



**Universidad
Europea**

Máster en Bioinformática

**MICROBIOTA Y ESQUIZOFRENIA:
REUSO DE DATOS PÚBLICOS
SEGÚN ESTÁNDARES FAIR**

Autora: Ana Belén Castro García

Tutora: María del Rocío González Soltero

Curso 2023-24

AGRADECIMIENTOS

Primero, quiero agradecer a María del Rocío González, tutora de este trabajo, por guiarme siempre que lo he necesitado, ayudarme y sacar tiempo de donde no había. También a mis familiares, por apoyarme siempre en todo lo que me propongo e interesarse en lo que estaba haciendo y escribiendo.

A Javi, mi compañero de cafés y chismes en el trabajo, y a mis compañeros de Máster, ese grupo de Whatsapp que ha servido de apoyo en reiteradas ocasiones para no tirar la toalla entre tanto estrés y trabajo. Aún no nos conocemos en persona muchos de los componentes de ese grupo, pero ojalá coincidir. Son personas increíbles.

Por último, dar las gracias a Lucía, por ser un apoyo fundamental en mi vida, y a Roberto, mi compañero de viajes y entrenamientos, mi desconexión y mi paz.

ÍNDICE

1. Introducción	8
1.1. Concepto de microbiota	9
1.2. Metagenómica aplicada al estudio de la microbiota.....	11
1.3. Métodos de secuenciación	12
1.4. Principios FAIR.....	14
2. Objetivos.....	15
3. Metodología	15
3.1. Búsqueda bibliográfica	15
3.2. Creación de tablas bibliográfica y principios FAIR.....	16
3.3. Selección de datos.....	19
3.4. Uso de programas informáticos	19
3.5. Datos utilizados y archivos generados	20
4. Resultados.....	20
4.1. Resultados búsqueda bibliográfica.....	20
4.2. Elección de base de datos	22
4.3. Protocolo de amplificación de muestras	24
4.4. Análisis descriptivo y estadístico.....	24
4.5. Diseño del pipeline	25
5. Discusión.....	30
6. Conclusiones.....	32
7. Bibliografía	33
8. Declaración obligatoria del uso de herramientas de IA.....	35

RESUMEN

La esquizofrenia es un trastorno mental que afecta a la función cognitiva y cerebral y conlleva complicaciones asociadas, como puede ser el suicidio. Entre los factores de riesgo encontramos los factores genéticos, ambientales y, recientemente, se ha observado el papel de la microbiota intestinal en la enfermedad debido a la existencia del eje microbiota-intestino-cerebro.

Esta influencia se aborda mediante el estudio de la microbiota con el uso de metagenómica y técnicas avanzadas de secuenciación, subrayando la importancia de los principios FAIR en los datos en investigación.

Objetivos:

Este trabajo aborda la relación entre la microbiota y la esquizofrenia en los estudios reportados mediante una búsqueda bibliográfica que cumpla con los principios FAIR. De este objetivo principal, analizaremos relaciones con las variables presentes en los metadatos asociados.

Material y métodos:

Extraemos de la búsqueda bibliográfica realizada en PubMed artículos a evaluar en los principios FAIR a través de una serie de preguntas depositadas en una encuesta de REDCap. Creamos una tabla bibliográfica, haciendo una selección de artículos que cumplan los principios FAIR para realizar un análisis estadístico posterior y diseñar un pipeline en RStudio utilizando Bioconductor.

Resultados:

La búsqueda realizada nos dio como resultado 526 artículos, de los cuales seleccionamos 13 tras aplicar ciertos filtros. Se evaluaron según los principios FAIR, revelando que varios no los cumplimentan.

Tras el análisis, se observan diferencias significativas donde las personas con esquizofrenia muestran una composición bacteriana diferente a las personas sin enfermedad, destacando una reducción de abundancia de *Firmicutes* y un

aumento de *Actinobacteria* y *Proteobacteria*. Además, hay una reducción en personas con esquizofrenia de la familia *Ruminococcacea*.

Discusión:

Existe una escasez de estudios que cumplan los principios FAIR, limitando las investigaciones futuras. Las diferencias observadas entre la composición de la microbiota de los dos grupos concuerda con los resultados del estudio escogido para el reuso de los datos. Estas diferencias pueden tener un papel importante en el diagnóstico y tratamiento, aun así, la relación causal sigue siendo incierta, por lo que es necesario un mayor tamaño muestral y llevar a cabo análisis estadísticos adicionales.

Conclusiones:

La alteración de la microbiota en personas con esquizofrenia puede utilizarse como un posible marcador en el diagnóstico. Además, es esencial seguir los principios FAIR para fomentar una investigación reproducible y accesible.

Es recomendable evaluar terapias basadas en la modulación de la microbiota para mejorar la salud mental y profundizar en el papel de la microbiota en el eje microbiota-intestino-cerebro.

Palabras clave: Disbiosis, esquizofrenia, metagenómica, microbiota, principios FAIR, secuenciación.

ABSTRACT

Schizophrenia is a mental disorder that affects cognitive and brain function and involves associated complications such as suicide. Among the risk factors we find genetic, environmental factors and, recently, the role of the intestinal microbiota in the disease has been observed due to the existence of the microbiota-gut-brain axis.

This influence is addressed through the study of the microbiota using metagenomics and advanced sequencing techniques, underlining the importance of FAIR principles in research data.

Objectives:

This paper addresses the relationship between microbiota and schizophrenia in reported studies through a literature search complying with FAIR principles. From this main objective, we will analyze the relationships between variables present in the associated metadata.

Methods:

We extracted from the bibliographic search performed in PubMed articles to be evaluated on the FAIR principles by means of a series of questions deposited in a REDCap survey. We created a bibliographic table, making a selection of articles that meet the FAIR principles to perform a subsequent statistical analysis and design a pipeline in RStudio using Bioconductor.

Results:

The search resulted in 526 articles, of which we selected 13 after applying certain filters. They were evaluated according to the FAIR principles, revealing that several did not comply with them.

After analysis, significant differences are observed where people with schizophrenia show a different bacterial composition than people without disease, highlighting a reduction in the abundance of *Firmicutes* and an increase

in *Actinobacteria* and *Proteobacteria*. In addition, there is a reduction in people with schizophrenia of the *Ruminococcacea* family.

Discussion:

There is a paucity of studies that comply with FAIR principles, limiting future research. The differences observed between the microbiota composition of the two groups is consistent with the study chosen for data reuse. These differences may play an important role in diagnosis and treatment, yet the causal relationship remains unclear, requiring larger sample sizes and further statistical analysis.

Conclusions:

Altered microbiota in people with schizophrenia can be used as a potential marker in diagnosis. In addition, it is essential to follow the FAIR principles to promote reproducible and accessible research.

It is advisable to evaluate therapies based on modulation of the microbiota to improve mental health and to further investigate the role of the microbiota in the gut-brain axis.

Keywords: Dysbiosis, schizophrenia, metagenomic, microbiota, FAIR principles, sequencing.

1. Introducción

En la actualidad, los trastornos mentales afectan a una de cada ocho personas en el mundo, siendo considerados un problema de salud global que afecta a toda la población, independientemente de la edad, sexo o país de procedencia.

La *Guía de consulta de los criterios de diagnósticos del DSM-5* (American Psychiatric Association, 2014) define trastorno mental como una alteración clínicamente significativa de cognición, de la regulación de las emociones o del comportamiento de un individuo. También abarca las discapacidades psicosociales, estados mentales asociados a una angustia considerable, discapacidad funcional o riesgo de conducta autolesiva.

Entre los diferentes trastornos mentales, podemos hablar de ansiedad, depresión, trastorno bipolar, estrés postraumático, esquizofrenia o trastornos del comportamiento alimentario.

La esquizofrenia es un trastorno mental caracterizada por presentar delirio y alucinaciones, una desorganización del pensamiento y disminución de la función cerebral y cognitiva. Comienza en la etapa adolescente o en la adultez temprana y presenta una prevalencia de 4 cada 1000 (Gejman & Sanders, 2012). Su diagnóstico es clínico y no existen tratamientos curativos, solo paliativos para poder disminuir los síntomas.

Presenta una alta tasa de mortalidad debido al suicidio en la fase temprana de la enfermedad y las complicaciones cardiovasculares derivadas a la adicción a la nicotina que tienen los enfermos. Por esto, la esperanza de vida disminuye 10-20 años por debajo de la media de la población general (Laursen *et al.*, 2014).

Tiene una etiología heterogénea por presentar diferentes mecanismos patofisiológicos con afección final en la función cerebral. En cuanto a factores de riesgos de padecer esquizofrenia, podemos hablar de un abuso de la marihuana, factores ambientales durante el embarazo o el parto, historial familiar con esquizofrenia y traumatismos encefálicos. Se han identificado ciertos genes específicos como *GRIN2A*, *SP4*, *STAG1* y *FAM120A* asociados a padecer esquizofrenia (Trubetskoy *et al.*, 2022).

