

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA
AGRONÒMICA I DEL MEDI NATURAL



ESTUDIO METATRANSCRIPTÓMICO DEL DESARROLLO DE LA MICROBIOTA INTESTINAL EN INFANTES

TRABAJO FIN DE GRADO EN BIOTECNOLOGÍA

ALUMNO: Joan Compte Barrón

TUTOR: José Gadea Vacas

COTUTORA: María Pilar Francino Puget

COTUTORA COLABORADORA: María José Gosalbes Soler

Curso Académico: 2014-2015

VALENCIA, septiembre 2015



Licencia Creative Commons "Reconocimiento no Comercial –Sin Obra Derivada"

Estudio metatranscriptómico del desarrollo de la microbiota intestinal en infantes

Autor: Joan Compte Barrón

Tutor: José Gadea Vacas

Cotutora: María Pilar Francino Puget

Cotutora colaboradora: María José Gosalbes Soler

Valencia, Septiembre 2015

RESUMEN:

El intestino humano es el hábitat natural de una comunidad bacteriana dinámica y diversa que tiene una gran importancia para la salud humana. La colonización del tracto gastrointestinal en infantes es compleja y con una gran variabilidad interindividual. Para comprender como se desarrolla la funcionalidad de la microbiota intestinal es importante comprender tanto su composición como la expresión génica de sus miembros. En este trabajo se ha procedido al análisis del metatranscriptoma de la microbiota intestinal de una cohorte de infantes a lo largo del primer año de vida a partir de muestras fecales. El RNA total se purificó y se eliminaron los RNA ribosomales. El RNA restante se retrotranscribió y secuenció mediante la tecnología Illumina, y las secuencias obtenidas se procesaron, filtraron y analizaron mediante técnicas bioinformáticas, obteniéndose un conjunto de secuencias correspondientes a RNA mensajeros bacterianos. Estas secuencias se encuentran listas para su anotación funcional y posterior análisis estadístico de modo que ayuden a comprender como evoluciona la expresión génica de la microbiota intestinal durante este periodo.

Palabras clave: Metatranscriptómica, microbiota, intestino humano, infantes, expresión génica

Metatranscriptomic study of the intestinal microbiota development in infants

Author: Joan Compte Barrón

Tutor: José Gadea Vacas

Co-tutor: María Pilar Francino Puget

Collaborator co-tutor: María José Gosalbes Soler

Valencia, September 2015

ABSTRACT

The human gut constitutes the natural habitat of a dynamic and diverse bacterial community, which acquires a great importance for human health. The gastrointestinal track colonization in infants is complex and comprises a great inter-individual variability. In order to understand how the gut's microbiota functionality develops, it is important to comprehend not only its composition but also the gene expression of its members. In the present work, the metatranscriptome, that is the total RNA produced by the intestinal microbiota, was analysed in a cohort of infants, from which stool samples were collected along the first year of life. Total RNA was purified and ribosomal RNAs were eliminated. The remaining RNAs were retrotranscribed and sequenced by means of Illumina technology and the recovered sequences were bioinformatically processed, filtered and analysed to obtain a dataset corresponding to bacterial messenger RNAs. These sequences are ready for further bioinformatics and statistical analysis aimed at addressing how the intestinal microbiota evolves during this period.

Keywords: Metatranscriptomics, microbiota, human gut, infants, gene expression

AGRADECIMIENTOS

Aprovecho estas líneas para dar las gracias a todas aquellas personas que han hecho posible la realización de este trabajo y a todas aquellas que han contribuido a rendir de este primer peregrinaje por el camino del conocimiento una aventura realmente fascinante.

En primer lugar, quisiera destacar mi más profundo agradecimiento a Pilar Francino por haberme dado la oportunidad de formar parte de su equipo y por haber confiado en mí para llevar a cabo este complejo y exhaustivo trabajo. También por su supervisión a lo largo de su realización y sus correcciones en los bocetos preliminares de la versión escrita. En el plano personal, también aprovecho para agradecerle que me haya enseñado que ser el jefe de un grupo no lleva inherente la característica de estar siempre dando órdenes, ni la de destilar mal humor, sino más bien todo lo contrario, llegando a establecer relaciones de verdadera fraternidad.

También, y como no podía ser de otra manera, dar las gracias a Pepa Gosalbes por su infinita paciencia y por la gran cantidad de horas que me ha dedicado, transmitiéndome toda su enseñanza y sabiduría. También por haber sido mi acompañante en este proceso de constante aprendizaje. Asimismo por sus repetidas correcciones y mejoras de la versión escrita de este estudio. En el plano personal, por haberme enseñado que en la ciencia la clave es la constancia y el no desfallecer nunca, aunque los resultados no se obtengan a la primera y aunque estos no suelen ser los esperados.

A José Gadea, por haber accedido a ser mi tutor y haber estado siempre a mi disposición.

A Nuria Jiménez, por su agradable compañía en la instrucción y realización de las librerías. También por haber estado siempre disponible para resolver cualquier duda y haberme dado consejos y detalles acerca del proceso general de secuenciación. En el plano personal, por haberme enseñado que del buen humor se puede hacer una seña de identidad.

A Alejandro Artacho, por la realización de los *pipelines* que han permitido el procesamiento bioinformático de las muestras y obtención de los resultados finales. En el plano personal, por enseñarme a ser condescendiente, a saber escuchar y a no perder nunca la calma.

A todos los pre-docs (Alba, María, Sandrine, Caterina, Jorge,...), por haberme ayudado siempre en todas las dudas que me han surgido a lo largo de estos meses de trabajo y por haber aportado su grano de arena para convertir mi estancia en el laboratorio en un segundo hogar.

A los demás trabajadores del Área de Genómica y Salud de FISABIO (Llúcia, Loreto, Manoli, Alma, Concha, Sandra, Bea,...), por su constante contribución a la generación de buen clima en la gran familia que forman todos ellos en el laboratorio.

También recorro a estas líneas para acordarme de todas las personas increíbles, en especial de Alfredo Quijano, que me han acompañado a lo largo de estos 4 años de carrera y que, al igual que el proceso de colonización de la microbiota, han ido aportándome pequeños detalles y enseñanzas que han ido moldeando mi carácter contribuyendo a expandir enormemente el mosaico vital de mi persona.

A mis padres, abuelos y el resto de familiares con amor sublime y a Dios por todo lo demás.

ÍNDICE DE ABREVIATURAS

cDNA: DNA complementario

DNA: ácido desoxirribonucleico

DNasa: desoxirribonucleasa

dNTPs: desoxirribonucleótidos trifosfato

EDTA: ácido etilendiaminotetraacético

FISH: fluorescent *in situ* hybridization

MetaHIT: metagenomics of the Human Intestinal Tract

MIP: par madre-infante

MBU: unidad de biología molecular

mRNA: RNA mensajero

NEC: enterocolitis necrotizante neonatal

PCR: reacción en cadena de la polimerasa

RFLP: restriction fragment length polymorphism

RNA: ácido ribonucleico

RNasa: ribonucleasa

rRNA: RNA ribosómico

rpm.: revoluciones por minuto

S: coeficiente de sedimentación de Svedberg

spp: especies

tampón TE: tampón formado por tris y EDTA

TBE: tris-borato-EDTA

TGI: tracto gastrointestinal

Tris: trishidroximetilaminometano

tRNA: RNA de transferencia

TTO: terminal-tagging oligo

ÍNDICE

1. INTRODUCCIÓN	1
1.1. La microbiota intestinal.....	1
1.1.1. Características generales.....	1
1.1.2. Dinámica y maduración funcional.....	2
1.1.3. Factores que afectan a la colonización del intestino.....	4
1.1.3.1. Parto vaginal o por cesárea.....	5
1.1.3.2. Parto prematuro y antibióticos.....	6
1.1.3.3. Influencia de la dieta.....	6
1.2. Las estrategias “Ómicas”.....	7
1.2.1. Historia anterior a ellas.....	7
1.2.2. La metagenómica.....	8
1.2.3. La metatranscriptómica.....	8
1.3. Relevancia del estudio en bebés.....	9
2. OBJETIVOS	10
3. MATERIALES Y MÉTODOS	11
3.1. Esquema general.....	11
3.2. Cohorte infantil, recogida de muestras y procesamiento.....	12
3.3. Extracción de RNA total.....	13
3.4. Cuantificación del RNA.....	14
3.5. Electroforesis en gel de agarosa.....	14
3.6. Tratamiento con DNAasa.....	15
3.7. Enriquecimiento en mRNA.....	15
3.8. Síntesis de librerías.....	16
3.9. Evaluación de las librerías.....	18
3.10. Diluciones y “pool” de las librerías.....	18
3.11. Preparación librerías para el secuenciador.....	19
3.12. Secuenciación.....	19
3.13. Análisis bioinformático.....	20
3.13.1. Procesamiento y limpieza de las secuencias.....	20
3.13.2. Eliminación de los rRNA bacterianos, eucariotas y humanos..	20
4. RESULTADOS Y DISCUSIÓN	21
4.1. Parte experimental.....	21
4.2. Procesamiento bioinformático.....	29
5. CONCLUSIONES	31
6. BIBLIOGRAFÍA	32
7. ANEXO	37

ÍNDICE DE FIGURAS

Figura 1. Variación longitudinal y latitudinal del TGI.....	2
Figura 2. Enterotipos de la microbiota intestinal.....	3
Figura 3. Árbol filogenético comparativo de la microbiota fecal.....	4
Figura 4. Proceso de colonización intestinal y sus factores circundantes.....	5
Figura 5. Diagrama de flujo del trabajo.....	11
Figura 6. Protocolo de eliminación de rRNA.....	15
Figura 7. Esquema de la preparación de una librería de DNA.....	17
Figura 8. Electroforesis de algunas de las muestras analizadas.....	23
Figura 9. Relación entre concentración de RNA y edad de los infantes.....	24
Figura 10. Eficiencia de distintos métodos de reducción de rRNA.....	25
Figura 11. Electroferogramas de algunas de las librerías.....	28
Figura 12. Fórmula para obtener una concentración en nM.....	28
Figura S1. Cambios en la microbiota asociados a distintas enfermedades.....	37
Figura S2. Distribución de la calidad por ciclo del secuenciador.....	38
Figura S3. Distribución de la longitud y calidad de los <i>reads</i> antes del <i>joining</i> ..	39
Figura S4. Número de <i>reads</i> obtenidos por muestra.....	40
Figura S5. Distribución de las calidades antes de la evaluación cualitativa.....	41
Figura S6. Distribución de la longitud media de los <i>reads</i> vs su número por muestra.....	42
Figura S7. Distribución de la longitud promedio de los <i>reads</i> después del <i>joining</i>	43
Figura S8. Número de <i>reads</i> unidos y no unidos por muestra.....	44

ÍNDICE DE TABLAS

Tabla 1. Información acerca de las madres y sus bebés.....	12
Tabla 2. Concentraciones finales de RNA total.....	21
Tabla 3. Concentraciones finales de mRNA y porcentajes de reducción.....	26
Tabla 4. Concentraciones de las librerías.....	27
Tabla 5. Número de <i>reads</i> generados por el secuenciador.....	29
Tabla S1. Número de <i>reads</i> por muestra y sus longitud y desviación estándar medias.....	45
Tabla S2. Resumen después del <i>joining</i>	46
Tabla S3. <i>Reads</i> tras la limpieza frente a los iniciales y el porcentaje de reducción..	47
Tabla S4. Tipos de secuencias eliminados durante el filtrado.....	48

1. INTRODUCCIÓN

1.1. La microbiota intestinal

1.1.1. Características generales

El tracto gastrointestinal (TGI) constituye un complejo y dinámico ecosistema que incluye bacterias, virus, arqueas, protozoos y hongos, donde factores tanto bióticos como abióticos desempeñan funciones determinantes para alcanzar y mantener un equilibrio homeostático (Mackie et al., 1999; Collado et al., 2013; Harzallah et Belhadj, 2013). Se estima que lo pueblan entre 10^{13} y 10^{14} de estos microorganismos, cifras que superan en 10 veces el número total de células en el cuerpo humano y que resultan en una cantidad de genes microbianos que superan en 150 veces el número total de genes de nuestro genoma (Gill et al., 2006; Qin et al., 2010). En concreto, la microbiota intestinal se alza como la más diversa y densa de todo el cuerpo desarrollando actividades fundamentales en la maduración intestinal, la digestión, la absorción de nutrientes, la angiogénesis, la modulación del sistema inmune, el desarrollo del tejido linfoide asociado al intestino (GALT), la protección frente a patógenos y la regulación del almacenamiento de grasas (Palmer et al., 2007; Sekirov et al., 2010; Moore et al., 2011, Grenham et al., 2012). Recientemente, además, se ha comprobado su capacidad para modular la homeostasis ósea (Sjörger et al., 2012) y, además, constituye un importante reservorio de genes que confieren resistencia frente a antibióticos denominado resistoma (Forslund et al., 2013).

A día de hoy, se ha demostrado que el intestino adulto presenta una composición, más o menos homogénea, de microorganismos predominando los filos bacterianos Bacteroides y Firmicutes en un 95% y, en menor abundancia, Proteobacteria y Actinobacteria (Palmer et al., 2007). No obstante, existe una cierta variabilidad a niveles filogenéticos más bajos tanto entre las distintas partes del TGI como entre individuos ocasionando que cada intestino presente un mosaico microbiano único (Sekirov et al., 2010; Vallès et al., 2014), (ver figura 1). Como se muestra en dicha figura, es destacable mencionar cómo a medida que se avanza a las partes más distales del TGI, el número de microorganismos aumenta y cómo la microbiota en el lumen intestinal difiere significativamente respecto a las adsorbidas y embebidas en la capa mucosa y a las del epitelio.

Por otro lado, el intestino infantil es mucho más variable en composición y menos estable en el tiempo progresando desde lo que se creía anteriormente como la esterilidad en el feto, a una colonización masiva durante el primer año de vida, al final del cual se alcanzan una estabilidad y una complejidad de la microbiota que se asemejan a las de la microbiota adulta (Mackie et al., 1999; Palmer et al., 2007; Dogra et al., 2015). No obstante, cabe puntualizar que recientes estudios (Gosalbes et al., 2013) han demostrado mediante el análisis del meconio (primera deposición) que el TGI no es estéril como se creía años atrás (Pietzak, 2004; Palmer et al., 2007).

Asimismo, una vez establecida, la composición de la microbiota intestinal ha sido caracterizada como relativamente estable a lo largo de la vida adulta, pero, no obstante, puede ser alterada como resultado de infecciones bacterianas patógenas, tratamiento con antibióticos, estilo de vida, cirugías y un cambio a largo plazo de la dieta (Rodríguez et al., 2015). Se ha demostrado que desviaciones estructurales en este complejo sistema microbiano pueden aumentar el riesgo de padecer enfermedades, incluyendo trastornos tan diversos como la diabetes de tipo 2, la enfermedad de Crohn, distintas alergias, la obesidad, la enfermedad inflamatoria intestinal, la anorexia, el cáncer gástrico o el autismo (ver figura 1 del anexo).

Respecto a aquellas desviaciones en la composición de la microbiota que causan disfunciones de las capacidades cognitivas, resulta muy llamativo destacar la estrecha relación entre el cerebro y el intestino. El término que describe esta interacción se denomina eje cerebro-intestino y hace referencia al complejo sistema de comunicación bidireccional que existe entre el sistema nervioso central y el TGI, el cual es vital para el mantenimiento de la homeostasis. De este modo, señales desde el cerebro pueden influenciar las capacidades motoras, sensoriales y secretoras del TGI y, reversamente, mensajes viscerales pueden modular funciones cerebrales (Grenham et al., 2011).

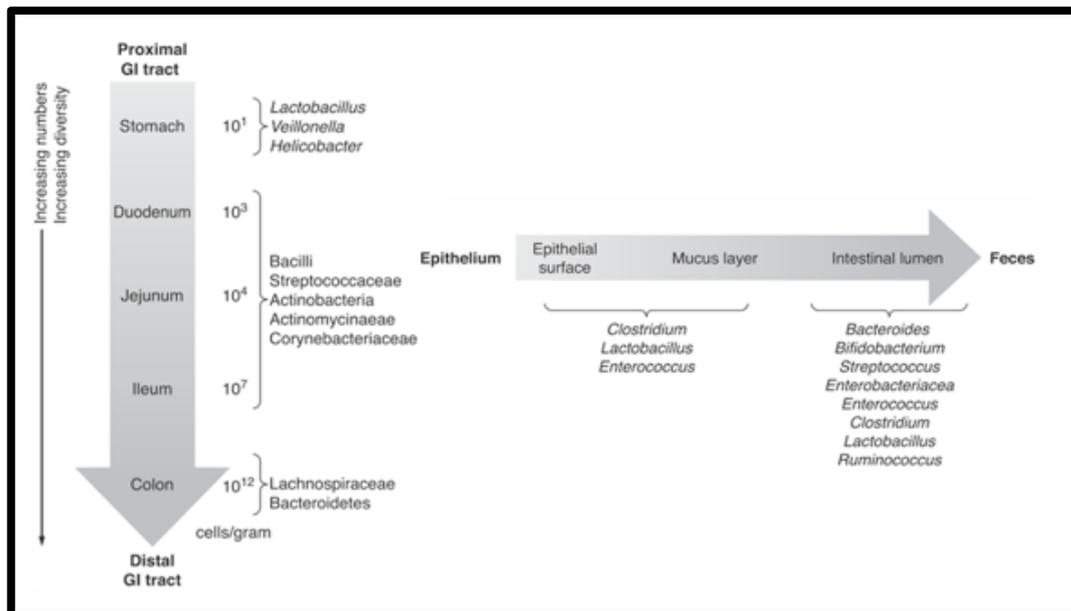


Figura 1: Izquierda) Variación longitudinal en el número y la composición bacteriana a lo largo del TGI. Derecha) Variación transversal en la composición del TGI. (Modificado de Sekirov et al., 2010).

1.1.2. Dinámica y maduración funcional

Como se ha comentado, la variabilidad interindividual hace muy compleja una descripción general de la microbiota intestinal. No obstante, sí que se ha establecido un microbioma base ("core") (Turnbaugh et al., 2007), ya que se ha demostrado que, una vez, alcanzada la estabilidad, si bien no a nivel de especies, la microbiota de distintos individuos adquiere una cierta similitud a nivel funcional, siendo cada una asignada a uno de los tres enterotipos propuestos por Arumugam (2011). El *core* supone el mínimo permitido para el correcto funcionamiento del ecosistema intestinal (ver figura 2), (Qin et al., 2010, Arumugam et al., 2011; Korecka et Arulampalam, 2012).

Desde un punto de vista ecológico, la primera etapa de la colonización intestinal representa un ensamblaje de novo (Costello et al., 2012) y se ve muy influenciado por factores externos que se comentarán más adelante. No obstante, todavía no está claro cómo éstos contribuyen a la composición general y a la función del microbioma infantil y cómo los diferentes microorganismos que lo conforman cooperan o compiten entre ellos mientras el ambiente intestinal va variando a lo largo del tiempo (Bäckhed et al., 2015).

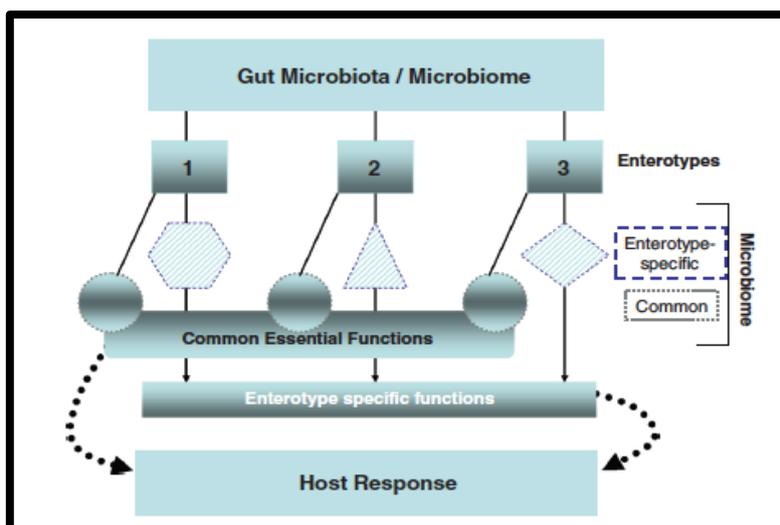


Figura 2: La microbiota puede ser clasificada en tres posibles enterotipos que presentan funciones comunes entre ellos, pero también específicas de cada uno. (Korecka et Arulampalam, 2012).

