



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

INSTITUTO DE BIOTECNOLOGÍA

Título

DIVERSIDAD Y DINÁMICA DEL VIROMA EUCARIONTE OROFARÍNGEO Y FECAL DEL
LACTANTE MENOR ASINTOMÁTICO

TESIS
QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:
XAIRA JIMENA RIVERA GUTIÉRREZ

DIRECTOR DE TESIS
DR. CARLOS FEDERICO ARIAS ORTIZ
INSTITUTO DE BIOTECNOLOGÍA

COMITÉ TUTOR
DRA. CECILIA XIMÉNEZ GARCÍA - FACULTAD DE MEDICINA
DR. JOSÉ LUIS PUENTE GARCÍA - INSTITUTO DE BIOTECNOLOGÍA

CUERNAVACA, MORELOS, ABRIL 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicatorias

A mis padres, quienes dejaron su familia y país en busca de mejores oportunidades y calidad de vida, y sentaron el privilegio a través del cual, logré llegar a donde estoy. Gracias por todo su esfuerzo, apoyo y cariño.

A mis hermanos, los tlacuaches más divertidos y abrazables del mundo mundial.

A los abuelxs, tíxs y primxs, quienes siempre mandan mensajes alentadores y me reciben con cariño y deliciosa comida cuando voy a ver a la familia.

A Gui y Jo, mis papás adoptivos y parte clave de mis primeros años en el doctorado, gracias por tanto cariño.

A mi yo del pasado, quién no le tuvo miedo a aprender cosas de cero, cambiarse de ciudad, conocer gente nueva y comenzar de cero. Gracias por tanta resiliencia.

A lxs amigxs, lxs nuevxs, lxs viejxs, lxs cercanxs y lxs que viven lejos. Todxs han sido, en algún punto de mi vida la familia escogida, el apoyo en la ansiedad y el acompañamiento necesario.

Al socio, por ser tan buen miembro del e-quipeo.

Agradecimientos

A la Universidad Nacional Autónoma de México por su apoyo con el financiamiento IN217821 de PAPIIT-DGAPA-UNAM y al CONACyT por el financiamiento 0272601 FONSEC/S008.

Al programa de becas de posgrado de Conacyt por su apoyo económico a través de la beca de posgrado 829902.

Al instituto Sabin por la beca otorgada para asistir al 14 Simposio Internacional de Rotavirus.

Al programa de posgrado por el apoyo económico otorgado para asistir a un Congreso Nacional de Virología.

A mi tutor, el doctor Carlos Arias, por su apoyo desde el momento en que recibí mi correo solicitando cambiar de tutor, gracias por los debates, por escuchar mis ideas, ayudarme a no perderme en el mar de información y darle un contexto a todo lo analizado.

A mi comité tutor, la doctora Cecilia y el doctor José Luis, por su constante apoyo y consejos durante el desarrollo de este proyecto.

A los doctores Pavel Isa, Tomás López y Carlos Sandoval, investigadores del laboratorio los cuales siempre estuvieron dispuestos a apoyarme y contestar todas mis dudas.

A la doctora Blanca Taboada por su constante apoyo y enseñanzas durante mi aprendizaje bioinformático, el cual comenzó desde cero.

A Marco A. Espinoza por su apoyo técnico durante el desarrollo del proyecto y mis experimentos.

A todos los integrantes del laboratorio Arias-López, gracias por la convivencia y la grata compañía. A Cat y a Jey, a quienes constantemente estaba llenando de preguntas sobre virología, que suerte haber llegado al mismo tiempo que ustedes y tenerlos como apoyo para descubrir el instituto y el laboratorio.

A Rodri y Mau, quienes me han enseñado la importancia del papel del post-doc como mentor para los doctorantes, gracias por el gran equipo de trabajo que hemos formado, además de una amistad basada en ñoñería y gusto por comer.

Al Instituto de Biotecnología de la UNAM por darnos acceso a su supercomputadora y a Jerome Verleyen y Juan Manuel Hurtado, por su apoyo técnico computacional.

Al equipo de trabajo de la doctora Cecilia Ximenez, especialmente a Xochiquetzalli Soto-Martínez por su apoyo con las familias de la comunidad de Xoxocotla, así como a los servicios de salud del estado de Morelos (Dirección de capacitación e investigación y el centro de salud de Xoxocolta Morelos), así como a la facultad de medicina de la UNAM.

Al grupo whatsapp de tardes de bioinfo, las retas de volley de los viernes y a la comunidad del IBT.

Índice de contenido

DEDICATORIAS	- 2 -
AGRADECIMIENTOS	- 3 -
ÍNDICE DE CONTENIDO	- 5 -
ÍNDICE DE TABLAS	- 7 -
ÍNDICE DE FIGURAS	- 9 -
PRÓLOGO	- 16 -
GLOSARIO	- 19 -
DICCIONARIO DE ABREVIACIONES	- 21 -
RESUMEN	- 23 -
ABSTRACT	- 25 -
INTRODUCCIÓN	- 27 -
El viroma	- 27 -
La interacción entre la orofaringe y el tracto gastrointestinal.....	- 29 -
PREGUNTA DE INVESTIGACIÓN	- 31 -
JUSTIFICACIÓN	- 31 -
HIPÓTESIS	- 31 -
OBJETIVOS	- 32 -
General:	- 32 -
Específicos:	- 32 -
MATERIAL Y MÉTODOS	- 33 -
Población de estudio:	- 33 -
Criterios de inclusión y exclusión:	- 34 -
Tamaño de muestra:.....	- 34 -
Consideraciones éticas:	- 34 -
Toma de las muestras:.....	- 34 -
Extracción de ácidos nucleicos:	- 35 -
Secuenciación:	- 35 -
Confirmación de presencia de virus en muestras:	- 36 -
Análisis bioinformático:	- 36 -
Análisis estadístico:.....	- 40 -

RESULTADOS Y DISCUSIÓN	- 42 -
Muestras.....	- 42 -
Secuenciación	- 43 -
Procesamiento.....	- 44 -
Normalización	- 45 -
Asignación taxonómica y análisis ecológico	- 45 -
Capítulo 1: El viroma de la orofaringe	- 46 -
Capítulo 2: El viroma de las muestras fecales	- 57 -
Capítulo 3: Diversidad y abundancia de calicivirus en las muestras fecales	- 68 -
Capítulo 4: El viroma compartido entre la orofaringe y las muestras fecales	- 74 -
Capítulo 5: Análisis filogenéticos de virus compartidos en ambos tractos	- 80 -
Capítulo 6: Identificación de rotavirus vacunal	- 84 -
Capítulo 7: Los niños están en constante contacto con los virus de plantas, principalmente de la familia <i>Virgaviridae</i> ..	- 90 -
DISCUSIÓN GENERAL.....	- 97 -
LIMITACIONES.....	- 101 -
CONCLUSIONES	- 102 -
PERSPECTIVAS.....	- 103 -
REFERENCIAS	- 104 -
MATERIAL SUPLEMENTARIO.....	- 111 -
Tablas.....	- 111 -
Figuras.....	- 115 -

Índice de tablas

Tabla 1. Información general de los niños y las condiciones de vida recolectados durante el estudio. - 33 -

Tabla 2. Porcentaje de frecuencia y abundancia de especies virales encontrados en las muestras orofaríngeas. Marcados en negrita se señalan las especies con más de 1% de abundancia de las lecturas de ADN detectadas. El porcentaje de frecuencia (% Frecuencia) se calculó considerando todas las muestras de la cohorte (n=90). El porcentaje de abundancia (% Abundancia) se calculó con el total de lecturas entre las muestras y el total de lecturas asignadas por especie. El porcentaje de abundancia por hospedero (% Abundancia por hospedero) se calculó considerando el total de lecturas de especies que replican en cada hospedero..... - 53 -

Tabla 3. Porcentaje de frecuencia y abundancia de especies virales encontrados en las muestras fecales. Marcados en negrita se señalan las especies con más de 1% de abundancia. El porcentaje de frecuencia (% Frecuencia) se calculó considerando todas las muestras de la cohorte (n=97). El porcentaje de abundancia (% Abundancia) se calculó con el total de lecturas entre las muestras y el total de lecturas asignadas por especie. El porcentaje de abundancia por hospedero (% Abundancia por hospedero) se calculó considerando el total de lecturas de especies que replican en cada hospedero..... - 59 -

Tabla 4. Genotipificación de genomas completos y parciales utilizando la herramienta “Calicivirus genotyping tool”. Los genotipos que tienen un () no pudieron ser tipificados. Durante el ensamblaje se utilizó una cobertura de 3x y para verificar posibles variantes, una cobertura de 10x. Si la posición de nucleótidos no tenía esa cobertura, se colocaba una N. Se colorearon los genotipos para mejor identificación y agrupamiento. Region B: RdRp. Región C: VP1..... - 71 -*

Tabla 5. Porcentaje de abundancia de lecturas de familias distribuidas entre los hospederos por cada nicho..... - 76 -

Tabla 6. Muestras en las cuales se identificaron genes asociados a la vacuna RotaTeq. Se puso la fecha de vacunación previa a una muestra identificada con genes de rotavirus. Todos los genes fueron identificados con la herramienta ViPR (ver la sección de materiales y métodos)..... - 89 -

Tabla S1: Lecturas secuenciadas y lecturas válidas, es decir, aquellas que quedaron después de eliminar lecturas de mala calidad, duplicadas y lecturas asignadas al hospedero y a genes

ribosomales. La tabla muestra cada niño y la edad en meses de la toma de muestra. Símbolos con ‘-’ representan ausencia de muestra en ese mes.- 111 -

Tabla S2: Estadística de las muestras orofaríngeas. Valores estadísticos de comparación de los datos de los infantes con los índices de alfa y beta diversidad en las lecturas de virus que infectan a humanos. En negrita, los valores que fueron estadísticamente significativos. Se realizaron pruebas no paramétricas.- 112 -

Tabla S3: Estadística de las muestras fecales. Valores estadísticos de comparación de los datos de los infantes con los índices de alfa y beta diversidad en las lecturas de virus que infectan a humanos. En negrita, los valores que fueron estadísticamente significativos. Se realizaron pruebas no paramétricas.- 113 -

Tabla S4. Muestras que de acuerdo al criterio de infección, fueron positivas a norovirus. El porcentaje de cobertura se calculó a partir del genoma original. El número de lecturas son la cantidad de lecturas que formaron ese genoma o contig.....- 114 -

Tabla S5. Muestras que de acuerdo al criterio de infección, fueron positivas a sapovirus. El porcentaje de cobertura se calculó a partir del genoma original. El número de lecturas son la cantidad de lecturas que formaron ese genoma o contig.....- 114 -

Índice de figuras

Figura 1. Diagrama de flujo de análisis bioinformático general. La palabra hit hace referencia a aquellas lecturas que mapearon contra la base de referencia empleada. - 38 -

Figura 2. Diagrama de flujo análisis bioinformático particular para las secuencias de familias virales - 40 -

Figura 3. Línea de tiempo de toma de muestras e introducción de alimentos. Los círculos señalan los días de la toma de muestras, los cuadrados el momento de introducción del alimento señalado. La línea inferior (eje x) muestra el mes de la toma de muestra, aunque de forma general etiquetamos las edades como mes completo o medio mes, de acuerdo con la semana de la toma de muestra. - 42 -

Figura 4. Distribución de lecturas mapeadas contra fagos, todos los genes registrados en GenBank (etiqueta todo nt) y bacterias en las muestras orofaríngeas y fecales, representada en porcentajes. Cada panel corresponde a un niño, y su etiqueta se observa en la parte superior. La edad asignada en cada muestra se observa en el eje inferior. VOI = viroma orofaríngeo infantil. VFI = Viroma fecal infantil. Cada número representa a un infante..... - 44 -

Figura 5. Porcentaje de lecturas identificadas a nivel taxonómico para virus procariontes y eucariontes en muestras orofaríngeas y fecales. - 45 -

Figura 6. Grafica de pie de las familias con mayor abundancia de lecturas en las muestras orofaríngeas (Izquierda), en gris claro con la etiqueta 'Others' están aquellas familias que tienen menos del 0.05% de abundancia de lecturas. Gráfica de barras de familias con un porcentaje menor a 0.1% de abundancia de lecturas (derecha)..... - 47 -

Figura 7. Abundancia de lecturas virales asignadas a la especie CMV en escala logarítmica base 10 para cada uno de los niños, mostrada en puntos. Las muestras dónde no hay puntos fueron negativas para CMV. VOI = viroma orofaríngeo infantil. Cada número representa a un infante y en la parte inferior (eje x) se observa la edad en meses de toma de cada muestra... - 50 -

Figura 8. Mapa de calor representando las abundancias de especies virales humanas que se replican en la orofaringe. Las abundancias están normalizadas y expresadas en escala de logaritmo 10. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Cada panel representa a un niño con el identificador en la parte superior de cada panel y la edad en meses en la parte inferior (eje x). VOI = viroma orofaríngeo infantil y cada número

representa a un niño de la cohorte. Las líneas punteadas dividen los meses entre el primer y segundo semestre de edad. Especies que se replican en humanos: TTV, torque teno virus; TTMV, torque teno mini virus; TTMDV, torque teno midi virus; CMV, citomegalovirus; HPV, human papillomavirus; HBoV, human bocavirus; RVA, rotavirus A; HPBV, human picobirnavirus; BetaCoV1, betacoronavirus 1; EVA, enterovirus A; AstV MLB3, Astrovirus MLB3; HRVA, rhinovirus A; HRVB, rhinovirus B; HRVC, rhinovirus C. - 51 -

Figura 9. Diversidad genética de los papillomavirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TVM2+F+I. Los genomas de referencia y raíz están en negritas.(b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura. - 53 -

Figura 10. Gráficas de barras comparando la diversidad de los virus eucariontes que infectan humanos presentes en la orofaringe, analizadas por trimestre de edad con datos normalizados. En la barra superior se identifica a cada niño según el caso. VOI=viroma orofaríngeo infantil. (A) Promedio de abundancia de lecturas asignadas a especies en todas las muestras de los niños, en escala logarítmica base 10; (B) Promedio de abundancia de lecturas asignadas a especies por niño, en escala logarítmica base 10; (C) Número de especies en todas las muestras de los niños, calculado con el índice de Chao; (D) Índice de Chao calculado por cada trimestre de cada niño; (E) Distribución de la abundancia entre especies en todas las muestras de los niños, calculada con el índice de Shannon; (F) Índice de Shannon calculado por trimestre para cada niño..... - 55 -

Figura 11. Grafica de pie de las familias con mayor abundancia de lecturas en las muestras fecales (Izquierda), en gris claro con la etiqueta 'Others' están aquellas familias que tienen menos del 0.05% de abundancia de lecturas. La gráfica de barras ilustra familias con un porcentaje menor a 0.1% de abundancia de lecturas (derecha)..... - 58 -

Figura 12. Diversidad genética de los parechovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM2+F+R4. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura. - 61 -

Figura 13. Mapa de calor representando las abundancias de especies virales humanas que se replican en la materia fecal. Las abundancias están normalizadas y expresadas en escala de logaritmo 10. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Cada panel representa a un niño con el identificador en la parte superior de cada panel y la edad en meses en la parte inferior (eje x). VFI = viroma fecal infantil. Las líneas punteadas dividen los meses entre el primer y segundo semestre de edad. Las especies que fueron

taxonómicamente asignadas como 'sp' se omitieron de la figura. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Especies que se replican en humanos: TTV, torque teno virus; TTMV, torque teno mini virus; TTMDV, torque teno midi virus; HPV, human papillomavirus; HBoV, human bocavirus; RVA, rotavirus A; NoV, norovirus; SaV, sapovirus; PeVA, parechovirus A; AstV, mastrovirus; EVA, enterovirus A; EVB, enterovirus B; EVC, enterovirus C; HRVA, rinovirus A; HRVB, rinovirus B; HRVC, rinovirus C. - 64 -

Figura 14. Gráficas de barras comparando la diversidad de los virus eucariontes que infectan humanos presentes en las muestras fecales, analizadas por trimestre de edad. En la barra superior se identifica a cada niño según el caso. VFI= viroma fecal infantil. (A) Promedio de abundancia de lecturas en todas las muestras de los niños, en escala logarítmica 10; (B) Promedio de abundancia de lecturas por niño, en escala logarítmica 10; (C) Número de especies en todas las muestras de los niños, calculado con el índice de Chao; (D) Índice de Chao calculado por cada trimestre de cada niño; (E) Distribución de la abundancia entre especies en todas las muestras de los niños, calculada con el índice de Shannon; (F) Índice de Shannon calculado por trimestre para cada niño. - 66 -

Figura 15. Frecuencia y abundancia de norovirus y sapovirus en los infantes. En el panel A se observan las muestras positivas a norovirus, en el panel B a sapovirus. La abundancia se encuentra normalizada y escalada en logaritmo base 10. Para simplificar la figura, se sumó la abundancia de las muestras tomadas a mitad del mes, con el mes posterior. Los cuadros blancos representan muestras negativas, los cuadros con 'NS' indican la ausencia de muestra en esa fecha. Se marcó con un (*) las muestras que no pudieron ser doblemente genotipificadas debido a una falta de cobertura de la región RdRp..... - 69 -

Figura 16. Detección por estaciones del año de norovirus. En cada infante se señaló con un (*) el mes y año de nacimiento; este se coloca para señalar también que el aumento de lecturas se asocia más hacia la temporalidad que a la edad de los niños. - 72 -

Figura 17. Análisis de coordenadas principales de la diversidad beta presente en los viomas. Los círculos representan las muestras de saliva y los rectángulos las muestras fecales. Tomado de Abeles et al, 2015, PLOS ONE. - 75 -

Figura 18. Diagrama de Venn ilustrando las especies compartidas y únicas de cada nicho que infectan a humanos (a) o a plantas y/u hongos (b). Las especies están ordenadas en orden descendente de acuerdo a su abundancia. En la esquina inferior se observa al nicho al que pertenece cada especie..... - 77 -

Figura 19. Gráficas de barras comparando la diversidad de los virus eucariontes que infectan humanos presentes en ambos nichos. (a) Promedio de abundancia de lecturas asignadas a especies en todas las muestras por nicho, en escala logarítmica base 10; (b) Promedio de abundancia de lecturas asignadas a especie por niño por nicho, en escala logarítmica base 10; (c) Número de especies en todas las muestras por nicho, calculado con el índice de Chao; (d) Índice de Chao calculado por cada nicho de cada niño; (e) Distribución de la abundancia entre especies en todas las muestras de cada nicho, calculada con el índice de Shannon; (f) Índice de Shannon calculado por nicho para cada niño. - 79 -

Figura 20. Diversidad genética de los astrovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM2+F+I. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura..... - 81 -

Figura 21. Diversidad genética de los enterovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo GTR+F+R2. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura..... - 82 -

Figura 22. Diversidad genética de los rinovirus. (a) Árbol filogenético de genomas completos y contigs (>500) construido con el modelo GTR+F+I+G4. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura. - 82 -

Figura 23. Diversidad genética de los anellovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM2+F+R2. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura..... - 83 -

Figura 24. Diversidad genética de los bocavirus. (a) Árbol filogenético de genomas completos y contigs (>500) construido con el modelo HKY+F. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura. - 84 -

Figura 25. Presencia de la lecturas asignadas a la especie rotavirus A en las muestras fecales (A) y orofaríngeas (B) en escala logaritmo base 10. En la parte superior se marca a cada infante; VFI = viomra fecal infantil, VOI= viroma rorofaríngeo infantil, seguido por el número de infante. En la parte inferior la fecha de toma de muestra y las fechas de vacunación con formato mes-día (eje x), adentro del mapa de calor, las vacunas están marcadas como '1, 2 o 3ra dosis'. En negritas, las fechas que tienen toma de muestra pareada en ambos nichos..... - 86 -

Figura 26. Diversidad genética de los segmentos VP4 y VP7 de rotavirus. (a) Árbol filogenético del segmento VP4 con genomas completos y secuencias parciales (contigs >500) construido con el modelo HKY. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura del gen VP4. (c) Árbol filogenético del segmento VP7 con genomas completos y secuencias parciales (contigs >500) construido con el modelo HKY+I. (d) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura del gen VP7. Los genomas de referencia y raíz están en negritas. - 87 -

Figura 27. Mapa de calor que representa las abundancias normalizadas expresadas en escala de logaritmo base 10 de las especies virales de plantas que están presentes en las muestras orofaríngeas (a) o fecales (b). Cada panel representa a un niño con el identificador en la parte superior de cada panel y la edad en meses en la parte inferior (eje x). VOI= viroma orofaríngeo infantil. VFI = viroma fecal infantil. Las especies que fueron taxonómicamente asignadas como 'sp' se omitieron de la figura. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Las líneas punteadas negras dividen los meses entre el primer y segundo semestre de edad, las azules indican la introducción de otros líquidos y las rojas de frutas y/o verduras. **Especies que se replican en plantas y/u hongos presentes en la orofaringe:** SsVL, Schffersomyces Segobiensis virus L; PsV, Puccinia Striiformis totivirus; RPaTVs, Red Clover Powdery mildew-associated totivirus; PACV, Persea Americana chrysovirus; BPEV, Bell Pepper alphaendornavirus; BPeMV, Bell Pepper mottle virus; PMMoV, Pepper Mild Mottle virus; RCNaV, Rattail Cactus necrosis-associated virus; TSAMV, Tropical Soda Apple mosaic virus; TMGMV, Tobacco Mild Green mosaic virus; ToBRFV, Tomato Brown Rugose Fruit virus; OpV2, Opuntia virus 2; GarVA, Garlic virus; ACLSV, Apple chlorotic leaf spot virus; AbV, Agaricus Bisporus virus. **Especies que se replican en plantas y/u hongos presentes en materia fecal:** ZMCV1, Zea mays chrysovirus; MATV, Maize associated totivirus; ToNSV, Tomato necrotic spot virus; BPeMV, Bell Pepper mottle virus; PaMMV, Paprika Mild mottle virus; PMMoV, Pepper Mild mottle virus; RCNaV, Rattail Cactus necrosis-associated virus; TMGMV, Tobacco Mild Green mosaic virus; ToMV, Tomato mosaic virus; TSAMV, Tropical Soda Apple mosaic virus; OpV2, Opuntia virus 2; MNSV, Melon necrotic spot virus; MeSMV, Melon severe mosaic topovirus. - 91 -

Figura 28. Abundancia de PMMV en la orofaringe y muestras fecales, expresada en logaritmo base 10 (eje y). En la parte superior está marcado el identificador de cada infante y en la inferior la edad en meses de cada infante en el momento de la toma de muestra (eje x). VI= Viroma infantil. - 94 -

Figura 29. Abundancia de TSAMV en la orofaringe y muestras fecales, expresada en logaritmo base 10 (eje y). En la parte superior está marcado el identificador de cada infante y en la inferior la edad en meses de cada infante en el momento de la toma de muestra (eje x). VI= Viroma infantil. - 95 -

Figura S1. Distancias que existen en la composición de virus eucariontes que infectan a humanos en las muestras orofaríngeas (a-c) Análisis no constreñidos (PCoA) mostrando los 3 ejes principales y porcentaje por eje. (d) análisis constreñidos (db-RDA) mostrando solo 2 ejes. Cada niño tiene un color y punto distintivo. Las elipses en color negro representan el error estándar del centroide.....- 115 -

Figura S2. Distancias que existen en la composición de virus eucariontes que infectan a humanos en las muestras fecales (a-c) Análisis no constreñidos (PCoA) mostrando los 3 ejes principales y porcentaje por eje. (d) análisis constreñidos (db-RDA) mostrando solo 2 ejes. Cada niño tiene un color y punto distintivo. Las elipses en color negro representan el error estándar del centroide.....- 115 -

Figura S3. Diversidad genética de norovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo GTR+I. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.....- 116 -

Figura S4. Diversidad genética de sapovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM1+F+G4. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.....- 116 -

Figura S5. Distancias que existen en la composición de virus eucariontes que infectan a humanos en ambos tractos. (a-c) Análisis no constreñidos (PCoA) mostrando los 3 ejes principales y porcentaje por eje. (d) análisis constreñidos (db-RDA) mostrando solo 1 eje. Cada niño tiene un color y se distinguen los nichos por las viñetas empleadas.....- 117 -

Figura S6. Diversidad genética de pepper mild mottle virus. Árbol filogenético creado con genomas completos y secuencias parciales (contigs > 500nt) construido con el modelo GTR+F+R10. Los genomas de referencia y raíz están en negritas.....- 118 -

Figura S7. Cambios de nucleótidos de pepper mild mottle virus. Mapa de calor hecho a partir de secuencias con un mínimo de cobertura del genoma del 40%.....- 119 -

Figura S8. Diversidad genética de tropical soda apple mosaic virus. Árbol filogenético creado con genomas completos y secuencias parciales (contigs > 500nt) construido con el modelo GTR+F+R6. Los genomas de referencia y raíz están en negritas.....- 120 -

Prólogo

Este trabajo forma parte de un proyecto para determinar el establecimiento del microbioma en infantes mexicanos, en el que se está analizando el bacterioma, el viroma y el eucarioma.

La parte del viroma ha sido analizada por diversos alumnos del laboratorio Arias-López. Previo a este trabajo se analizó, en un sub-grupo de la cohorte, la prevalencia y filogenia de una familia de virus que infectan plantas y que fue muy abundante y frecuente tanto en la orofaringe como en muestras fecales. De este análisis se publicó el siguiente artículo:

Tobamoviruses can be frequently present in the oropharynx and gut of infants during their first year of life. Aguado-García Y, Taboada B, Morán P, **Rivera-Gutiérrez X**, Serrano-Vázquez A, Iša P, Rojas-Velázquez L, Pérez-Juárez H, López S, Torres J, Ximénez C, Arias CF. Sci Rep. 2020 Aug 12;10(1):13595. doi: 10.1038/s41598-020-70684-w. PMID: 32788688; PMCID: PMC7423923.

Esta tesis extiende los análisis a la cohorte completa de niños, enfocándose en describir el viroma eucarionte encontrado tanto en la orofaringe como en las muestras fecales, y explorando familias que se destacan por ser frecuentes y abundantes. Por esta razón la tesis está organizada con una introducción general y después 6 capítulos, los cuales fueron englobados en 2 manuscritos donde soy autora principal, uno ya publicado y otro en revisión.

Cada capítulo presenta en estructura una breve introducción, resultados y discusión, finalizando la tesis con una discusión, conclusiones generales y perspectivas.

En el primer capítulo se presentan los resultados del viroma orofaríngeo de los 9 niños, comenzando con una descripción general de lo encontrado, para después describir los hallazgos a nivel de especies virales de interés y posteriormente, con el afán de proporcionar una visión más panorámica, se realizaron análisis ecológicos.

En el segundo capítulo se presenta una estructura similar al primer capítulo, pero enfocada a los virus presentes en las muestras fecales, seguido por el tercer capítulo donde se describe a detalle la familia más frecuente y abundante del viroma fecal: la familia *Caliciviridae*. De la descripción detallada de esta familia se publicó el siguiente artículo:

High Prevalence and Diversity of Caliciviruses in a Community Setting Determined by a Metagenomic Approach. **Rivera- Gutiérrez X***, Morán P, Taboada B, Serrano-Vázquez A, Iša P, Rojas-Velázquez L, Pérez-Juárez H, López S, Torres J, Ximénez C, Arias CF. Microbiol

Spectr. 2022 Feb 23;10(1):e0185321. doi: 10.1128/spectrum.01853-21. Epub 2022 Feb 23. PMID: 35196791; PMCID: PMC8865552.

En el cuarto capítulo se comparan los dos nichos estudiados haciendo énfasis en las especies compartidas y en las diferencias clave de ambos nichos, tales como la abundancia de lecturas y la diversidad observada.

El quinto capítulo es una extensión del cuarto capítulo, dónde se analiza la filogenia y diferencias de nucleótidos de especies virales encontradas en ambos nichos.

El sexto capítulo se describe a la especie más interesante que se compartió entre ambos nichos: rotavirus A, el cual en algunas muestras logramos identificar secuencias asociadas a una cepa derivada de la aplicación de vacuna. Este capítulo se presentó en formato cartel en el 14 Simposio Internacional de Rotavirus, gracias a una beca otorgada por el Instituto Sabin.

Finalmente, el séptimo capítulo ofrece una visión de los virus de plantas encontrados en el viroma. Este capítulo incluye la descripción de dos virus que pertenecen a la familia Virgaviridae y que fueron los más abundantes y frecuentes en ambos nichos: Pepper Mild mottle virus (PMMoV) y Tropical Soda Apple mosaic virus (TSAMV).

El siguiente manuscrito es un resumen de los resultados del primer, segundo, cuarto, quinto y sexto capítulo:

The fecal and oropharyngeal eukaryotic viromes of healthy infants during the first year of life are personal. **Rivera-Gutiérrez X***, Morán P, Taboada B, Serrano-Vázquez A, Iša P, Rojas-Velázquez L, Pérez-Juárez H, López S, Torres J, Ximénez C, Arias CF. Scientific reports. 2023 Jan 17;13(1):938. doi: 10.1038/s41598-022-26707-9. PMID: 36650178; PMCID: PMC9845211.

Además de los análisis relacionados a mi proyecto de tesis y en el marco de la pandemia, también participé como médico de apoyo para el campus UNAM-Morelos y en el análisis y seguimiento epidemiológico y genómico del virus SARS-CoV-2, donde fui co-autora de los siguientes artículos:

Dominance of 3 sublineages of the SARS-CoV 2 Delta Variant in México. Taboada B, Zárate S, García-López R, Muñoz-Medina JE, Sanchez-Flores A, Herrera-Estrella A, Boukadida C, Gómez-Gil B, Selem Mojica N, Rosales-Rivera M, Salas-Lais AG, Gutiérrez-Ríos RM, Loza A, **Rivera-Gutierrez X**, Vazquez-Perez JA, Matías-Florentino M, Pérez-García M, Ávila-Ríos S, Hurtado JM, Herrera-Nájera CI, Núñez-Contreras JdJ, Sarquiz-Martínez B, García-Arias VE, Santiago-Mauricio MG, Martínez-Miguel B, Enciso-Ibarra J, Cháidez-Quiróz C, Iša P, Wong-

Chew RM, Jiménez-Corona M-E, López S, Arias CF. Viruses. 2022; 14(6):1165. <https://doi.org/10.3390/v14061165>.

The Alpha Variant (B.1.1.7) of SARS-CoV-2 Failed to Become Dominant in México. Zárate S, Taboada B, Muñoz-Medina JE, Iša P, Sanchez-Flores A, Boukadida C, Herrera-Estrella A, Selem Mojica N, Rosales-Rivera M, Gómez-Gil B, Salas-Lais AG, Santacruz-Tinoco CE, Montoya-Fuentes H, Alvarado-Yaah JE, Molina-Salinas GM, Espinoza-Ayala GE, Enciso-Moreno JA, Gutiérrez-Ríos RM, Loza A, Moreno-Contreras J, García-López R, **Rivera-Gutierrez X**, Comas-García A, Wong-Chew RM, Jiménez-Corona ME, Del Angel RM, Vazquez-Perez JA, Matías-Florentino M, Pérez-García M, Ávila-Ríos S, Castelán-Sánchez HG, Delaye L, Martínez-Castilla LP, Escalera-Zamudio M, López S, Arias CF. Microbiol Spectr. 2022 Apr 27;10(2):e0224021. doi: 10.1128/spectrum.02240-21. Epub 2022 Apr 7. PMID: 35389245; PMCID: PMC9045257.

Two-year follow-up of the COVID-19 pandemic in Mexico. Loza A, Wong-Chew RM, Jimenez-Corona ME, Zarate S, Lopez S, Ciria R, Palomares D, García-López R, Isa P, Taboada B, Rosales-Rivera M, Boukadida C, Herrera-Estrella A, Selem Mojica N, **Rivera-Gutierrez X**, Muñoz-Medina JE, Salas-Lais AG, Sanchez-Flores A, Vazquez-Perez JA, Arias CF, Gutierrez-Rios RM. Front. Public Health. 2023. 10:1050673. doi: 10.3390/fpubh.2022.105673

Artículo sometido:

SARS-CoV-2 BW lineage, a fast-growing Omicron variant from southeast Mexico bearing relevant escape mutations. García-López R, **Rivera-Gutiérrez X***, Rosales-Rivera M, Taboada B, Zárate S, Muñoz-Medina JE, Roche B, Herrera-Estrella A, Gómez-Gil B, Sanchez-Flores A, Arias CF. Co-primer autor.

Glosario

- **Abundancia:** en ecología es la representación relativa de una especie en un nicho, es decir, el número de veces que está presente una especie en un nicho específico. En nuestro caso lo consideramos como el número de lecturas asignadas de forma taxonómica a una especie específica en cada muestra.
- **Bacteriota:** término que se utiliza cuando la descripción de la parte bacteriana de la microbiota se basa en técnicas metataxonómicas [1], es decir, secuenciación del gen ribosomal 16S.
- **Bacterioma:** término utilizado para definir los genes y taxones bacterianos encontrados en un nicho a través de técnicas de metagenómica.
- **Enterotipo:** término acuñado en los estudios de bacteriota intestinal para definir grupos de familias bacterianas prevalentes, independientes de edad, sexo, y hasta ubicación geográfica [2].
- **Frecuencia:** se define como número de repeticiones por unidad de tiempo de cualquier evento. Para finalidades de esta tesis, se define como el número de muestras, de uno o todos los niños, en las que una especie viral está presente.
- **Índice Chao1:** es un método no paramétrico el cual tiene la finalidad de estimar la riqueza, es decir, el número total de especies, incluyendo aquellas no observadas, en una muestra a partir de las especies que se observan en ésta. Se basa en la premisa de inferir especies no observadas a partir de especies con poca abundancia. Este índice es útil en datos que tienen en su mayoría especies poco abundantes [3].
- **Índice de disimilitud Bray-Curtis:** es una métrica de disimilitud que utiliza la abundancia de cada especie dentro de cada muestra como la base para hacer las comparaciones entre muestras, considerando las especies en común entre las muestras, dando como resultado una matriz de disimilitudes [4].
- **Índice Shannon:** es una medida de entropía (en este caso, incertidumbre) de la comunidad, tiene la finalidad de estimar la diversidad ecológica. La diversidad ecológica o biológica es la variabilidad entre organismos en un nicho ecológico. Considera dos factores: la riqueza, es decir el número de especies, y qué tan homogénea o heterogéneamente se distribuye la abundancia dentro de estas especies. Si hay una abundancia mayoritaria de una sola especie, el índice es bajo [3].
- **Individualidad:** conjunto de características propias del individuo. Para fines de esta tesis, cuando se menciona el término “individualidad”, se hace referencia a los componentes ambientales, sociales y posiblemente genéticos que tienen una dinámica peculiar en cada niño.
- **Lactante:** término que define los 24 meses de vida, los cuales están marcados por cambios drásticos en el crecimiento y desarrollo de los niños. Asimismo, comprende la etapa de

consumo principal de leche materna y/o fórmula. Operacionalmente se divide en lactante menor (1 – 12 meses de vida) y mayor (12 – 24 meses de vida) [5].

- **Metagenómica:** proceso de obtención de información genética o genes a partir de secuenciación masiva aleatoria de ADN o ADNc [1].
- **Metataxonómica:** proceso de alto rendimiento utilizado para caracterizar todos los microorganismos presentes en la microbiota a partir de un marcador genético taxonómico. (p. Ej. 16S, 18S) [1].
- **Microbioma:** término que considera los microorganismos (bacterias, arqueas, eucariontes y virus), sus genes y/o genomas y las condiciones ambientales que conforman un hábitat. También es utilizado de forma intercambiable con metagenoma, ambos usados para definir la colección de genes y genomas de los miembros de la microbiota. El microbioma se estudia a través de técnicas metagenómicas, metabolómicas, transcriptómicas y proteómicas [1].
- **Microbiota:** conjunto de microorganismos presentes en un ambiente o nicho determinado. Acuñado inicialmente por Lederberg y McCray, tenía como misión hacer énfasis en la importancia de los microorganismos, en particular las bacterias, que co-habitan en el cuerpo humano y su impacto en la salud y la enfermedad. Este se estudia principalmente analizando las sub-unidades ribosomales 16S y 18S [1].
- **Viroma:** conjunto de virus eucariontes, procariontes y retrovirus endógenos que están presentes en un nicho determinado en un tiempo fijo. Al no haber marcadores taxonómicos para estudiarlos, la asignación taxonómica depende de la identificación de sus genes a través de técnicas metagenómicas [1,6]. Para fines de esta tesis, los resultados hacen referencia a la parte eucarionte del viroma.

Diccionario de abreviaciones

- **AbV:** Agaricus Bisporus virus
- **ACLSV:** Apple chlorotic leaf spot virus
- **AstV MLB3:** Astrovirus MLB3
- **BPEV:** Bell Pepper alphaendornavirus
- **BPeMV:** Bell Pepper mottle virus
- **BetaCoV1:** Betacoronavirus 1
- **CMV:** Vitomegalovirus
- **EVA:** Enterovirus A
- **GarVA:** Garlic virus
- **HBoV:** Human bocavirus
- **HPV:** Human papilomavirus
- **HPBV:** Human picobirnavirus
- **HRVA:** Rhinovirus A
- **HRVB:** Rhinovirus B
- **HRVC:** Rhinovirus C
- **MATV:** Maize associated totivirus
- **MNSV:** Melon necrotic spot virus
- **MeSMV:** Melon severe mosaic toposvirus
- **OpV2:** Opuntia virus 2
- **PCR:** Reacción de cadena de polimerasa. Cuando tiene el prefijo “RT” se hace referencia a la técnica de transcripción reversa por reacción de cadena de polimerasa.
- **PaMMV:** Paprika Mild mottle virus
- **PMMoV:** Pepper Mild Mottle virus
- **PACV:** *Persea Americana* chrysovirus
- **PsV:** *Puccinia Striiformis* totivirus
- **RCNaV:** *Rattail Cactus* necrosis-associated virus
- **RPaTVs:** Red Clover Powdery mildew-associated totivirus
- **RVA:** Rotavirus A
- **SsVL:** *Schffersomyces Segobiensis* virus L
- **TSAMV:** Tropical Soda Apple mosaic virus
- **TMGMV:** Tobacco Mild Green mosaic virus
- **ToBRFV:** Tomato Brown Rugose Fruit virus
- **ToMV:** Tomato mosaic virus
- **ToNSV:** Tomato necrotic spot virus
- **TTV:** Torque teno virus
- **TTMV:** Torque teno mini virus
- **TTMDV:** Torque teno midi virus

- **VFI:** Viroma fecal infantil
- **VOI:** Viroma orofaríngeo infantil
- **ZMCV1:** *Zea mays* chrysovirus

Resumen

Utilizando métodos de secuenciación genómica describimos y comparamos la diversidad y la dinámica del viroma de la orofaringe y de muestras fecales de 9 niños asintomáticos en una comunidad semi-rural. La toma de muestras comenzó aproximadamente 2 semanas después del nacimiento y continuó de forma mensual hasta cumplir los 12 meses de edad, resultando en un total de 90 muestras orofaríngeas y 97 fecales.

En ambos nichos estudiados encontramos que más del 95% del total de las secuencias asignadas pertenecían a especies virales que se replican en humanos o en plantas. Respecto a las especies virales humanas más abundantes y frecuentes, éstas pertenecen a 3 familias en la orofaringe: *Herpesviridae*, *Picornaviridae* y *Reoviridae*. Como dato interesante, dentro de la familia herpesvirus se encuentra la especie betaherpesvirus 5, también llamada citomegalovirus (CMV), que se encontró en todos los niños estudiados.

En las muestras fecales, las especies más predominantes pertenecen a cuatro familias virales fueron las: *Caliciviridae*, *Picornaviridae*, *Reoviridae* y *Anelloviridae*. Los virus de la familia *Caliciviridae*, conocida por causar brotes de gastroenteritis, fueron los más frecuentes y abundantes a pesar de que los niños no reportaron síntomas gastrointestinales durante todo el tiempo de estudio.

También observamos ciertas especies compartidas entre nichos. Un análisis filogenético y de diferencias de nucleótidos nos sugiere que en algunas familias, hay una variedad de especies que están circulando en la comunidad y que los niños están en constante contacto con ellas, mientras que en otros escenarios, observamos una excreción prolongada de la misma especie viral, sugiriendo infecciones asintomáticas.

Uno de los casos más interesantes de familias compartidas, abundantes y prevalentes fue el de los rotavirus, donde se identificaron secuencias de virus vacunales, en específico a la vacuna RotaTeq. Por otro lado, ambos nichos mostraron una alta frecuencia y abundancia de dos especies de la familia *Virgaviridae*. Un análisis en un subgrupo de 3 niños sobre la presencia de estas especies se realizó previamente en nuestro laboratorio. Aquí extendemos esos hallazgos a los 6 niños restantes.

De forma general, observamos diferencias en la abundancia de secuencias y diversidad de las especies virales en ambos nichos y a través de todo el tiempo de estudio, con algunos virus presentes desde los primeros meses de vida. En ambos nichos pudimos ver que a pesar de que algunas variables podrían explicar la composición del viroma (alimentación, tipo de vivienda), su establecimiento y desarrollo en los niños no se pudo establecer por la alta

heterogeneidad observado en las diferentes muestras. En las muestras fecales encontramos tendencias de cambios en la frecuencia de algunas especies virales dependientes de la edad.

En conclusión, nuestros resultados sugieren que el comportamiento del viroma de los niños en cada uno de estos nichos es dinámico y muy probablemente influenciado por el ambiente, alimentación y la edad. Además, encontramos que la composición del viroma es principalmente individual. Nuestra descripción abre la puerta al interrogante de si esta exposición constante a diferentes especies virales tiene un impacto a largo plazo en la salud o el desarrollo del sistema inmune de los niños.

Abstract

Using a metagenomic sequencing approach we describe and compare the diversity and dynamics of the oropharyngeal and fecal eukaryotic virome of nine asymptomatic children in a semi-rural community setting. A total of 90 oropharyngeal swabs and 97 fecal samples were collected starting 2 weeks after birth and monthly thereafter until 12 months of age. In both niches, more than 95% of the total sequence reads obtained were represented by viruses that replicate either in humans or in plants.

Regarding human viruses, viral species within three families were most abundant and frequent in the oropharynx: *Herpesviridae*, *Picornaviridae* and *Reoviridae*. Interestingly, among the herpesvirus family, the species betaherpesvirus 5 (CMV) was found in all the children.

Regarding fecal samples, species belonging to four virus families predominated: *Caliciviridae*, *Picornaviridae*, *Reoviridae*, and *Anelloviridae*. Viruses in the calicivirus family, known for its gastrointestinal outbreaks, were the most abundantly and frequently found, which is an interesting finding considering that none of the children had gastrointestinal symptoms during the time of the study.