Actualmente, existen numerosas revisiones científicas donde se realiza un diagnóstico molecular, relacionando la diversidad microbiana del aparato digestivo y de la cavidad oral en personas con esquizofrenia frente a personas sanas (Yan *et al.*, 2022).

1.1. Concepto de microbiota

El cuerpo humano está colonizado por millones de microorganismos, localizados en múltiples zonas del cuerpo, como la piel, mucosas, vías respiratorias o vías urinarias. Todas estas comunidades de microorganismos se designan con el nombre de microbioma, basado en una relación beneficiosa tanto para los microorganismos como para los humanos (Uzcátegui, 2016).

La humedad, temperatura y nutrientes son características fisiológicas que hacen que aparezcan y prosperen ciertos microorganismos en las diferentes partes del cuerpo. De este, obtienen oxígeno y nutrientes necesarios para sus procesos vitales.

El tracto digestivo de humanos contiene 100 billones de microorganismos, constituyendo más de 1000 especies diferentes y más de 3 millones de genes. Dentro de este pool de microorganismos, que puede llegar a pesar hasta 2 kilogramos (Moya, 2017), podemos encontrar microorganismos beneficiosos para el huésped. Estos microorganismos forman comunidades en todo el aparato, denominándose microbiota intestinal, los cuales aportan genes y ayudan en ciertos procesos fisiológicos como la nutrición o la inmunidad frente a patógenos (J. Álvarez *et al.*, 2021).

La microbiota intestinal humana puede experimentar una alteración en su composición en cualquier momento del transcurso de la vida, denominada disbiosis, que puede presentar un incremento, una disminución o una pérdida de microorganismos y de diversidad microbiana (Moreno Calderón, 2022). Los factores que afectan son:

- **Factores genéticos:** Hay especies que son heredables. También, hay ciertos genes que codifican enzimas e influyen en la composición de la microbiota.

- **Edad:** La variedad de microorganismos varía con cada persona, según los hábitos que presente, y a lo largo de su vida, ya que el microbioma en recién nacidos difiere del microbioma en adultos. En cuanto a la microbiota, existe una influencia dependiente del tipo de parto que tenga lugar, de la alimentación que recibe el neonato en los primeros días de vida y de la utilización de antibióticos (Moreno Calderón, 2022).

Los neonatos presentan un sistema inmune poco desarrollado, el cual obtiene una transferencia microbiológica por parte de la madre a través de la placenta y el líquido amniótico. En el nacimiento, se observa una colonización intestinal de bacterias anaerobias como *Bifidobacterium*, *Bacteriodes* y *Clostridium* (La Rosa Hernández, Gómez Cabeza & Sánchez Castañeda, 2014).

Existe una transición entre la microbiota intestinal en neonatos y en adultos favorecida por un aumento de la ingestión de hidratos de carbono. En el aparato digestivo de adultos, se observa la predominancia de bacterias aerobias, pertenecientes a los filos Firmicutes, Bacteriodes, Actinobacteria y Proteobacteria, y de bacterias anaerobias, que pertenecen a los géneros *Bacteriodes*, *Bifidobacterium*, *Eubacterium*, *Clostridium*, *Lactobacillus*, *Fusobacterium* y cocos gran positivos (V. R. Álvarez *et al.*, 2010).

Las personas mayores presentan un intestino con una microbiota que difiere al de los adultos sanos, como consecuencia del cambio de estilo de vida, de la funcionalidad del aparato digestivo y de la variación del sistema inmune. Las bacterias oportunistas, como *C. difficile* y *C. perfringens*, aumentan y las bacterias comensales disminuyen.

- **Enfermedades y lesiones:** La infección de especies patógenas conlleva una inflamación intestinal, desencadenando una liberación de componentes que afecta a la microbiota existente.
- **Estilo de vida:** La higiene, la dieta y factores ambientales modulan la microbiota. En el caso de la dieta, alimentos ricos en azufre favorece la aparición de microorganismos patógenos y los azúcares refinados y

carbohidratos incrementan la actividad fermentativa de las bacterias (Arce-Hernández, 2020).

Disponemos de una comunicación bidireccional entre la microbiota y el sistema nervioso central, representando el eje microbiota-intestino-cerebro. El desajuste de la composición de la microbiota puede llevar a la persona a padecer un síndrome o enfermedad (Andreo-Martínez *et al.*, 2017).

En estudios, se observó que ciertas bacterias comensales pueden alterar la expresión del neurotransmisor GABA, con función de inhibir el sistema nervioso. La alteración de GABA conlleva ciertas patologías psicológicas como pueden ser la ansiedad y la depresión (Arce-Hernández, 2020).

Debido a esto, analizar y comprender la influencia de la microbiota en la enfermedad de esquizofrenia es un gran avance en el diagnóstico y posible tratamiento.

1.2. Metagenómica aplicada al estudio de la microbiota

Existen numerosas técnicas de secuenciación de alto rendimiento para analizar y comprender la microbiota. La metagenómica es una herramienta utilizada para el estudio del genoma de un entorno específico, sin tener la necesidad de aislar y realizar cultivos en laboratorio de microorganismos, y proporcionando una visión general no sesgada (Robles-Alonso & Guarner, 2013).

Para la identificación de microorganismos, se realiza una extracción del ADN de una muestra biológica para amplificar y secuenciar después los genes. Los métodos utilizados con mayor frecuencia son los basados en marcadores genéticos (como es el 16S rRNA) y la metagenómica *shotgun* (García-Mazcorro *et al.*, 2020).

El gen 16S rRNA es un componente ribosomal de organismos procariotas que consta de unos 1500 nucleótidos y es codificado por el ADN ribosomal 16S. Adquiere una estructura secundaria al plegarse caracterizada por una doble hebra que se mantiene a lo largo del tiempo en todos los organismos. Contiene 9 regiones (V1-V9) hipervariables, es decir, regiones con una variabilidad significativa alta en la secuencia de nucleótidos respecto a otras zonas de la

secuencia, que aportan información filogenética y taxonómica (Valenzuela-González *et al.*, 2015). Por ejemplo, las regiones V3 y V4 se utilizan para la identificación y clasificación de la microbiota intestinal por presentar una alta variabilidad entre bacterias (Sánchez-Salguero & Santos-Argumedo, 2018).

La secuenciación del gen 16S rRNA se basa en una amplificación de regiones específicas, donde la molécula 16S rRNA es utilizada como marcador molecular (Bharti & Grimm, 2021). Este nos proporciona información sobre procariontes, siendo los datos generados manejables y sencillos de analizar.

Además, también se utiliza la secuenciación *shotgun*, la cual proporciona mayor información, no solo del microbioma, si no de funciones genéticas y metabólicas de los microorganismos de la muestra. La metagenómica *shotgun* se basa en una secuenciación de todo el material genético de una muestra, ya sea bacteria, hongo, virus o cualquier organismo presente, de manera simultánea.

1.3. Métodos de secuenciación

Las tecnologías de secuenciación han avanzado a lo largo de los años presentando múltiples ventajas, aunque aún con ciertas limitaciones. La principal ventaja es que permiten secuenciar masivamente millones de fragmentos de ADN en numerosos individuos, disminuyendo el tiempo de realización y los costes, pero dando como resultado un elevado número de datos. Debido a esto, es necesario un requerimiento de sistemas computacionales complejos y un personal específico especializado en bioinformática y en el análisis de datos.

Las técnicas presentan dos fases: una fase *in vitro*, la cual se basa en construir las bibliotecas genómicas que serán utilizadas después, y una fase *in silico*, donde se realiza el análisis bioinformático de los datos obtenidos de la secuenciación (López De Heredia, 2016).

Existen diferentes métodos de secuenciación de ADN que se clasifican según su desarrollo en referencia a la tecnología utilizada y a la capacidad de secuenciación:

- **Secuenciación de Primera Generación:** Basado en la secuenciación de “terminación de cadena”, conocida como el método de Sanger. En este método se utilizan didesoxinucleótidos (ddNTPs), provocando la terminación de la cadena. Los fragmentos resultantes son separados según su longitud por electroforesis y el tamaño máximo de lectura es en torno a 1000 pares de bases.

Presenta una alta precisión, más aún en secuencias cortas, pero es un proceso lento y costoso.

- **Secuenciación de Segunda Generación (Secuenciación de Alto Rendimiento o NGS):** Basado en la secuenciación masiva en paralelo de millones de fragmentos de ADN de forma simultánea. Se divide según la longitud de las lecturas que se generan en plataformas de lecturas largas, basadas en más de 300 pares de bases, y de lecturas cortas, basadas en menos de 300 pares de bases.

La diferencia entre ambas radica en que una longitud de lecturas cortas conlleva necesariamente el mapeo con un genoma de referencia mientras que, una longitud de lecturas largas tiene una mayor tasa de error respecto a la identificación de los nucleótidos.

- **Secuenciación de Tercera Generación:** Tecnología más reciente en la cual no es necesario realizar una PCR (reacción en cadena de la polimerasa) previa a la secuenciación, siendo una diferencia respecto a los dos tipos anteriores. Esta diferencia hace que el tiempo de análisis se vea reducido y que la detección de nucleótidos sea realizada en tiempo real.

