



**Centro de Investigación en Alimentación y
Desarrollo, A.C.**

**ESTUDIO FILOGEOGRÁFICO DE CEPAS *DE Salmonella*
enterica AISLADAS DE MUESTRAS AMBIENTALES EN
CULIACÁN, SINALOA**

Por:

M.C. José Roberto Aguirre Sánchez

TESIS APROBADA POR LA

COORDINACIÓN DE CIENCIAS Y TECNOLOGÍAS DE PRODUCTOS AGRÍCOLAS DE
ZONAS TROPICALES Y SUBTROPICALES

Como requisito parcial para obtener el grado de

DOCTOR EN CIENCIAS

APROBACIÓN

Los miembros del comité designado para la revisión de la tesis de José Roberto Aguirre Sánchez la han encontrado satisfactoria y recomiendan que sea aceptada como requisito parcial para obtener el grado de Doctor en Ciencias.



Dr. Cristóbal Chaidez Quiroz
Director de Tesis



Dr. Jaime Martínez Urtaza
Co-director de Tesis



Dr. José Andrés Medrano Félix
Integrante del comité de tesis



Dra. Nohelia Castro del Campo
Integrante del comité de tesis

DECLARACIÓN INSTITUCIONAL

La información generada en la tesis "Estudio Filogeográfico de Cepas de *Salmonella enterica* Aisladas de Muestras Ambientales en Culiacán, Sinaloa" es propiedad intelectual del Centro de Investigación en Alimentación y Desarrollo, A.C. (CIAD). Se permiten y agradecen las citas breves del material contenido en esta tesis sin permiso especial del autor José Roberto Aguirre Sánchez, siempre y cuando se dé crédito correspondiente. Para la reproducción parcial o total de la tesis con fines académicos, se deberá contar con la autorización escrita de quien ocupe la titularidad de la Dirección General del CIAD.

La publicación en comunicaciones científicas o de divulgación popular de los datos contenidos en esta tesis, deberá dar los créditos al CIAD, previa autorización escrita del director(a) de tesis.



AGRADECIMIENTOS

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por su apoyo económico durante la realización de mi proyecto de doctorado.

Al Centro de Investigación en Alimentación y Desarrollo A.C. (CIAD) por permitirme ser parte de su equipo y brindarme los espacios físicos para la realización de mi investigación. Agradezco infinitamente el apoyo intelectual y humanístico que recibí de cada miembro de CIAD, hago mención especial a los miembros de mi comité: Dr. Cristóbal Chaidez Quiroz, Dr. Jaime Martínez Urtaza, Dr. José Andrés Medrano Félix y Dra. Nohelia Castro del Campo.

Agradezco a todas las personas con las que coincidí y me brindaron algún aprendizaje, forman parte esencial de mí. A mi familia, por apoyarme en cada paso de mi vida y alentarme a ser mejor persona, ustedes son mi motor para seguir adelante y vencer todos los obstáculos que se me presentan.

A todos ustedes, mi más sincero agradecimiento.

DEDICATORIA

Este trabajo y todos los emanados durante mi formación como doctor en ciencias van dedicados a mi familia, especialmente a mi mamá/abuela por creer en mí y siempre apoyarme en cada decisión. Sin lugar a dudas, no sería la persona que soy si no hubiera sido por ella. Por lo que le dedico este esfuerzo y logro, que no es nada comparado con lo que ella ha hecho; sin embargo, espero poder regresarle, aunque sea un poco de todo lo que me ha dado.

También me la auto dedico, toda la vida he sido una persona que nunca se reconoce los logros o que trata de minimizarlos. Sin embargo, el día de hoy me permito aplaudirme y sentirme orgulloso de lo que he logrado con mi perseverancia, trabajo y pasión. No ha sido un camino fácil, pero siempre he dado pasos con un corazón lleno de esperanza y sueños.

CONTENIDO

APROBACIÓN.....	2
DECLARACIÓN INSTITUCIONAL	3
AGRADECIMIENTOS	4
DEDICATORIA	5
CONTENIDO	6
LISTA DE FIGURAS	8
LISTA DE CUADROS.....	9
RESUMEN	10
ABSTRACT	11
1. SINOPSIS.....	12
1.1. Justificación	12
1.2. Antecedentes.....	13
1.2.1. Características del Género <i>Salmonella</i>	13
1.2.1.1. Características Biológicas.	13
1.2.1.2. Características genéticas.	15
1.2.1.3. Taxonomía del género <i>Salmonella</i>	15
1.2.2. Epidemiología	17
1.2.2.1. Datos Epidemiológicos	17
1.2.3. Resistencia a Antimicrobianos.....	18
1.2.3.1. Resistencia a antibióticos por parte de <i>Salmonella</i>	18
1.2.4. Virulencia.....	20
1.2.4.1. Mecanismo de patogenicidad por parte de <i>Salmonella enterica</i>	20
1.2.4.2. Islas de Patogenicidad	22
1.2.4.3. Sistema de Secreción Tipo III (T3SS).	24
1.2.5. Caracterización de <i>Salmonella</i>	25
1.2.5.1. Métodos Fenotípicos.	26
1.2.5.1.1. Serotipificación	26
1.2.5.1.2. Fagotipificación.....	27
1.2.5.2. Métodos genotípicos	27
1.2.5.2.1. Tipificación Multilocus de Secuencias (MLST).	27
1.2.5.2.2. Electroforesis de campos pulsados (PFGE)	29
1.2.5.2.3. Análisis multi-locus de secuencias repetitivas en tándem de número variable (MLVA).	29
1.2.5.2.4. Secuenciación de genomas.....	30
1.2.6. Secuenciación y Bioinformática	31
1.2.6.1. Plataformas de secuenciación	31
1.2.6.1.1. Illumina	31
1.2.6.1.2. Nanopore	32

CONTENIDO (continuación)

1.2.6.2. Análisis de Genomas Bacterianos	32
1.2.6.2.1. Ensamblado de lecturas	33
1.2.6.2.2. Anotación funcional y estructural	35
1.2.6.2.3. Construcción de árboles filogenéticos.....	35
1.2.6.2.4. Construcción de pangenomas	37
1.2.6.2.5. Reloj molecular.	38
1.3. Hipótesis	40
1.4. Objetivo General.....	40
1.5. Objetivos Específicos	40
1.6. Sección Integradora	41
1.7. Referencias	42
2. GENOMIC SIGNATURES OF ADAPTATION TO NATURAL SETTINGS IN NON-TYPHOIDAL <i>Salmonella enterica</i> SEROVARS SAINTPAUL, THOMPSON AND WELTEVREDEN.....	53
3. GENOMIC-WIDE ANALYSIS APPROACH REVEALED GENOMIC SIMILARITY FOR ENVIRONMENTAL MEXICAN <i>S. Oranienburg</i> GENOMES.....	66
4. CONCLUSIONES GENERALES	98
5. RECOMENDACIONES	99

LISTA DE FIGURAS

Figura		Página
1	Clasificación del género <i>Salmonella</i>	16
2	Antígenos de superficie en <i>Salmonella enterica</i> . Somático (O), flagelar (H) y virulencia (Vi)	17
3	Mecanismo de patogenicidad e invasión exhibido por <i>Salmonella enterica</i>	22
4	Esquema del T3SS en bacterias patogénicas. Se conforma principalmente de una estructura inmersa en la membrana bacteriana, prolongándose hasta terminar en una especie de punta o aguja.....	26
5	Protocolo simple para la serotipificación de <i>Salmonella</i> mediante el uso de antisueros	27
6	Análisis MLST	31
7	Ensamblado <i>de novo</i> mediante gráficas de Bruijn	35
8	Niveles de estructuración en el ensamblaje <i>de novo</i>	35
9	Designación de los integrantes de un grupo taxonómico de acuerdo a si comparten un ancestro común	36
10	Clasificación de los métodos de reconstrucción filogenética en base al tipo de dato y método de reconstrucción	37
11	Componentes principales para el análisis molecular mediante reloj	40

LISTA DE CUADROS

Cuadro		Página
1	Descripción de las primeras 5 islas de patogenicidad en <i>Salmonella enterica</i>	24
2	Genes constitutivos utilizados para la asignación taxonómica mediante la técnica MLST basada en la comparación de perfiles alélicos.....	29

RESUMEN

Salmonella enterica es una bacteria patogénica clasificada como Gram-negativa, responsable de brotes epidemiológicos a nivel mundial por la ingesta de agua o alimentos contaminados por materia fecal, siendo esta su principal vía de transmisión. Dentro de las principales enfermedades se encuentran la salmonelosis y fiebre no tifoidea, como sintomatología general se presenta fiebre, vómito, dolor de cabeza y diarrea, principalmente. Históricamente, las infecciones ocasionadas por *Salmonella* no tifoidea eran atribuidas al consumo de alimentos de origen animal, sin embargo, hoy en día, existen múltiples casos de brotes epidemiológicos en México y el mundo por el consumo de hortalizas frescas. Aunque esta bacteria residente en el intestino de humanos y animales, se ha encontrado viable en ambientes no hospederos como agua de río. Aunado a esto, la transferencia por vía fecal-oral, es cada vez más frecuente debido a la mala manipulación y procesamiento de alimentos. El desarrollo tecnológico ha permitido el surgimiento de técnicas basadas en secuenciación y bioinformática, logrando el estudio detallado de la identidad genética de cada organismo, mejorando el entendimiento de su biología básica. Por lo que el objetivo de este trabajo de investigación consiste en conocer más sobre la huella genética de esta bacteria y su movilidad en la región de Culiacán, Sinaloa. Para ello se exploraron las firmas genéticas de adaptación entre aislados ambientales y clínicos, se analizó la estructura poblacional y la similitud genética entre las cepas de la región; y por último se realizó un reloj molecular para conocer el flujo de *Salmonella* por la región. Como resultado se encontró un grupo clonal con alta similitud genética entre los aislados, lo que indica una constante circulación de estas cepas. También se identificaron genes específicos a cepas ambientales que les permiten sobrevivir en condiciones ambientales, mientras que las clínicas mostraron mayor contenido en genes de virulencia. La duplicación genética y pseudogenización son procesos evolutivos importantes en *Salmonella* para lograr adaptarse en diferentes ambientes.

Palabras claves: *Salmonella*, genómica, ambiente, adaptación, filogeografía

ABSTRACT

Salmonella enterica is a pathogenic bacterium classified as gram-negative, causing agent of epidemiological outbreaks worldwide due to the ingestion of water and food contaminated by fecal matter, this has been the main transmission route. Among its clinical manifestations are salmonellosis and non-typhoid fever, with general symptoms including headache, diarrhea, and fever. Historically, infections caused by *Salmonella* were linked to animal food consumption. Nevertheless, today there are multiple cases of epidemiological outbreaks in Mexico associated to fresh producers' consumption. Although this bacterium is native from the human and animals' intestine, it has been found to be viable in non-host environments as river water. In addition to this, the transmission of *Salmonella* between people is increasing due to the mobility and international trading. Technological development has allowed the emergence of techniques based on sequencing and bioinformatics, achieving the detailed study of genetic identity for each organism. This improves our understanding of the basic biology aspects. Therefore, the objective of this research work is to learn more about genetic footprint of *Salmonella* and its mobility in the region. For this reason, the genetic signatures of adaptation between environmental and clinical isolates were explored, the population structure and the genetic similarity between the strains were analyzed; and finally, molecular clock was carried out to know the flow of *Salmonella* through the region. A clonal group with high genetic similarity was found among isolates, indicating a constant circulation of these strains. Genetic specific markers were found for environmental isolates, helping them to survive under these unfavorable conditions, while clinical isolates showed higher virulence content for survival inside host. Genetic duplication and pseudogenization are important evolutionary process in *Salmonella* to achieve adaptation in different environments.

Keywords: *Salmonella*, genomics, environment, adaptation, phylogenetics.

1. SINOPSIS

1.1 Justificación

Salmonella enterica es una de las principales bacterias patogénicas reportadas a nivel mundial por su presencia en agua y alimentos contaminados, siendo responsable de producir brotes epidemiológicos resultantes en fiebre tifoidea y salmonelosis no tifoidea. Reportes indican que esta bacteria ocasiona 1.35 millones de infecciones, 26,500 hospitalizaciones y 420 muertes en Estados Unidos de América, mientras que en México la cifra considerable 149,571 casos anuales. Tradicionalmente, las infecciones se transmiten por vía fecal-oral y se han asociado al consumo de alimentos contaminados de origen animal; sin embargo, hoy en día, existe un incremento en el número de brotes epidemiológicos asociados al consumo de hortalizas frescas contaminadas con *S. enterica*. Esto puede deberse al uso de agua contaminada como parte del proceso de irrigación, así como deficiencias en las buenas prácticas agrícolas y de manufactura. Así mismo, es bien conocido que este patógeno es un habitante natural del intestino humano; no obstante, se ha observado una alta prevalencia en ambientes no hospederos como el agua de río permitiendo hipotetizar que *S. enterica* se encuentra en constante movimiento entre hospederos y ambientes no hospederos, y que además cuenta con elementos genéticos que le permiten desarrollar estrategias de supervivencia y adaptación en el ambiente. Lo que supone un riesgo importante a la salud pública y sector económico. Pese a todo, el número de trabajos enfocados en estudiar los aspectos genéticos de *S. enterica* como su homología, firmas genéticas de adaptación y filogeografía, siguen siendo temas poco explorados. Por lo que el presente trabajo de investigación se enfocó en construir un marco de referencia genómico con las principales características de *S. enterica* aisladas de muestras ambientales de la región de Culiacán, Sinaloa, mediante una aproximación bioinformática. Lo anterior, aportará información de relevancia para mejorar el entendimiento de la biología básica y dinámica poblacional de dicha bacteria en ambientes no hospederos; lo que resulta de importancia para el sector científico, salud y económico.

1.2 Antecedentes

1.2.1 Características del Género *Salmonella*

1.2.1.1 Características Biológicas. El género *Salmonella* pertenece a la familia *Enterobacteriaceae*, es un bacilo gran-negativo que se encuentra dentro del grupo de las bacterias entéricas; donde también se ubican *E. coli* y *Shigella* (Baker & Dougan, 2007). Este género fue nombrado así en honor al veterinario Daniel E. Salmon quien aisló por primera vez dicha bacteria en 1855 en el intestino de cerdos infectados con cólera, posteriormente esta bacteria fue nombrada *Salmonella choleraesius* (Le Minor & Popoff, 1987).

Esta bacteria es originaria del tracto gastrointestinal de humanos y animales de sangre caliente, caracterizándose por infectar un amplio rango de hospederos (Winfield & Groisman, 2003). *Salmonella* figura entre los principales patógenos responsables de enfermedades transmitidas por los alimentos (ETAs), siendo transmitida mediante la vía fecal-oral a través de la ingesta de alimentos contaminados, dentro de los que destacan huevos, productos lácteos y carnes crudas; aunque de forma general, comida proveniente de animales como el puerco, aves de corral y res (Voetsch et al., 2004).

Salmonella es serológicamente diversa; existen 6 especies y su adaptación a diferentes condiciones ambientales, nutrimentales, entre otras, han llevado a que se encuentre ampliamente distribuida en hospederos de sangre caliente tales como aves y mamíferos; así como de sangre fría, destacando reptiles y anfibios (Eng et al., 2015). La adaptación de *Salmonella* a diferentes ambientes hospederos y no hospederos puede ser multiespecie, pero en algunas ocasiones ciertos serotipos de *Salmonella* son restringidos a hospederos, como es el caso de *S. Typhi*, *S. Paratyphi* A, B y C; que residen solamente en humanos (White, 2002), así como en aves de corral donde los serotipos *S. Gallinarium* y *S. Pullorum* son restringidos a hospederos, responsables de causar tifoidea aviar y pullorosis (diarrea blanca) (Alvarado & Isac, 2022). Por otra parte, también existen serotipos no tifoideos, los cuales no representan especificidad por algún hospedero en particular, destacando *S. Typhimurium*, *S. Enteritidis* y *S. Heidelberg*, los cuales se han encontrado en el ambiente y distintos tipos de animales de granja tales como bovinos, porcinos y caprinos (Andino & Hanning, 2015).

Para poder sobrevivir, esta bacteria debe ser capaz de infectar a un hospedero, adaptarse y persistir en su paso por el ambiente para posteriormente entrar en un nuevo hospedero susceptible, desarrollándose y multiplicándose en el recién infectado (Palmer & Slauch, 2017). Una de las rutas más importantes para la transmisión de este patógeno es la vía fecal-oral, donde el patógeno abandona al huésped a través de las heces de humanos, animales de granja, mascotas y animales salvajes, donde otro hospedero entrará en contacto con el microorganismo mediante la ingesta; por lo que cualquier ruta que tome la materia contaminada (agua o alimentos) con heces y llegue directamente o indirectamente a la boca del alguien, podría transmitir el patógeno (Baudart et al., 2000; Winfield & Groisman, 2003).

La supervivencia de *Salmonella* en ambientes no hospederos (es decir fuera del hospedero), está determinada por una serie de factores, dentro de los cuales se incluyen: temperatura, humedad, tipo de suelo, presencia de plantas y exposición a la luz solar (UV) (Velge et al., 2012), depredación por protozoarios y número de microorganismos, así como su maleabilidad genómica que le permite incorporar, eliminar o reordenar genes para ajustarse a los diferentes ambientes (Mouslim et al., 2002). Además, esta bacteria puede ser acarreada o transportada a aguas superficiales por medio de lluvias y escorrentías superficiales donde permanece viable e inclusive reproducirse (Fatica & Schneider, 2011). Así, se cuenta con amplia información científica sobre su paso por ecosistemas acuáticos, los cuales sirven como reservorios y consecuentemente como medio de transmisión a un nuevo hospedero.

Comparada con otras enterobacterias, se ha reportado que *Salmonella* tiene una alta tasa de supervivencia en ambientes acuáticos; y aún más cuando hay contaminación o convergencia de tuberías de desagüe correspondiente a agua residual urbana a cuerpos acuáticos (Chao et al., 1987). Por ejemplo, en condiciones de temperatura que oscila entre 20 y 30°C, *Salmonella* puede permanecer viable hasta por 30 días (Steele & Odumeru, 2004); sin embargo, como resultado de las condiciones adversas propias de los ambientes no hospederos, el tiempo de supervivencia se reduce a menos de 2 semanas (Jacobsen & Bech, 2012). Estudios recientes han identificado que *Salmonella* es capaz de sobrevivir y multiplicarse en agua destinada a irrigación de cultivos que presentan rangos amplios de pH (4.5-9.5) y temperatura (7-48°C) (H. Liu et al., 2018).

El periodo de supervivencia de *Salmonella* en agua es mayor a las dos semanas, por lo que se puede encontrar en ríos, lagos y océanos que han sido contaminados con material fecal; mientras que en ambientes acuáticos es de menos de 30 días debido a las condiciones adversas de los factores

presentes en ambientes no hospederos. En caso de otro tipo de ambientes como por ejemplo, en agua estéril o buffer fosfatado, se ha reportado que *Salmonella* puede sobrevivir hasta por 5 años (H. Liu et al., 2018).

Sin embargo, *Salmonella* ha desarrollado una serie de estrategias que le permiten adaptarse y sobrevivir a condiciones adversas propias de ambientes no hospederos, resaltando mecanismos como la formación de biopelículas, siendo comunidades estructuradas de células bacterianas embebidas en una matriz polimérica las cuales se encuentran adheridas a superficies inertes (Høiby et al., 2010) confiriéndoles resistencia al estrés ambiental y la oportunidad de sobrevivir para su transmisión a nuevos huéspedes (Homøe et al., 2009).

1.2.1.2 Características genéticas. Con respecto a las características genéticas, *Salmonella* tiene un cromosoma circular con un tamaño de genoma de aproximadamente 4.3-5.0 Mpb, el cual puede variar de acuerdo a su serotipo. Este presenta un promedio de ~4,500 genes; sin embargo algunas cepas pueden tener material extracromosomal en forma de plásmidos, el cual generalmente contiene genes asociados a virulencia y resistencia a antibióticos (Baker & Dougan, 2007). Recientemente, se ha acuñado el término pangenoma para referirse al conjunto de los genes de una especie (Vernikos et al., 2015). Éste se encuentra integrado por el genoma core, agrupando los genes compartidos por la mayoría de los aislados que tienen implicaciones en la biología básica de *Salmonella*; y el genoma accesorio, siendo los genes exclusivos para un número limitado de cepas, que pueden tener un papel importante en la diferenciación y adaptación de este patógeno a una vasta diversidad de ambientes. El genoma core de *Salmonella* se encuentra constituido por una región de 3.8 Mbp compartidos por el 96% de 4,893 genomas analizados (Laing et al., 2017), lo que se traduce a un promedio aproximado de 2,811 genes compartidos. Por su parte, el genoma accesorio, se encuentra asociado a genes responsables de la supervivencia a ambientes específicos (Delmont & Eren, 2018).

