Dinámica de la microbiota en niños, púberes y adultos

PhD Thesis

Alejandra Rey Mariño

January 2022

Supervisor:

Dr. M. Pilar Francino Puget

VNIVERSITAT DE VALÈNCIA

Doctorado en Biomedicina y Biotecnología

Fundació Fisabio

La Dr. M. Pilar Francino Puget,

CERTIFICA

Que Doña Alejandra Rey Mariño, Licenciada en Biología por la *Universidade da Coruña*, ha realizado bajo su dirección el trabajo titulado: "Dinámica de la microbiota en niños, púberes y adultos", que presenta para optar al grado de Doctora en el programa de Biomedicina y Biotecnología por la *Universitat de València*. Asimismo, certifica haber dirigido y supervisado tanto los distintos aspectos del trabajo como su redacción.

Y para que conste a los efectos oportunos, en cumplimiento de la legislación vigente, firma el presente certificado en

Valencia, a 18 de enero de 2022

La directora:

Fdo. M. Pilar Francino Puget

This PhD Thesis work was supported by grant UPG-15-286 from Ministerio de Economía, Industria y Competitividad (Gobierno de España).

Agradecimientos

En primer lugar, quiero agradecer de todo corazón a las personas que han participado en este estudio, a mis voluntarios, porque sin ellos ahora no estaría escribiendo esta tesis. Muchas gracias por la paciencia que habéis tenido, por vuestro desinterés e interés en la ciencia. GRACIAS.

A mi directora de tesis, Pilar, por haberme dado la oportunidad de crecer profesionalmente (y también personalmente) en este tiempo. A Andrés por la confianza que depositó en mi desde mi TFM. A Pepa, Nuria, Javi y Artacho, porque esta tesis es de ellos también. A Dani y Alberto, los mejores compis de grupo que puede tener una al tercer año de tesis (¡ojalá hubierais llegado antes!). A Susana, por haber montado todo el tinglado antes de que yo llegara, por haberme enseñado todo desde el principio, por dejarme acudir a ella cuando estaba más perdida que un pulpo en un garaje, y por los xuxus de crema. A todos mis compis de FISABIO, los que estuvisteis y los que estáis, porque las comidas sin esas conversaciones *random* hubieran sido muy aburridas y porque los agobios si se llevan en grupo se llevan mejor.

A mis *nenis* Carlota y Vane, porque siempre estuvieron orgullosas de mi (¡XSP!). A Alba, Irene y Grecia por haberme preparado para este momento. Estaba todo pensado para que pasarais vosotras primero por esto (vosotras id yendo, yo ya tal…). A Anita por los paseítos cuando voy a Coruña. A Migui porque sé que SIEMPRE, SIEMPRE, SIEMPRE estarás ahí. A Anis, porque entre otras muchísimas cosas, sino fuera por ti no hubiera sabido que Valencia mola. A mis *monstros*: Majo, Ana, Pablo, Iván, Julen (*Poblet Punk Family forever!*), Laura y Carlos (mi familia valenciana) e Isa (L). Gracias por haber hecho que me sienta como en mi casa a 900 km de ella. A Muri por la turra que le di estos últimos 3 meses con LA TESIS (hoy en día lo que se valora es la tranquilidad y un buen meme). A mis norteños festivaleros, porque *non todo vai ser traballar*. Gracias a Ana Siscar por descubrirme una forma de expresar que en la vida hubiera imaginado que me llegaría a llenar tanto (se agradece descansar el cerebro y soltar un poco el corazón). Y gracias también a toda la gente que conocí en estos ya (*mi madriña*) casi ocho años. *Casalla Paradise*.

Por supuesto a mis compis de piso Albi y Punkiño, porque a ver quién iba a aguantar mis manías. Pats, muchas gracias por haber sido MI PERSONA este tiempo.

A mi familia por estar y por hacer que esté orgullosa de venir de dónde vengo. Y por supuesto a mis padres, por el apoyo, el amor y la paciencia que han tenido estos años, por ser mi faro y mi ancla.

Resumen amplio

Introducción

La microbiota intestinal se implanta a lo largo del desarrollo del individuo y su composición y funciones difieren según la edad. Conocer los cambios que se producen en la microbiota intestinal a lo largo de la vida puede ayudar al diagnóstico, tratamiento y prevención de enfermedades relacionadas con alteraciones metabólicas e inmunitarias. Pretendemos contribuir a la comprensión del proceso de desarrollo de la microbiota intestinal, centrándonos en dos períodos que han recibido poca atención: los primeros años y la adolescencia.

Objetivos

La presente tesis doctoral tiene como objetivos principales, por un lado, determinar y comparar los parámetros que caracterizan a las microbiotas sanas en niños, adolescentes y adultos, y por otro, determinar y comparar la estabilidad de la microbiota intestinal en niños, adolescentes y adultos en cuanto a su composición y/o función. Además, los objetivos secundarios son determinar si existen diferencias significativas en la

microbiota intestinal de niños antes y después del destete en cuanto a composición y/o función, así como determinar si existen diferencias significativas en adolescentes antes y después de la menarquia en cuanto a composición y/o función.

Métodos

Reclutamos a 60 participantes sanos de 24 familias de Valencia (España), divididos en tres grupos de edad: adultos (n = 35; 41,6 años de edad de media al inicio del estudio), adolescentes (n = 13; 11,9 años de edad de media al inicio del estudio) y niños (n = 12; 20 meses de edad de media al inicio del estudio). Se recogieron muestras fecales de cada uno de los participantes cada tres o cuatro meses durante tres años aproximadamente, con un total de 10 puntos temporales de muestreo. En total, se recolectaron 535 muestras (119 de niños, 101 de adolescentes y 315 de adultos). A continuación, se extrajo el ADN genómico total de las bacterias presentes en las muestras fecales. La región V3-V4 del gen 16S del ARN ribosomal se amplificó y se secuenció en la plataforma MiSeq (Illumina) realizando controles negativos en cada paso de los protocolos. El procesamiento, ensamblaje y anotación de las lecturas se realizó utilizando el flujo de trabajo del paquete DADA2 de R, con

asignación a nivel de especie al 100% de similitud o asignación al nivel taxonómico más detallado posible. Luego se empleó BLAST para proporcionar asignaciones a nivel de especie a los ASV que permanecieron sin asignar bajo este criterio estricto, utilizando un límite de similitud del 97 %. Finalmente, se obtuvo una tabla de contingencia con los ASV anotados en cada nivel de taxonomía y sus abundancias absolutas por muestra. Los datos de abundancia se normalizaron utilizando el método *Total Sum Scaling* (TSS). La secuenciación del metagenoma mediante estrategia *shotgun* también se realizó en MiSeq (Illumina) realizando controles negativos en cada paso de los protocolos. El procesamiento previo de las lecturas, el filtrado, el ensamblaje y la anotación se realizaron mediante herramientas informáticas disponibles públicamente. Se generó una tabla de contingencia de anotaciones TIGRFAM y sus abundancias absolutas por muestra y se normalizó mediante TSS. La anotación taxonómica se realizó mapeando las lecturas contra la base de datos de proteínas no redundantes de NCBI usando Kaiju. Además, hemos inferido la taxonomía para cada pauta de lectura abierta encontrada en el conjunto de datos, por lo que hemos podido cuantificar aquellos taxones que contribuyen genéticamente a las funciones codificadas en el metagenoma.

Se representaron gráficamente los perfiles composicionales y funcionales por grupo de edad calculando las abundancias relativas promedio de los datos a nivel de género y de categorías funcionales TIGRFAM por grupo y punto temporal. La diversidad alfa (índice de Shannon y estimador Chao1) se calculó a nivel de ASV para los datos del gen 16S del ARN ribosomal. También aplicamos el test U de Mann-Whitney para verificar diferencias significativas entre diferentes puntos temporales para cada grupo y entre grupos para cada punto temporal para ambos estimadores de diversidad. El análisis CCA (*Canonical Correspondence Analysis*) y el test ADONIS (PMANOVA) se realizaron con los datos funcionales y taxonómicos para mostrar la variación de los datos que puede explicarse por el grupo de edad. Además, se realizó un análisis de coordenadas principales (PCoA) sobre las distancias del índice de Jaccard en función de las abundancias relativas de los ASV a partir de datos de series temporales, incorporando el tiempo como tercer eje en el gráfico para explorar cómo las muestras cambian con el tiempo.

Para identificar biomarcadores taxonómicos y funcionales de los diferentes grupos de edad durante el período de muestreo, aplicamos el

algoritmo LEfSe (análisis discriminante lineal o LDA) sobre los datos a nivel de género, especie y anotación funcional, con el fin de detectar aquellos rasgos exclusivos de cada grupo de edad. El análisis sPLS-DA (*sparse Partial Least Squares Discriminant Analysis*) también se realizó sobre los datos de ASVs y anotaciones funcionales TIGRFAM con el fin de discriminar grupos de muestras e identificar el subconjunto de variables (taxones y funciones) más discriminante de grupo en los datos. Finalmente se construyeron mapas de calor (*Clustered Image Maps* o CIMs) y redes de correlaciones para visualizar los resultados del análisis. Para el análisis de múltiples ómicas se empleó el marco de trabajo DIABLO (implementado en mixOmics). DIABLO realiza la integración de datos de metagenomas y amplicones del gen 16S del ARN ribosomal medidos en las mismas N muestras biológicas (N-integración) de manera supervisada. De esta manera se identifican conjuntos distintivos de variables altamente correlacionadas entre los diferentes tipos de ómicas, mediante el modelado de relaciones entre los conjuntos de datos ómicos. Se construyeron redes de correlaciones con las variables seleccionadas y CIMs para visualizar los resultados de DIABLO. Además, se construyó un diagrama *Circos*, que muestra los diferentes tipos de variables seleccionadas en un

círculo, con enlaces entre ómicas que indican fuertes correlaciones positivas o negativas (r> 0.7), así como los niveles de abundancia relativa de cada variable para cada grupo de edad.

A partir de los datos de series temporales quisimos ver cómo cambia la microbiota con el tiempo en cada uno de los grupos de edad y cuantificar ese cambio, de modo que pudiéramos hacer comparaciones dentro de y entre los grupos. Para hacer esto, empleamos varias estrategias. Primero utilizamos la plataforma de análisis de microbiomas QIIME2 para recopilar información sobre el cambio y la estabilidad de la comunidad microbiana en los diferentes grupos de edad a partir de los datos de amplicones del gen 16S del ARN ribosomal de nuestro muestreo longitudinal. *q2-longitudinal* es una funcionalidad dentro de QIIME2 que incorpora múltiples métodos para realizar y visualizar este tipo de análisis. Con esta herramienta testamos la diferencia en los valores de diversidad alfa (índice de Shannon) entre el primer punto temporal (T1) y el último (T10) dentro del mismo grupo (test de Wilcoxon-signed rank), así como si estas diferencias son significativamente diferentes entre los grupos de edad (test U de Mann-Whitney). También se utilizaron modelos de efectos mixtos lineales para probar los cambios en la

diversidad alfa (índice de Shannon) a lo largo del tiempo y en respuesta a los diferentes grupos de edad. Finalmente, se empleó la regresión de aprendizaje automático (bosques aleatorios o *random forests*) para explorar cómo cambian las comunidades microbianas en los diferentes grupos de edad, aprendiendo de la estructura de los datos para identificar taxones (incluidos los de baja abundancia) que predigan las diferentes etapas o puntos temporales. Las variables importantes son aquellas que cambian con el tiempo y su abundancia predice el punto temporal específico de recogida de la muestra, aunque la importancia de esta variable no implica una significancia estadística, ya que se trata de un método meramente exploratorio. La abundancia longitudinal de cada variable se representó mediante gráficos de "inestabilidad" (*volatility plots*). La segunda estrategia para estudiar la estabilidad global de la microbiota en cuanto a composición y función en los diferentes grupos de edad fue utilizar *complexCruncher*. Este método se basa en el hecho observado de que la dinámica de la microbiota se ajusta bien a la ley de potencia de Taylor, lo que significa que existe una clara correspondencia entre la media relativa y la dispersión de los grupos taxonómicos. Los parámetros de la ley de Taylor son la variabilidad de los taxones (V), que es un

estimador directo de las fluctuaciones en el tiempo, y el índice de escala (β) de la ley de potencia, que proporciona información sobre las propiedades estadísticas del ecosistema, y ambos correlacionan con la estabilidad de la microbiota. El método calcula estos parámetros tanto a nivel de composición (ASV) como de función (subfunciones TIGRFAM) y construye el espacio de parámetros de la ley de Taylor, con la variabilidad de las características (V) en el eje x y el índice de escala (β) en el eje y, ambos en unidades de desviación estándar. La zona de "estabilidad" se estableció tomando como referencia el grupo de adultos y se delimitaron las áreas correspondientes a las regiones que contenían el 68% y el 98% de los sujetos adultos en el espacio de parámetros de Taylor. *ComplexCruncher* también calcula el índice de estabilidad de rango (RSI), una forma de cuantificar la estabilidad de cada una de las variables (ASVs y categorías funcionales) a lo largo del tiempo en cada grupo de edad. Realizamos un análisis CCA y un test ADONIS basados en los RSI para verificar si podemos separar las comunidades microbianas de los diferentes grupos de edad en función de la distribución de los valores de estabilidad entre las variables, y luego representamos gráficamente aquellos ASV y subfunciones cuyo RSI es significativamente diferente entre dos grupos

de edad. Finalmente, aplicamos una tercera estrategia para estudiar si la estabilidad de la microbiota dentro de los diferentes grupos de edad cambia a lo largo del tiempo de muestreo. Con este fin, calculamos el índice de Jaccard para todos los pares de muestras consecutivas del mismo individuo, tanto a nivel de ASV como de TIGRFAM. También realizamos test U de Mann-Whitney entre y dentro de los grupos de edad para probar la significancia de las diferencias en el índice de Jaccard 1) entre comparaciones de puntos de tiempo consecutivos dentro del mismo grupo de edad (comparación dentro del grupo) y 2) entre grupos para la misma comparación de puntos de tiempo consecutivos (comparación entre grupos).

Para investigar cómo el cese de la alimentación con leche materna afecta la composición y función de la microbiota de los niños, hemos comparado las muestras de niños antes y después del destete. Las llamadas muestras “antes del destete” (N=27) incluyen las de los niños antes del final del proceso de destete, y las muestras “después del destete” (N=72) incluyen las de los niños después del final del proceso de destete, incluidos los niños que habían sido destetados antes del inicio del estudio. Con estas muestras, se realizaron análisis CCA y test de ADONIS sobre los datos de ASVs y

TIGRFAMs, así como análisis de diversidad alfa (índice de Shannon y estimador Chao1 sobre ASVs) y análisis LEfSe. También analizamos por separado las muestras de adolescentes antes y después de la menarquia, ya que los cambios en los niveles hormonales asociados a la menarquia también podrían afectar la composición y función de la microbiota intestinal. Las muestras "antes de la menarquia" (N=18) incluyen las de las adolescentes en los puntos temporales anteriores a la aparición de la menarquia durante el estudio y las muestras "después de la menarquia" (N=18) incluyen las de las mismas adolescentes después de la menarquia, así como las de las adolescentes cuya menarquia apareció antes de que comenzara el estudio. Se realizó un análisis CCA y test de ADONIS sobre los datos de ASVs y TIGRFAMs, así como análisis de diversidad alfa (índice de Shannon y estimadores Chao1 sobre ASVs) y análisis LEfSe. Estos análisis se realizaron con los datos agregados de todas las muestras “antes” y “después” del destete o de la menarquia de cada niño o adolescente utilizando la mediana de las abundancias relativas de cada variable.

Resultados

Extrajimos el ADN bacteriano de cada muestra y realizamos la secuenciación en MiSeq (Illumina) de la región V3-V4 del gen 16S del ARN ribosomal y del metagenoma, lo que produjo una secuenciación de gran profundidad (un total de 59.533.430 lecturas para secuenciación del gen 16S del ARN ribosomal: una media de 124.063 lecturas por muestra en niños, 122.020 lecturas por muestra en adolescentes y 117.555 lecturas por muestra en adultos; un total de 216.779.268 lecturas para el metagenoma: una media de 485.157 lecturas por muestra en niños, 411.977 lecturas por muestra en adolescentes y 425.492 lecturas por muestra en adultos.

En cuanto a la composición global de la microbiota, hemos visto que, aunque existe variación interindividual en la abundancia relativa de los diferentes filos, Firmicutes es el filo más abundante en todos los individuos en los tres grupos de edad, seguido de Bacteroidota y Actinobacteriota y, en menor medida, Proteobacteria y Verrucomicrobia. En términos de diversidad alfa calculada a nivel de ASVs, los niños tienen los valores más bajos para el índice de Shannon y el estimador de riqueza Chao1 durante todo el período de muestreo. Los valores para ambos índices de diversidad son en su mayoría significativamente más bajos que los de los adolescentes y adultos durante los

primeros puntos temporales, y luego aumentan hasta volverse similares a los de los otros grupos de edad. Adolescentes y adultos tienen valores de diversidad alfa similares a lo largo del período de muestreo. Se observó un cambio significativo en la diversidad de la microbiota de los niños entre T1 y T10 con una diferencia promedio de casi un punto en el valor del índice de Shannon, mientras que el valor de la diferencia en la diversidad entre estos puntos es cercano a cero tanto en adolescentes como en adultos. Los modelos de efectos mixtos lineales confirman además que la diversidad de Shannon aumenta con el tiempo en los niños, mientras que en adultos y adolescentes permanece en valores muy similares durante el período de muestreo.

Los análisis de correspondencia canónica (CCA) y test de ADONIS (PMANOVA) muestran que las muestras de niños se agrupan lejos de las de adolescentes y adultos en todos los casos, excepto en la comparación de "niños frente a adolescentes" a nivel de filo. El análisis de coordenadas principales (PCoA) revela un cambio direccional de las muestras de niños hacia la composición de la microbiota intestinal adulta, de modo que, hacia el final del período de muestreo, niños, adolescentes y adultos comparten el mismo espacio.

El perfil funcional del metagenoma, basado en la clasificación jerárquica de TIGRFAM, revela que la variación interindividual e intraindividual es menos pronunciada a este nivel en comparación con la composición taxonómica. Las funciones relacionadas con la categoría "proteínas transportadoras y de unión" son las más abundantes en todas las muestras, seguidas de las relacionadas con "síntesis de proteínas", "destino de proteínas", "metabolismo energético" y "metabolismo del ADN". El análisis CCA y ADONIS para los tres grupos de edad y para cada comparación dos a dos a nivel funcional muestra que las muestras de los diferentes grupos de edad se agrupan separadas entre sí en todos los casos, excepto en la comparación de adolescentes y adultos.

Las comparaciones dos a dos utilizando el algoritmo LEfSe muestran que dos filos están significativamente sobrerrepresentados en los niños en comparación con los adultos, Actinobacteriota y Proteobacteria, encontrándose en cantidades intermedias en adolescentes. *Bifidobacterium* no está sobrerrepresentada en niños en comparación con adolescentes, sino que disminuye significativamente entre adolescentes y adultos. *Oxalobacter* está sobrerrepresentada en niños en relación a adolescentes y adultos, lo que

indica que su abundancia relativa aumenta antes de la adolescencia, mientras que *Olsenella* está sobrerrepresentada en adultos en relación a niños y adolescentes, lo que indica que su abundancia aumenta después de este período. El CIM basado en los ASV seleccionados en los primeros dos componentes muestra que las muestras del mismo grupo de edad se agrupan juntas. Las afiliaciones taxonómicas de los ASV asociados con cada grupo de edad son acordes con los resultados de LEfSe y se obtuvieron resultados similares al realizar el mismo análisis a nivel de género.

A nivel funcional, el análisis LEfSe muestra un enriquecimiento en los niños de varios subroles pertenecientes a diversas categorías con respecto a los adolescentes y a los adultos. Estas funciones están relacionadas principalmente con el metabolismo energético, incluyendo funciones asociadas con la respiración aeróbica. Otros subroles están enriquecidos en los niños solo en relación a los adultos, lo que sugiere que decaen después de la adolescencia. Por el contrario, existen numerosos subroles que aumentan en la adolescencia y perduran en la edad adulta. El CIM basado en las anotaciones TIGRFAM seleccionadas en los primeros cuatro componentes obtenidos a través de sPLS-DA muestra que las muestras de niños y adultos

están representadas principalmente en grupos únicos, mientras que las muestras de adolescentes pueden ser encontradas en varias agrupaciones que también contienen muestras de los otros grupos de edad. Las funciones asociadas con cada grupo son acordes con los resultados de LEfSe.

Además de buscar biomarcadores taxonómicos y funcionales en los conjuntos de datos del amplicones del gen 16S del ARN ribosomal y metagenoma por separado, intentamos identificar las relaciones entre estos biomarcadores usando DIABLO, un método de clasificación integradora multivariante, que busca firmas moleculares que incluyen diferentes tipos de datos. El CIM basado en las firmas moleculares identificadas en los primeros cuatro componentes muestra que la mayoría de las muestras de cada grupo de edad están contenidas en dos o tres agrupaciones distintas, cada una de las cuales contiene principalmente muestras del mismo grupo. Una red de correlaciones construida con las variables seleccionadas más relevantes (tanto ASV como TIGRFAM) muestra tres subredes. La primera subred contiene seis TIGRFAM asociadas a niños, la segunda contiene cuatro TIGRFAM asociadas a adolescentes y la tercera incluye cinco TIGRFAM asociadas a niños. Estas TIGRFAM de la tercera subred se asocian a una sola ASV (*Bifidobacterium* NA)

mientras que las de las otras subredes se asocian con ASVs pertenecientes a varios géneros.

Al aplicar el método *complexCruncher* a los datos del gen 16S del ARN ribosomal, podemos ver que los niños son el grupo más inestable, mientras que los adolescentes y los adultos tienen distribuciones similares de V y β. El método también calculó el índice de estabilidad de rango (RSI) de todos los taxones de todas las muestras en el conjunto de datos. Los análisis CCA y ADONIS basados en los valores de RSI muestran que los tres grupos de edad se agrupan lejos unos de otros, siendo los adolescentes los que muestran la mayor dispersión entre los individuos. El mismo análisis con datos del metagenoma (a nivel de subrol) no muestra una diferencia general significativa entre los diferentes grupos de edad en términos de estabilidad, pero algunos subroles específicos se comportan de manera diferente en los diferentes grupos.

Al analizar la cantidad de cambio por unidad de tiempo a lo largo del período de muestreo mediante el cálculo del índice de Jaccard para todos los pares de muestras consecutivas de un mismo individuo, hemos visto que la estabilidad en los tres grupos de edad es menor a nivel de composición que a

nivel funcional. Tanto a nivel composicional como funcional, adolescentes y adultos presentan valores similares del índice de Jaccard durante todo el período de muestreo. Por el contrario, los niños tienen una estabilidad significativamente menor que los adolescentes y adultos durante el primer intervalo que aumenta en los siguientes puntos, indicando que tanto la composición taxonómica como la funcional de la microbiota se estabilizan durante la niñez. Al analizar las tendencias temporales de la variabilidad de taxones dentro de los grupos de edad con *q2-longitudinal*, hemos identificado tres ASV en los que una sobrerrepresentación inicial en niños desaparece durante el período de muestreo: ASV00031 (*Blautia obeum*), ASV00089 (*Streptococcus* NA) y ASV00046 (*Bifidobacterium bifidum*).

El análisis CCA y el test de ADONIS muestran que las muestras antes y después del destete se separan unas de otras a nivel de ASV y de anotación de TIGRFAM, y la diversidad aumenta después del destete. El análisis LEfSe a nivel de composición muestra que antes del destete hay una sobrerrepresentación de ASV que se anotan como *Bifidobacterium*, *Veillonella*, *Gemella sanguinis*, Clostridia UCG-014 y *Erysipelatoclostridium ramosum*. Después del destete, los ASV anotados como *Ralstonia*, *Bacteroides*, *Odoribacter splanchnicus*,

Blautia, Lachnospira pectinoschiza y Clostridia grupo vadinBB60 están sobrerrepresentados, así como trece ASVs en la familia Oscillospiraceae. A nivel funcional, las anotaciones de TIGRFAM sobrerrepresentadas antes del destete (excepto una) estaban sobrerrepresentadas en los niños en las comparaciones anteriores de los tres grupos de edad. Después del destete varias de las anotaciones de TIGRFAM sobrerrepresentadas están casi todas ellas sobrerrepresentadas en adolescentes y adultos en la comparación de los tres grupos. En conjunto, estos resultados indican que el destete contribuye a generar una microbiota más similar a la de adolescentes y adultos tanto a nivel de composición taxonómica como funcional.

Los análisis CCA y ADONIS del efecto de la menarquia en adolescentes revela que las muestras antes y después de este evento no se separan unas de otras ni a nivel taxonómico ni funcional. En términos de diversidad, tanto el índice de Shannon como el estimador de Chao1 aumentan significativamente después de la menarquia. El análisis LEfSe muestra que *Coprococcus catus*, *Clostridium sensu stricto 1* NA y *Veillonella* NA, todos ellos dentro del filo de Firmicutes, están sobrerrepresentados antes de la menarquia, mientras que *Butyricicoccus faecihominis*, *Adlercreutzia equolifaciens* y Desulfovibrionaceae

NA están sobrerrepresentados después de la menarquia. A nivel funcional, antes de la menarquia están sobrerrepresentados varios subroles dentro de la categoría “proteínas transportadoras y de unión” y el subrol “envoltura celular/biosíntesis y degradación de polisacáridos y lipopolisacáridos superficiales”. Un grupo más heterogéneo de funciones está sobrerrepresentado después de la menarquia.

Conclusiones

Aunque la microbiota intestinal se estabiliza al principio de la infancia en términos de cambios rápidos y a gran escala en la diversidad, la composición y el perfil funcional, se siguen produciendo cambios tanto antes como después del período de la adolescencia. Por lo tanto, esta etapa todavía puede considerarse un período de transición, en el que algunos taxones y funciones aún no han alcanzado la abundancia observada en los adultos, por lo que las etapas finales del desarrollo que ocurren durante este tiempo aún pueden ser importantes para el establecimiento de una microbiota adulta saludable. También hemos observado que el destete tiene un impacto sustancial en la diversidad, composición y perfil funcional de la microbiota de los niños que han pasado por un largo período de lactancia materna, lo que

resulta en una convergencia hacia las características de la microbiota adulta. Por último, nuestros resultados muestran que la menarquia viene acompañada de un aumento en la diversidad de la microbiota, así como de cambios moderados en la composición y función, aunque se necesitan más análisis para caracterizar mejor la dinámica de la microbiota en este período.

Abstract

The intestinal microbiota is deployed along the development of the individual and its composition and functions differ depending on age. Knowing about the changes that occur in the gut microbiota throughout life can help the diagnosis, treatment and prevention of diseases related to metabolic and immune alterations. We intend to contribute to the understanding of the process of intestinal microbiota development, focusing on two periods that have received little attention: the toddler years and adolescence. To this aim, we analyzed diversity, taxonomic composition, functional profile, and temporal stability in a cohort of 12 toddlers, 13 adolescents and 35 adults. The toddlers had undergone a long period of breastfeeding and most of them were weaned during the study, allowing us to evaluate the effect of this event. We determined the composition and function of the microbiota through 16S ribosomal RNA-based and whole genome shotgun metagenomics approaches. We established how the microbiota varies among the different age groups and compared the degree of microbiota stability at different ages. We observed that the differences between the gut microbiota of toddlers and that of adults

progressively subsided during childhood and adolescence, with cessation of breastfeeding providing an important contribution to the convergence of the microbiota towards adult-like properties. The microbiota of the toddlers stabilized early in terms of large-scale, rapid changes in diversity, composition, and functional profile, but further changes did occur both before and after adolescence. Therefore, adolescence can be considered a transitional period, in which some taxa and functions have yet to attain the abundances observed in adults, so that the final stages of development occurring during this time may still be important for the establishment of a healthy adult microbiota.

A mis padres

CONTENTS

List of Figures

List of Tables

"Science and everyday life cannot and should not be separated"

Rosalind Franklin

1 INTRODUCTION

1.1 HUMAN GUT MICROBIOME

1.1.1 Overview of the human gut microbiome

A diverse collection of microorganisms, known as microbiota (Neish, 2009), inhabits the human body coexisting with their host, colonizing all the surfaces of the human body: the skin, the eye, oral cavity, and the respiratory, urinary, reproductive, and gastrointestinal tracts (Wilson, 2004). In the latter is where a larger amount of these microorganisms is concentrated, with 70% of all the human microbiota present in the colon (Whitman *et al.*, 1998), due to the large surface of this organ (200 m^2) and the molecules from the host's diet, secretions and epithelial cells that can be found here and used as nutrients (Gebbers & Laissue, 1989; Ley *et al*., 2006). The microorganisms that comprise the human gut microbiota are bacteria, archaea, viruses, and unicellular eukaryotes. Most of these are bacteria, gathering 100 trillion microorganisms with a density in the large intestine estimated at 10^{11} per milliliter, which results in a ratio of human:bacterial cells close to 1:1 (Sender

et al., 2016). All these microorganisms encode over 3 million genes, 100 times more than the human genomic content (Gill *et al.*, 2006), which in turn produce thousands of metabolites. The set of these microbial genes is known as microbiome or microbial metagenome, and contributes essential functionalities to human physiology such as food digestion, nutrient absorption, protection against pathogens and host immunity regulation (Collado *et al*., 2013; Sekirov *et al*., 2010). In this sense, human beings can be considered as "superorganisms" or holobionts in which a relationship of symbiosis is established between the host and its microbiota (Bäckhed *et al.*, 2005; Gill *et al.*, 2006; Rosenberg & Zilber-Rosenberg, 2013). This relationship is based on a coevolution of the microbiome with our ancestral lineages, in which evolutionary adaptations to the changing environment (dietary changes, starvation, etc.) led to an increased fitness of the microorganisms in the gut and in turn to an evolved host immune system that prevents the colonization of potentially pathogenic microbes (Ochman *et al.*, 2010; Bosch, 2014).