Las dos plataformas más utilizadas hoy en día en metagenómica son Illumina y Oxford Nanopore. La tecnología Illumina se considera una secuenciación por síntesis, con una longitud de lecturas cortas donde se lleva a cabo una amplificación de los fragmentos de ADN y una detección de secuencias mediante el marcaje con fluorescencia. En cambio, la tecnología Nanopore se basa en la secuenciación de nanoporos, capaz de llevar a cabo lecturas de secuencias largas y en tiempo real (Xia *et al.*, 2023).

En la figura 1 podemos observar un esquema de los pasos a llevar a cabo en un experimento de secuenciación masiva y los diferentes tipos de secuenciación de ADN.

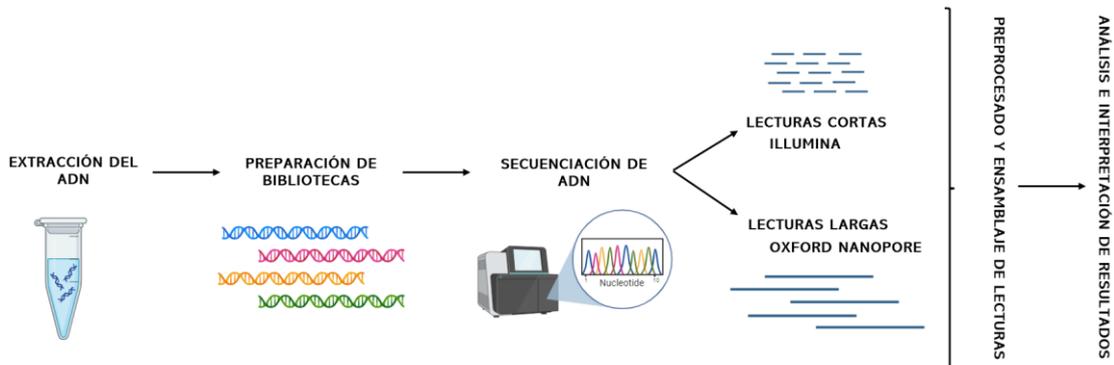


Figura 1 – Esquema general de los procesos en un experimento de secuenciación masiva. Creado en BioRender.

1.4. Principios FAIR

Todos los datos en el ámbito científico-tecnológico tienen que ser validados para los principios FAIR, es decir, tienen que ser localizables, accesibles, interoperables y reutilizables debido al progreso y el avance producido en la ciencia (García-Espinosa *et al.*, 2022).

Una forma de localizar los datos es asignando un identificador único, que tiene la función de identificador en investigación, sobre todo en bases de datos genómicas y biológicas. Un identificador utilizado en investigación es el número de acceso PRJ, basado en una secuencia alfanumérica. Los datos quedan asignados a un PRJ en BioProject de NCBI (National Center for Biotechnology Information), facilitando a otros investigadores el acceso a los datos, gestionarlos, utilizarlos y recuperar la información. Por esto, sin un identificador único o enlace específico a los datos, multitud de artículos no cumplen con los principios FAIR y es imposible usar los datos utilizados por otros investigadores para refutar o reusar los datos.

Además, tener un metadato asociado a los datos de la muestra que nos aporte una descripción o información sobre los datos en sí, es imprescindible para poder saber procedencia y tipo de dato, así como para realizar un análisis estadístico básico de las muestras y sectorizar los datos en diversas categorías si fuera necesario.

2. Objetivos

El objetivo general del presente trabajo es analizar si existe relación significativa entre la microbiota y la esquizofrenia en los estudios reportados, hasta la fecha, en humanos. Para ello, localizaremos una base de datos de estudios previamente publicados que cumplan estándares FAIR y elaboraremos un protocolo de análisis de los datos.

Del objetivo general observaremos varios objetivos secundarios:

O2: Identificar en bibliografía estudios que relacionen microbiota con esquizofrenia

O3: Analizar si estos estudios tienen depositados datos en abierto siguiendo estándares FAIR, es decir, que cumplan con ser encontrables, accesibles, interoperables y reusables.

O4: Preprocesar datos para posterior análisis.

O5: Buscar relaciones significativas en los datos teniendo en cuenta variables como sexo, aparición diferencial de bacterias, etc.

3. Metodología

3.1. Búsqueda bibliográfica

Se realizó una revisión bibliográfica sistemática en PubMed, que consiste en una base de datos de literatura científica gratuita y gestionada por la *National Library of Medicine*. Para la revisión bibliográfica, se usaron los operadores booleanos “AND” para incluir todos los términos y “OR” para buscar al menos un término en la búsqueda.

La fórmula introducida en la base de datos PubMed para obtener artículos científicos específicos que relacionen la enfermedad con la microbiota fue:

(microbiota OR microbiome) AND schizophrenia

La bibliografía seleccionada fue a criterio de la autora, priorizando artículos de la actualidad sobre otros más antiguos, y que la especie de estudio sea el humano.

3.2. Creación de tablas bibliográfica y principios FAIR

Se crea una tabla en Excel la cual contiene, además de título y autor, el número de acceso de los datos si lo tuviese, el tipo de dato, si hay metadata asociado, el tipo de secuenciación utilizado para la extracción de los datos y el tipo de muestra.

Para observar si se cumplen los principios FAIR, utilizamos una escala predeterminada. En este caso, tenemos a disposición del alumnado una herramienta para la autoevaluación de los datos llamada REDCap, donde obtenemos un valor numérico según la opción escogida en cada pregunta (Bellido Esteban *et al.*, 2020). El enlace a la autoevaluación es:

<https://redcap.universidadeuropea.com/surveys/?s=L89WYPJ3RFL3D7WP>

El puntaje se distribuye en varias preguntas que evalúan los principios de menor a mayor, siendo el número menor que cumple el principio FAIR en cuestión y el número mayor que no lo cumple.

Las preguntas de la encuesta se dividen en los cuatro principios y cada pregunta tiene 5 opciones a elegir:

Findable:

- ¿Tiene el conjunto de datos algún identificador asignado?
 - (1) Único, citable y persistente (ejemplo: DOI, ARK o Handle)
 - (2) Dirección web (URL)
 - (3) Identificador local
 - (4) Sin identificador

- ¿Se incluye el identificados del conjunto de datos en todos los registros/datos que describen?
 - (1) Sí
 - (2) No

- ¿Cómo se describen los datos con metadatos?
 - (1) Utilización exhaustiva de un esquema formal reconocido de metadatos legibles por máquina.
 - (2) De forma exhaustiva, pero en un formato de texto no estándar.
 - (3) Breve título y descripción.
 - (4) Los datos no están descritos.

- ¿En qué tipo de repositorio se encuentra el registro de metadatos?
 - (1) Los datos se encuentran en un único lugar, pero pueden consultarse a través de varios registros.
 - (2) Repositorio público general.
 - (3) Repositorio específico de área de estudio.
 - (4) Repositorio local institucional.
 - (5) Los datos no están descritos en ningún repositorio.

Accesible:

- ¿Hasta qué punto son accesibles los datos?
 - (1) Acceso público.
 - (2) Totalmente accesible a las personas que cumplan las condiciones explícitamente establecidas, por ejemplo, aprobación ética para datos sensibles.
 - (3) Un subconjunto no identificable o modificado de los datos es de acceso público.
 - (4) Acceso condicional no especificado, por ejemplo, "póngase en contacto con el responsable de la custodia de los datos para obtener acceso".
 - (5) Acceso sólo a metadatos.
 - (6) Sin acceso a datos o metadatos.

- ¿Están los datos disponibles en línea sin necesidad de protocolos una vez aprobado el acceso?
 - (1) API estándar de servicios web (p. ej., OGC)
 - (2) Servicio web no estándar (por ejemplo, OpenAPI/Swagger/API informal)
 - (3) Descarga de archivos en línea.
 - (4) Mediante acuerdo individual.
 - (5) Ningún acceso a los datos.

Interoperable:

- ¿En qué formato están disponible los datos?
 - (1) En un formato estructurado, de estándar abierto y legible por máquinas.
 - (2) En un formato estructurado, de estándar abierto y no legible por máquinas.
 - (3) Sobre todo en formato cerrado (proprietary format).
- ¿Qué describe mejor los tipos de vocabularios/ontologías/esquemas utilizados para definir los datos?
 - (1) Normalizado, abierto y universal mediante identificadores globales resolubles que enlazan con la explicación.
 - (2) Vocabularios/ontologías/esquemas normalizados sin identificadores globales.
 - (3) No se han aplicado normas en la descripción de los datos.
 - (4) Datos no descritos.
- ¿Cómo se vinculan los metadatos con otros datos y metadatos?
 - (1) Los metadatos se representan en un formato legible por máquina, por ejemplo, en un formato de datos enlazados como el Resource Description Framework (RDF).
 - (2) El registro de metadatos incluye enlaces URI (uniform resource identifier) a metadatos, datos y definiciones relacionados.
 - (3) No hay links a otros metadatos.