A grandes rasgos, cabe destacar varios procesos sucesivos que conllevan una maduración funcional y preparan al TGI infantil para alcanzar la plena actividad. Los análisis metagenómicos han revelado el empleo de distintas fuentes de energía a lo largo del tiempo. En recién nacidos abundan los genes del sistema fosfotransferasa implicados en el transporte y utilización de carbohidratos simples. Así, en los infantes de 4 meses de edad la microbiota presenta los genes de utilización del lactato, como el transportador específico de lactosa, propios de una dieta dominada por la leche (Koenig et al., 2011; Bäckhed et al., 2015). En común, los microbiomas tanto de recién nacidos como en los bebés de 4 meses están enriquecidos en genes requeridos para la degradación de los azúcares de la leche, principal fuente nutritiva en estos dos grupos.

Sin embargo, como resultado de la evolución sucesiva del TGI, los infantes de un año de edad, ya muestran genes implicados en la degradación de azúcares más complejos, almidón y pectina, resultado de la ingesta de alimentos ya sólidos a esta edad (Bäckhed et al., 2015). A su vez, sólo en bebés de esta edad, se observan genes implicados en la producción de metano; hecho asociado al incremento en la capacidad fermentativa propia del microbioma adulto (Hoffmann et al., 2013). Estas observaciones demuestran la progresión en el asentamiento bacteriano en el intestino infantil y arrojan datos reveladores acerca de su desarrollo (ver figura 3).

Por otro lado, comprender la estabilidad de la microbiota de un individuo a través del tiempo es un paso importante que posibilitará la predicción de enfermedades y el desarrollo de terapias que corrijan las correspondientes disbiosis (Clemente et al., 2012). Respecto a esta afirmación ya se han presentado estudios (Caporaso et al., 2011 y Costello et al., 2009) que demuestran que la composición de la microbiota es estable en el tiempo en individuos sanos. No obstante, esta consistencia temporal asume que numerosas variables descritas a continuación también se mantienen constantes (Clemente et al., 2012). De hecho, estudios tanto en ratones (Turnbaugh et al., 2009) como en humanos (Wu et al., 2011) demostraron que cambios bruscos en la composición de la dieta (pasar de una rica en grasas y baja en fibra a una pobre en grasas y alta en fibra, por ejemplo) son capaces de alterar la microbiota en tan sólo 24 horas. De este modo, estos estudios longitudinales permiten diferenciar aquellos microorganismos transitorios que provienen del ambiente de aquellos miembros permanentes de la microbiota y elucidar diferencias entre las

interacciones del huésped con cada uno de los dos grupos. Además sientan las bases para estudios de caracterización microbiana a gran escala y más exhaustivos.

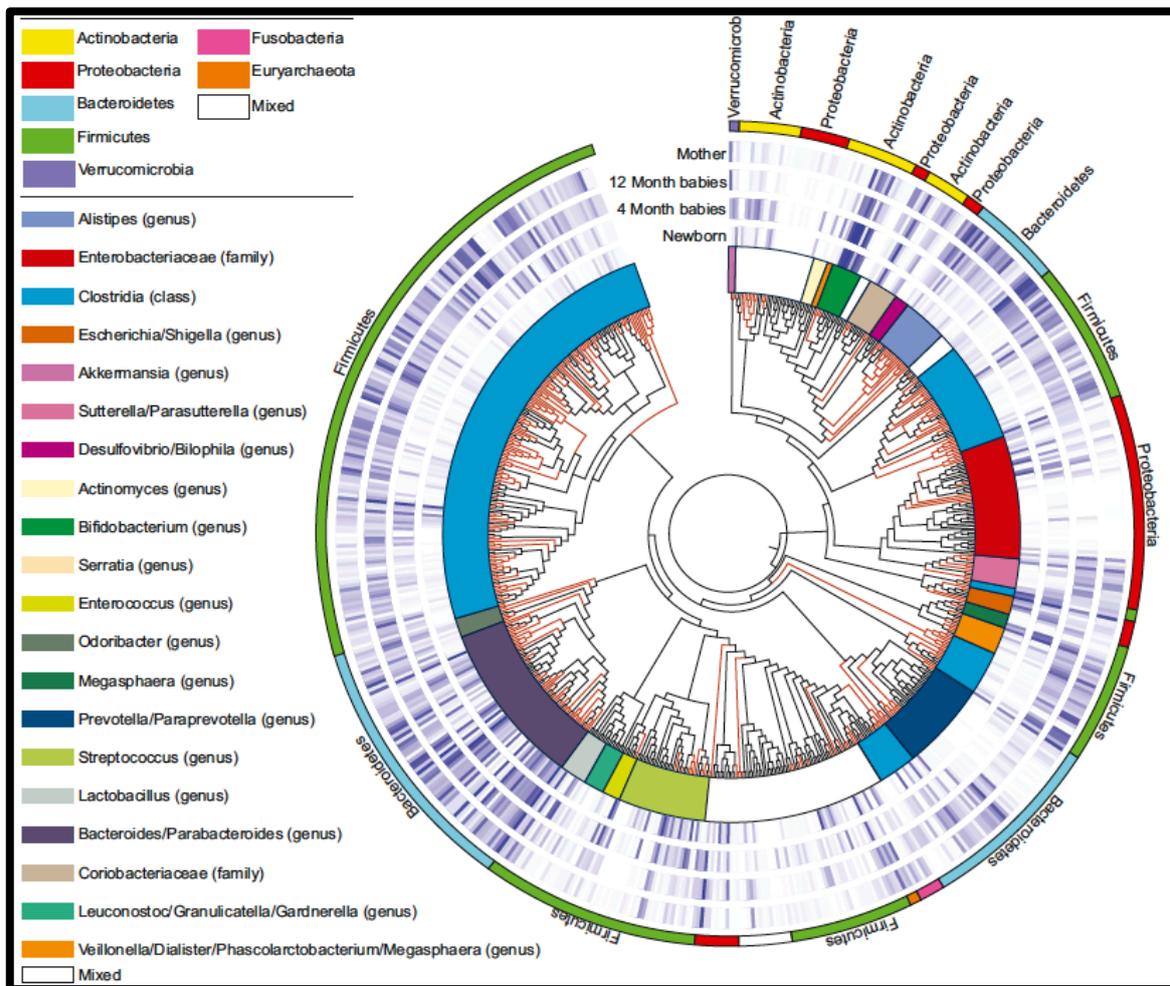


Figura 3: Árbol filogenético de la microbiota fecal presente en recién nacidos, en infantes de 4 meses y 1 año. También se representa la materna a modo de comparativa. (Bäckhed et al., 2015).

1.1.3. Factores que afectan a la colonización del intestino

Desde que empezaron los primeros estudios sobre la microbiota intestinal hasta hace una década escasa (Palmer et al., 2007), estaba comúnmente aceptada la teoría de que el TGI estaba exento de microorganismos al nacer y que la colonización comenzaba durante el parto (Tissier, 1900; Palmer et al., 2007). Sin embargo desde 2008 (Jiménez et al., 2008) ha sido constatado que el TGI es colonizado de modo prenatal, sugiriéndose diversas hipótesis al respecto (Rautava et al., 2012; Gosalbes et al., 2013; Rodríguez et al., 2015), pero la función y, principalmente, los mecanismos asociados a esta exposición microbiana permanecen aún irresolutos.

Por otro lado, sí que están bien establecidas las etapas por las que el TGI del recién nacido es colonizado y los factores involucrados en dicho proceso. Desde el paso a través del canal de parto y durante las primeras horas de vida, los microbiomas vaginal, fecal y dermal de la madre, así como la microbiota ambiental, constituyen la principal fuente de inóculo (Mandar

et Mikelsaar, 1996; Koenig et al., 2010; Domínguez-Bello et al., 2010). Hasta el primer año, la composición microbiana intestinal sufre de modo constante sucesiones de comunidades bacterianas, siendo la variación entre individuos muy notoria, hasta que, finalmente, tras este aumento gradual tanto en número como en diversidad de especies, alcanza una mayor estabilidad, asemejándose a la microbiota adulta (Mackie et al., 1999; Sekirov et al., 2010; Vallès et al 2014) (ver figura 4).

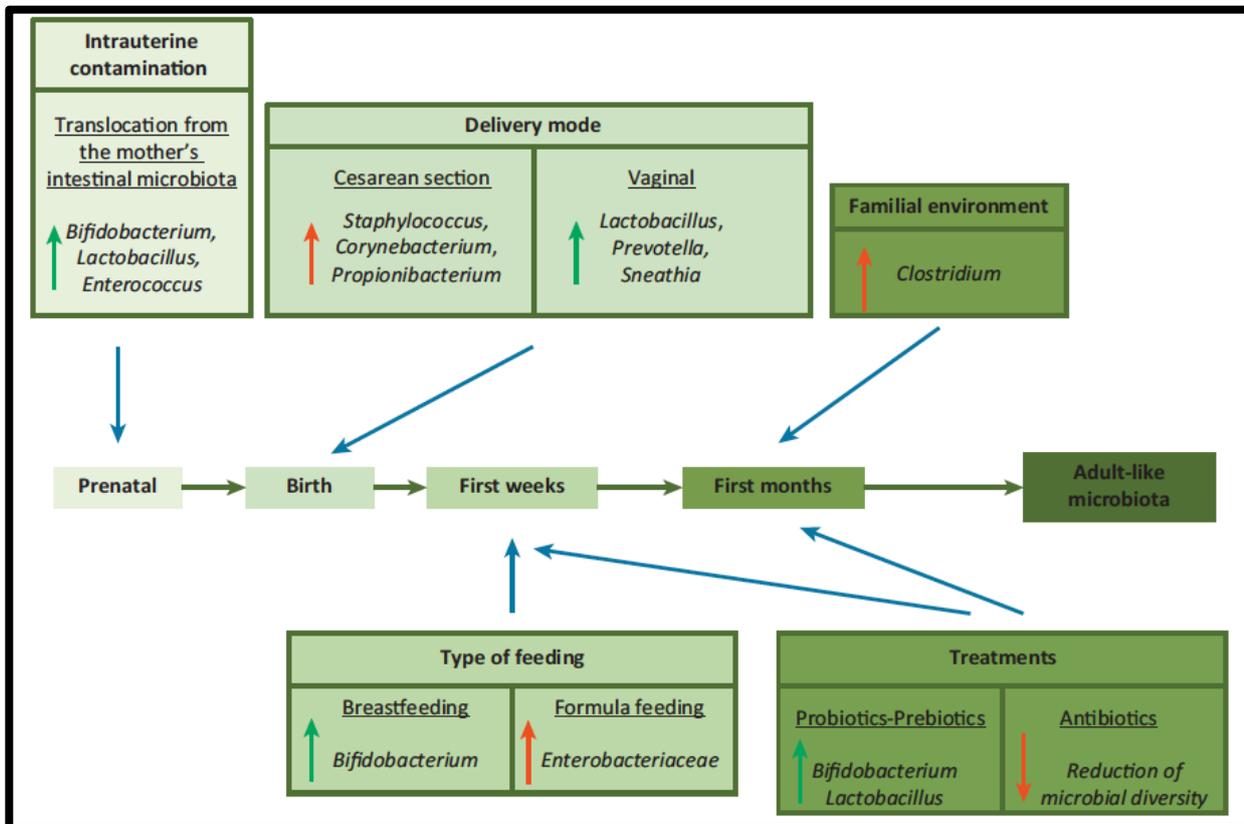


Figura 4: Proceso de colonización del intestino humano las flechas verdes indican modificaciones beneficiosas; en rojo las consideradas como negativas para un desarrollo saludable. (Matamoros et al., 2013).

Como muestra la figura 4, los factores que marcan la dinámica de colonización intestinal durante el primer año de vida son muy variados. Los más destacados se presentan a continuación.

1.1.3.1 Parto vaginal o por cesárea

El modo de parto es determinante para la colonización inicial del intestino. Los neonatos que nacen por parto vaginal son enfrentados a bacterias maternas vaginales y fecales, particularmente *Lactobacillus*, *Enterobacteriaceae* y *Bifidobacterium spp.* (Moore et al., 2014; Rodríguez et al., 2015). Sin embargo, los recién nacidos por cesárea no son expuestos directamente a los microbios maternos comentados y, en su lugar, son colonizados con los dermales maternos, los ambientales (*Staphylococcus*, *Corynebacterium* y *Propionibacterium spp.*) y los del entorno hospitalario (Domínguez-Bello et al., 2010; Marques et al., 2010; Savino et al., 2010). De hecho, ha sido demostrado que la microbiota de los neonatos por parto vaginal se parece en composición a la comunidad bacteriana

vaginal, indicando su masiva influencia, y que presenta una mayor diversidad que en las pertenecientes a recién nacidos por cesárea (Jakobsson et al., 2014). Cabe destacar la importancia de esta última característica ya que enfermedades asociadas al sistema inmune como las alergias, son más prevalentes en personas que nacieron mediante cesárea (Bager et al., 2008).

1.1.3.2. Parto prematuro y antibióticos

El desarrollo óptimo de la microbiota intestinal se ve perturbado a raíz de un parto prematuro y, en ocasiones, del suministro de antibióticos, teniendo implicaciones negativas a largo plazo tanto para la diversidad microbiana como para la salud (Rodríguez et al., 2015). En niños prematuros la microbiota se caracteriza por ser poco diversa y presentar altos niveles de bacterias potencialmente patógenas, así como una baja abundancia de *Bifidobacterium* y *Bacteroides* en comparación con niños nacidos tras gestación completa (Barrett et al., 2012). Esto se debe a que el TGI es inmaduro tanto funcional como morfológicamente, previniendo la correcta digestión y absorción de importantes nutrientes y, además, el pH, que constituye una barrera fisio-química clave en el propio TGI, es más elevado que en infantes de gestaciones completas, ocasionando la alteración del correcto establecimiento de la microbiota (Moore et al., 2011; Barrett et al., 2013). Es más, enfermedades inflamatorias severas como la enterocolitis necrotizante neonatal (NEC) están asociadas a partos prematuros (Rodríguez et al., 2015).

Asimismo cabe destacar que otro factor relacionado con el parto prematuro y que también conlleva importantes mermas y perturbaciones en la microbiota intestinal es el suministro de antibióticos en el periodo perinatal (Rautava et al., 2012). Éstos suelen aplicarse para tratar infecciones como la corioamnionitis y también de modo profiláctico para prevenir infecciones. No obstante, su uso prolongado está asociado a NEC y parálisis cerebral, así como riesgo de muerte en bebés prematuros (Cotten et al., 2009).

1.1.3.3. La dieta

Uno de los factores determinantes para el asentamiento de la microbiota intestinal, y de especial relevancia en el presente trabajo, lo constituye la dieta. En primer lugar, cabe destacar la leche como una fuente principal de colonización. Ésta, además de contener nutrientes y elementos protectores como la lisozima y la lactoferrina (Cabrera-Rubio et al., 2012), supone un constante suministro de bacterias (hasta 700 especies distintas han sido reveladas) comensales, mutualísticas y, potencialmente, probióticas, incluyendo principalmente bifidobacterias (Fernández et al., 2012; Cabrera-Rubio et al., 2012; Rautava et al., 2012; Rodríguez et al., 2015). Además, el crecimiento de algunas de ellas es estimulado por otros componentes de la leche como son los oligosacáridos complejos prebióticos (Rodríguez et al., 2015). Por otro lado, es interesante describir qué sucede en aquellos niños con condiciones que les impiden ser alimentados con leche materna y han de ser nutridos con fórmula. Recientes estudios (Adlerberth et Wold, 2009; Moore et al., 2011; Bezirtzoglou et al., 2011) han demostrado que las microbiotas intestinales de niños alimentados por fórmula son más diversas que las lactantes pero menos ricas en bifidobacterias. Además, los niños alimentados por leche materna están mejor protegidos frente a infecciones debido a que las fórmulas no atesoran todos los constituyentes inmunológicos de la leche. Sin embargo, el desarrollo de fórmulas que tratan de reproducir

fidedignamente la composición de la leche materna, incluyendo además un mayor aporte nutricional, está en constante evolución (Guaraldi et Salvatori, 2012).

El siguiente paso clave en el proceso de colonización es el que supone la introducción de alimentos sólidos a la dieta. Este hecho implica un gran cambio en las abundancias de las especies (es destacable el rápido aumento en Bacteroidetes y Firmicutes y de la carga bacteriana en general), un enriquecimiento en genes funcionales responsables de la degradación de carbohidratos y xenobióticos así como de la biosíntesis de vitaminas característicos del microbioma adulto, además de un notable incremento en ácidos grasos de cadena corta (Koenig et al., 2011; Matamoros et al., 2012). De hecho, en su estudio, Palmer et al. (2007) demostraron que la nutrición por alimentos sólidos conlleva un cambio en la microbiota intestinal, estableciéndose consorcios de bacterias propios de la adulta. A su vez, es muy notorio el hecho de que incluso antes del destete se han encontrado enzimas microbianas capaces de degradar polisacáridos de origen vegetal (Vaishampayan et al., 2010), sugiriendo que los cambios evolutivos de la microbiota preparan durante los primeros meses al intestino para estar metabólicamente listo para la introducción de alimentos sólidos (Koenig et al., 2011).

1.2. Las estrategias “Meta-ómicas”

1.2.1. Técnicas precedentes en el estudio del microbioma

Los primeros estudios realizados sobre la microbiota intestinal fueron llevados a cabo en los albores de la década de los 70 mediante técnicas de cultivo, siendo unos de los pioneros Moore y Holdeman en 1974. Si bien es cierto que dichos métodos identificaron más de 400 especies bacterianas en el intestino humano, el empleo de técnicas de cultivo presenta muchas desventajas (Dave et al., 2012). Por ejemplo, cabe destacar que el ambiente anóxico en el intestino provoca que sólo puedan habitar en él anaerobios facultativos o estrictos, tipos de bacterias difíciles de cultivar o directamente incultivables (entre el 60% y el 80%) (Adlerberth et Wold, 2009; Bezirtzoglou et al., 2011; Dave et al., 2012). Además, habitualmente los medios de cultivo no son verdaderamente específicos y son demasiado selectivos para algunas bacterias. Todo esto ocasionó que mediante técnicas de cultivo fueran inviables tanto la reproducción como el estudio del ecosistema microbiano intestinal (Bezirtzoglou et al., 2011; Dave et al., 2012). Por lo tanto, la necesidad de superar estos importantes inconvenientes hizo que en los años siguientes se desarrollaran las técnicas moleculares independientes de cultivo. Entre ellas destacan los geles de electroforesis desnaturalizantes en gradiente, la hibridación fluorescente *in situ* (FISH) y el análisis de RFLP terminales (Li et al., 2006; Nocker et al., 2007; Bezirtzoglou et al., 2011). A su vez, también se desarrollaban estudios empleando microarrays de DNA (Trosvik et al., 2010; Vebø et al., 2010) que, sin embargo, están cayendo en cierto desuso debido a las limitaciones en su aplicabilidad y al acotado rango de detección de los niveles de transcritos debido al ruido de fondo (“background”), así como a la saturación y densidad de los spots (van Vliet, 2009; Bloom et al., 2009; Reigstad et Kashyap, 2013). Posteriormente, se han generalizado las técnicas basadas en la secuenciación del gen ribosomal 16S (16S rRNA) de bacterias. Este gen presenta dominios muy conservados flanqueados por regiones hipervariables propias de cada especie lo que permite su asignación taxonómica (Turnbaugh et al 2007; Guaraldi et Salvatori, 2012; Dave et al., 2012) y el estudio de la filogenia bacteriana a un mayor nivel de profundidad (Young et Schmidt, 2003).

No obstante, a pesar del enorme progreso que supusieron las aproximaciones por secuenciación que empleaban el gen 16S rRNA en la determinación de la comunidad

bacteriana intestinal, no se aportaba información alguna sobre la fisiología bacteriana y el significado ecológico de dicha microbiota (DeLong et Pace, 2001). Con todo, tanto la percepción, como la magnitud de conocimiento y el análisis del microbioma intestinal, en particular, sufrieron una profunda mejora tras la aplicación de las tecnologías de secuenciación de nueva generación en dicho campo y el concomitante nacimiento de las “meta-ómicas” (Franzosa et al., 2014).