1.1.2 Composition and variations of the gut microbiota

The gut microbiota is composed mainly by strict anaerobes, with an abundance two to three orders of magnitude greater than that of aerobe and facultative anaerobe bacteria (Gordon & Dubos, 1970; Savage, 1970). The number of bacterial species that live in the gut of an individual is estimated to be more than 1,000 (Qin *et al.*, 2010); however, it has been suggested that the collective human gut microbiota could be composed of more than 4,000 bacterial and archaeal species (Almeida *et al.*, 2021). Of the 146 bacterial and archaeal phyla known to date (Parks *et al.*, 2021) only two are predominant in the gut: Bacteroidota and Firmicutes, and to a lesser extent Proteobacteria, Actinobacteriota, Verrucomicrobia, Fusobacteria, Cyanobacteria, Synergistetes, Tenericutes, Saccharibacteria, and Spirochaetes (Almeida *et al.*, 2019). Despite this, there is a great variety of microbiotas, in terms of composition and function, between individuals and even in the same individual over time due to various factors that are developed below.

1.1.2.1 Variations in the same individual

1.1.2.1.1 Infancy as a key period in acquisition and development of the gut microbiota

The gut microbiota colonizes the gut during the first days after birth, and the way in which this process takes place can condition its subsequent development, even years later. During the first 24 hours of life, a wide variety of microorganisms enter the baby's intestine from the mother and the environment, most of them transitory (Conway, 1997). However, in the following days, colonization patterns begin to appear depending on gestational age of delivery, mode of delivery (vaginal or cesarean section) and the method of feeding (breast milk or milk-based formula).

1.1.2.1.1.1 Gestational age at delivery

Numerous studies have found that the microbiota of preterm infants (< 37 weeks of gestation) is different from that of full-term, the former presenting less diversity and an increase in potentially pathogenic bacteria from the Enterobacteriaceae family (Arboleya *et al.*, 2011). In addition, a decrease in beneficial strict anaerobic bacteria has been found in these children, such as *Bifidobacterium* spp., *Bacteroides* spp., and *Lactobacillus* spp.

(Butel *et al.*, 2007; Arboleya *et al.*, 2012, 2015). These microbial perturbations on the gut microbiota of the preterm newborns could be due to organ immaturity and the environmental conditions of the incubator, such as intubation, enteral feeding, antibiotic use, and hospital stay (LaRosa *et al.*, 2014; Arboleya *et al.*, 2015). It has been suggested that gestational age could have a greater long-term influence on the colonization and development of the intestinal microbiota than other factors such as the use of antibiotics during labor or the first weeks of life, the mode of delivery or the method of feeding (Brooks *et al.*, 2014). One of these long-term effects is the development of necrotizing enterocolitis, a severe intestinal disorder that is associated with preterm birth and dysbiosis of the gut microbiota (Wang *et al.*, 2009; Mai *et al.*, 2011).

1.1.2.1.1.2 Mode of delivery

During birth, the newborns acquire microorganisms from the new environment they have just arrived in. In a vaginal delivery, these microorganisms come from the vaginal and intestinal microbiotas of the mother, which come into contact with the baby in the birth canal. Analysis of the meconium of these infants (the fecal material that was accumulated in the

intestines of the fetus and is discharged at or near the time of birth) shows a correlation between the composition of the newborn gut microbiota and that of the mother's vagina, with an enrichment of bacteria of the *Lactobacillus*, *Prevotella* and *Sneathia* genera (Dominguez-Bello *et al.*, 2010), as well as bacteria of the *Bifidobacterium* and *Bacteroides* genera and other facultative anaerobic species such as *Escherichia coli, Staphylococcus* spp*., Bacteroides fragilis*, and *Streptococcus* spp. (Biasucci *et al*., 2008, 2010; Grönlund *et al*., 1999; Jakobsson *et al*., 2014; Pantoja-Feliciano *et al*., 2013; Penders et al., 2006a). On the contrary, children born by caesarean section acquire the intestinal microbiota from bacteria of the mother's skin and the hospital environment, with a low abundance of bacteria of the *Bacteroides*, *Bifidobacterium*, and *Lactobacillus* genera, as well as a lower overall diversity compared to vaginally delivered infants (Dominguez-Bello *et al.*, 2010; Azad *et al.*, 2013). These differences were found in children up to 7 years later (Salminen *et al.*, 2004), relating cesarean birth with an increased risk of suffering from asthma, arthritis, obesity, and inflammatory bowel disease in these children (Sevelsted *et al.*, 2015; Chen *et al.*, 2017). However, a more recent study has seen a partial restoration of the intestinal microbiota of these

children compared to that of children born by vaginal delivery (Dominguez-Bello *et al.*, 2016), so the long-term impact on health of cesarean birth is yet to be determined.

1.1.2.1.1.3 Methods of milk feeding

Another factor that significantly influences the composition of the intestinal microbiota in infants is the mode of feeding. Although both breast-fed and formula-fed infants generally show high levels of Actinobacteriota, in the former a greater and more diverse number of *Bifidobacterium* spp. are observed (Roger *et al.*, 2010). Breast milk contains a large number of galactooligosaccharides, which are fermented by *Bifidobacterium* spp. producing short chain fatty acids (SCFAs) that in turn produce a prebiotic effect by preventing the colonization of pathogenic bacteria (Marcobal *et al.*, 2010; Underwood *et al.*, 2015). In contrast, formula-fed infants have higher numbers of potentially pathogenic bacteria compared to breast-fed infants, such as *Clostridium difficile*, *Bacteroides* spp., and *Escherichia coli* (Azad *et al*., 2013; Fallani *et al*., 2010; Penders *et al*., 2006a). Besides, bacteria have been found in breast milk that possibly have their origin in the maternal gastrointestinal tract, with the entero-mammary pathway being a mechanism

of vertical transmission of the intestinal microbiota (Perez *et al.*, 2007; LaTuga *et al.*, 2014).

1.1.2.1.2 Age

From birth, the microbiota undergoes a progressive process of implementation and heterogenization until reaching typical adult characteristics. Diversity increases with age and the microbiota stabilizes. It had been proposed that this process ended at three years of age (Yatsunenko *et al.*, 2012) but recent studies indicate that this process could be extended in time, specifically until adolescence (Hopkins, 2001; Agans *et al.*, 2011; Hollister *et al.*, 2015; Nafarin *et al.*, 2019). Hormonal changes produced at this stage, like those that occur during pregnancy, could be the cause of important changes in the composition of the microbiota. The changes that occur in each life stage will be discussed in detail later.

1.1.2.1.3 Antibiotics

Antibiotic treatments modify the gut microbiota composition depending on the antibiotic class, dose, time of exposure, bacterial target, and mode of action, acting selectively and leading to an increase of certain species and a decrease in others, reducing the overall diversity (Pérez-Cobas

et al., 2013; Iizumi *et al.*, 2017). These shifts in the microbiota can be restored after the antibiotic treatment due to the resilience of the community, but in some cases a repeated treatment over time can lead to permanent and irreversible changes, although its impact depends largely on the conditions of the host, with great variability between individuals (Dethlefsen & Relman, 2011; Lozupone *et al.*, 2012).

1.1.2.2 Variations between individuals

In addition to the variations in the intestinal microbiota that can be found in the same individual over time, there are also variations between individuals caused mainly by host genetics, geographical location, social context, or dietary habits and other environmental influences.

1.1.2.2.1 Enterotypes

Each individual has an intestinal microbiota composed of a specific pattern of bacteria that can be grouped into types, which are called enterotypes (Arumugam *et al.*, 2011). These enterotypes are reproducible patterns that are useful for stratifying the human microbiome. Enterotypes are independent of age, gender, cultural background, and geography and represent networks of co-occurring microorganisms that are centered around

one “driver” taxon: enterotype I (driven by the genus *Bacteroides*), enterotype II (driven by *Prevotella*), and enterotype III (driven by *Ruminococcus*). Each of them is characterized by a different way of obtaining energy from the substrates available in the intestine, so they represent clusters of functions rather than a specific taxonomic composition: enterotype I ferment carbohydrates by glycolysis and the pentose phosphate pathway, whereas enterotypes II and III degrade glycoproteins. Enterotypes characterize individuals since they remain stable over time during adulthood and can be restored after a disturbance (Costea *et al.*, 2018).

1.1.2.2.2 Environment, lifestyle, and geographical location

It is well known that diet can rapidly alter the gut microbiota composition. In a study in which the intestinal microbiota of individuals with an animal-based diet (low fiber and high fat and protein) was analyzed (David *et al.*, 2013), comparing it with that of individuals with a plant-based diet (high fiber and low fat and protein), it was seen that in the former there is an increase in bile-tolerant bacteria, such as *Alistipes* spp., *Bilophila* spp., and *Bacteroides* spp., and a decrease in bacteria that metabolize plant polysaccharides (*Roseburia* spp., *Eubacterium rectale*, and *Ruminococcus*

bromii). In addition, when diet was changed from one type to another in the same individual, changes were observed in a matter of days, even changing the microbiota's manner of obtaining energy, with a tradeoff between carbohydrate and protein fermentation. Other studies that compare the gut microbiota of geographically remote populations and rural *versus* urban areas show the relevance of cultural and lifestyle factors that influence diet and shape the composition of the microbiota (De Filippo *et al.*, 2010; Arumugam *et al.*, 2011; Wu *et al.*, 2011; Tyakht *et al.*, 2013; Schnorr *et al.*, 2014; Pasolli *et al.*, 2019).

Social context is another factor that influences the composition of the microbiota. A study based on the Home Microbiome Project data has found that the gut microbiota is similar in individuals living in the same household, including pets (Lax *et al.*, 2014). Educational level and other social factors that are more difficult to study also influence the microbiota composition (De Mello *et al.*, 2009; Ding & Schloss, 2014).

1.1.3 Functions of a healthy gut microbiota

The gut microbiota plays an important role in the maintenance of the host health, being involved in various aspects of the physiology of the host.

1.1.3.1 Protection against pathogens

Commensal bacteria in the gut act in many ways to prevent the colonization or growth of pathogenic microorganisms. Their mere presence acts as a physical barrier that prevents the invasion of pathogens by competitive exclusion, through the occupation of attachment sites in the intestinal epithelium, consumption of nutrient sources and production of antimicrobial substances or by stimulating the production of these molecules by the host. The expression of these antimicrobial peptides can be induced by the presence of commensal bacteria or the products of their metabolism (Liévin-Le Moal & Servin, 2006; Salzman *et al.*, 2007). *Lactobacillus* and *Bifidobacterium* contain well-known probiotic strains that prevent the invasion of bacterial pathogens, such as *Listeria* spp. or *E. coli* (Corr *et al.*, 2007; Medellin-Peña & Griffiths, 2008).

1.1.3.2 Metabolism

As mentioned above, the gut microbiota encodes a broader and more versatile metabolome than that produced by the human genome. Most of the host-beneficial metabolic functions developed by the microbiome are involved in the improvement of nutrient bioavailability by the degradation of

nondigestible dietary compounds and modulation of absorptive capacity of the intestinal epithelium to facilitate nutrient uptake. The processing of the indigestible dietary polysaccharides implies their break down into absorbable monosaccharides, affecting energy harvest from the diet. Some bacteria are involved in the transformation of dietary fiber into short chain fatty acids (SCFAs), including acetate, propionate, and butyrate, which constitute an additional source of energy and prevent the accumulation of metabolic products potentially toxic to the host (Vella & Farrugia, 1998; Macfarlane & Macfarlane, 2003; Duncan *et al.*, 2004). In addition, the microbiota is involved in other beneficial metabolic functions for the host, such as fat storage (Bäckhed *et al.*, 2004), improvement of the absorption of calcium, magnesium, and phosphorus (Andrieux & Sacquet, 1983), and production and supply of vitamin K and some vitamin B (Hill, 1997). Moreover, it contributes to human protein homeostasis (Davila *et al.*, 2013).

1.1.3.3 Immunomodulation

Bacteria of the intestine also play an important role in the development of the host immune system and in regulation of innate and adaptive immunity (Isolauri *et al*., 2008) due to the fact that intestinal mucosa

is the largest area in contact with a wide variety of antigens of the external environment. The immunomodulatory capacity of the gut microbiota is given by the exposure of the structural components of the members of the community to the immune cells of the host. That is the case of the recognition of the polysaccharides A (PSA) of *Bacteroides fragilis* by dendritic cells (Mazmanian *et al.*, 2005), which induces the proliferation of CD4+ T cells, of the activation of natural killer (NK) cells by *Lactobacilli* spp. (Fink *et al.*, 2007), of the induction of the differentiation of inflammatory cells by members of the Bacteroidota phylum (Ivanov *et al.*, 2008), or of the induction of the maturation of B-cells triggered by peptidoglycans of Gram-negative bacteria (Bouskra *et al.*, 2008).

The intestinal mucosa system has two functions in the immunomodulation of the host: prevention of the excessive immune response to normal microbiota and control of the overgrowth of the bacterial population to avoid detrimental conditions. The promotion of a more tolerogenic state in the gastrointestinal tract can be achieved by three ways. The first one is the physical separation of bacteria and host cells by the mucus layer of the colon, whose bottom tier is mostly free of bacteria, whereas the

luminal part is in contact with the members of the microbiota (Johansson *et al.*, 2008). The second is by decreasing the proinflammatory potential of the microbiota, by means, for instance, of the dephosphorylation by the intestinal alkaline phosphatase (Iap) of the endotoxic component of the bacterial lipopolysaccharides (Bates *et al.*, 2007). Finally, members of the microbiota can modulate the localized host immune response towards tolerance, as is the case of *B. thetaiotaomicron*, which prevents the activation of proinflammatory transcription factors of the host (Kelly *et al.*, 2004), or of different *Lactobacillus* spp. and certain *E. coli* strains, which condition the dendritic cells of the intestine towards a more tolerogenic state in the gastrointestinal tract (Zeuthen *et al.*, 2008).

Different strategies are employed by the host to prevent overgrowth of the bacterial population in the gut. On the one hand, dendritic cells of the intestine, activated by the presence of some components of the intestinal microbiota, induce the production of secretory IgA (sIgA), which coats bacteria allowing the local control of their population (Macpherson & Uhr, 2004; Tsuji *et al.*, 2008; Yanagibashi *et al.*, 2014). On the other hand, sIgA, as it is covering

the microbiota, minimizes its contact with the cells of the mucosal system, reducing the immune response from the host (Peterson *et al.*, 2007).

1.1.3.4 Gut-brain axis

There is a bidirectional signaling between the gastrointestinal tract and the brain, which is known as "gut-brain axis" (Rhee *et al.*, 2009). Bacteria of the gut microbiota may send signals to the central nervous system through the afferent vagus nerve, which originates in the brain stem and innervates abdominal organs, where it collects visceral sensory information and sends it to the brain (Collins & Bercik, 2009). It has been found that the intestinal microbiota may be controlling the levels of some signaling molecules, such as neurotrophic factor, norepinephrine and tryptophan, in areas of the central nervous system related to mood and behavior, suggesting a role in pathophysiology of mood disorders (Forsythe *et al.*, 2010; Foster & McVey Neufeld, 2013). Gut bacteria may also play a key role in the pathophysiology of eating disorders, as they have been shown to regulate the levels of the autoantibodies that target appetite-regulating hormones (Fetissov *et al.*, 2008).

1.1.4 Gut microbiota and disease

1.1.4.1 Health or disease: a matter of microbial balance

Healthy intestinal bacterial communities can recover from a disturbance keeping stable in composition and function. This capacity is known as resilience and is associated with a high diversity of species and a functional redundancy of the ecosystem, resulting in a stable equilibrium in response to perturbations such as treatment with antibiotics, dietary life events, or infections (Holling, 1973; Faith *et al.*, 2013; Greenhalgh *et al.*, 2016). However, if the disturbance is strong enough and exceeds the self-regeneration threshold of the microbiota, it can lead to a loss of intestinal-host balance associated with functions of the microbiota that are detrimental to the host, a state known as dysbiosis. This condition of dysbiosis contributes to the manifestation of a disease that cannot be attributed to a single bacterial species, but to an imbalance of the community as a whole, and that is often related to a loss of diversity and low-grade spontaneous inflammation at the mucosal barrier (Singh *et al*., 2016). Frequently, the state of dysbiosis can itself be resilient to external disturbances, such as therapy (Tamboli *et al.*, 2004). At

least 50 diseases are known to be associated with dysbiosis of the gut microbiota, some of them are detailed below (Shin *et al.*, 2015).

1.1.4.2 Intestinal disorders

- **Irritable Bowel Syndrome (IBS)**

IBS is a chronic heterogeneous gastrointestinal disorder of multifactorial origin which presents with a recurrent abdominal pain on average at least one day a week in the last three months. It is one of the most common gastrointestinal disorders, but its pathophysiology is not well understood (Bhattarai *et al*., 2017). However, several investigations that have studied the microbiota of patients with IBS and compared it with healthy patients have found a correlation between this disease and alterations in their microbiota (Krogius-Kurikka *et al.*, 2009; Salonen *et al.*, 2010). Compared to healthy controls, IBS patients have an increase in the Firmicutes-Bacteroidota ratio, as well as an increase in *Veillonella*, *Streptococci*, and *Ruminococcus* within the Firmicutes, and *Enterobacteriaceae* spp. in the Proteobacteria, and a decrease in other Firmicutes (*Lactobacilli* and *Faecalibacterium*) and Actinobacteriota (*Bifidobacterium* and *Collinsella*). In addition, a general decrease in microbial

richness is involved in the inflammatory response, prompting some of the symptoms observed in this disease.

- **Inflammatory Bowel Disease (IBD)**

IBDs are idiopathic relapsing disorders characterized by chronic inflammation of the intestinal tract including ulcerative colitis (UC) and Crohn's disease (CRD), in which gut microbial imbalances contribute to disease severity (Frank *et al.*, 2007). Thus, a decrease in *Lachnospiraceae* and Bacteroidota microbial populations has been observed in IBD patients compared to healthy controls. It seems that there are some differences regarding the variation of the microbial population between UC and CRD patients, although the results of the studies carried out are not conclusive (Joossens *et al.*, 2011; Hansen *et al.*, 2012; Machiels *et al.*, 2014).

- **Colorectal cancer (CRC)**

CRC is the third most common cause of cancer mortality in the world, after breast and lung cancer, and its incidence is increasing rapidly (Jemal *et al.*, 2011). There are differences in gut microbiota structure patterns in CRC patients compared with healthy persons: a reduction of butyrate-producing

bacteria (such as *Roseburia* sp. and other members of the *Lachnospiraceae* family) and an increase in opportunistic pathogens (*Bacteroides fragilis*, *Enterococcus*, *Escherichia/Shigella*, *Klebsiella*, *Streptococcus*, *Peptostreptococcus*) that may contribute to the etiology of the disease (Wang *et al*., 2011).

- ***Clostridium difficile* infection**

C. difficile is a nosocomial pathogen that causes antibiotic-associated diarrhea (CDAD) due to toxin production that leads to cellular death and tissue damage (Kelly & LaMont, 2008; Britton & Young, 2014). It has been observed that an antibiotic-induced dysbiosis in the gastrointestinal tract in patients with CDAD favors the germination and growth of *C. difficile*. The imbalance of the microbial community during *C. difficile* infection is associated with a reduction in the population of bacterial species involved in resistance to colonization by this potential pathogen, rather than to a specific bacterial composition (Knecht *et al.*, 2014).

1.1.4.3 Metabolic disorders

- **Obesity and overweight**

Obesity is one of the major public health problems due to the high risk that children who suffer from this disorder in childhood have of developing diseases in adulthood such as autoimmune diseases, type 2 diabetes, and early metabolic and cardiovascular diseases. In addition to risk factors such as a diet rich in carbohydrates and fats and a sedentary lifestyle, numerous studies have linked obesity with changes in the composition of the gut microbiota, observing a lower species diversity and shifts in the abundance of genes involved in metabolism. Specifically, the microbiota of obese individuals has an increased capacity to harvest energy from the diet (Bäckhed *et al.*, 2004; Tsai & Coyle, 2009; Turnbaugh & Gordon, 2009). Other studies have confirmed the increase of Proteobacteria in obese individuals, which would lead to the production of pro-inflammatory molecules by these bacteria (such as lipopolysaccharides) and play a role in energy harvesting and increasing host fat storage (Rizzatti *et al.*, 2017; Bai *et al.*, 2019). All this indicates that the gut microbiota is a contributing factor to the pathophysiology of obesity.

- **Type 2 Diabetes (T2D)**

Several studies have shown that the composition of the intestinal microbiota is altered in patients with T2D compared with healthy people, with

a decrease in Clostridia and butyrate-producing bacteria (such as *Roseburia*), and an increase in the Bacteroidota-Firmicutes ratio, Betaproteobacteria, various opportunistic pathogens (*Clostridium* spp. and *Bacteroides caccae*), and bacteria that carry out sulfate reduction, such as *Desulfovibrionaceae* (Larsen *et al.*, 2010; Musso *et al.*, 2011; Qin *et al.*, 2012).

1.1.4.4 Atopic diseases

Atopic diseases are chronic inflammatory disorders that occur in susceptible individuals because of aberrant immune responses against common environmental antigens. Numerous studies have linked a disruption in the early development of the gut microbiota to the development of atopic diseases (including eczema, food allergy, and asthma), due to the key role played by the gut microbiota in modulating the immune system and in the induction of tolerance (Björkstén *et al*., 1999; Kalliomäki *et al.*, 2001; Murray *et al*., 2005; Penders *et al.*, 2006b, 2007). However, the specific bacterial species that produce beneficial or harmful effects on the development of these diseases are not clear, nor are the molecular mechanisms by which the microbial community of the gut affects the development of immunotolerance (Rautava *et al.*, 2004; Romagnani, 2004; Sjögren *et al.*, 2009).

1.1.4.5 Central Nervous System-related diseases

As mentioned above, the gut-brain axis connects the brain and the gut in a bidirectional manner, enabling communication between both. The brain can influence gastrointestinal and immune functions, so that a disruption of the intestinal microbiota could lead to disorders related to the central nervous system.

- **Alzheimer's and Parkinson's diseases**

Alzheimer's disease (AD) is associated with impaired cognition and cerebral accumulation of amyloid-beta peptides. In a study comparing the relative abundance of certain bacteria in the intestinal microbiota in AD patients who showed cognitive impairment and brain amyloidosis with patients who did not show these signs, an increase in the abundance of *E. coli* and a reduction in *E. rectale* in the first group was found, and these changes in the intestinal microbiota were associated with a peripheral inflammatory state (Cattaneo *et al.*, 2017). In another study in which the composition of the intestinal microbiota of patients diagnosed with AD was compared with patients without a diagnosis, a lower microbial diversity was seen in the former, with a decrease in Firmicutes and *Bifidobacterium* abundances and an

increase in *Bacteroides* (Vogt *et al.*, 2017). In contrast, an alteration of alpha diversity has not been observed in patients with Parkinson's disease (PD) compared to healthy controls (Hopfner *et al.*, 2017). However, beta diversity is significantly different for three bacterial families (*Lactobacillaceae*, *Barnesiellaceae*, and *Enterococcaceae*). These changes in the microbiota of PD patients may give rise to the neuroinflammation mechanism observed in the pathology of this disease.

- **Autism spectrum disorders (ASD)**

Several studies have found that the gut microbiota of children with ASD is less diverse and has reduced microbial populations of *Bifidobacterium*, *Akkermansia muciniphila*, and *Sutterella* compared with control subjects, and an increase of the genera *Lactobacillus*, *Clostridium*, *Bacteroides*, *Desulfovibrio*, *Caloramator*, and *Sarcina* (Finegold, 2011; Finegold *et al*., 2002, 2012; Li *et al.*, 2017; Wang *et al*., 2011; Williams *et al*., 2012). These studies demonstrate a close relationship between the intestinal microbiota and the clinical manifestations of autism spectrum disorders.

- **Mood disorders**

The gut microbiota also influences behavior and is associated with mental problems such as schizophrenia, anxiety, stress, or depression (Neufeld & Foster, 2009). A study in mice showed that exposure to stress causes changes in the composition of the intestinal microbiota, such as a decrease in the abundance of *Bacteroides* spp. and an increase in *Clostridium* spp. (Bailey *et al.*, 2011). Another study found that relative abundances of *Lactobacilli* and *Bifidobacterium* are decreased in patients diagnosed with major depression (Dowlati *et al.*, 2010). Indeed, chronic physiological stress alters gut microbiota composition, with a reduction of *Lactobacilli* spp. and an increase in *E. coli* and *Pseudomonas* spp., and depression due to chronic stress is associated with an increase in members of the Enterobacteriaceae family (Bhattarai *et al*., 2017).

1.2 DEVELOPMENT OF HUMAN GUT MICROBIOME THROUGHOUT LIFE

As has been mentioned previously, the gastrointestinal microbiome is dynamic over a human lifetime. Colonization may begin *in utero*, and then the immature microbiota at birth undergoes a progressive process of implementation and heterogenization until reaching typical adult characteristics in terms of composition, functionality, diversity, and stability.

1.2.1 Perinatal period

Initial colonization may begin before birth, since, by means of culture-independent molecular methods, bacteria have been found in placenta, amniotic fluid, fetal membranes, cord blood, and meconium under non-pathogenic conditions (Aagaard *et al*., 2014; Ardissone *et al*., 2014; Gosalbes *et al*., 2013; Jiménez *et al*., 2005, 2008; Rautava *et al*., 2012; Steel *et al*., 2005). This intrauterine microbiota is diverse and similar to the mother's oral microbiota, with taxa such as *Fusobacterium* spp. and *Bacteroides* spp. (Aagaard *et al*., 2014; Rautava *et al*., 2012), so it is speculated that the colonization route could be through small wounds in the oral cavity through which the bacteria would enter the bloodstream. From there they would colonize the placenta, being the initial inoculum of the intestinal microbiota of the fetus (Lockhart *et al.*, 2008; Zheng *et al.*, 2015).

At birth, the gut microbiome has a lower diversity and a higher variability among individuals compared to adults (Dominguez-Bello *et al.*, 2010; Yatsunenko *et al.*, 2012). The neonatal gut microbiome is dominated by Firmicutes, Actinobacteriota (especially the *Bifidobacterium* genus), and Proteobacteria, with lower abundance of Bacteroidota compared to adults

(Ardissone *et al*., 2014; Del Chierico *et al*., 2015; Gosalbes *et al*., 2013). As mentioned previously, gestational age at delivery, mode of birth, and antibiotic use, are key factors that shape gut microbiota colonization in the neonatal gut.

1.2.2 Infancy and childhood

During the first year of life, the gut microbiota fluctuates and undergoes a maturation process, with a progressive increase in bacterial diversity that is acquired through diet and the environment (Schanche *et al.*, 2015; Thompson *et al.*, 2015). While the children are nursing, their microbiome is enriched in functions related to digestion of oligosaccharides found in breast milk. As already mentioned, there are differences in the composition of the microbiota of infants fed with breast milk and those fed with formula milk, the latter exhibiting an increase in potentially pathogenic bacteria (Azad *et al*., 2013; Fallani *et al*., 2010; Penders *et al*., 2006). At the time of weaning and the introduction of solid foods in the diet, the gut microbiome is enriched in genes involved in the digestion of polysaccharides (Ravcheev *et al*., 2013; Vallès *et al*. 2014) and vitamin biosynthesis (Koenig *et al.*, 2011; Bäckhed *et al.*, 2015).

Before the development of the molecular culture-independent techniques it was thought that the gut microbiota reached maturity at 4 years of age (Ellis-Pegler *et al.*, 1975). However, recent studies have observed that there are still differences in the microbiota of children and pre-adolescents compared to that of adults in terms of composition and function, showing that there is an increase of Firmicutes, Proteobacteria, and Actinobacteriota in children compared to the adults' microbiota, and a decrease in Bacteroidota (Cheng *et al*., 2015; Hollister *et al*., 2015; Hopkins *et al*., 2001; Saulnier *et al*., 2011). The differences found between the microbiota of pre-adolescents and that of adults seem to be at a lower taxonomic level: an increase in *Roseburia*, *Faecalibacterium*, *Ruminococcus*, *Alistipes*, and *Bacteroides* species was observed in pre-adolescents, as well as an increase in functions involved in development, such as the synthesis of vitamin B_{12} (Hollister *et al*., 2015). The same study found that with age there was an increase in the functions of the microbiome related to inflammation and metabolic dysfunction.

1.2.3 Adolescence

There are only a few studies that compare the microbiota of adolescents with that of adults, and they have observed that at this stage

there are still differences in the composition of the gut microbiota. An increase in *Clostridium* spp. and *Bifidobacterium* spp. has been observed in the microbiota of adolescents compared to that of adults (Hopkins *et al.*, 2001; Agans *et al.*, 2011). This period could be key in the development of the microbiota since major hormonal changes take place.

1.2.4 Adulthood

The adult microbiota is the best known, since most studies in healthy individuals have focused on this stage. In general, the adult microbiota is mainly composed of bacteria in the Firmicutes, Bacteroidota and Proteobacteria phyla, although there are differences due to geographic location and cultural habits (Huttenhower *et al.*, 2012; Yatsunenko *et al.*, 2012; Schnorr *et al.*, 2014). At this point, the bacterial community of the gut remains relatively stable throughout adulthood, except for disturbances due to infections, antibiotic treatments, or drastic changes in diet or environment (David *et al.*, 2013). Despite this, there are large interindividual differences at various taxonomic levels. Large strain level differences suggest that the human gut microbiome is very individual-specific (Zhu *et al.*, 2015).

1.2.5 Elderly

Several studies have shown that the intestinal microbiota in the elderly is different from that of younger adults, exhibiting a higher Firmicutes to Bacteroidota ratio and a reduction in *Bifidobacterium* spp., *Bacteroides* spp., *Prevotella* spp., and *Faecalibacterium prausnitzii* (Mariat *et al.*, 2009), as well as an increase in *Enterobacteriaceae* (Cho & Blaser, 2012), which results in a detriment to the quality of life (Van Tongeren *et al.*, 2005). Factors such as increased use of medication, a poor diet, and hormonal changes produced during this stage could be the cause of the changes in the composition of the microbiota (Voreades *et al.*, 2014).

1.2.6 Gut microbiome during pregnancy

During pregnancy the body undergoes numerous changes at the hormonal, immunological and metabolic levels that could be associated with changes in the intestinal microbiota (Koren *et al*., 2012; Mor & Cardenas, 2010; Newbern & Freemark, 2011; Gosalbes *et al*. 2019). In a study in which the intestinal microbiota of women was characterized throughout pregnancy (Koren *et al.*, 2012), a profound change was observed in the third trimester, while during the first trimester the microbiota is similar to that of healthy non-

pregnant women and healthy men. These changes include an increase in Proteobacteria and Actinobacteriota, as well as a reduction in *Fecalibacterium prausnitzii*, a butyrate-producing bacterium with anti-inflammatory effects (Sokol *et al.*, 2008). These changes in the taxonomic composition of the microbiota induce dysbiosis, inflammation, and weight gain, features of the metabolic syndrome (Vijay-Kumar *et al.*, 2010), but which during pregnancy are beneficial for the correct development of the fetus (Mor and Cardenas, 2010).

