VNIVERSITAT DE VALÈNCIA Escola Tècnica Superior d'Enginyeria

PRINCIPE FELIPE CENTRO DE INVESTIGACION

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA

VNIVERSITAT DE VALÈNCIA

TRABAJO DE FIN DE MÁSTER

Alineamientos de genoma completo y su uso en la identificación taxonómica de aislados bacterianos y especiación de cepas del género *Bifidobacterium*

AUTOR:
Antonio Bahilo Gómez

TUTORES:
Alfonso Benítez Páez
Rosario Gil García

Enero, 2023

Presidente del Tribunal Evaluador / *President del tribunal avaluador* via **CCA del Máster** / *CCA del Màster*

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA

TRABAJO DE FIN DE MÁSTER

Alineamientos de genoma completo y su uso en la identificación taxonómica de aislados bacterianos y especiación de cepas del género *Bifidobacterium*

AUTOR:
Antonio Bahilo Gómez

TUTORES:
Alfonso Benítez Páez
Rosario Gil García

Tribunal:

PRESIDENTE/A:

VOCAL 1:

VOCAL 2:

FECHA DE DEFENSA:

CALIFICACIÓN:

Presidente del Tribunal Evaluador / *President del tribunal avaluador* via **CCA del Máster** / *CCA del Màster*

Resumen

Las especies del género *Bifidobacterium* son bacterias grampositivas, con respiración anaerobia que fueron descubiertas por primera vez en las heces de lactantes, siendo uno de los taxones más abundantes y estudiados de la microbiota intestinal en humanos. El uso de las características morfológicas de las bacterias ha sido la principal herramienta para su clasificación hasta el surgimiento de las técnicas genéticas y genómicas. Actualmente, existen multitud de técnicas de comparación de identidades genómicas disponibles, siendo la herramienta ANI (identidad nucleotídica media) una de las medidas más robustas de identificación de especies bacterianas.

En este trabajo se ha realizado un análisis comparativo de genomas completos para un grupo de datos de más de 800 genomas de cepas pertenecientes al género *Bifidobacterium* con la herramienta FastANI, con tal de descubrir posibles errores de anotación en las especies y establecer relaciones taxonómicas más estables dentro del género *Bifidobacterium.*

Los valores obtenidos en el análisis comparativo de genomas han permitido excluir 24 genomas del estudio y reunificar 30 especies atendiendo al valor de ANI intraespecie obtenido (> 94%). Además, se ha mostrado que el intervalo de (80-94%) de ANI podría ser útil para distinguir la variación género específica con respecto a los valores reportados en la literatura para organismos bacterianos.

En conclusión, haría falta realizar más análisis basados en la herramienta bioinformática ANI para determinar un valor global y robusto, que permita a la comunidad científica poder establecer rangos estándares de valores ANI interespecies e interespecies, con el objetivo de fortalecer las anotaciones bacterianas y comprender mejor las relaciones beneficiosas entre este taxón bacteriano y el ser humano.

Palabras clave

FastANI, ANI, *Bifidobacterium*, R, análisis genómico, bifidobacterias, intraespecie.

Índice

Resumen ..... *i*

Índice ..... *ii*

Lista de tablas y figuras ..... *iii*

Lista de acrónimos ..... *iv*

Introducción *1*

1.1 Características del género *Bifidobacterium* **1**

1.1.1 Papel de *Bifidobacterium* en el microbiota intestinal de recién nacidos 1

1.1.2 Beneficios generados por las bifidobacterias en la salud humana 2

1.2 Análisis de la diversidad genética en procariotas **3**

1.2.1 Herramientas genéticas para la determinación de especies 3

1.2.2 Análisis de genomas completos para la determinación de especies 4

1.2.3 Genómica aplicada al estudio de las bifidobacterias 5

Objetivos *7*

Material y métodos *8*

3.1 Análisis bioinformático **8**

3.1.1 Diseño del estudio 8

3.1.2 Secuencias genómicas utilizadas 8

3.2 Estudio filogenético de los genomas **9**

3.2.1 Algoritmo de FastANI 9

3.2.2 FastANI aplicado a los genomas de *Bifidobacterium* 10

3.3 Filtrado estadístico con R **10**

3.4 Acceso al código **10**

Resultados y Discusión *11*

Conclusiones *17*

Referencias bibliográficas *18*

Lista de tablas y figuras

Figura 3.1. Pipeline del análisis bioinformático.....8

Figura 3.2. Representación gráfica del algoritmo de FastANI.....9

Figura 4.1. Plot de densidad de las diferentes especies analizadas.....12

Tabla 4.1 Relaciones ANI intraespecies.....13

Tabla 4.2. Relaciones ANI interespecies.....13

Figura 4.2. Mapa de calor de las relaciones entre especies.....14

Tabla 4.3. Relaciones ANI intraespecies tras la reunificación y filtrado.....15

Figura 4.3. Mapa de calor de las relaciones entre especies después de la reunificación y filtrado..15

Figura 4.4. Comparación de la distribución de ANI entre las diferentes especies.....16

Lista de acrónimos

ADN: ácido desoxirribonucleico

ANI: identidad nucleotídica media

ARNr: ácido ribonucleico ribosomal

BLAST: Herramienta de búsqueda de secuencias por alineamiento

BV-BRC: Base de datos de genomas bacterianos y virales

DDH: técnicas de hibridación de ADN-ADN

Mashmap: Herramienta de alineamiento de secuencias

Mb: Megabases, unidad de medida de longitudes de ADN que equivalen a 10^6 de bases

MLST: Método de análisis de variabilidad genética que usa determinados genes

MUMmer: Herramienta informática de reconocimiento de secuencias

NCBI: Biblioteca nacional americana de información sobre biotecnológica

R: Lenguaje de programación especializado en el análisis estadístico

Introducción

1.1 Características del género *Bifidobacterium*

Las especies del género *Bifidobacterium* son bacterias grampositivas, con respiración anaerobia, no formadoras de esporas, inmóviles y con un alto contenido de G+C (55-66%) (Scardovi, 1986). Respecto a su morfología, presentan forma de bacilos con irregularidades que les confiere forma bifurcada en forma de "Y". Esta disposición bífida puede ser producida por la ausencia o baja concentración de aminoácidos e iones Ca^{2+} en los medios de crecimiento (Klijn *et al;* 2005).

Pertenecen al filo Actinomyceota que es uno de los filos más grandes del reino Bacteria. Se clasifican dentro de la familia Bifidobacteriaceae, y en esta familia encontramos cinco géneros diferentes: *Gardnerella, Parascardovia, Scardovia, Aeriscardovia y Bifidobacterium.* Para el género de *Bifidobacterium* existen más de 30 especies descritas hasta la fecha, cabe destacar las especies *B.breve y B.longum* por ser las más prevalentes , las más estudiadas y secuenciadas de la microbiota intestinal en humanos adultos. (Lee & O´Sullivan, 2010). La clasificación taxonómica se ha realizado, comúnmente, mediante la comparación de la secuencia del gen del ARN ribosómico 16S, pero actualmente existen diferentes herramientas que permiten evaluar con más precisión la filogenia descrita dentro de cada género, esta idea se abordará más adelante en otro apartado.

1.1.1 Papel de *Bifidobacterium* en el microbiota intestinal de recién nacidos

Las bifidobacterias fueron descubiertas por primera vez en las heces de lactantes, siendo uno de los taxones más abundantes e investigados de la microbiota intestinal en los recién nacidos (Turroni *et al*; 2022). La microbiota intestinal en bebés está compuesta por miles de millones de células microbianas que habitan en el intestino delgado y grueso, siendo este último el lugar de mayor variedad y complejidad microbiana. Algunas investigaciones recientes apoyan la idea de que la colonización inicial del intestino ocurre poco después del nacimiento, y está influenciada por diversos factores perinatales como el tipo de parto, la alimentación o la edad del gestante (Wang *et al;* 2020). De hecho, algunos de los microorganismos que se introducen a través de la alimentación han desarrollado características fisiológicas y morfológicas específicas para su persistencia en el intestino. Ejemplo de ello son las estructuras extracelulares, como las fimbrias, que permiten la adhesión de las bacterias a la mucosa intestinal. La progresiva comprensión de estas interacciones ha revelado un patrón de coevolución a largo plazo entre el anfitrión humano y su microbiota huésped, resultando en una relación mutualista crucial para el correcto desarrollo del infante durante los primeros años de vida (Song *et al;* 2021).