1.2.2. La Metagenómica

Se ha estimado que en un cuerpo humano hay alrededor de 100 trillones de bacterias de las cuales la mayoría habita el intestino (Ley et al., 2006). Dada la complejidad que genera este macropoblado ecosistema, el avance de las tecnologías de secuenciación masiva y, sobretudo, el abaratamiento de sus costes, ha sido clave para poder abordar análisis de esta comunidad tan compleja (Dave et al., 2012). Este inexorable desarrollo permitió el nacimiento, entre otras, de la metagenómica. Éste término se acuñó para englobar la técnica que implica construir librerías de DNA de un ecosistema bacteriano y, posteriormente, analizar sus secuencias y funciones (Riesenfeld et al., 2004). Desde entonces, el constante progreso de los métodos tanto de preparación de librerías y secuenciación como de análisis bioinformático ha permitido la realización de numerosos estudios en el campo de la microbiota intestinal, tanto en laboratorios individuales como bajo el amparo de importantes proyectos internacionales como el Microbioma Humano, una rama especializada del proyecto del genoma humano, y el proyecto europeo MetaHIT. En conjunto, estos proyectos han logrado caracterizar en gran medida la composición y dinámica del microbioma humano (Dethlefsen et al., 2006; Turnbaugh et al., 2007; Palmer et al., 2007; Sekirov et al., 2010; Qin et al., 2010; Koenig et al., 2011; Yatsunencko et al., 2012; Vallès et al., 2014; Rodríguez et al., 2015).

1.2.3. La Metatranscriptómica

No obstante, pese a la gran utilidad de la metagenómica para obtener información acerca de la diversidad, contenido genético y sus potenciales funciones de comunidades bacterianas complejas, esta aproximación no permite establecer la expresión génica de la comunidad bacteriana intestinal (Gosalbes et al., 2011; Celaj et al., 2014). Por ello, paralelamente, y dada la necesidad de entender qué genes y qué rutas génicas se expresan *in vivo* en el TGI (tanto en condiciones normales como en las fisiológicamente alteradas), así como su regulación, se ha ido desarrollando la metatranscriptómica (Poroyko et al., 2010; Giannoukos et al. 2012; Franzosa et al., 2014). Esta reciente herramienta permite estudiar cómo cambian los fenotipos de las comunidades bacterianas con el tiempo, analizando su RNA, en respuesta a las distintas exposiciones ambientales y a distintas condiciones patológicas (Poroyko et al., 2010; Reigstad et Kashyap, 2013). Asimismo, estudios muy exhaustivos en muestras derivadas de ambientes oceánicos han demostrado que el análisis metatranscriptómico ha permitido describir nuevos transcritos no anotados anteriormente, correspondiendo a nuevos genes no caracterizados por las investigaciones metagenómicas (Poroyko et al., 2010; Franzosa et al., 2014). A día de hoy aún son escasos los estudios que emplean la metatranscriptómica como herramienta para caracterizar funcionalmente la microbiota intestinal (Booijink et al., 2010; Gosalbes et al., 2011; Franzosa et al., 2014) no habiéndose publicado aún estudios en infantes. No obstante, a pesar de la valiosa información que éstos van aportando, el análisis bioinformático supone la principal área de investigación y desarrollo. Muestra de ello es que se está consiguiendo pasar de la

asignación de secuencias a familias de genes mediante bases de datos a explorar los análisis de las rutas bioquímicas y redes metabólicas (Xiong et al., 2012; Mitra et al., 2011).

1.3. Relevancia del estudio en bebés

Como se destacaba en la introducción, la microbiota intestinal juega un papel clave en una gran variedad de funciones que contribuyen al bienestar y correcto funcionamiento del organismo que pueden ser clasificadas en 3 tipos principales: la metabólica, la trófica y la protectora (Guarner et Malagelada, 2003). Como el desarrollo de esta microbiota ocurre, primariamente, durante la infancia y una alteración de su equilibrio (disbiosis) o de sus funciones puede, potencialmente, contribuir a un amplio elenco de enfermedades (enfermedades atópicas, obesidad y diabetes), el estudio en infantes está cobrando, cada vez más, notable interés (Vael et al., 2009, Rautava et al., 2012; Korecka et Arulampalam, 2012; Prince et al., 2014). Asimismo, se ha constatado que las interacciones entre la microbiota y el huésped durante la infancia son un factor determinante para gozar de salud a largo plazo o padecer alguna enfermedad (Penders et al., 2007; Vallès et al., 2014). Sin embargo, pese al preponderante papel de la microbiota infantil en la salud humana (Bäckhed et al., 2015), los estudios metagenómicos focalizados en abordar el estudio de la diversidad microbiana en el TGI y detectar las desviaciones propias de enfermos, se realizan mayoritariamente en adultos (Turnbaugh et al., 2007; Turnbaugh et al., 2009; Arumugam et al., 2011; Collado et al., 2013). Si bien es cierto, en los últimos años se están realizando prometedores avances con cohortes infantiles (Vallès et al., 2014; Bäckhed et al., 2015).

Con todo, a pesar de su importancia, los procesos de colonización microbiana en el TGI de infantes, permanecen, a día de hoy, poco entendidos y se precisan de muchos más estudios para su comprensión a gran escala (Poroyko et al., 2010, Vaishampayan et al., 2010). Esta falta de información se acrecienta en el aspecto funcional de dicha microbiota, ya que hay escasos trabajos que aborden el metatranscriptoma durante la infancia.

2. OBJETIVOS

El presente trabajo constituye un estudio pionero ya que trata de caracterizar específicamente la evolución de la expresión génica (metatranscriptoma) de la microbiota intestinal de un infante a lo largo del primer año de vida.

Asimismo, está en perfecta sinergia con el estudio realizado por Vallès et al. (2014), que analizó la sucesión microbiana en el intestino haciendo hincapié en los patrones taxonómicos y contenido génico, pudiéndose obtener por primera vez, una visión global de la dinámica microbiana en el TGI durante el primer año de vida.

A su vez, para poder realizar un análisis más completo, también se procesaron muestras fecales de las madres una semana antes del parto y un año después. Gracias a ello, el presente trabajo también permite analizar cómo influye el estado gestante en la microbiota intestinal de las progenitoras.

3. MATERIALES Y MÉTODOS

3.1. Esquema general

El diagrama de flujo mostrado en la figura 5 hace referencia al esquema general seguido a lo largo del presente trabajo. En primer lugar, se partió de una suspensión bacteriana obtenida a partir de una muestra fecal, de la que se aisló el RNA total eliminando los rRNA y tRNA para obtener mRNA. Seguidamente, se sintetizó su cDNA mediante hexámeros aleatorios acompañado por la ligación de adaptadores. Posteriormente estas librerías de cDNA fueron secuenciadas empleado el equipo MiSeq (Illumina) y, por último, se inició el análisis bioinformático.

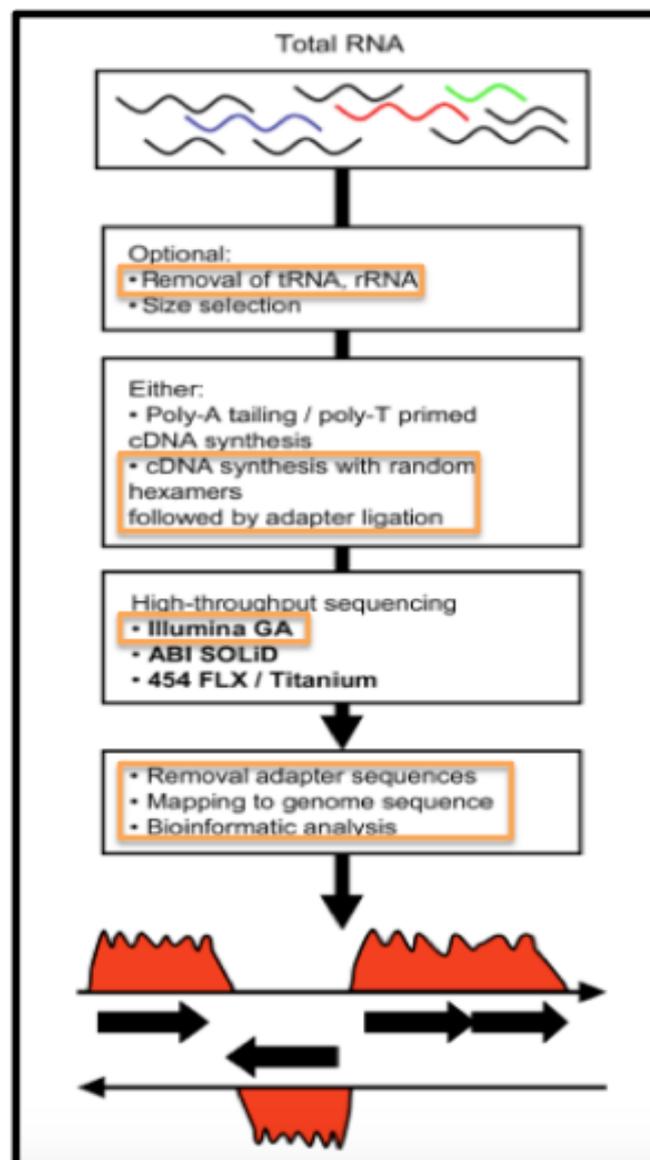


Figura 5: Diagrama de flujo de los pasos implicados en la secuenciación de un transcriptoma microbiano. Nótese que los pasos adoptados en el presente trabajo están bordeados en un recuadro naranja (van Vliet, 2009).

3.2. Cohorte infantil, recogida de muestras y procesamiento

El objetivo de este estudio es explorar los cambios que suceden en el metatranscriptoma intestinal de infantes durante su primer año de vida, complementando de este modo el trabajo de Vallès et al. (2014), donde se realizó el correspondiente análisis taxonómico y de composición génica de la comunidad bacteriana intestinal. Con este propósito, el presente estudio se ha desarrollado con las muestras empleadas en el trabajo de Vallès et al. (2014), para el cual se seleccionaron mujeres con embarazos sanos y que declararan su intención de alimentar a sus bebés, exclusivamente, con leche materna durante, al menos, tres meses.

Vallès et al. (2014) analizaron las muestras fecales de 13 pares madre-infante (MIPs en tabla 1) en dos puntos temporales concretos para la madre, una semana antes del parto (MA) y un año después de él (MB), y en cinco diferentes ocasiones para el infante, a la semana del nacimiento (I1), al mes (I2), a los tres meses (I3, antes de la introducción de alimentos sólidos), a los 7 meses (I4, después de la introducción de alimento sólidos) y al año (I5). No obstante, para el análisis del metatranscriptoma del presente trabajo no quedaba suficiente material para muchas de las muestras recogidas por lo que se pudieron seleccionar un reducido número de siete pares madre-infante, de los que las muestras I2 no estaban disponibles. Por tanto, fueron 42 las muestras seleccionadas para llevar a cabo este estudio, de las cuales todas menos una (MIP09-MB) llegaron al último paso de secuenciación.

Los 7 bebés de los MIPs analizados tuvieron un periodo de gestación normal (más de 37 semanas): 5 nacieron por parto vaginal y 2 por cesárea. Seis de ellos fueron alimentados exclusivamente con leche materna durante, al menos, tres meses y uno fue alimentado con leche materna durante el primer mes y mediante fórmula a partir de entonces.

Tabla 1: Información acerca de las madres y sus bebés. Tabla adaptada de Vallès et al. (2014).

MIP: Mother Infant Pair (par madre-infante)

a. Para las muestras MA se informa si se suministraron antibióticos durante el parto y cuál en particular. En el caso de cesáreas se reporta la administración de amoxicilina que supone la práctica estándar en hospitales españoles. Ninguna de las madres había tomado antibióticos antes del parto durante, al menos, tres meses.

b. Oftalmowell es una combinación de gramicidina, neomicina y polimixina B.

Muestra	Edad	Sexo	Parto	Antibióticos madre ^a	Antibióticos bebés	Dieta	Secuenciación realizada
MIP01-MA	29 años	F	-	no	-	-	Sí
MIP01-MB	30 años	F	-	-	-	-	Sí
MIP01-I1	1 sem.	M	vaginal	-	-	Leche	Sí
MIP01-I3	3 meses	M	vaginal	-	-	Leche	Sí
MIP01-I4	7 meses	M	vaginal	-	-	Sólidos	Sí
MIP01-I5	1 año	M	vaginal	-	-	Sólidos	Sí
MIP03-MA	30 años	F	-	no	-	-	Sí
MIP03-MB	31 años	F	-	-	-	-	Sí
MIP03-I1	1 sem.	F	vaginal	amoxicilina	Oftalmowell ^b	Leche	Sí
MIP03-I3	3 meses	F	vaginal	cefuroxima	-	Leche	Sí
MIP03-I4	7 meses	F	vaginal	-	-	Sólidos	Sí
MIP03-I5	1 año	F	vaginal	amoxicilina	cefuroxima	Sólidos	Sí
MIP06-MA	42 años	F	-	amoxicilina	-	-	Sí
MIP06-MB	43 años	F	-	-	-	-	Sí
MIP06-I1	1 sem.	F	cesárea	-	-	Leche	Sí

MIP06-I3	3 meses	F	cesárea	-	-	Leche	Sí
MIP06-I4	7 meses	F	cesárea	-	-	Sólidos	Sí
MIP06-I5	1 año	F	cesárea	-	amoxicilina	Sólidos	Sí
MIP07-MA	31	F	-	amoxicilina	-	-	Sí
MIP07-MB	32	F	-	-	-	-	Sí
MIP07-I1	1 sem.	M	cesárea	-	-	Leche	Sí
MIP07-I3	3 meses	M	cesárea	-	-	Leche	Sí
MIP07-I4	7 meses	M	cesárea	-	-	Sólidos	Sí
MIP07-I5	1 año	M	cesárea	-	-	Sólidos	Sí
MIP08-MA	30	F	-	no	-	-	Sí
MIP08-MB	31	F	-	-	-	-	Sí
MIP08-I1	1 sem.	F	vaginal	-	-	Leche	Sí
MIP08-I3	3 meses	F	vaginal	-	-	Leche	Sí
MIP08-I4	7 meses	F	vaginal	-	-	Sólidos	Sí
MIP08-I5	1 año	F	vaginal	-	-	Sólidos	Sí
MIP09-MA	30	F	-	no	-	-	Sí
MIP09-MB	31	F	-	-	-	-	No
MIP09-I1	1 sem.	M	vaginal	amoxicilina	-	Mezclado	Sí
MIP09-I3	3 meses	M	vaginal	-	-	Papilla	Sí
MIP09-I4	7 meses	M	vaginal	-	-	Sólidos	Sí
MIP09-I5	1 año	M	vaginal	-	-	Sólidos	Sí
MIP016-MA	39	F	-	amoxicilina	-	-	Sí
MIP016-MB	40	F	-	-	-	-	Sí
MIP016-I1	1 sem.	M	vaginal	-	-	Leche	Sí
MIP016-I3	3 meses	M	vaginal	-	-	Leche	Sí
MIP016-I4	7 meses	M	vaginal	-	-	Sólidos	Sí
MIP016-I5	1 año	M	vaginal	-	-	Sólidos	Sí

En el laboratorio, las muestras fueron homogeneizadas en una solución RNA later/tampón fosfato alcalino y centrifugadas a 4°C durante 2 minutos a 2500 rpm para eliminar residuos fecales. El sobrenadante resultante que contiene la comunidad bacteriana fecal se almacenó a -80°C hasta su uso.

3.3. Extracción del RNA total

Antes de detallar el protocolo cabe destacar que trabajar con RNA requiere un cuidado extremo, dado que es una molécula extremadamente lábil y que, por ello, es necesario desarrollar las técnicas y procedimientos correspondientes en ambientes libres de RNAasas ("RNase-free"). En nuestro trabajo, se habilitó un banco específico en el laboratorio que asegurara dichas condiciones para llevar a cabo las extracciones de las muestras fecales. Asimismo, cada vez, antes de comenzar, se lavaron tanto el banco como todos los utensilios que fueran a ser utilizados con etanol 70°.

Para extraer el RNA total de la suspensión bacteriana se empleó el kit RiboPure™ Bacteria (Ambion, Life Technologies). Este kit que se puede emplear para aislar RNA de casi cualquier bacteria gram-positiva o gram-negativa, combina la rotura de las paredes celulares bacterianas con fenol y una purificación del RNA mediante filtrado por columnas.

En primer lugar, para extraer el RNA de las bacterias, se debe lisar sus paredes celulares. Ello se consiguió agitando en un vórtex con soporte las bacterias, previamente tratadas con la solución RNAWiz que contiene fenol, y perlas de zirconio. Posteriormente el lisado se mezcló con cloroformo y se centrifugó. De este modo, se obtuvo una solución con fases distintas, de las que la superior era la acuosa que contenía el RNA y el DNA y en la orgánica inferior proteínas, ácidos grasos, polisacáridos y otros restos celulares. Cuidadosamente se extrajo la fase superior y se pasó a través de un filtro de sílice donde el RNA queda retenido. Para eliminar contaminantes que todavía estaban presentes, el filtro se lavó con distintas soluciones y finalmente, el RNA se eluyó con una solución de baja fuerza iónica. El resultado de este protocolo fue la obtención de un RNA de muy alta calidad, libre de compuestos orgánicos, preparado para ser enriquecido en RNA mensajero (mRNA).

3.4. Cuantificación del RNA

Para validar la calidad y la concentración del RNA total extraído, éste fue cuantificado mediante una medida fluorimétrica, empleando el reactivo Quant-iT RiboGreen de Invitrogen (LifeTechnologies). Éste es un reactivo fluorescente muy sensible que se intercala entre los nucleótidos del RNA y que presenta la gran ventaja que solo requiere 1 μ l de muestra.

Para llevar a cabo la cuantificación, en primer lugar, se preparó una dilución 1:200 de Ribogreen con TE 1X. Luego, se añadieron 25 μ l de dicha preparación, 24 μ l de TE 1X y 1 μ l de muestra. Esta mezcla, una vez agitada en el vórtex, fue incubada a temperatura ambiente en condiciones de oscuridad durante 4 minutos para que el Ribogreen se intercalase en la muestra. Por último, se trasvasó este volumen a una cuveta de cristal y ésta fue introducida en el fluorímetro de filtro QuantiFluor ST (Promega) para obtener la concentración.

Aquellas muestras que no superaron los 10 ng/ μ l fueron almacenadas y se procedió a repetir la extracción del RNA total de la suspensión bacteriana. Éste fue, nuevamente, cuantificado y en el caso de que tampoco superase el umbral descrito, ambas muestras fueron juntadas en un único tubo y concentradas en el SpeedVac (Thermoscientific). Siguiendo este procedimiento se logró que todas las muestras, excepto una, superaran la barrera de los 10 ng/ μ l.

3.5. Electroforesis en gel de agarosa

Para comprobar que la integridad del RNA total no estuviera alterada, además de la pertinente cuantificación, se realizó una electroforesis en gel de agarosa.

Para ello, se preparó un gel de agarosa al 0,8%, añadiendo 3 μ l de gel Red cada 20 μ l de agarosa. Tras introducir el gel, una vez solidificado, en la cubeta de electroforesis, se añadió solución tampón TBE 1X con la siguiente composición: 89 mM Tris (pH 7.6), 89mM borato, 2mM EDTA. El marcador de pesos moleculares empleado fue el GeneRuler (ThermoScientific).

En aquellos casos en los que se vieron nítidamente las dos bandas correspondientes al 23 S y 16 S rRNA, se valoró como apta la muestra para seguir con el tratamiento con DNasa.

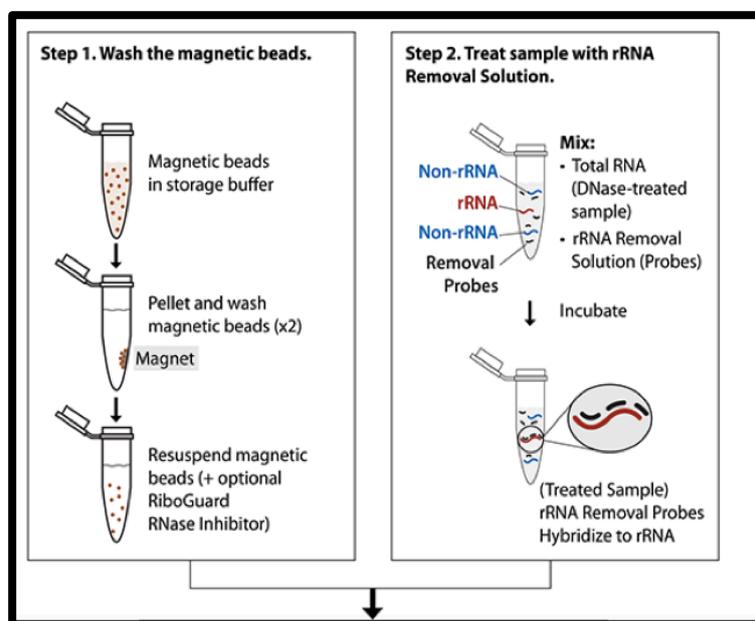
3.6. Tratamiento con DNasa

Previo al enriquecimiento en mRNA, para asegurar que el contenido eluído es sólo RNA, se eliminó cualquier traza de DNA mediante el kit Baseline-Zero™ DNase (Epicentre, Illumina). Éste sigue un protocolo relativamente sencillo en el que el RNA se resuspendió en 17µl de agua libre de RNAasa. Posteriormente se añadieron el tampón (2 µl) de reacción y la DNasa (1 µl, 1MBU). Para que la reacción tuviera lugar, ésta fue incubada a 37°C durante media hora. Una vez transcurrido este lapso de tiempo, se inactivó la DNasa adicionando la solución Stop (2 µl), incubando la muestra a 65°C durante 10 minutos.

