caso de las secuencias se trata acerca del ADN humano. Para esta eliminación se utiliza la librería de Bowtie2. Con esta librería se puede construir el

fastp report

Summary	
General	
fastp version:	0.23.4 ( <a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a> )
sequencing:	single-end (317 cycles)
mean length before filtering:	138bp
mean length after filtering:	138bp
duplication rate:	2.802355% (may be overestimated since this is SE data)
Before filtering	
total reads:	3,064886 M
total bases:	424,327576 M
Q20 bases:	364,674047 M (85,941633%)
Q30 bases:	15,561086 M (3,667234%)
GC content:	47,785244%
After filtering	
total reads:	3,064886 M
total bases:	424,327576 M
Q20 bases:	364,674047 M (85,941633%)
Q30 bases:	15,561086 M (3,667234%)
GC content:	47,785244%

**Figura 5.7:** Reporte generado por fastp por secuencia acerca del estado anterior de las secuencias y luego de ser filtradas

index personalizado que alineará las lecturas de nuestras secuencias al genoma humano de referencia encontrado en [https://ftp.ncbi.nlm.nih.gov/refseq/H\\_sapiens](https://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens). Esto dado a que al momento de intentar descargar el índice de humano ya desarrollado para Bowtie2 da errores de poca-identificación.

```

1 source activate bowtie2
2 conda install -c bioconda bowtie2 -y
3 mkdir bowtie2index
4 cd bowtie2index
5 wget https://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/annotation/GRCh38_latest/
   refseq_identifiers/GRCh38_latest_genomic.fna.gz
6 gzip -d GRCh38_latest_genomic.fna.gz
7 bowtie2-build GRCh38_latest_genomic.fna human
8 rm GRCh38_latest_genomic.fna
9 cd ..

```

**Listing 5.7:** Instalación de bowtie2 y creación de index

Por temas de espacio en el disco duro y de memoria RAM fue necesaria la creación de una máquina virtual en una nube pública. Para el presente proyecto se realizó la creación de una máquina virtual en Digital Ocean (llamada Droplet) con características de 32 GB de RAM, 8 AMD vCPUs y 400 GB de memoria en disco duro. Esta máquina virtual fue creada con un sistema operativo Ubuntu 20.04 (LTS) x64.

Una vez creada la máquina virtual se conecta a la máquina virtual a través de la IPv4 provisionada con ella. Se corre el código y este crea los 6 archivos necesarios para index de bowtie2 los cuales son ilustrados en la Figura [5.8](#).

Una vez creado el index se procede a realizar el alineamiento de las secuencias y el filtrado del ADN del hospedero.

```

1 mkdir no_host
2 cd no_host
3 for infile in ../filtered_reads/*.fastq.gz
4     do
5         base=$(basename ${infile})
6         newname="${base}%.*"
7         bowtie2 -x ../bowtie2index/human -U ${base} > ${newname}.sam
8     done
9 done
10 conda deactivate
11 rm -r ../bowtie2index

```

**Listing 5.8:** Filtrado de lecturas



```
human.1.bt2
human.2.bt2
human.3.bt2
human.4.bt2
human.rev.1.bt2
human.rev.2.bt2
```

Figura 5.8: Archivos creados por bowtie2-build

Este script crea un nuevo directorio llamado `no_host` y cambia el directorio actual a ese directorio. Luego recorre todos los archivos FASTQ.GZ del directorio `../filtered_reads`. Para cada archivo FASTQ.GZ, el script realiza los siguientes pasos:

1. Extrae el nombre base del archivo, que es el nombre del archivo sin la extensión `.fastq.gz`.
2. Crea un nuevo nombre para el archivo eliminando la extensión `.fastq.gz` y añadiendo la extensión `.sam`.
3. Ejecuta el comando `bowtie2` para alinear el archivo FASTQ.GZ con el índice del genoma humano de referencia creado en el paso anterior. La salida del comando `bowtie2` es un archivo SAM, que contiene las lecturas alineadas.

Una vez que el script ha terminado de procesar todos los archivos FASTQ.GZ, desactiva el entorno `bowtie2` `conda` y elimina el directorio `../bowtie2index`.

El resultado de correr este código será la creación del directorio `no_host` con las secuencias alineadas y no alineadas al ADN humano en un formato SAM. Estas secuencias deben de ser convertidas a un formato `fastq` para lo cual se utilizará la librería de `Samtools`.

## Samtools

```
1 #Samtools
2 source activate samtools
3 conda install -c bioconda samtools -y
4 sudo apt-get install pigz -y
5 for infile in *.sam
6     do
7         base=$(basename ${infile})
8         newname="${base%.*}"
9         #convertir Sam en Bam
10        samtools view -bS ${base} > ${newname}.bam
11        rm ${infile}
12        #filtrar secuencias que no alinearon al index
13        samtools view -b -f 4 ${newname}.bam > ${newname}.unmapped.bam
14        rm ${newname}.bam
15        #ordenar secuencias
16        samtools sort -n -@ 4 ${newname}.unmapped.bam -o ${newname}_sorted.bam
17        rm ${newname}.unmapped.bam
18        samtools fastq -@ 4 ${newname}_sorted.bam > ${newname}.fastq
```

```

19     rm -r *.bam
20     pigz ${newname}.fastq
21     done
22 conda deactivate samtools

```

**Listing 5.9:** Conversión de formato SAM a FASTQ.GZ

Este script primero activa un entorno conda llamado samtools. Este entorno contiene el paquete de software samtools, así como cualquier otra dependencia que samtools necesite para ejecutarse. A continuación, el script instala el paquete de software pigz. Pigz es un compresor gzip paralelo, lo que significa que puede comprimir archivos mucho más rápido que gzip cuando hay varios núcleos disponibles.

Luego el script recorre todos los archivos SAM del directorio actual. Para cada archivo SAM, el script realiza los siguientes pasos:

1. Extrae el nombre base del archivo, que es el nombre del archivo sin la extensión .sam.
2. Crea un nuevo nombre para el archivo eliminando la extensión .sam.
3. Con -bS se convierte el archivo SAM en un archivo BAM utilizando el comando samtools view.
4. Filtra el archivo BAM para eliminar cualquier secuencia que no se alinee con el genoma de referencia utilizando el comando samtools view con la opción -f 4.
  - -b: esta bandera en el código nos indica que el *output* sera en formato BAM.
  - -f: esta bandera nos indica sobre que criterio se basará samtools para extraer los datos. En este caso se usa 4 para conseguir los no alineados.
5. Ordena el archivo BAM utilizando el comando sort de samtools con la opción -n para ordenar las secuencias por nombre.
6. Convierte el archivo BAM en un archivo FASTQ utilizando el comando fastq de samtools.
7. Comprime el archivo FASTQ con el comando pigz.

Una vez que el script ha terminado de procesar todos los archivos SAM, elimina todos los archivos SAM y BAM del directorio actual. También desactiva el entorno conda de samtools.

Cabe recalcar que se hace una eliminación de archivos de extensión SAM y BAM por eficiencia de espacio, ya que se toma en cuenta que se puede estar tratando con una extensa base de datos para análisis de microbioma.

Una vez se obtiene una carpeta con las secuencias fastq.gz sin el ADN humano se procede a la clasificación de bacterias de las secuencias.

## 5.1.2. Asignación de taxonomía

### Kaiju

Una de las primeras herramientas esenciales para la clasificación y asignación de taxonomía es Kaiju. Para poder instalar y utilizar kaiju se es necesario correr el siguiente código:

```

1 conda create -n kaiju -y
2 source activate kaiju
3 conda install -c bioconda kaiju -y
4 mkdir kaijudb

```

```

5 mkdir kaiju
6 cd kaijudb
7 kaiju-makedb -s refseq
8 cd ../kaiju
9 for infile in ../no_host/*.fastq.gz
10 do
11     base=$(basename ${infile})
12     newname="${base%.*}_kaiju.out"
13     kaiju -t nodes.dmp -f ./kaijudb_refseq.fmi -i <(gunzip -c ${infile}) -z 4 -x
14     -s 75 -a greedy -e 5 -o ${newname}_kaiju.out -v
15 done
conda deactivate

```

**Listing 5.10:** Uso de Kaiju

Este script crea un nuevo entorno conda llamado kaiju e instala el paquete de software Kaiju en ese entorno. Luego crea dos directorios: kaijudb y kaiju. El directorio kaijudb almacenará la base de datos de referencia Kaiju. El directorio kaiju almacenará los archivos de salida de Kaiju.

A continuación, el script cambia el directorio actual al directorio kaijudb y utiliza el comando kaiju-makedb para crear la base de datos de referencia Kaiju. La base de datos de referencia se crea a partir de la base de datos RefSeq, que es una base de datos de genomas completos y anotados de Archaea, Bacteria y Virus de la base de datos de NCBI Refseq.

Luego el script cambia el directorio actual al directorio kaiju y repasa todos los archivos FASTQ.GZ del directorio no\_host. Para cada archivo FASTQ.GZ, el script realiza los siguientes pasos:

1. Extrae el nombre base del archivo, que es el nombre del archivo sin la extensión .fastq.gz.
2. Crea un nuevo nombre para el archivo eliminando la extensión .fastq.gz y añadiendo el sufijo \_kaiju.out.
3. Ejecuta el comando kaiju para clasificar las secuencias en el archivo FASTQ.GZ utilizando la base de datos de referencia Kaiju. La salida del comando kaiju es un archivo llamado \$newname\_kaiju.out.
  - La opción '-t' en el comando 'kaiju' especifica la ruta a la base de datos de taxonomía Kaiju.
  - La opción '-f' especifica la ruta a la base de datos de referencia Kaiju.
  - La opción '-i' especifica la ruta al archivo de entrada FASTQ.GZ.
  - La opción '-z' especifica el número de hilos a utilizar.
  - La opción '-x' especifica que el archivo FASTQ.GZ de entrada no está emparejado.
  - La opción '-s' especifica la puntuación mínima de alineación requerida para que una secuencia sea clasificada.
  - La opción '-a' especifica el algoritmo de clasificación a utilizar, la cual en este caso es *greedy*.
  - La opción '-e' especifica el número máximo de nodos taxonómicos a reportar para cada secuencia.
  - La opción '-o' especifica la ruta del fichero de salida.
  - La opción '-v' especifica que el comando 'kaiju' debe mostrar información detallada.

Una vez que el script ha terminado de procesar todos los archivos FASTQ.GZ, desactiva el entorno 'kaiju' conda. Este script puede utilizarse para clasificar las secuencias de un conjunto de datos metagenómicos o metatranscriptómicos utilizando Kaiju. La salida del script es un conjunto de

archivos llamados \$newname\_kaiju.out', que contienen la clasificación taxonómica de las secuencias. Estos archivos pueden utilizarse para análisis posteriores, como la identificación de la composición microbiana de una muestra o la determinación de la abundancia de genes específicos.

Sin embargo, dado a los recursos computacionales que se requiere para hacer la construcción del index para poder luego hacer la clasificación de bacterias, el cual es de 112 GB de RAM, se encontraron errores en la creación de la base de datos de referencia y se prosiguió con otro algoritmo de clasificación.

## Kraken2

Kraken fue el primer clasificador metagenómico basado en k-mer. Utiliza una tabla hash para indexar los k-mers de la base de datos de referencia y, a continuación, utiliza un algoritmo codicioso para clasificar las lecturas de la muestra de consulta. Kraken es muy rápido y eficiente, pero no es tan preciso como Kraken 2. Este es una versión más reciente de Kraken, más precisa y eficiente. Utiliza un modelo probabilístico para clasificar las lecturas en la muestra de consulta, y también tiene en cuenta la abundancia de los k-mers en la base de datos de referencia. Kraken 2 es más lento que Kraken, pero más preciso. Por lo que se hace uso de Kraken 2, para este software se puede descargar y construir la base de datos *standard*. Sin embargo esta base de datos *standard* necesita 100 GB de RAM por lo que se prosiguió con la creación de una base de datos "personalizada" que contenga los índices de bacteria y archaea.

```
1 conda create -n kraken2 -y
2 source activate kraken2
3 conda install -c bioconda kraken2 -y
4 kraken2-build --download-taxonomy --skip-maps --db 'krakendb'
5 kraken2-build --download-library bacteria --db 'krakendb'
6 kraken2-build --download-library archaea --db 'krakendb'
7 kraken2-build --build --max-db-size 8000000000 --db 'krakendb'
```

**Listing 5.11:** Instalación de Kraken2 y construcción de base de datos

Este código instala y construye el clasificador metagenómico Kraken2. La primera línea de código crea un nuevo entorno Conda llamado kraken2. La bandera -y le dice a Conda que responda automáticamente Sí a cualquier pregunta. La segunda instala el paquete Kraken2 desde el canal Bioconda. La bandera -y le dice a Conda que responda automáticamente Sí a cualquier pregunta.

La tercera, cuarta y quinta líneas de código descargan la taxonomía, la biblioteca de bacterias y la biblioteca de arqueas para Kraken2, respectivamente. La bandera --skip-maps le dice a Kraken2 que no construya los mapas de alineamiento, lo que puede ahorrar mucho espacio en disco.

La sexta línea de código construye la base de datos de Kraken2. La opción --max-db-size fija el tamaño máximo de la base de datos en 8 GB. A pesar, de haber indicado que la base de datos no excediera un espacio mayor a 8GB, se utiliza un espacio mayor a 60 GB de RAM para la construcción de "krakendb" por lo que se prosiguió con la descarga externa de la base de datos ya construida por un usuario subida en la nube de Amazon (AWS).

```
1 cd ..
2 mkdir dbdownload
3 cd dbdownload
4 curl https://genome-idx.s3.amazonaws.com/kraken/k2_standard_16gb_20230605.tar.gz --
  output dbdownloaded.tar.gz
5 gzip -d dbdownloaded.tar.gz
6 tar -xvf dbdownloaded.tar
7 rm dbdownloaded.tar
8 cd ..
```

**Listing 5.12:** Descarga de base de datos standard de Kraken2

Con este código se logra tener los archivos esenciales de hash.k2d, opts.k2d y taxo.k2d que necesita Kraken2 para realizar la clasificación taxonómica.

```
1 mkdir kraken2
2 cd kraken2
3 for infile in ../no_host/*.fastq.gz
4 do
5     base=$(basename ${infile})
6     newname="${base%.*}.kraken"
7     kraken2 --db ~/tesis/dbdownload ../no_host/${infile} --gzip-compressed --output ${
8         newname}.kraken --report ${newname}.report --threads 4
9 done
10 conda deactivate
11 cd ..
```

**Listing 5.13:** Clasificación de bacterias con Kraken2

Este script crea un directorio llamado kraken2, cambia el directorio actual a kraken2 y, a continuación, realiza un bucle sobre todos los archivos FASTQ.GZ del directorio ../no\_host. Para cada archivo FASTQ.GZ, el script realiza los siguientes pasos:

1. Extrae el nombre base del archivo, que es el nombre del archivo sin la extensión .fastq.gz.
2. Crea un nuevo nombre para el archivo eliminando la extensión .fastq.gz y añadiendo el sufijo .kraken.
3. Ejecuta el comando kraken2 para clasificar las secuencias en el archivo FASTQ.GZ utilizando la base de datos Kraken2 que se descargó en el ejemplo anterior.
  - La opción -db especifica la ruta a la base de datos Kraken2.
  - -gzip-compressed indica a Kraken2 que el archivo FASTQ.GZ de entrada está comprimido.
  - -output especifica la ruta al archivo de resultados de salida de Kraken2.
  - -report especifica la ruta del archivo de informe de Kraken2 de salida.
  - -threads especifica el número de hilos a utilizar.

Como resultado de este script se nos brinda la clasificación de bacterias en un formato kraken, en el archivo, como puede verse en la Figura 5.9 se muestra como por cada lectura dentro del archivo de secuencias .fastq.gz se logra, o no se logra, la clasificación de taxonomía. Para luego, si en dado se logra clasificar, con que índice de la base de datos se hizo la alineación.

### 5.1.3. Análisis de secuencias

#### Krona

El output de Kraken2 puede utilizarse para realizar análisis con Krona. Krona es una herramienta de visualización que puede generar gráficos interactivos de los resultados de Kraken2. Estos gráficos pueden ayudar a los investigadores a comprender la composición microbiana de una muestra o la abundancia de genes específicos.

