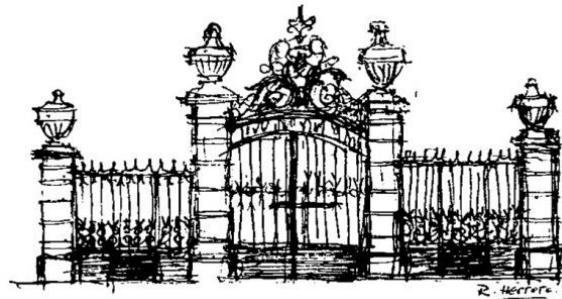


UNIVERSIDAD DE EXTREMADURA



ESCUELA DE INGENIERÍAS AGRARIAS



**REVISIÓN SOBRE LOS ESTUDIOS DEL USO DE LA
SECUENCIACIÓN DE ÚLTIMA GENERACIÓN PARA
MEJORAR LA SEGURIDAD DE LOS ALIMENTOS.**

TRABAJO FIN DE GRADO

CIENCIA Y TECNOLOGÍA DE LOS ALIMENTOS

María Gómez Cendrero

Badajoz, Noviembre 2021

*María
Gómez
Cendrero*

**REVISIÓN SOBRE LOS ESTUDIOS DEL USO DE LA SECUENCIACIÓN DE
ÚLTIMA GENERACIÓN PARA MEJORAR LA SEGURIDAD DE LOS
ALIMENTOS**

GRADO EN CIENCIA Y TECNOLOGÍA DE LOS ALIMENTOS

Noviembre, 2021

UNIVERSIDAD DE EXTREMADURA



ESCUELA DE INGENIERÍAS AGRARIAS



**REVISIÓN SOBRE LOS ESTUDIOS DEL USO DE LA
SECUENCIACIÓN DE ÚLTIMA GENERACIÓN PARA
MEJORAR LA SEGURIDAD DE LOS ALIMENTOS.**

TRABAJO FIN DE GRADO

CIENCIA Y TECNOLOGÍA DE LOS ALIMENTOS

María Gómez Cendrero

Badajoz, noviembre 2021

TRABAJO FIN DE GRADO

REVISIÓN SOBRE LOS ESTUDIOS DEL USO DE LA SECUENCIACIÓN DE ÚLTIMA GENERACIÓN PARA MEJORAR LA SEGURIDAD DE LOS ALIMENTOS.

GRADO EN CIENCIA Y TECNOLOGÍA DE LOS ALIMENTOS

AUTOR: María Gómez Cendrero

Fdo: María Gómez Cendrero

TUTOR/ES: María José Benito Bernaldez
Almudena Vázquez Merchán

Tutor

Cotutor

Fdo: María José Benito Bernáldez **Fdo:** Almudena Vázquez Merchán

Convocatoria: noviembre, 2021

ÍNDICE

1. INTRODUCCIÓN	1
2. OBJETIVOS	4
3. DESCRIPCIÓN DE LAS TECNOLOGÍAS.	6
4. SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS.....	13
4.1. Aplicaciones actuales.	13
4.2. Los principios del rastreo basados en el WGS.	14
4.2.1. El enfoque del SNP.	15
4.2.2. Enfoque genético.....	17
4.2.3. Análisis filogenético.....	18
4.2.4. Comparación entre el SNP y el cg/wgMLST.....	18
4.3. Interpretación de los resultados.	19
4.4. La necesidad de estandarización.....	20
4.5. Salud pública y medidas reglamentarias basadas en los resultados del WGS.....	21
4.6. Gestión de la seguridad alimentaria.....	23
4.7. Consideraciones sobre la aplicación en la industria.	24
4.8. Desafíos que deben abordarse.	25
5. SECUENCIACIÓN DE AMPLICONES, METAGENÓMICA Y METATRANSCRIPTÓMICA.....	28
5.1. Una definición de los términos.....	28
5.1.1. Perfil de la comunidad microbiana basada en amplicón (metabarcoding).....	28
5.1.2. Perfil del microbioma metagenómico	29
5.2. La meta-ómica para la caracterización funcional de los microbiomas.....	29
5.3. Herramientas computacionales para la caracterización de los microbiomas.	30
5.4. Aplicaciones de la metagenómica en la seguridad alimentaria.	31
5.5. Problemas y desafíos.	32
5.6. Validación y evaluación comparativa.....	35
6. CONSIDERACIONES Y DESAFÍOS RELACIONADOS CON EL INTERCAMBIO DE DATOS.	38
6.1. Interpretación correcta de los datos.	38
6.2. Claridad jurídica/debida diligencia.....	39
6.3. Los propietarios de los datos.	40
7. PERSPECTIVAS FUTURAS PARA MEJORAR LA SEGURIDAD ALIMENTARIA.....	42
7.1. Secuenciación del genoma completo.....	42
7.2. Análisis metagenómico.....	43

7.3. Las repercusiones de la aplicación de los NGS en el comercio y la industria alimentaria.	45
8. CONCLUSIONES.	47
9. REFERENCIAS.....	49

1. INTRODUCCIÓN

INTRODUCCIÓN

1. INTRODUCCIÓN

Las técnicas *Next Generation Sequencing* (NGS) tienen un alto rendimiento y producen miles o incluso millones de secuencias al mismo tiempo. Estas secuencias permiten la identificación precisa de taxones microbianos, incluidos los organismos no cultivables y los presentes en pequeñas cantidades. En aplicaciones específicas, NGS proporciona un inventario completo de todos los genes y operones microbianos presentes o que se expresan en diferentes condiciones de estudio. Las técnicas de NGS están revolucionando el campo de la ecología microbiana y recientemente se han utilizado para examinar varios ecosistemas alimentarios. La Secuenciación de Última Generación o *Next Generation Sequencing* (NGS) combinada con robustos enfoques bioinformáticos, están revolucionando la microbiología alimentaria.

En el último decenio, la Secuenciación de la Última Generación (NGS) ha dejado de ser únicamente una herramienta de investigación para convertirse en una aplicación rutinaria en muchos campos, como el diagnóstico, la investigación de brotes, la resistencia a los antimicrobianos, la medicina forense y la autenticidad de los alimentos (Allard et al., 2017, Goodwin et al., 2016, Quainoo et al., 2017). La tecnología se está desarrollando a un ritmo rápido, con una mejora continua de la calidad y una reducción de los costos (The National Human Research Institute, 2017) y está teniendo una gran influencia en la microbiología de los alimentos.

El NGS en la microbiología de los alimentos se utiliza predominantemente de dos maneras: I) la determinación de la secuencia del genoma completo de un solo aislado cultivado (por ejemplo, una colonia bacteriana, un virus o cualquier otro organismo), lo que se denomina comúnmente "secuenciación del genoma completo" (WGS), y II) la "metagenómica", en la que la NGS se aplica a una muestra biológica generando secuencias de múltiples (si no todos) microorganismos en esa muestra. El alto poder discriminatorio de la WGS en comparación con los instrumentos tradicionales de tipificación molecular está bien establecido y la WGS está ganando aceptación como instrumento de vigilancia prospectiva de las enfermedades transmitidas por los alimentos (Allard et al., 2016, Ashton et al., 2016, Jackson et al., 2016). La tecnología de NGS está sustituyendo cada vez más las técnicas tradicionales de tipificación y caracterización microbiana, proporcionando respuestas más rápidas y precisas.

INTRODUCCIÓN

La aplicación de la metagenómica para mejorar la inocuidad y la calidad de los alimentos está todavía en sus comienzos y ofrece interesantes oportunidades para predecir la presencia o la aparición de patógenos y microorganismos de deterioro sobre la base de los cambios observados en comunidades microbianas enteras, así como la posibilidad de caracterizar la microbiota desconocida.

El presente trabajo se centra en la utilización reciente y el potencial futuro de los NGS en la microbiología alimentaria, y también se examinan los desafíos actuales. El trabajo también tiene por objeto promover el uso de los NGS en la industria alimentaria, destacando al mismo tiempo las lagunas de conocimientos y las necesidades futuras de investigación para aumentar el valor generado por la aplicación de la tecnología de NGS a los usuarios.

2. OBJETIVOS

DESCRIPCIÓN DE LAS TECNOLOGÍAS

2. OBJETIVOS

El trabajo tiene por objeto promover la aplicación más amplia de las tecnologías de NGS en la industria alimentaria, así como poner de relieve las lagunas de conocimiento y las nuevas aplicaciones de las NGS con el fin de impulsar las investigaciones futuras y aumentar los resultados de la seguridad alimentaria gracias a su uso más amplio. Para ello el trabajo se centrará en:

1. Revisar los estudios de la optimización y aplicación de tecnologías de secuenciación masiva de última generación.
2. Revisar los trabajos de aplicación de estas tecnologías en la industria alimentaria.
3. Proponer nuevas aplicaciones o investigaciones para su utilización en seguridad alimentaria.

3. DESCRIPCIÓN DE LAS TECNOLOGÍAS

3. DESCRIPCIÓN DE LAS TECNOLOGÍAS.

La secuenciación del genoma microbiano se ha convertido en la corriente principal en el campo de la microbiología de los alimentos debido a que crece su asequibilidad y mejora la velocidad de secuenciación y calidad de los datos obtenidos. Esto es posible debido a los avances en las tecnologías conocidas como secuenciación de última generación. Esta tecnología engloba la secuenciación masiva que proporciona lecturas de secuenciación cortas, y la secuenciación de una sola molécula que proporciona lecturas de secuenciación largas. La lectura corta es muy precisa y produce lecturas de 100-300 pb que luego se ensamblan en genomas incompletos. Los genomas completos no pueden generarse a partir de las lecturas cortas debido a las dificultades para ensamblar regiones repetitivas y grandes reordenamientos genómicos. Esto puede ser un problema cuando se requieren genomas completos y para determinar regiones genómicas complejas, ya que se necesitan lecturas más largas (Loman y Pallen, 2015). La secuenciación del ADN microbiano se puede realizar con las tecnologías desarrolladas por Illumina, Ion Torrent, Pacific Biosciences y Oxford Nanopore.

Illumina.

Illumina/Solexa lanzaron el Genome Analyzer II en 2006, y los avances en la tecnología de Illumina en los años intermedios han marcado en gran medida el ritmo de los enormes aumentos en la producción y las reducciones de costes. Como consecuencia, las máquinas de Illumina dominan actualmente el mercado de HTS. El proceso de secuenciación implica la amplificación clonal de fragmentos de ADN ligados a adaptadores en la superficie de un portaobjetos de vidrio (Bentley et al., 2008). Las bases se leen utilizando una estrategia de terminación cíclica reversible, que secuencia la cadena molde un nucleótido a la vez a través de rondas progresivas de incorporación de bases, lavado, formación de imágenes y escisión. En esta estrategia, se utilizan 3'-O-azidometil-dNTPs marcados con fluorescencia para detener la reacción de polimerización, lo que permite eliminar las bases no incorporadas y obtener imágenes fluorescentes para determinar el nucleótido añadido (Guo et al., 2008). Tras el escaneo de la celda de flujo con una cámara de dispositivo de carga acoplada (CCD), se eliminan la fracción fluorescente y el bloque 3', y se repite el proceso. En todos los modelos de Illumina, las tasas de error globales son inferiores al 1%, y el tipo de error más común es la sustitución (Dohm et al., 2008). Illumina produce actualmente un conjunto de secuenciadores (MiSeq, NextSeq, HiSeq, NovaSeq y iSeq) optimizados para una variedad de

DESCRIPCIÓN DE LAS TECNOLOGÍAS

rendimientos y tiempos de entrega (Tabla 3.1). Los MiSeq y HiSeq son las plataformas más consolidadas (Figura 3.1.).

TABLA 3.1 Secuenciadores y sus características. Fuente: Elaboración propia.

Secuenciadores	Tiempo de lectura (Horas)	GB de Reads	Rango de lecturas (millones)	Longitud máxima de lectura (pb)
MiSeq	4-56	0,54-15	1-25	2x300
NextSeq	12-35	16,25-120	130-400	2x150
HiSeq	24-264	47-1800	2,1-6000	2x150
NovaSeq	13-44	4800-6000	32000-40000	2x150
iSeq	9-17,5	1,2	4	2x150

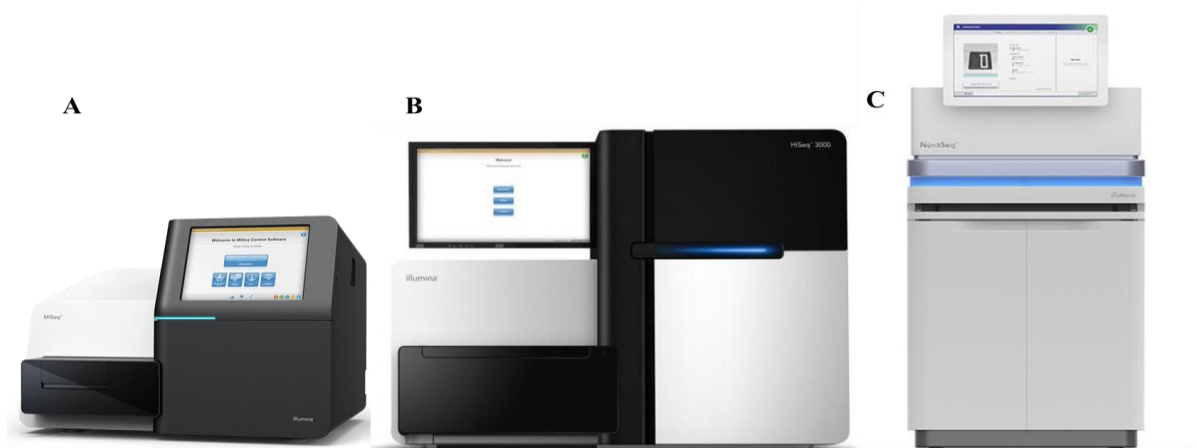


FIGURA 3.1. Equipos de secuenciación de Illumina: Miseq (A), HiSeq 2500 (B) y NovaSeq 6000 (C).

Ion Torrent.

Life Technologies comercializó la tecnología de secuenciación por semiconductores de Ion Torrent en 2010 en forma de secuenciador Ion PGM de sobremesa. En concreto, se utiliza la emulsión-PCR para amplificar clonalmente fragmentos de ADN ligados a adaptadores en la superficie de las perlas. Las perlas se distribuyen posteriormente en micropocillos donde se produce una reacción de secuenciación por síntesis. La secuenciación por semiconductores de Ion Torrent mide los cambios de pH inducidos por la liberación de iones de hidrógeno durante la extensión del ADN (Rothberg et al., 2011). Estos cambios de pH son detectados por un sensor colocado en el fondo del micropozo y convertidos en una señal de voltaje (Figura 3.2.). La señal de voltaje es proporcional al número de bases incorporadas, y la adición secuencial de nucleótidos individuales durante cada ciclo de secuenciación permite la discriminación de bases. Además, Ion Torrent evita el barrido óptico

DESCRIPCIÓN DE LAS TECNOLOGÍAS

para distinguir los nucleótidos durante los ciclos de secuenciación, una diferencia que acelera drásticamente las ejecuciones de secuenciación y reduce los costes. Ion Torrent lanzó una segunda máquina en 2012, la Ion Proton, que aumenta la producción, sin embargo, presenta actualmente un máximo de 200 pb de longitud de lectura, frente a los 400 pb del PGM. También hay disponibles varios chips para adaptar los resultados a las distintas aplicaciones. El PGM es más útil para proyectos de resecuenciación selectiva y análisis de genomas pequeños, mientras que el Proton es capaz de realizar la secuenciación del exoma y el análisis del transcriptoma completo. Las inserciones y deleciones son los tipos de error más comunes (Liu et al., 2012). La correlación entre el número de bases incorporadas y el cambio de tensión posterior no escala perfectamente, puede conducir a un aumento de las tasas de error (Rothberg et al., 2011).

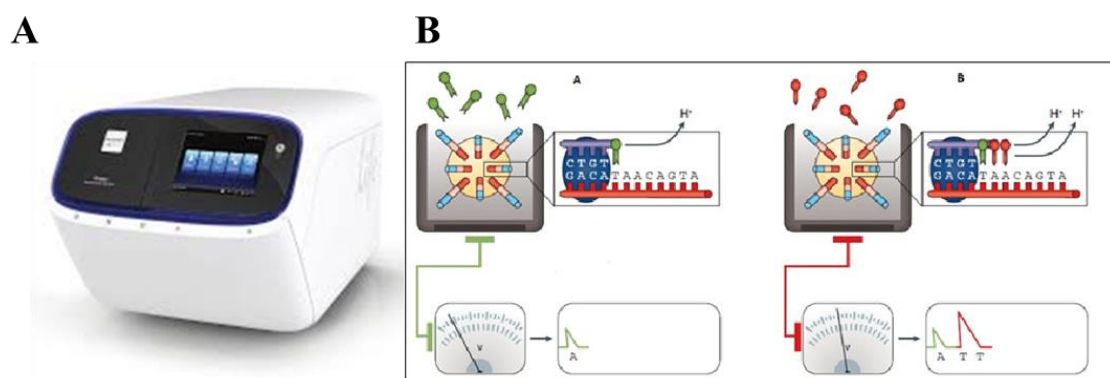


FIGURA 3.2. Ion Proton (A) y funcionamiento tecnología Ion Torrent (B). Fuente: Kchouk et al., 2017.