Shared viral species among both niches were also present. A phylogenetic and nucleotide differential analysis suggest that in some families, there are different species that are in close contact with the children, while in other scenarios, we observe and confirm a prolonged shedding of the same virus, suggesting asymptomatic infections.

An interesting case of a highly frequent and abundant family is that of rotaviruses, from the *Reoviridae* family, which in some samples were identified as components of the RotaTeq vaccine, which is part of the national vaccination campaign. On the other hand, both niches showed a high abundance of two species of plant viruses of the family *Virgaviridae*. A sub-analysis on a sub-set of 3 children was previously described by a researcher of our laboratory. Here we extend their findings into the 6 remaining children.

Overall, differences in the abundance of sequence reads and diversity of virus species were observed in both niches and throughout the year of study, with some viruses already present in the first months of life. In both niches, we observed that despite some socioeconomic variables could influence the composition of the virome (such as feeding or housing conditions), a high heterogeneity composition was present among samples, hindering possible explanations of how the infant virome establishes and develops. In fecal samples, tendencies regarding frequency changes between certain species were dependent on age.

In conclusion, our results suggest that the children's virome in each of these anatomical niches is dynamic and most likely shaped by the environment, feeding and age. Moreover, we found that the virome composition is mostly individual. Whether this constant exposition to different viruses has a long-term impact on the children's health or development, remains to be studied.

Introducción

Conocemos como microbiota a la comunidad de microorganismos formada por protistas, hongos, bacterias, arqueas y virus que cohabitan en un nicho específico [1]. Con poco más de 20 años de estudios de la parte bacteriana de la microbiota en humanos (bacteriota), hemos llegado a describir que la simbiosis entre la microbiota y su hospedero tiene un papel en el estado de salud y enfermedad. Este rol comienza justo después del nacimiento, que es cuando el ser humano es colonizado por las bacterias [7]. Se ha descrito que diversos factores modifican la relación entre la colonización del hospedero por la bacteriota; inicialmente la ruta de nacimiento, pero después de forma constante por la ingesta de alimentos, consumo de antibióticos, estilo de vida y medio ambiente, entre otros factores [8], con una tendencia a un comportamiento bastante dinámico, con cambios drásticos más o menos hasta los 3 años de vida, donde se estabiliza y se parece más a el bacteriota observado en adultos sanos [9].

De hecho, estas exposiciones externas se consideran tan relevantes que uno de los conceptos actuales más importantes es que el microbioma favorece diferentes aspectos de la fisiología del ser humano, y que en los niños pudiese modificar su desarrollo [10], durante el cual se cree que influyen factores tales como la forma de colonización y dinámica de los componentes del microbioma, es decir los microorganismos y sus genes presentes en un nicho [1]. Asimismo, se ha descrito que beneficia la modulación de la respuesta inmune (reacción a antígenos específicos y desarrollo de tolerancia) y ayuda a la diferenciación y maduración celular [11].

Como hemos mostrado, el componente de la microbiota del cual se tiene más conocimiento es la bacteriota; de hecho, en comparación la información que se tiene acerca del viroma es escasa [12]. Esta escasez de conocimiento incrementa cuando consideramos solamente el viroma eucarionte. Aquí decimos viroma y no virota, pues para poder identificar a los virus a nivel taxonómico se necesitan tecnologías de secuenciación de metagenomas. Y este dato, es quizá uno de los factores más importantes que explican porqué casi no hay estudios del viroma; mientras que las bacterias tienen marcadores universales que son relativamente económicos y que facilitan la asignación taxonómica, siendo el más común la subunidad ribosomal 16S [1], los virus no, haciendo su estudio más laborioso, complejo y costoso [13].

El viroma

El viroma es la parte viral del microbioma y se divide de forma muy general en virus que infectan a organismos procariontes o eucariontes. El viroma humano está compuesto por virus que son capaces de replicarse en células humanas y aquellos que infectan a bacterias (bacteriofágos), arqueas, o a otros organismos eucariontes [14], así como por retrovirus endógenos [6].

En el pasado, el estudio y detección de virus requería técnicas de propagación viral en cultivos

celulares, lo cual limitó el conocimiento e identificación viral. Actualmente los virus son considerados las entidades biológicas más numerosas y diversas en la tierra, y gracias a la introducción de tecnologías de secuenciación masiva y metagenómica ha sido posible describir la composición del viroma en diversos nichos, y cómo su composición y abundancia cambia respecto al nicho estudiado [6]. En el cuerpo humano se ha estimado que existe una población de hasta 10^8 virus por mililitro de fluidos provenientes de cavidad nasal, oral, faringe y saliva, mientras que en el intestino se tiene un estimado de 10^9 partículas virales por gramo de heces [15–17].

Es importante aclarar que los virus pueden ser de ADN o ARN y de cadena positiva o negativa y de cadena simple o doble, y esto pareciera ser importante al describir viromas eucariontes del cuerpo humano en diferentes nichos, con ciertos nichos teniendo una mayor prevalencia de virus de ARN, como el tracto gastrointestinal, y otros de ADN, como es el caso de la cavidad oral y vías respiratorias [18]. Además, los virus eucariontes de RNA que se han identificado en muestras de humanos han sido divididos también por sus rasgos clínicos, en patógenos y no patógenos. Dentro los virus de RNA no patógenos destacan aquellos encontrados en plantas, en familias como *Picobirnaviridae* y *Virgaviridae*, los cuales pareciera que tienen su origen en la comida. Dentro de los virus de RNA identificados como patógenos se encuentran las familias *Reoviridae*, *Coronaviridae*, *Caliciviridae*, *Flaviviridae* y *Picornaviridae*, entre otras. En el caso de los virus de DNA casi todas las familias han sido catalogadas como patógenas y las más descritas en humanos son *Herpesviridae*, *Papillomaviridae* y *Polyomaviridae*, [6]. La única excepción es quizá la familia *Anelloviridae*, la cual si bien no ha sido clasificada como patógena de forma directa, se ha relacionado con el desarrollo de enfermedades respiratorias en niños y se ha sugerido como marcador molecular de procesos inflamatorios [19].

A pesar de que hemos logrado avanzar en la identificación y predicción de nuevos virus, la realidad es que aún hay mucho que no ha sido clasificado y que actualmente se ha descrito como ‘materia negra viral’, la cual sigue siendo un reto al conocimiento. Además, a diferencia de los estudios en los que se clasifica a las bacterias a través de caracterizar el RNA ribosomal 16S, el costo computacional de trabajar con virus es mayor, y no hay flujos de análisis informáticos (conocidas en inglés como *pipelines*) bien establecidos, como sí sucede en estudios de 16S usando herramientas como QIIME [20], la cual es una plataforma amigable para el usuario, donde se puede semi-automatizar el análisis bioinformático y estadístico y que no demanda una gran cantidad de recursos computacionales. Aún hay mucho camino por recorrer en cuanto al viroma, aunque es muy probable que así como se ha descrito para el bacterioma, factores tales como dieta, ubicación geográfica, condiciones de vida, calidad del agua, higiene, zoonosis y estatus económico, entre otros, afecten la diversidad y dinámica de este [16,21].

La interacción entre la orofaringe y el tracto gastrointestinal

Durante el parto (sin importar si este es por vía vaginal o quirúrgica) la nariz y la boca, componentes del aparato respiratorio y del sistema digestivo respectivamente, son los portales de entrada de los microorganismos adquiridos a partir de la madre y del medio ambiente. Estos microorganismos posteriormente serán parte del microbioma [22–24].

El sistema digestivo está compuesto por diversas estructuras; boca, orofaringe, esófago, estómago, duodeno, yeyuno, íleon y colon [25]. Sin embargo, cuando pensamos en el sistema digestivo y su microbiota, nos referimos a los estudios de bacteriota en materia fecal, pero no consideramos la puerta de entrada: la boca. Al respecto, se ha sugerido que esta podría actuar como un reservorio constante de bacterias para el intestino. Algunos estudios han señalado una diferencia en la abundancia de la cantidad de bacterias compartidas entre la boca y el tracto gastrointestinal (materia fecal) en adultos [26] y también en ancianos [27]. Respecto al viroma, un estudio realizado en 9 adultos encontró familias de bacteriófagos compartidas entre muestras de saliva y heces, con abundancias que varían según el nicho, no así con familias de virus eucariontes [17].

Estas diferencias en abundancia de microorganismos pueden darse por diferentes motivos, dentro de los cuales destacan: 1) la naturaleza de la boca como un nicho constantemente expuesto al medio ambiente, contrario al resto del aparato intestinal que depende de lo que se deglute; 2) la saliva actúa como un acarreador de microorganismos, se estima que tragamos un estimado diario de 600 ml, con hasta 10^9 bacterias/ml [27] y; 3) la capacidad de los microorganismos para sobrevivir a las defensas naturales del tracto digestivo, tales como enzimas y cambios en el pH [28].

Por otra parte, la orofaringe se comporta como un espacio híbrido donde tanto lo respirado como lo ingerido tendrán contacto con ella, haciendo de este espacio quizá uno de los espacios internos más expuestos al medio ambiente. Por lo mismo, es lógico imaginar un escenario donde existan especies compartidas con el tracto gastrointestinal, pero a su vez, también especies propias de la orofaringe, que quizá podrían ser adquiridas a través de la respiración y no de la ingesta. Desafortunadamente, hasta el momento de escritura de esta tesis, no hay estudios de viroma en orofaringe de niños sanos por lo que la composición del viroma en esta área del cuerpo no está descrita aún.

Finalmente, podemos decir que tanto la orofaringe, como la materia fecal son nichos de interés para el estudio del viroma. Esta importancia es inclusive aún más relevante si consideramos que la mayoría de los estudios en niños sanos menores de un año de edad están enfocados a la descripción del viroma en materia fecal [22,29–35], con pocos estudios llevados a cabo en comunidades rurales [36–39]. Además, la gran mayoría enfoca su descripción a la presencia de bacteriófagos [29–32,34], y en ocasiones, a su interacción con bacterias como parte del

microbioma [35,39]. Respecto a los virus eucariontes, se describen como presentes de forma esporádica, con un incremento notorio en abundancia a nivel de ciertas familias después de los 6 meses de edad [31,34,36,39].

En este trabajo comparamos las comunidades de virus eucariontes presentes en la orofaringe y materia fecal de muestras obtenidas mensualmente, desde el nacimiento hasta el primer año de vida, de 9 niños asintomáticos que habitan en Xoxocotla, una comunidad rural de Morelos. Asimismo, describimos la diversidad y dinámica del viroma en ambos nichos, las especies particulares de cada nicho y aquellas compartidas, y analizamos las variables que podrían contribuir al establecimiento del viroma. Finalmente, establecemos un perfil longitudinal de los cambios ecológicos y a nivel de especie que ocurren durante esta fase importante del desarrollo del infante. El estudio de una población rural mexicana es una valiosa aportación a la literatura, pues la mayoría de lo poco publicado para niños sanos menores de un año ha sido reportada para comunidades urbanas y de países desarrollados y ha estado dirigida principalmente a la caracterización de virus bacterianos.

Pregunta de investigación

¿Cuáles son las poblaciones virales de origen eucarionte presentes en la orofaringe y materia fecal de lactantes menores sanos y cómo cambia la diversidad y dinámica de estos virus a lo largo del primer año de vida de los niños?

Justificación

Como un primer paso para determinar si existe un posible efecto del viroma eucarionte sobre el desarrollo y salud de los infantes, es importante identificar y describir cual es la diversidad y dinámica de las poblaciones de virus eucariontes durante el desarrollo de un lactante menor sano durante su primer año de vida y cómo pudiesen influir factores externos, tales como alimentación o vacunas, en el establecimiento del viroma.

Hipótesis

Se identificarán poblaciones virales de origen eucarionte en la orofaringe y materia fecal del lactante menor sano a lo largo de los primeros 12 meses de vida. La composición de estas poblaciones será diversa y cambiará con los períodos de desarrollo del menor sano y la interacción con factores externos.

Objetivos

General:

Caracterizar por medio de un enfoque metagenómico el viroma eucarionte de la orofaringe y materia fecal de lactantes menores sanos en un estudio longitudinal, con muestras a partir de los 15 días de nacimiento y hasta los 12 meses de edad, para describir la composición y dinámica del viroma en estos dos nichos.

Específicos:

1. Caracterizar la composición y dinámica del viroma eucarionte de la orofaringe y materia fecal en las etapas tempranas de la vida.
2. Determinar si existen especies compartidas o similitudes entre la composición del viroma orofaríngeo y el de la materia fecal (viroma gastrointestinal).
3. Describir si la presencia de algún virus se relaciona con alguna etapa de desarrollo importante del niño, entre otros, el reconocimiento de objetos, el gateo y la alimentación, o con algún factor externo.
4. Describir familias que puedan ser relevantes dentro de los hallazgos, ya sea por su relevancia clínica o abundancia.

Material y métodos

Este es un estudio longitudinal, observacional y descriptivo que se llevó a cabo en el laboratorio de virología molecular, en el Instituto de Biotecnología, UNAM campus Morelos, en colaboración la Dra. Cecilia Ximénez de la Facultad de Medicina de la UNAM, el Dr. Javier Torres del IMSS, y la participación de la Coordinación de Salud en Xoxocotla y dos centros de la secretaría de salud del poblado.

Población de estudio:

Este estudio se llevó a cabo con la cooperación de los servicios de salud y el centro de salud de la comunidad de Xoxocotla, Morelos. Se invitó a participar a mujeres embarazadas sanas que cursaban el tercer trimestre de embarazo y que habían acudido de forma regular a control de embarazo en la unidad de salud de la comunidad. Información detallada, tiempo de duración y la seguridad de poderse retirar del estudio en cualquier momento fue explicada a cada una de las participantes que aceptaron ingresar, después se obtuvo el consentimiento informado.

Una vez firmado el consentimiento, se obtuvo la siguiente información de la madre: cuestionario, historia clínica detallada con antecedentes sociodemográficos, socioeconómicos y antecedentes patológicos (Tabla 1; Información sociodemográfica). Además, se realizó el compromiso que si durante el tiempo de estudio el bebé presentaba algún síntoma de vías respiratorias o gastrointestinales, se debía notificar al centro de salud y el personal pasaría a la casa para tomar una muestra extra de heces y exudado faríngeo.

Al momento del parto se notificó al personal de la clínica y se programó la primer visita, la cual fue en promedio a los 22 ± 14 días de nacido para la toma de la primer muestra orofaríngea y fecal. Estas muestras se tomaron de forma mensual hasta que los niños cumplieron 12 meses de vida. Adicionalmente, se llevó un registro de hábitos de lactancia, alimenticios (una vez ocurrida la ablactación), vacunas, toma de antibióticos y desarrollo del lactante. También se realizaron mediciones de peso y talla, todos los registros se realizaron cada semana durante los primeros 3 meses y cada 15 días los siguientes 9 meses (Tabla 1; información general).

Tabla 1. Información general de los niños y las condiciones de vida recolectados durante el estudio.

Información General						Información sociodemográfica									
Infante	Fecha de nacimiento	Día de primer toma de muestra	Parto	Sexo	Percentil promedio de peso	Hacinamiento	Almacenamiento de agua	Consumo de agua	Drenaje	Cocina	Material del suelo	Cucarachas	Roedores	Animales domésticos	Edad de introducción de otros líquidos
2	04-03-2015	15	Vaginal	M	50th	Si	Contenedor	Llave	No	Estufa	Tierra	No	No	Si	4
4	15-04-2015	15	Césarea	M	15th	No	Contenedor	Llave	No	Madera	Concreto	Si	Si	Si	3.5
5	25-04-2015	16	Césarea	F	25th	No	Cisterna	Botella	Si	Estufa	Concreto	Si	No	Si	2
6	09-06-2015	13	Césarea	M	50th	Si	Contenedor	Llave	Si	Madera	Concreto	No	No	Si	2
7	23-06-2015	16	Vaginal	F	25th	Si	Cisterna	Botella	No	Madera	Tierra	No	No	Si	1
9	28-11-2015	53	Vaginal	F	75th	Si	Cisterna	Botella	No	Estufa	Tierra	Si	Si	Si	3.5
11	16-10-2015	26	Césarea	F	50th	No	Contenedor	Llave	No	Madera	Tierra	Si	Si	Si	5
15	29-12-2015	8	Vaginal	F	95th	No	Contenedor	Botella	Si	Estufa	Concreto	Si	Si	Si	0.5
19	1-05-2016	39	Vaginal	F	15th	No	Cisterna	Botella	Si	Estufa	Concreto	Si	Si	No	1

Criterios de inclusión y exclusión:

	<i>Inclusión</i>	<i>Exclusión</i>
<i>Madre</i>	<ul style="list-style-type: none">• Mujeres con embarazo normo-evolutivo hasta el tercer trimestre de embarazo y pertenecientes al seguro popular• Domicilio estable en la comunidad• Haber acudido con periodicidad a consultas de seguimiento y control de embarazo	<ul style="list-style-type: none">• Antecedentes de enfermedades crónico degenerativas o embarazo concomitante con enfermedades crónico degenerativas• Antecedente de embarazo de alto riesgo• Gestación concomitante con patologías
<i>Hijo</i>	<ul style="list-style-type: none">• Niños nacidos a término• Productos únicos	<ul style="list-style-type: none">• Enfermedades por malformaciones congénitas• Bajo peso al nacer• Prematuridad

Tamaño de muestra:

Inicialmente se reclutaron 20 mujeres embarazadas, para dar seguimiento a sus respectivos bebés en el período comprendido entre los años 2015 y 2017, pero solo fueron considerados para este estudio los niños que tuvieran más del 80% de las muestras, por lo que se analizaron únicamente 9 niños.

Consideraciones éticas:

Este estudio fue aprobado por los Servicios de Salud del Estado de Morelos, el Comité de Investigación y Ética de la Facultad de Medicina de la Universidad Nacional Autónoma de México (Proyecto #088/2014) y el Comité de Bioética del Instituto de Biotecnología (Proyecto #261). La aprobación de los comités se basó en la Norma Oficial Mexicana (NOM-012-SSA3-2007), así como en los acuerdos de Helsinki.

Toma de las muestras:

Las muestras fecales fueron tomadas por las madres en un recipiente estéril, de acuerdo con el día que estaba agendada la visita. Si para el momento de visita no había muestra fecal, el personal de la clínica realizaba otra visita al día siguiente; mientras tanto, la muestra se guardaba en el congelador del refrigerador casero. Para minimizar la posible contaminación de las muestras fecales, las madres fueron entrenadas por una enfermera capacitada para recolectar las muestras de los pañales, los cuales se utilizaron al revés para evitar que la muestra se absorbiera. Las muestras orofaríngeas se tomaron al momento de la visita a casa

con un hisopo estéril (rayon-tipped, BD BBL) y se colocaron en 1 ml de medio de transporte viral (Microtest M4-RT, Remel). Una vez colectadas, ambas muestras se guardaron a -20 °C en la clínica de la comunidad por una semana y después se almacenaron a -70°C en el Instituto de Biotecnología hasta el momento de la secuenciación.

Extracción de ácidos nucleicos:

El manejo y extracción de ácidos nucleicos de las muestras orofaríngeas, fecales y controles fue realizada por el Dr. Pavel Isa (Instituto de Biotecnología, UNAM), de acuerdo al método descrito por Sorber y *co/s.* [40]. De manera breve, las muestras fecales se homogeneizaron utilizando un batidor de perlas (Biospec Products, USA), con una mezcla de 100 mg de perlas de vidrio de 150 – 212 μm (Sigma, USA) y 10% de cloroformo en 1 ml de amortiguador de solución salina de fosfato. Después, las muestras, ya sean orofaríngeas o fecales homogeneizadas, se centrifugaron a 2000xg y el sobrenadante se filtró utilizando un filtro Spin-X de 0.45 μm (Costar, NY), centrifugando a 10,000xg por 10 a 15 min, según fuera necesario. Posteriormente, las muestras filtradas se trataron con 1 μl de turbo ADNasa (Ambion, USA) con amortiguador 10x del mismo kit de turbo DNAsa y 2 μl de ARNsa (2 mg/ml, Sigma, USA), y se incubaron durante 30 min a 37°C. Los ácidos nucleicos se extrajeron utilizando el kit PureLink viral RNA/DNA (Invitrogen, USA) y se diluyeron en agua libre de ARNsas.

Los ácidos nucleicos extraídos (ADN y ARN) se amplificaron de forma aleatoria; el cADN se obtuvo con transcripción reversa de ARN (RT-PCR) utilizando el kit SuperScript III (Invitrogen, USA) con los cebadores 5'-GTTTCCCAGTAGGTCTCN9-3'. La segunda hebra se generó con dos rondas de síntesis utilizando Secuenasa 2.0 (USB, USA). Se amplificaron el ADN y cADN utilizando el cebador 5'-GTTTCCCAGTAGGTCTC-3' y la polimerasa Phusion High Fidelity (Finnzymes) por 10 ciclos. Se limpió el material amplificado utilizando el kit ZYMO DNA. Como control de extracción, este mismo proceso se llevó a cabo con 2 muestras de agua estéril (control 1 y 2). No se observaron virus eucariontes en la secuenciación de estas 2 muestras control. Una descripción más detallada de esta metodología se puede encontrar en Martínez y *co/s.* [41].

Secuenciación:

A partir de los ácidos nucleicos extraídos se prepararon las bibliotecas de secuenciación utilizando el kit de Nextera XT DNA (Illumina), con el cual se fragmentaron los ácidos nucleicos en segmentos de 200 pares de bases. La secuenciación se llevó a cabo en la plataforma NextSeq 500 (Illumina) generando lecturas pareadas de 75 nucleótidos en la Unidad Universitaria de Secuenciación Masiva de DNA del Laboratorio Nacional de Apoyo Tecnológico a las Ciencias Genómicas localizado en el Instituto de Biotecnología. El llamado de bases se realizó utilizando el software de Illumina de Analisis de Tiempo Real v1.18.54 y el

demultiplexado, es decir, la recuperación de la información, se hizo con el programa bcl2fastq v2.15.0.4.

Confirmación de presencia de virus en muestras:

Para confirmar la presencia de algunas especies virales (*Tropical Soda Apple* mosaic virus, *Pepper Mild* mottle virus) se realizó PCR dirigida.

Análisis bioinformático:

Análisis bioinformático general:

Se utilizó un *pipeline* «hecho en casa» (Fig. 1);

- 1) Calidad y pre-procesamiento de las lecturas: se revisó la calidad de las lecturas con el software FastQC v0.11.7 [42], para después remover adaptadores, bases de baja calidad, lecturas de baja complejidad y lecturas menores a 40 pares de bases utilizando los programas Cut-Adapt v1.8 [43] y FastP v0.19.4 [44]. Como se realizó amplificación de los ácidos nucleicos para poder secuenciarlos, se eliminaron las secuencias de DNA (lecturas) duplicadas utilizando el programa CD-HIT-Dup v4.6.8 [45].
- 2) Lecturas Válidas: se utilizó el programa Bowtie2 v2.3.4.3 [46] para buscar (mapear) las lecturas utilizando como referencia el genoma humano, tomado de Genbank (GRCh38.p13) [47,48]. Posteriormente, aquellas lecturas que no mapearon contra humano se utilizaron para mapear contra la base de Silva (SSU Ref and LSU Parc 132) [49] para eliminar lo más posible los genes ribosomales (16S, 18S, 23S, 28S y 5.8S). Una vez eliminadas las lecturas de genes humanos y ribosomales, se consideró el set como lecturas validas.
- 3) Identificación de lecturas: utilizando como referencia las bases de datos actualizadas a Marzo 2020, se mapearon a nivel de nucleótido todas las lecturas validas de cada una de las muestras, utilizando el programa BbMap v38.26 [50]. Se mapeó contra la base de bacteriófagos; aquellas lecturas que no se identificaban como pertenecientes a los fagos, se mapearon contra la base de todos los virus registrados en Genbank, incluyendo virus de organismos procariontes, y aquellas que no mapearon contra esta base se utilizaron para mapear contra la base de secuencias de bacterias.
- 4) Ensamblés *de novo*: todas las lecturas identificadas, ya sea como posibles fagos, virus o bacterias, se utilizaron para hacer ensamblés *de novo* con el programa SPAdes v3.14.1 [51]. Para ahorrar tiempo computacional, y evitar un solapamiento de los contigs formados, se hizo una clusterización al 95% de identidad utilizando CD-HIT v4.6.8.

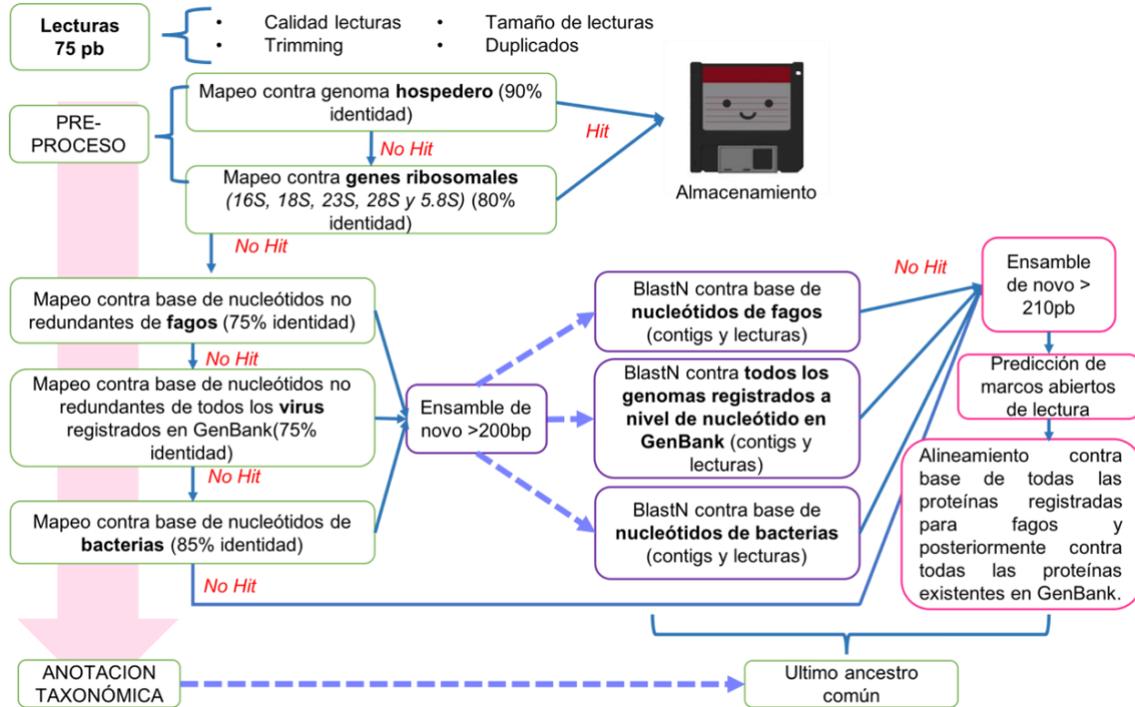
Después, para saber cuántas lecturas formaron cada contig representativo creé un programa en idioma Perl, el cual tiene la función de contar las lecturas que forman cada contig y asigna este conteo como magnitud. A las lecturas que no se ensamblaron se les asignó en automático “magnitud=1”.

- 5) Identificación taxonómica: posteriormente, se utilizó el alineador BLASTn v2.11.0 [52] para hacer una asignación taxonómica de los contigs y las lecturas a cada una de las siguientes bases a nivel de nucleótido no redundante: fagos, todos los genes descritos a nivel de nucleótido en NCBI (todo nt) [48] y bacterias con un valor-E de 0.001 80% de cobertura, 70% de identidad y pidiendo 20 alineamientos por cada lectura o contig para reducir falsos positivos.
- 6) Ensamblados *de novo* y predicción de marcos abiertos de lectura: aquellas lecturas que no tuvieron asignación taxonómica, se extrajeron con un programa en idioma Perl nuevamente hecho por mí y concatenadas con aquellas lecturas que en el paso 3 no pudieron ser asignadas. Posteriormente, se ensamblaron *de novo* con el programa SPAdes v3.14.1 [51] con un tamaño mínimo de 210 pares de bases. A partir de esos contigs se realizó una predicción de marcos abiertos de lectura con el programa Prodigal v2.6.3 [53].
- 7) Identificación taxonómica a nivel de proteínas: los contigs se alinearon contra fagos y posteriormente contra todas las proteínas no redundantes registradas en la base de datos ‘nr’ de NCBI [48], utilizando el programa Diamond v9.9.22 [54].
- 8) Construcción de matriz de especies virales: se utilizó el programa MEGAN v6.19.7 [55], el cual permite el análisis de set de datos metagenómicos de gran escala con un algoritmo que crea un árbol con el último ancestro común para analizar los resultados de las asignaciones a nivel de nucleótido y proteína de los contigs y las lecturas de cada una de las muestras. Aquí la magnitud se utilizó como un contador de observaciones, y la base de referencia utilizada fue la de Julio-2020 (creada por los desarrolladores del programa). Se utilizó el modo “nativo” con un 10% de corte para el puntaje del mejor hit, por lo tanto, el mejor hit tiene que tener una diferencia de más del 10% del puntaje para ser asignado a una especie diferente.

De esta forma, se reduce el riesgo de falsos positivos; contigs o lecturas asignadas a especies de forma muy específica llegan a las hojas del árbol, mientras que lecturas o contigs que son alineados con regiones conservadas se asignan a órdenes más altos (género, familia). Este programa se utilizó para crear una matriz de conteos crudos de lecturas y contigs a nivel de familia y especie para cada una de las muestras. Además

se realizó una curación manual aleatoria una vez obtenida la asignación taxonómica para confirmar que los alineamientos correspondieran a la información.

Figura 1. Diagrama de flujo de análisis bioinformático general. La palabra *hit* hace referencia a aquellas lecturas que mapearon contra la base de referencia empleada.



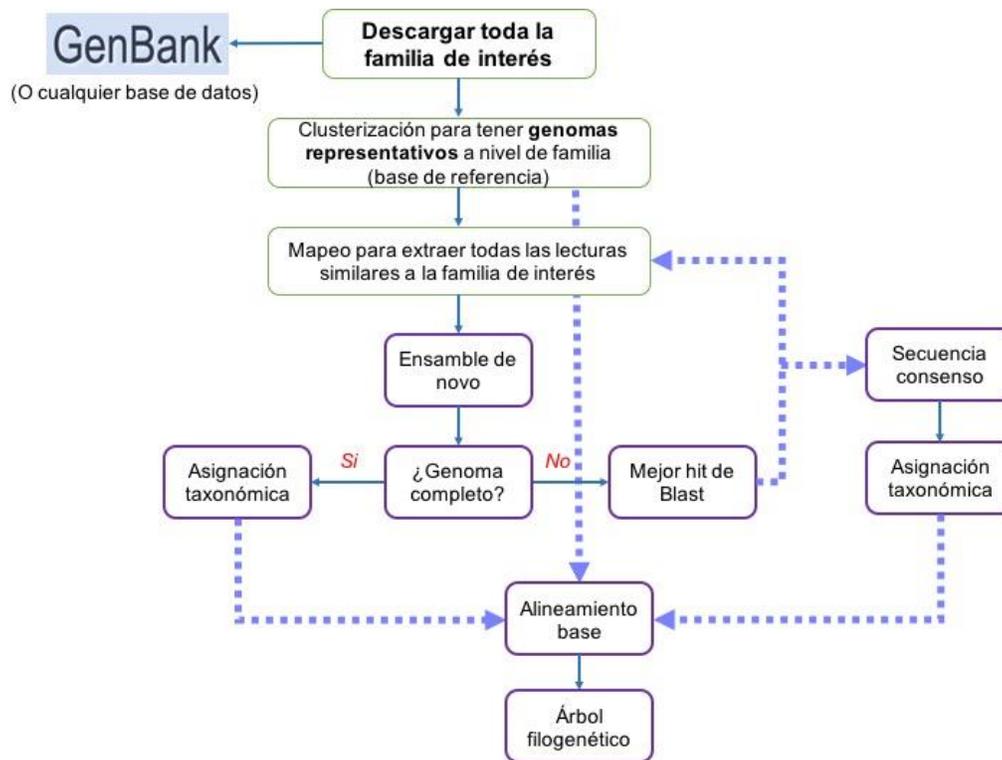
Análisis bioinformático particular:

Se realizaron análisis específicos para las siguientes familias; *Caliciviridae*, *Reoviridae*, *Anelloviridae*, *Astroviridae*, *Picornaviridae*, *Parvoviridae* y *Virgaviridae*. En todos los casos, se partió desde principios similares y después los análisis se afinaron de acuerdo con la información obtenida (Fig. 2).

- 1) Identificación de lecturas por familia de interés: primero se descargaron todos los genomas completos registrados en Genbank [48] para la familia de interés, posteriormente se hizo clusterización a diferentes porcentajes de identidad utilizando Cd-Hit v4.6.8 [45], con el afán de tener genomas representativos. Esta base clusterizada se utilizó para extraer todas las secuencias similares a la familia a partir de las lecturas válidas con el programa BbMap v38.26 [50].
- 2) Ensamblaje de novo:

- a. Se realizaron ensamblajes *de novo* con el programa SPAdes v3.14.1 [51] o IDBA v1.1.3 [56].
 - b. Secuencias consenso: si no se lograba obtener un genoma completo, se realizaba una secuencia consenso buscando (mapeando) el contig a la referencia más similar encontrada con la herramienta en línea de Blast [52], utilizando el programa Bowtie2 v2.3.4.3 [46]. Después, se utilizó Samtools v1.9 [57] para tener la primer secuencia consenso. Se realizó el llamado del consenso con iVar v1.3.1 [58] utilizando una calidad de base >20 y una profundidad de cobertura mínima de 2x; de lo contrario, la base se reemplazaba con una N. Para el llamado de polimorfismos de nucleótido único se utilizó un punto de corte mínimo de 10x de cobertura. Cada consenso o genoma completo se verificó utilizando BLAST v2.7.1 [52].
 - c. En algunas familias no se hizo búsqueda de secuencia consenso, en cambio se utilizó el programa MEGAN v 6.19.7[55], el cual ofrece una secuencia consenso a partir de los “mejores hits de blast” introducidos al programa. Estos se deben extraer por cada muestra.
- 3) Análisis filogenético: para realizar los análisis filogenéticos se realizó un alineamiento base utilizando el programa Mafft v7 [59] utilizando genomas completos obtenidos y genomas de referencia. En caso de tener genomas parciales, éstos se agregaron después al alineamiento, nuevamente utilizando Mafft v7 [59].
- a. Construcción árbol filogenético: Se realizó el cálculo del mejor modelo evolutivo para el alineamiento utilizando el programa MEGA vX [60]. Finalmente, se construyeron árboles filogenéticos con el método evolutivo indicado con 1000 permutaciones utilizando el programa RAXML v8.2.12 [61] o IQTree[62], según fuese necesario.
- 4) Detección de polimerasa para tipificación viral:
- a. Familia *Caliciviridae*: se utilizó la herramienta en línea ‘Human Calicivirus Typing Tool’ [63], la cual hace un alineamiento dirigido a las dos regiones que se utilizan para identificar a los calicivirus: RdRp para norovirus (región de la polimerasa) y la proteína VP1, la cual codifica la cápside, para norovirus y sapovirus.
 - b. Familia *Reoviridae*: se realizó con la herramienta en línea ‘determinación del genotipo de rotavirus A’ de la base de datos de patógenos del NIAID y ViPR [64], esta herramienta predice los genotipos de rotavirus utilizando un análisis de segmento por segmento del genoma de rotavirus. De forma computacional crea un alinamiento con los segmentos de referencia utilizando diferentes programas (Blast, ClustalW, Muscle y VIGOR) y da un porcentaje promedio de identidad.

Figura 2. Diagrama de flujo análisis bioinformático particular para las secuencias de familias virales



Análisis estadístico:

A pesar de que se describen todas las especies de virus eucariontes encontradas, el análisis ecológico y estadístico se enfocó a las especies de virus que se sabe se replican en humanos.

- 1) Lecturas sobrerrepresentadas: ya que algunas lecturas podrían estar sobrerrepresentadas debido a sesgos en la secuenciación, la matriz de conteos crudos se revisó de forma manual contra los alineamientos a nivel de especie utilizando el programa Megan [55].
- 2) Normalización: para eliminar posibles hits falsos, se eliminaron observaciones menores a 4 por cada especie en cada muestra, y se corrigieron las diferencias de profundidad de secuenciación de cada muestra dividiendo el número de conteo de lecturas virales asignadas entre las lecturas validas de cada muestra y después se normalizó a una profundidad de 5 millones por muestra, dando como resultado una matriz de datos normalizados a nivel de especie, y otra a nivel de familia.
- 3) Eliminación de valores atípicos: En algunas muestras se tuvieron valores totales atípicos a pesar de la normalización, por lo que se eliminaron las muestras con el 5% superior e

inferior en abundancia total de lecturas, a excepción de las muestras del primer trimestre de vida, las cuales se esperaban con abundancia baja.

- 4) Realización de gráficas: esta matriz de datos normalizada se utilizó para hacer las figuras. Las gráficas se hicieron con los paquetes ggplot2 [65] y cowplot [66] con el programa R v3.6.3 [67].
- 5) Medidas ecológicas: se realizaron con el paquete Vegan [68], en el programa R v3.6.3 [67].
 - a. Para las medidas de alfa diversidad se calculó la riqueza como el número esperado de especies 'Chao1', y la diversidad se calculó con el índice de Shannon.
 - b. Para la diversidad beta, se utilizó el índice de Bray-Curtis; las distancias generadas se utilizaron para los análisis de coordenadas principales. Los centroides y la dispersión se calcularon con la función ordispider del paquete de Vegan [68].
 - c. Para comparaciones individuales a nivel de especie se utilizó la herramienta LEfSe v1.39.5.0 [69] disponible en Galaxy [70].
- 6) Estadística aplicada: con toda la información colectada de los niños (Tabla 1) se realizaron análisis según fue necesario con las medidas de alfa diversidad: para comparar variables que tenían medidas repetidas (trimestres, semestres, tractos) se utilizó la función aovp del paquete lmp[71] con permutaciones por defecto (5000); para comparaciones entre y dentro de los grupos se consideró la variable como factor fijo, mientras que los niños y las muestras tomadas en cada mes fueron consideradas como factores aleatorios, con las muestras anidadas a los niños para considerar las muestras repetidas.

Las variables de alimentación se analizaron con un modelo de muestras de antes y después y se compararon utilizando la prueba de Wilcoxon para muestras pareadas. Variables que agrupaban a los niños (modo de nacimiento, peso promedio, consumo de agua, etc.) fueron analizadas con la prueba U-Mann-Whitney utilizando puntos específicos: primera muestra tomada, muestras del 3er, 6to, 9no y 12vo mes. Para la beta diversidad se utilizó una prueba no paramétrica multivariada (PERMANOVA) con 1000 permutaciones con estrata. Para contrastar más de dos grupos se utilizó la función pareada de Adonis. Para todas las pruebas la varianza fue verificada con la prueba de Levene o betadisper del paquete Vegan [68], según fue necesario. Se consideró significativa una $p < 0.05$.

Resultados y discusión

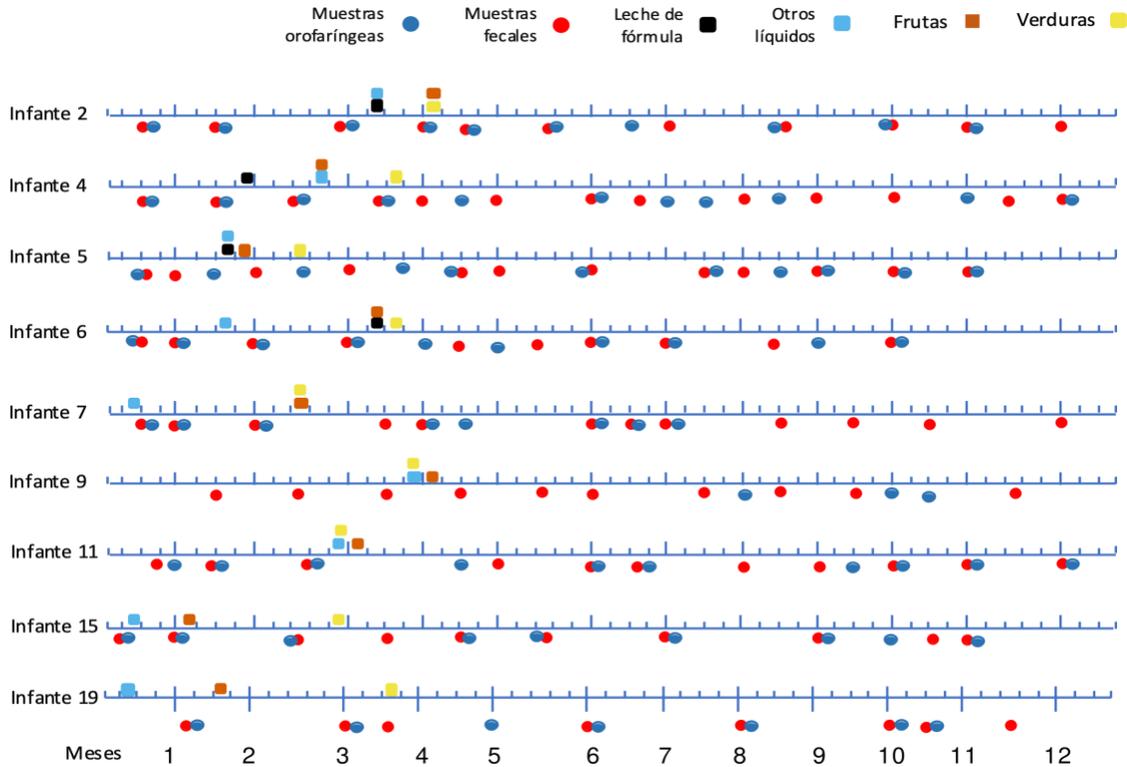
Muestras

De los 9 niños estudiados, se obtuvieron un total de 99 muestras orofaríngeas y 106 fecales, así como información respecto a su alimentación y vivienda; sin embargo, se analizaron solo 90 muestras orofaríngeas y 97 fecales, ya que las demás muestras fueron tomadas después del primer año de vida de los infantes. A pesar de los intentos y el excelente trabajo del personal de la clínica, no todas las muestras fueron recolectadas de forma pareada y el patrón fue más bien la colección de muestras fecales ligeramente desfasadas, en relación con el tiempo de toma de las muestras orofaríngeas.

Esto resulta lógico considerando el esquema de colección de muestras, en el que las muestras fecales se tomaron por las mamás y podían ser almacenadas en el congelador de la casa, así que probablemente aprovechaban cualquier oportunidad para tomar la muestra dentro de la semana agendada para la visita a domicilio. Asimismo, no se pudieron coleccionar todas las muestras de forma exactamente mensual (Fig. 3).