1.3 TECHNIQUES AND METHODS TO STUDY THE GUT MICROBIOTA

A variety of techniques are used to study gut microbial communities, and they can be classified into two groups: culture-based and culture-independent methods.

1.3.1 Culture-based techniques

Classically, the composition of the gut microbiota has been analyzed using quantitative culture techniques, that use different media to select specific groups of microorganisms based on their metabolic requirements. These methods are cost-effective and reproducible, but their use has many

disadvantages (Wilson, 2004). First, the detection of bacteria at the species or strain level is practically impossible, since selection media are often unable to distinguish between different closely related phylogenetic groups of bacteria. Second, most gut bacteria are anaerobic (> 80%), making it difficult to provide the optimum environmental conditions to culture them (Eckburg *et al.*, 2005). Thirdly, due to the laboriousness of these techniques the number of errors is inevitably increased, and the number of samples that can be processed is reduced, resulting in a decrease of the statistical reliability of the results. Finally, identifying isolates is labor-intensive, since it requires the use of numerous morphological, physiological, and metabolic tests, and is often not very discriminatory. Despite these shortcomings, the use of these methods is still necessary in certain experimental designs as a complementary technique.

1.3.2 Culture-independent techniques

Since culture-based methods of microbiota analysis are insufficient, plenty of molecular-based techniques, which were traditionally used in microbial ecology to characterize marine and soil microbial communities, are being used in the human microbiome field.

1.3.2.1 "Fingerprinting" methods

- **Denaturing Gradient Gel Electrophoresis (DGGE)**

In this technique, PCR-amplified gene products extracted from the bacterial community samples migrate on an acrylamide gel according to their G+C content, creating a banding pattern that indicates the diversity of the sample (Zoetendal *et al.*, 2002).

- **Terminal Restriction Fragment Length Polymorphisms (TRFLP)**

In this case the TRFLP profiles are generated by the digestion of the full-length rRNA PCR amplicons with a restriction endonuclease giving rise to fragments that vary in length depending on the particular sequence of the 16S gene. The fragments migrate in a gel creating a distinct pattern for each sample (Li *et al*., 2007).

- **Ribosomal Intergenic Spacer Analysis (RISA)**

This method involves PCR amplification of the ribosomal intergenic spacer (RIS) region existing between the 16S and 23S rRNA genes in prokaryotes (equivalent to the ITS region in eukaryotes). The 5' primer is fluorescently labeled, so the fragments can be detected producing a fingerprint of the RIS

fragments. This technique could resolve differences between species or strains (García-Martínez *et al.*, 1999), but there is not yet an extensive database containing RIS taxonomy for gut microorganisms.

The "fingerprinting" methods have a high limit of detection because of the ability of the fragments to be resolved on a gel, so other molecular techniques have been developed for the screening of human gut microbial communities.

1.3.2.2 DNA microarrays

Microarrays are a high-throughput screening molecular technology that contains probes designed to detect specific members of the gut microbiota, detecting up to 775 phylospecies, even those for which DNA is only present at 0.00025% of the total community DNA (Paliy *et al.*, 2009). It is a cost-effective and less time-consuming alternative to 16S sequencing methods, but it has a high limit of detection and presents hybridization biases.

1.3.2.3 Fluorescence In Situ Hybridization (FISH)

FISH is another tool that can be used to screen the gut bacterial community, but using fewer probes designed to hybridize to 16S rRNA sequences unique to the targeted bacterial groups (at phylum, genus, or

species level). Currently a set of 15 probes can identify about 90% of the normal intestinal microbiota members (Harmsen *et al.*, 2002).

1.3.2.4 Quantitative PCR (qPCR)

Quantitative PCR can also be used to target specific bacterial groups in complex samples, but the design of specific probes is needed for the taxonomic group to be detected, so it is not possible to identify novel species or strains. Additionally, it requires a reference strain to generate the standard curve, which cannot be possible when there is no suitable culturable strain available.

1.3.2.5 Sequencing methods

The first approach based on sequencing methods used for the characterization of the intestinal microbiota was the Sanger sequencing of the full-length 16S rRNA gene (Eckburg *et al.*, 2005). The 16S rRNA has the advantages of being universally distributed among all prokaryotes, having diverse species-specific domains (hypervariable regions V1-V9), and being reliable for inferring phylogenetic relationships. For the characterization of these genes, they were first amplified by PCR using broad-range bacterial and

archaeal primers, and then the PCR products were cloned and sequenced bidirectionally.

The Sanger method is very expensive as the number of samples to be analyzed increases, so more cost-effective alternatives have been developed, known as Next Generation Sequencing (NGS) methods. Using NGS technology, a large number of 16S rRNA gene fragments can be sequenced simultaneously thanks to the higher throughput. These are obtained through the selective amplification of one or more of the hypervariable regions of the gene, yielding smaller sequence reads than those obtained by Sanger but that still provide good taxonomic resolution (Margulies *et al.*, 2005). In addition, it is a faster method since the creation of clone libraries is not necessary, and it can use barcoded primers that allow the sequencing of multiple samples at the same time in a single run (Hamady *et al.*, 2008). Several NGS platforms can be used to study the human microbiota (Table 1.1), using different technologies that have strengths and weaknesses that must be considered when designing an experiment.

Table 1.1. Summary of the main NGS platforms for the study of microbial communities (Hodkinson & Grice, 2015).

Platform	Clonal amplification	Chemistry	Average read length
454	Emulsion PCR	Pyrosequencing	700 bp*
Illumina	Bridge amplification	Reversible dye terminator	300 bp^{+}
SOLiD	Emulsion PCR	Oligonucleotide 8-mer chained ligation	75 bp*
Ion Torrent	Emulsion PCR	Proton detection	400 bp†
PacBio	N/A (single molecule)	Phospholinked fluorescent nucleotide	8,500 bp

*Paired-end sequencing available

$^{+}$Overlapping paired-end sequencing available

†Bidirectional sequencing available

1.3.2.6 Meta-omics approaches

In addition to the taxonomic composition and the number of bacteria in the intestinal microbial community, it is necessary to know the functional contribution that they make to the host to have a complete vision of the ecology of the gut microbiome. For this purpose, other high-throughput technologies have been developed and applied to measure multiple "omic" data types.

- **Metagenomics**

Shotgun metagenomic sequencing gives sequence information from the collective genomes of the microbiota without a specific primer (as is the case of the amplicon-based sequencing methods) thus removing biases from primer choices. This method provides information on the gene composition and the functional capacity of the microbial community in a single experiment, although it is costlier and more time-consuming than marker gene amplification.

- **Metatranscriptomics**

As in the case of metagenomics, metatranscriptomics is based on the high-throughput sequencing of nucleic acids isolated directly from gut microbial communities, but instead of using DNA involves the characterization of the RNA content, providing information about the actively expressed genes in the community. Furthermore, as in metagenomics approaches, metatranscriptomics generates structural information of the community, allowing the identification of the active members of the microbiota.

- **Metaproteomics**

Metaproteomics, or whole community proteomics, applies non-targeted shotgun mass spectrometry to assess the diversity and abundance of the proteins in the gut metaproteome, allowing the quantification of the microbial protein expression levels directly. Moreover, the protein separation steps of the technique can be altered to select those host proteins that interact with specific microbial proteins, providing information about the interactions between host and microbial components. However, proteins identified by metaproteomics differ significantly from the proteins predicted by metagenomics, suggesting that functional gene analysis does not correlate with gene expression levels (Verberkmoes *et al.*, 2009), so more comparisons must be done to confirm the accuracy of the technique.

- **Metabolomics**

Metabolomics implies the simultaneous analysis of multiple small metabolites present in a sample to generate a metabolic profile of the microbial population, and in this way study its function and systemic impact on host metabolism. Although a comprehensive overview of all the metabolites present in a biological sample is not yet possible due to its high

complexity, metabolomics is a powerful method for understanding the role of the microbiota in the physiology of the host.

1.4 INTERVENTIONS TO MODULATE THE GUT MICROBIOTA

Recent advances in the knowledge of the intestinal microbiota have allowed the development of intervention methods to increase the benefits it confers to health or decrease its ability to cause diseases. Antibiotics have been widely used for the control of disease-causing bacteria, but the increasing development of resistance to a wide range of antibiotics by some pathogenic bacteria has prompted interest in developing alternative methods.

1.4.1 Probiotics, prebiotics, synbiotics and postbiotics

- **Probiotics**

Based on the Food Agriculture Organization of the United Nations - World Health Organization (FAO-WHO) 2001 definition, probiotics are "live microorganisms, which when administered in adequate amounts, confer a health benefit on the host". They must meet three requirements: they should be fully characterized (both genetically and physiologically), they must remain viable during storage and during their passage through the gastrointestinal

tract, and their beneficial effects on health must have been demonstrated in human studies. Their beneficial effects may rely on the inhibition of pathogenic bacteria and their role in immune system development, synthesis of vitamins, and in the maintenance of the gastrointestinal barrier integrity (Collado *et al.*, 2009; Bermudez-Brito *et al.*, 2012; Hemarajata & Versalovic, 2013). Most of them are lactic acid bacteria, such as *Lactobacillus* spp. and *Bifidobacterium* spp., although their effects on health are strain and disease specific (O'Toole *et al.*, 2017; Sánchez *et al.*, 2017). However, clinical results for many probiotics are conflicting.

- **Prebiotics**

A consensus statement defines a prebiotic as "a substrate that is selectively utilized by host microorganisms conferring a health benefit" (Gibson *et al.*, 2017), so they are able to change the composition of the microbiota by stimulating the growth of bacteria that confer a benefit on the health of the host (Vyas & Ranganathan, 2012). Compounds that have been reported to have prebiotic effects include human milk oligosaccharides (HMOs) and other oligosaccharides, dietary fiber, phenolics and phytochemicals, conjugated linoleic acid, and polyunsaturated fatty acids.

These compounds generally enhance the growth and activity of *Bifidobacterium* spp., which produces beneficial effects on health such as improvement of the gastrointestinal barrier function, increase in mineral absorption, modulation of energy metabolism and satiety, and a reduction of the risk of intestinal infections (Roberfroid *et al.*, 2010; Sanders *et al.*, 2019).

- **Synbiotics**

Synbiotics are defined as "synergistic mixtures of probiotics and prebiotics that beneficially affect the host by improving the survival and colonization of live beneficial microorganisms in the intestinal tract of the host" (Food and Agriculture Organization of the United Nations World Health Organization, 2002). These compounds have been shown to have beneficial effects in infants with cow's milk allergy and in asthma derived from atopic dermatitis (Van Der Aa *et al.*, 2011; Burks *et al.*, 2015), as well as in constipation, lowering high fasting blood glucose levels, and in the risk of developing postoperative sepsis after gastro-intestinal surgery in adults (Arumugam *et al*., 2016; Miller *et al*., 2017; Nikbakht *et al*., 2018).

- **Postbiotics**

Postbiotics are functional bioactive compounds that are generated in a matrix during fermentation, including microbial cells, cell constituents, and metabolites, which used in combination with nutritional components may promote health (Wegh *et al.*, 2019). These compounds include short-chain fatty acids (SCFAs) and components like microbial fractions, functional proteins, secreted polysaccharides, extracellular polysaccharides (EPS), cell lysates, teichoic acid, peptidoglycan-derived muropeptides and pili-type structures (Konstantinov *et al.*, 2013; Slavin, 2013; Markowiak & Ślizewska, 2017; Sánchez *et al.*, 2017; Wegh *et al.*, 2017; O'Grady *et al.*, 2019). The advantage of postbiotics over probiotics is that they maintain the viability of microorganisms and their stability in the product, which facilitates the delivery of the active compounds at the intestine, improves shelf-life and simplifies packaging and transport (Ouwehand *et al.*, 2000).

However, there are still many questions about the mechanism of action and the health effects of prebiotics, probiotics, synbiotics, and postbiotics, so future clinical studies should include data not only on the taxonomic composition of the microbiota or measurements of certain compounds such as SCFAs, but also data obtained through metagenomics,

metatranscriptomics, metabolomics and metaproteomics that allow us to obtain a holistic vision of the functioning of the intestinal microbiota and its interaction with the host.

1.4.2 Fecal microbiota transplantation

Fecal microbiota transplantation (FMT) comprises the administration of a solution of fecal matter from a donor into the intestinal tract of a recipient individual for the benefit of their health status (Smits *et al.*, 2013). The donor must not have a family history of autoimmune, metabolic, and malignant diseases, and obviously blood and fecal samples must be screened for any potential pathogens. Fecal infusions are prepared by mixing with water or normal saline and filtered to remove particulate matter. Then, fecal solution is infused by several methods: through a nasogastric tube, nasojejunal tube, esophagogastroduodenoscopy, colonoscopy, or retention enema (Gough *et al.*, 2011). FMT has been shown to be effective in treating recurrent *Clostridium difficile* infections in patients who do not respond to conventional antibiotic treatments, proving to be a safe, well-tolerated, and effective treatment with few adverse events (Kelly *et al*., 2015). However, the mechanisms by which the FMT exerts the therapeutic effects have not been

yet elucidated, although it is most likely that it is the competitive exclusion of the pathogen by the microorganisms from the administered fecal solution which prevents the growth of *C. difficile* (Kelly *et al.*, 2015). Although several studies have found FMT relatively free of adverse effects (Brandt *et al.*, 2012), there exists the possibility of an unrecognized infectious disease transfer or the development of chronic disease (e.g., obesity, diabetes). Therefore, several research groups are exploring the use of synthetic stool solutions that would contain defined bacterial populations which would avoid these adverse effects (Petrof *et al.*, 2013).

2 OBJECTIVES

The intestinal microbiota is deployed along the development of the individual and its composition and functions differ depending on age. Knowing about the changes that occur in the gut microbiota throughout life can help the diagnosis, treatment and prevention of diseases related to metabolic and immune alterations. The combination of the different high-throughput meta-omics is key to understand the process of intestinal microbiota development and at what level and to what extent these changes occur. Taking this into consideration, the present thesis is aimed at achieving the following objectives:

- **Main objectives:**

1. Determine and compare the parameters that characterize healthy microbiotas in toddlers, adolescents, and adults.
2. Determine and compare the stability of the intestinal microbiota in toddlers, adolescents, and adults in terms of composition and/or function.

- **Secondary objectives:**

3. Determine if there are significant differences in the intestinal microbiota of toddlers before and after weaning in terms of composition and/or function.
4. Determine if there are significant differences in the intestinal microbiota of adolescents before and after menarche in terms of composition and/or function.

3 METHODS

3.1 COHORT AND STUDY DESIGN

This study was approved by the Ethics Committee of the Center for Public Health Research (CSISP), Valencia, Spain. All participants in the study read and signed forms of informed consent specifically approved for this project by the Ethics Committee. We recruited 60 healthy participants of 24 families from Valencia (Spain), divided into three age groups: adults (W = woman; M = man), adolescents (AA = female; AO = male) and toddlers (NA = female; NO = male). Adults (n = 35), who are the parents of the other two, were 41.6 years of age on average at the start of the study, adolescents (n = 13) were 11.9 years of age on average, and toddlers 20 months of age (n = 12). In the case of toddlers and adolescents, none of them are siblings except participants NA-6 and NO-6, who are twins. No participants had any gastrointestinal diseases, diabetes, obesity, or any other metabolic disease. In addition, none of them had taken antibiotics in at least three months before the start of the study. Fecal samples were collected every three to four months for three years approximately, with a total of 10 sampling time points

(metadata of each sample are available in Table S1 in the Appendix section). All samples were collected by the participants (or by the parents in the case of the toddlers) in sterile containers (10 ml) and stored in 10 ml of RNAlater solution (Invitrogen) at room temperature and later at -80ºC in the laboratory until processing. Participants were asked to fill a questionnaire with information related to diet, general health, intake of antibiotics, pre- and probiotics, breastfeeding, lifestyle, etc. In total, 496 samples were collected (119 from toddlers, 101 from adolescents and 276 from adults; Table 3.1), representing an 82.7% of the initial sampling schedule (Figure 3.1).

Table 3.1. Sample distribution by age group and time point, and total number of samples.

	toddlers	adolescents	adults	
T1	12	13	31	
T2	12	13	31	
T3	12	13	31	
T4	12	9	26	
T5	12	12	30	
T6	12	9	26	
T7	12	11	29	
T8	12	7	24	
T9	12	7	24	
T10	11	7	24	
Total	*119*	*101*	*276*	***496***

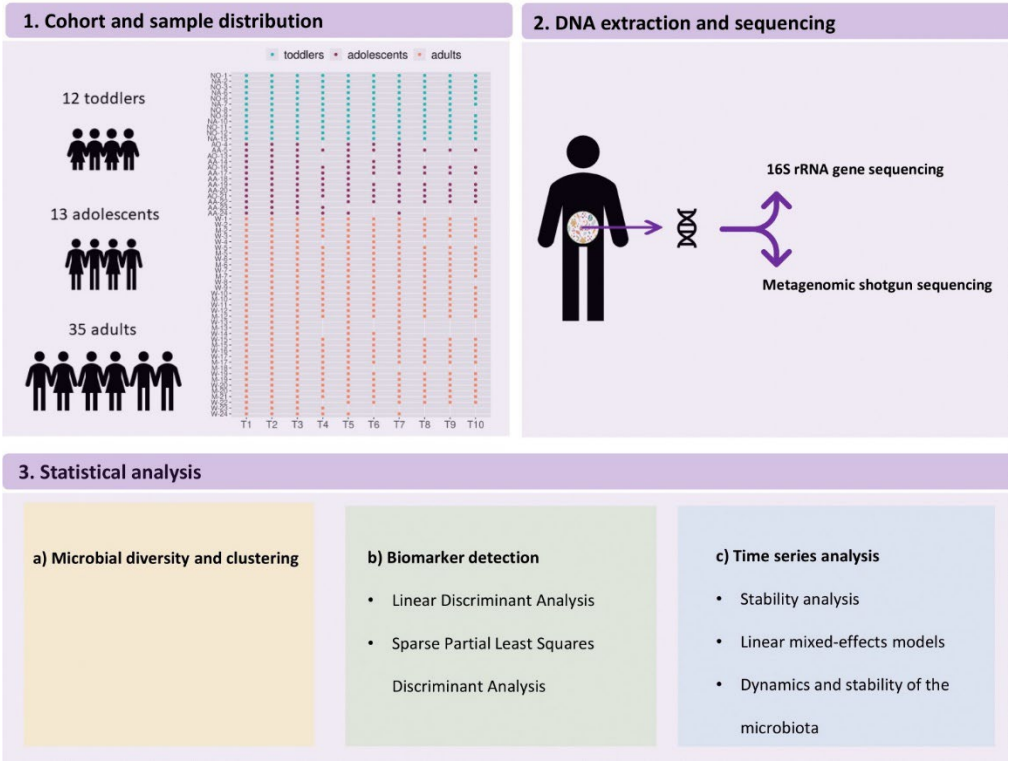


Figure 3.1. Methods summary.

3.2 BACTERIAL DNA EXTRACTION

Fecal samples were defrosted and resuspended in 10 ml of phosphate-buffered saline (PBS) (containing, per liter, 8 g of NaCl, 0.2 g of KCl, 1.44 g of Na_2HPO_4, and 0.24 g of KH_2PO_4 [pH 7.2]) and then centrifuged first at 2,000 rpm at 4ºC for 5 minutes to remove fecal debris and then at 13,000 rpm for 2 minutes to pellet bacterial cells. Total genomic DNA from bacteria was extracted using the MagNA Pure LC DNA isolation kit III for Bacteria and Fungi

using the MagNA Pure LC instrument (Roche) according to the instructions of the manufacturer, with addition of a digestion step at the beginning of the protocol using lysozyme for 30 minutes at 37ºC followed by cell lysing in 1 mm Zirconium bead tubes with 500 microliters of phenol:chloroform:isoamyl alcohol (25:24:1).

3.3 16S rRNA GENE SEQUENCING, READ PROCESSING AND ANNOTATION

Sequencing of samples was done in the MiSeq System (Illumina) and library preparation was done following the manufacturer's guidelines (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). The V3-V4 region of the 16S rRNA gene was amplified and sequenced with 2x300 bp paired-end reads (~120 samples per run yielding ~122,000 reads per sample on average; sequencing statistics are available in Table S2 in the Appendix section). Negative controls were performed at each step of the protocols. Processing, assembly and annotation of reads were done using the workflow of the DADA2 package (v1.8.0; Callahan *et al*., 2016) in the R Statistical Software (v1.8.0; R Core Team, 2018). Raw sequence data were filtered to remove low

quality reads and trimmed using the following parameters: maxN = 0, maxEE = c(5,5), truncQ = 0, rm.phix = TRUE and truncLEN = c(35,75). A sample inference algorithm was applied to the filtered and trimmed data, and the forward and reverse reads were merged with a minimum of 15 overlapping bases. An Amplicon Sequence Variant table (ASV) was then constructed. Chimeras and human sequences were removed, the latter through the alignment against the human genome (GRCh38.p13) using bowtie2 (v1.8.0; Langmead & Salzberg, 2012). Taxonomy was assigned by comparison against the SILVA Reference database (release 138; Quast *et al.*, 2013) in the DADA2 framework, with species-level assignment at 100% similarity or assignment at the deepest possible taxonomic level. BLAST v2.10.1 (Altschul *et al.*, 1997; Camacho *et al.*, 2009) was then employed to provide species-level assignations to ASVs that remained unassigned under this strict criterion, using a 97% similarity cutoff. Finally, a contingency table with the ASVs annotated at each taxonomy level and their absolute abundances per sample was obtained. Abundance data were normalized using the Total Sum Scaling method (TSS, dividing each feature count by the total library size, obtaining the relative proportion of counts for that feature).

3.4 METAGENOMIC SHOTGUN SEQUENCING, READ PROCESSING AND ANNOTATION

Shotgun metagenomic sequencing was done in the MiSeq System (Illumina) and library preparation was done following the manufacturer's guidelines using the 2x300 bp paired-end read protocol using the Nextera XT DNA kit (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-05.pdf) loading ~60 samples per run yielding ~437,000 reads per sample on average (sequencing statistics are available in Table S3 in the Appendix section). Negative controls were performed at each step of the protocols. Preprocessing of the reads, filtering, assembly, and annotation were done using an in-house pipeline that includes publicly available tools. Adapter sequences were removed with cutadapt (v2.9; Martin, 2011). Prinseq (v0.20.4; Schmieder & Edwards, 2011) was used to filter and trim the end or the reads by quality (parameters used: NS_MAX_P=1, MIN_LEN=50, MIN_QUAL_MEAN=30, TRIM_QUAL_RIGHT=30, TRIM_QUAL_TYPE=mean, TRIM_QUAL_WINDOW=10, TRIM_QUAL_STEP=10, LC_THRESHOLD=70,

REMOVE_LOW_ENTROPY_READS=TRUE). The joining of the overlapping pairs was done with FLASH (v0.20.4; Magoč & Salzberg, 2011) with a minimum overlap region of 12 nucleotides and a maximum of 65 nucleotides, and a maximum average value of mismatches of 0.1. Host and ribosomal RNA sequences were removed using bowtie2 (v2.3.5.1; Langmead & Salzberg, 2012) against the human genome (GRCh38.p13) and the SILVA Reference database release 138 (v2.3.5.1; Quast *et al*., 2013) respectively. Bacterial reads were then assembled into contigs using megahit (v2.3.5.1; Li *et al*., 2015) with a minimum percentage of identity of 95% and a minimum coverage of 80%, and then reads were mapped against the contigs. Codifying regions inside contigs were predicted with Prodigal (v2.6.3; Hyatt *et al*., 2010) with default parameters. For the functional annotation of the predicted open reading frames (ORFs), HMMER (v2.6.3; Finn *et al.*, 2011) was used to map these ORFs against the protein models in the TIGRFAM database (v2.6.3; Haft *et al*., 2001), setting the significance threshold or E-value at 0.001 for hit matches. A contingency table of TIGRFAM annotations and their absolute abundances per sample was generated and normalized using TSS. Taxonomical annotation was performed by mapping the reads against the

NCBI non-redundant protein database using Kaiju (v2.6.3; Menzel *et al*., 2016), with the following parameters: MIN_MATCH_LENGTH=20, MAX_MISMATCHES=5, MAX_EVALUE=1e-06. Furthermore, we developed an in-house pipeline that infers taxonomy for each ORF found in the data set, so we have been able to quantify those taxa that contribute genetically to the protein functions encoded in the metagenome.

3.5 Comparison of microbiota diversity, taxonomic composition and functional profile

Compositional and functional profiles were summarized by age group by plotting the average relative abundances of genera or of TIGRFAM subroles per time point. Alpha diversity (Shannon index and richness or Chao1 estimator) was calculated at ASV level for 16S rRNA gene amplicon sequencing data using the vegan R package (v2.5.6; Oksanen *et al*., 2019). The Shannon index was used because it uses the natural logarithms of relative species abundances, which reduces the weight of the more abundant species thus taking into account the changes in the abundance of rare species (Shannon, 1948), while the Chao1 estimators has been shown to be one of the most

reliable non-parametric estimators of species richness in species-rich samples (Chao, 1984). We also applied the Mann-Whitney U test to check for significant differences (p-value < 0.05) between different time points for each group and between groups for each time point for both alpha diversity estimators. Canonical Correspondence Analysis (CCA) and the ADONIS test were performed on compositional and functional data to display the variation in the data that can be explained by the age group. For this analysis, data previously aggregated by individual were used. The aggregation was performed by adding the abundances of each feature (ASVs and TIGRFAMs) of the samples from all time points for each individual and then the median of the relative abundances of each one was calculated. Additionally, Principal Coordinates Analysis (PCoA) was performed on the Jaccard index distances based on ASVs relative abundances from time series data, incorporating time as the third axis in the plot (in addition to principal coordinate 1 and principal coordinate 2) to explore how samples change over time.

3.6 BIOMARKER DETECTION

3.6.1 Linear discriminant analysis

To identify taxonomic and functional biomarkers from the different age groups during the sampling period we applied the linear discriminant analysis (LDA) effect size (LEfSe) algorithm (Segata *et al.*, 2011) on genus, species, and functional subrole data, previously aggregated as the same way as for CCA and ADONIS tests, in order to detect those features exclusively characteristic of the age group. The algorithm combines Kruskal-Wallis and pairwise Wilcoxon rank-sum tests for statistical significance and feature selection, using default parameters for significance (p-value < 0.05) and linear discriminant analysis threshold (> 2.0). Data were uploaded to the Galaxy web platform, and the public server was used to apply the algorithm (Afgan *et al.*, 2018).

3.6.2 Sparse Partial Least Squares discriminant analysis

3.6.2.1 Single 'omics

Analysis by the sparse Partial Least Squares-discriminant analysis (sPLS-DA), implemented in the mixOmics R package (v6.6.2; Rohart *et al.*,

2017), was also performed on ASV and TIGRFAM annotation data with the purpose of discriminating sample groups and identifying the most discriminant subset of features in the data set. Previously, data was filtered by removing low features (those with less than 10 counts), zero and near zero counts (features with less than 0.01% presence in the data set), as well as normalized using the Total Sum Scaling method to remove technical bias due to differences in sequencing depth by dividing each feature count with the total library size, yielding the relative proportion of counts. In addition, samples were previously aggregated by donor, as the sPLS-DA method is not intended for time series data. Performance and parameter tuning were carried out using 5-fold cross validation repeated 10 and 100 times for performance and tuning steps respectively. In this last step we set a maximum of ncomp = 6 for metagenome data and ncomp = 2 for 16S rRNA gene amplicons data, as suggested from the performance assessment. The maximal distance was specified for both datasets to predict class membership across all cross-validation runs. The final sPLS-DA model was run setting default parameters, except for a log ratio transformation (logratio = “CLR”) to deal with compositional data. To avoid divisions by zero in this step, a *pseudocount*

matrix was calculated by adding the smallest non-zero value to the normalized and aggregated count matrix. Clustered Image Maps (CIM) and relevance networks were constructed to visualize the results of the analysis. In CIMs (or heatmaps) a hierarchical clustering operates simultaneously on the rows and columns of the matrix, which is colored based on its value (Euclidean distance between subsets of variables). Rows and columns are reordered according to a hierarchical clustering, and dendrograms resulting of this clustering are added to the left side and to the top of the image. Relevance networks represent the correlation structure between variables. The function calculates a pair-wise similarity matrix directly obtained from the latent components of the integrative approaches. The similarity value between a pair of variables is obtained by calculating the sum of the correlations between the original variables and each of the latent components of the model. For every estimated correlation coefficient (which can be seen as a robust approximation of the Pearson correlation) exceeding a prespecified threshold between two variables (in absolute value) an edge is drawn between these two variables (nodes). The advantage of relevance networks is their ability to simultaneously represent positive and negative correlations, which are missed

by methods based on Euclidean distances or mutual information. Those networks are bipartite and thus only a link between two variables of different types can be represented. Networks were saved in gml format (Graph Markup Language) using the igraph R package (v1.2.4.1; Csardi & Nepusz, 2006) and input to Cytoscape (v3.8.2; Shannon *et al*., 2003) for representation.

3.6.2.2 Integrative 'omics

For the multiple 'omics analysis the DIABLO framework (v6.6.2; Rohart *et al*., 2017; Singh *et al*., 2016) was employed using the same data treated in the same way (ASV and TIGRFAM aggregated *pseudocounts* matrices as input). DIABLO performs the integration of metagenome and 16S rRNA gene amplicons datasets measured on the same N biological samples (N-integration) in a supervised context. It identifies a signature composed of highly correlated features across the different types of 'omics, by modelling relationships between the 'omics data sets. In the parameter choice step, we chose a design (matrix) where all the blocks (data sets) are connected with a link of 0.1. Then, we fit a DIABLO model without variable selection to assess the global performance and chose the number of components for the final DIABLO model. This step was run with 10-fold cross-validation repeated 100

times. The optimal number of components was 4 for the final DIABLO model. The tuning step identified a multi 'omics signature of 7, 30, 30 and 12 ASVs and 30, 5, 5 and 5 TIGRFAM annotations on the first 4 components. Relevance networks of the selected features and Clustered Image Maps were constructed to visualize DIABLO results. Furthermore, a *Circos* plot was constructed, which displays the different types of selected features on a circle, with links between 'omics indicating strong positive or negative correlations ($r > 0.7$) and relative abundance levels of each variable according to each age group.