1.2.1.3 Taxonomía del género *Salmonella*. El género *Salmonella* se encuentra dividido en dos especies: *S. enterica* y *S. bongori*; a su vez la especie *enterica* se encuentra dividida en 6 sub-especies que están designadas por nombre, o más comúnmente, por números romanos (Figura 1).

Sin embargo, la gran mayoría de los serotipos de *Salmonella enterica* se encuentran dentro de la sub-especie *enterica*, siendo ampliamente conocidos por infectar humanos y animales (Selander et al., 1996).

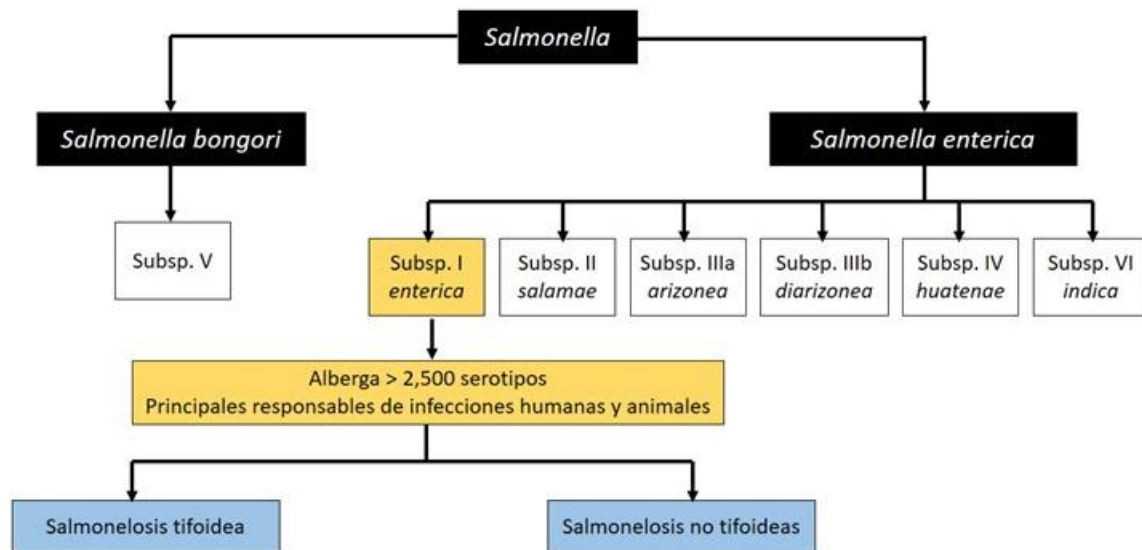


Figura 1. Clasificación del género *Salmonella* (creación propia)

A su vez, cada sub-especie está dividida en serotipos, de acuerdo a la reacción inmunitaria (aglutinación) con los antígenos de superficie (Fig. 2) mediante el uso de antisueros específicos, según el esquema propuesto por White-Kauffman. Dentro de los antígenos presentes en *Salmonella* podemos encontrar:

- Antígeno O (somático): es codificado por el gen *wba*, el cual está constituido por una cadena repetida de polisacáridos, formando parte del lipopolisacárido, que se genera y sobresale de la membrana externa actuando como barrera de protección a agentes externos en bacterias Gram-negativas (Grimont & Weill, 2007).
- Antígeno H (flagelar): está formado por la proteína más abundante del flagelo (estructura que permite movimiento) “la flagelina”, pudiéndose encontrar como mono o difásicas (cuando se encuentra la misma representación del antígeno o cuando el antígeno puede tener una forma alternativa). Las cuales son codificados por los genes *fliC* y *fliB* (Grimont & Weill, 2007).
- Antígeno Vi (virulencia): se refiere a la producción de un polisacárido capsular de

virulencia, el cual se ha asociado a la fiebre tifoidea en humanos (Wain et al., 2005). Este antígeno sólo está presente en *S. Typhi*, *S. Paratyphi* y *S. Dublin*, confiriendo resistencia contra la respuesta inmune celular y humoral del huésped, y se compone de un homopolímero lineal formado por unidades α -1,4-N acetilgalactosaminuronato (Zhang et al., 2003).

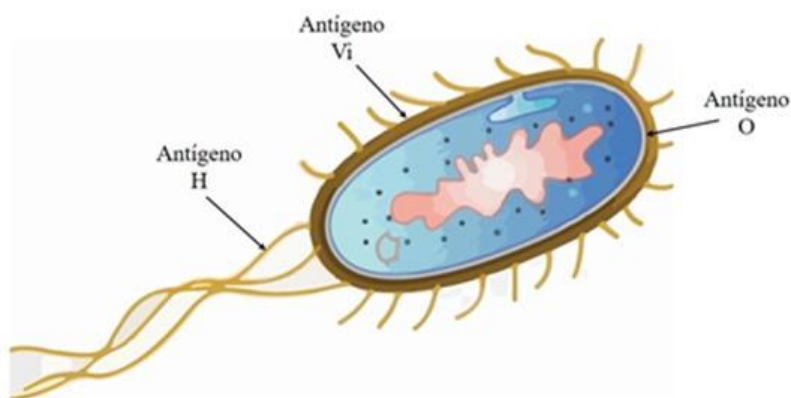


Figura 2. Antígenos de superficie en *Salmonella enterica*. Somático (O), flagelar (H) y virulencia (Vi).

1.2.2 Epidemiología

1.2.2.1 Datos Epidemiológicos. *Salmonella* es una bacteria responsable de generar enfermedades de transmisión alimentaria (ETAs) o también conocidas como intoxicaciones alimentarias, las cuales están ampliamente extendidas a nivel mundial (Jajere, 2019), encontrándose dentro de las 4 principales causas de diarrea a nivel mundial. En los últimos años, los brotes de salmonelosis se han asociado al consumo de alimentos contaminados de origen animal y vegetal.

Bajo una perspectiva global, la Organización Mundial de la Salud (2018), reporta que cada año aproximadamente una de cada 10 personas contrae alguna enfermedad asociada a *Salmonella*, perdiéndose aproximadamente 33 millones de vidas humanas anualmente. Por su parte, el Centro para el Control y Prevención de Enfermedades (2022), estima que dicha bacteria patógena es responsable de producir 1.35 millones de infecciones, 26,500 hospitalizaciones y 420 decesos

solamente en Estados Unidos de América, siendo los alimentos contaminados con heces fecales la principal fuente de brotes epidemiológicos. Por otra parte, la Unión Europea, a través de la Autoridad de Seguridad Alimentaria Europea (2021) ha registrado 100 mil casos de infecciones anuales, teniendo una carga económica de 3 mil millones de Euros al año asociados a tratamientos y estadías prolongadas en hospitales.

En lo que respecta a México, el Sistema Nacional de Vigilancia Epidemiológico (SINAVE) reporta que en el periodo comprendido entre 1984 a 2017, hubo un aumento considerable de fiebre tifoidea, pasando de 7,629 a 45,280 (Contreras-Soto et al., 2018); mientras que para fiebre paratifoidea y otras salmonelosis de 31,943 a 104,471 casos, respectivamente; mostrando a estados como Sinaloa y Tamaulipas como los estados de mayor índice de fiebre tifoidea con 12.9% y 10.25% respectivamente, mientras que para el caso de fiebre tifoidea y otras salmonelosis, destacan los estados de Chiapas (12.99%), Veracruz (9.37%) y Tabasco (8.31%) corresponden a los estados con mayor índice de casos (Contreras-Soto et al., 2018; Zaidi et al., 2006).

Adicionalmente, la segunda causa de morbilidad en México durante el año 2014, fue debido a las infecciones intestinales por otros microorganismos y las mal definidas con una tasa de 41.3 casos por cada 1,000 habitantes; donde podrían encontrarse las infecciones ocasionadas por *Salmonella*, ya que raramente se suele realizar un diagnóstico específico y la sintomatología ocasionada, hacen sospechar con la misma verosimilitud, dos o más enfermedades (Soto-Estrada et al., 2016).

De acuerdo al Boletín Epidemiológico Semanal No. 52 que comprende del 25 al 31 de diciembre del 2022, Sinaloa ocupa el primer lugar por concepto fiebre tifoidea con 3,334 casos reportados, también se tienen registrados 672 casos de paratifoidea y 1833 casos de otras salmonelosis. (SINAVE, 2022). Sin embargo, existen pocos reportes específicos que señalen a *Salmonella* no tifoidea como agente etiológico; y mucho menos, el serotipo responsable. Es por ello que la caracterización y trazabilidad de *Salmonella* juegan un papel importante en la prevención y contención de brotes epidemiológicos, y, por lo tanto, casos clínicos.

1.2.3 Resistencia a Antimicrobianos

1.2.3.1 Resistencia a antibióticos por parte de *Salmonella*. De acuerdo a CDC, la resistencia a antibióticos es uno de los más grandes problemas actuales que enfrentan las instituciones de salud

pública a nivel mundial. En Estados Unidos se ha estimado que al menos 2 millones de personas presentan alguna infección resistente a antibióticos y al menos 23,000 personas mueren a causa de ello (CDC, 2022). Por su parte la OMS reporta que la resistencia hace que se incrementen los costos médicos, que se prolonguen las estancias hospitalarias y que aumente la mortalidad. Adicionalmente, existe un creciente número de infecciones como la neumonía, tuberculosis y ETAs, que son cada vez más difíciles de tratar a medida que los antibióticos van perdiendo eficacia (Serra Valdés, 2017).

La resistencia a antibióticos se da cuando las bacterias u hongos han desarrollado la habilidad de suprimir o evadir el efecto de antibióticos (medicamentos utilizados para combatir las infecciones bacterianas); por lo que estos no mueren y continúan creciendo. Este estado puede afectar a cualquier persona en cualquier etapa de desarrollo.

Salmonella ha sido asociada a más de 1 millón de casos de enfermedad anualmente, de los cuales se estima que en 100,000 se deben a alguna resistencia de algunos de los antibióticos de primera línea para combatir esta bacteria como: ceftriaxona y ciprofloxacina. Sin embargo, la existencia de cepas con multirresistencia en más de 5 antibióticos han generado más de 66,000 casos de enfermedad del 2009 al 2011 en Estados Unidos (Ventola, 2015).

Entre algunos de los casos reportados de resistencia en *Salmonella* se encuentra el brote del 2015 en pepino, donde se determinó que los aislados de *S. Poona* presentaron resistencia a tetraciclina y ácido nalixídico. También se encontró una baja susceptibilidad a la ciprofloxacina, siendo un antibiótico importante para combatir las enfermedades ocasionadas en niños por *Salmonella* (Prevention, 2015).

Por otra parte, se detectó resistencia a ampicilina y tetraciclina por parte de *S. Enteritidis* aislada de pollo crudo y congelado (Smith et al., 2008). En el año 2014 *S. Heidelberg* se encontró involucrada en un brote epidemiológico en pollo, de los cuales el 67% de las cepas involucradas presentaron multirresistencia.

Gracias a la plasticidad genética que poseen las bacterias, les ha permitido adaptarse a diferentes condiciones ambientales, así como tener estrategias que les permiten evadir la acción de los antibióticos (Antony et al., 2018). Generalmente las estrategias utilizadas, se pueden englobar en la mutación de genes y la adquisición de material genético exógeno mediante transferencia horizontal (HGT) (Blair et al., 2015).

El primer mecanismo se refiere a cuando una población bacteriana da lugar a descendencia con algún cambio o mutación en ciertos genes; de tal forma que la nueva población al momento de ser

sometida a algún antibiótico, va a perpetuar; mientras que la población inicial morirá por falta de estos genes (McDermott et al., 2016). Por lo que, a partir de la generación sobreviviente, la mutación se fijará en la población. Las mutaciones en los genes pueden hacerles frente a los antibióticos mediante la disminución de la afinidad del antibiótico, la cual puede ocurrir mediante diferentes mecanismos tales activación de bombas de flujo, cambios metabólicos y secreción de enzimas específicas (Martinez & Baquero, 2000).

Mientras que HGT se refiere al aporte genético que tiene un organismo como producto del intercambio o aceptación de material genético de un organismo o fuente ajena a la población originaria. Este mecanismo juega un papel muy importante en la diseminación de la resistencia para las especies de *Salmonella* (Oladeinde et al., 2019). Estos genes de resistencia pueden estar presentes tanto en plásmidos (material genético circular extra en bacterias), o bien incluidos en el cromosoma bacteriano. La transmisión mediada por plásmidos es el mecanismo más eficiente para la transferencia de dichos genes, pudiendo involucrar la transferencia de más de un gen (Carattoli, 2003). Así mismo, existen otros mecanismos que podrán tener un aporte a la resistencia a antibióticos por parte de *Salmonella*; como lo son los transposones (material genético móvil), integrones y fagos (Domingues et al., 2012).

Debido a los riesgos por patogenicidad y resistencia a antibióticos expuestos anteriormente, resulta de importancia caracterizar a la cepa de *Salmonella* involucrada para contrarrestar sus efectos y establecer medidas que disminuyan la probabilidad de la generación de brotes epidemiológicos. Por lo que, a través del tiempo, se han utilizado métodos fenotípicos y genotípicos para realizar dicha tarea. Por ejemplo, en Texas la detección de *Salmonella* se ha realizado mediante el empleo de técnicas como la electroforesis de campo pulsado con la finalidad de generar una huella genética y determinar si las cepas son responsables de generar brotes epidemiológicos; o bien, detectar vehículos de transmisión (Long et al., 2010).

1.2.4 Virulencia

1.2.4.1 Mecanismo de patogenicidad por parte de *Salmonella enterica*. Las infecciones ocasionadas por *Salmonella* comienzan cuando es ingerida por medio de agua o algún alimento contaminado con material fecal que contenga al microorganismo. Aquí *Salmonella* experimenta severos cambios

ambientales cuando entran al hospedero por vía oral como, por ejemplo: pH ácido, aumento de temperatura, baja tensión de oxígeno y alta osmolaridad; a lo que responden mediante modulaciones en la expresión de los genes (Zhang et al., 2003).

Salmonella inicia su ciclo de infección invadiendo al hospedero a través del tejido linfoide y adhiriéndose a las células epiteliales de íleon y a las células M, ya que representan una entrada ideal para las enterobacterias. Esto lo logra mediante un mecanismo conocido como disparo, donde la bacteria envía señales a las células epiteliales, induciendo rearrreglos en el citoesqueleto para la formación de pliegues en la superficie (Figura 3). Para esto, se han identificado diversas proteínas efectoras de la isla de patogenicidad I (SPI) involucradas en dichos re-arreglos, entre las que destacan SipA, SopE y SopB (Eng et al., 2015).

Después de la invasión al epitelio, la bacteria se multiplica intracelularmente llegando a los ganglios linfáticos, donde se distribuye a todo el cuerpo mediante el sistema circulatorio. Dependiendo del serotipo y de la efectividad que esta tenga al hospedero, puede causar daños en hígado, vesícula, entre otros. Una vez invadido el intestino, se puede desencadenar una respuesta inflamatoria dando lugar a la formación de úlceras (Figueira & Holden, 2012).

En cada paso del proceso infeccioso, *Salmonella* responde a los cambios en su entorno mediante la expresión de ciertos genes que le permiten sobrevivir a las condiciones desfavorables (Jones, 1997). A estos genes se les conoce como factores de virulencia y en esencia le confieren la habilidad patogénica a *Salmonella* para que sea capaz de desencadenar un proceso infeccioso y colonizar a un hospedero en específico (Wu et al., 2008).

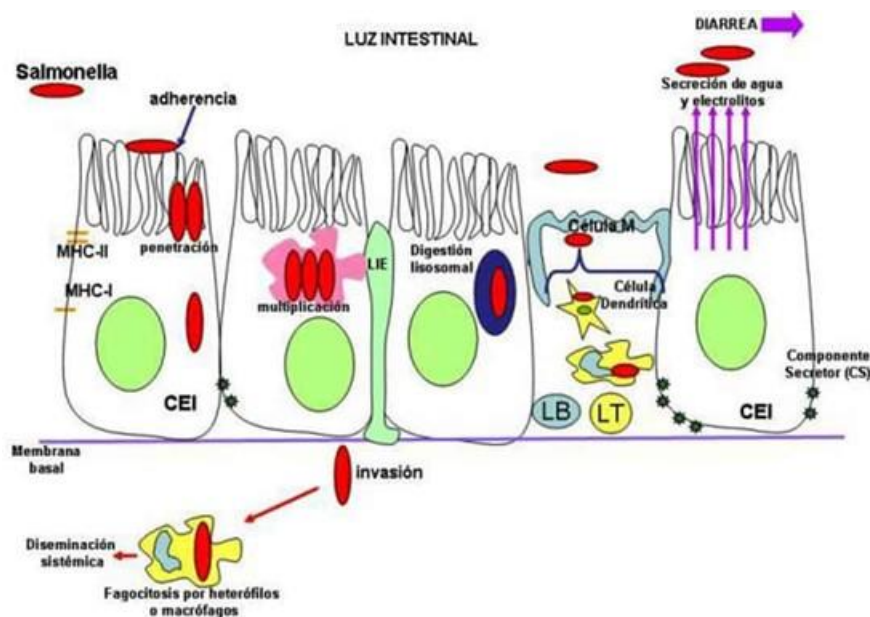


Figura 3. Mecanismo de patogenicidad e invasión exhibido por *Salmonella enterica* (tomada de <https://www.veterinariadigital.com/noticias/salmonella-colonizacion-del-tracto-gastrointestinal-por-dra-liliana-revolledo/>).

1.2.4.2 Islas de Patogenicidad. Virulencia es el grado de patogenicidad que presenta un microorganismo, el cual se indica por la severidad de la enfermedad producida y su habilidad de invadir tejidos de un hospedero (Ryan, 2016). Mientras que los factores de virulencia se refieren a las propiedades (productos génicos) que le permiten a un organismo establecerse en un huésped desencadenando una enfermedad (Chen et al., 2012). Entre estos factores de virulencia se incluyen proteínas de adhesión a superficies celulares, toxinas, enzimas hidrolíticas, carbohidratos y proteínas para la protección de bacterias (Al-Khafaji et al., 2021).

Los genes de virulencia más importantes son aquellos que se localizan en las Islas de Patogenicidad de *Salmonella* (SPI). Actualmente, se tiene conocimiento de 17 SPIs; sin embargo, de acuerdo a los estudios intensivos que se han hecho a través del tiempo en *S. Typhimurium*, se ha reportado que las primeras 5 SPIs (Tabla 1), corresponden a las que tienen mayor impacto en la interacción con hospederos (Marcus et al., 2000). Adicionalmente, se han reportado que existen otros mecanismos o elementos que pudieran tener un aporte a la virulencia de *Salmonella*; donde se destacan los plásmidos de virulencia y las proteínas relacionadas con la formación de biopelículas (Mazurkiewicz et al., 2008).

La mayoría de estas SPIs, se encuentran flanqueadas por genes de ARNt y difieren respecto al contenido G+C del resto del genoma; lo que sugiere que pudieron ser adquiridas por procesos de

transferencia horizontal mediante plásmidos, bacteriófagos, transposones, elementos de inserción o recombinación de ADN exógeno con el del hospedero (Hacker & Carniel, 2001).

A continuación, se describe la funcionalidad de las primeras 5 SPIs. La SPI-1 codifica para el sistema de secreción tipo III clase I (T3SS-I), que sirve para la translocación de proteínas efectoras a la célula eucariótica. Estas proteínas median la invasión de las células no fagocíticas por *Salmonella* (Galán, 2001), proceso que involucra la modificación del citoesqueleto y son expresadas en respuesta a las señales percibidas por las bacterias en el intestino del hospedero infectado (Lostroh & Lee, 2001). Adicionalmente, se ha identificado la función de genes que conforman esta isla de patogenicidad mediante la inducción de mutaciones. Por ejemplo, el gen *InvA* se determinó que tiene un papel importante en la invasión de este patógeno a las células epiteliales (Ochoa & Rodríguez, 2005).

SPI-2 se compone de dos elementos distintos:

- El segmento más grande (25 Kb): Siendo esencial para desencadenar la patogenicidad y se encuentra solamente en especie *Salmonella enterica*. Este segmento codifica para los genes de T3SS clase II, sistema regulatorio, de chaperonas y efectores.
- El segmento pequeño (15 Kb): Es responsable de la virulencia sistemática y codifica para 5 genes *ttr*, los cuales son responsables de la utilización del tetraionato como aceptor de electrones para llevar a cabo la respiración anaerobia de la bacteria (Figueira & Holden, 2012).

La distribución de SPI-3 entre las especies de *Salmonella* varía; sin embargo, todas consisten en 10 marcos de lectura abiertos (ORFs) organizados en 6 unidades transcripcionales. De los genes más importantes presentes en esta isla se encuentra *mgtC*, que codifica para el sistema de transporte de magnesio de alta afinidad, requerido para la supervivencia intramacrófago y el crecimiento bacteriano en condiciones limitantes de iones Mg^{2+} (Ochoa & Rodríguez, 2005).