3.7. Enriquecimiento en mRNA

Para eliminar el rRNA se empleó el kit Ribo- Zero™ Magnetic (Bacteria) (Epicentre, Illumina) que se basa en preparar unas bolas magnéticas a las que se unen unas cadenas monocatenarias que hibridan con regiones altamente conservadas de los RNA 16S, 23S y 5S. De este modo se elimina casi la totalidad del rRNA y permite enriquecer la preparación en mRNA para su secuenciación.

En primer lugar, se prepararon las bolas magnéticas, realizando dos lavados consecutivos con agua libre de RNAasa. Por otro lado, se trató la muestra de RNA con la solución rRNA removal la cual consta de pequeñas secuencias que hibridan con regiones muy conservadas de los rRNA citados. Entonces, tras un período de incubación de 10 minutos a 68° y de 5 a temperatura ambiente, se añadió la solución del RNA a las bolas magnéticas lavadas. Tras dos incubaciones de 5 minutos, una a 50°C y otra a temperatura ambiente, se puso dicha mezcla en un soporte magnético. Transcurrido un lapso de tiempo de 1 minuto se retiró el sobrenadante enriquecido en mRNA al ser retenidos los rRNA en el pellet de color marrón (ver figura 6). Posteriormente, se realizó una purificación de dicho mRNA mediante una precipitación con etanol 100% y dos posteriores lavados con etanol 70%. Por último, se dejó secar el pellet a temperatura ambiente y se resuspendió en 15µl de agua libre de RNAasa y se mantuvo a -70°C hasta su posterior procesamiento.



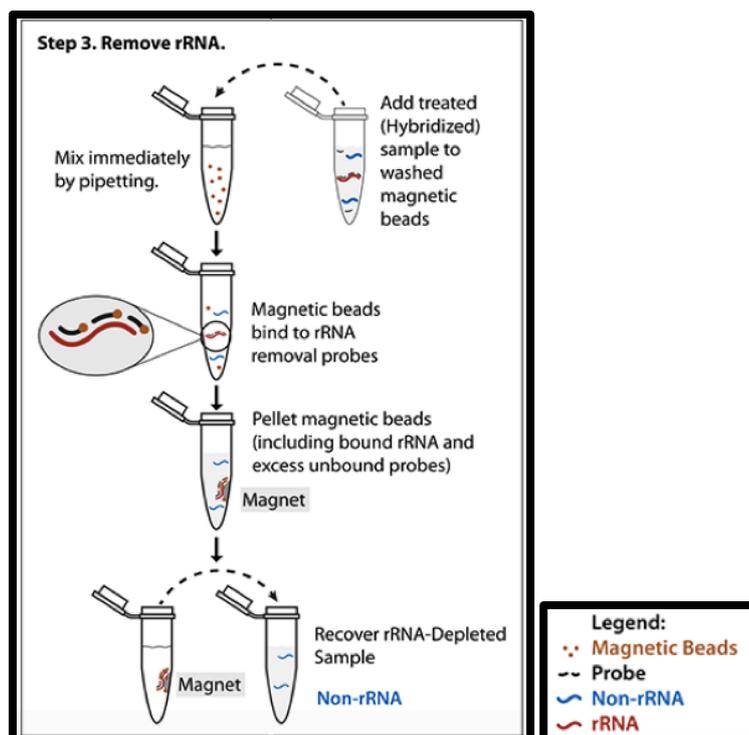


Figura 6: Protocolo de enriquecimiento en mRNA Ribo-Zero rRNA removal (Illumina)

3.8. Síntesis de librerías

Las librerías se obtuvieron mediante el kit ScriptSeq™ v2 RNA-Seq Library Preparation (Epicentre, Illumina), recurriendo al protocolo específicamente diseñado para mRNAs severamente fragmentados.

En primer lugar, se adicionaron el primer de cDNA (2 µl) y agua libre de nucleasa (cantidad en función de la concentración del mRNA de partida) al mRNA durante 5 minutos a 65°C. Una vez transcurridos, para parar la reacción y evitar hibridaciones inespecíficas, se depositaron los tubos en hielo.

Paralelamente se preparó una reacción máster mix. Para ello se mezclaron los siguientes reactivos: un pre-mix de síntesis (3,0 µl, constituido por los dNTPs, un conjunto de hexámeros aleatorios con *tags* (etiquetas) unidas al extremo 5 (ver figura 7), el tampón de reacción (un inhibidor de RNasa y un estabilizador), DTT (0,5 µl, 100mM), un agente reductor que elimina puentes disulfuro, relajando así la estructura secundaria del RNA y facilitando la retrotranscripción, la solución de fragmentación del RNA (1 µl) y la transcriptasa reversa (0,5 µl), que es la que permite la síntesis del cDNA. Esta solución fue añadida a cada uno de los tubos, en hielo, que contenían el mRNA preparado en el primer paso y estos fueron depositados en un termociclador. Una vez alcanzada la fase de pausa, se adicionó a los tubos la solución de finalización (1 µl) para parar la reacción y evitar la formación de subproductos. Por último, para inactivar la solución de finalización se aumentó la temperatura del termociclador hasta los 95°C durante 3 minutos. Transcurrido ese tiempo se obtuvo el cDNA que estaba sintetizado por completo.

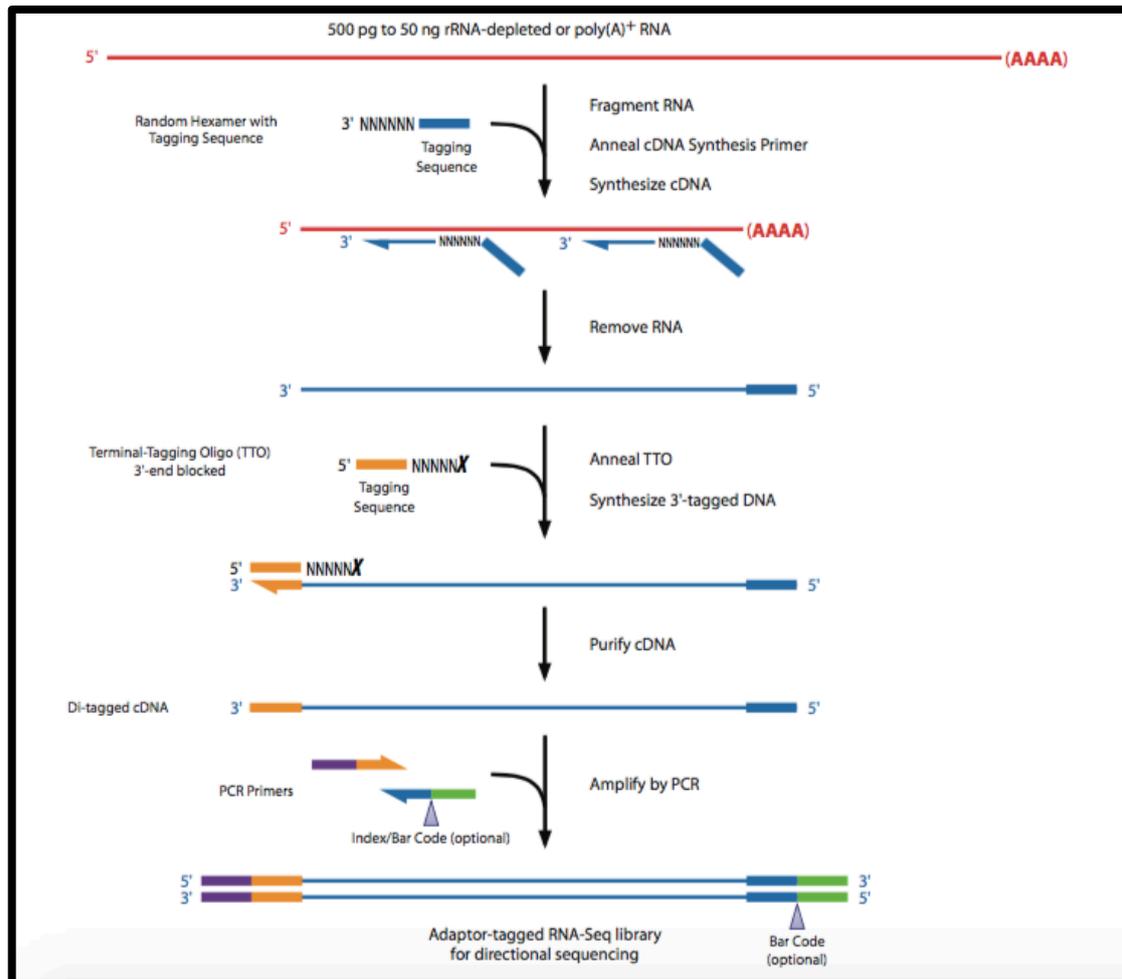


Figura 7: Protocolo de preparación de las librerías ScriptSeq v2 RNA-Seq (Illumina)

Una vez se obtuvo el cDNA, marcado en 5, se tuvo que añadir una *tag* en su extremo 3' también, como requisito para su posterior identificación. Con este propósito, se estableció una reacción master mix de marcaje terminal empleando una DNA polimerasa (0,5 µl) y el pre-mix de marcaje terminal (7,5 µl), constituido entre otros reactivos por dNTPs y un TTO (terminal tagging oligo) que presenta una tag conocida en 5' y en 3' una secuencia aleatoria cuyo último nucleótido está bloqueado impidiendo la síntesis de una cadena de DNA indeseada (ver figura 7). Esta solución fue añadida al cDNA del apartado anterior y depositada en el termociclador.

El resultado de esta reacción fue un cDNA con tags conocidas en ambos extremos ("di-tagged"), que tuvo que ser purificado para que la amplificación posterior fuese lo más eficiente posible, facilitando de este modo la posterior secuenciación. La purificación fue llevada a cabo mediante el kit Agencourt AMPure XP system (BeckmanCoulter) que se basa en sucesivos lavados con bolas magnéticas y con etanol 80%.

Por último, se amplificaron las librerías de cDNA, cada una con una secuencia específica denominada código de barras ("barcode"), mediante una PCR a fin de sintetizar la segunda hebra de DNA y de obtener librerías con secuencias identificativas únicas que faciliten el análisis bioinformático posterior. Con este propósito se dispusieron en tubos sobre hielo los siguientes reactivos: 22,5 µl del cDNA "di-tagged" del paso anterior, 1 µl del cebador Forward PCR Primer, 1 µl del "barcode" ScriptSeq Index PCR Primer (uno por

librería), 25 µl del pre-mix FailSafe PCR (constituido por el tampón de reacción, los dNTPs y otros componentes que aseguren una correcta amplificación) y 0,5 µl de la polimerasa FailSafe PCR Enzyme que con su actividad correctora 3' → 5' correctora ("proofreading") asegura una alta fidelidad en la incorporación nucleotídica. Seguidamente, tras una breve agitación que mezclara bien todos los componentes y obtuviera una suspensión homogénea, se colocaron los tubos en el termociclador dispuesto con una programación de 15 ciclos para RNA degradados. Transcurrido el programa, se obtuvo un DNA bicatenario marcado en ambos extremos y con un código de barras específico (ver figura 7) que, antes de ser secuenciado, volvió a ser purificado mediante el kit AMPure XP system para generar librerías constituidas de fragmentos de más de 200 nucleótidos y para eliminar los dímeros de cebadores que se hubieran podido formar durante la PCR.

3.9. Evaluación de las librerías

En primer lugar, tras la purificación y a modo de comprobación se cuantificaron todas las muestras mediante el fluorímetro Qubit 3.0 (Life Technologies). Para ello, se emplearon el reactivo de ensayo Qubit (1µl) y el tampón de dilución (199µl) del kit Qubit dsDNA HS (Life technologies). Entonces, para medir la concentración, se añadieron 199µl de dicha disolución y 1µl de la muestra a un tubo de 500µl que fue introducido en el fluorímetro.

Seguidamente, se realizó una PCR test y un consiguiente gel para comprobar que el cDNA se había amplificado correctamente. La PCR se preparó recurriendo a 10µl de master mix GoTaq (Promega), que incluye una Taq polimerasa, dNTPs, MgCl₂ y tampones de reacción a concentraciones óptimas para una amplificación eficiente, 0,5µl de ambos cebadores (10 µM), 8µl de agua libre de nucleasas y 1µl del DNA correspondiente. El gel se preparó siguiendo el procedimiento descrito, excepto que en este caso se empleó agarosa al 1,4% para asegurar una correcta migración del DNA.

Por último, para una evaluación más exhaustiva y precisa de la librerías se empleó el bioanalizador 2100 (Agilent Technologies), que es un instrumento que emplea la tecnología de microfluidos Lab-on-a-Chip basada en electroforesis capilar para analizar la integridad, concentración y otros parámetros del DNA con un chip de alta sensibilidad. En su aplicación para el estudio del DNA de las librerías permitió comprobar muestra por muestra si había contaminaciones y, principalmente, si no había un exceso de cebador, ya que en la mayoría de ellas la concentración de partida fue baja. En aquellos casos en los que se advirtió dicha situación, la muestra volvió a ser purificada mediante el kit AMPure XP Purification a fin de evitar un exceso de amplificaciones no deseadas durante la secuenciación.

3.10. Dilución y "pool" de las librerías

Una vez todas las librerías fueron evaluadas, para obtener una mezcla equimolecular de todas las muestras y asegurar que habría una cantidad suficiente de todas las muestras en las siguientes etapas, se realizaron diluciones con agua miliQ para llevarlas a todas hasta la concentración de la muestra con menor cantidad que resultó ser la MIP06-MA con 1ng/µl. Seguidamente, al estar las muestras a dicha concentración, se mezclaron todas en un único tubo ("pool") (empleando alícuotas de 5 µl de cada una) para posibilitar su posterior secuenciación.

3.11. Preparación de librerías para el secuenciador MiSeq

Una vez todas las librerías fueron evaluadas, se prepararon para ser cargadas en el secuenciador MiSeq (Illumina) empleando la guía *Preparing Libraries for Sequencing on the Miseq*[®] (Illumina). El fundamento de este protocolo es desnaturalizar y diluir las librerías para permitir y optimizar su secuenciación.

Con este propósito, en primer lugar, el tampón de hibridación HT1 se descongeló y se mantuvo a 4°C hasta su uso. Paralelamente, se preparó una dilución de 1 ml de 0.2 N NaOH combinando 200µl del stock 1.0 N NaOH con 800µl de agua libre de iones. Con este propósito se mezclaron 5µl de la librería “pool” preparada en el paso anterior con 5µl de dicha solución. Tras mezclar brevemente los reactivos, se centrifugó el tubo a velocidad media durante un minuto y se dejó incubando a temperatura ambiente 5 minutos más. De esta manera se completó la desnaturalización del DNA en cadena simples para poder, posteriormente, en el secuenciador generar los “clusters” que permitan la secuenciación.

Seguidamente, se añadieron 990µl del tampón HT1 enfriado en hielo previamente, a los 10µl del DNA desnaturalizado a fin de disminuir la concentración de sales generada en la solución. De este modo se obtuvo una disolución con una concentración de 20pM de la librería desnaturalizada en 1mM de NaOH y ésta se depositó en hielo. A continuación, ésta fue nuevamente diluida hasta obtener la concentración final de 15pM añadiendo 150µl del tampón HT1 a 450µl del DNA desnaturalizado.

Paralelamente se preparó el control PhiX, que es una librería ligada con adaptadores derivada del genoma del bacteriófago Phi-X174. Ésta se emplea en los “runs” del secuenciador como control de calidad para la generación de los “clusters”, el proceso de secuenciación y el posterior alineamiento. Para ello se combinaron 2µl de la librería 10nM PhiX y 3µl de la solución Tris-Cl a pH 8.5 con 0,1% Tween 20 de modo que la solución permanezca a un pH óptimo. Luego, el control PhiX fue desnaturalizado siguiendo los mismos pasos que se realizó con la librería “pool” de muestras de DNA resultando en una librería de 20 pM. Asimismo, luego fue diluido hasta los 12,5 pM añadiendo 225µl del tampón HT1 a 375µl del control.

Por último, se combinaron la librería “pool” de las muestras desnaturalizadas (594µl) con la del control PhiX (6µl) y esta suspensión (600µl) fue cargada en el cartucho de los reactivos del secuenciador.

3.12. Secuenciación

La secuenciación se realizó en el secuenciador MiSeq Gene & Small Genome de Illumina, especialmente diseñado para la secuenciación de genomas pequeños y metagenomas / metatranscriptomas, empleando una placa (“flow cell”) de un solo carril (“lane”) y programado para correr 230 ciclos. El cartucho de los reactivos empleado fue el Miseq Reagent Kit V2 que permite la obtención de *reads* de hasta de 250 pares de base por cada extremos de las secuencias, es decir, de hasta 500 pares de bases por fragmento, que pueden o no sobrelaparse, en mayor o menor medida, dependiendo de la longitud del mismo.

3.13. Análisis bioinformático

3.13.1. Procesamiento y limpieza de las secuencias

Una vez finalizado el proceso de secuenciación, el software del secuenciador proporcionó una carpeta con todas las muestras ya demultiplexadas pero sin haber sido sometidas a ningún tipo de filtro. Por ello, para agilizar y hacer que el análisis bioinformático sea viable, aquellas secuencias correspondientes a los adaptadores y a los códigos de barras, las bases que no superaron los 20 puntos de calidad (ver figura 2 del anexo) y las lecturas que no llegaban a las 50 pares de bases, fueron eliminadas. Esta limpieza fue llevada a cabo mediante el programa *prinseq-lite* (Scmieder et al., 2011).

Por otro lado, cabe destacar que en la carpeta que generó el secuenciador, para cada muestra se encontraron dos archivos denominados R1 y R2. Éstos contenían los *reads* en las direcciones directa (*forward*, en la dirección 5' → 3') y reversa (*reverse*, en la dirección 3' → 5') respectivamente de cada uno de los extremos de los fragmentos secuenciados. Éstos, en el caso de existir solapamiento entre ellos, fueron unidos empleando el programa *fastq-join* (Aronesty, 2011); los que no pudieron ser juntados se anotaron en 2 archivos denominados un1 y un2 respectivamente.

Ambos procesamientos fueron realizados por el servicio de secuenciación del área de Genómica y Salud de FISABIO-Salud Pública.

3.13.2. Eliminación de los rRNA bacterianos, eucariotas y humanos

El siguiente paso lo constituyó el eliminar aquellos *reads* pertenecientes a las secuencias de RNA ribosómicos de bacterias o eucariotas de modo que en el posterior alineamiento solo se interrogaran aquellas secuencias que originalmente fueron mRNA. Este filtrado fue realizado siguiendo una línea de comandos o *pipeline* desarrollada por los técnicos del área de secuenciación.

Para ello, en primer lugar, se fraccionaron los archivos de cada muestra en *chunks*, que permiten una mayor versatilidad y velocidad de procesamiento de las secuencias. Éstos fueron nuevamente filtrados cualitativamente por el programa *prinseq*. Posteriormente, con los *chunks* generados como buenos, se llevó a cabo una primera criba, eliminando los RNA ribosómicos (realizando alineamientos con la base de datos de rRNAs *SILVA*) y, específicamente, los humanos (mediante el software *Burrows-Wheeler Alignment tool* optimizado para secuencias de Illumina que permite el mapeado de secuencias poco divergentes frente a una genoma de referencia). Luego, se realizó un segundo filtrado de los ribosomales bacterianos empleando el algoritmo Mega Blast.