Reusable:

- ¿Cuál de las siguientes opciones describe mejor los derechos de licencia/uso asociado a los datos?
 - (1) Licencia estándar legible por máquina (por ejemplo, Creative Commons)
 - (2) Licencia estándar basada en texto.
 - (3) Licencia no estándar legible por máquina (que indique claramente en qué condiciones se pueden reutilizar los datos)
 - (4) Licencia no estándar basada en texto.
 - (5) Sin licencia.

- ¿Cuánta información de procedencia se ha capturado para facilitar la reutilización de los datos?
 - (1) Totalmente registrado en un formato legible por máquina.
 - (2) Totalmente registrado en formato de texto.
 - (3) Parcialmente registrado.
 - (4) No se ha registrado información sobre la procedencia.

De todos los artículos seleccionados, escogeremos el que menor puntaje presenta en la encuesta de principios FAIR.

3.3. Selección de datos

Una vez definidos los artículos que cumplen los principios FAIR, la selección de datos la realizamos según el tipo de secuencia que viene asociado al estudio. Actualmente, los estudios los podemos dividir en si utilizan secuencias largas o secuencias cortas, por lo que es un filtro a utilizar.

3.4. Uso de programas informáticos

En la realización del análisis estadístico de variables, ver correlaciones significativas entre los datos obtenidos del artículo y el diseño de un pipeline haremos uso de RStudio 4.3.2.

En cuanto a repositorio de paquetes utilizados en RStudio, emplearemos Bioconductor, ya que nos realiza un análisis de datos biológicos y para visualizar los resultados mediante gráficos, la librería ggplot2.

3.5. Datos utilizados y archivos generados

Los datos utilizados en este trabajo son dispuestos en una carpeta compartida en OneDrive en el siguiente enlace, así como un archivo con el código seguido para el análisis en formato R para su posible revisión y reproducibilidad:

https://liveuem-my.sharepoint.com/:f/g/personal/22361475_live_uem_es/EiShyuKnjg5OqIPhCiA8T9IBhAxdex3c6lO-WmKGINrz9Q?e=SWCicd

4. Resultados

4.1. Resultados búsqueda bibliográfica

Al realizar la búsqueda en PubMed, obtenemos 526 resultados de los cuales, utilizando la herramienta Covidence para generar un diagrama de flujo (Figura 2), detectamos que un artículo científico se encuentra duplicado. Tras esto, seleccionamos el filtro “Humans”, reduciendo los resultados a 337 búsquedas y si seleccionamos el filtro de artículos en los últimos 10 años, se reduce a 330 búsquedas. El rango de años con mayor número de publicaciones se distribuye desde 2020 a la actualidad. Esto nos muestra que es un campo de investigación actual.

De los resultados de la búsqueda final, seleccionamos 13 artículos al azar para crear una tabla bibliográfica (Tabla 1).

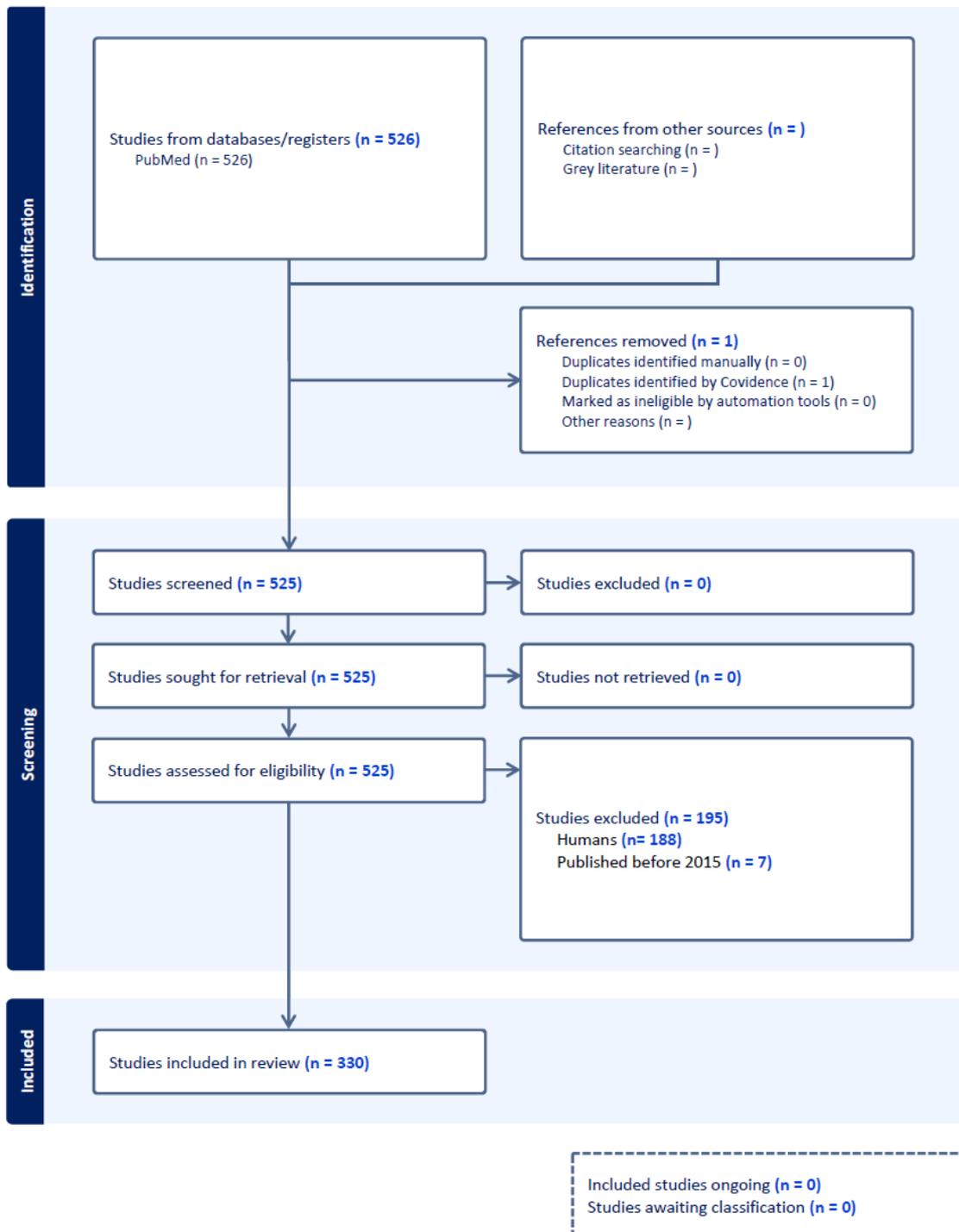


Figura 2 - Diagrama de flujo bibliográfico seguido. Creado en Covidence.

4.2. Elección de base de datos

Observamos que ciertos artículos no siguen los principios FAIR (Tabla 1), ya que no existe un metadata asociado más allá del lugar donde se realizó el estudio en cuestión, ni siquiera si la muestra es de hombre o mujer o la edad de la persona.

La encuesta de principios FAIR solo fue realizada en los artículos científicos que tienen número de acceso o enlace al que poder acceder y extraer los datos debido a que, sin este requisito previo, se incumplen los cuatro puntos de los principios FAIR.

En cuanto a los artículos que tienen número de acceso, podemos observar dos puntuaciones diferentes, siendo la puntuación más alta una peor valoración respecto a los principios FAIR. Destacan dos artículos científicos con una puntuación de 25 en la encuesta realizada en REDCap: Castro-Nallar *et al.*, (2015); y Li *et al.*, (2020).

Como no disponemos de capacidad de cómputo, por realizar esta revisión y análisis en un dispositivo poco potente en cuanto a software para utilizar secuencias largas, y lo que queremos analizar es si una enfermedad está relacionada con la microbiota, escogemos el método de secuenciación del gen 16S rRNA frente al método *shotgun*. Además, son más sencillos de reusar respecto a los datos obtenidos del método de secuenciación *shotgun* de Castro-Nallar.

Tabla 1 - Resultados escogidos de búsqueda bibliográfica en PubMed y Google Scholar.