La microbiota intestinal infantil se desarrolla a través de un proceso dinámico que comienza en el nacimiento y continúa hasta los 2 o 3 años, que es cuando se asemeja a la composición de un adulto. Es sabido que la transmisión vertical desde la microbiota materna es una vía importante para la colonización intestinal infantil, junto con otros taxones microbianos adquiridos por otras rutas (Miliani *et al;* 2017).

Una de las principales vías de colonización vertical de *Bifidobacterium* es la alimentación neonatal con leche materna. Esta ruta representa la mejor dieta para el correcto desarrollo de la microbiota en los recién nacidos. Además, permite la colonización temprana y abundante de bifidobacterias, lo que a su vez lleva a la producción de sus principales metabolitos de fermentación, ácido acético y láctico, que tienen efectos inhibitorios en el crecimiento de microorganismos perjudiciales como *Salmonella spp* y *Listeria spp* (Devika & Raman; 2019).

1.1.2 Beneficios generados por las bifidobacterias en la salud humana

Las bifidobacterias parecen tener varios efectos beneficiosos en el cuerpo humano y su abundancia en la microbiota intestinal infantil parece estar inversamente relacionada con la aparición de varias enfermedades, dando paso a un posible uso de estas bacterias como biomarcadores asociados al bienestar humano. De hecho, la falta de estos microorganismos en el intestino infantil se ha relacionado con el desarrollo del asma y de algunas enfermedades autoinmunes (Cukrowska *et al;* 2020). También, se ha estudiado que su presencia durante la infancia parece mejorar la eficacia protectora de las vacunas infantiles al fortalecer la memoria inmunitaria (Huda *et al;* 2019).

Por otro lado, han surgido diferentes hipótesis que parecen defender la implicación de las bifidobacterias en las conexiones entre el intestino y el cerebro, ya que se cree que pueden ayudar a la organización funcional de los circuitos neuronales durante los primeros años de vida. De hecho, se ha descubierto que estas bacterias están implicadas en la regulación de la expresión genética de la sinapsis y en la modulación de funciones despeñadas por las células de la microglía (Luck *et al;* 2020).

Debido a las supuestas propiedades beneficiosas que tienen las bifidobacterias para la salud, estas se han incorporado en muchos alimentos como ingredientes probióticos. Por esto, es crucial conocer cuál es la microbiota natural del ser humano ya que existen estudios, como el desarrollado por Fragiadakis y colaboradores, donde hallaron diferencias significativas entre las poblaciones microbianas humanas en diferentes regiones del mundo, y cuyas diferencias residían, fundamentalmente, en la dieta, el estilo de vida, edad y sexo de la población (Fragiadakis *et al;* 2019).

Por último, es importante comentar que para poder sobrevivir en nichos tan diversos como la mucosa oral o la mucosa intestinal, las bifidobacterias deben poseer adaptaciones genéticas que les permitan ser competitivas con el resto de bacterias. Por ello, la determinación de las secuencias genómicas permite explicar diferentes peculiaridades

como sus capacidades metabólicas, su evasión del sistema inmunitario adaptativo del hospedador y su adhesión al hospedador mediante apéndices específicos. Gracias al uso de herramientas bioinformáticas se pueden llegar a conocer las divergencias genéticas entre todas las especies de este género bacteriano (O'Callaghan & van Sinderen 2016).

1.2 Análisis de la diversidad genética en procariotas

Desde los inicios de la microbiología, los científicos han tenido la necesidad de clasificar taxonómicamente todos los microorganismos que podían ser aislados en cultivos. Para poder generar estas clasificaciones han necesitados construirlas en base a sus diferencias fenotípicas y genotípicas. Por ello, el objetivo del análisis de la diversidad microbiana ha sido la construcción de taxones estables que permitieran objetivamente conocer la procedencia de una cepa recién descubierta.

1.2.1 Herramientas genéticas para la determinación de especies

El uso de las características morfológicas de las bacterias ha sido la principal herramienta para su clasificación hasta el surgimiento de la genética y de sus técnicas de análisis. El uso de las técnicas de hibridación de ADN-ADN (DDH, del inglés) permitió mejorar las agrupaciones taxonómicas determinadas anteriormente por sus características fenotípicas. Las comparaciones de cadenas de ADN con similitudes mayores al 70%, con la técnica DDH, permitían conocer si una cepa pertenecía a la especie comparada (Tindal *et al;* 2010). Esta técnica necesitaba de mucho tiempo de preparación y era propensa a errores, y debido al descubrimiento de la secuenciación, surgieron nuevos métodos de análisis de comparación de secuencias.

El uso de las regiones hipervariables del gen del ARN ribosómico (ARNr) 16S ha sido la técnica más ampliamente utilizada en la identificación de bacterias. Algunos investigadores han podido relacionar los resultados del análisis del ARNr 16S con los datos obtenidos de la técnica DDH, siendo un valor del 97% en ARNr 16S extrapolable a un 70% en el análisis por DDH (Stackerbrandt & Goebel, 1994). Otra herramienta utilizada para determinaciones evolutivas en procariotas es la técnica de MLST (*multi-locus sequence typing*), esta permite comparar las secuencias de genes codificantes con una base de datos, cuya elección de genes dependerá del grupo bacteriano al que pertenezca el aislado, siendo muy útil cuando el análisis de ADN ribosómico no es suficiente resolutivo. Sin embargo, aunque estás dos técnicas siguen empleándose en diferentes campos de la genómica bacteriana, actualmente ha surgido una nueva metodología que combina las nuevas técnicas de secuenciación masiva con el uso de herramientas bioinformáticas, cosa que permite comparar los genomas completos de las bacterias bajos estudio.

1.2.2 Análisis de genomas completos para la determinación de especies

Dentro de todas las técnicas de comparación de identidades genómicas disponibles, la herramienta ANI (identidad nucleotídica media) es una de las medidas más robustas de identificación de especies bacterianas. De hecho, algunos autores la han propuesta como la medida más útil y sencilla para la determinación de diferentes especies a partir de una gran variedad de datos secuenciados (Konstantinidis & Tiedje, 2005).

Al igual que se realizó la comparación de DHH con el análisis del gen ARNr 16S, algunos autores han comparado la técnica ANI con los resultados por DHH, dando como resultado que un 70% de DHH correspondería a un 95% en ANI. Además, observaron que cuando solo se analizaban las zonas codificantes de proteínas el valor podría bajar hasta un 85% en ANI. Estos resultados mostraron la amplia diversidad que existe dentro de taxón especie y concluyeron que la herramienta ANI puede sustituir perfectamente al resto de técnicas descritas anteriormente (Goris *et al;* 2007).

Dentro de la herramienta ANI existen multitud de algoritmos que permiten compara nucleótidos promedios mediante el mapeo de secuencias sin alineación previa. Estos se pueden dividir en dos tipos dependiendo de la estrategia algorítmica utilizada.