Pacific Biosciences.

La secuenciación de una sola molécula en tiempo real (SMRT) fue iniciada por Nanofluidics, Inc. y comercializada por Pacific Biosciences. Cabe destacar que esta tecnología no necesita una PCR previa. La preparación del patrón implica la ligación de adaptadores de horquilla de una sola hebra en los extremos de moléculas de ADN digeridas o de ADNc, generando un patrón tapado (campana SMRT). Mediante el uso de una polimerasa que desplaza la hebra, la molécula de ADN original puede secuenciarse varias veces, aumentando así la precisión (Travers et al., 2010). Es importante destacar que se evita la amplificación clonal, lo que permite la secuenciación directa del ADN nativo y potencialmente modificado. La síntesis de ADN se produce en cámaras denominadas guías de onda de modo cero (ZMW), en las que una sola polimerasa está inmovilizada en el fondo de la cámara (Levene et al., 2003). Estas cámaras reducen el ruido de fondo de manera que las

DESCRIPCIÓN DE LAS TECNOLOGÍAS

versiones marcadas con fosfato de los cuatro nucleótidos pueden estar presentes simultáneamente. Así, la polimerización se produce de forma continua, y la secuencia de ADN puede leerse en tiempo real a partir de las señales fluorescentes grabadas en un vídeo (Eid et al., 2009). Lanzado en 2010, el RS II sigue siendo la única máquina de Pacific Biosciences disponible en el mercado, sin embargo, han mejorado el rendimiento. Al igual que en la mayoría de las plataformas de secuenciación de moléculas individuales, las tasas de error (11%) son evidentes para las lecturas de una sola pasada. Sin embargo, los errores de secuenciación se distribuyen de forma aleatoria, lo que permite realizar llamadas de consenso precisas con una cobertura creciente o con múltiples pasadas alrededor de la misma plantilla, las llamadas secuencias de consenso circulares (Carneiro et al., 2012; Koren et al., 2012). Al evitar la amplificación clonal, la secuenciación SMRT es también mucho menos sensible al contenido de la secuencia GC que otras plataformas (Loomis et al., 2013). La secuenciación SMRT es particularmente útil para proyectos que implican el ensamblaje de pequeños genomas bacterianos y virales, así como el acabado de grandes genomas (English et al., 2012). La reconstrucción de la variación estructural (VS) en el genoma (Chaisson et al., 2015) y el uso de isoformas en el transcriptoma (Sharon et al., 2013) son también áreas clave en las que la secuenciación SMRT tiene claras ventajas sobre las tecnologías de lectura corta. Sin embargo, el menor rendimiento y los mayores costes de secuenciación por base limitan actualmente el alcance de la mayoría de los estudios de genoma. Además de proporcionar lecturas largas e imparciales, la reacción de polimerización se monitoriza en tiempo real, lo que permite recoger datos relativos tanto a la composición de las bases como a la cinética enzimática. Se producen distintos perfiles cinéticos a medida que la polimerasa encuentra varios tipos de metilación del ADN (Figura 3.3.) (Flusberg et al., 2010). Los instrumentos de secuenciación SMRT no se limitan a estudiar únicamente el ADN, ya que otras moléculas, como los ribosomas, pueden anclarse al fondo de la ZMW y monitorizarse con una resolución de una sola molécula (Uemura et al., 2010).

DESCRIPCIÓN DE LAS TECNOLOGÍAS

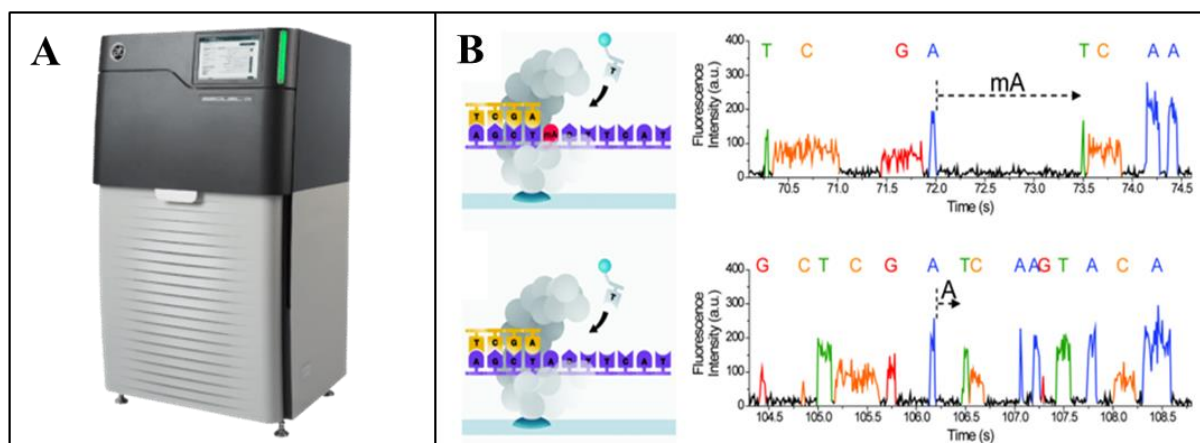


FIGURA 3.3. Equipo de secuenciación Sequel II (A) y principio y ejemplo de detección de la metilación del ADN durante la secuenciación SMRT (B). Fuente: Flusberg et al., 2010.

Oxford Nanopore Technologies.

La secuenciación basada en nanoporos es una estrategia emergente de una sola molécula que ha avanzado considerablemente en los últimos años, y Oxford Nanopore Technologies lidera el desarrollo y la comercialización de este método. La secuenciación por nanoporos puede adoptar diversas formas, pero se basa principalmente en la transición del ADN o de nucleótidos individuales a través de un pequeño canal (Wang et al., 2015). En la tecnología actual de Oxford Nanopore, una celda de flujo de secuenciación comprende cientos de micropozos independientes, cada uno de los cuales contiene una bicapa sintética perforada por nanoporos biológicos. La secuenciación se lleva a cabo midiendo los cambios característicos de la corriente que se inducen a medida que las bases son enhebradas a través del poro por una proteína de motor molecular. La preparación es mínima, ya que implica la fragmentación del ADN y la ligadura de los adaptadores. Esta metodología puede realizarse con o sin amplificación por PCR. El primer adaptador se une con una enzima motora propia, así como con un anclaje molecular, mientras que el segundo adaptador es un oligonucleótido en forma de horquilla que se une con una segunda proteína motora denominada HP (Quick et al., 2014). Este diseño permite secuenciar ambas cadenas de ADN a partir de una única molécula, lo que aumenta la precisión (Ashton et al., 2015; Quick et al., 2014).

El primer dispositivo comercialmente disponible es el MinION, un secuenciador portátil alimentado por USB, que Oxford Nanopore Technologies lanzó a principios de 2014 como parte de un programa de acceso temprano. Como ocurre con todas las metodologías de secuenciación de una sola molécula, las tasas de error son elevadas, sin embargo, las lecturas

DESCRIPCIÓN DE LAS TECNOLOGÍAS

de MinION se han utilizado con éxito para determinar la posición y la estructura de una isla de resistencia bacteriana en combinación con lecturas derivadas de Illumina (Ashton et al., 2015) y para resolver una brecha de ensamblaje en Xq24 humano (Jain et al., 2015). Es la tecnología junto con la SMRT que más avances presenta y la más prometedora por su versatilidad.

La Tabla 3.2. recoge la comparativa de las tecnologías de secuenciación masiva que existen hasta ahora, en función del tamaño de lectura, tiempo de ejecución y tasa de error.

TABLA 3.2. Comparativa de las tecnologías. Fuente: Elaboración propia.

	Tecnología de secuenciación	Longitud de lectura	Tasa de error	Tipo de instrumento	Tiempo de ejecución
Illumina	Secuenciación por síntesis	Lecturas cortas	Bajo	Sobremesa	2-29h
Ion Torrent	Secuenciación por síntesis	Lecturas cortas	Bajo	Sobremesa	2-4h
Pacific Biosciences	Secuenciación de una sola molécula por síntesis	Lecturas largas	Alto	Gran escala	0,5-4h
Oxford Nanopore Technologies	Molécula única	Lecturas largas	Alto	Portátil	1min-48h

4. SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

4. SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS.

4.1. Aplicaciones actuales.

La WGS de patógenos microbianos se ha introducido en la vigilancia de la salud pública con relativa rapidez en comparación con los avances metodológicos anteriores, con informes de su uso por parte de los primeros adoptantes desde aproximadamente 2011 (Koser et al., 2012; Lienau et al., 2011). Aunque inicialmente se utilizó para los análisis retrospectivos de brotes de enfermedades transmitidas por los alimentos detectados por tecnologías de tipificación como la electroforesis en gel de campo pulsado (PFGE), la WGS de patógenos microbianos se ha introducido ahora para la vigilancia prospectiva de patógenos bacterianos transmitidos por los alimentos en al menos cuatro países: Reino Unido, Dinamarca, Francia y Estados Unidos (Allard et al., 2016; Ashton et al., 2016; Jackson et al., 2016; Kvistholm Jensen et al., 2016; Moura et al., 2016). El año siguiente a la implementación de WGS para la vigilancia de evaluación prospectiva de la listeriosis en los Estados Unidos, se detectaron más brotes y más pequeños, los brotes se detectaron antes, la fuente de los brotes se identificó con más frecuencia y el número total de casos relacionados con los brotes identificados aumentó (Jackson et al., 2016).

En el ámbito de la salud pública, la WGS se está introduciendo como una tecnología de sustitución, es decir, sustituirá a la mayoría de los métodos actuales de identificación y caracterización en el laboratorio de microbiología, como el serotipado, el perfil de virulencia, la determinación de la resistencia a los antimicrobianos y los métodos de tipificación molecular anteriores. En un entorno de salud pública, la sustitución del conjunto de métodos tradicionales de identificación y tipificación microbiológica por un único flujo de trabajo analítico eficiente de WGS hace que la implementación sea rentable, además de proporcionar a la salud pública datos más precisos y procesables que los recogidos anteriormente (Figura 4.1.1.) (Grant et al., 2018).

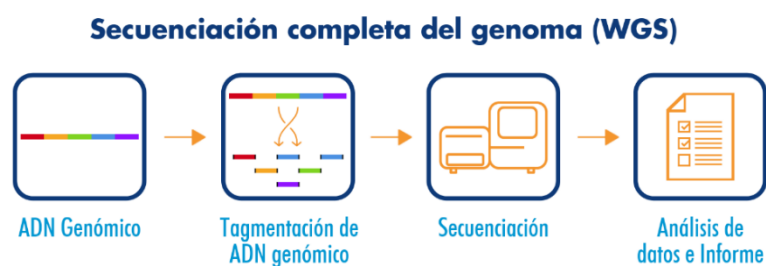


FIGURA 4.1.1. Secuenciación completa del genoma, WGS. Fuente: Merieux Nutrisciences.

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

Siguiendo el ejemplo del sector de la salud pública, cada vez se considera más la posibilidad de aplicar la WGS en la industria alimentaria. Esto no sólo se debe a la necesidad de comprender los planteamientos de la salud pública, sino también a las enormes ventajas y promesas de mejora de la calidad y la seguridad de los alimentos que ofrece esta tecnología. Un beneficio clave e inmediato para la industria alimentaria es la mejora del análisis de la causa raíz en un caso de contaminación por patógenos o deterioro. Por ejemplo, el WGS puede ayudar a distinguir entre introducción nueva y recurrente de un organismo en el entorno de producción. También puede utilizarse para predecir rasgos como la virulencia o la resistencia antimicrobiana de un patógeno o la capacidad de un organismo de deterioro para romper las barreras de conservación de un producto. Si bien, las pruebas de seguridad alimentaria de la industria no exigen la caracterización microbiana detallada que requieren los laboratorios de referencia, el WGS se está explorando cada vez más para rastrear la fuente de contaminación microbiana (Rantsiou et al., 2017; Hoorde y Butler, 2018). A medida que el coste de la secuenciación disminuye con las mejoras tecnológicas, hace más factible que la industria se plantee incorporar su uso.

4.2. Los principios del rastreo basados en el WGS.

Los métodos de subtipificación molecular han demostrado ser muy valiosos para el seguimiento y la localización de patógenos a lo largo de la cadena alimentaria, ayudando a identificar las fuentes de infección y la ruta de transmisión. Esto incluye cuando la fuente de infección se debe a la consecuencia de una mala práctica del manipulador de alimentos, ya que la tipificación molecular puede mostrar que los aislados de los casos, el manipulador de alimentos o el entorno de los servicios alimentarios proceden de una fuente común. La información adicional disponible a través de WGS mejora en gran medida nuestra capacidad para determinar la fuente de infección. En tiempo, las bacterias acumulan cambios en su ADN y esto puede utilizarse para medir su evolución. Mientras que los anteriores métodos de subtipificación molecular detectan los cambios de secuencia en una pequeña porción del genoma microbiano, la WGS los capta en todo el genoma y, por tanto, describe con mayor precisión el parentesco genético de las cepas. En el rastreo y la localización, se evalúa el parentesco de las secuencias bacterianas de los brotes y de la cadena de producción de alimentos para determinar si podrían formar parte de la misma cadena de transmisión (Gerner-Smidt et al., 2013).

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

En la actualidad existen dos enfoques principales para analizar los datos genómicos con el fin de determinar el parentesco entre cepas; los enfoques basados en SNP y los basados en genes. El análisis de los datos WGS mediante cualquiera de los dos enfoques es un proceso complejo en el que se combinan múltiples pasos para producir resultados finales, como matrices de SNP o de alelos y árboles filogenéticos (Timme et al., 2017). La gran cantidad de datos generados en WGS conlleva retos para su análisis (Deurenberg et al., 2017; Wyres et al., 2014). Esto ha llevado a que se desarrollen múltiples soluciones de software, principalmente a través de esfuerzos académicos, que en general requieren conocimientos especializados y experiencia para desplegar y ejecutar. Sin embargo, recientemente se han desarrollado programas informáticos comerciales que aportan una interfaz fácil de usar, lo que permite a los no expertos en bioinformática, con la formación adecuada tanto en el software bioinformático como en la interpretación de los resultados finales de WGS, llevar a cabo los análisis. El software comercial puede ser caro, pero dado que se necesitan pocos conocimientos bioinformáticos, puede ser una solución más rentable para muchos usuarios de la industria alimentaria.

La bioinformática es un recurso útil para el análisis de datos de HTS basados en software libre con múltiples aplicaciones, que aplica las matemáticas, la estadística y la computación al estudio biológico, por tanto, requiere de unos conocimientos básicos en lenguajes de programación (Bash, Perl, Python y R, entre otros) y, preferentemente, también en el uso de sistemas operativos basados en UNIX®, como Linux y MacOS®. Al ser de licencia y código libres, Linux se ha convertido en la opción de preferencia para los investigadores. Dos herramientas bioinformáticas han popularizado el análisis de la microbiota presente en una muestra: Mothur y QIIME, que recientemente han lanzado su segunda versión (Hernández et al., 2019).

4.2.1. El enfoque del SNP.

SNP (*Single Nucleotide Polymorphism*) es un polimorfismo de un único nucleótido que puede interrumpir la función de un gen o bien ser fuente de variación del ADN y usarse como marcador. En el enfoque basado en los SNP, las lecturas de secuenciación se alinean o mapean con un genoma de referencia secuenciado conocido, y se determinan las diferencias de nucleótidos en las regiones codificantes y no codificantes (Davis et al., 2015). Para cada aislado, cada SNP relativo al genoma de referencia se registra y luego se utiliza para

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

cuantificar el parentesco genético entre las cepas. La selección del genoma de referencia es un paso crítico, ya que debe estar completamente secuenciado, es decir, lo más contiguo posible, y estar estrechamente relacionado con los genomas que se analizan (por ejemplo, el mismo serotipo). Un genoma de referencia lejanamente relacionado puede dar lugar a una subestimación del parentesco genético de los aislados investigados, ya que aumenta la probabilidad de un mapeo erróneo y disminuye las regiones a las que se pueden mapear las lecturas (Figura 4.2.1.1) (Carrico et al., 2018; Schürch et al., 2018). La variación de los elementos genéticos móviles, como los plásmidos y los profagos, no se limita, por definición, a la herencia vertical y, por tanto, no siempre refleja la verdadera historia evolutiva entre cepas y por lo que no es un indicador fiable de la relación epidemiológica. Las regiones repetitivas, como los profagos y los elementos de inserción, a menudo se excluyen implícitamente debido a un mapeo ambiguo (es decir, las lecturas de secuenciación pueden mapearse en múltiples lugares en el genoma de referencia y, por lo tanto, se ignoran) o explícitamente mediante el enmascaramiento de las regiones de alta densidad de SNP. A pesar de estas exclusiones, el análisis de SNP suele realizarse utilizando más del 95% del genoma secuenciado.