Para utilizar Krona con el output de Kraken2, primero debe convertir el output de Kraken2 a un formato compatible con Krona. Esto se puede hacer utilizando el comando ktImportTaxonomy de Krona. Una vez que el output de Kraken2 está en un formato compatible con Krona, puede ser utilizado para generar gráficos interactivos con Krona.

```

C ERR8212210.25783 751585 168 0:10 751585:5 0:22 751585:6 0:9 751585:4 0:12 751585:3 0:12 751585:1 0:11 751585:2 0:16 751585:1 0:9 751585:5 0:5 751585:1
U ERR8212210.25784 0 126 0:92
C ERR8212210.25785 165179 153 0:7 165179:5 0:1 165179:5 0:2 165179:9 0:16 165179:2 0:15 2:2 0:19 165179:6 0:1 165179:5 0:12 165179:2 0:10
C ERR8212210.25786 165179 177 0:61 165179:1 0:7 165179:2 0:5 165179:3 0:16 165179:8 0:40
U ERR8212210.25787 0 130 0:96
U ERR8212210.25788 0 136 0:102
U ERR8212210.25789 0 143 0:109
C ERR8212210.25790 301301 133 301301:4 0:10 301301:5 0:50 301301:6 0:1 301301:2 0:3 301301:5 0:13
C ERR8212210.25791 537011 163 0:7 165179:6 0:9 165179:1 0:5 165179:5 0:13 165179:5 0:12 537011:5 0:6 537011:4 0:35 537011:5 0:11
C ERR8212210.25792 165179 126 0:10 165179:8 0:10 165179:1 0:12 165179:5 0:11 165179:3 0:5 165179:5 0:2 165179:5 0:3 165179:1 0:8 165179:3
U ERR8212210.25793 0 150 0:116
U ERR8212210.25794 0 163 0:129
U ERR8212210.25795 0 191 0:157
U ERR8212210.25796 0 205 0:171
C ERR8212210.25797 165179 185 0:41 165179:4 0:6 165179:5 0:1 165179:2 0:12
C ERR8212210.25798 537011 111 0:17 537011:2 0:5 537011:1 0:5 165179:5 0:1 537011:2 165179:5 0:1 165179:6 0:27
C ERR8212210.25799 165179 148 0:15 165179:5 0:5 165179:5 0:7 165179:2 0:2 165179:8 0:3 838:1 0:61
U ERR8212210.25800 0 138 0:101 537011:3
C ERR8212210.25801 165179 138 0:1 165179:5 0:26 165179:3 0:16 165179:15 0:10 165179:4 0:24
C ERR8212210.25802 186802 116 2:1 0:41 186802:5 0:12 186802:2 0:1 2:5 0:11 186802:3 0:1
C ERR8212210.25803 216851 142 0:4 216851:4 0:13 216851:5 0:7 216851:2 0:7 216851:3 0:10 216851:4 0:1 216851:5 0:2 216851:1 0:16 216851:1 0:12 216851:5 0:1 216851:2 0:3
U ERR8212210.25804 0 155 0:121
C ERR8212210.25805 537011 140 0:23 537011:3 0:25 165179:4 0:12 165179:3 0:10 165179:1 0:8 165179:6 0:8 165179:3
U ERR8212210.25806 0 128 0:94
C ERR8212210.25807 0 80 0:46
U ERR8212210.25808 165179 193 0:2 165179:2 0:15 165179:3 0:5 165179:1 0:4 165179:7 0:2 165179:5 0:15 165179:14 0:5 165179:5 0:37 165179:7 0:7 165179:6 0:17
C ERR8212210.25809 537011 162 0:25 171549:3 0:6 165179:5 537011:2 0:11 537011:2 0:15 537011:4 0:3 537011:5 0:44 537011:1 0:2
U ERR8212210.25810 0 126 0:92
C ERR8212210.25811 537011 138 0:38 537011:2 0:25 838:4 0:35
C ERR8212210.25812 165179 191 0:3 165179:10 0:33 165179:1 0:10 165179:3 0:14 838:5 0:5 165179:5 0:16 576514:3 0:44 165179:5
U ERR8212210.25813 0 177 0:138 537011:5
U ERR8212210.25814 0 117 0:1 2763672:5 0:77
U ERR8212210.25815 0 146 0:112
C ERR8212210.25816 537011 172 0:27 165179:6 0:45 537011:5 0:28 165179:5 0:22
C ERR8212210.25817 165179 137 0:11 165179:2 0:2 838:2 0:5 165179:3 0:6 165179:2 0:1 838:5 0:38 165179:5 0:16
U ERR8212210.25818 820 168 0:20:7 0:11 0:20:5 0:12 0:16:1 0:29 0:16:4 0:19 0:16:5 0:13 0:16:1 0:37 0:16:1 0:2 0:16:5 0:15
C ERR8212210.25819 0 138 0:23 963:4 0:77
C ERR8212210.25820 165179 178 0:82 165179:5 0:5 165179:5 0:47
U ERR8212210.25821 0 163 0:129
C ERR8212210.25822 33043 148 0:1 33042:2 0:5 33043:3 0:7 33042:5 0:22 33042:6 0:18 33042:1 0:8 33042:5 0:16 33043:2 0:5 33042:1 0:7
U ERR8212210.25823 0 156 0:102
C ERR8212210.25824 165179 126 0:5 165179:5 0:1 165179:5 0:44 838:1 0:31
U ERR8212210.25825 0 149 0:110 537011:2 0:3
C ERR8212210.25826 165179 140 0:19 165179:2 0:8 165179:4 0:73
C ERR8212210.25827 165179 118 0:1 165179:3 0:5 165179:2 0:18 165179:3 0:43 165179:1 0:8
U ERR8212210.25828 537011 140 0:40 537011:3 0:27 537011:5 0:19 537011:4
C ERR8212210.25829 0 126 0:92
C ERR8212210.25830 537011 151 0:74 165179:2 0:12 165179:1 0:22 537011:5 0:1
U ERR8212210.25831 0 119 0:85
C ERR8212210.25832 2763672 158 0:65 2763672:2 0:10 2763672:3 0:44
U ERR8212210.25833 0 144 0:110
U ERR8212210.25834 0 139 0:105
C ERR8212210.25835 537011 142 0:1 537011:1 0:16 537011:2 0:21 537011:4 0:10 537011:9 0:1 537011:5 0:1 165179:2 0:20 165179:4 0:11
U ERR8212210.25836 0 143 0:109

```

Figura 5.9: Ejemplo de archivo .kraken

```

0.00 9 0 K 2732005 Bamfordvirae
0.00 6 0 P 2732007 Nucleocytoviricota
0.00 6 0 C 2732523 Megaviricetes
0.00 6 0 O 2732554 Imitervirales
0.00 6 0 F 549779 Mimiviridae
0.00 6 0 F1 985780 unclassified Mimiviridae
0.00 4 0 S 985782 Moumouvirus
0.00 4 4 S1 2109587 Moumouvirus australiensis
0.00 2 2 S 2811091 Cottonvirus japonicus
0.00 3 0 P 2732008 Preplasmiviricota
0.00 3 0 C 2732529 Tectiliviricetes
0.00 3 0 O 2732559 Rowavirales
0.00 3 0 F 10508 Adenoviridae
0.00 3 0 G 10552 Aviadenovirus
0.00 3 0 S 1548417 Pigeon aviadenovirus A
0.00 3 3 S1 764030 Pigeon adenovirus 1
0.00 8 0 D1 2559587 Riboviria
0.00 8 0 K 2732396 Orthornavirae
0.00 7 0 P 2732407 Lenarviricota
0.00 7 0 C 2732499 Howeltoviricetes
0.00 7 0 O 2732503 Cryppavirales
0.00 7 0 F 2732892 Mitoviridae
0.00 7 0 G 2948686 Duamitovirus
0.00 7 0 S 2955788 Duamitovirus hean3
0.00 7 7 S1 2603543 Heterobasidion mitovirus 3
0.00 1 0 P 2497569 Negarnaviricota
0.00 1 0 P1 2497571 Polyploviricota
0.00 1 0 C 2497576 Elliioviricetes
0.00 1 0 O 1980410 Bunyavirales
0.00 1 0 F 1980412 Fimoviridae
0.00 1 0 G 675845 Emaravirus
0.00 1 0 G1 1086556 unclassified Emaravirus
0.00 1 1 S 2507689 Pistacia emaravirus

```

Figura 5.10: Reporte generado por clasificación Kraken2

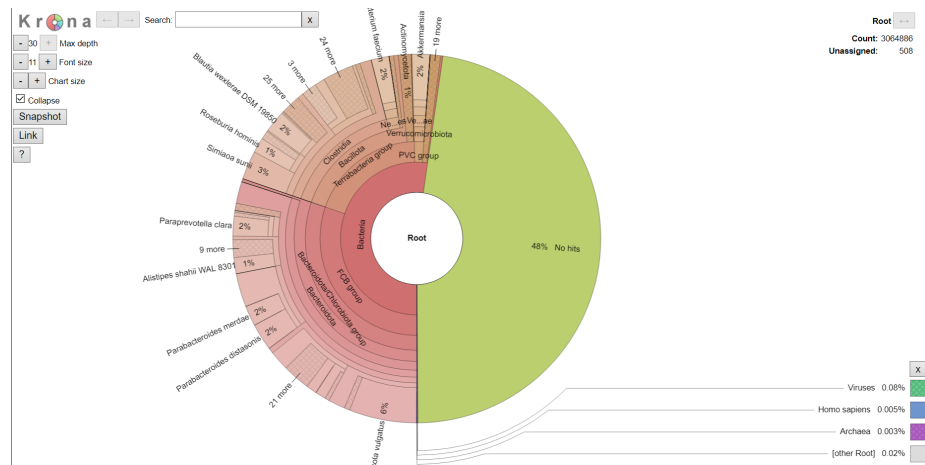


Figura 5.11: Archivo HTML creado por Krona

```

1 conda create -n krona -y
2 conda activate krona
3 conda install -c bioconda krona -y
4 sudo apt-get install make
5 ktUpdateTaxonomy.sh
6 mkdir kronaout
7 for infile in ./kraken2/*kraken
8 do
9   krakenname+=${infile}
10  krakenname+=" "
11 done
12 ktImportTaxonomy -q 2 -t 3 ${krakenname} -o ./kronaout/all_samples.kraken.html

```

Listing 5.14: Visualización con krona

El objetivo del script es el de generar un gráfico de Krona para todos los archivos de salida de Kraken2 en el directorio actual. Esto se hace a través del bucle *for* en donde se crea una variable con el nombre de todos estos archivos. A partir de esto, el comando `ktImportTaxonomy` genera un gráfico Krona a partir de los archivos de salida de Kraken2. La opción `-q` especifica el número de hilos a utilizar. La opción `-t` especifica el nivel taxonómico que se utilizará para el gráfico. La opción `-o` especifica el nombre del archivo de salida para el gráfico Krona.

Una vez que el script haya terminado de ejecutarse, tendrá un gráfico Krona llamado `all_samples.kraken.html` en el directorio `kronaout`. Puede abrir este archivo en un navegador web para ver el gráfico Krona.

El gráfico Krona mostrará la composición taxonómica de las muestras que ha analizado con Kraken2. Se puede utilizar el gráfico Krona para identificar los taxones más abundantes en las muestras, así como para identificar cualquier taxón inusual que pueda estar presente. Un ejemplo de estos gráficos se puede observar en la Figura 5.11. El output de Krona es un gráfico interactivo que muestra la composición taxonómica de una muestra. El gráfico se compone de un árbol jerárquico que representa las relaciones entre los diferentes taxa. El tamaño de cada nodo del árbol representa la abundancia del taxon correspondiente en la muestra.

## KrakenTools

KrakenTools es un conjunto de scripts que se pueden utilizar después de usar el software Kraken o Kraken2 para analizar datos metagenómicos. Proporciona una variedad de funciones, incluyendo:

- Filtrado de resultados
- Clasificación taxonómica
- Estimación de la abundancia

Sin embargo, se utilizará para hacer unión de los reportes de las secuencias analizadas con el *pipeline* para tener un archivo donde se concentre la información.

```

1 #krakenTools
2 conda create --name python3-env python -y
3 source activate python3-env
4 pip install numpy
5 pip install scipy
6 pip install argparse
7 pip install pandas
8 pip install collection
9 sudo apt-get install unzip
10
11 wget https://github.com/jenniferlu717/KrakenTools/archive/refs/heads/master.zip
12 unzip master.zip
13 rm master.zip
14 for infile in ./kraken2/*report
15 do
16 reportname+=${infile}
17 reportname+=" "
18 done
19 python KrakenTools-master/combine_kreports.py -r ${reportname} --only-combine -o
    combine.report
20 conda deactivate

```

**Listing 5.15:** Utilización de KrakenTools

Este código primero crea un nuevo entorno conda llamado python3-env. Luego activa el nuevo entorno e instala los siguientes paquetes de Python, los cuales son dependencias de KrakenTools:

- NumPy: Una biblioteca para la computación científica
- SciPy: Una biblioteca para la computación científica
- Argparse: Una biblioteca para analizar argumentos de línea de comandos
- Pandas: Una biblioteca para el análisis y la manipulación de datos
- Collection (Colección): Una biblioteca para trabajar con colecciones de datos

A continuación, el código instala el paquete unzip, necesario para descomprimir el archivo KrakenTools.

Luego, el código itera sobre todos los informes de Kraken en el directorio actual y los añade a una lista llamada reportname. A continuación, utiliza el script combine\_kreports.py de KrakenTools para combinar los informes en un único informe llamado combine.report. Este informe tendrá la misma estructura que la Figura 5.10. Esta sección puede ser útil si en dado caso el usuario desea usar un software ya creado para visualización de reportes generados por Kraken, como lo puede ser Pavian (<https://fbreitwieser.shinyapps.io/pavian/>). Sin embargo, en el proyecto se utilizará Shiny apps para el desarrollo de la visualización.

## Kraken-biom

Debido a que el formato kraken o de reporte resultante del software Kraken2, se es necesario convertir estos archivos a un formato más eficiente (json) y que sea interpretable por los programas que se desea usar. En este caso, el programa sería R.

Se utiliza Kraken-biom para la transformación deseada.

```
1 conda create -n kraken-biom -y
2 source activate kraken-biom
3 conda install -c bioconda kraken-biom -y
4 kraken-biom ${reportname} --fmt json -o combine.biom
```

**Listing 5.16:** Utilización de Kraken-biom

Este es un script bash que crea un entorno conda llamado kraken-biom, instala el paquete kraken-biom en ese entorno, y luego utiliza el paquete kraken-biom para convertir un informe Kraken o Kraken2 a un archivo BIOM en formato JSON.

Lo cual sería el último paso para la generación del *pipeline* deseado.

Para el caso de aplicación de pacientes de esquizofrenia se ve necesaria la descarga de un archivo TSV de cada uno de los códigos de acceso del proyecto PRJEB41217 de las muestras de pacientes con esquizofrenia y PRJEB41786 para las secuencias control.

Para obtener las secuencias se puede utilizar el comando de wget junto con el link de descarga de las secuencias proporcionado por la página de <https://www.ebi.ac.uk/ena/browser/view/PRJEB41786> para automatizar este proceso. Una vez se tenga en la misma carpeta las secuencias control y de esquizofrenia se debe de correr

```
1 ./pipeline.sh #nombre del archivo .sh creado con los codigos necesarios
```

**Listing 5.17:** Ejecución del archivo que contiene el *pipeline* creado

Para obtener los resultados descritos anteriormente, destacando la importancia de la creación del archivo .biom. Esta estructura final se puede observar en la Figura [5.12](#)

### 5.1.4. Visualización de resultados

Para la visualización de los archivos se creo una Shiny App, en donde se permite la carga de dos archivos de formato BIOM, de los cuales se podrá tener la visualización de gráficas de:

- Alfa diversidad
- Beta diversidad con diferentes metodos para medir distancia y la ordinación.
- Abundancia relativa y absoluta de las muestras ingresadas.

El *dashboard* desarrollado tiene un layout general presentado en la Figura [5.13](#) y se puede encontrar en el repositorio <https://github.com/marisaaaai/dashboardBiom>.

```

.:
KrakenTools-master combine.report fastp.html fastqc kraken2 no_host pipeline.sh stats.txt
combine.biom dbdownload fastp.json filtered_reads kronaout nohup.out sG.sh

./KrakenTools-master:
DiversityTools combine_kreports.py filter_bracken.out.py kreport2mpa.py
LICENSE combine_mpa.py fix_unmapped.py make_kreport.py
README.md extract_kraken_reads.py kreport2krona.py make_ktaxonomy.py

./KrakenTools-master/DiversityTools:
README.md alpha_diversity.py beta_diversity.py

./dbdownload:
database100mers.kmer_distrib database300mers.kmer_distrib inspect.txt taxo.k2d
database150mers.kmer_distrib database50mers.kmer_distrib ktaxonomy.tsv
database200mers.kmer_distrib database75mers.kmer_distrib opts.k2d
database250mers.kmer_distrib hash.k2d seqid2taxid.map

./fastqc:
ERR8212209_fastqc.html ERR8212256_fastqc.html ERR8212338_fastqc.html ERR8212446_fastqc.html
ERR8212209_fastqc.zip ERR8212256_fastqc.zip ERR8212338_fastqc.zip ERR8212446_fastqc.zip
ERR8212212_fastqc.html ERR8212280_fastqc.html ERR8212346_fastqc.html ERR8212452_fastqc.html
ERR8212212_fastqc.zip ERR8212280_fastqc.zip ERR8212346_fastqc.zip ERR8212452_fastqc.zip
ERR8212214_fastqc.html ERR8212284_fastqc.html ERR8212351_fastqc.html ERR8212471_fastqc.html
ERR8212214_fastqc.zip ERR8212284_fastqc.zip ERR8212351_fastqc.zip ERR8212471_fastqc.zip

./filtered_reads:
ERR8212209_filt.fastq.gz ERR8212271_filt.fastq.gz ERR8212346_filt.fastq.gz ERR8212471_filt.fastq.gz
ERR8212212_filt.fastq.gz ERR8212280_filt.fastq.gz ERR8212351_filt.fastq.gz ERR8212475_filt.fastq.gz
ERR8212214_filt.fastq.gz ERR8212284_filt.fastq.gz ERR8212356_filt.fastq.gz ERR8212479_filt.fastq.gz
ERR8212216_filt.fastq.gz ERR8212286_filt.fastq.gz ERR8212364_filt.fastq.gz ERR8212484_filt.fastq.gz
ERR8212218_filt.fastq.gz ERR8212287_filt.fastq.gz ERR8212370_filt.fastq.gz ERR8212504_filt.fastq.gz
ERR8212221_filt.fastq.gz ERR8212290_filt.fastq.gz ERR8212375_filt.fastq.gz ERR8212506_filt.fastq.gz
ERR8212222_filt.fastq.gz ERR8212305_filt.fastq.gz ERR8212384_filt.fastq.gz ERR8212516_filt.fastq.gz
ERR8212226_filt.fastq.gz ERR8212306_filt.fastq.gz ERR8212387_filt.fastq.gz ERR8212520_filt.fastq.gz
ERR8212230_filt.fastq.gz ERR8212311_filt.fastq.gz ERR8212389_filt.fastq.gz ERR8212523_filt.fastq.gz
ERR8212242_filt.fastq.gz ERR8212320_filt.fastq.gz ERR8212391_filt.fastq.gz ERR8212525_filt.fastq.gz
ERR8212250_filt.fastq.gz ERR8212325_filt.fastq.gz ERR8212420_filt.fastq.gz ERR8212527_filt.fastq.gz
ERR8212251_filt.fastq.gz ERR8212331_filt.fastq.gz ERR8212434_filt.fastq.gz
ERR8212254_filt.fastq.gz ERR8212337_filt.fastq.gz ERR8212446_filt.fastq.gz
ERR8212256_filt.fastq.gz ERR8212338_filt.fastq.gz ERR8212452_filt.fastq.gz

./kraken2:
ERR8212209.kraken ERR8212251.kraken ERR8212311.kraken ERR8212375.kraken ERR8212479.kraken
ERR8212209.report ERR8212251.report ERR8212311.report ERR8212375.report ERR8212479.report
ERR8212212.kraken ERR8212254.kraken ERR8212320.kraken ERR8212384.kraken ERR8212484.kraken
ERR8212212.report ERR8212254.report ERR8212320.report ERR8212384.report ERR8212484.report
ERR8212214.kraken ERR8212256.kraken ERR8212325.kraken ERR8212387.kraken ERR8212504.kraken
ERR8212214.report ERR8212256.report ERR8212325.report ERR8212387.report ERR8212504.kraken
ERR8212216.kraken ERR8212271.kraken ERR8212331.kraken ERR8212389.kraken ERR8212506.kraken
ERR8212216.report ERR8212271.report ERR8212331.report ERR8212389.report ERR8212506.report
ERR8212218.kraken ERR8212280.kraken ERR8212337.kraken ERR8212391.kraken ERR8212516.kraken
ERR8212218.report ERR8212280.report ERR8212337.report ERR8212391.report ERR8212516.report
ERR8212221.kraken ERR8212284.kraken ERR8212338.kraken ERR8212420.kraken ERR8212520.kraken
ERR8212221.report ERR8212284.report ERR8212338.report ERR8212420.report ERR8212520.report

./no_host:
ERR8212209_fastq.gz ERR8212251_fastq.gz ERR8212311_fastq.gz ERR8212375_fastq.gz ERR8212479_fastq.gz
ERR8212212_fastq.gz ERR8212254_fastq.gz ERR8212320_fastq.gz ERR8212384_fastq.gz ERR8212484_fastq.gz
ERR8212214_fastq.gz ERR8212256_fastq.gz ERR8212325_fastq.gz ERR8212387_fastq.gz ERR8212504_fastq.gz
ERR8212216_fastq.gz ERR8212271_fastq.gz ERR8212331_fastq.gz ERR8212389_fastq.gz ERR8212506_fastq.gz
ERR8212218_fastq.gz ERR8212280_fastq.gz ERR8212337_fastq.gz ERR8212391_fastq.gz ERR8212516_fastq.gz
ERR8212221_fastq.gz ERR8212284_fastq.gz ERR8212338_fastq.gz ERR8212420_fastq.gz ERR8212520_fastq.gz
ERR8212222_fastq.gz ERR8212286_fastq.gz ERR8212346_fastq.gz ERR8212434_fastq.gz ERR8212523_fastq.gz
ERR8212226_fastq.gz ERR8212287_fastq.gz ERR8212346_fastq.gz ERR8212446_fastq.gz ERR8212525_fastq.gz
ERR8212230_fastq.gz ERR8212290_fastq.gz ERR8212356_fastq.gz ERR8212452_fastq.gz ERR8212527_fastq.gz