Por un lado, tenemos las herramientas ANI basadas en comparaciones por cálculos en BLAST, donde primero se realiza una búsqueda de los genes codificante ortólogos (Konstantinidis & Tiedje, 2005) o cortando el genoma aleatoriamente en fragmentos de 1020 nucleótidos (Goris *et al;* 2007). A esta metodología de ANI se le conoce como ANIb (basada en BLAST). Por otro lado, existen un algoritmo de comparación de secuencias basado en árboles de sufijos que permite calcular y comparar la alineación rápida de secuencias en genomas con longitudes de millones de nucleótidos (Mb). Este tipo de software es conocido como MUMmer (Kurtz *et al;* 2004). A estas herramientas se les conoce como ANIm (basadas en MUMmer).

Al comparar los tiempos de trabajo de los diferentes algoritmos (ANIb y ANIm) se observó que ANIm es mucho más rápido que ANIb e igual de preciso. Esto es debido a que ANIm no necesita de cortes previos del genoma o del cribado de los genes ortólogos entre las especies a comparar (Richter & Rosselló-Móra, 2009). Estas diferencias de tiempos se acentúan conforme más genomas se añadan al proceso de comparación.

En la actualidad, se siguen desarrollando nuevos algoritmos para disminuir los tiempos de análisis de genomas completos. Ejemplo de esto es un nuevo algoritmo basado en MinHash (algoritmo que estima la similitud entre dos conjuntos de elementos a partir del cálculo de la similitud de Jaccard) conocido como Mashmap (Jain *et al;* 2018). Este algoritmo es utilizado por el método FastANI, el cual se comparó con el algoritmo ANIb siendo FastANI hasta tres órdenes de magnitud más rápido que ANIb. Este estudio llevado

a cabo por Chirag Jain y colaboradores, comparó 1675 genomas extraído de NCBI contra el genoma de referencia *E.coli K-12 MG1655*, siendo FastANI (782x) más rápido que ANIb (Jain *et al;* 2018).

1.2.3 Genómica aplicada al estudio de las bifidobacterias

Durante años, la taxonomía bacteriana se ha basado en el uso de técnicas de biología molecular y en la comparación de secuencias individuales para medir el grado de similitud y definir las relaciones filogenéticas de nuevas especies. Sin embargo, debido a la reciente explosión de datos genómicos y de herramientas de análisis y uso en el campo de la microbiología, se ha facilitado el acceso al contenido completo de los genomas bacterianos, lo que ha brindado la oportunidad de investigar con precisión la diversidad genética de los microorganismos (Lugli *et al;* 2018).

El análisis *in silico* de secuencias de genomas completos se ha utilizado para analizar las diferentes especies del género *Bifidobacterium*. Un claro ejemplo de esto, fue un estudio donde se analizaron 13 genomas de *B.breve* aislados de diferentes nichos (tracto vaginal, leche materna y heces humanas). Se encontró que existía gran variabilidad genética entre los genomas de la misma especie y observaron que todos los organismos compartían genes relacionados con la adaptación al medio del hospedador y con la colonización del intestino (Bottacini *et al;* 2014).

En otro estudio de análisis de genomas de *Bifidobacterium* mediante el método ANIm, se analizaron 191 cepas de *Bifidobacterium longum* para la generación de un pangenoma (conjunto de todos los genes de un determinado taxon). Se observó que *B.longum* forma un pangenoma abierto con un total de 17.000 genes, de los cuales solo el 3% forman parte del genoma central, es decir, son compartidos por todos los genomas analizados (Albert *et al;* 2019).

Las anteriores investigaciones se han centrado en análisis de la diversidad genética de las bifidobacterias a nivel de especies concretas. Sin embargo, existen otros estudios que han analizado las diferencias evolutivas entre las especies del género *Bifodobacterium* mediante el uso de ANI como herramienta bioinformática.

En un estudio de reciente publicación se analizaron 75 genomas de bifidobacterias con el método ANIb y, junto con el filtrado previo de genes ortólogos, se concluyó que la transferencia horizontal de genes influye en la diversidad de las funciones metabólicas (*p ej.* metabolismos de carbohidratos.) del género. Además, esta investigación generó pistas sobre las especies que podrían conferir mayor beneficio para la salud humana y ser estas utilizadas como posibles probióticos (Deb, 2022).

Otro claro ejemplo es el estudio realizado por Lugli y colaboradores donde se evaluó la variabilidad intraespecie (diversidad de genomas dentro de cada especie) del género

Bifidobacterium para 10 especies diferentes repartidas en 258 genomas. Se observó en la mayoría de las especies, con varios organismos anotados dentro de ellas, que los valores de ANI fluctuaban entre 95-99.98%, y que las especies (*B. adolescentis, B. animalis, B. bifidum, B. breve, B. dentium, B. longum, and B. pseudocatenulatum*) tenían en todos los casos un valor de ANI intraespecie mayor a 95%. En cambio, los genomas de las especies anotadas como (*B. asteroides, B. pseudolongum, and B. thermophilum*) poseían valores de ANI menores de 93%. Los resultados de este estudio mostraron el gran potencial del análisis de genomas completos mediante técnicas bioinformáticas para la correcta clasificación taxonómica de bacterias. Además, mostraron la cantidad de errores o malas anotaciones que existen en las bases de datos de genomas bacterianos actuales (Lugli *et al;* 2018).

Las estrategias de comparación de genomas completos, aplicadas a la gran cantidad de información genómica disponible en las bases de datos, pueden ser de gran utilidad para estudios de diversidad genómica y para establecer relaciones taxonómicas más fiables. Además, estas herramientas bioinformáticas pueden servir para detectar inconsistencias en las bases de datos (e.g. anotaciones erróneas) y permiten el filtrado de información para la consecución y establecimiento de estudios más robustos sobre las relaciones filogenómicas entre especies de este grupo de bacterias, las cuales son determinantes para nuestra salud.

En el presente trabajo, se ha realizado un análisis de más de 800 genomas de bifidobacterias extraídos de la base de datos BV-BRC (del inglés, *Bacterial and Viral Bioinformatics Resource Center*) con la herramienta FastANI, dada su demostrada capacidad biocomputacional para la comparación de cientos de genomas simultáneamente. Se han estudiado, gracias al lenguaje de programación de R, las relaciones intraespecie para descubrir posibles errores de anotación en las especies detectadas y generar reunificaciones de especies atendiendo a los valores interespecie obtenidos con FastANI. Este trabajo servirá como referencia al filtrado previo que debe hacerse en los genomas secuenciados antes de proceder a la realización de cualquier pangenoma o de análisis genéticos de especies bacterianas.

Objetivos

El objetivo general del presente trabajo final de máster es realizar un análisis comparativo de genoma completo para un grupo de datos de más de 800 genomas de cepas pertenecientes al género *Bifidobacterium* con el estándar ANI, para descubrir posibles errores de anotación en las especies y establecer relaciones taxonómicas más estables dentro del género *Bifidobacterium*. Los objetivos específicos son:

- Establecer relaciones filogenéticas más robustas entre las diferentes especies de bifidobacterias con la herramienta FastANI.
- Detectar inconsistencias generadas a partir de anotaciones taxonómicas erróneas entre los genomas descargados de la base de datos BV-BCR.
- Establecer distribución de valores ANI intra- e interespecie en el grupo de datos analizado, con el propósito de distinguir variación género específica con respecto a los valores reportados en la literatura para especies de bacteria.

Material y métodos

3.1 Análisis bioinformático

3.1.1 Diseño del estudio

Para el estudio se desarrollaron *scripts* en lenguaje *Bash* y en *R* (se adjuntan en los anexos) para la automatización del proceso. La configuración del recurso computacional usado fue un procesador de 3GHz Intel Core i5 de 4 núcleos y con una memoria de 16 GB 2400 MHz DDR4. En la figura 3.1 se muestra el protocolo general del análisis de los genomas de las bifidobacterias.