Esta diferencia de muestreo era también un poco esperada, pues tener seguimiento y apego en poblaciones infantiles puede llegar a ser complicado; en ocasiones las muestras orofaríngeas no podían tomarse porque los infantes acababan de comer, lo cual es una variable difícil de controlar en el primer año de vida. Sin embargo, como grupo, se tuvo un promedio de 10 ± 1 muestras por infante en cada uno de los nichos (Tabla S1).

Figura 3. Línea de tiempo de toma de muestras e introducción de alimentos. Los círculos señalan los días de la toma de muestras, los cuadrados el momento de introducción del alimento señalado. La línea inferior (eje x) muestra el mes de la toma de muestra, aunque de forma general etiquetamos las edades como mes completo o medio mes, de acuerdo con la semana de la toma de muestra.



Secuenciación

Se obtuvo un promedio de 29,015,329 y 20,464,999 lecturas de las muestras orofaríngeas y fecales, respectivamente. Después de verificar calidad y eliminar lecturas asociadas al hospedero y a genes ribosomales (16S, 18S, 23S, 28S y 5.8S), se terminó con un total de $1,204,867 \pm 868,988$ lecturas válidas de las muestras orofaríngeas y $4,049,622 \pm 2,612,729$ lecturas de las muestras fecales (Tabla S1).

Esta pérdida de lecturas durante el pre-procesamiento viene principalmente del paso de eliminar lecturas duplicadas (75% y 76% de lecturas eliminadas para las muestras orofaríngeas y fecales, respectivamente), las cuales estaban presentes debido a los pasos de amplificación necesarios para realizar la secuenciación de los ácidos nucleicos extraídos. Considerando todas las lecturas perdidas en los pasos de pre-procesamiento, las muestras orofaríngeas tuvieron una pérdida mayor de lecturas (96%), en comparación a las fecales (85.7%). Creemos que se debe principalmente por la dificultad que representa extraer ácidos nucleicos a partir de un hisopado con mucosa y saliva, así como a la posible interacción constante de la orofaringe con el medio ambiente, lo cual explicaría el porcentaje tan alto de genes asociados a secuencias ribosomales (16% contra 3%, respectivamente).

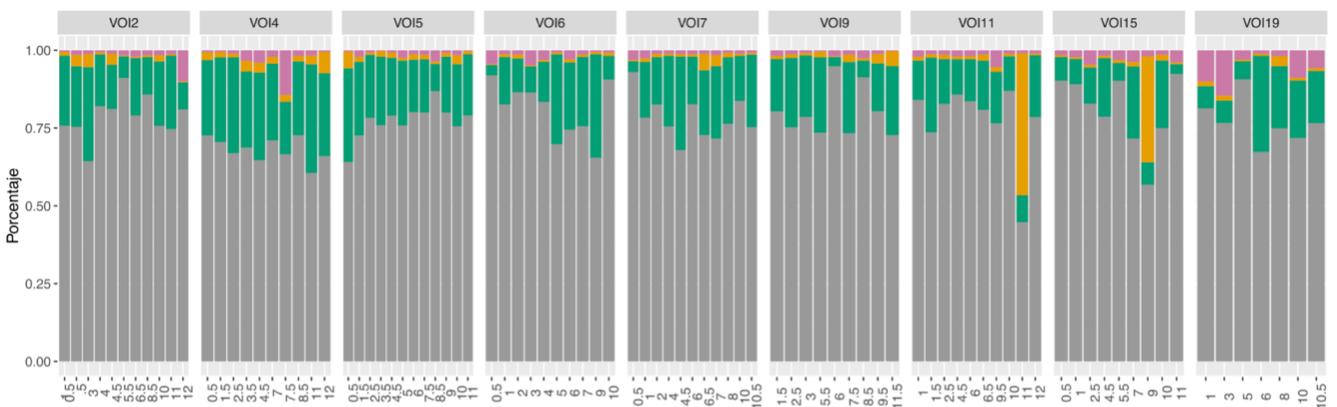
Procesamiento

A partir de las lecturas válidas, y con el afán de hacer el *pipeline* menos costoso en tiempo y recursos computacionales, se hicieron mapeos dirigidos hacia fagos, todos los virus registrados en GenBank y bacterias. Aquellas lecturas que no mapearon contra ninguna de estas bases de secuencias se asumieron como 'otros', sabiendo que éstas podrían contener genes de otros microorganismos, ARN no codificante o inclusive genes de virus no descritos.

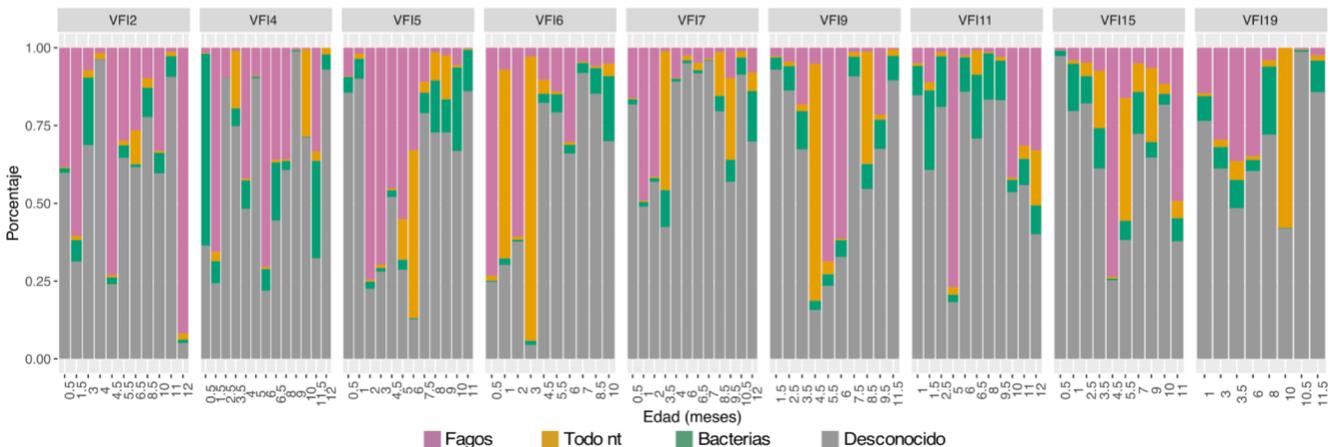
Esta separación con mapeos nos ayudó a generar contigs y hacer asignación taxonómica de forma más rápida y ordenada. Si bien el hacer un mapeo laxo (75%-80% de identidad) no garantiza que todas las lecturas mapeadas se asignaron con un alto porcentaje de identidad de forma taxonómica, sí proporciona una idea inicial sobre cómo está distribuida la población de microorganismos presentes en cada muestra (Fig. 4).

Figura 4. Distribución de lecturas mapeadas contra fagos, todos los genes registrados en GenBank (etiqueta todo nt) y bacterias en las muestras orofaríngeas y fecales, representada en porcentajes. Cada panel corresponde a un niño, y su etiqueta se observa en la parte superior. La edad asignada en cada muestra se observa en el eje inferior. VOI = viroma orofaríngeo infantil. VFI = Viroma fecal infantil. Cada número representa a un infante.

Orofaringeo



Fecal



Resulta interesante observar que, a diferencia de las muestras fecales, las muestras orofaríngeas tenían mayor cantidad de lecturas asignadas a 'desconocido', así como a bacterias. Esto quizá podría reflejar la constante exposición de la orofaringe al medio ambiente, pues al momento de asignar las lecturas clasificadas como 'desconocidas' a nivel de proteína, encontramos genes de plantas, hongos e insectos. En el caso de las muestras gástricas, no es sorprendente encontrar una mayor cantidad de fagos, como se ha descrito previamente en estudios de viroma en infantes, en los que se ha sugerido una correlación inversa entre la presencia del número de especies de fagos (riqueza y diversidad) con el número de bacterias [22,29,36,72,73]. En ambos nichos se tienen algunas muestras donde pareciera haber un sesgo de secuenciación hacia los virus en comparación a las demás muestras, sin embargo esas diferencias se normalizaron *a posteriori*.

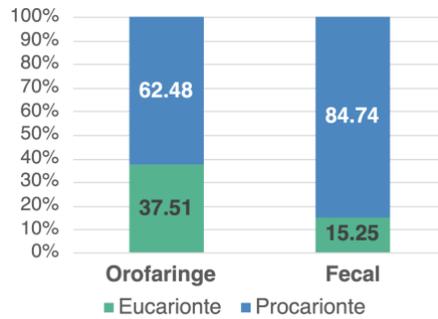
Normalización

A pesar de ser procedimientos estandarizados y hechos con cuidado, sabemos que siempre habrán posibles diferencias al momento del muestreo, extracción y secuenciación de los ácidos nucleicos entre las muestras, y por eso la normalización es importante. Para normalizar se consideraron diferentes estrategias y se fueron descartando en relación a nuestros datos, considerando como mejor opción normalizar de acuerdo con la profundidad de secuenciación, para así respetar los posibles cambios ecológicos que tendrían que suceder en la evolución del tiempo de las muestras, el tamaño de los genes secuenciados y también para poder comparar entre los dos nichos.

Asignación taxonómica y análisis ecológico

A nivel taxonómico una gran proporción de lecturas fueron asignadas a virus procariontes, siendo estos 2 veces más abundantes en la orofaringe y 6 veces más abundantes en las muestras fecales, al compararlos con los virus eucariontes (Fig. 5). Como se mencionó, diversos estudios han descrito la gran abundancia y dinámica de virus procariontes en muestras fecales, por lo que estas diferencias de proporciones de virus no nos parece descabellada. A partir de esta sección, se verán por pequeños capítulos los análisis por cada nicho, análisis de familias particulares y la comparación de ambos nichos. Para hacer más sencilla la historia, cada segmento incluirá introducción, resultados y discusión de éstos.

Figura 5. Porcentaje de lecturas identificadas a nivel taxonómico para virus procariontes y eucariontes en muestras orofaríngeas y fecales.



Capítulo 1: El viroma de la orofaringe

Conocida como la parte oral de la faringe, la orofaringe es una zona anatómica híbrida, es decir, tiene funciones respiratorias y digestivas; facilita el pasaje de aire, limita el flujo faríngeo-nasal, propulsa el bolo alimenticio como parte de la deglución y participa en la fonación [74]. Anatómicamente, se extiende desde el paladar blando hacia la punta superior de la epiglotis [75].

Debido a esta función híbrida, es de esperarse que los microorganismos que se encuentran en la orofaringe tengan un origen mixto, es decir, microorganismos que se respiren, así como aquellos que puedan ser ingeridos, a través de la comida o la saliva. De hecho, la saliva ha sido descrita como un acarreador de microorganismos de la boca hacia otras partes del cuerpo, y que la adherencia a los epitelios o mucosas de estos microorganismos depende de las características de cada uno de ellos [28]; sin embargo, no solo tiene el papel de acarreadora, también favorece el aclaramiento de microorganismos durante la deglución. En el caso de los virus, los cuales también pueden ser acarreados por la saliva, la colonización en algún nicho dependerá de su tropismo hacia las células dónde se encuentren [76].

En una revisión reciente, se estableció que las comunidades bacterianas de la microbiota son similares en los recién nacidos en diferentes partes del cuerpo: nariz, boca, intestinos y piel, y que esto está relacionado a la exposición durante el parto. Asimismo, se ha descrito que a los 6 meses la bacteriota oral y fecal de los niños es muy similar al de la boca, leche y areola de sus mamás, mostrando la influencia que tienen las exposiciones a la formación de la comunidad bacteriana. En el caso de las bacterias se cree que esta colonización compartida entre nichos se debe al pH menos ácido del estómago en los lactantes, lo cual permite que los microorganismos sobrevivan y colonicen el intestino [28]. En el caso de los virus, estos cambios de pH no afectan su capacidad para ser acarreados a diferentes nichos.

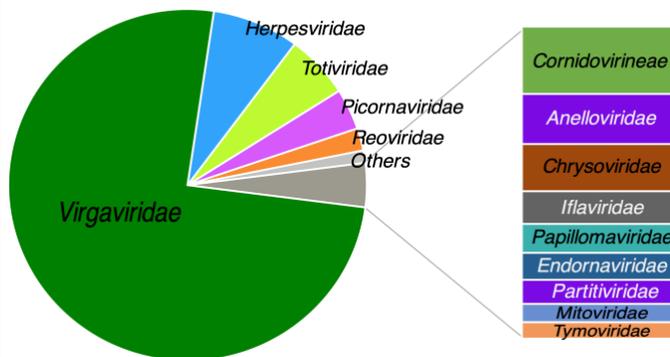
Respecto al viroma, no hay estudios enfocados a la orofaringe o a la cavidad oral en niños sanos menores de un año que describan, ya sea de forma transversal o longitudinal, los virus que podrían estar presentes [28]. Sin embargo, se ha establecido que las infecciones primarias en la cavidad oral están relacionadas a especies de la familias *Herpesviridae* y

Papillomaviridae [28]. En población adulta, los virus eucariontes de ADN más prevalentes son roseolovirus, lymphocryptovirus, betapapillomavirus, mastadenovirus y polyomavirus. Asimismo, se ha observado que en esta población el viroma oral varía mucho entre individuos y es estable con el paso del tiempo, pues suele ser influenciado por el ambiente donde las personas viven. Respecto a esto, se han observado viromas similares entre personas que viven en la misma casa [76].

Por otro lado, los viromas nasal, tanto de niños [77,78] como de adultos [79], de saliva en niños [80,81] y de placa periodontal adulta [76,81] han sido descritos con predominancia de las familias *Anelloviridae*, *Papillomaviridae* y *Herpesviridae*.

En nuestro trabajo se encontró un total de 52 familias como parte del viroma orofaríngeo, sin embargo, solo 5 de éstas representaban más del 1% del total de abundancia de las lecturas, y sumadas representaban poco más del 95% de las lecturas totales (Fig. 6). Tres de estas familias son conocidas por infectar humanos: *Herpesviridae*, *Picornaviridae* y *Reoviridae*, mientras que las otras dos, *Virgaviridae* y *Totiviridae*, infectan plantas y hongos (Tabla 2).

Figura 6. Grafica de pie de las familias con mayor abundancia de lecturas en las muestras orofaríngeas (Izquierda), en gris claro con la etiqueta 'Others' están aquellas familias que tienen menos del 0.05% de abundancia de lecturas. Gráfica de barras de familias con un porcentaje menor a 0.1% de abundancia de lecturas (derecha).



A nivel de especie, se encontraron 50 diferentes en las muestras orofaríngeas, con un promedio de 8 ± 5 especies virales y una media de 5,563 (187 - 1,980,824) lecturas por muestra. De estas 50 especies, 15 se conocen por replicarse en humanos y 27 en plantas. Basados en la frecuencia de especies, siete especies humanas y 14 de plantas tuvieron una baja frecuencia (menor al 10%) y una abundancia sumada de todas estas especies de 2.4% en todas las muestras. Creemos que estas especies podrían encontrarse en el ambiente o en la comida y fueron respiradas o ingeridas y por lo tanto no son ubicuas de la cavidad oral.

Por otro lado, 7 especies virales fueron las más abundantes, aunque a nivel de frecuencia de detección, 3 especies que se replican en plantas fueron las más comunes, es decir, se encontraron en más de la mitad de las muestras (Tabla 2) y se hablará de ellas más adelante.

Las otras 4 especies con mayor abundancia de lecturas infectan humanos, y su frecuencia es variable;

- 1) Betaherpesvirus 5 (citomegalovirus, CMV) fue la especie más abundante, representando 7.8% de las lecturas virales. Este virus se detectó en todos los niños de la cohorte, en algunos durante los primeros meses de vida (niños 2, 4, 15 y 19), mientras que en otros (niños 5, 7 y 15) se encontraron muestras positivas durante todo el año de estudio (Fig. 7). Se ha descrito al CMV como el patógeno viral más común en los recién nacidos, de hecho, se considera como una infección tan ubicua que se ha definido como una forma de inmunización natural en niños sanos [82] y no se ha relacionado con enfermedades en niños vigilados hasta sus 4 años de vida [83].

CMV es un virus ubicuo que genera una excreción prolongada, con una seroprevalencia de anticuerpos que varía entre 39.3% a 48% en hombres adultos, mientras que en mujeres varía entre el 24.6% y 95.7%, dependiendo en gran medida de la zona geográfica; la seroprevalencia es menor en países desarrollados[84]. El mecanismo de infección primaria posiblemente ocurre vía transmucosa, iniciando ya sea en el epitelio de la naso u orofarínge, para después pasar a los nódulos linfáticos [82]. Esta infección se puede dar por infecciones vía *in utero*, conocido como síndrome de TORCH, o en el período perinatal, ya sea por vía vaginal durante el parto o a través de fluidos como la saliva o la leche materna (LM), donde en niños sanos nacidos a término cursan la infección de forma asintomática.

Interesantemente, se demostró que la abundancia de CMV incrementa en la leche materna con el tiempo; en un estudio longitudinal se reportó un 8.8% de muestras positivas en las 2 primeras semanas post parto, y después incrementó al 26% de muestras entre los 45 y 90 días post-parto [85].

Sin embargo, a pesar de estos cambios en la leche, la prevalencia de excreción de CMV es bastante variable; en un estudio se reportó que de 278 mujeres sólo el 13% de ellas presentaron CMV en leche materna [83]. Asimismo, la prevalencia reportada de CMV en leche materna varía entre 13 y 50% en diferentes estudios, dependiendo de la región estudiada [86]. Por otro lado, se ha reportado que solo el 20% de los niños de mujeres que excretan CMV en leche materna se «infectan», sugiriendo que los infectados podrían tener déficit de inmunidad adaptativa [82]. Además, se cree que esta «baja»

prevalencia puede deberse a la protección conferida por la IgG materna [86]. Es importante tener en cuenta que en los estudios en los que se determina la presencia de un patógeno, únicamente a través de detectar parte de su material genético en realidad no se sabe si hay una infección activa. En estos estudios mencionados los porcentajes se basan en la detección del virus vía PCR y en la cantidad de virus infeccioso encontrado [82,86]. En un estudio de seroprevalencia en muestras de saliva de niños entre 0 y 5 años encontraron que 11% de los niños excretan citomegalovirus, teniendo un pico de prevalencia a los 3 años, pero mayor título viral al año de vida[87].

Si la prevalencia reportada de excreción es del 11% en saliva infantil, ¿por qué todos los niños de nuestro estudio tienen lecturas asociadas a esta especie viral? Asimismo, si los niños 4, 15 y 19 tienen muestras positivas a CMV en su primer mes de vida y el niño 4 nació de forma quirúrgica, ¿podríamos pensar en diversos factores que expliquen la presencia de este virus? Antes de intentar contestar esta pregunta es importante mencionar que no consideramos la presencia de CMV en las muestras como infecciones, pues no podemos confirmar si hay replicación viral, particularmente si el virus es ingerido junto con componentes de la respuesta inmune antagónica, como anticuerpos anti-CMV, leucocitos o citosinas anti-virales, los cuales se pueden encontrar en la leche materna.

Una posible explicación para la presencia de CMV durante el primer mes de vida es la transmisión de CMV vía leche materna y así tendríamos una prevalencia del 30% sobre el total de niños estudiados, ¿qué pasa con los otros 6 niños los cuales presentan muestras positivas mes y medio y hasta 6 meses después del nacimiento, mostrándonos que todos los niños tuvieron contacto con CMV? (Fig. 7). Una posible explicación es que hayan aspirado el virus, pues se ha reportado que CMV puede estar presente en el aire; un estudio llevado a cabo en una guardería infantil donde tomaron muestras de bioaerosol describió una abundancia de lecturas asignadas a CMV del 7.8% en muestras tomadas durante la primavera y un incremento en la abundancia relativa de esta especie durante el invierno [88], considerando que las pruebas de monitoreo para CMV se hacen en saliva, es adecuado asumir que los genes de CMV encontrados en el aire llegaron ahí por gotitas de saliva en el aire.

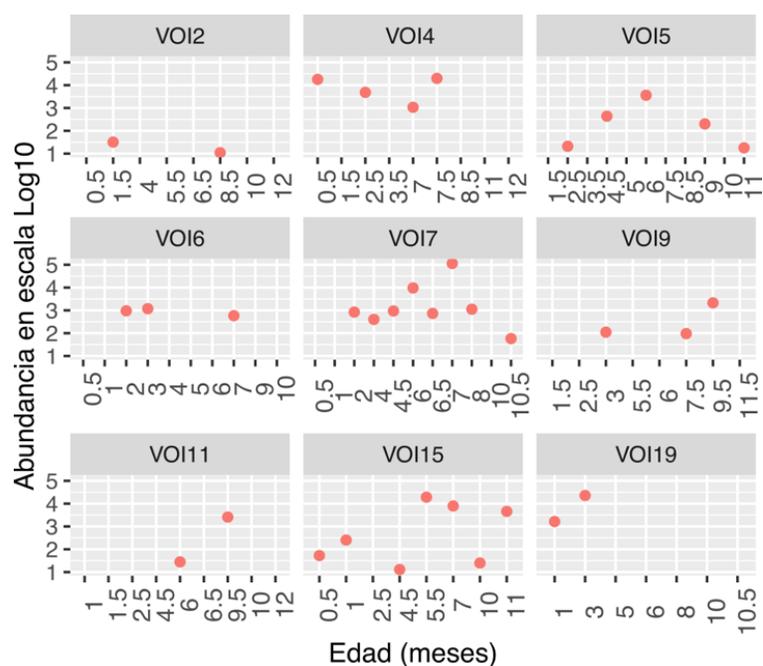
Nuestros niños no van a una guardería, pero es posible que convivan con niños mayores, asumiendo que tienen hermanos o primos con los que cohabitan. Asimismo, podríamos creer que hay un sesgo de muestreo y que quizá con más exudado faríngeo y mayor profundidad de secuenciación podríamos identificar más muestras positivas. Desde luego, estas son especulaciones, pues al analizar si la abundancia de CMV

podiese variar entre las condiciones de nacimiento o socioeconómicas de los niños, no encontramos ninguna diferencia significativa.

Otra observación interesante es que no se observaron patrones de aumento o disminución de la abundancia de este virus a lo largo del año de vida de los niños, lo cual podría deberse, desde una falta de profundidad en la secuenciación, hasta el hecho que al ser posibles «infecciones asintomáticas» su título viral es bajo, y por lo tanto la abundancia de lecturas relacionadas a CMV también, y en el caso de los niños que no tuvieron muestras continuas se debe considerar que podría tratarse de virus no infecciosos que estaban de paso.

También es posible que hubiera una toma de leche por parte del niño algunas horas previas antes de la toma de la muestra de la orofaringe, por lo cual estuviéramos viendo una mezcla de virus del niño, pero también de la madre (Fig. 7). Para determinar cuál es el origen de las secuencias virales tendrían que hacerse estudios pareados y enfocados a resolver este punto.

Figura 7. Abundancia de lecturas virales asignadas a la especie CMV en escala logarítmica base 10 para cada uno de los niños, mostrada en puntos. Las muestras dónde no hay puntos fueron negativas para CMV. VOI = viroma orofaríngeo infantil. Cada número representa a un infante y en la parte inferior (eje x) se observa la edad en meses de toma de cada muestra.



- 2) Rotavirus A se encontró en casi todas las muestras de los niños 2, 15 y 19 y fue la segunda especie más abundante en las muestras (2%) (Fig. 8). Al ser una especie

conocida por su replicación intestinal, que además fue identificada en nuestras muestras fecales, la asumimos como compartida entre ambos nichos, por lo que profundizaremos más adelante sobre los hallazgos de esta especie.

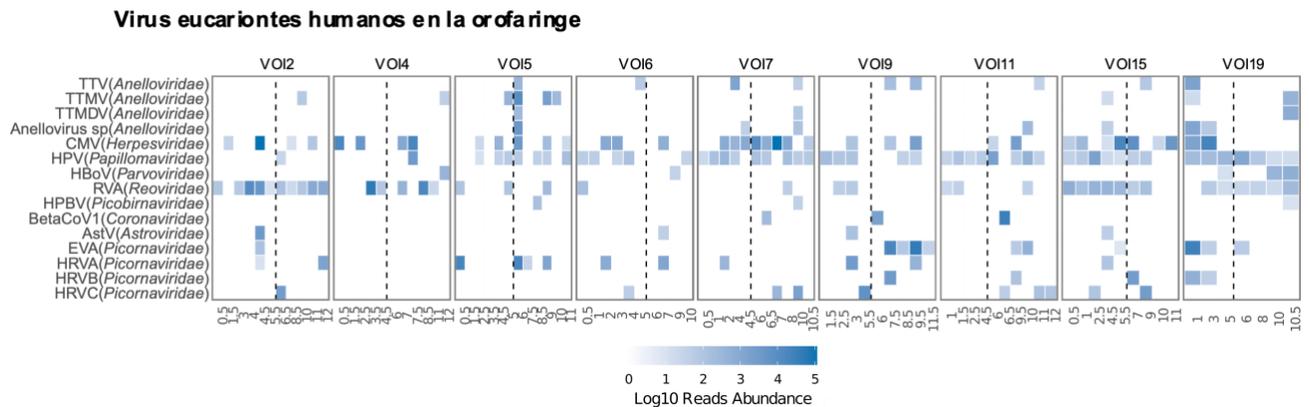
- 3) Enterovirus A (EVA) se identificó en 5 niños, con casos interesantes, como el caso del niño 9 donde fue detectado en cuatro muestras consecutivas, lo que nos sugiere una posible infección asintomática con excreción viral prolongada. La consideramos asintomática ya que no se reportaron síntomas de los niños durante el período de estudio; además, es interesante encontrar muestras continuas en la orofaringe, ya que no se encontraron en las muestras fecales tomadas en tiempos similares. Otro caso interesante es el del niño 19, donde esta misma especie fue encontrada de forma abundante durante los primeros meses de vida (Fig. 8).

La excreción asintomática de EVA ha sido descrita en estudios metagenómicos de muestras de saliva en niños con enfermedad pie-mano-boca, pero también en niños asintomáticos, con una menor abundancia de lecturas virales en los niños asintomáticos [80]. Aunque nuestra población permaneció asintomática y pudimos asignar taxonómicamente las lecturas, no podemos decir si se trata de infecciones productivas o si las secuencias virales encontradas representan virus infecciosos. Sin embargo, este estudio si es indicativo de los primeros contactos virales que se están estableciendo durante el desarrollo de los infantes.

- 4) Rinovirus A se encontró en 6 de los 9 niños (2, 5, 6, 7, 9 y 15), con una frecuencia de entre 1 a 4 muestras por niño. Un patrón similar, (niños 2, 6, 7, 9, 11 y 15), fue observado con la especie rinovirus C (Fig. 8). Debido a nuestra metodología no podemos asumir la presencia de genes que se asocian a especies virales como infecciones activas, sin embargo, la presencia asintomática de rinovirus humanos ha sido reportada previamente en muestras nasofaríngeas de niños, así como la excreción hasta 30 días post-infección en infantes infectados menores de un año[89], así que también es factible que estemos viendo posibles infecciones no esterilizantes.

Figura 8. Mapa de calor representando las abundancias de especies virales humanas que se replican en la orofaringe. Las abundancias están normalizadas y expresadas en escala de logaritmo 10. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Cada panel representa a un niño con el identificador en la parte superior de cada panel y la edad en meses en la parte inferior (eje x). VOI = viroma orofaríngeo infantil y cada número representa a un niño de la cohorte. Las líneas punteadas dividen los meses entre el primer y segundo semestre de edad. Especies que se replican en humanos: TTV, torque teno virus; TTMV, torque teno mini virus; TTMDV, torque teno midi virus; CMV, citomegalovirus; HPV, human papillomavirus; HBoV, human bocavirus; RVA, rotavirus A; HPBV, human picobirnavirus; BetaCoV1,

betacoronavirus 1; EVA, enterovirus A; AstV MLB3, Astrovirus MLB3; HRVA, rinovirus A; HRVB, rinovirus B; HRVC, rinovirus C.



Además de las 4 especies más abundantes descritas anteriormente, la especie papilomavirus humano (VPH) no tuvo una abundancia que llamara la atención, y fue más bien baja (0.27%), pero mostró la frecuencia más alta de todas las especies virales que infectan a humanos (presente en 56% de las muestras analizadas) (Tabla 2). Un análisis más fino nos permitió identificar 2 tipos de VPH: 165 y 93, ambos descritos en muestras cutáneas y por lo tanto, de bajo riesgo patogénico[90,91]. Asimismo, fue detectada en más del 80% de las muestras de los niños 7, 11, 15 y 19, incluyendo aquellas tomadas durante los primeros meses de vida (Fig. 8).

Esta llamativa frecuencia en las muestras mostró una decreciente tendencia, siendo más frecuente durante el primer trimestre (68% de las muestras) y menos frecuente después; 62%, 50% y 47%, respectivamente. La presencia de VPH en la mucosa oral de los recién nacidos es algo ya descrito, donde la vía de contagio se considera vertical, es decir de la madre al infante durante el parto, con una prevalencia de VPH oral de 22.5% al nacimiento y que disminuye hasta el 16.9% a los 2 meses de vida [92]. Estas prevalencias que decrecen nos sugieren que podríamos estar detectando virus de paso, los cuales no se están replicando, o bien, la adaptabilidad del sistema inmune el cual no permite mayor replicación conforme va creciendo el niño.

Aunque esta pareciera ser la explicación de la prevalencia en los primeros meses de vida de los niños de nuestra cohorte, las especies identificadas en nuestras muestras no corresponden a este tipo de infección al menos en los niños 11, 15 y 19 (Fig. 9). Además, en el caso de la niña 7, es posible que se trate del mismo virus, pero en caso de que no fuese, podríamos sugerir otra explicación: un estudio del viroma nasal en niños menores de 5 años en comunidades indígenas aisladas de las comunidades urbanas [77], en 3 lugares diferentes en Suramérica, describió la presencia de papilomavirus en 28 (44%) de las muestras, siendo la

segunda familia de virus más frecuentemente detectada. En un escenario similar al nuestro, también tuvo una baja abundancia, representando el 0.087% del total de lecturas.

De forma interesante, los niños que tuvieron más presencia de papilomavirus fueron aquellos que vivían en la comunidad menos aislada, sugiriendo que la presencia de éstos está correlacionada con la exposición a gente externa a la familia directa o en el caso de este estudio, de la comunidad [77]. Respecto a la exposición, se ha descrito que el VPH es una de las especies virales más abundantes encontradas en superficies de baños públicos [88], reafirmando que es un virus que podría estar circulando en el aire y su ingesta puede ser común y temprana.

Figura 9. Diversidad genética de los papillomavirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TVM2+F+I. Los genomas de referencia y raíz están en negritas.(b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.

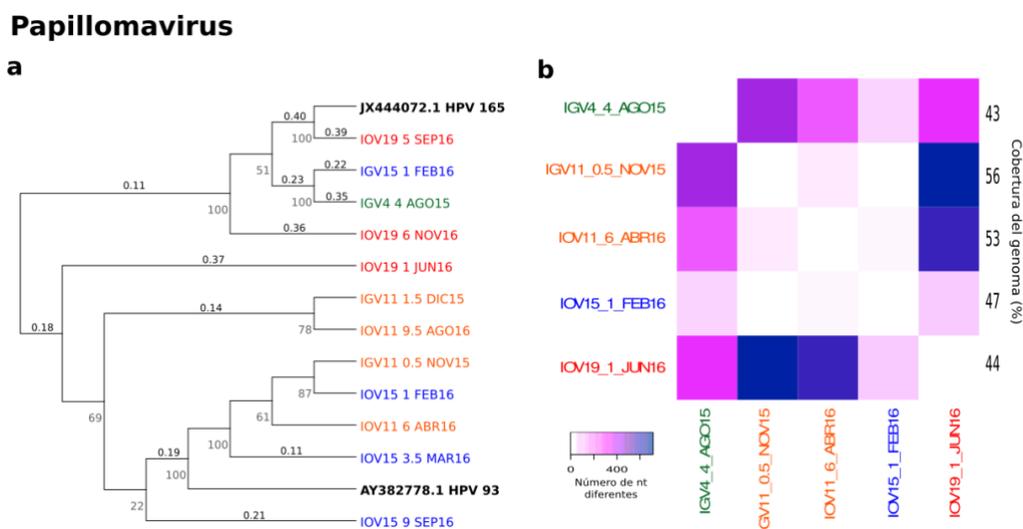


Tabla 2. Porcentaje de frecuencia y abundancia de especies virales encontrados en las muestras orofaríngeas. Marcados en negrita se señalan las especies con más de 1% de abundancia de las lecturas de ADN detectadas. El porcentaje de frecuencia (% Frecuencia) se calculó considerando todas las muestras de la cohorte (n=90). El porcentaje de abundancia (% Abundancia) se calculó con el total de lecturas entre las muestras y el total de lecturas asignadas por especie. El porcentaje de abundancia por hospedero (% Abundancia por hospedero) se calculó considerando el total de lecturas de especies que replican en cada hospedero.

Hospedero	Especies virales	Familia	% Frecuencia	% Abundancia	% Abundancia por hospedero
Humano	Torque teno virus	<i>Anelloviridae</i>	10	0.09	0.61
	Torque teno mini virus		10	0.28	1.89
	Torque teno midi virus		3	0.02	0.15
	Anellovirus sp		7.8	0.14	0.92
	Betaherpesvirus 5	<i>Herpesviridae</i>	42.2	7.85	51.72
	Human papillomavirus	<i>Papillomaviridae</i>	55.6	0.27	1.85
	Human bocavirus	<i>Parvoviridae</i>	5.6	0.03	0.22
	Rotavirus A	<i>Reoviridae</i>	42.2	2.05	13.52
	Human picobinavirus	<i>Picobinaviridae</i>	3.3	<0.01	0.07
	Betacoronavirus	<i>Coronaviridae</i>	3.3	0.78	5.24
	Astrovirus	<i>Astroviridae</i>	4.4	0.09	0.59
	Enterovirus A	<i>Picomaviridae</i>	12.2	2.14	13.50
	Rhinovirus A		13.3	1.15	7.38
	Rhinovirus B		5.6	0.15	1.03
Rhinovirus C	11.1		0.33	2.21	
Planta y/u Hongos	Genomoviridae sp.	<i>Genomoviridae</i>	4.4	0.04	0.05
	Persea americana chrysovirus	<i>Chrysoviridae</i>	16.7	0.51	0.62
	Scheffersomyces segobiensis virus L	<i>Totiviridae</i>	50.0	5.52	6.70
	Totivirus sp		14.4	0.05	0.06
	Puccinia striiformis totivirus		5.6	0.02	0.02
	Red clover powdery mildew-associated totivirus		15.6	0.19	0.24
	Bell pepper alphaendomavirus	<i>Endomaviridae</i>	15.6	0.13	0.16
	Endomaviridae sp		8.9	0.12	0.15
	Bell pepper mottle virus	<i>Virgaviridae</i>	4.4	0.05	0.06
	Pepper mild mottle virus		65.6	31.54	38.33
	Rattail cactus necrosis-associated virus		12.2	0.11	0.13
	Tropical soda apple mosaic virus		77.8	42.85	52.08
	Tobacco mild green mosaic virus		6.7	0.27	0.33
	Tomato brown rugose fruit virus		11.1	0.02	0.02
	Opuntia virus 2	<i>Alphaflexiviridae</i>	3.3	0.10	0.12
	Garlic virus		4.4	0.02	0.03
	Alphaflexiviridae sp		5.6	0.04	0.05
	Apple chlorotic leaf spot virus		8.9	<0.01	<0.01
	Betaflexiviridae sp	<i>Betaflexiviridae</i>	12.2	0.02	0.03
	Tymovirales sp.		5.6	0.12	0.15
	Procedovirinae sp	<i>Tombusviridae</i>	5.6	<0.01	<0.01
	Mitovirus sp.	<i>Mitoviridae</i>	21.1	0.13	0.16
	Amalgaviridae sp	<i>Amalgaviridae</i>	5.6	0.02	0.02
	Agaricus bisporus virus	<i>Partitiviridae</i>	6.0	0.07	0.08
	Partitiviridae sp		13.3	0.21	0.26
	Secoviridae sp	<i>Secoviridae</i>	7.8	0.03	0.04
Potyvirus sp	<i>Potyviridae</i>	10.0	0.04	0.05	
Hospedero desconocido	Dicistroviridae sp.	<i>Riboviria</i>	9	0.039	2.851
	Iflaviridae sp		18	0.335	24.469
	Picomavirales sp.		9	0.253	18.491
	Hubei sp		23	0.300	21.910
	Bat posalivirus	<i>Unclassified viruses</i>	14	0.302	22.056

Es importante resaltar que el hallazgo de secuencias virales no permite determinar si existen virus infecciosos en las muestras y tampoco si los virus se están replicando como resultado de una infección, o si sólo están presentes en la orofaringe como consecuencia de exposición ambiental o de rastros de su paso hacia el tracto gastrointestinal.

Análisis a nivel ecológico

Se hicieron análisis con los datos disponibles, tanto generales como sociodemográficos (Tabla 1, tabla S2), considerando como uno de los factores más relevantes la edad, pues la edad ha sido asociada con las etapas del desarrollo infantil. Al analizar el promedio de lecturas asignadas a especies por muestra, considerando a todos los niños en un sólo grupo, observamos una posible tendencia hacia el incremento de lecturas por trimestre (Fig. 10a). Sin embargo, una visualización individual de los niños nos mostró que esa tendencia no es generalizable en los infantes y que cada niño tiene más bien patrones individuales, con picos en el número de lecturas que se comportan de forma independiente a la edad (Fig. 10b).

Además de lo anterior, cuando se analizaron el número de especies por muestra, es decir la riqueza (Fig. 10c y d), y la distribución de la abundancia de lecturas entre el número de especies presentes, es decir la diversidad (Figs. 10e y f), no observamos ninguna tendencia clara de cambios entre los trimestres y tampoco se encontró significancia estadística al comparar los trimestres. Por otro lado, en las muestras de todos los niños se observó un a diferencia en el número de especies presentes en cada muestra, pero no en su abundancia. Esto sugiere que la dinámica de exposición de cada niño es bastante personalizada.

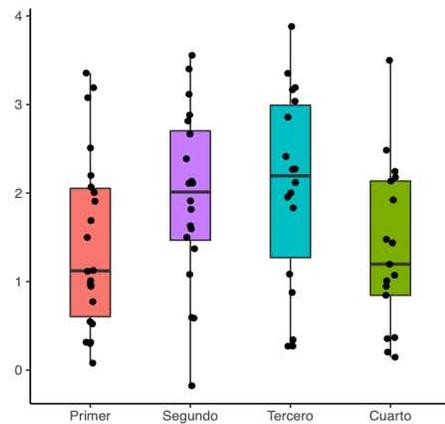
A nivel sociodemográfico análisis específicos por especie viral mostraron que Rotavirus A y VPH son más abundantes en niños donde el agua se almacena en recipientes al aire libre; asimismo, observamos más abundancia de VPH en niños que consumen agua de la llave (p ajustada > 0.05). La presencia en aguas contaminadas de virus que tienen vías de transmisión fecal-oral, como es el caso de rotavirus, ha sido descrita con anterioridad, aunque se ha reportado que la transmisión por aguas contaminadas no es tan común como el contagio persona a persona; por otro lado, se ha observado que en aguas estancadas el riesgo de contagio incrementa [93].

Si bien en algunos casos pudimos asociar la presencia de RVA con la vacunación (ver capítulo 6), no sabemos si todas las secuencias virales encontradas en la orofaringe tienen su origen en los virus vacunales. Por esta razón, no podemos descartar el agua contaminada como una posible fuente del virus, más aún cuando observamos la diferencia de abundancia de lecturas virales entre los niños que viven en casas donde se almacena el agua en recipientes de plástico afuera de la vivienda. Algo similar podría también suceder con VPH; un estudio de aguas residuales en Italia identificó una gran variedad de lecturas asociadas a VPH de mucosas y cutáneo en 81% de las muestras tomadas [94].

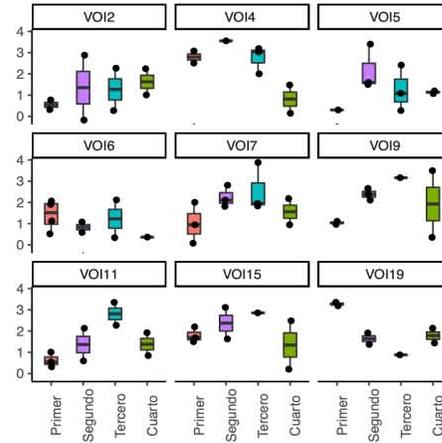
Figura 10. Gráficas de barras comparando la diversidad de los virus eucariontes que infectan humanos presentes en la orofaringe, analizadas por trimestre de edad con datos normalizados. En la barra superior se identifica a cada niño según el caso. VOI=viroma orofaríngeo infantil. (A) Promedio de abundancia de

lecturas asignadas a especies en todas las muestras de los niños, en escala logarítmica base 10; (B) Promedio de abundancia de lecturas asignadas a especies por niño, en escala logarítmica base 10; (C) Número de especies en todas las muestras de los niños, calculado con el índice de Chao; (D) Índice de Chao calculado por cada trimestre de cada niño; (E) Distribución de la abundancia entre especies en todas las muestras de los niños, calculada con el índice de Shannon; (F) Índice de Shannon calculado por trimestre para cada niño.