```

Figura 5.12: Ejemplo de la estructura final de archivos a obtener

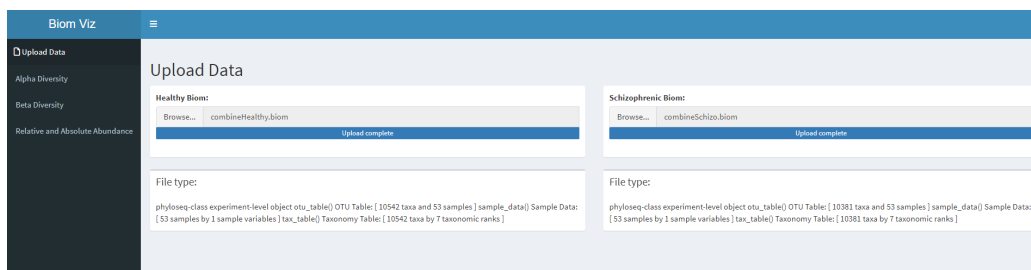


Figura 5.13: Layout principal de *Dashboard* generado por Shiny App

A partir de los dos archivos .BIOM y del *dashboard* creado en Shiny se obtuvieron gráficas que hablan acerca de la alfa y beta diversidad. Como también de la abundancia relativa y absoluta.

## 6.1. Alfa diversidad

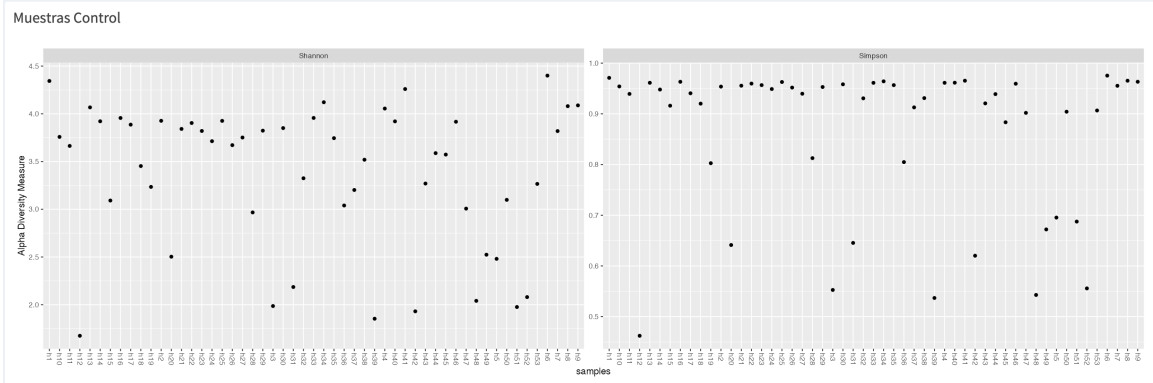
En las Figuras [6.1](#) y [6.2](#) podemos observar las gráficas de índices de Shannon y Simpson de las muestras control y esquizofrénicas respectivamente. que cuentan acerca de la alfa diversidad de las muestras. Se puede observar que para las muestras control, se encuentra que para el índice de Shannon la mayoría se encuentra dentro del rango de 3 -4. Mientras que para el índice de Simpson la mayoría se ve que estan entre 0.9 - 1. Por otro lado, las muestras con esquizofrenía muestran ser diferentes, en el índice de Shannon las muestras estar esparcidas entre valores 2.5 y 4, la mayoría estando  $< 3$ ; mientras que en el índice de Simpson también se ven dentro del ranago de 0.9 y 1.

## 6.2. Beta diversidad

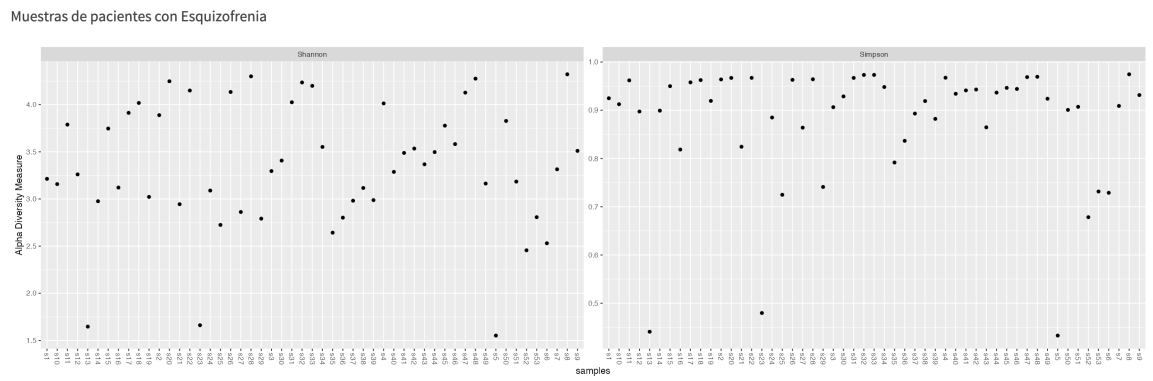
Al momento de realizar las gráficas de diversidad beta estas pueden ser construidas con diferentes métodos y con diferentes mediciones de disimilaridad. En la aplicación presente se utilizó el método de Bray-Curtis con los métodos de análisis de:

- Correspondencias destrendado (DCA)
- Análisis de correspondencia (CA)
- Método de escalamiento multidimensional no métrico (NMDS).

Como resultado se obtuvo las gráficas que se describen a continuación. En ambos conjuntos de las muestras se puede observar que en las graficas del DCA en la Figura [6.3](#) las muestras se ven

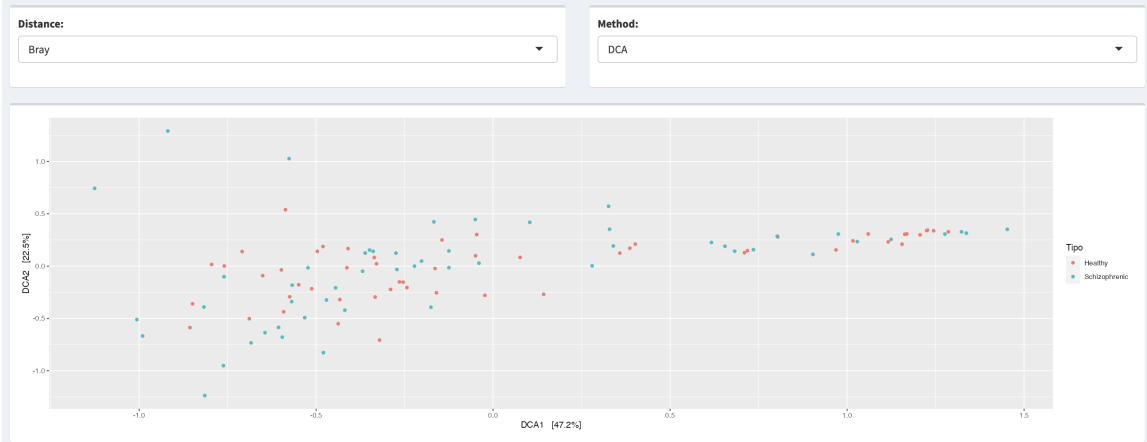


**Figura 6.1:** Alfa diversidad para muestras control. Graficando en eje y la medida de alfa diversidad y en el eje x las muestras control. Teniendo del lado izquierdo medida del índice de Shannon y del lado derecho del índice de Simpson.

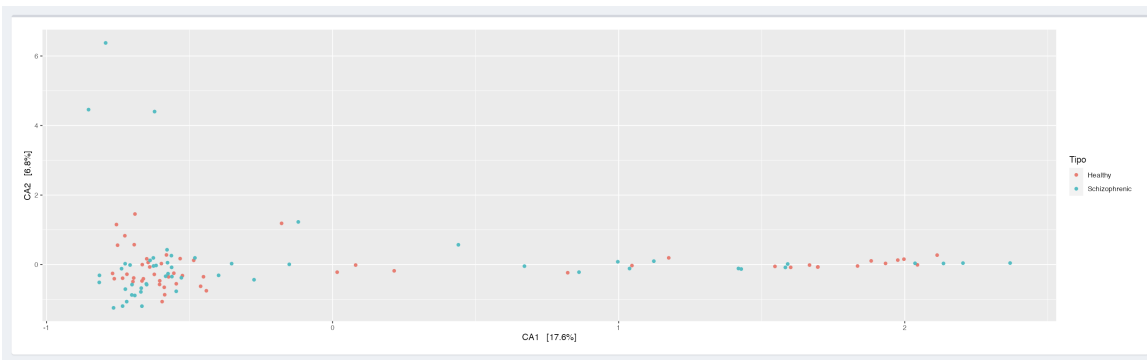


**Figura 6.2:** Alfa diversidad para muestras de pacientes esquizofrénicos. Graficando en eje y la medida de alfa diversidad y en el eje x las muestras esquizofrénicas. Teniendo del lado izquierdo medida del índice de Shannon y del lado derecho del índice de Simpson.

## Gráficas de Beta Diversidad



**Figura 6.3:** Comparación de Beta diversidad entre los grupos de muestras control (rosa) y muestras esquizofrénicas (celeste) utilizando el método DCA encontrando una variabilidad entre las muestras de 69.7%.



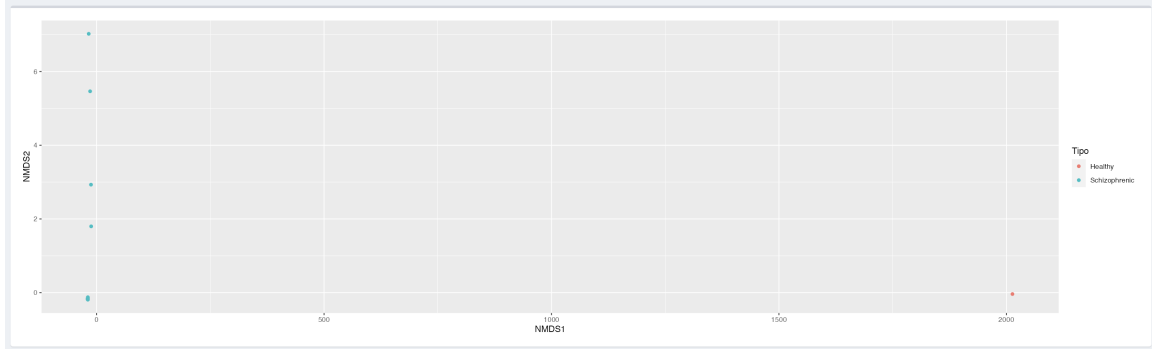
**Figura 6.4:** Comparación de Beta diversidad entre los grupos de muestras control (rosa) y muestras esquizofrénicas (celeste) utilizando el método CCA encontrando una variabilidad entre las muestras de 24.4%

esparcidas a lo largo del gráfico, siguiendo un patrón similar. Mientras que por otro lado, se comienza a ver un grado de agrupación en las gráficas del CA en la Figura 6.4 de las muestras. Finalmente con el método de NMDS la agrupación de muestras es más extrema en la Figura 6.5

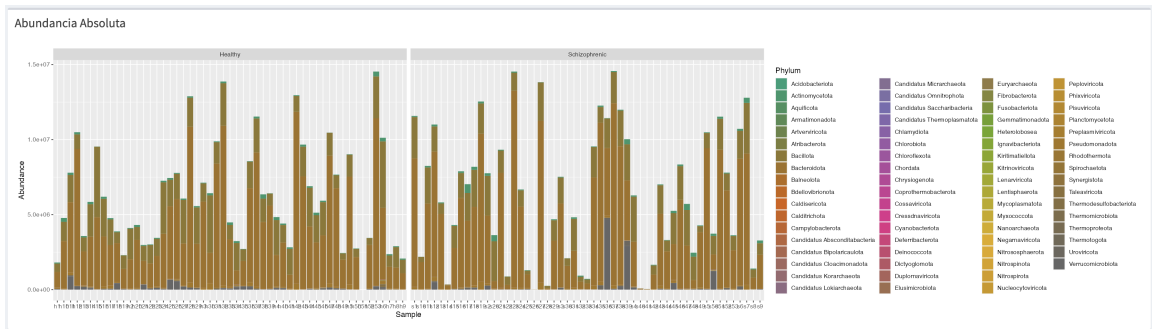
## 6.3. Abundancia

### 6.3.1. Abundancia absoluta de filos

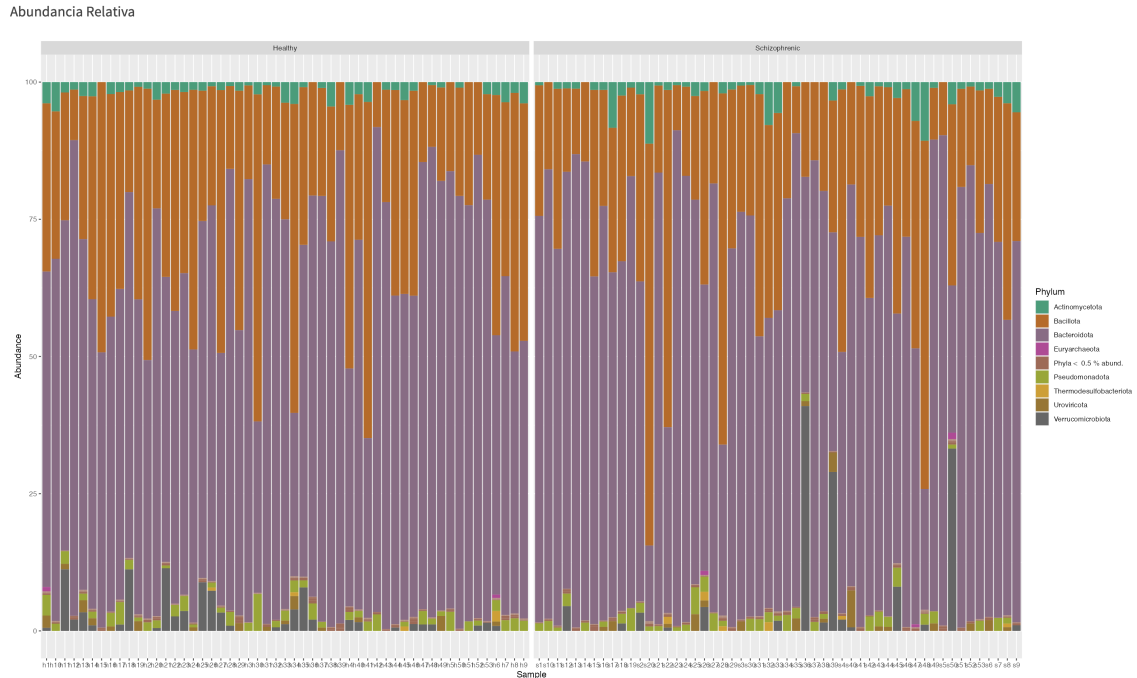
Por otro lado, con el *dashboard* de Shiny, se obtuvieron los resultados vistos en las Figura 6.6 que explican la abundancia absoluta de los dos conjuntos de muestras. Sin embargo, dado a que se clasificó muchas taxonomías de bacterias, es esencial hacer un filtro en la operación de diversidad relativa para tener una mejor visualización de los datos.



**Figura 6.5:** Comparación de Beta diversidad entre los grupos de muestras control (rosa) y muestras esquizofrénicas (celeste) utilizando el método NMDS.



**Figura 6.6:** Abundancia absoluta de las muestras por filos, agrupados por tipo de muestras en conjunto con la leyenda generada con los filos de bacterias identificadas.