3.7 TIME SERIES ANALYSIS

From the time series data, we wanted to see how the microbiota changes over time in each of the age groups and quantify that change, so that we can make comparisons within and between groups. To do this, we employed a variety of strategies described below.

3.7.1 Analyses of diversity and taxonomic composition dynamics with q2-longitudinal

We used the QIIME2 v2020 microbiome analysis platform (Bolyen *et al.*, 2019) to collect information on the change and stability of the microbial community in different age groups from 16S rRNA gene amplicon data from our longitudinal sampling. q2-longitudinal (Bokulich *et al.*, 2018) is a plugin for QIIME2 that incorporates multiple methods including interactive plotting, linear mixed-effects models, paired differences and distances, longitudinal feature selection and volatility analysis, among others. With this tool we tested the difference in alpha diversity values (Shannon index) between the first time point (T1) and the last one (T10) within the same group (Wilcoxon-signed rank test), as well as whether these paired T1-T10 differences were significantly different between age groups (Mann-Whitney U test). Linear Mixed Effects models were also used to test the changes in alpha diversity (Shannon index) over time and in response to the different age groups. Finally, machine learning regression (random forests) was employed to explore how microbial communities change in the different age groups, by learning the structure of the data and identifying any taxa (including low-abundance ones)

that are predictive of the different stages or time points. Important features are those that change over time and their abundance is predictive of the specific time point when a sample is collected, although this feature importance does not imply statistical importance, as this is only an exploratory method. The longitudinal abundance of each feature was plotted using volatility plots.

3.7.2 Stability of the microbiota

In order to study the overall stability in composition and function of the microbiota within children, adolescents and adults, we have applied *complexCruncher* (Martí *et al.*, 2017). This method is based on the observed fact that the dynamics of the microbiota fit well with Taylor's power law, which means that there is a clear correspondence between the relative mean and the dispersion of the taxonomic groups. The parameters of the Taylor's law are the variability of taxa (V), that is a direct estimator of fluctuations over time, and the scale index (β) of the power law, that provides information about the statistical properties of the ecosystem, and both correlate with the stability of the microbiota. V represents the maximum variability attainable by a hypothetically dominant taxa (with relative abundance close to 1). If V is

small, the microbial community is stable; if it is large, the microbiota might be unstable or resilient. Further, if the scaling index β is 1/2, the system behaves like a Poisson distribution, and if it is 1, the system behaves like an exponential distribution. However, metagenomes, in general, undergo time course variations with β between these two universal classes, so the fact that the scaling index is always smaller than 1 means that the most abundant taxa are less volatile than the less abundant ones. The pipeline calculates these parameters both at compositional (ASVs) and functional (TIGRFAM subroles) levels and constructs the Taylor's law parameter space, with the variability of the features (V) in the x-axis and the scale index (β) in the y-axis, both in standard deviation units. The "stability" zone was established taking the group of adults as reference; areas corresponding to the regions containing 68% and 98% of the adult subjects in the Taylor parameter space were delimited. *ComplexCruncher* also calculates the rank stability index (RSI), a way to quantify the stability of each of the features (ASVs and subroles) over time in each age group. The RSI is calculated, per feature, as 1 minus the quotient of the number of true rank hops taken divided by the number of maximum possible rank hops, all to the p power:

$$RSI = \left(1 - \frac{D}{(N-1)(t-1)}\right)p$$

where *D* is the total number of rank hops taken by the element studied, *N* is the number of elements that have been ranked, and *t* is the number of time samples. RSI is strictly 1 for a feature whose range never changes over time and strictly 0 for an element whose range oscillates between the extremes. We performed a Canonical Correspondence Analysis (CCA) and an ADONIS test based on RSIs to check if we can separate the microbial communities of the different age groups based on the distribution of stability values among features, and then plot those ASVs and subroles whose RSI is significantly different between two age groups (Wilcoxon signed-rank test p-value < 0.1).

Finally, we applied another strategy in order to study whether microbiota stability within the different age groups changes throughout the sampling time span. To this aim, we calculated the Jaccard index for all pairs of consecutive samples from the same subject, both at ASV and TIGRFAM levels. The Jaccard index is a beta-diversity metric that indicates the similarity between two samples or data sets by computing the fraction of shared features between them, and it ranges from 0 to 1: the closer to 1, the more

similar the two samples. We also performed Mann-Whitney U tests both between and within age groups to test the significance of the differences in the Jaccard index 1) between comparisons of consecutive time points within the same age group (within-group comparison) and 2) between groups for the same consecutive time points comparison (between-groups comparison).

3.8 WEANING ANALYSIS

To investigate how the cessation of breast milk feeding impacts the composition and function of toddlers' microbiota we have compared the samples of toddlers before and after weaning. The so-called "before weaning" samples (N = 27) include those of toddlers before the end of the weaning process, and the "after weaning" samples (N = 72) include those of toddlers after the end of the weaning process, including those of toddlers who had been weaned before the start of the study (Table S1 in the Appendix section). Using these samples, Canonical Correspondence Analysis and Adonis test were performed on ASVs and TIGRFAMs data, as well as alpha-diversity analysis (Shannon index and Chao1 estimators on ASVs) and LefSe analysis using default parameters. These analyses were performed with the aggregated data of all the samples "before" and "after" weaning for each

toddler, and then the median of the relative abundances of each feature was calculated.

3.9 MENARCHE ANALYSIS

We analyzed separately the samples of female adolescents before and after the menarche event, since changes in hormone levels associated to menarche could also affect the composition and function of the gut microbiota. Samples “before menarche” (N = 18) include those from girls who had menarche during the study and samples “after menarche” (N = 18) include those from the same girls after menarche as well as those from girls whose menarche appeared before the study started (Table S1 in the Appendix section). Canonical Correspondence Analysis and Adonis test were performed on ASVs and TIGRFAMs data, as well as alpha-diversity analysis (Shannon index and Chao1 estimators on ASVs) and LefSe analysis using default parameters. These analyses were performed with the aggregated data of all the samples “before” and “after” menarche for each female adolescent, and then the median of the relative abundances of each feature was calculated.

4 RESULTS

4.1 GUT MICROBIOTA 16S RRNA GENE AND METAGENOMIC SEQUENCING IN TODDLERS, ADOLESCENTS, AND ADULTS

To study the process of intestinal microbiota development, we collected stool samples from 60 Spanish volunteers recruited based on two exclusion criteria: suffering any gastrointestinal or metabolic disease and having taken antibiotics in the three months preceding the start of the study. The initial cohort was comprised of 12 toddlers, 13 adolescents, and 35 adults distributed over a total of 24 families, since adults are the parents of the toddlers and adolescents (Table S1 in the Appendix section). Ten stool samples, self-collected or collected by the parents, were obtained from each participant, every 3-4 months. Compliance with the sample collection schedule was high, so that we obtained a total of 496 samples over the 10 timepoints (119 samples from toddlers, 101 samples from adolescents and 276 samples from adults). To analyze the composition and function of the microbial communities in this cohort, we extracted bacterial DNA from each

sample and performed Illumina sequencing of the V3-V4 region of the 16S rRNA gene and shotgun metagenomics, using paired-end, partially overlapping reads on the MiSeq platform, yielding a high depth of sequencing (a total of 59,533,430 reads for 16S rRNA gene sequencing: mean of 124,063 reads per sample in toddlers, 122,020 reads per sample in adolescents and 117,555 reads per sample in adults; and a total of 216,779,268 reads for shotgun metagenomics: mean of 485,157 reads per sample in toddlers, 411,977 reads per sample in adolescents and 425,492 reads per sample in adults; Tables S2-S3 in the Appendix section, respectively). To avoid confounding factors, samples from women who became pregnant during sampling time (subjects W-7, W-8, W-11, and W-12) were not included in subsequent analyses.

4.2 GLOBAL COMPOSITION AND DIVERSITY IN THE GUT MICROBIOTA AT DIFFERENT AGES

Although there is interindividual variation in the relative abundance of different phyla, Firmicutes is the most abundant phylum in all subjects in the

three age groups, followed by Bacteroidota and Actinobacteriota and, to a lesser extent, Proteobacteria and Verrucomicrobia (Figure 4.1).

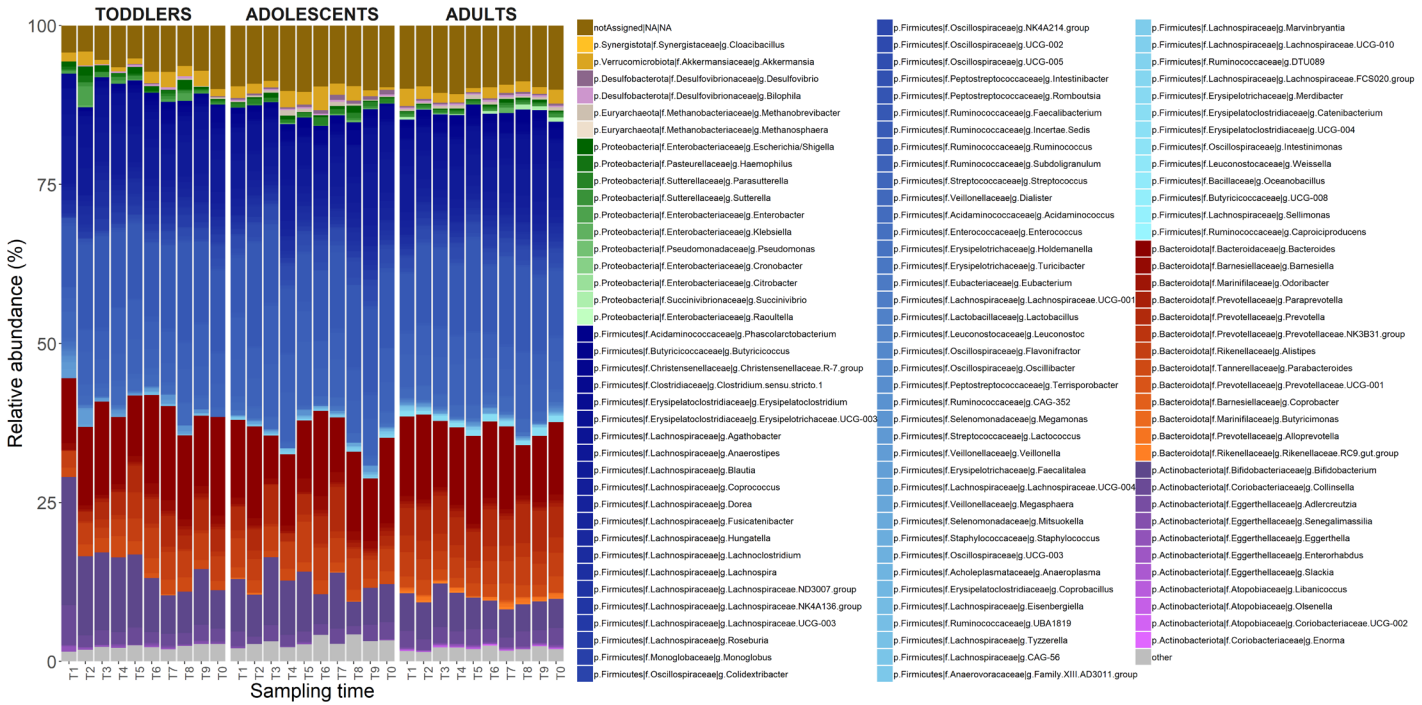


Figure 4.1. Microbial community composition in fecal samples of toddlers, adolescents, and adults. Relative abundances of bacterial genera (y-axis). Each bar represents the mean relative abundance of all individuals from the same age group for each sampling time point (x-axis).

In terms of alpha diversity computed at the level of Amplicon Sequence Variants (ASVs), toddlers have the lowest values for the Shannon index and the Chao1 richness estimator throughout the sampling period. Values for both diversity indexes are mostly significantly lower than those of adolescents and adults during the first timepoints (Figure 4.2a; Mann-Whitney U test; p-value < 0,05. All p-values available in Table S4 in the Appendix section), and then increase until becoming similar to those of the other age groups by T3 (28 months of age in average). Adolescents and adults have similar alpha diversity values throughout the sampling period (Mann-Whitney U test; p-value > 0,05). For the Shannon index, we tested whether diversity values change significantly between the first and the last sampling time points in the different groups (Figure 4.2b). There is a significant change in the microbiota diversity of toddlers between T1 and T10 (Wilcoxon signed-rank test; FDR p-value = 0.00293), with an average difference of almost one point in the Shannon index value, whereas the difference in diversity between these points is close to 0 in both adolescents and adults. Accordingly, the T10-T1 difference is significantly higher in toddlers compared to the other two groups (Mann-Whitney U test; FDR p-value = 0.00121 for "toddlers *versus*

adolescents" comparison and FDR p-value = 0.00040 for "toddlers *versus* adults" comparison). Linear Mixed Effects models further confirm that Shannon diversity increases with time in toddlers, while in adults and adolescents it remains at very similar values during the sampling period (Figure 4.2c).

A

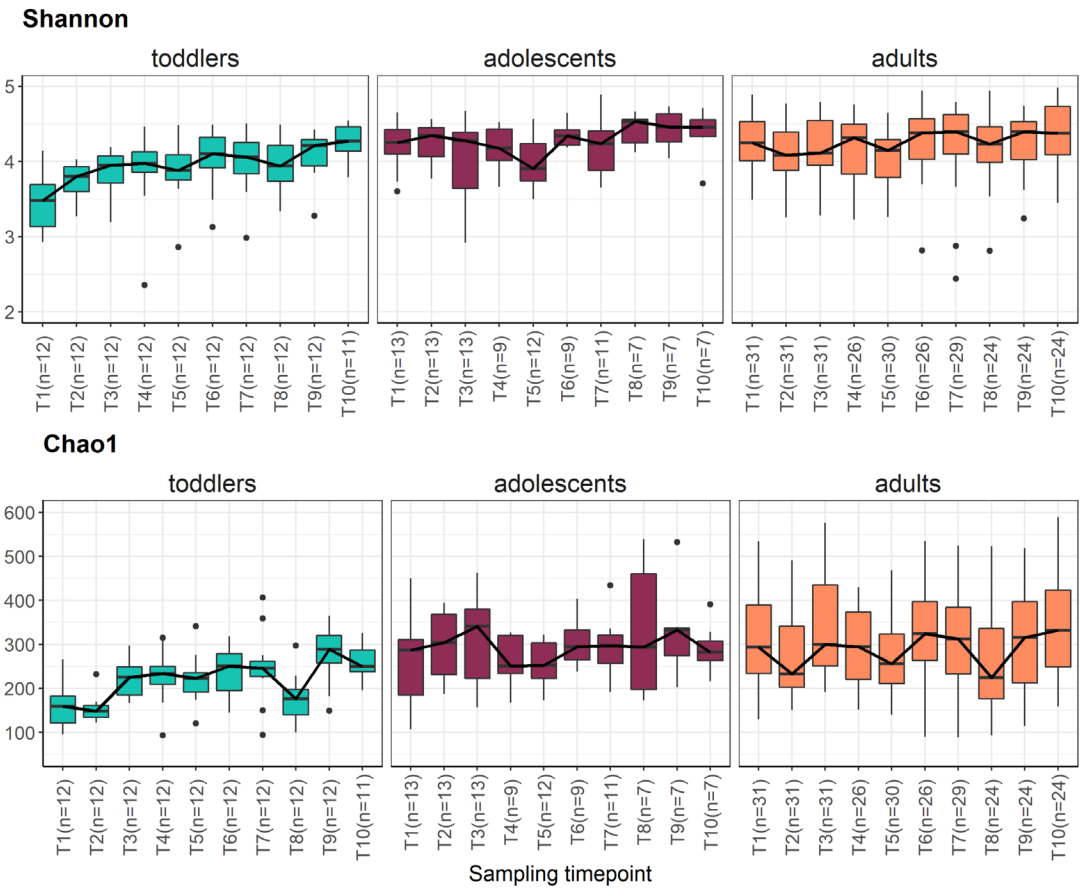
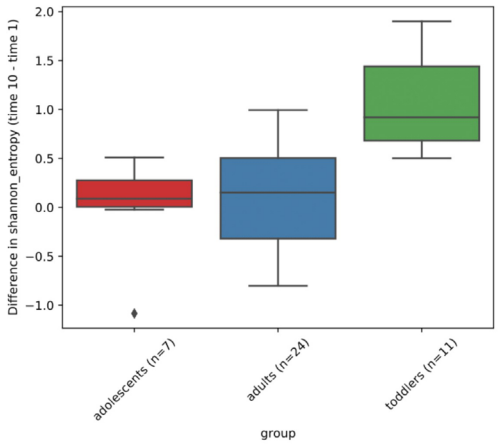


Figure continues on next page

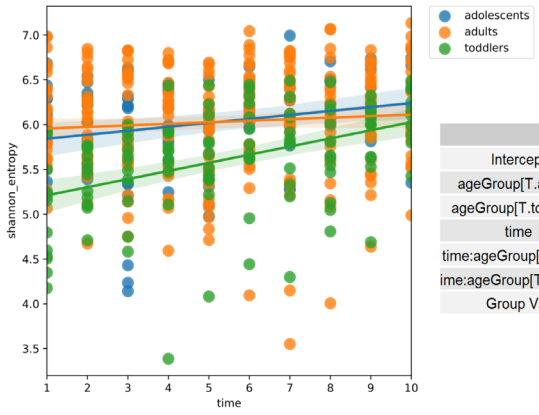
B



Group	W (wilcoxon signed-rank test)	P – value	FDR P-value
toddlers	0	0.00098*	0.00293*
adolescents	8	0.37500	0.37500
adults	118	0.37475	0.37500

Group A	Group B	Mann-Whitney U	P – value	FDR P-value
adolescents	adults	79	0.83164	0.83164
adolescents	toddlers	1	0.00081*	0.00121*
adults	toddlers	24	0.00013*	0.00040*

C



	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	5.862	0.142	41.269	0.000	5.584	6.140
ageGroup[T.adults]	0.085	0.169	0.505	0.613	-0.245	0.416
ageGroup[T.toddlers]	-0.747	0.203	-3.681	0.000	-1.145	-0.349
time	0.021	0.015	1.394	0.163	-0.009	0.051
time:ageGroup[T.adults]	-0.010	0.018	-0.581	0.561	-0.045	0.024
ime:ageGroup[T.toddlers]	0.071	0.020	3.535	0.000	0.032	0.110
Group Var	0.175	0.099	NA	NA	NA	NA

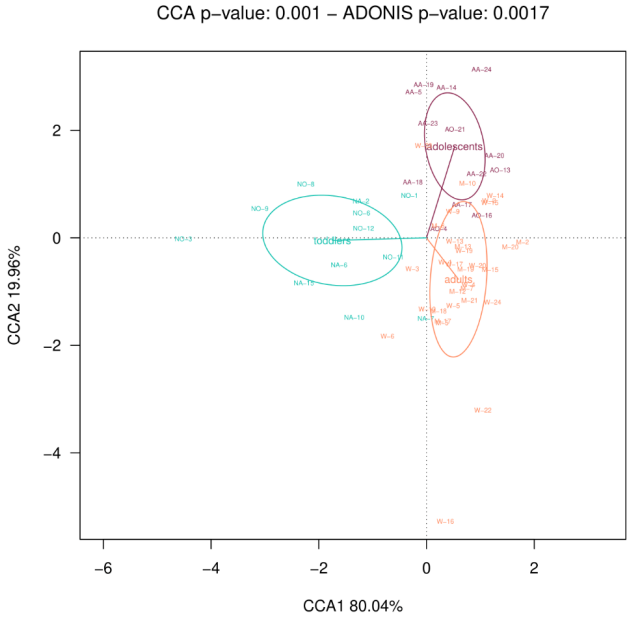
Figure 4.2. Longitudinal development of gut icrobiota diversity. (a) Shannon index and Chao1 diversity and richness measurements respectively for each age group during sampling time. **(b)** Difference in alpha diversity values (Shannon index) between the first time point (T1) and the last one (T10) (intra-group Wilcoxon signed-rank test) and comparison of the T10-T1

difference between age groups (inter-group Mann-Whitney U test). **(c)** Linear Mixed Effects models to test the changes in alpha diversity (Shannon index) over time.

In order to visualize whether age drives differences in gut microbiota composition, we performed Canonical Correspondence Analyses (CCA) and ADONIS tests (PMANOVA) for the three age groups (Figure 4.3a) and for each pairwise age comparison (Table S5 in the Appendix section) at the levels of ASV, species, genus, and phylum relative abundances. Fecal samples from toddlers cluster away from those of adolescents and adults (ADONIS test and CCA p-value < 0.01) in all cases, except in the "toddlers *versus* adolescents" comparison at phylum level (CCA p-value = 0.45; ADONIS p-value = 0.57). When the three groups are considered, the first CCA axis separates toddlers from adolescents and adults, representing 80.04% of the age-associated variability at ASV level, whereas the second axis separates adolescents from adults (19.96% of the variability explained). In this case, CCA and ADONIS p-values are significant at ASV, species and genus levels, but not at phylum level. Furthermore, to visualize the temporal dynamics within each age group, we performed a Principal Coordinates Analysis (PCoA) based on Jaccard index distances and plotted the values of the first and second axes against time

(Figure 4.3b). This plot reveals that, while the location of the samples from adolescents and adults is conserved through time, the toddlers' samples initially occupy higher values on the first axis but over time are displaced towards those of the older subjects. Therefore, we observe a directional change of the toddlers' samples towards the adult gut microbiota composition, so that, towards the end of the sampling period, toddlers, adolescents, and adults share similar values on this axis. Analyses based on Bray-Curtis and UniFrac (weighted and unweighted) distances yielded similar results.

A


CCA p−value: 0.001 − ADONIS p−value: 0.0017
adolescents
toddlers
adults
CCA2 19.96%
CCA1 80.04%
2
0
−2
−4
−6
−4
−2
0
2

B

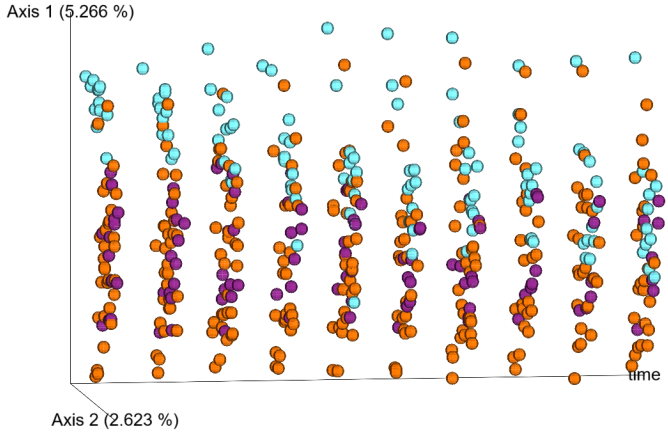

Axis 1 (5.266 %)
time
Axis 2 (2.623 %)

Figure 4.3. Multivariate statistical ordination analysis on taxonomic data. (a) Canonical Correspondence Analysis (CCA) and ADONIS test, and **(b)** Principal Coordinates Analysis (PCoA) on the Jaccard index distances, based on ASV relative abundances from time series data of all subjects from the three age groups.

The functional profile of the metagenome, based on the TIGRFAM hierarchical classification, reveals that inter- and intra-individual variation are less pronounced compared to the taxonomic composition (Figure 4.4a). Functions related to the category (or main role) “transport and binding proteins” are the most abundant across all samples, followed by those related to “protein synthesis”, “protein fate”, “energy metabolism” and “DNA metabolism”.

As in the case of the 16S amplicon data, we performed CCA and ADONIS tests for the three age groups and for each pairwise age comparison (Figure 4.4b; statistical values in Table S5 in the Appendix section) at the level of TIGRFAM subroles and at the lower level of individual TIGRFAM protein annotations, in this case to visualize whether age drives differences in the gut microbiota metagenome. Samples from the three different age groups cluster away from each other (ADONIS test and CCA p-value < 0.05) in all cases except

in the adolescents-adults comparison (ADONIS test p-value = 0.42 and CCA p-value = 0.29 at TIGRFAM protein annotation level; ADONIS test p-value = 0.94 and CCA p-value = 0.64 at TIGRFAM subrole annotation level); therefore, differences between adolescents and adults are less pronounced for function than for taxonomic composition.

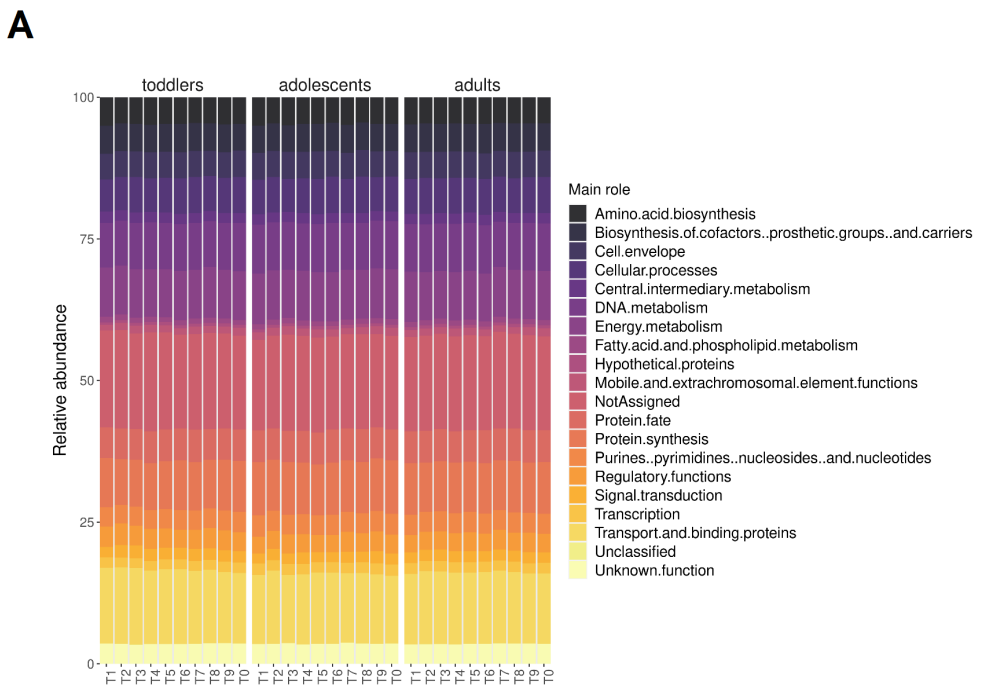


Figure continues on next page

B

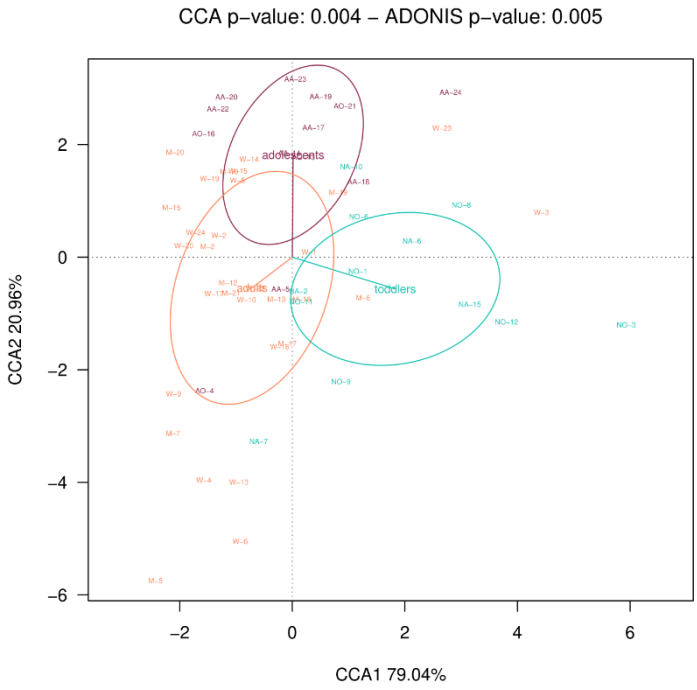


Figure 4.4. Functional profile in fecal samples of toddlers, adolescents, and adults. (a) Relative abundances (y-axis) at TIGRFAM main role level for each sampling timepoint (x-axis) in each age group. **(b)** Canonical Correspondence Analysis (CCA) and ADONIS test based on TIGRFAMs subrole relative abundances from time series data of all subjects from the three age groups.

4.3 DETECTION OF CHARACTERISTIC BACTERIA OF EACH AGE GROUP

Pairwise comparisons using the linear discriminant analysis effect size (LEfSe) algorithm (Segata *et al.*, 2011) show that two phyla are significantly over-represented in toddlers compared with adults, Actinobacteriota (LDA score = 4.5; p-value = 0.01) and Proteobacteria (LDA score = 3.99; p-value = 0.001), while they occur at intermediate abundances in adolescents. Analyses at the genus and species level (Figure 4.5a and Figure 4.5b respectively) indicate that these over-representations in toddlers are mainly due to *Eggerthella* (*E. lenta*), *Gordonibacter* (*G. pamelaeae*), and *Bifidobacterium* (*B. longum*, *B. bifidum*, *B. breve*) in the Actinobacteriota, and to *Enterobacter*, *Actinobacillus*, and *Haemophilus* in the Proteobacteria. *Haemophilus* is also over-represented in toddlers when compared with adolescents, indicating that its relative abundance has already decreased by this period. In contrast, *Bifidobacterium* is not over-represented in toddlers compared to adolescents, but rather decreases significantly between adolescents and adults. Within the Actinobacteriota and Proteobacteria there are also some organisms that are under-represented in toddlers, as is the case of *Olsenella* and Eggerthellaceae DNF00809 in the Actinobacteriota, and of *Parasutterella* (*P.*

excrementihominis) and *Oxalobacter* in the Proteobacteria. *Oxalobacter* is under-represented in toddlers relative to both adolescents and adults, indicating that its relative abundance increases before adolescence, whereas *Olsenella* is over-represented in adults relative to both toddlers and adolescents, indicating that its abundance increases after this period. On the other hand, the relative abundances of Eggerthellaceae DNF00809 and *Parasutterella* have their highest values in adolescents but are only significantly over-represented in this group relative to toddlers, not to adults.