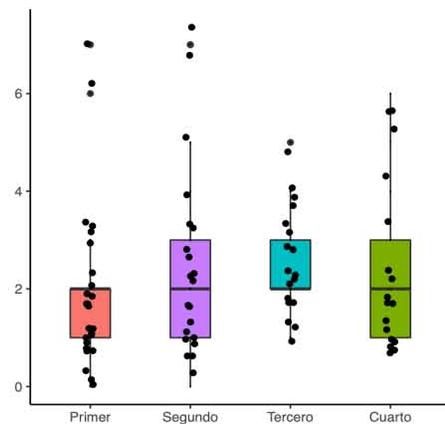
A Promedio de abundancia de lecturas



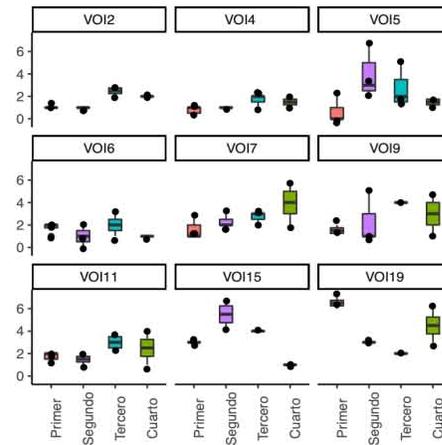
B Promedio de abundancia de lecturas



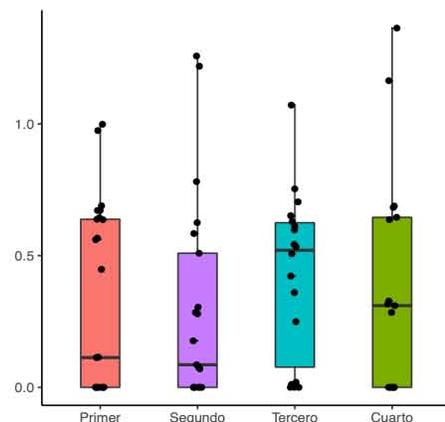
C Índice Chao



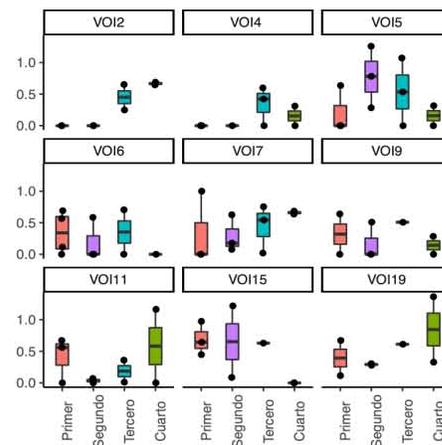
D Índice Chao



E Índice Shannon



F Índice Shannon



Al no haber observado tendencias o cambios específicos en la abundancia de lecturas asignadas a especie, riqueza y diversidad, consideramos que el viroma de cada individuo podría tener una composición y dinámica independiente. Esto se confirmó al realizar los análisis de composición viral de las muestras y observar que la variable que más explicaba la variación fue el viroma particular de cada niño (PERMANOVA $R^2= 21.7\%$, $p=0.001$). Otras variables que fueron estadísticamente significativas y cuya dispersión no representa un posible factor que afecte la interpretación estadística, (tales como forma de nacimiento, el consumo de leche de fórmula, forma de almacenar el agua y la presencia de roedores, entre otros) muestran influencia muy bajas sobre la composición de las muestras (R^2 sumado de todas las variables > 0.23 , es decir, 23%), mostrando que el viroma está moldeado por diferentes variables, donde lo que más peso tiene es la individualidad de los niños, es decir la mezcla peculiar de factores ambientales, genéticos y hasta sociales que son propios de cada niño.

Al graficar estas posibles diferencias, nos encontramos con que las gráficas de ordenamiento, esto es la representación gráfica y dimensional de las diferencias que existen entre cada una de las muestras, no apoyan algún tipo de agrupamiento. En el análisis de coordenadas principales (PCoA) no observamos que las muestras sean cercanas a los centroides, lo que significa que no se parecen entre ellas a pesar de venir del mismo individuo. Tampoco se observa una separación clara entre cada individuo en ninguno de los ejes (Fig. S1 a-c). Hasta cierto punto esto es esperado, pues son muestras que van evolucionando con el paso del tiempo y que permanecen independientes de las muestras de los otros individuos.

Al complementar con un análisis constreñido de coordenadas principales (db-RDA, Fig. S1d) se observa una separación un poco más clara, con los niños 2, 15 y 19 más alejados de los demás niños, y el infante 4 más cercano a estos 3 niños. Al tratar de encontrar posibles explicaciones a estas separaciones notamos que los infantes 2, 15 y 19 se distinguen por tener el mayor número de muestras positivas a RVA (más del 70%), y con mayor abundancia de secuencias para este virus, y que el infante 4 también tiene más muestras colectadas (50%) en comparación de los otros 5 infantes, sugiriendo que posiblemente sea la presencia de RVA la que marca más diferencias en la composición del viroma de los infantes.

Capítulo 2: El viroma de las muestras fecales

El sistema en el que más se ha estudiado el microbioma es el gastrointestinal, el cual ha sido caracterizado a través de muestras de materia fecal [12] enfocándose en gran parte en el bacteriota. Esto se ha dado principalmente por la relevancia clínica del sistema gastrointestinal, la facilidad para obtener las muestras y la abundancia de microorganismos que habitan en éste (es el sistema que más biomasa representa hasta ahora en el cuerpo humano), además de la facilidad económica y computacional de hacer estudios basados en la caracterización del gene ribosomal 16S [13].

Se ha sugerido que el intestino es colonizado posterior al nacimiento de una forma escalonada, dinámica y dependiente de su exposición, hasta que llega a tener una configuración similar a la de los adultos, es decir, estable [95]. Diversos estudios de bacterioma han señalado cómo el desarrollo de ésta en un recién nacido a término está influenciada por la leche materna, el tipo de nacimiento, localización geográfica y la presencia de hermanos o mascotas en casa. Por este motivo, ha sido considerada como un factor importantísimo para el desarrollo y crecimiento del lactante menor, el cual que tiene un papel crucial en la salud y enfermedad [25].

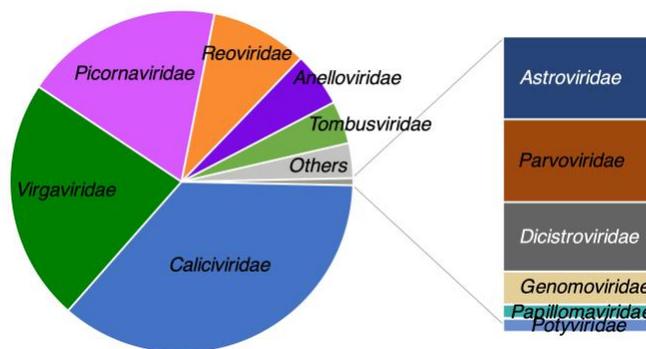
En comparación a los estudios del bacterioma, la información sobre los estudios del viroma, cómo evoluciona y se ve afectado por diversos factores es escasa, y muchos de los trabajos de viroma intestinal se han enfocado directamente a caracterizar los virus procariontes y sus interacciones con el bacterioma. En niños menores de un año existen pocos trabajos en los que se considera al viroma eucarionte [22,29,30,32–35,72,95] y pocos de ellos son estudios longitudinales que describen cómo cambia la composición del viroma a lo largo del tiempo [30–36,38,95]. Además, solo algunos han tenido cohortes exclusivamente de niños sanos/asintomáticos [29–31,35,37,38,73,95]. Aún menos de estos estudios se han llevado a cabo en niños que habitan en comunidades rurales [36–38,73].

Además, en muchos de los estudios del viroma en niños se ha descrito sobre todo la diversidad a nivel de familia, en los cuales la familia *Anelloviridae* ha sido la más abundante y relacionada con el desarrollo [22,29–31,36]; algunos estudios han señalado el aumento de abundancia de esta familia durante los 12 a 18 meses de edad [31,36], y otros los han relacionado con la forma de nacimiento, sugiriendo que los niños que nacen por parto tienen mayor abundancia de anellovirus [22]. Las familias *Adenoviridae*, *Astroviridae*, *Parvoviridae* y *Reoviridae* también han sido descritos en infantes sin presentar síntomas [95].

En las muestras de heces de nuestra cohorte encontramos 46 familias de virus eucariontes, aunque el 95% de abundancia de lecturas asociadas a estos virus se reparten entre 6 familias. De éstas, cuatro infectan a humanos: *Caliciviridae*, *Picornaviridae*, *Reoviridae* y *Anelloviridae*, mientras que los virus de las otras dos infectan plantas: *Virgaviridae* y *Tombusviridae* (Fig. 11). Los virus que se replican en humanos tuvieron la mayor proporción de lecturas virales comparados con los de plantas, con un 46% a nivel de familia (Tabla 5) y un 68% a nivel de especie.

Figura 11. Grafica de pie de las familias con mayor abundancia de lecturas en las muestras fecales (Izquierda), en gris claro con la etiqueta 'Others' están aquellas familias que tienen menos del 0.05% de

abundancia de lecturas. La gráfica de barras ilustra familias con un porcentaje menor a 0.1% de abundancia de lecturas (derecha).



Como se mencionó, las familias *Caliciviridae*, *Picornaviridae*, *Reoviridae* y *Anelloviridae* fueron las más abundantes en nuestra cohorte. A pesar de que existen estudios de seguimiento en niños menores de un año [30,31,33–36,38,95], la pregunta de investigación, y por lo tanto la metodología, varía en las diferentes publicaciones, lo que dificulta comparar entre trabajos. Aun así, como referencia, Mogotsi y cols. [38] reportaron a la familia *Picornaviridae* como la más prevalente, pero al analizar los virus presentes a nivel de especie se dieron cuenta se debía a la presencia de virus de polio vacunal. En nuestro caso, en seis muestras se identificaron secuencias de poliovirus. En otro estudio llevado a cabo en 8 infantes en Bélgica, los más abundantes fueron los virus de las familias *Picornaviridae*, *Reoviridae* y *Anelloviridae* [95]. Por otro lado, también se han encontrado en estudios transversales a virus de las familias *Picornaviridae*, *Caliciviridae* y *Circoviridae* como las más abundantes en niños sanos (con una edad promedio de 4.5 años) en comunidades amazónicas [37].

En nuestro trabajo se identificaron 46 especies con un promedio de 11 ± 5 especies, y 5,877 (121-5,288,674) lecturas por muestra. De estas 46 especies, 17 son especies de virus que se replican en humanos y 18 en plantas. De forma general, los virus que se replican en humanos y en plantas se observaron de forma frecuente, con excepción de 10 especies que se encontraron en menos del 10% de las muestras (Tabla 3). Por otro lado, 11 especies virales representaron el 93% de abundancia de lecturas virales: 8 son virus de humano y 3 de plantas, con una abundancia predominante de los virus que infectan humanos (Tabla 3).

Tabla 3. Porcentaje de frecuencia y abundancia de especies virales encontrados en las muestras fecales. Marcados en negrita se señalan las especies con más de 1% de abundancia. El porcentaje de frecuencia (% Frecuencia) se calculó considerando todas las muestras de la cohorte (n=97). El porcentaje de abundancia (% Abundancia) se calculó con el total de lecturas entre las muestras y el total de lecturas asignadas por especie. El porcentaje de abundancia por hospedero (% Abundancia por hospedero) se calculó considerando el total de lecturas de especies que replican en cada hospedero.

Hospedero	Especies virales	Familia	% Frecuencia	% Abundancia	% Abundancia por hospedero
Human	Torque teno virus	<i>Anelloviridae</i>	5.2	0.03	0.05
	Torque teno mini virus		49.5	3.73	5.47
	Torque teno midi virus		22.7	0.10	0.15
	Anellovirus sp		49.0	1.25	6.14
	Human papillomavirus	<i>Papillomaviridae</i>	17.5	0.03	0.04
	Human bocavirus	<i>Parvoviridae</i>	14.4	0.12	0.18
	Rotavirus A	<i>Reoviridae</i>	67.0	9.12	13.38
	Norwalk virus	<i>Caliciviridae</i>	58.8	32.27	47.36
	Sapporo virus		28.9	3.83	5.61
	Mamastrovirus 1	<i>Astroviridae</i>	32.0	0.18	0.27
	Enterovirus A	<i>Picomaviridae</i>	20.6	2.00	2.94
	Enterovirus B		39.2	3.99	5.85
	Enterovirus C		9.3	0.26	0.37
	Rhinovirus A		14.4	0.25	0.36
	Rhinovirus B		5.2	<0.01	0.01
	Rhinovirus C		5.2	<0.01	0.01
	Parechovirus A		58.8	12.22	17.94
Plants and/or Fungi	Gemycirculavirus sp.	<i>Genomoviridae</i>	40.2	0.05	0.18
	Genomoviridae sp.		37.1	0.02	0.08
	Zea mays chrysovirus 1	<i>Chrysoviridae</i>	5.2	<0.01	0.04
	Maize-associated totivirus 3	<i>Totiviridae</i>	4.1	<0.01	0.04
	Tomato necrotic spot virus	<i>Bromoviridae</i>	4.1	<0.01	0.04
	Bell pepper mottle virus	<i>Virgaviridae</i>	12.4	<0.01	0.04
	Paprika mild mottle virus		14.4	0.02	0.08
	Pepper mild mottle virus		86.6	10.39	38.37
	Rattail cactus necrosis-associated virus		25.8	0.55	2.02
	Tobacco mild green mosaic virus		38.1	0.28	1.03
	Tomato mosaic virus		18.6	<0.01	0.04
	Tropical soda apple mosaic virus		97.9	11.52	42.54
	Opuntia virus 2		27.8	0.05	0.18
	Tymovirales sp.	<i>Tymovirales</i>	9.3	<0.01	0.04
	Melon necrotic spot virus	<i>Tombusviridae</i>	34.0	4.03	14.86
	Melon severe mosaic tospovirus	<i>Tospoviridae</i>	11.3	0.03	0.11
	Solemoviridae sp.	<i>Solemoviridae</i>	2.1	<0.01	0.04
Potyvirus sp	<i>Potyviridae</i>	8.2	0.02	0.06	
Insectos	Blattodean pefuambidensovirus 1	<i>Parvoviridae</i>	15	0.012	7.138
	Bemisia-associated dicistrovirus 1	<i>Dicistroviridae</i>	16	0.154	92.862
	Adeno-associated virus	<i>Parvoviridae</i>	36	0.035	0.778
	Circoviridae sp.	<i>Circoviridae</i>	7	0.007	0.162
	Picomavirales sp.	<i>Orthomavirae</i>	15	0.036	0.801
No clasificados	unclassified Densovirinae	<i>Parvoviridae</i>	9	0.007	10.075
	unclassified insect virus	<i>Virus no clasificados</i>	9	0.007	9.059
	Riboviria sp.		4	0.024	32.760
	Hubei sp		15	0.029	39.878
	unclassified animal virus		6	0.006	8.228

1) Los virus de la familia *Caliciviridae* fueron los que se encontraron de manera más frecuente y con mayor abundancia, en particular los norovirus y sapovirus, de los cuales hablaremos más adelante, en una sección independiente (capítulo 3).

2) Dentro de la familia *Picornaviridae* se detectaron 4 especies con alta abundancia: Parechovirus A, Enterovirus, Rotavirus A y Torque teno mini virus. De interés, un patrón de presencia de familias similar al nuestro fue reportado en un estudio de seguimiento por dos años en una cohorte de 19 infantes [32]. En nuestro caso:

- a. Parechovirus A (PeVA) tuvo una abundancia del 12.2% y también fue bastante frecuente, estando presente en el 59% de las muestras totales (Tabla 3), y en más del 75% de las muestras de los niños 5, 6, 7 y 15 (Fig. 13). Al realizar un análisis filogenético observamos que los virus fueron genéticamente diversos tanto en muestras de un mismo niño, como de diferentes niños. Solo dos especies similares fueron detectadas, infectando a dos niños en un mismo mes (5 y 6) (Fig. 12).

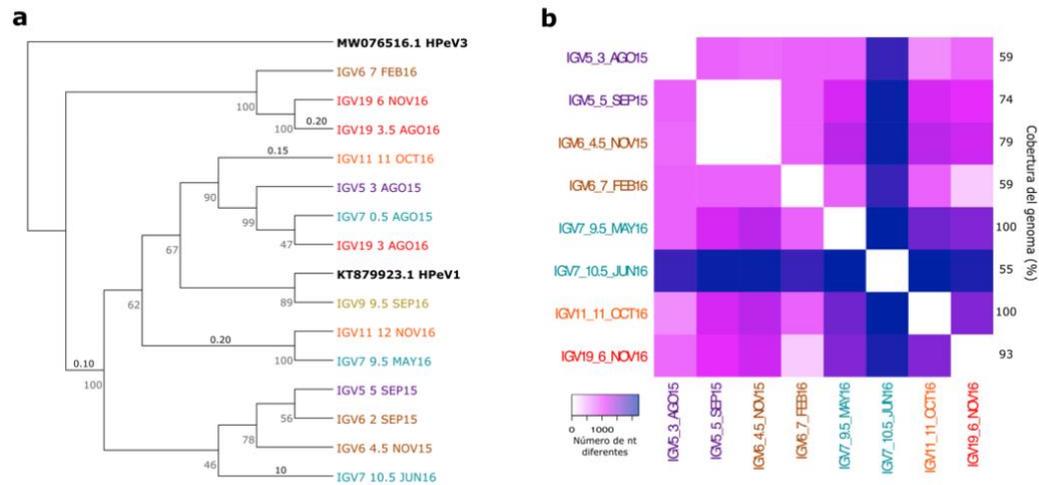
A pesar de que los parechovirus humanos han sido asociados a enfermedades gastrointestinales o respiratorias, la presencia en tan alto número de muestras de heces pudiera ser reflejo en algunos niños de una infección asintomática con excreción prolongada, lo cual se ha reportado previamente en infantes menores de 1 año. Esta excreción se ha reportado con una duración promedio de 51 con un máximo de 93 días [30].

Respecto al 12% de abundancia total encontrada en nuestra cohorte (Tabla 3), en un estudio longitudinal en 8 infantes con 21 muestras tomadas durante el primer año de vida encontraron un porcentaje similar al nuestro, aunque la identificación taxonómica en ese estudio se hizo a nivel de género (11.6%) y reportó una menor frecuencia (16.8%) [95] en comparación a nuestro estudio (59%). Diversos estudios longitudinales en niños sanos menores de un año han reportado la presencia de parechovirus humanos, aunque en su mayoría a nivel de género [30–33,37,73,96,97]. Por otro lado, los parechovirus han sido asociados a la presencia de mascotas en las casas [33], aunque nosotros no pudimos explorar esta variable, ya que 8 de los 9 infantes tienen mascotas. Respecto a la prevalencia de PeVA, se ha descrito entre el 23 y 57% en niños en Malawi. Esta es una prevalencia similar a la nuestra, y en ambos casos mayor a la reportada mundialmente (2%) [98], lo cual sugiere que la prevalencia podría verse influenciada por la región o estatus socioeconómico dónde se encuentran los niños.

Figura 12. Diversidad genética de los parechovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM2+F+R4. Los genomas de referencia y

raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.

Parechovirus



- b. Enterovirus; Encontramos 3 especies diferentes de enterovirus. Dos de ellas tuvieron una abundancia superior al 1%, es decir, fueron de las especies más abundantes; enterovirus A con un 2% de abundancia de lecturas y enterovirus B con 3.9%. Enterovirus B se observó en todos los infantes, con abundancias bastante variables, sin embargo, en 2 infantes (6 y 9) estuvo presente durante casi todo el año de vida. Para enterovirus A, solo 6 de los 9 infantes tuvieron muestras positivas a esta especie, estando presente en 50% o más de las muestras de los infantes 2 y 15 (Fig. 13). Se identificaron secuencias como poliovirus en 5 infantes a diferentes edades (niños 2, 4, 5, 9 y 19, y se hablará de estos hallazgos más adelante). Estos virus presumiblemente son vacunales, derivados de la vacuna, u otros enterovirus "similares a poliovirus".

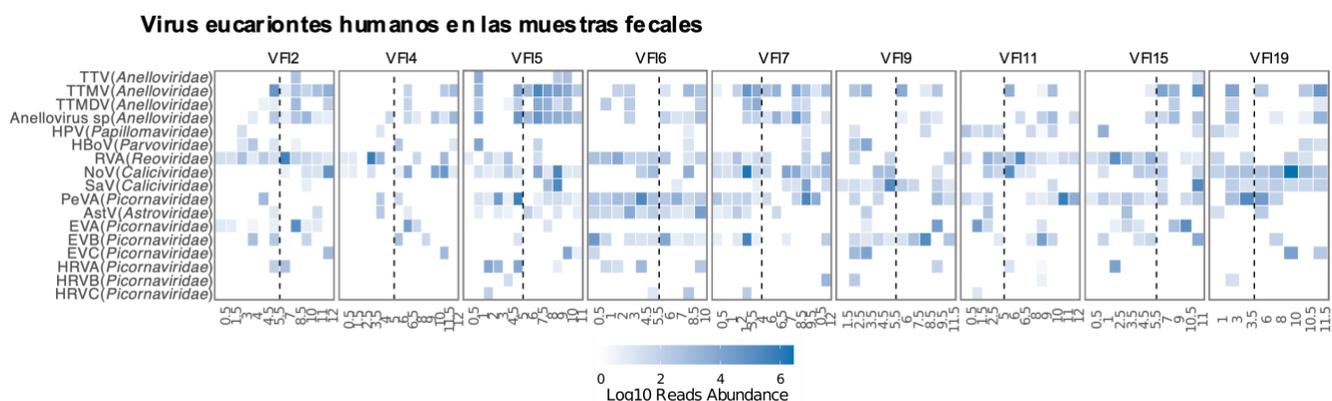
Las infecciones por enterovirus han sido asociadas a riesgo de padecer autoinmunidad de islotes pancreáticos y también con enfermedad celíaca [32,34], lo cual resulta interesante ya que ambas enfermedades son autoinmunes y estas observaciones se hicieron en niños menores de 2 años, una etapa donde el moldeado del sistema inmune a través de la exposición es crucial [99]. Por otro lado, se ha reportado la presencia de estas especies en niños asintomáticos menores de un año [30,31,35,100], y en algunos casos inclusive se ha sugerido que esta especie aumenta en frecuencia y abundancia con la edad [33]; sin embargo, nosotros no observamos ninguna tendencia o estadística que sugiriera un aumento con la edad. Dentro de los factores que podrían explicar la presencia

de los enterovirus en nuestro estudio se encuentra el agua contaminada (se observó una mayor abundancia de enterovirus A asociado a agua de contenedor (ANOVA con permutaciones para muestras repetidas < 0.05), y posiblemente con un bajo estatus socioeconómico (enterovirus B más abundante en niños que viven en hacinamiento, ANOVA con permutaciones para muestras repetidas < 0.05). Estos resultados concuerdan con la disminución de enterovirus en muestras fecales de niños menores de 5 años después de una intervención de filtrado de agua en comunidades en Etiopía [97].

- c. Finalmente, también se encontraron las especies rinovirus A, B y C, aunque con una abundancia sumada de 0.26% y una frecuencia sumada del 25%. Rinovirus A fue la especie más frecuente (14%) y abundante (0.24%), en los infantes 5 y 6 (Fig. 13).
- 3) La especie rotavirus A tuvo una abundancia de 9.11% y una frecuencia del 67% en las muestras. Los rotavirus se encontraron de manera compartida con la orofaringe, por lo que se hablará de estos hallazgos más adelante, y con mayor detalle (capítulo 6).
 - 4) En la familia *Anelloviridae*, dos especies tuvieron una alta frecuencia (49% y 50%) con una baja abundancia (3.7% y 1.2%):
 - a. La especie torque teno mini virus (TTMV), perteneciente al género *Betatorquevirus*, tuvo un porcentaje de abundancia de 3.73%, con una frecuencia del 49% en las muestras. Esta especie se encontró en todos los niños, con una frecuencia mayor al 50% de las muestras de los infantes 2, 5, 6, 7 y 9. Además, un incremento de frecuencia de TTMV de forma trimestral fue observado con 26%, 44%, 61% y 72% de muestras por trimestre, respectivamente; esta tendencia se confirmó estadísticamente, con mayores lecturas en el segundo semestre de vida en comparación con el primero (ANOVA con permutaciones para medidas repetidas < 0.05). Respecto a su frecuencia, un estudio reciente en el que buscaron betatorquevirus por qPCR en muestras de sangre de infantes de 0 a 12 meses, encontraron 50% de positividad [101]. En conjunto, estos datos sugieren que este género es muy común durante el primer año de vida.
 - b. La especie *anelloviridae* sp también estuvo frecuentemente encontrada en todos los niños y con un patrón muy similar a TTMV (Fig. 13) y también mostrando un aumento gradual de frecuencia por trimestre con 11%, 40%, 85% y 77%.

La alta abundancia de TTMV, de la familia *Anelloviridae*, se refleja en el hecho de que esta familia de virus fue la cuarta más abundante en las muestras fecales de los infantes. En contraste con nuestros hallazgos, algunos estudios [30,31,36] han descrito que los virus de esta familia son los más frecuentes y abundantes en niños sanos menores de 1 año. Es posible que estas diferencias se deban a un sesgo metodológico durante el procesamiento de las muestras, en el paso de enriquecimiento viral, pues algunos de estos estudios [31,36] han utilizado la tecnología de amplificación por desplazamiento múltiple, la cual se ha reportado que tiene un sesgo de amplificación para virus de ADN de cadena sencilla [102,103]. Otro factor que se ha descrito a nivel de familia, es una mayor presencia de anellovirus en niños nacidos por vía vaginal [22], sin embargo, nosotros no encontramos diferencias, ni a nivel de familia, ni de especie, lo cual puede deberse al bajo número de niños reclutados en nuestro estudio.

Figura 13. Mapa de calor representando las abundancias de especies virales humanas que se replican en la materia fecal. Las abundancias están normalizadas y expresadas en escala de logaritmo 10. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Cada panel representa a un niño con el identificador en la parte superior de cada panel y la edad en meses en la parte inferior (eje x). VFI = viroma fecal infantil. Las líneas punteadas dividen los meses entre el primer y segundo semestre de edad. Las especies que fueron taxonómicamente asignadas como 'sp' se omitieron de la figura. Dentro del paréntesis está escrita la familia a la que pertenece cada especie. Especies que se replican en humanos: TTV, torque teno virus; TTMV, torque teno mini virus; TTMDV, torque teno midi virus; HPV, human papillomavirus; HBoV, human bocavirus; RVA, rotavirus A; NoV, norovirus; SaV, sapovirus; PeVA, parechovirus A; AstV, mamastrovirus; EVA, enterovirus A; EVB, enterovirus B; EVC, enterovirus C; HRVA, rinovirus A; HRVB, rinovirus B; HRVC, rinovirus C.



Respecto a las especies encontradas durante el primer mes de vida, resultó interesante observar que todos los niños con muestra tomada antes o en los primeros 16 días de vida (infantes 2, 4, 5, 6, 7, y 15) estaban presentes de 1 a 5 especies virales: rotavirus, norovirus, enterovirus, PMMoV and TSAMV. Además, en el infante 15 encontramos 2 especies (rotavirus A y parechovirus A) en la primer muestra, la cual fue tomada a los 8 días de nacimiento. En un

estudio donde la búsqueda de virus gastrointestinales se llevó a cabo utilizando RT-qPCR durante el primer mes de vida (tomas de meconio, y posteriormente a los 2, 7 y 28 días de nacidos) de 15 niños en Gabón y Ghana, no encontraron virus en el meconio, pero uno de los niños fue positivo a rotavirus A en el segundo día de vida, y cuatro niños en las muestras de 7 y 28 días. En este estudio enterovirus también estuvo presente en 6 niños en las muestras a los 7 y 28 días [104].

Por otro lado, algunos análisis metagenómicos de virus de ADN en niños con 1 semana de nacidos mostraron la presencia de bacteriófagos, pero no de virus eucariontes [29]. De acuerdo a nuestros hallazgos y al trabajo de Brazier y cols. [104], podríamos sugerir que los virus eucariontes de RNA son los primeros en encontrarse en el intestino de los infantes.

Análisis a nivel ecológico

Similar a los análisis de la orofaringe, se hicieron análisis a nivel ecológico considerando todas las variables obtenidas a partir de los registros de historia clínica y seguimiento de los niños (Tabla 1, tabla S3). De forma general, observamos que el promedio de abundancia de lecturas asignadas a especies y el número esperado de especies por muestra, es decir, el índice de Chao1, aumenta después de la ingesta de frutas, verduras y otras bebidas diferentes a la leche materna o de fórmula, lo cual es hasta cierto punto, esperado, pues como se ha descrito antes, la alimentación influye en el bacterioma y algo similar podría esperarse del viroma. Además, la ingesta de genes virales a través de la alimentación se ha reportado previamente en diversos virus de plantas [105]. Que exista un aumento de especies presentes, pero no de su abundancia nos sugiere que podría existir un mecanismo de control ecológico, posiblemente los anticuerpos que se encuentran en la leche materna [106,107].

Los análisis de diversidad alfa con la media de abundancia de lecturas asignadas a especies e índice de Shannon, entre y dentro de los trimestres de los infantes no fue estadísticamente significativo (Fig. 14 a y e). Una visualización por infante sugiere que en cinco infantes (2, 4, 5, 7, 15 y 19) había un aumento de abundancia de lecturas asignadas a especies hasta el tercer trimestre de vida (Fig. 14b). Sin embargo, un análisis más fino demostró que solo hay un aumento de abundancia entre y dentro del primer y el cuarto trimestre de vida (ANOVA con permutaciones para medidas repetidas $p=0.03$ y $p=0.04$, respectivamente). En comparación, no se observaron alzas de lecturas por trimestre en las muestras orofaríngeas, lo que pudiera deberse a que el intestino, al estar menos expuesto al medio ambiente, tiene un tránsito del bolo alimenticio más lento y acumular el bolo de la deglución, tiene, posiblemente, dinámicas ecológicas más estables en comparación a otros nichos.

De manera similar, el número de especies, es decir la riqueza, fue diferente entre y dentro de cada trimestre de cada niño (ANOVA con permutaciones para medidas repetidas $p=0.01$ y

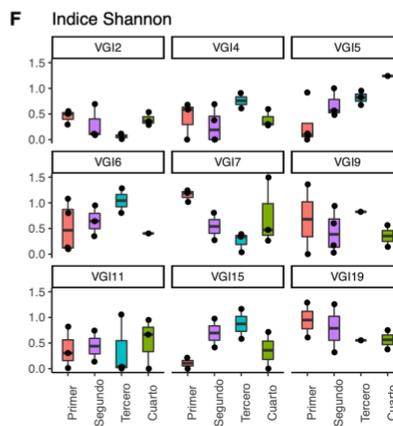
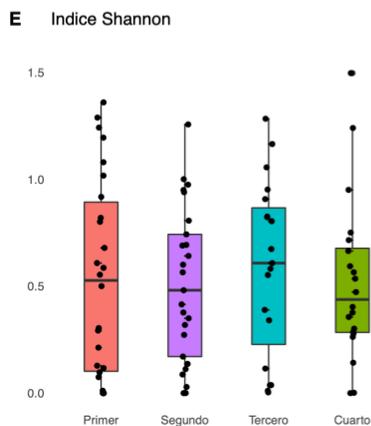
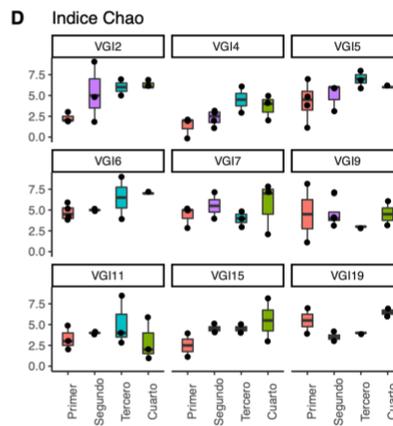
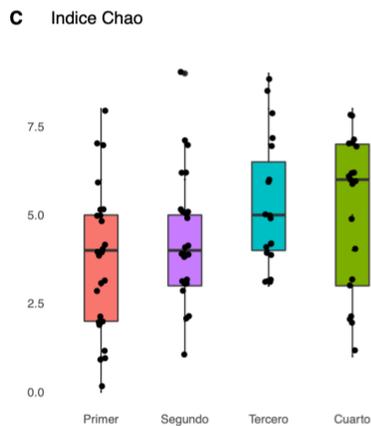
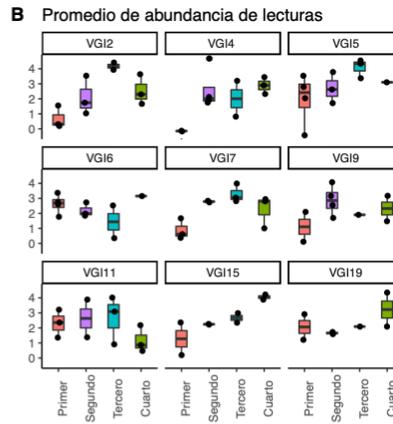
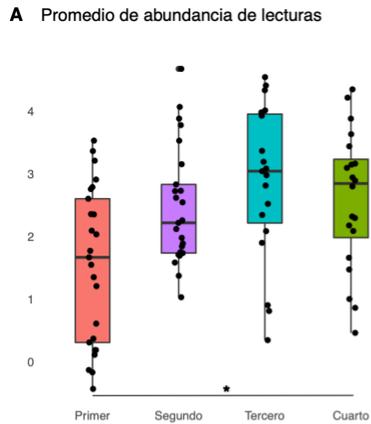
$p=0.001$, respectivamente). La visualización global sugiere una tendencia al incremento de riqueza a través del tiempo, sin embargo, al verlos de forma individual se observa que los cambios en cada infante son independientes de la edad (Fig. 14 c y d).

Mientras no encontramos tendencias específicas entre la información general y sociodemográfica de los niños y la riqueza y diversidad de virus encontrada, a excepción del aumento de especies en muestras posterior a la alimentación, los análisis de composición del viroma, es decir, las diferencias que existen entre las muestras, resultaron ser bastante más ilustrativos, mostrando que la variación de las muestras puede verse modificada por factores tales como la edad y alimentación (Tabla S3).

Al intentar identificar el porcentaje de variación de cada factor, esto es, el peso que tiene cada una de las variables en la heterogeneidad de los datos obtenidos de los niños, vimos que cada factor tiene muy poca influencia, con un total sumado de 49%. Las diferencias entre los trimestres son las que más explican estas variaciones ($R^2= 10.6\%$, $p=0.001$) y la diferencia de composición del primer trimestre el que más peso tiene ($R^2= 6.8\%$, $p=0.001$). Sin embargo, la variación más grande en la composición de las comunidades virales en las muestras fue explicado, como en la orofaringe, por la individualidad de las muestras de cada uno de los niños ($R^2=25.2\%$, $p=0.001$). Esta individualidad se ha observado también en distintas partes del cuerpo de adultos sanos, describiendo que al analizar virus eucariontes de ADN, cada persona tenía “huellas virales únicas” [28].

Estos resultados nos sugieren que la composición y dinámica del viroma de la materia fecal también se comporta de forma única para sus hospederos y que está altamente influenciada por el medio ambiente dónde habitan. Al respecto, se ha descrito que el viroma es más similar entre infantes que cohabitan en la misma casa, que entre infantes de la misma edad [36,72].

Figura 14. Gráficas de barras comparando la diversidad de los virus eucariontes que infectan humanos presentes en las muestras fecales, analizadas por trimestre de edad. En la barra superior se identifica a cada niño según el caso. VFI= viroma fecal infantil. (A) Promedio de abundancia de lecturas en todas las muestras de los niños, en escala logarítmica 10; (B) Promedio de abundancia de lecturas por niño, en escala logarítmica 10; (C) Número de especies en todas las muestras de los niños, calculado con el índice de Chao; (D) Índice de Chao calculado por cada trimestre de cada niño; (E) Distribución de la abundancia entre especies en todas las muestras de los niños, calculada con el índice de Shannon; (F) Índice de Shannon calculado por trimestre para cada niño.



Considerando que diversos factores afectan como se establece el viroma fecal, la observación de tendencias que sugieren una población viral más rica después del primer trimestre de vida de los infantes es congruente. Durante los primeros 3 meses de vida los niños tienen contacto principalmente con la madre o cuidadora. Después comienzan a explorar el ambiente que habitan utilizando su cuerpo y la ingesta de otros tipos de comida comienza. El aumento de especies virales con la edad ya ha sido descrito previamente, aunque fue enfocado a cambios por semestre y principalmente describiendo a la familia de anellovirus [31,34,36].

Similar a la orofaringe, encontramos que al hacer un análisis principal de coordenadas (PCoA), no observamos ningún tipo de agrupamiento, con muestras lejanas a los centroides y sin separación clara entre los infantes en ninguno de los ejes (Fig. S2A-C). Al complementar con un análisis constreñido de coordenadas principales (db-RDA, Fig. S2D), seguimos sin poder observar separaciones drásticas entre los infantes, sin embargo los infantes 5, 6, 9 y 19 tienen sus centroides más alejados de los demás. Analizando porque podríamos tener este comportamiento, observamos que el infante 5 tiene la mayor cantidad de muestras y abundancia de TTMV, el infante 6 de astrovirus y los infantes 9 y 19 de sapovirus (Fig. 13). Al hacer análisis de forma individual por especie, no se encontraron factores que pudieran explicar por qué estos niños tenían mayor abundancia de alguna de las especies mencionadas.

Capítulo 3: Diversidad y abundancia de calicivirus en las muestras fecales

Los calicivirus humanos son considerados la principal causa de gastroenteritis no bacteriana en países donde la vacunación para rotavirus ha sido efectivamente implementada [108,109]. Estos virus tienen una cadena sencilla positiva de ARN [ssARN(+)] y se clasifican en dos géneros: *Norovirus* y *Sapovirus* [110], teniendo como especies prototipo el virus de Norwalk y el virus Sapporo, respectivamente. Actualmente los norovirus se clasifican en 10 genogrupos (GI al GX), cinco de los cuales se han identificado en humanos; GI, GII, GIV, GVIII and GIX [111]. A nivel mundial, los dos genogrupos más frecuentemente detectados son GI (con 9 genotipos) y GII (22 genotipos), siendo GII.4 el genotipo más predominante [112].

Por otro lado, los sapovirus están relacionados con gastroenteritis esporádicas y rara vez relacionados con brotes de diarrea. Asimismo, han sido comúnmente reportados como una causa de infección sintomática y asintomática en diferentes comunidades infantiles [113,114]. Se clasifican en al menos 15 genogrupos (GI al GXV), de los cuales GI (7 genotipos), GII (8 genotipos), GIV (1 genotipo) y GV (2 genotipos) infectan a humanos [115].

Los virus de la familia *Caliciviridae* fueron los más abundantes (32.7% norovirus y 3.8% sapovirus) y frecuentes en las muestras fecales de nuestra cohorte, con 58 (60%) de las 97 muestras positivas a virus de esta familia; en tres infantes (5, 7 y 19) más del 80% de las muestras tuvieron lecturas asignadas a esta familia. De forma interesante, el infante 7 tuvo una muestra positiva a calicivirus a los 16 días de haber nacido, y los niños 5, 7, 15 y 19 al primer mes. Sin embargo, de forma general la frecuencia de lecturas específicas para esta familia fue baja durante el primer trimestre de vida (45% para norovirus y 11% en sapovirus), dificultado el poder discernir si estos casos representan una infección o una ingestión pasiva del virus. La frecuencia de lecturas asociadas a esta familia incrementó después del cuarto mes de vida (ANOVA con permutaciones para medidas repetidas $p = 0.016$) y se mantuvo hasta los 12 meses de vida (72% y 45%, respectivamente), sugiriendo una interacción mayor de los infantes con fómites y con el medio ambiente, aunque también podría relacionarse con infecciones

abortadas o restringidas por anticuerpos de calicivirus presentes en la leche materna [116], o como consecuencia de una infección transitoria que no pudo colonizar un epitelio intestinal inmaduro.

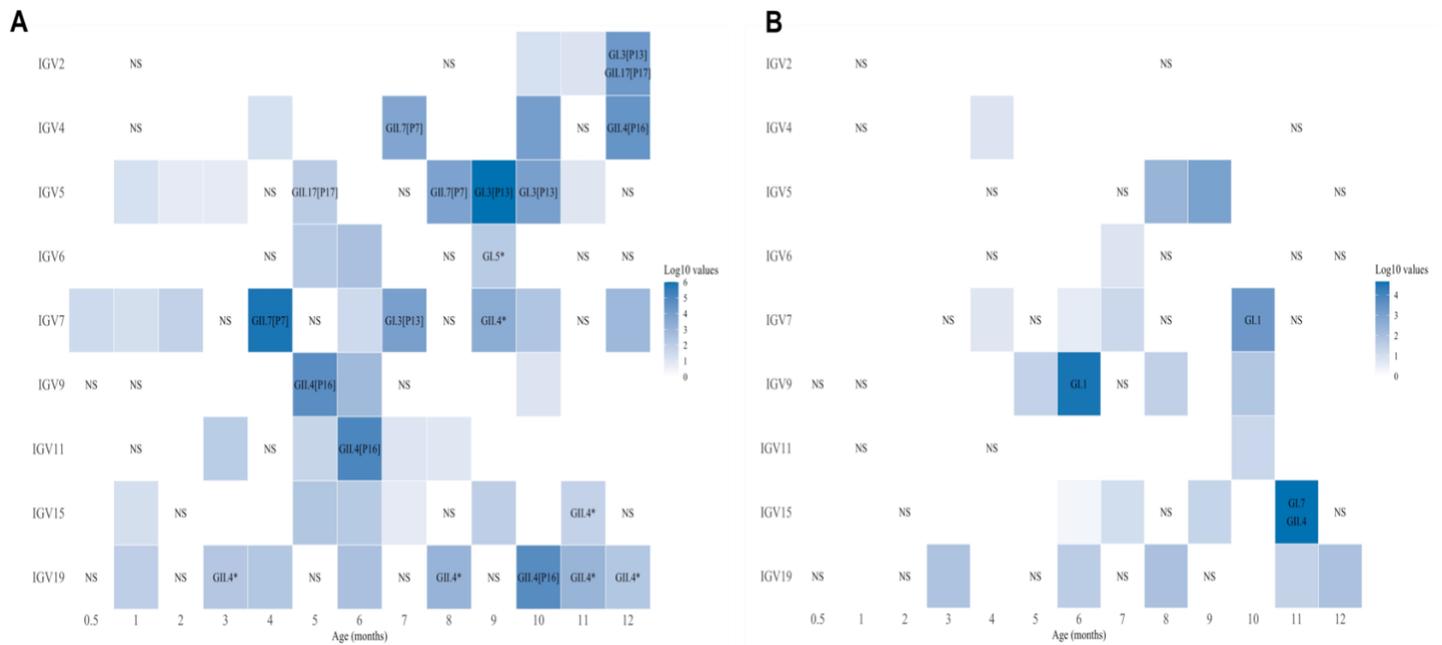
Para poder identificar posibles infecciones, seleccionamos de forma arbitraria como punto de corte haber obtenido cuando menos una cobertura del genoma de más de 20%. Considerando que el genoma de estos virus es de aproximadamente 7500 pares de bases para norovirus y 7400 pares de bases para sapovirus, cubrir amplios segmentos del genoma a través de secuenciación aleatoria sugiere que están presentes de manera abundante. Con este punto de corte consideramos como posibles infecciones 26 muestras para norovirus (Tabla S4) y 5 para sapovirus (Tabla S5), por lo que cada niño tuvo al menos una y hasta seis infecciones para norovirus.

En general, lecturas de norovirus se encontraron en 54 (56%) de las muestras (Fig. 15A), mientras que lecturas de los sapovirus se encontraron en 24 (25%) (Fig. 15B). Diversos estudios han caracterizado la presencia de calicivirus en niños hospitalizados por gastroenteritis aguda, pero muy pocos se han llevado a cabo en esquemas de seguimiento en comunidades rurales. En un meta-estudio de 8 países con ingreso económico bajo o medio, se colectaron muestras fecales de forma mensual durante el primer año de vida, y después a los 15, 18, 21 y 24 meses [117], describiendo una prevalencia del 19% en infección asintomática de norovirus en un grupo de 100 niños [118]. De forma similar, un meta-análisis en Latinoamérica [109] y un estudio mexicano [110] describieron una prevalencia de 8 y 7%, respectivamente, de infecciones asintomáticas de norovirus. Igualmente, un meta-análisis que incluyó a 81 países describió una prevalencia global de 7% en comunidades rurales [119].