**Figura 6.7:** Abundancia relativa con filtro aplicado de abundancia  $0.5\% >$  de las muestras por filos, agrupados por tipo de muestras en conjunto con la leyenda generada con los filos de bacterias identificada.

### 6.3.2. Abundancia relativa de filos

Para tener un mejor filtrado y visualización de los porcentajes de las bacterias dentro de los conjuntos de muestras se puede establecer un filtro que solo se muestren los filos de bacterias con una abundancia mayor a una cantidad entre 0 - 10. En el caso del presente proyecto se escogió el porcentaje inicial de 0.5 % lo cual dió como resultado lo visto en la Figura 6.7 para las muestras control y de pacientes esquizofrénicos respectivamente. De estos resultados se puede observar que para las muestras control el filo de bacterias más presentes son *Actinomycetota*, *Bacillota*, *Bacteroidota* y *Euryarchaeota*. Mientras que los menos abundantes son *Pseudomonadota*, *Thermodesulfobacteriota*, *Uroviricota* y *Verrucomicrobiota*.

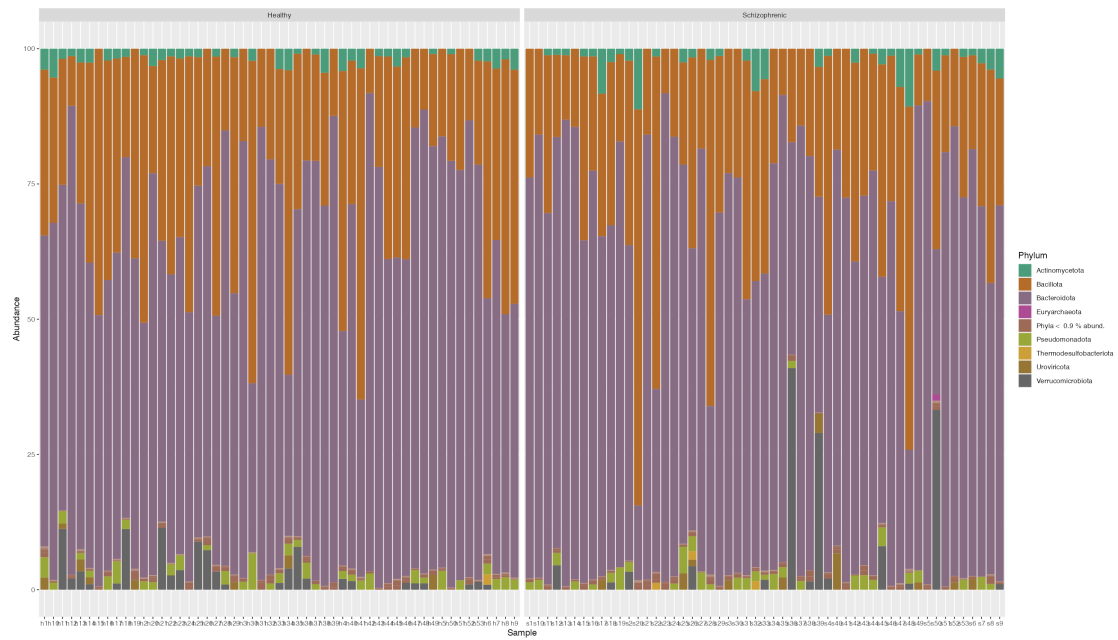
Estos filos prevalecieron con porcentajes de abundancia desde el 10 % al 80 %. Sin embargo al acercarse al 90 % se obtiene que el filo *Euryarchaeota* disminuye bastante su presencia en la muestra control y todavía prevalece en las muestras de pacientes con esquizofrenía.

### 6.3.3. Abundancia relativa de géneros

Se realizaron gráficos y tablas con el que se pueda obtener una abundancia relativa basada en el Género de las bacteria. Las tablas desarrollados con Rshiny pueden observarse en la Figura 6.9 Sin embargo, por motivos de una comprensión más sencilla se insertaran como tablas en este archivo y no como imagen.

En la Figura 6.10 y en la tabla de la Figura 6.11 se encuentran los géneros de bacteria que se encuentran en aproximadamente el 95 % de la comunidad de las muestras. En la tabla se puede observar el tipo de muestra al que pertenece y su frecuencia. Las celdas coloreadas de verde opaco significa que el género se encuentra en ambos tipos de muestras, el verde más intenso es un caso

Abundancia Relativa



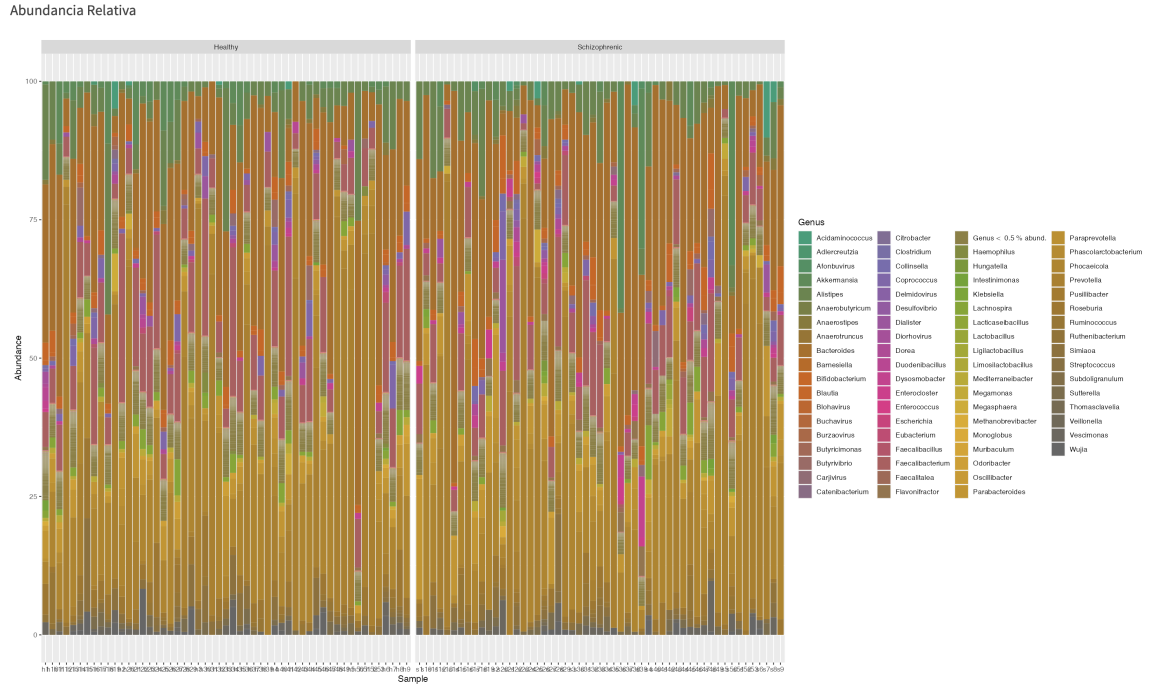
**Figura 6.8:** Abundancia relativa con filtro aplicado de abundancia  $0.9\% >$  de las muestras por filos, agrupados por tipo de muestras en conjunto con la leyenda generada con los filos de bacterias identificada.

Tabla Resumen

Show 100 entidades Search: schizo

Genus	Tipo	rec_count
Genus > 0.5 % abund.	Schizophrenic	132582
Bacteroides	Schizophrenic	53
Parabacteroides	Schizophrenic	51
Phocaeicola	Schizophrenic	51
Alistipes	Schizophrenic	50
Faecalibacterium	Schizophrenic	49
Blautia	Schizophrenic	45
Roseburia	Schizophrenic	45
Blifidobacterium	Schizophrenic	32
Dysosmobacter	Schizophrenic	32
Odoribacter	Schizophrenic	30
Simiaoa	Schizophrenic	28
Lachnospira	Schizophrenic	25
Vesicimonas	Schizophrenic	25
Flavonifractor	Schizophrenic	24
Phascolarctobacterium	Schizophrenic	24
Mediterraneibacter	Schizophrenic	23
Prevotella	Schizophrenic	23
Coprococcus	Schizophrenic	21

**Figura 6.9:** Ejemplo de tabla en Shiny



**Figura 6.10:** Abundancia relativa con filtro aplicado de abundancia 0.5 %> de las muestras por géneros, agrupados por tipo de muestras en conjunto con la leyenda generada con los géneros de bacterias identificada.

interesante y las celdas no coloreadas presentan géneros que solo se encuentran en un tipo de muestra.

De esta tabla se puede observar que los géneros *Haemophilus*, *Adlercreutzia*, *Barnesiella*, *Blohavirus*, *Diorhavirus* y *Muribaculum* solo se encuentran en las muestras control. Por otro lado, los géneros *Escheria*, *Alfonbuvirus*, *Oscillibacter*, *Thomasclavelia*, *Buchavirus*, *Citrobacter*, *Delmidovirus*, *Enterococcus*, *Faecalitalea*, *Hungatella*, *Lacticaseibacillus*, *LactoBacillus*, *limosilactobacillus* y *Monoglobus* las que solamente se encuentran en las muestras de pacientes con esquizofrenia. Teniendo solamente a *Flavonifactor* como un caso peculiar.

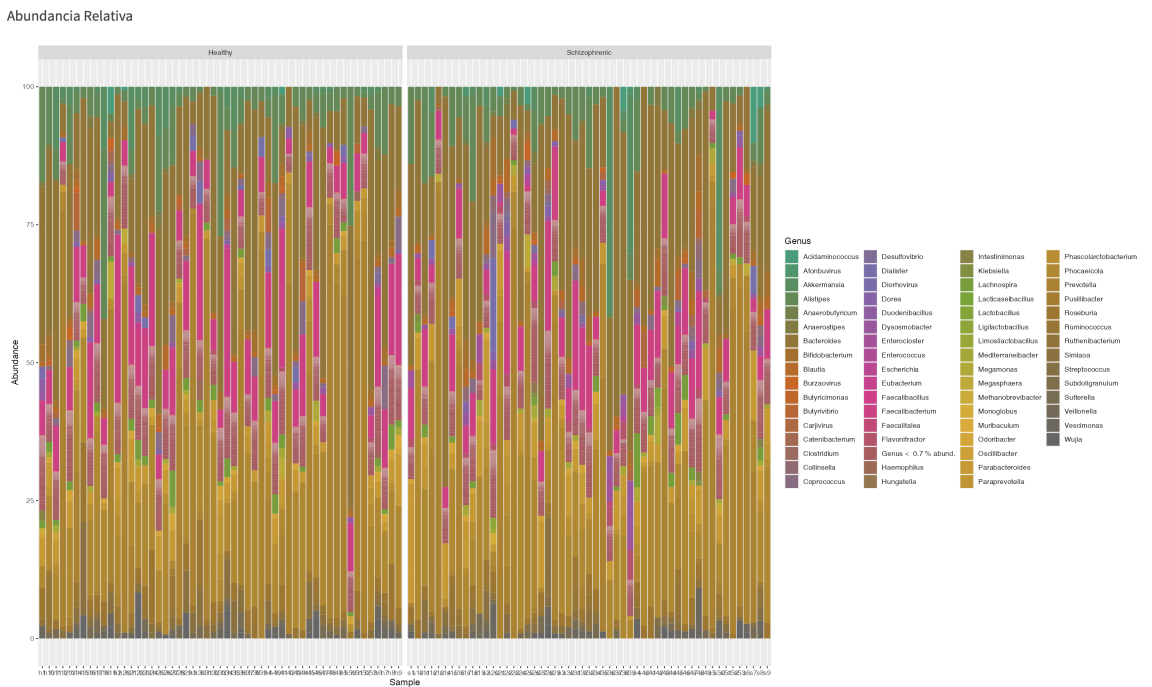
En las figuras 6.12 y 6.13 se pueden observar los géneros que pasaron un filtro aún mayor, de 0.7 %. Aquí se puede observar los géneros *Haemophilus*, *Diorhavirus* y *Muribaculum* presentaron una prevalencia al filtro del conjunto de datos control. Mientras que los géneros *Escheria*, *Alfonbuvirus*, *Oscillibacter*, *Enterococcus*, *Faecalitalea*, *Hungatella*, *LacticaeiBacillus*, *Lactobacillus*, *Ligilactobacillus*, *Limosilactobacillus* y *Monoglobus* son los géneros que no aparecen en la muestra control. En este caso existen celdas marcadas en rojo ya que son los géneros que pasaron un segundo filtro. Manteniendo a *Flavonifactor* como un caso peculiar.

En las figuras 6.14 y 6.15 encontramos a los géneros que presentan una abundancia relativa mayor al 2% de la muestra. De este filtro se nota que los géneros *Duodenibacillus*, *Diorhavirus*, *Haemophilus* y *Muribaculum* no se presentan en el grupo de muestras esquizofrénicas, siendo la tercera vez de los últimos tres mencionados que están ausentes. En la muestra de los pacientes con esquizofrenia se nota que los géneros *Dysosmobacter*, *Flavinofactor*, *Enterocloster*, *Dorea*, *Mediterraneibacter*, *Streptococcus*, *Catenibacterium*, *Escheria*, *Lactobacillus* y *Ruthenibacterium* son los ausentes en la muestra control. Notando que sería el segundo filtro para *Dysosmobacter* y *Enterocloster* y el tercero para *Escheria* y *Lactobacillus* por el que pasan.

Finalmente, el último filtro aplicado fue el de que los géneros deben de tener una abundancia mayor del 5%. Estos resultados pueden ser visualizados en las figuras 6.16 y 6.17. Aquí se puede

Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia
Bacteroides	Healthy	53	Anaerostipes	Healthy	15	Faecalibacillus	Healthy	2	Bacteroides	Schizophrenic	53	Collinsella	Schizophrenic	18
Faecalibacterium	Healthy	53	Dysosmobacter	Healthy	15	Megamonas	Healthy	2	Parabacteroides	Schizophrenic	51	Paraprevotella	Schizophrenic	18
Alitipes	Healthy	51	Collinsella	Healthy	12	Methanobrevibacter	Healthy	2	Phocaeicola	Schizophrenic	51	Akkermansia	Schizophrenic	18
Parabacteroides	Healthy	51	Dialister	Healthy	12	Puillibacter	Healthy	2	Alitipes	Schizophrenic	50	Butyrivibrio	Schizophrenic	18
Phocaeicola	Healthy	51	Butyrivibrio	Healthy	12	Adereocetza	Healthy	1	Faecalibacterium	Schizophrenic	48	Wuja	Schizophrenic	18
Roseburia	Healthy	51	Ruthenibacterium	Healthy	12	Anaerobutylicum	Healthy	1	Blaustia	Schizophrenic	48	Enterocloster	Schizophrenic	11
Blaustia	Healthy	43	Streptococcus	Healthy	8	Anaerotruncus	Healthy	1	Roseburia	Schizophrenic	45	Ruthenibacterium	Schizophrenic	11
Simiaca	Healthy	46	Eubacterium	Healthy	6	Barnesiella	Healthy	1	Bifidobacterium	Schizophrenic	32	Anaerobutylicum	Schizophrenic	11
Bifidobacterium	Healthy	31	Flavonifractor	Healthy	6	Biohavirus	Healthy	1	Dysosmobacter	Schizophrenic	32	Streptococcus	Schizophrenic	6
Lachnospira	Healthy	36	Haemophilus	Healthy	6	Diorhavirus	Healthy	1	Odoribacter	Schizophrenic	30	Acidaminococcus	Schizophrenic	6
Odoribacter	Healthy	36	Butyrivibrio	Healthy	5	Klebsiella	Healthy	1	Simiaca	Schizophrenic	28	Carjivirus	Schizophrenic	6
Vesicimonas	Healthy	30	Carjivirus	Healthy	5	Ligilactobacillus	Healthy	1	Lachnospira	Schizophrenic	25	Catenibacterium	Schizophrenic	6
Coprococcus	Healthy	33	Duodenibacillus	Healthy	5	Muribaculum	Healthy	1	Vesicimonas	Schizophrenic	25	Desulfobivrio	Schizophrenic	6
Akkermansia	Healthy	29	Acidaminococcus	Healthy	4	Subdoligranulum	Healthy	1	Megamonas	Schizophrenic	24	Dialister	Schizophrenic	6
Prevotella	Healthy	21	Burzaovirus	Healthy	4				Phascolarctobacterium	Schizophrenic	24	Megamonas	Schizophrenic	6
Sutterella	Healthy	21	Desulfobivrio	Healthy	4				Mediteraneibacter	Schizophrenic	22	Megaphaera	Schizophrenic	6
Mediteraneibacter	Healthy	25	Megaphaera	Healthy	4				Prevotella	Schizophrenic	23	Burzaovirus	Schizophrenic	5
Phascolarctobacterium	Healthy	21	Catenibacterium	Healthy	3				Coprococcus	Schizophrenic	21	Clostridium	Schizophrenic	5
Dorea	Healthy	19	Clostridium	Healthy	3				Sutterella	Schizophrenic	21	Escherichia	Schizophrenic	5
Ruminococcus	Healthy	18	Intestinimonas	Healthy	3				Dorea	Schizophrenic	17	Eubacterium	Schizophrenic	5
Wuja	Healthy	18	Valloniella	Healthy	3				Ruminococcus	Schizophrenic	17	Intestinimonas	Schizophrenic	4
Paraprevotella	Healthy	18	Enterocloster	Healthy	2				Anaerostipes	Schizophrenic	16	Klebsiella	Schizophrenic	4
												Monoglobus	Schizophrenic	1

**Figura 6.11:** Tabla del resumen generado por Shiny de los géneros de bacteria identificados que cumplen con el filtro de abundancia 0.5% >. En la tabla los géneros resaltados con verde son los géneros presentes en ambos grupos de análisis (Control y esquizofrénicos); con verde oscuro los géneros presentes en ambos pero con una diferencia de frecuencia interesante; y los que no están resaltados son los géneros que no están presentes en el otro grupo.