Cita	Título	Tecnología de secuenciación	Número de acceso	BBDD	Metadato asociado	Tamaño muestral	Tipo de datos	Secuencia	Tipo de muestra	Formato del dato	FAIR
Sen et al., 2024	Dysregulation of Microbiota in Patients With First-Episode Psychosis Is Associated With Symptom Severity and Treatment Response	Illumina HiSeq 4000	PRJNA1044118	NCBI	NA	66	Shotgun	DNA	Fecal	FASTA/FASTQ	31
Yolken et al., 2021	The oropharyngeal microbiome is altered in individuals with schizophrenia and mania	Illumina	NA	NA	NA	316	Amplicones	16S rRNA	Orofaringea	NA	No
Cui et al., 2021	Salivary Metabolomics Reveals that Metabolic Alterations Precede the Onset of Schizophrenia	NA	NA	NA	NA	208	NA	NA	Saliva	NA	No
Castro-Nallar et al., 2015	Composition, taxonomy and functional diversity of the oropharynx microbiome in individuals with schizophrenia and controls	Illumina HiSeq	PRJNA255439	NCBI	Si	32	Shotgun	DNA	Orofaringea	FASTA/FASTQ	25
Qing et al., 2021	Salivary microbiome profiling reveals a dysbiotic schizophrenia-associated microbiota	Illumina HiSeq 2500	PRJNA647054	NCBI	NA	208	Amplicones	16S rRNA	Saliva	FASTA/FASTQ	31
Krzysciak et al., 2023	The Association of the Oral Microbiota with the Effects of Acid Stress Induced by an Increase of Brain Lactate in Schizophrenia Patients	NA	NA	NA	NA	40	NA	NA	Orofaringea	NA	No
Lee et al., 2023	Gut and oral microbiome modulate molecular and clinical markers of schizophrenia-related symptoms: A transdiagnostic, multilevel pilot study	Illumina MiSeq	PRJNA978061	NCBI	NA	23	Amplicones	16S rRNA	Fecal y oral	FASTA/FASTQ	31
Thirion et al., 2022	Alteration of Gut Microbiome in Patients With Schizophrenia Indicates Links Between Bacterial Tyrosine Biosynthesis and Cognitive Dysfunction	Ion torrent	PRJEB41217, PRJEB41786, PRJEB41787	NCBI	NA	396	Shotgun	DNA	Fecal	FASTA/FASTQ	31
Yan et al., 2022	A comparative study to determine the association of gut microbiome with schizophrenia in Zhejiang, China	Illumina Miseq	PRJNA880407	NCBI	NA	100	Amplicones	16S rRNA	Fecal	FASTA/FASTQ	31
Shi et al., 2023	Intricate role of intestinal microbe and metabolite in schizophrenia	Illumina HiSeq 2500	NA	NCBI	NA	124	Amplicones	16S rRNA	Fecal	NA	No
O'Donnell et al., 2022	The Role of the Microbiome in the Metabolic Health of People with Schizophrenia and Related Psychoses: Cross-Sectional and Pre-Post Lifestyle Intervention Analyses	Illumina MiSeq	PRJNA896324	NCBI	NA	44	Amplicones	16S rRNA	Fecal	FASTA/FASTQ	31
Munawar et al., 2022	Modulation of Gut Microbial Diversity through Non-Pharmaceutical Approaches to Treat Schizophrenia	NA	NA	NA	NA	NA	NA	16S rRNA	Fecal	NA	No
Li et al., 2020	Altered gut microbiota associated with symptom severity in schizophrenia	Illumina MiSeq	https://peerj.com/articles/9574/#supplementary-material	peerj	Si	162	Amplicones	16S rRNA	Fecal	FASTA	25

NA: No aparece.

4.3. Protocolo de amplificación de muestras

Los datos utilizados del artículo científico escogido se obtuvieron de muestras fecales de los diferentes participantes, del cual se extrajo el DNA y se realizó la amplificación de la región V4. Se utilizó la tecnología Illumina MiSeq en la secuenciación de alto rendimiento.

4.4. Análisis descriptivo y estadístico

Disponemos de un archivo con metadata asociada a la muestra de estudio, así como una tabla con la riqueza taxonómica de cada muestra a estudio y un archivo con las secuencias extraídas en formato FASTA:

La base de datos está compuesta por 162 individuos, de los cuales se dividen según el sexo y si presentan o no esquizofrenia:

	<i>CONTROL (NC)</i>	<i>ESQUIZOFRENIA (SZ)</i>
<i>MUJERES</i>	41	35
<i>HOMBRES</i>	39	47

Las variables que encontramos en los datos generales son ID, grupo al que pertenecen (sano o esquizofrenia), edad, sexo, altura, peso, si es fumador, si bebe alcohol, niveles de colesterol, glucosa, lipoproteínas y si tienen tratamiento.

Utilizamos la correlación punto-biserial para ver si existe relación alguna entre una variable continua y una variable binaria. Esta correlación nos proporciona como resultado un coeficiente entre el rango -1 y 1, donde los valores cercanos a 0 no presentan relación alguna.

Con los datos obtenidos de Li *et al.* obtenemos como resultado que:

- Existe una correlación muy débil de presentar esquizofrenia en edades más jóvenes, siendo el coeficiente de correlación -0.0410059.

- Existe una correlación positiva entre la esquizofrenia y tener altos los valores de colesterol total y glucosa en sangre, siendo los coeficientes de correlación 0.5689503 y 0.4279271 respectivamente.
- No existe correlación entre la esquizofrenia y triglicéridos, HbA1c (hemoglobina glicosilada) y lipoproteínas, obtenido coeficientes bajos.

4.5. Diseño del pipeline

Partimos de tres archivos diferentes:

- Un archivo con secuencias de ADN en formato FASTA.
- Un archivo con una tabla de unidades taxonómicas operativas.
- Un archivo con datos descriptivos de cada muestra.

Previo al diseño, tenemos la necesidad de instalar varias librerías:

- Biostrings: es una librería de Bioconductor con la funcionalidad de buscar, alinear y comparar secuencias. Útil en análisis de datos de secuencias biológicas y en manipulación datos genómicos de gran volumen.
- DADA2: diseñado para analizar datos de secuencias procedentes de 16S rRNA, donde se puede detectar y corregir los errores que contengan los datos.
- Phyloseq: especializada en microbioma y filogenia de esta mediante el análisis de datos metagenómicos.

Además, descargamos el archivo SILVA del repositorio de acceso abierto Zenodo. Se trata de un dataset de referencia tanto de lecturas largas como cortas, el cual contiene información taxonómica de los dominios Bacteria, Arquea y Eucariota. SILVA nos ayuda en la asignación taxonómica de las secuencias de nuestro archivo FASTA, clasificándolas en Reino, Filo, Clase, Orden, Familia, Género y Especie.

Para realizar una determinación de una posible alteración de la microbiota entre grupo sano y grupo con esquizofrenia, aplicamos un análisis de índice de diversidad alfa y beta con sus respectivas pruebas estadísticas.

El análisis de diversidad alfa se utiliza para evaluar la diferencia en la diversidad entre el grupo sano y el grupo con esquizofrenia con la utilización de las métricas Shannon y Simpson. El índice de Shannon considera la riqueza y la uniformidad de las especies mientras que el índice de Simpson mide principalmente la abundancia de especies.

Las muestras de enfermos con esquizofrenia tienden a valores más bajos en el índice Shannon (Figura 3) sugiriendo que existe una menor riqueza en la microbiota intestinal respecto a las personas sanas. En el índice Simpson, el grupo esquizofrenia alcanza valores mayores, corroborando el resultado del índice Shannon.

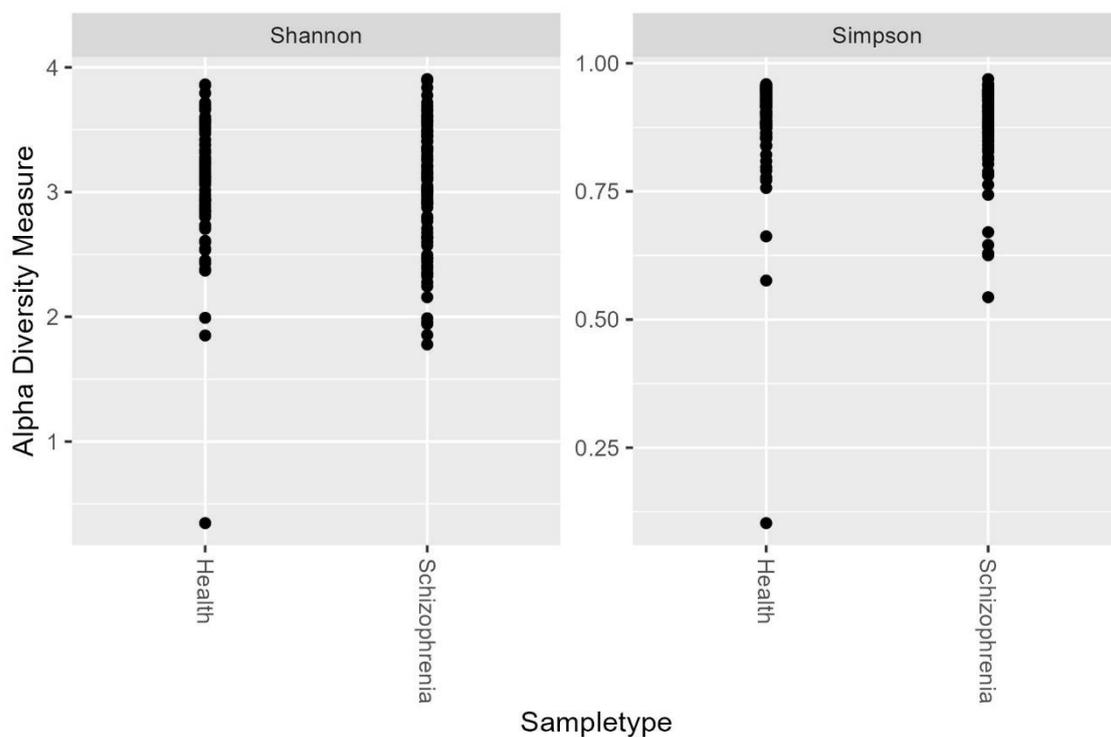


Figura 3 - Análisis de diversidad alfa de grupo sano frente a grupo con esquizofrenia utilizando las métricas Shannon y Simpson.