El papel de la SPI-4 aún no se ha analizado completamente, sin embargo, se han realizado trabajos en modelos animales, observando que esta isla contribuye a la inflamación intestinal en modelo murino y como factor de colonización durante la fase de infección en becerros (Gerlach et al., 2007). También se ha determinado que dicha isla media la adhesión a células epiteliales. SPI-4 se compone aproximadamente de 6 genes, los cuales se han designado como *siiA-F* por ser genes que tiene un rol central en las infecciones intestinales (Gerlach et al., 2007).

En lo que respecta a SPI-5, fue identificada por primera vez en *Salmonella* Dublin y se compone

de 5 genes *pipA-B*, involucrados en la de la respuesta inmune en ganado; sin embargo, se ha identificado que no tiene una función significativa en gallinas (Stevens et al., 2009). Adicionalmente, SPI-5 codifica proteínas efectoras necesarias para SPI-1 y SPI-2. Por ejemplo; el gen *sopB* codifica para una proteína efectora de T3SS (Hensel, 2004).

Cuado 1. Descripción de las primeras 5 islas de patogenicidad en *Salmonella enterica*.

Isla de patogenicidad	Tamaño (Kb)	Contenido G+C	Fenotipo	Función	ARNt adyacentes
SPI-1	46	52%	Codifica las proteínas necesarias para ensamblar el complejo de secreción tipo III	Translocación de proteínas efectoras	No se encuentra asociada a un ARNt
SPI-2	41	43%	Codifica para un segundo sistema de secreción	-Patogénesis sistémica -Respiración anaerobia -Crecimiento intracelular	<i>valK</i> ARNt
SPI-3	17	47.5%	Codifica el gen <i>MgtCB</i> para aumentar la afinidad por el Mg^{2+}	Adaptación a las condiciones limitantes de nutrición	<i>serC</i> ARNt
SPI-4	28	50%	Codifica para el tipo de secreción I	Secreción de toxinas	<i>ssb</i> ARNt
SPI-5	9	43.6-52%	Codifica proteínas efectoras para SPI-1 y SPI-2	Asociada a enteropatogénesis	<i>serT</i> ARNt

Fuente: (Hensel, 2004; B. Liu et al., 2019)

1.2.4.3 Sistema de Secreción Tipo III (T3SS). Este tipo de sistema de secreción se encuentra en bacterias Gram-negativas, permitiéndoles la translocación de proteínas efectoras dentro de la célula huésped (Cabañas-Romero & Huerta-Saquero, 2014). T3SS se componen de una estructura especializada parecida a una jeringa compuesta por múltiples anillos que se extienden a través de la envoltura bacteriana (Figura 4) (Galán, 2001). Fue identificada originalmente en *Salmonella Typhimurium*, pero se ha reportado que dicho sistema es codificado por otras bacterias (Blocker et al., 2001).

El T3SS consiste principalmente en una base que cubre la envoltura bacteriana y una aguja que termina en estructura de punta. La bacteria secreta dos proteínas translocatorias, que se introducen en la membrana de la célula huésped para la formación de perforaciones o poros, permitiendo la transferencia de otras proteínas necesarias para el patógeno (Lee et al., 2014).

El principal objetivo del sistema T3SS es inyectar proteínas bacterianas denominadas efectoras, las cuales pasan al citoplasma del hospedero, afectándola señalización intracelular, la dinámica del citoesqueleto y el tráfico vesicular de la célula a favor del patógeno, lo que constituye un paso importante para el desarrollo de la enfermedad en el huésped (Lee et al., 2014).

Este es un sistema complejo que se compone de más de 20 proteínas estructurales. Así mismo, se ha reportado que este sistema de secreción se encuentra conservado en *Salmonella enterica*, por lo que la estructura está presente de manera generalizada en todo el género (González-Pedrajo & Dreyfus, 2003).

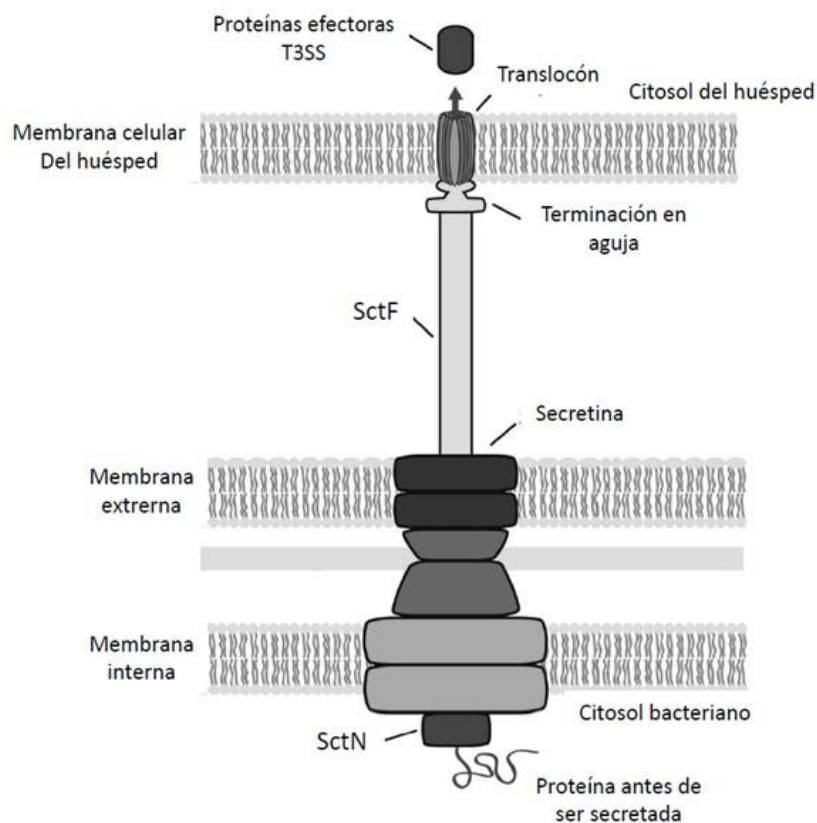


Figura 4. Esquema del T3SS en bacterias patógenas. Se conforma principalmente de una estructura inmersa en la membrana bacteriana, prolongándose hasta terminar en una especie de punta o aguja (Duncan et al., 2012).

1.2.5 Caracterización de *Salmonella*

1.2.5.1 Métodos Fenotípicos. Se refieren al conjunto de técnicas basadas en demostrar presencia o ausencia de actividad biológica o metabólica expresada por el organismo de interés, resultando de utilidad para la caracterización de agentes etiológicos. Entre las técnicas más utilizadas para la tipificación de este género bacteriano se encuentran las reacciones inmunitarias (serotipificación), fagotipificación y la utilización de marcadores bioquímicos (biotipificación) (Ferrari et al., 2017).

1.2.5.1.1 Serotipificación. Se refiere a la clasificación de los serotipos de acuerdo a la combinación de los antígenos superficiales O, H y Vi según el esquema propuesto por White-Kauffmann. Según dicho esquema, el género se divide en 67 serogrupos, designados con letras y posteriormente con números. La fórmula antigénica de una *Salmonella* se expresa como: antígeno somático O: antígeno flagelar 1: antígeno flagelar de la fase 2; por lo que la estructura antigénica para un serotipo Aberdeen sería O:11 H: i : 1,2. (Grimont & Weill, 2007)

La serotipificación se realiza después de la identificación de la especie, en un cultivo de aislado de *Salmonella* en un medio no selectivo u otro agar recomendado para dicho fin. Una vez confirmada la identidad del cultivo se procede a dejar caer una gota del antisuero (anti-O y anti-H) en una tira de aglutinación y posteriormente se agrega el cultivo puro (Fig. 5). Se mezcla ligeramente en forma circular por aproximadamente 30 segundos y se observa si ocurrió aglutinación (Schrader et al., 2008).

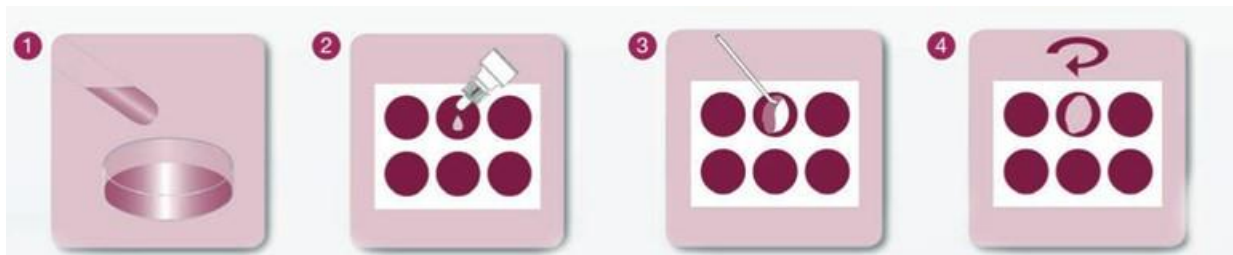


Figura 5. Protocolo simple para la serotipificación de *Salmonella* mediante el uso de antisueros. 1) Confirmación del cultivo puro. 2) Agregado de antisuero y colonia microbiana en tira de aglutinación. 3) Mezclado de elementos agregados. 4) Movimientos circulares para observar aglutinación. Imagen tomada de Bacterial Serotyping Guide for *Salmonella* BIO-RAD (https://www.bio-rad.com/webroot/web/pdf/fsd/literature/FSD_14-0699.pdf).

La serotipificación es una técnica sencilla y estable que permite la identificación del serotipo y uno

de los principales marcadores epidemiológicos de este género; sin embargo, los resultados pueden ser difíciles de interpretar, por lo que puede haber conflictos en la estandarización, inclusive entre los laboratorios de referencia (Gonzalez Pedraza et al., 2014).

1.2.5.1.2 Fagotipificación. Se le conoce como fagos a los virus que infectan bacterias, produciendo en la mayoría de los casos, su eliminación. Debido a su alta especificidad, cada fago es capaz de infectar y lisar solamente a una cepa determinada de una bacteria o cepas muy similares dependiendo de las características moleculares del fago y de los receptores presentes en la superficie de la bacteria, incluyéndose en un mismo grupo denominado fagotipo (Schmieger, 1999). Para el caso del género de *Salmonella*, se ha utilizado esta técnica para tipificar las cepas en función de que sean receptivas a la lisis de uno u otro fago.

Sin embargo, la realización de esta técnica requiere un amplio catálogo de fagos que no suele comercializarse y cuyo acceso se limita a los laboratorios de referencia. Además, no todos los aislados pueden ser caracterizados, ya que existen diferentes mecanismos que pueden provocar un cambio en el fagotipo. Actualmente el fagotipado se utiliza en los laboratorios de referencia como marcador epidemiológico secundario (Rabsch, 2007).

1.2.5.2 Métodos genotípicos. Permiten la diferenciación de especies bacterias, así como su parentesco genético. Estos métodos presentan mayor sensibilidad, mejores niveles de estandarización, reproducibilidad y poder discriminatorio. Entre los métodos más utilizados se encuentra la tipificación multilocus de secuencias (MLST), electroforesis de campos pulsados (PFGE), análisis de repeticiones en tándem en multilocus de número variable (MLVA), repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) y secuenciación de genomas completos (WGS) (Ferrari et al., 2017).

1.2.5.2.1 Tipificación Multilocus de Secuencias (MLST). MLST es una técnica genética para la caracterización taxonómica de bacterias, la cual se basa en la obtención de 7 fragmentos específicos de genes constituyentes o también conocidos como house-keeping; los cuales son vitales para la

supervivencia de *Salmonella* (Tabla 2) o de algún otro género bacteriano y por lo tanto, dichos genes siempre se deberán estar expresando. Las secuencias de cada fragmento de los genes son comparadas con las secuencias previamente identificadas (alelos), asignándoles un valor numérico para cada uno de los 7 alelos. La combinación de esos 7 valores numéricos define el perfil alélico al cual se le asigna un tipo de secuencia (ST), siendo este utilizado para describir a la cepa (Aanensen & Spratt, 2005).

Esta técnica se ha utilizado en estudios de epidemiología por al menos durante dos décadas para categorizar los diferentes grupos clonales (grupo de organismos genéticamente emparentados) de grupos bacterianos basándose en los comparativos de perfiles alélicos. Por lo que se han desarrollado un número importante de herramientas computacionales para calcular el ST a partir de datos provenientes de secuenciación de siguiente generación (NGS) (Page et al., 2017).

Cuadro 2. Genes constitutivos utilizados para la asignación taxonómica mediante la técnica MLST basada en la comparación de perfiles alélicos.

Loci	aroC	dnaN	hemD	hisD	purE	sucA	thrA
# Alelos	840	828	787	1167	879	858	983
Nombre del gen	Chorismate synthase	Beta sliding clamp	Uroporphyrinogen III methyltransferase	Histidinol dehydrogenase	Phosphoribosylaminoimidazole carboxylase	2-oxoglutarate dehydrogenase	Aspartokinase

La ventaja de MLST es que los datos provenientes de las secuencias no son ambiguos y los perfiles alélicos de los aislamientos pueden comprarse fácilmente con bases de datos de libre acceso. Adicionalmente, los perfiles alélicos también se pueden obtener a partir de material clínico mediante la amplificación por PCR de los genes constitutivos, por lo que los aislados pueden cultivarse incluso cuando no pueden cultivarse a partir de material clínico (Urwin & Maiden, 2003). También, se ha reportado que este tipo de análisis tiene mayor poder de resolución y objetividad con respecto a técnicas tradicionales como se serotipificación, debido a la naturaleza de basarse en comparativos de genes y no en la migración de proteínas. Actualmente se ha propuesto el uso de esta serotipificación *in silico* (basado en secuencias), como remplazo a la serotipificación tradicional basada en el uso de antisueros específicos (Achtman et al., 2012).

1.2.5.2.2 Electroforesis de campos pulsados (PFGE). PFGE es una técnica molecular considerada como el método estándar de oro en la genotipificación en estudios epidemiológicos, debido a su sensibilidad y alto índice discriminatorio (Neoh et al., 2019). Esta técnica corta el genoma generando fragmentos los cuales se someten al fundamento de la electroforesis convencional: aplicar un campo eléctrico para hacer que viaje el material genético logrando así su separación de acuerdo a su tamaño durante su migración en una matriz (gel) con un tamaño de poro determinado. Sin embargo, la PFGE pretende resolver la limitante de la técnica de electroforesis convencional; donde no era posible trabajar con fragmentos de restricción de grandes tamaños. Adicionalmente, utiliza enzimas de restricción que generan alrededor de 20 o 30 fragmentos grandes de ADN que son utilizados para generar un patrón electroforético particular, mientras que la separación de los fragmentos en PFGE se da por cambios intermitentes en la orientación del campo eléctrico, causando una reorientación y pudiendo alcanzar separaciones de hasta 12 Mb (Herschleb et al., 2007).

No obstante, PFGE presenta una serie de limitantes al momento de predecir serotipos, ya que múltiples aislados pueden tener patrones electroforéticos idénticos por haber emergido de un ancestro común; o bien puede presentar un alto nivel en la diversidad de patrones electroforéticos por el análisis de cepas polifiléticas (cuando las cepas no emergen o provienen de un ancestro común). Se ha reportado que los resultados de PFGE para la precisión en la predicción de los serotipos va desde el 75 al 99%, sin embargo, la mayoría de las investigaciones no incluyen una diversidad adecuada de serotipos (Shi et al., 2015).

1.2.5.2.3 Análisis multi-locus de secuencias repetitivas en tándem de número variable (MLVA). MLVA es un método utilizado para realizar tipificación molecular de microorganismos, el cual utiliza la variación natural en el número de repeticiones en tándem (secuencias de dos o más pares de bases de ADN que se repiten a lo largo del cromosoma) encontradas en diferentes *loci* en el genoma de los organismos. Esta técnica se desarrolló por primera vez en bacterias patógenas como *Haemophilus influenzae* y *Mycobacterium tuberculosis*, basándose en técnicas tradicionales de biología molecular como la reacción en cadena de la polimerasa (PCR) y posteriormente obtener los resultados mediante geles electroforéticos (Van Belkum, 2007).

El primer paso consiste en el aislamiento del ADN del microorganismo de interés utilizando algún

kit comercial o preparaciones de laboratorio. Lo siguiente es determinar la calidad del ADN y cuantificarlo; generalmente la estrategia preferida para la cuantificación es el uso de Nanodrop a una longitud de onda de 260 nm. De acuerdo a la concentración obtenida de la lectura, se deberá ajustar la cantidad medida para llevar a cabo una reacción de amplificación de la región de interés mediante PCR, para posteriormente comparar los resultados (Fig. 6).

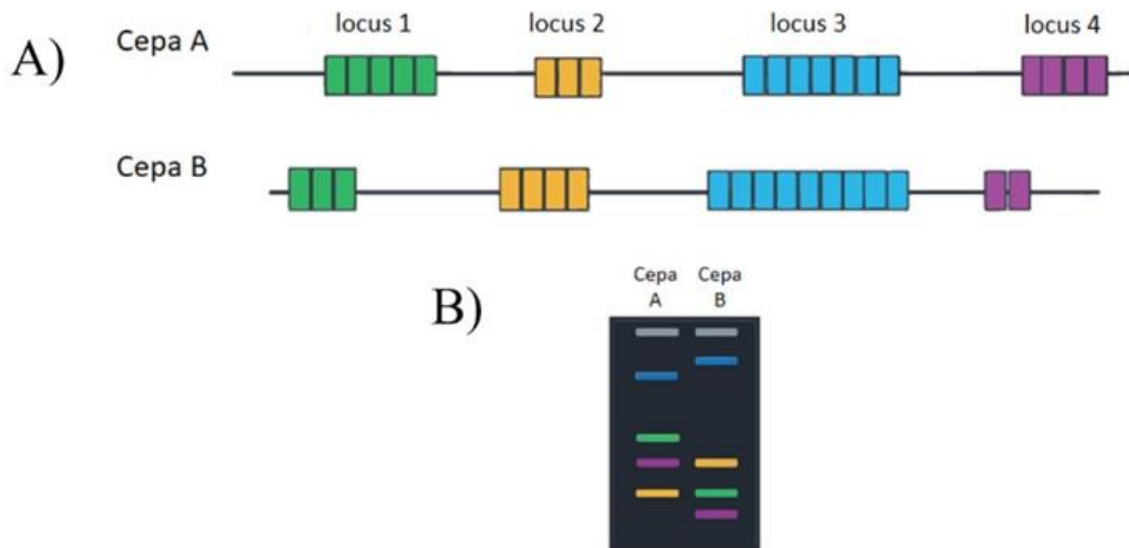


Figura 6. Análisis MLST. A) Se pueden observar los fragmentos amplificados para 4 loci en tándem pertenecientes a las cepas A y B. B) Gel electroforético con el bandeo obtenido para cada perfil. Como se puede observar, no presentan un perfil similar, por lo que se trata de especies no relacionadas.

1.2.5.2.4 Secuenciación de genomas. Método de caracterización genotípico que consiste en encontrar el orden nucleotídico de cada una de las bases que componen el genoma. Esto se puede emplear para conocer la secuencia de un gen hasta el genoma completo, a lo que se le conoce como secuenciación de genomas completos (WGS).

De manera importante, WGS no sólo tiene el potencial para la identificación de marcadores específicos de serotipos; sino tiene el potencial para generar una gran cantidad de información acerca de los aislados, incluyendo: grupos clonales, resistencia a antibióticos, virulencia, filogenias e información acerca de brotes epidemiológicos (Besser et al., 2018).

El primer método de secuenciación (1975 a 2005) fue el propuesto por Sanger, el cual producía segmentos de secuencias (reads) de 500 a 1000 pb de longitud. Dicho método se basa en hacer copias de la región específica, adicionando ddNTPs que carecían del grupo funcional hidroxilo

(OH⁻) en el carbono 3', lo que impide el alargamiento de la cadena, por lo que se genera una serie de fragmentos que son corridos en electroforesis para determinar el orden de las bases (Metzker, 2010).

Posteriormente, con la automatización del proceso de secuenciación y el surgimiento de la pirosecuenciación dio inicio a la era de la “secuenciación de siguiente generación” (NGS), caracterizada por la generación de grandes cantidades de datos y el procesamiento de muestras en paralelo. Bajo este concepto, aparecen distintas plataformas de secuenciación como Illumina, PacBio, IonTorrent; diferenciadas por el principio y la longitud de los reads generados (Besser et al., 2018; Metzker, 2010)

El uso de la secuenciación NGS se ha empezado a utilizar ampliamente para la elucidación de las rutas de transmisión de brotes epidemiológicos en hospitales, como los causados por *Staphylococcus aureus* y *Klebsiella pneumonia* (Harris et al., 2010).

El uso de la secuenciación se vislumbra como un método para la tipificación de especies de *Salmonella*; así como su aplicación en el estudio de brotes epidemiológicos y caracterización genética (Shi et al., 2015). Por lo que a continuación se presenta una revisión bibliográfica de las plataformas más empleadas por la comunidad científica.