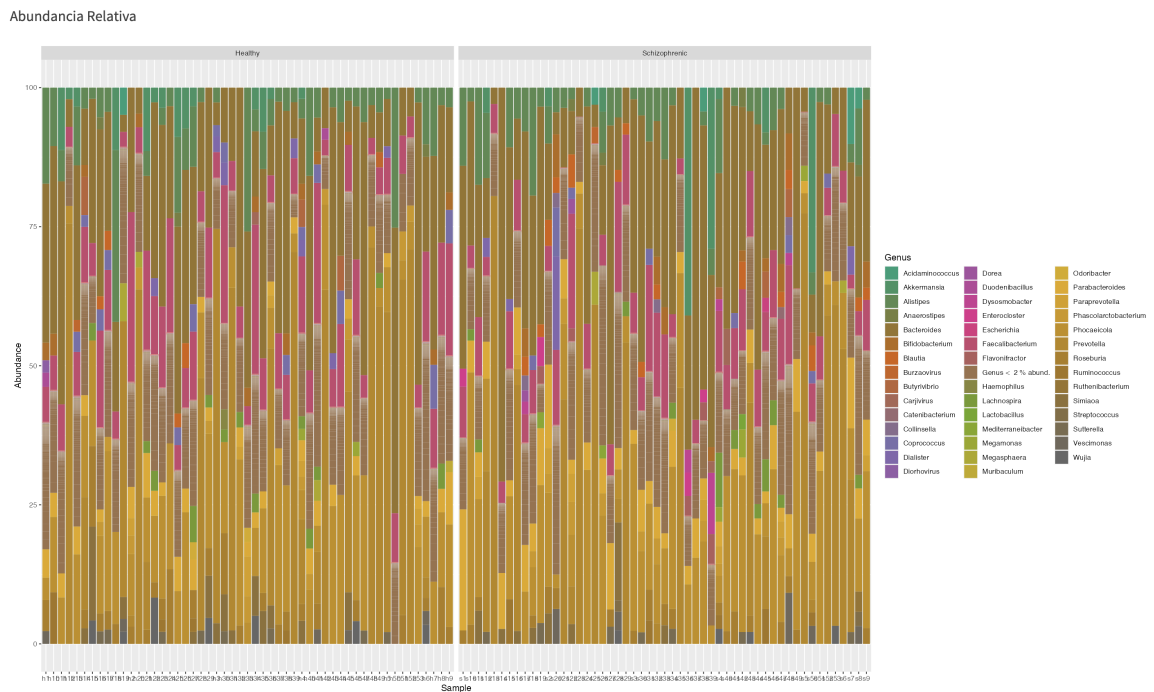


**Figura 6.12:** Abundancia relativa con filtro aplicado de abundancia 0.7% > de las muestras por géneros, agrupados por tipo de muestras en conjunto con la leyenda generada con los géneros de bacterias identificada.

Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia
Bacteroides	Healthy	53	Dialister	Healthy	10	Clostridium	Healthy	1
Faecalibacterium	Healthy	53	Dorea	Healthy	10	Diorhiovirus	Healthy	1
Alitipes	Healthy	53	Dysosmobacter	Healthy	9	Intestinimonas	Healthy	1
Phocaicola	Healthy	51	Collinsella	Healthy	7	Methanodoreaibacter	Healthy	1
Parabacteroides	Healthy	40	Butyrivibrio	Healthy	6	Muribaculum	Healthy	1
Roseburia	Healthy	47	Butyrivibrio	Healthy	5			
Blautia	Healthy	39	Duodenibacillus	Healthy	5			
Bifidobacterium	Healthy	34	Ruthenibacterium	Healthy	5			
Simiaoa	Healthy	33	Burzaovirus	Healthy	4			
Lachnospira	Healthy	32	Carjivirus	Healthy	4			
Odoribacter	Healthy	29	Haemophilus	Healthy	4			
Coprococcus	Healthy	26	Acidaminococcus	Healthy	3			
Prevotella	Healthy	26	Eubacterium	Healthy	3			
Akkermansia	Healthy	25	Flavonifractor	Healthy	3			
Vesicimonas	Healthy	25	Megasphaera	Healthy	3			
Sutterella	Healthy	24	Streptococcus	Healthy	3			
Phascolarctobacterium	Healthy	21	Desulfovibrio	Healthy	2			
Mediterraneibacter	Healthy	15	Megamonas	Healthy	2			
Paraprevotella	Healthy	15	Pullibacter	Healthy	2			
Ruminococcus	Healthy	14	Veillonella	Healthy	2			
Wujiang	Healthy	13	Anaerobutyricum	Healthy	1			
Anaerostipes	Healthy	10	Catenibacterium	Healthy	1			

Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia
Bacteroides	Schizophrenic	53	Anaerostipes	Schizophrenic	10	Afonavirus	Schizophrenic	2
Phocaicola	Schizophrenic	51	Collinsella	Schizophrenic	10	Clostridium	Schizophrenic	2
Parabacteroides	Schizophrenic	49	Dorea	Schizophrenic	10	Faecalibacillus	Schizophrenic	2
Faecalibacterium	Schizophrenic	48	Enterococcus	Schizophrenic	10	Intestinimonas	Schizophrenic	2
Alitipes	Schizophrenic	47	Wujiang	Schizophrenic	10	Methanobrevibacter	Schizophrenic	2
Roseburia	Schizophrenic	42	Butyrivibrio	Schizophrenic	9	Oscillibacter	Schizophrenic	2
Blautia	Schizophrenic	40	Ruthenibacterium	Schizophrenic	7	Veillonella	Schizophrenic	2
Bifidobacterium	Schizophrenic	26	Streptococcus	Schizophrenic	7	Enterococcus	Schizophrenic	1
Dysosmobacter	Schizophrenic	24	Acidaminococcus	Schizophrenic	6	Faecalitalea	Schizophrenic	1
Simiaoa	Schizophrenic	24	Burzaovirus	Schizophrenic	6	Hungatella	Schizophrenic	1
Phascolarctobacterium	Schizophrenic	23	Carjivirus	Schizophrenic	6	Lactocaseibacillus	Schizophrenic	1
Prevotella	Schizophrenic	23	Dialister	Schizophrenic	6	Lactobacillus	Schizophrenic	1
Odobacter	Schizophrenic	22	Megamonas	Schizophrenic	6	Lagobacterium	Schizophrenic	1
Vesicimonas	Schizophrenic	22	Anaerobutyricum	Schizophrenic	6	Uromolactobacillus	Schizophrenic	1
Lachnospira	Schizophrenic	21	Catenibacterium	Schizophrenic	4	Monoglobus	Schizophrenic	1
Flavonifractor	Schizophrenic	18	Desulfovibrio	Schizophrenic	4	Pullibacter	Schizophrenic	1
Mediterraneibacter	Schizophrenic	18	Escherichia	Schizophrenic	4	Subdoligranulum	Schizophrenic	1
Sutterella	Schizophrenic	18	Klebsiella	Schizophrenic	4			
Coprococcus	Schizophrenic	17	Megasphaera	Schizophrenic	4			
Ruminococcus	Schizophrenic	17	Butyrivibrio	Schizophrenic	3			
Paraprevotella	Schizophrenic	15	Duodenibacillus	Schizophrenic	3			
Akkermansia	Schizophrenic	13	Eubacterium	Schizophrenic	3			

**Figura 6.13:** Tabla del resumen generado por Shiny de los géneros de bacteria identificados que cumplen con el filtro de abundancia  $0.7\% >$ . En la tabla los géneros resaltados con verde son los géneros presentes en ambos grupos de análisis (Control y esquizofrénicos); con rojo los géneros ausentes del otro grupo por segunda vez; y los que no están resaltados son los géneros que no están presentes en el otro grupo.



**Figura 6.14:** Abundancia relativa con filtro aplicado de abundancia  $2\% >$  de las muestras por géneros, agrupados por tipo de muestras en conjunto con la leyenda generada con los géneros de bacterias identificada.

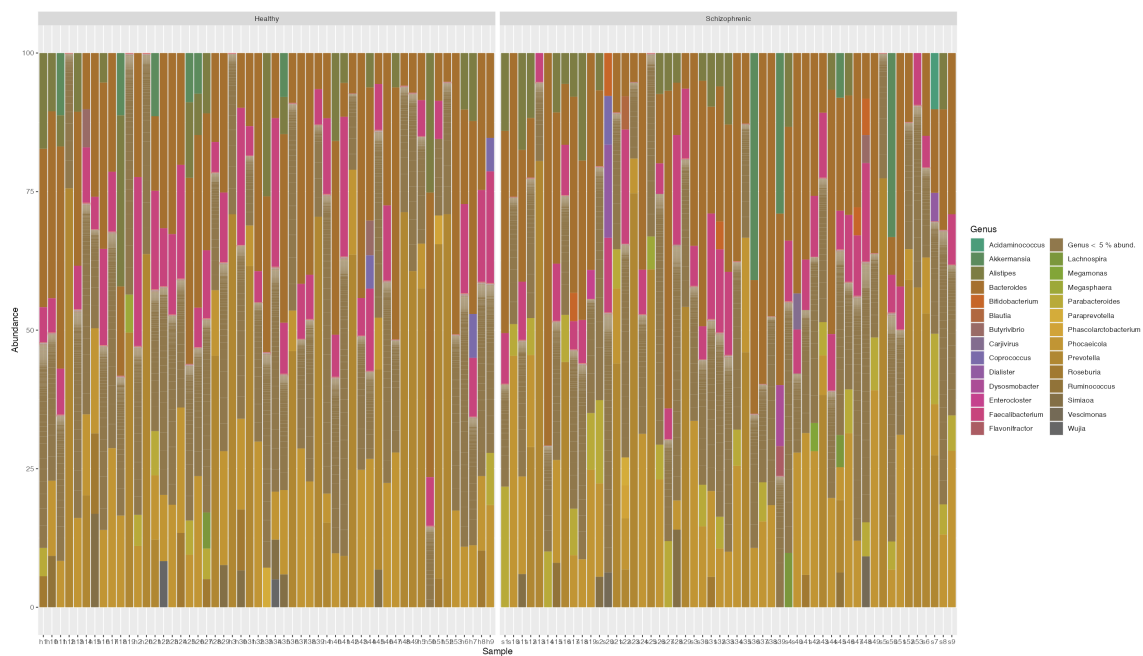
Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia
Bacteroides	Healthy	53	Megasphaera	Healthy	3
Faecalibacterium	Healthy	53	Burzaovirus	Healthy	2
Phocaeicola	Healthy	48	Duodenibacillus	Healthy	2
Alistipes	Healthy	41	Megamonas	Healthy	2
Parabacteroides	Healthy	32	Odoribacter	Healthy	2
Roseburia	Healthy	25	Acidaminococcus	Healthy	1
Prevotella	Healthy	21	Carjivirus	Healthy	1
Akkermansia	Healthy	14	Diorhavirus	Healthy	1
Coproccoccus	Healthy	14	Haemophilus	Healthy	1
Lachnospira	Healthy	13	Muribaculum	Healthy	1
Phascolarctobacterium	Healthy	13			
Simiaoa	Healthy	13			
Bifidobacterium	Healthy	12			
Sutterella	Healthy	8			
Dialister	Healthy	7			
Paraprevotella	Healthy	7			
Wujia	Healthy	7			
Vescimonas	Healthy	5			
Blautia	Healthy	4			
Ruminococcus	Healthy	4			
Anaerostipes	Healthy	3			
Butyrivibrio	Healthy	3			

Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia
Bacteroides	Schizophrenic	53	Collinsella	Schizophrenic	4
Phocaeicola	Schizophrenic	48	Dialister	Schizophrenic	4
Parabacteroides	Schizophrenic	45	Enterocloster	Schizophrenic	4
Faecalibacterium	Schizophrenic	44	Wujia	Schizophrenic	4
Alistipes	Schizophrenic	38	Butyrivibrio	Schizophrenic	3
Prevotella	Schizophrenic	19	Megamonas	Schizophrenic	3
Roseburia	Schizophrenic	19	Carjivirus	Schizophrenic	2
Blautia	Schizophrenic	15	Dorea	Schizophrenic	2
Bifidobacterium	Schizophrenic	11	Mediterraneibacte	Schizophrenic	2
Lachnospira	Schizophrenic	11	Megasphaera	Schizophrenic	2
Phascolarctobacterium	Schizophrenic	10	Streptococcus	Schizophrenic	2
Akkermansia	Schizophrenic	8	Burzaovirus	Schizophrenic	1
Coproccoccus	Schizophrenic	8	Catenibacterium	Schizophrenic	1
Paraprevotella	Schizophrenic	8	Escherichia	Schizophrenic	1
Simiaoa	Schizophrenic	8	Lactobacillus	Schizophrenic	1
Sutterella	Schizophrenic	7	Odoribacter	Schizophrenic	1
Dysosmobacter	Schizophrenic	5	Ruthenibacterium	Schizophrenic	1
Flavinifactor	Schizophrenic	5			
Ruminococcus	Schizophrenic	5			
Vescimonas	Schizophrenic	5			
Acidaminococcus	Schizophrenic	4			
Anaerostipes	Schizophrenic	4			

**Figura 6.15:** Tabla del resumen generado por Shiny de los géneros de bacteria identificados que cumplen con el filtro de abundancia  $2\% >$ . En la tabla los géneros resaltados con verde son los géneros presentes en ambos grupos de análisis (Control y esquizofrénicos); con rojo los géneros ausentes del otro grupo por segunda vez; con amarillo los géneros ausentes por tres filtros; y los que no están resaltados son los géneros que no están presentes en el otro grupo.

observar que solamente el género Wujia no aparece en la muestra de pacientes esquizofrénicos. Sin embargo, también se obtiene que los géneros Bifidobacterium, Dialister, Vescimonas, Acidaminococcus, Blautia, Carjivirus, Dysosmobacter, Enterocloster, Flavinifactor y Megamonas los que faltan del grupo control. De los mencionados anteriormente destacan *Dysosmobacter* y *Flavinifactor* que han sido prevalente durante los 3 filtros.

Abundancia Relativa



**Figura 6.16:** Abundancia relativa con filtro aplicado de abundancia 5% > de las muestras por géneros, agrupados por tipo de muestras en conjunto con la leyenda generada con los géneros de bacterias identificada.

Genero	Tipo	Frecuencia
Bacteroides	Healthy	49
Faecalibacterium	Healthy	39
Phocaeicola	Healthy	39
Prevotella	Healthy	19
Alistipes	Healthy	18
Roseburia	Healthy	7
Akkermansia	Healthy	6
Parabacteroides	Healthy	6
Simiaoa	Healthy	6
Coprococcus	Healthy	3
Butyrivibrio	Healthy	2
Wujia	Healthy	2
Lachnospira	Healthy	1
Megasphaera	Healthy	1
Paraprevotella	Healthy	1
Phascolarctobacterium	Healthy	1
Ruminococcus	Healthy	1

Genero	Tipo	Frecuencia	Genero	Tipo	Frecuencia
Bacteroides	Schizophrenic	48	Megamonas	Schizophrenic	1
Phocaeicola	Schizophrenic	40	Megasphaera	Schizophrenic	1
Faecalibacterium	Schizophrenic	34	Paraprevotella	Schizophrenic	1
Parabacteroides	Schizophrenic	23	Phascolarctobacterium	Schizophrenic	1
Alistipes	Schizophrenic	19			
Prevotella	Schizophrenic	16			
Bifidobacterium	Schizophrenic	5			
Akkermansia	Schizophrenic	4			
Simiaoa	Schizophrenic	3			
Dialister	Schizophrenic	2			
Lachnospira	Schizophrenic	2			
Roseburia	Schizophrenic	2			
Ruminococcus	Schizophrenic	2			
Vescimonas	Schizophrenic	2			
Acidaminococcus	Schizophrenic	1			
Blautia	Schizophrenic	1			
Butyrivibrio	Schizophrenic	1			
Carjivirus	Schizophrenic	1			
Coprococcus	Schizophrenic	1			
Dysosmobacter	Schizophrenic	1			
Enterocloster	Schizophrenic	1			
Flavonifractor	Schizophrenic	1			

**Figura 6.17:** Tabla del resumen generado por Shiny de los géneros de bacteria identificados que cumplen con el filtro de abundancia 5% >. En la tabla los géneros resaltados con verde son los géneros presentes en ambos grupos de análisis (Control y esquizofrénicos); con amarillo los géneros ausentes por tres filtros; y los que no están resaltados son los géneros que no están presentes en el otro grupo.

El *pipeline* realizado efectuó con éxito la verificación de calidad y filtrado de secuencias, como también realizó la remoción de la secuencia del hospedero de las lecturas (en este caso removió la secuencia de humano) y la clasificación de taxonomía de bacterias encontradas en las muestras de heces intestinales de sujetos de estudio de Dinamarca. A pesar de que el proyecto no contó con un control positivo para comparar los resultados y verificar la veracidad, se reconoce el éxito del *pipeline* dado que se llegó a una clasificación similar al del estudio '*Alteration of gut microbiome in patients with schizophrenia indicates links between bacterial tyrosine biosynthesis and cognitive dysfunction*' escrito por Thirion F, et al. de donde se obtuvieron las secuencias para crear el *pipeline*. El resultado de ejecutar el *pipeline* 2 veces, una vez para cada conjunto de datos, fue de dos archivos BIOM que tuvieran la información necesaria para su análisis en R/R shiny.

Las primeras gráficas creadas y observadas en el *dashboard* son las de Alfa diversidad. Estas nos dicen que tanto el microbioma de las muestras de sujetos control y las muestras de pacientes esquizofrénicos tienen una diversidad alta. Una alfa diversidad alta significa que una comunidad tiene una gran cantidad de especies diferentes. Tanto el índice de Shannon como el de Simpson ayudan en esto. El índice de Shannon tiene un rango de valores de 0 a 8. Un valor de 0 indica que hay una sola especie en la comunidad, mientras que un valor de 8 indica que todas las especies son igualmente abundantes (Roswell, Dushoff, y Winfree, 2021). El índice de Simpson, por otro lado, se basa en la dominancia de una comunidad. Este índice tiene un rango de valores de 0 a 1. Un valor de 0 indica que todas las especies son igualmente abundantes, mientras que un valor de 1 indica que una sola especie es la única abundante (Roswell y cols., 2021).