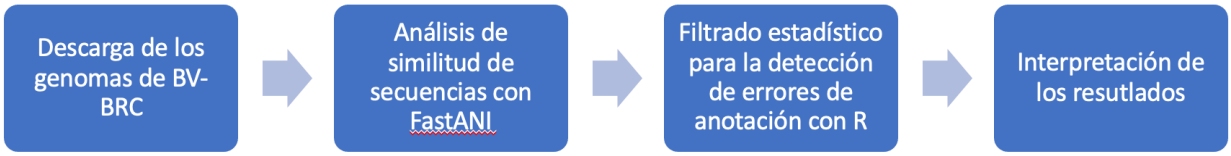

Figura 3.1. Pasos utilizados en el estudio. En la imagen se muestra el flujo de trabajo realizado durante el estudio.

3.1.2 Secuencias genómicas utilizadas

Todos los genomas utilizados en este estudio fueron descargados del repositorio público BV-BRC [https://www.bv-brc.org/]. Para descargarlos se aplicaron los filtros (*Genome_quality,"Good",Host_group."Human"*), obteniendo un total de 822 genomas completos del género *Bifidobacterium* relacionados la microbiota humana. Las tecnologías de secuenciación utilizadas la obtención de los genomas fueron muy diversas, se utilizaron diferentes técnicas de NGS, como *Illimunia*, *Roche 454* o *Ion Torrent*, y métodos de tercera generación como Oxoford Nanopore y PacBio. Los nichos de donde se asilaron las muestras de bacterias fueron muy heterogéneos (heces, leche humana, intestino y suplementos alimenticios).

Los números identificativos y el resto de las características de cada uno de los genomas usados se muestran en el archivo "*BVBRC_genome.txt",* que se puede descargar del siguiente enlace de GitHub [https://github.com/Tonibg2/TFMbioinformatica.git]

3.2 Estudio filogenético de los genomas

3.2.1 Algoritmo de FastANI

Los alineamientos de las secuencias genómicas completas fueron realizados a nivel de ADN usando la herramienta bioinformática FastANI. Esta se basa en el algoritmo Mashmap, que realiza un análisis de similitud entre un genoma de referencia y genoma problema. Para ello, primero divide el genoma problema en pequeños fragmentos no superpuestos de tamaño fijo (*K*-meros) indicado por el usuario, en nuestro caso 1Kb de longitud (*l*). Estos k-meros son indexados con una función hash, conocida como Minhash, para convertirlos en un valor único de 64 bits que permitan un rápido almacenamiento y acceso a estas secuencias en la memoria evitando posibles colisiones. A continuación, se comparan cada *k*-mero de la secuencia problema con el genoma de referencia, utilizando para ello el coeficiente de similitud de *Jaccard*. El resultado de la comparación es un valor de identidad, entre 0-1, siendo valores muy similares los cercanos a 1. Por último, se guardan los resultados en un conjunto *M* formado por tripletes de $\langle f, i, p \rangle$, donde *f* es la identificación de fragmento, *i* es la estimación de identidad entre secuencias y *p* es la posición del genoma donde *f* se asigna con el genoma de referencia. Para obtener el valor final de FastANI, se filtran todos los conjuntos *M* que tengan el valor máximo de identidad (*i*) para una determinada posición en el genoma de referencia *(f*) y se realiza una media de los valores identidad filtrados de todos los conjuntos *M (véase la figura 1).*

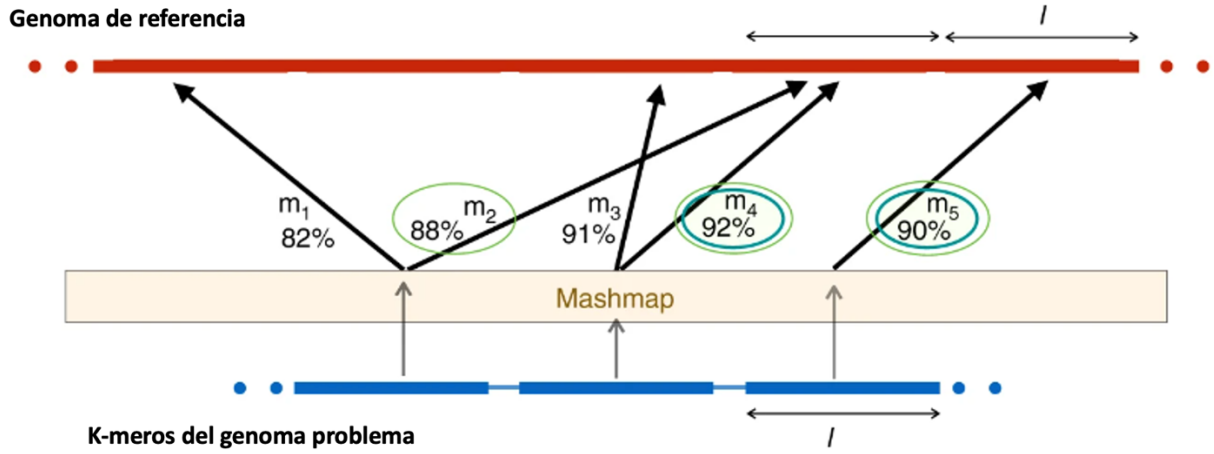

Figura 3.2. Representación gráfica del algoritmo de FastANI. En la imagen se muestra el flujo de trabajo para el cálculo de ANI entre el genoma de referencia y el genoma problema. Se obtienen cinco asignaciones a partir de tres fragmentos (k-meros) del genoma problema usando el algoritmo Mashmap. El conjunto *M* estaría representado por $M=\{m_2,m_4,m_5\}$ tras filtrar los K-meros con mayor valor de identidad. El conjunto M final de la región del genoma de referencia estaría formado por el *k*-mero de mayor identidad

para una posición *f* determinada, dando como resultado un conjunto $M=\{m_4,m_5\}$ de donde se extraerá la identidad media. Imagen editada del artículo (Jain *et al;* 2018).

3.2.2 FastANI aplicado a los genomas de *Bifidobacterium*

El estudio de los 822 genomas con la herramienta ANI se realizó de tal modo que todos genomas fueron comparados contra ellos mismos y contra el resto, es decir, se realizaron un total 675.684 alineamientos globales con un tiempo de duración 21 horas. Además, destacar que se añadieron al análisis un genoma de *E.coli K-12 MG1655* y otro de *Lactobacillus gasseri ATCC 33323* para ser utilizados como controles (*outliers*) en los resultados finales de FastANI.

3.3 Filtrado estadístico con R

Los resultados obtenidos con la herramienta FastANI fueron cargados en el entorno *R 4.0.2* como un matriz de dimensiones 824x824. En primer lugar, se procedió a la eliminación de los géneros y especies no vinculadas con las bifidobacterias relacionadas con el humano. Las especies nos descritas se reasignaron con aquellas especies cuyo valor de identidad media eran mayor al 95%. En segundo lugar, se evaluaron las relaciones intraespecies e interespecie analizando para cada especie una serie de parámetros estadísticos (media, desviación estándar, intervalo de confianza y valor mínimo de identidad). En último lugar, se eliminaron las especies que no cumplían los parámetros estadísticos establecidos y se reasignaron especies mal anotadas.

3.4 Acceso al código

Todos los *scripts* de *Bash* y *R* utilizados en el presente trabajo aparecen detallados en el documento "*Filtrado_TFM_Bifido_ANI_Humano.Rmd*" que puede ser descargado del siguiente enlace [https://github.com/Tonibg2/TFMbioinformatica.git]

Resultados y Discusión

En este estudio se ha realizado una comparación de genomas completos de *Bifidobacterium* con la herramienta FastANI, obteniéndose una matriz de relaciones de valores ANI, de dimensiones 824x824, que ha sido analizada con el lenguaje de programación R.