4. RESULTADOS Y DISCUSIÓN

4.1. PARTE EXPERIMENTAL

Si bien es cierto que el RNA constituye el primer paso de la expresión génica y, por tanto, aporta una información acerca de la funcionalidad mucho más específica que el DNA, su procesamiento y manejo resultan ser verdaderamente complejos debido a lo delicado de su naturaleza. La fragilidad del RNA radica primariamente en dos aspectos. En primer lugar, cabe destacar que la actividad ribonucleasa es ubicua en cualquier ambiente y dado que las RNasas, que son las que la llevan a cabo, son extremadamente inalterables (no requieren cofactores y son termoestables), las preparaciones de RNA son fácilmente degradables por ellas (Tan et Yiap, 2009). Por otro lado, atendiendo a la estructura molecular, el RNA es termodinámicamente menos estable que el DNA debido al grupo hidroxilo en 2' del anillo de ribosa que promueve un ataque hidrofílico al enlace fosfodiéster 5' → 3' para formar un fosfato cíclico 2' → 3'. Además, este intermediario cíclico se estabiliza por cationes divalente como el Mg²⁺, por lo que éstos, que también son muy abundantes, facilitan la degradación del RNA. A su vez, altas temperaturas y pH aumentan su hidrólisis. De hecho, se ha comprobado que la inestabilidad de esta molécula es tal, que su vida media en un procarionota como *E. coli* es de tan sólo 6,8 minutos (Selinger et al., 2003).

Por todos estos motivos, a lo largo de todo el trabajo, se puso un especial esmero en minimizar al máximo cualquier tipo de contaminación trabajando en un ambiente libre de RNAsas “RNase-free”. Con todo, a pesar de las dificultades técnicas que conlleva el manejo de RNA, finalmente, tras varias extracciones de la mayoría de muestras, se obtuvieron concentraciones aceptables (>10 ng/μl) para seguir con el proyecto de todas ellas excepto de la M09B (ver tabla 2). Esto pudo deberse a que estuviera degradada o a algún error técnico durante el procesamiento de la muestra fecal.

Tabla 2: Concentraciones finales de RNA total expresadas en ng/μl.

M01 A	M03A	M06 A	M07 A	M08 A	M09 A	MI6 A
32,2	29,96	15,235	48,29	18,47	10,105	48,97
M01 B	M03B	M06 B	M07 B	M08 B	M09 B	MI6 B
11,03	26,76	57,45	24,895	42,36	0,7155	29,635
M01 I1	M03 I1	M06 I1	M07 I1	M08 I1	M09 I1	MI6 I1
43,35	17,555	16,425	9,07	15,045	12,075	34,23
M01 I3	M03 I3	M06 I3	M07 I3	M08 I3	M09 I3	MI6 I3
45,645	17	25,12	27,385	19,82	23,32	41,33
M01 I4	M03 I4	M06 I4	M07 I4	M08 I4	M09 I4	MI6 I4
32,915	9,125	35,7	14,2	38,38	63,45	54,95
M01 I5	M03 I5	M06 I5	M07 I5	M08 I5	M09 I5	MI6 I5
43,455	62,75	40,125	34,23	42,205	39,33	35,265

No obstante, una alta concentración no garantiza la integridad del RNA total extraído o la ausencia de contaminantes. Como se puede observar en la figura 8, la

evaluación de las extracciones fue realizada mediante electroforesis en gel de agarosa y ésta pone de manifiesto lo que se acaba de comentar: muestras, como por ejemplo la M0314, con altas concentraciones (43,02 ng/μl en este caso), no implican que la extracción haya eliminado la mayoría de los contaminantes (como restos celulares, DNA, etc.) ni que el RNA no esté parcialmente degradado. De hecho, dicha muestra tuvo que ser extraída hasta en 3 ocasiones y, paradójicamente, el RNA de mayor calidad se visualizó en la muestra con menor concentración de partida (9,125 ng/μl). Por este motivo, todas aquellas muestras que no mostraran una relativa integridad (i.e. en las que no se vieran nítidamente las bandas correspondientes al 23S y al 16S rRNAs, que son los más abundantes en procariotas) fueron repetidas, realizando nuevamente las respectivas extracciones y cuantificaciones. Respecto a estas reflexiones es importante destacar que la necesidad de repetir las extracciones hizo que de algunas muestras (M09B, M0711 y la propia M0314) se acabaran las muestras fecales de partida y no se alcanzara el umbral de los 10ng/μl. No obstante, tanto la segunda como la tercera sobrepasaron los 9ng/μl (ver tabla 2) y sus geles evidenciaron una buena integridad del RNA. Por el contrario, la primera, al no llegar ni a 1ng/μl hizo inviable su consideración para seguir con el procedimiento.

Por otro lado, también es destacable comentar el rango de concentraciones desde un significado biológico. Como se destacó al principio, las muestras infantiles fueron tomadas en distintos intervalos de tiempo y, por tanto, este hecho tendría que quedar patente ya que a más edad, la microbiota que va colonizando el TGI es más abundante y diversa, expresándose de este modo más genes, lo que tendría que conllevar un claro aumento en la concentración de RNA. Esta teórica linealidad se pudo apreciar sólo en las muestras de las series M06 y M08 (ver figura 9), aunque en varias de las demás series la concentración final al año fue claramente superior a la obtenida en las dos o tres primeras muestras (M03, M07, M09) a pesar de no observarse una linealidad en el proceso. Es posible que haya existido variabilidad no controlada en alguno de los numerosos factores implicados en la obtención del RNA (por ejemplo, la conservación y procesamiento de las muestras fecales) que haya contribuido a este resultado.

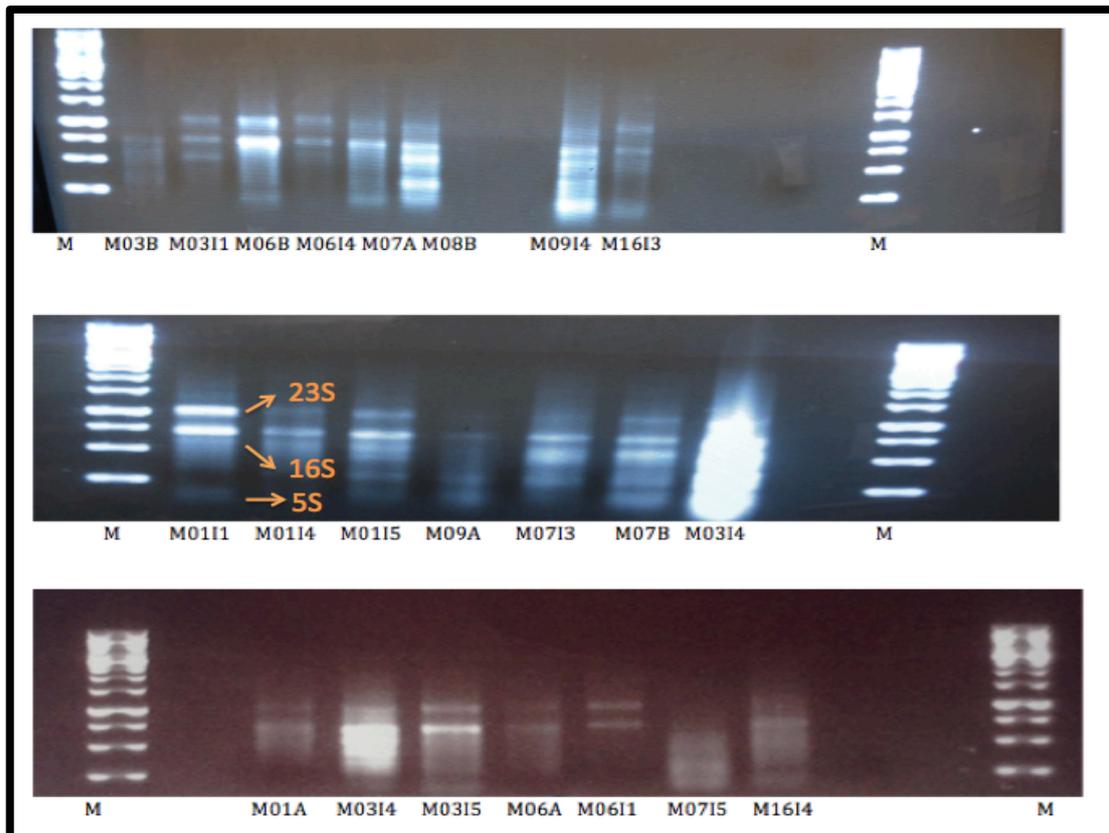


Figura 8: Electroforesis de algunas de las extracciones realizadas. Nótese que se ha tomado la muestra M0111 como referencia para indicar a qué ácidos nucleicos corresponden cada una de las bandas observadas.

Por otro lado, como uno de los objetivos de este estudio es establecer comparaciones entre los infantes y sus progenitoras, cabe comentar brevemente, antes de evaluar la expresión génica, de modo preliminar, si la concentración de RNA en las muestras maternas antes y después del embarazo sigue también algún tipo de relación. A día de hoy, hay pocos estudios que hayan caracterizado los cambios en la microbiota materna durante y después del embarazo, si bien es cierto que se ha comprobado que el número de bacterias aumenta durante el embarazo (Collado et al., 2008; Koren et al., 2012). No obstante, se ha postulado que, tras el parto y coincidiendo con el retorno a una condición fisiológica normal, la microbiota se asemeja a la de antes del embarazo (Prince et al., 2014).

Por este razonamiento, todas las muestras MIP-MA (una semana antes del parto), en teoría, deberían presentar mayores concentraciones que las MIP-MB (un año después del parto), ya que al aumentar el número de los conformantes de la microbiota la expresión génica haría lo propio y esto se traduciría en aumento de la concentración de RNA. 5 (M01A, M03A, M07A, M09A y M16A) de las 7 muestras recogidas una semana antes del parto sí mostraron mayor concentración que las recogidas un año después (ver tabla 2). Sin embargo, como en la comparación anterior de los bebés, es posible que factores técnicos hayan impedido que este razonamiento se cumpliera para todas las muestras.

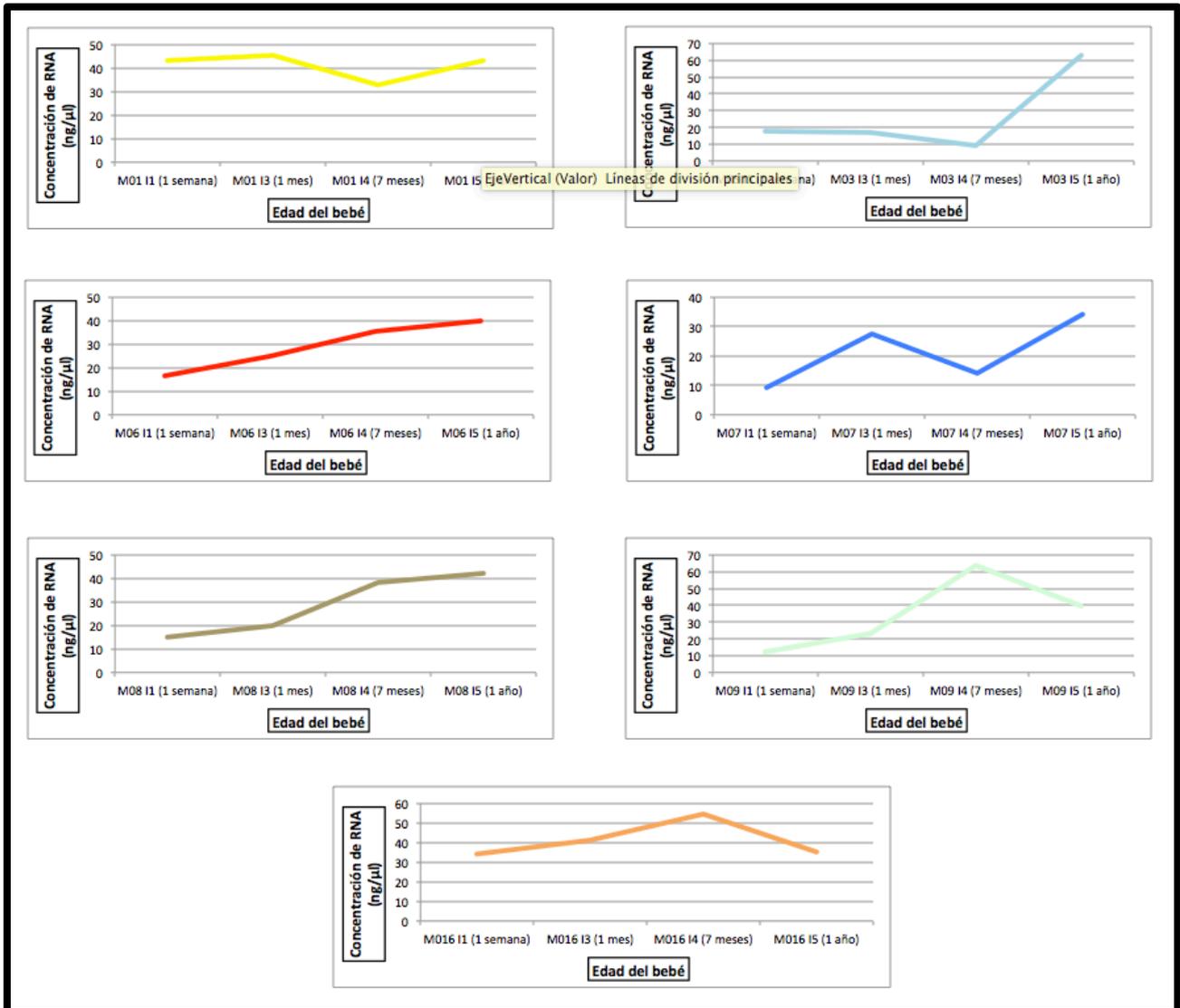


Figura 9: Relación entre la concentración de RNA y la edad de los infantes. (Nótese que la linealidad sólo está presente en las muestras MIP06 y MIP08 (rojo y marrón claro respectivamente).

Dado que alrededor del 97% del RNA total procarionta está constituido por RNA ribosomal, el enriquecimiento en mRNA de las muestras de RNA total constituye un verdadero desafío al carecer éste de cola poly(A) en su extremo 3' y no permitir ser aislado mediante selección por secuencias oligo-dT como se suele hacer en eucariotas (Sorek et Cossart, 2009; Giannoukos et al., 2012; Bhagwat et al., 2014). Asimismo, esta característica lo hace muy sensible a la degradación por RNasas ubicuas como la PNPasa (Peano et al., 2013). Es por este motivo que la eliminación del rRNA se antoja como un paso crítico y se requieren estrategias altamente especializadas. En este trabajo se empleó el Ribo-ZeroTM Magnetic (Bacteria) de Epicentre. Al revés que otros, utilizados ampliamente pero con limitado éxito en propósitos metatranscriptómicos, este kit está constituyendo una auténtica revolución en el sector ya que es capaz de eliminar más del 95 % del rRNA (ver figura 10).

Además, en contraste con la selección mediante el enriquecimiento de la cola poly(A), este procedimiento preserva información sobre RNAs no adenilados, no codificantes

y regulatorios permitiendo la investigación sobre la regulación, la edición del RNA y otros fenómenos que aumentan la comprensión acerca de la complejidad del transcriptoma.

El éxito de este kit en la eliminación de RNA ribosómico radica en el empleo de largos transcritos biotinilados de rRNA como sondas (*probes*) que hibridan con el rRNA presente en la muestra de RNA total. Los híbridos resultantes son descartados mediante su adición a unas bolas magnéticas que llevan unidas estreptavidina lo que deja en la suspensión un RNA libre de rRNA (ver figura 6).

En la figura 10 se muestran los resultados obtenidos por Giannoukos et al. (2012) en un estudio comparativo entre los métodos disponibles, donde el Ribo-Zero™ fue el único que realmente consiguió eliminar casi por completo las secuencias del RNA ribosomal. Además se demostró que dicha eliminación es altamente reproducible y que también es muy efectivo en muestras de RNA de baja calidad o degradado.

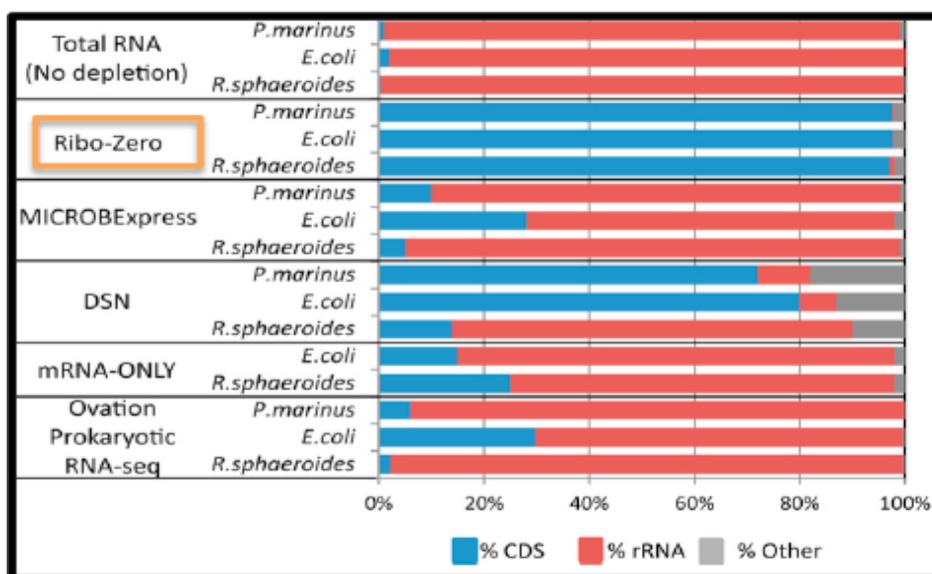


Figura 10: Comparación de la eficiencia de distintos métodos de reducción de rRNA (Giannoukos et al., 2012). Nótese que el empleado en el presente trabajo está resaltado en un cuadro naranja y que CDS representa las siglas de *coding DNA sequence*, es decir, las secuencias de mRNA una vez retrotranscritas.

La sorprendente capacidad de este kit para eliminar ribosómico también fue demostrada en el presente trabajo. Observando la tabla 3, donde se muestran las concentraciones obtenidas tras haber realizado el procedimiento de enriquecimiento en mRNA, queda patente que la disminución de las cantidades es muy notable y que la cantidad de mensajero recuperada es ínfima.

Para percibir de modo más global la efectividad del Ribo-Zero™, en la tabla 3 se muestran los porcentajes de reducción del RNA entre las muestras antes y después de ser procesadas y los resultados obtenidos son realmente asombrosos: todas las muestras excepto 3 (MIP06-I5, MIP09-I4, MIP16-MA) superaron el 90% de reducción y hasta 17 mostraron más de un 99%. En cualquier caso, aunque los porcentajes de reducción disminuyesen a un 90% o a un 80%, ya sería un notable éxito comparado con los otros métodos disponibles.

Tabla 3: Concentraciones finales de mRNA expresadas en ng/ μ l y los porcentajes de reducción entre las muestras de RNA total y las de mRNA entre paréntesis.

M01 A	M03A	M06 A	M07 A	M08 A	M09 A	MI6 A
0,10895 (99,66%)	0,00185 (99,99%)	0,555 (96,36%)	0,07935 (99,84%)	0,3861 (97,91%)	0,04655 (99,54)	8,85 (81,93%)
M01 B	M03B	M06 B	M07 B	M08 B	M09 B	MI6 B
0,2717 (97,54%)	0,27185 (98,98%)	1,518 (97,36%)	0,089 (99,64%)	0,206 (99,51%)	-	2,7625 (81,93%)
M01 I1	M03 I1	M06 I1	M07 I1	M08 I1	M09 I1	MI6 I1
0,0269 (99,94%)	0,3437 (98,04%)	0,1307 (99,20%)	0,5155 (94,32%)	0,33595 (97,77%)	0,45055 (96,27%)	0,01555 (99,95%)
M01 I3	M03 I3	M06 I3	M07 I3	M08 I3	M09 I3	MI6 I3
0,32845 (99,28%)	0,011 (99,94%)	1,554 (93,81%)	0,1713 (99,37%)	0,01525 (99,92%)	0,6985 (97,00%)	2,078 (94,97%)
M01 I4	M03 I4	M06 I4	M07 I4	M08 I4	M09 I4	MI6 I4
0,4062 (98,77%)	0,4714 (94,83%)	0,1463 (99,59%)	0,18255 (98,71%)	0,8885 (97,68%)	6,515 (89,73%)	3,3365 (93,93%)
M01 I5	M03 I5	M06 I5	M07 I5	M08 I5	M09 I5	MI6 I5
0,33875 (99,22%)	3,598 (94,27%)	4,417 (88,99%)	0,1506 (99,56%)	2,837 (93,28%)	7,3 (92,79%)	0,852 (97,58%)

Posteriormente, la preparación de las librerías de cada una de las muestras rindió las concentraciones apreciables en la tabla 4. Para evaluarlas en cuanto a calidad y ver que no presentaban restos de terminadores y adaptadores no unidos fueron analizadas en el bioanalizador.