Para la realización de una prueba estadística de evaluación de diversidad alfa utilizamos el test de Wilcoxon, ya que el grupo sano no presenta una distribución normal. El p-valor es de 0.3738 para el índice de Shannon y de 0.4224 para el índice de Simpson, lo cual nos indica que no hay una diferencia estadística significativa y, por tanto, la diversidad interna de la microbiota entre los dos grupos es similar.

El análisis de diversidad beta utiliza una representación de PCoA (Principal Coordinates Analysis), basándose en una matriz de distancias Bray-Curtis, con la finalidad de representar las diferencias en la composición microbiana entre sanos y enfermos.

Las muestras del grupo de personas sanas se superponen con el grupo de enfermos (Figura 4) y no se observa un agrupamiento claro, ya que existe una gran variabilidad en la composición de la microbiota

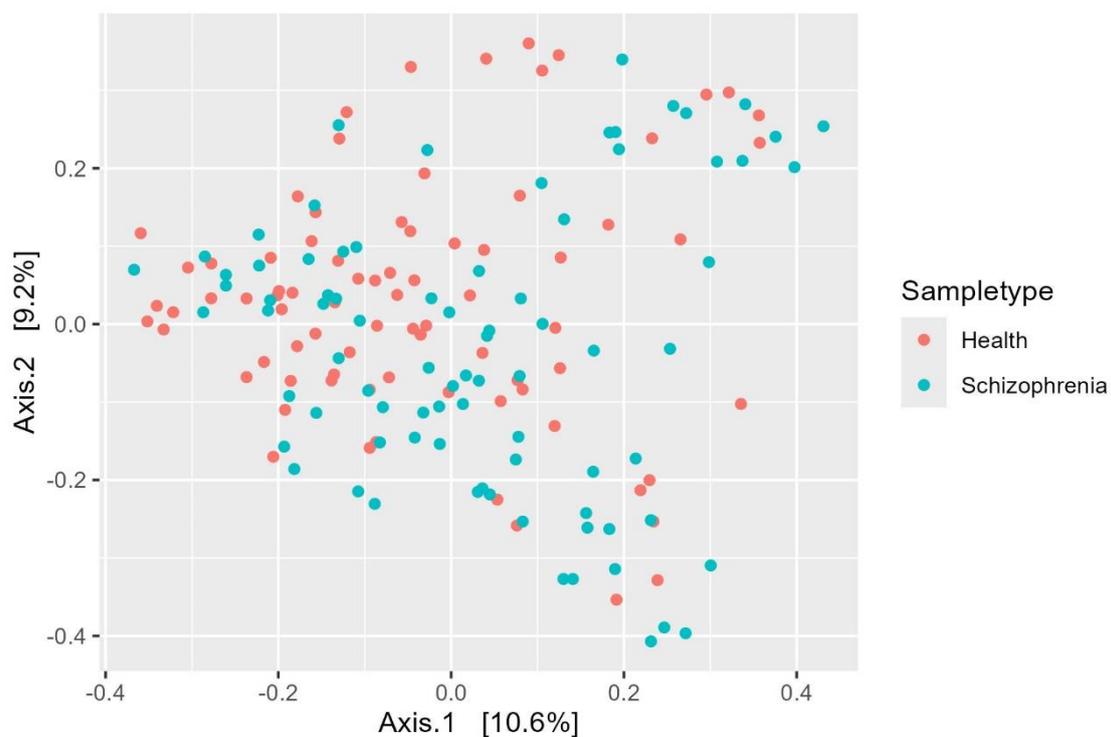


Figura 4 – Gráfico de Análisis de Coordenadas Principales (PCoA) donde muestra la diversidad beta en grupo sano y con esquizofrenia.

La interpretación del análisis de diversidad beta la realizamos mediante la prueba estadística PERMANOVA, evaluando la significancia entre grupos. El p-valor es de 0.001, indicando que es significativo, y R^2 es igual a 0.02021. Los datos obtenidos sugieren que se presentan alteraciones en la composición en la microbiota de personas con esquizofrenia en comparación con la composición de la microbiota de personas sanas.

Al analizar las abundancias relativas tanto por filo (Figura 5) como por familia (Figura 6) en los dos grupos, observamos ciertas diferencias:

En ambos grupos, el filo *Firmicutes* tiene dominancia, presente en más del 70% del total de las bacterias, aunque el porcentaje de este es mayor en personas sanas. También existe una diferencia en *Actinobacterias* y *Proteobacterias* donde el porcentaje es mayor en el grupo de enfermos de esquizofrenia.

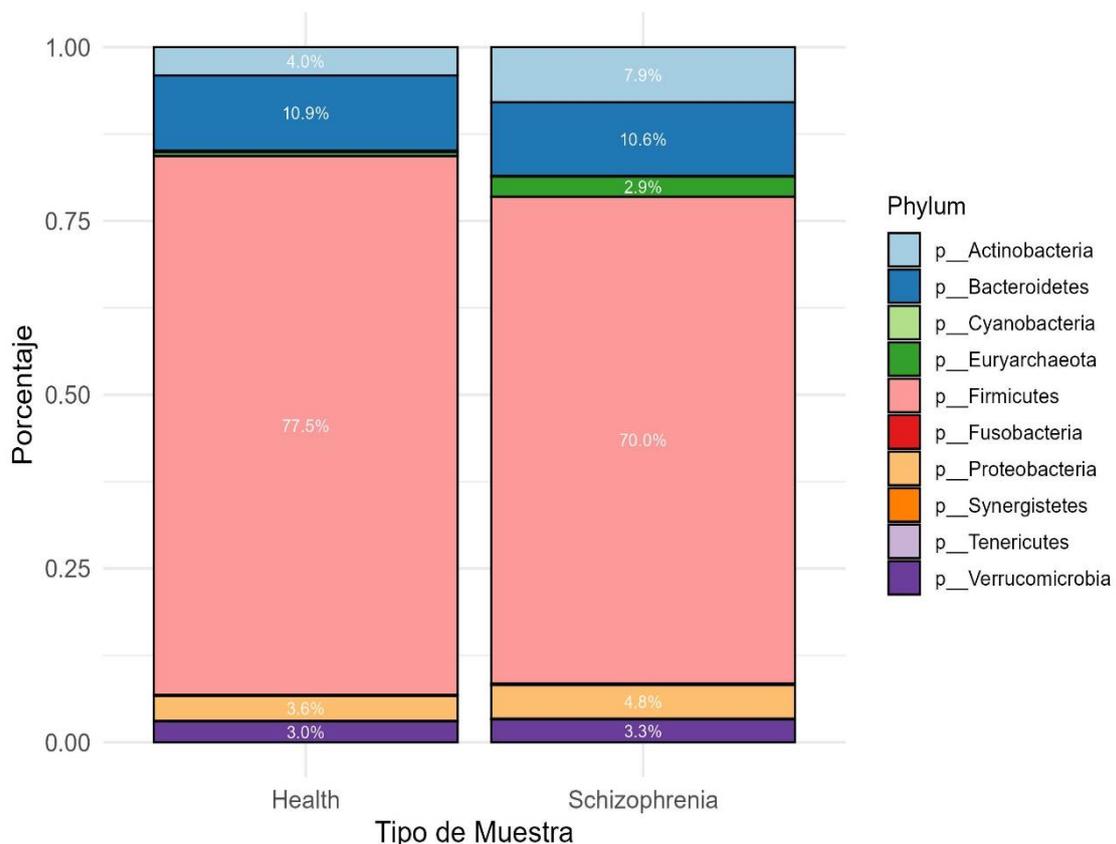


Figura 5 - Gráfico de abundancia microbiana agrupado por filo en grupo sano y grupo esquizofrenia.

Si profundizamos en las abundancias agrupadas por familia, observamos una abundancia de *Lachnospiraceae*, *Ruminococcaceae* y *Veillonellaceae*, pertenecientes las tres al filo *Firmicutes*. La mayor diferencia la podemos observar en la familia *Ruminococcaceae*, presentando un porcentaje de 34.3% en el grupo sano y reduciéndose a 24.1% en el grupo de personas con esquizofrenia.