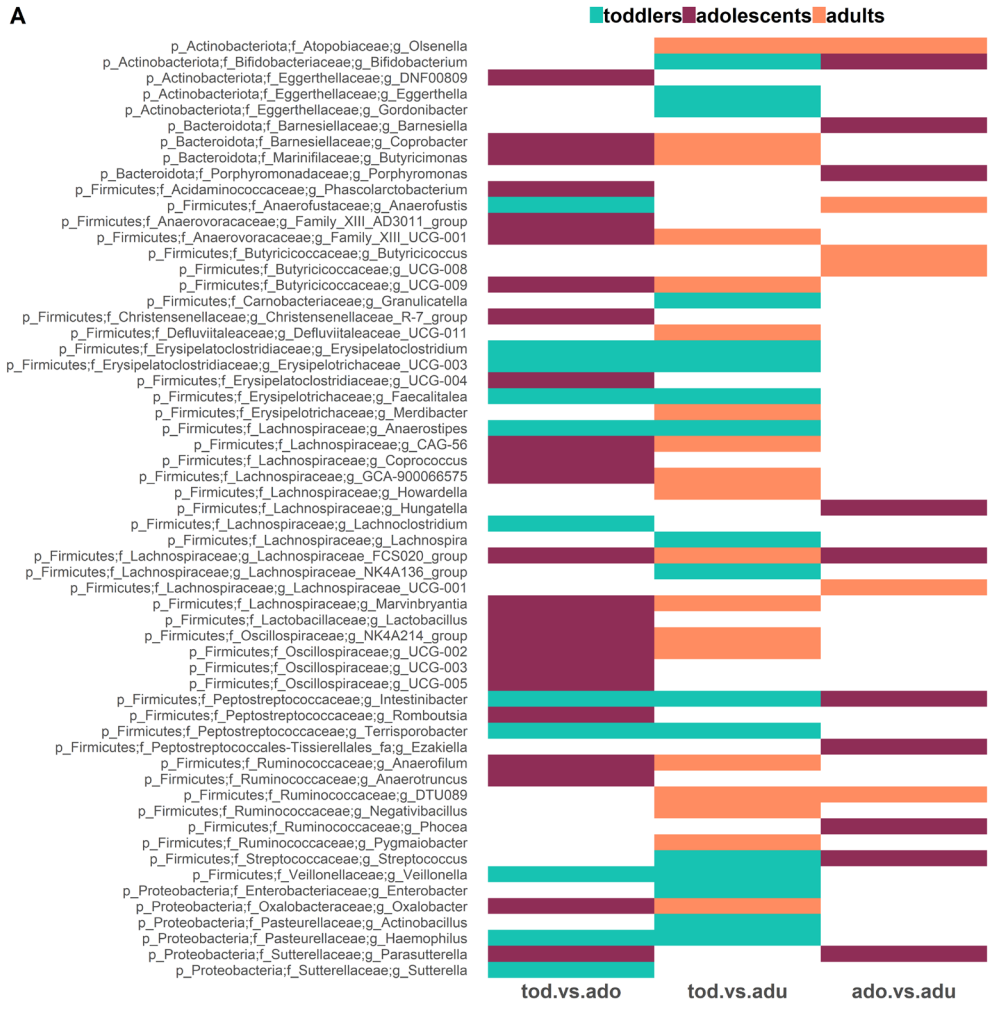


Figure continues on next page

Figure 4.5. Taxonomical biomarkers of the different age groups (I). LEfSe analyses of differences between the three age groups showing the age group in which the feature is significantly over-represented according to the LDA score (log 10) in each pair-wise comparison ("toddlers *versus* adolescents", "adolescents *versus* adults" and "toddlers *versus* adults") for the most discriminative features at genus **(a)** and species **(b)** levels.

Within the Firmicutes phylum, there are also several bacterial groups that are over-represented in toddlers with respect to both adolescents and adults, including *Erysipelatoclostridium* (*E. ramosum*), Erysipelotrichaceae UCG-003, *Faecalitalea* (*F. cylindroides*), *Anaerostipes* (*A. hadrus*), *Intestinibacter* (*I. bartlettii*), *Terrisporobacter*, and *Veilonella* (*V. dispar*). Several others still are over-represented only in relation to adults (*Granulicatella*, *Lachnospira*, Lachnospiraceae NK4A136 group, *Streptococcus*), indicating that they only decrease after adolescence. On the other hand, many Firmicutes are under-represented in toddlers with respect to adolescents and adults, as is the case of *Marvinbryantia*, *Anaerofilum* and various unclassified genus-level groups within the families Anaerovoracaceae, Butyricicoccaceae, Lachnospiraceae, and Oscillospiraceae. There is also a large group of taxa that appear to have increased gradually between the toddler years and adulthood, since they are under-represented in toddlers only with respect to adults, while not differing significantly between the latter and adolescents. This group includes *Howardella* (*H. ureilytica*), *Merdibacter*, *Negativibacillus*, *Pygmaiobacter*, and Defluviitaleaceae UCG011. In addition, Ruminococcaceae DTU089 and *Butyricicoccus faecihominis* are over-represented in adults with respect to

both toddlers and adolescents, but do not differ significantly between adolescents and toddlers, suggesting that they started increasing only after the adolescence period. Finally, there are also some organisms within the Firmicutes having their highest or lowest relative abundances in the adolescents group. These patterns are most pronounced for Lachnospiraceae FCS020 and *Anaerofustis*, which are over- and under-represented, respectively, in relation to both other age groups.

In the Bacteroidota phylum there are no genera over-represented in toddlers, but the species *Bacteroides fragilis* and *B. ovatus* are over-represented relative to both adolescents and adults, while *B. salyersiae* is over-represented only in relation to adolescents. In contrast, *B. coprocola* is over-represented in adults in relation to both toddlers and adolescents, while *B. eggerthii* and *B. massiliensis* are over-represented only in relation to toddlers. In addition, *Coprobacter* (*C. fastidiosus*) and *Butyricimonas*, as well as *Coprococcus comes* and several species of the genus *Alistipes* (*A. ihumii*, *A. indistinctus*, *A. obesi*) are under-represented in toddlers with respect to both adolescents and adults, while not differing between adults and adolescents, indicating that they have increased to adult-like abundances by the

adolescence period. Finally, there are also some Bacteroidota that appear to peak in the adolescents group. This is the case for *Barnesiella* (*B. intestinihominis*) and *Porphyromonas,* which are over-represented in adolescents with respect to adults, although not with respect to toddlers, and for the species *P. bennonis*, which is over-represented in adolescents relative to both.

To further analyze the compositional differences among age groups, we employed the sPLS-DA method (sparse Partial Least Squares-discriminant analysis, implemented in the mixOmics R package; Rohart *et al*., 2017). A Clustered Image Map (CIM) based on the ASVs selected on the first two components (10 and 110 features selected respectively) shows that samples from the same age group cluster together (Figure 4.6a). The taxonomic affiliations of the ASVs associated with each age group are in accordance with LEfSe results (Table S6 in the Appendix section) and similar results were obtained when performing the same analysis at genus level. Relevant correlations ($r > 0.4$) were used to construct networks including the discriminant ASVs for each age group (Figure 4.6b). In toddlers there are positive correlations (blue edges) with ASVs annotated as *Veillonella dispar*,

Bifidobacterium bifidum, *Parabacteroides distasonis*, *Lachnoclostridium* NA, Erysipelotrichaceae NA, and *Haemophilus* NA, with these last two also negatively correlated with adults (red edges). Negative correlations with toddlers involve three ASVs annotated as Oscillospiraceae and one annotated as *Faecalibacterium prausnitzii* (positively correlated with adolescents). ASVs positively correlated with adults are annotated in the Ruminococcaceae, Christensenellaceae and Lachnospiraceae families, in addition to one ASV annotated as *Bacteroides coprocola*. On the other hand, negative correlations with adults include, in addition to the Erysipelotrichaceae NA and *Haemophilus* NA ASVs mentioned above, two ASVs in the Ruminococcaceae family and one assigned to *Desulfovibrio desulfuricans* that are positively correlated with adolescents, as well as two ASVs assigned to *Intestinibacter bartlettii* and *Gordonibater pamelaeae*, which do not have correlations with the other groups. Finally, adolescents show only positive correlations with various ASVs, mainly annotated within the Firmicutes families *Ruminococcaceae* and *Lachnospiraceae*, including two ASVs annotated to *Faecalibacterium prausnitzii* and two to the genus *Lachnoclostridium*. In addition, there are also positive correlations between adolescents and ASVs

annotated to the Firmicutes families Lactobacillaceae, Christensenellaceae, Anaerovoracaceae, and Oscillobacteriaceae, and with ASVs annotated to the Bacteroidota *Butyricimonas* and to the above-mentioned Desulfobacterota *Desulfovibrio desulfuricans.*

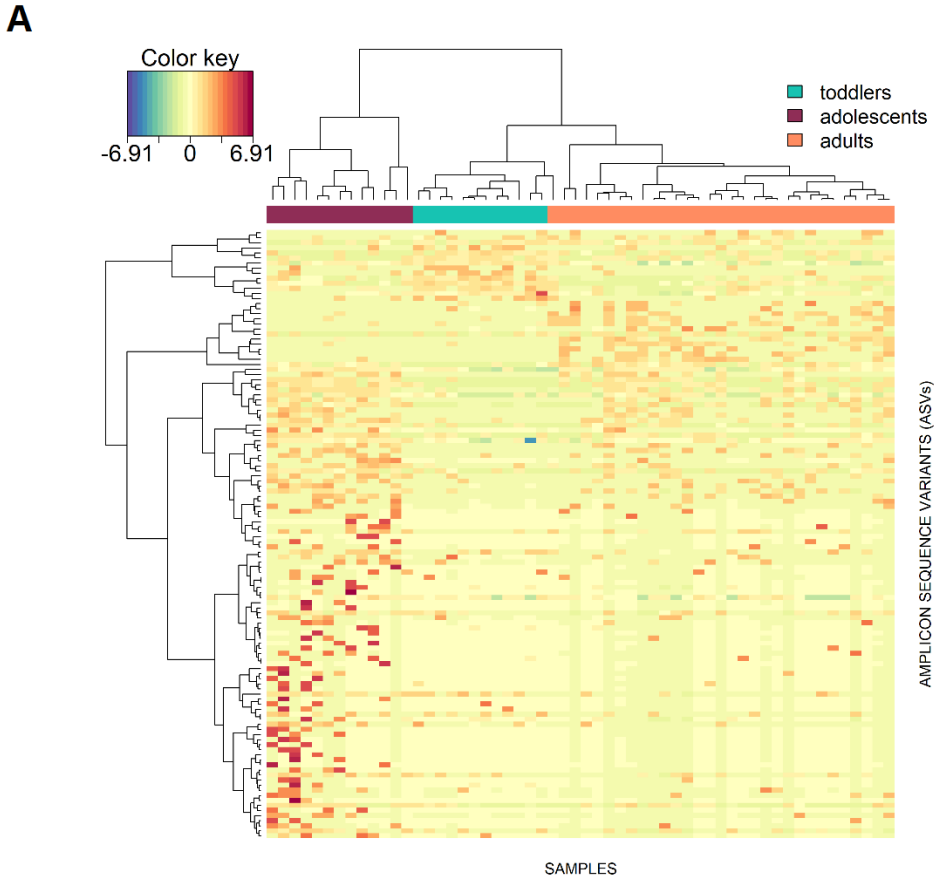


Figure continues on next page

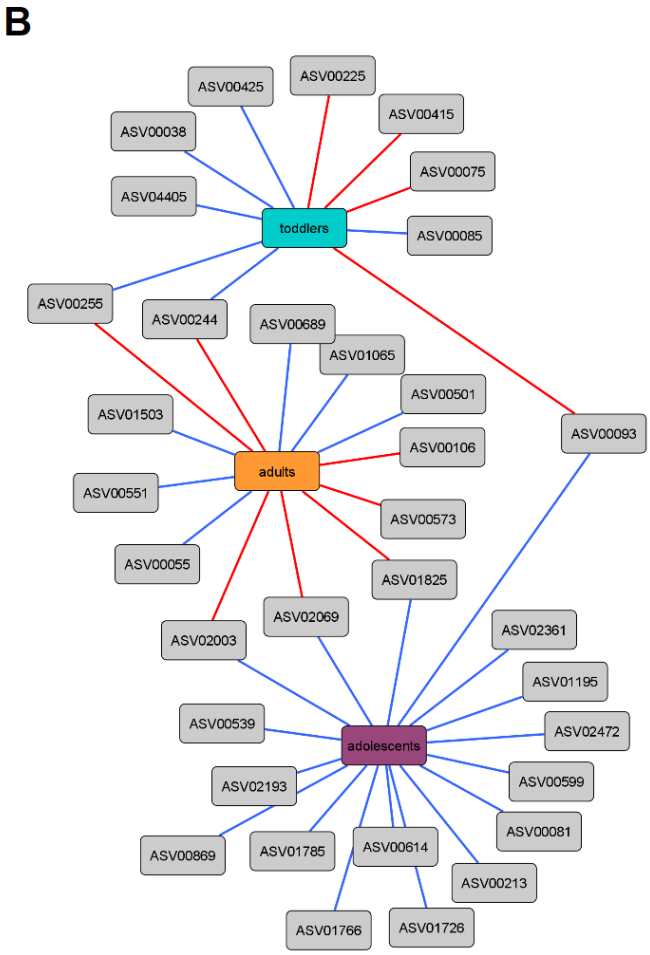


Figure 4.6. Taxonomical biomarkers of the different age groups (II). (a) Clustered Image Map (CIM) at ASV level obtained by sPLS-DA (sparse Partial Least Squares Discriminant Analysis) where samples are represented in columns and selected features in rows (10 and 110 ASVs selected on each component). **(b)** Relevance network of the selected ASVs ($r > 0.4$) by sPLS-DA, showing positive (blue) and negative (red) correlations between features in each age group.

4.4 DETECTION OF CHARACTERISTIC FUNCTIONS OF EACH AGE GROUP

At functional level, LEfSe analysis (Figure 4.7) shows an enrichment in toddlers of various subroles belonging to diverse functional categories with respect to both adolescents and adults. These include both the “anaerobic” and “aerobic” subroles within the “energy metabolism” category, as well as the “DNA interactions” and “small molecule interactions” subroles within “regulatory functions”, the “carbohydrates, organic alcohols and acids” and “cations and iron carrying compounds” subroles within “transport and binding proteins”, the “nitrogen metabolism” subrole within “central intermediary metabolism”, and the “other” subrole within “signal transduction”. In addition, several other subroles are enriched in toddlers only in relation to adults, suggesting that they decline after adolescence. These include the “menaquinone and ubiquinone” subrole within the “biosynthesis of cofactors, prosthetic groups and carriers” category, the “biosynthesis and degradation of polysaccharides”, “pentose phosphate pathway” and “TCA cycle” subroles within the “energy metabolism” category, and the “amino acids, peptides and amines” and “anions” subroles within the “transport and binding proteins” category. In contrast, there are numerous subroles that rather increase in

adolescence and endure in adulthood: “thiamine” within the “biosynthesis of cofactors, prosthetic groups and carriers” category, “chemoautotrophy”, “glycolysis/gluconeogenesis” and “other” within the “energy metabolism” category, “DNA replication, recombination and repair” and “other” within the “DNA metabolism” category, and “salvage of nucleosides and nucleotides” within the “purines, pyrimidines, nucleosides and nucleotides” category, as well as several subroles within the “transcription”, “protein synthesis”, and “protein fate” categories.

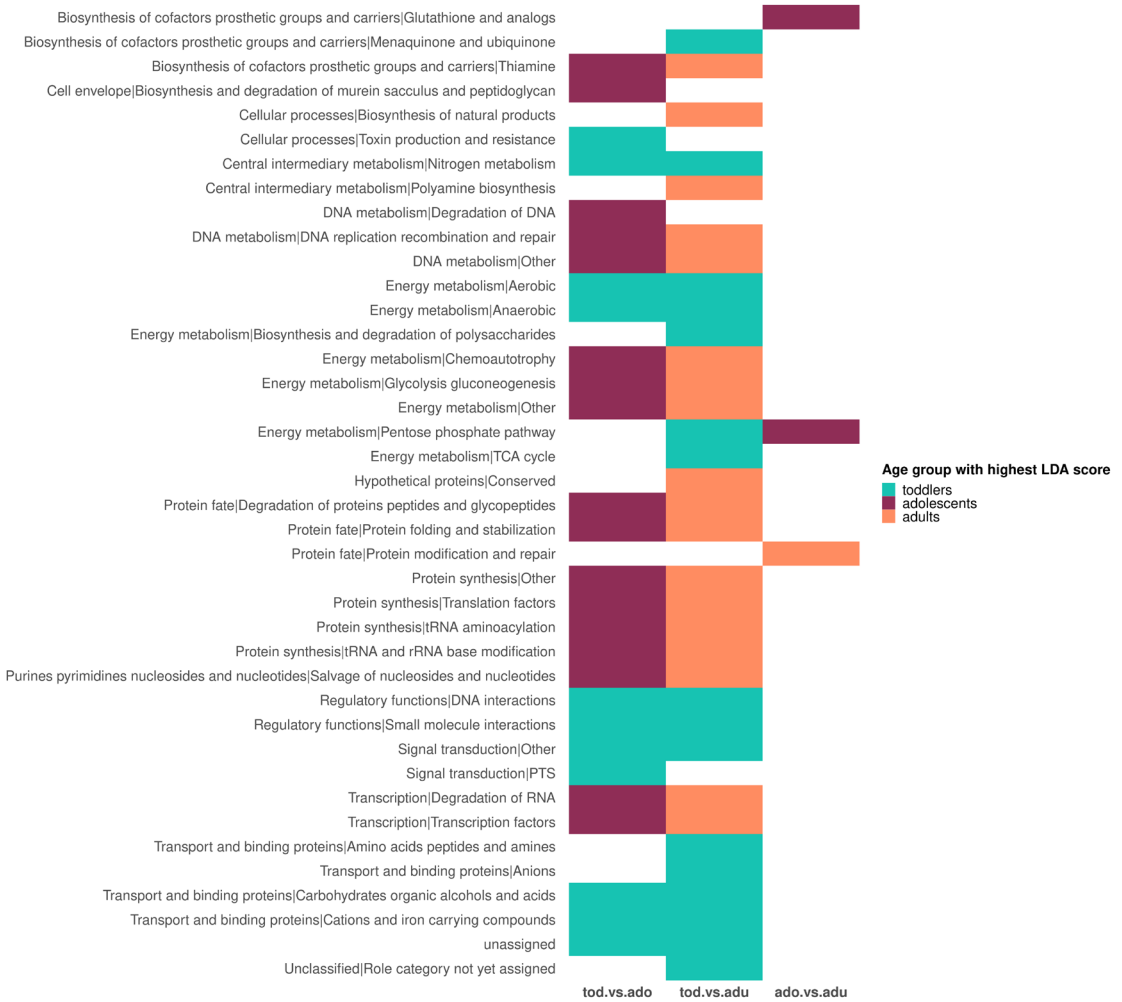

Biosynthesis of cofactors prosthetic groups and carriers|Glutathione and analogs
Biosynthesis of cofactors prosthetic groups and carriers|Menaquinone and ubiquinone
Biosynthesis of cofactors prosthetic groups and carriers|Thiamine
Cell envelope|Biosynthesis and degradation of murein sacculus and peptidoglycan
Cellular processes|Biosynthesis of natural products
Cellular processes|Toxin production and resistance
Central intermediary metabolism|Nitrogen metabolism
Central intermediary metabolism|Polyamine biosynthesis
DNA metabolism|Degradation of DNA
DNA metabolism|DNA replication recombination and repair
DNA metabolism|Other
Energy metabolism|Aerobic
Energy metabolism|Anaerobic
Energy metabolism|Biosynthesis and degradation of polysaccharides
Energy metabolism|Chemoautotrophy
Energy metabolism|Glycolysis gluconeogenesis
Energy metabolism|Other
Energy metabolism|Pentose phosphate pathway
Energy metabolism|TCA cycle
Hypothetical proteins|Conserved
Protein fate|Degradation of proteins peptides and glycopeptides
Protein fate|Protein folding and stabilization
Protein fate|Protein modification and repair
Protein synthesis|Other
Protein synthesis|Translation factors
Protein synthesis|tRNA aminoacylation
Protein synthesis|tRNA and rRNA base modification
Purines pyrimidines nucleosides and nucleotides|Salvage of nucleosides and nucleotides
Regulatory functions|DNA interactions
Regulatory functions|Small molecule interactions
Signal transduction|Other
Signal transduction|PTS
Transcription|Degradation of RNA
Transcription|Transcription factors
Transport and binding proteins|Amino acids peptides and amines
Transport and binding proteins|Anions
Transport and binding proteins|Carbohydrates organic alcohols and acids
Transport and binding proteins|Cations and iron carrying compounds
unassigned
Unclassified|Role category not yet assigned
tod.vs.ado
tod.vs.adu
ado.vs.adu
Age group with highest LDA score
toddlers
adolescents
adults

Figure 4.7. Functional biomarkers of the different age groups (I). LEfSe analyses of differences between the three age groups showing the age group in which the feature is significantly over-represented according to the LDA score (log 10) in each pair-wise comparison ("toddlers *versus* adolescents", "adolescents *versus* adults" and "toddlers *versus* adults") for the most discriminative TIGRFAM subroles.

A CIM based on the TIGRFAM annotations selected on the first four components obtained through sPLS-DA (300, 210, 240 and 280 features selected respectively) shows that samples from toddlers and adults are mainly represented in single clusters, whereas samples from adolescents can be found in various clusters also containing samples from the other age groups (Figure 4.8a). Functions associated with each group are mostly in accordance with LEfSe results (Table S7). Relevant correlations ($r > 0.4$) were used to construct networks at TIGRFAM level (Figure 4.8b): in toddlers there is a positive correlation with functions related to "amino acid biosynthesis/pyruvate family" (subrole 74), "biosynthesis of cofactors, prosthetic groups and carriers/glutathione and analogs" (subrole 86), and "cellular processes/other" (subrole 92), as well as to "protein synthesis/translation factors" (subrole 169), "energy metabolism" ("aerobic" and "anaerobic"; subroles 108 and 110 respectively), "transport and binding

proteins" ("amino acids, peptides, and amines", "anions", and "carbohydrates, organic alcohols and acids"; subroles 142, 143, and 144 respectively), "regulatory functions" ("DNA interactions" and "small molecule interactions"; subroles 261 and 264), and "signal transduction" ("PTS" and "other"; subroles 700 and 710). Functions in some of these subroles are negatively correlated with adolescents ("regulatory functions/small interactions", "signal transduction/PTS", and "signal transduction/other"; subroles 264, 700, and 710) or adults ("biosynthesis of cofactors, prosthetic groups and carriers/glutathione and analogs" and "energy metabolism/aerobic"; subroles 86 and 108). Subroles containing functions that are negatively correlated with toddlers mainly belong to "transcription" ("degradation of RNA", "DNA-dependent RNA polymerase", and "transcription factors"; subroles 134, 135, and 165), "protein synthesis" ("other", "tRNA aminoacylation", "tRNA and rRNA base modification", and "translation factors"; subroles 136, 137, 168, and 169) and "protein fate" ("protein folding and stabilization" and "degradation of proteins, peptides, and glycopeptides"; subroles 95 and 138), "energy metabolism" ("amino acids and amines", "glycolysis/gluconeogenesis", "chemoautotrophy", and "other";

subroles 109, 116, 180 and 184), “DNA metabolism” (“other”, “degradation of DNA”, and “DNA replication, recombination and repair”; subroles 130, 131, and 132), and “purines, pyrimidines, nucleosides and nucleotides” (“nucleotide and nucleoside interconversions” and “salvage of nucleosides and nucleotides”; subroles 124 and 127). Functions in many of these subroles are positively correlated with adults, including subroles “nucleotide and nucleoside interconversions” and “salvage of nucleosides and nucleotides” within “purines, pyrimidines, nucleosides and nucleotides” main role (subroles 124 and 127), “degradation of DNA” and “other” subroles within “DNA metabolism” main role (subroles 130 and 131), “transcription/degradation of RNA” (subrole 134), “protein synthesis/other” (subrole 136), “protein fate/ degradation of proteins, peptides, and glycopeptides” (subrole 138), and “energy metabolism/other” (subrole 184). In addition, adults also have positive correlations with functions within “protein fate/protein modification and repair” (subrole 140), “central intermediary metabolism/nitrogen fixation” (subrole 179) and “biosynthesis of cofactors, prosthetic groups and carriers/thiamine” (subrole 162), as well as a negative correlation with functions in the subrole “energy

metabolism/pentose phosphate pathway" (subrole 117). Finally, there are no functions positively correlated with adolescents, and only four of them are negatively correlated with this age group, i. e., the three previously mentioned functions that are positively correlated with toddlers and a fourth function belonging to "biosynthesis of cofactors, prosthetic groups and carriers/heme, porphyrin, and cobalamin" (subrole 79).

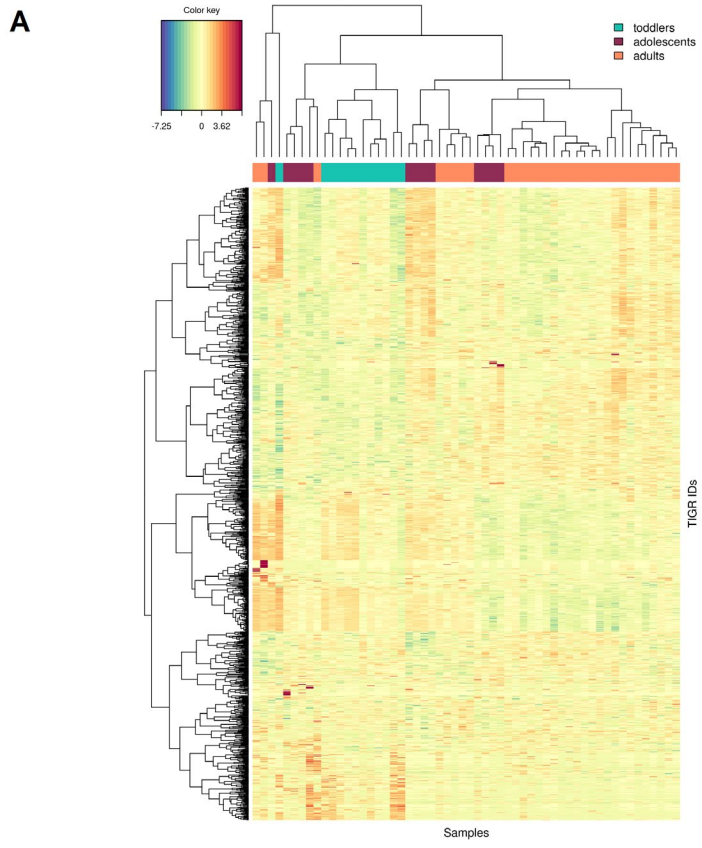


Figure continues on next page

B

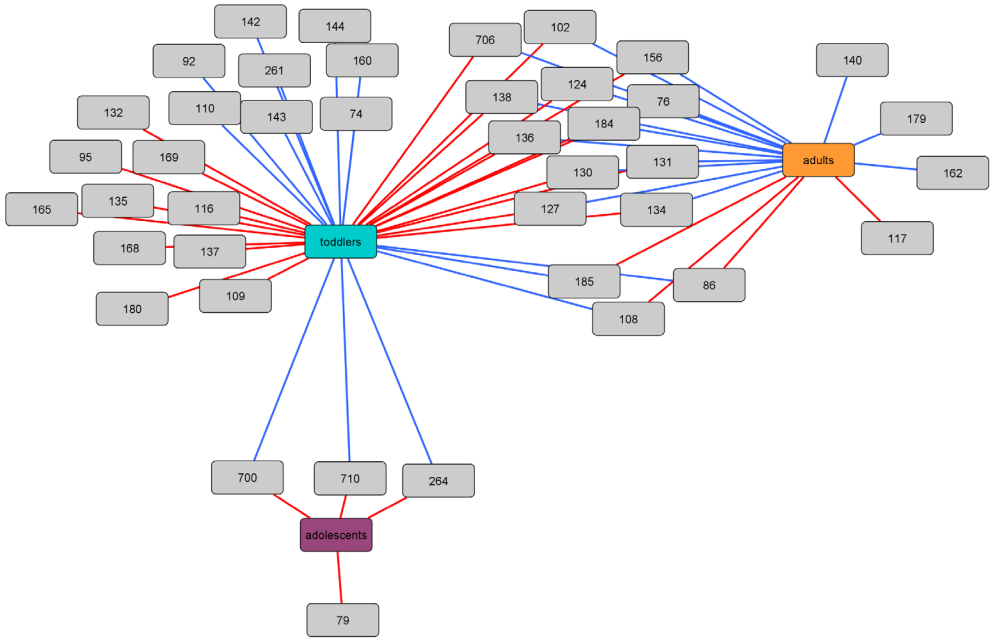


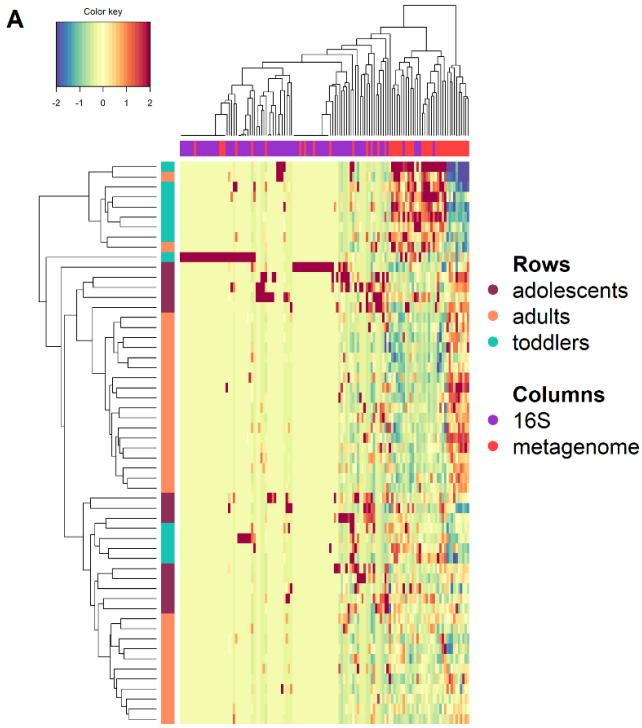
Figure 4.8. Functional biomarkers of the different age groups (II). (a) Clustered Image Map at TIGRFAM level obtained by sPLS-DA (sparse Partial Least Squares Discriminant Analysis) where samples are represented in columns and selected features in rows (300, 210, 240 and 280 features selected on each component). **(b)** Relevance network of the selected TIGRFAMs (r>0.4) by sPLS-DA, showing positive (blue) and negative (red) correlations between features in each age group. For each TIGRFAM, a number representing the subrole to which it belongs is shown (see text and Table S7 for the correspondence to subrole names).