1.2.6 Secuenciación y Bioinformática

1.2.6.1 Plataformas de Secuenciación.

1.2.6.1.1 Illumina. Para llevar a cabo la secuenciación es necesario realizar la preparación de la librería, que se refiere a la generación de los fragmentos de ADN, aproximadamente de 250 pb de acuerdo al equipo y metodología con la que se secuenciarán (ya que los secuenciadores no pueden trabajar con genomas completos por su longitud); seguido, el material genético fragmentado se unirá a la celda de flujo (flow cell) (Tritt et al., 2012). La plataforma de Illumina, Inc., con base en San Diego (CA, USA) funciona mediante secuenciación por síntesis, es decir utiliza bases nitrogenadas marcadas con fluorescencia denominadas dNTPs (deoxinucleótidos trifosfatos) que

se añaden a las hebras crecientes mediante una polimerasa. El proceso de secuenciación inicia con la generación de clústers (generación y agrupación de reads), por lo que se lleva a cabo la inmovilización de los iniciadores (primers) forward (Fw) y reversos (Rv) en las superficies de las celdas de flujo, haciendo asequible la participación de enzimas polimerasas. Posteriormente, se realiza una amplificación llamada PCR por puente (bridge PCR), donde se crean más de 1,000 copias idénticas a las muestras inmovilizadas. Para la identificación de las bases en los distintos clústeres, se realiza una serie de ciclos donde se van agregando una mezcla de primers, enzimas, dNTPs marcados; los cuales son excitados con láser para la emisión de la fluorescencia. Esta es captada por el lector del equipo y traducida a la identidad de la base nucleotídica para ir generando una cadena creciente de ADN (Steemers & Gunderson¹ Kevin L, 2005).

1.2.6.1.2 Nanopore. Consiste en una plataforma de secuenciación que permite el monitoreo en tiempo real de las secuencias amplificadas. Entre sus características principales destacan su portabilidad (con un tamaño similar a un dispositivo móvil) y la capacidad de amplificación de lecturas de mayor tamaño en comparación a las generadas por otros equipos de secuenciación como Illumina y IonTorrent. Esta plataforma se fundamenta en hacer pasar los fragmentos de material genético a través de proteínas ancladas a las celdas de flujo. La detección específica de las bases nitrogenadas se realiza cuantificando la interrupción de corriente eléctrica basal en la placa conforme va pasando la cadena de ADN por el poro, resultando en una codificación específica para cada base nitrogenada y posterior identificación (Deamer et al., 2016).

A diferencia de las plataformas tradicionales o más usadas por la comunidad científica, que muestran el total de la información de secuenciación en la etapa final, Nanopore permite el acceso en tiempo real a la información. Con esto, se podrían identificar patógenos durante la etapa inicial de la secuenciación y tomar decisiones basadas en evidencias científicas para minimizar el riesgo de infección. Otro aspecto interesante, es el bajo costo asociado a reactivos y equipos, haciéndolo accesible para la comunidad científica (Branton et al., 2008).

1.2.6.2 Análisis de Genomas Bacterianos. La estrategia general que se sigue para el análisis de genomas inicia una vez que se recibe la secuencia nucleotídica de cada uno de los aislados. Estos

representan los fragmentos o lecturas que se originaron durante el proceso de secuenciación; por lo que el primer paso implica la reconstrucción del genoma bacteriano. Una vez reconstruido, se hace una identificación genética para identificar los genes presentes (Loman & Pallen, 2015). Mientras que, para la construcción de árboles filogenéticos, se hace un alineamiento o comparación de todas las secuencias utilizando el estimador de máxima verosimilitud.

No obstante, se presenta una descripción más detallada de las implicaciones que conlleva cada etapa; empezando por el procesamiento de las lecturas obtenidas mediante secuenciación, la caracterización genética, relaciones de parentesco, construcción de pangenomas y establecimiento del reloj molecular (Didelot et al., 2012).

Sin embargo, para poder realizar el procesamiento de las lecturas y asignarles un significado biológico, es importante considerar aspectos básicos asociados a la calidad y profundidad. La calidad se refiere al valor numérico asociado a cada base expresado como la medida estadística de la certeza de cada base dada una probabilidad de error. Por lo regular, una calidad mínima de 30, es la recomendable para realizar investigaciones bajo una aproximación de SNPs (Pfeifer, 2017). Por su parte, la profundidad se refiere al número de veces que una base en particular fue secuenciada, en cuanto mayor sea el valor de profundidad (generalmente referido como un valor numérico acompañado de una X), mayor seguridad se tendrá con respecto a que la variante si corresponde a una real (Trost et al., 2018).

1.2.6.2.1 Ensamblado de lecturas. Se le conoce como ensamble de genomas al proceso de descifrar la secuencia genómica de los pequeños fragmentos obtenidos por el proceso de secuenciación. Existen dos estrategias para el ensamblaje de los genomas: por comparación y *de novo* (Aguilar-Bultet & Falquet, 2015). El primero de ellos utiliza un genoma como referencia, mientras que el *de novo*, sólo utiliza la información obtenida de la secuenciación sin conocimiento previo de la secuencia (Wain et al., 2005). Por lo que para la segunda aproximación es importante conocer previamente parámetros como el tamaño de genoma y contenido GC; adicionalmente, este tipo de ensamble es el recomendado para genomas bacterianos debido a sus genomas pequeños.

Para encontrar la mejor aproximación a la secuencia original, se utilizan algoritmos informáticos específicos que se enfocan en encontrar regiones consenso sin que se repita ningún fragmento. Dentro de las estrategias empleadas por los programas ensambladores de secuencias se encuentran

los gráficos de Bruijn; siendo el algoritmo más utilizado (Figura 7).

Este se basa en el modelamiento de las relaciones entre la sub-cadenas de longitud k dentro de los reads. De manera que los nodos en el gráfico, representan los k -mer, y los conectores indican que hay solapación entre los k -mers; permitiendo la formación de secuencias contiguas de mayor longitud, hasta la reconstrucción del genoma (Aguilar-Bultet & Falquet, 2015).

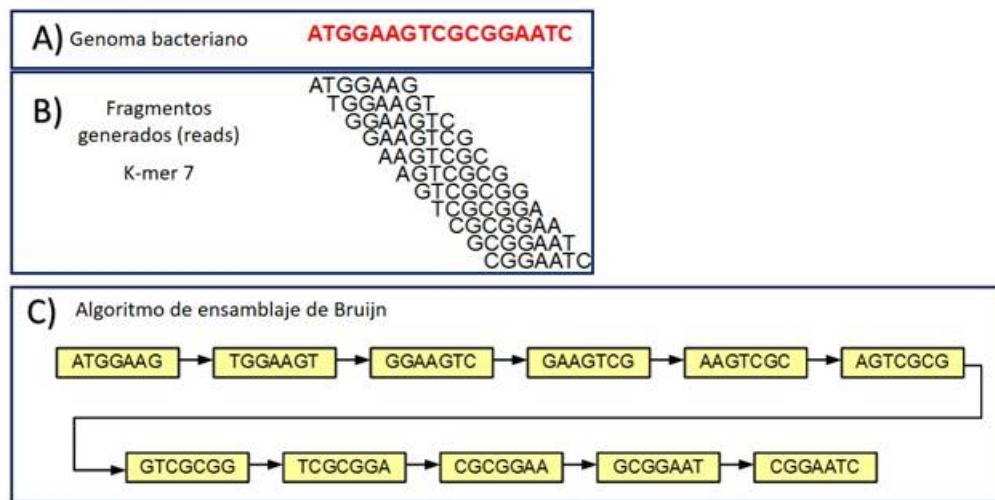


Figura 7. Ensamblado *de novo* mediante gráficas de Bruijn. A) Se puede apreciar la secuencia original del genoma bacteriano. B) Una vez fragmentado y secuenciado se obtienen fragmentos que son acomodados de acuerdo a la identidad entre los mismos. C) Gráfica generada con el acomodo para cada read según su posición en el genoma (Figura adaptada de HOMOLOG.US – FRONTIER IN BIOINFORMATICS <https://homolog.us/Tutorials/book4/p2.1.html>).

En la Figura 8 se presenta un diagrama en el que se puede observar los niveles de estructuración. A partir del ADN bacteriano se hace el proceso de secuenciación, el cual necesita fragmentar dicho ADN generando fragmentos denominados reads (siendo estos los datos que te entrega un secuenciador). Posteriormente se estos fragmentos forman secuencias contiguas más largas denominadas contigs, que estos a su vez forman regiones más largas denominadas scaffolds (Ekblom & Wolf, 2014).

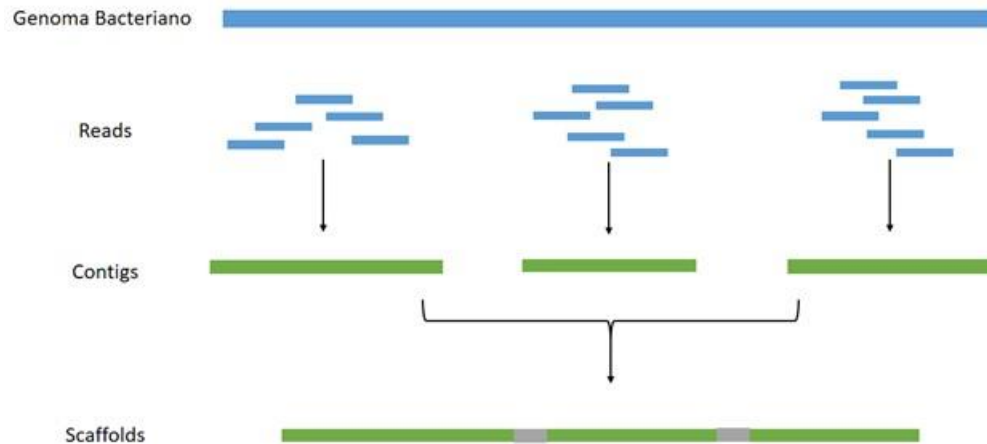


Figura 8. Niveles de estructuración en el ensamblaje *de novo*

1.2.6.2.2 Anotación funcional y estructural. Se le conoce como anotación al proceso por el cual se identifican y se les asigna una función a los genes en un genoma (anotación funcional y estructural, respectivamente). Este proceso puede incluir una pequeña descripción de la evidencia de esta asignación, como por ejemplo el porcentaje de identidad, que se refiere a la similitud de la secuencia propia contra la reportada en las bases de datos (Koonin & Galperin, 2002).

El proceso de anotación debe de incluir como mínimo la predicción de regiones codificantes y sus productos, pero es deseable que también se incluya la predicción de RNAs y péptidos de señal. Para llevar a cabo dicho proceso se necesita de una secuencia genómica previamente ensamblada (Torsten Seemann, 2014).

1.2.6.2.3 Construcción de árboles filogenéticos. Se le conoce como árbol filogenético al diagrama que representa las relaciones evolutivas entre las especies o unidades taxonómicas operativas (OTU) incluidas en el análisis. Una forma fácil para describir la disposición de los OTUs en el árbol es su designación en clados, los cuales pueden ser monofiléticos, parafilético y polifilético. El primero de ellos se refiere a la agrupación de un ancestro y todos sus descendientes; mientras que el clado parafilético es aquel que incluye a un ancestro, pero no a todos sus descendientes. Además, se le conoce como grupo polifilético a aquel que no incluye al ancestro de los miembros del clado (Fig 9).

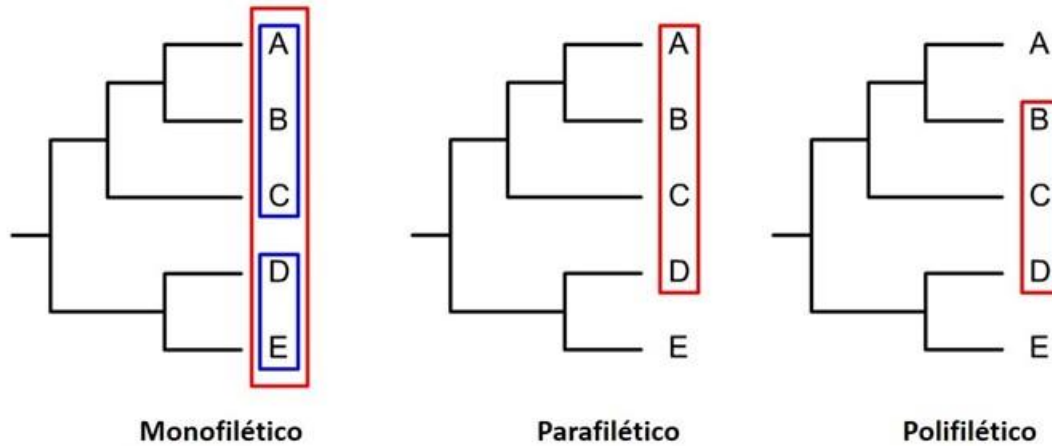


Figura 9. Designación de los integrantes de un grupo taxonómico de acuerdo a si comparten un ancestro común. (Figura adaptada de <http://www.plantasyhongos.es/>).

La construcción de dichos árboles sirve para realizar inferencias filogenéticas, que se refiere al proceso hipotético para establecer los parentescos evolutivos entre las especies; en otras palabras, es el proceso de estimación para obtener una historia evolutiva basada en los datos que se tienen (Brocchieri, 2001). Tradicionalmente los análisis filogenéticos se basaban en la anatomía de las especies; sin embargo, el uso de datos moleculares se ha convertido en una práctica muy común. Esto es debido a la universalidad del ADN en los organismos vivos, los costos de secuenciación han disminuido y además las secuencias genéticas son altamente específicas, contando con gran cantidad de información (Baum & Smith, 2013).

Los métodos de reconstrucción filogenética se pueden clasificar (Fig 10) en base al tipo de datos que emplean (caracteres discretos o matrices de distancia) y la forma en la que encuentran la topología (método algorítmico o criterio de optimización) (Caballero & Suárez, 1999).

		TIPO DE DATOS	
		Matrices de Distancias	Caracteres discretos
MÉTODO DE RECONSTRUCCIÓN	Algoritmo de agrupamiento	<u>UPGMA</u> <u>Unión de vecinos (NJ)</u>	
	Criterio de optimización	<u>Evolución mínima</u>	<u>Máxima parsimonia</u> <u>Máxima verosimilitud</u>

Figura 10. Clasificación de los métodos de reconstrucción filogenética en base al tipo de dato y método de reconstrucción.

Los métodos discretos utilizan cada sitio del alineamiento (una función probabilística de cada sitio), de tal forma que cada carácter es analizado de forma independiente en el alineamiento y generando así el proceso de evolución (Lemey et al., 2009).

En contraste, los métodos de matrices de distancia, primero convierten los alineamientos de las secuencias en una matriz de distancias de acuerdo a algún modelo evolutivo seleccionado, siendo esta matriz utilizada por el algoritmo para calcular y construir el árbol (Lemey et al., 2009).

Los criterios de optimización permiten evaluar la topología del árbol filogenético mediante asignación de un puntaje (score) con la que poder comparar cuantitativamente el mérito de cada árbol en base al criterio de optimización. Así mismo utiliza algoritmos de búsqueda para calcular el valor de la función y encontrar el árbol con el mejor score acorde a dicho criterio (Lemey et al., 2009).

Para la realización de este trabajo de investigación se utilizaron los métodos de optimización puesto que tienen la ventaja de requerir una función probabilística explícita que relaciona los datos con la topología (modelo de sustitución). Lo que permite evaluar la calidad del árbol generado mediante el uso de distintas técnicas estadísticas.

1.2.6.2.4 Construcción de pangenomas. Se le conoce como pangenoma al conjunto de genes compartidos por un grupo de cepas bacterianas, normalmente de la misma especie. Este se encuentra integrado o subclasificado en genoma núcleo (core) y accesorio (Delmont & Eren, 2018)

El primero de ellos, se refiere a la colección de genes exclusiva compartida para un conjunto de individuos de una población. Generalmente, los genes encontrados en dichos segmentos tienen un rol central en la biología básica del organismo en estudio, tales como en el metabolismo de carbohidratos y transporte, replicación, estructura de la pared celular, traducción, entre otros. Por otra parte, el genoma accesorio se refiere al conjunto de genes compartidos por un grupo reducido de cepas bacterianas (Anjum et al., 2005). Estos genes se pueden encontrar asociados a resistencia a antibióticos, virulencia y adaptación a diferentes ambientes. También aquí encontramos genes característicos para ciertas subespecies o serotipos, dependiendo de la asignación taxonómica (Subedi et al., 2018).

La clasificación de los genes dentro de los genomas está dada por la frecuencia de los mismos en los datos analizados. Adicionalmente, diferentes autores han propuesto clasificaciones adicionales. Sin embargo, dentro de las más utilizadas se encuentra la propuesta por Page et al., 2015, donde establece que el pan genoma se componer de los genes core (frecuencia en genes del 99 al 100% de las cepas), soft core (95 al 99%), genes Shell (15 al 95%) y genes cloud (0 al 15%) (Page et al., 2015).

1.2.6.2.5 Reloj molecular. Este término fue propuesto inicialmente por Emile Zuckerkandl y Linus Pauling, al estudiar las fechas en las que se originaron las globulinas (Ho, 2008). La hipótesis del reloj molecular establece que las secuencias de ADN y aminoacídicas, evolucionan a una tasa relativamente constante en un periodo de tiempo entre diferentes organismos (Zuckerkandl & Pauling, 1965), por lo que la diferencia genética entre dos especies será proporcional al tiempo en el que estas compartieron un ancestro común (ser vivo del cual descienden dos organismos). Esta información sirve para estimar las escalas evolutivas cuando la información paleontológica o fósil no es suficiente o deja huecos en la reconstrucción evolutiva (M. Wang & Caetano-Anollés, 2013). Sin embargo, la validez de la hipótesis del reloj molecular en cuanto a que las especies evolucionan a una tasa constante fue en su etapa inicial cuestionada, ya que la utilización de un modelo estadístico era poco probable que ayudará a moldear el comportamiento, siendo la evolución molecular un proceso complejo (Simpson, 1964).

Con el tiempo, se observó una alta tasa de mutación en aminoácidos, lo que ayudó a establecer que la porción de estos aminoácidos que se ven afectados, tenía un impacto en la adaptación. Como

respuesta, se propuso la teoría neutral de la evolución molecular, estableciendo que muchas mutaciones no tenían impactos medibles en los organismos, es decir eran neutrales (Kimura, 1968), indicando que la tasa de mutación es constante entre los linajes.

No obstante, estudios posteriores encontraron variaciones entre los linajes, contradiciendo las observaciones o postulados más sencillos acerca de tasa de cambio entre los organismos (Simpson, 1964). Lo que trajo como consecuencia la aparición y adopción de una gran cantidad de asunciones en la tasa de heterogeneidad entre los linajes, denominados modelos de reloj molecular (Kumar, 2005).

El análisis del reloj molecular se encuentra constituido por diversas etapas (Fig 11), que incluyen la obtención del material genético, ensamblado de lecturas, inferencias filogenéticas (construcción de árboles), determinar tasa de evolución, elegir calibradores para el reloj, seleccionar el método de estimación y, por último, correr el análisis.

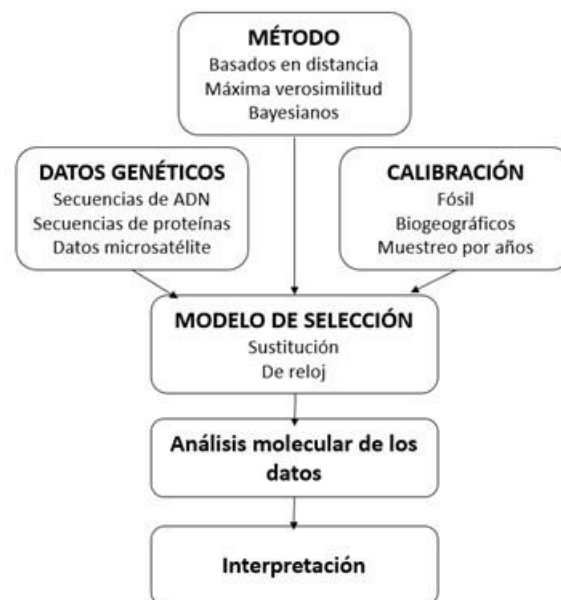


Figura 11. Componentes principales para el análisis molecular mediante reloj. Se inicia colectando los datos genéticos, identificando calibradores, eligiendo método de análisis, seleccionando modelo estadístico para identificar tasa de sustitución nucleotídica. Figura tomada y adaptada de (Van Der Wal & Ho, 2019).

Los relojes moleculares han revolucionado el conocimiento sobre la divergencia de las especies, por consiguiente, la de los genes en términos de espacio geográfico y temporalidad de su evolución. En la ausencia de otros factores para datar la historia evolutiva de las especies, como el registro

fósil o biogeográfico, el reloj molecular nos permite inferir la duplicación genética y los eventos de especiación en el tiempo. Con la continua adopción de la secuenciación de genomas y el entendimiento de las tasas de mutación, esta técnica tendrá un rol más importante en estudios de evolución (Kumar, 2005).