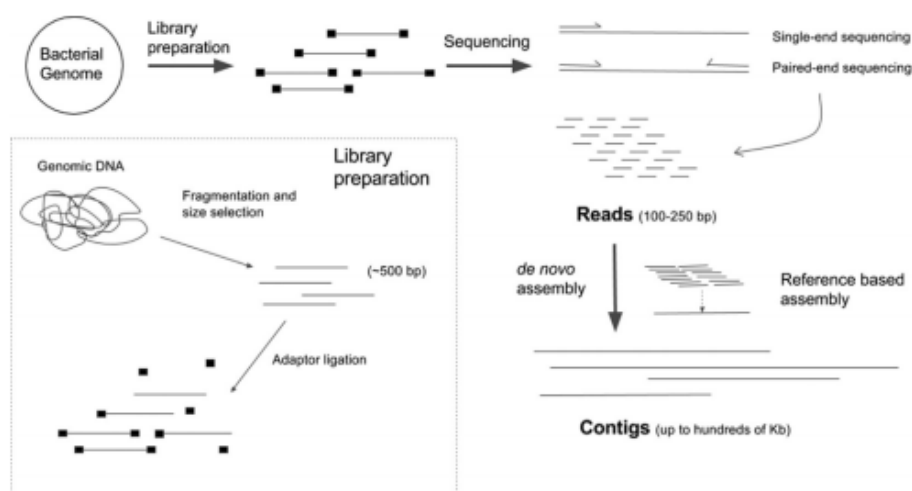


FIGURA 4.2.1.1. Conceptos de secuenciación de genomas bacterianos a contigs. Fuente: Carrico 2018.

El número de diferencias de SNP puede variar dependiendo de la cepa de referencia, el mapeo de referencia, así como el método de llamada de SNP utilizado (Pightling et al., 2015). Existen varias herramientas de análisis de SNP de dominio público que se están desarrollando activamente, además de otras nuevas que están apareciendo en línea. Esto hace que sea difícil

compararlas, sobre todo porque no se han desarrollado directrices o normas ampliamente aceptadas para seleccionar herramientas de análisis de SNP. Se recomienda a los usuarios que utilicen herramientas basadas en SNP previamente validadas, como las desarrolladas por la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA), los Centros para el Control y la Prevención de Enfermedades (CDC), el Centro de Salud Pública de Inglaterra (PHE) y el Centro de Epidemiología Genómica (CGE) que están disponibles en plataformas de código libre gratuitas y realizar una verificación interna, idealmente utilizando conjuntos de datos de referencia que están cada vez más disponibles (Timme et al., 2017).

4.2.2. Enfoque genético.

El análisis gen por gen consiste en evaluar la variación en las regiones codificantes, es decir, los genes de un genoma bacteriano (Maiden et al., 2013). En una extensión de la tipificación tradicional de secuencias de 7 loci, Multilocus Sequence Typing (MLST), los genes de un genoma central definido, core genome Multi-Locus Sequence Typing (cgMLST), o de todo el genoma, whole genome Multi-Locus Sequence Typing (wgMLST), que incluye los genes accesorios más variables, se comparan con una base de datos de referencia de todas las variantes genéticas conocidas (alelos) para una especie concreta. Cada secuencia de genes o alelos se reduce a un número y los genomas se comparan en función del número de diferencias alélicas que hay, de forma comparable a como se utiliza el número de diferencias de SNP. Dado que la referencia es una base de datos de loci y alelos de numerosas cepas, el análisis no depende de la selección de una cepa de referencia estrechamente relacionada para la evaluación precisa del parentesco de aislados genéticamente similares. A menudo, antes del análisis gen por gen, las lecturas de secuenciación se ensamblan, normalmente utilizando el enfoque basado en *de novo*, en secuencias contiguas más largas (llamadas contigs) que constituyen un borrador del genoma (es decir, uno que todavía contiene huecos). Para asignar un tipo de MLST, las lecturas cortas ensambladas se comparan mediante BLAST con una base de datos de alelos de referencia (esquema MLST) que contiene todas las variantes alélicas conocidas para cada locus definido para una especie. Las variaciones, incluidos los SNP, los indels (inserciones y deleciones) y las recombinaciones en el mismo gen se consideran como una única diferencia alélica. En algunas líneas de MLST, la llamada de alelos se completa con la llamada de alelos sin ensamblaje, en la que las lecturas de secuenciación en bruto se asignan a los alelos en una base de datos. La elección del

ensamblaje o de la llamada de alelos sin ensamblaje suele depender de si ya existe un ensamblaje *de novo* o si las lecturas se han mapeado a un genoma de referencia. Una evaluación de diferentes softwares MLST para datos de secuenciación NGS ha sido realizada utilizando un conjunto de datos validados que proporciona información sobre la precisión, las limitaciones y el rendimiento computacional (Page et al., 2017).

4.2.3. Análisis filogenético.

La variación genética detectada por el análisis SNP o gen por gen puede utilizarse para inferir las relaciones filogenéticas entre los aislados bacterianos y esto se suele mostrar en forma de árbol filogenético. El árbol representa el modelo evolutivo calculado (obtenido mediante diferentes algoritmos de inferencia de árboles posibles, como la parsimonia, la máxima verosimilitud y los métodos bayesianos o de distancia) de los aislados como una serie de ramas a partir de la raíz o ancestro común. Los aislados agrupados cerca de las hojas del árbol están más estrechamente relacionados que otros aislados en el resto del árbol. Diversos autores (Ajawatanawong, 2017; Baldauf, 2003; Hedge y Wilson, 2016; Yang y Rannala, 2012) han realizado algunas revisiones acerca de los principios que sustentan la construcción e interpretación de los árboles filogenéticos, observando que pueden utilizarse para identificar las relaciones evolutivas entre organismos, así como las secuencias de genes. Para analizar una vía evolutiva, hay que partir de secuencias ortólogas y realizar el análisis correctamente. Sin embargo, las filogenias de un solo gen suelen tener menos señal evolutiva.

4.2.4. Comparación entre el SNP y el cg/wgMLST.

La elección del enfoque genómico comparativo a utilizar depende de las necesidades del usuario final y del contexto epidemiológico. Mientras que los enfoques SNP o gen por gen pueden utilizarse para investigar un número fijo de cepas asociadas a un evento de contaminación concreto, el cgMLST podría ser más apropiado si varios usuarios necesitan analizar sistemáticamente cada nuevo aislado añadido a una base de datos común, por ejemplo, en una red de vigilancia de brotes, especialmente si la información de la secuencia de secuencias no puede ser de dominio público. Para investigar la filogenia, el uso de cgMLST o cgSNP puede proporcionar análisis más robustos que wgMLST o wgSNP, ya que incluye sólo regiones del genoma presentes en todas las cepas, sin embargo, el uso de

wgMLST o wgSNP puede dar una mayor resolución para la discriminación de cepas. Los enfoques SNP y gen por gen evalúan la variación genética de formas ligeramente diferentes y deben considerarse complementarios y utilizarse ambos cuando un método por sí solo no proporciona una respuesta clara o para respaldar con más fuerza una asociación entre aislados, por ejemplo, para confirmar el origen de un brote y respaldar la acción reguladora. Hasta la fecha, ambos métodos han demostrado ser igual de discriminatorios a la hora de determinar el parentesco entre cepas y epidemiológicamente concordantes para las investigaciones de brotes (Chen et al., 2017; Cunningham et al., 2017; Katz et al., 2017).

4.3. Interpretación de los resultados.

La interpretación biológica del parentesco genético de los aislados mediante los datos de la secuencia suele ser sencilla, siempre que todos los parámetros de control de calidad de la secuencia estén dentro de los valores esperados y se conozca la estabilidad genética de las bacterias en cuestión, por ejemplo, sus tasas de mutación espontánea. En el análisis WGS, el número de diferencias SNP/alelo se utiliza para construir árboles filogenéticos que proporcionan información sobre la historia evolutiva de los aislados. En un sentido biológico, una alta similitud de secuencia por el análisis WGS significa que los aislados comparten un ancestro común reciente, y una baja similitud significa que no lo hacen (Pightling et al., 2018). Es un supuesto fundamental en la epidemiología molecular que la filogenia refleja el parentesco epidemiológico, es decir, los aislados clínicos o los aislados clínicos y alimentarios o ambientales que están estrechamente relacionados filogenéticamente es probable que estén epidemiológicamente o causalmente relacionados (Besser et al., 2018). Aunque esta suposición suele ser cierta, no siempre lo es debido a las conexiones complejas o indirectas que pueden darse en cualquier punto del continuo de la granja a la mesa. Por lo tanto, es fundamental que las pruebas epidemiológicas y de rastreo de alimentos se utilicen para apoyar y facilitar la correcta interpretación de los análisis de WGS. El análisis WGS proporciona pruebas sólidas de que los aislados están relacionados genéticamente, pero no significa necesariamente que un caso clínico se haya infectado directamente a partir de un alimento o de un local concreto en el que se hayan obtenido aislados coincidentes con el WGS. Es esencial que se disponga de pruebas epidemiológicas para apoyar los hallazgos filogenéticos, determinar el vehículo alimentario, la fuente original de contaminación y el modo de transmisión.

4.4. La necesidad de estandarización.

Para maximizar los beneficios de la WGS, los datos generados deben ser precisos, fiables y globalmente comparables, independientemente de la plataforma de secuenciación, el enfoque bioinformático y el software utilizado. La estandarización es el proceso por el que se consigue esto y, mientras que existen normas y directrices para la secuenciación genética humana, hay pocas disponibles para la WGS microbiana. Esto se debe principalmente a que la genómica de patógenos es un campo en rápido desarrollo y comprende especialidades, como la bioinformática, que no han sido sometidas a los procedimientos de normalización del laboratorio de microbiología anteriormente. Sin embargo, muchos de los principios y prácticas de calidad desarrollados para la secuenciación humana son igualmente aplicables al análisis WGS microbiano (Gargis et al., 2016) y los criterios y estándares de rendimiento específicos de WGS microbiano se están haciendo disponibles (Kozyreva et al., 2017; Portmann et al., 2018). Al igual que los métodos de subtipificación microbiana a los que sustituye, la WGS microbiana requiere validación y verificación y debe someterse a todos los procedimientos de garantía de calidad que constituyen un buen sistema de gestión de la calidad del laboratorio.

El flujo de trabajo de la WGS consta de tres componentes: la preparación de las muestras, la secuenciación y el análisis de los datos y todo el proceso, de principio a fin, debe ser validado con respecto a los métodos de tipificación existentes, por ejemplo, PFGE o el análisis de repetición en tándem de número variable de múltiples focos (MLVA), con un conjunto bien definido de cepas para garantizar que el método funciona para el propósito previsto por el usuario final; esto también facilita la generación de directrices interpretativas para la interpretación coherente de los resultados. La validación establece especificaciones de rendimiento como la exactitud, la precisión, la reproducibilidad, la repetibilidad, la sensibilidad y la especificidad, así como la capacidad discriminatoria y la concordancia epidemiológica. Los procedimientos de control de calidad son necesarios para todos los componentes del proceso de WGS, incluyendo la calidad y cantidad de ADN de la muestra, las puntuaciones de calidad de la secuencia, incluyendo la profundidad de la cobertura de la secuencia, la longitud de la lectura y la calidad de la secuencia, así como el uso de controles de muestra positivos y negativos conocidos. Al igual que con otros componentes de WGS, el proceso de análisis bioinformático, una vez optimizado, es necesario controlar la versión y cualquier modificación posterior requerirá algún tipo de revalidación. Una vez validado todo el proceso de WGS, es necesario realizar una evaluación periódica e independiente de su

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

rendimiento, es decir, una verificación, que puede lograrse mediante el uso de controles de calidad internos, controles de calidad externos y la participación en pruebas de aptitud. Se realizan esquemas que proporcionan cepas bacterianas para las pruebas de extremo a extremo, el ADN extraído para la secuenciación y la evaluación del análisis de datos y los datos de la secuencia de la misma cepa para el análisis bioinformático.

Otras iniciativas de calidad incluyen actividades de evaluación comparativa en las que se dispone de conjuntos de cepas bien caracterizados para evaluar el rendimiento de los procesos bioinformáticos. También se ha trabajado en el marco del proyecto Engage, financiado por la EFSA (<http://www.engage-europe.eu>), para evaluar herramientas bioinformáticas específicas. Se ha utilizado un conjunto estándar de datos de secuenciación para evaluar diferentes herramientas *de novo* para predecir los serotipos de *Salmonella* spp, así como herramientas de perfilado de genes de resistencia a los antimicrobianos. Los resultados de estos estudios de evaluación comparativa demuestran que el serotipado y la predicción de la resistencia a los antimicrobianos en *Salmonella* utilizando datos de WGS es una opción muy factible.

4.5. Salud pública y medidas reglamentarias basadas en los resultados del WGS.

Cada vez más, la ciencia de los alimentos y la salud pública controlan las bases de datos de secuencias para identificar los aislados indistinguibles de los pacientes, la cadena alimentaria y los aislados clínicos agrupados que podrían indicar un brote de origen alimentario. Estos hallazgos justifican la exploración de la posible relación entre los casos y el o los aislados alimentarios. El uso de un perfil WGS como parte de la definición de casos en la investigación de un brote permite descartar casos dentro o fuera del brote con un grado de resolución mayor que el que era posible anteriormente. Las pruebas de WGS de que los aislados son la misma cepa están permitiendo atribuir los casos a brotes durante períodos de tiempo más largos y vincular los casos de áreas geográficas más amplias de lo que era posible con los métodos de tipificación anteriores, por ejemplo, se puede demostrar que los aislados de *L. monocytogenes* de los casos de listeriosis ocurridos durante varios años son la misma cepa (Chen et al., 2017; Gillesberg Lassen et al., 2016; Kleta et al., 2017; Wilson et al., 2016); Se ha demostrado que los aislamientos de *Salmonella enteritidis* procedentes de casos en diferentes países europeos son los mismos mediante el análisis de SNP y que han evolucionado a partir de un ancestro común (Dallman et al., 2016). Una definición de caso

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

más robusta da mayor potencia a los análisis epidemiológicos posteriores, como los estudios de casos y controles, ya que los casos no relacionados que pueden haberse incluido previamente como parte del brote, ya no confunden los análisis (Lienau et al., 2011). Los datos de secuencias de los aislamientos de los brotes pueden compararse con bases de datos de secuencias conocidas y pueden coincidir con los aislamientos asociados a distintas señales geográficas que pueden dar indicaciones sobre la posible fuente original de contaminación y, por tanto, ayudar a orientar las investigaciones sobre la cadena alimentaria y el medio ambiente (Hoffmann et al., 2016).

El mayor poder del análisis WGS para demostrar un parentesco genético inequívoco proporciona pruebas más sólidas para la adopción de medidas de salud pública y puede permitir la intervención en una fase más temprana. Sin embargo, como se ha reiterado anteriormente, las pruebas epidemiológicas son vitales junto con las pruebas de WGS para garantizar que se adopten las medidas de salud pública y reglamentarias adecuadas.

Los brotes detectados por WGS se investigan utilizando enfoques similares a los utilizados anteriormente, con entrevistas a los casos sobre su exposición a los alimentos y estudios de casos y controles realizados como parte de las investigaciones epidemiológicas analíticas para proporcionar pruebas de apoyo a la posible fuente de alimentos. Las autoridades alimentarias llevarán a cabo investigaciones sobre la trazabilidad de los productos alimentarios implicados para confirmar o refutar su vinculación con el brote y, en caso de que esté vinculada, para identificar la causa fundamental del brote, de modo que puedan aplicarse medidas de control eficaces. Además de las abrumadoras pruebas que el WGS aporta a las investigaciones de brotes, también proporciona apoyo para evitar la asociación de falsos positivos de un alimento a un brote. Cuando se ha identificado un patógeno en un producto alimentario o en el entorno de producción de alimentos, la secuencia del aislado puede compararse con una base de datos de aislados humanos para ver si hay alguna coincidencia. Hasta la fecha, la industria alimentaria se ha puesto en contacto con el Centro de Salud Pública de Inglaterra (PHE) en dos ocasiones distintas para comparar los perfiles WGS de los patógenos con los de los casos de enfermedad humana; en ambas ocasiones no se encontraron coincidencias (Grant, 2018). Sin embargo, independientemente de que se identifique alguna enfermedad humana, la presencia del patógeno en un producto acabado (alimento) o en un entorno crítico de procesamiento de alimentos significa un fallo en los controles preventivos o en las condiciones higiénicas y puede desencadenar una investigación y/o acciones de cumplimiento del regulador (Jagadeesan et al., 2019).