Como se muestra en la figura 11, en la cual las bandas altas del principio (35 pares de bases) y del final (10380 pares de bases) pertenecían al marcador inferior y superior respectivamente, la mayor cantidad de los fragmentos se localizaban en la zona de alrededor de las 300 pares de bases. Es de importante relevancia destacar la banda sobre las 50 pares de bases, ya que ésta indica la presencia de cebadores no unidos que, luego, en el proceso de secuenciación podrían generar problemas de hibridaciones inespecíficas. En la figura 11 se han seleccionado 4 electroferogramas correspondientes a 4 librerías que son representativas de los aspectos comentados. Los dos superiores (M08B y M03A) muestran un patrón electroforético normal, es decir, un pico sobre las 250 pares de bases bien definido de y de área relativamente estrecha, y no se aprecia ningún pico alrededor de los 50 pares de bases en ellos.

Sin embargo, en los dos inferiores (M01I1 y M03I3) si se observan anomalías. La M01I1 pese a tener el patrón normal, sí presenta el pico perteneciente a los cebadores. El caso de la M03I3 es aún más notorio ya que no sólo no presenta un patrón aceptable sino que el pico de los cebadores es mucho mayor que incluso el de los marcadores. Esto puede ser debido a una mala preparación de la muestra o que al tener poca cantidad, en el paso de la amplificación hubiera muchos más cebadores que muestra y por tanto resultarían en un

número mucho mayor. Por estos motivos, para eliminar los excesos de cebadores y tratar de obtener un patrón normal en aquellas como la M03I3, las muestras que presentaban electroferogramas con estas características tuvieron que ser purificadas nuevamente (ver tabla 4).

Por otro lado la evaluación de los electroferogramas permitió comprobar que la garantía del protocolo de generar fragmentos de más de 200 pares de bases se cumplió en todos los casos.

Tabla 4: Concentraciones (en ng/μl) de las librerías después de la primera purificación [Librería] y, en los casos necesarios, después de la segunda.

Índice/barcode	Muestra	[Librería]	[Librería 2ª purific]
1	M01 A	2,94	
2	M03 A	1,56	
3	M06 A	1,26	1,10
4	M07 A	3,04	
5	M08 A	4,54	
6	M09 A	1,91	
7	M16 A	11,30	
8	M01 B	1,43	1,26
9	M03 B	2,58	1,32
10	M06 B	10,00	
11	M07 B	8,70	
12	M08 B	15,00	
13	M16 B	2,40	1,17
14	M01 I1	12,30	7,46
15	M03 I1	1,86	1,47
16	M06 I1	14,70	8,68
17	M07 I1	3,04	1,66
19	M08 I1	3,83	
20	M09 I1	1,66	1,35
22	M16 I1	12,50	
23	M01 I3	1,99	
30	M03 I3	2,23	1,39
31	M06 I3	27,40	13,33
46	M07 I3	2,44	1,47
18	M08 I3	7,84	
25	M09 I3	9,32	6,22
27	M16 I3	2,23	1,73
29	M01 I4	8,78	12,90
33	M03 I4	4,62	4,48
35	M06 I4	6,92	5,46
37	M07 I4	2,34	1,38
39	M08 I4	5,80	
41	M09 I4	18,10	11,90
43	M16 I4	10,80	
45	M01 I5	14,00	9,84
47	M03 I5	14,10	11,00
21	M06 I5	2,78	2,20

24	M07 I5	1,85	1,48
26	M08 I5	14,20	10,40
28	M09 I5	14,50	10,20
32	M16 I5	10,70	8,72

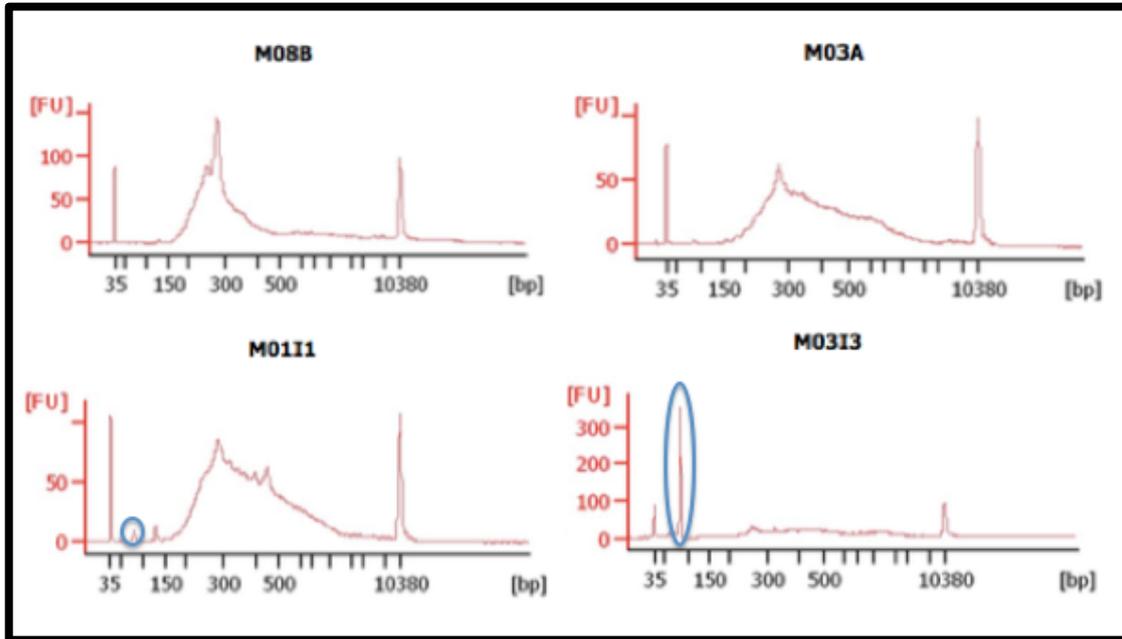


Figura 11: Electroferogramas de algunas librerías. Nótese las dos superiores presentan un patrón normal y por lo tanto no tuvieron que volver a ser purificadas, mientras que las dos inferiores que presentaban el pico característico de un exceso de cebador (sobre las 50 pares de bases, redondeado en azul) sí.

Una vez todas las muestras que lo requerían fueron nuevamente purificadas, se realizó una dilución de todas ellas para que estuvieran a la concentración de la M06A (ver tabla 5), que fue aquella con menor concentración (1,1 ng/μl) de modo que se asegurara que hubiese la misma cantidad de cada una de ellas y que aquellas con unos valores bajos no se perdieran en los procesos posteriores ya que la diferencia de concentración entre algunas superaba los 10 ng/μl. Al estar todas diluidas, se realizó una librería “pool” con alícuotas de todas ellas. Esta nueva librería total fue cuantificada para comprobar que su concentración se asemejara al 1,1ng/ μ l de cada una de sus librerías constituyentes. El resultado fue altamente satisfactorio ya que se obtuvo una concentración de 1ng/μl.

Empleando la ecuación de la figura 12 se calculó el valor correspondiente a esta concentración en unidades en nM, dando éste 6,06nM. Como el kit para preparar las librerías está diseñado para librerías de 2nM o 4nM, se siguió el protocolo de las de 4nM.

$$X \frac{ng}{\mu l} \times \frac{10^6 \mu l}{1L} \times \frac{1nmol}{660ng} \times \frac{1}{N} = nM$$

Figura 12: Fórmula para obtener una concentración en nM de las librerías de Illumina partiendo de ng/μl. Nótese que N hace referencia al tamaño del fragmento en pares de bases.

4.2 Procesamiento bioinformático

El secuenciador rindió una cantidad de más de 47 millones de *reads* repartidos en aquellos correspondientes a las lecturas *forward* y a las *reverse* de las secuencias muestrales y a las respectivas de la librería control de PhiX (ver tabla 5). Asimismo, en el anexo I, se muestra un desglose de los resultados del proceso de secuenciación y los posteriores procesos de filtración y limpieza de las secuencias, en forma de tablas y gráficos.

Tabla 5: Número de *reads* obtenidos por el secuenciador

RESUMEN DE LOS READS	
Nº total de <i>reads</i> R1	23525999
Nº total de <i>reads</i> R2	23525999
Nº total de <i>reads</i> del PhiX 1	114130
Nº total de <i>reads</i> del PhiX 2	114130

Sobre la calidad cabe destacar que, como se muestra en la figura 3 del anexo, antes de unirlos, los *reads* R1 tuvieron una media de calidad mayor (37 sobre 40) que los R2 (35 sobre 40) (ver además figura 5 del anexo). Resaltar que calidades de 30 ya se consideran muy buenas, de modo que en el presente trabajo se obtuvieron valores cualitativos excelentes.

Sobre la longitud de los *reads* (ver tabla 1 y figuras 3 y 6 del anexo) cabe resaltar que, de media, todos se encontraron entre las 229 y las 230 pares de bases. Esto fue indicativo de que la preparación de las librerías, en la cual se buscaba obtener fragmentos de más de 200 pares de bases, fue óptima.

Apuntar que en las figuras del anexo, el término *undetermined* hace referencia a aquellos *reads* que no pudieron ser asignados a ninguna muestra bien por una lectura incorrecta o ausencia del código de barras, bien por su pérdida o bien porque son los pertenecientes al control PhiX.

En cuanto a la cantidad de *reads* por muestra (ver figura 4 del anexo) hay que comentar varios aspectos. Es muy notorio el hecho de que, a pesar de partir de una librería *pool* equimolecular (todas las librerías fueron diluidas a 1 ng/ μ l) y que, por tanto, todas las muestras deberían haber estado representadas por un número similar de *reads*, las diferencias entre ellas son altamente llamativas. El ejemplo más extremo lo representan la muestra con más *reads* (M09I4) que tiene más de un millón de *reads* y la M06A que no llega siquiera a los 200000. Esta observación fue discutida con los miembros del departamento de secuenciación y ellos afirmaron que en todos los *runs* que habían corrido a lo largo de su experiencia laboral, tanto en Illumina como en 454, se observaba esta distribución dispar.

Entre las posibles razones para este hecho, argumentaron acerca de la influencia del tamaño de los fragmentos (que siendo pequeños en ocasiones, tienen mayor facilidad para formar *clusters*), de la calidad de la librería, que a su vez viene determinada por la calidad de la muestra (por ejemplo, una gran fragmentación disminuye la calidad de las librerías) y del sesgo técnico al que está sometida la preparación de las librerías (errores de pipeteo, volúmenes inexactos, etc.). Por otro lado, hicieron referencia al límite de detección del fluorímetro Qubit, que es de 0,1 ng/ μ l y al estar tratando, en bastantes casos (ver tabla 5) con muestras de muy baja concentración, la precisión de las medidas carece de la precisión

que sí atesoran para mayores concentraciones. De hecho, una observación reveladora supone que aquellas muestras cuyas librerías rindieron concentraciones menores son las que menos *reads* generaron como M0311 y M06A.

A su vez, es importante analizar el resultado de la unión de los *reads* R1 y R2. Como muestra la figura 7 del anexo la longitud de los *reads* ya unidos (*joined*), cuya media es 279,5 pares de bases, es apenas unas 50 pares de bases más larga que la media de sus R1 y R2 constituyentes, que es 229,2 (ver tablas 1 y 2 del anexo). Este hecho se debe a que muchos de los fragmentos de partida tenían una longitud inferior a los 300 pares de bases, de manera que existe un elevado grado de solapamiento entre sus R1 y R2. No obstante, como se muestra en la figura 8 del anexo, el número de *reads unjoined*, es decir, aquellos que, por no haber solapamiento o por ser éste muy escaso (inferior a las 50 pares de bases), no se pudieron juntar, es muy superior a los *joined* en todas las muestras, llegando a ser hasta más de 20 veces superior como es el caso de la muestra M0914. Por este motivo se deduce que los fragmentos de las librerías fueron, en su mayoría, superiores a las 450 pares de bases.

En la tabla 3 del anexo se muestra el resultado de la evaluación cualitativa de las muestras siendo resaltable observar que la cantidad de *reads* eliminados tras el filtrado de secuencias correspondientes a rRNAs no es muy alta (casi todas se sitúan por debajo del 10% de reducción, excepto cinco de ellas), lo que permite justificar la gran eficiencia del método de eliminación de RNAs ribosomales. Respecto a ello, en la tabla 4 del anexo se observa que en su mayoría, los *reads* eliminados corresponden a secuencias de rRNA bacterianas que no fueron eliminadas previamente. Es reseñable comentar que el número de *reads* eliminados pertenecientes tanto a rRNAs de microorganismos eucariotas como de humanos fue muy reducido dado que son las bacterias las principales pobladoras del intestino humano y, por tanto, las generadoras de una mayor cantidad de secuencias. A su vez, cabe destacar que el número de *reads* humanos es muy pequeño dado que al tratarse de individuos sanos, sus heces no presentan descamación y, por lo tanto, su contaminación es mínima.

El paso final en el análisis bioinformático de las secuencias de cDNAs obtenidas será su anotación funcional, es decir, la asignación y/o predicción de la función de los *reads* correspondientes a los genes expresados en base a búsquedas por homología contra bases de datos de genomas de referencia. Debido al elevado número de *unjoined reads* obtenidas en nuestras muestras, un paso importante posterior a la anotación será el de establecer si un determinado *read* R2 forma parte de un sólo gen junto a su par R1 o, por el contrario, si proviene de un gen distinto al asignado para el R1 expresado en el mismo transcrito policistrónico. En el primer caso, el gen deberá ser contabilizado tan solo una vez en el perfil global de expresión génica de la muestra, ya que de otra forma la abundancia de algunos transcritos, principalmente los de los genes más largos, sería sobreestimada. Finalmente, en base a la anotación funcional obtenida, se podrá proceder a la comparación de las funcionalidades que se expresan en la microbiota intestinal en diferentes momentos de su desarrollo durante el primer año de vida.

5. Conclusiones

A falta de la realización de la anotación, del presente trabajo se extraen las siguientes conclusiones:

1. La variabilidad no controlada en alguno de los numerosos factores implicados en la obtención del RNA han impedido la teórica linealidad entre la edad del infante y concentración de RNA en sus muestras fecales.
2. La eficiencia del enriquecimiento en mRNA del RNA total extraído de una suspensión bacteriana empleando el kit Ribo- ZeroTM Magnetic (Bacteria) es excelente, asegurando, en la mayoría de casos, reducciones de más del 90% de rRNA, incluso tratándose de un RNA altamente degradado.
3. El proceso de secuenciación fue un rotundo éxito, obteniéndose más de 47 millones de *reads* lo cual supone una importante cantidad tratándose de un metatranscriptoma, además de presentar todos ellos calidades superiores a los 35 puntos de valoración.
4. La alta proporción de *reads* no unidos (*unjoined*) implica que muchos de los fragmentos generados durante la retrotranscripción fueron de un tamaño considerable, abriendo esto la puerta a la anotación tanto de genes ya conocidos como de nuevos genes.
5. Este trabajo de secuenciación ha generado un gran número de secuencias de cDNAs de alta calidad que han sido completamente procesadas y están listas para la anotación tanto de genes ya conocidos como de nuevos genes. Esto permitirá establecer la primera descripción de como evoluciona la expresión génica global de la comunidad microbiana intestinal en los infantes durante su primer año de vida.

6. Bibliografía

- Adlerberth, I., & Wold, a. E. (2009). Establishment of the gut microbiota in Western infants. *Acta Paediatrica, International Journal of Paediatrics*, 98(2), 229–238. <http://doi.org/10.1111/j.1651-2227.2008.01060.x>
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, 473(7346), 174–180. <http://doi.org/10.1038/nature10187>
- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Jun, W. (2015). Dynamics and stabilization of the human gut microbiome in the first year of life. *Under Consideration at Science*, 690–703. <http://doi.org/10.1016/j.chom.2015.04.004>
- Bager, P., Wohlfahrt, J., & Westergaard, T. (2008). Caesarean delivery and risk of atopy and allergic diseases: Meta-analyses. *Clinical and Experimental Allergy*, 38(4), 634–642. <http://doi.org/10.1111/j.1365-2222.2008.02939.x>
- Balzola, F., Bernstein, C., Ho, G. T., & Lees, C. (2010). A human gut microbial gene catalogue established by metagenomic sequencing: Commentary. *Inflammatory Bowel Disease Monitor*, 11(1), 28. <http://doi.org/10.1038/nature08821>
- Barrett, E., Kerr, C., Murphy, K., O’Sullivan, O., Ryan, C. A., Dempsey, E. M., ... Stanton, C. (2013). The individual-specific and diverse nature of the preterm infant microbiota. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 98(4), F334–40. <http://doi.org/10.1136/archdischild-2012-303035>
- Bezirtzoglou, E., Tsitsias, A., & Welling, G. W. (2011). Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe*, 17(6), 478–482. <http://doi.org/10.1016/j.anaerobe.2011.03.009>
- Bhagwat, A. a, Ying, Z. I., & Smith, A. (2014). Evaluation of Ribosomal RNA Removal Protocols for Salmonella RNA-Seq Projects. *Advances in Microbiology*, 4(January), 25–32.
- Bloom, J. S., Khan, Z., Kruglyak, L., Singh, M., & Caudy, A. a. (2009). Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics*, 10, 221. <http://doi.org/10.1186/1471-2164-10-221>
- Booijink, C. C. G. M., Boekhorst, J., Zoetendal, E. G., Smidt, H., Kleerebezem, M., & De Vos, W. M. (2010). Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. *Applied and Environmental Microbiology*, 76(16), 5533–5540. <http://doi.org/10.1128/AEM.00502-10>
- Cabrera-Rubio, R., Collado, M. C., Laitinen, K., Salminen, S., Isolauri, E., & Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *American Journal of Clinical Nutrition*, 96(3), 544–551. <http://doi.org/10.3945/ajcn.112.037382>
- Celaj, A., Markle, J., Danska, J., & Parkinson, J. (2014). Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome*, 2(1), 39. <http://doi.org/10.1186/2049-2618-2-39>
- Church, G. M., Church, G. M., Rosenow, C., & Rosenow, C. (2003). Global RNA Half-Life Analysis in. *Genome Research*, (January), 216–223. <http://doi.org/10.1101/gr.912603>.
- Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). Review The Impact of the Gut Microbiota on Human Health : An Integrative View. *Cell*, 148(6), 1258–1270. <http://doi.org/10.1016/j.cell.2012.01.035>
- Collado, M. C., Isolauri, E., Laitinen, K., & Salminen, S. (2008). Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *American Journal of Clinical Nutrition*, 88(4), 894–899. <http://doi.org/88/4/894>

- Cotten, C. M., Taylor, S., Stoll, B., Goldberg, R. N., Hansen, N. I., Sánchez, P. J., Benjamin, D. K. (2009). Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. *Pediatrics*, *123*(1), 58–66. <http://doi.org/10.1542/peds.2007-3423>
- Dave, M., Higgins, P. D., Middha, S., & Rioux, K. P. (2012). The human gut microbiome: Current knowledge, challenges, and future directions. *Translational Research*, *160*(4), 246–257. <http://doi.org/10.1016/j.trsl.2012.05.003>
- DeLong, E. F., & Pace, N. R. (2001). Environmental diversity of bacteria and archaea. *Systematic Biology*, *50*(4), 470–478. <http://doi.org/10.1080/10635150118513>
- Dethlefsen, L., Eckburg, P. B., Bik, E. M., & Relman, D. a. (2006). Assembly of the human intestinal microbiota. *Trends in Ecology and Evolution*, *21*(9), 517–523. <http://doi.org/10.1016/j.tree.2006.06.013>
- Dogra, S., Sakwinska, O., Soh, S.-E., Ngo-Bru, C., Bruck, W. M., Berger, B., ... Group, G. S. (2015). Dynamics of Infant Gut Microbiota Are Influenced by Delivery Mode and Gestational Duration and Are Associated with Subsequent. *mBio*, *6*(1), 1–9. <http://doi.org/10.1128/mBio.02419-14>.Invited
- Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(26), 11971–11975. <http://doi.org/10.1073/pnas.1002601107>
- Fernández, L., Langa, S., Martín, V., Maldonado, A., Jiménez, E., Martín, R., & Rodríguez, J. M. (2013). The human milk microbiota: Origin and potential roles in health and disease. *Pharmacological Research*, *69*(1), 1–10. <http://doi.org/10.1016/j.phrs.2012.09.001>
- Forslund, K., Sunagawa, S., Kultima, J. R., Mende, D. R., Arumugam, M., Typas, a, & Bork, P. (2013). Country-specific antibiotic use practices impact the human gut resistome. *Genome Res*, *23*(7), 1163–1169. <http://doi.org/10.1101/gr.155465.113>
- Franzosa, E. a, Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Huttenhower, C. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), E2329–38. <http://doi.org/10.1073/pnas.1319284111>
- Giannoukos, G., Ciulla, D. M., Huang, K., Haas, B. J., IZard, J., Levin, J. Z., Gnirke, A. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology*, *13*(3), R23. <http://doi.org/10.1186/gb-2012-13-3-r23>
- Gosalbes, M. J., Durbán, A., Pignatelli, M., Abellan, J. J., Jiménez-Hernández, N., Pérez-Cobas, A. E., Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS ONE*, *6*(3), 1–9. <http://doi.org/10.1371/journal.pone.0017447>
- Gosalbes, M. J., Llop, S., Vallès, Y., Moya, a., Ballester, F., & Francino, M. P. (2013). Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical and Experimental Allergy*, *43*(2), 198–211. <http://doi.org/10.1111/cea.12063>
- Grenham, S., Clarke, G., Cryan, J. F., & Dinan, T. G. (2011). Brain-gut-microbe communication in health and disease. *Frontiers in Physiology*, *2* DEC(December), 1–15. <http://doi.org/10.3389/fphys.2011.00094>
- Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Peter, J., Samuel, B. S., Nelson, K. E. (2011). NIH Public Access, *312*(5778), 1355–1359. <http://doi.org/10.1126/science.1124234>.Metagenomic
- Guaraldi, F., & Salvatori, G. (2012). Effect of Breast and Formula Feeding on Gut Microbiota Shaping in Newborns. *Frontiers in Cellular and Infection Microbiology*, *2*(October), 1–4. <http://doi.org/10.3389/fcimb.2012.00094>
- Guarner, F., & Malagelada, J.-R. (2003). Gut flora in health and disease. *The Lancet*, *360*(9356), 512–519. [http://doi.org/10.1016/S0140-6736\(03\)12489-0](http://doi.org/10.1016/S0140-6736(03)12489-0)