Para poder llevar a cabo este análisis, el primer paso fue cambiar los identificadores numéricos de los genomas comparados por sus nombres científicos, de todas las filas y columnas de la matriz de datos, a partir de un documento de texto generado tras la descarga de los genomas de BV-BRC. Se comprobó que los valores ANI de los *outliers* fueran menores que los valores establecidos por la bibliografía como propios del género *Bifidobacterium,* donde se considera que un organismo está relacionado con un género si tiene un valor entre 80-100% de ANI (Jain *et al;* 2018). Se obtuvieron valores ANI para los *outliers* menores a 74%, y en la mayoría de los casos, resulto de valores no identificados (*NA*) ya que el algoritmo no otorga valores a las comparaciones si son menores al 74%. Además, en un primer filtrado de los datos se eliminaron 14 genomas identificados dentro del género *Enterococcus* y un genoma de *Cutibacterium* con valores de ANI no identificados. Tras identificar las diferentes especies dentro del género *Bifidobacterium*, se eliminaron 5 genomas clasificados en tres diferentes especies (*B.scardovii, B.gallicum y B.thermophilum*) que no estaban relacionadas como posibles hospedares del organismo humano. De hecho, *B.thermophilum* se relaciona como bacteria hospedadora en heces de cerdo (*Sus scrofa*), *B.scardovii* se ha asociado a infecciones humanas y *B.gallicum,* a pesar de aislarse en heces humanas, se piensa que no son hospedadores naturales del hombre porque muestran una baja relación genética con el resto de especies del género *Bifidobacterium* (Carl & Pradip, 2014; Lee & O´Sullivan, 2010)

En segundo lugar, se reasignaron 24 especies no descritas (identificadas como *B.sp*) según su valor más alto de ANI obtenido (todas las reasignaciones con valores ANI por encima 94%) tras la comparación con el resto de especies, siendo estas especies no identificadas finalmente anotadas dentro de *B.adolescentis, B.longum y B.kashiwanohense.* Este prefiltrado inicial permitió obtener un total de 12 especies diferentes (*B.breve, B.kashiwanohense, B.animalis, B.bifidum, B.adolescentis, B.longum, B.pseudocatenulatum, B.pseudolongum, B.angulatum, B.catenulatum, B.dentium y B.stercoris*), repartidos en 802 genomas, pertenecientes al género *Bifidobacterium* y relacionadas con la microbiota humana (Carl & Pradip, 2014; Lee & O´Sullivan, 2010; Lugli *et al;* 2018).

Tras realizar un primer prefiltrado, se realizó un análisis de las distribuciones ANI intraespecie a partir de los valores ANI resultantes de las comparaciones entre cada uno de

los genomas de las diferentes especies. En la figura 4.1 se representan las diferentes distribuciones de ANIs intraespecies, observándose que *B.longum* tiene algunos genomas mal asignados debido a que existen valores ANI por debajo del 90%. Se determinó que las especies *B.longum (*N=366*)* y *B.breve (*N=100*)* eran las más abundantes de conjunto de genomas estudiados, siendo el 58% del total de genomas analizados, en congruencia con el estudio de Lee y O´Sullivan (2010), donde mostraron que ambas especies eran las prevalentes en la microbiota intestinal en humanos adultos.

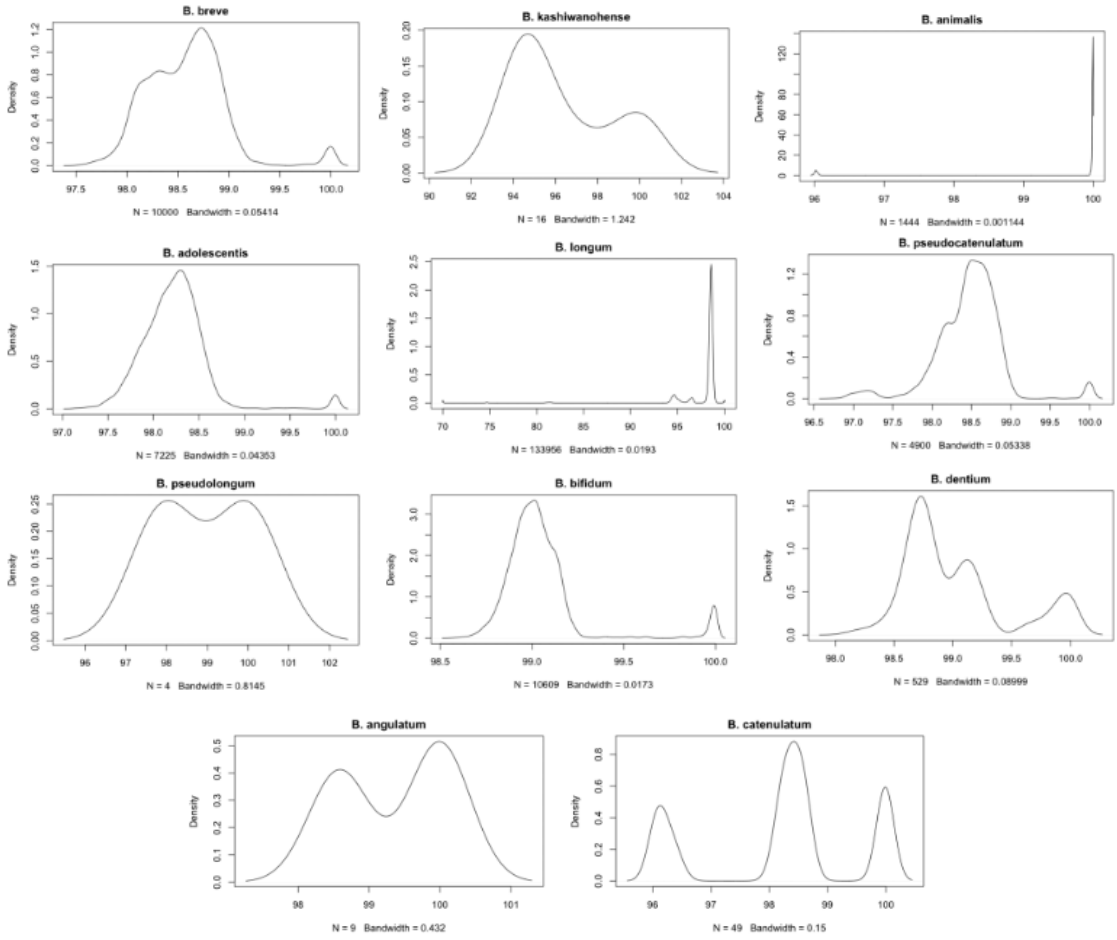

Figura 4.1. Plot de densidad de las diferentes especies analizadas. En la imagen se muestra las distribuciones de las densidades de los diferentes valores ANI obtenidos intraespecie. Para *B.stercoris* no se ha realizado plot ya que solo está compuesto por un único genoma. El parámetro *N* hace referencia a la cantidad de ANIs evaluados, siendo $\sqrt{N}$ el número total genomas estudiados para cada especie. El valor *bandwith* se refiere al grosor de las curvas de densidades representadas.

Para conseguir una mayor confianza entre los valores ANI resultantes, se realizó un análisis estadístico para las relaciones intraespecie e interespecie.

Tabla 4.1. Relaciones ANI intraespecies. En la imagen se muestra las distribuciones intraespecie para las diferentes especies de *Bifidobacterium*.

