



UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE CIENCIA E INGENIERÍA EN ALIMENTOS Y BIOTECNOLOGÍA

CARRERA DE BIOTECNOLOGÍA

Tema: Evaluación *in silico* de microbiomas intestinales para identificar biomarcadores específicos en enfermedades de Crohn y Celíaca usando técnicas de metagenómica comparativa.

Informe Final del Trabajo de Titulación, opción Proyecto de Investigación, previo la obtención del Título de Ingenieros Biotecnólogos, otorgado por la Universidad Técnica de Ambato, a través de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Autores: Alan Joshua López Herrera

Israel Joel Lucero Chisaguano

Tutor: Mg. Cristian Fernando Galarza Galarza

Ambato - Ecuador

Febrero - 2025

APROBACIÓN DEL TUTOR

Mg. Cristian Fernando Galarza Galarza

CERTIFICA:

Que el presente Informe Final del Trabajo de Titulación ha sido prolijamente revisado. Por lo tanto, autorizo la presentación de este Informe Final del Trabajo de Titulación, opción Proyecto de Investigación el mismo que responde a las normas establecidas en el reglamento de Títulos y Grados de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Ambato, 16 de Enero de 2025



Firmado electrónicamente por:
**CRISTIAN FERNANDO
GALARZA GALARZA**

Mg. Cristian Fernando Galarza Galarza

1803160272

TUTOR

AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Alan Joshua López Herrera y Israel Joel Lucero Chisaguano manifestamos que los resultados obtenidos en el presente Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, previo a la obtención del título de Ingenieros Biotecnólogos, son absolutamente originales, auténticos y personales; a excepción de las citas bibliográficas.



Alan Joshua López Herrera

1850254770

AUTOR



Israel Joel Lucero Chisaguano

1850626068

AUTOR

DERECHOS DE AUTOR

Autorizamos a la Universidad Técnica de Ambato, para que haga uso de este Informe Final del Trabajo de Titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedemos los Derechos en línea patrimoniales de nuestro Informe Final del Trabajo de Titulación, con fines de difusión pública, además aprobamos la reproducción de este, dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando nuestros derechos de autor.



Alan Joshua López Herrera

1850254770

AUTOR



Israel Joel Lucero Chisaguano

1850626068

AUTOR

APROBACIÓN DEL TRIBUNAL DE GRADO

Los suscritos Profesores Calificadores, aprueban el presente Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, el mismo que ha sido elaborado de conformidad con las disposiciones emitidas por la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología de la Universidad Técnica de Ambato.

Para constancia firman:



firmado electrónicamente por:
**DOLORES DEL ROCIO
ROBALINO MARTINEZ**

Presidente de Tribunal



firmado electrónicamente por:
**JUAN DE DIOS
ESPINOZA MOYA**

Mg. Juan de Dios Espinoza Moya

1803201431



firmado electrónicamente por:
**SANTIAGO ESMIRO
CADENA CARRERA**

Dr. Santiago Esmiro Cadena Carrera

1715602593

Ambato, 30 de enero de 2025

DEDICATORIA

“Dedico este trabajo a los pilares fundamentales de mi vida:

A mi madre y padre, cuya luz me ha guiado en cada uno de mis pasos en este camino.

*A mis hermanos Jessica y Michael, cuya presencia y palabras de aliento me han dado fuerzas para
seguir adelante.*

Sin su apoyo incondicional, sus sacrificios y su fe en mí, este sueño nunca habría tomado forma.

Este logro es tan suyo como mío; nunca lo habría conseguido sin ustedes”.

-Alan Joshua Lopez Herrera

DEDICATORIA

“A mis padres y familia, por ser mi mayor ejemplo de perseverancia, amor y esfuerzo, y por brindarme su apoyo incondicional en cada paso de este camino.

A mis profesores y mentores, cuya guía y conocimiento me han inspirado a alcanzar este logro.

A mis amigos y compañeros, por compartir conmigo los desafíos y alegrías que este proyecto ha traído.

Y, sobre todo, al Creador, por darme fortaleza y sabiduría en los momentos más difíciles.

Este triunfo es mío, pero la victoria es suya. Con profundo agradecimiento y cariño”.

- Israel Joel Lucero Chisaguano

AGRADECIMIENTO

En primer lugar, agradezco a Dios por dame salud y vida, permitiéndome alcanzar uno de mis sueños y guiándome a conocer a personas extraordinarias en el camino.

A mis padres, Luis y María, por estar siempre a mi lado y guiarme con amor en cada paso que he dado. Cada uno de sus consejos han sido fundamental convertirme en quien soy hoy. Su esfuerzo y dedicación, al brindarme la oportunidad de estudiar y salir adelante, son el motor que me impulsa a superar cada día más. A mis hermanos, Jessica y Michael gracias por su constante apoyo y por estar siempre presentes en cada etapa de mi vida.

A Joel, mi compañero de tesis, gracias por tu compañía y paciencia en esta emocionante etapa. Valoro profundamente tu amistad y consejos. Compartir este trabajo contigo fue una experiencia inolvidable y espero que la vida nos permita reencontrarnos en el futuro.

A mis amigos ya graduados, Wendy, Ketsi, Gonza, Cayo, Carito, Licha, Kevin y Erick gracias por los momentos compartidos y por su apoyo incondicional durante los semestres que pudimos disfrutar juntos. Su amistad ha sido un pilar importante en mi vida universitaria.

Al grupo que me acompañó durante este proceso, Cris, Monse, John, Diana, Juss y Gaby, gracias por los momentos divertidos y por su apoyo en los momentos difíciles. Ustedes hicieron que el último semestre fuera único. Llevaré siempre en mi corazón los recuerdos vividos juntos.

A mi tutor Cristian, agradezco profundamente su guía y apoyo incondicional durante esta gran travesía. Su compromiso y paciencia fueron clave para llevar este proyecto adelante. Espero que la vida nos permita trabajar juntos nuevamente como colegas.

Con cariño Joshua

AGRADECIMIENTO

Quiero comenzar expresando mi más profunda gratitud al Creador, quien me ha guiado y fortalecido en cada paso de este camino. Su presencia ha sido mi refugio y mi fuente de inspiración constante. Reconozco que mi vida y todo lo que haga nunca podrán compensar todo lo que Él ha hecho por mí.

Tehilah (Salmos) 116:12

מִהָאֵשֶׁב לַיהוָה כָּל־תַּגְמוּלוֹתַי עָלַי:

¿Cómo puedo pagarle al Creador por tanta bondad que me ha mostrado?

A mis padres, Ramiro y Marisol, por su ejemplo de dedicación y fe. A mis hermanos, Abigail, Jair, Ariel y Keyla, por su compañía en los momentos difíciles. A mi abuelita, María “Milita”, cuyo cariño siempre ha estado para mí. A toda mi familia, por estar siempre presentes y ser pilares en mi vida.

A mi tutor de tesis, Cristian Galarza, por su paciencia, guía y valiosos aportes, que fueron cruciales para la culminación de este trabajo. También quiero agradecer a los docentes que fueron fuente de inspiración Mirari Arancibia, Alex Valencia, Liliana Lalaleo, Alejandra Sánchez, Danae Fernández, Mario García, Liliana Cerda, Lorena Núñez e Irving Tubón.

A mi compañero de tesis Joshua, por haber compartido esta experiencia única y por haber formado una amistad valiosa, espero que la vida nos permita seguir trabajando juntos en otros proyectos.

A los amigos que hice en todo este camino. A Fabricio, Alex y Luis por haber compartido momentos desde el inicio de la carrera. A Mery, Kathy, Ale, Erick, por su compañía y todas las risas que pasamos. A Monse, Chris, Andi, Lili, Jae y Pame, amigos de clase con los que compartí momentos y experiencias únicas. A John, Diana, Gaby y Justin amigos que con sus locuras alegraron mis días. Tal vez me falten amigos por mencionar, pero a todos los llevo en mi corazón y espero que nuestra amistad perdure.

Al alma mater, la Universidad Técnica de Ambato y a la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología, por abrirme sus puertas y en cuyas aulas, pasillos y patios pude formarme académica y personalmente.

A Sociedad Estudiantil de Renovación (SER), cuya pasión y compromiso dejaron una huella imborrable en mi vida. Su incansable dedicación con el servicio desinteresado y la búsqueda de un cambio positivo en la universidad me enseñaron el verdadero significado del liderazgo.

*Culmino esta etapa con un corazón lleno de gratitud hacia todos los que formaron parte de este proceso, pero sobre todo hacia Dios, quien me recuerda que "Todo lo puedo en Mashiaj que me fortalece" (**Filipenses 4:13**).*

ÍNDICE GENERAL DE CONTENIDOS

APROBACIÓN DEL TUTOR.....	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN	iii
DERECHOS DE AUTOR	iv
APROBACIÓN DEL TRIBUNAL DE GRADO.....	v
DEDICATORIA.....	vi
AGRADECIMIENTO	viii
ÍNDICE DE TABLAS	xii
ÍNDICE DE FIGURAS	xiv
RESUMEN EJECUTIVO.....	xv
ABSTRACT	xvi
CAPÍTULO I.- MARCO TEÓRICO	1
1.1. Antecedentes Investigativos	1
1.1.1. Introducción a las Enfermedades Inflamatorias Intestinales (EII) y la Enfermedad de Crohn.	1
1.1.2. Epidemiología de las Enfermedades Inflamatorias Intestinales (EII).	5
1.1.3. Microbiota intestinal y su papel en las EII.	6
1.1.4. Bases de Datos Genómicas y Herramientas Bioinformáticas.	7
1.1.5. Metagenómica.	10
1.1.6. Biomarcadores Microbianos en las EII.....	12
1.2. Objetivos.....	16
1.2.1 Objetivo General	16
1.2.2 Objetivos Específicos	16
CAPÍTULO II.- METODOLOGÍA.....	17
2.1 Materiales.....	17
2.2 Métodos	18
2.2.1. Clasificación de los datos de las muestras de las enfermedades Celíaca y de Crohn.	18
2.2.1.1 Obtención de las secuencias (SRR) de la enfermedad de Crohn y Celíaca.	18
2.2.1.2 Descarga de las secuencias (SRA) de las enfermedades de Crohn y Celíaca.	20
2.2.1.4 Instalación de programas para el análisis metagenómico.....	22

2.2.1.5	Importación y visualización de las secuencias.....	22
2.2.1.6	Control de calidad de las secuencias.....	23
2.2.2	Ejecución del análisis metagenómico.....	24
2.2.2.1	Análisis de Abundancia (FeatureTable and FeatureData).....	24
2.2.2.2	Análisis de diversidad alfa y beta.	24
2.2.2.3	Análisis taxonómico.....	24
2.2.2.4	Creación de los árboles filogenéticos.....	25
2.2.3	Evaluación y comparación de los perfiles metagenómicos.	25
2.2.3.1	Tabulación de los datos de acuerdo con el sexo y el tipo de enfermedad.	25
2.2.3.2	Comparativa de la diversidad alfa y beta de cada muestra.	25
2.2.3.3	Cuadros comparativos de abundancia entre los tipos de enfermedades.	26
2.2.3.4	Comparativa del análisis taxonómico.	26
2.2.3.5	Identificación de posibles biomarcadores.	26
CAPÍTULO III.- RESULTADOS Y DISCUSIÓN		27
3.1.	Análisis y discusión de los resultados.....	27
CAPÍTULO IV.- CONCLUSIONES Y RECOMENDACIONES		52
4.1.	Conclusiones.....	52
4.2.	Recomendaciones	53
ANEXOS.....		62

ÍNDICE DE TABLAS

Tabla 1. Programas bioinformáticos.	17
Tabla 2. Equipos.	17
Tabla 3. Información de los SRAs de cada muestra para enfermedad.	18
Tabla 4. Estándares de calidad de Illumina.	21
Tabla 5. Parámetros de corte en Trimmomatic.	21
Tabla 6. Formato del archivo manifest.tsv	23
Tabla 7. Resumen de los archivos demux.qzv de cada enfermedad.	30
Tabla 8. Parámetros de corte para el proceso de DADA2.	32
Tabla 9. Resumen estadístico de filtered-table.qzv de las enfermedades.	33
Tabla 10. Frecuencia por características de filtered-table.qzv de las enfermedades.	34
Tabla 11. Estadísticas de longitudes de secuencia de las enfermedades.	34
Tabla 12. Resumen de siete números de longitudes (nts) de secuencias de las enfermedades.	35
Tabla 13. Resumen estadístico de cluster-table.qzv de las enfermedades.	36
Tabla 14. Frecuencia por características de cluster-table.qzv de las enfermedades.	37
Tabla 15. Estadísticas de longitudes de secuencia de las enfermedades.	38
Tabla 16. Resumen de siete números de longitudes (nts) de secuencias de las enfermedades.	38
Tabla 17. Resumen de las tablas de las asignaciones taxonómicas.	41
Tabla 18. Abundancia relativa de los taxones (Filo) por sexo en la EC.	43
Tabla 19. Abundancia relativa de los taxones (Filo) por sexo en la enfermedad de Crohn.	44
Tabla 20. P_value de los índices de diversidad Alfa entre los sexos de cada muestra.	45

Tabla 21. P_value de los índices de diversidad Beta entre los sexos de cada muestra.....	46
Tabla 22. Frecuencia relativa de las funciones biológicas de los taxones en cada enfermedad	48
Tabla 23. Funciones significativas de ambas enfermedades.....	50
Tabla 24. Biomarcadores individuales de cada enfermedad.	51

ÍNDICE DE FIGURAS

Figura 1. Distribución de longitud de los reads en las muestras de: E. Celíaca, Control y Crohn.	27
Figura 2. FASTQC de las secuencias crudas de todas las muestras.	28
Figura 3. Distribución de la longitud de las secuencias de las muestras de Celíaca, Control y Crohn después del corte de Trimmomatic.....	29
Figura 4. Visualización de las secuencias tras el control de calidad en FASTQC.....	30
Figura 5. Histograma de frecuencia de las lecturas forward y reverse. (colocar en el gráfico FW y RV).....	31
Figura 6. Calidad de las lecturas forward y reverse del archivo demux.qzv.....	32
Figura 7. Gráfico PCOA Weighted Unifrac (PCOA).....	40
Figura 8. Abundancia relativa y árboles filogenéticos de las muestras Celíaca, Control y Crohn.	42
Figura 9. Indicador Shannon de diversidad Alfa.	45
Figura 10. Foresplot de Crohn vs Control y Celíaca vs Control.....	47
Figura 11. Rutas metabólicas más representativas de la EC y ECr.	49

RESUMEN EJECUTIVO

La comprensión de la microbiota intestinal es crucial para el avance en el diagnóstico y tratamiento de enfermedades inflamatorias como la enfermedad de Crohn (ECr) y la enfermedad Celíaca (EC). Este estudio emplea técnicas de metagenómica comparativa para identificar biomarcadores específicos que permitan diferenciar estas patologías. Los resultados resaltan la relevancia de la disbiosis microbiana como factor común, pero también destacan diferencias significativas en la composición y funciones de la microbiota intestinal en cada enfermedad, lo que permitirá avanzar en el conocimiento de estas patologías, y crear oportunidades para el desarrollo de herramientas diagnósticas y terapias innovadoras. El análisis se realizó mediante muestras del gen 16S rRNA, procesándolas con la herramienta QIIME2 bajo estrictos criterios de calidad de las muestras. Paralelamente se evaluó el impacto del sexo analizando las variaciones significativas en la diversidad microbiana beta en pacientes con ECr. El análisis funcional realizado con PICRUST reveló rutas metabólicas específicas relacionadas con la inflamación en la EC y el estrés oxidativo en la ECr. Las diferencias microbianas observadas mostraron un aumento de *Proteobacterias*, junto con reducciones en *Firmicutes* y *Actinobacteria*. Además, se identificaron biomarcadores en común como, *Faecalibacterium* y *Lachnospiraceae*, y otros específicos como *Alistipes* en ECr y *Lachnoanaerobaculum* en EC. Estos resultados validan la metagenómica como una herramienta para explorar la relación microbioma-enfermedad, facilitando el diagnóstico temprano y el diseño de terapias personalizadas dirigidas a restaurar el equilibrio microbiano, lo que marca un hito significativo en el desarrollo de la medicina de precisión para enfermedades inflamatorias intestinales.

Palabras claves: Crohn, Celíaca, QIIME2, microbiota intestinal, metagenómica, 16S rRNA.

ABSTRACT

Understanding the intestinal microbiota is crucial for advancing the diagnosis and treatment of inflammatory diseases such as Crohn's disease (CrD) and Celiac disease (CD). This study employs comparative metagenomics techniques to identify specific biomarkers to differentiate these pathologies. The results highlight the relevance of microbial dysbiosis as a common factor, but also highlight significant differences in the composition and functions of the intestinal microbiota in each disease, which will advance the understanding of these pathologies, and create opportunities for the development of diagnostic tools and innovative therapies. The analysis was performed using 16S rRNA gene samples, processed with the QIIME2 tool under strict sample quality criteria. In parallel, the impact of sex was evaluated by analyzing significant variations in beta microbial diversity in patients with CrD. Functional analysis performed with PICRUSt revealed specific metabolic pathways related to inflammation in CD and oxidative stress in CrD. Microbial differences observed showed an increase in Proteobacteria, along with reductions in Firmicutes and Actinobacteria. In addition, biomarkers were identified in common such as, *Faecalibacterium* and *Lachnospiraceae*, and specific ones such as *Alistipes* in CrD and *Lachnoanaerobaculum* in CD. These results validate metagenomics as a tool to explore the microbiome-disease relationship, facilitating early diagnosis and the design of personalized therapies aimed at restoring microbial balance, marking a significant milestone in the development of precision medicine for inflammatory bowel diseases.

Keywords: Crohn, Celiac, QIIME2, gut microbiota, metagenomics, 16S rRNA.

CAPÍTULO I.- MARCO TEÓRICO

1.1. Antecedentes Investigativos

1.1.1. Introducción a las Enfermedades Inflamatorias Intestinales (EII) y la Enfermedad de Crohn.

1.1.1.1. Definición de EII.

Las Enfermedades Inflamatorias Intestinales (EII) son trastornos crónicos del tracto gastrointestinal causados por un desequilibrio entre el sistema inmune, la microbiota intestinal y los factores ambientales **(Silva et al., 2019)**. Las EII presentan una inflamación característica que puede extenderse a otras áreas del cuerpo, afectando órganos como la piel, los ojos y el hígado **(Seyedian et al., 2019)**. Además, se ha observado que este proceso inflamatorio está mediado por una compleja interacción entre factores genéticos, ambientales y el microbioma intestinal, lo que contribuye a la perpetuación de la inflamación crónica **(Bruner et al., 2023)**. Entre los factores ambientales que contribuyen al desarrollo de las EII se puede citar al tabaquismo, dieta rica en alimentos ultraprocesados y el uso temprano de antibióticos, todos estos afectan negativamente el equilibrio del microbioma intestinal y predisponen el desarrollo de estas enfermedades **(Singh & Bernstein, 2022)**.

El desarrollo de las EII presenta una respuesta inmune desregulada frente a la disbiosis intestinal, esta se ve caracterizada por alteraciones en la composición de la microbiota, como la disminución de bacterias beneficiosas como *Firmicutes* y *Ruminococcaceae*, además, de un incremento de bacterias patógenas como *Escherichia/Shigella* **(Bruner et al., 2023; Russo et al., 2021)**. Por otra parte, ciertas condiciones externas, como el uso de medicamentos inmunomoduladores, también pueden desencadenar EII secundarias, lo que resalta la importancia de los factores externos en la etiología de estas enfermedades **(Ghouri et al., 2022)**. Las principales EII son la enfermedad de Crohn (ECr) y la colitis ulcerosa (CU) **(Russo et al., 2021)**, también existen otras enfermedades inmunomediadas como la enfermedad Celíaca (EC) **(Ricci et al., 2022)**.

1.1.1.2. Enfermedad de Crohn

La ECr es un trastorno inflamatorio crónico que afecta principalmente al tracto gastrointestinal, específicamente al íleon y colon **(Zhang et al., 2024)**. Su patogénesis involucra una compleja interacción de factores genéticos, ambientales y microbianos. El sistema inmune desempeña un papel crucial en el desarrollo de la enfermedad (patogénesis), ya que una respuesta anormal contribuye al daño progresivo del tracto gastrointestinal. Entre los factores de riesgo más importantes se encuentran la predisposición genética, la dieta, el tabaquismo, y la disbiosis microbiana. **(Gordon et al., 2024; Zhang et al., 2024)**.

La ECr está influenciada por factores genéticos, como las mutaciones en el gen NOD2, además de factores ambientales, lo que aumenta el riesgo de desarrollar la enfermedad **(Seyedian et al., 2019; Jarmakiewicz et al., 2022; Singh & Bernstein, 2022)**. Según **Gordon et al. (2024)**, los síntomas de esta enfermedad varían dependiendo de su localización y extensión, siendo los más comunes el dolor abdominal, diarrea crónica, pérdida de peso y fatiga. La inflamación también va acompañada de una disbiosis significativa del microbioma intestinal, caracterizada por un aumento de bacterias patógenas y una disminución de bacterias beneficiosas. Este desequilibrio favorece la persistencia de la inflamación crónica y el recrudecimiento de los síntomas **(Russo et al., 2021; Singh & Bernstein, 2022)**.

1.1.1.3. Enfermedad Celíaca

La EC es una enteropatía autoinmune desencadenada por la exposición al gluten, una proteína presente en cereales como el trigo, la cebada y el centeno **(Bruner et al., 2023)**. Esta condición afecta principalmente la mucosa del intestino delgado, causando daño a los enterocitos y atrofia de las vellosidades intestinales **(Seyedian et al., 2019)**. La patogénesis de la EC implica una interacción compleja entre factores genéticos y ambientales. Un componente genético importante es la presencia de los alelos HLA-DQ2 y HLA-DQ8, que están presentes en la mayoría de los pacientes con EC y sus familias **(Aljada et al., 2021)**. Los factores ambientales que más influyen representan son la introducción temprana del gluten, infecciones gastrointestinales y cambios en la microbiota intestinal **(Silvester et al., 2021; Singh & Bernstein, 2022)**.

Los síntomas principales de la EC incluyen diarrea, esteatorrea, dolor y distensión abdominal, pérdida de peso, así como deficiencias nutricionales derivadas de la mala absorción de electrolitos, vitaminas y minerales **(Silvester et al., 2021)**. La EC presenta manifestaciones extraintestinales que pueden involucrar síntomas neurológicos como dolores de cabeza y migrañas, afecciones en otros sistemas, como el endocrino, renal, hepático y pulmonar, entre otros **(Silvester et al., 2021; Therrien et al., 2020)**. Estas manifestaciones extraintestinales a menudo complican el diagnóstico de la enfermedad, ya que pueden confundirse con otras patologías, esto puede retrasar el diagnóstico temprano de la enfermedad **(Therrien et al., 2020)**.

1.1.1.4. Relación entre enfermedad Celíaca y Crohn

La relación entre la EC y la ECr se manifiesta a través de varios factores comunes, a pesar de ser patologías distintas. Ambas enfermedades comparten una predisposición genética que contribuye al desarrollo de respuestas inflamatorias anormales en el intestino, a pesar de que aún no se ha establecido una relación causal directa, algunos estudios experimentales han mostrado que el gluten puede desencadenar inflamación intestinal y aumentar la permeabilidad de la barrera epitelial, lo cual se observa también en la ECr **(Caio et al., 2021)**. La disbiosis intestinal es otro factor relevante, ya que se ha observado que, tanto en la EC como en la ECr la disminución de bacterias beneficiosas y el aumento de microorganismos patógenos afectan la función de la barrera intestinal y contribuyen a la inflamación crónica **(Skoufou et al., 2024; Roskowska et al., 2024)**.