4.6. Gestión de la seguridad alimentaria.

El rastreo preciso de la fuente durante la investigación de un evento de contaminación es una de las principales aplicaciones de la WGS en la gestión de la seguridad alimentaria. Comprender si el agente patógeno o de deterioro detectado es el resultado de un evento de contaminación esporádica o de uno recurrente es esencial para entender la causa raíz de la contaminación y facilitará la aplicación o verificación de las medidas de control. Esto permitirá a la industria centrarse en las áreas prioritarias de intervención, ya sea en la fábrica o en el nivel de los proveedores, y permitirá un seguimiento eficaz para determinar si la acción ha tenido éxito. WGS puede utilizarse para mejorar la gestión de proveedores y materias primas y optimizar los esfuerzos en los programas de verificación de patógenos ambientales. Un mejor análisis de las causas fundamentales permitirá comprender mejor las vías de transmisión e identificar nuevas fuentes de contaminación. Los resultados y las mejoras resultantes en las prácticas de fabricación y cultivo pueden compartirse con todo el sector alimentario, no sólo con la instalación implicada en el suceso de contaminación (Jagadeesan et al., 2019).

Además del cotejo directo de los aislados ambientales en relación con una contaminación, también es posible que la industria compare los aislados con las entradas de las bases de datos públicas utilizadas por las autoridades sanitarias y reguladores de los alimentos. Dependiendo de la base de datos utilizada para la comparación, puede obtenerse información como la identificación de posibles fuentes nuevas, señales geográficas sobre el posible origen de la contaminación y la asociación con enfermedades humanas. La WGS también puede aportar información valiosa para perfeccionar la etapa de "identificación del peligro" en el proceso de evaluación del riesgo microbiano. Los conocimientos existentes sobre los organismos suelen obtenerse estudiando cepas de laboratorio bien caracterizadas que no necesariamente representan la diversidad fenotípica de la población en general. Por ejemplo, Maury et al., 2016, identificaron recientemente nuevos factores de virulencia en *L. monocytogenes* comparando los genomas de cepas clínicas y asociadas a los alimentos. (Yahara et al., 2017) examinaron el impacto de varias etapas de la cadena de producción avícola en las poblaciones de *Campylobacter* spp. utilizando WGS y estudios de asociación del genoma completo (GWAS).

Es probable que muchas disciplinas, como la microbiología alimentaria predictiva y el procesamiento térmico, se beneficien del uso de datos de WGS para predicción fenotípica.

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

Hay una serie de herramientas basadas en la web y bases de datos que identifican los genes de interés alineando los borradores de los genomas con una base de datos de genes. Por ejemplo, los datos del genoma obtenidos mediante la secuenciación rutinaria de los aislados cotidianos pueden consultarse para predecir rasgos como el perfil de virulencia, la resistencia al calor, la respuesta al estrés, la formación de biopelículas, la resistencia a los antimicrobianos y a los biocidas mediante el estudio paralelo de sus características fenotípicas (Rantsiou et al., 2017).

El uso de la WGS para la evaluación de riesgos transmitidos por los alimentos, hasta ahora, esta actividad en el laboratorio se ha basado en métodos fenotípicos y en métodos de subtipificación molecular más antiguos, sin embargo, recientemente se han utilizado datos genómicos para identificar las fuentes probables de infección. Con la relevancia filogenética de la WGS, se pueden hacer inferencias más fiables sobre el origen común y, por tanto, también sobre la fuente de cepas con perfiles WGS similares (Franz et al., 2016). Sin embargo, para lograr esto, es necesario desarrollar nuevos enfoques de modelización que puedan manejar las enormes cantidades de datos de secuencias. Una vez establecida, esta atribución de la fuente se convertirá en una herramienta extremadamente poderosa para identificar las áreas de la producción de alimentos que están asociadas con la mayoría de las enfermedades humanas. Esto ayudará a la industria alimentaria y a otros a dar prioridad a las actividades de seguridad alimentaria que tienen más probabilidades de dar lugar a alimentos más seguros y, por lo tanto, también a reducir la carga de las enfermedades transmitidas por los alimentos.

4.7. Consideraciones sobre la aplicación en la industria.

En el caso de las industrias y los minoristas que cuentan con microbiólogos con formación clásica y recursos limitados para gastar, no sólo hay que tener en cuenta la precisión, sino también la practicidad, la sencillez y el coste de un método antes de aplicar el WGS. Lo ideal es que un método novedoso sea más barato o, al menos, esté a la altura de los que se utilizan actualmente. La ruta más probable para la adopción por parte de la industria es a través de un enfoque de nivel de entrada utilizando cg/wg MLST con WGS de terceros o análisis completos de terceros. Hay varias soluciones comerciales disponibles y algunas tienen tanto MLST cg/wg como análisis SNP en sus líneas de trabajo con el objetivo de identificar grupos primarios utilizando el enfoque MLST y el análisis SNP para confirmar la relación entre los aislados en un grupo. La clave para permitir la adopción de WGS en la

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

aplicación rutinaria es la simplificación del análisis y, sobre todo, la simplificación de los informes finitos. El informe finito del análisis de tipificación WGS debería decir idealmente: coincidencia Sí/No/Tal vez y análisis Éxito/Fracaso, que son parámetros que una persona no especializada puede interpretar. El informe también debería incluir una explicación de los resultados que describiera las advertencias y el razonamiento en el que se basa la interpretación final.

Una consideración importante para la adopción del WGS por parte de la industria es que las pruebas microbiológicas rutinarias de los alimentos no siempre requieren la caracterización detallada que ofrece la secuenciación y que exige la salud pública. Por lo tanto, es más probable que su adopción se produzca en función de las necesidades en lugar de una sustitución total de los métodos existentes. La industria está utilizando cada vez más la WGS para rastrear y localizar el origen de la contaminación; se espera que su éxito en este ámbito, junto con la disminución de los costes de secuenciación, haga que se adopte la WGS. Se espera que su éxito en este ámbito, junto con la disminución de los costes de secuenciación, fomente su uso más amplio.

4.8. Desafíos que deben abordarse.

Aunque la WGS ha revolucionado la tipificación molecular de los patógenos, existen varias lagunas y desafíos científicos que deben ser abordados para mejorar la interpretación de los datos de la WGS y permitir su uso generalizado en la gestión de la seguridad alimentaria para la industria alimentaria, entre ellos:

- Seguir trabajando en la estandarización del protocolo de extremo a extremo para permitir el intercambio y las comparaciones globales de los datos de WGS.
- Investigación para mejorar la comprensión de los aislados indistinguibles de fuentes epidemiológicamente no relacionadas para fortalecer la interpretación de los datos de WGS.
- Investigación del papel de los nichos ambientales en las tasas de mutación de los patógenos para apoyar las nociones de parentesco. Esto mejoraría la interpretación de los datos de WGS, específicamente para desarrollar una orientación sobre los valores de corte de SNP/alelos y también para las cepas que pueden provenir de diferentes entornos y soportar diferentes tasas de crecimiento, pero que deben ser consideradas en una investigación.

SECUENCIACIÓN DEL GENOMA COMPLETO DE LOS AISLADOS

- Exploración del valor del análisis WGS de elementos genéticos móviles (MGE). En general, los MGE se excluyen del análisis WGS aunque es bien sabido que a menudo contribuyen a la virulencia y a la resistencia antimicrobiana.
- La WGS de aislados bacterianos es una tecnología disruptiva, ya que cambia por completo la forma en que tradicionalmente se ha realizado la microbiología, en particular la subtipificación. Esto, junto con los importantes costes analíticos y los requisitos de conocimiento y competencia, son actualmente obstáculos para su uso generalizado por parte de la industria.

5. SECUENCIACIÓN DE AMPLICONES, METAGENÓMICA Y METATRANSCRIPTÓMICA

5. SECUENCIACIÓN DE AMPLICONES, METAGENÓMICA Y METATRANSCRIPTÓMICA.

5.1. Una definición de los términos.

Para sondear las especies y la diversidad funcional de las comunidades microbianas sin necesidad de un cultivo bacteriano se utilizan dos enfoques que emplean tecnologías NGS: la secuenciación de amplicones o metabarcoding, que implica la amplificación y secuenciación de familias de genes marcadores específicos; y la metagenómica, la secuenciación aleatoria de Shotgun de todo el contenido genómico de las comunidades.

Se recomienda utilizar el término 'metabarcoding' cuando se aplican técnicas basadas en amplicones y el término 'metagenómica' solo cuando se aplica la secuenciación Shotgun no dirigida. Ambas técnicas han sido muy exitosas para identificar e investigar microorganismos no cultivables (Cao et al., 2017; Forbes et al., 2017).

5.1.1. Perfil de la comunidad microbiana basada en amplicón (metabarcoding).

Esta tecnología requiere el aislamiento del ADN directamente de muestras que pueden incluir cultivos iniciadores, muestras tomadas durante el procesamiento de la producción, el producto alimentario final y muestras ambientales. El ADN extraído se somete a la amplificación por PCR de los genes marcadores filogenéticos; por lo general, el gen 16S rRNA en el caso de las arqueas y las bacterias y; el gen 18S rRNA o la región ITS (espaciador transcrito interno) en el caso de los eucariotas. La secuenciación masiva en paralelo de estos amplicones genera un conjunto de información de perfiles sobre la microbiota, a menudo compleja, asociada a los productos alimentarios. A continuación, los datos de secuenciación se procesan con pipelines bioinformáticos específicos (descritos en la sección 4.3) para estructurar y anotar esta información bruta en conocimiento.

Una de las ventajas del enfoque de metabarcoding es la capacidad de seguir la sucesión de las poblaciones microbianas a lo largo del tiempo en varios niveles taxonómicos. Por ejemplo, el oligotipado permite diferenciar taxones microbianos estrechamente relacionados mediante datos de la secuencia del gen 16S rRNA (Eren et al., 2013). En comparación con la secuenciación aleatoria de shotgun (metagenómica), el metabarcoding

proporciona una visión general rentable de la composición taxonómica de una muestra y ya se ha aplicado a diversos productos alimentarios.

5.1.2. Perfil del microbioma metagenómico

La metagenómica genera información de secuenciación del material genético de una muestra, permite la identificación de cepas individuales y puede permitir la predicción de las funciones codificadas por las comunidades microbianas. Este enfoque ya ha permitido medir los niveles de diversidad de la población *in situ* (Baker et al., 2006; Venter et al., 2004) y determinar las familias de genes específicas o enriquecidas en un hábitat (Tyson et al., 2004).

La metagenómica también se está explorando para la detección, identificación y caracterización de patógenos en los alimentos (Aw et al., 2016; Leonard et al., 2015, 2016) y en el entorno de la cadena alimentaria (Yang et al., 2016). Aunque se han notificado límites de detección bajos para los patógenos bacterianos introducidos en los alimentos, esto se produce después de varias horas de enriquecimiento basado en cultivos, junto con una gran profundidad de secuenciación para garantizar la captura de la diversidad genómica dentro de la muestra (Sekse et al., 2017). Sin embargo, la metagenómica ofrece la oportunidad de estudiar la diversidad de los microorganismos de una muestra de forma menos sesgada que la metabarcodificación y se está utilizando para mejorar los métodos de enriquecimiento basados en cultivos (Forbes et al., 2017). La metagenómica shotgun puede proporcionar una visión valiosa y rápida de la presencia de marcadores genéticos especificando especies, serotipos, genes de virulencia y resistencia antimicrobiana (AMR), etc. El futuro desarrollo de la bioinformática en la metagenómica hace posible la investigación de la filogenia (Ottesen et al., 2016; Truong et al., 2017).

5.2. La meta-ómica para la caracterización funcional de los microbiomas.

El campo de la ómica ambiental (o meta-ómica) ha ampliado drásticamente nuestros conocimientos sobre las comunidades microbianas (Waldor et al., 2015), impulsando un cambio de paradigma en el que se considera la comunidad microbiana completa en lugar de las especies individuales. La importancia de las interacciones ecológicas entre los microorganismos se reconoce ahora y debe incluirse en un marco global para seguir desarrollando modelos de la función de los ecosistemas comunitarios (Raes y Bork, 2008). La

metagenómica por sí sola es un potente enfoque para caracterizar las comunidades microbianas, pero tiene un potencial aún mayor cuando se combina con otras tecnologías "ómicas" complementarias, como la medición de la expresión de ARNm (metatranscriptómica), la detección y categorización de proteínas (proteómica) y la concentración de metabolitos (metabolómica) (Warnecke y Hugenholtz, 2007). El término "food-omic" se ha acuñado para referirse a la aplicación de las tecnologías 'ómicas en el procesamiento de alimentos, la nutrición y la seguridad alimentaria (Cifuentes, 2009). En particular, la combinación de la metagenómica y la proteómica tiene un gran potencial para el estudio de la producción de alimentos, la evaluación de la seguridad, la autenticidad y la calidad de los mismos (Josic et al., 2017). Es posible utilizar métodos proteómicos basados en la espectrometría de masas (EM) para evaluar la abundancia de proteínas y la distribución de las funciones metabólicas en las comunidades microbianas naturales (Ram et al., 2005). Sin duda, la traslación de las tecnologías 'ómicas a la microbiología de los alimentos tendrá un impacto importante en la industria alimentaria (Brown et al., 2017; Walsh et al., 2017). Cabe destacar que los avances en biología computacional que permiten describir los genomas ambientales y su expresión *in situ* han acompañado a estas nuevas tecnologías (Segata et al., 2013).

5.3. Herramientas computacionales para la caracterización de los microbiomas.

La mayoría de los procesos bioinformáticos comienzan con la limpieza y el filtrado de calidad de las secuencias obtenidas tras la secuenciación, antes de su agrupación en unidades taxonómicas operativas (OTU), normalmente con una similitud del 97% (Konstantinidis y Tiedje, 2005). Los pipelines elaborados para software libre como othur (Schloss et al., 2009) y QIIME 2 (<http://qiime.org/>; Caporaso et al., 2010) realizan todo el análisis desde las secuencias brutas hasta las matrices de abundancia de OTUs. La delimitación de las OTUs es útil para detectar lineajes distintos, estimar la diversidad y evaluar la estructura de la comunidad microbiana.

Los análisis de co-ocurrencia y correlación aplicados a los datos de metabarcodificación y metagenómica se utilizan cada vez más para la predicción de las interacciones entre especies y el análisis de las estructuras de las comunidades microbianas (Faust y Raes, 2012). Actualmente se dispone de una variedad de herramientas para reconstruir las redes ecológicas y los análisis de redes están revelando especies clave

inesperadas que participan en las funciones clave de los ecosistemas a nivel mundial (Guidi et al., 2016).

Para los análisis bioinformáticos genómicos/metatranscriptómicos existen varios pipelines para el preprocesamiento, el ensamblaje, la agrupación y los análisis, como MOCAT2 (Kultima et al., 2016), MetAMOS (Treangen et al., 2013) e IMP (Narayanasamy et al., 2016) como marcos independientes y MG-RAST (Wilke et al., 2016) y Anvi'o (Eren et al., 2015b) como plataformas basadas en la web. Para las anotaciones funcionales de datos meta-ómicos, las bases de datos más utilizadas siguen siendo KEGG (Kanehisa et al., 2017), COG (Huerta-Cepas et al., 2016), Pfam (Finn et al., 2016), LDA Effect Size (LefSe) (Segata et al., 2010) y PICRUST (Langille et al., 2013) para las clasificaciones funcionales. Entre las técnicas de bases de datos de asignación taxonómica destacan SILVA, para bases de datos de ARN ribosómico de alta calidad o GreenGenes, que permite el acceso y descarga de las secuencias registradas. Por último, las plataformas bioinformáticas que implementan flujos de trabajo completos, como Galaxy (Afgan et al., 2016; Bornich et al., 2016) y EDGE (Li et al., 2017) permiten el desarrollo y el despliegue de pipelines personalizados y adaptados a las necesidades de los investigadores. Se necesitan conocimientos bioinformáticos amplios para utilizar estas herramientas e interpretar los resultados obtenidos, aunque existen opciones de personalización y soluciones comerciales para simplificar estos pasos y hacerlos más accesibles.