Respecto a los sapovirus, se describió una prevalencia de 9.7% en muestras asintomáticas [120] en niños vigilados hasta los 2 años de vida. Otros estudios han reportado que la incidencia de sapovirus es mayor durante el segundo año de vida [121], con 1.33 casos por niño/año, comparado con el primer año de vida (0.51 casos por niño/año). Esto podría explicar la baja proporción y abundancia de lecturas en muestras asociadas a sapovirus cuando la comparamos con la frecuencia y abundancia de norovirus en nuestro estudio. Por otro lado, si consideramos el criterio arbitrario de infección, tenemos que la prevalencia es del 29% para norovirus y 5% para sapovirus, lo cual se acerca a lo reportado en algunos estudios, con valores menores a los nuestros [118]. Con este criterio, todos los niños hubieran tenido al menos una infección asintomática de norovirus (3 casos por niño/año), y 3 niños hubieran tenido al menos una infección de sapovirus (0.33 casos por niño/año).

Figura 15. Frecuencia y abundancia de norovirus y sapovirus en los infantes. En el panel A se observan las muestras positivas a norovirus, en el panel B a sapovirus. La abundancia se encuentra normalizada y escalada en logaritmo base 10. Para simplificar la figura, se sumó la abundancia de las muestras tomadas

a mitad del mes, con el mes posterior. Los cuadros blancos representan muestras negativas, los cuadros con 'NS' indican la ausencia de muestra en esa fecha. Se marcó con un (*) las muestras que no pudieron ser doblemente genotipificadas debido a una falta de cobertura de la región RdRp.



Nuestros hallazgos de una prevalencia más alta que la descrita previamente para ambas virus puede deberse entre otras razones, a la sensibilidad de la secuenciación de nueva generación empleada en nuestro trabajo, ya que varios otros estudios que han reportado una prevalencia más baja de calicivirus han utilizado RT-qPCR para su detección [109,110,117–122]. A pesar de ser considerada como el estándar de oro en la detección de norovirus, se ha planteado la dificultad para poder interpretar resultados positivos en sujetos asintomáticos utilizando RT-qPCR [123,124]. Por otro lado, los métodos de secuenciación de nueva generación son considerados como el estado del arte para la detección de virus, con la desventaja de su alto costo, lo cual explica porque no se utilizan de forma rutinaria.

Para hacer un análisis más completo, se realizaron ensamblajes *de novo*, y las secuencias consenso se utilizaron para realizar análisis filogenéticos y para genotipificar los virus utilizando la herramienta en línea de genotipificación de calicivirus humanos [63] (Tabla 4). De manera interesante, los resultados filogenéticos son bastante similares a la tipificación con la herramienta, a pesar de que ésta se basa en dos regiones del genoma de norovirus para su tipificación: la que codifica por la polimerasa (RdRp) y la cápside (VP1); en el caso de los sapovirus la tipificación se basa sólo en la cápside (VP1). Por otro lado, los árboles se generaron con alineamientos de genoma completo y después se agregaron los genomas parciales (contigs de al menos 1500 pares de bases) (Fig. S3a y 4a). También se hicieron análisis de diferencias de nucleótidos, donde tanto para norovirus, como para sapovirus, no

observamos diferencias marcadas entre mismos genotipos o genogrupos (Fig. S3b y S4b). En ambos casos observamos que para norovirus, el genotipo más común fue GII, siendo identificado en 15 muestras, seguido por GI en 6 muestras. De interés, encontramos al menos 5 genotipos diferentes circulando en la comunidad, correspondientes a GI.3[P13] (4 muestras), GI.5 (1 muestra), GII.4[P16] (10 muestras), GII.7[P7] (3 muestras) y GII.17[P17] (2 muestras). La alta prevalencia de GII.4 ya ha sido reportada previamente [116,118,125]. El genotipo GI.3 ha sido asociado a brotes relacionados con agua contaminada [126,127], lo cual nos sugiere que el agua contaminada podría ser una fuente de esta especie en nuestra comunidad, sobre todo cuando el agua de la comunidad suele ser tomada de un pozo central en envases, o bien, comprada en pipas.

Tabla 4. Genotipificación de genomas completos y parciales utilizando la herramienta “Calicivirus genotyping tool”. Los genotipos que tienen un (*) no pudieron ser tipificados. Durante el ensamblaje se utilizó una cobertura de 3x y para verificar posibles variantes, una cobertura de 10x. Si la posición de nucleótidos no tenía esa cobertura, se colocaba una N. Se colorearon los genotipos para mejor identificación y agrupamiento. Region B: RdRp. Región C: VP1.

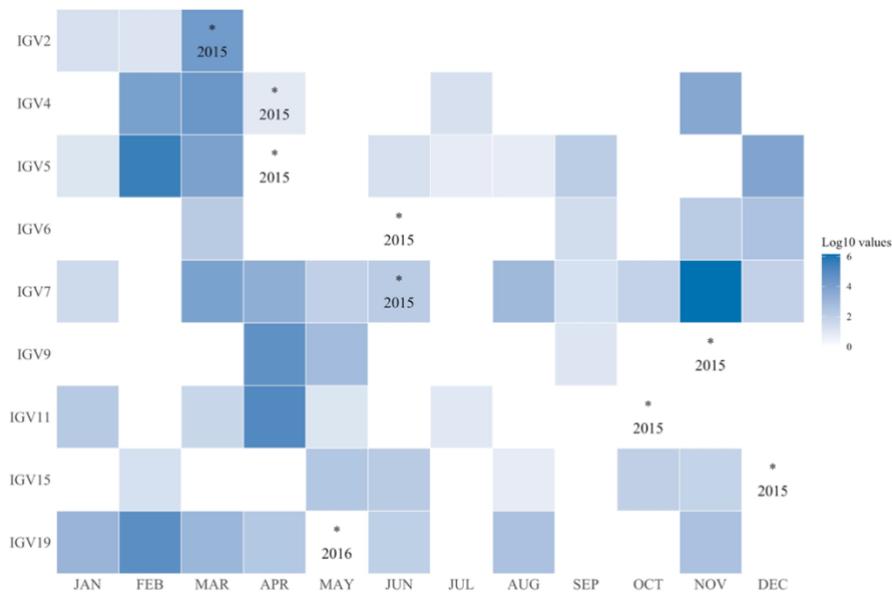
Genero	Infante	Edad en meses	Mes de la muestra	Año de la muestra	Genotipo	Region-B	Region-C	Longitud del genoma de referencia	# de N	Pares de bases en el consenso
NoV	IGV2	12	Mar	2016	GI.3[P13]	93%	96%	7,556	135	7,421
		12	Mar	2016	GII.17[P17]	97%	98%	7,470	102	7,368
		13	Apr	2016	GII.17[P17]	99%	99%	7,470	86	7,384
	IGV4	7	Nov	2015	GII.7[P7]	97%	97%	7,374	0	7,374
		11.5	Mar	2016	GII.4[P16]	98%	99%	7,500	1	7,499
	IGV5	4.5	Sep	2015	GII.17[P17]	-	71%	7,559	1,638	5,921
		7.5	Dec	2015	GII.7[P7]	97%	97%	7,374	0	7,374
		9	Feb	2016	GI.3[P13]	93%	96%	7,691	0	7,691
		10	Mar	2016	GI.3[P13]	93%	96%	7,572	119	7,453
	IGV6	8.5	Mar	2016	GI.5*	-	85%	7,701	6,368	1,333
	IGV7	3.5	Nov	2015	GII.7[P7]	97%	98%	7,374	0	7,374
		7	Mar	2016	GI.3[P13]	93%	96%	7,691	0	7,691
		8.5	Apr	2016	GII.4*	-	94%	7,501	188	7,313
	IGV9	4.5	Apr	2016	GII.4[P16]	98%	99%	7,456	45	7,411
	IGV11	5.5	Apr	2016	GII.4[P16]	98%	99%	7,467	34	7,433
	IGV15	11	Nov	2017	GII.4*	-	70%	7,501	3,608	3,893
	IGV19	3	Aug	2016	GII.4*	-	74%	7,502	3,420	4,082
		8	Jan	2017	GII.4*	-	70%	7,502	2,497	5,005
		10	Feb	2017	GII.4[P16]	98%	99%	7,502	23	7,479
10.5		Mar	2017	GII.4*	-	86%	7,502	1,109	6,393	
11.5		Abr	2017	GII.4*	-	81%	7,502	3,543	3,959	
SaV	IGV7	9.5	May	2016	GI.1	-	99%	7,375	13	7,362
	IGV9	5.5	May	2016	GI.1	-	99%	7,386	2	7,384
	IGV15	10.5	Nov	2016	GI.7	-	96%	7,131	321	6,810
		11	Dec	2017	GII.4	-	97%	7,446	14	7,432

Los de los infantes (2 y 4) fueron infectados con diferentes genotipos de norovirus durante su primer año de vida, y dos infantes (5 y 7) con tres (Tabla 4, Fig. 15A). Esto nos sugiere que es posible tener infecciones múltiples con genotipos distintos. En el infante 19 se encontraron 3 muestras consecutivas genotipificadas como GII.4 (Fig. 15A), sugiriendo una infección prolongada con excreción continua del virus, como se ha reportado previamente [116,128,129]. Asimismo, el infante 5 tuvo dos muestras consecutivas genotipificadas como GI.3[P13], que también podría reflejar una infección prolongada. Por otro lado, también hubieron co-infecciones; en el infante 2, observamos una con dos norovirus diferentes (GI.3[P13] and GII.17[P17]). Nuestros hallazgos respaldan lo descrito previamente, que sugiere que las reinfecciones de norovirus son comunes con genotipos no relacionados de forma cercana [116,118,120,121,125], al respecto, un modelo propone que los genotipos se agrupan en inmunitipos, sugiriendo que la protección entre genotipos cercanos podría existir, mientras que aquella entre genotipos más distantes pudiera ser baja y permitir reinfecciones [130]. Los sapovirus fueron identificados como G1.1 (2 muestras), GI.7 (1 muestra) y GII.4 (1 muestra) (Tabla 4, Fig. 15B); el infante 15 presentó una co-infección con los genotipos GI.7 y GII.4.

A pesar de la baja cantidad de muestras en las que pudimos identificar el genotipo, pareciera que existe un patrón relacionado con las estaciones del año. Todas las muestras del genotipo GII.7[P7] se detectaron entre noviembre y diciembre 2015, mientras que el genotipo GI.3[P13] se encontró entre febrero y marzo 2016. Asimismo, el genotipo más frecuente, es decir el GII.4, estuvo más presente en los meses de marzo y abril tanto de 2016 como 2015 (Tabla 4, Fig. S3). Además, sin considerar la genotipificación, observamos una gran frecuencia y abundancia de norovirus detectados entre noviembre y abril, tanto en el período 2015 – 2016, como 2016 – 2017, aparentemente independiente de la edad de los infantes (Fig. 16). Esta observación concuerda con lo ya descrito en brotes epidemiológicos donde en el hemisferio norte la mayoría de los casos ocurren durante el invierno y en los países templados, como México, durante la temporada invierno-primavera. Sin embargo, estos reportes de temporalidad no reportan prevalencia de genotipos y el genotipo más prevalente reportado en los brotes usualmente es el GII.4 [112]. A pesar de que nuestro muestreo es limitado, es probable que estemos observando un desplazamiento de genotipos debido a una posible competencia en infectividad o a un montaje de respuesta inmune que es dependiente de cada genotipo, similar a la dinámica de competencia que hemos observado entre las variantes que han circulado en nuestro país durante la pandemia de SARS-COV-2 [131,132].

En el caso de los sapovirus, el genotipo GI.1 estuvo presente en dos infantes (7 y 9) durante mayo 2016 (Tabla 4). No se observaron otro tipo de tendencias para este género.

Figura 16. Detección por estaciones del año de norovirus. En cada infante se señaló con un (*) el mes y año de nacimiento; este se coloca para señalar también que el aumento de lecturas se asocia más hacia la temporalidad que a la edad de los niños.



Dentro de las variables de los niños, una de las que más nos llamó la atención fue el peso bajo. Cuatro de los infantes estudiados (4, 5, 7 y 19) tuvieron un percentil promedio igual o menor a 25 (Tabla 1), y también fueron los que mostraron mayor frecuencia de infección por calicivirus al compararla con aquellos con un percentil promedio de 50 o mayor (71% versus 44% de positividad, $\chi^2 p=0.007$). Asimismo, también encontramos diferencias al comparar las abundancias, (ANOVA con permutaciones para medidas repetidas, $p=0.001$). Esta misma tendencia la pudimos observar cuando consideramos solo las «infecciones» (ANOVA con permutaciones para medidas repetidas, $p=0.007$). También fue interesante observar que la genotipificación pudo determinarse en dos y hasta cinco muestras diferentes en los infantes de peso bajo; los infantes 5 y 19, que describimos con posible excreción viral prolongada tienen un percentil promedio menor al percentil 15. A pesar de que estos percentiles no pueden ser considerados como un reflejo de enfermedad, estos comportamientos de excreción y mayor abundancia de carga viral ya han sido reportados en niños malnutridos [118].

En este estudio detallado sobre la familia más abundante en las muestras fecales de nuestra cohorte se mostró la importancia del uso de técnicas de secuenciación masiva para explorar con más detalle la prevalencia y dinámicas de las infecciones virales. Una aproximación que hemos aprendido y mejorado en los últimos años con la pandemia por COVID-19.

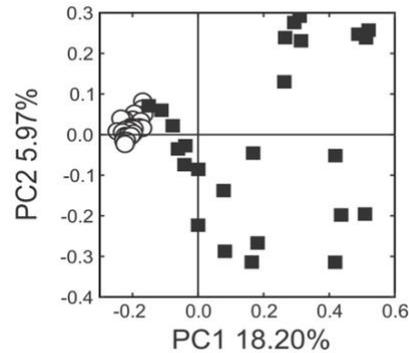
Capítulo 4: El viroma compartido entre la orofaringe y las muestras fecales

Como ya se ha mencionado, la orofaringe tiene un rol híbrido, formando parte, tanto del sistema digestivo como del respiratorio, cuyas funciones varían dependiendo de cada nicho, por lo cual pensar que podrían existir especies o familias compartidas entre la orofaringe y el intestino no parece descabellado. Estudios previos de bacterioma han mostrado, en adultos, que existen ciertas comunidades que se comparten entre la boca y el intestino, con diferencias en la abundancia; en la boca hay ciertas familias que son menos abundantes que en el intestino y otras que son predominantes de la boca, pero no se observan en el intestino. Esto sugiere que: 1) la boca actúa como un reservorio constante de microorganismos y 2) ciertos microorganismos no sobreviven o no se replican en el intestino a pesar de ser transportados ahí [26,27].

Esta comparación de nichos casi no se ha hecho para el viroma eucarionte. Un estudio en población adulta mostró cambios en el viroma posterior a la toma de antibióticos [17], así como diferencias entre la boca y el intestino; al hacer un análisis de composición de muestras (PCoA) observaron diferencias en la composición de las comunidades de acuerdo a los nichos estudiados. Ellos describen poca heterogeneidad entre las comunidades salivales cuando son comparadas con las fecales, sugiriendo que hay mayor conservación o similitud entre las comunidades orales (Fig. 17). También encontraron que los viromas de ambos nichos reflejan el ambiente individual del hospedero, es decir, las muestras no tenían patrones de composición homogéneos entre ellas a pesar de venir de gente de edades y características sociales similares.

Así mismo, ellos sugieren que la estabilidad del viroma oral, el cual no mostró grandes cambios después de la introducción de antibióticos, se debe a la dinámica natural del ambiente oral, el cual está constantemente expuesto a perturbaciones como comida, bebidas e higiene dental, y a la par, una comunidad que es capaz de re-poblar de forma rápida una gran porción de sus virus. Algo que nosotros mismos sugerimos en el capítulo 1 al destacar la individualidad de los viromas, pero también la presencia de ciertos virus «transitorios», tales como rinovirus B o torque teno midi virus.

Figura 17. Análisis de coordenadas principales de la diversidad beta presente en los viomas. Los círculos representan las muestras de saliva y los rectángulos las muestras fecales. Tomado de Abeles *et al*, 2015, PLOS ONE.



En nuestro estudio se encontró que más del 95% de las lecturas en ambos tractos son asignadas ya sea a virus que infectan humanos o a plantas y/u hongos, sin embargo, la proporción en cada tracto fue diferente. A nivel de familia, los virus humanos representaron una mayor abundancia (46% - 11,166,900 lecturas) en las muestras fecales al compararlas con las de la orofaringe (17% - 683,899). Esta distribución fue inversa en los virus de plantas, con un menor porcentaje en las muestras fecales (54% - 4,437,512) en comparación a las orofaríngeas (80% - 3,822,694) (Tabla 5).

Esto sugiere que la principal diferencia en el vioma de estos dos nichos está dada por los virus presentes en el ambiente que entran en contacto con el hospedero ya sea ingeridos o respirados. La orofaringe, al ser un nicho híbrido, está mucho más expuesto que el intestino al medio ambiente, lo cual explicaría la diferencia en proporción de virus de plantas, los cuales posiblemente se encuentran presentes en el ambiente, como el aire, agua, superficies e inclusive la piel y ropa de las personas, aunque también estén presentes en la comida y agua de consumo.

Respecto a los virus que infectan humanos, existe aún una brecha, formada en parte por las limitaciones para describir virus de ARN y el sesgo de amplificación hacia los virus de ADN. Sin embargo, los dos virus más encontradas en la boca de adultos y que se han identificado como los primeros virus en ser encontrados en la boca, son los virus del herpes y papiloma [17,81]. Por otro lado, en el intestino se encontraron virus de humano en un mayor número, por el tropismo de ciertos virus hacia el epitelio intestinal, y menos cantidad de virus de plantas que asumimos que fueron ingeridas.

Por otro lado, se ha sugerido en modelos animales que ciertos virus entéricos (norovirus, astrovirus y rotavirus) tienen la capacidad de replicarse en diferentes zonas anatómicas, incluyendo las glándulas salivales mayores, es decir glándulas sublinguales, submandibulares y parótidas, y también de ser excretados a través de la saliva de forma infecciosa. Si bien este

estudio se hizo en modelos murinos con especies virales murinas, también sugieren que la boca podría actuar como un reservorio para virus conocidos por causar infecciones entéricas [133].

Tabla 5. Porcentaje de abundancia de lecturas de familias distribuidas entre los hospederos por cada nicho.

Muestras	Hospedero	% Abundancia
Orofaríngeas	Animales	1.94
	Plantas y/u hongos	79.76
	Humanos	17.72
	Insectos	0.55
Fecales	Animales	0.11
	Plantas y/u hongos	53.79
	Humanos	46.09
	Insectos	0.01

Aún con las diferencias entre ambos nichos, la realidad es que ambos son parte de un mismo sistema por donde la comida, saliva y líquidos pasan. En nuestro trabajo observamos que 10 especies de virus humanos fueron compartidas entre ambos tractos (Fig. 18a), aunque algunas de estas especies no tuvieron las mismas abundancias o no estuvieron en muestras de ambos tractos tomadas al mismo tiempo o en fechas cercanas. Dentro de estos virus, rotavirus A fue el más abundante, por esta razón, y porque es uno de los patógenos gastrointestinales más importantes, se describirán las observaciones en el siguiente capítulo.

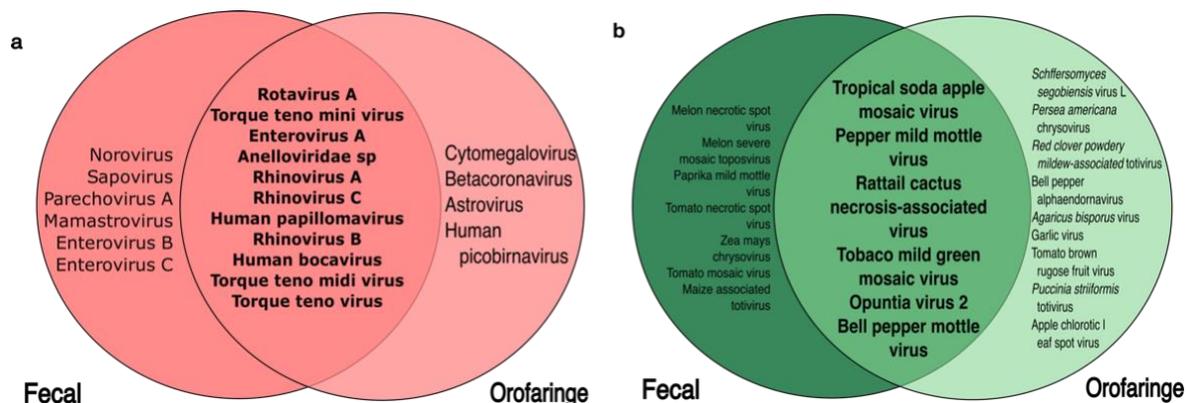
EVA también se encontró en ambos nichos, con una abundancia similar (2%), aunque al analizar cuándo estuvo presente, notamos que no hubo concordancia en el tiempo entre las muestras de los dos nichos. Por otro lado, al tratarse de un virus respiratorio, HRVA representó un 7.3% de la proporción de virus que infectan humanos en la orofaringe, mientras que en las muestras fecales fue un 0.3%. TTMV estuvo presente en 5 de los 9 niños de las muestras orofaríngeas, con muy poca frecuencia y abundancia (1.8%), mientras que en las muestras fecales estuvo presente en todos los niños con una mayor abundancia (5.4%). Con un comportamiento similar, anelloviridae sp fue menos frecuente y abundante (7.8% y 0.14%, respectivamente) en la orofaringe en comparación con las muestras fecales (49% y 1.25%) (Tabla 2 y 3). Los otros virus en ambos tractos se encontraron de forma esporádica entre las muestras, sin mostrar abundancias o patrones de frecuencia.

También se observaron virus que se encontraron de forma única en uno u otro de los nichos, con 4 especies únicas en la orofaringe y 6 en las muestras fecales (Fig. 18a). Algunas de estas especies destacaron por su alta frecuencia y abundancia, como la especie norovirus en las muestras fecales y citomegalovirus en las orofaríngeas. Esto sugiere que, bajo ciertas

condiciones, estos virus pueden estar presentes o replicarse en sus nichos correspondientes sin causar enfermedad aparente.

Respecto a los virus de plantas y/u hongos, se observó una gran variedad (26 especies), pero de éstas sólo 5 fueron compartidas entre ambos nichos, todas pertenecientes a la familia *Virgaviridae* (Fig. 18b). Esta familia está muy presente en vegetales, particularmente de la familia Solanacea, tales como el chile, pimienta o tomate. Si bien no consideramos que los niños consuman picante a una edad tan temprana, no podemos descartar que se use por ejemplo como una alternativa herbolaria para el dolor en encías por la dentición, aunque esto no explicaría porqué esta familia de virus está tan presente inclusive en los primeros meses de vida de los infantes. Algo que quizá pudiera explicarlo sería la excreción de estos virus a través de la leche materna, aunque esto no ha sido estudiado. Debido a que su frecuencia y abundancia es considerable en el viroma de ambos nichos, se hablará más a detalle sobre los virus de plantas en un capítulo posterior.

Figura 18. Diagrama de Venn ilustrando las especies compartidas y únicas de cada nicho que infectan a humanos (a) o a plantas y/u hongos (b). Las especies están ordenadas en orden descendente de acuerdo a su abundancia. En la esquina inferior se observa al nicho al que pertenece cada especie.



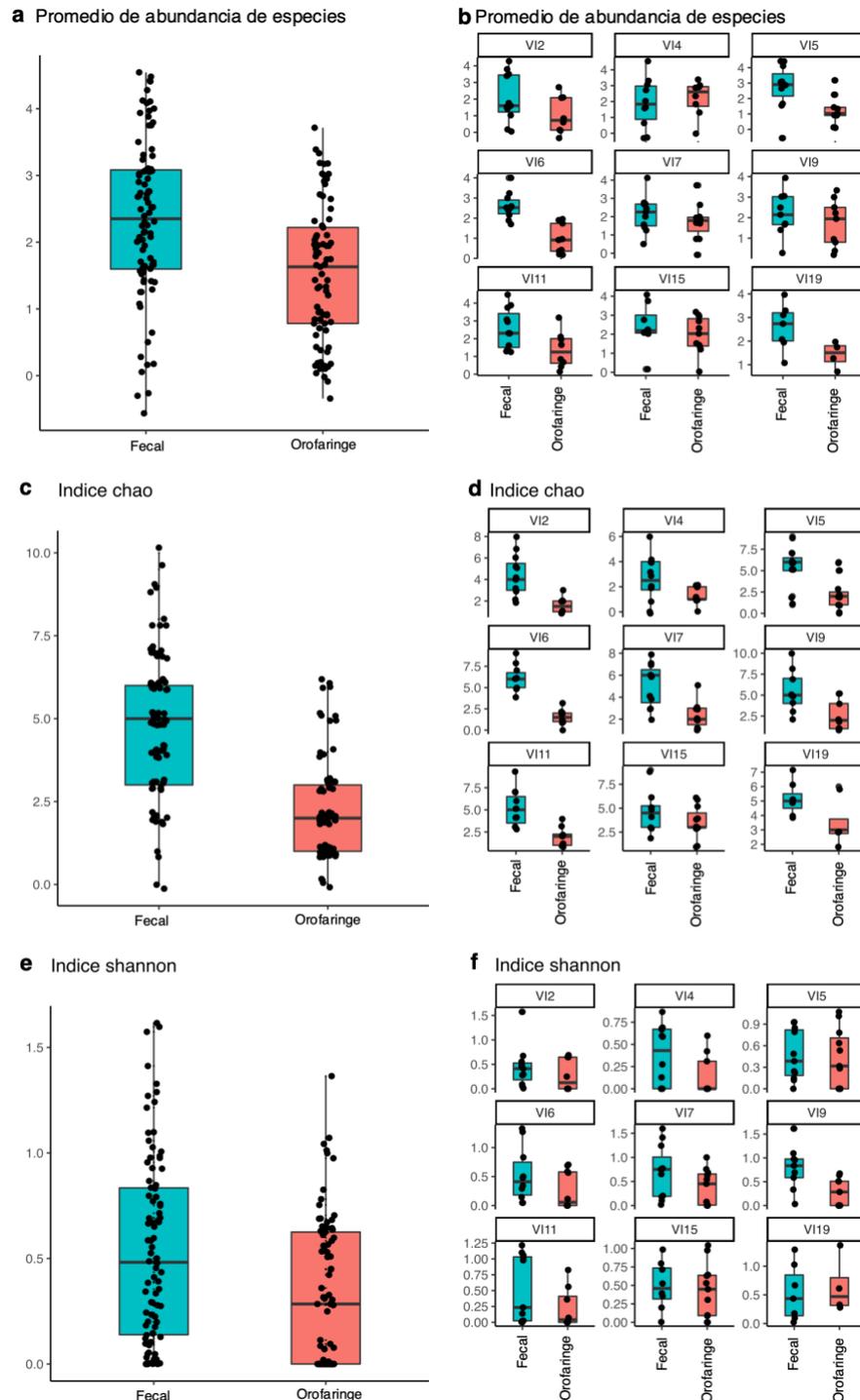
Al comparar las abundancias de las lecturas virales en cada nicho se encontró que fueron significativamente menores en las muestras orofaríngeas (ANOVA con permutaciones para medidas repetidas $p=0.001$) (Fig. 19a) y esta observación también se vio en la mayoría de los niños (Fig. 19b). Uno de los factores que podrían explicar estas diferencias es el total de ácidos nucleicos extraídos, los cuales fueron más difíciles de extraer a partir de un hisopado faríngeo que de una muestra fecal.

Además, la riqueza también fue diferente entre los nichos, teniendo las muestras fecales un mayor número de especies virales por muestra (Fig. 19c), a pesar de que el total de número de especies encontradas fue mayor en la orofaringe (50 especies versus 46 en las muestras fecales) (ANOVA con permutaciones para medidas repetidas $p=0.001$). Estas diferencias

también se observaron en la mayoría de los infantes (Fig. 19d). Esto nos sugiere que a pesar de haber menos especies en las muestras fecales, éstas se comparten entre muestras, y fueron más estables por lo que la riqueza es mayor. Por otro lado, también nos sugiere que hay más especies circulando en la orofaringe y que ésta, debido a su alta exposición ambiental, tiene un viroma más dinámico.

Aunque no fue estadísticamente significativo, la diversidad de las muestras fue mayor en las muestras fecales (Fig. 19e); algunos de los infantes mostraron una diferencia acentuada (4, 6, 7, 9 y 11), mientras que otros (2, 5, 15 y 19) no mostraron diferencias marcadas, sugiriéndonos que en nuestra cohorte hubieron infantes con especies dominantes (menos diversos), así como infantes con una dominancia o distribución similar de especies (más diverso) en ambos nichos (Fig.19f).

Figura 19. Gráficas de barras comparando la diversidad de los virus eucariontes que infectan humanos presentes en ambos nichos. (a) Promedio de abundancia de lecturas asignadas a especies en todas las muestras por nicho, en escala logarítmica base 10; (b) Promedio de abundancia de lecturas asignadas a especie por niño por nicho, en escala logarítmica base 10; (c) Número de especies en todas las muestras por nicho, calculado con el índice de Chao; (d) Índice de Chao calculado por cada nicho de cada niño; (e) Distribución de la abundancia entre especies en todas las muestras de cada nicho, calculada con el índice de Shannon; (f) Índice de Shannon calculado por nicho para cada niño.



Al analizar las composiciones virales entre ambos tractos, se encontró que éste factor explica un 25.4% de variación (PERMANOVA $R^2= 25.4\%$, $p= 0.001$, Fig. S5). Esta compleja interacción entre la composición viene, como hemos establecido, de las dinámicas e individualidad de cada infante y en cada tracto ($R^2=12.8\%$, $p=0.001$). Por esta razón sugerimos que aunque podemos encontrar diferencias en la composición entre ambos nichos, algunos infantes tendrán más diferencias que otros, como observamos con los índices de riqueza y diversidad y los análisis por nicho (Fig. 10 y 14).

Encontramos también, al hacer un análisis no constreñido (PCoA), que hay separaciones entre los nichos dependiendo de los ejes que se vean. Sin embargo, observando los ejes 1 y 2, que explican 54.2% de cómo se observan las muestras de forma reducida en dos dimensiones (Fig. S5A), observamos una separación discreta de las muestras, con algunas sobrelapando. Al complementar con un análisis constreñido de coordenadas principales (db-RDA, Fig. S5D) observamos una separación mucho más clara entre las muestras, sin embargo, no las vemos todas agrupadas; algunas muestras orofaríngeas de los niños 5, 15 y 19 se separan de las demás; al revisar estas muestras, son las que tienen más abundancia o riqueza.

Por otro lado, las muestras fecales están bastante dispersas en el plano dimensional observado (db-RDA, Fig. S5D), lo que creemos se debe a las diferencias de cada niño, pero también a la evolución esperada de la composición viral de las muestras durante el primer año de vida. Resulta interesante observar en la figura 16 la distribución de las muestras, pues pareciera que a pesar de haber sido realizadas en adultos, y tener diferentes especies reportadas, el comportamiento de las muestras es similar, teniendo las muestras orofaríngeas una composición viral más similar entre sus muestras que las muestras fecales.

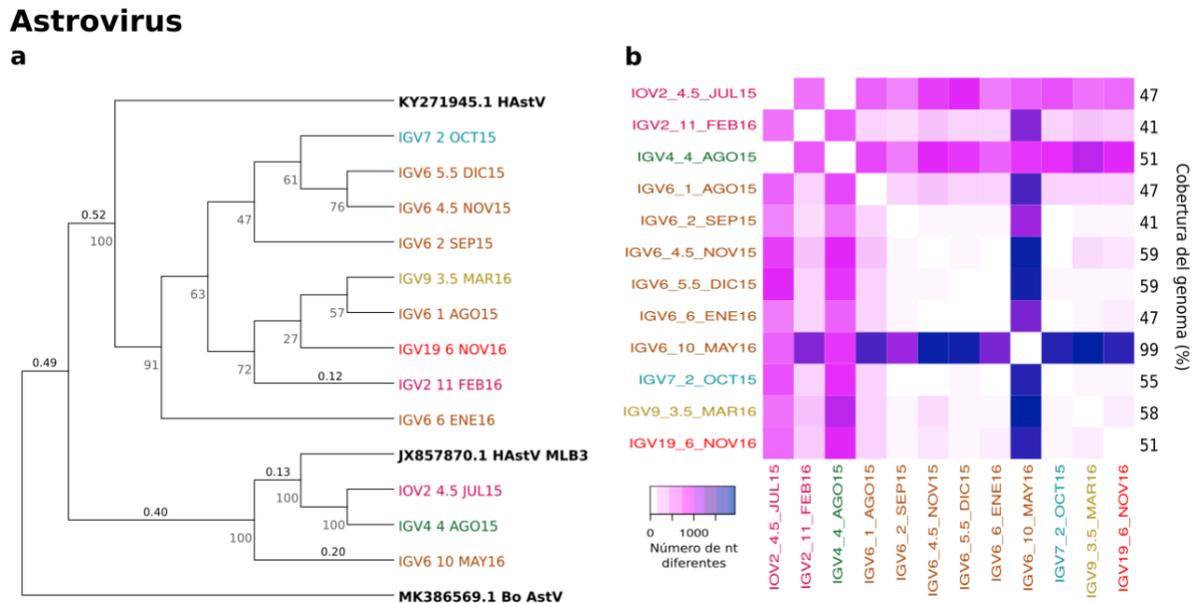
Capítulo 5: Análisis filogenéticos de virus compartidos en ambos tractos

Para evaluar la diversidad genómica de algunas especies que se encontraban en ambos tractos se realizó un análisis que consta de una reconstrucción filogenética y una matriz de diferencias de nucleótidos, la primera, con el afán de ver que tan cercanas son las especies identificadas, ya sea con genoma completo o contigs de mínimo 500 pares de bases, mientras que la segunda compara los posibles cambios de nucleótidos que puedan existir entre regiones en común entre las especies del árbol que tengan un mínimo de 40% de cobertura del genoma.

A pesar de ser una especie con poca abundancia (0.09% y 0.18%) o frecuencia (4.4% y 32%) en la orofaringe y materia fecal, se lograron identificar 12 muestras con astrovirus humano, las cuales pertenecieron a 6 niños. La reconstrucción filogenética nos muestra que 2 especies estaban presentes en la comunidad: astrovirus humano y astrovirus humano MLB3 (Fig. 20a). De forma interesante, el niño 6 tuvo todas sus muestras positivas a este virus, y logramos ensamblar 6 muestras para ser utilizadas en el análisis filogenético, donde 5 de ellas fueron

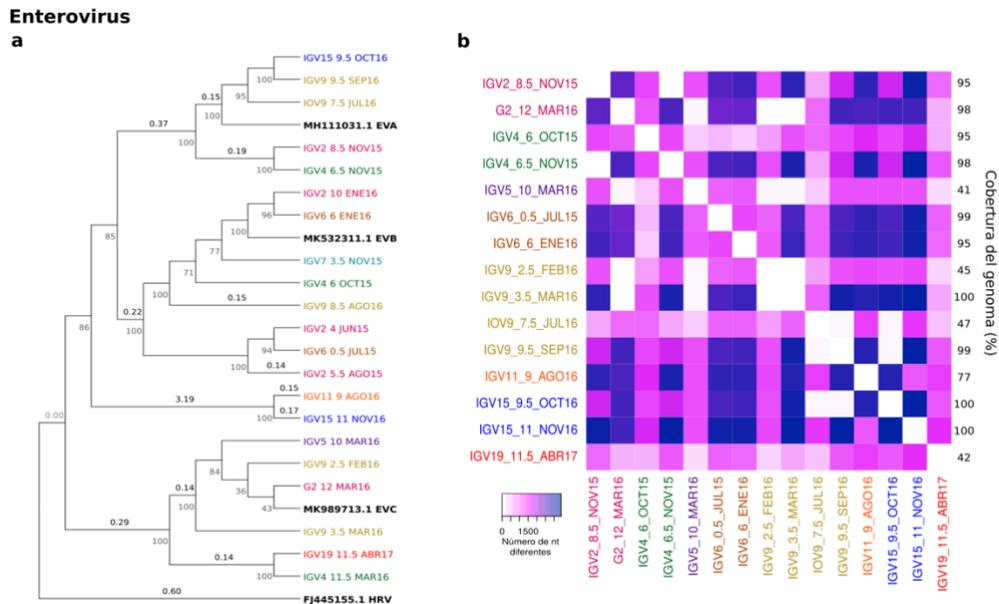
posiblemente asociadas al mismo virus (Fig. 20a) y sin muchos cambios de nucleótidos entre ellas (Fig. 20b). Esto nos sugiere una infección persistente de aproximadamente 6 meses, mientras que la muestra del 10mo mes, fue más similar al astrovirus MLB3.

Figura 20. Diversidad genética de los astrovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM2+F+I. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.



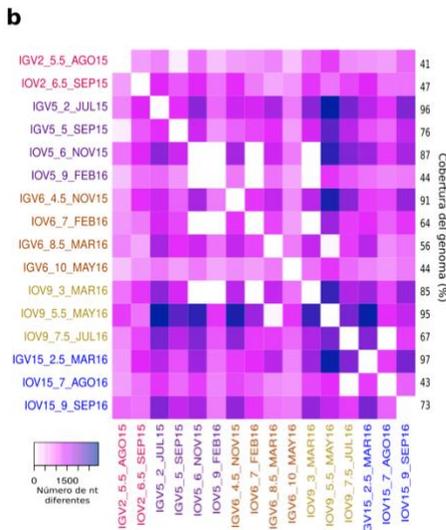
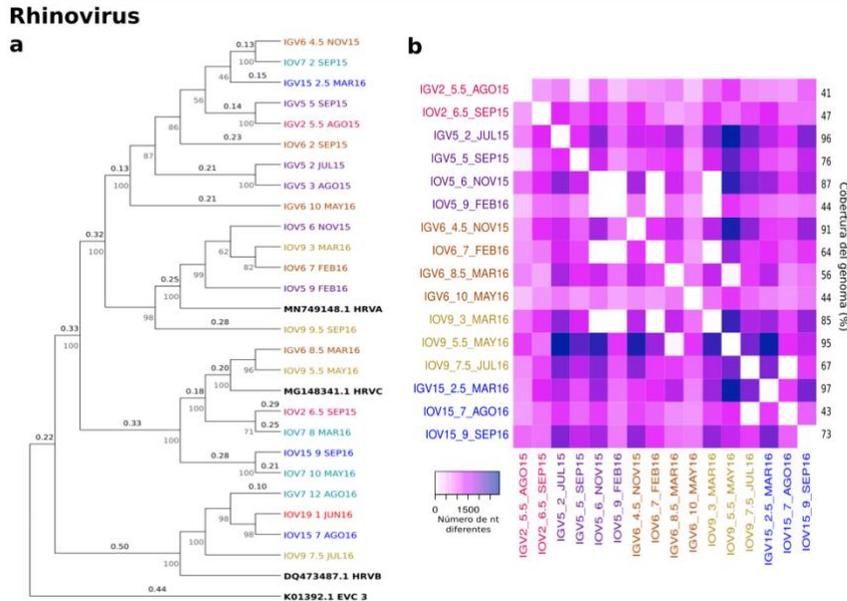
En el caso de los enterovirus, observamos una gran diversidad entre especies, con una división de clados entre enterovirus A, B y C (Fig. 21a), como nota, estos virus suelen tener una identidad del 70% con una cobertura aproximada del 60% entre ellos, es decir, tienen regiones bastante similares, pero al mismo tiempo, son lo suficiente diferentes para poder distinguirlos. Uno de los clados más llamativos es el de enterovirus B, el cual pareciera dividirse en subespecies, pues al observar el mapa de calor de diferencia de nucleótidos (muestra IGV4-6, IGV6-0.5 e IGV6-6) no se observan grandes diferencias nucleotídicas y ambos son genomas casi completos (Fig. 21b). El clado de enterovirus C llama la atención porque las muestras son asociadas a poliovirus. Uno de los casos quizá más interesantes es de la niña 9, quien tiene 5 muestras con la suficiente cobertura para ser analizadas filogenéticamente. Observamos entonces que tiene 2 muestras asociadas a enterovirus A; una muestra orofaríngea (tomada a los 7.5 meses de edad) y una fecal (9.5 meses) sin una diferencia marcada de nucleótidos (Fig. 21b); es posible que se trate de la entrada de este virus y su posterior permanencia. Un muestra intermedia, tomada a los 8.5 meses es más cercana a enterovirus B, mientras que dos muestras tomadas mucho antes (2.5 y 3.5 meses) son de enterovirus C.

Figura 21. Diversidad genética de los enterovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo GTR+F+R2. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.



Respecto a los rinovirus, la especie A fue la más frecuente y abundante en ambos nichos (13.3% y 14.4% de frecuencia y 1.15% y 0.25% de abundancia en muestras orofaríngeas y fecales, respectivamente). Considerando diferencias entre las especies, se observó una gran variabilidad entre los rinovirus (Fig. 22b), lo cual es esperado si consideramos que entre las especies (A, B y C) solo hay un estimado de 40% de cobertura entre ellos, con identidades que rondan entre el 60 y 70%, es decir, son bastante diferentes. A pesar que en 6 niños se encontraron muestras consecutivas (2, 5, 6, 7, 9 y 15), no se observaron especies similares de un mismo o ambos tractos, a excepción de 2 muestras gástricas del infante 5 (2 y 3) que están en el mismo clado (Fig. 22a). En el caso de rinovirus A se observa una división en dos clados, lo que nos sugiere que quizá estaban circulando diferentes sub-especies en la comunidad.