Una característica común entre ambas enfermedades es la presencia de sobrecrecimiento bacteriano del intestino delgado (SIBO), que puede desempeñar un papel importante en la patogénesis de la inflamación intestinal y en el empeoramiento de los síntomas. Los estudios recientes sugieren que el SIBO es más prevalente en pacientes con enfermedades gastrointestinales como la EC y la ECr, contribuyendo a la disfunción de la barrera intestinal y la producción de moléculas proinflamatorias **(Roskowska et al., 2024)**. La EC puede presentarse como una causa subyacente en pacientes con enfermedad inflamatoria intestinal (EII) que no responde a tratamientos convencionales, lo que indica una posible conexión entre ambas enfermedades en términos de disfunción inmunitaria y vulnerabilidad a infecciones oportunistas **(Wieser et al., 2024)**.

1.1.1.5. Tratamientos Actuales y complicaciones

Para el tratamiento óptimo de la ECr, es esencial considerar la estratificación del riesgo del paciente, preferencias personales y factores clínicos, como la edad y extensión de las lesiones intestinales **(Cushing & Higgins, 2021; Petagna et al., 2020)**. Sin embargo, la evidencia que respalda este enfoque es limitada **(Torres et al., 2020)**. Los tratamientos comunes incluyen antibióticos, corticoides e inmunosupresores como el ácido 5-aminosalicílico (5-ASA), la sulfasalazina, la azatioprina, la 6-mercaptopurina y el metotrexato **(Petagna et al., 2020; Cushing & Higgins, 2021)**. Por otra parte, los anticuerpos monoclonales (infliximab, adalimumab, certolizumab) son eficaces, pero aumentan el riesgo de infecciones y cáncer **(Bruner et al., 2023)**.

Las cirugías como la resección intestinal son opciones frecuentes con beneficios en la recuperación, pero afectan la microbiota intestinal, aumentando la tendencia a las recaídas **(Petagna et al., 2020; Bruner et al., 2023)**. La desnutrición es frecuente en la ECr, por lo que el apoyo nutricional es importante. Se utiliza la nutrición enteral (NE), nutrición parenteral (NP), dietas FODMAP y semivegetariana para reducir inflamación, aunque no todos los pacientes mejoran y los beneficios se pierden al reintroducir alimentos sólidos **(Caio et al., 2021; Seyedian et al., 2019)**. A pesar de que todos los tratamientos antes mencionados mejoran los síntomas a corto plazo, ninguno de ellos ha demostrado ser curativo o completamente seguro **(Petagna et al., 2020; Seyedian et al., 2019)**.

El tratamiento más común para la EC es una dieta estricta sin gluten, cuyo objetivo es eliminar los antígenos que desencadenan la inflamación **(Machado, 2023)**, aunque es efectiva para mejorar la calidad de vida y prevenir la progresión de la enfermedad, presenta desventajas significativas. Mantener la dieta resulta difícil debido a la presencia ubicua del gluten en alimentos y productos inesperados, lo cual afecta la calidad de vida de los pacientes y genera miedo a la contaminación accidental. Alrededor del 40% de los pacientes con EC no están satisfechos con la dieta y buscan alternativas **(Caio et al., 2019)**. Además, la recuperación histológica del intestino es lenta, especialmente en adultos, llegando inclusive a tardar entre 2 a 5 años **(Aljada et al., 2021)**.

Frente a estas limitantes, se han venido investigando tratamientos farmacológicos que complementen la dieta sin gluten. Entre estos se encuentran el acetato de larazótido, que reduce la permeabilidad intestinal, las proteasas específicas del gluten y degradan los péptidos inmunogénicos del gluten **(Caio et al., 2019; Wei et al., 2020)**. Otras estrategias incluyen el bloqueo de la exposición inmunogénica del

gluten, la inhibición de la transglutaminasa (TG-2) y la inducción de la tolerancia inmunitaria mediante desensibilización o vacunación **(Machado, 2023)**. Aunque estas terapias han mostrado potencial, ninguna ha demostrado ser completamente efectiva o segura para reemplazar la dieta sin gluten, por lo que, su desarrollo sigue en fases de investigación **(Caio et al., 2019)**.

1.1.2. Epidemiología de las Enfermedades Inflammatorias Intestinales (EII).

En un inicio se consideró que las EII eran más prevalentes en la población occidental, especialmente en América del Norte y Europa Occidental, debido a factores relacionados con el estilo de vida y la industrialización **(Mak et al., 2020)**. Sin embargo, la creciente industrialización y los cambios socioeconómicos han incrementado significativamente la incidencia de estas enfermedades en países en desarrollo y muchos países orientales **(Feuerstein & Cheifetz, 2017)**. Este aumento puede atribuirse a factores como la urbanización, la adopción de dietas occidentales ricas en alimentos ultraprocesados y bajas en fibra y las alteraciones del microbioma intestinal debido al uso temprano de antibióticos **(Singh & Bernstein, 2022)**.

La epidemiología de las EII varía dependiendo de la región. La ECr tiene una incidencia de 3 a 20 casos por cada 100.000 habitantes, siendo más prevalente en países industrializados como América del Norte y Europa Occidental, sin embargo, en las últimas décadas, se ha observado un aumento en la incidencia de esta en regiones de América del Sur y Asia, probablemente debido a la creciente industrialización y cambios en el estilo de vida **(Feuerstein & Cheifetz, 2017; Mak et al., 2020)**.

En América Latina, la incidencia y prevalencia de las EII sigue siendo menor en comparación con las regiones más desarrolladas como América del Norte y Europa Occidental **(Selvaratnam et al., 2019)**. Informes recientes han señalado un incremento en la aparición de estas enfermedades en algunos países de esta región, lo cual podría estar relacionado con factores como la urbanización, el cambio en los patrones dietéticos y de estilo de vida **(Simian et al., 2016)**.

Por otro lado, la EC que anteriormente se consideraba limitada a ciertas poblaciones, actualmente se reconoce como una enfermedad global, con una prevalencia del 1,4% en todo el mundo y de 0,7% confirmada por biopsia **(Silvester et al., 2021)**. No obstante, esta prevalencia puede variar considerablemente entre continentes y países, se puede considerar una probable sobreestimación debido

a la falta de estudios de evaluación exhaustivos (endoscopías completas) **(Lebwohl & Rubio, 2021)**, a pesar de esto, se sabe que la prevalencia y la incidencia de la EC está en aumento.

1.1.3. Microbiota intestinal y su papel en las EII.

1.1.3.1. Concepto del microbioma y microbiota intestinal.

El microbioma humano es el conjunto de microorganismos (bacterias, virus, hongos y otros microbios) que habitan en y sobre el cuerpo humano, así como los genes que estos microorganismos contienen. Si bien todos los sitios del cuerpo están colonizados, el sitio con mayor cantidad de microbios es el intestino **(de Vos et al., 2022)**. Solo el tracto gastrointestinal humano puede albergar una microbiota considerablemente amplia y compleja, que incluye en su composición a arqueas, bacterias, virus y eucariotas **(Wang et al., 2015)**.

El microbioma intestinal desempeña un papel fundamental en la función inmunitaria del organismo y en distintas enfermedades, el poder manipular el microbioma intestinal podría tener un beneficio clínico en el manejo de estas enfermedades, siempre y cuando se demuestre la relación entre la composición del microbioma y la enfermedad **(Sanna et al., 2019)**. La microbiota intestinal es importante en la protección del huésped contra microbios patógenos, modulando la inmunidad y regulando los procesos metabólicos **(Wang et al., 2015)**. En algunos estudios se reporta el desequilibrio de los microorganismos intestinales en pacientes con EII, los cuales presentaban una menor diversidad microbiana y un crecimiento excesivo de anaerobios facultativos, a este desequilibrio en la microbiota se lo conoce como disbiosis **(Li et al., 2022)**.

1.1.3.2. Disbiosis y las EII.

La microbiota intestinal juega un papel fundamental en la regulación de la barrea intestinal y el sistema inmune, contribuyendo con la homeostasis del sistema. No obstante, esto puede presentar una alteración en la composición de la microbiota debido a diversos factores ambientales, genéticos y dietéticos, dando como resultado la disbiosis y contribuyendo así al progreso de las EII **(Silva et al., 2019; Singh & Bernstein, 2022)**.

Se ha evidenciado una disminución en el número de bacterias beneficiosas (*Firmicutes*), así como también un aumento en el número de bacterias patógenas (*Proteobacteria*) en pacientes con EII. Este desequilibrio conlleva la pérdida de funciones protectoras y es la responsable del apareamiento de procesos inflamatorios, debido a que las bacterias patógenas producen compuestos que perjudican la mucosa intestinal (**Russo et al., 2021**).

1.1.4. Bases de Datos Genómicas y Herramientas Bioinformáticas.

1.1.4.1. Fuentes de datos.

Las bases de datos biológicas juegan un papel fundamental en la bioinformática, ofrecen la facilidad de acceder a diversos datos de importancia, entre los que se incluyen variedad de secuencias genómicas de distintos organismos (**Zhao et al., 2023**). Entre los consorcios y las bases más representativas se pueden encontrar, GenBank del NCBI, Banco de Datos de ADN de Japón (DDBJ) y Laboratorio Europeo de Biología Molecular (EMBL), los cuales recopilan y difunden un gran catálogo de secuencias de ADN y ARN. NCBI proporciona una amplia variedad de datos de ADN de diversas especies, incluyendo genomas de referencia, metagenomas del microbioma humano y perfiles de variación genéticos (**Rangwala et al., 2021**).

Por la naturaleza que tienen los metagenomas, estos permiten que se realice estudios metagenómicos en los cuales se emplean distintos tipos de datos; ADN genómico, ADN mitocondrial, ARN ribosomal, ARN mensajero y ARN transcriptómico. El ADN genómico otorga información de las muestras extraídas y proporciona un perfil de la diversidad genética de los microorganismos que se encuentran en la muestra, permitiendo revelar la presencia y características de diversas especies microbianas (**Bag et al., 2016**).

1.1.4.2. QIIME2

A pesar de la gran cantidad de herramientas bioinformáticas disponibles para el análisis de datos del microbioma, Quantitative Insights Into Microbial Ecology (QIIME) es una de las más populares (**Kioroglou et al., 2019**). QIIME es una plataforma de ciencia de datos biológicos de código abierto, diseñada originalmente para facilitar el análisis de amplicones del microbioma. No obstante, esta ha evolucionado para abarcar nuevos dominios, como el análisis de ensayos serológicos altamente

multiplexados, la genómica de patógenos, secuenciación de genes como el 16S rRNA , 18S rRNA, metagenómica shotgun del microbioma y metabolómica no dirigida (**Bolyen et al., 2019; Estaki et al., 2020**).

La versión más reciente QIIME2, es una plataforma completamente rediseñada que ha reemplazado a su predecesora (**Estaki et al., 2020**). QIIME2 ofrece una amplia gama de complementos y herramientas desarrolladas por la comunidad, que mejoran diferentes aspectos del análisis del microbioma, tales como controles de calidad, asignación taxonómica, inserción filogenética, herramientas de visualización y generación de informes mediante su extensión web QIIME2 View (<https://view.qiime2.org/>) (**Thompson et al., 2022; Estaki et al., 2020**).

QIIME2 puede instalarse en la mayoría de los sistemas operativos, incluyendo Mac OS, Linux y Windows (a través de una máquina virtual) (**Wang et al., 2015; Donovan et al., 2015**). Su instalación requiere al menos de 6 a 7 GB de espacio en disco y un mínimo de 4 GB de RAM para conjuntos de datos pequeños, aunque se recomienda contar con 8 GB de RAM para manejar datos más grandes (**Estaki et al., 2020**).

La complejidad de los procedimientos de laboratorio y los marcos bioinformáticos, con sus múltiples protocolos y algoritmos, pueden resultar abrumadores para los investigadores (**Kioroglou et al., 2019**). En este sentido, QIIME2 destaca por integrar un seguimiento detallado de la procedencia de los datos, un formato de archivo estandarizado, visualizaciones interactivas y una arquitectura de complementos extensibles, lo que contribuye a su popularidad en investigaciones del microbioma (**Thompson et al. , 2022**). Las herramientas de QIIME2 permiten rastrear de manera precisa cómo se generó un resultado y se realizó un análisis, lo que garantiza que los análisis bioinformáticos en QIIME2 sean reproducibles, replicables y robustos, sin esfuerzo adicional por parte de los usuarios (**Keefe et al., 2023**).

Según **Bolyen et al. (2019)**, los avances en la investigación del microbioma tienen el potencial de mejorar numerosos aspectos de la salud y el bienestar, QIIME2 contribuirá a impulsar dichos avances al permitir un análisis accesible y respaldado por una amplia comunidad de colaboradores.

1.1.4.3. Plataforma EzBioCloud.

EzBioCloud (<https://www.ezbiocloud.net/apps>) es un portal público de análisis de datos que se enfoca en taxonomía, ecología, genómica, metagenómica y microbioma de genes y genomas del gen 16S rRNA. Proporciona herramientas bioinformáticas complementarias que facilitan la clasificación e identificación de miembros de los dominios Bacteria y Archaea (Yoon et al., 2017). Su plataforma en la nube es utilizada en más de 50 países por más de 22.000 usuarios, incluidos académicos, organizaciones de investigación sin fines de lucro, comunidades médicas, agencias gubernamentales y empresas globales (ChunLab, Inc., 2024).

EzBioCloud se emplea principalmente para identificar características genómicas asociadas a diferentes grupos utilizando datos de 16S rRNA (Tango et al., 2020). Estudios como los de Sánchez et al. (2020a, 2020b) demuestran la validez de su uso en análisis bioinformáticos basados en datos de secuenciación del gen bacteriano 16S rRNA (región V3-V4) con el objetivo de identificar bacterias y analizar la composición de la microbiota en las muestras de estudio. Además, puede integrarse con herramientas como KEGG para caracterizar vías funcionales, como se realizó en el estudio de Tango et al. (2020).

1.1.4.4. Análisis funcional (PICRUSt2).

En cuanto al análisis funcional, la secuenciación del gen 16S rRNA y la secuenciación metagenómica shotgun (MGS) son métodos fundamentales que difieren significativamente. La secuenciación 16S rRNA se centra en un marcador genético específico, lo que facilita la identificación taxonómica de bacterias y arqueas, pero no proporciona perfiles funcionales de las comunidades microbianas, en contraste con la MGS, que secuencia todo el material genético presente en la muestra, permitiendo caracterizar la composición taxonómica y las funciones metabólicas de la comunidad, por lo que se corre el riesgo de generar contaminación en el huésped o de tener una baja biomasa microbiana en ciertas muestras (Douglas et al. , 2020). Para abordar esta limitación, se han desarrollado métodos que permiten predecir funciones a partir de los datos de secuenciación del 16S rRNA, para lo que se utilizan herramientas como PICRUSt , Tax4Fun2 y MicFunPred.

PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) es una de las herramientas más populares para la predicción de perfiles funcionales, facilita la generación de abundancia de vías integrales dentro de las comunidades microbianas, lo que proporciona información

clave sobre sus roles funcionales (**Yang et al., 2023**). Fue desarrollado en 2013, su versión más reciente, PICRUSt2, integra herramientas de código abierto y una base de datos de 41,926 genomas bacterianos y arqueales, es compatible de forma predeterminada con bases como KEGG (**Douglas et al., 2020**). Para la colocación filogenética en PICRUSt2 se utilizan tres herramientas principales: HMMER que coloca las secuencias variantes de amplicones (ASV), EPA-ng para determinar su posición óptima en una filogenia de referencia, y GAPPa para generar un nuevo árbol que incorpore dichas colocaciones (**Yang et al., 2023**).

PICRUSt2 cuenta con el paquete ggpicrust2 para R, el cual permite realizar análisis avanzados de diferenciación de abundancias y visualización los resultados, además permite convertir las tablas de abundancia de salida de PICRUSt2 en tablas de vías KEGG (**Yang et al., 2023**).

Estudios como los de **Hu et al. (2021)** y **Yu et al. (2021)** han demostrado la eficacia de **PICRUSt2** en el análisis de la microbiota humana, mediante la predicción de composiciones funcionales metagenómicas en múltiples patologías, como el síndrome coronario agudo y el carcinoma de tiroides.

1.1.5. Metagenómica.

1.1.5.1. Fundamentos de la metagenómica.

Los análisis metagenómicos permiten estudiar diversas comunidades microbianas no cultivables, como la microbiota intestinal, lo que permite entender la estructura y función de las comunidades microbianas y catalogar sus genes a partir de la secuenciación de una muestra de ADN (**Wang et al., 2015**). Como ya se mencionó, la disbiosis microbiana es característica de la ECr y EC, por lo tanto, la metagenómica puede ser utilizada para estudiar este desequilibrio en la diversidad del microbioma intestinal, para relacionar estos hallazgos con la salud y la enfermedad del paciente (**Tiffany & Bäumler, 2019**).

Para estudiar la diversidad microbiana se utilizan dos metodologías: secuenciación shotgun y secuenciación de amplicones del gen 16S (**Brumfield et al., 2020**). La primera es útil para identificar una gran diversidad de especies microbianas debido a que procesa el genoma completo, mientras que la segunda es económica y solo secuencia una parte del genoma, llegando así a limitar el alcance genómico del método (**Beaudry et al., 2021**).

1.1.5.2 Secuenciación metagenómica de shotgun.

La secuenciación shotgun posee un enfoque dirigido a la secuenciación del ADN completo, otorgando una visión amplia de la composición microbiana y su potencial funcional. Este método permite secuenciar todo el material genético, dando como resultado una resolución taxonómica elevada **(Yoshitake et al., 2021)**. Además, este es un método ideal para el estudio de la diversidad microbiana ya que permite revelar una gran variedad de microorganismos **(Datta et al., 2020)**. Este método consta de diversos pasos como la recolección de muestras, extracción de ADN, construcción de bibliotecas y el análisis de datos **(Lapidus & Korobeynikov, 2021)**.

La metagenómica utiliza muestras provenientes de la secuenciación shotgun, posee la capacidad de detectar varios tipos de microorganismo bacterias, eucariotas, virus y arqueas, proporcionando así una visión completa del microbioma de la muestra, además, permite el análisis funcional facilitando la identificación de rutas metabólicas y funcionales presentes en la comunidad microbiana **(Christensen & Olsen, 2018; Taş et al., 2021)**. Sin embargo, esta técnica presenta distintas limitaciones, como su elevado costo, especialmente cuando se necesita una elevada profundidad al detectar microorganismos en bajas concentraciones y, además, presenta problemas de contaminación por el ADN del huésped dificultando así la detección precisa de microorganismos **(Odom et al., 2023; Zhang et al., 2020)**.

1.1.5.3 Secuenciación del amplicón 16S.

La secuenciación del amplicón 16S rRNA, o metataxonómica, es útil para caracterizar la diversidad microbiana y predecir su función metabólica **(Pérez et al., 2022)**. Esta secuenciación se basa fundamentalmente en amplificar y secuenciar una región específica del gen que codifica la subunidad ribosomal 16S, la que se halla presente en casi todas las arqueas y bacterias. Este gen contiene regiones conservadas e hipervariables, que permiten hallar diferencias entre diversos taxones microbianos **(Gupta et al., 2019)**.

El gen 16S tiene en total 9 regiones hipervariables, de las cuales la V3-V4 son las más utilizadas debido a su gran longitud y su elevada heterogeneidad de nucleótidos, otorgando así un elevado poder discriminatorio para diferenciar especies **(Bukin et al., 2019; Wei et al., 2021)**. La secuenciación del amplicón 16S se considera útil para caracterizar taxones debido a su bajo costo, así como también a la

facilidad de uso de bases taxonómicas como SILVA, que permiten asignar las secuencias a sus taxones correspondientes (**Christensen et al., 2018**). Este método puede ser utilizado junto con herramientas bioinformáticas como el PICRUST2 para predecir las funciones metabólicas de la comunidad microbiana (**Douglas et al., 2019**).

1.1.6. Biomarcadores Microbianos en las EII.

1.1.6.1. Definición de biomarcadores.

Los biomarcadores son indicadores medibles que proporcionan información objetiva sobre un proceso biológico, fisiológico o patológico, permitiendo evaluar de manera precisa el estado de dicho proceso en un momento específico (**Russo, 2022**). Estos actúan como indicadores clave en el diagnóstico y tratamiento de diversas enfermedades, ayudando a evaluar la actividad de la enfermedad y a predecir recaídas, especialmente en el caso de las EII facilitando la diferenciación entre distintas patologías (**Stoppe et al., 2020**).

Según **Russo (2022)**, existen varios tipos de biomarcadores, como los de susceptibilidad/riesgo, diagnóstico, monitoreo, pronóstico, predictivos, farmacodinámicos de respuesta, y de seguridad. Entre ellos, los biomarcadores de diagnóstico se utilizan para identificar a individuos con una enfermedad o condición de interés, o para definir un subconjunto específico de dicha enfermedad. Estos biomarcadores potenciales pueden ser fundamentales para el diagnóstico y pronóstico temprano de enfermedades, así como para el desarrollo de nuevas herramientas terapéuticas en distintos trastornos (**Agus et al., 2021**).

En estudios recientes como los realizados por **Duttagupta et al. (2023)** y **Aljarra et al. (2024)**, se ha demostrado como la microbiota intestinal puede utilizarse como biomarcador en distintas patologías, incluyendo el cáncer de pulmón y los gliomas.

1.1.6.2. Biomarcadores específicos para enfermedad de Crohn y Celíaca.

Liu et al. (2022) mencionan que existen tres tipos principales de biomarcadores de las EII, en la ECR: serológicos, fecales y urinarios. Entre los biomarcadores más destacados se encuentran la proteína C

reactiva (CRP), la velocidad de sedimentación globular (VSG), los anticuerpos anti-*Saccharomyces cerevisiae* (ASCA), los anticuerpos perinucleares anti-neutrófilos (p-ANCA), el óxido nítrico (NO), varias citocinas (como TNF- α), la calprotectina, la calgranulina C, la lactoferrina, la lipocalina-2/NGAL, la mieloperoxidasa (MP) y microARNs como el miR-21 y miR-155 que han sido identificados en la progresión de las EII.

En las últimas décadas, la calprotectina fecal ha recibido considerable atención como biomarcador no invasivo para el diagnóstico y tratamiento de la ECr. Numerosos estudios han evaluado la eficacia de la calprotectina fecal en la diferenciación entre la EII y el síndrome del intestino irritable (SII), si bien algunos de estos estudios han mostrado resultados prometedores, otros han encontrado limitaciones significativas (**Khalki et al., 2020**). A pesar de su uso global, presenta limitaciones debido a su especificidad relativamente baja debido a que este biomarcador se correlaciona con el número de neutrófilos presentes en el lumen intestinal, que no solo está relacionada con la ECr sino con otras infecciones de la mucosa, traumatismos o estrés ambiental en el intestino, lo que limita la diferenciación entre diferentes etiologías. Otra desventaja se da en ciertas situaciones de inflamación, la calprotectina fecal contribuye a la lesión y la inflamación de la mucosa en otros tejidos, como la piel y los pulmones, lo cual también se observa en el intestino (**Jukic et al., 2021**).

Otras limitaciones de la calprotectina fecal incluyen la confiabilidad de las muestras almacenadas a lo largo del tiempo. Se ha demostrado que, después de seis días a temperatura ambiente, la concentración de calprotectina fecal puede disminuir hasta un 35 %, lo que afecta la precisión de los resultados. Además, el uso de ciertos fármacos, como los antiinflamatorios no esteroideos (AINE), puede aumentar los niveles de calprotectina fecal, disminuyendo la confiabilidad de los resultados (**Khalki et al., 2020**).