5.4. Aplicaciones de la metagenómica en la seguridad alimentaria.

Grupos como el Consorcio para la Secuenciación de la Cadena de Suministro de Alimentos, fundado por IBM y Mars Incorporated, están dedicando esfuerzos a la recopilación de información sobre el genoma de las bacterias patógenas a lo largo de la cadena de suministro de alimentos, así como a la caracterización y cuantificación del microbioma antes y después del procesamiento para utilizar los datos genómicos y metagenómicos para garantizar la seguridad, la autenticidad y la trazabilidad de los alimentos. La información sobre la secuencia de ADN y ARN recogida de muestras de alimentos por el CSFSC se utilizará para describir una línea de base microbiana que represente a las comunidades microbianas normales, que puede aplicarse para rastrear la fuente de contaminación y para la autenticación de alimentos. Utilizando datos de la investigación del CSFSC, IBM está desarrollando un banco de trabajo bioinformático

escalable basado en la web, el Metagenomics Computation and Analytics Workbench, diseñado para analizar datos de secuencias metagenómicas y metatranscriptómicas para evaluar los peligros microbiológicos y para la autenticación de alimentos en la cadena de suministro. El trabajo realizado hasta la fecha en el CSFSC y su herramienta bioinformática MCAW relacionada ofrecen un modelo de recopilación de bases de datos genómicos y metagenómicos de alta calidad, así como un banco de trabajo bioinformático que puede llegar a aplicar la NGS a la seguridad alimentaria. Los proveedores de servicios más pequeños están aplicando enfoques similares, con el objetivo de utilizar la NGS para caracterizar los patógenos en los ingredientes y productos alimentarios. Estos estudios y esfuerzos combinados podrían aportar una nueva perspectiva sobre la evaluación del riesgo microbiológico y una base para las estrategias de mitigación, así como implicaciones relacionadas con las normas actuales de gestión de la seguridad alimentaria. También se está utilizando NGS para conocer la microbiota de multitud de matrices alimentarias a lo largo del proceso de producción, como por ejemplo en el estudio de Filipis et al., 2014, en el que el fin era proporcionar información sobre la microbiota implicada en la maduración de quesos. Además, se llevó a cabo la secuenciación de alto rendimiento independiente del cultivo de los amplicones del gen *lacS* para evaluar la biodiversidad que se produce dentro de la especie *S. thermophilus* (Filipis et al., 2014).

5.5. Problemas y desafíos.

La evaluación del repertorio funcional completo de una población microbiana sigue siendo difícil debido a la naturaleza incompleta de la anotación funcional de genes o proteínas individuales en las bases de datos públicas. En los últimos años, se han utilizado categorías funcionales detalladas presentes en las bases de datos KEGG (Kanehisa et al., 2017) y SEED (Aziz et al., 2012; Overbeek et al., 2005) para anotar y comparar genomas y metagenomas mediante los sistemas KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007), Metagenomics Rapid Annotation using Subsystem Technology (Wilke et al., 2016) y Metagenome Analyzer (Huson et al., 2007, 2016). Sin embargo, estas categorías funcionales suelen ser amplias y no permiten distinguir las características metabólicas y fisiológicas. Se necesitan nuevas herramientas para caracterizar posibles vías fisiológicas y metabólicas (De Filippo et al., 2012), como el sistema MAPLE (Takami et al., 2016), que utiliza anotaciones

SECUENCIACIÓN DE AMPLICONES, METAGENÓMICA Y METATRANSCRIPTÓMICA

de módulos KEGG y permite estimar la abundancia funcional e indica la probabilidad de funcionamiento del módulo KEGG en función de los resultados de la tasa de finalización.

Al igual que con los métodos microbiológicos tradicionales, el muestreo es un primer paso extremadamente importante para recopilar información microbiológica relevante del entorno de procesamiento de alimentos y de los productos finales (Comisión Internacional de Especificaciones Microbiológicas para los Alimentos y Christian y Roberts, 1986; Ni et al., 2013). La diversidad de tipos de muestras se reflejará en las variaciones de las densidades celulares, la viabilidad celular y la presencia de biofilms. Lamentablemente, la gran variedad de matrices en la producción de alimentos no permite una solución única para todos. Por lo tanto, es necesario diseñar esquemas de muestreo específicos para cada proceso y producto. La interpretación errónea de los resultados, especialmente en muestras que contienen un bajo número de células microbianas, puede deberse a la contaminación que puede provenir de los reactivos utilizados para la extracción de ADN (Biesbroek et al., 2012). El ADN de las células muertas también puede dar una falsa impresión de la carga microbiana en un producto alimentario o en un entorno de procesamiento. El precultivo puede utilizarse para el enriquecimiento de células viables. Sin embargo, esto debe tener en cuenta los microorganismos que requieren condiciones de crecimiento específicas, como una temperatura más alta, la disponibilidad de oxígeno y/o factores nutricionales específicos (Zhao et al., 2013) y no se conocen los requisitos de crecimiento de cada microorganismo. En el caso del análisis metatranscriptómico, el precultivo es, por supuesto, indeseable, ya que afectaría al estado fisiológico de las células. Además, las muestras deben procesarse lo más rápidamente posible para la extracción de ARN, almacenarse a $-80\text{ }^{\circ}\text{C}$ o fijarse utilizando soluciones como RNALater (ThermoScientific). Esto es crucial para obtener una imagen precisa de la actividad microbiana en una muestra.

Los métodos de extracción de ácido nucleico afectan indudablemente a la naturaleza, así como la calidad y la cantidad de ADN/ARN obtenido de los microorganismos presentes en una muestra, y por tanto influyen en los resultados experimentales. Es esencial tener esto en cuenta durante la interpretación de los datos y destaca la necesidad de utilizar métodos de extracción que sean óptimos para un determinado estudio o conocer los sesgos que el método de extracción puede introducir (Bag et al., 2016; Klenner et al., 2017; Cottier et al., 2018; Panek et al., 2018; Vaidya et al., 2018). La matriz de la que se purifica el ADN o el ARN para el análisis metagenómico/metatranscriptómico también requiere una atención especial. En el caso del aislamiento de ADN, el producto a menudo contiene material genómico vegetal o

animal que también arrojaría información de la secuencia, diluyendo así la información relevante de la secuencia microbiana. Para superar esto existen protocolos para eliminar el ADN no microbiano (Feehery et al., 2013, Gosiewski et al., 2014, 2017). El contenido de la matriz también puede interferir en el rendimiento del análisis molecular, ya que puede inhibir las reacciones bioquímicas necesarias (de Boer et al., 2015). Un posible enfoque para eliminar los componentes de la matriz es recuperar los microbios mediante centrifugación diferencial y filtración de soluciones acuosas. Los biofilms son a veces muy rígidos, lo que hace que estas complejas comunidades microbianas sean difíciles de homogeneizar (Corcoll et al., 2017). Las opciones para abrir estas comunidades incluyen el tratamiento enzimático combinado con fuertes fuerzas de cizallamiento como la sonicación y el beadbeating.

La cuestión de los enfoques metagenómicos para detectar y caracterizar cepas y rasgos específicos en muestras clínicas sin necesidad de utilizar el cultivo es cada vez más apremiante en la salud pública, ya que los laboratorios clínicos están dejando de cultivar patógenos bacterianos para detectarlos directamente en las muestras mediante PCR o inmunoensayos enzimáticos (Marder et al., 2017). La metabarcodificación tras la amplificación de uno o varios genes conservados puede utilizarse para detectar diferentes especies en un espécimen, pero no detectará los serotipos dentro de una especie que incluya comensales, por ejemplo, *E. coli* que incluye los serotipos productores de verotoxina (productores de toxina Shiga, VTEC/STEC), enteroagregantes (EAEC), enteropatógenos (EPEC), enteroinvasivos (EIEC) y *Shigella* y variantes menos virulentas de especies patógenas, por ejemplo, los serotipos no O1 y no O139 de *Vibrio cholerae*. Este problema podría resolverse dirigiéndose a los genes que codifican los factores de virulencia asociados a estos serotipos, pero, aunque esto podría ser factible con los genes que codifican el serotipo, a menudo no lo es con los genes asociados a la virulencia que suelen estar presentes en elementos genéticos móviles, por ejemplo, plásmidos y fagos, ya que podría ser imposible determinar a cuál de las múltiples bacterias de la muestra pertenecen. Esta es un área activa de investigación actual (Spencer et al., 2015).

El metabarcoding tradicional no suele proporcionar una resolución suficiente para diferenciar entre distintos aislados o entre muestras. Esto es necesario para el rastreo de fuentes similar al WGS de los aislados cultivados. Una solución a este problema es utilizar un enfoque similar y potencialmente compatible con wgMLST para analizar las secuencias de los aislados cultivados. Se seleccionan tantos loci como sea posible (hasta unos cuantos miles) de los esquemas de wgMLST para su amplificación y secuenciación directamente desde el

espécimen. Este enfoque se está probando actualmente para la detección y subtipificación de *Salmonella* con el objetivo de diseñar un sistema de detección y subtipificación independiente del cultivo que se aproxime a la resolución del esquema wgMLST.

También se está llevando a cabo la secuenciación Shotgun para la detección simultánea y la subtipificación de los patógenos sin necesidad de cultivo. Ha funcionado en estudios retrospectivos de muestras procedentes de brotes en los que el patógeno implicado ya había sido identificado mediante cultivo (Huang et al., 2017; Loman et al., 2013). Sin embargo, sin un conocimiento previo del patógeno, es necesario resolver una serie de cuestiones como la mencionada vinculación de los genes de los elementos genéticos móviles con las cepas a las que pertenecen. Los recientes avances en la secuenciación de células individuales parecen prometedores para abordar esta cuestión tanto para el metabarcoding como para la metagenómica (Lan et al., 2017; Spencer et al., 2015).

Además de las cuestiones que aquí se discuten, se necesitan mejoras críticas en las tecnologías de secuenciación y la bioinformática antes de que el metabarcoding o el Shotgun puedan aplicarse de forma rentable para el diagnóstico y la subtipificación de los patógenos transmitidos por los alimentos en apoyo de la salud pública y la seguridad alimentaria. Sin embargo, es probable que el rápido progreso de los avances en NGS anuncie la desaparición del cultivo dependiente como uno de los principales métodos en microbiología alimentaria.

5.6. Validación y evaluación comparativa.

Como ocurre con cualquier nueva tecnología en rápido desarrollo, la validación y estandarización de extremo a extremo de la NGS es un reto. Sin embargo, la necesidad de validación, evaluación comparativa y normalización es crucial para definir las directrices y las mejores prácticas para su aplicación en la gestión de la seguridad y la calidad de los alimentos.

A pesar de la disponibilidad de varios protocolos de laboratorio y de muchas herramientas dedicadas al análisis de datos de secuenciación de amplicones y metagenómica, su validación suele ser limitada debido a la naturaleza compleja de las muestras ambientales o alimentarias. La variedad de protocolos y soluciones de software para las aplicaciones de NGS sigue ampliándose, lo que hace que la validación y la estandarización sean un obstáculo para aplicaciones específicas. Sin embargo, se han llevado a cabo varios estudios

SECUENCIACIÓN DE AMPLICONES, METAGENÓMICA Y METATRANSCRIPTÓMICA

comparativos para comprobar el rendimiento y comparar diversos métodos y herramientas en los diferentes pasos de un estudio meta-ómico; a saber, la preparación de la muestra (Lewandowska et al., 2017), la extracción de ADN/ARN (Knudsen et al., 2016; Yuan et al., 2012), la preparación de la biblioteca (Jones et al., 2015; Schirmer et al., 2015), la plataforma de secuenciación utilizada (Tremblay et al., 2015) y el enfoque bioinformático aplicado (Siegwald et al., 2017). Sin embargo, la estandarización en este campo está todavía en desarrollo y la comparación y validación de estos protocolos y herramientas son esenciales para obtener información significativa y para hacer que el intercambio de información intra e inter-laboratorio sea efectivo (Costea et al., 2017).

Con respecto a los análisis bioinformáticos, existen pipelines de última generación que incluyen pasos cruciales como la eliminación de adaptadores, la eliminación de la secuencia del genoma matriz (carne, verduras, frutas, etc.), el filtrado de lecturas de baja calidad, el ensamblaje de contigs y, finalmente, la realización de búsquedas contra bases de datos regularmente actualizadas (Olson et al., 2017; Schlaberg et al., 2017). Singer et al. (2016) informaron del uso de una comunidad simulada definida con genomas de referencia completos para la evaluación comparativa y la validación de la secuenciación metagenómica, y recientemente se ha creado un recurso público para la evaluación comparativa bioinformática del microbioma (Bokulich et al., 2016; Singer et al., 2016). La importancia de la validación y la evaluación comparativa a menudo se pasa por alto, pero es esencial para una buena interpretación de los datos en el contexto de la seguridad alimentaria (por ejemplo, la identificación de patógenos).

El estado actual de validación y estandarización con respecto a la detección de cepas, así como la asignación de marcadores de virulencia y resistencia a especies o cepas específicas, está más avanzado en la WGS en comparación con la metagenómica. Esto puede explicarse fácilmente por las diferencias inherentes a ambos enfoques: La WGS permite acceder fácilmente a los genomas de uno en uno, a bajo rendimiento, mientras que la metagenómica está adaptada para evaluar genomas fragmentados de muestras complejas a un alto rendimiento. Sin embargo, nuevos enfoques bioinformáticos están permitiendo ahora la identificación de cepas coespecíficas (es decir, pertenecientes a la misma especie) a partir de datos de secuencias metagenómicas (Luo et al., 2015; Zolfo et al., 2017), aunque estos enfoques suelen depender de la información completa del genoma disponible en bases de datos públicas.

6. CONSIDERACIONES Y DESAFÍOS RELACIONADOS CON EL INTERCAMBIO DE DATOS

6. CONSIDERACIONES Y DESAFÍOS RELACIONADOS CON EL INTERCAMBIO DE DATOS.

La industria alimentaria es global, ya que produce y comercializa artículos en todo el mundo. Los productos procesados y las materias primas se transportan entre continentes y se someten a diversas investigaciones por parte de países exportadores e importadores. Esto da lugar a la generación de datos en varias etapas y en distintos países por parte de diferentes organizaciones y empresas. En este contexto, la NGS se aplica cada vez más, como se ha descrito en detalle en las secciones anteriores. Se reconoce ampliamente que el máximo beneficio de la NGS se obtendrá mediante la obtención de datos de secuencias junto con un conjunto mínimo acordado de metadatos descriptivos (FAO, 2016). La industria se beneficia si sus aislados se incluyen en los análisis científicos que, en última instancia, conducen a una comprensión más profunda de la diversidad microbiana mundial, la ecología y la distribución de los organismos. La salud pública se beneficiará tanto de la mejora de la detección y resolución de brotes como de la aplicación proactiva por parte de la industria de medidas de prevención y control más eficaces basadas en la inteligencia de la NGS.

En la actualidad, la industria está preocupada porque no existen salvaguardias para proteger a las empresas de las acciones reguladoras, así como para proteger la reputación de la empresa y el valor de la marca, lo que está obligando a las empresas a limitar el intercambio al mínimo legal, a pesar de que los beneficios del intercambio de datos son fácilmente reconocidos. Por lo tanto, para fomentar la compartición, es necesario reducir el riesgo mientras se mejoran los beneficios y se demuestra el valor (FAO, 2016). En las siguientes secciones se describen algunos de los aspectos clave que deben abordarse para fomentar el intercambio de datos.

6.1. Interpretación correcta de los datos.

Los datos de WGS susceptibles de ser malinterpretados por personal mal formado pueden suponer graves riesgos para la industria alimentaria, especialmente en la era de las redes sociales. Para que la industria se comprometa con un modelo de datos abiertos, deben abordarse mecanismos para prevenir y abordar estos problemas (FAO, 2016; Taboada et al., 2017). Así lo puso de manifiesto recientemente la Universidad Técnica de Dinamarca, donde un análisis preliminar informó de la presencia de ADN de mono en las hamburguesas. Tras un análisis posterior, se demostró que se trataba de ADN de ganado

(<https://www.food.dtu.dk/english/news/food-safety/Nyhed?id={800739D1-F72D-4C57-BAB1-4376E0A87BC7}>}, acceso 12 de marzo de 2021). Las limitaciones de la base de datos y las lecturas cortas utilizadas para la comparación de datos se identificaron como razones para la interpretación errónea de los resultados de la secuencia, lo que pone de relieve la importancia crítica de los conocimientos especializados para analizar e interpretar los datos de WGS.

Además, especialmente en el campo de la metagenómica microbiana, no se dispone de normas para la interpretación de los datos ni se ha llegado a un acuerdo al respecto, lo que puede dar lugar a informes contradictorios sobre los mismos resultados (Clooney et al., 2016). Esto se aplica no solo a los diferentes enfoques y métodos de análisis de datos, sino también cuando se utiliza el mismo enfoque, pero las conclusiones difieren.