	B.breve	*B.kashiwanohense*	*B.animalis*	*B.bifidum*	*B.adolescentis*	*B.longum*
Tamaño_N	10000	16	1444	10609	7225	133956
Media_ANI_intra	98,57	96,31	99,60	99,05	98,22	97,89
Desviación	0,39	2,40	1,19	0,24	0,36	2,59
CI 95%	98,56-98,57	95,13-97,48	99,53-99,65	99,04-99,05	98,21-98,22	97,87-97,90
Mínimo	97,54	93,99	95,95	98,56	97,15	70,00
	B.pseudocatenulatum	***B.pseudolongum***	***B.angulatum***	***B.catenulatum***	***B.dentium***	***B.stercoris***
Tamaño_N	4900,00	4,00	9,00	49,00	529,00	1,00
Media_ANI_intra	98,47	98,97	99,37	98,22	99,01	100,00
Desviación	0,42	1,19	0,74	1,34	0,45	NA
CI 95%	98,45-98,48	97,79-100	98,88-99,85	97,83-98,59	98,97-99,05	NA-NA
Mínimo	96,73	97,92	98,56	96,01	98,14	100,00

En la relaciones intraespecies los valores de las medias y de los intervalos de confianzas entraron dentro de los valores esperados, pero se detectó que la variabilidad entre los datos de ANI eran demasiado elevados en las especies *B.longum* (SD=2,59) y *B.kashiwanohense* (SD=2,4). Además, se obtuvo un valor mínimo de ANI menor a 94 en ambas especies, siendo la media de los valores ANI intraespecie de un 98,6% para el conjunto de especies, como se observa en la tabla 4.1. La media de valor ANI interespecie (98,6%) obtenida entra dentro de los resultados alcanzados por otros estudios que describen a organismo dentro de una especie si su valor de ANI es mayor al 95% (Goris *et al;* 2007; Jain *et al;* 2018).

Respecto a las relaciones interespecies, se obtuvo un valor medio de ANI de un 80,42 %, un dato un poco más bajo que el obtenido en el estudio de análisis de 90K de genomas de procariotas por Jain y colaboradores (2018), donde obtuvieron que los valores de genomas relacionados con una determinada especie fluctuaban entre 83-95% de ANI. Cabe destacar que no se ha realizado todavía ningún análisis preciso, con la herramienta ANI, que determine qué valor es el aconsejado en las comparaciones interespecies de genomas pertenecientes al mismo género taxonómico.

Por otro lado, y a pesar de tener desviación muy alta entre algunos datos, esta es esperada debido a la gran variabilidad de muestras comparadas entre las diferentes especies (véase a la Tabla 4.2).

Tabla 4.2. Relaciones ANI interespecies. En la imagen se muestra las distribuciones interespecie para las diferentes especies de *Bifidobacterium.*

	B.breve	*B.kashiwanohense*	*B.animalis*	*B.bifidum*	*B.adolescentis*	*B.longum*
Tamaño_N	70200,00	3192,00	29032,00	71997,00	60945,00	159576,00
Media_ANI_intra	83,15	81,16	77,60	80,13	80,54	81,42
Desviación	3,60	4,00	0,44	1,01	1,68	3,02
CI 95%	83,12-83,18	81,02-81,29	77,59-77,60	80,11-80,13	80,52-80,54	81,40-81,43
Mínimo	70,00	74,81	70,00	70,00	70,00	70,00
	B.pseudocatenulatum	***B.pseudolongum***	***B.angulatum***	***B.catenulatum***	***B.dentium***	***B.stercoris***
Tamaño_N	51240,00	1600,00	2397,00	5565,00	17917,00	801,00
Media_ANI_intra	80,18	78,11	79,95	80,93	79,56	82,38
Desviación	2,26	0,81	0,97	3,97	1,54	5,63
CI 95%	80,16-80,19	78,07-78,15	79,90-79,98	80,82-81,03	79,53-79,58	81,98-82,76
Mínimo	70,00	70,00	70,00	70,00	70,00	74,71

A continuación, se realizó un análisis de la matriz de datos para poder detectar posibles reunificaciones entre las especies. Para ello, se compararon todos los valores ANI mediante un mapa de calor o *heatmap* a partir de las medias de los valores ANI para cada una de las especies.

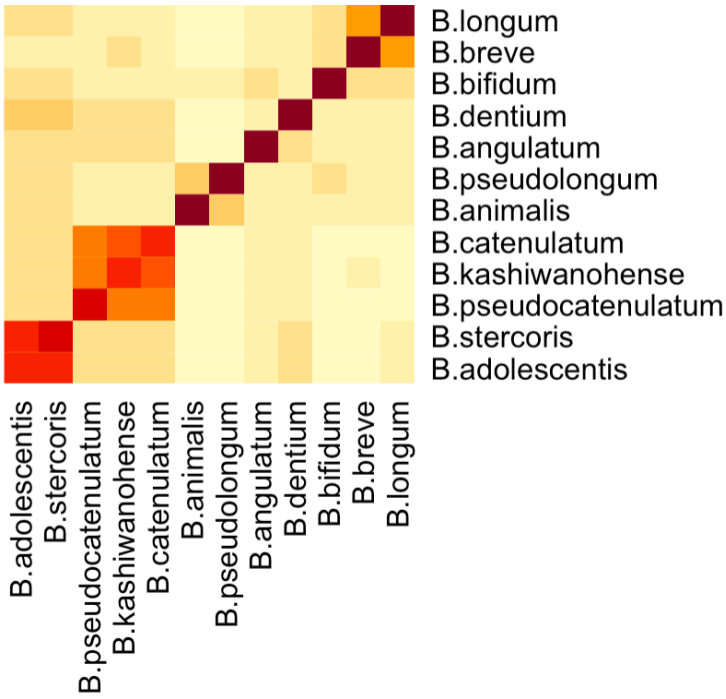

Figura 4.2. Mapa de calor de las relaciones entre especies. Las relaciones de mayor ANI entre 12 especies se muestran con colores más oscuros, mientras que las relaciones con menor ANI se observan con los cuadros más claros. En la diagonal aparecen los valores de la comparación de cada especie consigo misma. Destacan la alta similitud entre las especies (*B.kashiwanohense y B.catenulatum*) y (*B.adolescentis y B.stercoris*).

Se observó que las especies *B.kashiwanohense y B.catenulatum* tenían un valor de ANI interespecie del 95%, mientras que la relación entre las especies *B.adolescentis y B.stercoris* obtuvieron un ANI del 98,1%. En cambio, *B.longum y B.breve* solo compartían un 86% de ANI. Los resultados obtenidos entre *B.kashiwanohense y B.catenulatum* son coherentes con los estudios que muestran que la especie *B.kashiwanohense* es en realidad una subespecie de *B.catenulatum.* (Lui *et al*; 2022). Además, otros estudios de comparación de secuencias del gen 16S del ARNr (con el método DDH) entre *B.adolescentis y B.stercoris,* muestran que *B.stercoris* debe ser reclasificada dentro de *B.adolescentis* debido a su alta similitud de secuencias 99% y 97,7% entre algunos genes (16S ARNr y dnaJ1, respectivamente) (Killer *et al*; 2013)

Tras el análisis de los resultados observados en la figura 4.2, se realizó una nueva reunificación de las especies que tuvieran un valor de ANI interespecie mayor o igual al 94%. Pero antes de unificar las especies *B.kashiwanohense* con *B.catenulatum*, y las especies *B.adolescentis con B.stercoris,* se realizó un filtrado para eliminar todos los genomas con un valor intraespecie menor a 94% para así excluir las especies mal

identificadas, como las que se muestran en la figura 4.1.

Tabla 4.3. Relaciones ANI intraespecies tras la reunificación y filtrado. En la imagen se muestra las distribuciones interespecie para las diferentes especies de *Bifidobacterium.*

	B.breve	***B.catenulatum***	***B.animalis***	***B.bifidum***	***B.adolescentis***
Tamaño_N	10000,00	121,00	1444,00	10609,00	7396,00
Media_ANI_intra	98,57	96,44	99,60	99,05	98,22
Desviación	0,39	2,10	1,19	0,24	0,36
CI 95%	98,56-98,57	96,06-96,81	99,53-99,65	99,04-99,05	98,21-98,22
Mínimo	97,5	94,0	95,9	98,6	97,1
	B.longum	***B.pseudocatenulatum***	***B.pseudolongum***	***B.angulatum***	***B.dentium***
Tamaño_N	132496,00	4900,00	4,00	9,00	529,00
Media_ANI_intra	98,12	98,47	98,97	99,37	99,01
Desviación	1,19	0,42	1,19	0,74	0,45
CI 95%	98,11-98,12	98,45-98,48	97,79-100,13	98,88-99,85	98,97-99,05
Mínimo	94,1	96,7	97,9	98,6	98,1

Después del filtrado para la reunificación de especies y la eliminación de especies mal anotadas, se eliminaron 2 genomas de *B.longum* y se realizó, nuevamente, un estudio de las relaciones de cada una de las especies con el resto de estas. Los resultados se observan en el mapa de calor de la figura 4.3.