4.5 RELATING TAXONOMIC AND FUNCTIONAL SIGNATURES OF THE MICROBIOTA

In addition to looking for taxonomic and functional biomarkers in the 16S rRNA gene amplicon and metagenome data sets separately, we tried to identify the relationships between these biomarkers using DIABLO (*D*ata *I*ntegration *A*nalysis for *B*iomarker discovery using *L*atent c*O*mponents), a multivariate integrative classification method, also implemented in the mixOmics R package (Singh *et al*., 2016), which seeks for molecular signatures across different data types. The CIM based on the molecular signatures identified in the first four components (7, 30, 30 and 12 from the 16S rRNA gene amplicon data, and 30, 5, 5 and 5 from the metagenome data) shows that most samples from each age group are contained in two or three distinct clusters, each containing mostly samples from the same group (Figure 4.9a). A relevance network constructed with the most relevant of the selected features (both ASVs and TIGRFAMs) highlights three subnetworks (Figure 4.9b; Pearson's correlation coefficient > 0.7). The first subnetwork (left) contains 6 TIGRFAMs associated with toddlers (Table S8 in the Appendix section) belonging to the subrole "energy metabolism/electron transport", the subrole "cellular processes/detoxification" and the subrole 719 (not

assigned, but mostly containing functions associated to CRISPR proteins). These first subnetwork functions positively correlate with ASVs annotated as *Prevotella copri*, *Phascolarctobacterium succinatutens*, *Veillonella atypica*, Erysipelotrichaceae UCG-003 NA, *Paraprevotella clara*, *Bacteroides uniformis*, *Bacteroides* NA, Lachnospiraceae NK4A136 group NA, Incertae Sedis NA, *Butyricicoccus faecihominis*, Oscillospiraceae NA, Flavobacteriales NA, *Blautia* NA, and *Actinobacillus* NA. The second subnetwork (center) contains 4 TIGRFAMs associated with adolescents (Table S8 in the Appendix section) belonging to subroles "regulatory functions/protein interactions" and "central intermediary metabolism/other," and two other functions that have not been assigned to any subrole. The second subnetwork functions positively correlate with ASVs annotated as *Prevotella copri*, *Ruminococcus callidus*, *Veillonella* NA, *Holdemanella biformis*, *Subdoligranulum variabile*, Peptococcaceae NA, *Bacteroides* NA, Ruminococcaceae NA, Flavobacteriales NA, Anaerovoracaceae Family XIII AD3011 group NA and *Alistipes* NA. The third subnet (right) includes a single ASV belonging to an unassigned *Bifidobacterium* (ASV00007) that is associated with toddlers (Table S8 in the Appendix section). This *Bifidobacterium* ASV correlates positively with

functions related to subroles “purines, pyrimidines, nucleosides, and nucleotides/2-deoxyribonucleotide metabolism”, “transport and binding proteins/cations and iron carrying compounds”, and “cellular processes/chemotaxis and motility”, and negatively with TIGR03591 within “transcription/degradation of RNA”; this last function is over-represented in adults, where bifidobacteria are under-represented. The C*ircos* plot (Figure 4.9c) highlights the mainly positive correlations between ‘omics features (Pearson’s correlation coefficient > 0.7), showing the relative abundance of each feature in the three age groups (ASV and TIGRFAM IDs and their associated taxa and subroles, as well as the corresponding age group in which the relative abundance of each feature is highest, are shown in Table S8 in the Appendix section); most of the features are enriched in toddlers and/or adolescents.



B

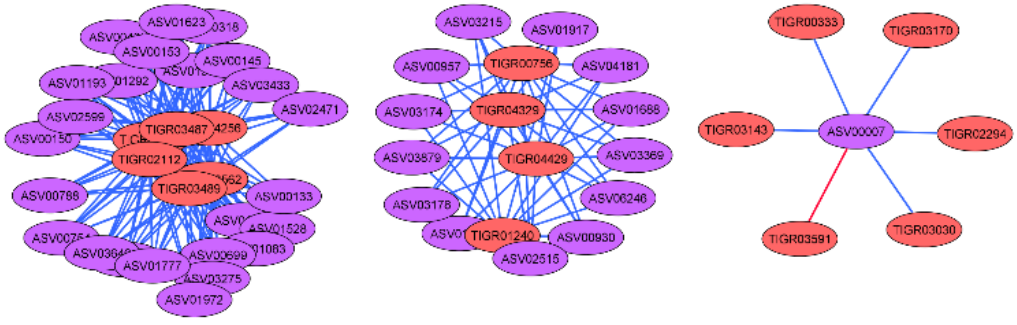


Figure continues on next page

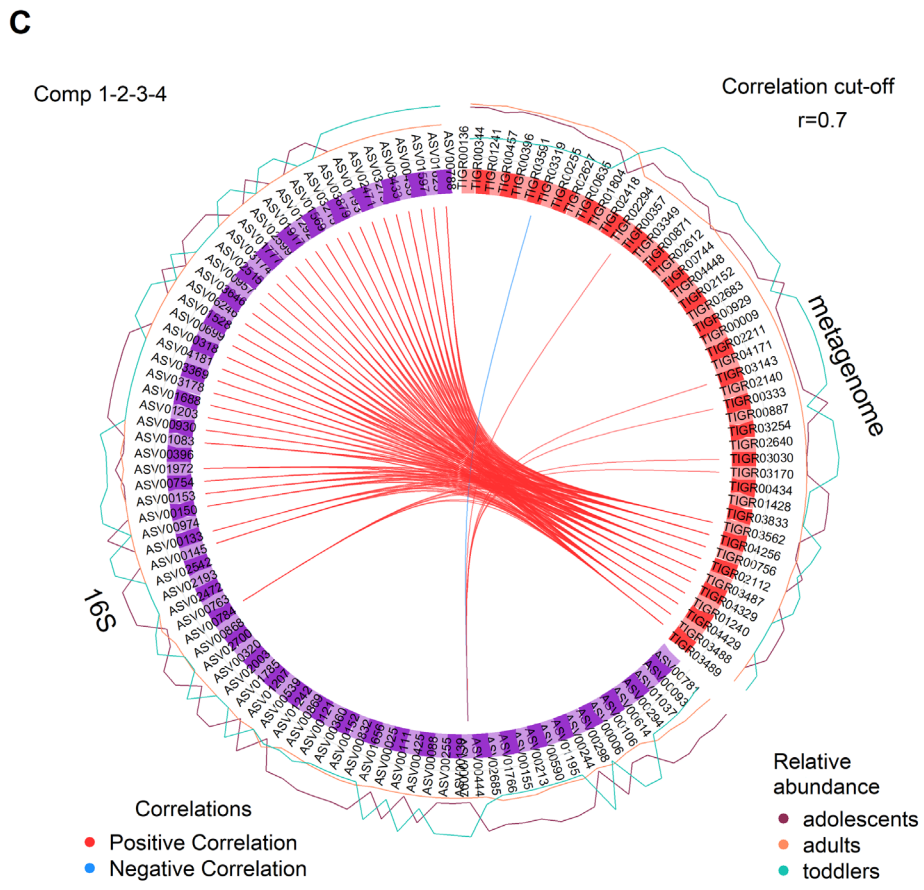


Figure 4.9. Multi 'omics integrative analysis. (a) Clustered Image Map (CIM) of the multi-omics molecular signature, where samples are represented in rows and selected features on the first four components in columns. **(b)** Relevance network visualization of the selected features. **(c)** *Circos* plot showing the positive (blue) and negative (red) correlations (r > 0.7) between selected features and the age group in which each of the features has a greater presence.

Furthermore, we checked whether the taxonomical annotations of metagenomic reads supported the associations between taxons and functions indicated by DIABLO. Although most of the ORFs corresponding to the TIGRFAMs present in the DIABLO networks were not taxonomically annotated, we could confirm some of the associations. In particular, regarding the TIGRFAMs positively associated with *Bifidobacterium* ASV00007 in the third subnetwork, reads mapping to ORFs annotated as TIGR00333 (*nrdl* protein, subrole "2-deoxyribonucleotide metabolism") were present only in toddlers and adolescents and were mostly assigned to *Bifidobacterium*. In contrast, reads mapping to ORFs annotated as TIGR03591 (polyribonucleotide nucleotidyltransferase, subrole "degradation of RNA"), negatively associated with *Bifidobacterium* ASV00007 and over-represented in adults (Table S8), were mainly assigned to *Faecalibacterium* (data not shown).

4.6 TEMPORAL STABILITY OF THE MICROBIOTA AT DIFFERENT AGES

To study the overall stability of microbiota composition within our three age groups, we have applied *complexCruncher* (Martí *et al.*, 2017). Applying this method to 16S rRNA amplicon data we can see that toddlers are the most unstable group (Figure 4.10a), while adolescents and adults have similar

distributions of V and β. The method also calculates the rank stability index (RSI) of all the taxa from all samples in the data set. A CCA and an ADONIS test based on the RSI values show that the three age groups cluster away from each other (Figure 4.10c), with adolescents showing the most dispersion among individuals. Figure 4.11a shows those ASVs whose RSI is significantly different between two age groups (Wilcoxon signed-rank test p-value < 0.1). In toddlers, the following annotated ASVs have a significantly higher RSI than in adults and adolescents: *Bacteroides* NA, *Alistipes* NA, *Acidaminococcus intestini*, Clostridium sensu stricto 1 NA, *Turicibacter sanguinis*, *Agathobacter* NA, and *Streptococcus* NA, and only *Subdoligranulum* NA has a lower RSI in this group than in the other two; *Coprococcus comes* and *Subdoligranulum* NA have a higher RSI in adolescents than in toddlers, but not in respect to adults. Some taxa have high values of RSI in toddlers and adults, but significantly lower values in adolescents, as is the case of *Adlercreutzia equolifaciens*, *Parabacteroides distasonis*, *Anaerostipes hadrus*, *Faecalibacterium prausnitzii*, and *Subdoligranulum variabile*. The same analysis with metagenome data (at subrole level) does not show a significant overall difference between the different age groups in terms of stability (Figures

4.10b and 4.10d). However, some specific subroles do behave differently in the different age groups. Figure 4.11b shows those subroles whose RSI is significantly different between two age groups (Wilcoxon signed-rank test p-value < 0.1): subroles “cellular processes/adaptations to atypical conditions”, “energy metabolism/sugars”, “transport and binding proteins/cations and iron carrying compounds”, and “degradation of proteins, peptides and glycopeptides” and “protein modification and repair” in the “protein fate” category have higher RSI values in toddlers than in adolescents and adults; in adolescents, the subroles “cell envelope/biosynthesis and degradation of murein sacculus and peptidoglycan”, “cellular processes/sporulation and germination”, and “protein synthesis/tRNA aminoacylation” have higher RSI values than in toddlers, but not than in adults. In this latter group, only one subrole (“energy metabolism/electron transport”) has higher values of RSI with respect to toddlers, but not to adolescents.

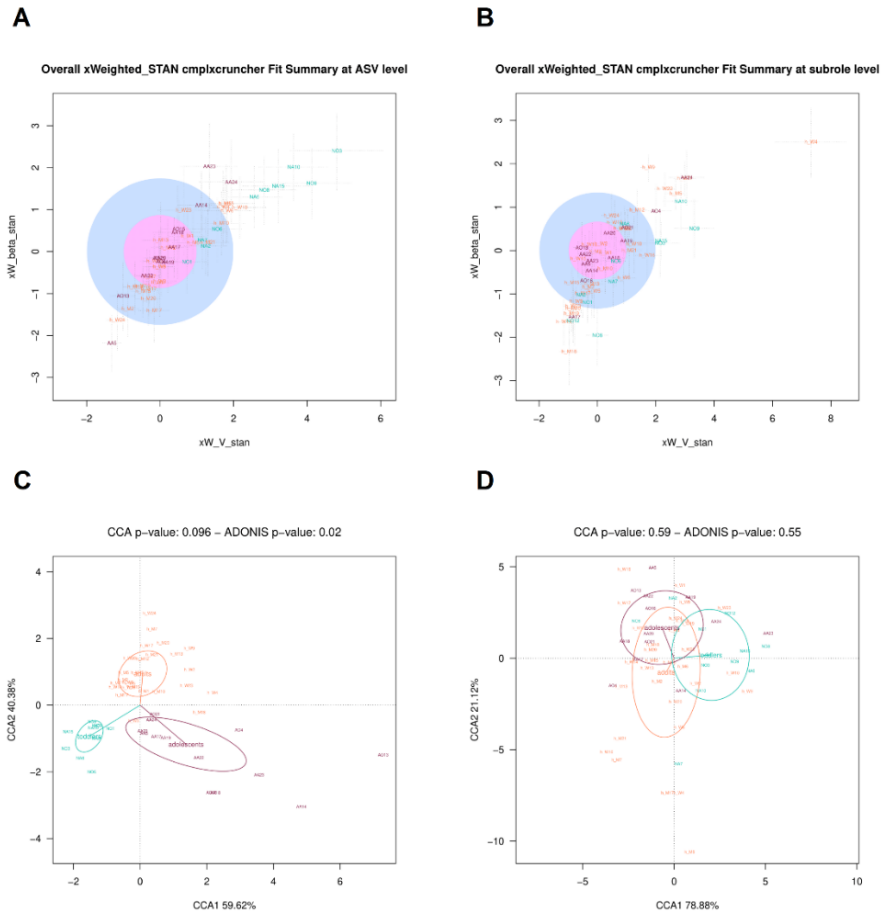


Figure 4.10. Temporal stability of the microbiota (I). (a,b) Taylor's law parameters space, with the variability (V) of the ASVs (a) or functional subroles (b) in the x-axis and the scale index (β) in the y-axis, both in standard deviation units; the pink circle corresponds to the region containing 68% of the adult subjects (taken as reference group) in the Taylor parameters space and the blue circumference around it delimits the 98% region. Dashed lines represent the error of the calculated parameters for each subject. **(c,d)** Canonical Correspondence Analysis (CCA) and ADONIS test based on RSI data for annotated ASVs (c) and subroles (d).

Figure continues on next page

Figure 4.11. Temporal stability of the microbiota (II). Features (annotated ASVs and subroles, respectively) whose RSI is significantly different between two age groups (Wilcoxon signed-rank test p-value < 0.1).

In addition, to further examine the stability of the microbiota over time within the different age groups, we analyzed the amount of change per unit of time throughout the sampling period. For this purpose, we calculated for all pairs of consecutive samples from the same subject the Jaccard index, a beta-diversity metric that indicates the fraction of shared features between them, both at compositional and functional levels, considering ASV and TIGRFAM annotations, respectively. Stability in all three age groups is lower at the compositional level than at the functional level (Figure 4.12). At both the compositional and functional levels, adolescents and adults have similar values of the Jaccard index throughout the sampling period. In contrast, toddlers have a significantly lower stability than adolescents and adults during the first-time interval (Mann-Whitney U test, p-value < 0,01 for the T1-T2 comparison at both levels. All p-values available in Table S9 in the Appendix section), with the fraction of shared features between consecutive samples increasing after that. This indicates that both the taxonomic and the functional composition of the microbiota stabilize during the toddler period.

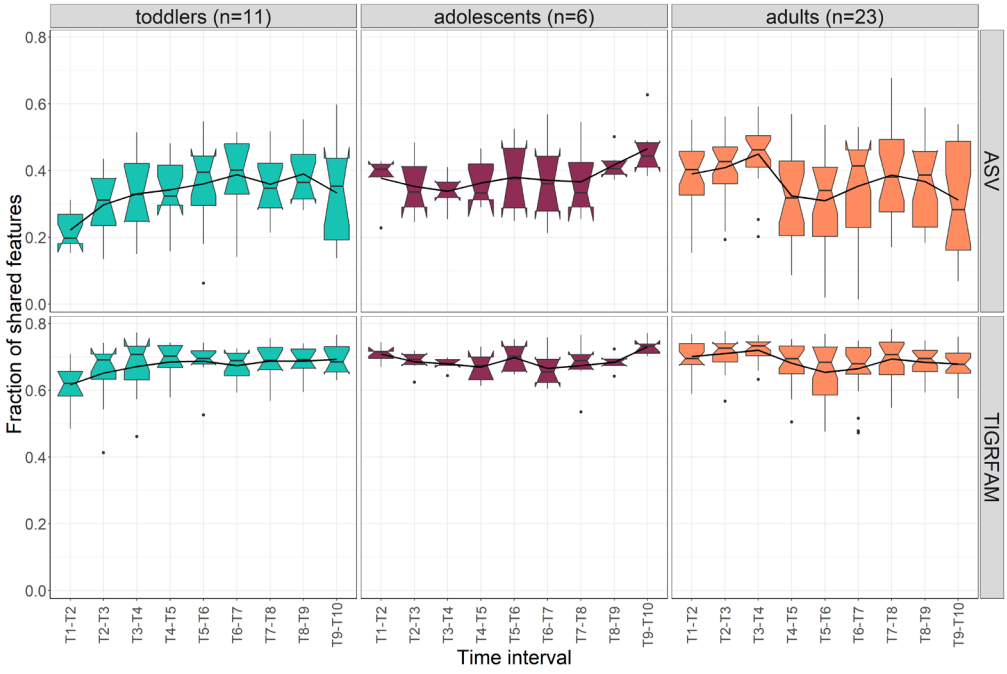


Figure 4.12. Temporal stability of the microbiota (III). Stability analysis at compositional and functional levels. Jaccard index for each consecutive sample pair from the same subject. Each box represents the mean Jaccard index for each set of consecutive sample pairs in each age group at ASV level (top) and TIGRFAM level (bottom); lines connect the means of each box.

Furthermore, in order to detect temporal trends of taxon variability within age groups we used a supervised learning regressor within the q2-longitudinal pipeline included in the QIIME 2 microbiome analysis platform (Bokulich *et al.*, 2018). This allowed us to identify several cases in which an initial over-representation of an ASV in toddlers disappears during the

sampling period (Figure 4.13). This is the case of ASV00031 (*Blautia obeum*), ASV00089 (*Streptococcus* NA), and ASV00046 (*Bifidobacterium bifidum*). ASV00031 is found at very low abundance in adolescents and adults throughout the sampling period, while in toddlers it has a higher abundance at T1 but drops close to 0% at T2 (24 months of age in average; Figure 4.13a). ASV00089 is also highest in toddlers at T1 and is nearly absent in adolescents and adults, but, after a decrease from T1 to T2, it then remains at low levels until T9 (4.17 years of age in average), at which point it practically disappears (Figure 4.13b). Also noteworthy is the case of ASV00046, which appears in similar relative abundances in the three groups at T1 (although highest in toddlers and lowest in adults) and then rises in toddlers until T4 (32 months of age in average), to decrease to values similar to those of adults and adolescents by T5 (36 months of age in average; Figure 4.13c).

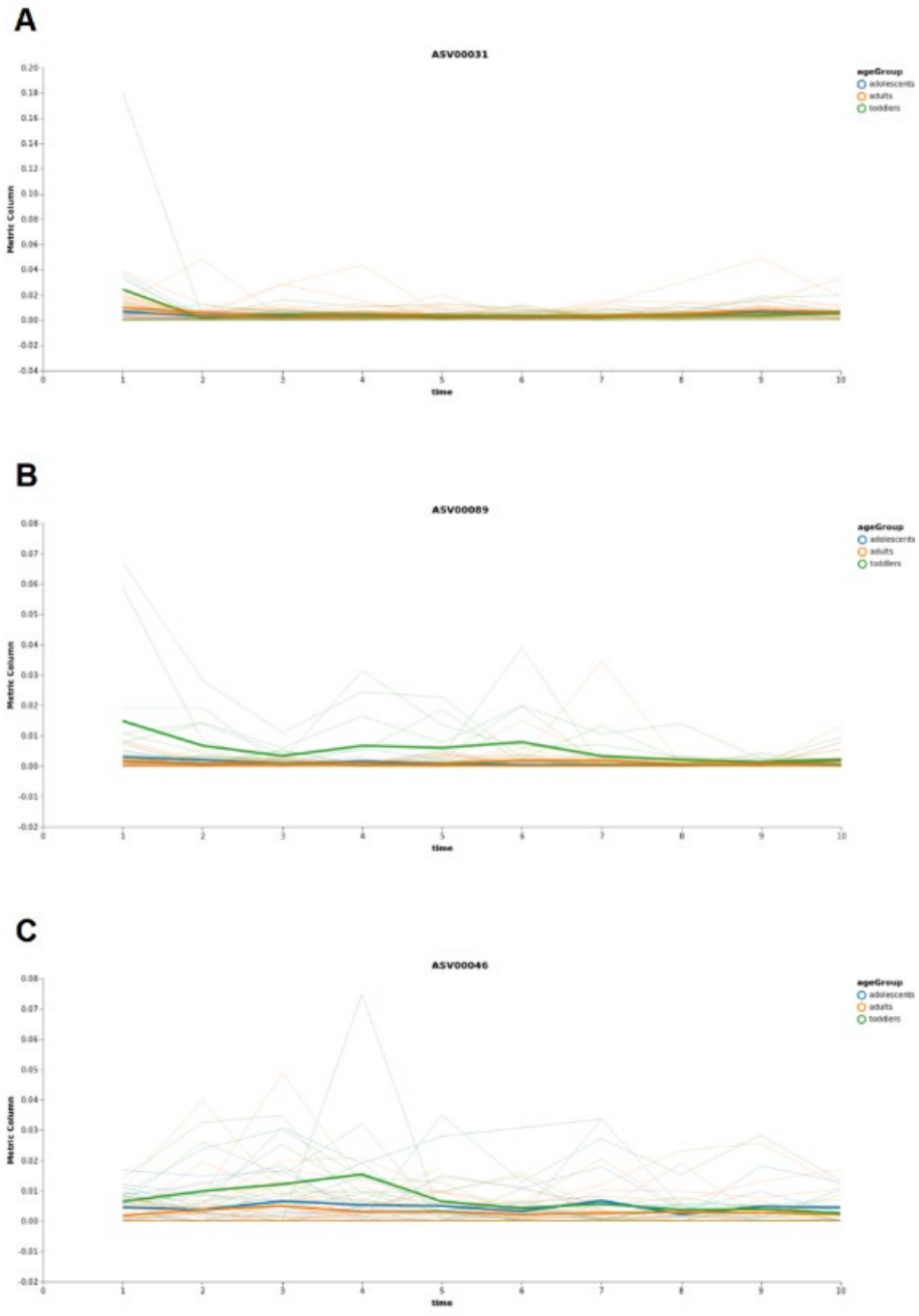

A
ASV00031
ageGroup
adolescents
adults
toddlers
Metric Column
time
B
ASV00089
ageGroup
adolescents
adults
toddlers
Metric Column
time
C
ASV00046
ageGroup
adolescents
adults
toddlers
Metric Column
time

Figure 4.13. Feature volatility plots with those features that are predictive of the time point for each group. Longitudinal relative abundances (y-axis) at each time point (x-axis) are represented for **(a)** ASV00031 (*Blautia obeum*), **(b)** ASV00089 (*Streptococcus* NA) and **(c)** ASV00046 (*Bifidobacterium bifidum*).

4.7 CHANGES IN THE GUT MICROBIOTA AFTER WEANING IN TODDLERS

Significant changes in diet may produce shifts in the gut microbiome. We already know that breast milk feeding in infants has a major impact in the development of the gut microbiota during this period. To investigate how the cessation of breast milk feeding impacts the composition and function of toddlers' microbiota we have compared the samples of toddlers before and after weaning (N=27 and N=72 respectively; Table 4.1 and Table S1 in the Appendix section).

Table 4.1. Time points at which weaning occurred in each toddler.

Subject	Weaning time point
NO-1	T2
NA-2	T4
NO-3	T7
NA-6	--
NO-6	--
NA-7	T9
NO-8	T5
NO-9	T9

NA-10	T4
NO-11	T2
NO-12	T4
NA-15	--

-- Weaning occurred before the start of the study.

Canonical Correspondence Analysis (CCA) and ADONIS test (PMANOVA) show that samples before and after weaning cluster away from each other at ASV level (Figure 4.14a; ADONIS test and CCA p-value < 0.05) and TIGRFAM annotation level (Figure 4.14b; ADONIS test and CCA p-values = 0.053 and 0.056 respectively), and diversity increases after weaning (Mann-Whitney U test: Shannon index p-value = 0.00039; Chao1 p-value = 0.00025; Figure 4.14c and 4.14d respectively).

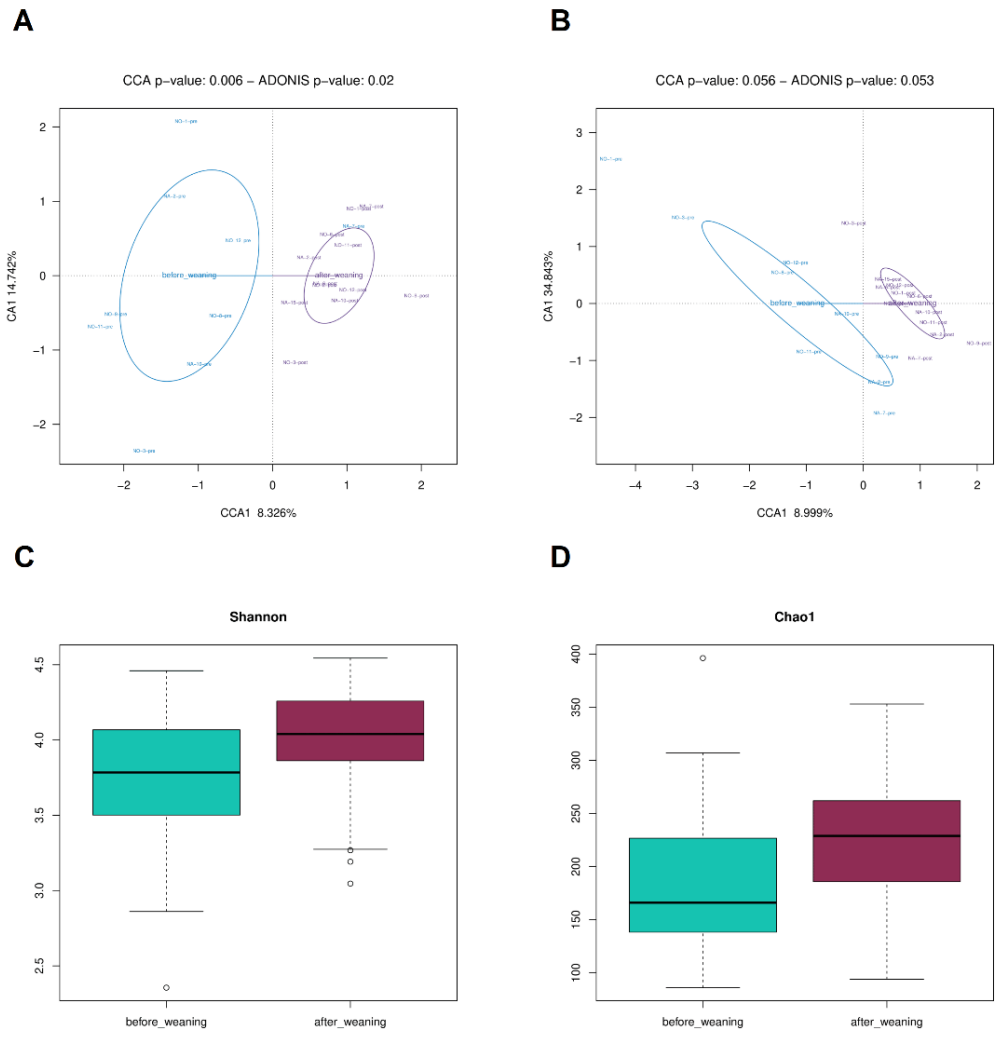


Figure 4.14. (a,b) Canonical Correspondence Analysis (CCA) at the level of **(a)** ASVs and **(b)** TIGRFAMs relative abundances, and **(c,d)** alpha diversity of samples from toddlers before and after weaning.

LEfSe analysis (Figure 4.15) at compositional level shows that before weaning (Figure 4.15a) there is an over-representation of ASVs that are

annotated as *Bifidobacterium* (ASV00007, ASV00254), *Veillonella* (ASV00594, ASV00354, ASV00435), *Gemella sanguinis* (ASV02921), Clostridia UCG-014 (ASV00494) and *Erysipelatoclostridium ramosum* (ASV00321); the latter also appeared to be over-represented in toddlers compared to adolescents and adults in the LEfSe analysis above. After weaning, ASVs annotated as *Ralstonia* (ASV01061), *Bacteroides* (ASV00157), *Odoribacter splanchnicus* (ASV00143), *Blautia* (ASV00165), *Lachnospira pectinoschiza* (ASV00140), and the Clostridia vadinBB60 group (ASV00749) are over-represented, as well as 13 ASVs in the Oscillospiraceae family; members of the Oscillospiraceae were under-represented in toddlers as a whole in the analyses above. At functional level, TIGRFAM annotations over-represented before weaning (Figure 4.15b) are in the subroles "cellular processes/toxin production and resistance" (TIGR00711), "protein fate/protein and peptide secretion and trafficking" (TIGR03924), "transport and binding proteins/amino acids, peptides and amines" (TIGR03262), "transport and binding proteins/cations and iron carrying compounds" (TIGR02770), "biosynthesis of cofactors, prosthetic groups, and carriers/menaquinone and ubiquinone" (TIGR01923), and "DNA metabolism/restriction/modification" (TIGR00348). All these subroles (except

the last) were over-represented in toddlers in the LEfSe age-groups comparisons above. After weaning, over-represented TIGRFAM annotations are in the subroles “protein fate/degradation of proteins, peptides, and glycopeptides” (TIGR00763), “DNA metabolism/DNA replication, recombination, and repair” (TIGR02785, TIGR02773), “energy metabolism/amino acids and amines” (TIGR00461), “cellular processes/DNA transformation ” (TIGR03783, TIGR03782), “transport and binding proteins/cations and iron carrying compounds” (TIGR00437), “transport and binding proteins/carbohydrates, organic alcohols, and acids” (TIGR02633), “unknown function/general” (TIGR00762), “energy metabolism/other” (TIGR01828) and “energy metabolism/fermentation” (TIGR03948). Of these, the first three are under-represented in toddlers compared to adolescents and adults in the LEfSe age-groups comparison.