6.2. Claridad jurídica/debida diligencia.

En la mayoría de las investigaciones de rastreo de fuentes de WGS, se incluyen en el análisis datos de secuencias de cepas estrechamente relacionadas para comprender con precisión el parentesco de los aislados estudiados. Esto se suele conseguir consultando las secuencias de interés en una base de datos de secuencias pública que comprende cepas aisladas de múltiples fuentes. Esto puede dar lugar a la agrupación de un aislado alimentario/ambiental que se está analizando con un aislado clínico. La situación se complica cuando se encuentra un vínculo entre un paciente histórico y un aislado interno reciente, y viceversa, con respecto a los pasos subsiguientes que debe dar el procesador de alimentos desde la perspectiva de la diligencia debida. En los Estados Unidos, todos los patógenos transmitidos por los alimentos que se obtienen a través de la vigilancia y la inspección se secuencian y las secuencias se cargan en el dominio público, donde residirán durante toda la vida de la base de datos. Las coincidencias con cualquier aislado que comparta un ancestro común reciente pueden dar lugar a nuevas investigaciones por parte de las instituciones de salud pública. En la mayoría de los casos, no se tomarán medidas reglamentarias sin información adicional, ya sea en relación con la exposición a los alimentos o con las observaciones antihigiénicas en el proceso continuo de la granja a la mesa. La respuesta reglamentaria depende de lo que se encuentre durante la inspección y de cómo responda la industria de acuerdo con las a largo plazo a la inspección y la regulación. El WGS es sólo la herramienta de subtipificación más reciente que se está aplicando, pero fundamentalmente la

CONSIDERACIONES Y DESAFÍOS RELACIONADOS CON EL INTERCAMBIO DE DATOS

toma de decisiones y las acciones reguladoras no han cambiado en gran medida. El WGS ayuda a los reguladores a reconocer antes los posibles problemas gracias a la mayor precisión de la tecnología, lo que permite una respuesta más rápida para mejorar la seguridad alimentaria y la salud pública. Los reguladores están interesados en saber cuándo una empresa presenta un problema de contaminación y qué se ha hecho para paliarlo y evitar que se repita. Sin embargo, la adopción de las tecnologías NGS por parte de la industria también le permitirá investigar más a fondo los posibles problemas de higiene o contaminación en sus instalaciones, facilitando el análisis de la causa raíz y proporcionándoles la oportunidad de ser mucho más proactivos a la hora de abordar dichos problemas de contaminación (Amini, 2017; FAO, 2016b). El uso rutinario de WGS en los Análisis de Peligros y Puntos de Control Crítico (APPCC) significará que las empresas alimentarias son mucho más conscientes de lo que ocurre en sus entornos de producción y sean más preventivas a la hora de evitar las enfermedades transmitidas por los alimentos en lugar de reaccionar ante ellas.

6.3. Los propietarios de los datos.

Existe la preocupación de que el uso de datos WGS disponibles públicamente pueda dar lugar a barreras comerciales e incluso a acciones legales locales debido a que los países operan dentro de marcos legales diferentes. Por lo tanto, existe un fuerte deseo de establecer y acordar un marco legal global y armonizado para facilitar el intercambio abierto (FAO, 2016). Las posibles soluciones a algunos de los problemas podrían ser retrasos definidos de común acuerdo en la puesta en común de datos o incluso un "periodo de gracia" sin consecuencias legales para promover la puesta en común activa de datos. Se requiere un esfuerzo considerable en términos de cooperación y coordinación en este ámbito para lograr el objetivo de compartir abiertamente los datos de WGS. Es importante que la industria desarrolle mecanismos tanto para compartir como para proteger la información sensible, de modo que pueda contribuir a las bases de datos de WGS con mayor comodidad (Jagadeesan et al., 2019).

7. PERSPECTIVAS FUTURAS PARA MEJORAR LA SEGURIDAD ALIMENTARIA

7. PERSPECTIVAS FUTURAS PARA MEJORAR LA SEGURIDAD ALIMENTARIA.

La industria alimentaria adopta cada vez más las tecnologías NGS para una amplia variedad de investigaciones microbiológicas alimentarias y la Figura 7.1. resume las diferentes posibilidades que ofrece a la industria en función de los requisitos, los recursos disponibles y el interés y la experiencia de cada empresa.

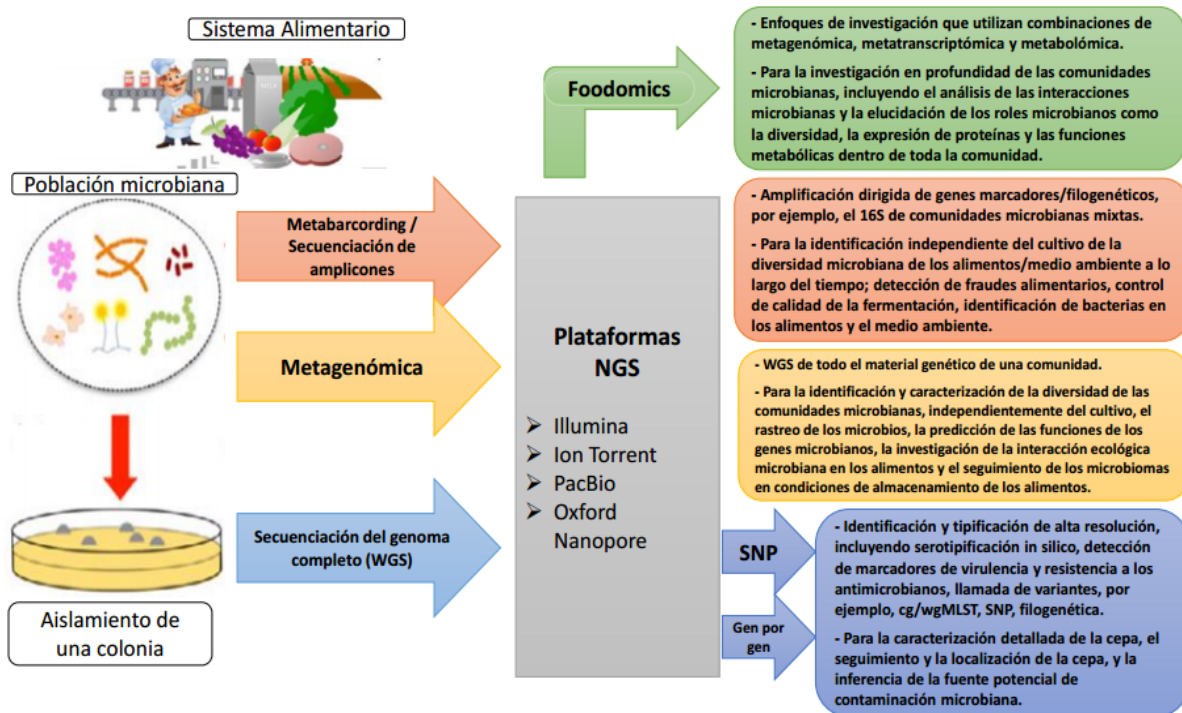


FIGURA 7.1. Resumen del uso potencial de la NGS por parte de la industria alimentaria. Fuente: Elaboración propia.

7.1. Secuenciación del genoma completo.

Una de las aplicaciones clave del WGS en la industria alimentaria será la de comprender la causa raíz de un evento de contaminación para poder abordarlo rápidamente. El proceso completo de WGS tiene que ser cómodo, rápido y asequible para que el WGS se utilice de forma rutinaria. El desarrollo de pipelines bioinformáticos fáciles de usar y la armonización de los métodos de análisis contribuirán a facilitarlos. La WGS debe adoptarse no como un complemento de las técnicas de caracterización microbiológica existentes, sino como un sustituto de los métodos de identificación y tipificación existentes, para que el beneficio de los costes se haga realidad.

PERSPECTIVAS FUTURAS PARA MEJORAR LA SEGURIDAD ALIMENTARIA

La industria se beneficiará enormemente si las características fenotípicas, como los perfiles de crecimiento e inactivación, pueden predecirse a partir del análisis del genoma. Sin embargo, dado que las respuestas fenotípicas a menudo también se controlan a nivel transcripcional y post-transcripcional, los enfoques multiómicos desempeñarán un papel clave para la caracterización de patógenos en el futuro. Además, es probable que los datos generados por la WGS y la metagenómica se integren con la microbiología predictiva para un mayor control de la seguridad y la calidad de los alimentos a lo largo de la cadena alimentaria. En el futuro, las bases de datos genómicos podrán vincularse a sitios web dedicados a la microbiología predictiva, como ComBase (<http://www.combase.cc/index.php/es/>; acceso 18 de octubre de 2021).

El máximo beneficio para la seguridad alimentaria de la WGS depende de que se compartan los datos y se prevé que la industria desarrolle un mecanismo para compartir y proteger la información sensible, de modo que pueda contribuir a las bases de datos de la WGS con mayor comodidad.

7.2. Análisis metagenómico.

Las herramientas metagenómicas pueden mejorar la comprensión de la ecología microbiana de las líneas de procesamiento de alimentos. Dentro de una comunidad microbiana, las interacciones entre los patógenos y el microbioma asociado pueden indicar la existencia de una especie patógena específica o afectar a su colonización. Las variaciones en los factores ambientales, como el pH, la concentración de sal y la actividad del agua, causadas por los tratamientos de procesamiento y manipulación pueden provocar los correspondientes cambios en la comunidad microbiana (Weimer et al., 2016). Los productores de alimentos podrán validar o mejorar la gestión actual de los riesgos microbianos utilizando el enfoque metagenómico para supervisar la aparición y la abundancia de microbios y genes en la comunidad microbiana de las líneas de procesamiento de alimentos.

Para la gestión del riesgo de deterioro microbiano, es importante controlar los cambios en la comunidad microbiana durante el almacenamiento para planificar las condiciones adecuadas de procesamiento, tratamiento y almacenamiento de los productos alimentarios (Ercolini, 2013). Las herramientas metagenómicas pueden ayudar a anticipar el deterioro microbiano mediante el estudio de los cambios en la diversidad o la proporción de los microbios asociados al deterioro en la microbiota de los productos alimenticios (Ercolini et

al., 2011; Kable et al., 2016), así como el seguimiento del comportamiento de las poblaciones asociadas a los fermentos/al deterioro en los alimentos cultivados (Masoud et al., 2012). Estas herramientas han permitido a los investigadores desarrollar la comprensión de los defectos de origen desconocido y desarrollar estrategias para eliminar esos defectos, como los que afectan a la carne y el marisco (Chaillou et al., 2015), la carne de salchicha (Hultman et al., 2015), el vino de arroz chino (Hong et al., 2016) y los quesos continentales (Quigley et al., 2016). La información obtenida de estas aplicaciones se ha utilizado para seleccionar cultivos iniciadores utilizados para producir alimentos fermentados con una calidad más homogénea (Galimberti et al., 2015), para identificar biomarcadores de madurez y calidad, y para optimizar las condiciones ambientales durante la producción de quesos (Wolfe et al., 2014), impulsando la formación de comunidades microbianas para producir alimentos con las propiedades deseadas. Las aplicaciones de los estudios metagenómicos han revelado que las diferencias en la microbiota del suelo influyen en los sabores de los vinos producidos en diferentes regiones geográficas (Zarraonaindia et al., 2015).

Los enfoques metagenómicos y metatranscriptómicos también tienen un gran potencial para convertirse en opciones valiosas para detectar la autenticidad e integridad de los alimentos mediante la descripción precisa de la comunidad microbiana de un producto alimentario específico. Las metodologías tradicionales de barcoding de ADN basadas en la PCR y la secuenciación de Sanger están limitadas por su bajo rendimiento y la necesidad de una alta pureza y concentración del ADN de las muestras de alimentos (Shokralla et al., 2014). Estas limitaciones están siendo abordadas por las tecnologías NGS de alto rendimiento, incluyendo los enfoques metagenómicos, que proporcionan más información sobre las poblaciones de la comunidad microbiana y los ingredientes biológicos de un producto alimenticio, además de permitir pruebas independientes de los cultivos. El software de predicción del metagenoma también se ha utilizado para comprender el impacto de las atmósferas modificadas en las vías metabólicas, para ayudar al diseño de los sistemas de conservación (Ferrocino y Cocolin, 2017). Estos enfoques metagenómicos, cuando se combinan con otras tecnologías -ómicas como la proteómica y la metabolómica tienen el potencial de vincular especies particulares en una comunidad con características funcionales, como la producción de sabor o la producción de metabolitos dañinos como las aminos biogénicas en el vino de arroz (Liu et al., 2016).

PERSPECTIVAS FUTURAS PARA MEJORAR LA SEGURIDAD ALIMENTARIA

La utilización del metagenoma en la industria alimentaria se enfrenta a retos como la detección de ADN procedente de microbios muertos, la baja sensibilidad de la detección en comparación con los métodos basados en cultivos y los costes relativamente elevados.

7.3. Las repercusiones de la aplicación de los NGS en el comercio y la industria alimentaria.

La aplicación de la NGS en la gestión de la seguridad alimentaria puede convertirse en un factor de cambio para el comercio mundial de alimentos. Mientras los principales actores siguen impulsando las tecnologías NGS para la gestión de la seguridad alimentaria mundial, también es urgente cerrar la brecha tecnológica entre los países productores de alimentos menos avanzados para facilitar el comercio mundial de alimentos. Los países en desarrollo tienen una importante preocupación por el posible desequilibrio de las oportunidades comerciales, ya que podrían no ser capaces de proporcionar el mismo nivel de datos basados en WGS que otros (FAO, 2016). Los obstáculos para el uso de WGS incluyen la falta de infraestructura, por ejemplo, servicios básicos y/o acceso a Internet, y la necesidad de desarrollar una mano de obra capacitada, tanto a nivel normativo como de la industria alimentaria, para realizar e interpretar los datos de WGS.

Es importante que los esfuerzos internacionales para facilitar la transición de las antiguas tecnologías a la NGS a nivel mundial sigan ofreciendo oportunidades a estos países, en términos de tecnología y formación, intercambio de conocimientos, reestructuración del sistema de seguridad alimentaria dentro del país y también mejorando la industria alimentaria local del país. La aparición de la tecnología NGS podría ser un punto de inflexión para salvar la distancia entre los países productores de alimentos menos avanzados y las naciones desarrolladas.

Por último, la extensión final del impacto de la NGS será la reducción de los costes de la industria alimentaria. El coste de la generación de secuencias genómicas bacterianas sigue disminuyendo rápidamente y en los próximos años se espera que el coste de la aplicación de la tecnología NGS supere fácilmente el coste del cultivo microbiológico y el examen fisiológico. Esta reducción de costes se suma a los beneficios de transformación de la industria alimentaria que esta nueva tecnología está llamada a aportar.

8. CONCLUSIONES

CONCLUSIONES

8. CONCLUSIONES.

PRIMERA. Las nuevas tecnologías de secuenciación masiva pueden aportar información muy clara sobre diferentes peligros asociados a puntos de control crítico, además, tienen un alto rendimiento y son muy precisas. Recientemente se están utilizando para examinar ecosistemas alimentarios, lo que produce una continua mejora de la calidad y seguridad alimentaria, viéndose esto reflejado en un gran avance en la industria agroalimentaria. Se puede afirmar que las técnicas tradicionales de tipificación y caracterización microbiana son sustituidas cada vez más por la tecnología de NGS debido a que son más rápidas y precisas.

SEGUNDA. La tecnología NGS está revolucionando la microbiología alimentaria, ya que supone un avance técnico tanto para el ámbito de la investigación como para la industria. Su incorporación en sistemas APPCC permitiría reducir pérdidas económicas debido a problemas en la cadena alimentaria, como puede ser la aparición de microorganismos patógenos como *L. monocytogenes* o *Salmonella* spp.

TERCERA. La Secuenciación de genoma completo o WGS junto con estudios epidemiológicos permite vincular casos de brotes producidos por microorganismos patógenos en zonas geográficas más amplias que las tecnologías dependientes de cultivo.

CUARTA. A pesar de que la Secuenciación de la Última Generación (NGS) comenzó siendo una herramienta únicamente de investigación, en los últimos años se ha convertido en una aplicación utilizada en muchos campos, especialmente en el ámbito agroalimentario y clínico. Las perspectivas futuras para esta tecnología indican que seguirá evolucionando, disminuyendo las tasas de error, tiempos de ejecución y costes, incrementándose su especificidad e implementación. Por tanto, su aplicación en las industrias agroalimentarias permitiría mejorar la seguridad en el mercado alimentario.