Aunque estos biomarcadores se emplean comúnmente para el diagnóstico de la ECr, presentan varias limitaciones. **Liu et al. (2022)** explica que algunas de estas limitaciones son, la elevación de ciertos biomarcadores en presencia de otras condiciones inflamatorias, la baja sensibilidad y especificidad, y la falta de efectividad cuando se usan de manera aislada. Además, muchos de estos biomarcadores están influenciados por factores externos, lo que reduce su especificidad y confiabilidad.

Con respecto a la EC, en la actualidad, existen diversos biomarcadores multimodales que podrían ayudar a diagnosticar, monitorear y mejorar la calidad de vida de los pacientes con esta enfermedad (**Giuffrè et al., 2022**). Según **Ramírez et al. (2020)**, estos biomarcadores pueden ser obtenidos de diferentes tipos de muestras, como heces, biopsias, sangre y orina.

Uno de los biomarcadores más comunes en la EC es la detección de anticuerpos específicos, como los anti-transglutaminasa (Anti-TGs), anti-TG2 y anti-TG6, que se obtienen de muestras sanguíneas (**Giuffrè et al., 2022**). Sin embargo, una de las principales limitaciones de estas pruebas es su posible confusión con otras patologías, como la ataxia por gluten. La precisión y sensibilidad de estas pruebas requieren ser optimizadas antes de su implementación generalizada (**Ramírez et al., 2020**).

Otro biomarcador importante es la identificación de los haplotipos HLA-DQ2 y HLA-DQ8, los cuales están asociados con una predisposición genética a la EC. Estos haplotipos se detectan principalmente a partir de muestras de biopsia duodenal (**Giuffrè et al., 2022**). Sin embargo, más del 99 % de los pacientes con EC son positivos para estas moléculas, lo que significa que la genotipificación de HLA no aporta un valor añadido significativo en el diagnóstico de la enfermedad (**Ramírez et al., 2020**).

Los microARNs (miRNAs) también han sido propuestos como biomarcadores para la EC. Según **Gnodi et al. (2022)**, los miRNAs como miR-449a, miR-124a, miR-189, miR-299-5p y miR-379, que regulan genes involucrados en la proliferación y diferenciación de células intestinales (como Notch1 y KLF4) se pueden detectar en el plasma. Aunque estos resultados son prometedores, la identificación de una firma específica probablemente requerirá la combinación de varios miRNAs para lograr una mayor sensibilidad y especificidad.

El estudio de **Wood et al. (2020)**, también destaca el potencial de la zonulina como biomarcador, dado que su papel está involucrado en el aumento de la permeabilidad intestinal, lo cual es relevante tanto para la EC como para la diabetes tipo 1. Sin embargo, aún se requieren investigaciones adicionales para validar la zonulina como biomarcador. Otros biomarcadores relevantes incluyen la proteína de unión a ácidos grasos intestinales (i-FABP), la citrulina, diversas citoquinas como IL-17 y TWEAK, la proteína HMGB1, y los péptidos inmunogénicos de gliadina. Cada uno de estos biomarcadores presenta limitaciones específicas, como falta de especificidad, influencia de factores externos, o dificultades para su aplicación en la práctica clínica rutinaria. Aunque estos biomarcadores tienen un gran potencial para mejorar el manejo de la EC, se requieren más investigaciones para validar su efectividad y estandarizar su uso, garantizando así un diagnóstico y monitoreo más preciso (**Rodríguez et al., 2020**).

Hasta el momento, el diagnóstico de la EC se basa principalmente en pruebas serológicas y biopsias invasivas. A pesar de los avances en el diagnóstico todavía existen razones importantes para seguir buscando nuevos biomarcadores que mejoren el diagnóstico y el seguimiento de la enfermedad (**Giuffrè et al., 2022**).

1.1.6.2. Potencial diagnóstico de los biomarcadores microbianos

Las investigaciones recientes han resaltado el potencial de los microbios intestinales como biomarcadores para distintas patologías complejas. Un ejemplo de ello es el estudio de **Zheng et al. (2023)**, que subraya la forma en que los microbios intestinales pueden actuar como biomarcadores efectivos para entender los mecanismos implicados en el desarrollo de enfermedades como el carcinoma hepatocelular.

En el caso específico de la Cr y la EC, los biomarcadores juegan un papel fundamental para guiar los enfoques terapéuticos y predecir la respuesta a tratamientos. Sin embargo, la identificación de biomarcadores específicos para estas enfermedades sigue siendo un desafío no resuelto (**Lee et al., 2021**). El análisis metagenómico se ha mostrado prometedor para enfrentar esta limitación, ya que permite identificar diferencias microbianas en el tracto intestinal que podrían actuar como biomarcadores para cada patología. Estos biomarcadores intestinales se presentan como herramientas valiosas tanto para el diagnóstico como para la comprensión de la patogénesis de las enfermedades, ofreciendo una mejor base para el desarrollo de tratamientos específicos y personalizados (**Hu et al., 2023**).

Además, los biomarcadores tienen el potencial de mejorar significativamente el diagnóstico y la predicción de respuestas terapéuticas en las EII. Sin embargo, la diferenciación eficaz entre estas enfermedades aún representa un desafío considerable (**Lee et al., 2021**). Los estudios metagenómicos, junto con técnicas avanzadas como las medidas de centralidad y la detección comunitaria en redes, proporcionan un enfoque poderoso para analizar datos complejos del microbioma intestinal (**Hu et al., 2023**). Estas técnicas permiten resaltar biomarcadores específicos que permitan mejorar el diagnóstico diferencial entre la ECr y la EC, optimizando la precisión y eficacia de los tratamientos.

1.2. Objetivos

1.2.1 Objetivo General

Evaluar *in silico* los microbiomas intestinales para identificar biomarcadores específicos en enfermedades de Crohn y Celíaca usando técnicas de metagenómica comparativa.

1.2.2 Objetivos Específicos

- Clasificar muestras del microbioma provenientes de experimentos relacionados con la enfermedad Celíaca y de Crohn almacenados en bases de datos biológicas públicas.
- Ejecutar un análisis metagenómico de las muestras obtenidas de las enfermedades Celíaca y Crohn.
- Evaluar perfiles metagenómicos de las enfermedades Celíaca y Crohn para la identificación de biomarcadores específicos en cada enfermedad.

CAPÍTULO II.- METODOLOGÍA

2.1 Materiales

El presente trabajo de titulación se realizó con un enfoque experimental *in silico*, se llevó a cabo en las instalaciones de la Facultad de Ciencias e Ingeniería en Alimentos y Biotecnología de la Universidad Técnica de Ambato. En la Tabla 1 se detallan los programas y recursos bioinformáticos utilizados en el desarrollo de la investigación y en la Tabla 2 se especifican los equipos utilizados en la investigación.

Tabla 1.

Programas bioinformáticos.

Recursos	Enlaces
NCBI	https://www.ncbi.nlm.nih.gov/
PubMed	https://pubmed.ncbi.nlm.nih.gov/
UBUNTU	https://ubuntu.com/
QIIME2 View	https://view.qiime2.org/
iTOL	https://itol.embl.de/
EZBioCloud	https://www.ezbiocloud.net/

Tabla 2.

Equipos

Recursos	Cantidad
Computadora portátil	2
Servicio de internet o Router	1

2.2 Métodos

2.2.1. Clasificación de los datos de las muestras de las enfermedades Celíaca y de Crohn.

2.2.1.1 Obtención de las secuencias (SRR) de la enfermedad de Crohn y Celíaca.

Para obtener las secuencias de las ECr y EC, se utilizó la base de datos GEO (Gen Expression Omnibus) administrada por el consorcio NCBI (National Center for Biotechnology Information). Se utilizaron los siguientes criterios de búsqueda: Inflammatory intestinal diseases, Celiac disease, Crohn disease, Metabarcoding, 16S rARN.

Para clasificar de mejor forma los datos se aplicaron filtros adicionales, entre ellos clasificación por sexo de los pacientes, procedencia (mucosa ileal y mucosa intestinal), tipo de datos (ADN genómico) y técnica de secuenciación (Illumina MiSeq). Una vez realizada esta búsqueda y filtrado se escogió las secuencias del bioproyecto SRP330345 (PMID: 36246269) para la EC y el bioproyecto SRP297049 (PMID: 35967326) para la ECr.

En la siguiente tabla se especifica el registro SRA completo de cada una de las muestras seleccionadas.

Tabla 3.

Información de los SRAs de cada muestra para enfermedad.

Enfermedad Celíaca			Enfermedad de Crohn		
SRA ID	Estado	Sexo	SRA ID	Estado	Sexo
SRR15291910	Celíaco	Femenino	SRR13216189	No inflamado	Masculino
SRR15291911		Masculino	SRR13216190	No inflamado	Masculino
SRR15291912	Celíaco	Femenino	SRR13216191	No inflamado	Femenino
SRR15291913	Celíaco	Femenino	SRR13216192	No inflamado	Masculino
SRR15291914	Celíaco	Masculino	SRR13216193	No inflamado	Masculino
SRR15291915	Posible celíaco	Masculino	SRR13216194	No inflamado	Femenino
SRR15291916		Femenino	SRR13216195	No inflamado	Masculino
SRR15291917		Femenino	SRR13216196	No inflamado	Femenino

Las muestras seleccionadas se dividieron en tres grupos:

- **Muestras de EC:** Se tomaron todas las muestras exceptuando los SRAs SRR15291915 al SRR15291924, debido a que estos son posibles casos de celiaquía, más no son casos comprobados.
- **Muestras Control:** En este grupo se incluyeron a los SRAs SRR13216189 al SRR13216198, ya que estos provienen de pacientes que no tenían ningún tipo de inflamación intestinal.
- **Muestras de Enfermedad de Crohn:** Se tomaron en cuenta a las muestras de SRR13216199 al SRR13216208, ya que estos pacientes tenían ECr diagnosticada.

2.2.1.2 Descarga de las secuencias (SRA) de las enfermedades de Crohn y Celíaca.

Se utilizó el sistema operativo linux Ubuntu 24.04, mediante línea de comando, se instaló el SRA Toolkit (**Anexo 1**). Este software facilitó el acceso a datos de secuenciación genómica almacenados en el SRA (Sequence Read Archive), los cuales forman parte de la International Nucleotide Sequence Database Collaboration (INSDC) y son un formato estándar definido por el NCBI para datos de secuenciación de nueva generación (NGS), desde aquí se gestionan los datos para su descarga, extracción y formateo (**Abouelkhair, 2020**).

Para descargar los SRAs se ejecutó la línea de códigos detallada en el **Anexo 2**.

2.2.1.3 Control de calidad de las secuencias (FASTQC)

Se instaló el programa FASTQC, se procedió a la visualización de la calidad de las lecturas crudas de cada una de las muestras mencionadas, para ello fue necesario correr el comando detallado en el **Anexo 3**. En este proceso se obtienen varios archivos, el más importante es el archivo .html que permite realizar un análisis gráfico de la calidad de las secuencias y un archivo .txt con la información resumida del análisis.

Los cortes en aquellas secuencias que presentan baja calidad se realizaron con el programa Trimmomatic (**Anexo 4**), para lo cual se tomó como parámetro un quality score de 30, de acuerdo con los estándares de calidad de **Illumina (2024), Tabla 4**.

Tabla 4.

Estándares de calidad de Illumina.

Quality Score	Probabilidad de una llamada de base incorrecta	Precisión de llamada de base inferida
10	1 de cada 10	90%
20	1 de cada 100	99%
30	1 de cada 1000	99,9%

Nota: Relación entre la puntuación de calidad de la secuenciación y la precisión de la llamada base.

Fuente: Illumina (2024).

Una vez analizada la calidad de calidad inicial de las secuencias se trabajó con los siguientes parámetros para el corte:

Tabla 5.

Parámetros de corte en Trimmomatic.

Muestras	Parámetros de corte
Enfermedad Celíaca	- Adapters TruSeq3-PE-2.fa - Slidwindow: 4:25 - Minlen: 245
Muestras Control Enfermedad de Crohn	- Adapters TruSeq3-PE-2.fa - Slidwindow: 4:30 - Minlen: 50

Nota: Se visualizan los parámetros utilizados en la herramienta para mejorar la calidad de las secuencias.

En las muestras de EC la mayoría de las secuencias tenían la longitud mayor a 245, por esta razón el corte se realizó en esta posición, en el caso de la ECr y el control la distribución de secuencias por longitud era menos uniforme, pero con menos cantidad de secuencias hasta la posición 50. Para los cortes de las muestras de control y ECr se realizó un archivo .bash mediante el cual se realizó un bucle para automatizar el corte, el código de este bash está descrito en el **Anexo 5**. Posteriormente se efectuó nuevamente la verificación de la calidad de las secuencias de cada muestra, el resultado de este procedimiento se encuentra un archivo descargable tipo .html en una tabla índice en el **Anexo 6**.

2.2.1.4 Instalación de programas para el análisis metagenómico.

Para llevar a cabo el análisis metagenómico se empleó la plataforma bioinformática QIIME 2™, la cual permite el análisis de secuencias de microbiomas mediante la técnica de próxima generación, es de acceso gratuito y código abierto (Bolyen et al., 2019). Actualmente, la plataforma QIIME 2 cuenta con tres distribuciones oficiales (Amplicon, Shotgun y Tiny), para este estudio se empleó la distribución QIIME2 Amplicon por la naturaleza de las muestras (16S rRNA).

Se instaló QIIME2 a través de Miniconda, que proporciona un entorno para la administración de paquetes de forma ligera (Bolyen et al., 2019). Miniconda se ejecutó sobre el sistema operativo linux Ubuntu LTS a través de líneas de comando, la versión que se instaló fue V2023.5 sobre un *environment* específico para ejecutar el análisis metagenómico (“qiime2-2023.5”).

2.2.1.5 Importación y visualización de las secuencias.

Se importaron los datos de las secuencias al entorno de trabajo de QIIME2, este paso es importante ya que permite transformar los distintos datos que se tienen a archivos tipo artefact QIIME (extensión .qza), que sirven para efectuar el análisis metagenómico. Este proceso se realizó mediante el plugin de importación de QIIME2 (Anexo 7) (Bolyen et al., 2019).

Para este paso es necesario contar con un archivo manifest.tsv que incluya el identificador de las secuencias y la dirección tanto del forward como el reverse, para ello es necesario que este archivo *manifest* cuente con una estructura predeterminada, ver en la **Tabla 6**. Además, debe existir un archivo *manifest* por cada muestra y cada comparación que se va a realizar, los archivos *manifest* completos se pueden observar en el **Anexo 8**.

Tabla 6.

Formato del archivo manifest.tsv

sample-id	forward-absolute-path	reverse-absolute-path
SRR.....	/home/../../..	/home/../../..
SRR.....	/home/../../..	/home/../../..
....		

Nota: Es necesario que todos los ID de las muestras así como la dirección de las secuencias (forward y reverse) sea la correcta, caso contrario el proceso no se ejecutará.

Mediante este proceso se han demultiplexado las secuencias, obteniendo el archivo paired-end-demux.qza, luego se realizó un resumen de las secuencias demultiplexadas y la transformación de formato .qza a .qzv (**Anexo 9**), para su visualización en QIIME 2 View (**Telatin, 2021**).

2.2.1.6 Control de calidad de las secuencias.

Una vez obtenido los archivos demultiplexados, se realizó el control de calidad de las secuencias, utilizando DADA2 (Divide Amplicon Denoising Algorithm 2). Esta herramienta bioinformática implementada en lenguaje de programación R e integrada a través de la interfaz de QIIME2, es ampliamente utilizada para detección y corrección de errores en secuencias de amplicones de ADN (**Callahan et al., 2016**). A partir de esta, se generaron archivos esenciales para las etapas posteriores del análisis.

- **Table.qza:** Contiene información sobre la cantidad de veces que cada característica aparece en cada una de las muestras.
- **Rep-seqs.qza:** Incluye las secuencias representativas de cada característica identificada, estas representan variantes de secuencias únicas (ASVs) obtenidas tras el procesamiento de DADA2.
- **Denoising-stats.qza:** Proporciona detalles sobre el número de secuencias iniciales, las que pasan el filtro, aquellas corregidas y las retenidas tras el proceso.

(**Bolyen et al., 2019**).

Posteriormente, se llevó a cabo un reprocesamiento del filtrado para eliminar secuencias duplicadas y minimizar errores, asegurando la representatividad y la calidad en el análisis taxonómico. Todos estos procedimientos se describen en el Pipeline del **Anexo 10**.

2.2.2 Ejecución del análisis metagenómico.

2.2.2.1 Análisis de Abundancia (FeatureTable and FeatureData).

El análisis de abundancia se realizó utilizando la técnica de clustering, se utilizó los archivos obtenidos tras el filtrado y eliminación de duplicados (filtered-table.qza y filtered-rep-seqs.qza) empleando el comando detallado en el **Anexo 11**. Este procedimiento generó los archivos clustered_table.qza (FeatureTable) y clustered_sequences.qza (FeatureData). Posteriormente, se transformaron estos archivos del formato .qza a .qzv mediante el plugin especificado en el **Anexo 12**. Con esto se obtuvieron gráficas de frecuencias de las secuencias representativas y una tabla que asocia dichas secuencias con un identificador único (ID) (**Bolyen et al., 2019**).

2.2.2.2 Análisis de diversidad alfa y beta.

Se resumió el contenido del archivo clustered_table (**Anexo 13**), para ajustar el valor de la profundidad de muestreo en los análisis de diversidad Alfa, Beta y Alfa rarefacción. Posteriormente, se calcularon indicadores de diversidad, evaluación de significancia, gráficos de Alfa rarefacción y transformaciones de formato .qza a .qzv siguiendo el Pipeline descrito en el **Anexo 14**.

2.2.2.3 Análisis taxonómico.

Para la asignación taxonómica de los identificadores de las secuencias representativas se utilizó el clasificador pre entrenado SILVA, disponible en QIIME2. Este clasificador, basado en el algoritmo de Naive Bayes es reconocido por su eficiencia y actualización continua. La descarga del clasificador se realizó mediante el comando detallado en el **Anexo 15 (Bolyen et al., 2019)**.

Para ejecutar la asignación taxonómica y visualización se utilizó el código descrito en el **Anexo 16**. Los identificadores únicos de cada secuencia son reemplazados por taxones clasificados por SILVA, con lo que se facilitó la visualización y el uso en el análisis filogenético.

2.2.2.4 Creación de los árboles filogenéticos.

Para la creación de los árboles filogenéticos se utilizó la Tabla de Frecuencias y las secuencias representativas, a partir de estos se creó un árbol filogenético mediante el cual se identificaron las comunidades microbianas presentes en las muestras de forma gráfica. Para este proceso se utilizó el plugin detallado en el **Anexo 17**, en este proceso se crean cuatro archivos importantes: secuencias alineadas, secuencias alineadas con filtrado de posiciones altamente variables, árbol no enraizado y el árbol enraizado (**Bolyen et al., 2019**). Se exportó el árbol enraizado mediante el comando del **Anexo 18**, obteniendo un archivo de tipo .nwk el cual se visualizó en la plataforma iTOL.

2.2.3 Evaluación y comparación de los perfiles metagenómicos.

2.2.3.1 Tabulación de los datos de acuerdo con el sexo y el tipo de enfermedad.

Se elaboró una tabla en la que se clasificó los datos por sexo y tipo de enfermedad del paciente, esto facilita el análisis de las correlaciones presentes entre estas variables, así como la visualización de tendencias relevantes para la identificación de biomarcadores.

2.2.3.2 Comparativa de la diversidad alfa y beta de cada muestra.

Se evaluó también la diversidad Alfa (riqueza y equidad de especies dentro de cada muestra) y la diversidad Beta (diferencias en composición microbiana entre grupos). Este análisis permitió identificar similitudes y diferencias en las comunidades microbianas en función de la enfermedad y el sexo del paciente.

2.2.3.3 Cuadros comparativos de abundancia entre los tipos de enfermedades.

Se generaron tablas para efectuar los análisis comparativos de abundancia de las secuencias en las muestras de ambas enfermedades. Esto facilitó la identificación de diferencias significativas en la composición microbiana asociada a las enfermedades de ECr y EC.

2.2.3.4 Comparativa del análisis taxonómico.

Se compararon las especies identificadas en cada muestra, evaluando similitudes y diferencias taxonómicas entre los grupos de estudio. Esto permitió identificar de patrones taxonómicos específicos asociados a cada enfermedad y la posible influencia del sexo en la distribución de especies.

2.2.3.5 Identificación de posibles biomarcadores.

Se identificaron especies representativas que variaron en función de la enfermedad y el sexo del paciente, las cuales se considerarán como posibles biomarcadores de diagnóstico para las ECr y EC.

CAPÍTULO III.- RESULTADOS Y DISCUSIÓN

3.1. Análisis y discusión de los resultados

3.1.1. Clasificación de los datos de las muestras de las enfermedades Celiaca y de Crohn.

El control de calidad efectuado sobre las muestras secuenciadas permitió analizar la distribución de las longitudes de los reads en archivos .txt., diferenciando las secuencias forward y reverse (**Figura 1**).

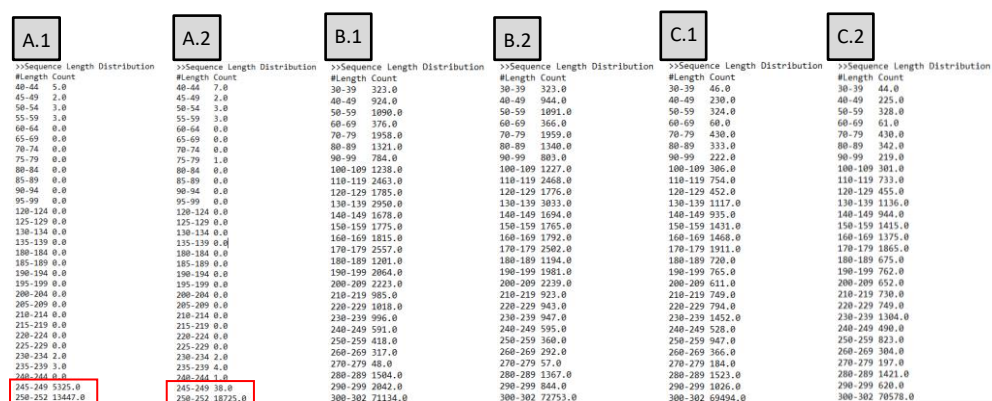


Figura 1. Distribución de longitud de los reads en las muestras de: E. Celiaca, Control y Crohn.

Nota: En las muestras asociadas al a EC (**A.1** y **A.2**) se observó una mayor distribución a partir de la longitud 245. Las muestras control (**B.1**, **B.2**) y las asociadas a la ECr (**C.1** y **C.2**) mostraron una distribución uniforme a lo largo de todas las longitudes.

Adicionalmente, los gráficos de *quality score* obtenidos de las secuencias crudas, como se aprecia en la **Figura 2**, evidenciaron que la calidad disminuía significativamente en todas las muestras analizadas, particularmente en las posiciones finales de los reads.