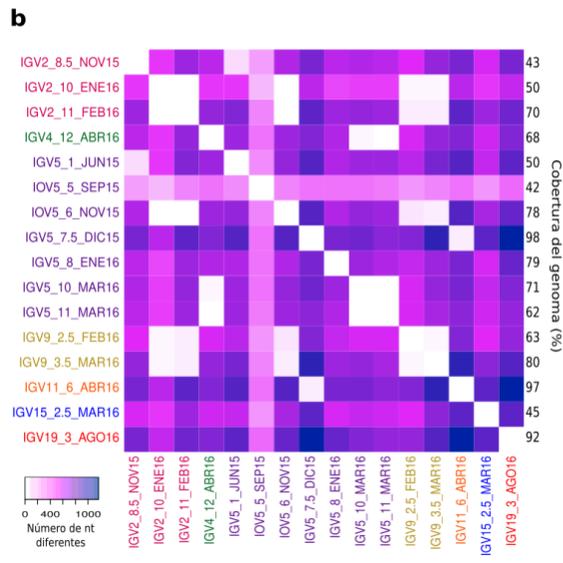
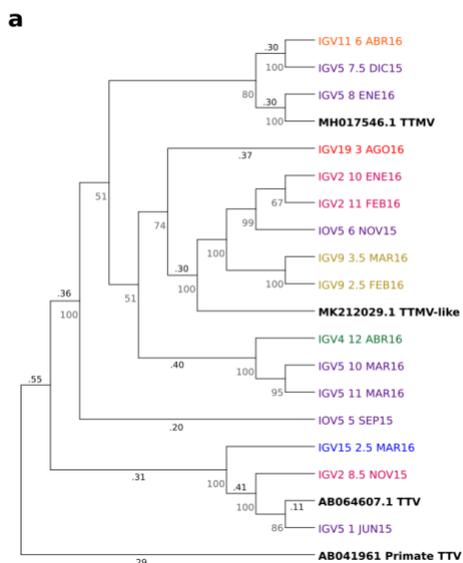
Figura 22. Diversidad genética de los rinovirus. (a) Árbol filogenético de genomas completos y contigs (>500) construido con el modelo GTR+F+I+G4. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.



Los anellovirus tuvieron una gran variabilidad genética (Fig. 23b), y en tres niños (2, 5, y 9) se observaron muestras fecales tomadas de forma consecutiva entre los meses de Enero y Marzo de 2016 que mostraban el mismo virus en cada niño (Fig. 23a), sugiriendo una excreción prolongada con dinámicas asociadas a una temporalidad estacional. De forma más interesante, en la niña 5 se encontraron diversas muestras, tanto fecales como orofaríngeas, que sugieren una exposición constante a diferentes especies de anellovirus.

Figura 23. Diversidad genética de los anellovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM2+F+R2. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.

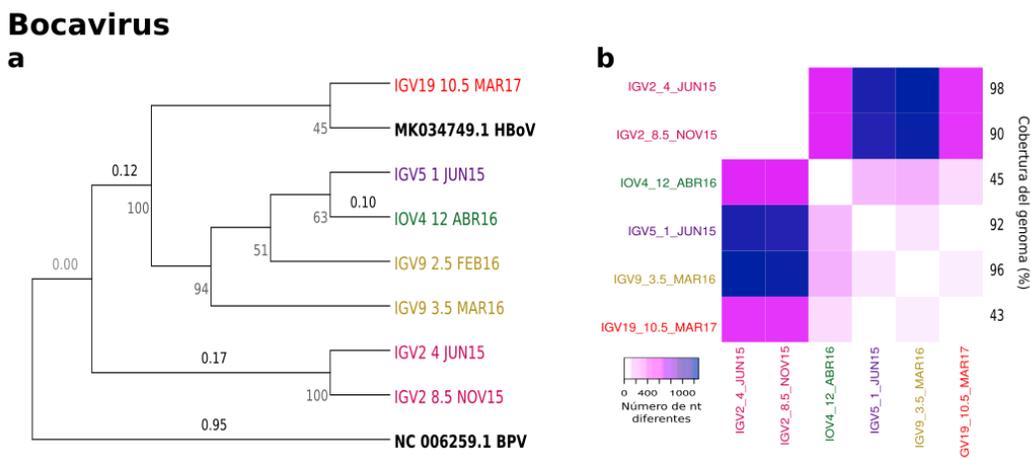
Anellovirus



A pesar de tener una frecuencia considerable en la orofaringe y contrariamente, una baja frecuencia en las muestras fecales, la abundancia de los papillomavirus fue baja y por lo tanto, no hubieron demasiadas muestras donde se lograron ensamblar contigs de más de 500 pares de bases o tener coberturas mayores al 40% del genoma (Fig. 8a). Sin embargo, dentro de estas pocas muestras destacaron las de dos niñas; en la niña 11 se observó una relación cercana entre dos muestras orofaríngeas (6 y 9.5) y dos fecales (0.5 y 1.5) (Fig. 8b), mientras que en la niña 15, tres muestras orofaríngeas fueron relativamente cercanas (1, 3.5 y 9). En ambos casos creemos que existe la posibilidad que las demás muestras pudieran estar relacionadas genéticamente, pues los papillomavirus son bastante variables entre sus géneros y especies.

Finalmente, los bocavirus se identificaron en 5 niños, donde las muestras tomadas del niño 2 están separadas en otro clado (Fig. 24a) y al ver el número de diferencias de nucleótidos, se observa que entre ellos son idénticos (Fig. 24b), sugiriendo una posible excreción prolongada del mismo virus. En los otros 4 niños, se observa una diferencia de nucleótidos mucho menos marcada a pesar de ser muestras tomadas en diferentes puntos del tiempo, en algunos casos, inclusive con 1 año de diferencia.

Figura 24. Diversidad genética de los bocavirus. (a) Árbol filogenético de genomas completos y contigs (>500) construido con el modelo HKY+F. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.



Capítulo 6: Identificación de rotavirus vacunal

Los rotavirus son virus de ARN de doble cadena, el cual está compuesto por 11 segmentos y rodeado por tres capas concéntricas de proteína. En la cápside del virus hay dos proteínas,

VP4 y VP7, que definen el serotipo del virus e inducen anticuerpos que neutralizan la infección viral.

Los rotavirus son conocidos por ser la primer causa de gastroenteritis no bacteriana en niños menores de 5 años en países donde no se ha establecido un programa de vacunación, teniendo una alta tasa de mortalidad en países en vías de desarrollo [134].

Actualmente existen dos vacunas de administración oral que previenen la gravedad clínica de rotavirus: 1) RotaTeq (Merck, Whitehouse Station, NJ, USA) aprobada en 2006, es una vacuna pentavalente, recombinante, basada el rotavirus de bovino WC3. Cuatro de los virus tienen todos los genes del rotavirus bovino, excepto el gen que codifica para VP7, los cuales son de origen humano y representan los genotipos G1, G2, G3 y G4 que son los más frecuentes en la población infantil. El quinto virus de la vacuna tiene también todos los genes del rotavirus WC3, excepto el gen que codifica para la proteína VP4, genotipo P1A[8], que es el genotipo de VP4 más frecuente en humanos [135]; y 2) Rotarix, vacuna monovalente de rotavirus humano atenuado (GlaxoSmithKline, Brentford, UK), aprobada en 2008 [136]. La administración de RotaTeq consta de 3 dosis, iniciando entre las 6 y 12 semanas de vida, con dosis subsecuentes con intervalos de 4 a 10 semanas.

RVA fue la especie más abundante de virus de humano en los dos nichos, teniendo casi la misma proporción de lecturas en la orofaringe (13.5%) y en las muestras fecales (13.3%) (Tabla 2 y 3). En dos infantes, rotavirus estuvo presente durante casi todo el año en ambos nichos (infantes 2 y 15), mientras que en otros se encontró más frecuentemente en las muestras fecales, en comparación con la orofaringe; en este último nicho este virus se detectó de forma esporádica, con excepción de los infantes 4 y 9 (Fig. 25).

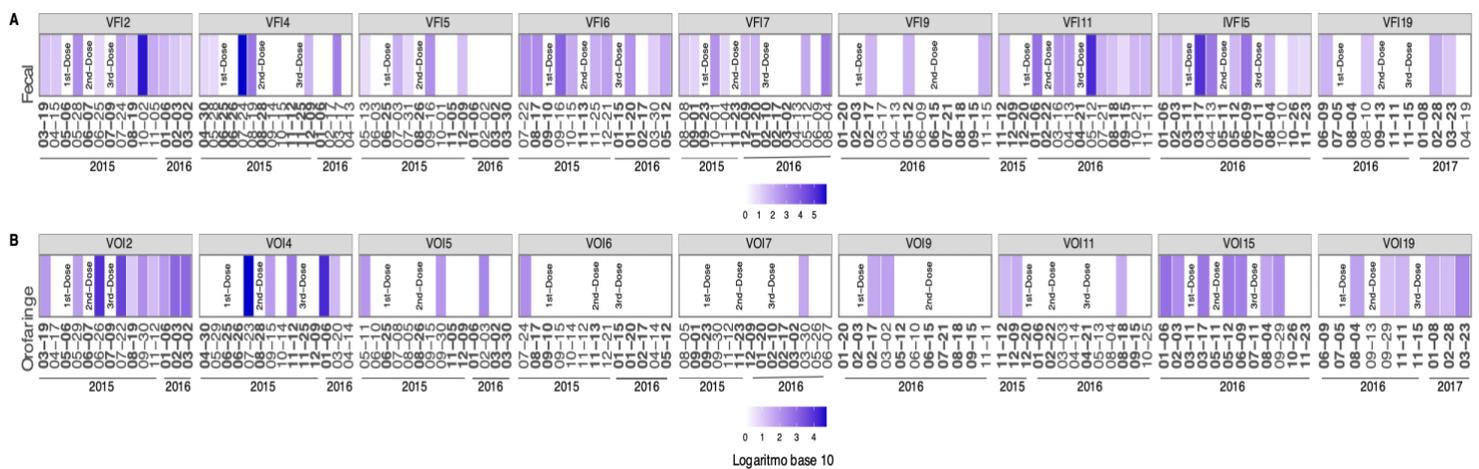
Como señalamos previamente, la abundancia de lecturas en las muestras orofaríngeas fue menor que en las fecales, por lo que no fue sorprendente encontrar una diferencia 16 veces menor de lecturas en la orofaringe (92,583 contra 1,494,082) para este virus. A pesar de que en algunos infantes pudimos observar la presencia de rotavirus en ambos nichos, no pudimos determinar si el virus encontrado en tiempos similares fue consecuencia de la misma o diferentes exposiciones.

En el análisis por nicho, RVA tuvo una abundancia general del 2% en la orofaringe y se encontró en casi todas las muestras orofaríngeas de los niños 2, 15 y 19 (Fig. 25a). En las muestras fecales, RVA se encontró en todos los niños, representando el 9.1% de las lecturas y estando presente en 65 (67%) de las muestras, siendo más frecuente en los infantes 2, 6, 7, 11 y 15 (Fig. 25b). Además, en las muestras fecales encontramos una disminución de la prevalencia desde el primer hasta el tercer trimestre de vida (81%, 70% y 42%, respectivamente) y también de la abundancia, donde fue más abundante durante el primer

semestre de vida que en el segundo (ANOVA con permutaciones para muestras repetidas $p=.001$), lo cual muy probablemente está relacionado a la vacunación.

Debido a que teníamos los datos de vacunación de los infantes con la vacuna RotaTeq, nos pareció interesante analizar, primero de forma general, si el incremento de lecturas asociadas a rotavirus podría deberse a la vacunación (Fig. 25), y después de forma más detallada, si podríamos identificar secuencias asociadas a la vacuna para confirmar si realmente estábamos recuperando el virus vacunal.

Figura 25. Presencia de las lecturas asignadas a la especie rotavirus A en las muestras fecales (A) y orofaríngeas (B) en escala logaritmo base 10. En la parte superior se marca a cada infante; VFI = viroma fecal infantil, VOI= viroma orofaríngeo infantil, seguido por el número de infante. En la parte inferior la fecha de toma de muestra y las fechas de vacunación con formato mes-día (eje x), adentro del mapa de calor, las vacunas están marcadas como '1, 2 o 3ra dosis'. En negritas, las fechas que tienen toma de muestra pareada en ambos nichos.



Para hacer un análisis más fino se hizo una búsqueda intencionada de las lecturas de rotavirus, para después hacer ensamblados *de novo*, y en algunos casos obtener secuencias consenso dirigidas hacia los segmentos de los virus de la vacuna RotaTeq, que es la que se administró en el poblado de Xoxocotla durante el periodo de toma de muestras para nuestro estudio. Estos consensos se analizaron de dos formas diferentes, primero con la herramienta en línea ViPR [64], que permite identificar a qué segmento y genotipo pertenece cada gen y después, haciendo un análisis filogenético y de diferencia de nucleótidos (Fig. 26).

La reconstrucción filogenética del gen VP4 dividió las secuencias en 3 grupos (Fig. 26a); serotipo P1A, asociado a los serotipos del VP7 (G1 a G4), los cuales pueden tener su origen de cepas locales o vacunales; serotipo P8, asociado al serotipo de VP7 G12, probablemente de cepas locales y finalmente, el serotipo P7, de origen bovino y derivado de la vacuna. En el caso del gen VP7 estos se agruparon en cuatro serotipos: G1, G3 y G4 derivados de la vacuna

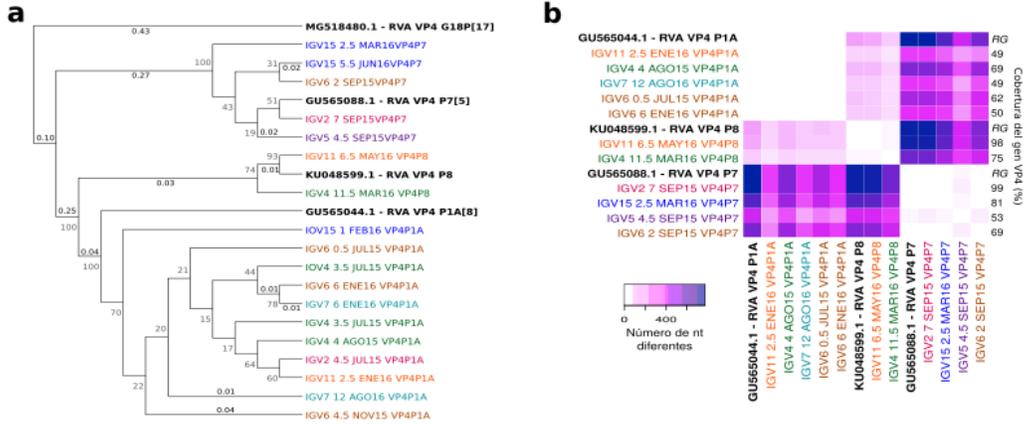
o de cepas locales, y el serotipo G6 el cual es bovino y está relacionado a la vacuna (Fig. 26c). En ambos casos, no se observaron grandes diferencias de nucleótidos dentro de cada serotipo (Fig. 26b y d).

En la orofaringe, utilizando la herramienta ViPR se identificaron dos muestras (infante 2, mes 4 e infante 4 mes 3.5) con rotavirus asociado a la vacuna RotaTeq; ya sea porque se identificaron genes bovinos o bien, porque tenían diversos genes de VP7. Adicionalmente, la reconstrucción filogenética sugiere que la primer muestra del infante 15, tomada a los 8 días de nacimiento, también está relacionada a la vacunación (Fig. 26c).

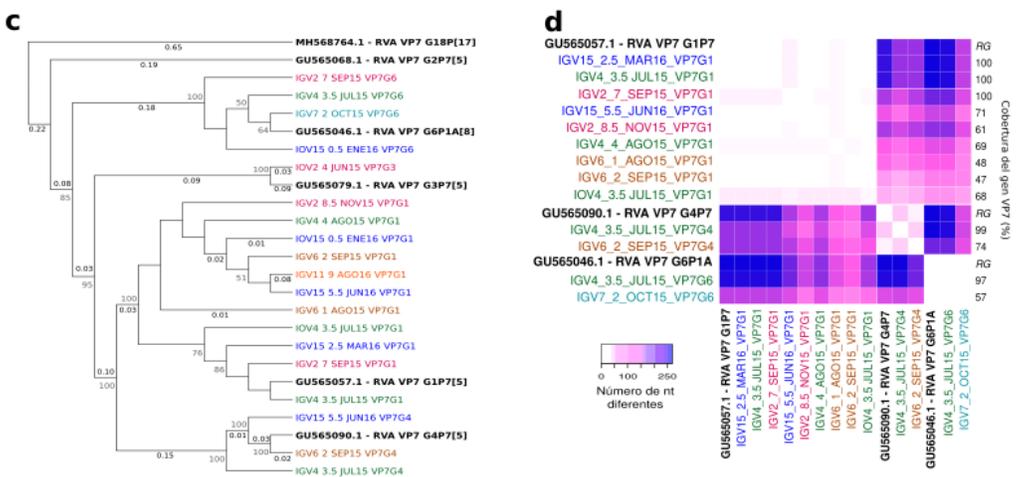
Como dato interesante, la muestra del niño 2 fue tomada 19 días después de la segunda dosis vacunal (Tabla 6), sugiriendo una posible replicación en la orofaringe, como se demostró en las glándulas salivales en un modelo murino [133]. De interés, se ha descrito también la presencia de rotavirus en secreciones respiratorias por transmisiones caseras [137] y en aspirados orofaríngeos de niños hospitalizados por diarrea aguda con síntomas respiratorios [138].

Figura 26. Diversidad genética de los segmentos VP4 y VP7 de rotavirus. (a) Árbol filogenético del segmento VP4 con genomas completos y secuencias parciales (contigs >500) construido con el modelo HKY. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura del gen VP4. (c) Árbol filogenético del segmento VP7 con genomas completos y secuencias parciales (contigs >500) construido con el modelo HKY+I. (d) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura del gen VP7. Los genomas de referencia y raíz están en negritas.

Segmento VP4 de Rotavirus



Segmento VP7 de Rotavirus



En el trabajo reciente de Gosh y cols [133], se demostró que rotavirus, norovirus y astrovirus, virus con tropismo intestinal, tienen la capacidad de replicarse en las glándulas salivales y se mostró también que el virus en la saliva es una vía de transmisión eficiente en su modelo de estudio. Es interesante considerar la posibilidad de que la vía oral pueda también representar una vía de transmisión en humanos a través de secreciones como la saliva o aerosoles; es tentador también especular que este mecanismo podría funcionar también como una forma pasiva de inmunización para los niños no vacunados que estuvieran en contacto con uno vacunado.

En las muestras fecales, utilizando la herramienta ViPR, se logró identificar una muestra asociada a virus vacunal en cada uno de estos cuatro infantes (2, 4, 5 y 6), y en dos muestras del infante 15 (Tabla 6), por otro lado, la reconstrucción filogenética identificó una muestra más, en esta ocasión, de la niña 7 en su segundo mes de vida (Fig. 26c). Es importante resaltar que cuatro de estas muestras fueron tomadas más de 20 días después de que la vacuna fue administrada, sugiriendo que quizá se replican y que probablemente estamos observando una excreción prolongada de virus vacunal, lo que también podría contribuir a la inmunidad

poblacional de manera pasiva. Finalmente, en dos muestras fecales de niños diferentes (4 y 11) se observó un VP4 diferente al vacunal, sugiriendo infecciones con una cepa local.

Tabla 6. Muestras en las cuales se identificaron genes asociados a la vacuna RotaTeq. Se puso la fecha de vacunación previa a una muestra identificada con genes de rotavirus. Todos los genes fueron identificados con la herramienta ViPR (ver la sección de materiales y métodos).

Tipo de muestra	Infante	Edad (meses)	Fecha de muestreo	Fecha de vacunación	Días después de vacunación	Dosis de la vacuna	Origen de los genes de rotavirus	
							Bovina	Humano
Orofaringe	2	4	26-06-2015	07-06-2015	19	2	VP1	VP7(G1), VP7(G2), VP7(G3), VP7(G4)
	4	3.5	23-07-2015	25-06-2015	28	1	-	VP7 (G1), VP7(G2), VP7(G3), VP7(G4), VP4(P1A)
Fecal	2	7	02-10-2015	09-07-2015	85	3	VP1, VP4(P7), VP6, NSP1, NSP2, NSP4, NSP5	VP7(G1), VP7(G2), VP7(G3), VP7(G4)
	2	8.5	05-11-2015	09-07-2015	119	3	-	VP7(G1), VP7(G4), VP4(P1A)
	4	3.5	24-07-2015	25-06-2015	29	1	VP1, VP2, VP6, NSP1, NSP2, NSP3, NSP4, NSP5, VP7	VP3, VP7(G1), VP7(G2), VP7(G4), VP4(P1A)
	5	4.5	16-09-2015	26-08-2015	21	2	VP1, VP4(P7)	VP7(G1)
	6	2	16-09-2015	10-09-2015	6	1	VP1, VP2, VP4(P7), VP6, NSP1, NSP2	VP7(G1), VP7(G4)
	15	2.5	17-03-2016	11-03-2016	6	1	NSP1, NSP2, VP1, VP6	VP7(G1), VP7(G4)
	15	5.5	09-06-2016	11-05-2016	29	2	NSP1, VP1, VP6	VP7(G1)

La presencia de RVA ha sido descrita tanto de forma frecuente [33,37,96] como en número limitado de muestras fecales [32,34,100,104] en estudios llevados a cabo en niños sanos de diversos países. En dos de estos estudios, se detectó la presencia de genes asociados a la vacuna Rotarix [14,38] o RotaTeq [37].

En nuestro estudio, las lecturas asociadas a genes vacunales bovinos o humanos se identificaron en 11 (6%) muestras. En algunas muestras no se logró la genotipificación por falta de cobertura de acuerdo al valor de corte de la herramienta ViPR, pero si en el árbol filogenético, lo cual nos sugiere que con mayor profundidad o un método dirigido a encontrar virus vacunales, podríamos quizá observar más genes asociados a RotaTeq en nuestras muestras. Las muestras dónde se identificó RotaTeq fueron detectadas desde los 6 hasta los 90 días posteriores a la vacunación. No podemos establecer si esto representa una excreción prolongada, o una transmisión pasiva de otro infante recientemente vacunado. Se ha reportado previamente la transmisión de rotavirus vacunal entre hermanos [139], así como la excreción prolongada de virus vacunal, pero no después de los 15 días posteriores a la vacunación [140].

Nuestros hallazgos sugieren que la excreción podría ser más allá de los 15 días post-vacunación, y de forma interesante, que más allá de monitorizar la presencia de rotavirus en aguas residuales como una de las formas más comunes de gastroenteritis por transmisión de rotavirus, es tentador plantearnos si también se pueden identificar secuencias vacunales en aguas residuales, demostrar si estas pudiesen ser infecciosas y que efecto podría tener la ingesta de virus vacunal en la salud o estado inmune de las personas.

Capítulo 7: Los niños están en constante contacto con los virus de plantas, principalmente de la familia *Virgaviridae*

Los virus de plantas están presentes de forma abundante y ubicua en la naturaleza, tanto en plantas silvestres, como aquellas de consumo humano. Al respecto, recientemente se ha reportado respuesta inmune hacia virus de plantas en cohortes humanas, así como presencia de estos virus en muestras fecales de adultos [105] y de niños [31,33,38,95]. Su origen en los humanos se ha relacionado principalmente a la alimentación [95,105], e igualmente se ha demostrado que estos virus están presentes en el aire [88]. La mayoría de los virus de plantas o fitovirus son de ARN y han sido descritos como los virus de ARN más prevalentes en las heces humanas [105].

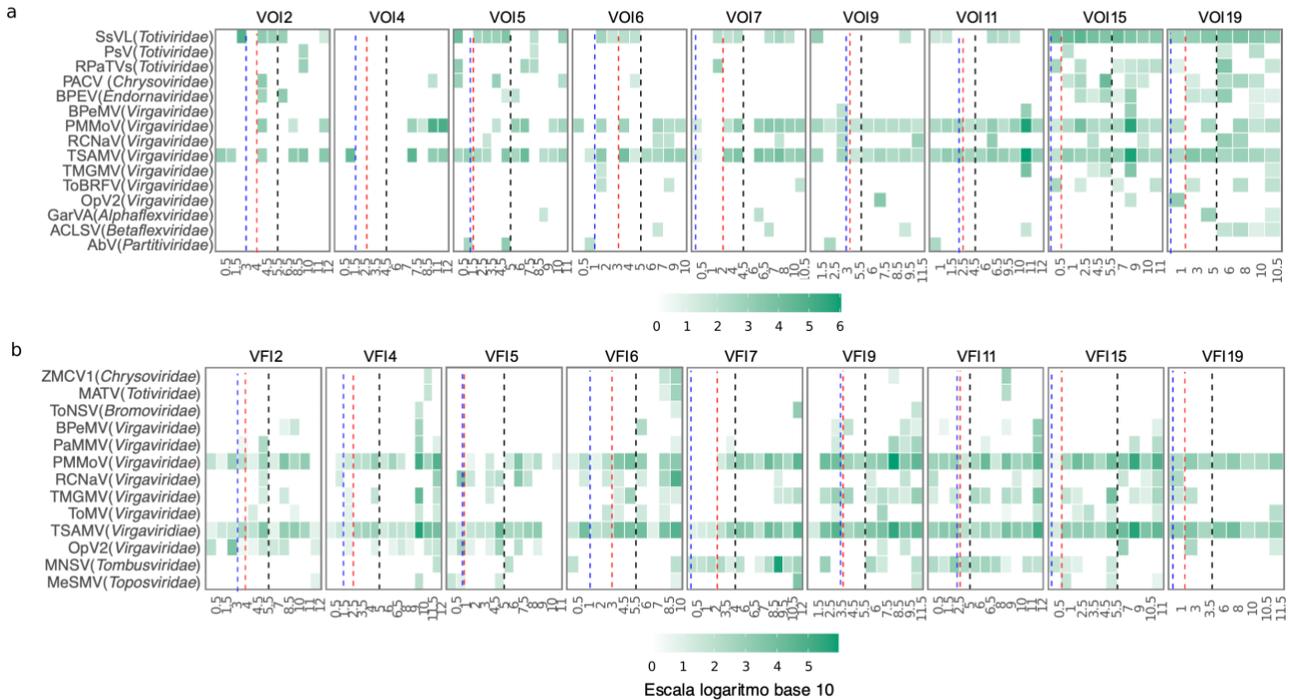
En nuestro estudio se encontró una gran variedad de virus de plantas en la orofaringe, representando a nivel de especie casi el 83% de abundancia total (Tabla 2, Fig. 6). La mayoría de la abundancia viene de 3 especies virales, las dos primeras de la familia *Virgaviridae* : pepper mild mottle virus (PMMoV) y tropical soda apple mosaic virus (TSAMV). La tercer especie es un virus nuevo, pero cercano taxonómicamente a nivel de proteína al virus *Scheffersomyces segobiensis* L de la familia *Totiviridae*. Esta última especie se encontró con gran abundancia y frecuencia principalmente en las niñas 15 y 19 (Fig. 27a). Al revisar los datos sociodemográficos y de crecimiento no encontramos algún factor común entre estas dos niñas que pudiera explicar la presencia de este virus. De forma interesante, mientras que en general podríamos decir que las demás especies de virus de plantas, independiente de la familia viral a la que pertenecen, se observaron de forma esporádica, con poca abundancia y frecuencia entre las muestras de los infantes, nuevamente hay un comportamiento anormal en las niñas 15 y 19 con mayor frecuencia de virus de plantas durante su primer año de vida (Fig. 27a).

El patrón esporádico observado pudiera deberse a la exposición constante de la orofaringe al medio ambiente, como se ha reportado previamente [28,76]. Asimismo, también consideramos que estos virus se encuentran circulando en el aire y podrían ser inhalados; un estudio de muestras de aire de una guardería infantil reportó que los virus de plantas representaron el 94% de frecuencia de los virus de ARN descritos, con una alta prevalencia durante primavera, verano y otoño, decreciendo en invierno, cuando los virus humanos tienen una mayor presencia [88]. Esto es interesante, pues en invierno es cuando hay menos contacto con el exterior, usualmente se cierran las ventanas y los lugares se ventilan menos, sugiriendo que de no ser por este tipo de comportamientos, pudiera haber una exposición continua a virus de plantas.

Por otro lado, en las muestras fecales la abundancia de lecturas asignadas a especies de virus de plantas fue mucho más baja en comparación a la orofaringe, con 27% de abundancia de lecturas en las muestras fecales (Fig. 10) versus 83% en la orofaringe. Tres especies fueron las más abundantes; 2 de la familia *Virgaviridae* (igual a lo que se observó en la orofaringe), y otra especie perteneciente a la familia *Tombusviridae*, llamada melón necrotic spot virus (MNSV) (Tabla 3, Fig. 27b), el cual ha sido descrito en sistemas de irrigación agrícola en Eslovenia y España [101]. De 19 especies de virus de plantas detectadas, 8 pertenecen a la familia *Virgaviridae* y éstas son las más frecuentes y abundantes en las muestras (Tabla 3, Fig. 27b), las demás especies son bastante esporádicas a excepción de MNSV.

Figura 27. Mapa de calor que representa las abundancias normalizadas expresadas en escala de logaritmo base 10 de las especies virales de plantas que están presentes en las muestras orofaríngeas (a) o fecales (b). Cada panel representa a un niño con el identificador en la parte superior de cada panel y la edad en meses en la parte inferior (eje x). VOI= viroma orofaríngeo infantil. VFI = viroma fecal infantil. Las especies que fueron taxonómicamente asignadas como 'sp' se omitieron de la figura. Dentro del paréntesis está escrita la familia a la que pertenece casa especie. Las líneas punteadas negras dividen los meses entre el primer y segundo semestre de edad, las azules indican la introducción de otros líquidos y las rojas de frutas y/o verduras. **Especies que se replican en plantas y/u hongos presentes en la orofaringe:** SsVL, *Schffersomyces Segobiensis* virus L; PsV, *Puccinia Striiformis* totivirus; RPaTVs, *Red Clover Powdery mildew-associated* totivirus; PACV, *Persea Americana* chrysovirus; BPEV, *Bell Pepper alphaendornavirus*; BPeMV, *Bell Pepper mottle virus*; PMMoV, *Pepper Mild Mottle virus*; RCNaV, *Rattail Cactus necrosis-associated virus*; TSAMV, *Tropical Soda Apple* mosaic virus; TMGMV, *Tobacco Mild Green* mosaic virus; ToBRFV, *Tomato Brown Rugose Fruit* virus; OpV2, *Opuntia* virus 2; GarVA, *Garlic virus*; ACLSV, *Apple chlorotic leaf spot virus*; AbV, *Agaricus Bisporus* virus. **Especies que se replican en plantas y/u hongos presentes en materia fecal:** ZMCV1, *Zea mays* chrysovirus; MATV, *Maize associated totivirus*; ToNSV, *Tomato necrotic spot virus*; BPeMV, *Bell Pepper mottle virus*; PaMMV, *Paprika Mild mottle virus*; PMMoV, *Pepper Mild mottle virus*; RCNaV, *Rattail Cactus necrosis-associated virus*; TMGMV, *Tobacco Mild Green* mosaic virus; ToMV, *Tomato mosaic virus*; TSAMV, *Tropical Soda Apple* mosaic virus; OpV2, *Opuntia* virus 2; MNSV, *Melon necrotic spot virus*; MeSMV, *Melon severe mosaic toposvirus*.

Virus eucariontes de plantas



Es importante destacar que los niños de nuestro estudio viven en una población semi-rural que no está completamente pavimentada, por lo que están en contacto constante con el polvo y tierra. Así mismo, la principal actividad económica es la agricultura y los campos de cultivo están cerca de la zona poblada, lo cual podría explicar la presencia de virus de plantas en ambos nichos; en la orofaringe al ser transportadas de forma aérea o bien, en el polvo de la ropa de sus padres o cuidadores, mientras que en el intestino por la ingesta de alimentos, té o agua; en México existen tradiciones herbolarias utilizadas para aliviar molestias en los bebés, tales como el consumo de té o infusiones los cuales suelen darse inclusive antes del inicio de ingesta de otros alimentos que no sean leche materna [141].

En particular, las especies más abundantes descritas en nuestra cohorte pertenecen al género *Tobamovirus*, de la familia *Virgaviridae*. En las fases iniciales de este estudio, y considerando su alta frecuencia en nuestra población, hicimos un reporte preliminar de su presencia en muestras fecales de 3 de los niños participantes [142]. Dos especies fueron de especial interés por su abundancia y frecuencia: PMMoV y TSAMV. El PMMoV se había descrito previamente en muestras de heces con hasta 10^9 partículas virales por gramo de heces; mientras que no había reportes previos de la presencia del TSAMV en humanos, lo que puede deberse a que al momento de hacer el estudio este virus había sido recientemente descrito, con solo dos genomas completos en GenBank [143].

El ensamble de genomas completos y contigs de estos virus y su análisis filogenético (Fig. S6 y 8) indicaron que no hay relación entre las especies identificadas en la orofaringe y la materia

fecal, sugiriendo que las cepas de virus encontradas en la orofaringe no están directamente relacionadas a las encontradas en la materia fecal. Igualmente, al observar las diferencias entre nucleótidos, observamos que hay una gran variabilidad entre las muestras (Fig. S7 y 9). Solo en casos particulares se observó la misma cepa en muestras consecutivas.

El análisis de la cohorte completa extendió estos resultados, mostrando que en las muestras fecales de los 9 niños estuvieron presentes de manera frecuente y abundante los virus TSMoV y PMMoV. El análisis de la presencia de virus de plantas en la orofaringe, que se realizó posteriormente para todos los infantes, mostró también la detección frecuente de TSMoV y PMMoV. Los hallazgos realizados se resumen a continuación:

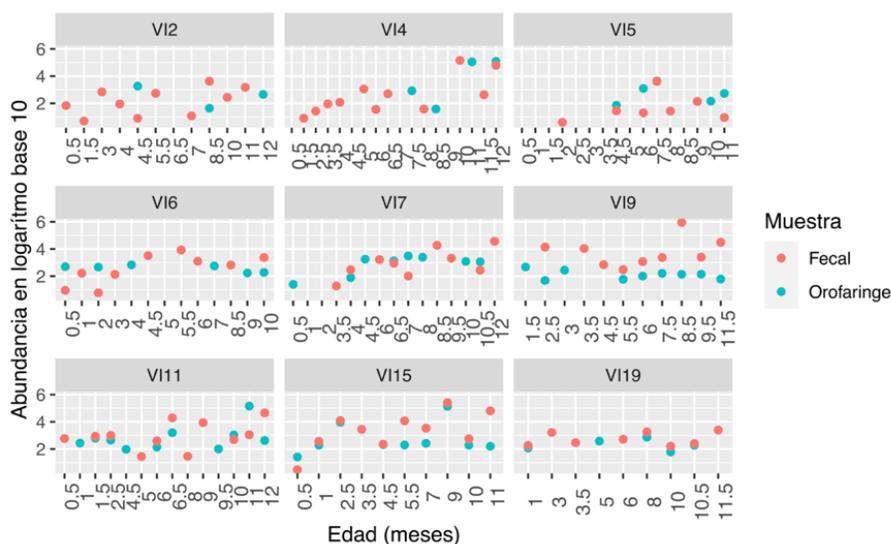
- 1) PMMoV tuvo una abundancia de 31.5% de lecturas en la orofaringe y 10.3% en muestras fecales. Esta abundancia es en proporción a las lecturas totales de cada nicho, por lo que a pesar de que es más abundante en la orofaringe, al ver lecturas crudas, en realidad hay más lecturas de las muestras fecales que las orofaríngeas (Fig. 28). Respecto a la frecuencia, ésta fue más alta en las muestras fecales (86.6% en comparación a 65.6%). Al analizarla como especie, encontramos que su abundancia incrementa después del primer semestre de vida (ANOVA con permutaciones para medidas repetidas $p < 0.05$). Con excepción de los infantes 2, 4 y 5 (Fig. 28), en todos los demás niños este virus estuvo presente en ambos nichos casi durante todo su primer año de vida. El análisis filogenético y de nucleótidos mostró una gran diversidad (Fig. S6 y 7), con los virus del infante 4 teniendo la menor diferencia de nucleótidos (Fig. S7), mientras que en el árbol, solo 2 niños tuvieron muestras consecutivas dentro de un mismo clado; niño 6, muestras a los 4.5 y 5.5 meses de edad y niña 15, con muestras a los 2.5 y 3.5 meses (Fig. S6), mientras que la niña 7 tuvo un muestra orofaríngea (10 meses) y una gástrica (12 meses) dentro de un mismo clado.

PMMoV ha sido descrito como un marcador de agua contaminada por su alta estabilidad ambiental [105]. Sin embargo, no encontramos diferencia en abundancia de esta especie al comparar el tipo de consumo de agua en los infantes.

Esto sugiere que, aunque el agua podría ser una fuente de este virus, en los primeros meses de vida pareciera más probable que su presencia constante provenga de la ingesta de alimentos o exposición ambiental, lo cual explicaría también la diferencia de abundancias entre semestres. PMMoV ha sido descrito como el virus más encontrado en comida a base de pimiento (57% de abundancia) y también en alimentos como chile, tomate y papa; en algunos casos se ha demostrado que los virus presentes en estos alimentos aún pueden ser infecciosos [144]. El PMMoV también ha sido descrito en heces de animales de granja como pollos, vacas y gansos [105], lo que sugiere que

podría ser un virus que está ampliamente presente en una comunidad rural, tanto en el agua, como alimentos, así como en múltiples animales.

Figura 28. Abundancia de PMMV en la orofaringe y muestras fecales, expresada en logaritmo base 10 (eje y). En la parte superior está marcado el identificador de cada infante y en la inferior la edad en meses de cada infante en el momento de la toma de muestra (eje x). VI= Viroma infantil.

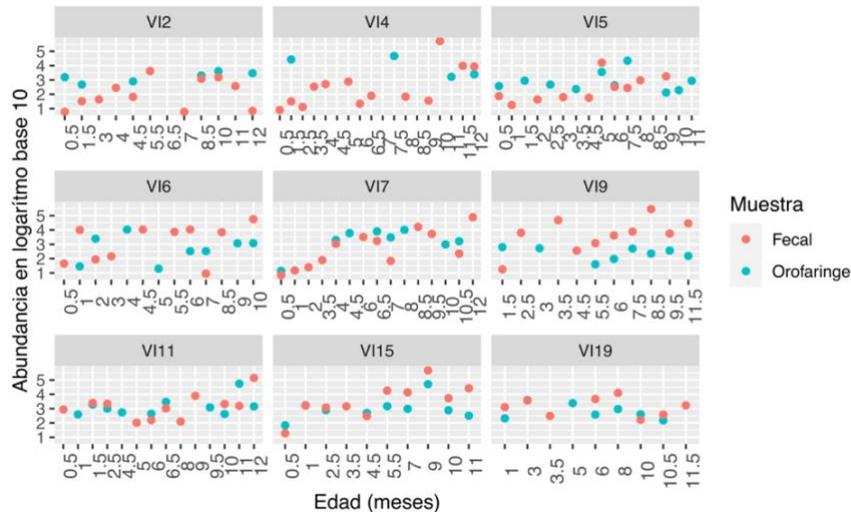


2) TSAMV tuvo un porcentaje de abundancia de lecturas de 42.8% en la orofaringe y 11.5% en las muestras fecales. En contraste, la positividad fue más alta en las muestras fecales (97.9% contra 77.8%). Al igual que con PMMV, encontramos que su abundancia aumenta después del primer semestre de vida (ANOVA con permutaciones para medidas repetidas, $p < 0.05$) y también encontramos menor frecuencia de virus en la orofaringe de los infantes 2 y 4 (Fig. 29). Esta especie no ha sido descrita tan ampliamente; de hecho fue descrita por primera vez en 2016 [143] y posteriormente por nosotros [142], por lo que no existe mucha información sobre sus dinámicas o la presencia en humanos.

Al extender los análisis a los 9 niños, encontramos una gran variabilidad de virus (Fig. S9) y en el análisis filogenético encontramos muestras en 3 niños con genomas de TSAMV que son similares, a pesar de tener edades en meses muy espaciados o de venir de nichos diferentes; en la niña 7 se encontró similitud de genoma entre una muestra fecal (7 meses) y una orofaríngea (10 meses); en la niña 19, igual, una orofaríngea (1 mes) y una fecal (8 meses) mientras que en la niña 15, las muestras orales del mes 0.5 y 11 son similares entre ellas, así como las muestras fecales de los meses 10.5 y 11 (Fig. S8). De forma interesante, la muestra oral del mes 11 no se

parece a la fecal del mismo mes (Fig. S7 y 8), mostrando la diversidad de virus que se encuentran en la comunidad, inclusive en el mismo período de tiempo.

Figura 29. Abundancia de TSAMV en la orofaringe y muestras fecales, expresada en logaritmo base 10 (eje y). En la parte superior está marcado el identificador de cada infante y en la inferior la edad en meses de cada infante en el momento de la toma de muestra (eje x). VI= Viroma infantil.



Es interesante destacar que en ambas especies la abundancia de lecturas fue mayor en la orofaringe que en las muestras fecales. No así cuando observamos la frecuencia de muestras por nicho; en el caso de TSAMV tenemos que la frecuencia en las muestras orofaríngeas es del 51% (Tabla 2), mientras que en las fecales fue del 42% (Tabla 3). En el caso de PMMoV, en ambos nichos la frecuencia es del 38%. El tener frecuencias similares nos hace pensar que la exposición de los niños a estos dos virus de plantas es constante y que la variación entre nichos está dada principalmente por la abundancia.

No podemos determinar el origen exacto de estos virus, pero estas diferencias de abundancia pudieran ser explicadas por la posibilidad de que 1) PMMoV y TSAMV circulen en el ambiente a través del polvo, aire, agua contaminada e inclusive la ropa de los cuidadores de cada infante y por lo tanto son aspirados y deglutidos de forma constante, posiblemente en cantidades bajas pero que ante la exposición continua logran estar en cantidades mayores en la orofaringe en comparación a las heces; 2) PMMoV y TSAMV son virus que se ingieren a través de alimentos, agua utilizada para hacer la comida y bebidas, tés medicinales y quizá pudieran aun estar presentes en la areola materna, y lo que vemos como una menor abundancia podría ser resultado de la pérdida de 'virus' durante el tránsito intestinal. Cabe mencionar que no se ha reportado la presencia de virus de plantas en piel o leche materna.

Al igual que PMMoV, de forma general los virus pertenecientes al género *Tobamovirus* han sido descritos en alimentos como chile, pimienta, papa, tomate, pepino, entre otros [144]. Asimismo, se ha intentado relacionar la presencia de este género de virus de plantas con patologías humanas, tales como cáncer de pulmón, aunque han sido estudios de posible asociación en los que no se obtuvieron resultados concluyentes [145,146]. Específicamente, PMMoV ha sido asociado a patologías intestinales; un estudio en población adulta observó una correlación entre la presencia de PMMoV y fiebre, dolor abdominal y prurito, aunque no se identificó una relación causal [147]. Asimismo, se han encontrado anticuerpos contra PMMoV en sangre de adultos expuestos a este virus [105].

A partir de nuestros resultados, podemos plantear que posiblemente los adultos con anticuerpos contra PMMoV han estado expuestos a este virus desde la infancia, y si ese fuera el caso, habría que preguntarse si se podrían encontrar anticuerpos expresados para otras especies de virus de plantas o si PMMoV tiene alguna particularidad. Es destacable la presencia de anticuerpos hacia un virus que, hasta ahora, no se ha demostrado que tenga la capacidad de replicarse en células humanas, pero que es identificado por nuestros mecanismos de defensa quizá por su alta abundancia y frecuencia.