- Harzallah, D., & Belhadj, H. (2013). Lactic Acid Bacteria as Probiotics : Characteristics , Selection Criteria and Role in Immunomodulation of Human GI Muccosal Barrier. *Characteristics, Selection Criteria and Role in Immunomodulation of Human GI Muccosal Barrier*, 197–216. <http://doi.org/http://dx.doi.org/10.5772/50732>
- Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., Bushman, F. D. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE*, 8(6). <http://doi.org/10.1371/journal.pone.0066019>
- Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., ... Andersson, A. F. (2014). Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut*, 63(4), 559–66. <http://doi.org/10.1136/gutjnl-2012-303249>
- Jiménez, E., Marín, M. L., Martín, R., Odriozola, J. M., Olivares, M., Xaus, J., Rodríguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Research in Microbiology*, 159(3), 187–193. <http://doi.org/10.1016/j.resmic.2007.12.007>
- Karlsson, F., Tremaroli, V., Nielsen, J., & Bäckhed, F. (2013). Assessing the human gut microbiota in metabolic diseases. *Diabetes*, 62(10), 3341–3349. <http://doi.org/10.2337/db13-0844>
- Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl , 4578–4585. <http://doi.org/10.1073/pnas.1000081107>
- Korecka, A., & Arulampalam, V. (2012). The gut microbiome: Scourge, sentinel or spectator? *Journal of Oral Microbiology*, 4(2012), 1–14. <http://doi.org/10.3402/jom.v4i0.9367>
- Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Kling Bäckhed, H., Ley, R. E. (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell*, 150(3), 470–480. <http://doi.org/10.1016/j.cell.2012.07.008>
- Letunic, I., Yamada, T., Kanehisa, M., & Bork, P. (2008). NCBI HomoloGene/viPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci*, 33(3), 101–103. Retrieved from <http://www.ncbi.nlm.nih.gov/homologene>
- Ley, R. E., Peterson, D. a., & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124(4), 837–848. <http://doi.org/10.1016/j.cell.2006.02.017>
- Li, F., Hullar, M. a, & Lampe, J. W. (2007). Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. *J Microbiol Methods*, 68(2), 303–311. <http://doi.org/10.1016/j.mimet.2006.09.006>
- Mackie, R. I., Sghir, A., & Gaskins, H. R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *American Journal of Clinical Nutrition*, 69(5).
- Marques, T. M., Wall, R., Ross, R. P., Fitzgerald, G. F., Ryan, C. A., & Stanton, C. (2010). Programming infant gut microbiota: Influence of dietary and environmental factors. *Current Opinion in Biotechnology*, 21(2), 149–156. <http://doi.org/10.1016/j.copbio.2010.03.020>
- Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., & De La Cochetiere, M. F. (2013). Development of intestinal microbiota in infants and its impact on health. *Trends in Microbiology*, 21(4), 167–173. <http://doi.org/10.1016/j.tim.2012.12.001>
- Mitra, S., Rupek, P., Richter, D. C., Urich, T., Gilbert, J. a, Meyer, F., Huson, D. H. (2011). Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. *BMC Bioinformatics*, 12 Suppl 1(Suppl 1), S21. <http://doi.org/10.1186/1471-2105-12-S1-S21>
- Moore, T. a., Hanson, C. K., & Anderson-Berry, a. (2011). Colonization of the Gastrointestinal Tract in Neonates: A Review. *ICAN: Infant, Child, & Adolescent Nutrition*, 3(5), 291–295. <http://doi.org/10.1177/1941406411421629>

- Nocker, A., Burr, M., & Camper, A. K. (2007). Genotypic microbial community profiling: A critical technical review. *Microbial Ecology*, *54*(2), 276–289. <http://doi.org/10.1007/s00248-006-9199-5>
- Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. a., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biology*, *5*(7), 1556–1573. <http://doi.org/10.1371/journal.pbio.0050177>
- Peano, C., Pietrelli, A., Consolandi, C., Rossi, E., Petiti, L., Tagliabue, L., Landini, P. (2013). An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microbial Informatics and Experimentation*, *3*(1), 1. <http://doi.org/10.1186/2042-5783-3-1>
- Penders, J., Stobberingh, E. E., Brandt, P. a V. D., & Thijs, C. (2007). The role of the intestinal microbiota in the development of atopic disorders. *Allergy: European Journal of Allergy and Clinical Immunology*, *62*(11), 1223–1236. <http://doi.org/10.1111/j.1398-9995.2007.01462.x>
- Poroyko, V., White, J. R., Wang, M., Donovan, S., Alverdy, J., Liu, D. C., & Morowitz, M. J. (2010). Gut microbial gene expression in mother-fed and formula-fed piglets. *PLoS ONE*, *5*(8). <http://doi.org/10.1371/journal.pone.0012459>
- Prince, A. L., Antony, K. M., Ma, J., & Aagaard, K. M. (2014). The microbiome and development: A mother's perspective. *Seminars in Reproductive Medicine*, *32*(1), 14–22. <http://doi.org/10.1055/s-0033-1361818>
- Rautava, S., Luoto, R., Salminen, S., & Isolauri, E. (2012). Microbial contact during pregnancy, intestinal colonization and human disease. *Nature Reviews Gastroenterology & Hepatology*, *9*(10), 565–576. <http://doi.org/10.1038/nrgastro.2012.144>
- Reigstad, C. S., & Kashyap, P. C. (2013). Beyond phylotyping: Understanding the impact of gut microbiota on host biology. *Neurogastroenterology and Motility*, *25*(5), 358–372. <http://doi.org/10.1111/nmo.12134>
- Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics*, *38*, 525–552. <http://doi.org/10.1146/annurev.genet.38.072902.091216>
- Rodri, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health & Disease*, *1*(26), 26050.
- Savino, F., Roana, J., Mandras, N., Tarasco, V., Locatelli, E., & Tullio, V. (2011). Faecal microbiota in breast-fed infants after antibiotic therapy. *Acta Paediatrica, International Journal of Paediatrics*, *100*(1), 75–78. <http://doi.org/10.1111/j.1651-2227.2010.01988.x>
- Sekirov, I., Russell, S., & Antunes, L. (2010). Gut microbiota in health and disease. *Physiological Reviews*, *90*(3), 859–904. <http://doi.org/10.1152/physrev.00045.2009>
- Sjögren, K., Engdahl, C., Henning, P., Lerner, U. H., Tremaroli, V., Lagerquist, M. K., Ohlsson, C. (2012). The gut microbiota regulates bone mass in mice. *Journal of Bone and Mineral Research*, *27*(6), 1357–1367. <http://doi.org/10.1002/jbmr.1588>
- Sorek, R., & Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature Reviews. Genetics*, *11*(1), 9–16. <http://doi.org/10.1038/nrg2695>
- Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and protein extraction: The past and the present. *Journal of Biomedicine and Biotechnology*, 2009. <http://doi.org/10.1155/2009/574398>
- Trosvik, P., Stenseth, N. C., & Rudi, K. (2010). Convergent temporal dynamics of the human infant gut microbiota. *The ISME Journal*, *4*(2), 151–158. <http://doi.org/10.1038/ismej.2009.96>
- Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Ley, R. E., Sogin, M. L., Gordon, J. I. (2009). A core gut microbiom in obese and lean twins. *Nature*, *457*(32089), 480–484. <http://doi.org/10.1038/nature07540.A>

- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. <http://doi.org/10.1038/nature06244>
- Vael, C., & Desager, K. (2009). The importance of the development of the intestinal microbiota in infancy. *Current Opinion in Pediatrics*, *21*(6), 794–800. <http://doi.org/10.1097/MOP.0b013e328332351b>
- Vaishampayan, P. a., Kuehl, J. V., Froula, J. L., Morgan, J. L., Ochman, H., & Francino, M. P. (2010). Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biology and Evolution*, *2*(1), 53–66. <http://doi.org/10.1093/gbe/evp057>
- Vallès, Y., Artacho, A., Pascual-García, A., Ferrús, M. L., Gosalbes, M. J., Abellán, J. J., & Francino, M. P. (2014). Microbial Succession in the Gut: Directional Trends of Taxonomic and Functional Change in a Birth Cohort of Spanish Infants. *PLoS Genetics*, *10*(6). <http://doi.org/10.1371/journal.pgen.1004406>
- Van Vliet, a. H. M. (2010). Next generation sequencing of microbial transcriptomes: Challenges and opportunities. *FEMS Microbiology Letters*, *302*(1), 1–7. <http://doi.org/10.1111/j.1574-6968.2009.01767.x>
- Xiong, X., Frank, D. N., Robertson, C. E., Hung, S. S., Markle, J., Canty, A. J., Parkinson, J. (2012). Generation and analysis of a mouse intestinal metatranscriptome through Illumina based RNA-sequencing. *PLoS ONE*, *7*(4). <http://doi.org/10.1371/journal.pone.0036009>
- Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222–227. <http://doi.org/10.1038/nature11053>
- Young, V. B., & Schmidt, T. M. (2004). Antibiotic-Associated Diarrhea Accompanied by Large-Scale Alterations in the Composition of the Fecal Microbiota Antibiotic-Associated Diarrhea Accompanied by Large-Scale Alterations in the Composition of the Fecal Microbiota, *42*(3), 1203–1206. <http://doi.org/10.1128/JCM.42.3.1203>
- Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y., Sue, a, ... Li, H. (2012). NIH Public Access. *Science*, *334*(6052), 105–108. <http://doi.org/10.1126/science.1208344>.Linking

7. ANEXO

Implicated Microbiota ^a	Changes in Microbiota Presence/Function	References ^b
Allergies		
<i>Lactobacillus</i> spp. ↓	early colonization with <i>Lactobacillus</i> associated w/decreased allergies	Round et al., 2011
<i>Bifidobacterium adolescentis</i> ↓	early colonization with more diverse microbiota	Round and Mazmanian, 2009
<i>Clostridium difficile</i> ↓	might prevent allergies	
<i>Helicobacter pylori</i> ↓	<i>H. pylori</i> tolerance mediated by Tregs that suppress asthma	Arnold et al., 2011
Celiac's disease		
<i>Bacteroides vulgatus</i> ↑	higher diversity (Shannon-Wiener index) in	Elinav et al., 2011
<i>Escherichia coli</i> ↓	Celiac's disease patients versus controls	
<i>Clostridium coccooides</i> ↓		
Gastric Cancer		
<i>H. pylori</i> ↑	important element in carcinogenic pathway for developing gastric adenocarcinomas	Lathrop et al., 2011
Austim		
Bacteroidetes ↑	increased bacterial diversity in feces of autistic children compared	Robinson et al., 2010
Proteobacteria ↑	to controls	
Actinobacteria ↓		
Firmicutes ↓		
Obesity		
Bacteroidetes ↓	significant changes in gut microbiota are associated with	Ley et al., 2005; Pflughoeft and Versalovic, 2011
<i>Lactobacillus</i> ↑	increasing obesity	
Firmicutes/Bacteroidetes ratio ↓		Ley et al., 2005
<i>Methanobrevibacter smithii</i> ↓		Turnbaugh et al., 2009b
Anorexia		
<i>Methanobrevibacter smithii</i> ↑	Bacteroidetes, Firmicutes, and <i>Lactobacillus</i> similar to lean patients, though <i>M. smithii</i> significantly increased	Armougom et al., 2009; Pflughoeft and Versalovic, 2011
IBD— Crohn's Disease		
<i>Bacteroides ovatus</i> ↑	less diversity in patients with Crohn's disease compared to	Dicksved et al., 2008
<i>Bacteroides vulgatus</i> ↑	healthy patients	
<i>Bacteroides uniformis</i> ↓		
IBD (General)		
Bacteroidetes ↓	IBD associated with overall community dysbiosis rather	Spor et al., 2011;
<i>Lachnospiraceae</i> ↓	than single causal bacterial species	Perry et al., 2006
Actinobacteria ↑		
Proteobacteria ↑		
<i>Clostridium leptum</i> ↓		
<i>Clostridium coccooides</i> ↓		
<i>Faecalibacterium prasnitzii</i> ↓		
Firmicutes/Bacteroidetes ratio ↓		
<i>Bifidobacteria</i> ↓		
Type 2 Diabetes		
Firmicutes ↓	shifts in gut microbiota associated with increases in plasma	Brown, 2000
<i>Clostridia</i> ↓	glucose concentrations	
<i>Bacteroides-Prevotella</i> ↑ versus		
<i>Clostridia coccooides-Eubacterium</i>		
<i>rectale</i> ↓		
<i>Betaproteobacteria</i> ↑		
Bacteroidetes/Firmicutes ratio ↑		

^aChanges relative to healthy subjects. Increase: ↑. Decrease: ↓.

^bReferences are exemplary rather than exhaustive and focus on studies that compare healthy versus diseased individuals.

Figura S1: Cambios en la microbiota intestinal asociados a distintas enfermedades (Clemente et al., 2012)

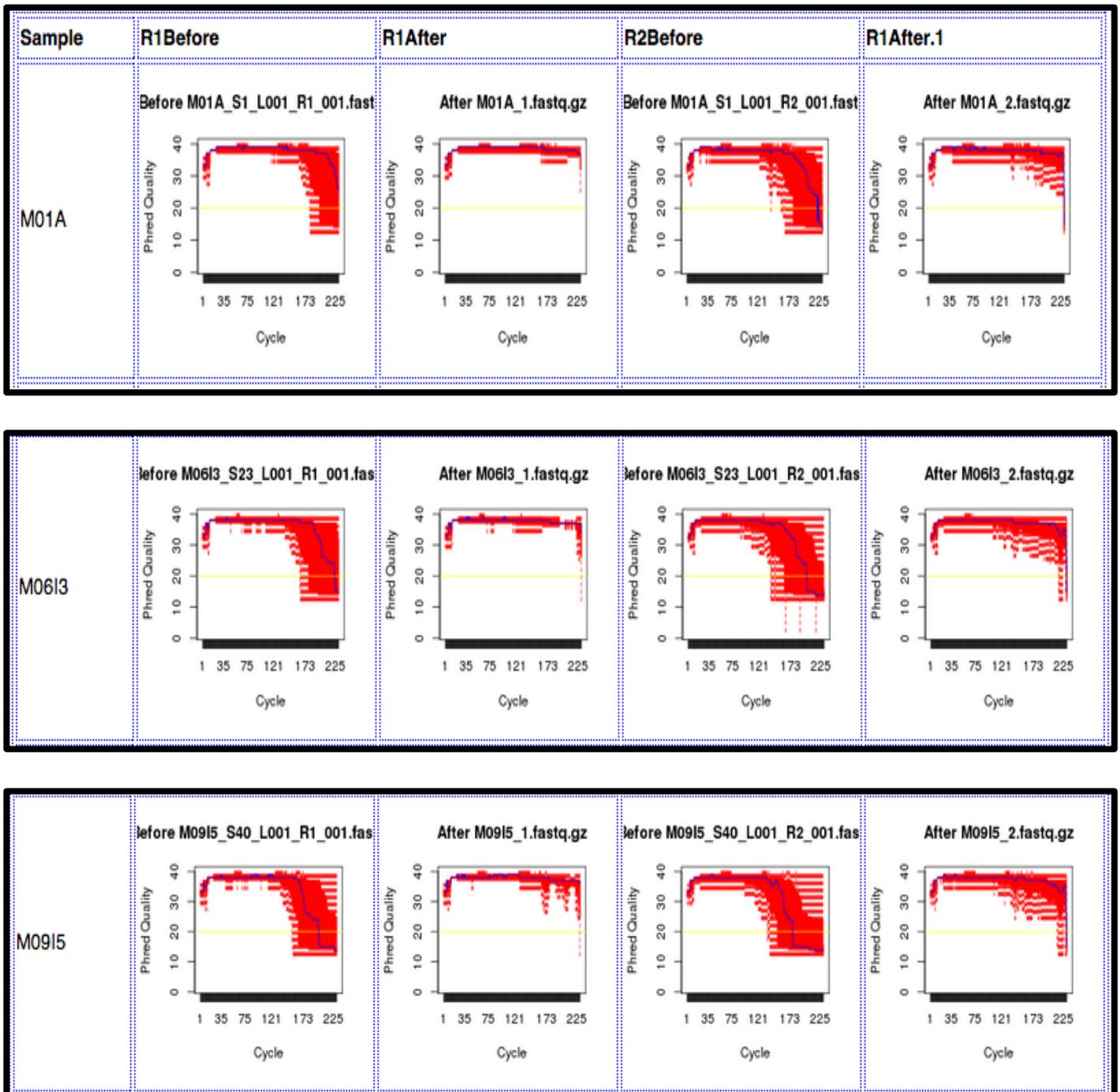


Figura S2: Distribución de la calidad por ciclo del secuenciador de algunas muestras antes y después de la evaluación cualitativa

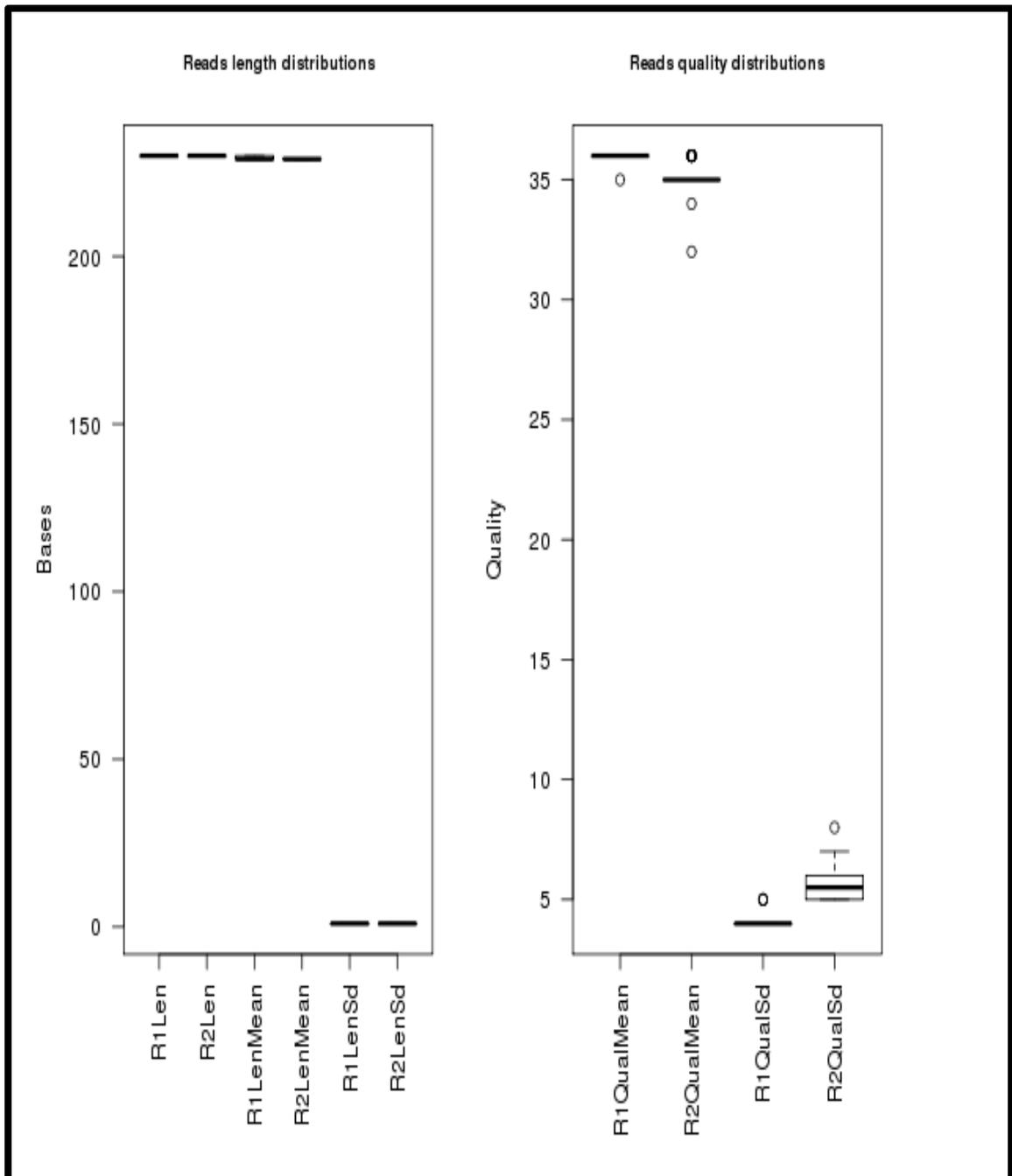


Figura S3: Distribución de la longitud (izquierda) y calidad (derecha) de los reads antes del joining.