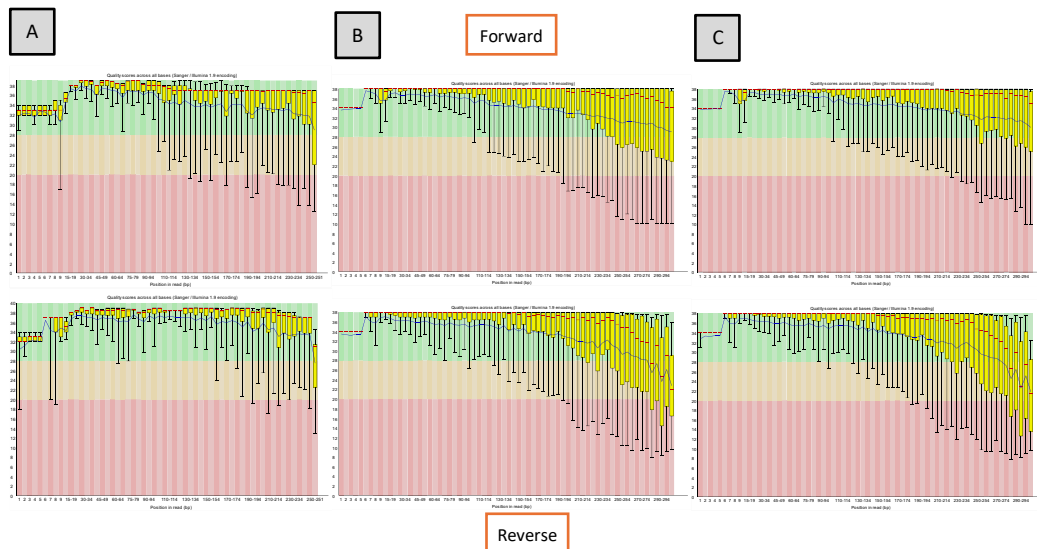


Figura 2. FASTQC de las secuencias crudas de todas las muestras.

Nota: Se puede observar como la calidad promedio de cada una de las muestras: A) Celíaca, B) Control y C) Crohn, se encontraba por debajo de un *quality score* de 30, lo que comprometía la confiabilidad de los resultados, según los estándares de **Illumina (2024)**. Este deterioro fue especialmente notable en las posiciones finales, tanto en las secuencias *forward* como en las *reverse*.

En estos gráficos observamos la calidad de los reads de las muestras, esto permite establecer los parámetros de corte que se mencionaron en la **Tabla 5**, con la finalidad de no perder una gran cantidad de secuencias y que las mismas cuenten con un *quality score* óptimo de 30 a 38.

Tras este proceso, se verificó la eficacia de los cortes mediante archivos .txt y .html. La **Figura 3** muestra los archivos .txt después del corte, en donde se cortaron las secuencias que no mantuvieron la calidad adecuada.

A.1	A.2	B.1	B.2	C.1	C.2
>>Sequence Length Distribution #Length Count	>>Sequence Length Distribution #Length Count	>>Sequence Length Distribution #Length Count	>>Sequence Length Distribution #Length Count	>>Sequence Length Distribution #Length Count	>>Sequence Length Distribution #Length Count
245 14.0	245 16.0	50-54 472.0	50-54 688.0	50-54 1020.0	50-54 1096.0
246 27.0	246 135.0	55-59 491.0	55-59 1045.0	55-59 672.0	55-59 1374.0
247 25.0	247 70.0	60-64 238.0	60-64 969.0	60-64 374.0	60-64 1067.0
248 24.0	248 59.0	65-69 503.0	65-69 918.0	65-69 686.0	65-69 1142.0
249 1607.0	249 102.0	70-74 445.0	70-74 730.0	70-74 1135.0	70-74 1634.0
250 3088.0	250 3118.0	75-79 610.0	75-79 890.0	75-79 1231.0	75-79 1402.0
251 500.0	251 1785.0	80-84 649.0	80-84 1207.0	80-84 834.0	80-84 1461.0
		85-89 501.0	85-89 997.0	85-89 943.0	85-89 1324.0
		90-94 445.0	90-94 847.0	90-94 673.0	90-94 1062.0
		95-99 654.0	95-99 894.0	95-99 1051.0	95-99 1275.0
		160-164 858.0	160-164 1099.0	160-164 1090.0	160-164 1255.0
		165-169 773.0	165-169 1099.0	165-169 846.0	165-169 1155.0
		170-174 781.0	170-174 924.0	170-174 1200.0	170-174 1189.0
		175-179 1295.0	175-179 1232.0	175-179 1612.0	175-179 1393.0
		180-184 675.0	180-184 1122.0	180-184 897.0	180-184 1068.0
		185-189 511.0	185-189 1070.0	185-189 817.0	185-189 1040.0
		190-194 772.0	190-194 803.0	190-194 1075.0	190-194 1003.0
		195-199 953.0	195-199 1416.0	195-199 1902.0	195-199 1682.0
		235-239 1162.0	235-239 610.0	235-239 807.0	235-239 602.0
		240-244 694.0	240-244 1609.0	240-244 877.0	240-244 1734.0
		245-249 1154.0	245-249 1451.0	245-249 1200.0	245-249 1237.0
		250-254 1266.0	250-254 1537.0	250-254 1159.0	250-254 1765.0
		255-259 855.0	255-259 1739.0	255-259 741.0	255-259 1434.0
		260-264 609.0	260-264 1003.0	260-264 621.0	260-264 950.0
		265-269 679.0	265-269 1909.0	265-269 766.0	265-269 1847.0
		270-274 830.0	270-274 1116.0	270-274 825.0	270-274 997.0
		275-279 799.0	275-279 1870.0	275-279 822.0	275-279 1120.0
		280-284 635.0	280-284 1676.0	280-284 805.0	280-284 1351.0
		285-289 1355.0	285-289 954.0	285-289 1391.0	285-289 752.0
		290-294 727.0	290-294 1049.0	290-294 972.0	290-294 1005.0
		295-299 1105.0	295-299 135.0	295-299 1432.0	295-299 178.0
		300-302 16581.0	300-302 1844.0	300-302 13026.0	300-302 2267.0

Figura 3. Distribución de la longitud de las secuencias de las muestras de Celíaca, Control y Crohn después del corte de Trimmomatic.

Nota: Las figuras **A.1** y **A.2** (forward y reverse de Celíaca) muestran que el corte mantuvo las secuencias desde la longitud 245 en adelante; mientras que las figuras **B.1**, **B.2** (forward y reverse de Control), **C.1** y **C.2** (forward y reverse de Crohn) mantuvieron la distribución uniforme menorando la cantidad de las secuencias de manera mínima y no significativa.

Mientras que las gráficas de los archivos .html mostraron como la calidad mejoró tras los cortes realizados (**Figura 4**) en comparación con las gráficas de las secuencias crudas (**Figura 2**).

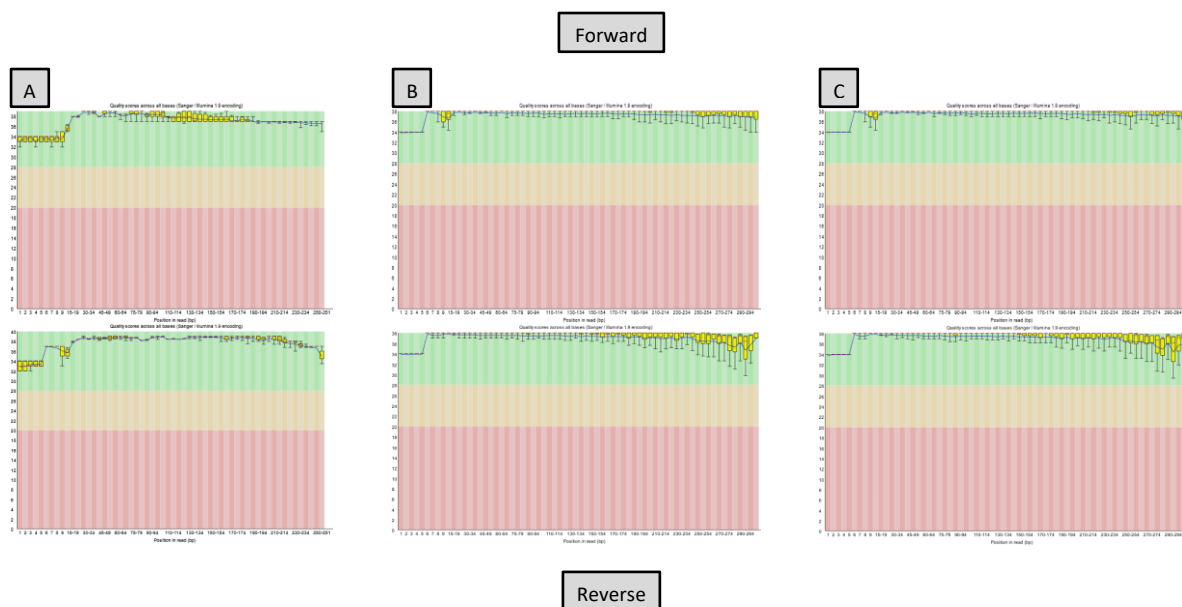


Figura 4. Visualización de las secuencias tras el control de calidad en FASTQC.

Nota: Se observa que la calidad en cada una de las muestras: A) Celíaca, B) Control y C) Crohn, es superior a un quality score de 30, lo que garantiza la calidad de las lecturas indicando una probabilidad de error 1 en 1000 (99.9%) (Illumina, 2024). También se observó un ligero descenso en la calidad de las secuencias reverse en las últimas posiciones, aunque estas se mantuvieron dentro de los límites aceptables para el análisis metagenómico.

Posteriormente, se generaron los archivos *paired-end-demux.qza* y *demux.qzv*, que permitieron obtener estadísticas detalladas de las secuencias por grupo (Tabla 7).

Tabla 7.

Resumen de los archivos demux.qzv de cada enfermedad.

	Celíaca		Control		Crohn	
	Forward reads	Reverse reads	Forward reads	Reverse reads	Forward reads	Reverse reads
Mínimo	461,00	461,00	46337,00	46337,00	40318,00	40318,00
Mediana	5521,00	5521,00	72923,00	72923,00	64872,50	64872,50
Media	6936,705882	6936,705882	76398,30	76398,30	71248,50	71248,50
Máximo	23464,00	23464,00	105607,00	105607,00	116938,00	116938,00
Total	117924,00	117924,00	763983,00	763983,00	712485,00	712485,00

Nota: Todas las muestras presentan una cantidad suficiente de secuencias y un número igual de lecturas en cadena forward y reverse, lo que cumple con los requisitos necesarios para análisis posteriores utilizando algoritmos como DADA2 (Odom et al., 2023). Además, las muestras de EC mostraron una

menor cantidad de secuencias en comparación con las muestras Control y ECr, lo que podría estar relacionado con una menor diversidad microbiana.

También se obtuvieron histogramas de la distribución del total de las secuencias en cada muestra (Figura 5).

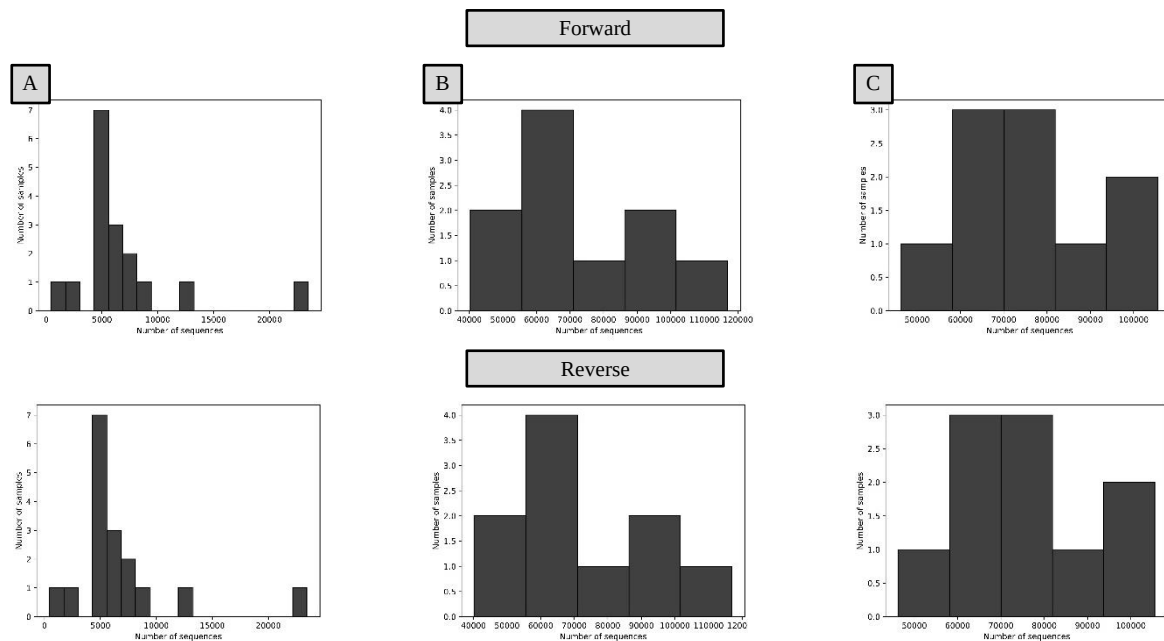


Figura 5. Histograma de frecuencia de las lecturas forward y reverse. (colocar en el gráfico FW y RV).

Nota: En los histogramas de frecuencia de lecturas forward (superior) y reverse (inferior), se observa que el grupo Celíaca (A) presenta una mayor proporción de muestras con una menor cantidad de lecturas entre 5000 y 10000, lo que indica una cobertura menor en comparación con los grupos Control (B) y Crohn (C). Por otro lado, los grupos Control y Crohn muestran distribuciones más uniformes, con la mayoría de las muestras concentradas entre 40.000 y 80.000 lecturas. No se observan diferencias significativas entre las lecturas forward y reverse en ningún grupo, lo que indica un buen emparejamiento de las secuencias.

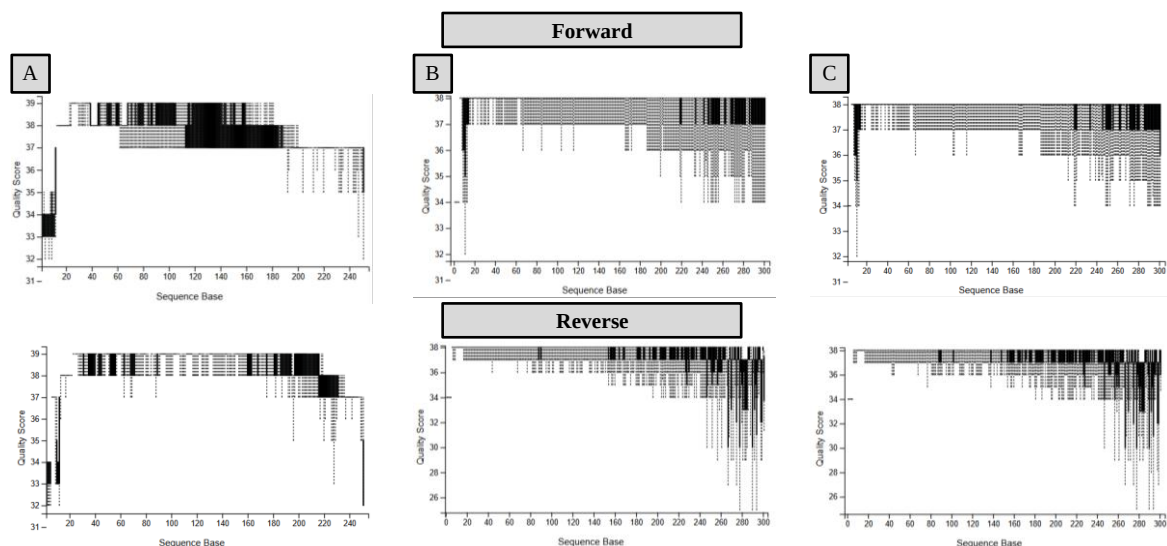


Figura 6. Calidad de las lecturas forward y reverse del archivo demux.qzv.

Nota: La gráfica de calidad presentada muestra la distribución de los puntajes de calidad de las secuencias forward (gráficas superiores) y reverse (gráficas inferiores) para cada grupo analizado: Celíaca (A), Control (B) y Crohn (C). Estas gráficas presentan como la calidad es inferior en las primeras posiciones (<20) y en posiciones finales.

El procesamiento posterior con el algoritmo DADA2 (**Anexo 10**) permitió optimizar la calidad de las secuencias mediante parámetros de truncamiento y recorte adaptados a cada grupo (**Tabla 5**), en base a las gráficas de calidad del archivo demux.qzv.

Tabla 8.

Parámetros de corte para el proceso de DADA2.

	Celíaca	Control	Crohn
p-trunc-len-f	220	280	240
p-trunc-len-r	240	246	220
p-trim-left-f	20	20	20
p-trim-left-r	20	20	20

Nota: Los valores trunc indican la posición final de las secuencias donde se realiza el truncamiento, es decir, las bases posteriores a este punto son descartadas. Por otro lado, los valores trim especifican la posición inicial desde la cual se recortan las secuencias, eliminando las bases previas a este punto. Las letras f y r hacen referencia a las secuencias forward y reverse respectivamente. Estos parámetros son

clave para garantizar que las secuencias utilizadas en el análisis mantengan una calidad óptima, considerando las características específicas de cada grupo de muestras (Celíaca, Control y Crohn).

Como resultado, se generaron los archivos *table.qza*, *rep-seqs.qza* y *denoising-stats.qza* (resumen de los archivos en los **Anexos 19, 20, 21 y 22**). Posteriormente se filtraron las secuencias para eliminar aquellas que no eran significativas (**Edgar, 2016**), con la finalidad de mejorar la calidad, asegurar que los datos sean representativos, no presenten errores y evitar sesgo los resultados. Tras este proceso se obtuvieron los archivos *filtered-table.qza* y *filtered-rep-seqs.qza*, cuyos resultados se resumen en la

Tabla 9.

Resumen estadístico de *filtered-table.qzv* de las enfermedades.

	Celíaca	Control	Crohn
Numero de muestras	17	10	10
Numero de Características	72	742	527
Frecuencia total	12,198	110,559	72,205

Nota: La tabla resume los principales resultados obtenidos del archivo *filtered-table.qzv*, mostrando que en la muestra control presenta una mayor cantidad de características y frecuencia en comparación con Celíaca y Crohn.

La **Tabla 9** presenta el número de muestras, características y frecuencias totales asociadas a la microbiota intestinal. Se analizaron 17 muestras para Celíaca, 10 para Crohn y 10 del grupo Control. El grupo Control muestra una cantidad significativamente mayor de características (742) en comparación con Celíaca (72) y Crohn (527), esto podría relacionarse con la diversidad microbiana propia de individuos sanos. En términos de frecuencia total, en la EC se identificó un menor número de características y frecuencia total (12,198) en comparación con los grupos control (110,559) y Crohn (72,205). Este patrón sugiere una asociación entre las enfermedades inflamatorias intestinales y la reducción de la diversidad microbiana, lo que coincide con estudios previos que describen el impacto de la disbiosis en estas patologías (**Silva et al., 2019**).

Tabla 10.

Frecuencia por características de filtered-table.qzv de las enfermedades.

	Celiaca	Control	Crohn
Frecuencia mínima	10	10	10
Primer cuartil	14	20	18
Frecuencia media	27	40	33
Tercer cuartil	69,5	86,75	71,5
Frecuencia máxima	3,104	13,938	7,220
Frecuencia media	169,4	149,00134770889488	137,01138519924098

Nota: La tabla presenta las frecuencias estadísticas asociadas a las características de la microbiota intestinal para cada uno de los grupos (Celiaca, Control, y Crohn) en la cual se incluye valores mínimos, máximos medias y cuartiles.

En la **Tabla 10** se puede inferir que la frecuencia mínima es constante en los tres grupos (10), sin embargo, las diferencias se hacen presentes al analizar los cuartiles y los promedios. El grupo Control muestra una frecuencia media de 40, superior a la de Celiaca (27) y Crohn (33), lo que indica una mayor diversidad en la microbiota de los individuos sanos. Además, la frecuencia máxima del Control (13,938) es mayor que la de Celiaca (3,104) y Crohn (7,220), lo que refuerza que las enfermedades inflamatorias intestinales están asociadas con la disbiosis. Por otro lado, el tercer cuartil muestra un patrón similar, siendo el Control más alto (86,75), seguido de Crohn (71,5) y Celiaca (69,5).

Tabla 11.

Estadísticas de longitudes de secuencia de las enfermedades.

	Recuento de secuencias	Longitud mínima	Longitud máxima	Longitud media	Rango	Desviación estándar
Celiaca	72	245	406	380,29	161	53,07
Control	742	260	436	381,33	176	59,75
Crohn	527	260	466	384,42	206	58,5

Nota: La tabla presenta las estadísticas descriptivas de las longitudes de las secuencias representativas obtenidas del archivo filtered-reps-seqs.qzv para las muestras de Celiaca, Crohn y el grupo control. Estas métricas permiten evaluar la calidad y consistencia de las secuencias utilizadas en el análisis.

En la **Tabla 11** se presenta el recuento de secuencias, el cual varía significativamente entre los grupos, siendo notablemente mayor en el Control (742) y Crohn (527), en comparación con Celíaca (72). Esto podría indicar una mayor pérdida de secuencias en el proceso de filtrado en el grupo Celíaca. En cuanto a la longitud mínima las secuencias de Celíaca tienen un valor ligeramente inferior (245) frente al Control y Crohn (260). La longitud máxima también presenta diferencias, siendo mayor en Crohn (466), seguido del Control (436) y Celíaca (406), sugiriendo mayor variabilidad en la microbiota de la enfermedad de Crohn. Además, la longitud media es similar entre los tres grupos, con valores que oscilan entre 380,29 (Celíaca) y 384,42 (Crohn).

El rango de las longitudes de las secuencias varía entre los grupos, siendo mayor en Crohn (206), seguido por el Control (176) y Celíaca (161). Este mayor rango puede indicar una mayor variabilidad en las longitudes de las secuencias representativas, lo que podría reflejar una mayor diversidad estructural en su microbiota. Por otro lado, la desviación estándar es mayor en el Control (59,75), seguido por Crohn (58,5) y Celíaca (53,07). Una mayor desviación estándar sugiere que las longitudes de las secuencias están más dispersas.

Tabla 12.

Resumen de siete números de longitudes (nts) de secuencias de las enfermedades.

Percentil	2%	9%	25%	50%	75%	91%	98%
Celíaca	246	249	400	400	403	405	406
Control	260	260	395	403	425	425	426
Crohn	260	260	400	405	425	425	426

Nota: La tabla presenta un resumen de los percentiles clave de las longitudes de las secuencias representativas (en nucleótidos, nts) para las enfermedades, basado en el archivo filtered-reps-seqs.qzv.

Mediante la **Tabla 12** se puede inferir que en los percentiles más bajos (2% y 9%), las secuencias de Celíaca muestran longitudes menores (246 y 249 nt, respectivamente) en comparación con Crohn y control (ambos con 260 nt), lo que sugiere una mayor proporción de secuencias más cortas en el grupo Celíaca. En los percentiles intermedios (25% y 50%), Celíaca y Crohn presentan valores similares mientras que el control muestra un valor ligeramente menor en el percentil 25% (395 nt). En los

percentiles superiores (75%, 91% y 98%), las secuencias del grupo Control y Crohn alcanzan longitudes mayores (425-426 nt) en comparación con Celíaca, cuyo máximo llega a 406 nt. Esto indica que las secuencias de los grupos Control y Crohn tienen una distribución más amplia hacia las longitudes mayores, lo que podría reflejar una mayor complejidad o diversidad en estas comunidades microbianas

3.1.2. Ejecución del análisis metagenómico

Tras el proceso de clustering, las características se agruparon en unidades taxonómicas operativas (OTUS) con un porcentaje de similitud del 97%. Este agrupamiento generó los archivos *clustered_table.qzv* y *clustered_sequences.qzv*, los cuales contienen la información sobre las características y las agrupadas.

Tabla 13.

Resumen estadístico de cluster-table.qzv de las enfermedades.

	Celíaca	Control	Crohn
Numero de muestras	17	10	10
Numero de Características	42	472	355
Frecuencia total	12,198	110,559	72,205

Nota: La tabla resume los principales resultados obtenidos del archivo clustered-table.qzv, mostrando que en la muestra control presenta una mayor cantidad de características y frecuencia en comparación con Celíaca y Crohn.