Dado a que en ambas figuras 6.1 y 6.2 se encuentran en los rangos considerados altos en ambos índices esto puede significar que el microbioma es diverso, pero que también está dominado por unas pocas especies. Esto puede ser un signo de un microbioma saludable, ya que la diversidad específica es importante para la salud, pero la dominancia también puede ser importante para la función (Lozupone, Stombaugh, Gordon, Jansson, y Knight, 2012). En el caso específico de aplicación tratado puede significar que estas especies dominantes que evidencian el índice de Simpson en las figuras 6.1 y 6.2 podrían ser importantes para la función del microbioma intestinal, como la digestión o la absorción de nutrientes (Lozupone y cols., 2012). Sin embargo hay que denotar la cantidad de puntos fuera del rango de  $> 3.5$  para el índice de Shannon y  $> 0.9$  para el índice de Simpson. Ya que las muestras de pacientes esquizofrénicos se encuentran más esparcidas en las gráficas, lo cual

puede significar menos diversidad en sí.

La beta diversidad puede ser medida con diferentes métodos y con diferentes índices que miden la distancia, en el presente caso de estudio se usó la distancia de Bray-Curtis. Se utilizó este índice ya que además de poseer una fórmula sencilla de entender

$$\beta_{bc} = \frac{\sum |x_i - y_i|}{\sum (x_i + y_i)}$$

En donde:

- $\beta_{bc}$  es el índice Bray-Curtis
- $x_i$  es la abundancia relativa de la especie  $i$  en la comunidad  $x$
- $y_i$  es la abundancia relativa de la especie  $i$  en la comunidad  $y$

(Beals, 1984).

Este índice de Bray-Curtis es una medida robusta de la disimilitud entre las comunidades porque es relativamente insensible a los cambios en la abundancia de las especies. Esto significa que el índice de Bray-Curtis puede ser utilizado para comparar la beta diversidad de comunidades con diferentes tamaños o estructuras. Lo cuál es útil ya que hay muestras con una menor cantidad de lecturas y de longitudes variantes (Beals, 1984). La Figura 6.3 nos muestra que no hay gradiente específico en la abundancia de bacterias en el bioma intestinal en ambos conjuntos de datos, lo cuál nos indican que las muestras son variables y que la composición de las comunidades son complejas y que pueden contener varias especies (Eilertsen, Økland, Økland, y Pedersen, 1990). Esto porque ambos conjuntos de muestras tienen sus puntos esparcidos a lo largo de los dos ejes.

La Figura 6.4 muestra ser interesante ya que las muestras comienzan a tener un patrón y se encuentran más agrupadas que las gráficas de DCA. La principal diferencia entre DCA y CA es que DCA se basa en la distancia entre las muestras, mientras que CA se basa en la correlación entre las especies (Eilertsen y cols., 1990). Esto significa que DCA es más sensible a las diferencias en la abundancia de las especies, mientras que CA es más sensible a las asociaciones entre las especies. Esto puede indicar que, a pesar de contener varias especies, estas especies no son tan diferentes entre sí, que tienen una relación entre sí.

Con respecto a la Figura 6.5, esta gráfica muestra ser diferente a las gráficas generadas por el método de DCA y CA. Esto se debe a que, NMDS a diferencia de DCA no utiliza una distancia euclidiana entre las muestras, sino que se basa en la similaridad de los datos. Por el otro lado se diferencia del CA porque NMDS se basa en la composición de las especies (NMDS y un poquito más, 2019). La diferencia entre estos gráficos puede basarse en que las muestras que se están analizando son similares en composición de especies, pero que no hay una asociación significativa entre las especies.

Dada a la escases de diferencia entre las gráficas de diversidad beta podemos inferir que es posible que las comunidades microbianas en el bioma intestinal de pacientes esquizofrénicos y del grupo control sean bastante similares en términos de composición y estructura (Eilertsen y cols., 1990) (Beals, 1984). Esta homogeneidad podría indicar que la microbiota intestinal no está directamente relacionada con la patogénesis de la esquizofrenia en este contexto particular que se desarrolló el estudio. Sin embargo, cabe agregar que existe la posibilidad de que hayan factores externos, los cuales influyan más en la composición de la microbiota intestinal que la condición de esquizofrenia en sí misma. Estos factores podrían incluir la dieta, el estilo de vida, el uso de medicamentos, entre otros, que podrían estar equilibrando las diferencias microbianas entre los grupos.

En los análisis de diversidad de microbioma entre los dos grupos a analizar se tiene que considerar de que la variabilidad individual en la microbiota intestinal puede ser alta, lo que dificulta la detección

de diferencias entre grupos, especialmente si las muestras son relativamente pequeñas. Es posible que se necesiten estudios más grandes y exhaustivos para capturar las diferencias sutiles en la composición microbiana entre los grupos. Por lo que, se prosiguió con el análisis de abundancia.

Las gráficas de abundancia ayudan a concretar las suposiciones hechas en los análisis anteriores, o inclusive realizar descubrimientos específicos que ayudan a diferenciar los microbiotas. En la Figura 6.6 se observa todos los filos clasificados de los dos conjuntos de muestras. Sin embargo, dado a que las muestras dentro de los dos grupos pueden llegar a ser muy diferentes entre ellas y complejas es por lo que se hacen las figuras de abundancia relativa. Para estas gráficas se puede hacer modificaciones acerca del porcentaje de abundancia que se debe cumplir, lo cuál ayuda a tener una mejora en la visualización de los datos. La Figura 6.7 hace un filtrado acerca del 40% de abundancia. Los filos clasificados en ambos conjuntos son los mismos, con una abundancia relativa similar, sin tomar en cuenta las muestras atípicas, las cuáles casi no llegaron a tener clasificaciones de filo de bacterias dentro de ellas.

Por esta razón, se decidió cambiar el porcentaje de filtro a los diferentes valores posibles dentro del rango de 0 a 1 con saltos de 0.1%. La diferencia se vió hasta que se llegó al 0.9%, en donde las muestras de pacientes con esquizofrenia todavía presentan la clasificación del filo *Euryarchaeota*. A pesar de este encuentro, no es un dato significativo ya que el cambio entre 0.8% y 0.9% en términos de abundancia es poco. Además que solo se muestra este género de bacteria sobre una muestra tomada de un paciente con esquizofrenia. Lo cual no es una muestra significativa dentro del conjunto de datos.

Por el otro lado, este cambio, motivó a que se realizarán las gráficas de abundancia relativas de los géneros de las bacterias de los dos conjuntos de muestras con su respectivas tablas. Estas Figuras son las que se pueden visualizar en la sección de **Abundancia relativa por género**. De todas las bacterias mencionadas en resultados las interesantes son las que prevalecen aún después de los filtros, pero no se encuentran en el otro grupo de muestras.

En las tablas de abundancia relativa realizadas se puede observar que, las muestras de pacientes con esquizofrenia suelen tener más géneros identificados, por ende una diversidad más alta que las muestras control. Lo cuál puede indicar una complejidad de comunidad microbiana más alta. Este hallazgo también fue realizado por (F. Zhu y cols., 2020). Sin embargo, otros estudios de microbiota en pacientes con esquizofrenia como (Yolken, Prandovszky, Severance, Hatfield, y Dickerson 2021), (Yuan y cols., 2018), (Z. Zhang y cols., 2022) no presentan estos resultados, por lo que no se debe de teorizar este resultado.

En este estudio aplicativo, se llega a observar una marcada diferencia en la composición microbiana entre el grupo de pacientes con esquizofrenia y el grupo control. Es notable que los géneros Diorhavirus, Haemophilus y Muribaculum estén ausentes en las muestras de pacientes esquizofrénicos, mientras que predominan en el grupo control tras dos filtros. Por otro lado, los géneros Enterocloster, Faecalitalea, Hungatella, Lacticaseibacillus, Limosilactobacillus, Monoglobus y Oscillibacter son prominentes en las muestras de pacientes esquizofrénicos después de un filtro, mientras que los géneros Lactobacillus, Dysosmobacter y Escherichia son más prevalentes después de dos filtros. El género Flavonifractor es interesante ya que su presencia es significativamente menor en el grupo control y desaparece después del segundo filtro.

La presencia o ausencia de estos géneros microbianos puede tener importantes implicaciones biológicas. Se sabe que muchos de estos géneros están involucrados en la producción de metabolitos bioactivos, como ácidos grasos de cadena corta, neurotransmisores y compuestos antiinflamatorios (Haas-Neill, Iwashita, Dvorkin-Gheva, y Forsythe, 2022). Estos metabolitos pueden desempeñar un papel crucial en la comunicación entre el intestino y el cerebro a través del eje intestino-cerebro, influenciando la función cerebral y el estado de ánimo (Javitt, 2007).

La ausencia de ciertos géneros en los pacientes con esquizofrenia podría sugerir una disbiosis microbiana que podría estar relacionada con la fisiopatología de la enfermedad. Por ejemplo, se ha

sugerido que la producción reducida de ciertos metabolitos por parte de la microbiota intestinal podría contribuir a la inflamación crónica y al estrés oxidativo, procesos que se han implicado en la etiología de la esquizofrenia (Javitt, 2007) (Howes y Kapur, 2009).

En este punto se debe de proseguir con un soporte fuerte literario ya que muchas especies de bacterias aún no se le ha brindado una función específica en el microbioma intestinal e inclusive, aún son desconocidas como para tener datos curados de su secuencia. Además, se debe de considerar la influencia de factores externos como la dieta, el estilo de vida y el uso de medicamentos puede ayudar a contextualizar mejor estos hallazgos y a identificar posibles estrategias terapéuticas basadas en la modulación de la microbiota intestinal.

Este es el caso de *Diorhavirus*, se cree que parte de la familia a la que pertenece *Intestiviridae* son estabilizadores de microbioma intestinal del ser humano, sin embargo no hay suficientes estudios que respalden los datos (Cao y cols., 2022). *Escheria* es un género bastante popular dado a la famosa infección por *E. coli*, la cual es nociva por sus malestares estomacales, sin embargo, se ha realizado un estudio acerca de las secuelas que puede tener la infección por *E. coli* y entre ellas estaban los ataques de pánico, disminución en las funciones cognitivas y un pequeño porcentaje de alucinaciones (Kleimann y cols., 2014), este es el mismo caso que no hay suficientes datos para respaldar los resultados de este estudio. En cuanto a los metabolitos relacionados con *Escherichia*, esta bacteria es conocida por producir una variedad de productos metabólicos durante su crecimiento y metabolismo, incluyendo ácidos grasos de cadena corta (AGCC) como acetato, propionato y butirato, así como metabolitos nitrogenados como amoníaco y compuestos sulfurados (Wang, Jin, Ji, Huang, y Xie, 2022). En relación con la esquizofrenia, la investigación sobre los metabolitos específicos asociados con *Escheria* y su impacto en la enfermedad es limitada y no hay estudios concluyentes que demuestren una relación causal directa.

Por otro lado, se encuentran los que han sido estudiados exhaustivamente por su implicación en la salud humana. Por ejemplo *Haemophilus* es un género de bacterias gram negativas que son conocidas por causar infecciones (Musher, 1996). Sin embargo, una alta abundancia se ha visto relacionada con el desarrollo de síntomas de psicosis, lo cuál es importante para el presente estudio (C. Zhu y cols., 2021). Asimismo, *Faecalitalea* es un género que se ha visto altamente relacionado con síndrome de depresión severa y con problemas de Insomnio (Q. Zhang y cols., 2021). Como géneros independientes podemos encontrar que no hay una relación directa con la patogénesis de la esquizofrenia o la presentación de síntomas activos en las personas diagnosticadas. Sin embargo, se ha estudiado que uno de los antibióticos betalactámicos como las aminopenicilinas cuyo espectro incluye *Haemophilus*, *Escheria* y *Faecalitalea* afectan neurológicamente al sistema nervioso central (Morales y cols., s.f.). Estas afecciones se deben al efecto inhibitorio en la transmisión del neurotransmisor GABA, ya que la estructura del anillo betalactámico comparte similitudes con este neurotransmisor. Además, se produce la liberación de citoquinas y endotoxinas, lo que resulta en un aumento en la excitabilidad de los receptores N-metil-D-aspartato (NMDA) y alpha-amino-3-hidroxi-5-metilisoaxazol propionato (AMPA), acumulación de metabolitos epileptogénicos y reacciones debidas a edema cerebral alérgico (Mattappalil y Mergenhagen, 2014). Esta adversión en el SNC puede reflejarse con la manifestación de convulsiones, encefalopatía, mioclonos, temblor, hiperexcitabilidad e hiperactividad, incluso se ha evidenciado cambios de la personalidad, confusión, psicosis, ataques de pánico y parkinsonismo (Vardakas, Kalimeris, Triarides, y Falagas, 2018).

Los géneros *Hungatella* y *Oscillibacter* se han tratado de relacionar con el desarrollo de enfermedades mentales desde edad temprana y con el desarrollo de síntomas psicóticos, sin embargo, no hay datos suficientes para encontrar una relación significativa entre la abundancia de estas dos bacterias y el desarrollo de estas enfermedades (Kraaij y cols., 2023).

Otro género cuya alta abundancia se ve relacionado con los síntomas de la depresión es *Enterocloster*, en un estudio de 333 personas realizado en china, se encontró una abundancia alta de *Enterocloster bolteae* *Enterocloster clostridioformis* en los individuos diagnosticados con depresión (Park, Li, Wu, y Zhang, 2023). Por otro lado, el género *Enterocloster* si bien no es un género bien

establecido, los géneros relacionados con *Enterococcus* pueden estar implicados en la producción de ácidos grasos de cadena corta, que son metabolitos importantes para la salud intestinal y la función cerebral (Burckhardt, Chong, Pett, y Tropini, 2023).

Los géneros *Lacticaseibacillus* y *Limosilactobacillus* son dos géneros bacterianos relacionados que pertenecen a la familia *Lactobacillaceae* y están estrechamente asociados con la producción de ácido láctico y otros metabolitos beneficiosos en el intestino. También han demostrado ser psicoactivos, en especial *Lacticaseibacillus rhamnosus* y *Limosilactobacillus reuteri* son utilizados como psicobióticos que funcionan como ansiolíticos y como desinflamadores del hipocampo para ratones (Haas-Neill y cols., 2022). Por otro lado, *Lactobacillus* ha sido usado como un efectivo antidepresivo que reduce los síntomas del trastorno de Depresión severa (Johnson, Thurairajasingam, Letchumanan, Chan, y Lee, 2021). Otro género que ha sido altamente relacionado con asistir en la lidiación del estrés y combate con el síntoma de incapacidad de experimentar placer, el cuál ha sido relacionado con depresión, ansiedad y esquizofrenia; es el género *Muribaculum*, las bacterias de este género han demostrado ser enriquecidas por probióticos recetados a pacientes que sufren de estos síntomas (Z. Zhang y cols., 2022). El género *Dysosmobacter* se ha visto relacionada positivamente con la calidad de salud física de personas en un estudio realizado acerca del análisis del microbioma intestinal luego de la pandemia (Malan-Müller, Valles-Colomer, Palomo, y Leza, s.f.) y también ha demostrado ser un buen candidato para probióticos contra la Diabetes 2 (? ?). Finalmente, el género de los más interesantes del estudio es el *Flavonifractor*, este género inclusive antes de hacer el primer filtrado mostró tener una abundancia relativa más alta en el grupo de muestras de pacientes esquizofrénicos que en las muestras control.

*Flavonifractor* son un grupo de bacterias intestinales que tienen la capacidad de metabolizar los flavonoides, compuestos antioxidantes y antiinflamatorios presentes en plantas. Estudios han revelado que estas bacterias facilitan la absorción de flavonoides en el intestino al hacer que los metabolitos resultantes sean más solubles en agua (Sankaranarayanan y cols., 2021). Aunque la investigación sobre los efectos de *Flavonifractor* en la salud humana aún está en curso, estudios preliminares sugieren su potencial en la prevención y tratamiento de enfermedades como la diabetes, el cáncer y las enfermedades cardíacas (Sankaranarayanan y cols., 2021). Sin embargo, su relación con la salud mental, específicamente la esquizofrenia, es notable. Investigaciones recientes (X. Zhang y cols., 2020) (F. Zhu y cols., 2020) han vinculado la abundancia relativa de *Flavonifractor* con un mayor riesgo de desarrollar esquizofrenia y la manifestación de síntomas más severos de este trastorno mental.

Es importante destacar que los flavonoides y sus metabolitos pueden influir en la producción de neurotransmisores en el cerebro. Por ejemplo, el triptófano, un aminoácido presente en los flavonoides, es un precursor del neurotransmisor serotonina, que juega un papel crucial en la regulación del estado de ánimo y el bienestar emocional (Coello y cols., 2021). Además, el triptófano también es precursor de la síntesis de dopamina y acetilcolina, neurotransmisores asociados con la función cognitiva y el estado de ánimo (Coello y cols., 2021). Esta conexión entre los flavonoides, las bacterias intestinales y la producción de neurotransmisores sugiere un posible vínculo entre la dieta, la microbiota intestinal y la salud mental.

Estos resultados son importantes para la demostración de la efectividad del *pipeline* creado para la clasificación taxonómica de microbioma. Sin embargo, cabe agregar que estos datos no son significativos para establecer una relación directa entre las bacterias identificadas con la patogénesis de la esquizofrenia. A pesar de esto, sí puede ser un indicio para estudios futuros acerca de este desorden mental y la microbiota humana. Además que las muestras utilizadas en el presente trabajo pertenecen a personas de Dinamarca, por lo que su aplicación en Guatemala debe de tomar en cuenta que por diferentes tipos de dieta, el microbioma humano pueda ser diferente.

Del estudio se concluye que el *pipeline* desarrollado logra realizar la clasificación de taxonomía de microbiota intestinal humano con éxito y de forma automatizada sin la necesidad de interacción humana. Teniendo así como resultados archivos que sean legibles y utilizables para la generación de visualizaciones para el futuro análisis por parte del investigador.

En el estudio aplicado con las muestras de pacientes con esquizofrenia y de muestras control, se logra identificar que para ambos conjuntos de datos se tiene que hay una diversidad alta y no difieren significativamente entre ellos en estos términos. Teniendo índices de Shannon y Simpson altos en las gráficas de alfa diversidad y no proporcionando una diferencia significativa en las gráficas de beta diversidad. Lo que significa que no hay una disimilitud entre ellas.