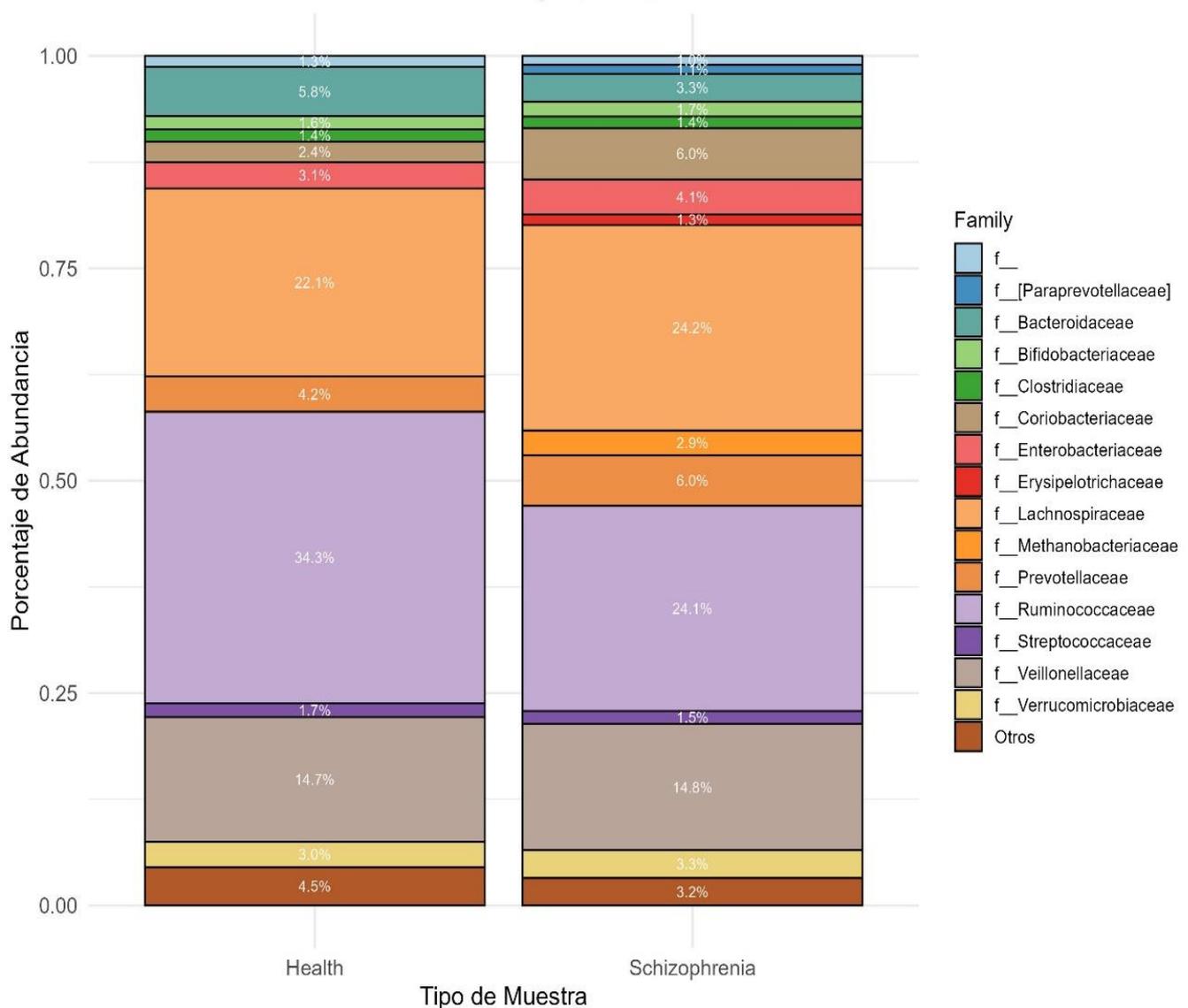


Figura 6 - Gráfico de abundancia microbiana agrupado por familias en grupo sano y grupo esquizofrenia.

5. Discusión

Tras la revisión bibliográfica realizada, es notoria la falta de los principios FAIR en los datos asociados a los artículos científicos. Dentro de un artículo científico, en general, es complicado localizar dónde se encuentran los datos o dónde los podemos encontrar, pese a que múltiples artículos tienen un apartado de información complementaria o similar.

Una vez se localiza la referencia a esos datos, hemos observado que, o no hay datos, o se utilizan bases de datos de hospitales que no las hacen públicas por tener una base de datos propia y por privacidad, lo que disminuye la interoperabilidad a cero. En el caso que haya un número de acceso, hay que probar que sea válido.

Esto genera un gran problema de reproducibilidad, incluso de fiabilidad de los datos, ya que es imposible reproducir el estudio en cuestión. Se crea una barrera para la mejora en la investigación al no poder acceder de forma abierta a esos datos.

En mayo del 2022, se propuso un Reglamento para el Espacio Europeo de Datos de la Salud (EHDS) con la finalidad de regular la gestión, acceso y uso seguro de los datos en toda la Unión Europea (Marelli *et al.*, 2023). La propuesta se apoyó en reglamentos preexistentes como el Reglamento General de Protección de Datos (GDPR) y, recientemente, se aprobó la regulación como lo que se facilitará el uso de datos en el ámbito clínico y de salud pública (Li & Quinn, 2024).

Otra limitación, a nivel personal, es la capacidad de cómputo a la hora de utilizar los datos. Dependiendo del software que se disponga, se podrá utilizar un tipo u otro de datos. En nuestro caso, utilizamos datos de secuenciación del gen 16S rRNA para realizar el análisis e interpretación de resultados, siendo efectivo a nivel de filo y familia mientras que, si se quiere analizar a nivel de especie, es recomendable utilizar datos *shotgun* (Ranjan *et al.*, 2016).

Una vez elaborada la revisión bibliográfica y obtener unos datos válidos, realizamos un análisis estadístico en el que obtenemos como resultado una correlación entre presentar esquizofrenia y tener altos los valores de colesterol

total y glucosa en sangre. Se ha observado que determinados antipsicóticos producen una alteración en la función metabólica, siendo una posible variable para analizar y tener en cuenta en estudios futuros (Pillinger *et al.*, 2020; Lindenmayer *et al.*, 2003).

Los análisis posteriores sugieren que la microbiota intestinal presenta una alteración en enfermos de esquizofrenia frente a personas sanas, pero no es un marcador definitivo, ya que es necesario análisis estadísticos adicionales a este estudio, así como un mayor tamaño muestral, para poder determinar con certeza la relación. Existen investigaciones que acotan y demuestran que se puede presentar biomarcadores significantes y que son capaces de distinguir muestras de esquizofrenia frente a muestras control (Shen *et al.*, 2018; Pan *et al.*, 2020)

Li *et al.*, (2020) obtuvo como resultado que el grupo sano presentaba un enriquecimiento mayor de *Firmicutes*, mientras que el grupo de esquizofrenia presentaba un enriquecimiento mayor de *Actinobacteria*, concordando con nuestros resultados. Detectaron, también, cambios en la abundancia de la bacteria *Succinivibrio*, apareciendo en muestras de enfermos de esquizofrenia, pero no en muestras de personas sanas.

Bacterias como *Haemophilus*, *Sutterella* y *Clostridium* disminuyen en muestras de esquizofrenia (Nguyen *et al.*, 2019) y, al igual que observamos en nuestros resultados, la disminución de la abundancia de *Ruminococcaceae* está relacionada con las muestras de personas sanas (Thirion *et al.*, 2023).

Todos estos hallazgos diferenciales entre los dos grupos podrían aplicarse al ámbito clínico como prevención, ajustando la dieta o teniendo conciencia de ello, para hacer que estas diferencias en la microbiota sean mínimas o nulas.

Si es cierto que la microbiota presenta un papel importante en la salud mental y cada vez existe más evidencia científica que lo respalda, pero es un ámbito de la ciencia que está en continua investigación actualmente.

6. Conclusiones

En investigación es fundamental que los datos utilizados en los artículos científicos sigan los principios FAIR para garantizar su reuso, reproducibilidad e interoperabilidad, especialmente en estudios sobre la microbiota intestinal, siendo un potencial marcador a estudiar en enfermedades como la esquizofrenia.

Las alteraciones observadas en la microbiota, tanto en su diversidad interna como en su composición, indican la necesidad de contar con modelos válidos que puedan detectar estas modificaciones, permitiendo un diagnóstico temprano antes de realizar otras pruebas más específicas.

Además, es imprescindible aumentar el número de estudios que exploren la relación entre el eje microbiota-intestino-cerebro, lo que podría abrir la posibilidad de terapias basadas en la modulación de la microbiota. También resulta crucial que investigaciones futuras profundicen en la determinación de la causalidad entre la microbiota y la esquizofrenia.

7. Bibliografía

American Psychiatric Association. (2014). Guía de consulta de los criterios diagnósticos del DSM-5. Editorial Médica Panamericana.

Gejman, P. V., & Sanders, A. R. (2012). La etiología de la esquizofrenia. *Medicina (Buenos Aires)*, 72(3), 227-234.

Laursen, T. M., Nordentoft, M., & Mortensen, P. B. (2014). Excess early mortality in schizophrenia. *Annual Review of Clinical Psychology*, 10(1), 425–448. <https://doi.org/10.1146/annurev-clinpsy-032813-153657>

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., Lam, M., Watanabe, K., Frei, O., Ge, T., Harwood, J. C., Koopmans, F., Magnusson, S., Richards, A. L., Sidorenko, J., Wu, Y., ... Schizophrenia Working Group of the Psychiatric Genomics Consortium (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, 604(7906), 502–508. <https://doi.org/10.1038/s41586-022-04434-5>

Yan, F., Xia, L., Xu, L., Deng, L., & Jin, G. (2022). A comparative study to determine the association of gut microbiome with schizophrenia in Zhejiang, China. *BMC Psychiatry*, 22(1). <https://doi.org/10.1186/s12888-022-04328-w>

Uzcátegui U, Ofelia. (2016). Microbioma humano. *Revista de Obstetricia y Ginecología de Venezuela*, 76(1), 1-3. Recuperado en 07 de agosto de 2024, de http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0048-77322016000100001&lng=es&tlng=es.