La **Tabla 13** presenta el número de muestras, características y frecuencias totales de las OTUs asociadas a la microbiota intestinal. Se analizaron 17 muestras para Celíaca, 10 para Crohn y 10 del grupo Control. Aunque el grupo Control muestra una cantidad significativamente mayor de características (472) en comparación con Celíaca (42) y Crohn (355), esto podría deberse a la diversidad microbiana propia de pacientes sanos. En términos de frecuencia total, el Control también presenta un valor considerablemente mayor (110,559) frente a Crohn (72,205) y Celíaca (12,198). Estos datos evidencian que, tanto en la tabla filtrada como en la tabla agrupada, el grupo Control supera en diversidad y

frecuencia a los grupos de Celíaca y Crohn. Esta reducción en diversidad microbiana en los grupos de enfermedad está asociada con un estado de disbiosis, que podría desempeñar un papel clave en el desarrollo y progresión de estas patologías.

Tabla 14.

Frecuencia por características de cluster-table.qzv de las enfermedades.

	Celíaca	Control	Crohn
Frecuencia mínima	10	10	10
Primer cuartil	14,25	19	18
Frecuencia media	57,5	42	33
Tercer cuartil	106,75	124,00	71,5
Frecuencia máxima	3270,00	21065,00	7,220
Frecuencia media	290,428571	149,00134770889488	137,01138519924098

Nota: La tabla presenta las frecuencias estadísticas asociadas a las características de la microbiota intestinal para cada uno de los grupos (Celíaca, Control, y Crohn) en la cual se incluye valores mínimos, máximos medias y cuartiles.

Mediante la **Tabla 14** se puede inferir que la frecuencia mínima es constante en los tres grupos (10), sin embargo, las diferencias se presentan al analizar los cuartiles y los promedios. Celíaca presenta una frecuencia mínima de 57,5 superior a la del Control (42) y Crohn (33). Además, la frecuencia máxima del Control (21065) es mayor que la de Celíaca (3270) y Crohn (7,220). Por otro lado, el tercer cuartil muestra un patrón similar, siendo el Control más alto (124), seguido de Crohn (71,5) y Celíaca (106,75).

Tabla 15.

Estadísticas de longitudes de secuencia de las enfermedades.

	Recuento de secuencias	Longitud mínima	Longitud máxima	Longitud media	Rango	Desviación estándar
Celiaca	42	245	406	379,86	161	54,06
Control	472	260	436	370,67	176	64,3
Crohn	355	260	466	378,68	206	61,37

Nota: La tabla presenta las estadísticas descriptivas de las longitudes de las secuencias representativas obtenidas del archivo clustered-reps-seqs.qzv para las muestras de Celiaca, Crohn y el grupo control. Estas métricas permiten evaluar la calidad y consistencia de las secuencias utilizadas en el análisis.

Como se puede observar en la **Tabla 15** en comparación de la **Tabla 11**, el recuento de secuencias por características disminuye, siendo en el caso de Celiaca que pasan de 72 a 42 características, Control de 742 a 472 y Crohn de 527 a 355. Esto muestra que el proceso de asignación de OTUs tuvo éxito, ya que agrupa las secuencias con una similitud del 97%, a pesar de esto, la frecuencia de las secuencias se mantiene en todos los tipos de muestra con ciertas variaciones mínimas, esto asegura que en el proceso de asignación de OTUs no se han perdido secuencias.

Esta evidencia se refuerza con los datos mostrados en la **Tabla 16**, en esta se observa que la frecuencia de las características en los distintos percentiles mantiene las mismas frecuencias con una pequeña variación en ciertos percentiles, por otro lado, la distribución de las características no ha variado, lo que asegura que no hay cambios significativos en la representatividad de cada una de las secuencias.

Tabla 16.

Resumen de siete números de longitudes (nts) de secuencias de las enfermedades.

Percentil	2%	9%	25%	50%	75%	91%	98%
Celiaca	245	249	400	400	403	405	405
Control	260	260	298	400	420	425	426

Crohn	260	260	350	402	425	425	426
--------------	-----	-----	-----	-----	-----	-----	-----

Nota: La tabla presenta un resumen de los percentiles clave de las longitudes de las secuencias representativas (en nucleótidos, nt) para las enfermedades, basado en el archivo clustered-sequences.

El análisis de diversidad Alfa y Beta se efectuó utilizando los archivos clustered_table.qza, y metadata_enfermedad.tsv. Inicialmente, se muestran las curvas de rarefacción correspondientes a las EC y ECr (**ver Anexos 23 y 24**), estas permiten evaluar si la profundidad de secuenciación es adecuada. Como lo menciona **Icaza (2013)**, estas curvas son útiles para determinar si el muestreo es suficiente para capturar la mayor cantidad de diversidad microbiana.

En el **Anexo 23** se muestra la curva de rarefacción de la EC donde se observa distintos índices, cada uno evalúa diferentes aspectos de la diversidad microbiana. Shannon se enfoca en la diversidad microbiana (riqueza y equidad), esta alcanza la estabilización alrededor de 3,5 a una profundidad de secuenciación mayor a 800, siendo mayor en mujeres que hombres. Faith's PD mide la diversidad considerando la distancia filogenética entre especies, se estabiliza en un valor de 3,5 en valores de secuenciación superior a 600, se muestra una tendencia similar entre ambos sexos. Observed Features mide la riqueza observada (especies identificadas), esta se estabiliza en un valor mayor a 27, en mujeres ocurre a profundidades de secuenciación intermedias a 600, mientras que en hombres se requiere profundidades mayores a 600. En cuanto a la curva de rarefacción de la ECr (**ver Anexo 24**) el índice Shannon alcanza su estabilización alrededor de un valor de 5 a profundidades mayores de 2000, con una ligera variación en hombres a mayores profundidades. Faith's PD presenta valores altos en mujeres (24) en comparación con hombres (18), estabilizándose a profundidades de secuenciación cercanos a 3000. Finalmente, el índice Observed Features presenta valores similares a los anteriores en donde las mujeres alcanzan valores cercanos a 70 estabilizándose a valores de profundidad de 2000 lecturas, mientras que los hombres presentan valores bajos a profundidades mayores.

En cuanto a la diferenciación de la variabilidad de las comunidades microbianas en cada una de las condiciones, se obtuvo un PCOA (Análisis de coordenadas principales) que indica las muestras de cada paciente con EC y ECr y su relación con las muestras Control.

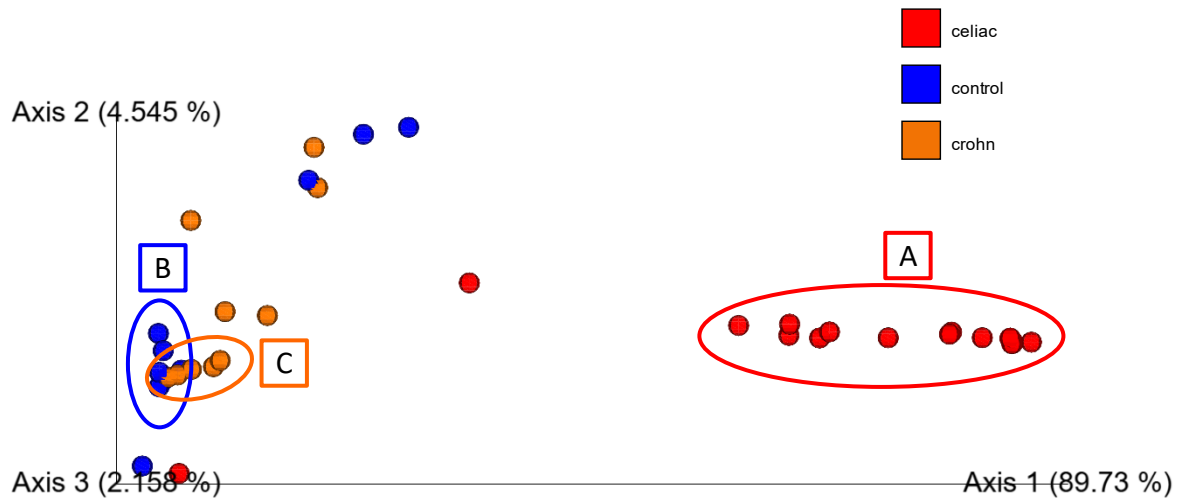


Figura 7. Gráfico PCOA Weighted Unifrac (PCOA).

Nota: Las muestras de Celíaca (A) son el grupo que más se diferencian del resto. Mientras que Control (B) y Crohn (C) son grupos que se sobreponen pero tienen una distribución diferente. Además de contar con puntos de muestras dispersas.

El análisis PCOA con la métrica Weighted UniFrac muestra diferencias significativas entre las condiciones estudiadas. Las muestras de EC (rojo) están marcadamente separadas del resto, especialmente a lo largo del eje 1 (89.73%), indicando una disbiosis severa y específica, posiblemente relacionada con una reducción en bacterias beneficiosas para la microbiota intestinal, como lo menciona **Russo et al., (2021) & Quaglio et al. (2022)**, aquí se presenta una disminución de *Firmicutes*, *Bacteroidotas* y *Actinobacterium*. En contraste, las muestras de Crohn (naranja) muestran cierto solapamiento con las muestras Control (azul), aunque con una ligera diferenciación a lo largo del eje 2 (4.545%), lo que sugiere alteraciones más sutiles en la composición microbiana. Con estas observaciones se destaca que la EC presenta un impacto microbiano más drástico, mientras que Crohn comparte más similitudes con la microbiota de individuos sanos. Esto también podría deberse a que la patogénesis en ambas enfermedades es distinta (**Aljada et al., 2021; Gordon et al., 2024**).

La asignación taxonómica se realizó sobre las 3 muestras (Celíaca, Control y Crohn, taxonomy.tsv **ANEXO 25**) los cuales contienen el Feature ID con el código de las características que también están presentes en el resumen de la tabla de características, los taxones asignados para cada característica y un valor de confianza (entre 0 y 1) que indica que tan segura es la asignación realizada por el clasificador.

Además, QIIME2 permitió colapsar estas características, agrupando cada taxón en varios niveles taxonómicos, de los que se utilizaron 5 de los 7 niveles que QIIME2 ofrece, estos son: 2) Filo, 3) Clase, 4) Orden, 5) Familia y 6) Género; con el objetivo de mantener la fidelidad de los datos con la tecnología 16S rRNA manteniendo una potencia aceptable de reconocimiento taxonómico, tal como lo menciona **Durazzi et al. (2021)**. Las asignaciones taxonómicas en los niveles mencionados de las muestras se pueden observar en el **ANEXO 26**. En la **Tabla 17** se presentan los datos resumidos de dichas asignaciones.

Tabla 17.

Resumen de las tablas de las asignaciones taxonómicas.

	Frecuencia	Filo	Clase	Orden	Familia	Género
Celiaca	(1) 8589	(1) 1	(1) 1	(1) 1	(1) 1	(1) 1
	(2) 3609	(2) 8	(2) 9	(2) 16	(2) 22	(2) 31
Control	(1) 10727	(1) 1	(1) 1	(1) 1	(1) 1	(1) 1
	(2) 99832	(2) 18	(2) 30	(2) 74	(2) 124	(2) 227
Crohn	(1) 10727	(1) 1	(1) 1	(1) 1	(1) 1	(1) 1
	(2) 99832	(2) 16	(2) 24	(2) 65	(2) 113	(2) 187

Nota: (1) son las características no asignadas por el clasificador y (2) aquellas que lograron ser asignadas.

Se puede destacar que existen diferencias, aunque Silva sea un clasificador comúnmente empleado, tiene ciertas limitaciones ya que no logra clasificar todas las características. Esto puede deberse a distintas razones, como alteraciones en las lecturas debido al control de calidad, falta de potencia del clasificador, además, la existencia de ciertas secuencias de las características que sean taxones y aún no están anotados en bases de datos como NCBI.

A partir de los taxones y la frecuencia de estos en las muestras, se graficaron Barplots para representar la abundancia relativa de los filos clasificados en cada muestra y árboles filogenéticos que contienen los géneros de las especies presentes y su relación con el filo al que pertenecen.

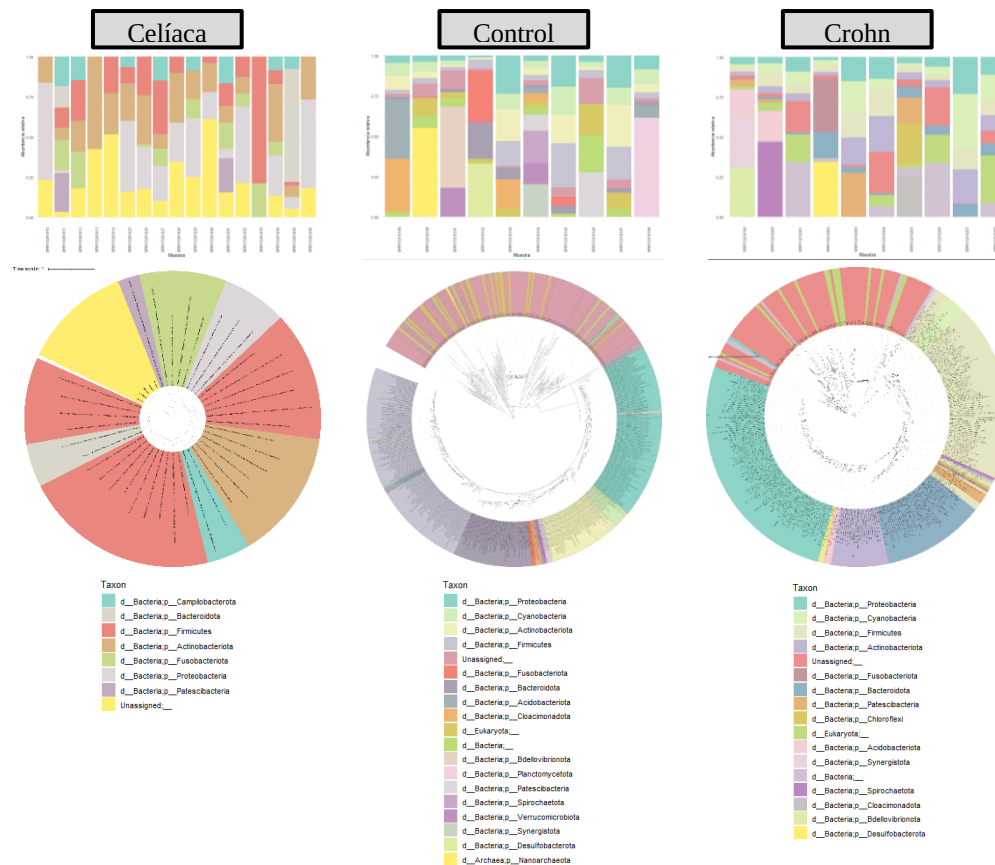


Figura 8. Abundancia relativa y árboles filogenéticos de las muestras Celíaca, Control y Crohn.

Nota: El gráfico muestra los árboles filogenéticos y abundancia relativa de cada uno de los grupos de estudio.

En la **Figura 8** se observa que el grupo control presenta mayor homogeneidad en la distribución de los filos microbianos dominantes, lo que sugiere una microbiota intestinal más equilibrada en comparación con las muestras de EC y ECr. Por el contrario, estas muestras indican una distribución menos homogénea, caracterizada por un aumento en filos como *Proteobacterias* y una disminución de *Firmicutes*, *Bacteroidota* y *Actinobacteriota*, lo que representa claramente un evento de disbiosis e inflamación intestinal, caracterizados por la disminución de filos beneficiosos y aumento de filos patógenos u oportunistas.

Según **Quaglio et al. (2022)**, la microbiota intestinal humana evoluciona a lo largo de la vida, durante los primeros años, su composición es relativamente simple, con una baja diversidad y un predominio relativo de *Proteobacterias* y *Actinobacterias*. Con el paso del tiempo la microbiota se diversifica y los filos *Firmicutes* y *Bacteroidota* emergen como los más prevalentes. En adultos, más del 90% de las

bacterias intestinales pertenecen a los filos *Firmicutes*, *Bacteroidota*, *Proteobacterias* o *Actinobacteriota*.

Vester et al. (2019), explican el aumento de *Proteobacterias* en condiciones inflamatorias a través de dos mecanismos principales: la hipótesis del oxígeno y la presencia de nitrato. La hipótesis del oxígeno propone que, en un colon saludable, las células epiteliales consumen el oxígeno en el lumen a través de la beta oxidación, creando un ambiente anaerobio. Durante la inflamación, esta capacidad disminuye, aumentando la disponibilidad de oxígeno y favoreciendo el crecimiento de *Proteobacterias*, lo que contribuye al desequilibrio microbiano. Así mismo, la presencia de nitrato generado como subproducto de la inflamación, proporciona una ventaja competitiva a bacterias comensales oportunistas, promoviendo su crecimiento y eventual predominio en la microbiota intestinal.

En cambio, filos como *Firmicutes*, *Bacteroidotas* y *Actinobacterium*, son considerados benéficos ya que están relacionados con procesos de regulación inflamatoria. **Gao et al., 2022**, demostraron como las bacterias *Firmicutes* producen DL-endopeptidasas que se relacionan con el ligando NOD2, el mismo que está involucrado en distintos procesos fisiológicos como el entrenamiento inmunológico, regeneración epitelial, resistencia a la insulina, respuesta a la vacunación, además, el deterioro de este ligando se asocia directamente con la enfermedad de Crohn.

3.1.3. Evaluación y comparación de los perfiles metagenómicos.

Para la tabulación en función al sexo y enfermedad se calculó promedios de los taxones a nivel de Filo, en estos se puede diferenciar los taxones más abundantes dependiendo del sexo.

Tabla 18.

Abundancia relativa de los taxones (Filo) por sexo en la EC.

Hombres		Mujeres	
Taxón	Promedio	Taxón	Promedio
Unassigned;__	464,75	Unassigned;__	517,7
d__Bacteria;p__Fusobacteriota	220,5	d__Bacteria;p__Fusobacteriota	44,8
d__Bacteria;p__Firmicutes	109,5	d__Bacteria;p__Firmicutes	35,6
d__Bacteria;p__Campilobacterot a	60,5	d__Bacteria;p__Actinobacteriota	27,2

Hombres		Mujeres	
Taxón	Promedio	Taxón	Promedio
d__Bacteria;p__Actinobacteriota	53,5	d__Bacteria;p__Campilobacterot a	12,8
d__Bacteria;p__Proteobacteria	11,5	d__Bacteria;p__Proteobacteria	12,3
d__Bacteria;p__Patescibacteria	7	d__Bacteria;p__Bacteroidota	1,1
d__Bacteria;p__Bacteroidota	2,5	d__Bacteria;p__Patescibacteria	0,7

Nota. La tabla evidencia diferencias significativas en los perfiles metagenómicos a nivel de Filo entre hombres y mujeres con EC. En ambos sexos, el grupo Unassigned muestra la mayor abundancia, destacando la necesidad de explorar posibles taxones no identificados. Los hombres presentan mayor abundancia de *Fusobacteriota* (220,5) y *Firmicutes* (109,5) en comparación con las mujeres, donde son menos prevalentes (44,8 y 35,6 respectivamente). Además, los hombres muestran una mayor diversidad en los taxones representados, mientras que en mujeres la abundancia se concentra en menos grupos.

Tabla 19.

Abundancia relativa de los taxones (Filo) por sexo en la enfermedad de Crohn.

Hombres		Mujeres	
Taxón	Promedio	Taxón	Promedio
d__Bacteria;p__Proteobacteria	3797,3	d__Bacteria;p__Proteobacteria	2852,5
d__Bacteria;p__Firmicutes	1931,2	d__Bacteria;p__Firmicutes	1037,5
d__Bacteria;p__Cyanobacteria	793	Unassigned;__	903,8
d__Bacteria;p__Bacteroidota	518,7	d__Bacteria;p__Cyanobacteria	517,3
d__Bacteria;p__Actinobacteriota	444,5	d__Bacteria;p__Actinobacteriota	353,5
Unassigned;__	435,8	d__Bacteria;p__Bacteroidota	89
d__Bacteria;p__Fusobacteriota	73	d__Eukaryota;__	35,5
d__Bacteria;p__Patescibacteria	69,3	d__Bacteria;__	15
d__Eukaryota;__	42,2	d__Bacteria;p__Fusobacteriota	0
d__Bacteria;p__Acidobacteriota	20,3	d__Bacteria;p__Patescibacteria	0
d__Bacteria;p__Chloroflexi	14,7	d__Bacteria;p__Chloroflexi	0
d__Bacteria;p__Synergistota	8,3	d__Bacteria;p__Acidobacteriota	0
d__Bacteria;__	5,8	d__Bacteria;p__Synergistota	0
d__Bacteria;p__Spirochaetota	3,7	d__Bacteria;p__Spirochaetota	0
d__Bacteria;p__Cloacimonadota	3	d__Bacteria;p__Cloacimonadota	0
d__Bacteria;p__Bdellovibrionota	2	d__Bacteria;p__Bdellovibrionota	0
d__Bacteria;p__Desulfobacterota	2	d__Bacteria;p__Desulfobacterota	0

Nota. La tabla muestra diferencias marcadas en la abundancia relativa de los taxones a nivel de filo entre hombres y mujeres con enfermedad de Crohn. En ambos sexos, *Proteobacteria* es el filo predominante, siendo notablemente mayor en hombres (3797,3) que en mujeres (2852,5), lo que podría indicar que juega un papel clave en la enfermedad. Los hombres presentan mayores abundancias en

otros filos como *Firmicutes* (1931,2), *Cyanobacteria* (793) y *Bacteroidota* (518,7), mientras que en mujeres estos mismos taxones son menos prevalentes. Cabe destacar que algunos filos, como *Fusobacteriota*, *Patescibacteria*, *Chloroflexi*, y otros, están ausentes en las mujeres y tienen baja representación en hombres. Además, el grupo Unassigned es más abundante en mujeres (903,8) que en hombres (435,8), lo que sugiere la posibilidad de taxones no caracterizados relevantes en mujeres.

Para diversidad alfa se calcularon distintos índices: Shannon, Evenness, Faith 's PD y Observed features, los cuales proporcionan una visión detallada sobre la riqueza y equidad de las comunidades presentes en las muestras.

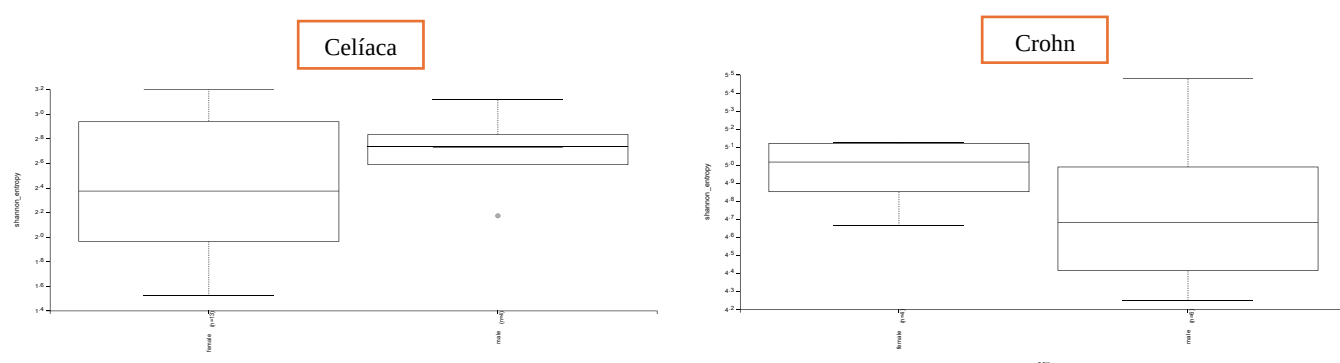


Figura 9. Indicador Shannon de diversidad Alfa.