9. REFERENCIAS

REFERENCIAS

9. REFERENCIAS.

- Ajawatanawong, P. (2017). Molecular Phylogenetics: Concepts for a Newcomer. *Advances in Biochemical Engineering/Biotechnology*, 160(October 2016), 185–196. https://doi.org/10.1007/10_2016_49
- Allard, M. W., Bell, R., Ferreira, C. M., Gonzalez-Escalona, N., Hoffmann, M., Muruvanda, T., Ottesen, A., Ramachandran, P., Reed, E., Sharma, S., Stevens, E., Timme, R., Zheng, J., & Brown, E. W. (2018). Genomics of foodborne pathogens for microbial food safety. *Current Opinion in Biotechnology*, 49, 224–229. <https://doi.org/10.1016/j.copbio.2017.11.002>
- Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., & Timme, R. (2016). Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *Journal of Clinical Microbiology*, 54(8), 1975–1983. <https://doi.org/10.1128/JCM.00081-16>
- Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J., & O’Grady, J. (2015). MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature Biotechnology*, 33(3), 296–302. <https://doi.org/10.1038/nbt.3103>
- Ashton, P. M., Nair, S., Peters, T. M., Bale, J. A., Powell, D. G., Painset, A., Tewolde, R., Schaefer, U., Jenkins, C., Dallman, T. J., De Pinna, E. M., & Grant, K. A. (2016). Identification of Salmonella for public health surveillance using whole genome sequencing. *PeerJ*, 2016(4), 1–18. <https://doi.org/10.7717/peerj.1752>
- Aw, T. G., Wengert, S., & Rose, J. B. (2016). Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses. *International Journal of Food Microbiology*, 223, 50–56. <https://doi.org/10.1016/j.ijfoodmicro.2016.02.008>
- Aziz, R. K., Devoid, S., Disz, T., Edwards, R. A., Henry, C. S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Stevens, R. L., Vonstein, V., & Xia, F. (2012). SEED Servers: High-Performance Access to the SEED Genomes, Annotations, and Metabolic Models. *PLoS ONE*, 7(10), 1–10. <https://doi.org/10.1371/journal.pone.0048053>

REFERENCIAS

- Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., Saxena, S., Allin, K. H., Hansen, T., Arumugam, M., Vestergaard, H., Pedersen, O., Pereira, V., Abraham, P., Tripathi, R., Wadhwa, N., Bhatnagar, S., ... Das, B. (2016). An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Scientific Reports*, 6(June). <https://doi.org/10.1038/srep26775>
- Baker, B. J., Tyson, G. W., Webb, R. I., Flanagan, J., Hugenholtz, P., Allen, E. E., & Banfield, J. F. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. *Science*, 314(5807), 1933–1935. <https://doi.org/10.1126/science.1132690>
- Baldauf, S. L. (2003). The deep roots of eukaryotes. *Science*, 300(5626), 1703–1706. <https://doi.org/10.1126/science.1085544>
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., ... Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59. <https://doi.org/10.1038/nature07517>
- Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, 24(4), 335–341. <https://doi.org/10.1016/j.cmi.2017.10.013>
- Biesbroek, G., Sanders, E. A. M., Roeselers, G., Wang, X., Caspers, M. P. M., Trzciński, K., Bogaert, D., & Keijser, B. J. F. (2012). Deep sequencing analyses of low density microbial communities: Working at the boundary of accurate microbiota detection. *PLoS ONE*, 7(3). <https://doi.org/10.1371/journal.pone.0032942>
- Bokulich, N. A., Rideout, J. R., Mercurio, W. G., Shiffer, A., Wolfe, B., Maurice, C. F., Dutton, R. J., Turnbaugh, P. J., Knight, R., & Caporaso, J. G. (2016). mockrobiota: a Public Resource for Microbiome Bioinformatics Benchmarking. *MSystems*, 1(5), 1–7. <https://doi.org/10.1128/msystems.00062-16>
- Børnich, C., Grytten, I., Hovig, E., Paulsen, J., Čech, M., & Sandve, G. K. (2016). Galaxy Portal: Interacting with the galaxy platform through mobile devices. *Bioinformatics*, 32(11), 1743–1745. <https://doi.org/10.1093/bioinformatics/btw042>

REFERENCIAS

- Carrigo, J. A., Rossi, M., Moran-Gilad, J., Van Domselaar, G., & Ramirez, M. (2018). A primer on microbial bioinformatics for nonbioinformaticians. *Clinical Microbiology and Infection*, 24(4), 342–349. <https://doi.org/10.1016/j.cmi.2017.12.015>
- Chen, Y., Luo, Y., Carleton, H., Timme, R., Melka, D., Muruvanda, T., Wang, C., Kastanis, G., Katz, L. S., Turner, L., Fritzinger, A., Moore, T., Stones, R., Blankenship, J., Salter, M., Parish, M., Hammack, T. S., Evans, P. S., Tarr, C. L., ... Brown, E. W. (2017). Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of *Listeria monocytogenes* isolates associated with an outbreak linked to cheese, United States, 2013. *Applied and Environmental Microbiology*, 83(15), 1–15. <https://doi.org/10.1128/AEM.00633-17>
- Cifuentes, A. (2009). Food analysis and foodomics. *Journal of Chromatography A*, 1216(43), 7109. <https://doi.org/10.1016/j.chroma.2009.09.018>
- Clooney, A. G., Fouhy, F., Sleator, R. D., O'Driscoll, A., Stanton, C., Cotter, P. D., & Claesson, M. J. (2016). Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. *PLoS ONE*, 11(2), 1–16. <https://doi.org/10.1371/journal.pone.0148028>
- Corcoll, N., Österlund, T., Sinclair, L., Eiler, A., Kristiansson, E., Backhaus, T., & Martin Eriksson, K. (2017). Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing. *FEMS Microbiology Letters*, 364(14), 1–9. <https://doi.org/10.1093/femsle/fnx139>
- Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F. E., Kultima, J. R., Hayward, M. R., Coelho, L. P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J. R. M., Carton, T., Cools-Portier, S., ... Bork, P. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology*, 35(11), 1069–1076. <https://doi.org/10.1038/nbt.3960>
- Cottier, F., Srinivasan, K. G., Yurieva, M., Liao, W., Poidinger, M., Zolezzi, F., & Pavelka, N. (2018). Advantages of meta-total RNA sequencing (MeTRS) over shotgun metagenomics and amplicon-based sequencing in the profiling of complex microbial communities. *Npj Biofilms and Microbiomes*, 4(1). <https://doi.org/10.1038/s41522-017->

REFERENCIAS

0046-x

- Dallman, T., Inns, T., Jombart, T., Ashton, P., Loman, N., Chatt, C., Messelhaeusser, U., Rabsch, W., Simon, S., Nikisins, S., Bernard, H., le Hello, S., Jourdan da-Silva, N., Kornschober, C., Mossong, J., Hawkey, P., de Pinna, E., Grant, K., & Cleary, P. (2016). Phylogenetic structure of European Salmonella Enteritidis outbreak correlates with national and international egg distribution network. *Microbial Genomics*, 2(8), e000070. <https://doi.org/10.1099/mgen.0.000070>
- Davis, S., Pettengill, J. B., Luo, Y., Payne, J., Shpuntoff, A., Rand, H., & Strain, E. (2015). CFSAN SNP pipeline: An automated method for constructing snp matrices from next-generation sequence data. *PeerJ Computer Science*, 2015(8), 1–11. <https://doi.org/10.7717/peerj-cs.20>
- de Boer, P., Caspers, M., Sanders, J.-W., Kemperman, R., Wijman, J., Lommerse, G., Roeselers, G., Montijn, R., Abee, T., & Kort, R. (2015). Amplicon sequencing for the quantification of spoilage microbiota in complex foods including bacterial spores. *Microbiome*, 3(1), 1–13. <https://doi.org/10.1186/s40168-015-0096-3>
- De Filippo, C., Ramazzotti, M., Fontana, P., & Cavalieri, D. (2012). Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. *Briefings in Bioinformatics*, 13(6), 696–710. <https://doi.org/10.1093/bib/bbs070>
- Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., García-Cobos, S., Kooistra-Smid, A. M. D., Raangs, E. C., Rosema, S., Veloo, A. C. M., Zhou, K., Friedrich, A. W., & Rossen, J. W. A. (2017). Application of next generation sequencing in clinical microbiology and infection prevention. *Journal of Biotechnology*, 243, 16–24. <https://doi.org/10.1016/j.jbiotec.2016.12.022>
- Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, 36(16). <https://doi.org/10.1093/nar/gkn425>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., ... Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), 133–138. <https://doi.org/10.1126/science.1162986>

REFERENCIAS

- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D. M., Reid, J. G., Worley, K. C., & Gibbs, R. A. (2012). Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE*, 7(11), 1–12. <https://doi.org/10.1371/journal.pone.0047768>
- Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, 2015(10). <https://doi.org/10.7717/peerj.1319>
- Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., & Sogin, M. L. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, 4(12), 1111–1119. <https://doi.org/10.1111/2041-210X.12114>
- Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., & Sogin, M. L. (2015). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME Journal*, 9, 968–979. <https://doi.org/10.1038/ismej.2014.195>
- Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology*, 10(8), 538–550. <https://doi.org/10.1038/nrmicro2832>
- Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A., & Pradhan, S. (2013). A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA. *PLoS ONE*, 8(10). <https://doi.org/10.1371/journal.pone.0076096>
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, 44(D1), D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, 7(6), 461–465. <https://doi.org/10.1038/nmeth.1459>
- Forbes, J. D., Knox, N. C., Ronholm, J., Pagotto, F., & Reimer, A. (2017). Metagenomics: The next culture-independent game changer. *Frontiers in Microbiology*, 8(JUL), 1–21.

REFERENCIAS

- <https://doi.org/10.3389/fmicb.2017.01069>
- Franz, E., Gras, L. M., & Dallman, T. (2016). Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. *Current Opinion in Food Science*, 8, 74–79. <https://doi.org/10.1016/j.cofs.2016.04.004>
- Gargis, A. S., Kalman, L., & Lubin, I. M. (2016). Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. *Journal of Clinical Microbiology*, 54(12), 2857–2865. <https://doi.org/10.1128/JCM.00949-16>
- Gillesberg Lassen, S., Ethelberg, S., Björkman, J. T., Jensen, T., Sørensen, G., Kvistholm Jensen, A., Müller, L., Nielsen, E. M., & Mølbak, K. (2016). Two listeria outbreaks caused by smoked fish consumption—using whole-genome sequencing for outbreak investigations. *Clinical Microbiology and Infection*, 22(7), 620–624. <https://doi.org/10.1016/j.cmi.2016.04.017>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Gosiewski, T., Ludwig-Galezowska, A. H., Huminska, K., Sroka-Oleksiak, A., Radkowski, P., Salamon, D., Wojciechowicz, J., Kus-Slowinska, M., Bulanda, M., & Wolkow, P. P. (2017). Comprehensive detection and identification of bacterial DNA in the blood of patients with sepsis and healthy volunteers using next-generation sequencing method - the observation of DNAemia. *European Journal of Clinical Microbiology and Infectious Diseases*, 36(2), 329–336. <https://doi.org/10.1007/s10096-016-2805-7>
- Gosiewski, Tomasz, Jurkiewicz-Badacz, D., Sroka, A., Brzywczy-Włoch, M., & Bulanda, M. (2014). A novel, nested, multiplex, real-time PCR for detection of bacteria and fungi in blood. *BMC Microbiology*, 14(1), 1–7. <https://doi.org/10.1186/1471-2180-14-144>
- Grant, K., Jenkins, C., Arnold, C., Green, J., & Zambon, M. (2018). Implementing pathogen genomics. *Phe*, 31. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/731057/implementing_pathogen_genomics_a_case_study.pdf
- Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J. R., Coelho, L. P., Espinoza, J. C. I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., ... Gorsky, G.

REFERENCIAS

- (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, 532(7600), 465–470. <https://doi.org/10.1038/nature16942>
- Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Dae, H. K., Mong, S. M., Meng, Q., Cao, H., Li, X., Shi, S., Yu, L., Kalachikov, S., Russo, J. J., Turro, N. J., & Ju, J. (2008). Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 105(27), 9145–9150. <https://doi.org/10.1073/pnas.0804023105>
- Hedge, J., & Wilson, D. J. (2016). Practical Approaches for Detecting Selection in Microbial Genomes. *PLoS Computational Biology*, 12(2), 1–12. <https://doi.org/10.1371/journal.pcbi.1004739>
- Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., & Eiros, J. M. (2020). Bioinformatics of next generation sequencing in clinical microbiology diagnosis. *Revista Argentina de Microbiología*, 52(2), 150–161. <https://doi.org/10.1016/j.ram.2019.06.003>
- Hoffmann, M., Luo, Y., Monday, S. R., Gonzalez-Escalona, N., Ottesen, A. R., Muruvanda, T., Wang, C., Kastanis, G., Keys, C., Janies, D., Senturk, I. F., Catalyurek, U. V., Wang, H., Hammack, T. S., Wolfgang, W. J., Schoonmaker-Bopp, D., Chu, A., Myers, R., Haendiges, J., ... Brown, E. W. (2016). Tracing origins of the salmonella bareilly strain causing a food-borne outbreak in the United States. *Journal of Infectious Diseases*, 213(4), 502–508. <https://doi.org/10.1093/infdis/jiv297>
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., Von Mering, C., & Bork, P. (2016). EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1), D286–D293. <https://doi.org/10.1093/nar/gkv1248>
- Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3), 377–386. <https://doi.org/10.1101/gr.5969107>
- Huson, D. H., Beier, S., Flade, I., Górška, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H. J., & Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology*, 12(6), 1–

REFERENCIAS

12. <https://doi.org/10.1371/journal.pcbi.1004957>
- Jackson, B. R., Tarr, C., Strain, E., Jackson, K. A., Conrad, A., Carleton, H., Katz, L. S., Stroika, S., Gould, L. H., Mody, R. K., Silk, B. J., Beal, J., Chen, Y., Timme, R., Doyle, M., Fields, A., Wise, M., Tillman, G., Defibaugh-Chavez, S., ... Gerner-Smidt, P. (2016). Implementation of Nationwide Real-time Whole-genome Sequencing to Enhance Listeriosis Outbreak Detection and Investigation. *Clinical Infectious Diseases*, *63*(3), 380–386. <https://doi.org/10.1093/cid/ciw242>
- Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., Chaffron, S., Van Der Vossen, J., Tang, S., Katase, M., McClure, P., Kimura, B., Ching Chai, L., Chapman, J., & Grant, K. (2019). The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiology*, *79*(June 2018), 96–115. <https://doi.org/10.1016/j.fm.2018.11.005>
- Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., & Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nature Methods*, *12*(4), 351–356. <https://doi.org/10.1038/nmeth.3290>
- Jensen, A. K., Nielsen, E. M., Björkman, J. T., Jensen, T., Müller, L., Persson, S., Bjerager, G., Perge, A., Krause, T. G., Kiil, K., Sørensen, G., Andersen, J. K., Mølbak, K., & Ethelberg, S. (2016). Whole-genome sequencing used to investigate a nationwide outbreak of listeriosis caused by ready-to-eat delicatessen meat, Denmark, 2014. *Clinical Infectious Diseases*, *63*(1), 64–70. <https://doi.org/10.1093/cid/ciw192>
- Jones, M. B., Highlander, S. K., Anderson, E. L., Li, W., Dayrit, M., Klitgord, N., Fabani, M. M., Seguritan, V., Green, J., Pride, D. T., Yooseph, S., Biggs, W., Nelson, K. E., & Craig Venter, J. (2015). Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 14024–14029. <https://doi.org/10.1073/pnas.1519288112>
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- Klenner, J., Kohl, C., Dabrowski, P. W., & Nitsche, A. (2017). Comparing Viral Metagenomic Extraction Methods. *Metagenomics: Current Advances and Emerging*

REFERENCIAS

- Concepts*, August, 4–6. <https://doi.org/10.21775/9781910190593.04>
- Kleta, S., Hammerl, J. A., Dieckmann, R., Malorny, B., Borowiak, M., Halbedel, S., Prager, R., Trost, E., Flieger, A., Wilking, H., Vygen-Bonnet, S., Busch, U., Messelhäuser, U., Horlacher, S., Schönberger, K., Lohr, D., Aichinger, E., Lubner, P., Hensel, A., & Al Dahouk, S. (2017). Molecular tracing to find source of protracted invasive listeriosis outbreak, Southern Germany, 2012–2016. *Emerging Infectious Diseases*, 23(10), 1680–1683. <https://doi.org/10.3201/eid2310.161623>
- Knudsen, B. E., Bergmark, L., Munk, P., Lukjancenko, O., Priemé, A., Aarestrup, F. M., & Pamp, S. J. (2016). Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. *MSystems*, 1(5). <https://doi.org/10.1128/msystems.00095-16>
- Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2567–2572. <https://doi.org/10.1073/pnas.0409727102>
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., & Phillippy, A. M. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology*, 30(7), 693–700. <https://doi.org/10.1038/nbt.2280>
- Köser, C. U., Ellington, M. J., Cartwright, E. J. P., Gillespie, S. H., Brown, N. M., Farrington, M., Holden, M. T. G., Dougan, G., Bentley, S. D., Parkhill, J., & Peacock, S. J. (2012). Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. *PLoS Pathogens*, 8(8). <https://doi.org/10.1371/journal.ppat.1002824>
- Kozyreva VK, Truong C_H, Greninger A, Crandall J, Mukhopadhyay R, C. V. (2017). crossm Compliant Whole-Genome Sequencing in the Public Health Microbiology. *Journal of Clinical Microbiology*, 55(8), 2502–2520.
- Kultima, J. R., Coelho, L. P., Forslund, K., Huerta-Cepas, J., Li, S. S., Driessen, M., Voigt, A. Y., Zeller, G., Sunagawa, S., & Bork, P. (2016). MOCAT2: A metagenomic assembly, annotation and profiling framework. *Bioinformatics*, 32(16), 2520–2523. <https://doi.org/10.1093/bioinformatics/btw183>
- Langille, M., Zaneveld, J., Caporaso, J. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* 31, 814–

REFERENCIAS

- 821 (2013). <https://doi.org/10.1038/nbt.2676>
- Leonard, S. R., Mammel, M. K., Lacher, D. W., & Elkins, C. A. (2015). Application of metagenomic sequencing to food safety: Detection of shiga toxin-producing *Escherichia coli* on fresh bagged spinach. *Applied and Environmental Microbiology*, *81*(23), 8183–8191. <https://doi.org/10.1128/AEM.02601-15>
- Levene, H. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, *299*(5607), 682–686. <https://doi.org/10.1126/science.1079700>
- Lewandowska, D. W., Zagordi, O., Geissberger, F. D., Kufner, V., Schmutz, S., Böni, J., Metzner, K. J., Trkola, A., & Huber, M. (2017). Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. *Microbiome*, *5*(1), 94. <https://doi.org/10.1186/s40168-017-0317-z>
- Li, P. E., Lo, C. C., Anderson, J. J., Davenport, K. W., Bishop-Lilly, K. A., Xu, Y., Ahmed, S., Feng, S., Mokashi, V. P., & Chain, P. S. G. (2017). Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Research*, *45*(1), 67–80. <https://doi.org/10.1093/nar/gkw1027>
- Lienau, E. K., Strain, E., Wang, C., Zheng, J., Ottesen, A. R., Keys, C. E., Hammack, T. S., Musser, S. M., Brown, E. W., Allard, M. W., Cao, G., Meng, J., & Stones, R. (2011). Identification of a Salmonellosis Outbreak by Means of Molecular Sequencing. *New England Journal of Medicine*, *364*(10), 981–982. <https://doi.org/10.1056/nejmc1100443>
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, *2012*. <https://doi.org/10.1155/2012/251364>
- Loman, N. J., Constantinidou, C., Christner, M., Chan, J. Z. M., Quick, J., Weir, J. C., Quince, C., Smith, G. P., Betley, J. R., Aepfelbacher, M., & Pallen, M. J. (2013). A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of shiga-toxigenic *Escherichia coli* O104:H4. *JAMA - Journal of the American Medical Association*, *309*(14), 1502–1510. <https://doi.org/10.1001/jama.2013.3231>
- Loomis, E. W., Eid, J. S., Peluso, P., Yin, J., Hickey, L., Rank, D., McCalmon, S., Hagerman, R. J., Tassone, F., & Hagerman, P. J. (2013). Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene. *Genome Research*, *23*(1), 121–128.