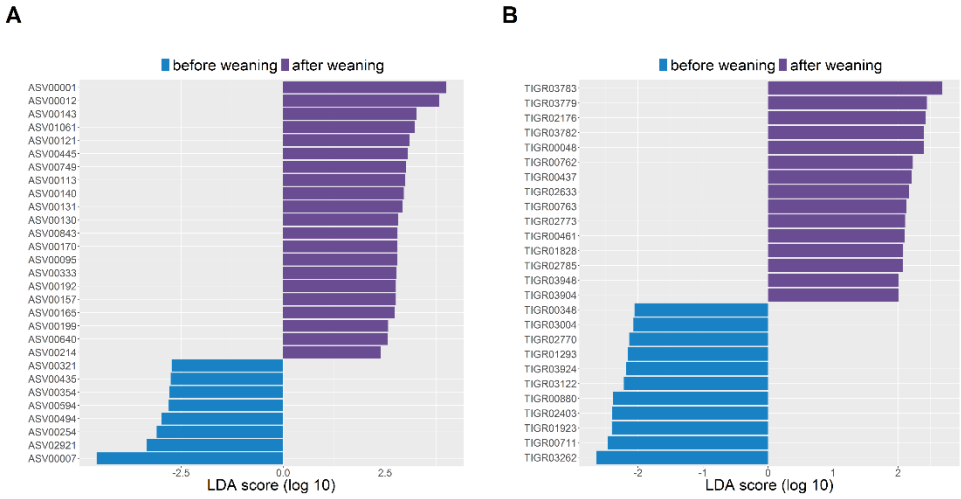


Figure 4.15. Biomarker analysis of samples from toddlers before and after weaning at **(a)** ASV and **(b)** TIGRFAM annotation levels.

4.8 CHANGES IN THE GUT MICROBIOTA AFTER MENARCHE IN ADOLESCENTS

The changes in hormone levels associated to menarche could also affect the composition and function of the gut microbiota, as has been observed for other life events characterized by great hormonal changes, such as pregnancy. For this reason, we have also analyzed separately the samples of female adolescents before (N=18) and after (N=18) the menarche event (Table 4.2 and Table S1 in the Appendix section).

Table 4.2. Time points at which menarche occurred in each female adolescent.

Subject	Menarche time point
AA-5	--
AA-14	T3
AA-17	T10
AA-18	T4
AA-19	T8
AA-20	T6
AA-22	--
AA-23	†
AA-24	T4

-- Menarche occurred before the start of the study.

†Menarche occurred after the end of the study

CCA and ADONIS tests show that samples before and after menarche do not cluster away from each other either at taxonomical or at functional levels (ADONIS test and CCA p-values > 0.05 at ASV, species, genus, TIGRFAM and subrole levels; figures and data not shown). In terms of diversity, both the Shannon index and the Chao1 estimators increase significantly after menarche (Figure 4.16; p-values = 0.00018 and 0.022 respectively).

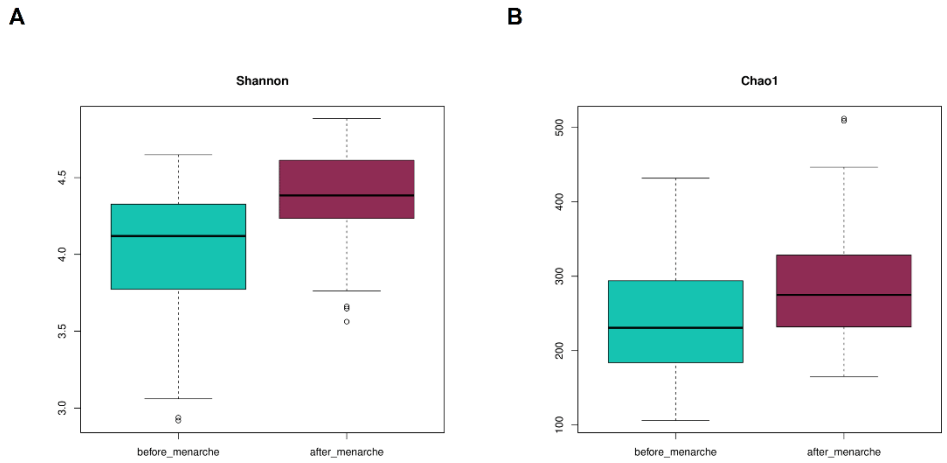


Figure 4.16. Alpha diversity of samples from adolescents before and after menarche measured using **(a)** Shannon index and **(b)** Chao1.

LEfSe analysis (Figure 4.17) shows that ASVs annotated as *Coprococcus catus* (ASV00102), Clostridium sensu stricto 1 NA (ASV00203) and *Veillonella* NA (ASV00234), all of them within the phylum of Firmicutes, are over-represented before menarche in female adolescents, while ASVs annotated as *Butyricicoccus faecihominis*, *Adlercreutzia equolifaciens* and *Desulfovibrionaceae* NA are over-represented after menarche (Figure 4.17a). At functional level (Figure 4.17b), TIGRFAMs that are over-represented before menarche belong mainly in the "transport and binding proteins" category (TIGR00792: subrole "carbohydrates, organic alcohols, and acids"; TIGR00836:

subrole “cations and iron carrying compounds”; TIGR02141: subrole “anions”), while a single TIGRFAM annotation belongs to the “cell envelope” category (TIGR01181: subrole “biosynthesis and degradation of surface polysaccharides and lipopolysaccharides”). A more heterogeneous group of TIGRFAM annotations are over-represented after menarche, belonging to a variety of functional categories and subroles: “DNA metabolism/DNA replication, recombination, and repair” (TIGR00630), “purines, pyrimidines, nucleosides, and nucleotides/salvage of nucleosides and nucleotides” (TIGR00235), “transcription/degradation of RNA” (TIGR03591), and “energy metabolism/electron transport” (TIGR01944).

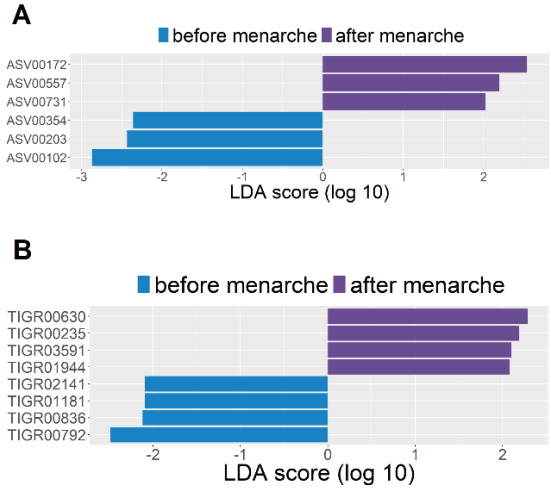


Figure 4.17. LEfSe analyses of differences between samples of adolescents before and after menarche at **(a)** ASV and **(b)** TIGRFAM annotation levels.

5 DISCUSSION

The structure of the human gut microbiota has been extensively studied in healthy adult individuals and infants, as well as in numerous health/disease conditions, and it has been thought that its adult composition is established around 1-4 years of age (Ellis-Pegler *et al.*, 1975). Recently, some studies have suggested that the microbiota of adolescents and pre-adolescents could continue to develop at this age, but few studies have been carried out in this age group and most of them are associated with some disorder or disease. In addition, most of them are based on the relative taxonomic composition of the microbiota, so they lack information on gene composition and functional capacity of the gut microbiome.

In the present study we have analyzed the changes that occur in the intestinal microbiota with development to better characterize and distinguish the healthy microbiota of children, adolescents, and adults. To this end, we have used fecal samples from three well-defined age groups and analyzed 16S rRNA gene amplicons and metagenomes over time. We have compared diversity, composition, functional profile, stability, and dynamics of the

microbiota in toddlers, adolescents, and adults, and we have identified biomarkers from each of the three age groups. At the taxonomic level, toddlers show a microbiota with less diversity than adolescents and adults at the beginning of the study, but diversity increases to levels similar to those of adolescents and adults by the time they are 28 months of age in average. In terms of taxonomic composition, differences among toddlers and adults are still substantial enough to be detected even at the level of phyla, with an over-representation of Actinobacteriota and Proteobacteria. *Bifidobacterium* is one of the main taxa that distinguish toddlers from adults, but, interestingly, we cannot consider it uniquely characteristic of toddlers since it is still abundant in adolescents. Regarding the Firmicutes phylum, although not over- or under-represented overall in toddlers, it is interesting to note that many taxa do have significantly different abundances in this age group, with, among others, over-representations of *Veillonella*, *Intestinibacter* and various Erysipelotrichaceae taxa and under-representations of numerous members of the Lachnospiraceae and Oscillospiraceae, including one of the ASVs of *Faecalibacterium prausnitzii*. The abundance of some of these taxa is likely related to dietary factors. For instance, *Erysipelatoclostridium* has been

positively associated with dairy intake (Smith-Brown *et al.*, 2016) and our toddler group is characterized by the fact that many of its members still complement their diet with breast milk. In addition, it is also noteworthy that for some taxa, although overall relative abundance does not differ in toddlers, some particular species are over- or under-represented in this group. This is the case for Bacteroides, in which *B. fragilis* and *B. ovatus* are over-represented while *B. coprocola*, *B. eggerthii* and *B. massiliensis* are under-represented in relation to adults. This pattern suggests age specialization at the species level.

The intestinal microbiota of adolescents appears to be in an intermediate stage of development in terms of taxonomic composition, as shown by the high relative abundance of bacteria such as *Bifidobacterium*, which are highly abundant in infants and toddlers (Hopkins *et al.*, 2001; Tannock *et al.*, 2013). Moreover, the fact that we are able to analyze children, adolescents and adults within the same study allows us to define when some compositional changes occur relative to adolescence. Some bacteria have increased or decreased in abundance before this stage, while others have increased or decreased between adolescence and adulthood. For instance, our results

indicate that some microaerophilic or facultatively anerobic bacteria, such as *Veillonella* and the potentially pathogenic *Haemophilus*, which causes infections mainly in childhood, have already decreased in abundance by adolescence. On the other hand, some bacteria of the intestinal microbiota that are considered indicators of health in adulthood have increased to adult-like levels by adolescence, as is the case of *Faecalibacterium prausnitzii*, the most important butyrate-producing bacterium, the decrease of which in the intestinal tract has been related to the triggering of inflammatory processes (Ferreira-Halder *et al.*, 2017). In contrast, bacteria that change in abundance after adolescence include *Streptococcus* and various Lachnospiraceae, which decrease after this period, as well as Ruminococcaceae DTU089 and *Butyricicoccus faecihominis*, which rather increase between adolescence and adulthood. Interestingly, a recent study showed that the relative abundance of the family Ruminococcaceae increased with age, being significantly higher in the elderly population and the highest in a high longevity group (> 90 years old) (Yang *et al.*, 2020), suggesting that members of this family of butyrate-producers may continue to increase throughout life. Finally, in some cases it is during adolescence that some groups of bacteria reach their peak of

abundance, suggesting that these organisms may be particularly adapted to this period. Such bacteria include Lachnospiraceae FCS020 and *Porphyromonas bennonis*. In adults, the abundance of Lachnospiraceae FCS020 has been correlated with exercise training (Verheggen *et al.*, 2021) and with a specific distribution of lipoprotein particle types (Vojinovic *et al.*, 2019), suggesting that the high abundance of these bacteria in adolescents may relate to their activity patterns or to specificities of their lipid metabolism.

In terms of functional composition, the differences between the age groups are less pronounced, with toddlers presenting the greatest difference with respect to the other two. In this group, the microbiome is enriched in energy metabolism functions, including some related to aerobic respiration. Network analyses allowed us to identify that *Bifidobacterium* is associated with some of these over-represented functions. Interestingly, some of the energy metabolism functions decline only after adolescence, such as "biosynthesis and degradation of polysaccharides", "pentose phosphate pathway" and "TCA cycle". In contrast, other functions, including many related to "DNA metabolism", "transcription", or "protein synthesis", among others, which are more abundant in adults than in children, have already increased

before adolescence. Therefore, also at a functional level, adolescence seems to represent a stage of transition towards an adult microbiome, where some functions have attained adult-like relative abundances whereas others are still present at the levels seen in young children.

Studies of the long-term stability of the microbiota are rare due to the difficulty in following the same individual through long periods of time. Here, we have been able to analyze stability throughout a 3-year period in children, adolescents and adults of the same population, at the level of taxonomic composition and functional profile. Our analyses show that the stability of the microbiota is lower at taxonomical than at functional level in all three age groups, indicating the existence of functional redundancy through time in all of them. Overall, when the three age groups were considered over the entire period of sampling, *complexCruncher* detected that toddlers are significantly more unstable than the other two groups only at the level of taxonomic composition (Figure 4.10). Further, local stability analyses revealed that, for both taxonomic composition and functional profile, it is only the first time interval in toddlers (T1-T2, between 20 and 24 months in average) that shows a drastically lower stability, while after this point stability levels are similar to

those of adolescents and adults. A supervised learning regressor approach allowed us to identify those ASVs that changed drastically in abundance in toddlers between T1 and T2 (Figure 4.13), belonging to *Blautia obeum* and *Streptococcus*. These results indicate that the microbiota of the toddlers in our study stabilized at around two years of age, after which the overall amount of change per unit of time was not significantly higher than that in the other age groups, neither in taxonomic composition, nor in functional profile.

The range of ages sampled in our study encompass two main events that could strongly influence the composition and stability of the gut microbiota. In the toddlers age group, most participants (9/12) were still consuming breast milk when they entered the study and were weaned during the study period. In the adolescents age group, most female participants (7/9) had not undergone menarche when they entered the study and most of them (6/7) underwent this event during the study. Although our study was not specifically designed to assess the impact of weaning or menarche, we compared the characteristics of the microbiota before and after these events.

Analysis of the intestinal microbiota before and after weaning shows that diversity increases after this event. *Bifidobacterium* is more abundant

before weaning, in accordance with the notion that many species of this genus are adapted to the utilization of breast milk oligosaccharides. In addition, several Firmicutes taxa are also more abundant before weaning. Most of these, such as *Veillonella*, were over-represented in the group of toddlers as a whole when compared to adolescents and adults, indicating that lactating toddlers likely contributed the most to this over-representation. On the other hand, after weaning there is an increase in other members of the Firmicutes, Bacteroidota and Proteobacteria, suggesting a diversification of the microbiota after this event. In particular, numerous ASVs belonging to the family Oscillospiraceae, several members of which were under-represented in toddlers as a whole, increase after weaning, suggesting that, in this case, lactating toddlers likely contributed the most to the under-representation of members of this family. Similarly, at functional level, nearly all functions that are over-represented before weaning were under-represented in toddlers as a whole when compared to adolescents and adults, whereas some of the functions that are over-represented after weaning, i. e. “protein fate/degradation of proteins, peptides and glycopeptides”, “DNA metabolism/DNA replication, recombination and repair”, and “energy

metabolism/amino acids and amines", were under-represented in the entire toddler group. Therefore, weaning contributes to generate a functional profile more similar to that in the older age groups. Overall, these results indicate a very substantial impact of weaning on the diversity, composition, and functional profile of the microbiota of toddlers that have undergone a long period of breastfeeding.

Since we know that major hormonal changes can affect the gut microbiota (Koren *et al.*, 2012), we have attempted to analyze the composition and function of the gut microbiota before and after menarche in adolescents. Due to the small number of samples, we have not been able to carry out reliable statistical tests to support our results. However, we have seen an increase in diversity after menarche, as well as a functional diversification with the increase of a variety of distinct functional groups. More studies are needed to understand the changes that occur in the intestinal microbiota before, during and after this period.

Taken together, our results suggest that, although the intestinal microbiota stabilizes early on during childhood in terms of large-scale, rapid changes in diversity, composition, and functional profile, further changes do

occur both before and after the adolescence period. Therefore, adolescence can still be considered a transitional period, in which some taxa and functions have yet to attain the abundances observed in adults, so that the final stages of development occurring during this time may still be important for the establishment of a healthy adult microbiota. Furthermore, the specificities of the gut microbiota in adolescents should be taken into account when analyzing possible alterations due to disease or other factors in this age group.

6 CONCLUSIONS

1. Toddlers show a microbiota with less diversity than adolescents and adults at the beginning of the study, but diversity increases to levels similar to those of adolescents and adults by the time they are 28 months of age in average.
2. Differences among toddlers and adults are still substantial enough to be detected even at the level of phyla, with an over-representation of Actinobacteriota and Proteobacteria.
3. *Bifidobacterium* is one of the main taxa that distinguishes toddlers from adults, but it is still abundant in adolescents.
4. Within some genera, such as *Bacteroides*, particular species are over- or under-represented in toddlers, suggesting an age specialization at the species level.
5. The intestinal microbiota of adolescents appears to be in an intermediate stage of development in terms of taxonomic composition, as shown by the high relative abundance of bacteria such as *Bifidobacterium*, which are highly abundant in infants and toddlers.

6. Some bacteria have increased or decreased in abundance before adolescence, while others have increased or decreased between adolescence and adulthood.
7. During adolescence some groups of bacteria reach their peak of abundance, such as Lachnospiraceae FCS020 and *Porphyromonas bennonis*, suggesting that these organisms may be particularly adapted to this period.
8. Differences between the age groups are less pronounced at functional level, with toddlers presenting the greatest difference with respect to the other two, being enriched in energy metabolism functions, including some related to aerobic respiration.
9. Some metabolism functions are enriched in both toddlers and adolescents and decline after this period.
10. Functions related to “DNA metabolism”, “transcription”, and “protein synthesis”, among others, are more abundant in adults and adolescents than in toddlers.
11. Adolescence seems to represent a stage of transition towards an adult microbiome also at a functional level, since some functions have

attained adult-like relative abundances whereas others are still present at the levels seen in young children.

12. *Bifidobacterium* contributes significantly to functions over-represented in toddlers.
13. The stability of the microbiota is lower at taxonomical than at functional level in all three age groups, indicating the existence of functional redundancy through time in all of them.
14. Over the entire period of sampling, toddlers are significantly more unstable than the other two groups only at the level of taxonomic composition.
15. The microbiota stabilizes at around two years of age, after which the overall amount of change per unit of time in children is not significantly higher than that in the other age groups, neither in taxonomic composition, nor in functional profile.
16. ASVs that changed drastically in abundance in toddlers between T1 (20 months) and T2 (24 months) belong to *Blautia obeum* and *Streptococcus*.

17. *Bifidobacterium* is more abundant in toddlers before weaning compared to toddlers after weaning, as well as many Firmicutes taxa, such as *Veillonella*, that were over-represented in the group of toddlers as a whole when compared to adolescents and adults.
18. After weaning there is an increase of numerous bacteria belonging to the Bacteroidota, Proteobacteria, and Firmicutes, particularly to the Oscillospiraceae family.
19. Weaning contributes to generate a functional profile more similar to that in adolescents and adults.
20. There is an increase in microbiota diversity after menarche, as well as a functional diversification with increases in the relative abundance of a variety of functions.

Final conclusions:

1. Although the intestinal microbiota stabilizes early on during childhood in terms of large-scale, rapid changes in diversity, composition, and functional profile, further changes do occur both before and after the adolescence period.

2. Adolescence can still be considered a transitional period, in which some taxa and functions have yet to attain the abundances observed in adults, so that the final stages of development occurring during this time may still be important for the establishment of a healthy adult microbiota.
3. Weaning has a substantial impact on the diversity, composition, and functional profile of the microbiota of toddlers that have undergone a long period of breastfeeding, resulting in a convergence towards the characteristics of the adult microbiota.
4. Menarche is accompanied by an increase in diversity in the microbiota, as well as moderate changes in composition and function, but further analyses are needed to better characterize the dynamics of the microbiota in this period.

7 REFERENCES

Aagaard, K. *et al.* (2014) 'The Placenta Harbors a Unique Microbiome', *Science Translational Medicine*, 6(237), pp. 237-265.

Afgan, E. *et al.* (2018) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update', *Nucleic Acids Research*, 46(W1), pp. W537–W544.

Agans, R. *et al.* (2011) 'Distal gut microbiota of adolescent children is different from that of adults', *FEMS Microbiology Ecology*, 77(2), pp. 404–412.

Almeida, A. *et al.* (2019) 'A new genomic blueprint of the human gut microbiota', *Nature*, 568(7753), pp. 499–504.

Almeida, A. *et al.* (2021) 'A unified catalog of 204,938 reference genomes from the human gut microbiome', *Nature Biotechnology*, 39(1), pp. 105–114.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic acids research*, 25(17), pp. 3389–3402.

Andrieux, C. and Sacquet, E. (1983) 'Effect of microflora and lactose on the absorption of calcium, phosphorus and magnesium in the hindgut of the rat', *Reproduction, nutrition, developpement*, 23(2a), pp. 259–271.

Arboleya, S. *et al.* (2011) 'Establishment and development of intestinal microbiota in preterm neonates', *FEMS Microbiology Ecology*, 79(3), pp. 763–772.

Arboleya, S. *et al.* (2012) 'Deep 16S rRNA metagenomics and quantitative PCR analyses of the premature infant fecal microbiota', *Anaerobe*, 18(3), pp. 378–380.

Arboleya, S. *et al.* (2015) 'Intestinal microbiota development in preterm neonates and effect of perinatal antibiotics', *The Journal of pediatrics*, 166(3), pp. 538–544.

Ardissone, A. N. *et al.* (2014) 'Meconium microbiome analysis identifies bacteria correlated with premature birth', *PloS one*, 9(3), pp. e90784.

Arumugam, M. *et al.* (2011) 'Enterotypes of the human gut microbiome'. *Nature*, 473(7346), pp. 174-180.

Arumugam, S., Lau, C. S. M. and Chamberlain, R. S. (2016) 'Probiotics and Synbiotics Decrease Postoperative Sepsis in Elective Gastrointestinal Surgical Patients: a Meta-Analysis', *Journal of Gastrointestinal Surgery* 20(6), pp. 1123–1131.

Azad, M. B. *et al.* (2013) 'Gut microbiota of healthy Canadian infants: Profiles by mode of delivery and infant diet at 4 months', *CMAJ*, 185(5), pp. 385–394.

Bäckhed, F. *et al.* (2004) 'The gut microbiota as an environmental factor that regulates fat storage', *Proceedings of the National Academy of Sciences of the United States of America*, 101(44), p. 15718.

Bäckhed, F. *et al.* (2005) 'Host-bacterial mutualism in the human intestine', *Science* 307(5717), pp. 1915–1920.

Bäckhed, F. *et al.* (2015) 'Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life', *Cell Host & Microbe*, 17(5), pp. 690–703.

Bai, J., Hu, Y. and Bruner, D. W. (2019) 'Composition of gut microbiota and its association with body mass index and lifestyle factors in a cohort of 7–18 years old children from the American Gut Project', *Pediatric Obesity*, 14(4), p. e12480.

Bailey, M. T. *et al.* (2011) 'Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation', *Brain, Behavior, and Immunity*, 25(3), pp. 397–407.

Bates, J. M. *et al.* (2007) 'Intestinal Alkaline Phosphatase Detoxifies Lipopolysaccharide and Prevents Inflammation in Response to the Gut Microbiota', *Cell host & microbe*, 2(6), p. 371.

Bermudez-Brito, M. *et al.* (2012) 'Probiotic Mechanisms of Action', *Annals of Nutrition and Metabolism*, 61(2), pp. 160–174.

Bhattarai, Y., Muniz Pedrogo, D. A. and Kashyap, P. C. (2017) 'Irritable bowel syndrome: a gut microbiota-related disorder?', *American journal of physiology. Gastrointestinal and liver physiology*, 312(1), pp. G52–G62.

Biasucci, G. *et al.* (2008) 'Cesarean Delivery May Affect the Early Biodiversity of Intestinal Bacteria', *The Journal of Nutrition*, 138(9), pp. 1796S-1800S.

Biasucci, G. *et al.* (2010) 'Mode of delivery affects the bacterial community in the newborn gut', *Early Human Development*, 86(1), pp. 13–15.

Björkstén, B. *et al.* (1999) 'The intestinal microflora in allergic Estonian and Swedish 2-year-old children', *Clinical & Experimental Allergy*, 29(3), pp. 342–346.

Bokulich, N. A. *et al.* (2018) 'q2-longitudinal: Longitudinal and paired-sample analyses of microbiome data.' *Msystems*, 3(6), pp. e00219-18.

Bolyen, E. *et al.* (2019) 'Reproducible, interactive, scalable and extensible

microbiome data science using QIIME 2', *Nature Biotechnology* 37(8), pp. 852–857.

Bosch, T. C. G. (2014) 'Rethinking the role of immunity: lessons from Hydra', *Trends in Immunology*, 35(10), pp. 495–502.

Bouskra, D. *et al.* (2008) 'Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis', *Nature*, 456(7221), pp. 507–510.

Brandt, L. J. *et al.* (2012) 'Long-term follow-up of colonoscopic fecal microbiota transplant for recurrent clostridium difficile infection', *American Journal of Gastroenterology*, 107(7), pp. 1079–1087.

Britton, R. A. and Young, V. B. (2014) 'Role of the intestinal microbiota in resistance to colonization by Clostridium difficile', *Gastroenterology*, 146(6), pp. 1547–1553.

Brooks, B. *et al.* (2014) 'Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants', *Microbiome*, 2(1), pp. 1–16.

Burks, A. W. *et al.* (2015) 'Synbiotics-supplemented amino acid-based formula supports adequate growth in cow's milk allergic infants', *Pediatric Allergy and Immunology*, 26(4), pp. 316–322.

Butel, M. J. *et al.* (2007) 'Conditions of bifidobacterial colonization in preterm infants: A prospective analysis', *Journal of Pediatric Gastroenterology and Nutrition*, 44(5), pp. 577–582.

Camacho, C. *et al.* (2009) 'BLAST+: architecture and applications', *BMC bioinformatics*, 10:421.

Quast, C. *et al.* (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic acids research*, 41(Database issue) , pp. D590-6.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583.

Cattaneo, A. *et al.* (2017) 'Association of brain amyloidosis with pro-inflammatory gut bacterial taxa and peripheral inflammation markers in cognitively impaired elderly', *Neurobiology of Aging*, 49, pp. 60–68.

Chao, A. (1984) 'Nonparametric Estimation of the Number of Classes in a Population', *Scandinavian Journal of Statistics,* 11, pp. 265-270.

Chen, G. *et al.* (2017) 'Associations of caesarean delivery and the occurrence of neurodevelopmental disorders, asthma or obesity in childhood based on Taiwan

birth cohort study', *BMJ Open*, 7(9), p. e017086.

Cheng, J. *et al.* (2015) 'Discordant temporal development of bacterial phyla and the emergence of core in the fecal microbiota of young children', *The ISME Journal*, 10(4), pp. 1002–1014.

Del Chierico, F. *et al.* (2015) 'Phylogenetic and Metabolic Tracking of Gut Microbiota during Perinatal Development', *PloS one*, 10(9), p. e0137347.

Cho, I. and Blaser, M. J. (2012) 'The human microbiome: at the interface of health and disease', *Nature reviews. Genetics*, 13(4), pp. 260–270.

Collado, M. *et al.* (2009) 'The Impact of Probiotic on Gut Health', *Current Drug Metabolism*, 10(1), pp. 68–78.

Collado, M. C. *et al.* (2013) 'Human microbiome and diseases: a metagenomic approach', *Bioactive Food as Dietary Interventions for Liver and Gastrointestinal Disease*, Chapter 15 (Academic Press San Diego), pp. 235–249.

Collins, S. M. and Bercik, P. (2009) 'The relationship between intestinal microbiota and the central nervous system in normal gastrointestinal function and disease', *Gastroenterology*, 136(6), pp. 2003–2014.

Conway, P. (1997) 'Development of intestinal microbiota', *Gastrointestinal microbiology*. 2 (Chapman and Hall London UK), pp. 3-38.

Corr, S. C., Gahan, C. G. M. and Hill, C. (2007) 'Impact of selected *Lactobacillus* and *Bifidobacterium* species on *Listeria monocytogenes* infection and the mucosal immune response', *FEMS immunology and medical microbiology*, 50(3), pp. 380–388.

Costea, P. I. *et al.* (2018) 'Enterotypes in the landscape of gut microbial community composition', *Nature Microbiology*, 3(1), pp. 8–16.

Csardi, G. and Nepusz, T. (2006) 'The igraph software package for complex network research', *InterJournal, Complex Systems*. Available at: http://igraph.org (Accessed: 15 November 2021).

David, L. A. *et al.* (2013) 'Diet rapidly and reproducibly alters the human gut microbiome', *Nature*, 505(7484), pp. 559–563.

Davila, A. M. *et al.* (2013) 'Intestinal luminal nitrogen metabolism: role of the gut microbiota and consequences for the host', *Pharmacological research*, 68(1), pp. 95–107.

De Filippo, C. *et al.* (2010) 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa', *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), pp. 14691–14696.

Dethlefsen, L. and Relman, D. A. (2011) 'Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.', *Proceedings of the National Academy of Sciences of the United States of America*, 108 (Suppl 1), pp. 4554–61.

Ding, T. and Schloss, P. D. (2014) 'Dynamics and associations of microbial community types across the human body', *Nature*, 509(7500), pp. 357–360.

Dominguez-Bello, M. G. *et al.* (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), pp. 11971–11975.

Dominguez-Bello, M. G. *et al.* (2016) 'Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer', *Nature Medicine*, 22(3). pp. 250-253.

Dowlati, Y. *et al.* (2010) 'A meta-analysis of cytokines in major depression', *Biological psychiatry*, 67(5), pp. 446–457.

Duncan, S. H., Louis, P. and Flint, H. J. (2004) 'Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product', *Applied and Environmental Microbiology*, 70(10), pp. 5810–5817.

Isolauri, E. *et al.* (2008) 'Modulation of the maturing gut barrier and microbiota: a novel target in allergic disease', *Current pharmaceutical design*, 14(14), pp. 146–147.

Eckburg, P. B. *et al.* (2005) 'Diversity of the human intestinal microbial flora', *Science*, 308(5728), pp. 1635–1638.

Ellis-Pegler, R. B., Crabtree, C. and Lambert, H. P. (1975) 'The faecal flora of children in the United Kingdom', *The Journal of hygiene*, 75(1), pp. 135–142.