1.3 Hipótesis

1. *Salmonella enterica* presenta una gran diversidad genética de acuerdo al número de serotipos presentes en la región de Culiacán, Sinaloa los cuales están repartidos en tipos de secuencias (ST) únicos para la mayoría de los aislados estudiados.
2. Los aislados de *Salmonella enterica* de origen ambiental presentan firmas genéticas relacionadas con mecanismos de respuesta a estrés y utilización de fuentes de carbono de origen no hospedante; mientras que los aislados de origen clínico poseen mayor contenido genético relacionado con virulencia y sistemas de secreción necesarios para la invasión y secreción de proteínas efectoras causando enfermedades en hospederos.

1.4 Objetivo general

Analizar la filogeografía y dinámica poblacional de *S. enterica* aisladas de agua de río en la región de Culiacán, Sinaloa mediante una aproximación genómica.

1.5 Objetivos específicos

1. Determinar la diversidad genética de los aislados de *S. enterica* en términos de grupos filogenéticos
2. Identificar firmas genéticas de adaptación ambiental por parte de *S. enterica* aisladas de agua

de río.

3. Reconstruir la filogenia y dinámica espacio-temporal de los aislados ambientales.

1.6 Sección integradora

En esta tesis se incluyen dos artículos originales. El primer artículo de investigación titulado “Genomic signatures of adaptation to natural settings in non-typhoidal *Salmonella enterica* Serovars Saintpaul, Thompson and Weltevreden” se encuentra publicado en el *Journal Infection, Genetics and Evolution*, el cual abordó información sobre las estrategias de supervivencia y adaptación exhibidas por cepas de *Salmonella enterica* aisladas de agua de río de la región de Culiacán, Sinaloa. Para ello se realizó una aproximación mediante genómica comparativa y herramientas bioinformáticas. Como resultados se identificaron a la duplicación genética y la acumulación de pseudogenes como las principales estrategias adaptativas para la supervivencia de *S. enterica* en los ambientes no hospederos. Así mismo, diferencias importantes entre el origen de los aislados pudieron ser determinadas; por ejemplo, los genes asociados a la resistencia a estés oxidativo y metabolismo de fuentes de carbono ambiental fueron identificados como firmas genéticas para los aislados de origen ambiental. Por su parte, los genes relacionados con los sistemas de secreción, flagelos y genes de virulencia se encuentran conservados entre los genomas de origen clínico, siendo esenciales para la invasión, internalización y patogénesis.

Con respecto al segundo artículo, titulado “Genomic approach revealed genomic similarity for environmental Mexican *S. Oranienburg* genomes”, abordó un estudio genético de la estructura poblacional de *S. enterica* serotipo Oranienburg, ya que es uno de los predominantes en la región de Culiacán, Sinaloa. Además, se encuentra ampliamente distribuido a nivel mundial y es responsable de brotes epidemiológicos en hortalizas frescas. De acuerdo a una aproximación bioinformática y genómica comparativa, se determinó que este serotipo presenta alta similitud genética con un amplio repertorio de genes de virulencia reportados como cruciales en el proceso de invasión y establecimiento por parte de este patógeno. El grupo clonal designado como ST23 fue encontrado como prevalente, y de acuerdo al número de SNPs, se encuentra recirculando en la región.

1.7 Referencias

- Aanensen, D. M., & Spratt, B. G. (2005). The multilocus sequence typing network: mlst.net. *Nucleic Acids Research*, 33(suppl_2), W728–W733. <https://doi.org/10.1093/nar/gki415>
- Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krauland, M. G., Hale, J. L., Harbottle, H., & Uesbeck, A. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog*, 8(6), e1002776.
- Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*, 37(2), 125–132.
- Al-Khafaji, N. S. K., Al-Bayati, A. M. K., & Al-Dahmoshi, H. O. M. (2021). Virulence Factors of *Salmonella Typhi*. In *Salmonella spp.-A Global Challenge*. IntechOpen.
- Alvarado, O., & Isac, W. (2022). *Caracterización del agente causal de Salmonelosis aviar” mediante revisión bibliográfica*. BABAHOYO: UTB, 2022.
- Andino, A., & Hanning, I. (2015). *Salmonella enterica*: survival, colonization, and virulence differences among serovars. *The Scientific World Journal*, 2015.
- Andrews, S. (2017). *FastQC: a quality control tool for high throughput sequence data*. 2010.
- Anjum, M. F., Marooney, C., Fookes, M., Baker, S., Dougan, G., Ivens, A., & Woodward, M. J. (2005). Identification of Core and Variable Components of the *Salmonella enterica* Subspecies I Genome by Microarray. *Infection and Immunity*, 73(12), 7894 LP – 7905. <https://doi.org/10.1128/IAI.73.12.7894-7905.2005>
- Antony, L., Behr, M., Sockett, D., Miskimins, D., Aulik, N., Christopher-Hennings, J., Nelson, E., Allard, M. W., & Scaria, J. (2018). Genome divergence and increased virulence of outbreak associated *Salmonella enterica* subspecies *enterica* serovar Heidelberg. *Gut Pathogens*, 10(1), 1–9.
- Baker, S., & Dougan, G. (2007). The genome of *Salmonella enterica* serovar Typhi. *Clinical Infectious Diseases*, 45(Supplement_1), S29–S33.
- Baudart, J., Lemarchand, K., Brisabois, A., & Lebaron, P. (2000). Diversity of *Salmonella* strains isolated from the aquatic environment as determined by serotyping and amplification of the ribosomal DNA spacer regions. *Applied and Environmental Microbiology*, 66(4), 1544–1552.
- Baum, D. A., & Smith, S. D. (2013). Tree thinking. *An Introduction to Phylogenetic Biology*. Roberts and Company Publishers.
- 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.
- Blair, J., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13(1), 42–51.

- Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P., & Allaoui, A. (2001). Structure and composition of the *Shigella flexneri* ‘needle complex’, a part of its type III secreton. *Molecular Microbiology*, 39(3), 652–663.
- Bloomfield, S. J., Benschop, J., Biggs, P. J., Marshall, J. C., Hayman, D. T. S., Carter, P. E., Midwinter, A. C., Mather, A. E., & French, N. P. (2017). Genomic analysis of *Salmonella enterica* serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand, 1998–2012. *Emerging Infectious Diseases*, 23(6), 906.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., & Huang, X. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology*, 26(10), 1146–1153.
- Brocchieri, L. (2001). Phylogenetic inferences from molecular sequences: review and critique. *Theoretical Population Biology*, 59(1), 27–40.
- Caballero, E. J. L., & Suárez, G. P. (1999). Métodos de análisis en la Reconstrucción Filogenética. *Bol. SEA*, 26, 45–56.
- Cabañas-Romero, P. V., & Huerta-Saquero, A. (2014). Nanomáquinas biológicas: los sistemas de secreción bacterianos. *Mundo Nano. Revista Interdisciplinaria En Nanociencias y Nanotecnología*, 7(13), 28–36.
- Carattoli, A. (2003). Plasmid-mediated antimicrobial resistance in *Salmonella enterica*. *Current Issues in Molecular Biology*, 5(4), 113–122.
- CDC. (2022). *Salmonella Outbreak Linked to Onions*. <https://www.cdc.gov/salmonella/oranienburg-09-21/index.html#:~:text=Multiple companies recalled onions,information about each company's recall>.
- Chao, W. L., Ding, R. J., & Chen, R. S. (1987). Survival of pathogenic bacteria in environmental microcosms. *Zhonghua Minguo Wei Sheng Wu Ji Mian Yi Xue Za Zhi= Chinese Journal of Microbiology and Immunology*, 20(4), 339–348.
- Chen, L., Xiong, Z., Sun, L., Yang, J., & Jin, Q. (2012). VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research*, 40(D1), D641–D645.
- Cherchame, E., Guillier, L., Lailler, R., Vignaud, M.-L., Silva, J.-D., Le Hello, S., Weill, F.-X., & Cadel-Six, S. (2022). *Salmonella enterica* subsp. *enterica* Welikade: guideline for phylogenetic analysis of serovars rarely involved in foodborne outbreaks. *BMC Genomics*, 23(1), 1–13.
- Cherchame, E., Ilango, G., Noël, V., & Cadel Six, S. (2022). Polyphyly in widespread *Salmonella enterica* serovars and using genomic proximity to choose the best reference genome for bioinformatics analyses. *Frontiers in Public Health*, 2820.
- Coil, D., Jospin, G., & Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, 31(4), 587–589.
- Coipan, C. E., Dallman, T. J., Brown, D., Hartman, H., van der Voort, M., van den Berg, R. R.,

- Palm, D., Kotila, S., van Wijk, T., & Franz, E. (2020). Concordance of SNP-and allele-based typing workflows in the context of a large-scale international *Salmonella* Enteritidis outbreak investigation. *Microbial Genomics*, 6(3).
- Connor, T. R., Barker, C. R., Baker, K. S., Weill, F.-X., Talukder, K. A., Smith, A. M., Baker, S., Gouali, M., Thanh, D. P., & Azmi, I. J. (2015). Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *Elife*, 4, e07335.
- Contreras-Soto, M. B., Medrano-Félix, J. A., Chaidez, C., Ibarra-Rodríguez, J. R., Martínez-Urtaza, J., & Castro-del Campo, N. (2018). Los últimos 50 años de *Salmonella* en México: Fuentes de aislamiento y factores que influyen en su prevalencia y diversidad. *Revista Bio Ciencias*, 6, 26.
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518–524.
- Delmont, T. O., & Eren, A. M. (2018). Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*, 6, e4320.
- den Bakker, H. C., Moreno Switt, A. I., Cummings, C. A., Hoelzer, K., Degoricija, L., Rodriguez-Rivera, L. D., Wright, E. M., Fang, R., Davis, M., & Root, T. (2011). A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type. *Applied and Environmental Microbiology*, 77(24), 8648–8655.
- Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, 13(9), 601–612.
- Domingues, S., da Silva, G. J., & Nielsen, K. M. (2012). Integrons: vehicles and pathways for horizontal dissemination in bacteria. *Mobile Genetic Elements*, 2(5), 211–223.
- Dos Santos, A. M. P., Ferrari, R. G., & Conte-Junior, C. A. (2019). Virulence factors in *Salmonella* Typhimurium: the sagacity of a bacterium. *Current Microbiology*, 76(6), 762–773.
- Duncan, M. C., Linington, R. G., & Auerbuch, V. (2012). Chemical inhibitors of the type three secretion system: disarming bacterial pathogens. *Antimicrobial Agents and Chemotherapy*, 56(11), 5433–5441.
- Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7(9), 1026–1042.
- Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G., & Lee, L.-H. (2015). *Salmonella*: a review on pathogenesis, epidemiology and antibiotic resistance. *Frontiers in Life Science*, 8(3), 284–293.
- Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., & Spratt, B. G. (2002). The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proceedings of the National Academy of Sciences*, 99(11), 7687–7692.
- FAO. (2020). *Agricultural production statistics 2000–2020*. Agricultural Production Statistics 2000–2020. <https://www.fao.org/3/cb9180en/cb9180en.pdf>
- Fatica, M. K., & Schneider, K. R. (2011). *Salmonella* and produce: survival in the plant

environment and implications in food safety. *Virulence*, 2(6), 573–579.

- Ferrari, R. G., Panzenhagen, P. H. N., & Conte-Junior, C. A. (2017). Phenotypic and genotypic eligible methods for Salmonella Typhimurium source tracking. *Frontiers in Microbiology*, 8, 2587.
- Figueira, R., & Holden, D. W. (2012). Functions of the Salmonella pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiology*, 158(5), 1147–1161.
- Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., & Carriço, J. A. (2012). PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, 13(1), 1–10.
- Galán, J. E. (2001). Salmonella interactions with host cells: type III secretion at work. *Annual Review of Cell and Developmental Biology*, 17(1), 53–86.
- Gerlach, R. G., Jäckel, D., Stecher, B., Wagner, C., Lupas, A., Hardt, W., & Hensel, M. (2007). Salmonella Pathogenicity Island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. *Cellular Microbiology*, 9(7), 1834–1850.
- González-Pedraja, B., & Dreyfus, G. (2003). Sistemas de secreción de proteínas en las bacterias Gram negativas: Biogénesis flagelar y translocación de factores de virulencia. *Mensaje Bioquímico*, 27, 45–63.
- Gonzalez Pedraza, J., Pereira Sanandres, N., Soto Varela, Z., Hernández Aguirre, E., & Villarreal Camacho, J. (2014). Aislamiento microbiológico de Salmonella spp. y herramientas moleculares para su detección. *Revista Salud Uninorte*, 30(1), 73–94.
- Grimont, P. A. D., & Weill, F.-X. (2007). Antigenic formulae of the Salmonella serovars. *WHO Collaborating Centre for Reference and Research on Salmonella*, 9, 1–166.
- Guard, J., Morales, C. A., Fedorka-Cray, P., & Gast, R. K. (2011). Single nucleotide polymorphisms that differentiate two subpopulations of Salmonella enteritidis within phage type. *BMC Research Notes*, 4(1), 1–15.
- Gymoese, P., Sørensen, G., Litrup, E., Olsen, J. E., Nielsen, E. M., & Torpdahl, M. (2017). Investigation of outbreaks of Salmonella enterica serovar Typhimurium and its monophasic variants using whole-genome sequencing, Denmark. *Emerging Infectious Diseases*, 23(10), 1631.
- Hacker, J., & Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Reports*, 2(5), 376–381.
- Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., & Lindsay, J. A. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, 327(5964), 469–474.
- Hedberg, C. W., Smith, K. E., Besser, J. M., Boxrud, D. J., Hennessy, T. W., Bender, J. B., Anderson, F. A., & Osterholm, M. T. (2001). Limitations of pulsed-field gel electrophoresis for the routine surveillance of Campylobacter infections. *The Journal of Infectious Diseases*, 184(2), 242–243.
- Hensel, M. (2004). Evolution of pathogenicity islands of Salmonella enterica. *International Journal of Medical Microbiology*, 294(2–3), 95–102.

- Herschleb, J., Ananiev, G., & Schwartz, D. C. (2007). Pulsed-field gel electrophoresis. *Nature Protocols*, 2(3), 677–684.
- Ho, S. (2008). The molecular clock and estimating species divergence. *Nature Education*, 1(1), 168.
- Højby, N., Bjarnsholt, T., Givskov, M., Molin, S., & Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *International Journal of Antimicrobial Agents*, 35(4), 322–332.
- Homøe, P., Bjarnsholt, T., Wessman, M., Sørensen, H. C. F., & Johansen, H. K. (2009). Morphological evidence of biofilm formation in Greenlanders with chronic suppurative otitis media. *European Archives of Oto-Rhino-Laryngology*, 266(10), 1533–1538.
- Jacobsen, C. S., & Bech, T. B. (2012). Soil survival of *Salmonella* and transfer to freshwater and fresh produce. *Food Research International*, 45(2), 557–566.
- Jajere, S. M. (2019). A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Veterinary World*, 12(4), 504.
- Johnson, R., Mylona, E., & Frankel, G. (2018). Typhoidal *Salmonella*: Distinctive virulence factors and pathogenesis. *Cellular Microbiology*, 20(9), e12939.
- Jolley, K. A., Bray, J. E., & Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. *Wellcome Open Research*, 3.
- Jones, B. D. (1997). Host responses to pathogenic *Salmonella* infection. *Genes & Development*, 11(6), 679–687.
- Kaur, J., & Jain, S. K. (2012). Role of antigens and virulence factors of *Salmonella enterica* serovar Typhi in its pathogenesis. *Microbiological Research*, 167(4), 199–210.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624–626.
- Kolenda, R., Ugorski, M., & Grzymajlo, K. (2019). Everything you always wanted to know about *Salmonella* type 1 fimbriae, but were afraid to ask. *Frontiers in Microbiology*, 10, 1017.
- Koonin, E., & Galperin, M. Y. (2002). *Sequence—evolution—function: computational approaches in comparative genomics*.
- Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, 6(8), 654–662.
- Lago, M., Monteil, V., Douche, T., Guglielmini, J., Criscuolo, A., Maufrais, C., Matondo, M., & Norel, F. (2017). Proteome remodelling by the stress sigma factor RpoS/ σ S in *Salmonella*: identification of small proteins and evidence for post-transcriptional regulation. *Scientific Reports*, 7(1), 1–15.
- Laing, C. R., Whiteside, M. D., & Gannon, V. P. J. (2017). Pan-genome analyses of the species *Salmonella enterica*, and identification of genomic markers predictive for species, subspecies, and serovar. *Frontiers in Microbiology*, 8, 1345.
- Le Minor, L., & Popoff, M. Y. (1987). Designation of *Salmonella enterica* sp. nov., nom. rev., as the Type and Only Species of the Genus *Salmonella*: Request for an Opinion. *International Journal of Systematic and Evolutionary Microbiology*, 37(4), 465–468.

- Le, V. T. M., & Diep, B. A. (2013). Selected insights from application of whole genome sequencing for outbreak investigations. *Current Opinion in Critical Care*, 19(5), 432.
- Lee, P.-C., Zmina, S. E., Stopford, C. M., Toska, J., & Rietsch, A. (2014). Control of type III secretion activity and substrate specificity by the cytoplasmic regulator PcrG. *Proceedings of the National Academy of Sciences*, 111(19), E2027–E2036.
- Lemey, P., Salemi, M., & Vandamme, A.-M. (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press.
- Lengai, G. M. W., Fulano, A. M., & Muthomi, J. W. (2022). Improving Access to Export Market for Fresh Vegetables through Reduction of Phytosanitary and Pesticide Residue Constraints. *Sustainability*, 14(13), 8183.
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Lipkin, W. I. (2013). The changing face of pathogen discovery and surveillance. *Nature Reviews Microbiology*, 11(2), 133–141.
- Lisa, G., Samarpita, W., Diana, C., & Ashley, A. N. (2022). *Salmonella enterica* serovar diversity, distribution, and prevalence in public access waters from a central California coastal leafy green growing region during 2011–2016.
- Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, 47(D1), D687–D692.
- Liu, H., Whitehouse, C. A., & Li, B. (2018). Presence and persistence of Salmonella in water: the impact on microbial quality of water and food safety. *Frontiers in Public Health*, 6, 159.
- Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, 13(12), 787–794.
- Long, S. G., DuPont, H. L., Gaul, L., Arafat, R. R., Selwyn, B. J., Rogers, J., & Casey, E. (2010). Pulsed-field gel electrophoresis for Salmonella infection surveillance, Texas, USA, 2007. *Emerging Infectious Diseases*, 16(6), 983.
- Lostroh, C. P., & Lee, C. A. (2001). The Salmonella pathogenicity island-1 type III secretion system. *Microbes and Infection*, 3(14–15), 1281–1291.
- Marcus, S. L., Brumell, J. H., Pfeifer, C. G., & Finlay, B. B. (2000). Salmonella pathogenicity islands: big virulence in small packages. *Microbes and Infection*, 2(2), 145–156.
- Martinez, J. L., & Baquero, F. (2000). Mutation frequencies and antibiotic resistance. *Antimicrobial Agents and Chemotherapy*, 44(7), 1771–1777.
- Mazurkiewicz, P., Thomas, J., Thompson, J. A., Liu, M., Arbibe, L., Sansonetti, P., & Holden, D. W. (2008). SpvC is a Salmonella effector with phosphothreonine lyase activity on host mitogen-activated protein kinases. *Molecular Microbiology*, 67(6), 1371–1383.
- McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., Ayers, S. L., Lam, C., Tate, H. P., & Zhao, S. (2016). Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal Salmonella. *Antimicrobial Agents and Chemotherapy*, 60(9), 5515–5520.

- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31–46.
- Morpheus. (n.d.). *MORPHEUS: Versatile matrix visualization and analysis software*. <https://software.broadinstitute.org/morpheus/>
- Mouslim, C., Hilbert, F., Huang, H., & Groisman, E. A. (2002). Conflicting needs for a *Salmonella* hypervirulence gene in host and non-host environments. *Molecular Microbiology*, 45(4), 1019–1027.
- Neoh, H., Tan, X.-E., Sapri, H. F., & Tan, T. L. (2019). Pulsed-field gel electrophoresis (PFGE): A review of the “gold standard” for bacteria typing and current alternatives. *Infection, Genetics and Evolution*, 74, 103935.
- Ochoa, I. M. F., & Rodríguez, A. V. (2005). Mecanismos moleculares de patogenicidad de *Salmonella* sp. *Revista Latinoamericana de Microbiología*, 47(1–2), 25–42.
- Oladeinde, A., Cook, K., Lakin, S. M., Woyda, R., Abdo, Z., Looft, T., Herrington, K., Zock, G., Lawrence, J. P., & Thomas IV, J. C. (2019). Horizontal gene transfer and acquired antibiotic resistance in *Salmonella enterica* serovar Heidelberg following in vitro incubation in broiler ceca. *Applied and Environmental Microbiology*, 85(22), e01903-19.
- Pacífico, A. del. (2009). *Alianza del Pacífico*.
- 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).
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22), 3691–3693.
- Palmer, A. D., & Slauch, J. M. (2017). Mechanisms of *Salmonella* pathogenesis in animal models. *Human and Ecological Risk Assessment: An International Journal*, 23(8), 1877–1892.
- Pfeifer, S. P. (2017). From next-generation resequencing reads to a high-quality variant data set. *Heredity*, 118(2), 111–124.
- Prevention, C. for D. C. and. (2015). *Multistate outbreak of Salmonella Poona infections linked to imported cucumbers*.
- Rabsch, W. (2007). *Salmonella typhimurium* phage typing for pathogens. In *Salmonella* (pp. 177–211). Springer.
- Rahman, M.-M., Lim, S.-J., & Park, Y.-C. (2022a). Development of Single Nucleotide Polymorphism (SNP)-Based Triplex PCR Marker for Serotype-specific *Escherichia coli* Detection. *Pathogens*, 11(2), 115.
- Rahman, M.-M., Lim, S.-J., & Park, Y.-C. (2022b). Genome-Wide Searching Single Nucleotide-Polymorphisms (SNPs) and SNPs-Targeting a Multiplex Primer for Identification of Common *Salmonella* Serotypes. *Pathogens*, 11(10), 1075.
- Ryan, J. (2016). *Biosecurity and bioterrorism: containing and preventing biological threats*. Butterworth-Heinemann.
- Sandt, C. H., Krouse, D. A., Cook, C. R., Hackman, A. L., Chmielecki, W. A., & Warren, N. G.

- (2006). The key role of pulsed-field gel electrophoresis in investigation of a large multiserotype and multistate food-borne outbreak of *Salmonella* infections centered in Pennsylvania. *Journal of Clinical Microbiology*, 44(9), 3208–3212.
- Savelli, C. J., Bradshaw, A., Ben Embarek, P., & Mateus, C. (2019). The FAO/WHO international food safety authorities network in review, 2004–2018: learning from the past and looking to the future. *Foodborne Pathogens and Disease*, 16(7), 480–488.
- Scaltriti, E., Sasser, D., Comandatore, F., Morganti, M., Mandalari, C., Gaiarsa, S., Bandi, C., Zehender, G., Bolzoni, L., & Casadei, G. (2015). Differential single nucleotide polymorphism-based analysis of an outbreak caused by *Salmonella enterica* serovar Manhattan reveals epidemiological details missed by standard pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*, 53(4), 1227–1238.
- Schmiegner, H. (1999). Molecular survey of the *Salmonella* phage typing system of Anderson. *Journal of Bacteriology*, 181(5), 1630–1635.
- Schrader, K. N., Fernandez-Castro, A., Cheung, W. K. W., Crandall, C. M., & Abbott, S. L. (2008). Evaluation of commercial antisera for *Salmonella* serotyping. *Journal of Clinical Microbiology*, 46(2), 685–688.
- Seemann, T. (2018). ABRicate: mass screening of contigs for antimicrobial and virulence genes. *Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Australia*. Available Online: <https://Github.Com/Tseemann/Abriicate> (Accessed on 28 February 2019).
- Seemann, Torsten. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069.
- Seemann, Torsten. (2018a). Snippy: fast bacterial variant calling from NGS reads. <https://github.com/tseemann/snippy>
- Seemann, Torsten. (2018b). Snp-dists: Convert a FASTA alignment to SNP distance matrix. <https://github.com/tseemann/snp-dists>
- Selander, R. K., Li, J., & Nelson, K. (1996). Evolutionary genetics of *Salmonella enterica*. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, 2, 2691–2707.
- Serra Valdés, M. Á. (2017). La resistencia microbiana en el contexto actual y la importancia del conocimiento y aplicación en la política antimicrobiana. *Revista Habanera de Ciencias Médicas*, 16(3), 402–419.
- Shi, C., Singh, P., Ranieri, M. L., Wiedmann, M., & Moreno Switt, A. I. (2015). Molecular methods for serovar determination of *Salmonella*. *Critical Reviews in Microbiology*, 41(3), 309–325.
- Shimizu, T. (2022). The Growth of the Fruit and Vegetable Export Industry in Peru. *SpringerBriefs in Economics*.
- Silva-Hidalgo, G., Ortiz-Navarrete, V. F., Alpuche-Aranda, C. M., Rendón-Maldonado, J. G., López-Valenzuela, M., Juárez-Barranco, F., & López-Moreno, H. S. (2012). Non-typhi *Salmonella* serovars found in Mexican zoo animals. *Research in Veterinary Science*, 93(3), 1132–1135.
- Simpson, G. G. (1964). Organisms and Molecules in Evolution: Studies of evolution at the molecular level lead to greater understanding and a balancing of viewpoints. *Science*,

146(3651), 1535–1538.

- Smith, K. E., Medus, C., Meyer, S. D., Boxrud, D. J., Leano, F. E., Hedberg, C. W., Elfering, K., Braymen, C., Bender, J. B., & Danila, R. N. (2008). Outbreaks of salmonellosis in Minnesota (1998 through 2006) associated with frozen, microwaveable, breaded, stuffed chicken products. *Journal of Food Protection*, 71(10), 2153–2160.
- Soto-Estrada, G., Moreno-Altamirano, L., & Pahua Díaz, D. (2016). Panorama epidemiológico de México, principales causas de morbilidad y mortalidad. *Revista de La Facultad de Medicina (México)*, 59(6), 8–22.
- Steele, M., & Odumeru, J. (2004). Irrigation water as source of foodborne pathogens on fruit and vegetables. *Journal of Food Protection*, 67(12), 2839–2849.
- Steemers, F. J., & Gunderson1 Kevin L, 2. (2005). *Illumina, Inc.*
- Stevens, M. P., Humphrey, T. J., & Maskell, D. J. (2009). Molecular insights into farm animal and zoonotic salmonella infections. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 364, Issue 1530, pp. 2708–2723). Royal Society. <https://doi.org/10.1098/rstb.2009.0094>
- Subedi, D., Kohli, G. S., Vijay, A. K., Willcox, M. D. P., & Rice, S. A. (2018). Accessory genome contributes to the virulence and resistance of the ocular isolate of *Pseudomonas aeruginosa*: A complete genome analysis. *BioRxiv*, 497974.
- Tomastikova, Z., Romero, S. B., Knotek, Z., & Karpiskova, R. (2017). Prevalence and characteristics of *Salmonella* species isolated from captive reptiles in the Czech Republic. *Veterinární Medicina*, 62(8), 456–469.
- Toro, M., Retamal, P., Ayers, S., Barreto, M., Allard, M., Brown, E. W., & Gonzalez-Escalona, N. (2016). Whole-genome sequencing analysis of *Salmonella enterica* serovar Enteritidis isolates in Chile provides insights into possible transmission between gulls, poultry, and humans. *Applied and Environmental Microbiology*, 82(20), 6223–6232.
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(W1), W232–W235.
- Tritt, A., Eisen, J. A., Facciotti, M. T., & Darling, A. E. (2012). An Integrated Pipeline for de Novo Assembly of Microbial Genomes. *PLOS ONE*, 7(9), e42304. <https://doi.org/10.1371/journal.pone.0042304>
- Trost, B., Walker, S., Wang, Z., Thiruvahindrapuram, B., MacDonald, J. R., Sung, W. W. L., Pereira, S. L., Whitney, J., Chan, A. J. S., & Pellicchia, G. (2018). A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *The American Journal of Human Genetics*, 102(1), 142–155.
- Urwin, R., & Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends in Microbiology*, 11(10), 479–487.
- Van Belkum, A. (2007). Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunology & Medical Microbiology*, 49(1), 22–27.
- Van Der Wal, C., & Ho, S. Y. W. (2019). *Molecular Clock* (S. Ranganathan, M. Gribskov, K.

- Nakai, & C. B. T.-E. of B. and C. B. Schönbach (eds.); pp. 719–726). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-809633-8.20257-4>
- Vázquez-Garcidueñas, M. S., Romero-Pérez, N. L., Figueroa-Aguilar, G. A., Jaime-Sánchez, J. L., & Vázquez-Marrufo, G. (2014). Investigation of a food-borne Salmonella Oranienburg outbreak in a Mexican prison. *The Journal of Infection in Developing Countries*, 8(02), 143–153.
- Velge, P., Wiedemann, A., Rosselin, M., Abed, N., Boumart, Z., Chaussé, A. M., Grépinet, O., Namdari, F., Roche, S. M., Rossignol, A., & Virlogeux-Payant, I. (2012). Multiplicity of Salmonella entry mechanisms, a new paradigm for Salmonella pathogenesis. *MicrobiologyOpen*, 1(3), 243–258. <https://doi.org/10.1002/mbo3.28>
- Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics*, 40(4), 277.
- Vernikos, G., Medini, D., Riley, D. R., & Tettelin, H. (2015). Ten years of pan-genome analyses. *Current Opinion in Microbiology*, 23, 148–154.
- Voetsch, A. C., Van Gilder, T. J., Angulo, F. J., Farley, M. M., Shallow, S., Marcus, R., Cieslak, P. R., Deneen, V. C., Tauxe, R. V., & Group, E. I. P. F. W. (2004). FoodNet estimate of the burden of illness caused by nontyphoidal Salmonella infections in the United States. *Clinical Infectious Diseases*, 38(Supplement_3), S127–S134.
- Wain, J., House, D., Zafar, A., Baker, S., Nair, S., Kidgell, C., Bhutta, Z., Dougan, G., & Hasan, R. (2005). Vi antigen expression in Salmonella enterica serovar Typhi clinical isolates from Pakistan. *Journal of Clinical Microbiology*, 43(3), 1158–1165.
- Wang, M., & Caetano-Anollés, G. (2013). *Molecular Clock* (S. Maloy & K. B. T.-B. E. of G. (Second E. Hughes (eds.); pp. 461–462). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374984-0.00964-5>
- Wang, Q., Holmes, N., Martinez, E., Howard, P., Hill-Cawthorne, G., & Sintchenko, V. (2015). It is not all about single nucleotide polymorphisms: comparison of mobile genetic elements and deletions in Listeria monocytogenes genomes links cases of hospital-acquired listeriosis to the environmental source. *Journal of Clinical Microbiology*, 53(11), 3492–3500.
- White, N. J. (2002). Salmonella typhi (typhoid fever) and S. paratyphi (paratyphoid fever). In *Antimicrobial Therapy and Vaccines* (pp. 583–603). Apple Tree Productions, New York.
- Winfield, M. D., & Groisman, E. A. (2003). Role of nonhost environments in the lifestyles of Salmonella and Escherichia coli. *Appl. Environ. Microbiol.*, 69(7), 3687–3694.
- Wu, H. J., Wang, A. H. J., & Jennings, M. P. (2008). Discovery of virulence factors of pathogenic bacteria. In *Current Opinion in Chemical Biology* (Vol. 12, Issue 1, pp. 93–101). Elsevier Current Trends. <https://doi.org/10.1016/j.cbpa.2008.01.023>
- Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P. J., Nash, J. H. E., & Taboada, E. N. (2016). The Salmonella in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft Salmonella genome assemblies. *PloS One*, 11(1), e0147101.
- Zaidi, M. B., Calva, J. J., Estrada-Garcia, M. T., Leon, V., Vazquez, G., Figueroa, G., Lopez, E., Contreras, J., Abbott, J., & Zhao, S. (2008). Integrated food chain surveillance system for

Salmonella spp. in Mexico. *Emerging Infectious Diseases*, 14(3), 429.

Zaidi, M. B., López, C., & Calva, E. (2006). Estudios mexicanos sobre Salmonella: epidemiología, vacunas y biología molecular. *Rev Latinoam Microbiol*, 48(2), 121–125.

Zhang, S., Kingsley, R. A., Santos, R. L., Andrews-Polymenis, H., Raffatellu, M., Figueiredo, J., Nunes, J., Tsois, R. M., Adams, L. G., & Bäumler, A. J. (2003). Molecular pathogenesis of Salmonella enterica serotype Typhimurium-induced diarrhea. *Infection and Immunity*, 71(1), 1–12.

Zuckerandl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In V. Bryson & H. J. B. T.-E. G. and P. Vogel (Eds.), *Evolving Genes and Proteins* (pp. 97–166). Academic Press. <https://doi.org/10.1016/b978-1-4832-2734-4.50017-6>

2. GENOMIC SIGNATURES OF ADAPTATION TO NATURAL SETTINGS IN NON-TYPHOIDAL *Salmonella enterica* SEROVARS SAINTPAUL, THOMPSON AND WELTEVREDEN

J. R. Aguirre-Sanchez¹, J. R. Ibarra-Rodriguez¹, I. F. Vega-Lopez², J. Martínez-Urtaza³, C. Chaidez-Quiroz¹

¹ Centro de Investigación en Alimentación y Desarrollo, Culiacán, México.

² Parque de Innovación Tecnológica de la Universidad Autónoma de Sinaloa, Culiacán, México.

³ Departement of Genetics and Microbiology, Universitat Autònoma de Barcelona, Barcelona, Spain.

*Correspondence:

Dr. Cristobal Chaidez Quiroz

CIAD, Culiacán

Carretera El Dorado Km 5.5

Campo el Diez, 80110

Culiacán, Sinaloa, México

email: chaqui@ciad.mx

Published: 13 February 2021

Journal of Infection, Genetic and Evolution

<https://doi.org/10.1016/j.meegid.2021.104771>



Research paper

Genomic signatures of adaptation to natural settings in non-typhoidal *Salmonella enterica* Serovars Saintpaul, Thompson and Weltevreden

J.R. Aguirre-Sanchez^a, J.R. Ibarra-Rodriguez^a, I.F. Vega-Lopez^b, J. Martínez-Urtaza^c,
C. Chaidez-Quiroz^{a,*}

^a Centro de Investigación en Alimentación y Desarrollo, Culiacán, Mexico

^b Parque de Innovación Tecnológica de la Universidad Autónoma de Sinaloa, Culiacán, Mexico

^c Departament of Genetics and Microbiology, Universitat Autònoma de Barcelona, Barcelona, Spain



ARTICLE INFO

Keywords:

Salmonella

Comparative genomics

Sequencing

Environmental

ABSTRACT

Salmonella enterica is a pathogenic bacterium responsible for intestinal illness and systemic diseases such as typhoid and paratyphoid fevers. Among clinical manifestation classification, non-typhoidal *Salmonella* is mainly known as foodborne pathogen associated with the consumption of fecal contaminated food and water. Even though *Salmonella* hosts include humans and warm-blooded animals, it has been found in non-host environments as river water where the bacteria use different strategies to fitness the environment persisting and establishment. Now with the availability of WGS and bioinformatics tools, we can explore bacterial genomes with higher resolution to increase our understanding of specific genetic signatures among environmental and clinical isolates, being the goal of this work. Pangenome construction allowed the detection of specific environmental and clinical gene clusters related to metabolism and secretion systems as the main signature respectively. Specifically, D-galactonate degradation pathway was observed mainly in environmental genomes while T3SS and flagellum genes were detected for all clinical but not for all environmental isolates. Gene duplication and pseudogenes accumulation were detected as the main adaptation strategy for environmental isolates; thus, isolation source may play an important role in genome plasticity, conferring a selective advantage to survive and persist for environmental *Salmonella* isolates. Intact prophage sequences with cargo genes were observable for both isolation sources playing an important role in virulence contribution.

1. Introduction

Salmonella is a worldwide prevalent pathogenic bacterium with disease symptoms ranging from enterocolitis to typhoid fever (Stevens et al., 2009). It has been estimated that *Salmonella* causes around 1.35 million infections resulting in 420 deaths in the United States each year (CDC, 2018) and it's typically acquired through the consumption of contaminated food or water with fecal material (Wray and Wray, 2000).

Salmonella life cycle comprises the colonization of animals and humans, although it is also ubiquitous in natural settings where it can reside until reaching a new host. The journey of *Salmonella* begins when contaminated food is ingested and reaches a low-pH stomach receptacle. Those able to survive the stomach acidic conditions make their way to the small intestine specifically in the ileum where *Salmonella* starts colonizing enterocytes, M cells, and dendritic cells in the intestinal epithelium by gaining nutrients and space (Garai et al., 2012; Velge

et al., 2012). After colonization takes place, *Salmonella* reproduces by binary fission to a large scale and the resulted clonal complex is released into the environment through fecal material to start the cycle again (Liu et al., 2018; Ranjbar et al., 2018).

Once excreted, *Salmonella* finds itself battling for survival facing limited nutrient availability, osmotic stress, variations in temperature, and pH (Winfield and Groisman, 2003). However, *Salmonella* is capable of sensing and responding to these stressful conditions. Otherwise, the bacterium could face potentially life-threatening conditions like those reported by Winfield and Groisman (2003) (Kenyon et al., 2002). These adverse conditions not only have a negative impact on *Salmonella* growth and survival but also influence virulence and resistance to several antimicrobials agents (Beceiro et al., 2013). Once the risk of perishing is overcome, the aquatic environment becomes the best niche for *Salmonella* to reside and reproduce (Raudales et al., 2014; Steele and Odumeru, 2004). Consequently, water becomes a major source of food

* Corresponding author at: CIAD, Culiacán, Carretera El Dorado Km 5.5, Campo el Diez, 80110 Culiacán, Sinaloa, Mexico.
E-mail address: chaqui@ciad.mx (C. Chaidez-Quiroz).

<https://doi.org/10.1016/j.meegid.2021.104771>

Received 27 August 2020; Received in revised form 4 February 2021; Accepted 11 February 2021

Available online 13 February 2021

1567-1348/© 2021 Elsevier B.V. All rights reserved.

contamination.

Several authors have proposed that *Salmonella* overcomes such stressful conditions by different survival strategies; for example, starvation-stress response (SSR) modulated by the scarcity of nutrients, carbon, and energy source shortage (C-starvation) (Spector, 1998; Spector and Cubitt, 1992). SSR has demonstrated that it accomplishes two main functions: combats the long-term effects of C-starvation and provides general resistance to other environmental stresses such as extreme temperature and pH (Kenyon et al., 2002). Two other conditions that exert protection to *Salmonella* in the environment are the formation of biofilm and the viable but nonculturable state. These conditions promote many cellular changes, including cell leakage, depletion of energy pools, altered expression of genes, and DNA replication (Morishige et al., 2017; Winfield and Groisman, 2003) increasing the survival rate in the environment.

The horizontal gene transfer (HGT) mediated by phages also allows *Salmonella* to diversify and survive in a variety of ecological niches by the incorporation of foreign genetic material into a recipient cell. This is known to be responsible for the acquisition of environmental, antimicrobial, and virulence genes playing an important role in the evolution of many bacterial species (Porwollik and McClelland, 2003).

Once *Salmonella* is settled down in the environment it can reach back human hosts causing food-borne diseases due to the consumption of fresh produce and animal products that come in contact with contaminated water (Greene et al., 2008; Hanning et al., 2009). Nevertheless, genetic differences between isolates from environmental and clinical settings have been poorly explored due to the limitations of the available techniques so far (Achtman et al., 2012; Page et al., 2017). The advent of whole-genome sequencing (WGS) technologies as well as the increasing number of bacterial genomes deposited in public repositories has enabled a deeper investigation to identify differential genomic signatures between strains isolated from distinct ecological niches and explore how genome plasticity and HGT are contributing to the survival of *Salmonella* outside the host (Balloux et al., 2018; Kwong et al., 2015).

Under this context, the present study aims to identify and characterize differences at a genomic level among *Salmonella enterica* isolated from environmental and clinical settings using different bioinformatics approaches and comparative genomics to gain a better understanding about the strategies of *Salmonella* to survive under the adverse conditions outside the host as a critical step in its life cycle and an intermediary stage to reach the human host and cause disease.

2. Materials and methods

2.1. Bacterial isolates and sequencing

A total of 34 *Salmonella enterica* strains isolated from Culiacan river and asymptomatic domestic animals belonging to serovars Saintpaul, Thompson, and Weltevreden (20, 6, and 8, respectively) were used. These strains belonged to the collection of the National Laboratory for Food Safety and Research (LANIA) and were isolated in the course of previous studies (Jiménez et al., 2011, 2014). Additionally, 3 clinical genomes for each of the serovars Saintpaul (CFSAN004173, CFSAN004174 and CFSAN004175), Thompson (RM1984, PNU-SAS019450 and PNU-SAS019186) and Weltevreden (2511STDY5462388, 2511STDY5462413 and 25STDY5712385) were downloaded from National Center for Biotechnology Information (NCBI) (Yao et al., 2017). DNA extraction for strains from LANIA's private collection was performed using the kit DNA easy Blood & Tissue Culture according to the manufacturer's instructions. After extraction, libraries were prepared with the Nextera XT DNA sample kit, and genomes were sequenced using Illumina Miseq platform. Three *S. enterica* belonging to serovars Saintpaul (CFSAN047351), Thompson (CFSAN047352) and Weltevreden (CFSAN047349) from the environmental source was also sequenced using nanopore technologies and assembled *de novo* by default settings in Canu program (Koren et al.,

2017). A hybrid assembly using both Nanopore and Miseq data was performed by comparing SPAdes (Bankevich et al., 2012) and Canu assemblies using Mauve to generate complete genomes for these isolates.