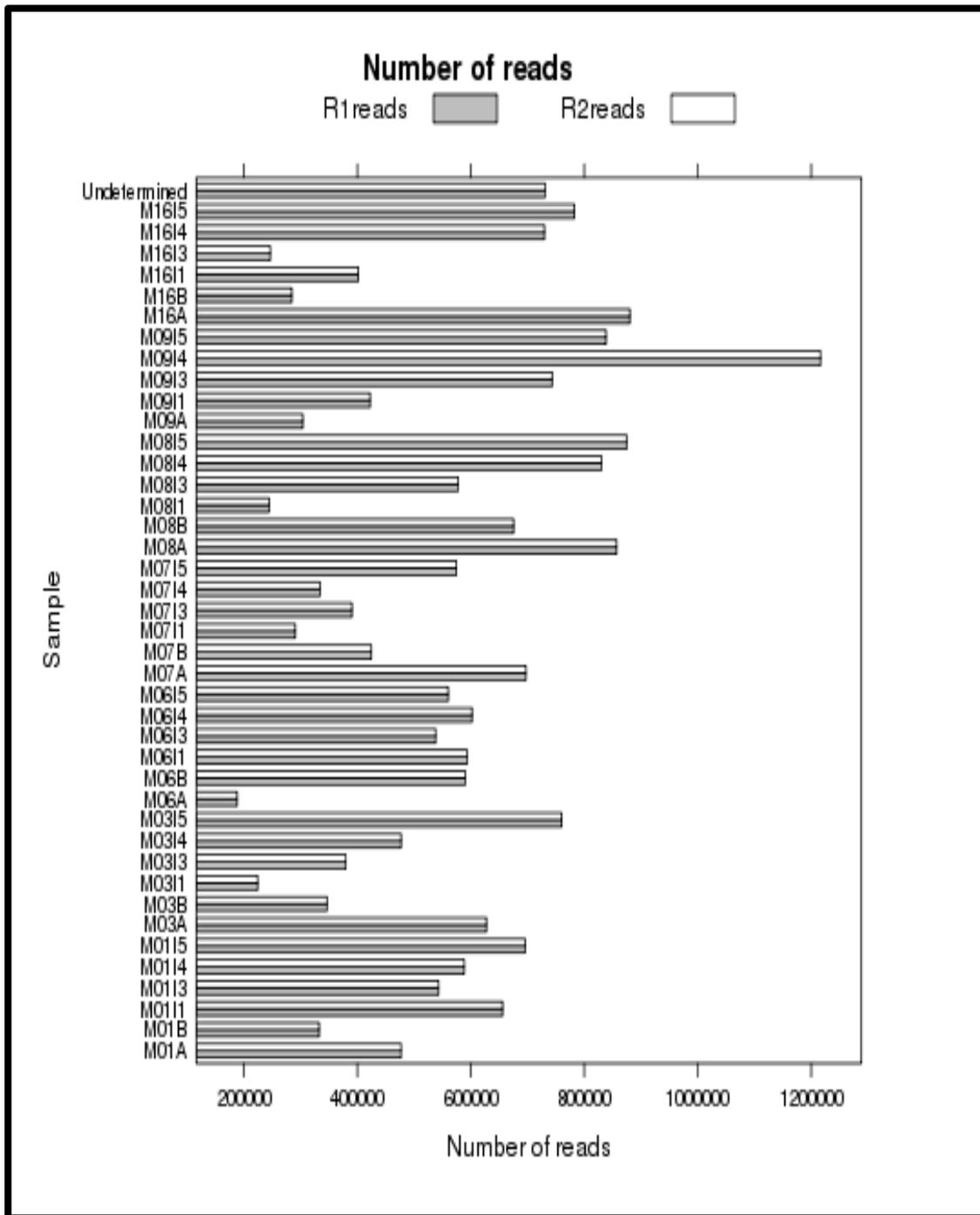


Figura S4: Número de reads (R1 y R2) obtenidos por muestra.

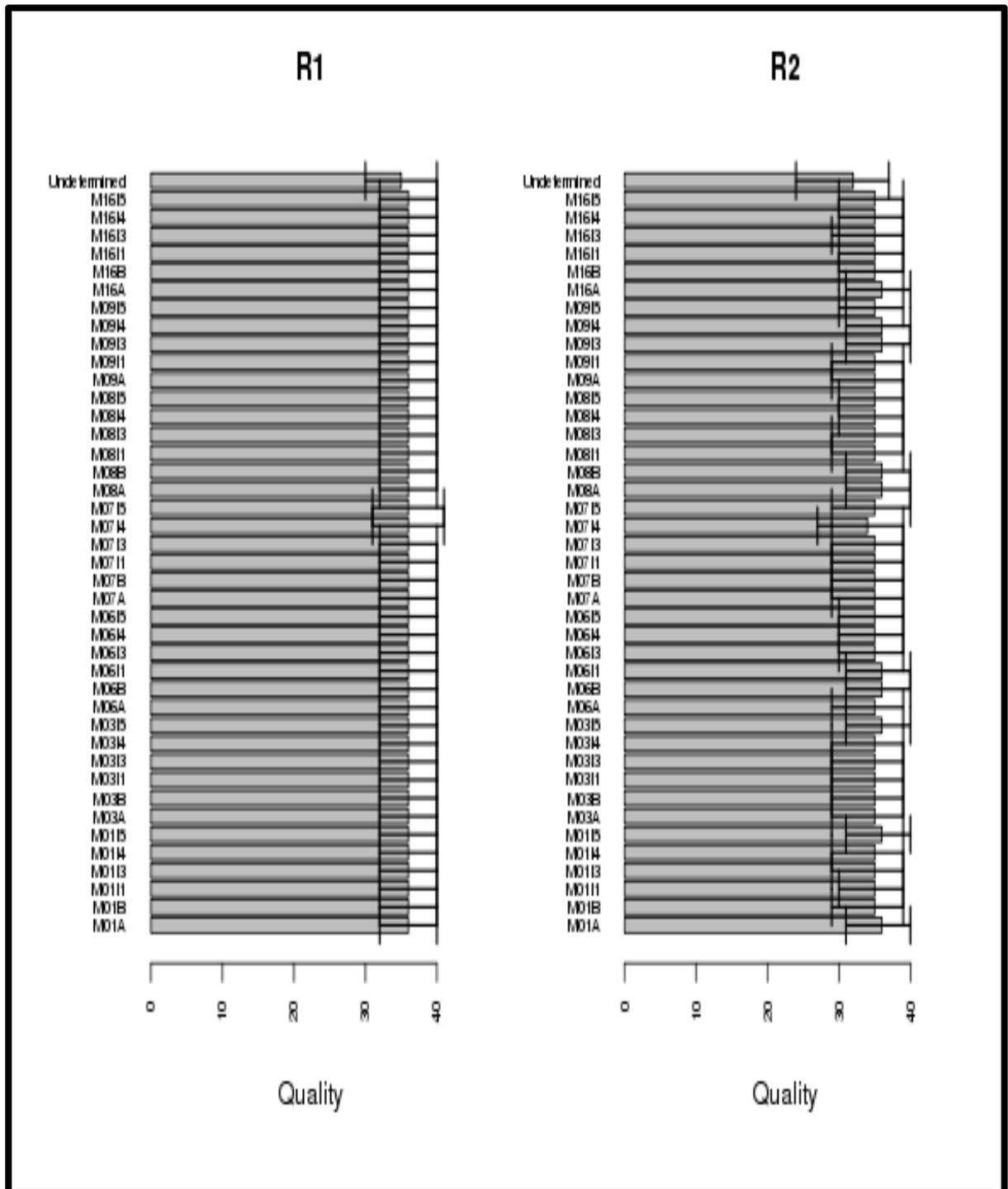


Figura S5: Distribución de las calidades antes de la evaluación cualitativa.

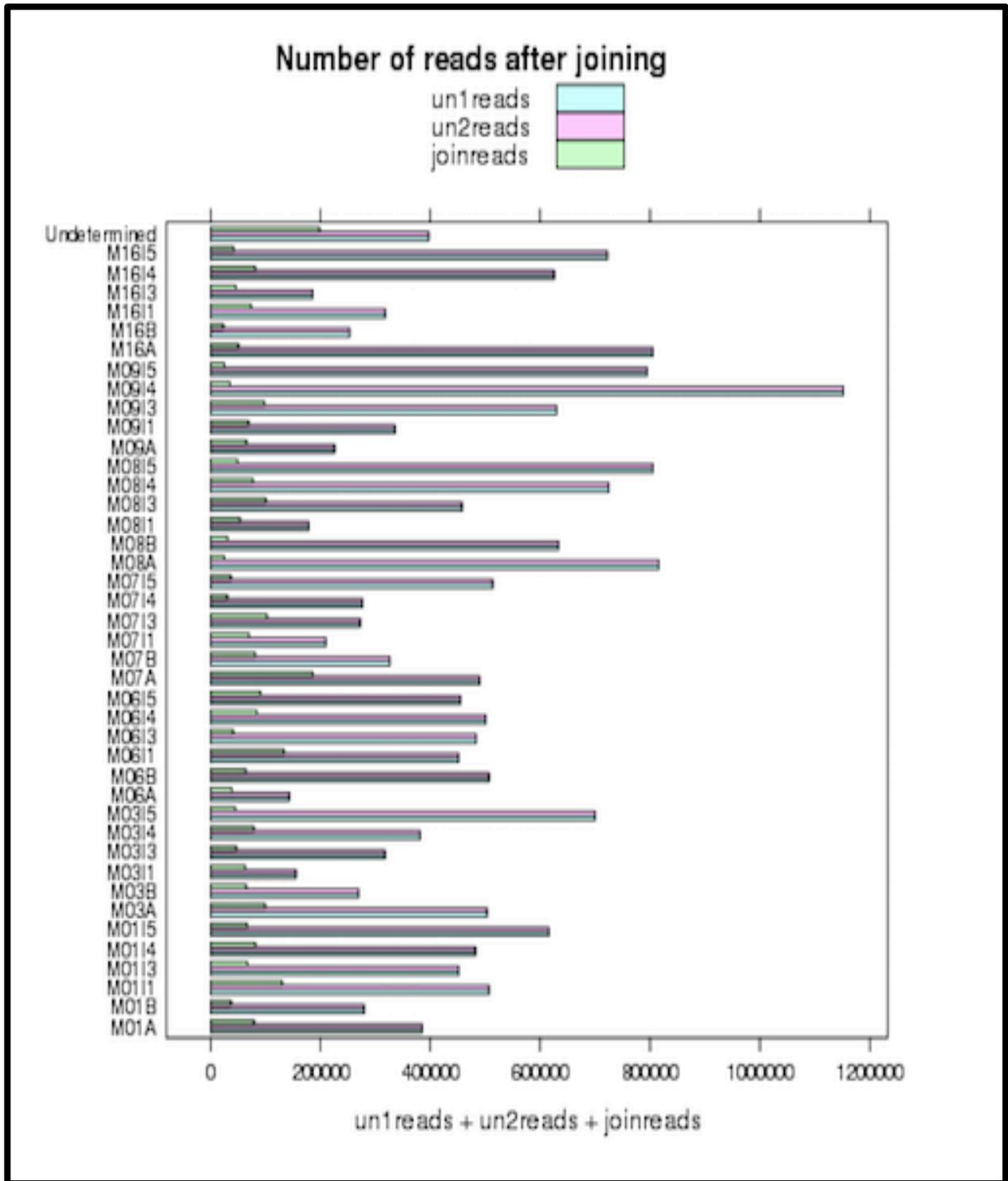


Figura S8: Número de reads unidos (*joined*) y no unidos (*unjoined*) por muestra.

Tabla S1: Número de *reads* por muestra con sus longitud y desviación estándar medias.

Sample	R1 reads	R1Len Mean	R1Len Sd	R2 reads	R2LenMean	R2LenSd
M01A	476914	229	1	476914	229	1
M01B	332618	229	1	332618	229	1
M01I1	656051	229	1	656051	229	1
M01I3	543191	229	1	543191	229	1
M01I4	588320	229	1	588320	229	1
M01I5	695439	229	1	695439	229	1
M03A	627517	229	1	627517	229	1
M03B	346245	230	1	346245	229	1
M03I1	224180	230	1	224180	229	1
M03I3	379589	229	1	379589	229	1
M03I4	476951	229	1	476951	229	1
M03I5	759856	229	1	759856	229	1
M06A	187940	230	1	187940	229	1
M06B	589773	229	1	589773	229	1
M06I1	593029	230	1	593029	229	1
M06I3	537933	229	1	537933	229	1
M06I4	602054	230	1	602054	229	1
M06I5	560006	230	1	560006	229	1
M07A	697166	229	1	697166	229	1
M07B	424508	229	1	424508	229	1
M07I1	290469	230	1	290469	229	1
M07I3	390364	230	1	390364	229	1
M07I4	333711	229	1	333711	229	1
M07I5	574462	229	1	574462	229	1
M08A	856995	230	1	856995	229	1
M08B	675394	229	1	675394	229	1
M08I1	244098	230	1	244098	229	1
M08I3	577273	229	1	577273	229	1
M08I4	829869	229	1	829869	229	1
M08I5	874430	229	1	874430	229	1
M09A	303734	230	1	303734	229	1
M09I1	422439	229	1	422439	229	1
M09I3	743253	229	1	743253	229	1
M09I4	1216474	229	1	1216474	229	1
M09I5	838283	229	1	838283	229	1
M16A	880391	229	1	880391	229	1
M16B	284420	229	1	284420	229	1
M16I1	401509	229	1	401509	229	1
M16I3	246336	229	1	246336	229	1
M16I4	729737	229	1	729737	229	1
M16I5	782370	229	1	782370	229	1
Undetermined	730708	229	1	730708	229	1

Tabla S2: Resumen después del *joining*.

Muestra	joinreads	joinLenMean	joinLenSd	joinQualMean	joinQualSd
M01A	79227	282	61	36	4
M01B	37612	273	63	36	4
M01I1	131047	288	55	36	4
M01I3	67641	275	69	36	4
M01I4	82155	275	54	36	4
M01I5	67336	279	54	36	4
M03A	99836	279	61	36	4
M03B	65107	282	57	36	4
M03I1	62699	300	60	36	4
M03I3	47189	273	58	36	4
M03I4	79177	284	53	36	4
M03I5	44968	283	55	36	4
M06A	38190	283	57	36	4
M06B	64561	276	52	36	4
M06I1	134267	296	56	36	4
M06I3	41348	281	57	36	4
M06I4	84309	285	62	36	4
M06I5	90685	284	60	36	4
M07A	186468	298	58	36	4
M07B	81747	283	52	36	4
M07I1	69672	288	67	36	4
M07I3	102956	278	58	36	4
M07I4	30744	236	75	36	4
M07I5	37488	272	59	36	4
M08A	24739	255	63	36	4
M08B	30805	279	58	36	4
M08I1	53953	297	61	36	4
M08I3	101015	288	58	36	4
M08I4	77032	274	61	36	4
M08I5	48684	278	56	36	4
M09A	65433	288	63	36	4
M09I1	69438	287	59	36	4
M09I3	98138	291	62	36	4
M09I4	35055	273	61	36	4
M09I5	25288	265	60	36	4
M16A	51121	266	50	36	4
M16B	23604	287	59	36	4
M16I1	74461	286	54	36	4
M16I3	46304	268	67	36	4
M16I4	81967	275	51	36	4
M16I5	42912	270	56	36	4
Undetermined	198639	316	86	36	4

Tabla S3: Reads obtenidos tras la limpieza frente a los iniciales y el respectivo porcentaje de reducción.

Muestra	Reads iniciales	Reads después de la limpieza	Porcentaje de reducción
M01A	850547	825074	2,99%
M01B	596704	569342	4,59%
M01I1	1144827	1100887	3,84%
M01I3	972549	924015	4,99%
M01I4	1047021	966942	7,65%
M01I5	1299256	1196736	7,89%
M03A	1106354	1053944	4,74%
M03B	603721	575237	4,72%
M03I1	374261	359175	4,03%
M03I3	682335	642794	5,79%
M03I4	841705	799936	4,96%
M03I5	1444630	1294981	10,36%
M06A	324670	300057	7,58%
M06B	1078709	1041661	3,43%
M06I1	1038429	1016322	2,13%
M06I3	1009628	571335	43,41%
M06I4	1087355	1048910	3,54%
M06I5	1001241	971252	3,00%
M07A	1166470	919189	21,20%
M07B	735139	617230	16,04%
M07I1	488724	471374	3,55%
M07I3	648110	620653	4,24%
M07I4	585238	546109	6,69%
M07I5	1064730	1011217	5,03%
M08A	1656491	1597439	3,56%
M08B	1298255	1259829	2,96%
M08I1	411749	391484	4,92%
M08I3	1016871	975572	4,06%
M08I4	1527468	1440445	5,70%
M08I5	1658992	1598705	3,63%
M09A	517549	498314	3,72%
M09I1	740598	705846	4,69%
M09I3	1357624	1291901	4,84%
M09I4	2339701	2215858	5,29%
M09I5	1615364	1543828	4,43%
M16A	1663003	1571281	5,52%
M16B	530740	479513	9,65%
M16I1	710321	685069	3,56%
M16I3	418564	397009	5,15%
M16I4	1333723	1196168	10,31%
M16I5	1488422	1432609	3,75%

Tabla S4: Tipos de secuencias eliminados durante el procesamiento bioinformático.

Muestra	Ribosomales bacterianas	Ribosomales humanas	Ribosomales eucariotas	Reads humanas
M01A	2680	77	137	358
M01B	3653	245	921	289
M01I1	885	35	4	26
M01I3	4638	736	227	1222
M01I4	32381	2826	559	958
M01I5	68404	33	45	25
M03A	10237	99	148	299
M03B	6240	1360	144	863
M03I1	2234	66	21	775
M03I3	9727	61	17	227
M03I4	13022	119	226	277
M03I5	103159	108	85	15
M06A	11788	52	72	516
M06B	2702	122	109	100
M06I1	1060	37	20	61
M06I3	396780	3	12	2
M06I4	8401	47	23	274
M06I5	2545	39	10	154
M07A	167828	112	36806	86
M07B	84060	23	42	20
M07I1	3068	14	13	267
M07I3	3424	2525	159	391
M07I4	6001	56	21	324
M07I5	4233	161	84	103
M08A	6881	66	71	104
M08B	3414	22	36	23
M08I1	899	151	38	490
M08I3	3040	204	35	97
M08I4	30205	902	1269	707
M08I5	2939	185	113	42
M09A	1589	70	265	306
M09I1	3063	723	107	455
M09I3	24748	1	14	151
M09I4	42791	152	277	27
M09I5	14319	20	24	67
M16A	37056	55	99	53
M16B	30061	38	7	306
M16I1	2387	33	4	58
M16I3	3224	89	18	1015
M16I4	89235	785	95	178
M16I5	7049	58	109	68