Nota: El gráfico compara la diversidad alfa, del índice Shannon, entre pacientes con EC y ECr, diferenciados por sexo. En el caso de la EC, se observa una mayor variabilidad en la diversidad entre las muestras femeninas en comparación con las masculinas, aquí los valores parecen estar más concentrados. Por otro lado, en la ECr, las muestras masculinas muestran una mayor dispersión en los valores de Shannon, mientras que las femeninas presentan una diversidad más homogénea.

Tabla 20.

P_value de los índices de diversidad Alfa entre los sexos de cada muestra.

	Shannon	Evenness	Faith's PD	Observed features
Celíaca	0,496906	0,36503	0,496906	0,230151
Crohn	0,286422	0,393769	0,393769	0,519891

Nota: Se analiza los p-values mediante la prueba de Kruskal-Wallis de los índices de la diversidad alfa con un nivel de significancia del 95%.

Los resultados presentados en la **Tabla 20** representan el análisis de diversidad Alfa tanto de la EC como de ECr, Los resultados muestran que no existe diferencias significativas ($p > 0.05$) en ninguno de los índices analizados, es decir, la riqueza, equidad, uniformidad y diversidad filogenética de las comunidades microbianas presentes en cada uno de los grupos (masculino y femenino) no muestran una separación clara, esto quiere decir que el sexo no tiene influencia para el desarrollo de las enfermedades, un comportamiento similar reportó **Pascal et al. (2017)**, donde concluyen que los pacientes con EII entre distintas enfermedades presentaron una diversidad Alfa microbiana menor puesto que el valor de significancia mostró un valor- $p > 0,05$.

En cuanto al análisis Beta se calcularon los índices Bray curtis, Jaccard, Unweighted Unifrac y Weighted Unifrac, estos permiten comparar la composición microbiana entre diferentes muestras.

Tabla 21.

P_value de los índices de diversidad Beta entre los sexos de cada muestra.

	Bray Curtis	Jaccard	Unweighted Unifrac	Weighted Unifrac
Celíaca	0,779	0,593	0,854	0,736
Crohn	0,005	0,284	0,482	0,03

Nota: Se analiza los p_value mediante la prueba PERMANOVA de los índices de la diversidad beta con un nivel de significancia del 95%.

En la **Tabla 21** se muestran los p-values para cada índice de la diversidad Beta. En el caso de la EC los resultados indican que no existe diferencias significativas ($p > 0.05$) en ninguno de los índices analizados, de forma similar en la ECr, los índices Jaccard y Unweighted Unifrac tampoco mostraron diferencias significativas, todo esto indica que no hay variaciones relacionadas a la presencia o ausencia de taxones o filogenia no ponderada. Sin embargo, se observaron diferencias significativas ($p < 0.05$) en los índices Bray Curtis y Weighted Unifrac, lo que significa que existen discrepancias en la composición microbiana entre los sexos, tanto en abundancia relativa como en relaciones filogenéticas ponderadas, respectivamente (**Xia & Sun, 2023**).

Con relación a que las enfermedades presentaron diferencias significativas con el control, y también entre estas, se procedió a ejecutar la comparación entre taxones a nivel de género, mediante R se ejecutó el código para analizar diferencias (**ver Anexo 27**), realizando las siguientes comparaciones:

- Celíaca vs Control
- Celíaca vs Crohn

Estas comparaciones nos permitieron obtener los taxones más significativos, mismos que podrían ser considerados como posibles biomarcadores para las enfermedades. Estos taxones se representaron en un Foresplot donde se representan de mejor forma los taxones más significativos.

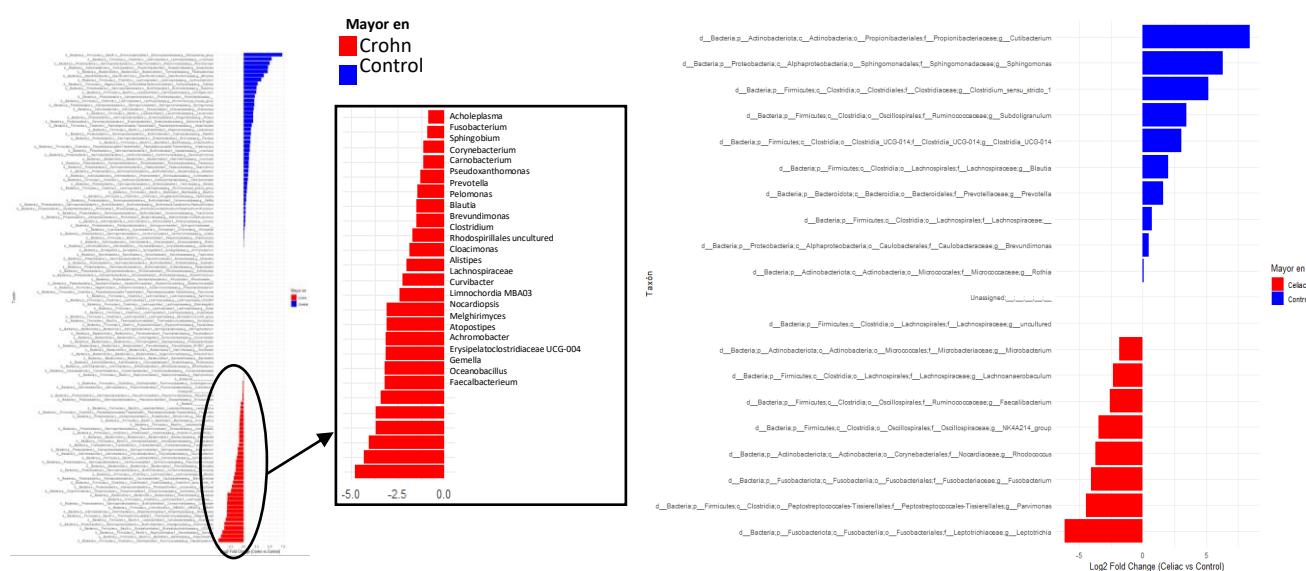


Figura 10. Foresplot Crohn vs Control y Celíaca vs Control.

Nota: La gráfica muestra los taxones que presentan mayor presencia en el control (>0) y en la enfermedad (<0). Para objeto de estudio los taxones se representan en rojo para cada enfermedad.

A partir de estos taxones se efectuó el análisis funcional para identificar si estos se encuentran identificados con funciones biológicas relacionadas a cada enfermedad. En el análisis se utilizaron dos archivos, un archivo con extensión .fasta que contiene el ID de las características (OTUs), otro que contiene las secuencias y otro archivo .biom obtenido a partir del archivo clustered-sequences.qza mediante el código mostrado en el **Anexo 28**.

Con esta información se ejecutó el análisis funcional mediante Picrust2, utilizando el script detallado en el **Anexo 29**. Se obtienen varios archivos donde constan: las rutas metabólicas y las funciones

biológicas relacionadas a los taxones más representativos. Para identificar las rutas más significativas se empleó el archivo `pred_metagenome_unstrat.tsv`, que contiene la frecuencia de las funciones en la muestra, estas funciones presentan formato de código KEGG (Ontology) como filas y las columnas con las muestras de cada enfermedad. De esta tabla se escogieron de manera general, aquellas funciones con mayor frecuencia en las muestras, la tabla completa de las funciones se encuentra en el **Anexo 30**, a continuación, se presentan las funciones con mayor frecuencia en cada enfermedad.

Tabla 22.

Frecuencia relativa de las funciones biológicas de los taxones en cada enfermedad

Enfermedad Celíaca		Crohn	
Función	Frecuencia	Función	Frecuencia
K00059	5,3687E+15	K01990	660,1
K08884	4,4158E+15	K06147	597,5
K04487	4,0275E+15	K01992	554,6
K03574	3,6275E+15	K03088	513
K06213	3,5628E+15	K02003	450
K11189	3,4511E+15	K02004	426,3
K00558	3,3805E+15	K02529	349,8
K03657	2,4726E+15	K07024	344,7
K03499	2,3726E+15	K00059	310,7
K06889	2,3294E+15	K03091	272,7

Nota: En la tabla se enlistan las funciones con mayor frecuencia de las muestras de cada enfermedad.

Con estos datos se obtuvo un heatmap donde se identificaron las funciones más significativas en el conjunto. El código R se ha descrito en el **Anexo 31**. Los heatmaps de cada enfermedad se muestran a continuación:

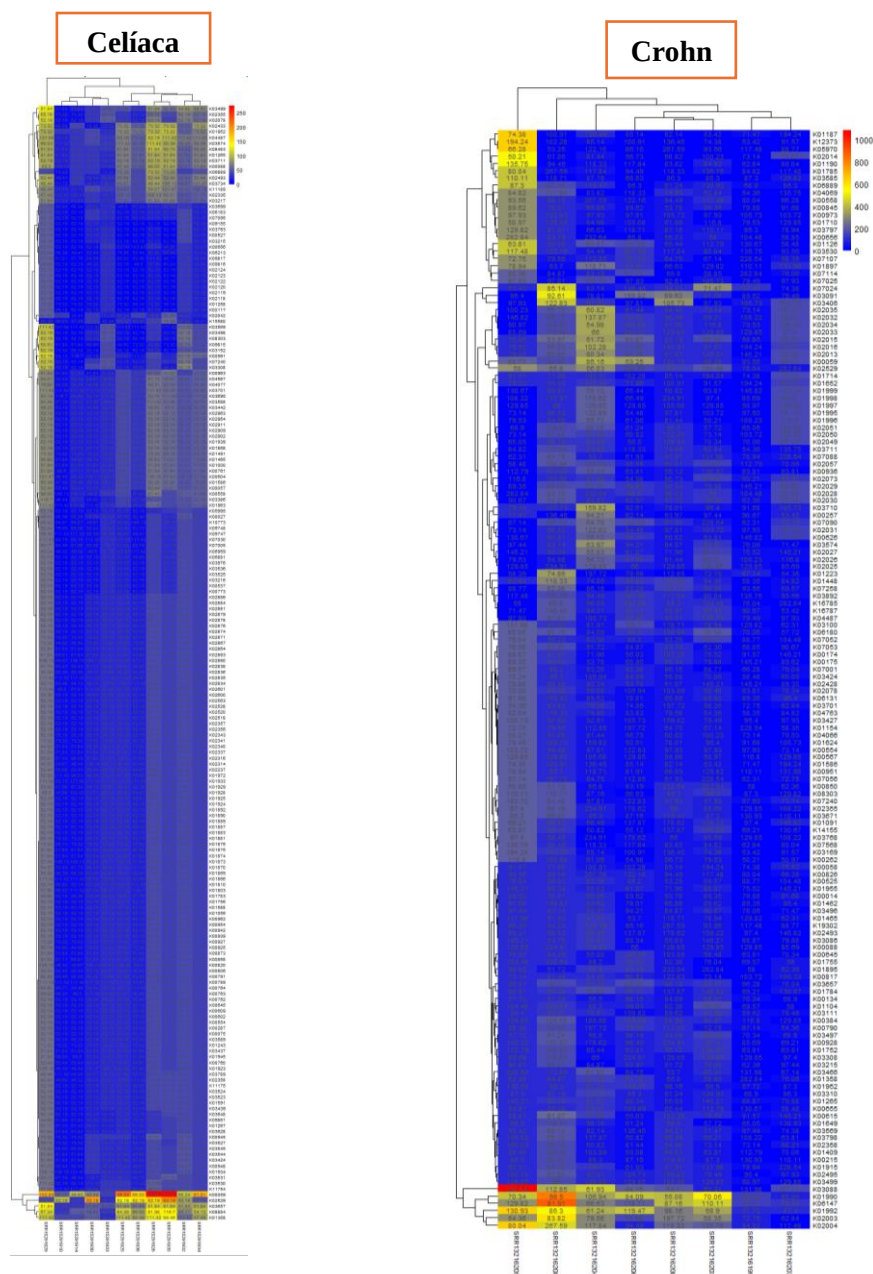


Figura 11. Rutas metabólicas más representativas de la EC y ECr.

Nota. Para el mapa de calor se utiliza la correlación de Pearson, a partir de esto se puede identificar distintas rutas metabólicas que son más representativas en cada una de las enfermedades.

Las funciones más representativas se representan en la siguiente tabla.

Tabla 23.

Funciones significativas de ambas enfermedades.

Enfermedad Celíaca		Enfermedad de Crohn	
Código	Funciones	Código	Funciones
K00059	Biosíntesis de ácidos grasos	K01990	Proteína de unión a ATP del sistema de transporte tipo ABC-2.
K02529	Regulador transcripcional de la familia LacI, represor del operón galactosa.	K06147	Casete de unión a ATP, subfamilia B, bacteriano (ABCB-BAC).
K03657	Helicasa de ADN dependiente de ATP UvrD/PcrA.	K01992	Proteína permeasa del sistema de transporte de tipo ABC-2.
K08884	Proteína quinasa de serina/treonina, bacteriana.	K02003	Proteína de unión a ATP del supuesto sistema de transporte ABC.
		K02004	Proteína permeasa del sistema de transporte ABC putativa.

Nota: Estos códigos se visualizaron en la base de datos KEGG para identificar las funciones a las que están relacionadas.

En la **Tabla 23** se puede observar las funciones más significativas de cada una de las enfermedades (EC y ECr) con sus respectivas funciones direccionadas a la incidencia de la enfermedad. En la EC se presentaron funciones relacionadas con rutas metabólicas de la síntesis de ácidos grasos y saturados. Esto se relaciona mucho con la inducción en la inflamación colónica, esto aumenta la expresión de citoquinas proinflamatorias y la permeabilidad intestinal, lo que desencadena en un mayor riesgo de desarrollar la EC o agravar el estado de la inflamación (**Escaffi et al., 2021**).

En cuanto a la ECr, se muestran en su mayoría funciones biológicas relacionadas con proteínas de transporte ABC. De acuerdo con el estudio realizado por **Yuan et al., (2022)**, estas proteínas están en el grupo denominado CuDEG, y juegan un papel importante en el desarrollo del estrés oxidativo, lo que podría verse reflejado en el posible desarrollo de ECr o agravamiento de las condiciones de la enfermedad aumentando la inflamación intestinal.

Una vez analizado el aspecto funcional de cada una de las enfermedades se puede determinar de una forma adecuada los biomarcadores para cada una de ellas, esto se observa en la **Figura 10** en la que se presentan patrones idénticos entre la ECr y EC, lo que sugiere que existen posibles biomarcadores de disbiosis e inflamación intestinal, tales como: *Faecalibacterium* *Fusobacterium* y *Lachnospiraceae*. El resto de los microorganismos se podrían considerar como biomarcadores particulares de cada enfermedad, estos podrían servir como diagnóstico y tratamientos especializados dirigidos a cada paciente con la finalidad de reconstituir la flora intestinal normal.

Tabla 24.

Biomarcadores individuales de cada enfermedad.

Celiaca	Crohn
<i>Microbacterium</i>	<i>Alistipes</i>
<i>Lachnoanaerobaculum</i>	<i>Curvibacter</i>
<i>Rhodococcus</i>	<i>Limnochordia</i>
<i>Parvimonas</i>	<i>Nocardiopsis</i>
<i>Leptotrichia</i>	<i>Melghirimyces</i>
	<i>Atopostipes</i>
	<i>Achromobacter</i>
	<i>Erysipelatoclostridiaceae</i>
	<i>Gemella</i>
	<i>Oceanobacillus</i>

Nota. Biomarcadores específicos para cada enfermedad.

CAPÍTULO IV.- CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones.

En el presente estudio se evaluó *in silico* los microbiomas intestinales, se identificaron posibles biomarcadores microbianos específicos asociados a la ECr y EC. Estos hallazgos contribuyen al entendimiento de la composición y función diferencial del microbioma en estas patologías, así como también ofrecen un potencial significativo de aplicación en el diagnóstico y tratamiento personalizado de estas enfermedades.

Se clasificó con éxito las muestras representativas del microbioma intestinal de pacientes con ECr y EC. Este proceso aseguró la validez y representatividad del análisis, destacándose la importancia de trabajar con datos curados. Sin embargo, se identificaron desafíos asociados a la heterogeneidad de las muestras, los cuales fueron mitigados mediante la aplicación de criterios estrictos de calidad. Este enfoque aseguró que se efectúe un análisis robusto y confiable de las características del microbioma en estas enfermedades.

El análisis metagenómico permitió evaluar las diferencias significativas en la diversidad microbiana presente en la EC y ECr. La disbiosis microbiana se identificó como una de las principales causas de estas diferencias. Además, se exploró la influencia del sexo en la microbiota para ambas patologías. El análisis de diversidad Alfa arrojó p-valores mayores a 0,05 tanto para EC y ECr como se describe en la **Tabla 20**, lo que demuestra que la composición microbiana dentro de las muestras no varió significativamente entre sexos. Además, en el análisis de diversidad Beta (**Tabla 21**) se obtuvieron resultados similares, en la EC todas las pruebas mostraron valores mayores a 0,05, sin embargo, los resultados mostraron diferencias significativas en ECr para todas las pruebas exceptuando Bray Curtis y Weighted-Unifrac donde se obtuvieron p-valores menores a 0,05 (0,005 y 0,03), esto sugiere que la abundancia y las relaciones filogenéticas de los microorganismos varían según el sexo.

Los perfiles metagenómicos de la EC y la ECr permitieron identificar biomarcadores microbianos específicos para cada una de ellas. Entre los microorganismos destacados tenemos *Faecalibacterium*, *Fusobacterium* y *Lachnospiraceae* los cuales están presentes en ambas patologías, posiblemente debido a su amplia distribución en la microbiota intestinal. Estos microorganismos podrían considerarse como

biomarcadores generales de disbiosis e inflamación. Por otro lado, los microorganismos descritos en la **Tabla 24** se consideran como biomarcadores específicos para el diagnóstico de cada enfermedad. Además, estos biomarcadores pueden servir para aplicaciones clínicas potenciales como son las terapias dirigidas, enfocadas en la recomposición del microbioma intestinal de los pacientes. Los resultados obtenidos confirman la utilidad del análisis metagenómico para explorar la relación entre microbiota intestinal y EII, destacando su relevancia en la identificación de biomarcadores y aplicación en el ámbito de la salud.

4.2. Recomendaciones

- Se debería considerar aumentar el tamaño de la muestra y diversificar las cohortes estudiadas (incluyendo más regiones geográficas o diferentes condiciones clínicas) para validar y generalizar los biomarcadores microbianos identificados. Incluir muestras adicionales de individuos sanos como controles robustos para fortalecer las comparaciones estadísticas y biológicas.
- Complementar el análisis metagenómico con datos de transcriptómica, metabolómica o proteómica para obtener una comprensión más integral de las funciones biológicas asociadas a los microorganismos identificados como biomarcadores, así como también evaluar las correlaciones funcionales entre la microbiota y los perfiles metabólicos de los pacientes.
- Investigar la sensibilidad y especificidad de los biomarcadores microbianos identificados para su aplicabilidad en pruebas diagnósticas explorando estrategias terapéuticas basadas en la recomposición del microbioma, trasplantes de microbiota fecal o probióticos dirigidos.
- Realizar estudios experimentales en modelos animales o cohortes humanas que validen la utilidad de los biomarcadores propuestos en el diagnóstico temprano de las EC y ECr.

REFERENCIAS BIBLIOGRÁFICAS

- Abouelkhair M. (2020). Non-SARS-CoV-2 genome sequences identified in clinical samples from COVID-19 infected patients: Evidence for co-infections. *PeerJ*, 8, e10246. <https://doi.org/10.7717/peerj.10246>
- Aljada, B., Zohni, A., & El-Matary, W. (2021). The Gluten-Free Diet for Celiac Disease and Beyond. *Nutrients*, 13(11), 3993. <https://doi.org/10.3390/nu13113993>
- Aljarrah, D., Chalour, N., Zorgani, A., Nissan, T., & Pranjol, M. Z. I. (2024). Exploring the gut microbiota and its potential as a biomarker in gliomas. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 173, 116420. <https://doi.org/10.1016/j.biopha.2024.116420>
- Agus, A., Clément, K., & Sokol, H. (2021). Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut*, 70(6), 1174–1182. <https://doi.org/10.1136/gutjnl-2020-323071>
- Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., ... & Das, B. (2016). An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Scientific reports*, 6(1), 26775.
- Beaudry, M., Wang, J., Kieran, T., Thomas, J., Bayona N., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, J., Rhodes, O., & Glenn, T. (2021). Improved Microbial Community Characterization of 16S rRNA via Metagenome Hybridization Capture Enrichment. *Frontiers in Microbiology*, 12, 644662. <https://doi.org/10.3389/FMICB.2021.644662/BIBTEX>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E. K., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology*, 37(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L., & Leddy, M. B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLOS ONE*, 15(2), e0228899. <https://doi.org/10.1371/JOURNAL.PONE.0228899>

- Caio, G., Volta, U., Sapone, A., Leffler, D. A., De Giorgio, R., Catassi, C., & Fasano, A. (2019). Celiac disease: a comprehensive current review. *BMC medicine*, 17(1), 142. <https://doi.org/10.1186/s12916-019-1380-z>
- Caio, G., Lungaro, L., Caputo, F., Zoli, E., Giancola, F., Chiarioni, G., De Giorgio, R., & Zoli, G. (2021). Nutritional Treatment in Crohn's Disease. *Nutrients*, 13(5), 1628. <https://doi.org/10.3390/nu13051628>
- Cushing, K., & Higgins, P. (2021). Management of Crohn Disease: A Review. *JAMA*, 325(1), 69–80. <https://doi.org/10.1001/jama.2020.18936>
- ChunLab, Inc. (2024). *About EzBioCloud*. EzBioCloud Help Center. <https://help.ezbiocloud.net/about-ezbiocloud-net/>
- de Vos, W., Tilg, H., Van Hul, M., & Cani, P. (2022). Gut microbiome and health: mechanistic insights. *Gut*, 71(5), 1020–1032. <https://doi.org/10.1136/gutjnl-2021-326789>
- Douglas, G., Maffei, V., Zaneveld, J., Yurgel, S., Brown, J., Taylor, C., Huttenhower, C., & Langille, M. (2020). PICRUSt2 for prediction of metagenome functions. *Nature biotechnology*, 38(6), 685–688. <https://doi.org/10.1038/s41587-020-0548-6>
- Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific reports*, 11(1), 3030.
- Duttagupta, S., Hakozaki, T., Routy, B., & Messaoudene, M. (2023). The Gut Microbiome from a Biomarker to a Novel Therapeutic Strategy for Immunotherapy Response in Patients with Lung Cancer. *Current oncology (Toronto, Ont.)*, 30(11), 9406–9427. <https://doi.org/10.3390/curroncol30110681>
- Edgar, R. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv (Cold Spring Harbor Laboratory). <https://doi.org/10.1101/081257>
- Estaki, M., Jiang, L., Bokulich, N. A., McDonald, D., González, A., Kosciulek, T., Martino, C., Zhu, Q., Birmingham, A., Vázquez-Baeza, Y., Dillon, M. R., Bolyen, E., Caporaso, J. G., & Knight, R. (2020). QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. *Current protocols in bioinformatics*, 70(1), e100. <https://doi.org/10.1002/cpbi.100>