2.2. Assembling and annotation

Reads were quality trimmed using Trimomatic v0.32 (Bolger et al., 2014). The first 20 pb were removed from each read and a 4-base wide sliding window was used to cut when the average Phred quality score per base was below 15. Reads less than 50 bp were removed from the dataset. Draft genomes were assembled *de novo* for each strain using A5-miseq v20160825 (Coil et al., 2015). Moreover, the sequence of three *S. enterica* belonging to serovars Saintpaul (CFSAN047351), Thompson (CFSAN047352) and Weltevreden (CFSAN047349) from the environmental source were closed by a combination of long-read (using MinION platform), and short-read (Illumina Miseq) sequencing as described in the previous work (Gonzalez-Escalona et al., 2019).

The generated assemblies were annotated using Prokka v1.11 (Seemann, 2014) and the ST groups identification was performed using Ariba v2.14.4 (Hunt et al., 2017) based on the PubMLST database scheme for *Salmonella* spp. (Jolley et al., 2018). To explore *Salmonella* metabolic pathways in clinical and environmental isolates the amino acid sequence in FASTA format for each genome was annotated with BlastKOALA (Kanehisa et al., 2016) then KEGG Mapper (Kanehisa and Sato, 2020) was used to reconstruct the metabolic pathways based on the previous genome annotation.

2.3. Pan-genome construction

We used the anvio's pan-genome workflow (Eren et al., 2015) to construct the pan-genome for each serovar based on the identification of gene clusters among the genomes. This workflow consisted of three main steps: the first one included the generation on anvio's genomes storage using the program "Anvi-gen-genomes-storage" to store DNA and amino acid sequences from external genomes, as well the annotation of genes. The second step consisted of the generation of an Anvi'o pan database using the program "anvi-pan-genome" to run the pangenomic analysis and finally the distribution visualization of gene clusters across genomes. To compute the *Salmonella enterica* serovars Saintpaul, Thompson, and Weltevreden pangenome we first generated an anvio's genomes storage database from the FASTA file of 23, 9 and 11 genomes for each serovar respectively using the environmental and clinical genomes. We ran the pangenome analysis with the genomes storage database for each serovar using the flag `-use-ncbi-blast` for amino acid sequence similarity search. Additionally, the parameters `-minbit 0.5` and `-mcl-inflation 10` were used to define the sensitivity of the algorithm during the identification of gene clusters for the comparison of very closely related genomes. Finally, the anvio's pangenome that stored all results was visualized. Within the context of anvio's analysis, proteins clusters refer to the group of translated DNA sequences from predicted open read frames that share a close degree of homology among all genomes, while gene clusters represent a set of genes found within the genomes that encode similar proteins. Gene calls were clustered into bins based on their frequency across genomes: the first one corresponds to core bin (genes presented in 95 to 100% of the genomes), shell bin (genes between 15 and 95%) and finally cloud bin (presented in 0 to 15%) (Page et al., 2015).

2.4. SNP calling and phylogenetic tree

A core genome alignment based on SNPs was constructed for each of the three serovars using Parsnp with a random clinical genome as a reference. The resulting alignments were explored visually with Gingr to identify the core percentage shared among all strains. A multi-FASTA file was generated for each alignment with HarvestTools (Treangen

et al., 2014) using as input the gnr file. A maximum-likelihood tree was built from the core genome SNP alignment using RAXML (Stamatakis, 2014) with the general time-reversible GTR model and gamma distribution with 100 bootstrap replicates to assess statistical support. Each resulting tree was midpoint rooting, visualized, and edited on iTOL v4 (Letunic and Bork, 2007). Additionally, a SNP difference matrix among each serovar was performed using CSI Phylogenetic v1.4. (Kass et al., 2014) selecting a minimum SNP quality of 30 with a minimum depth of 30× for each SNP position and ignoring heterozygous SNPs.

2.5. Antimicrobial, virulence gene and plasmid identification

AMRFinderPlus v3.1.1 (Feldgarden et al., 2019) command-line version was used for the detection of acquired or intrinsic antimicrobial resistance gene products using the NCBI database as part of its Pathogen Detection Pipeline. For this purpose, we ran the recommended command line using as input the FASTA assembly, the annotation file (GFF), and the protein sequence. Also, we added the plus option to detect virulence factors and stress-response genes. To detect point mutations in antimicrobial resistance genes (AMR) we specified *Salmonella* spp. as taxon. Additionally, ResFinder v3.2 (Zankari et al., 2012) was used to corroborate and analyze the acquired AMR and chromosomal mutations. For this purpose, we selected genus *Salmonella* with the option to show only known mutations above 90% of identity with a minimum alignment of 60%. This was tested in all closed environmental and clinical isolates. Nevertheless, 2 draft clinical genomes were used for serovar Thompson due to the lack of closed genomes.

To accurately detect virulence genes in the genomes we used the VFAnalyzer to construct orthologous groups and then compare it to reference genomes from The Virulence Factors Database (Liu et al., 2019). We used option 3 to upload private data selecting *Salmonella* as a genus. PlasmidFinder v2.1 (Carattoli et al., 2014) was used for the rapid identification of plasmids in *Enterobacteriaceae* species that are of interest for epidemiological investigations considering plasmid associated with the spread of antimicrobial resistance. For the analysis, we select a 95% identity and 60% as coverage as a minimum.

2.6. Phage identification and CRISPR region prediction

To explore HGT mediated by phages in the closed environmental and clinical isolates PHASTER web service (Arndt et al., 2016) was used. Intact phage sequences (with score > 90) for all genomes were selected to analyze the contribution to virulence or antimicrobial resistance using the VFDB database (Liu et al., 2019) and ResFinder v3.2 (Zankari et al., 2012). CRISPR regions were predicted using CRISPRCasFinder (Couvin et al., 2018) with 100 bp of flanking regions and performing CAS gene detection.

2.7. Secretion system search

Secretion system associated proteins were detected by MacSyFinder (Abby et al., 2014) using the TXSScan (Abby et al., 2016) model proteins profile selecting ordered replicon as type of dataset and all protein secretion systems to perform the detection (Falgellum, pT4SSi, pT4SSi, T1SS, T2SS, T3SS, T4P, T4SS_typeB, T4SS_typeC, T4SS_typeF, T4SS_typeG, T4SS_typeI, T4SS_typeT, T5aSS, T5bSS, T5cSS, T6SSiii, T6SSii, T6SSi, T9SS and Tad). Additionally, maximal e-value was adjusted to 1×10^{-4} .

3. Results

3.1. Pan-genome analysis

Pangenomic analysis of the 23, 9, and 11 isolated genomes from both environmental and clinical sources belonging to serovars Saintpaul, Thompson, and Weltevreden resulted in 5135, 4428 and 5499 protein

clusters, respectively (Table 1). For serovar Saintpaul, we detected 4275 and 860 protein clusters for bins core and cloud respectively. *S. Thompson* protein clusters were distributed in core, shell, and cloud bins with 4200, 71, and 157 clusters. Finally, Weltevreden holds 4273, 295, and 917 protein clusters associated with bins core, shell, and cloud (Table 1). A similar percentage of annotated genes with clusters of orthologous groups (COGs) functions was found for serovars Saintpaul and Thompson (80.7 and 83.6%) which most of that annotations are harbored in the core bin. On the other hand, 71.5% percent of annotated clusters were detected for Weltevreden which are also included in the core bin principally.

A total of 101,793, 39,120, and 53,223 genes for serovars Saintpaul, Thompson, and Weltevreden were identified by the gene call annotation in the contig databases for each genome (Table 1) respectively. More than 92% of the genes detected are included in the core.

Specific genes were identified in *S. Saintpaul* genome of isolates from environmental sources (Fig. 1). These genes were related to iron and magnesium acquisition systems ABC type, which are important for enteropathogenic bacteria by the catalysis of oxide-reduction reactions and the iron uptake for enzyme cofactors (Drago-Serrano, 2009). Moreover, additional copies of metabolic genes implied in the Entner-Doudoroff pathway as triosephosphate isomerase (tpiA) and phosphoglycerate mutase (gpmA) (Flaumholz et al., 2013) were detected. Concerning to clinical genomes, specific segments shared for all genomes were identified. These segments included genes related to phages, integrases, transposases, pilus, and type II secretory pathway fundamentally.

Genomes of *S. Thompson* isolates were more uniform with a reduced number of accessory genes (Fig. 2). Nevertheless, a specific region was detected in genomes from environmental sources which was not present in the two clinical genomes. This region was represented by the shell bin and included genes like YdaS, a member of Cro superfamily coding for a DNA binding protein to the xenobiotic response which is responsible for signal transduction of exogenous environment pollutants (Kuramoto et al., 2003), as well as genes related to the mobilome mediated by prophages (structural proteins for tail and capsid) and transposons. Also, genes associated with carbohydrate transport and metabolism (designed as G category for COGs) required for transcriptional repression during the growth on glycerol (Sherwood et al., 2009) were found. Additionally, genomes from environmental isolates also showed the presence of DNA replication proteins that protect cells against DNA damage resulted from alkylating and supercoiling activity were identify (Huang and Huang, 2018). The clpP gene was identified, which plays a central role in *Salmonella* to grow under various stressful conditions such as low pH, elevated temperature, and high salt concentrations (Thomsen et al., 2002). Just for one clinical genome a specific segment in the cloud bin was detected with genes associated with type II and IV secretory pathway, extracellular structures for pilus assembly, and phage related genes.

Finally, serovar Weltevreden pangenome showed in shell bin a specific gene content for only the environmental genomes. This unique block consisted of genes that codify transposases, recombinases for cutting and rejoining of DNA, SOS-response transcriptional repressors, and phage related genes. Moreover, a collinear block is representative for all environmental and one clinical genome, the genes in the segment are related with mobilome.

Concerning to the cloud bin, it comprises many specific gene clusters for clinical and environmental strains. The specific bits for clinical isolates included genes for T3SS as well as genes related to phage structure and integrases. The specific singleton in the clinical genome 2511STDY5712388 holds the gene RadD involved in radiation resistance and double-strand break repair (Chen et al., 2015), as well as the cytolethal distending toxin subunit B been a virulence factor in promoting the infection persistence (Pons et al., 2019). In this component of the genome of environmental isolates were identified some specific regions which genes categorized as hypothetical proteins due to the short

Table 1The number of gene clusters and genes detected across *S. enterica* serovars Saintpaul, Thompson, and Weltevreden from environmental and clinical isolates.

Serovar pan-genome	Number of protein clusters				Number of genes calls			
	Bins			Total	Bins			Total
	Core	Shell	Cloud		Core	Shell	Cloud	
<i>S. enterica</i> Saintpaul	4275	0	860	5135	99,985	0	1808	101,793
<i>S. enterica</i> Thompson	4200	71	157	4428	38,421	523	176	39,120
<i>S. enterica</i> Weltevreden	4273	295	917	5499	48,915	2871	1283	53,069

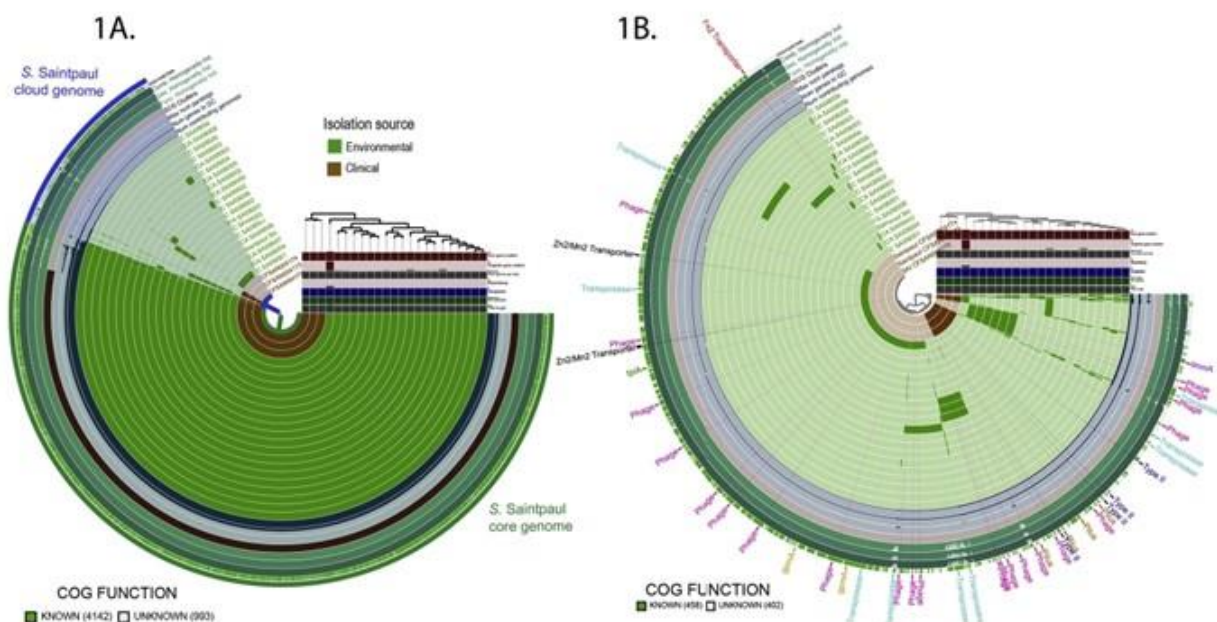


Fig. 1. (A) Pangenome of *S. Saintpaul* genomes isolated from environmental and clinical sources. Rings in green represent environmental genomes and those in brown the clinical. The outer colored ring indicates the classification of gene clusters. The number of the known and unknown gene function are indicated below. (B) Cloud split pangenome. Labels in the outer ring represent the annotated genes included in cloud pangenome. The outer colored ring indicates the classification of gene clusters. The number of the known and unknown gene functions are indicated below as COG function.

fragment of the ORF detected (20 bp). The only annotated genes corresponded to copies of genes *lexA*, which main function are the transcriptional repressor for the regulation of DNA repair in many bacteria (Smollett et al., 2012). Besides, the activator or repressor for mercury resistance operon *merR* was identified.

Interestingly, all environmental genomes displayed multiple copies of genes that play key roles in resistance to environmental stress in *Salmonella enterica* serovars. For example, genes *soxS* and *OxyR*, which have been reported in mechanism associated with oxidative, and bile resistance or genes *SlyA* and *RcsC* involved in antimicrobial peptide resistance (Spector and Kenyon, 2012). Additionally, copies of genes related to biofilm formation as *fimA* were identified. As part of core genome for all the serovars for both environmental and clinical sources, genes associated with C-starvation (*rpoE*, *rpoS*, *cprX* and *cAMP-CRP*) and biofilm formation (*csgDEFG-scgBAC* and *bcsABZC*) VBNC (*rpoS*) were detected but in single copies.

Until 230 metabolic pathways associated with carbohydrate, lipid, energy, nucleotide, amino acids, cofactors and vitamins, glycan, terpenoids and polyketides metabolism were annotated and identified according to KEGG Mapper for all genomes (Fig. 3). Module D-galactonate degradation belonging to galactose metabolism was differentially detected in all environmental strains and some few clinical isolates (33%).

3.2. Antimicrobial resistance gene and plasmid replicons identification

All the clinical and the environmental isolates were analyzed *in silico* to determine antimicrobial resistance genes and point mutations inducing resistance. A total of 6 genes (with identity >95%) for antimicrobial resistance, efflux pump, and stress genes were detected by Resfinder and ARMFinder (Table 2). Genes *GolS*, *GolT*, *mdsA*, *mdsB*, and *aac(6')-Iaa* (gold stress response, efflux pump complex, and aminoglycoside resistance) were found in all isolates. *arsD* gene was the only one distinctively detected in environmental Thompson and Weltevreden genomes. Its role is the response to metal arsenic stress.

Chromosomal point mutations were explored for genes *pmrB*, *pmrA*, *gyrB*, *16S.rrsD*, *parC*, *gyrA*, and *parE*. Only one-point mutation was detected in *parC* gene inducing a codon change from threonine to serine conferring resistance to nalidixic acid and ciprofloxacin. This mutation was identified in serovars Thompson and Weltevreden for both sources. No point mutations for genes listed previously were determined for serovar Saintpaul in either source.

An *in silico* typing and identification of plasmid replicons from a collection of 5599 fully sequenced plasmids associated with the members of the family Enterobacteriaceae was performed, including those ones for *Salmonella* Typhimurium. Plasmid replicons *IncI1* and *IncFII(S)* were respectively identified in all the three clinical genomes of serovars Saintpaul and Weltevreden. Only one clinical genome of serovar

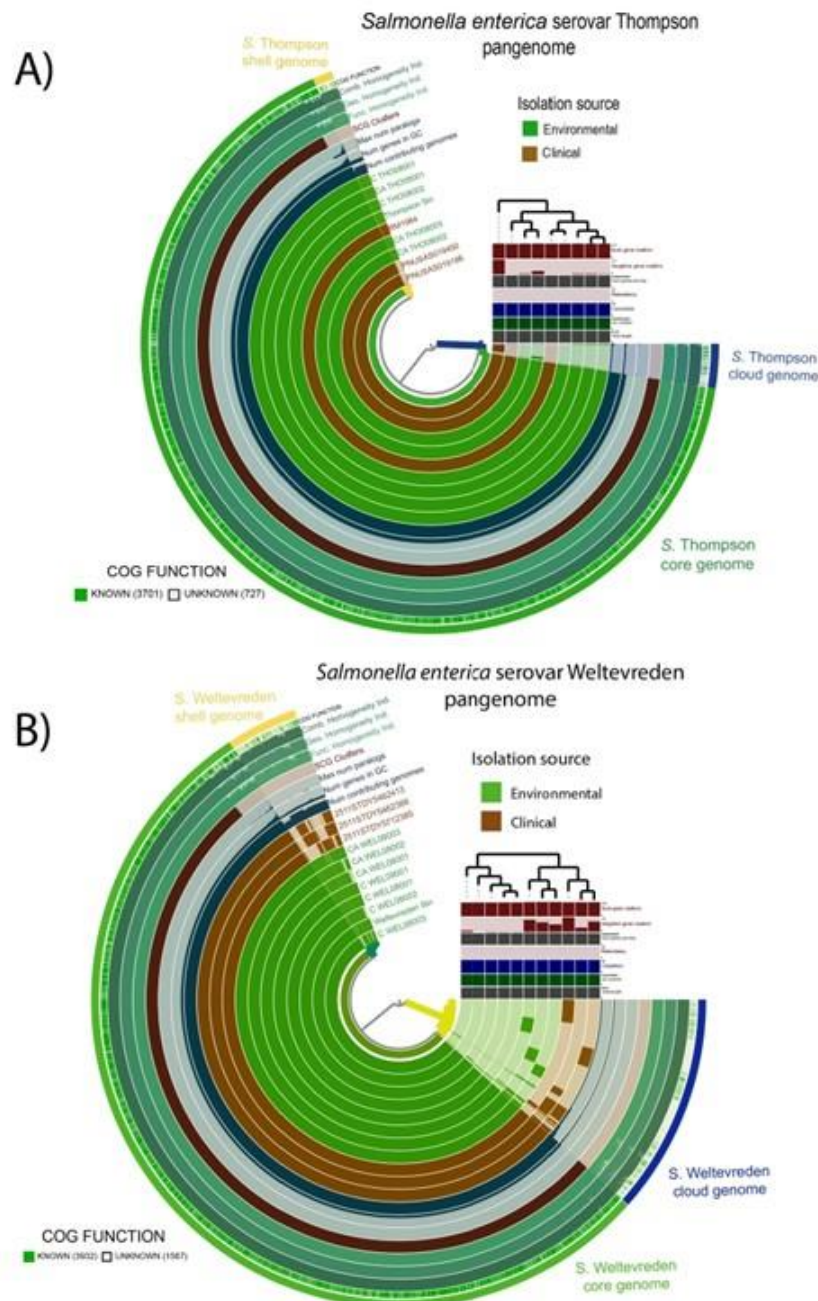


Fig. 2. (A) Pangenome of *S. Thompson* genomes isolated from clinical and environmental sources. (B) Pangenome of *S. Weltevreden* genomes isolated from clinical and environmental sources. Rings in green represent environmental genomes and those in brown the clinical.

Thompson showed the presence of replicon IncF(p96A). For environmental isolates, the replicon IncFII(S) was only detected for serovar Weltevreden. All matched replicons had an identity percentage higher than 97%.

3.3. Identification of virulence genes

Virulence genes related to fimbrial adherence determinants, macrophage inducible genes, magnesium uptake, non-fimbrial

adherence determinants, secretion system (SPI-1 and SPI-2), serum resistance and stress adaptation were analyzed for all genomes (Fig. 4). A similar virulence profile was determined for genomes belonging to the same serovar. Interestingly, the *SodCI* stress adaptation gene was not found in clinical genomes but was present in all the environmental isolates. Additionally, genes *cdtB* and *plxA* (codifies for typhoid toxin) were detected in clinical genomes (2 in *S. Weltevreden* and 1 in *S. Thompson*) and in one environmental genomes of serovar Weltevreden.