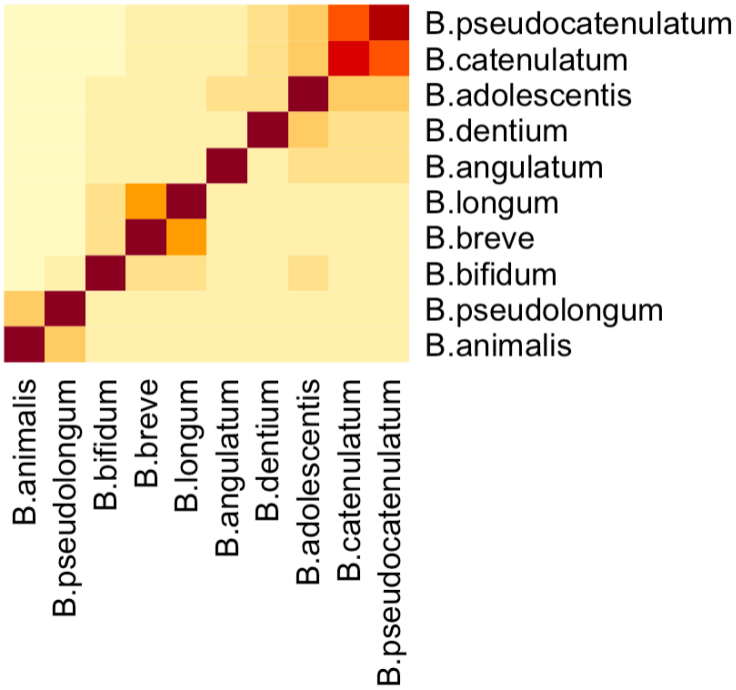

Figura 4.3. Mapa de calor de las relaciones entre especies después de la reunificación y filtrado. Las relaciones de mayor ANI entre 10 especies se muestran con colores más oscuros, mientras que las relaciones con menor ANI se observan con los cuadros más claros. En la diagonal aparecen los valores de la comparación de cada especie con ella misma. Destacan la alta similitud entre las especies (*B.catenulatum y B.pseudocatenulatum)* pero con un valor de ANI < 94%.

En último lugar, se muestran las distribuciones intraespecies para cada una de las especies filtradas donde se observan las frecuencias de los diferentes valores de ANI. Todas las especies tienen una distribución normal y no sé detectan valores atípicos dentro de cada especie.

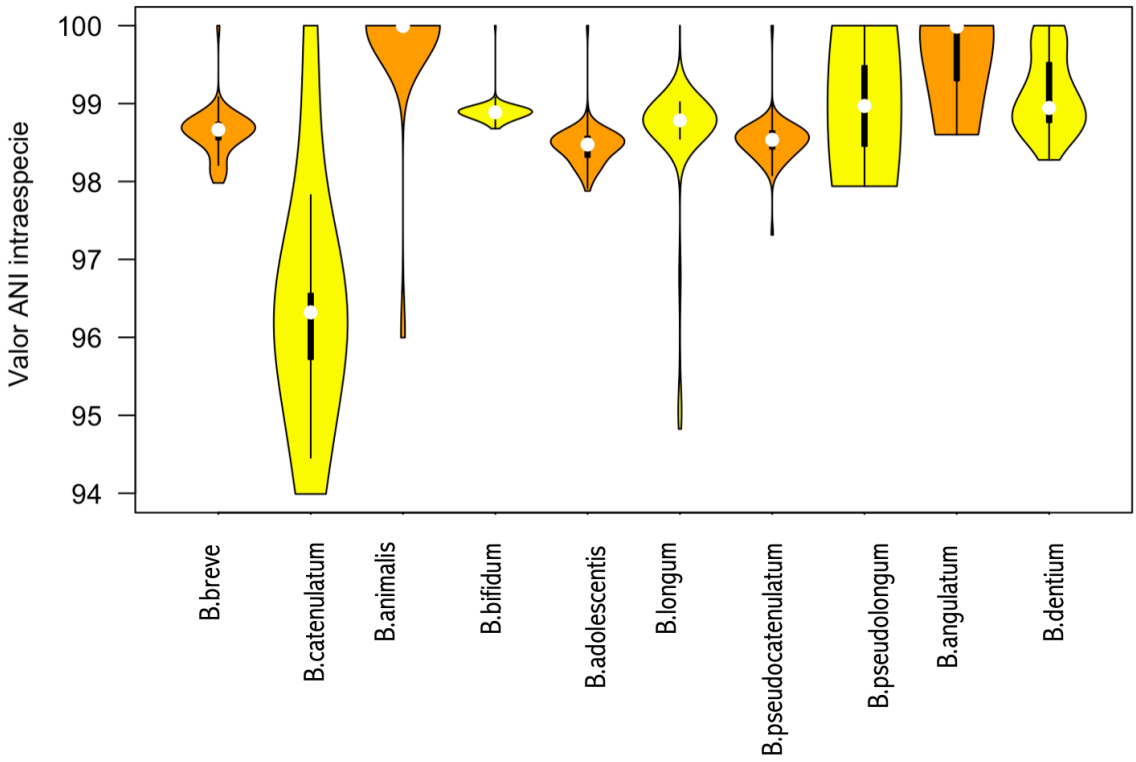

Figura 4.4. Comparación de la distribución de ANI entre las diferentes especies. Se muestra una representación de un *violinplot* para observar las distribuciones de los valores ANI intraespecie. Siendo el punto blanco la mediana de cada muestra. Todas las especies entran dentro de los valores ANI establecidos en este estudio (>94%).

El filtrado de las especies mal anotadas dentro de cada especie junto con la reunificación de varias especies (*B.kashiwanohense* y *B.stercoris*) han permitido excluir 24 genomas del estudio y reunificar 30 especies atendiendo al valor de ANI intraespecie (mayor al 94%). Estos resultados confirman la utilidad de la herramienta bioinformática ANI para la correcta demarcación de especies como proceso previo a la realización de cualquier estudio de análisis genético a nivel de especie bacteriano. Pero todavía queda por consolidar si nuestros resultados de valores ANI interespecie (80-94%) pueden servir como herramienta útil en la demarcación de especies perteneciente a un determinado género bacteriano, ya que existen pocos estudios que hayan utilizado esta técnica para dicho propósito y hayan determinado un valor fiable de ANI para las relaciones interespecie.

Conclusiones

A partir de los resultados, se puede concluir:

1. A pesar de descargar genomas de base de datos bien documentadas y fiables, como es el caso BV-BCR, debería de ser necesario realizar siempre un filtrado inicial antes de realizar cualquier estudio genómico para evitar la presencia de *outliers* en los resultados.

2. El uso de la herramienta bioinformática FastANI ha resultado ser un excelente método de análisis para las relaciones intraespecies, siendo el valor ANI obtenido en este estudio mayor a 94% para la clasificación de genomas dentro de una misma especie.

3. La reasignación de especies será una fase necesaria para cualquier análisis genómico que use genomas descargados de repositorios públicos, debido al crecimiento exponencial de técnicas bioinformáticas que están permitiendo obtener relaciones más robustas entre genomas dentro de una misma especie y corregir antiguas designaciones de especies basadas en pruebas obsoletas.