Asímismo, para la visualización centralizada de resultados, se logró desarrollar un *dashboard* en R Shiny que permitió el análisis visual de los resultados del *pipeline*, pudiendo así inferir acerca de la clasificación de microbioma de los dos conjuntos de muestras que se tenían y poder llegar a obtener resultados y conclusiones acerca de las diferentes comunidades microbianas.

Luego del análisis realizado de los datos presentados en la visualización se identifica que las muestras de pacientes con esquizofrenia presentan una alta abundancia relativa de bacterias del género *Flavonifractor*, *Dysosmobacter*, *Escheria* y *Lactobacillus*. De los cuales independientemente puede que no tengan relación con el desorden mental, sin embargo, una presencia alta de estos géneros en conjunto si puede indicar un efecto adversario en la salud mental del individuo. Por lo que la regulación de estos puede ser una estrategia importante para el tratamiento de la enfermedad.

De primera instancia se puede recomendar la creación de un archivo YAML en donde se encuentren todos los paquetes necesarios para la ejecución correcta del *pipeline*. De esta manera evitando que sea necesaria la creación, activación y desactivación de una cantidad considerable de ambientes. Además de si se desea aplicar el *pipeline* para muestras cuyo origen no es humano, se debe de cambiar el genoma a descargar, reemplazándolo por el correcto. Se recomienda así mismo que si el *pipeline* esta siendo ejecutado a través de una máquina virtual, se utilice el comando de *nohup* para poder mover la computadora de red sin que se interrumpa la ejecución. Asimismo se recomienda hacer pruebas con máquinas virtuales más grandes a la usada en el presente estudio y realizar la clasificación de taxonomía de bacterias con Kaiju, el cuál es más preciso que kraken.

En tema de la creación de un Dashboard en Shiny App se recomienda el uso de la librería "shinydashboard" de R, y de la actualización de R a la versión más reciente. Esta librería tiene funciones que son intuitivas y con una menor curva de aprendizaje.

En estudios de aplicación de comparación de comunidades microbianas puede resultar ser interesante juntar las muestras de un mismo conjunto en una sola y luego crear un solo BIOM file con estas dos nuevas quimeras de muestra para tener una comparación más directa y puede ayudar en los tiempos de ejecución y renderización del dashboard a usar.

En el caso de estudios de enfermedades mentales, como la esquizofrenia, se recomienda el impulso de análisis genéticos de estos pacientes, como también promover estudios donde se estudie el efecto de la medicina antipsicótica en la microbiota de los pacientes y los estudios diferenciales entre los pacientes con diferentes síntomas. Asimismo, se recomienda incluir en el análisis la oportunidad de construir árboles genéticos de las especies de bacterias encontradas, para poder así tener una diferentes agrupación. Sin embargo, se debe de tomar en cuenta que es necesario un genoma conocido y curado de estas bacterias para tener un alineamiento y árbol correcto. Además del enriquecimiento de los datos para obtener funcionalidades de las bacterias registradas en bases de datos curadas.

Finalmente se recomienda que se continuen con los estudios del microbioma y sus funciones e implicaciones en la vida ya que puede establecer un buen punto de partida para la creación de soluciones sostenibles que respondan a una necesidad específica.

- Abel, K. M., Drake, R., y Goldstein, J. M. (2010, octubre). Sex differences in schizophrenia. *International Review of Psychiatry*, 22(5), 417–428. doi: 10.3109/09540261.2010.515205
- Afshinnekoo, E., Chou, C., Alexander, N., Ahsanuddin, S., Schuetz, A. N., y Mason, C. E. (2017, abril). Precision metagenomics: Rapid metagenomic analyses for infectious disease diagnostics and public health surveillance. *Journal of Biomolecular Techniques: JBT*, 28(1), 40–45. doi: 10.7171/jbt.17-2801-007
- Aggarwal, N., Kitano, S., Pua, G. R. Y., Kittelmann, S., Hwang, I. Y., y Chang, M. W. (2022, noviembre). Microbiome and human health: Current understanding, engineering, and enabling technologies. *Chemical Reviews*, 123(1), 31–72. doi: 10.1021/acs.chemrev.2c00431
- Akhondzadeh, S. (2019). Microbiome and schizophrenia. *Avicenna J Med Biotechnol*, 11(4), 269.
- APA. (2013). *Diagnostic and statistical manual of mental disorders: Dsm-5™, 5th ed.* Arlington, VA, US: American Psychiatric Publishing, Inc. doi: 10.1176/appi.books.9780890425596
- (APA), A. P. A. (2019). *Guía de práctica clínica para el tratamiento de pacientes con esquizofrenia*. Descargado de <https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/Clinical%20Practice%20Guidelines/APA-Draft-Schizophrenia-Treatment-Guideline-October-2019.pdf>
- Australia, H. (2022). Schizophrenia. *Healthdirect Australia*.
- Babraham bioinformatics—fastqc a quality control tool for high throughput sequence data. (s.f.). Descargado de <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bash-scripts. (s.f.). Descargado de <https://data-skills.github.io/unix-and-bash/03-bash-scripts/index.html>
- Beals, E. W. (1984, enero). Bray-curtis ordination: An effective strategy for analysis of multivariate ecological data. En A. MacFadyen y E. D. Ford (Eds.), *Advances in ecological research* (Vol. 14, p. 1–55). Academic Press. Descargado de <https://www.sciencedirect.com/science/article/pii/S0065250408601683> doi: 10.1016/S0065-2504(08)60168-3
- Bellack, A. S., y Hersen, M. (Eds.). (2014). *Behavioral assessment: A practical handbook*. Springer Science & Business Media.
- Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., ... Segata, N. (2023, febrero). Extending and improving metagenomic taxonomic profiling with uncharacterized species using metaphlan 4. *Nature Biotechnology*, 1–12. doi: 10.1038/s41587-023-01688-w
- Burckhardt, J., Chong, D., Pett, N., y Tropini, C. (2023, marzo). Gut commensal enterocloster

- species host inoviruses that are secreted in vitro and in vivo. *Microbiome*, 11. doi: 10.1186/s40168-023-01496-z
- Cao, Z., Sugimura, N., Burgermeister, E., Ebert, M. P., Zuo, T., y Lan, P. (2022, julio). The gut virome: A new microbiome component in health and disease. *eBioMedicine*, 81. Descargado de [https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964\(22\)00294-8/fulltext](https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(22)00294-8/fulltext) doi: 10.1016/j.ebiom.2022.104113
- Carabotti, M., Scirocco, A., Maselli, M. A., y Severi, C. (2015). The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Annals of Gastroenterology: Quarterly Publication of the Hellenic Society of Gastroenterology*, 28(2), 203–209.
- Chen, S., y cols. (2018, September). fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17), i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, Y., Fan, L., Chai, Y., y Xu, J. (2022, septiembre). Advantages and challenges of metagenomic sequencing for the diagnosis of pulmonary infectious diseases. *The Clinical Respiratory Journal*, 16(10), 646–656. doi: 10.1111/crj.13538
- Chong, J., Liu, P., Zhou, G., y Xia, J. (2020, marzo). Using microbiomeanalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols*, 15(33), 799–821. doi: 10.1038/s41596-019-0264-1
- Cloutier, M., Aigbogun, M. S., Guerin, A., Nitulescu, R., Ramanakumar, A. V., Kamat, S. A., ... Wu, E. (2016, junio). The economic burden of schizophrenia in the united states in 2013. *The Journal of Clinical Psychiatry*, 77(6), 5379. doi: 10.4088/JCP.15m10278
- Coello, K., Hansen, T. H., Sørensen, N., Ottesen, N. M., Miskowiak, K. W., Pedersen, O., ... Vinberg, M. (2021, agosto). Affective disorders impact prevalence of flavonifractor and abundance of christensenellaceae in gut microbiota. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 110, 110300. doi: 10.1016/j.pnpbp.2021.110300
- Contributors, W. (s.f.). *Mental health and schizophrenia*. <https://www.webmd.com/schizophrenia/mental-health-schizophrenia>. (Accessed: March 2, 2023)
- Cryan, J., y Dinan, T. (2012, 09). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *nat. rev. neurosci.* 13, 701-712. *Nature reviews. Neuroscience*, 13, 701-12. doi: 10.1038/nrn3346
- Cummings, J. (2021, abril). The role of neuropsychiatric symptoms in research diagnostic criteria for neurodegenerative diseases. *The American Journal of Geriatric Psychiatry*, 29(4), 375–383. doi: 10.1016/j.jagp.2020.07.011
- de Theije, C. G. M., Wu, J., da Silva, S. L., Kamphuis, P. J., Garssen, J., Korte, S. M., y Kraneveld, A. D. (2011, septiembre). Pathways underlying the gut-to-brain connection in autism spectrum disorders as future targets for disease management. *European Journal of Pharmacology*, 668 Suppl 1, S70-80. doi: 10.1016/j.ejphar.2011.07.013
- Dinan, T. G., y Cryan, J. F. (2017). The microbiome-gut-brain axis in health and disease. *Gastroenterology Clinics*, 46(1), 77-89.
- Eilertsen, O., Økland, R. H., Økland, T., y Pedersen, O. (1990). Data manipulation and gradient length estimation in dca ordination. *Journal of Vegetation Science*, 1(2), 261–270. doi: 10.2307/3235663
- Foster, J. A., Rinaman, L., y Cryan, J. F. (2017, March 19). Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiology of Stress*, 7, 124-136. doi: 10.1016/j.ynstr.2017.03.001
- Galloway-Peña, J., y Hanson, B. (2020, marzo). Tools for analysis of the microbiome. *Digestive Diseases and Sciences*, 65(3), 674–685. doi: 10.1007/s10620-020-06091-y
- García Muñoz, O. A. (2014). *Manejo de la esquizofrenia*. Instituto Guatemalteco de Seguridad Social (IGSS). Descargado de <https://www.igssgt.org/wp-content/uploads/images/gpc-be/psiquiatrAa/GPC-BE%20No.%2058%20Esquizofrenia.pdf>
- Gareau, M. G., Sherman, P. M., y Walker, W. A. (2010, septiembre). Probiotics and the gut microbiota in intestinal health and disease. *Nature Reviews. Gastroenterology Hepatology*, 7(9), 503–514. doi: 10.1038/nrgastro.2010.117
- Gejman, P. V., Sanders, A. R., y Kendler, K. S. (2011). Genetics of schizophrenia: New findings and challenges. *Annual Review of Genomics and Human Genetics*, 12(1), 121-144. Descargado

- de <https://doi.org/10.1146/annurev-genom-082410-101459> (PMID: 21639796) doi: 10.1146/annurev-genom-082410-101459
- Goodman, L. S., y Brunton, L. L. (2018). *Goodman & Gilman's: Las bases farmacológicas de la terapéutica* (13.a ed. ed.). Editorial Médica Panamericana.
- Haas-Neill, S., Iwashita, E., Dvorkin-Gheva, A., y Forsythe, P. (2022, agosto). Effects of two distinct psychoactive microbes, lacticaseibacillus rhamnosus jb-1 and limosilactobacillus reuteri 6475, on circulating and hippocampal mrna in male mice. *International Journal of Molecular Sciences*, *23*(17), 9653. doi: 10.3390/ijms23179653
- Hoffman, K. W., Lee, J. J., Corcoran, C. M., Kimhy, D., Kranz, T. M., y Malaspina, D. (2020). Considering the microbiome in stress-related and neurodevelopmental trajectories to schizophrenia. *Frontiers in Psychiatry*, *11*. Descargado de <https://www.frontiersin.org/articles/10.3389/fpsy.2020.00629>
- Howes, O. D., y Kapur, S. (2009, mayo). The dopamine hypothesis of schizophrenia: version iii—the final common pathway. *Schizophrenia Bulletin*, *35*(3), 549–562. doi: 10.1093/schbul/sbp006
- Javitt, D. C. (2007). Glutamate and schizophrenia: phencyclidine, n-methyl-d-aspartate receptors, and dopamine-glutamate interactions. *International Review of Neurobiology*, *78*, 69–108. doi: 10.1016/S0074-7742(06)78003-5
- Johnson, D., Thuraiasingam, S., Letchumanan, V., Chan, K.-G., y Lee, L.-H. (2021, mayo). Exploring the role and potential of probiotics in the field of mental health: Major depressive disorder. *Nutrients*, *13*(5), 1728. doi: 10.3390/nu13051728
- Joo, M.-K., Ma, X., Yoo, J.-W., Shin, Y.-J., Kim, H.-J., y Kim, D.-H. (2023, julio). Patient-derived enterococcus mundtii and its capsular polysaccharides cause depression through the downregulation of nf-b-involved serotonin and bdnf expression. *Microbes and Infection*, *25*(6), 105116. doi: 10.1016/j.micinf.2023.105116
- Kahn, R. S., Sommer, I. E., Murray, R. M., Meyer-Lindenberg, A., Weinberger, D. R., Cannon, T. D., ... Insel, T. R. (2015). Schizophrenia. *Nature Reviews Disease Primers*, *1*(1), Article 1. Descargado de <https://doi.org/10.1038/nrdp.2015.67>
- Karl, J. P., Margolis, L. M., Madslie, E. H., Murphy, N. E., Castellani, J. W., Gundersen, Y., ... Pasiakos, S. M. (2017, junio). Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *312*(6), G559–G571. doi: 10.1152/ajpgi.00066.2017
- Kelly, J. R., Borre, Y., O' Brien, C., Patterson, E., El Aidy, S., Deane, J., ... Dinan, T. G. (2016, noviembre). Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. *Journal of Psychiatric Research*, *82*, 109–118. doi: 10.1016/j.jpsychires.2016.07.019
- Keshavan, M. S., Collin, G., Guimond, S., Kelly, S., Prasad, K. M., y Lizano, P. (2020, febrero). Neuroimaging in schizophrenia. *Neuroimaging clinics of North America*, *30*(1), 73–83. doi: 10.1016/j.nic.2019.09.007
- Khavari, B., y Cairns, M. J. (2020). Epigenomic dysregulation in schizophrenia: In search of disease etiology and biomarkers. *Cells*, *9*(8), 1837. doi: 10.3390/cells9081837
- Kleimann, A., Toto, S., Eberlein, C. K., Kielstein, J. T., Bleich, S., Frieling, H., y Sieberer, M. (2014, julio). Psychiatric symptoms in patients with shiga toxin-producing e. coli o104:h4 induced haemolytic-uraemic syndrome. *PLoS ONE*, *9*(7), e101839. doi: 10.1371/journal.pone.0101839
- Kraaij, R., Schuurmans, I. K., Radjabzadeh, D., Tiemeier, H., Dinan, T. G., Uitterlinden, A. G., ... Cecil, C. A. (2023, febrero). The gut microbiome and child mental health: A population-based study. *Brain, behavior, and immunity*, *108*, 188–196. doi: 10.1016/j.bbi.2022.12.006
- Levenson, R. W., Sturm, V. E., y Haase, C. M. (2014). Emotional and behavioral symptoms in neurodegenerative disease: A model for studying the neural bases of psychopathology. *Annual Review of Clinical Psychology*, *10*(1), 581–606. doi: 10.1146/annurev-clinpsy-032813-153653
- Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., y Knight, R. (2012, septiembre). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*(7415), 220–230. doi: 10.1038/nature11550
- Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., ... Steinegger,

- M. (2022, diciembre). Metagenome analysis using the kraken software suite. *Nature Protocols*, 17(1212), 2815–2839. doi: 10.1038/s41596-022-00738-y
- Ma, Y., Chen, H., Lan, C., y Ren, J. (2018). Help, hope and hype: Ethical considerations of human microbiome research and applications. *Protein & Cell*, 9(5), 404–415. Descargado de <https://doi.org/10.1007/s13238-018-0537-4>
- Malan-Müller, S., Valles-Colomer, M., Palomo, T., y Leza, J. C. (s.f.). The gut-microbiota-brain axis in a spanish population in the aftermath of the covid-19 pandemic: microbiota composition linked to anxiety, trauma, and depression profiles. *Gut Microbes*, 15(1), 2162306. doi: 10.1080/19490976.2022.2162306
- Mandal, P. K., Gaur, S., Roy, R. G., Samkaria, A., Ingole, R., y Goel, A. (2022, octubre). Schizophrenia, bipolar and major depressive disorders: Overview of clinical features, neurotransmitter alterations, pharmacological interventions, and impact of oxidative stress in the disease process. *ACS Chemical Neuroscience*, 13(19), 2784–2802. doi: 10.1021/acscchemneuro.2c00420
- Mangandid, C. A. M. (s.f.). Observatorio de salud mental en guatemala anÁlisis de la situaciÓn de la inversiÓn en salud mental - proyecto piloto en la ciudad de guatemala-. 2009.
- Marwaha, T. A. U. R. . B. M. R., S. (2016). Fifteen years on—early intervention for a new generation. *The British Journal of Psychiatry: The Journal of Mental Science*, 209(3), 186–188. Descargado de <https://doi.org/10.1192/bjp.bp.115.170035>
- Mattappalil, A., y Mergenhagen, K. A. (2014, noviembre). Neurotoxicity with antimicrobials in the elderly: a review. *Clinical Therapeutics*, 36(11), 1489-1511.e4. doi: 10.1016/j.clinthera.2014.09.020
- Mawe, G. M., Browning, K. N., Manfredsson, F. P., Camilleri, M., Hamilton, F. A., Hollander, J. A., ... Wiley, J. W. (2022, mayo). 2021 workshop: Neurodegenerative diseases in the gut-brain axis-parkinson’s disease. *Gastroenterology*, 162(6), 1574–1582. doi: 10.1053/j.gastro.2022.02.004
- McBride, B. (2022, enero). *Genomics sequencing data analysis: A cost timeline*. Descargado de <https://www.fiosgenomics.com/genomics-sequencing-data-analysis-a-cost-timeline/>
- Menzel, P., Ng, K. L., y Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with kaiju. *Nature Communications*, 7, 11257. Descargado de <https://doi.org/10.1038/ncomms11257>
- Metagenomics*. (s.f.). Descargado de <https://www.genome.gov/es/genetics-glossary/Metagenomica>
- Metagenomics next generation sequencing*. (s.f.). Descargado de <https://asm.org:443/Articles/2019/November/Metagenomic-Next-Generation-Sequencing-How-Does-It>
- Morales, P. A. H., Bastos, J. L. G., Marín, D. G., Londoño, L. M., Tamayo, A. H., Cárdenas, P. A. U., ... Mejía, M. C. B. (s.f.). Reacciones adversas a betalactámicos: una revisión de tema.
- Munawar, N., Ahsan, K., Muhammad, K., y et al. (2021). Hidden role of gut microbiome dysbiosis in schizophrenia: Antipsychotics or psychobiotics as therapeutics? *International Journal of Molecular Sciences*, 22(14), 7671. doi: 10.3390/ijms22147671
- Musher, D. M. (1996). Haemophilus species. En S. Baron (Ed.), *Medical microbiology* (4th ed.). Galveston (TX): University of Texas Medical Branch at Galveston. Descargado de <http://www.ncbi.nlm.nih.gov/books/NBK8458/>
- Nemani, K., Hosseini Ghomi, R., McCormick, B., y Fan, X. (2015). Schizophrenia and the gut-brain axis. *Prog Neuropsychopharmacol Biol Psychiatry*, 56, 155-160. doi: 10.1016/j.pnpbp.2014.08.018
- Nguyen, T. T. T., Kosciolk, T., Maldonado, Y., Daly, R. E., Martin, A. S., McDonald, D., y Knight, R. (2019). Differences in gut microbiome composition between persons with chronic schizophrenia and healthy comparison subjects. *Schizophrenia Research*, 204, 23-29.
- Nmds y un poquito mas*. (2019, enero). Descargado de <https://www.datanalytics.com/2019/01/24/nmds-y-un-poquito-mas-alla/>
- of Mental Health (NIMH), N. I. (2020). *Esquizofrenia*. Descargado de <https://www.nimh.nih.gov/health/topics/esquizofrenia/index.shtml>