REFERENCIAS

- <https://doi.org/10.1101/gr.141705.112>
- Maiden, M. C. J., Van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., & McCarthy, N. D. (2013). MLST revisited: The gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*, *11*(10), 728–736. <https://doi.org/10.1038/nrmicro3093>
- Marder, E. P., Cieslak, P. R., Cronquist, A. B., Dunn, J., Lathrop, S., Rabatsky-Ehr, T., Ryan, P., Smith, K., Tobin-D'Angelo, M., Vugia, D. J., Zansky, S., Holt, K. G., Wolpert, B. J., Lynch, M., Tauxe, R., & Geissler, A. L. (2017). Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food and the Effect of Increasing Use of Culture-Independent Diagnostic Tests on Surveillance — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2013–2016. *MMWR. Morbidity and Mortality Weekly Report*, *66*(15), 397–403. <https://doi.org/10.15585/mmwr.mm6615a1>
- Maury, M. M., Tsai, Y. H., Charlier, C., Touchon, M., Chenal-Francisque, V., Leclercq, A., Criscuolo, A., Gaultier, C., Roussel, S., Brisabois, A., Disson, O., Rocha, E. P. C., Brisse, S., & Lecuit, M. (2016). Uncovering *Listeria monocytogenes* hypervirulence by harnessing its biodiversity. *Nature Genetics*, *48*(3), 308–313. <https://doi.org/10.1038/ng.3501>
- MO, C., C, R., MG, R., SB, G., C, N., & MA, D. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics*, *13*, 375.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, *35*(SUPPL.2), 182–185. <https://doi.org/10.1093/nar/gkm321>
- Moura, A., Criscuolo, A., Pouseele, H., Maury, M. M., Leclercq, A., Tarr, C., Björkman, J. T., Dallman, T., Reimer, A., Enouf, V., Larssonneur, E., Carleton, H., Bracq-Dieye, H., Katz, L. S., Jones, L., Touchon, M., Tourdjman, M., Walker, M., Stroika, S., ... Brisse, S. (2016). Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nature Microbiology*, *2*(2). <https://doi.org/10.1038/nmicrobiol.2016.185>
- Narayanasamy, S., Jarosz, Y., Muller, E. E. L., Heintz-Buschart, A., Herold, M., Kaysen, A., Laczny, C. C., Pinel, N., May, P., & Wilmes, P. (2016). IMP: A pipeline for reproducible

REFERENCIAS

- reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biology*, 17(1), 1–21. <https://doi.org/10.1186/s13059-016-1116-8>
- Ni, J., Yan, Q., & Yu, Y. (2013). How much metagenomic sequencing is enough to achieve a given goal? *Scientific Reports*, 3(December). <https://doi.org/10.1038/srep01968>
- Olson, N. D., Treangen, T. J., Hill, C. M., Cepeda-Espinoza, V., Ghurye, J., Koren, S., & Pop, M. (2018). Metagenomic assembly through the lens of validation: Recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Briefings in Bioinformatics*, 20(4), 1140–1150. <https://doi.org/10.1093/bib/bbx098>
- Ottesen, A., Ramachandran, P., Reed, E., White, J. R., Hasan, N., Subramanian, P., Ryan, G., Jarvis, K., Grim, C., Daquigan, N., Hanes, D., Allard, M., Colwell, R., Brown, E., & Chen, Y. (2016). Enrichment dynamics of *Listeria monocytogenes* and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak. *BMC Microbiology*, 16(1), 1–11. <https://doi.org/10.1186/s12866-016-0894-1>
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., ... Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17), 5691–5702. <https://doi.org/10.1093/nar/gki866>
- Page, A. J., Alikhan, N. F., Carleton, H. A., Seemann, T., Keane, J. A., & Katz, L. S. (2017). Comparison of classical multi-locus sequence typing software for next-generation sequencing data. *Microbial Genomics*, 3(8), 1–8. <https://doi.org/10.1099/mgen.0.000124>
- Panek, M., Čipčić Paljetak, H., Barešić, A., Perić, M., Matijašić, M., Lojkić, I., Bender, D. V., Krznarić, Ž., & Verbanac, D. (2018). Methodology challenges in studying human gut microbiota-Effects of collection, storage, DNA extraction and next generation sequencing technologies. *Scientific Reports*, 8(1), 1–13. <https://doi.org/10.1038/s41598-018-23296-4>
- Pightling, A. W., Petronella, N., & Pagotto, F. (2015). Choice of reference-guided sequence assembler and SNP caller for analysis of *Listeria monocytogenes* short-read sequence data greatly influences rates of error Genomics. *BMC Research Notes*, 8(1), 1–9. <https://doi.org/10.1186/s13104-015-1689-4>

REFERENCIAS

- Pightling, A. W., Pettengill, J. B., Luo, Y., Baugher, J. D., Rand, H., & Strain, E. (2018). Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Frontiers in Microbiology*, 9(JUL), 1–13. <https://doi.org/10.3389/fmicb.2018.01482>
- Portmann, A. C., Fournier, C., Gimonet, J., Ngom-Bru, C., Barretto, C., & Baert, L. (2018). A validation approach of an end-to-end whole genome sequencing workflow for source tracking of *Listeria monocytogenes* and *Salmonella enterica*. *Frontiers in Microbiology*, 9(MAR), 1–13. <https://doi.org/10.3389/fmicb.2018.00446>
- Quainoo, S., Coolen, J. P. M., Sacha A. F. T. van Hijum, C., Martijn A. Huynen, c W. J. G. M., Willem van Schaik, E., & Wertheim, H. F. L. (2017). crossm Whole-Genome Sequencing of Bacterial Pathogens : the Future of Nosocomial. *Clinical Microbiology Reviews*, 30(4), 1015–1064.
- Quick, J., Quinlan, A. R., & Loman, N. J. (2014). A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *GigaScience*, 3(1), 1–6. <https://doi.org/10.1186/2047-217X-3-22>
- Raes, J., & Bork, P. (2008). Molecular eco-systems biology: Towards an understanding of community function. *Nature Reviews Microbiology*, 6(9), 693–699. <https://doi.org/10.1038/nrmicro1935>
- Rantsiou, K., Kathariou, S., Winkler, A., Skandamis, P., Saint-Cyr, M. J., Rouzeau-Szynalski, K., & Amézquita, A. (2018). Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment. *International Journal of Food Microbiology*, 287(December 2017), 3–9. <https://doi.org/10.1016/j.ijfoodmicro.2017.11.007>
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., ... Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348–352. <https://doi.org/10.1038/nature10242>
- Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, 43(6). <https://doi.org/10.1093/nar/gku1341>

REFERENCIAS

- Schlaberg, R., Chiu, C. Y., Miller, S., Procop, G. W., & Weinstock, G. (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Archives of Pathology and Laboratory Medicine*, *141*(6), 776–786. <https://doi.org/10.5858/arpa.2016-0539-RA>
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. <https://doi.org/10.1128/AEM.01541-09>
- Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L., & Goering, R. V. (2018). Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clinical Microbiology and Infection*, *24*(4), 350–354. <https://doi.org/10.1016/j.cmi.2017.12.016>
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome biology*, *12*(6), 1-18. <https://doi.org/10.1186/gb-2011-12-6-r60>
- Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., & Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Molecular Systems Biology*, *9*(1), 1–15. <https://doi.org/10.1038/msb.2013.22>
- Sekse, C., Holst-Jensen, A., Dobrindt, U., Johannessen, G. S., Li, W., Spilberg, B., & Shi, J. (2017). High throughput sequencing for detection of foodborne pathogens. *Frontiers in Microbiology*, *8*(OCT), 1–26. <https://doi.org/10.3389/fmicb.2017.02029>
- Sharon, D., Tilgner, H., Grubert, F., & Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology*, *31*(11), 1009–1014. <https://doi.org/10.1038/nbt.2705>
- Siegwald, L., Touzet, H., Lemoine, Y., Hot, D., Audebert, C., & Caboche, S. (2017). Assessment of common and emerging bioinformatics pipelines for targeted metagenomics. *PLoS ONE*, *12*(1), 1–26. <https://doi.org/10.1371/journal.pone.0169563>
- Singer, E., Andreopoulos, B., Bowers, R. M., Lee, J., Deshpande, S., Chiniquy, J., Ciobanu, D., Klenk, H. P., Zane, M., Daum, C., Clum, A., Cheng, J. F., Copeland, A., & Woyke,

REFERENCIAS

- T. (2016). Next generation sequencing data of a defined microbial mock community. *Scientific Data*, 3, 1–8. <https://doi.org/10.1038/sdata.2016.81>
- Spencer, S. J., Tamminen, M. V., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., A Weitz, D., Pitkänen, L. K., Vigneault, F., Virta, M. P., & Alm, E. J. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME Journal*, 10(2), 427–436. <https://doi.org/10.1038/ismej.2015.124>
- Taboada, E. N., Graham, M. R., Carriço, J. A., & Van Domselaar, G. (2017). Food safety in the age of next generation sequencing, bioinformatics, and open data access. *Frontiers in Microbiology*, 8(MAY), 1–10. <https://doi.org/10.3389/fmicb.2017.00909>
- Takami, H., Taniguchi, T., Arai, W., Takemoto, K., Moriya, Y., & Goto, S. (2016). An automated system for evaluation of the potential functionome: MAPLE version 2.1.0. *DNA Research*, 23(5), 467–475. <https://doi.org/10.1093/dnares/dsw030>
- Timme, R. E., Rand, H., Shumway, M., Trees, E. K., Simmons, M., Agarwala, R., Davis, S., Tillman, G. E., Defibaugh-Chavez, S., Carleton, H. A., Klimke, W. A., & Katz, L. S. (2017). Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. *PeerJ*, 2017(10), 1–13. <https://doi.org/10.7717/peerj.3893>
- Travers, K. J., Chin, C. S., Rank, D. R., Eid, J. S., & Turner, S. W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research*, 38(15). <https://doi.org/10.1093/nar/gkq543>
- Treangen, T. J., Koren, S., Sommer, D. D., Liu, B., Astrovskaia, I., Ondov, B., Darling, A. E., Phillippy, A. M., & Pop, M. (2013). MetAMOS: A modular and open source metagenomic assembly and analysis pipeline. *Genome Biology*, 14(1), 1–20. <https://doi.org/10.1186/gb-2013-14-1-r2>
- Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., Lee, J., Chen, F., Dangl, J. L., & Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology*, 6(AUG), 1–15. <https://doi.org/10.3389/fmicb.2015.00771>
- Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C., & Segata, N. (2017). Microbial strain-level population structure & genetic diversity from metagenomes. *Genome Research*, 27(4), 626–638. <https://doi.org/10.1101/gr.216242.116>

REFERENCIAS

- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., & Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, *428*(6978), 37–43. <https://doi.org/10.1038/nature02340>
- Uemura, S., Aitken, C. E., Korlach, J., Flusberg, B. A., Turner, S. W., & Puglisi, J. D. (2015). *resolution*. *464*(7291), 1012–1017. <https://doi.org/10.1038/nature08925>. Real-time
- Vaidya, J. D., van den Bogert, B., Edwards, J. E., Boekhorst, J., van Gastelen, S., Saccenti, E., Plugge, C. M., & Smidt, H. (2018). The effect of DNA extraction methods on observed microbial communities from fibrous and liquid rumen fractions of dairy cows. *Frontiers in Microbiology*, *9*(JAN), 1–16. <https://doi.org/10.3389/fmicb.2018.00092>
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., ... Smith, H. O. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*, *304*(5667), 66–74. <https://doi.org/10.1126/science.1093857>
- Waldor, M. K., Tyson, G., Borenstein, E., Ochman, H., Moeller, A., Finlay, B. B., Kong, H. H., Gordon, J. I., Nelson, K. E., Dabbagh, K., & Smith, H. (2015). Where Next for Microbiome Research? *PLoS Biology*, *13*(1), 1–9. <https://doi.org/10.1371/journal.pbio.1002050>
- Wang, Y., Wang, H., Mi, D., Gu, X., & Hu, W. (2015). Periodical assessment of electrophysiological recovery following sciatic nerve crush via surface stimulation in rats. *Neurological Sciences*, *36*(3), 449–456. <https://doi.org/10.1007/s10072-014-2005-0>
- Warnecke, F., & Hugenholtz, P. (2007). Building on basic metagenomics with complementary technologies. *Genome Biology*, *8*(12). <https://doi.org/10.1186/gb-2007-8-12-231>
- Weimer, B. C., Storey, D. B., Elkins, C. A., Baker, R. C., Markwell, P., Chambliss, D. D., Edlund, S. B., & Kaufman, J. H. (2016). Defining the food microbiome for authentication, safety, and process management. *IBM Journal of Research and Development*, *60*(5–6). <https://doi.org/10.1147/JRD.2016.2582598>
- Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., Paczian, T., Trimble, W. L., Bagchi, S., Grama, A., Chaterji, S., & Meyer, F. (2016). The MG-RAST

REFERENCIAS

- metagenomics database and portal in 2015. *Nucleic Acids Research*, 44(D1), D590–D594. <https://doi.org/10.1093/nar/gkv1322>
- Wilson, M. R., Brown, E., Keys, C., Strain, E., Luo, Y., Muruvanda, T., Grim, C., Beaubrun, J. J. G., Jarvis, K., Ewing, L., Gopinath, G., Hanes, D., Allard, M. W., & Musser, S. (2016). Whole genome DNA sequence analysis of *Salmonella* subspecies enterica serotype Tennessee obtained from related peanut butter foodborne outbreaks. *PLoS ONE*, 11(6), 1–21. <https://doi.org/10.1371/journal.pone.0146929>
- Wyres, K. L., Conway, T. C., Garg, S., Queiroz, C., Reumann, M., Holt, K., & Rusu, L. I. (2014). WGS analysis and interpretation in clinical and public health microbiology laboratories: What are the requirements and how do existing tools compare? *Pathogens*, 3(2), 437–458. <https://doi.org/10.3390/pathogens3020437>
- Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., Yang, H., Geornaras, I., Woerner, D. R., Jones, K. L., Ruiz, J., Boucher, C., Morley, P. S., & Belk, K. E. (2016). Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. *Applied and Environmental Microbiology*, 82(8), 2433–2443. <https://doi.org/10.1128/AEM.00078-16>
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics*, 13(5), 303–314. <https://doi.org/10.1038/nrg3186>
- Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE*, 7(3). <https://doi.org/10.1371/journal.pone.0033865>
- Zhao, Y., Caspers, M. P. M., Metselaar, K. I., De Boer, P., Roeselers, G., Moezelaar, R., Groot, M. N., Montijn, R. C., Abee, T., & Korta, R. (2013). Abiotic and Microbiotic Factors Controlling Biofilm Formation by Thermophilic Sporeformers. *Applied and Environmental Microbiology*, 79(18), 5652–5660. <https://doi.org/10.1128/AEM.00949-13>
- Zolfo, M., Tett, A., Jousson, O., Donati, C., & Segata, N. (2017). MetaMLST: Multi-locus strain-level bacterial typing from metagenomic samples. *Nucleic Acids Research*, 45(2), e7. <https://doi.org/10.1093/nar/gkw837>