4. El establecimiento de la distribución de valores ANI interespecie en el grupo de datos analizado, ha mostrado que el intervalo de (80-94%) podría ser útil para distinguir la variación género específica con respecto a los valores reportados en la literatura para organismos bacterianos.

Respecto a futuros estudios, sería importante tener en cuenta que son necesarios más análisis basados en la técnica ANI para determinar un valor global y robusto, que permita a la comunidad científica poder establecer unos rangos de ANI interespecies estándares. También serán necesarias nuevas investigaciones que permitan mejorar y optimizar el algoritmo para que se pueden clasificar los genomas en taxones superiores, ya que actualmente el corte se encuentra en la clasificación del género bacteriano. Además, el aumento del número de genomas completos, secuenciados con técnicas de tercera generación y compartidos en bases de datos, podría suponer una ayuda para mejorar los análisis basados en la herramienta ANI y fortalecer las anotaciones bacterianas actuales y futuras dentro del género *Bifidobacterium,* con el objetivo de comprender mejor las relaciones beneficiosas entre este grupo bacteriano y el ser humano.

Referencias bibliográficas

Albert, K., Rani, A., & Sela, D. A. (2019). Comparative Pangenomics of the Mammalian Gut Commensal Bifidobacterium longum. Microorganisms, 8(1), 7. https://doi.org/10.3390/microorganisms8010007

Batt Carl, Patel Pradip. (2014). Encyclopedia of Food Microbiology. 2nd Edition. Elsevier.

Bottacini, F., O'Connell Motherway, M., Kuczynski, J., O'Connell, K. J., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G. A., Zomer, A., Zhurina, D., Riedel, C., Ventura, M., & van Sinderen, D. (2014). Comparative genomics of the Bifidobacterium breve taxon. BMC genomics, 15(1), 170. https://doi.org/10.1186/1471-2164-15-170

Cukrowska, B., Bierła, J. B., Zakrzewska, M., Klukowski, M., & Maciorkowska, E. (2020). The Relationship between the Infant Gut Microbiota and Allergy. The Role of Bifidobacterium breve and Prebiotic Oligosaccharides in the Activation of Anti-Allergic Mechanisms in Early Life. Nutrients, 12(4), 946. https://doi.org/10.3390/nu12040946

Deb S. (2022). Pan-genome evolution and its association with divergence of metabolic functions in Bifidobacterium genus. World journal of microbiology & biotechnology, 38(12), 231. https://doi.org/10.1007/s11274-022-03430-1

Devika, N. T., & Raman, K. (2019). Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models. Scientific reports, 9(1), 18222. https://doi.org/10.1038/s41598-019-54696-9

Fragiadakis, G. K., Smits, S. A., Sonnenburg, E. D., Van Treuren, W., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Dominguez-Bello, M. G., Leach, J., & Sonnenburg, J. L. (2019). Links between environment, diet, and the hunter-gatherer microbiome. Gut microbes, 10(2), 216–227. https://doi.org/10.1080/19490976.2018.1494103

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. International journal of systematic and evolutionary microbiology, 57(Pt 1), 81–91. https://doi.org/10.1099/ijs.0.64483-0

Huda, M. N., Ahmad, S. M., Alam, M. J., Khanam, A., Kalanetra, K. M., Taft, D. H., Raqib, R., Underwood, M. A., Mills, D. A., & Stephensen, C. B. (2019). Bifidobacterium Abundance in Early Infancy and Vaccine Response at 2 Years of Age. Pediatrics, 143(2), e20181489. https://doi.org/10.1542/peds.2018-1489

Jain C, Dilthey A, Koren S, Aluru S, Phillippy AM. A Fast Approximate Algorithm for Mapping Long Reads to Large Reference Databases. J Comput Biol. 2018 Jul;25(7):766-779. doi: 10.1089/cmb.2018.0036. Epub 2018 Apr 30. PMID: 29708767; PMCID: PMC6067103.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature communications, 9(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Killer, J., Sedláček, I., Rada, V., Havlík, J., & Kopečný, J. (2013). Reclassification of Bifidobacterium stercoris Kim et al. 2010 as a later heterotypic synonym of Bifidobacterium adolescentis. International journal of systematic and evolutionary microbiology, 63(Pt 11), 4350–4353. https://doi.org/10.1099/ijs.0.054957-0

Klijn, A., A. Mercenier, and F. Arigoni.2005. Lessons from the genomes of bifidobacteria. FEMS Microbiol. Rev.29:491-509.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. Genome biology, 5(2), R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lee, J. H., & O'Sullivan, D. J. (2010). Genomic insights into bifidobacteria. Microbiology and molecular biology reviews : MMBR, 74(3), 378–416. https://doi.org/10.1128/MMBR.00004-10

Liu, J., Li, W., Yao, C., Yu, J., & Zhang, H. (2022). Comparative genomic analysis revealed genetic divergence between Bifidobacterium catenulatum subspecies present in infant versus adult guts. BMC microbiology, 22(1), 158. https://doi.org/10.1186/s12866-022-02573-3

Luck, B., Engevik, M. A., Ganesh, B. P., Lackey, E. P., Lin, T., Balderas, M., Major, A., Runge, J., Luna, R. A., Sillitoe, R. V., & Versalovic, J. (2020). Bifidobacteria shape host neural circuits during postnatal development by promoting synapse formation and microglial function. Scientific reports, 10(1), 7737. https://doi.org/10.1038/s41598-020-64173-3

Lugli, G. A., Milani, C., Duranti, S., Mancabelli, L., Mangifesta, M., Turroni, F.,

Viappiani, A., van Sinderen, D., & Ventura, M. (2018). Tracking the Taxonomy of the Genus Bifidobacterium Based on a Phylogenomic Approach. Applied and environmental microbiology, 84(4), e02249-17. https://doi.org/10.1128/AEM.02249-17

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., Lugli, G. A., Rodriguez, J. M., Bode, L., de Vos, W., Gueimonde, M., Margolles, A., van Sinderen, D., & Ventura, M. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiology and molecular biology reviews : MMBR, 81(4), e00036-17. https://doi.org/10.1128/MMBR.00036-17

O'Callaghan A, van Sinderen D. Bifidobacteria and Their Role as Members of the Human Gut Microbiota. Front Microbiol. 2016 Jun 15;7:925. doi: 10.3389/fmicb.2016.00925. PMID: 27379055; PMCID: PMC4908950.

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences of the United States of America, 106(45), 19126–19131. https://doi.org/10.1073/pnas.0906412106

Scardovi, V.1986. The genus Bifidobacterium Orla-jensen 1924, 472AL, p. 1418-1434. In P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 2. Williams & Wilkins, Baltimore, MD.

Song, Q., Wang, Y., Huang, L., Shen, M., Yu, Y., Yu, Q., Chen, Y., & Xie, J. (2021). Review of the relationships among polysaccharides, gut microbiota, and human health. Food research international (Ottawa, Ont.), 140, 109858. https://doi.org/10.1016/j.foodres.2020.109858

Stackebrandt E., Goebel B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. . Int J Syst Bacteriol 44, 846–849

Tindall, B. J., Rosselló-Móra, R., Busse, H. J., Ludwig, W., & Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. International journal of systematic and evolutionary microbiology, 60(Pt 1), 249–266. https://doi.org/10.1099/ijs.0.016949-0

Turroni, F., Milani, C., Ventura, M., & van Sinderen, D. (2022). The human gut microbiota during the initial stages of life: insights from bifidobacteria. Current

opinion in biotechnology, 73, 81–87. https://doi.org/10.1016/j.copbio.2021.07.012

Wang, S., Ryan, C. A., Boyaval, P., Dempsey, E. M., Ross, R. P., & Stanton, C. (2020). Maternal Vertical Transmission Affecting Early-life Microbiota Development. Trends in microbiology, 28(1), 28–45. https://doi.org/10.1016/j.tim.2019.07.010