Si trasladamos esta información a los niños de nuestra cohorte y recordamos que estos virus han estado presentes desde las primeras muestras, surgen nuevas interrogantes: 1) al ser ingeridos durante la etapa neonatal ¿los virus de plantas podrían, debido a su presencia prolongada, generar tolerancia inmune en una etapa de la vida durante la cual los mecanismos inmunes no están bien montados?; 2) en virus como RVA y CMV se ha establecido que el mecanismo de succión puede depositar virus en los ductos y glándulas mamarias generando una respuesta inmune, por lo que se han detectado anticuerpos en leche materna que no estaban presentes previo a la lactación [106,133,148,149], ¿es posible que los virus de plantas presentes en la saliva del bebé ingresen a la glándula y ductos mamarios, generen una respuesta inmune y esta se pueda observar en la LM?; 3) ¿la exposición constante a virus de plantas podría tener algún papel en la modulación y entrenamiento del sistema inmune?; 4) ¿los virus de plantas tendrán un papel a nivel ecológico en relación con otros virus y la posible modificación del microbioma podría afectar al hospedero?

Si bien estas preguntas no tienen respuesta hasta ahora, pareciera que entender el impacto al sistema inmune y a la dinámica ecológica que pudieran tener los virus de plantas se convierten cada vez más en preguntas que debieran tener respuesta urgente, como nos ha ido enseñando el COVID-19 y las secuelas a largo plazo, las cuales pareciera que tienen como origen una mala regulación del sistema inmune [150].

Discusión general

En este trabajo se analizó la dinámica y la diversidad del viroma eucarionte orofaríngeo y fecal de 9 niños aparentemente sanos durante su primer año de vida en una población semirural. A pesar de que no observamos en los dos nichos la presencia de un viroma eucarionte específico, esto es, especies en común entre todos los niños con patrones similares de distribución en tiempo o abundancia en cada nicho, logramos identificar que lo más importante, como una huella digital, es la individualidad de cada infante. Esta individualidad está desde luego influenciada, entre otros factores, por el medio ambiente que los rodea y sus interacciones con éste, así como por su alimentación. Aunque podrían existir factores en común en los infantes, tales como tipo de suelo o el origen del agua que consumen, el peso que tienen estas variables pareciera no ser tan relevante cuando se consideran como factores individuales. Cabe mencionar que en el diseño de estudio se consideró que los factores ambientales podrían ser similares ya que todos los niños habitan en la misma comunidad, y por lo mismo podrían tener un efecto directo y generalizado en el viroma de los niños. Sin embargo, es muy posible que el aire, agua, suelo y hasta la leña utilizada, por poner ejemplos, tengan variaciones desde su composición química hasta los microorganismos que se pueden encontrar en ellas. De forma interesante, en los últimos años (2020 y 2021) la enfermedad viral más reportada en la comunidad de Xoxocotla, de acuerdo al censo de salud[151], es dengue, seguida por influenza, dos especies virales que no identificamos en nuestros sets de datos.

Estudios previos de viroma han reportado mayor similitud de virus en niños que habitan en la misma vivienda en comparación con niños de la misma edad[36]. Resultados similares también se han descrito en el bacterioma, donde al analizarlo a nivel de especie o cepa, no se encuentran similitudes de poblaciones bacterianas excepto entre padre e hijos, no así entre niños de la misma edad[152]. Curiosamente, estos hallazgos, que se asemejan a los nuestros, fueron reportados en una cohorte con mayor número de personas analizadas y más muestras. Parte de esta individualidad también parte del nivel del taxón analizado, ya que al llegar a nivel de especie, hay tanta variabilidad que es complejo encontrar similitudes, como se podría hacer si el análisis se hiciera únicamente a niveles de género o familia [31,36,153].

Uno de nuestros propósitos fue tratar de definir si existe un viroma humano eucarionte *core*, es decir, si hay virus de humanos que sean predominantes en frecuencia y abundancia de forma generalizada en todos los niños. En nuestro estudio no se encontró un viroma eucarionte con las características mencionadas, compartido o común entre los infantes estudiados. Sin embargo, es importante recordar que se define como *viroma* a los genes de virus presentes en un nicho, así que lo que no tenemos, es una versión de *enterotipos virales*.

Si bien se observó una gran abundancia especies que se replican en humanos en cada nicho (norovirus en muestras fecales y citomegalovirus en las orofaríngeas), éstas no estuvieron

presentes de forma continua durante el primer año de vida, y tampoco se observaron en todos los niños. Mas bien se observó una dinámica entre virus dominantes y virus que se señalaron como esporádicos. Sin embargo, si consideramos la frecuencia de virus encontrados, es destacable que los 9 niños excretaron una gran variedad de virus entéricos de forma asintomática durante su primer año de vida, teniendo hasta 10 virus en una misma muestra. Además, algunos virus fueron detectados con una frecuencia considerable, como TMMV, anelloviridae sp, rotavirus A, parechovirus, y enterovirus en las muestras fecales (Fig. 13), así como papilomavirus en las orofaríngeas (Fig. 9).

Sospechamos que en el caso de los virus que se replican en humanos estas dinámicas se relacionan con la respuesta inmune del hospedero ante patógenos, y a su vez con las características fisiológicas de cada nicho, así mismo, la excreción prolongada de virus en sujetos asintomáticos ha sido descrita previamente [30,96,100,104,154]. Contrariamente, al analizar los virus de plantas, los cuales no se ha reportado que se repliquen en animales, observamos que particularmente *pepper mild mottle virus* y *tropical soda apple mosaic virus* se encuentran de forma constante y en todos los niños.

La forma en que influyen las especies dominantes en la composición del viroma fue interesante, pues en las gráficas de análisis constreñidos (db-RDA) pudimos observar que a pesar de tener estas dominancias la composición del viroma no cambia lo suficiente para observar una separación drástica de todas las muestras. Lo que se observó fue una separación de los centroides, es decir, la media de las disimilitudes entre las muestras de cada infante.

Resulta interesante notar que existen virus que infectan a humanos que podrían 'dominar' la ecología del viroma, como es el caso de CMV en la orofaringe y NoV en las muestras fecales, o RVA en ambos nichos. Así como se ha reportado en ciertas bacterias, como *Helicobacter pylori*, estos virus dominantes podrían tener mecanismos que modifiquen la ecología del nicho para permanecer dominantes. Utilizamos a *Helicobacter pylori*, como ejemplo, ya que esta bacteria tiene la capacidad de cambiar la composición de la microbiota estomacal gracias a que es capaz de regular el ambiente estomacal (microambiente) modificando el pH, citosinas y péptidos antimicrobianos, lo que en ciertas ocasiones termina afectando la respuesta inmune, mientras que en otras su comportamiento puede ser más comensal a pesar de estar presente de forma dominante[155]. En el caso de los virus no se han reportado mecanismos tan específicos en el microambiente pero pareciera lógico cuestionar si las especies que observamos de forma tan abundante y frecuente tienen algún efecto a nivel de microambiente o si sólo se están replicando de forma pasiva.

También podríamos preguntarnos ¿cuál es el efecto a nivel inmunológico de tener especies virales dominantes, como CMV o NoV por períodos prolongados de tiempo? La hipótesis de la higiene postula que a mayor exposición al ambiente, y por lo tanto a microorganismos, existe

un mejor entrenamiento del sistema inmunológico[156]. Sin embargo, esta hipótesis basa sus planteamientos en observaciones de la bacteriota y no se ha extendido a otros microorganismos y sus genes, lo cuales conforman el microbioma. Por lo tanto, es de interés entender cuál podría ser la respuesta inmune o celular ante la exposición inicial a especies virales que no generan sintomatología, pero que están de forma prolongada. Al respecto, se ha sugerido que la presencia de un aumento transitorio de alguna especie viral eucarionte sin la presencia de síntomas está relacionado con un sistema que lo facilita, es decir, un sistema inmune inmaduro[14]. Asimismo, se ha sugerido la presencia de algunas especies o familias virales pudiesen tener un papel benéfico en la salud[101].

Asimismo, pareciera también urgente entender dinámicas que han salido a la luz como amenazas a la salud durante la pandemia de COVID-19: reactivación de virus que estaban presentes de forma asintomática y que durante o posterior a la infección por SARS-CoV-2 se volvieron patogénicos [157,158]. Esto nos lleva a pensar que la composición del viroma puede cambiar de forma brusca ante una infección, afectando la respuesta inflamatoria así como la dinámica de virus que previamente se encontraban de forma dominante pero asintomática. En este sentido podríamos preguntarnos si pudiese haber una reactivación de CMV después de una infección que modifique la respuesta inflamatoria del cuerpo, la cual rompería la homeostasis y diera como resultados síntomas, ya sea de la infección viral o, inclusive, de una mala o ineficiente regulación inmunitaria o epitelial que afecte la ecología microbiana, como se ha sugerido que sucede después de la infección por SARS-CoV-2 [159].

Por otro lado, encontramos especies que se comparten entre los dos nichos estudiados. Uno de los casos más interesantes es quizá el de *rotavirus A*, el cual es conocido por ser un virus entérico que en nuestra cohorte se encontró también presente en muestras orofaríngeas. No es descabellado pensar que quizá podría replicarse en las glándulas salivales en humanos, como se demostró recientemente en modelos murinos infectados con *rotavirus A* murino (EDIM). En ese mismo estudio también demostraron, a través de cultivos celulares y esféroides de células de glándulas salivales murinas, que norovirus humano puede replicarse en células de las glándulas salivales [133].

La detección temprana de secuencias virales durante el desarrollo del niño, es decir, en las primeras muestras colectadas, inclusive antes de la ingesta de alimentos diferentes a la leche materna, nos sugiere que la interacción de los niños con virus eucariontes tiene diversos orígenes, los cuales sería interesante explorar para entender cómo se establece y mantiene el viroma.

Consideramos que nuestros hallazgos avanzan el conocimiento de la diversidad y dinámica del viroma infantil y sientan la base para preguntas importantes como las planteadas a lo largo de esta tesis, además de proveer una descripción detallada de las dinámicas y diversidad del

viroma. Finalmente, sugerimos que en este estudio cada niño presentó un viroma individual y único en cada nicho, con algunas especies específicas y otras compartidas entre ambos nichos, demostrando que la excreción asintomática de diversos virus es bastante común en infantes. Debe estudiarse si esta excreción y contacto constante con una gran variedad de virus humanos y de plantas durante la etapa temprana de vida, un período donde el desarrollo fisiológico e inmunológico es crucial, tiene algún efecto en estos procesos.

Limitaciones

Dentro de las limitantes de este estudio está el tratamiento de las muestras fecales con cloroformo al 10% para remover contaminación durante el aislamiento de ácidos nucleicos, ya que los virus que están envueltos en capas lipídicas se pueden ver afectados por este tratamiento y, aunque no son muy comunes en el intestino, podríamos haber perdido virus encontrados en la orofaringe que quizá fuesen ingeridos y por lo tanto, estar presentes en la materia fecal, como es el caso de citomegalovirus.

Asimismo, la obtención de cantidades equivalentes de ácidos nucleicos a partir de la diferencia de material orgánico y cantidad de muestra de cada nicho dificultan la profundidad de secuenciación y, por lo tanto, información que se pueda obtener de estas muestras. Otra limitación es que con esta metodología solo estamos viendo genes virales. No podemos definir si hay infección activa.

El número de niños incluidos en el estudio, la dificultad de mantener una cohorte a largo plazo y la colecta de muestras pareadas de acuerdo a un plan, son limitaciones propias no sólo de este estudio, sino de cualquier estudio de seguimiento de cohorte.

El diseño de este estudio no fue enfocado a detectar la fuente de los virus que podrían encontrarse en los nichos estudiados, sin embargo, se está iniciando un seguimiento dónde el afán es contestar este tipo de preguntas, con muestreos más exhaustivos que incluyen la leche materna, piel de areola del pezón, e inclusive agua y suelo. Este estudio podría darnos mayor información sobre las fuentes de contagio del viroma, pero también acerca de cómo este podría ir cambiando y así quizá responder si una vez establecido un virus de forma dominante, esta dominancia es dependiente o independiente del medio ambiente.

En general estas dificultades previnieron tener asociaciones fuertes y conclusiones claras acerca de cómo se establecen las comunidades virales.

Conclusiones

- Tanto el viroma de la orofaringe como el de la materia fecal tienen un patrón de que los hace únicos en cada hospedero.
- A diferencia de trabajos previos donde la descripción del viroma infantil se limitaba a nivel de familia o de forma poco descriptiva a nivel de especie, aquí se describe la presencia de diversas especies virales de forma frecuente y abundante.
- La formación y dinámica del viroma se ve influenciada por el medio ambiente, pero el peso de este factor es menor a la contribución que hace la individualidad de cada infante.
- El viroma de la orofaringe, en comparación con el de heces, presenta un número mayor de especies virales (tanto de humanos como de plantas) con un patrón de frecuencia esporádico, posiblemente por estar expuesta de forma más constante al medio ambiente.
- Después del primer trimestre de vida se observa un aumento de abundancia de lecturas virales en el viroma fecal, no así en el orofaríngeo. Creemos que esto está relacionado con la alimentación.
- El origen de las especies compartidas entre ambos nichos requiere de más estudio; éste pudiera estar ligado a la deglución, como es el caso de los virus de plantas, o bien, en casos específicos como RVA, la posibilidad de que se replique en ambos nichos.
- Los virus de plantas están presentes de manera común durante el primer año de vida tanto en la orofaringe como en el intestino. El posible impacto de esta presencia en el desarrollo del sistema inmunitario de los niños queda por explorarse.

Perspectivas

- Estudiar el posible papel de los virus eucariontes en el entrenamiento del sistema inmune durante los primeros años de vida de los infantes; esto podría analizarse utilizando transcriptómica para asociar la respuesta de células inmunes o mediadores de señalización celular con especies virales en modelos animales, o bien organoides. Esta aproximación se ha reportado previamente en tejidos humanos [160].
- Identificar cuál es la respuesta celular del hospedero (ya sea a través de la expresión genética de factores de inmunidad o transductores de señales celulares asociados con la inflamación) ante la presencia asintomática de virus dominantes, y cuál podría ser la diferencia ante la presencia sintomática de los virus. Asimismo, estudiar el efecto de esta respuesta sobre la barrera celular y por lo mismo, modificar los microorganismos presentes y su ecología.
- Hacer un seguimiento de niños recién vacunados con rotavirus con el afán de determinar el tiempo de excreción de este virus en materia fecal, y explorar su posible concurrente detección en la orofaringe y en la saliva. Asimismo, tratar de definir si estos virus vacunales excretados tienen el potencial de replicarse en otros individuos, lo que nos ayudaría a entender mejor si pudiese existir una inmunidad pasiva transferible entre personas. Podríamos abordar las preguntas con una metodología similar a desarrollada Gosh y cols., quienes mostraron transmisión de RVA en glándulas salivales de modelos murinos, pero también en monocapas de células salivales y en esferoides de células salivales [133]. Otra opción es seguir a los parientes cercanos y buscar rotavirus vacunal y posible presencia de anticuerpos.
- Estudiar si los virus de plantas tienen la capacidad de replicarse en humanos, lo que podría evaluarse utilizando enteroides humanos. Asimismo, se podría estudiar la respuesta celular (buscando expresión de transductores de señales celulares asociados con inflamación, factores de crecimiento o respuesta inmune) ante los virus de plantas, a través de la secuenciación de RNA mensajero, o secuenciación de célula única.
- Identificar si los virus de plantas encontrados en las muestras conservan su capacidad infecciosa. Esto podría lograrse infectando las plantas a través de las lesiones del epitelio vegetal y supervisando si hay infección posterior.
- Actualmente se da seguimiento a una nueva cohorte cuya metodología está enfocada en contestar cual es el origen de los diversos virus observados de forma abundante y/o frecuente en los infantes, así como mejorar el método de obtención de muestras. A partir de estas muestras, también podríamos intentar describir si existe una respuesta inmune dirigida hacia virus de interés desde los primeros días de vida, incluyendo los de plantas, midiendo anticuerpos en la saliva de los infantes.
- Extender el estudio a niños de distintas zonas geográficas.

Referencias

1. Marchesi, J. R. & Ravel, J. The vocabulary of microbiome research: a proposal. *Microbiome* **3**, 1–3 (2015).
2. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
3. Kim, B. R. *et al.* Deciphering diversity indices for a better understanding of microbial communities. *J. Microbiol. Biotechnol.* **27**, 2089–2093 (2017).
4. Roger Bray, J. & Curtis T., J. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
5. Perret, C., Pérez, C. & Poblete, M. Manual de Pediatría 1ed. *Esc. Med. Chile* p.17-19. (2018).
6. Geng-Hao, B., Sheng-Chieh, L., Yi-Hsiang, H. & Shih-Yen, C. The Human Virome: Viral Metagenomics, Relations with Human Disease, and Therapeutic Applications. *Viruses* **14**, 1–29 (2022).
7. Turrioni, F. *et al.* The infant gut microbiome as a microbial organ influencing host well-being. *Ital. J. Pediatr.* **46**, 1–13 (2020).
8. Christian Milani, Sabrina Duranti, Francesca Bottacini, B. *et al.* The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiol. Mol. Biol. Rev.* **81**, 1–67 (2017).
9. Yatsunenkov, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
10. Moossavi, S. *et al.* Composition and Variation of the Human Milk Microbiota Are Influenced by Maternal and Early-Life Factors. *Cell Host Microbe* **25**, 324-335.e4 (2019).
11. Ximenez, C. & Torres, J. Development of Microbiota in Infants and its Role in Maturation of Gut Mucosa and Immune System. *Arch. Med. Res.* **48**, 666–680 (2022).
12. NIH Human Microbiome Portfolio Analysis, T. NIH Human Microbiome Portfolio Analysis Team*. *Microbiome* **31** (2019).
13. Smith, S. E. *et al.* Emerging technologies in the study of the virome. *Curr. Opin. Virol.* **54**, 101231 (2022).
14. Fulci, V., Stronati, L., Cucchiara, S., Laudadio, I. & Carissimi, C. Emerging roles of gut virome in pediatric diseases. *Int. J. Mol. Sci.* **22**, 1–12 (2021).
15. Battista, J. R., Moya, A., Pride, D. T., Ogilvie, L. A. & Jones, B. V. The human gut virome: a multifaceted majority. (2015) doi:10.3389/fmicb.2015.00918.
16. Ly, M. *et al.* Transmission of viruses via our microbiomes. *Microbiome* **4**, 64 (2016).
17. Abeles, S. R., Ly, M., Santiago-Rodriguez, T. M. & Pride, D. T. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLoS One* **10**, e0134941 (2015).
18. Koonin, E. V., Dolja, V. V. & Krupovic, M. The healthy human virome: from virus–host symbiosis to disease. *Curr. Opin. Virol.* **47**, 86–94 (2021).
19. Freer, G. *et al.* The virome and its major component, Anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. *Front. Microbiol.* **9**, 1–7 (2018).
20. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
21. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
22. McCann, A. *et al.* Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* **2018**, 1–13 (2018).
23. Duranti, S. *et al.* Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. doi:10.1186/s40168-017-0282-6.
24. Maqsood, R. *et al.* Discordant transmission of bacteria and viruses from mothers to babies at birth. *Microbiome* **7**, 1–13 (2019).
25. Indrio, F. *et al.* Development of the Gastrointestinal Tract in Newborns as a Challenge for an Appropriate Nutrition: A Narrative Review. *Nutrients* **14**, 1–15 (2022).

26. Schmidt, T. S. B. *et al.* Extensive transmission of microbes along the gastrointestinal tract. *Elife* **8**, (2019).
27. Iwachi, M. *et al.* Relationship between oral and gut microbiota in elderly people. *Immunity, Inflamm. Dis.* **7**, 229–236 (2019).
28. Kaan, A. M., Kahharova, D. & Zaura, E. Acquisition and establishment of the oral microbiota. *Periodontol. 2000* **86**, 123–141 (2021).
29. Breitbart, M. *et al.* Viral diversity and dynamics in an infant gut. *Res. Microbiol.* **159**, 367–373 (2008).
30. Kapusinszky, B., Minor, P. & Delwart, E. Nearly constant shedding of diverse enteric viruses by two healthy infants. *J. Clin. Microbiol.* **50**, 3427–3434 (2012).
31. Lim, E. S. *et al.* Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
32. Kramná, L. *et al.* Gut virome sequencing in children with early islet autoimmunity. *Diabetes Care* **38**, 930–933 (2015).
33. Kim, K. W. *et al.* Higher frequency of vertebrate-infecting viruses in the gut of infants born to mothers with type 1 diabetes. *Pediatr. Diabetes* **21**, 271–279 (2020).
34. Lindfors, K. *et al.* Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: The TEDDY study. *Gut* **69**, 1416–1422 (2020).
35. Liang, G. *et al.* The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470–474 (2020).
36. Reyes, A. *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci.* **112**, 11941–11946 (2015).
37. Siqueira, J. D. *et al.* Complex virome in feces from Amerindian children in isolated Amazonian villages. *Nat. Commun.* **9**, 1–11 (2018).
38. Province, N. W. *et al.* Metagenomic Analysis of the Enteric RNA Virome of. *Viruses* **12**, 1–14 (2020).
39. Taboada, B. *et al.* The gut virome of healthy children during the first year of life is diverse and dynamic. *PLoS One* **16**, 1–18 (2021).
40. Sorber, K. *et al.* The long march: A sample preparation technique that enhances contig length and coverage by high-throughput short-read sequencing. *PLoS One* **3**, (2008).
41. Martínez, M. A. *et al.* DNA microarray for detection of gastrointestinal viruses. *J. Clin. Microbiol.* **53**, 136–145 (2015).
42. Andrews, S. FASTQC. A quality control tool for high throughput sequence data. at (2010).
43. Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
44. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
45. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
48. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
49. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
50. Bushnell, B. BBMap : A Fast , Accurate , Splice-Aware Aligner. 3–5 (2014).
51. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
52. Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct* **7**, 12 (2012).
53. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site

- identification. *BMC Bioinformatics* **11**, 119 (2010).
54. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
 55. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
 56. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA -- A Practical Iterative de Bruijn Graph De Novo Assembler. in *Research in Computational Molecular Biology* (ed. Berger, B.) 426–440 (Springer Berlin Heidelberg, 2010).
 57. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 58. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *bioRxiv* (2018) doi:10.1101/383513.
 59. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
 60. Stecher, G., Tamura, K. & Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol. Biol. Evol.* **37**, 1237–1239 (2020).
 61. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
 62. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
 63. Tatusov, R. L. *et al.* Human Calicivirus Typing tool: A web-based tool for genotyping human norovirus and sapovirus sequences. *J. Clin. Virol.* **134**, 104718 (2021).
 64. Pickett, B. E. *et al.* ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **40**, D593-8 (2012).
 65. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
 66. Wilke, C. O. cowplot: Streamlined Plot Theme and Plot Annotations for ‘ggplot2’. (2018).
 67. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* at (2020).
 68. Oksanen, J. *et al.* The vegan package. *Community Ecol. Packag.* **10**, 631–637 (2007).
 69. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
 70. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).
 71. Wheeler, B., T. M. ImPerm: permutation tests for linear models. at (2016).
 72. Lim, E. S., Wang, D. & Holtz, L. R. The Bacterial Microbiome and Virome Milestones of Infant Development. *Trends Microbiol.* **24**, 801–810 (2016).
 73. Taboada, B., Rojas-, L., Torres, J., Ximenez, C. & Arias, C. F. The gut virome of healthy children during the first. (2020).
 74. Dehesdin, D. & Choussy, O. Anatomía de la faringe. *EMC - Otorrinolaringol.* **29**, 1–12 (2000).
 75. Netter, F. H. Orofaringe, laringofaringe y laringe. in *Atlas de Anatomía Humana Netter* 77–79 (Elsevier, 2019).
 76. Baker, J. L., Bor, B., Agnello, M., Shi, W. & He, X. Ecology of the Oral Microbiome: Beyond Bacteria. *Trends Microbiol.* **25**, 362–374 (2017).
 77. Altan, E. *et al.* Effect of Geographic Isolation on the Nasal Virome of Indigenous Children. *J. Virol.* **93**, (2019).
 78. Tay, C. J. X. *et al.* Role of Upper Respiratory Microbiota and Virome in Childhood Rhinitis and Wheeze: Collegium Internationale Allergologicum Update 2021. *Int. Arch. Allergy Immunol.* **182**, 265–276 (2021).
 79. Lorio, A. *et al.* Cross-correlation of virome – bacteriome – host – metabolome to study respiratory health. *Trends Microbiol.* 1–13 (2021) doi:10.1016/j.tim.2021.04.011.
 80. Ho, S. X., Min, N., Wong, E. P. Y., Chong, C. Y. & Chu, J. J. H. Characterization of oral

- virome and microbiome revealed distinctive microbiome disruptions in paediatric patients with hand, foot and mouth disease. *npj Biofilms Microbiomes* **7**, (2021).
81. Martínez, A., Kuraji, R. & Kapila, Y. L. The human oral virome: Shedding light on the dark matter. *Periodontol. 2000* **87**, 282–298 (2021).
 82. Erin A. Osterholm, M. R. S. Impact of breast milk-acquired cytomegalovirus infection in premature infants: Pathogenesis, prevention, and clinical consequences? *Rev Med Virol* **30**, 1–11 (2020).
 83. Schleiss, M. R. Acquisition of human cytomegalovirus infection in infants via breast milk: Natural immunization or cause for concern? *Rev. Med. Virol.* **16**, 73–82 (2006).
 84. Fowler, K. *et al.* A systematic literature review of the global seroprevalence of cytomegalovirus: possible implications for treatment, screening, and vaccine development. *BMC Public Health* **22**, 1659 (2022).
 85. Meltem Dinleyici, Vicente Pérez-Brocal, Sertac Arslanoglu, Ozge Aydemir, Sibel Sevuk Ozumut, Neslihan Tekin, Yvan Vandenplas, Andrés Moya, E. C. D. Human Milk Virome Analysis: Changing Pattern Regarding Mode of Delivery, Birth Weight, and Lactational Stage. *Nutrients* **13**, (2021).
 86. Bardanzellu, F., Fanos, V. & Reali, A. Human Breast Milk-acquired Cytomegalovirus Infection: Certainties, Doubts and Perspectives. *Curr. Pediatr. Rev.* **15**, 30–41 (2018).
 87. Geris, J. M. *et al.* High prevalence of asymptomatic CMV shedding in healthy children attending the minnesota state fair. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **148**, 105102 (2022).
 88. Prussin, A. J. *et al.* Seasonal dynamics of DNA and RNA viral bioaerosol communities in a daycare setting. *15th Conf. Int. Soc. Indoor Air Qual. Clim. INDOOR AIR 2018* 1–14 (2018).
 89. Loeffelholz, M. J. *et al.* Duration of rhinovirus shedding in the upper respiratory tract in the first year of life. *Pediatrics* **134**, 1144–1150 (2014).
 90. Vasiljević, N. *et al.* Characterization of two novel cutaneous human papillomaviruses, HPV93 and HPV96. *J. Gen. Virol.* **88**, 1479–1483 (2007).
 91. Li, J. *et al.* Nine complete genome sequences of cutaneous human papillomavirus genotypes isolated from healthy skin of individuals living in rural He Nan province, China. *J. Virol.* **86**, 11936 (2012).
 92. Koskimaa, H. M. *et al.* Human papillomavirus genotypes present in the oral mucosa of newborns and their concordance with maternal cervical human papillomavirus genotypes. *J. Pediatr.* **160**, 837–843 (2012).
 93. Omatola, C. A. & Olaniran, A. O. Epidemiological significance of the occurrence and persistence of rotaviruses in water and sewage: a critical review and proposal for routine microbiological monitoring. *Environ. Sci. Process. Impacts* **24**, 380–399 (2022).
 94. la Rosa, G. *et al.* Mucosal and Cutaneous Human Papillomaviruses Detected in Raw Sewages. *PLoS One* **8**, 75–77 (2013).
 95. Beller, L. *et al.* The virota and its transkingdom interactions in the healthy infant gut. *PNAS* **119**, 1–12 (2022).
 96. Yinda, C. K. *et al.* Gut Virome Analysis of Cameroonians Reveals High Diversity of Enteric Viruses, Including Potential Interspecies Transmitted Viruses. *mSphere* **4**, (2019).
 97. Altan, E. *et al.* Enteric virome of Ethiopian children participating in a clean water intervention trial. *PLoS One* **13**, 1–20 (2018).
 98. Sridhar, A., Karelehto, E., Brouwer, L., Pajkrt, D. & Wolthers, K. C. Parechovirus a pathogenesis and the enigma of genotype A-3. *Viruses* **11**, 1–18 (2019).
 99. Simon, A. K., Hollander, G. A. & McMichael, A. Evolution of the immune system in humans from infancy to old age. *Proc. R. Soc. B Biol. Sci.* **282**, (2015).
 100. Rodríguez-Díaz, J. *et al.* Presence of human enteric viruses in the stools of healthy malawian 6-month-old infants. *J. Pediatr. Gastroenterol. Nutr.* **58**, 502–504 (2014).
 101. Kaczorowska, J. *et al.* Early-Life Colonization by Anelloviruses in Infants. *Viruses* **14**, 2–9 (2022).
 102. Gregory, A. C. *et al.* Resource The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut Resource The Gut Virome Database

- Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe* 1–17 (2020) doi:10.1016/j.chom.2020.08.003.
103. Carding, S. R., Davis, N. & Hoyles, L. Review article: the human intestinal virome in health and disease. *Aliment. Pharmacol. Ther.* **46**, 800–815 (2017).
 104. Brazier, L. *et al.* Evolution in fecal bacterial/viral composition in infants of two central African countries (Gabon and Republic of the Congo) during their first month of life. *PLoS One* **12**, 1–19 (2017).
 105. Balique, F., Lecoq, H., Raoult, D. & Colson, P. Can plant viruses cross the kingdom border and be pathogenic to humans? *Viruses* **7**, 2074–98 (2015).
 106. Maschmann, J. *et al.* Characterization of human breast milk leukocytes and their potential role in cytomegalovirus transmission to newborns. *Neonatology* **107**, 213–219 (2015).
 107. Doare, K. Le, Holder, B., Bassett, A. & Pannaraj, P. S. Mother’s Milk: A purposeful contribution to the development of the infant microbiota and immunity. *Front. Immunol.* **9**, (2018).
 108. Platts-Mills, J. A. *et al.* Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *Lancet Glob. Heal.* **6**, e1309–e1318 (2018).
 109. O’Ryan, M., Riera-Montes, M. & Lopman, B. Norovirus in Latin America. *Pediatr. Infect. Dis. J.* **36**, 127–134 (2017).
 110. Farkas, T. *et al.* Prevalence and genetic diversity of human caliciviruses (HuCVs) in Mexican children. *J. Med. Virol.* **62**, 217–223 (2000).
 111. Chhabra, P. *et al.* Updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* **100**, 1393–1406 (2019).
 112. Robilotti, E., Deresinski, S. & Pinsky, B. A. Norovirus. *Clin. Microbiol. Rev.* **28**, 134–164 (2015).
 113. Lyman, W. H. *et al.* Prospective Study of Etiologic Agents of Acute Gastroenteritis Outbreaks in Child Care Centers. *J. Pediatr.* **154**, 253–257 (2009).
 114. Mann, P., Pietsch, C. & Liebert, U. G. Genetic diversity of sapoviruses among inpatients in germany, 2008–2018. *Viruses* **11**, 1–14 (2019).
 115. Oka, T. *et al.* Human sapovirus classification based on complete capsid nucleotide sequences. *Arch. Virol.* **157**, 349–352 (2012).
 116. Saito, M. *et al.* Multiple norovirus infections in a birth cohort in a peruvian periurban community. *Clin. Infect. Dis.* **58**, 483–491 (2014).
 117. Rogawski, E. T. *et al.* Use of quantitative molecular diagnostic methods to investigate the effect of enteropathogen infections on linear growth in children in low-resource settings: longitudinal analysis of results from the MAL-ED cohort study. *Lancet Glob. Heal.* **6**, e1319–e1328 (2018).
 118. Rouhani, S. *et al.* Norovirus infection and acquired immunity in 8 countries: Results from the MAL-ED study. *Clin. Infect. Dis.* **62**, 1210–1217 (2016).
 119. Qi, R. *et al.* Global Prevalence of Asymptomatic Norovirus Infection: A Meta-analysis. *EClinicalMedicine* **2–3**, 50–58 (2018).
 120. Sánchez, G. J. *et al.* Epidemiology of Sapovirus Infections in a Birth Cohort in Peru. *Clin. Infect. Dis.* **66**, 1858–1863 (2018).
 121. Liu, X. *et al.* Etiological role and repeated infections of sapovirus among children aged less than 2 years in a cohort study in a peri-urban community of Peru. *J. Clin. Microbiol.* **54**, 1598–1604 (2016).
 122. Platts-Mills, J. A. *et al.* Pathogen-specific burdens of community diarrhoea in developing countries: A multisite birth cohort study (MAL-ED). *Lancet Glob. Heal.* **3**, e564–e575 (2015).
 123. Zaczek-Moczydlowska, M. A., Beizaei, A., Dillon, M. & Campbell, K. Current state-of-the-art diagnostics for Norovirus detection: Model approaches for point-of-care analysis. *Trends Food Sci. Technol.* **114**, 684–695 (2021).
 124. Liu, L. & Moore, M. D. A survey of analytical techniques for noroviruses. *Foods* **9**, (2020).

125. Nelson, M. I. *et al.* Genetic diversity of noroviruses circulating in a pediatric cohort in Bangladesh. *J. Infect. Dis.* **218**, 1937–1942 (2018).
126. Maunula, L., Miettinen, I. T. & Von Bonsdorff, C. H. Norovirus outbreaks from drinking water. *Emerg. Infect. Dis.* **11**, 1716–1721 (2005).
127. Maunula, L. & Von Bonsdorff, C. H. Norovirus genotypes causing gastroenteritis outbreaks in Finland 1998-2002. *J. Clin. Virol.* **34**, 186–194 (2005).
128. Cantelli, C. P. *et al.* Norovirus infection and HBGA host genetic susceptibility in a birth community-cohort, Rio de Janeiro, Brazil. *Infect. Genet. Evol.* **82**, 104280 (2020).
129. O’Ryan, M. L. *et al.* Symptomatic and asymptomatic rotavirus and norovirus infections during infancy in a Chilean birth cohort. *Pediatr. Infect. Dis. J.* **28**, 879–884 (2009).
130. Parra, G. I. *et al.* Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity. *PLoS Pathog.* **13**, 1–22 (2017).
131. Zárate, S. *et al.* The Alpha Variant (B.1.1.7) of SARS-CoV-2 Failed to Become Dominant in Mexico. *Microbiol. Spectr.* **10**, (2022).
132. Taboada, B. *et al.* Dominance of Three Sublineages of the SARS-CoV-2 Delta Variant in Mexico. *Viruses* **14**, 1–15 (2022).
133. Ghosh, S. *et al.* Enteric viruses replicate in salivary glands and infect through saliva. *Nat.* **2022** 1–6 (2022) doi:10.1038/s41586-022-04895-8.
134. Simsek, C. *et al.* High prevalence of coinfecting enteropathogens in suspected rotavirus vaccine breakthrough cases. *J. Clin. Microbiol.* **59**, (2021).
135. Matthijssens, J. *et al.* Molecular and biological characterization of the 5 human-bovine rotavirus (WC3)-based reassortant strains of the pentavalent rotavirus vaccine, RotaTeq®. *Virology* **403**, 111–127 (2010).
136. Anderson, E. J. Rotavirus vaccines: viral shedding and risk of transmission. *Lancet Infect. Dis.* **8**, 642–649 (2008).
137. Dennehy, P. H. Transmission of rotavirus and other enteric pathogens in the home. *Pediatr. Infect. Dis. J.* **19**, 103–105 (2000).
138. Zheng, B. J. *et al.* Rotavirus infection of the oropharynx and respiratory tract in young children. *J. Med. Virol.* **34**, 29–37 (1991).
139. Payne, D. C. *et al.* Sibling transmission of vaccine-derived rotavirus (RotaTeq) associated with rotavirus gastroenteritis. *Pediatrics* **125**, (2010).
140. Higashimoto, Y. *et al.* Monitoring Shedding of Five Genotypes of RotaTeq Vaccine Viruses by Genotype-Specific Real-Time Reverse Transcription-PCR Assays. *J. Clin. Microbiol.* **56**, (2018).
141. Cirilo Aguilar, B. G., Cantú Martínez, P. C., Verde Star, M. J. & Mata Cárdenas, B. USO DE LA HERBOLARIA EN EL CUIDADO DE LA SALUD. *RESPYN Rev. Salud Pública y Nutr.* **12**, (2011).
142. Aguado-García, Y. *et al.* Tobamoviruses can be frequently present in the oropharynx and gut of infants during their first year of life. *Sci. Rep.* **10**, 1–11 (2020).
143. Filmer, K., Adkins, S., Pongam, P. & D’Elia, T. The complete nucleotide sequence and genomic characterization of tropical soda apple mosaic virus. *Arch. Virol.* **161**, 2317–2320 (2016).
144. Dombrovsky, A. & Smith, E. Seed transmission of Tobamoviruses: Aspects of global disease distribution. in *Advances in seed biology* 233–260 (2017).
145. Bothwell, P. W. Lung cancer and tobacco mosaic virus. *Lancet* **1**, 657–658 (1960).
146. Pirié, N. W. Lung cancer and tobacco mosaic virus. *Lancet* **1**, 707 (1960).
147. Colson, P. *et al.* Pepper mild mottle virus, a plant virus associated with specific immune responses, Fever, abdominal pains, and pruritus in humans. *PLoS One* **5**, e10041 (2010).
148. Moylan, D. C. *et al.* Breast milk human cytomegalovirus (CMV) viral load and the establishment of breast milk CMV-pp65-Specific CD8 T Cells in Human CMV infected mothers. *J. Infect. Dis.* **216**, 1176–1179 (2017).
149. Kaur, A. *et al.* Natural history of postnatal rhesus cytomegalovirus shedding by dams and acquisition by infant rhesus monkeys. *PLoS One* **13**, 1–14 (2018).
150. Couzin-frankel, B. J. Scientists strive to unravel what is driving disabling symptoms. **376**, 1261–1266 (2022).

151. Morelos, G. del estado de. *Sintesis estadística municipal 2021*.
152. Valles-Colomer, M. *et al.* The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* **614**, (2023).
153. Granados, A.C., Ley, C., Walters, W.A., Federman, S., Santos, Y., Haggerty, T., Sotomayor-Gonzalez, A., Servellita, V., Ley, R.E., Parsonnet, J. and Chiu, C. Y. Longitudinal comparison of the developing gut virome in infants and their mothers. *Cell Host Microbe* **31**, 187-198.e3 (2022).
154. Okitsu, S. *et al.* Molecular detection of enteric viruses in the stool samples of children without diarrhea in Bangladesh. *Infect. Genet. Evol.* **77**, 104055 (2020).
155. Sun, Y. Y. & Jing, D. D. Relationship between Helicobacter pylori infection and gastrointestinal microecology. *World Chinese J. Dig.* **28**, 1261–1265 (2020).
156. Bach, J. F. Revisiting the Hygiene Hypothesis in the Context of Autoimmunity. *Front. Immunol.* **11**, 1–9 (2021).
157. Gold, J. E., Okyay, R. A., Licht, W. E. & Hurley, D. J. Investigation of Long COVID Prevalence and Its Relationship to Epstein-Barr Virus Reactivation. *Pathogens* vol. 10 at <https://doi.org/10.3390/pathogens10060763> (2021).
158. Simonnet, A. *et al.* High incidence of Epstein–Barr virus, cytomegalovirus, and human-herpes virus-6 reactivations in critically ill patients with COVID-19. *Infect. Dis. Now* **51**, 296–299 (2021).
159. Cardinale, V. *et al.* Intestinal permeability changes with bacterial translocation as key events modulating systemic host immune response to SARS-CoV-2: A working hypothesis. *Dig. liver Dis. Off. J. Ital. Soc. Gastroenterol. Ital. Assoc. Study Liver* **52**, 1383–1389 (2020).
160. Kumata, R., Ito, J., Takahashi, K., Suzuki, T. & Sato, K. A tissue level atlas of the healthy human virome. *BMC Biol.* **18**, 1–15 (2020).

Material suplementario

Tablas

Tabla S1: Lecturas secuenciadas y lecturas válidas, es decir, aquellas que quedaron después de eliminar lecturas de mala calidad, duplicadas y lecturas asignadas al hospedero y a genes ribosomales. La tabla muestra cada niño y la edad en meses de la toma de muestra. Símbolos con '-' representan ausencia de muestra en ese mes.