- Gao, J., Zhao, X., Hu, S., Huang, Z., Hu, M., Jin, S., Lu, B., Sun, K., Wang, Z., Fu, J., Weersma, R. K., He, X., & Zhou, H. (2022). Gut microbial DL-endopeptidase alleviates Crohn's disease via the NOD2 pathway. *Cell host & microbe*, 30(10), 1435–1449.e9. <https://doi.org/10.1016/j.chom.2022.08.002>
- Giuffrè, M., Gazzin, S., Zoratti, C., Llido, J. P., Lanza, G., Tiribelli, C., & Moretti, R. (2022). Celiac Disease and Neurological Manifestations: From Gluten to Neuroinflammation. *International journal of molecular sciences*, 23(24), 15564. <https://doi.org/10.3390/ijms232415564>
- Gordon, H., Minozzi, S., Kopylov, U., Verstockt, B., Chaparro, M., Buskens, C., Warusavitarne, J., Agrawal, M., Allocca, M., Atreya, R., Battat, R., Bettenworth, D., Bislenghi, G., Brown, S. R., Burisch, J., Casanova, M. J., Czuber-Dochan, W., de Groof, J., El-Hussuna, A., Ellul, P., ... Raine, T. (2024). ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment. *Journal of Crohn's & colitis*, jjae091. Advance online publication. <https://doi.org/10.1093/ecco-jcc/jjae091>
- Gupta, S., Mortensen, M. S., Schjørring, S., Trivedi, U., Vestergaard, G., Stokholm, J., Bisgaard, H., Krogfelt, K. A., & Sørensen, S. J. (2019). Amplicon sequencing 21 provides more accurate microbiome information in healthy children compared to culturing. *Communications Biology* 2019 2:1, 2(1), 1–7. <https://doi.org/10.1038/s42003-019-0540-1>
- Gnodi, E., Meneveri, R., & Barisani, D. (2022). Celiac disease: From genetics to epigenetics. *World journal of gastroenterology*, 28(4), 449–463. <https://doi.org/10.3748/wjg.v28.i4.449>
- Hu, X., Li, H., Zhao, X., Zhou, R., Liu, H., Sun, Y., Fan, Y., Shi, Y., Qiao, S., Liu, S., Liu, H., & Zhang, S. (2021). Multi-omics study reveals that statin therapy is associated with restoration of gut microbiota homeostasis and improvement in outcomes in patients with acute coronary syndrome. *Theranostics*, 11(12), 5778–5793. <https://doi.org/10.7150/thno.55946>
- Hu, M., Caldarelli, G., & Gili, T. (2023). Inflammatory bowel disease biomarkers revealed by the human gut microbiome network. *Scientific reports*, 13(1), 19428. <https://doi.org/10.1038/s41598-023-46184-y>
- Illumina. (2024). Quality scores for next-generation sequencing. Illumina. <https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html>
- Icaza, M. (2013). *Microbiota intestinal en la salud y la enfermedad*. *Revista de Gastroenterología de México*, 78(4), 240–248. doi:10.1016/j.rgm.2013.04.004

- Jarmakiewicz, S., Zielińska, M., Sokal, A., & Filip, R. (2022). Genetic and Epigenetic Etiology of Inflammatory Bowel Disease: An Update. *Genes*, 13(12), 2388. <https://doi.org/10.3390/genes13122388>
- Jukic, A., Bakiri, L., Wagner, E. F., Tilg, H., & Adolph, T. E. (2021). Calprotectin: from biomarker to biological function. *Gut*, 70(10), 1978–1988. <https://doi.org/10.1136/gutjnl-2021-324855>
- Keefe, C. R., Dillon, M. R., Gehret, E., Herman, C., Jewell, M., Wood, C. V., Bolyen, E., & Caporaso, J. G. (2023). Facilitating bioinformatics reproducibility with QIIME 2 Provenance Replay. *PLoS computational biology*, 19(11), e1011676. <https://doi.org/10.1371/journal.pcbi.1011676>
- Kioroglou, D., Mas, A., & Portillo, M. D. C. (2019). Evaluating the Effect of QIIME Balanced Default Parameters on Metataxonomic Analysis Workflows With a Mock Community. *Frontiers in microbiology*, 10, 1084. <https://doi.org/10.3389/fmicb.2019.01084>
- Li, M., Yang, L., Mu, C., Sun, Y., Gu, Y., Chen, D., Liu, T., & Cao, H. (2022). Gut microbial metabolome in inflammatory bowel disease: From association to therapeutic perspectives. *Computational and structural biotechnology journal*, 20, 2402–2414. <https://doi.org/10.1016/j.csbj.2022.03.038>
- Lebwohl, B., & Rubio-Tapia, A. (2021). Epidemiology, Presentation, and Diagnosis of Celiac Disease. *Gastroenterology*, 160(1), 63–75. <https://doi.org/10.1053/j.gastro.2020.06.098>
- Lee, J. W. J., Plichta, D., Hogstrom, L., Borren, N. Z., Lau, H., Gregory, S. M., Tan, W., Khalili, H., Clish, C., Vlamakis, H., Xavier, R. J., & Ananthakrishnan, A. N. (2021). Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. *Cell host & microbe*, 29(8), 1294–1304.e4. <https://doi.org/10.1016/j.chom.2021.06.019>
- Machado M. (2023). New Developments in Celiac Disease Treatment. *International journal of molecular sciences*, 24(2), 945. <https://doi.org/10.3390/ijms24020945>
- Mak, W., Zhao, M., Ng, S., & Burisch, J. (2020). The epidemiology of inflammatory bowel disease: East meets west. *Journal of Gastroenterology and Hepatology*, 35(3), 380–389.
- Quaglio, A., Grillo, T., De Oliveira, E., Di Stasi, L., & Sassaki, L. (2022). Gut microbiota, inflammatory bowel disease and colorectal cancer. *World journal of gastroenterology*, 28(30), 4053–4060. <https://doi.org/10.3748/wjg.v28.i30.4053>

- Odom, A., Faits, T., Castro, E., Crandall, K., & Johnson, W. (2023). Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. *Scientific reports*, 13(1), 13957. <https://doi.org/10.1038/s41598-023-40799-x>
- Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., ... & Manichanh, C. (2017). A microbial signature for Crohn's disease. *Gut*, 66(5), 813-822.
- Petagna, L., Antonelli, A., Ganini, C., Bellato, V., Campanelli, M., Divizia, A., Efrati, C., Franceschilli, M., Guida, A., Ingallinella, S., Montagnese, F., Sensi, B., Siragusa, L., & Sica, G. (2020). Pathophysiology of Crohn's disease inflammation and recurrence. *Biology direct*, 15(1), 23. <https://doi.org/10.1186/s13062-020-00280-5>
- Ramírez, A., Tan, I., Gonera, B., Visschedijk, M., Jonkers, I., & Withoff, S. (2020). Molecular Biomarkers for Celiac Disease: Past, Present and Future. *International journal of molecular sciences*, 21(22), 8528. <https://doi.org/10.3390/ijms21228528>
- Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., Joukov, V., Lotov, V., Pannu, R., Rudnev, D., Shkeda, A., Weitz, E. M., & Schneider, V. A. (2021). Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV). *Genome research*, 31(1), 159–169. <https://doi.org/10.1101/gr.266932.120>
- Ricci, F., Russo, E., Renzi, D., Baldi, S., Nannini, G., Lami, G., & Calabrò, A. (2022). Characterization of the “gut microbiota-immunity axis” and microbial lipid metabolites in atrophic and potential celiac disease. *Frontiers in Microbiology*, 13, 886008.
- Rodríguez, L., Vaquero, L., & Vivas, S. (2020). New celiac disease biomarkers. *Revista española de enfermedades digestivas*, 112(10), 792–796. <https://doi.org/10.17235/reed.2020.7217/2020>
- Roszkowska, P., Klimczak, E., Ostrycharz, E., Rączka, A., Wojciechowska-Koszko, I., Dybus, A., Cheng, Y. H., Yu, Y. H., Mazgaj, S., & Hukowska-Szematowicz, B. (2024). Small Intestinal Bacterial Overgrowth (SIBO) and Twelve Groups of Related Diseases-Current State of Knowledge. *Biomedicines*, 12(5), 1030. <https://doi.org/10.3390/biomedicines12051030>
- Russo, E., Giudici, F., Ricci, F., Scaringi, S., Nannini, G., Ficari, F., & Amedei, A. (2021). Diving into inflammation: a pilot study exploring the dynamics of the immune–microbiota axis in ileal tissue layers of patients with Crohn’s disease. *Journal of Crohn's and Colitis*, 15(9), 1500-1516.
- Russo E. (2022). The gut microbiota as a biomarker in epilepsy. *Neurobiology of disease*, 163, 105598. <https://doi.org/10.1016/j.nbd.2021.105598>

- Sánchez, S., Comas, O., Duelo, A., Veciana, M., Berlanga, M., Latorre, M., & Vidal, M. (2022). Intestinal Dysbiosis in Patients with Histamine Intolerance. *Nutrients*, 14(9), 1774. <https://doi.org/10.3390/nu14091774>
- Sánchez-Pérez, S., Comas-Basté, O., Duelo, A., Veciana-Nogués, M. T., Berlanga, M., Vidal-Carou, M. C., & Latorre-Moratalla, M. L. (2022). The dietary treatment of histamine intolerance reduces the abundance of some histamine-secreting bacteria of the gut microbiota in histamine intolerant women. A pilot study. *Frontiers in nutrition*, 9, 1018463. <https://doi.org/10.3389/fnut.2022.1018463>
- Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Vösa, U., Mujagic, Z., Masclee, A. A. M., Jonkers, D. M. A. E., Oosting, M., Joosten, L. A. B., Netea, M. G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., & McCarthy, M. I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature genetics*, 51(4), 600–605. <https://doi.org/10.1038/s41588-019-0350-x>
- Seyedian, S., Nokhostin, F., & Malamir, M. (2019). A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. *Journal of medicine and life*, 12(2), 113. doi: 10.25122/jml-2018-0075. PMID: 31406511; PMCID: PMC6685307.
- Silva, F., Gatica, T., & Pavez, C. (2019). Etiología y fisiopatología de la enfermedad inflamatoria intestinal. *Revista Médica Clínica Las Condes*, 30(4), 262-272.
- Silvester, J. A., Therrien, A., & Kelly, C. P. (2021). Celiac Disease: Fallacies and Facts. *The American journal of gastroenterology*, 116(6), 1148–1155. <https://doi.org/10.14309/ajg.0000000000001218>
- Skoufou, M., Tsigalou, C., Vradelis, S., & Bezirtzoglou, E. (2024). The Networked Interaction between Probiotics and Intestine in Health and Disease: A Promising Success Story. *Microorganisms*, 12(1), 194. <https://doi.org/10.3390/microorganisms12010194>
- Stoppe, C., Wendt, S., Mehta, N. M., Compher, C., Preiser, J. C., Heyland, D. K., & Kristof, A. S. (2020). Biomarkers in critical care nutrition. *Critical care (London, England)*, 24(1), 499. <https://doi.org/10.1186/s13054-020-03208-7>
- Tango, C., Seo, S., Kwon, M., Lee, D., Chang, H., & Kim, M. (2020). Taxonomic and Functional Differences in Cervical Microbiome Associated with Cervical Cancer Development. *Scientific reports*, 10(1), 9720. <https://doi.org/10.1038/s41598-020-66607-4>

- Taş, N., de Jong, A., Li, Y., Trubl, G., Xue, Y., & Dove, N. C. (2021). Metagenomic tools in microbial ecology research. *Current Opinion in Biotechnology*, 67, 184-191.
- Torres, J., Bonovas, S., Doherty, G., Kucharzik, T., Gisbert, J. P., Raine, T., Adamina, M., Armuzzi, A., Bachmann, O., Bager, P., Biancone, L., Bokemeyer, B., Bossuyt, P., Burisch, J., Collins, P., El-Hussuna, A., Ellul, P., Frei-Lanter, C., Furfaro, F., Gingert, C., ... Fiorino, G. (2020). ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment. *Journal of Crohn's & colitis*, 14(1), 4–22. <https://doi.org/10.1093/ecco-jcc/jjz180>
- Therrien, A., Kelly, C. P., & Silvester, J. A. (2020). Celiac Disease: Extraintestinal Manifestations and Associated Conditions. *Journal of clinical gastroenterology*, 54(1), 8–21. <https://doi.org/10.1097/MCG.0000000000001267>
- Thompson, L. R., Anderson, S. R., Den Uyl, P. A., Patin, N. V., Lim, S. J., Sanderson, G., & Goodwin, K. D. (2022). Tourmaline: A containerized workflow for rapid and iterable amplicon sequence analysis using QIIME 2 and Snakemake. *GigaScience*, 11, giac066. <https://doi.org/10.1093/gigascience/giac066>
- Vester, M., Mirsepasi, H., Prosberg, M., Mortensen, C., Träger, C., Skovsen, K., Thorkilgaard, T., Nøjgaard, C., Vind, I., Krogfelt, K., Sørensen, N., Bendtsen, F., & Petersen, A. (2019). Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Scientific reports*, 9(1), 13473. <https://doi.org/10.1038/s41598-019-49833-3>
- Wang, W. L., Xu, S. Y., Ren, Z. G., Tao, L., Jiang, J. W., & Zheng, S. S. (2015). Application of metagenomics in the human gut microbiome. *World journal of gastroenterology: WJG*, 21(3), 803–814. <https://doi.org/10.3748/wjg.v21.i3.803>
- Wei, G., Helmerhorst, E. J., Darwish, G., Blumenkranz, G., & Schuppan, D. (2020). Gluten Degrading Enzymes for Treatment of Celiac Disease. *Nutrients*, 12(7), 2095. <https://doi.org/10.3390/nu12072095>
- Wieser, H., Ciacci, C., Soldaini, C., Gizzi, C., & Santonicola, A. (2024). Gastrointestinal and Hepatobiliary Manifestations Associated with Untreated Celiac Disease in Adults and Children: A Narrative Overview. *Journal of clinical medicine*, 13(15), 4579. <https://doi.org/10.3390/jcm13154579>
- Wood, L., DeBoer, M., & Fasano, A. (2020). Zonulin as a potential putative biomarker of risk for shared type 1 diabetes and celiac disease autoimmunity. *Diabetes/metabolism research and reviews*, 36(5), e3309. <https://doi.org/10.1002/dmrr.3309>

- Xia, Y., & Sun, J. (2023). Beta Diversity Metrics and Ordination. In *Bioinformatic and Statistical Analysis of Microbiome Data: From Raw Sequences to Advanced Modeling with QIIME 2 and R* (pp. 335-395). Cham: Springer International Publishing.
- Yang, C., Mai, J., Cao, X., Burberry, A., Cominelli, F., & Zhang, L. (2023). ggpicrust2: an R package for PICRUST2 predicted functional profile analysis and visualization. *Bioinformatics (Oxford, England)*, 39(8), btad470. <https://doi.org/10.1093/bioinformatics/btad470>
- Yoon, S., Ha, S., Kwon, S., Lim, J., Kim, Y., Seo, H., & Chun, J. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International journal of systematic and evolutionary microbiology*, 67(5), 1613–1617. <https://doi.org/10.1099/ijsem.0.001755>
- Yu, X., Jiang, W., Kosik, R. O., Song, Y., Luo, Q., Qiao, T., Tong, J., Liu, S., Deng, C., Qin, S., Lv, Z., & Li, D. (2021). Gut microbiota changes and its potential relations with thyroid carcinoma. *Journal of advanced research*, 35, 61–70. <https://doi.org/10.1016/j.jare.2021.04.001>
- Yuan, Y., Fu, M., Li, N., & Ye, M. (2022). Identification of immune infiltration and cuproptosis-related subgroups in Crohn's disease. *Frontiers in immunology*, 13, 1074271. <https://doi.org/10.3389/fimmu.2022.1074271>
- Zhang, J., Vootukuru, N., & Niewiadomski, O. (2024). The effect of solid food diet therapies on the induction and maintenance of remission in Crohn's disease: a systematic review. *BMC gastroenterology*, 24(1), 250. <https://doi.org/10.1186/s12876-024-03315-7>
- Zhao, X., Zhao, H., Liu, Y., & Guo, Z. (2023). Methods, bioinformatics tools and databases in ecDNA research: An overview. *Computers in biology and medicine*, 167, 107680. <https://doi.org/10.1016/j.compbimed.2023.107680>
- Zheng, C., Lu, F., Chen, B., Yang, J., Yu, H., Wang, D., Xie, H., Chen, K., Xie, Y., Li, J., Bo, Z., Wang, Y., Chen, G., & Deng, T. (2023). Gut microbiome as a biomarker for predicting early recurrence of HBV-related hepatocellular carcinoma. *Cancer science*, 114(12), 4717–4731. <https://doi.org/10.1111/cas.15983>
- Escaffi, M., Navia, C., Quera, R., & Simian, D. (2021). *Nutrición y enfermedad inflamatoria intestinal: posibles mecanismos en la incidencia y manejo*. *Revista Médica Clínica Las Condes*, 32(4), 491–501. doi:10.1016/j.rmcl.2021.01.013

ANEXOS

Anexo 1. Script para instalar SRA Toolkit.

```
mkdir SRA_Toolkit
cd SRA_Toolkit
ls
wget --output-document sratoolkit.tar.gz https://ftp-
trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
tar -vzxvf sratoolkit.tar.gz
export PATH=$PATH:$PWD/sratoolkit.3.0.0-mac64/bin
which fastq-dump
vdb-config -i
```

Anexo 2. Script para la descarga de las secuencias.

```
prefetch SRR(seguido del número de la secuencia)
ls
ls SRR(seguido del número de la secuencia)/
fasterq-dump SRR(seguido del número de la secuencia)
ls
```

Anexo 3. Script para la instalación de FASTQC.

```
sudo apt update
sudo apt install fastqc
fastqc --version
```

Anexo 4. Script para la instalación de Trimmomatic.

```
wget Trimmomatic-0.39.zip
unzip Trimmomatic-0.39.zip
cd Trimmomatic-0.39.zip
java -jar trimmomatic-0.39.jar
```

Anexo 5. Enlace para los códigos de los archivos .bsh para el corte en bucle de Trimmomatic.

https://utaedu-my.sharepoint.com/:f/g/personal/ilucero6068_uta_edu_ec/En30OuwOVvJHhWj2IwnV0KkBWt-QgPvN1DGsHlBbm22x6g?e=5VKa2D

Anexo 6. Archivo e índice para visualizar los fastqc.

[qcvisualization.html](#) (ctrl + click - para poder visualizar)

Enfermedad	Raw Data		QC Data	
	Forward	Reverse	Forward	Reverse
Celiaca	SRR15291910 raw forward	SRR15291910 raw reverse	SRR15291910_qc forward	SRR15291910_qc reverse
	SRR15291911 raw forward	SRR15291911 raw reverse	SRR15291911_qc forward	SRR15291911_qc reverse
	SRR15291912 raw forward	SRR15291912 raw reverse	SRR15291912_qc forward	SRR15291912_qc reverse
	SRR15291913 raw forward	SRR15291913 raw reverse	SRR15291913_qc forward	SRR15291913_qc reverse
	SRR15291914 raw forward	SRR15291914 raw reverse	SRR15291914_qc forward	SRR15291914_qc reverse
	SRR15291925 raw forward	SRR15291925 raw reverse	SRR15291925_qc forward	SRR15291925_qc reverse
	SRR15291926 raw forward	SRR15291926 raw reverse	SRR15291926_qc forward	SRR15291926_qc reverse
	SRR15291927 raw forward	SRR15291927 raw reverse	SRR15291927_qc forward	SRR15291927_qc reverse
	SRR15291928 raw forward	SRR15291928 raw reverse	SRR15291928_qc forward	SRR15291928_qc reverse
	SRR15291929 raw forward	SRR15291929 raw reverse	SRR15291929_qc forward	SRR15291929_qc reverse
	SRR15291930 raw forward	SRR15291930 raw reverse	SRR15291930_qc forward	SRR15291930_qc reverse
	SRR15291931 raw forward	SRR15291931 raw reverse	SRR15291931_qc forward	SRR15291931_qc reverse
	SRR15291932 raw forward	SRR15291932 raw reverse	SRR15291932_qc forward	SRR15291932_qc reverse
	SRR15291933 raw forward	SRR15291933 raw reverse	SRR15291933_qc forward	SRR15291933_qc reverse
	SRR15291934 raw forward	SRR15291934 raw reverse	SRR15291934_qc forward	SRR15291934_qc reverse
	SRR15291935 raw forward	SRR15291935 raw reverse	SRR15291935_qc forward	SRR15291935_qc reverse
	SRR15291936 raw forward	SRR15291936 raw reverse	SRR15291936_qc forward	SRR15291936_qc reverse
Control	SRR13216189 raw forward	SRR13216189 raw reverse	SRR13216189_qc forward	SRR13216189_qc reverse
	SRR13216190 raw forward	SRR13216190 raw reverse	SRR13216190_qc forward	SRR13216190_qc reverse
	SRR13216191 raw forward	SRR13216191 raw reverse	SRR13216191_qc forward	SRR13216191_qc reverse
	SRR13216192 raw forward	SRR13216192 raw reverse	SRR13216192_qc forward	SRR13216192_qc reverse
	SRR13216193 raw forward	SRR13216193 raw reverse	SRR13216193_qc forward	SRR13216193_qc reverse
	SRR13216194 raw forward	SRR13216194 raw reverse	SRR13216194_qc forward	SRR13216194_qc reverse
	SRR13216195 raw forward	SRR13216195 raw reverse	SRR13216195_qc forward	SRR13216195_qc reverse
	SRR13216196 raw forward	SRR13216196 raw reverse	SRR13216196_qc forward	SRR13216196_qc reverse
	SRR13216197 raw forward	SRR13216197 raw reverse	SRR13216197_qc forward	SRR13216197_qc reverse
	SRR13216198 raw forward	SRR13216198 raw reverse	SRR13216198_qc forward	SRR13216198_qc reverse
Crohn	SRR13216199 raw forward	SRR13216199 raw reverse	SRR13216199_qc forward	SRR13216199_qc reverse
	SRR13216200 raw forward	SRR13216200 raw reverse	SRR13216200_qc forward	SRR13216200_qc reverse
	SRR13216201 raw forward	SRR13216201 raw reverse	SRR13216201_qc forward	SRR13216201_qc reverse
	SRR13216202 raw forward	SRR13216202 raw reverse	SRR13216202_qc forward	SRR13216202_qc reverse
	SRR13216203 raw forward	SRR13216203 raw reverse	SRR13216203_qc forward	SRR13216203_qc reverse
	SRR13216204 raw forward	SRR13216204 raw reverse	SRR13216204_qc forward	SRR13216204_qc reverse
	SRR13216205 raw forward	SRR13216205 raw reverse	SRR13216205_qc forward	SRR13216205_qc reverse
	SRR13216206 raw forward	SRR13216206 raw reverse	SRR13216206_qc forward	SRR13216206_qc reverse
	SRR13216207 raw forward	SRR13216207 raw reverse	SRR13216207_qc forward	SRR13216207_qc reverse
	SRR13216208 raw forward	SRR13216208 raw reverse	SRR13216208_qc forward	SRR13216208_qc reverse

Anexo 7. Script para la importación de secuencias.

```
qiime tools import \  
  --type 'SampleData[PairedEndSequencesWithQuality]' \  
  --input-path manifest(enfermedad).tsv \  
  --output-path paired-end-demux.qza \  
  --input-format PairedEndFastqManifestPhred33V2
```

Anexo 8. Archivos manifest de cada una de las muestras y comparaciones del estudio.

https://utaedu-my.sharepoint.com/:x:/g/personal/ilucero6068_uta_edu_ec/ESHOQZUJe0lBidRHKMjEiewBTRLHs8PRzgR3t2JOSY2R-Q?e=0TTjYT

Anexo 9. Script para transformar el archivo paired-end-demux.qza a .qzv.

```
qiime demux summarize \  
  --i-data paired-end-demux.qza \  
  --o-visualization demux.qzv
```

Anexo 10. Pipeline usado en el control de calidad de las secuencias (triplicado por caso de estudio).