- Owen, S. A. . M. P. B., M. J. (2016). Schizophrenia. *The Lancet*, 388(10039), 86–97. Descargado de [https://doi.org/10.1016/S0140-6736\(15\)01121-6](https://doi.org/10.1016/S0140-6736(15)01121-6)
- Park, S., Li, C., Wu, X., y Zhang, T. (2023, agosto). Gut microbiota alterations and their functional differences in depression according to enterotypes in asian individuals. *International Journal of Molecular Sciences*, 24(17), 13329. doi: 10.3390/ijms241713329
- Pereira, F. (2019, enero). Chapter 28 - metagenomics: A gateway to drug discovery. En S. N. Meena y M. M. Naik (Eds.), *Advances in biological science research* (p. 453–468). Academic Press. Descargado de <https://www.sciencedirect.com/science/article/pii/B9780128174975000288> doi: 10.1016/B978-0-12-817497-5.00028-8
- Pipeline-pitfalls.* (s.f.). Descargado de <https://www.educative.io/courses/guide-to-bash-programming/pipeline-pitfalls>
- Proquest platform.* (s.f.). Descargado de <https://www.proquest.com/openview/dd766724c6ed3ba3e104a39f49861305/1?pq-origsite=gscholar&cbl=44156>
- Roswell, M., Dushoff, J., y Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, 130(3), 321–338. doi: 10.1111/oik.07202
- Saha, C. D. W. J. . M. J., S. (2005). A systematic review of the prevalence of schizophrenia. *PLOS Medicine*, 2(5), e141. Descargado de <https://doi.org/10.1371/journal.pmed.0020141>
- Salahuddin, P., Fatima, M. T., Uversky, V. N., Khan, R. H., Islam, Z., y Furkan, M. (2021, noviembre). The role of amyloids in alzheimer’s and parkinson’s diseases. *International Journal of Biological Macromolecules*, 190, 44–55. doi: 10.1016/j.ijbiomac.2021.08.197
- Sankaranarayanan, R., Sekhon, P. K., Ambat, A., Nelson, J., Jose, D., Bhat, G. J., y Scaria, J. (2021, junio). Screening of human gut bacterial culture collection identifies species that biotransform quercetin into metabolites with anticancer properties. *International Journal of Molecular Sciences*, 22(13), 7045. doi: 10.3390/ijms22137045
- Schizophrenia.* (s.f.-a). National Institute of Mental Health (NIMH). Descargado de <https://www.nimh.nih.gov/health/topics/schizophrenia>
- Schizophrenia.* (s.f.-b). <https://www.who.int/news-room/fact-sheets/detail/schizophrenia>. (Retrieved May 30, 2023)
- Schrag, Z. M. (2011). The case against ethics review in the social sciences. *Research Ethics*, 7(4), 120-131. Descargado de <https://doi.org/10.1177/174701611100700402> doi: 10.1177/174701611100700402
- Severance, E. G., Prandovszky, E., Castiglione, J., y Yolken, R. H. (2015). Gastroenterology issues in schizophrenia: why the gut matters. *Current Psychiatry Reports*, 17(5), 27.
- Shabani, A., Masoumian, S., Zamirinejad, S., Hejri, M., Pirmorad, T., y Yaghmaeezadeh, H. (2021, marzo). Psychometric properties of structured clinical interview for dsm-5 disorders-clinician version (scid-5-cv). *Brain and Behavior*, 11(5), e01894. doi: 10.1002/brb3.1894
- Shen, W., Le, S., Li, Y., y Hu, F. (2016). Seqkit: a cross-platform and ultrafast toolkit for fasta/q file manipulation. *PLOS ONE*. Descargado de <https://doi.org/10.1371/journal.pone.0163962>
- Sherwin, E., Sandhu, K. V., Dinan, T. G., y Cryan, J. F. (2016, noviembre). May the force be with you: The light and dark sides of the microbiota-gut-brain axis in neuropsychiatry. *CNS drugs*, 30(11), 1019–1041. doi: 10.1007/s40263-016-0370-3
- Side effects of antipsychotics.* (s.f.). <https://www.mind.org.uk/information-support/drugs-and-treatments/antipsychotics/side-effects/>. (Retrieved March 3, 2023)
- Storr, W. (2020). *The science of storytelling: Why stories make us human and how to tell them better*. Abrams. (Google-Books-ID: G9WqDwAAQBAJ)
- Stroup, . G. N., T. S. (2018). Management of common adverse effects of antipsychotic medications. *World Psychiatry*, 17(3), 341–356. Descargado de <https://doi.org/10.1002/wps.20567>
- Taxonomic analysis with r – data processing and visualization for metagenomics.* (s.f.). <https://carpentries-lab.github.io/metagenomics-analysis/09-abundance-analyses/index.html>. (Retrieved May 3, 2023)
- Touitou, Y., Portaluppi, F., Smolensky, M. H., y Rensing, L. (2004). Ethical principles and standards for the conduct of human and animal biological rhythm research. *Chronobiology Internatio-*

- nal*, 21(1), 161-170. Descargado de <https://doi.org/10.1081/CBI-120030045> (PMID: 15129830) doi: 10.1081/CBI-120030045
- Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E. F., Wang, J., Tito, R. Y., ... Raes, J. (2019, Apr). The neuroactive potential of the human gut microbiota in quality of life and depression. *Nature Microbiology*, 4(4), 623–632. doi: 10.1038/s41564-018-0337-x
- Vardakas, K. Z., Kalimeris, G. D., Triarides, N. A., y Falagas, M. E. (2018, mayo). An update on adverse drug reactions related to -lactam antibiotics. *Expert Opinion on Drug Safety*, 17(5), 499–508. doi: 10.1080/14740338.2018.1462334
- Vásquez, M. M. . C. P., H. (s.f.). *Palmitato de paliperidona para el tratamiento de pacientes con esquizofrenia*. <https://docs.bvsalud.org/biblioref/2019/02/979653/ets-026-2018-palmitato-de-paliperidona.pdf>.
- Wang, Y., Jin, Y., Ji, X., Huang, M., y Xie, B. (2022, diciembre). Metabonomic analysis of metabolites produced by *Escherichia coli* in patients with and without sepsis. *Infection and Drug Resistance*, 15, 7339–7350. doi: 10.2147/IDR.S388034
- Weickert, C. S., Weickert, T. W., Pillai, A., y Buckley, P. F. (2013a). Biomarkers in schizophrenia: A brief conceptual consideration. *Disease Markers*, 35(1), 3–9. Descargado de <https://doi.org/10.1155/2013/510402>
- Weickert, C. S., Weickert, T. W., Pillai, A., y Buckley, P. F. (2013b, julio). Biomarkers in schizophrenia: A brief conceptual consideration. *Disease Markers*, 35(1), 3–9. doi: 10.1155/2013/510402
- Wood, D. E., Lu, J., y Langmead, B. (2019, noviembre). Improved metagenomic analysis with kraken 2. *Genome Biology*, 20(1), 257. doi: 10.1186/s13059-019-1891-0
- Wykes, T., Steel, C., Everitt, B., y Tarrier, N. (2008). Cognitive behavior therapy for schizophrenia: Effect sizes, clinical models, and methodological rigor. *Schizophrenia Bulletin*, 34(3), 523–537. Descargado de <https://doi.org/10.1093/schbul/sbm114>
- Yolken, R., Prandovszky, E., Severance, E. G., Hatfield, G., y Dickerson, F. (2021, agosto). The oropharyngeal microbiome is altered in individuals with schizophrenia and mania. *Schizophrenia Research*, 234, 51–57. doi: 10.1016/j.schres.2020.03.010
- Yuan, X., Zhang, P., Wang, Y., Liu, Y., Li, X., Kumar, B. U., ... Song, X. (2018, noviembre). Changes in metabolism and microbiota after 24-week risperidone treatment in drug naïve, normal weight patients with first episode schizophrenia. *Schizophrenia Research*, 201, 299–306. doi: 10.1016/j.schres.2018.05.017
- Zachariae, R., Pedersen, C. G., Jensen, A. B., Ehrnrooth, E., Rossen, P. B., von der Maase, H., y Lund, C. (2003). Association of perceived physician communication style with patient satisfaction, distress, cancer-related self-efficacy, and perceived control over the disease. *British Journal of Cancer*, 88(5), 658–665.
- Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., ... Xie, Y. (2021). Advances in metagenomics and its application in environmental microorganisms. *Frontiers in Microbiology*, 12. Descargado de <https://www.frontiersin.org/articles/10.3389/fmicb.2021.766364>
- Zhang, Q., Yun, Y., An, H., Zhao, W., Ma, T., Wang, Z., y Yang, F. (2021, mayo). Gut microbiome composition associated with major depressive disorder and sleep quality. *Frontiers in Psychiatry*, 12, 645045. doi: 10.3389/fpsy.2021.645045
- Zhang, X., Pan, L.-y., Zhang, Z., Zhou, Y.-y., Jiang, H.-y., y Ruan, B. (2020, febrero). Analysis of gut mycobiota in first-episode, drug-naïve Chinese patients with schizophrenia: A pilot study. *Behavioural Brain Research*, 379, 112374. doi: 10.1016/j.bbr.2019.112374
- Zhang, Z., Yao, C., Li, M., Wang, L., Huang, W., y Chen, Q. (2022, noviembre). Prophylactic effects of hyperforin on anhedonia-like phenotype in chronic restraint stress model: A role of gut microbiota. *Letters in Applied Microbiology*, 75(5), 1103–1110. doi: 10.1111/lam.13710
- Zhu, C., Zheng, M., Ali, U., Xia, Q., Wang, Z., Chenlong, ... Zhang, X. (2021). Association between abundance of *Haemophilus* in the gut microbiota and negative symptoms of schizophrenia. *Frontiers in Psychiatry*, 12. Descargado de <https://www.frontiersin.org/articles/10.3389/fpsy.2021.685910>
- Zhu, F., Ju, Y., Wang, W., Wang, Q., Guo, R., Ma, Q., ... Ma, X. (2020, marzo). Metagenome-wide association of gut microbiome features for schizophrenia. *Nature Communications*, 11(11), 1612. doi: 10.1038/s41467-020-15457-9

## 11.1. Archivo pipeline.sh creado

```
1 #!/usr/bin/bash
2 #Instalacion de canales
3 conda config --add channels defaults
4 conda config --add channels bioconda
5 conda config --add channels conda-forge
6 #activacion de ambiente conda
7 conda init
8
9
10 #creacion de ambientes
11 #-----Control de Calidad
12 -----
13 conda create -n fastqc -y
14 conda create -n fastp -y
15 conda create -n bowtie2 -y
16 conda create -n samtools -y
17
18 # Fastqc
19 source activate fastqc
20 conda install -c bioconda seqkit -y
21 seqkit stats *fastq.gz > stats.txt
22 conda install -c bioconda fastqc -y
23 mkdir fastqc
24 fastqc --noextract -o fastqc *.fastq.gz
25 conda deactivate
26
27 # Fastp
28 source activate fastp
29 conda install -c bioconda fastp -y
30 mkdir filtered_reads
31 for infile in *.fastq.gz
32 do
33     base=$(basename ${infile})
34     #Extrae el nombre del archivo sin la extension ".fastq.gz"
35     newname="${base%.*}"
36     fastp -i $infile -o filtered_reads/${newname}.filt.fastq.gz
```

```

36     done
37 conda deactivate
38 rm *fastq.gz
39
40 #-----Remover host
41     -----
42 # Bowtie2
43 ##creaci n de indice para Bowtie2
44 source activate bowtie2
45 conda install -c bioconda bowtie2 -y
46 mkdir bowtie2index
47 cd bowtie2index
48 wget https://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/annotation/GRCh38_latest/
49   refseq_identifiers/GRCh38_latest_genomic.fna.gz
50 gzip -d GRCh38_latest_genomic.fna.gz
51 bowtie2-build GRCh38_latest_genomic.fna human
52 rm GRCh38_latest_genomic.fna
53 cd ..
54
55 ##Alinear secuencias con host
56 mkdir no_host
57 cd no_host
58 for infile in ../filtered_reads/*.fastq.gz
59 do
60     base=$(basename ${infile})
61     newname="${base%.*}"
62     bowtie2 -x ../bowtie2index/human -U ../filtered_reads/${base} > ${newname}.
63     sam
64 done
65 conda deactivate
66 rm -r ../bowtie2index
67
68 #Samtools
69 source activate samtools
70 conda install -c bioconda samtools -y
71 sudo apt-get install pigz -y
72 for infile in *.sam
73 do
74     base=$(basename ${infile})
75     newname="${base%.*}"
76     #convertir Sam en Bam
77     samtools view -bS ${base} > ${newname}.bam
78     rm ${base}
79     #filtrar secuencias que no alinearon al index
80     samtools view -b -f 4 ${newname}.bam > ${newname}.unmapped.bam
81     rm ${newname}.bam
82     #ordenar secuencias
83     samtools sort -n -@ 4 ${newname}.unmapped.bam -o ${newname}_sorted.bam
84     rm ${newname}.unmapped.bam
85     samtools fastq -@ 4 ${newname}_sorted.bam > ${newname}.fastq
86     rm -r *.bam
87     pigz ${newname}.fastq
88 done
89 conda deactivate
90 #-----Clasificacion de taxonomia-----
91
92 #kaiju -t nodes.dmp -f kaiju_db_refseq.fmi -i <(gunzip -c ERR8212210.filt.fastq.gz) -
93   z 4 -x -s 75 -a greedy -e 5 -o sample1_proGenomes_kaiju.out -v
94
95 #kraken2-build --download-taxonomy --skip-maps --db 'krakendb'
96 #kraken2-build --download-library bacteria --db 'krakendb'
97 #kraken2-build --download-library archaea --db 'krakendb'
98 #kraken2-build --build --max-db-size 8000000000 --db 'krakendb'
99
100 #conda create -n krakenuniq -y

```

```

100 #source activate krakenuniq
101 #conda install -c bioconda krakenuniq=0.5.8 -y
102 #conda install -c bioconda blast=2.9.0 -y
103 #mkdir krakendb
104 #krakenuniq-download --db krakendb --threads 4 --dust refseq/bacteria refseq/archaea
105 #krakenuniq-build --db krakendb --kmer-len 31 --threads 6 --taxids-for-genomes --
    taxids-for-sequences
106
107 cd ..
108 mkdir dbdownload
109 cd dbdownload
110 curl https://genome-idx.s3.amazonaws.com/kraken/k2_standard_16gb_20230605.tar.gz --
    output dbdownloaded.tar.gz
111 gzip -d dbdownloaded.tar.gz
112 tar -xvf dbdownloaded.tar
113 rm dbdownloaded.tar
114 cd ..
115
116 conda create -n kraken2 -y
117 source activate kraken2
118 conda install -c bioconda kraken2 -y
119 mkdir kraken2
120 cd kraken2
121 for infile in ../no_host/*.fastq.gz
122 do
123     base=$(basename ${infile})
124     newname="${base%.*}"
125     kraken2 --db ~/tesis/dbdownload ../no_host/${infile} --gzip-compressed --output ${
        newname}.kraken --report ${newname}.report --threads 4
126 done
127 conda deactivate
128 cd ..
129
130
131 #-----Analisis de Clasificacion
    -----
132 #krona
133 conda create -n krona -y
134 source activate krona
135 conda install -c bioconda krona -y
136 sudo apt-get install make
137 ktUpdateTaxonomy.sh
138 mkdir kronaout
139 for infile in ./kraken2/*kraken
140 do
141     krakenname+=${infile}
142     krakenname+=" "
143 done
144 ktImportTaxonomy -q 2 -t 3 ${krakenname} -o ./kronaout/all_samples.kraken.html
145 conda deactivate
146
147 #krakenTools
148 conda create --name python3-env python -y
149 source activate python3-env
150 pip install numpy
151 pip install scipy
152 pip install argparse
153 pip install pandas
154 pip install collection
155 sudo apt-get install unzip
156
157 wget https://github.com/jenniferlu717/KrakenTools/archive/refs/heads/master.zip
158 unzip master.zip
159 rm master.zip
160 for infile in ./kraken2/*report
161 do
162     reportname+=${infile}
163     reportname+=" "

```

```
164 done
165 python KrakenTools-master/combine_kreports.py -r ${reportname} --only-combine -o
    combine.report
166 conda deactivate
167
168 #create json
169 conda create -n kraken-biom -y
170 source activate kraken-biom
171 conda install -c bioconda kraken-biom -y
172 kraken-biom ${reportname} --fmt json -o combine.biom
```

**Listing 11.1:** Contenido del archivo pipeline.sh