Infante	ID de muestra	Muestras orofaríngeas						Muestras fecales						
		Secuencias crudas	Secuencias de baja calidad	Secuencias de duplicadas	Secuencias asociadas a genes humanos	Secuencias asociadas a genes ribosomales	Secuencias válidas	Secuencias crudas	Secuencias de baja calidad	Secuencias de duplicadas	Secuencias asociadas a genes humanos	Secuencias asociadas a genes ribosomales	Secuencias válidas	
2	R2_0.5_MAR15	21,976,120	21,298,502	4,133,528	520,852	3,311,224	224,241	G2_0.5_MAR15	22,389,670	21,703,448	5,771,907	21,707	2,707,147	5,478,240
	R2_1.5_ABR15	23,151,008	22,165,824	5,280,153	482,473	3,844,390	647,786	G2_1.5_ABR15	35,032,158	33,850,900	6,353,194	19,915	6,060	6,324,663
	R2_3_MAY15	24,679,278	24,022,112	6,032,428	2,877,715	4,202,380	590,508	G2_3_MAY15	25,582,524	24,700,120	4,605,587	333,322	2,669,791	1,391,286
	R2_4_JUN15	25,073,416	23,812,304	8,003,276	1,019,336	4,062,235	2,092,242	G2_4_JUN15	44,855,392	42,913,312	10,747,167	2,178	17,532	10,728,652
	R2_4.5_JUL15	28,053,232	25,903,474	8,564,682	1,629,665	3,418,143	2,314,691	G2_4.5_JUL15	27,131,510	26,222,048	6,768,223	5,096	55,449	6,706,697
	R2_5.5_AGO15	22,250,236	21,328,780	6,023,911	445,314	2,725,057	2,432,677	G2_5.5_AGO15	18,034,312	17,408,942	6,206,671	61,949	699,278	5,443,606
	R2_6.5_SEPT15	14,133,516	13,149,596	5,277,860	1,120,753	1,816,077	1,801,925	G2_7_OCT15	23,130,918	22,418,532	6,124,479	5,013	64,509	6,053,107
	R2_8.5_NOV15	21,151,438	20,237,932	6,337,356	485,371	3,754,091	1,964,786	G2_8.5_NOV15	20,892,828	20,223,618	8,023,938	134,079	498,259	7,387,994
	R2_10_ENE16	25,104,964	24,392,750	5,898,754	284,512	4,858,388	653,847	G2_10_ENE16	20,466,646	19,498,240	4,965,742	15,450	22,359	5,201,316
	R2_11_FEB16	23,201,082	22,334,260	4,675,508	405,957	3,066,736	834,875	G2_11_FEB16	18,568,934	17,929,674	2,989,547	3,449	5,107	2,979,368
	R2_12_MAR16	23,025,382	22,819,594	7,204,416	543,870	2,535,019	3,502,498	G2_12_MAR16	29,819,606	29,060,642	6,363,543	18,090	62,458	6,279,691
	R4_0.5_ABR15	22,623,324	21,706,966	5,685,662	1,321,575	3,784,994	505,143	G4_0.5_ABR15	23,445,520	22,344,630	7,101,216	368,508	1,113,362	6,613,711
R4_1.5_MAY15	24,687,758	24,041,442	4,133,385	146,548	3,582,845	310,529	G4_1.5_MAY15	19,433,662	18,767,662	5,137,953	1,141	4,598	5,130,847	
R4_2.5_JUN15	22,388,592	21,640,620	4,905,173	1,172,266	3,104,624	492,066	G4_2.5_JUN15	21,162,978	20,545,114	5,418,843	3,014	32,273	5,382,677	
R4_3.5_JUL15	20,300,186	19,713,054	3,896,143	191,975	3,275,876	338,439	G4_3.5_JUL15	15,005,614	14,799,974	5,730,079	322,526	1,680,859	3,661,144	
4	R4_4.5_SEPT15	28,639,464	27,545,642	5,117,878	545,819	3,027,321	734,017	G4_4.5_SEPT15	29,814,574	29,115,648	7,236,158	15,427	253,261	6,963,796
	R4_6_OCT15	26,369,492	25,324,252	5,641,886	743,773	2,799,803	1,080,157	G4_6_OCT15	24,690,886	24,063,294	4,584,842	20,671	6,960	4,553,829
	R4_7_NOV15	22,516,512	21,830,826	5,554,905	237,068	4,688,506	535,662	G4_7_NOV15	15,223,150	13,661,644	5,440,491	16,637	123,870	5,295,862
	R4_7.5_DIC15	23,104,964	22,265,858	5,346,327	593,088	3,582,140	1,069,792	G4_7.5_DIC15	16,295,722	15,759,068	4,965,742	9,932	300,320	4,652,027
	R4_8.5_ENE16	26,899,436	25,888,390	5,851,276	476,001	4,050,192	818,694	G4_8.5_ENE16	25,529,218	24,783,022	4,740,721	6,597	65,980	4,666,886
	R4_9.5_FEB16	23,825,382	23,824,354	5,890,334	734,073	4,481,350	587,158	G4_9.5_FEB16	16,648,040	16,178,460	2,555,266	7,171	31,764	2,515,487
	R4_11_MAR16	15,909,362	15,382,404	5,031,819	1,211,103	1,966,633	1,395,882	G4_11.5_MAR16	23,278,964	19,971,646	7,437,396	169,259	1,060,287	6,189,602
	R4_12_ABR16	25,127,858	24,092,444	5,463,095	296,764	4,541,729	477,575	G4_12_ABR16	19,328,336	18,708,318	4,855,171	74,893	941,481	3,834,988
	R5_0.5_MAY15	23,975,684	23,210,610	4,370,135	489,532	3,320,603	496,819	G5_0.5_MAY15	18,324,532	17,753,346	4,816,326	312,530	120,719	4,379,807
	R5_1.5_JUN15	24,471,764	23,718,752	4,804,563	411,922	3,869,830	450,667	G5_1.5_JUN15	33,370,638	31,454,216	6,475,912	703,650	6,403,167	4,341,996
	R5_2.5_JUL15	22,866,988	21,813,130	8,901,100	2,540,843	3,820,938	2,459,141	G5_2.5_JUL15	26,109,586	25,128,336	7,076,351	21,989	560,015	6,488,246
	R5_3.5_AGO15	24,984,482	23,824,354	5,890,334	734,073	4,481,350	587,158	G5_3.5_AGO15	14,307,080	13,670,308	4,641,051	7,373	364,736	4,266,636
R5_4.5_SEPT15	23,751,012	22,561,026	7,607,556	2,375,810	3,876,992	1,275,062	G5_4.5_SEPT15	21,017,366	20,168,458	7,239,850	12,705	344,857	6,876,930	
R5_5_SEPT15	27,586,126	26,060,432	9,349,993	2,461,284	4,954,218	1,840,564	G5_5_SEPT15	19,652,194	18,808,992	5,806,011	507,936	982,399	4,305,271	
R5_6_NOV15	23,455,272	22,740,798	6,308,376	900,893	4,373,184	980,911	G5_6_NOV15	11,077,840	10,588,092	3,718,960	5,545	12,056	3,699,627	
R5_7.5_DIC15	20,214,358	19,405,044	4,944,599	553,582	3,627,127	712,280	G5_7.5_DIC15	18,744,410	18,058,420	6,285,983	292,635	402,722	5,580,891	
R5_8.5_ENE16	19,131,270	18,139,012	8,692,307	2,686,438	2,706,112	3,257,223	G5_8.5_ENE16	17,104,174	16,250,766	5,718,073	297,643	3,497,014	1,899,998	
R5_9_FEB16	23,193,950	22,368,970	5,421,335	986,936	3,721,189	648,915	G5_9_FEB16	22,195,006	21,166,360	6,878,513	79,192	990,352	5,803,589	
R5_10_MAR16	21,611,574	20,970,538	4,941,650	488,535	3,707,220	699,395	G5_10_MAR16	23,129,960	22,143,840	6,555,765	27,239	2,081,770	4,443,311	
R5_11_MAR16	21,371,926	20,630,980	7,362,667	2,235,029	3,892,930	1,184,992	G5_11_MAR16	18,286,016	15,963,142	5,734,455	4,353	4,002	5,722,003	
R6_0.5_JUL15	27,180,946	26,005,436	5,708,599	119,902	2,202,693	1,265,983	G6_0.5_JUL15	12,980,873	12,738,299	2,387,506	7,050	3,401	2,374,547	
R6_1_AGO15	32,465,606	32,266,916	6,126,802	197,567	5,099,936	706,552	G6_1_AGO15	19,166,362	18,837,757	2,353,867	50,125	40,284	2,231,301	
R6_2_SEPT15	26,643,066	26,459,828	7,137,024	299,076	5,637,635	958,218	G6_2_SEPT15	32,601,919	19,636,860	4,308,060	208,618	328,476	3,804,013	
R6_3_OCT15	25,916,676	24,962,964	7,913,934	76,475	2,238,402	689,374	G6_3_OCT15	22,175,458	16,487,744	2,642,440	61,071	54,908	2,496,542	
R6_4_NOV15	21,746,878	21,558,584	5,122,243	93,376	4,274,980	608,089	G6_4.5_NOV15	16,847,692	11,878,854	3,402,296	71,948	394,831	2,926,458	
R6_5_DIC15	26,441,926	26,278,136	7,786,466	591,698	6,091,145	1,043,311	G6_5_DIC15	15,246,512	14,877,981	2,328,332	52,610	149,837	2,032,338	
R6_6_ENE16	28,082,882	27,815,598	4,912,143	88,826	4,199,872	600,380	G6_6_ENE16	15,935,264	11,277,060	2,458,708	17,843	252,765	2,184,520	
R6_7_FEB16	32,792,600	32,566,008	7,788,143	155,508	6,850,910	692,421	G6_7_FEB16	16,516,566	15,481,282	7,698,751	165,409	172,002	5,329,136	
R6_9_ABR16	29,062,414	28,866,638	7,884,415	1,540,075	5,664,519	613,930	G6_8.5_MAR16	19,578,686	18,265,450	8,425,589	193,791	409,326	6,681,447	
R6_10_MAY16	25,641,730	24,474,226	6,034,275	15,563	4,918,477	849,866	G6_10_MAY16	25,699,940	24,638,548	9,612,052	490,457	570,306	6,743,209	
R7_0.5_AGO15	29,305,992	28,585,142	8,646,407	400,028	4,066,793	2,247,729	G7_0.5_AGO15	37,118,500	36,482,138	4,801,367	19,586	39,862	4,739,112	
R7_1_SEPT15	29,701,736	29,489,950	5,792,030	160,907	4,731,147	678,388	G7_1_SEPT15	35,985,055	35,985,266	5,358,858	17,091	56,935	5,281,548	
R7_2_SEPT15	30,684,936	30,462,796	7,616,325	314,914	6,114,222	1,024,285	G7_2_OCT15	10,955,756	10,816,206	2,329,705	7,374	28,059	2,296,167	
R7_4_NOV15	43,382,746	43,133,362	8,043,115	152,119	7,229,006	584,214	G7_3.5_NOV15	11,620,338	11,455,897	2,824,761	11,808	561,189	2,236,803	
R7_5.5_DIC15	42,502,832	42,310,632	6,531,579	89,562	5,910,552	478,648	G7_4_DIC15	52,577,147	5,662,646	5,662,646	9,922	63,595	5,585,226	
R7_6_ENE16	26,231,416	25,971,852	2,665,124	1,290,409	991,558	345,408	G7_6_ENE16	49,915,755	48,887,127	6,851,320	212,451	2,726,961	3,871,905	
R7_7.5_FEB16	32,466,294	32,267,764	7,515,363	110,938	6,655,447	666,264	G7_6.5_FEB16	7,879,022	7,704,810	1,119,811	4,940	18,400	1,043,087	
R7_8_MAR16	27,841,466	27,649,774	5,322,253	169,927	4,509,595	660,867	G7_7_MAR16	53,935,102	27,438,646	7,117,451	3,192	10,785	7,091,587	
R7_9_ABR16	36,708,006	36,481,832	8,264,443	157,259	7,399,952	631,355	G7_8.5_ABR16	13,285,066	13,096,945	2,133,898	29,040	249,597	1,840,490	
R7_10_MAY16	28,486,392	28,285,652	7,961,431	170,831	6,810,626	814,919	G7_9.5_MAY16	30,242,109	29,604,335	5,549,426	186,607	1,107,434	3,916,285	
R7_10.5_JUN16	32,063,202	31,863,672	6,251,534	165,776	5,468,960	524,770	G7_10.5_JUN16	19,510,346	12,426,702	3,897,276	240,901	1,585,924	2,046,899	
R9_1.5_ENE16	24,393,254	24,235,746	5,178,095	241,051	4,256,620	561,622	G7_12_AGO16							

Tabla S2: Estadística de las muestras orofaríngeas. Valores estadísticos de comparación de los datos de los infantes con los índices de alfa y beta diversidad en las lecturas de virus que infectan a humanos. En negrita, los valores que fueron estadísticamente significativos. Se realizaron pruebas no paramétricas.

	Variable / Valor de p	Diversidad alfa			Diversidad beta	
		Promedio de abundancia de especies	Indice Chao1	Indice Shannon	Bray-Curtis	R2
Información general	Trimestres	0.218	0.52	0.4878	0.427	0.038
	Primer Trimestre	0.355	0.087	0.508	0.689	0.007
	Segundo Trimestre	0.221	0.785	0.413	0.938	0.002
	Tercer Trimestre	0.06	0.100	0.149	0.105	0.022
	Cuarto Trimestre	0.359	0.971	0.872	0.224	0.018
	Semestres	0.380	0.177	0.180	0.447	0.011
Infantes	0.190	0.008	0.333	0.001	0.217	
Alimentación	Formula	0.164	0.356	0.507	0.002	0.045
	Otras bebidas	0.145	0.423	0.336	0.158	0.021
	Frutas	0.063	0.57	0.306	0.701	0.007
	Verduras	0.350	0.182	0.485	0.589	0.009
Información socio demográfica	Nacimiento				0.014	0.035
	1	0.555	0.266	0.73		
	3	0.285	0.019	0.904		
	6	0.412	0.387	0.48		
	9	0.73	0.176	0.512		
	12	0.415	0.209	0.904		
	Percentil promedio de peso				0.245	0.017
	1	1	0.536	0.555		
	3	0.555	0.622	0.904		
	6	0.412	0.901	0.368		
	9	1	1	1		
	12	0.318	0.642	0.51		
	Almacenamiento de agua				0.009	0.035
	1	0.800	0.711	0.417		
	3	0.400	0.46	0.111		
	6	0.900	0.621	0.73		
	9	0.904	0.014	0.78		
	12	0.300	0.73	0.9		
	Consumo de agua				0.001	0.051
	1	0.730	0.266	0.555		
	3	1.000	0.085	0.285		
	6	0.063	0.216	0.412		
	9	0.730	0.109	0.321		
	12	1.000	0.15	0.42		
	Material para cocinar				0.021	0.035
	1	0.285	0.536	0.904		
	3	0.555	0.14	0.304		
	6	0.111	0.058	0.063		
	9	1.000	0.325	0.73		
	12	1.000	0.403	0.415		
	Tipo de suelo				0.368	0.013
	1	0.412	1	0.285		
	3	0.555	1	0.412		
	6	0.730	0.621	0.512		
	9	0.111	0.46	0.73		
	12	0.285	0.502	0.56		
	Drenaje				0.275	0.016
	1	0.39	0.131	0.087		
	3	0.7	1	1		
	6	1	0.648	0.514		
	9	1	1	1		
	12	0.512	0.61	0.8		
	Cucarachas				0.225	0.016
	1	0.42	0.642	0.278		
	3	1	1	1		
	6	0.2	0.377	0.657		
	9	0.2	0.236	0.8		
	12	1	1	1		
Roedores				0.026	0.030	
1	0.115	0.199	0.142			
3	0.266	0.481	0.805			
6	1	0.648	0.541			
9	0.4	1	0.8			
12	1	1	1			
Hacinamiento				0.175	0.019	
1	0.519	0.497	0.142			
3	0.417	0.508	0.205			
6	0.75	0.325	0.456			
9	1	0.608	1			
12	0.666	1	0.675			

Tabla S3: Estadística de las muestras fecales. Valores estadísticos de comparación de los datos de los infantes con los índices de alfa y beta diversidad en las lecturas de virus que infectan a humanos. En negrita, los valores que fueron estadísticamente significativos. Se realizaron pruebas no paramétricas.

	Variable / Valor de p	Diversidad alfa			Diversidad beta	
		Promedio de abundancia de especies	Indice Chao1	Indice Shannon	Bray-Curtis	R2
Información general	Trimestres	0.252	0.072	0.64	0.001	0.106
	Primer Trimestre	0.003	0.021	0.775	0.001	0.068
	Segundo Trimestre	0.783	0.607	0.842	0.152	0.014
	Tercer Trimestre	0.510	0.135	0.414	0.005	0.033
	Cuarto Trimestre	0.217	0.108	0.728	0.003	0.042
	Semestres	0.002	0.010	0.660	0.001	0.082
	Infantes	0.824	0.044	0.361	0.001	0.263
Alimentación	Formula	0.194	0.179	0.635	0.266	0.034
	Otras bebidas	0.001	0.009	0.723	0.002	0.047
	Frutas	0.001	0.006	0.742	0.001	0.055
	Verduras	0.008	0.009	0.747	0.001	0.058
Información socio demográfica	Nacimiento				1	0.026
	1	1	0.485	0.8		
	3	0.8	0.857	1		
	6	0.4	0.21	0.2		
	9	0.56	0.345	1		
	12	1.000	0.456	0.3		
	Percentil promedio de peso				1	0.062
	1	0.2	0.485	0.4		
	3	1	0.857	1		
	6	0.2	0.103	0.8		
	9	1	0.15	0.9		
	12	0.8	0.315	0.5		
	Almacenamiento de agua				1	0.063
	1	0.800	1	0.417		
	3	0.400	0.628	0.111		
	6	0.800	0.9	0.73		
	9	0.900	0.15	0.78		
	12	1.000	0.84	0.9		
	Consumo de agua				1	0.044
	1	0.800	0.685	1		
	3	0.400	0.628	0.2		
	6	0.800	0.531	0.8		
	9	1.000	0.432	0.9		
	12	0.600	0.635	1		
	Material para cocinar				1	0.026
	1	0.800	0.885	0.8		
	3	0.200	1	0.4		
	6	0.800	0.316	1		
	9	0.600	0.451	0.926		
	12	1.000	0.605	0.305		
	Tipo de suelo				1	0.019
	1	1.000	1	0.8		
	3	0.800	1	1		
	6	0.400	0.802	0.2		
	9	0.720	0.915	0.4		
	12	0.653	0.895	0.63		
	Drenaje				0.275	0.016
	1	0.164	0.242	0.949		
	3	0.4	0.506	0.4		
	6	1	0.586	0.609		
	9	0.8	1	0.2		
	12	0.571	0.058	0.392		
	Cucarachas				0.225	0.016
	1	0.239	0.155	0.504		
	3	0.266	0.638	0.533		
	6	1	0.383	1		
	9	1	0.8	1		
	12	0.428	0.737	0.642		
	Roedores				1	0.030
	1	0.317	0.153	0.749		
	3	1	0.268	0.1		
	6	0.69	1	0.42		
	9	0.4	0.8	1		
	12	0.428	0.737	0.642		
	Hacinamiento				0.175	0.019
	1	0.49	0.555	1		
	3	0.7	1	1		
6	1	0.593	0.69			
9	1	0.8	1			
12	0.071	0.453	0.785			

Tabla S4. Muestras que de acuerdo al criterio de infección, fueron positivas a norovirus. El porcentaje de cobertura se calculó a partir del genoma original. El número de lecturas son la cantidad de lecturas que formaron ese genoma o contig.

Infante	Identificador de muestra	Genotipificación por alineamiento	Mejor hit de Blast	Tamaño en pares de bases	# de lecturas	% cobertura	Profundidad promedio
IGV2	G2_12_MAR16	GII.17[P17]	NC_039475	7556	60,165	98.9	586.3
		GI.3[P13]	MK073887.1	7691	21,301	98.2	206.1
IGV4	G4_6.5_NOV15	GII.7[P7]	MN038126.1	7374	6,491	100.0	63.7
	G4_10_FEB16	GI.3[P13]	MK073887.1	7691	5,675	40.2	134.1
	G4_11.5_MAR16	GII.4[P16]	MK762567.1	7501	55,760	100.0	451.9
IGV5	G5_4.5_SEP15	GII.17[P17]	NC_039475	7556	328	78.4	4.0
	G5_7.5_DIC15	GII.7[P7]	MN038126.1	7374	8,365	100.0	82.9
	G5_9_FEB16	GI.3[P13]	MK073887.1	7691	1,035,653	100.0	5670.5
	G5_10_MAR16	GI.3[P13]	MK073887.1	7691	4,258	98.5	40.9
IGV6	G6_4.5_NOV15	GII.7[P7]	MN038126.1	7374	80	40.9	2.9
	G6_5.5_DIC15	GII.4[P16]	MK762567.1	7501	275	71.1	4.6
IGV7	G7_3.5_NOV15	GII.7[P7]	MN038126.1	7374	856,374	100.0	5912.0
	G7_4_DIC15	GII.4[P16]	MK762567.1	7501	52	35.8	2.3
	G7_7_MAR16	GI.3[P13]	MK073887.1	7691	15,061	100.0	141.3
	G7_8.5_ABR16	GII.4[P16]	MK762567.1	7501	1,915	97.5	19.2
	G7_9.5_MAY16	GII.4[P16]	MK762567.1	7501	60	27.2	3.0
IGV9	G9_4.5_ABR16	GII.4[P16]	MK762567.1	7501	10,431	99.4	102.9
	G9_5.5_MAY16	GII.4[P16]	MK762567.1	7501	419	82.1	5.8
IGV11	G11_6_ABR16	GII.4[P16]	MK762567.1	7501	56,844	99.5	552.7
IGV15	G15_4.5_MAY16	GII.7[P7]	MN038126.1	7374	90	49.9	2.6
	G15_5.5_JUN16	GII.4[P16]	MK762567.1	7501	156	50.9	3.7
	G15_11_NOV16	GII.4[P16]	MK762567.1	7501	101	56.1	2.8
IGV19	G19_3_AGO16	GII.4[P16]	MK762567.1	7501	133	54.4	3.3
	G19_6_NOV16	GII.4[P16]	MK762567.1	7501	210	73.6	3.6
	G19_8_ENE17	GII.4[P16]	MK762567.1	7501	173	66.7	3.4
	G19_10_FEB17	GII.4[P16]	MK762567.1	7501	1,919,291	99.7	7286.6
	G19_10.5_MAR17	GII.4[P16]	MK762567.1	7501	552	85.2	6.2
G19_11.5_ABR17	GII.4[P16]	MK762567.1	7501	107	52.8	2.8	

Tabla S5. Muestras que de acuerdo al criterio de infección, fueron positivas a sapovirus. El porcentaje de cobertura se calculó a partir del genoma original. El número de lecturas son la cantidad de lecturas que formaron ese genoma o contig.

Infante	Identificador de muestra	Genotipificación por alineamiento	Mejor hit de Blast	Tamaño en pares de bases	# de lecturas	% cobertura	Profundidad promedio
IGV5	G5_8_ENE16	GI.1	MG012437.1	7388	331	71	4
	G5_9_FEB16	GII.1	AY237419.2	7490	55,859	43	832
IGV7	G7_9.5_MAY16	GI.1	MG012437.1	7388	9,917	100	99
IGV9	G9_5_MAY16	GI.1	MG012437.1	7388	126,253	100	1,238
IGV15	G15_11_NOV16	GI.7	AB522390.3	7452	49,264	96	478
		GII.4	KP067444.1	7460	39,553	100	385

Figuras

Figura S1. Distancias que existen en la composición de virus eucariontes que infectan a humanos en las muestras orofaríngeas (a-c) Análisis no constreñidos (PCoA) mostrando los 3 ejes principales y porcentaje por eje. (d) análisis constreñidos (db-RDA) mostrando solo 2 ejes. Cada niño tiene un color y punto distintivo. Las elipses en color negro representan el error estándar del centroide.

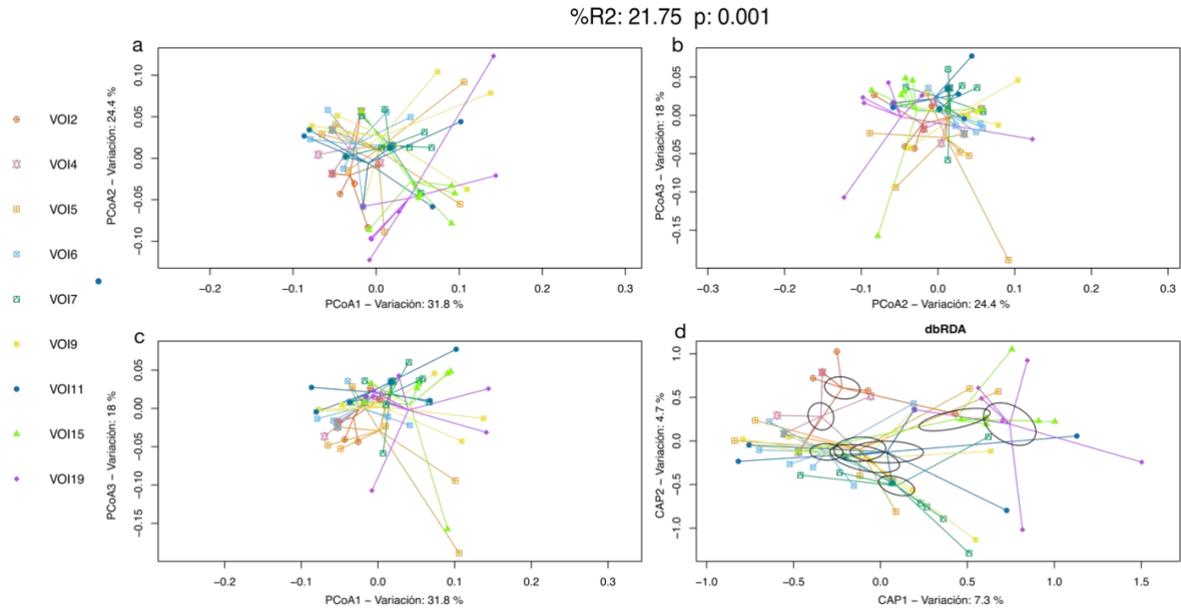


Figura S2. Distancias que existen en la composición de virus eucariontes que infectan a humanos en las muestras fecales (a-c) Análisis no constreñidos (PCoA) mostrando los 3 ejes principales y porcentaje por eje. (d) análisis constreñidos (db-RDA) mostrando solo 2 ejes. Cada niño tiene un color y punto distintivo. Las elipses en color negro representan el error estándar del centroide.

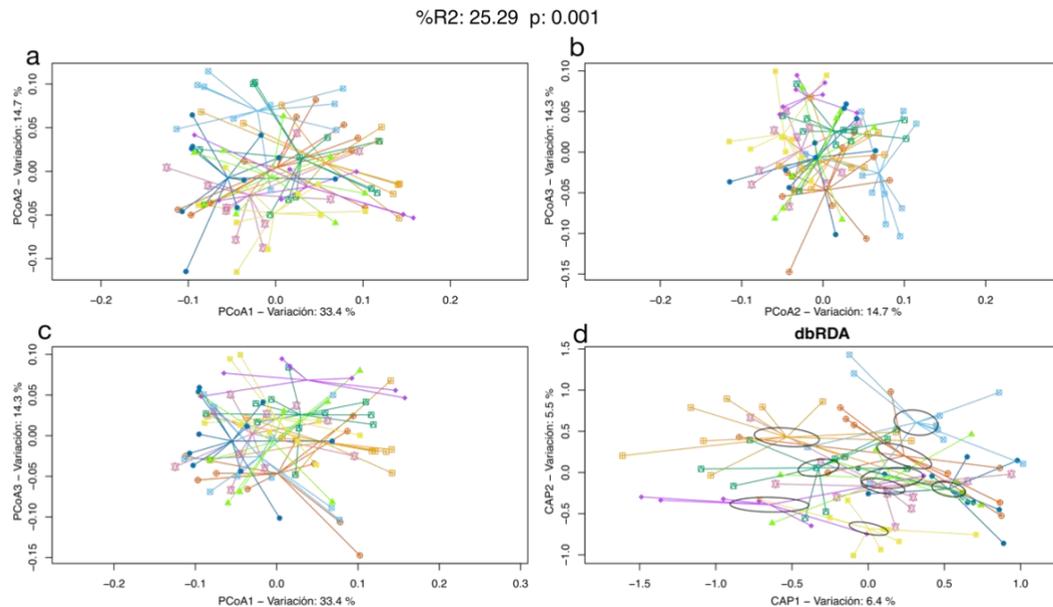


Figura S3. Diversidad genética de norovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo GTR+I. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.

Norovirus

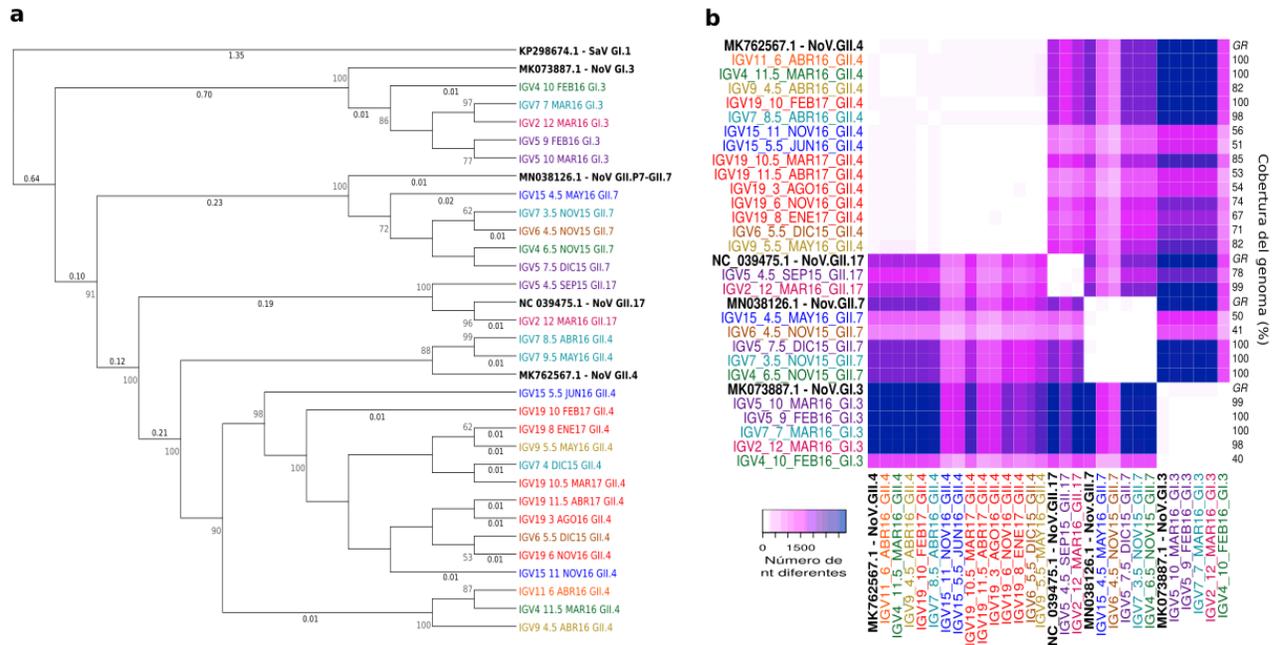


Figura S4. Diversidad genética de sapovirus. (a) Árbol filogenético de genomas completos y secuencias parciales (contigs >500) construido con el modelo TIM1+F+G4. Los genomas de referencia y raíz están en negritas. (b) Mapa de calor mostrando diferencias pareadas de cambios de nucleótidos en secuencias con al menos 40% de cobertura.

Sapovirus

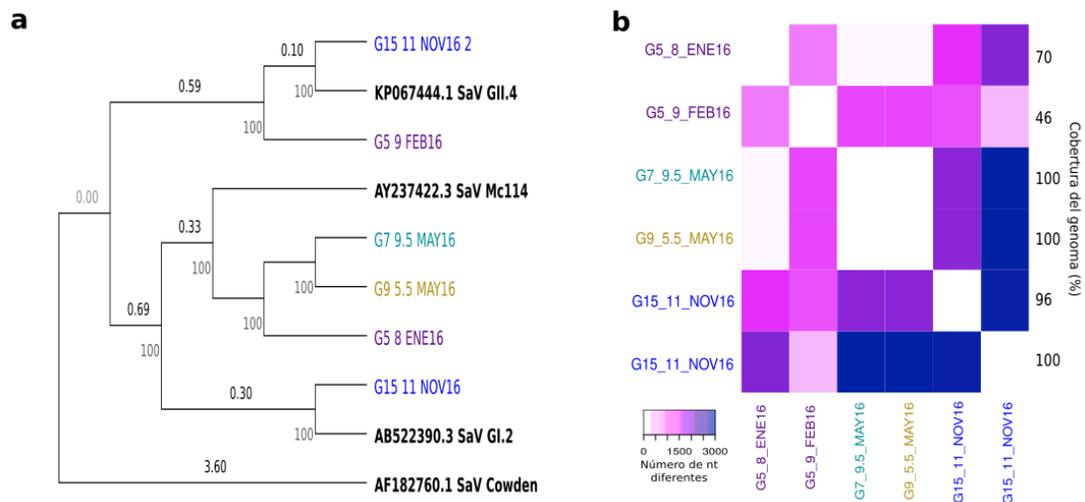


Figura S5. Distancias que existen en la composición de virus eucarionotes que infectan a humanos en ambos tractos. (a-c) Análisis no constreñidos (PCoA) mostrando los 3 ejes principales y porcentaje por eje. (d) análisis constreñidos (db-RDA) mostrando solo 1 eje. Cada niño tiene un color y se distinguen los nichos por las viñetas empleadas.

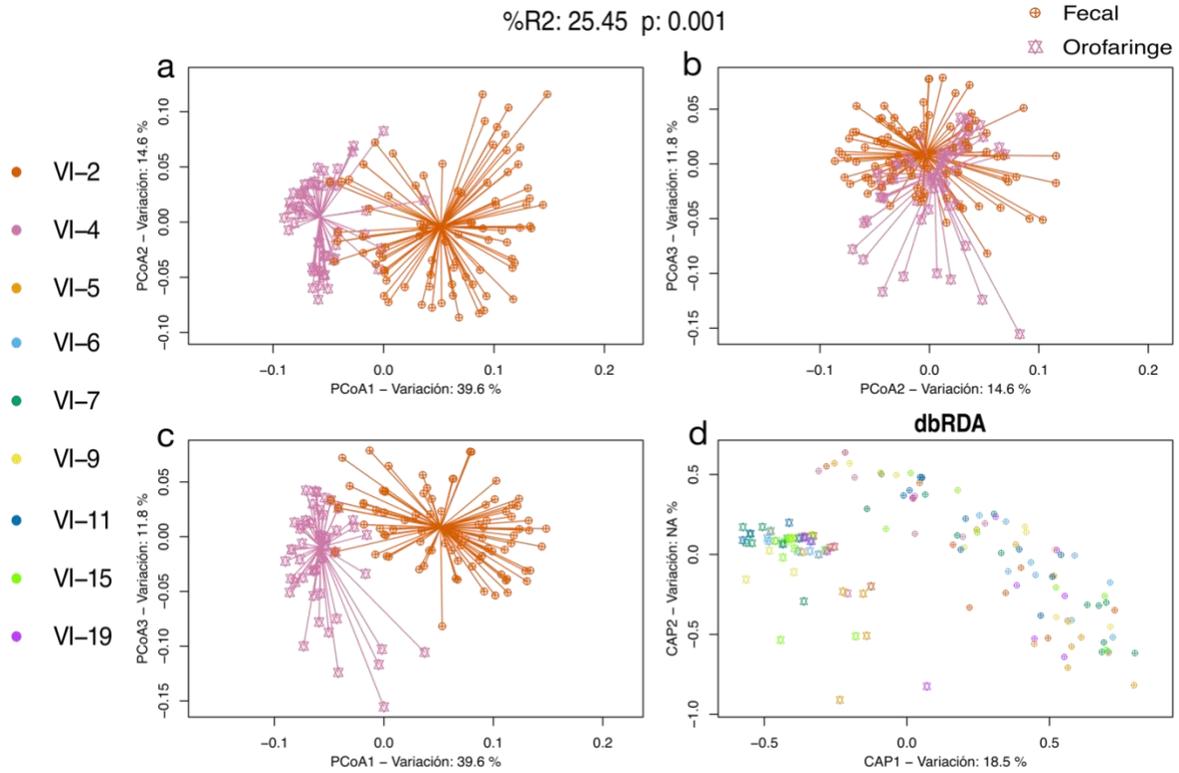


Figura S6. Diversidad genética de pepper mild mottle virus. Árbol filogenético creado con genomas completos y secuencias parciales (contigs > 500nt) construido con el modelo GTR+F+R10. Los genomas de referencia y raíz están en negritas.

Pepper mild mottle virus

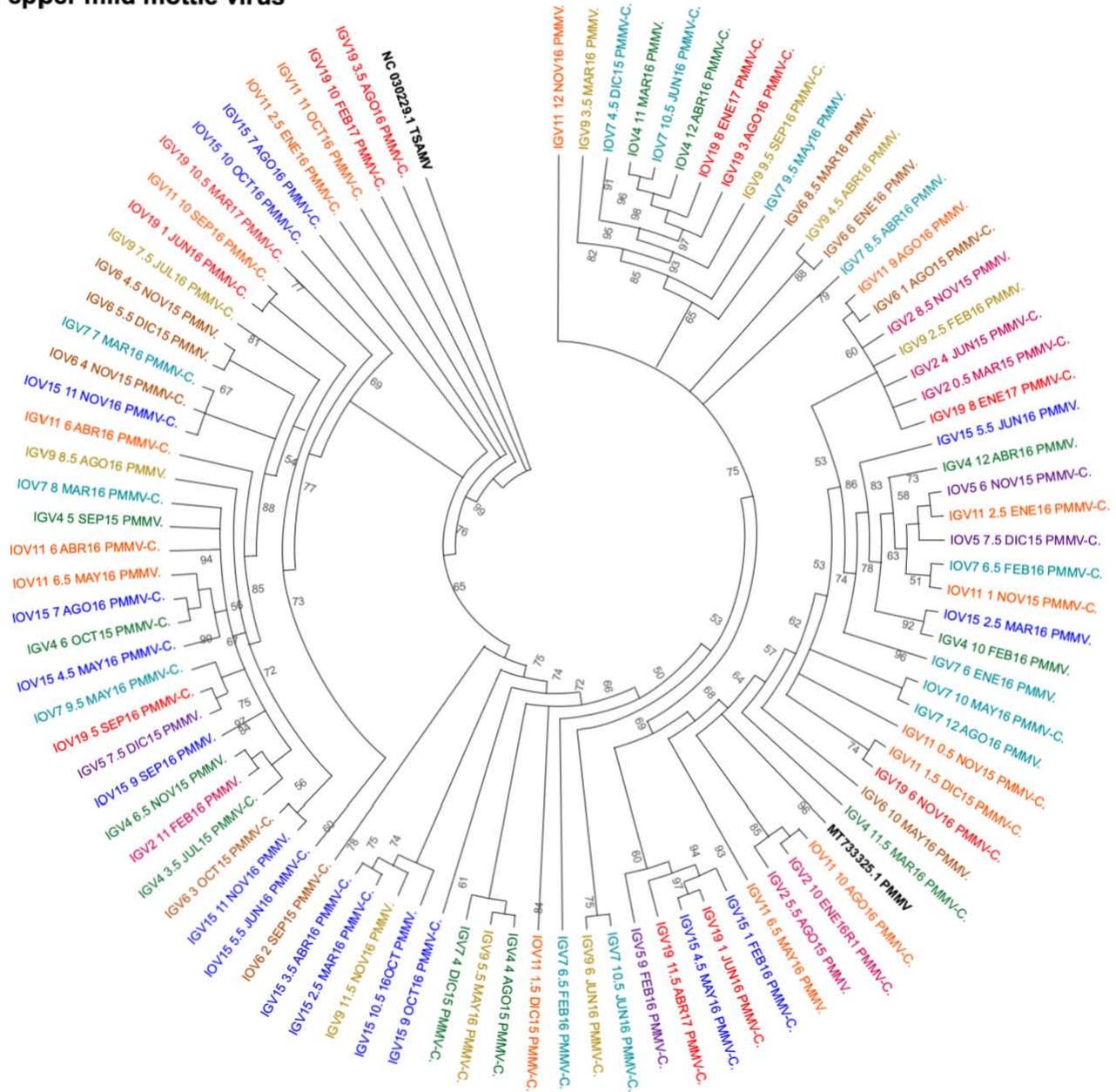


Figura S7. Cambios de nucleótidos de pepper mild mottle virus. Mapa de calor hecho a partir de secuencias con un mínimo de cobertura del genoma del 40%.

Pepper mild mottle virus

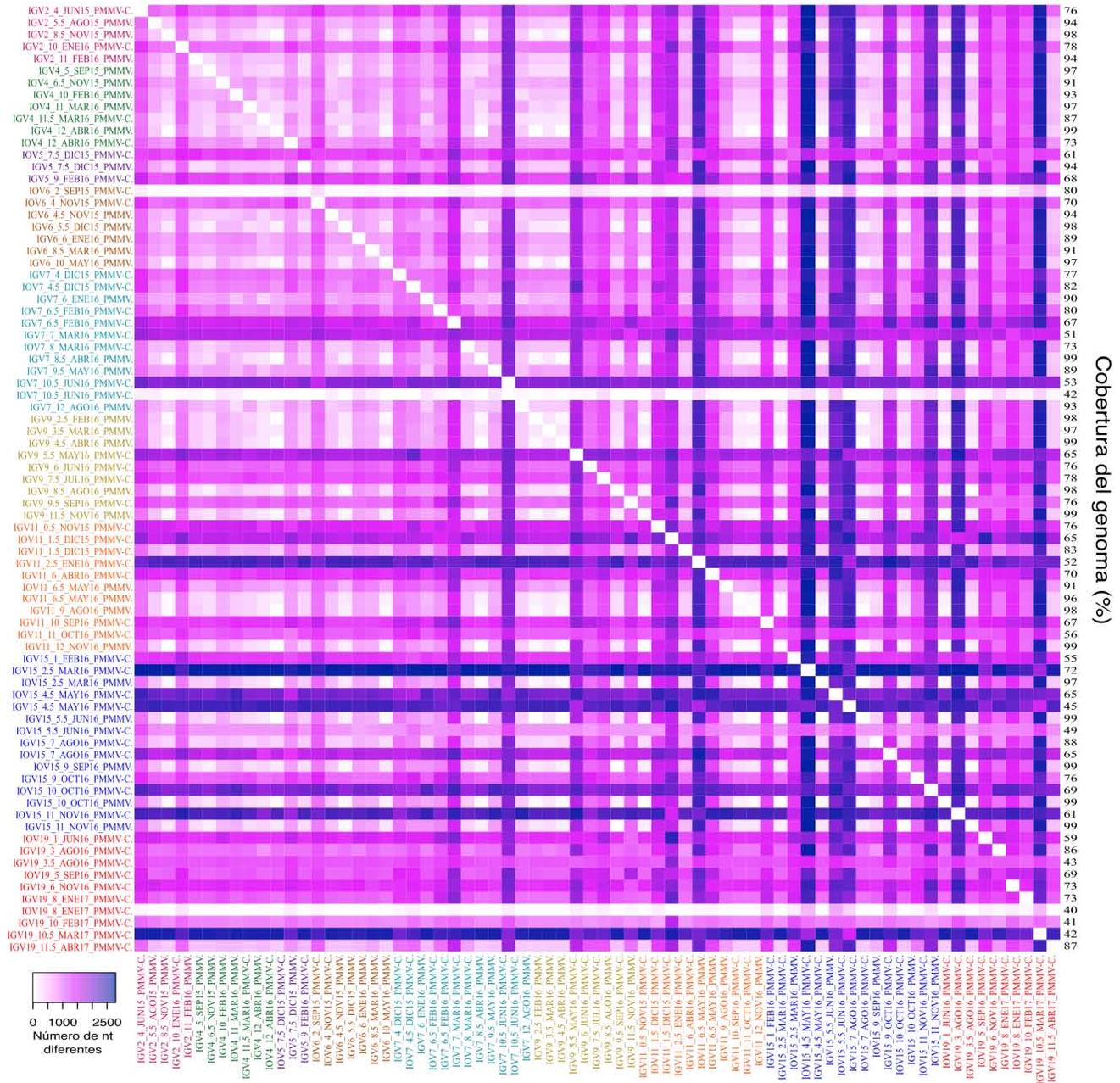


Figura S8. Diversidad genética de tropical soda apple mosaic virus. Árbol filogenético creado con genomas completos y secuencias parciales (contigs > 500nt) construido con el modelo GTR+F+R6. Los genomas de referencia y raíz están en negritas.

Tropical soda apple mosaic virus