*** DADA2**

```
qiime dada2 denoise-paired \  
  --i-demultiplexed-seqs paired-end-demux.qza \  
  --p-trunc-len-f 280 \  
  --p-trunc-len-r 200 \  
  --p-trim-left-f 20 \  
  --p-trim-left-r 20 \  
  --o-table table.qza \  
  --o-representative-sequences rep-seqs.qza \  
  --o-denoising-stats denoising-stats.qza
```

*** Visualización tabla de características**

```
qiime feature-table summarize \  
  --i-table table.qza \  
  --o-visualization table.qzv \  
  --m-sample-metadata-file metadata.tsv
```

*** Visualización tabla de secuencias representativas**

```
qiime feature-table tabulate-seqs \  
  --i-data rep-seqs.qza \  
  --o-visualization rep-seqs.qzv
```

*** Estadísticas del proceso de denoising**

```
qiime metadata tabulate \  
  --m-input-file denoising-stats.qza \  
  --o-visualization denoising-stats.qzv
```

*** Reproceso de filtrado de las secuencias**

```
qiime feature-table filter-features \  
  --i-table table.qza \  
  --p-min-frequency 10 \  
  --o-filtered-table filtered-table.qza
```

*** Eliminación de las secuencias duplicadas**

```
qiime feature-table filter-seqs \  
  --i-data rep-seqs.qza \  
  --i-table filtered-table.qza \  
  --o-filtered-data filtered-rep-seqs.qza
```

Anexo 11. Script utilizado para el clustering.

```
qiime vsearch cluster-features-de-novo \  
  --i-table filtered-table.qza \  
  --i-sequences filtered-rep-seqs.qza \  
  --p-perc-identity 0.97 \  
  --output-dir cluster \  
  --verbose
```

Anexo 12. Script para la transformación de formato (.qza a .qzv) de archivos obtenidos en el clustering.

```
*qiime feature-table summarize \  
  --i-table cluster/clustered_table.qza \  
  --o-visualization cluster/clustered_table.qzv \  
  --m-sample-metadata-file metadata(enfermedad).tsv
```

```
*qiime feature-table tabulate-seqs \  
  --i-table cluster/clustered_table.qza
```

```
--i-data cluster/clustered_sequences.qza \
--o-visualization cluster/clustered_sequences.qzv
```

Anexo 13. Script utilizado para el resumen de clustered_table.qza.

```
qiime feature-table summarize \
  --i-table clustered_table.qza \
  --o-visualization table-summary.qzv
```

Anexo 14. Pipeline utilizado para el análisis de diversidad Alfa y Beta.

*** Calculo de la diversidad alfa y beta:**

```
qiime diversity core-metrics-phylogenetic \
  --i-phylogeny rooted-tree.qza \
  --i-table clustered_table.qza \
  --p-sampling-depth # \
  --m-metadata-file metadata_celiac.tsv \
  --output-dir core-metrics-results
```

***Transformar de .qza a .zv (Alfa)**

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/shannon_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/shannon_vector.qzv
```

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/evenness_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/evenness_vector.qzv
```

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/observed_features_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/observed_features_vector.qzv
```

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/faith_pd_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/faith_pd_vector.qzv
```

***Evaluación de significancia de diversidad Alfa**

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/faith_pd_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/faith_pd_significance.qzv
```

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/shannon_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/shannon_significance.qzv
```

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/evenness_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/evenness_significance.qzv
```

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/observed_features_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/observed_features_significance.qzv
```

***Para análisis de diversidad Alfa**

```
qiime diversity alpha-group-significance \
  --i-alpha-diversity core-metrics-results/faith_pd_vector.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/alpha-significance.qzv
```

***Para Alfa rarefacción**

```
qiime diversity alpha-rarefaction \
  --i-table table.qza \
  --i-phylogeny rooted-tree.qza \
  --p-max-depth <max-depth> \
  --m-metadata-file metadata(enfermedad).tsv \
  --o-visualization core-metrics-results/Alfa/alpha-rarefaction.qzv
```

***Evaluación de significancia de diversidad Beta:**

```
qiime diversity beta-group-significance \
  --i-distance-matrix core-metrics-results/unweighted_unifrac_distance_matrix.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --m-metadata-column GroupColumn \
  --o-visualization core-metrics-results/Beta/unweighted_unifrac_significance.qzv \
  --p-pairwise
```

```
qiime diversity beta-group-significance \
  --i-distance-matrix core-metrics-results/weighted_unifrac_distance_matrix.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --m-metadata-column GroupColumn \
  --o-visualization core-metrics-results/Beta/weighted_unifrac_significance.qzv \
  --p-pairwise
```

```
qiime diversity beta-group-significance \
  --i-distance-matrix core-metrics-results/jaccard_distance_matrix.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --m-metadata-column GroupColumn \
  --o-visualization core-metrics-results/Beta/jaccard_significance.qzv \
  --p-pairwise
```

```
qiime diversity beta-group-significance \
  --i-distance-matrix core-metrics-results/bray_curtis_distance_matrix.qza \
  --m-metadata-file metadata(enfermedad).tsv \
  --m-metadata-column sex \
  --o-visualization core-metrics-results/Beta/bray_curtis_significance.qzv \
```

--p-pairwise

***Para análisis de diversidad Beta**

```
qiime diversity beta-group-significance \  
  --i-distance-matrix core-metrics-results/unweighted_unifrac_distance_matrix.qza \  
  --m-metadata-file metadata(enfermedad).tsv \  
  --m-metadata-column sex \  
  --p-method permanova \  
  --o-visualization core-metrics-results/Beta/beta-significance.qzv
```

***Transformar de .qza a .zv (Beta)**

```
qiime emperor plot \  
  --i-pcoa core-metrics-results/unweighted_unifrac_pcoa_results.qza \  
  --m-metadata-file metadata(enfermedad).tsv \  
  --o-visualization core-metrics-results/Beta/unweighted_unifrac_emperor.qzv
```

```
qiime emperor plot \  
  --i-pcoa core-metrics-results/weighted_unifrac_pcoa_results.qza \  
  --m-metadata-file metadata(enfermedad).tsv \  
  --o-visualization core-metrics-results/Beta/weighted_unifrac_emperor.qzv
```

```
qiime emperor plot \  
  --i-pcoa core-metrics-results/jaccard_pcoa_results.qza \  
  --m-metadata-file metadata(enfermedad).tsv \  
  --o-visualization core-metrics-results/Beta/jaccard_emperor.qzv
```

```
qiime emperor plot \  
  --i-pcoa core-metrics-results/bray_curtis_pcoa_results.qza \  
  --m-metadata-file metadata(enfermedad).tsv \  
  --o-visualization core-metrics-results/Beta/bray_curtis_emperor.qzv
```

Anexo 15. Código para la descarga del clasificador SILVA.

wget <https://data.qiime2.org/2023.5/common/silva-138-99-nb-classifier.qza>

Anexo 16. Script para la asignación taxonómica.

```
qiime feature-classifier classify-sklearn \  
  --i-classifier silva-138-99-nb-classifier.qza \  
  --i-reads clustered_sequences.qza \  
  --o-classification taxonomy.qza
```

***Explotación para visualizar la taxonomía**

```
qiime metadata tabulate \  
  --m-input-file taxonomy.qza \  
  --o-visualization taxonomy.qzv
```

Anexo 17. Script utilizado para la generación del árbol filogenético.

```
qiime phylogeny align-to-tree-mafft-fasttree \  
  --i-sequences clustered_sequences.qza \  
  --o-alignment aligned-rep-seqs.qza \  
  --o-masked-alignment masked-aligned-rep-seqs.qza \  
  --o-tree unrooted-tree.qza \  
  --o-rooted-tree rooted-tree.qza
```

Anexo 18. Script para la exportación del árbol filogenético en archive .nwk.

```
qiime tools export \  
  --input-path rooted-tree.qza \  
  --output-path exported-tree
```

Anexo 19. Tabla de las características (table.qzv y rep-seqs.qzv) de la EC.

https://utaedu-my.sharepoint.com/:u:/g/personal/ilucero6068_uta_edu_ec/ERuymHkzCgJLmuoGQvb4zZkBVZ8pfWOjcloCZC78pqCWSA?e=fD1Nzu

Anexo 20. Tabla de las características (table.qzv y rep-seqs.qzv) del Control.

https://utaedu-my.sharepoint.com/:u:/g/personal/ilucero6068_uta_edu_ec/ERS1QLFyqP5Krqzwbm7RAIUBFtz_uTSHGow-1vC5qvifRg?e=lhGRIw

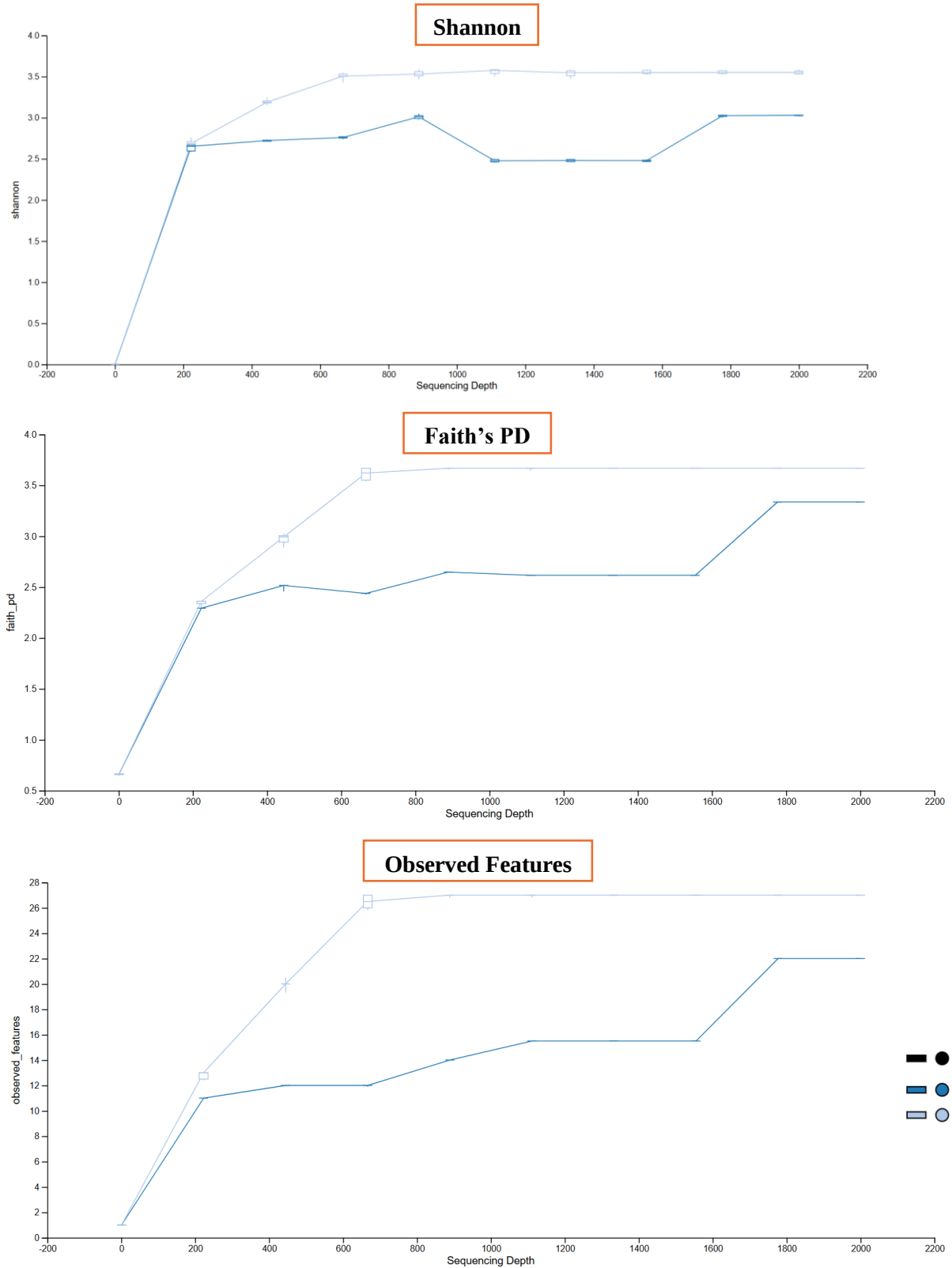
Anexo 21. Tabla de las características (table.qzv y rep-seqs.qzv) del Crohn.

https://utaedu-my.sharepoint.com/:u:/g/personal/ilucero6068_uta_edu_ec/EcxmhGwBf8dDquNDM1CHayEBnjEAwMUYYvtDGPXVpXCXUOw?e=Vp4UCg

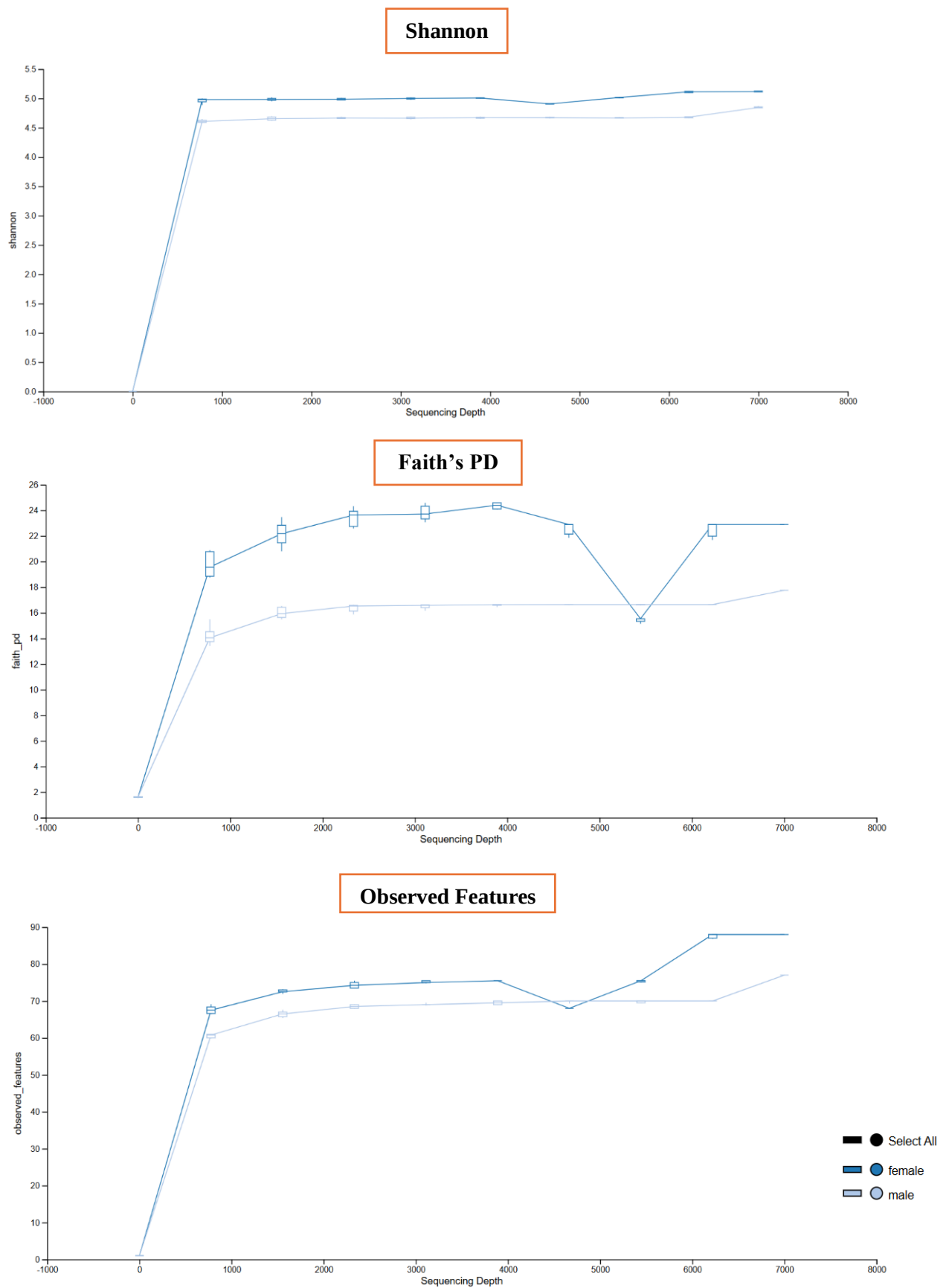
Anexo 22. Tabla resumen denoising-stats.

https://utaedu-my.sharepoint.com/:x:/g/personal/ilucero6068_uta_edu_ec/ESxXqqGGxkVKi7KXQfjgu80B_KJmqmNK3VXx-tIgXQkFw?e=0AsqR9

Anexo 23. Curva de rarefacción de la enfermedad EC.



Anexo 24. Curva de rarefacción de la ECr.



Anexo 25. Tablas taxonomy.tsv con la asignación taxonómica de cada muestra.

https://utaedu-my.sharepoint.com/:x:/g/personal/ilucero6068_uta_edu_ec/EfuBZ52iWmtHs5heOoDhlaoBDJgHJRqzfiEWGcBGmHJbPw?e=mhZDzi

Anexo 26. Tablas de la frecuencia de los taxones en cada nivel taxonómico de los grupos de estudio.

-EC

https://utaedu-my.sharepoint.com/:x:/g/personal/ilucero6068_uta_edu_ec/Eeo-uCYygsJCnMJNxJYNJJgBKl9PxvLploTziYrJGKK_Fg?e=Itex7u

-Control

https://utaedu-my.sharepoint.com/:x:/g/personal/ilucero6068_uta_edu_ec/EdldR2dZafNNuHDKTdZc_E4BACI98eiCZHbg9tDRIDTTEA?e=0ggDOQ

-ECr

https://utaedu-my.sharepoint.com/:x:/g/personal/ilucero6068_uta_edu_ec/ERbsuNJ4wUFCidtk5WUwdUBNQLtxzFCGF8PhekOdN0ibA?e=foty2b

Anexo 27. Script utilizado para el Foresplot.

Cargar librerías necesarias

```
library(DESeq2)
```

```
library(ggplot2)
```

Setear el directorio

```
setwd("dirección del directorio")
```

Cargar el archivo .tsv tanto de la EC y ECr

```
Celiac/Crohn_datos <- read.csv("dirección del archivo .tsv EC/ECr", sep = "\t", row.names = 1)
```

Cargar el archivo control.tsv

```
control_datos <- read.csv("dirección del archivo .tsv del control", sep = "\t", row.names = 1)
```

Combinar los archivos manteniendo solo los taxones comunes

```
datos_combinados <- merge(celiac/crohn_datos, control_datos, by = "row.names", all = FALSE)
```

Restaurar los nombres de los taxones como nombres de fila

```
rownames(datos_combinados) <- datos_combinados$Row.names
```

```
datos_combinados <- datos_combinados[, -1] # Eliminar la columna extra con nombres de taxones
```

Crear el vector que identifica las condiciones (Celiac/Crohn y Control)

```
condiciones <- data.frame(  
  row.names = colnames(datos_combinados),  
  Condicion = c(rep("Celiac/Crohn", ncol(celiac/crohn_datos)), rep("Control", ncol(control_datos)))  
)
```

Crear el objeto DESeq2

```
dds <- DESeqDataSetFromMatrix(countData = datos_combinados,  
                              colData = condiciones,  
                              design = ~ Condicion)
```

Realizar el análisis de abundancia diferencial

```
dds <- DESeq(dds)  
resultados <- results(dds)
```

Crear el dataframe de resultados con todos los taxones, sin filtrar por padj

```
resultados_df <- as.data.frame(resultados)  
resultados_df$Taxon <- rownames(resultados_df)
```

Crear el Forest Plot con log2 Fold Change para todos los taxones

```
ggplot(resultados_df, aes(x = log2FoldChange, y = reorder(Taxon, log2FoldChange))) +  
  geom_bar(stat = "identity", aes(fill = log2FoldChange > 0)) +  
  theme_minimal() +  
  labs(title = "Abundancia Diferencial entre Celiac/Crohn y Control",  
        x = "Log2 Fold Change (Celiac/Crohn vs Control)",  
        y = "Taxón",  
        fill = "Mayor en") +  
  scale_fill_manual(values = c("red", "blue"), labels = c("Celiac/Crohn", "Control")) +  
  • theme(axis.text.y = element_text(size = 8))
```

Anexo 28. Script para la obtención del archivo fasta y .biom.

Para obtener biom table:

```
qiime tools export \  
  --input-path clustered_table.qza \  
  --output-path ./  
  :: feature-table.biom
```

Para obtener .fasta:

```
qiime tools export \
  --input-path clustered_sequences.qza \
  --output-path
    :: dna-sequences.fasta
```

Anexo 29. Script utilizado para el análisis funcional.

```
picrust2_pipeline.py \
  -s dna-sequences.fasta \
  -i feature-table.biom \
  -o picrust2_out_pipeline \
  -p 12
```

Anexo 30. Tabla de funciones.

https://utaedu-my.sharepoint.com/:x/g/personal/ilucero6068_uta_edu_ec/Eal7HULRL4BMgnu05Sff0RkBnJFvSMc tmSRoVtvagD-L7A?e=UTZe2B

Anexo 31. Script utilizado para la creación del heatmap.

Cargar las bibliotecas necesarias

```
library(pheatmap)
```

```
library(dplyr)
```

Setear el directorio

```
setwd("dirección del directorio")
```

Paso 1: Cargar los datos de frecuencia de las rutas metabólicas

```
data <- read.table("heatmap.tsv", header = TRUE, row.names = 1, sep = "\t")
```

Paso 2: Filtrar solo columnas numéricas

```
data <- data %>% select(where(is.numeric))
```

Paso 3: Remover filas con datos faltantes

```
data <- na.omit(data)
```

Paso 4: Calcular la variabilidad de Pearson (coeficiente de variación)

Remover filas donde el promedio es cero para evitar división por cero

```
mean_values <- rowMeans(data)
```

```
data <- data[mean_values != 0, ]
```

```
mean_values <- mean_values[mean_values != 0]
```

Calcular la desviación estándar y luego el coeficiente de variación de Pearson

```
sd_values <- apply(data, 1, sd)
```

```
pearson_variability <- (sd_values / mean_values) * 100
```

Paso 5: Seleccionar las rutas con alta variabilidad de Pearson y alta abundancia promedio

```
top_pearson_threshold <- quantile(pearson_variability, 0.01, na.rm = TRUE) # Top 10% en  
variabilidad de Pearson
```

```
top_abundance_threshold <- quantile(mean_values, 0.01, na.rm = TRUE) # Top 50% en  
abundancia
```

Filtrar las rutas que cumplan ambos criterios

```
data_significant <- data[pearson_variability >= top_pearson_threshold & mean_values >=  
top_abundance_threshold, ]
```

Verificar si hay suficientes filas para el clustering

```
if (nrow(data_significant) >= 2) {
```

```
  # Paso 6: Crear una matriz de etiquetas para el heatmap con la variabilidad de Pearson
```

```
  labels <- matrix(round(pearson_variability[rownames(data_significant)], 2),
```

```
    nrow = nrow(data_significant),
```

```
    ncol = ncol(data_significant),
```

```
    byrow = TRUE)
```

```
  rownames(labels) <- rownames(data_significant)
```

```
  colnames(labels) <- colnames(data_significant)
```

```
  # Crear el heatmap con clustering
```

```
  pheatmap(data_significant,
```

```
    color = colorRampPalette(c("blue", "yellow", "red"))(50),
```

```
    main = "Rutas metabólicas más representativas en Celiac/Crohn (Variabilidad de Pearson)",
```

```
    fontsize_row = 8,
```

```
    fontsize_col = 8,
```

```
    border_color = NA,
```

```
    cluster_rows = TRUE,
```

```
    cluster_cols = TRUE,
```

```
    display_numbers = labels) # Mostrar el coeficiente de variación de Pearson
```

```
}
```