Moya, A. S. (2017). Microbioma y secuenciación masiva. *Rev Esp Quimioter*, 30(5), 305-311.

Álvarez, J., Fernández Real, J. M., Guarner, F., Gueimonde, M., Rodríguez, J. M., Saenz de Pipaon, M., & Sanz, Y. (2021). Microbiota intestinal y salud. *Gastroenterología y Hepatología*, 44(6), 519-535. <https://doi.org/10.1016/j.gastrohep.2021.01.009>

Moreno Calderón, X. (2022). Disbiosis en la microbiota intestinal. *Revista GEN*, 76(1), 17–23. http://saber.ucv.ve/ojs/index.php/rev_gen/article/view/24183

La Rosa Hernández, D., Gómez Cabeza, E. J., & Sánchez Castañeda, N. (2014). Intestinal microbiota in the development of the neonate's immune system. *Revista Cubana de Pediatría*, 86(4), 502-513. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-75312014000400011&lng=es&tlng=en

Ruiz Álvarez, V., Puig Peña, Y., & Rodríguez Acosta, M. (2010). Microbiota intestinal, sistema inmune y obesidad. *Revista Cubana de Investigaciones Biomédicas*, 29(3), 364-397. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-03002010000300007&lng=es&tlng=es

Arce-Hernández, W. (2020). Disbiosis intestinal: Alteración de la relación mutualista entre microbiota y sistema inmune. *Acta Académica*, 67, 171-182. <http://webservertest.uaca.ac.cr/index.php/actas/article/view/913/1191>

Andreo-Martínez, P., García-Martínez, N., & Sánchez-Samper, E. P. (2017). La microbiota intestinal y su relación con las enfermedades mentales. *Revista de Investigación Biomédica y de la Salud*, 1(1), 1-20. <http://riberdis.cedid.es/handle/11181/5361>

Robles-Alonso, V., & Guarner, F. (2013). Progreso en el conocimiento de la microbiota intestinal humana. *Nutrición hospitalaria*, 28(3), 553-557. <https://dx.doi.org/10.3305/nh.2013.28.3.6601>

García-Mazcorro, J. F., Rodríguez-Castañeda, D., García-Peña, M., & Jiménez, M. (2020). Bacterial communities and their effect on the immune system in dogs and cats. *Veterinaria México OA*, 7(4), 1-16. <https://doi.org/10.22201/fmvz.24486760e.2020.4.819>

Valenzuela-González, F., Casillas-Hernández, R., Villalpando, E., Vargas-Albores, F., & Harris, C. (2015). El gen ARNr 16S en el estudio de comunidades microbianas marinas. *Ciencias Marinas*, 41(4), 297-313. <https://doi.org/10.7773/cm.v41i4.2492>

Sánchez-Salguero, E. S., & Santos-Argumedo, L. (2018). La asociación de la microbiota humana con la inmunoglobulina A y su participación en la respuesta inmunológica. *Revista Alergia México*, 65(3), 264-278. <https://doi.org/10.29262/ram.v65i3.519>

Bharti, R., & Grimm, D. G. (2021). Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics*, 22(1), 178–193. <https://doi.org/10.1093/bib/bbz155>

López De Heredia Larrea, U. (2016). Las técnicas de secuenciación masiva en el estudio de la diversidad biológica. *Munibe Ciencias Naturales*, 64, 7-31. <https://doi.org/10.21630/mcn.2016.64.07>

Xia, Y., Li, X., Wu, Z., Nie, C., Cheng, Z., Sun, Y., Liu, L., & Zhang, T. (2023). Strategies and tools in Illumina and Nanopore-integrated metagenomic analysis of microbiome data. *iMeta*, 2, e72. <https://doi.org/10.1002/imt2.72>

García-Espinosa, E., Rojas-Concepción, A. A., Vitón-Castillo, A. A., & Lepez, C. O. (2022). Principios FAIR de gestión de datos de investigación en ciencias de la salud. *Revista Información Científica*, 101(6), e3865. <http://scielo.sld.cu/pdf/ric/v101n6/1028-9933-ric-101-06-e3865.pdf>

Bellido Esteban, A., Rodríguez Learte, A. I., Ryan, P., & González Soltero, M. del R. (2020). Desarrollo de una herramienta para la autoevaluación de datos de investigación según los estándares FAIR en entorno universitario. En *Construyendo el futuro de la educación y la ciencia* (pp. 477-483). Dykinson.

Li, S., Zhuo, M., Huang, X., Huang, Y., Zhou, J., Xiong, D., Li, J., Liu, Y., Pan, Z., Li, H., Chen, J., Li, X., Xiang, Z., Wu, F., & Wu, K. (2020). Altered gut microbiota associated with symptom severity in schizophrenia. *PeerJ*, 8, e9574. <https://doi.org/10.7717/peerj.9574>

Marelli, L., Stevens, M., Sharon, T., Van Hoyweghen, I., Boeckhout, M., Colussi, I., Degelsegger-Márquez, A., El-Sayed, S., Hoeyer, K., van Kessel, R., Krekora Zajac, D., Matei, M., Roda, S., Prainsack, B., Schlünder, I., Shabani, M., & Southerington, T. (2023). The European health data space: Too big to succeed? *Health Policy*, 135, 104861. <https://doi.org/10.1016/j.healthpol.2023.104861>

Li, W., & Quinn, P. (2024). The European Health Data Space: An expanded right to data portability? *Computer Law & Security Review*, 52, 105913. <https://doi.org/10.1016/j.clsr.2023.105913>

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and biophysical research communications*, 469(4), 967–977. <https://doi.org/10.1016/j.bbrc.2015.12.083>

Pillinger, T., McCutcheon, R. A., Vano, L., Mizuno, Y., Arumuham, A., Hindley, G., ... & Howes, O. D. (2020). Comparative effects of 18 antipsychotics on metabolic function in patients with schizophrenia, predictors of metabolic dysregulation, and association with psychopathology: A systematic review and network meta-analysis. *The Lancet Psychiatry*, 7(1), 64–77. [https://doi.org/10.1016/S2215-0366\(19\)30416-X](https://doi.org/10.1016/S2215-0366(19)30416-X)

Lindenmayer, J.-P., Czobor, P., Volavka, J., Citrome, L., Sheitman, B., McEvoy, J. P., ... Lieberman, J. A. (2003). Changes in glucose and cholesterol levels in patients with schizophrenia treated with typical or atypical antipsychotics. *American Journal of Psychiatry*, 160(2), 290–296. <https://doi.org/10.1176/appi.ajp.160.2.290>

Shen, Y., Xu, J., Li, Z., Huang, Y., Yuan, Y., Wang, J., Zhang, M., Hu, S., & Liang, Y. (2018). Analysis of gut microbiota diversity and auxiliary diagnosis as a biomarker in patients with schizophrenia: A cross-sectional study. *Schizophrenia Research*, 197, 470–477. <https://doi.org/10.1016/j.schres.2018.01.002>

Pan, R., Zhang, X., Gao, J., Yi, W., Wei, Q., & Su, H. (2020). Analysis of the diversity of intestinal microbiome and its potential value as a biomarker in patients with schizophrenia: A cohort study. *Psychiatry Research*, 291, 113260. <https://doi.org/10.1016/j.psychres.2020.113260>

Nguyen, T. T., Kosciolk, T., Maldonado, Y., Daly, R. E., Sirkin Martin, A., McDonald, D., Knight, R., & Jeste, D. V. (2019). Differences in gut microbiome composition between persons with chronic schizophrenia and healthy comparison subjects. *Schizophrenia Research*, 204, 23–29. <https://doi.org/10.1016/j.schres.2018.09.014>

Thirion, F., Speyer, H., Haldor Hansen, T., Nielsen, T., Fan, Y., Le Chatelier, E., Fromentin, S., Berland, M., Plaza Oñate, F., Pons, N., Galleron, N., Levenez, F., Markó, L., Birkner, T., Jørgensen, T., Forslund, S. K., Vestergaard, H., Hansen, T., Nordentoft, M., Mors, O., Benros, M. E., Pedersen, O., & Ehrlich, S. D. (2023). Alteration of gut microbiome in patients with schizophrenia indicates links between bacterial tyrosine biosynthesis and cognitive dysfunction. *Biological Psychiatry Global Open Science*, 3(2), 283–291. <https://doi.org/10.1016/j.bpsgos.2022.01.009>

8. Declaración obligatoria del uso de herramientas de IA

En el presente trabajo, se ha hecho un uso responsable de herramientas de IA (ChatGPT) con la finalidad de aportar ideas, siempre bajo revisión.