Faith, J. J. *et al.* (2013) 'The long-term stability of the human gut microbiota', *Science*, 341(6141), p. 1237439.

Fallani, M. *et al.* (2010) 'Intestinal microbiota of 6-week-old infants across Europe: Geographic influence beyond delivery mode, breast-feeding, and antibiotics',

Journal of Pediatric Gastroenterology and Nutrition, 51(1), pp. 77–84.

Ferreira-Halder, C. V., Faria, A. V. de S. and Andrade, S. S. (2017) 'Action and function of *Faecalibacterium prausnitzii* in health and disease', *Best Practice & Research Clinical Gastroenterology*, 31(6), pp. 643–648.

Fetissov, S. O. *et al.* (2008) 'Emerging role of autoantibodies against appetite-regulating neuropeptides in eating disorders', *Nutrition*, 24(9), pp. 854–859.

Finegold, S. M. *et al.* (2002) 'Gastrointestinal Microflora Studies in Late-Onset Autism', *Clinical Infectious Diseases*, 35(Supplement_1), pp. S6–S16.

Finegold, S. M. (2011) '*Desulfovibrio* species are potentially important in regressive autism', *Medical Hypotheses*, 77(2), pp. 270–274.

Finegold, S. M., Downes, J. and Summanen, P. H. (2012) 'Microbiology of regressive autism', *Anaerobe*, 18(2), pp. 260–262.

Fink, L. N. *et al.* (2007) 'Distinct gut-derived lactic acid bacteria elicit divergent dendritic cell-mediated NK cell responses', *International immunology*, 19(12), pp. 1319–1327.

Finn, R. D., Clements, J. and Eddy, S. R. (2011) 'HMMER web server: interactive sequence similarity searching', *Nucleic Acids Research*, 39(suppl_2), pp. W29–W37.

Food and Agriculture Organization of the United Nations World Health Organization (2002) 'Guidelines for the Evaluation of Probiotics in Food', *Joint Fao/Who Working Group on Drafting Guidelines for the Evaluation of Probiotics in Food*.

Forsythe, P. *et al.* (2010) 'Mood and gut feelings', *Brain, behavior, and immunity*, 24(1), pp. 9–16.

Foster, J. A. and McVey Neufeld, K. A. (2013) 'Gut-brain axis: how the microbiome influences anxiety and depression', *Trends in neurosciences*, 36(5), pp. 305–312.

Frank, D. N. *et al.* (2007) 'Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases', *Proceedings of the National Academy of Sciences*, 104(34), pp. 13780–13785.

García-Martínez, J. *et al.* (1999) 'Use of the 16S--23S ribosomal genes spacer region in studies of prokaryotic diversity', *Journal of microbiological methods*, 36(1–2), pp. 55–64.

Gibson, G. R. *et al.* (2017) 'Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the

definition and scope of prebiotics', *Nature Reviews Gastroenterology & Hepatology*, 14(8), pp. 491–502.

Gill, S. R. *et al.* (2006) 'Metagenomic Analysis of the Human Distal Gut Microbiome', *Science*, 312(5778), p. 1355.

Gordon, J. H. and Dubos, R. (1970) 'The anaerobic bacterial flora of the mouse cecum', *Journal of Experimental Medicine*, 132(2), pp. 251–260.

Gosalbes, M. J. *et al.* (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 : journal of the British Society for Allergy and Clinical Immunology*, 43(2), pp. 198–211.

Gough, E., Shaikh, H. and Manges, A. R. (2011) 'Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection', *Clinical Infectious Diseases*, 53(10), pp. 994–1002.

Greenhalgh, K. *et al.* (2016) 'The human gut microbiome in health: establishment and resilience of microbiota over a lifetime', *Environmental Microbiology*, 18(7), pp. 2103–2116.

Grönlund, M. M. *et al.* (1999) 'Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery', *Journal of Pediatric Gastroenterology & Nutrition*, 8(1), pp. 19–25.

Haft, D. H. *et al.* (2001) 'TIGRFAMs: a protein family resource for the functional identification of proteins', *Nucleic Acids Research*, 29(1), p. 41.

Hamady, M. *et al.* (2008) 'Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex', *Nature methods*, 5(3), pp. 235–237.

Hansen, R. *et al.* (2012) 'Microbiota of de-novo pediatric IBD: Increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis', *American Journal of Gastroenterology*, 107(12), pp. 1913–1922.

Harmsen, H. J. M. *et al.* (2002) 'Extensive set of 16S rRNA-based probes for detection of bacteria in human feces', *Applied and environmental microbiology*, 68(6), pp. 2982–2990.

Hemarajata, P. and Versalovic, J. (2013) 'Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation', *Therapeutic advances in gastroenterology*, 6(1), pp. 39–51.

Hill, M. J. (1997) 'Intestinal flora and endogenous vitamin synthesis', *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP)*, 6 (Suppl. 1), pp. S43-5.

Hodkinson, B. P. and Grice, E. A. (2015) 'Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research', *Advances in Wound Care*, 4(1), p. 50.

Holling, C. S. (1973) 'Resilience and stability of ecological systems', *Annual Review of Ecology and Systematics*, 4, pp. 1–23.

Hollister, E. B. *et al.* (2015) 'Structure and function of the healthy pre-adolescent pediatric gut microbiome', *Microbiome*, 3, p. 36.

Hopfner, F. *et al.* (2017) 'Gut microbiota in Parkinson disease in a northern German cohort', *Brain Research*, 1667, pp. 41–45.

Hopkins, M. J. (2001) 'Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles', *Gut*, 48(2), pp. 198–205.

Hopkins, M. J., Sharp, R. and Macfarlane, G. T. (2001) 'Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles', *Gut*, 48(2), p. 198.

Huttenhower, C. *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*, 486(7402), pp. 207–214.

Hyatt, D. *et al.* (2010) 'Prodigal: Prokaryotic gene recognition and translation initiation site identification', *BMC Bioinformatics*, 11(1), pp. 1–11.

Iizumi, T. *et al.* (2017) 'Gut Microbiome and Antibiotics', *Archives of medical research*, 48(8), pp. 727–734.

Ivanov, I. I. *et al.* (2008) 'Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine', *Cell host & microbe*, 4(4), pp. 337–349.

Jakobsson, H. E. *et al.* (2014) 'Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section', *Gut*, 63(4), pp. 559–566.

Jemal, A. *et al.* (2011) 'Global cancer statistics', *CA: A Cancer Journal for Clinicians*, 61(2), pp. 69–90.

Jiménez, E. *et al.* (2005) 'Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section', *Current Microbiology*, 51(4), pp. 270–274.

Jiménez, E. *et al.* (2008) 'Is meconium from healthy newborns actually sterile?', *Research in microbiology*, 159(3), pp. 187–193.

Gebbers, J. O. and Laissue, J. A. (1989) 'Immunologic structures and functions of the gut.', *Swiss archive for veterinary medicine*, 131(5), pp. 221-38.

Johansson, M. E. V. *et al.* (2008) 'The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), pp. 15064–15069.

Joossens, M. *et al.* (2011) 'Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives', *Gut*, 60(5), pp. 631–637.

Kalliomäki, M. *et al.* (2001) 'Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing.', *The Journal of Allergy and Clinical Immunology*, 107(1), pp. 129–134.

Kelly, C. P. and LaMont, J. T. (2008) '*Clostridium difficile*--more difficult than ever', *The New England journal of medicine*, 359(18), pp. 1932–1940.

Kelly, C. R. *et al.* (2015) 'Update on fecal microbiota transplantation 2015: indications, methodologies, mechanisms, and outlook', *Gastroenterology*, 149(1), pp. 223–237.

Kelly, D. *et al.* (2004) 'Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA', *Nature immunology*, 5(1), pp. 104–112.

Knecht, H. *et al.* (2014) 'Effects of b-lactam antibiotics and fluoroquinolones on human gut microbiota in relation to *Clostridium difficile* associated diarrhea', *PLoS ONE*, 9(2), p. e89417.

Koenig, J. E. *et al.* (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 1), pp. 4578–85.

Konstantinov, S. R., Kuipers, E. J. and Peppelenbosch, M. P. (2013) 'Functional genomic analyses of the gut microbiota for CRC screening', *Nature Reviews Gastroenterology & Hepatology*, 10(12), pp. 741–745.

Koren, O. *et al.* (2012) 'Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy', *Cell*, 150(3), pp. 470–480.

Krogius-Kurikka, L. *et al.* (2009) 'Microbial community analysis reveals high level phylogenetic alterations in the overall gastrointestinal microbiota of diarrhoea-predominant irritable bowel syndrome sufferers', *BMC Gastroenterology*, 9(1), pp. 1–11.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature methods*, 9(4), pp. 357–359.

La Rosa, P. S. *et al.* (2014) 'Patterned progression of bacterial populations in the premature infant gut', *Proceedings of the National Academy of Sciences of the United States of America*, 111(34), pp. 12522–12527.

Larsen, N. *et al.* (2010) 'Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults', *PLoS ONE*, 5(2), p. e9085.

LaTuga, M. S., Stuebe, A. and Seed, P. C. (2014) 'A review of the source and function of microbiota in breast milk', *Seminars in reproductive medicine*, 32(1), pp. 68–73.

Lax, S. *et al.* (2014) 'Longitudinal analysis of microbial interaction between humans and the indoor environment', *Science*, 345(6200), pp. 1048–1052.

Ley, R. E., Peterson, D. A. and Gordon, J. I. (2006) 'Ecological and evolutionary forces shaping microbial diversity in the human intestine', *Cell*, 124(4), pp. 837–848.

Li, D. *et al.* (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics*, 31(10), pp. 1674–1676.

Li, F., Hullar, M. A. J. and Lampe, J. W. (2007) 'Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota', *Journal of microbiological methods*, 68(2), pp. 303–311.

Li, Q. *et al.* (2017) 'The gut microbiota and autism spectrum disorders', *Frontiers in Cellular Neuroscience*, 11, p. 120.

Liévin-Le Moal, V. and Servin, A. L. (2006) 'The front line of enteric host defense against unwelcome intrusion of harmful microorganisms: mucins, antimicrobial peptides, and microbiota', *Clinical Microbiology Reviews*, 19(2), p. 315.

Lockhart, P. B. *et al.* (2008) 'Bacteremia associated with toothbrushing and dental extraction', *Circulation*, 117(24), pp. 3118–3125.

Lozupone, C. A. *et al.* (2012) 'Diversity, stability and resilience of the human gut microbiota.', *Nature*, 489(7415), pp. 220–30.

Macfarlane, S. and Macfarlane, G. T. (2003) 'Regulation of short-chain fatty acid production', *The Proceedings of the Nutrition Society*, 62(1), pp. 67–72.

Machiels, K. *et al.* (2014) 'A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis', *Gut*, 63(8), pp. 1275–1283.

Macpherson, A. J. and Uhr, T. (2004) 'Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria', *Science,* 303(5664), pp. 1662–1665.

Magoč, T. and Salzberg, S. L. (2011) 'FLASH: fast length adjustment of short reads to improve genome assemblies', *Bioinformatics*, 27(21), pp. 2957–2963.

Mai, V. *et al.* (2011) 'Fecal microbiota in premature infants prior to necrotizing enterocolitis', *PloS One*, 6(6), p. e20647.

Marcobal, A. *et al.* (2010) 'Consumption of human milk oligosaccharides by gut-related microbes', *Journal of agricultural and food chemistry*, 58(9), pp. 5334–5340.

Margulies, M. *et al.* (2005) 'Genome sequencing in microfabricated high-density picolitre reactors', *Nature*, 437(7057), pp. 376–380.

Mariat, D. *et al.* (2009) 'The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age', *BMC microbiology*, 9, p. 123.

Markowiak, P. and Ślizewska, K. (2017) 'Effects of probiotics, prebiotics, and synbiotics on human health', *Nutrients*, 9(9), p. 1021.

Martí, J. M. *et al.* (2017) 'Health and disease imprinted in the time variability of the human microbiome', *mSystems*, 2(2), pp. e00144-16.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12.

Mazmanian, S. K. *et al.* (2005) 'An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system', *Cell*, 122(1), pp. 107–118.

Medellin-Peña, M. J. and Griffiths, M. W. (2008) 'Effect of molecules secreted by *Lactobacillus acidophilus* strain La-5 on *Escherichia coli* O157:H7 colonization', *Applied and Enrironmental Microbiology*, 75(4), pp. 1165–1172.

De Mello, R. M. P. *et al.* (2009) '*Lactobacilli* and bifidobacteria in the feces of

schoolchildren of two different socioeconomic groups: children from a favela and children from a private school', *Jornal de Pediatria*, 85(4), pp. 307–314.

Menzel, P., Ng, K. L. and Krogh, A. (2016) 'Fast and sensitive taxonomic classification for metagenomics with Kaiju', *Nature Communications*, 7(1), pp. 1–9.

Miller, L. E., Ouwehand, A. C. and Ibarra, A. (2017) 'Effects of probiotic-containing products on stool frequency and intestinal transit in constipated adults: systematic review and meta-analysis of randomized controlled trials', *Annals of gastroenterology*, 30(6), pp. 629–639.

Mor, G. and Cardenas, I. (2010) 'The immune system in pregnancy: a unique complexity', *American journal of reproductive immunology*, 63(6), p. 425.

Murray, C. S. *et al.* (2005) 'Fecal microbiota in sensitized wheezy and non-sensitized non-wheezy children: a nested case-control study', *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, 35(6), pp. 741–745.

Musso, G., Gambino, R. and Cassader, M. (2011) 'Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes', *Annual review of medicine*, 62, pp. 361–380.

Nafarin, A. R. *et al.* (2019) 'Gut microbiome pattern in adolescents with functional gastrointestinal disease.', *International journal of pediatrics & adolescent medicine*, 6(1), pp. 12–15.

Neish, A. S. (2009) 'Microbes in gastrointestinal health and disease', *Gastroenterology*, 136(1), p. 65.

Neufeld, K. A. and Foster, J. A. (2009) 'Effects of gut microbiota on the brain: implications for psychiatry', *Journal of Psychiatry & Neuroscience : JPN*, 34(3), p. 230.

Newbern, D. and Freemark, M. (2011) 'Placental hormones and the control of maternal metabolism and fetal growth', *Current opinion in endocrinology, diabetes, and obesity*, 18(6), pp. 409–416.

Nikbakht, E. *et al.* (2018) 'Effect of probiotics and synbiotics on blood glucose: a systematic review and meta-analysis of controlled trials', *European journal of nutrition*, 57(1), pp. 95–106.

O'Grady, J., O'Connor, E. M. and Shanahan, F. (2019) 'Review article: dietary fibre in the era of microbiome science', *Alimentary Pharmacology & Therapeutics*, 49(5),

pp. 506–515.

O'Toole, P. W., Marchesi, J. R. and Hill, C. (2017) 'Next-generation probiotics: the spectrum from probiotics to live biotherapeutics', *Nature Microbiology*, 2(5), pp. 1–6.

Ochman, H. *et al.* (2010) 'Evolutionary relationships of wild hominids recapitulated by gut microbial communities', *PLOS Biology*, 8(11), p. e1000546.

Oksanen, J. *et al* (2019) 'vegan: Community Ecology Package. R package version 2.5-6. ', https://CRAN.R-project.org/package=vegan

Ouwehand, A. C. *et al.* (2000) 'Adhesion of inactivated probiotic strains to intestinal mucus', *Letters in applied microbiology*, 31(1), pp. 82–86.

Paliy, O. *et al.* (2009) 'High-throughput quantitative analysis of the human intestinal microbiota with a phylogenetic microarray.', *Applied and environmental microbiology*, 75(11), pp. 3572–9.

Pantoja-Feliciano, I. G. *et al.* (2013) 'Biphasic assembly of the murine intestinal microbiota during early development', *The ISME Journal*, 7(6), pp. 1112–1115.

Parks, D. H. *et al.* (2021) 'GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy', *Nucleic acids research*, 50(D1), pp. D785-D794.

Pasolli, E. *et al.* (2019) 'Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle', *Cell*, 176(3), pp. 649-662.

Penders, J. *et al.* (2006a) 'Factors influencing the composition of the intestinal microbiota in early infancy', *Pediatrics*, 118(2), pp. 511–521.

Penders, J. *et al.* (2006b) 'Molecular fingerprinting of the intestinal microbiota of infants in whom atopic eczema was or was not developing', *Clinical and Experimental Allergy*, 36(12), pp. 1602–1608.

Penders, J. *et al.* (2007) 'The role of the intestinal microbiota in the development of atopic disorders', *Allergy*, 62(11), pp. 1223–1236.

Pérez-Cobas, A. E. *et al.* (2013) 'Differential effects of antibiotic therapy on the structure and function of human gut microbiota', *PloS One*, 8(11), p. e80201.

Perez, P. F. *et al.* (2007) 'Bacterial imprinting of the neonatal immune system: Lessons from maternal cells?', *Pediatrics*, 119(3), pp. e724-32.

Peterson, D. A. *et al.* (2007) 'IgA response to symbiotic bacteria as a mediator of gut homeostasis', *Cell Host & Microbe*, 2(5), pp. 328–339.

Petrof, E. O. *et al.* (2013) 'Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: "RePOOPulating" the gut', *Microbiome*, 1(1), p. 3.

Qin, J. *et al.* (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*, 464(7285), pp. 59–65.

Qin, J. *et al.* (2012) 'A metagenome-wide association study of gut microbiota in type 2 diabetes', *Nature*, 490(7418), pp. 55–60.

R Core Team (2018) *R: A language and environment for statistical computing*, *R Foundation for Statistical Computing*. Available at: https://www.r-project.org/ (Accessed: 12 November 2021).

Rautava, S. *et al.* (2004) 'The hygiene hypothesis of atopic disease--an extended version', *Journal of pediatric gastroenterology and nutrition*, 38(4), pp. 378–388.

Rautava, S. *et al.* (2012) 'Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial.', *Neonatology*, 102(3), pp. 178–184.

Ravcheev, D. A. *et al.* (2013) 'Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: Comparative genomics reconstruction of metabolic and regulatory networks', *BMC Genomics*, 14(1), p. 873.

Rhee, S. H., Pothoulakis, C. and Mayer, E. A. (2009) 'Principles and clinical implications of the brain-gut-enteric microbiota axis', *Nature reviews. Gastroenterology & hepatology*, 6(5), pp. 306–314.

Rizzatti, G. *et al.* (2017) 'Proteobacteria: A common factor in human diseases', *BioMed Research International*, 2017, pp. 1-7.

Roberfroid, M. *et al.* (2010) 'Prebiotic effects: metabolic and health benefits', *British Journal of Nutrition*, 104(S2), pp. S1–S63.

Roger, L. C. *et al.* (2010) 'Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life', *Microbiology*, 156(11), pp. 3329–3341.

Rohart, F. *et al.* (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLoS computational biology*, 13(11), p. e1005752.

Romagnani, S. (2004) 'The increased prevalence of allergy and the hygiene

hypothesis: missing immune deviation, reduced immune suppression, or both?', *Immunology*, 112(3), pp. 352–363.

Rosenberg, E. and Zilber-Rosenberg, I. (2013) 'Microbiotas are transmitted between holobiont generations', *The Hologenome Concept: Human, Animal and Plant Microbiota,* Chapter (Springer), pp. 41–54.

Salminen, S. *et al.* (2004) 'Influence of mode of delivery on gut microbiota composition in seven year old children', *Gut*, 53(9), p. 1388.

Salonen, A., De Vos, W. M. and Palva, A. (2010) 'Gastrointestinal microbiota in irritable bowel syndrome: Present state and perspectives', *Microbiology*, 156(11), pp. 3205–3215.

Salzman, N. H., Underwood, M. A. and Bevins, C. L. (2007) 'Paneth cells, defensins, and the commensal microbiota: a hypothesis on intimate interplay at the intestinal mucosa', *Seminars in immunology*, 19(2), pp. 70–83.

Sánchez, B. *et al.* (2017) 'Probiotics, gut microbiota, and their influence on host health and disease', *Molecular nutrition & food research*, 61(1).

Sanders, M. E. *et al.* (2019) 'Probiotics and prebiotics in intestinal health and disease: from biology to the clinic', *Nature reviews. Gastroenterology & hepatology*, 16(10), pp. 605–616.

Saulnier, D. M. *et al.* (2011) 'Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome', *Gastroenterology*, 141(5), p. 1782.

Savage, D. C. (1970) 'Associations of Indigenous Microorganisms with Gastrointestinal Mucosal Epithelia', *The American Journal of Clinical Nutrition*, 23(11), pp. 1495–1501.

Schanche, M. *et al.* (2015) 'High-resolution analyses of overlap in the microbiota between mothers and their children', *Current Microbiology*, 71(2), pp. 283–290.

Schmieder, R. and Edwards, R. (2011) 'Quality control and preprocessing of metagenomic datasets', *Bioinformatics*, 27(6), pp. 863–864.

Schnorr, S. L. *et al.* (2014) 'Gut microbiome of the Hadza hunter-gatherers', *Nature Communications*, 5(1), pp. 1–12.

Segata, N. *et al.* (2011) 'Metagenomic biomarker discovery and explanation', *Genome biology*, 12(6), p. R60.

Sekirov, I. *et al.* (2010) 'Gut microbiota in health and disease', *Physiological reviews*,

90(3), pp. 859–904.

Sender, R., Fuchs, S. and Milo, R. (2016) 'Revised estimates for the number of human and bacteria cells in the body.', *PLoS biology*, 14(8), p. e1002533.

Sevelsted, A. *et al.* (2015) 'Cesarean section and chronic immune disorders', *Pediatrics*, 135(1), pp. e92–e98.

Shannon, C. E. (1948) 'A mathematical theory of communication', *Bell System Technical Journal*, 27(3), pp. 379–423.

Shannon, P. *et al.* (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome research*, 13(11), pp. 2498–2504.

Shin, N-R., Whon, T. W. and Bae, J-W. (2015) 'Proteobacteria: microbial signature of dysbiosis in gut microbiota', *Trends in biotechnology*, 33(9), pp. 496–503.

Singh, A. *et al.* (2016) 'DIABLO: from multi-omics assays to biomarker discovery, an integrative approach', *bioRxiv*, 35, pp. 3055-3062.

Singh, V. P., Proctor, S. D. and Willing, B. P. (2016) 'Koch's postulates, microbial dysbiosis and inflammatory bowel disease', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 22(7), pp. 594–599.

Sjögren, Y. M. *et al.* (2009) 'Altered early infant gut microbiota in children developing allergy up to 5 years of age', *Clinical & Experimental Allergy*, 39(4), pp. 518–526.

Slavin, J. (2013) 'Fiber and prebiotics: mechanisms and health benefits', *Nutrients 2013*, 5(4), pp. 1417–1435.

Smith-Brown, P. *et al.* (2016) 'Mothers secretor status affects development of childrens microbiota composition and function: a pilot study', *PLoS One*, 11(9), p. e0161211.

Smits, L. P. *et al.* (2013) 'Therapeutic potential of fecal microbiota transplantation', *Gastroenterology*, 145(5), pp. 946–953.

Sokol, H. *et al.* (2008) '*Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients', *Proceedings of the National Academy of Sciences*, 105(43), pp. 16731–16736.

Steel, J. H. *et al.* (2005) 'Bacteria and inflammatory cells in fetal membranes do not

always cause preterm labor', *Pediatric research*, 57(3), pp. 404–411.

Tamboli, C. P. *et al.* (2004) 'Dysbiosis in inflammatory bowel disease', *Gut*, 53(1), pp. 1–4.

Tannock, G. W. *et al.* (2013) 'Comparison of the compositions of the stool microbiotas of infants fed goat milk formula, cow milk-based formula, or breast milk', *Applied and Environmental Microbiology*, 79(9), p. 3040.

Thompson, A. L. *et al.* (2015) 'Milk- and solid-feeding practices and daycare attendance are associated with differences in bacterial diversity, predominant communities, and metabolic and immune function of the infant gut microbiome', *Frontiers in Cellular and Infection Microbiology*, 5, p. 3.

Tsai, F. and Coyle, W. J. (2009) 'The microbiome and obesity: Is obesity linked to our gut flora?', *Current Gastroenterology Reports*, 11(4), pp. 307–313.

Tsuji, M. *et al.* (2008) 'Dynamic interactions between bacteria and immune cells leading to intestinal IgA synthesis', *Seminars in immunology*, 20(1), pp. 59–66.

Turnbaugh, P. J. and Gordon, J. I. (2009) 'The core gut microbiome, energy balance and obesity', *The Journal of physiology*, 587(17), pp. 4153–4158.

Tyakht, A. V. *et al.* (2013) 'Human gut microbiota community structures in urban and rural populations in Russia', *Nature Communications,* 4(1), pp. 1–9.

Underwood, M. A. *et al.* (2015) 'Human milk oligosaccharides in premature infants: absorption, excretion, and influence on the intestinal microbiota', *Pediatric Research,* 78(6), pp. 670–677.

Van Der Aa, L. B. *et al.* (2011) 'Synbiotics prevent asthma-like symptoms in infants with atopic dermatitis', *Allergy*, 66(2), pp. 170–177.

Van Tongeren, S. P. *et al.* (2005) 'Fecal microbiota composition and frailty', *Applied and Environmental Microbiology*, 71(10), p. 6438.

Vella, A. and Farrugia, G. (1998) 'D-lactic acidosis: pathologic consequence of saprophytism', *Mayo Clinic proceedings*, 73(5), pp. 451–456.

Verberkmoes, N. C. *et al.* (2009) 'Shotgun metaproteomics of the human distal gut microbiota', *The ISME Journal*, 3(2), pp. 179–189.

Verheggen, R. J. H. M. *et al.* (2021) 'Eight-week exercise training in humans with obesity: Marked improvements in insulin sensitivity and modest changes in gut microbiome', *Obesity (Silver Spring, Md.)*, 29(10), pp. 1615–1624.

Vijay-Kumar, M. *et al.* (2010) 'Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5', *Science*, 328(5975), pp. 228–231.

Vogt, N. M. *et al.* (2017) 'Gut microbiome alterations in Alzheimer's disease', *Scientific Reports*, 7(1), pp. 1–11.

Vojinovic, D. *et al.* (2019) 'Relationship between gut microbiota and circulating metabolites in population-based cohorts', *Nature communications*, 10(1), p. 5813.

Voreades, N., Kozil, A. and Weir, T. L. (2014) 'Diet and the development of the human intestinal microbiome', *Frontiers in microbiology*, 5(494), pp. 1-9.

Vyas, U. and Ranganathan, N. (2012) 'Probiotics, prebiotics, and synbiotics: Gut and beyond', *Gastroenterology Research and Practice*, 2012(872716), pp. 1-16.

Wang, L. *et al.* (2011) 'Low relative abundances of the mucolytic Bacterium *Akkermansia muciniphila* and *Bifidobacterium* spp. in feces of children with autism', *Applied and Environmental Microbiology*, 77(18), pp. 6718–6721.

Wang, T. *et al.* (2011) 'Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers', *The ISME Journal*, 6(2), pp. 320–329.

Wang, Y. *et al.* (2009) '16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis', *The ISME Journal*, 3(8), pp. 944–954.

Wegh, C. A. M. *et al.* (2017) 'The effect of fiber and prebiotics on children's gastrointestinal disorders and microbiome', *Expert review of gastroenterology & hepatology*, 11(11), pp. 1031–1045.

Wegh, C. A. M. *et al.* (2019) 'Postbiotics and their potential applications in early life nutrition and beyond', *International Journal of Molecular Sciences*, 20(19), p. 4673.

Whitman, W. B., Coleman, D. C. and Wiebe, W. J. (1998) 'Prokaryotes: The unseen majority', *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), pp. 6578–6583.

Williams, B. L. *et al.* (2012) 'Application of novel PCR-based methods for detection, quantitation, and phylogenetic characterization of Sutterella species in intestinal biopsy samples from children with autism and gastrointestinal disturbances', *mBio*, 3(1), pp. 1–11.

Wilson, M. (2004) 'Microbial Inhabitants of Humans', Cambridge University Press.

Wu, G. D. *et al.* (2011) 'Linking long-term dietary patterns with gut microbial

enterotypes', *Science*, 334(6052), pp. 105–108.

Yanagibashi, T. *et al.* (2014) '*Bacteroides* induce higher IgA production than *Lactobacillus* by increasing activation-induced cytidine deaminase expression in B cells in murine Peyer's patches', *Bioscience, biotechnology, and biochemistry*, 73(2), pp. 372–377.

Yang, B. *et al.* (2020) 'Diversity of gut microbiota and bifidobacterial community of Chinese subjects of different ages and from different Regions', *Microorganisms*, 8(8), pp. 1–18.

Yatsunenko, T. *et al.* (2012) 'Human gut microbiome viewed across age and geography.', *Nature*, 486(7402), pp. 222–7.

Zeuthen, L. H., Fink, L. N. and Frokiaer, H. (2008) 'Epithelial cells prime the immune response to an array of gut-derived commensals towards a tolerogenic phenotype through distinct actions of thymic stromal lymphopoietin and transforming growth factor-beta', *Immunology*, 123(2), pp. 197–208.

Zheng, J. *et al.* (2015) 'The placental microbiome varies in association with low birth weight in full-term neonates', *Nutrients*, 7(8), pp. 6924–6937.

Zhu, A. *et al.* (2015) 'Inter-individual differences in the gene content of human gut bacterial species', *Genome biology*, 16(1), p. 82.

Zoetendal, E. G. *et al.* (2002) 'Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces', *Applied and environmental microbiology*, 68(7), pp. 3401–3407.

8 APPENDIX

Supplementary Tables and all figures are available at:

https://drive.google.com/drive/folders/1sj9igNPhXd785ZR4XFhsIOqgezsjB4kt?usp=sharing

This link can also be obtained by scanning the following **QR code**:

Supplementary Tables contents:

S1. Metadata for volunteers and samples.

S2. 16S rRNA sequencing statistics per sample.

S3. Metagenome sequencing statistics per sample.

S4. p-values of diversity (Shannon index and Chao1 estimator) pairwise comparisons between groups and between timepoints.

S5. Canonical Correspondence Analysis (CCA) and ADONIS test p-values.

S6. List of ASVs considering as discriminative of each age group by the sPLS-DA analysis.

S7. List of TIGRFAMs considering as discriminative of each age group by the sPLS-DA analysis.

S8. List of ASVs and TIGRFAMs IDs and their features associated of the *Circos* plot.

S9. p-values of stability (fraction of shared features between consecutive samples of the same individual) pairwise comparisons between groups and between timepoints.