Differences in the virulence profiles among serovars were identified.

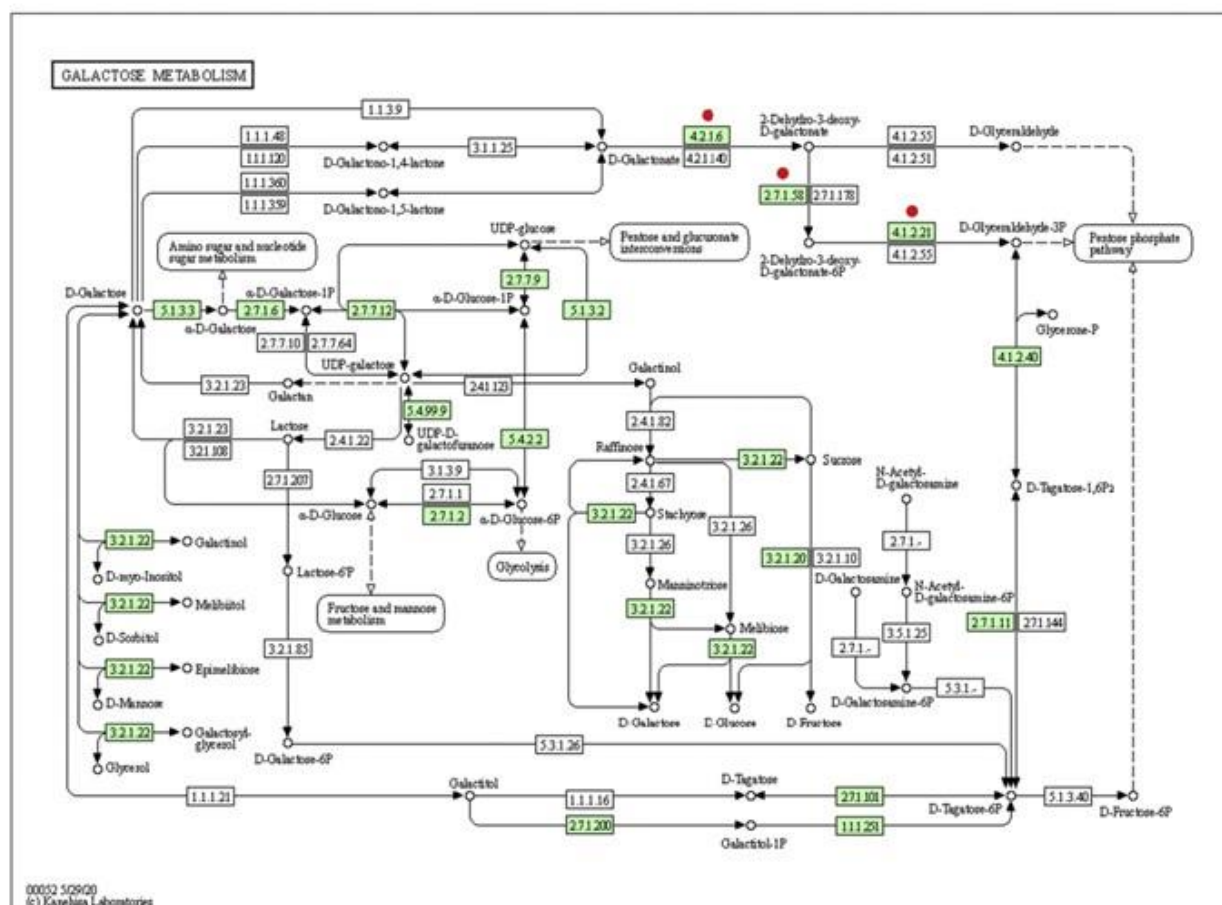


Fig. 3. Galactose metabolism pathway. Green colored labels represent participation in galactose metabolism. Blue circles above labels indicate enzymes coded by operon *ldg* involved in D-galactonate degradation in environmental genomes.

Table 2

Antibiotic resistance genes detected for closed clinical and environmental isolates for serovars Saintpaul, Thompson and Weltevreden. Symbol “+” indicates presence and “-” means absence.

Genomes			Resistance and stress response genes					
			<i>GolS</i>	<i>GolT</i>	<i>mdsA</i>	<i>mdsB</i>	<i>arsD</i>	<i>aac(6')-Iaa</i>
S. Saintpaul	Environmental	CFSAN047351	+	+	+	+	-	+
	Clinical	CFSAN004173	+	+	+	+	-	+
		CFSAN004174	+	+	+	+	-	+
S. Thompson	Environmental	CFSAN004175	+	+	+	+	-	+
	Clinical	CFSAN047352	+	+	+	+	+	+
		PNUSA019186	+	+	+	+	-	+
S. Weltevreden	Environmental	PNUSA019450	+	+	+	+	-	+
	Clinical	RM1984	+	+	+	+	-	+
		CFSAN047349	+	+	+	+	+	+
		2511STDY5462413	+	+	+	+	-	+
		2511STDY5462413	+	+	+	+	-	+
		2511STDY5712385	+	+	+	+	-	+

For example, *S. Weltevreden* lacked the long polar fimbria (*lpfABCDEF*); the operon *peg* (*pegABCD*) was detected exclusively for serovar Thompson; the fimbrial adherence operon *stk* (*stk ABCDEFG*) was present in Thompson and Saintpaul; operon *stc* (*stc ABCD*) was found in serovars Saintpaul and Weltevreden. Nevertheless, all the essential genes related to the SPI-1, SPI-2, and translocated effectors for both pathogenic islands were found conserved among all three serovars.

The serum resistance gene (*rck*) was identified in only one clinical

Thompson genome. All of the genomes had macrophage inducible genes (*mig-14* and *mig-5*), magnesium uptake (*mgtB* and *mgtC*), nonfimbrial adherence determinants (*misL*, *ratB*, *shdA*, and *sinH*) and regulation genes (*phoP* and *phoQ*). No capsular Vi antigen genes were detected for any genome.

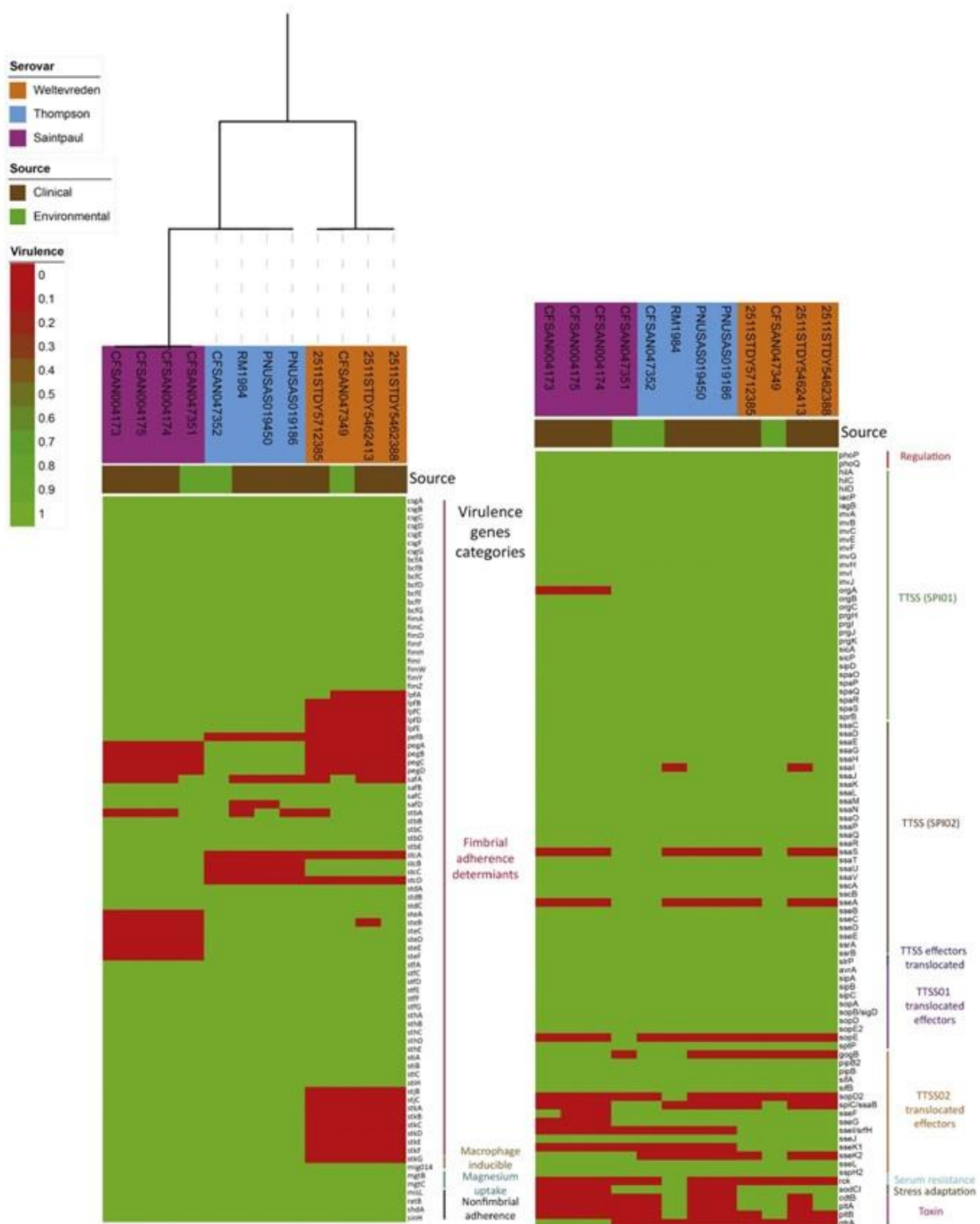


Fig. 4. Phylogenetic tree with the presence of virulence genes. Colored labels indicated *Salmonella* serovars. The identity scale is represented in red (absence) to green (presence) range. Virulence genes and virulence category labels are indicated in next to the heatmap.

3.4. Phage identification and CRISPR region prediction

A total of 37 intact lysogenic phages were identified by PHASTER (Table 3). The most common families included relatives of Gifsy-1 (NC_010392) and Gifsy-2 (NC_010393) phages. Gifsy-1 like phage was found in all clinical and environmental *Salmonella* serovars and Gifsy-2 like phage were present in all the Weltevreden genomes and one environmental Saintpaul genome. Additionally, some particular phage sequences were found exclusively in a specific serovar. For example, the Entero_PsP3 like phage (NC_0053408) and Salmon_Fels_1 (NC_010391) were identified in all the Weltevreden genomes. Additionally, the Salmon_ST64T (NC_004143) was found in the Saintpaul genomes. To gain a better understanding of the role of phages in virulence and environmental adaptation, we searched for carried non-phage cargo genes (Table 4). We detected 7 cargo genes related to *Salmonella* virulence, these include *sopE*, *ssrL*/*srfH*, *ctdB*, *plxA*, *shdA*, *zur*/*yjbK* and *gogB*. Interestingly, the cytolethal distending toxin subunit B was detected for almost all genomes belonging to serovar Weltevreden. *S. Typhi* toxin (pertussis-like toxin) *plxA* was detected for one Weltevreden genome belonging to both strain sources. Genes *spoE* and *gogB* consider as TTSS-1 and TTSS-2 translocated effector were detected in one environmental Saintpaul genome, and in one each clinical and environmental Thompson genomes. Clinical Weltevreden genomes hold the biggest virulence cargo genes with additionally *sdhA*, a nonfimbrial adherence determinant and *ssrL*/*srfH* a TTSS-2 translocated effector.

Type I-E class 1 CRISPR-Cas system which uses a multisubunit crRNA-effector for immunity and characterized by the presence of Cas3, Cas10, and CSF1 was detected for all *Salmonella* genomes (Table 4). Two CRISPR arrays with different numbers of spacers were detected across all genomes. Spacer sequences were also analyzed to detect target prophages with Blastp using UniProt phage sequences (<https://www.uniprot.org/>). Prophage targets were observed in 38% (9/24) of the total genomes with the most common prophage targets being tail and holins proteins.

3.5. Detection of secretion systems in *Salmonella enterica* genomes

An exploration of 22 models of the protein secretion system was performed for all clinical and environmental genomes. We detected 7 models with the presence of mandatory genes (essential in a system to be functional) related to flagellum, T1SS, T3SS, T4SS_typeF, T5SS, T5Bss, and T5cSS (Fig. 5). Differences among environmental and clinical

genomes were observed. For example, all clinical isolates held the mandatory genes related with the flagellum (*flhF*, *flhI*, *flhN/M*, *flp*, *flhQ*, *flhR*, *flhB* and *flhA* genes) and T3SS (*setC*, *setJ*, *setN*, *setQ*, *setR*, *setS*, *setT*, *setU* and *setV* genes) which are a crucial requirement for *Salmonella enterica* and other Enterobacteriaceae members to be able to infect a host. Nevertheless, the T4SS_typeF had a 10% higher frequency in genomes from environmental isolated than from the clinical ones. This secretion system encodes proteins involve in signaling and host cell interaction and is essentially characterized by various stages of pseudogenization revealed by gene duplication and deletions. System T5SS subtypes T5bSS (two-partner passenger-translocators) and T5cSS (trimeric auto-transporters) were identified for one Weltevreden clinical isolate. These subtypes are characterized by the use of the Sec translocase to pass through the inner membrane to the periplasmic space. Finally, mandatory genes *abc* and *onf* belonging to T1SS were detected for all both environmental and clinical genomes.

4. Discussion

Surface water is an inhospitable environment for *Salmonella*, therefore the ability to survive, resist, and adapt are conditioning factors that can determine the continuity of its life cycle (Liu et al., 2018). During its passage through the environment, *Salmonella* manages to be part of the microbial community of diverse environments, among which stand out the soil and the water where it has been reported that it manages to stay for long periods (Gorski et al., 2011; Martinez-Urtaza et al., 2004). However, the behavior of *Salmonella* in adverse aquatic environments has been poorly studied (Levantesi et al., 2012). The metabolic and genetic changes that adverse aquatic environments exert on the *Salmonella* colonization and adaptation processes are yet to be discovered. There have been efforts to address this question. Medrano-Félix et al. (2018) reported the metabolic profiles and diversity of carbohydrates used by *Salmonella* isolates under river water conditions (Medrano-Félix et al., 2018). Also, Estrada-Acosta et al. (2018) determined how aquatic environments affect *Salmonella*'s capacity to infect hosts (Estrada-Acosta et al., 2018). High-resolution genomic signatures could be another piece in the puzzle to understand *Salmonella*'s ability to colonize natural environments as river water and at what cost.

The construction of pangenome enabled the identification of serovars-specific genetic signature in isolates from both environmental and clinical sources. Also, accessory genome detection and quantification increased the understanding of the isolation sources, virulence and

Table 3
Intact lysogenic phages and cargo genes detected in clinical and environmental genomes.

<i>Salmonella</i> serovar	Sample	Source	No of phage	Best match (kb)	Virulence genes on phage
Saintpaul	CFSAN047351	Environmental	2	Entero_P88_NC_026014(21), Phage_Gifsy_2_NC_010393(47)	<i>sopE</i>
	CFSAN004173	Clinical	2	Phage_Gifsy_1_NC_010392(39), Salmon_ST64T_NC_004348(24)	–
	CFSAN004174	Clinical	2	Phage_Gifsy_1_NC_010392(39), Salmon_ST64T_NC_004348(24)	–
	CFSAN004175	Clinical	2	Phage_Gifsy_1_NC_010392(39), Salmon_ST64T_NC_004348(24)	–
Weltevreden	CFSAN047349	Environmental	6	Entero_PsP3_NC_005340(35), Gifsy_1_NC_01039(26), Haemop_HP1_NC_001697(18), Haemop_HP1_NC_001697(19), Phage_Gifsy_2_NC_010393(46), Salmon_Fels_1_NC_010391(27)	<i>ssrL</i> / <i>srfH</i> , <i>ctdB</i> , <i>plxA</i>
	2511STDY5462388	Clinical	8	Entero_PsP3_NC_005340(36), Phage_Gifsy_1_NC_010392(25), Phage_Gifsy_2_NC_010393(47), Salmon_118970_sal3_NC_031940(27), Salmon_epsilon15_NC_004775(36), Salmon_Fels_1_NC_010391(20), Salmon_g341c_NC_013059(48), Salmon_SEN34_NC_028699(33)	<i>ssrL</i> / <i>srfH</i> , <i>ctdB</i>
	2511STDY5662413	Clinical	8	Entero_PsP3_NC_005340(36), Escher_phiV10_NC_007804(36), Gifsy_1_NC_01039(25), Phage_Gifsy_2_NC_010393(46), Salmon_118970_sal3_NC_031940(61), Salmon_Fels_1_NC_010391(20), Salmon_SEN34_NC_028699(33), Shigel_SAV_NC_022749(1)	<i>shdA</i> , <i>ssrL</i> / <i>srfH</i> , <i>zur</i> / <i>yjbK</i>
	2511STDY5712385	Clinical	4	Entero_PsP3_NC_005340(35), Phage_Gifsy_1_NC_010392(23), Phage_Gifsy_2_NC_010393(36), Salmon_Fels_1_NC_010391(29)	<i>ssrL</i> / <i>srfH</i> , <i>ctdB</i> , <i>plxA</i>
Thompson	CFSAN047352	Environmental	1	Phage_Gifsy_1_NC_010392(39)	<i>gogB</i>
	PNUSAS19186	Clinical	1	Salmon_Fels_2_NC_010463(28)	–
	RM1984	Clinical	1	Phage_Gifsy_1_NC_010392(39)	<i>gogB</i>
	PNUSAS19450	Clinical	0	–	–

Table 4
Prophage target identified in spacer sequences in CRISPR arrays in *Salmonella* serovars.

<i>S. enterica</i> serovar	Sample	Source	CRISPR number	Cas cluster	Spacer		
					no	target	gene
Saintpaul	CFSAN047351	Environmental	2	CAS-TypeIE	22	prophage	Phage tail protein
					13	-	-
	CFSAN004173	Clinical	2	CAS-TypeIE	20	prophage	Phage tail protein
					15	-	-
	CFSAN004174	Clinical	2	CAS-TypeIE	20	prophage	Phage tail protein
Wелtevređen	CFSAN004175	Clinical	2	CAS-TypeIE	15	-	-
					20	prophage	Phage tail protein
					15	-	-
	CFSAN047349	Environmental	2	CAS-TypeIE	34	-	-
					44	-	-
Thompson	2511STDY5462388	Clinical	2	CAS-TypeIE	40	prophage	phage holin
					25	prophage	Phage tail protein
	2511STDY5662413	Clinical	2	CAS-TypeIE	32	prophage	phage holin
					29	prophage	Phage tail protein
	2511STDY5712385	Clinical	2	CAS-TypeIE	38	-	-
Thompson	CFSAN047352	Environmental	2	CAS-TypeIE	25	prophage	Phage tail protein
					14	-	-
					18	-	-
	PNUSAS19186	Clinical	2	CAS-TypeIE	17	-	-
					28	-	-
Thompson	PNUSAS19450	Clinical	2	CAS-TypeIE	19	-	-
					23	-	-
					26	-	-
	RM1984	Clinical	2	CAS-TypeIE	18	-	-
					18	-	-

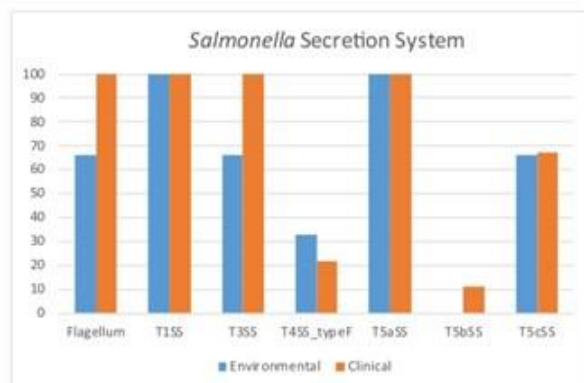


Fig. 5. *Salmonella enterica* serovars Saintpaul, Thompson, and Weltevreden secretion system detection. Environmental genomes are denoted in blue and clinical in orange. Each bar is labeled with the secretion systems detected.

resistance potential difference (Guimarães et al., 2015). To this end, we determined that serovar Saintpaul holds the biggest number of cloud-related genes with detectable environmental and clinical gene clusters, followed by *S. Weltevreden* and serovar Thompson. Serovar Weltevreden showed the biggest genetic variation in the pangenome among clinical and environmental isolates, this could be to the submissive genome structure characterized by the ability to routinely acquired and lose genetic material as previously reported for this serovar (Makendi et al., 2016). However, a similar number of core genes clustered among and within all serovars were detected. Therefore, a basic set of genes present in all *Salmonella* could be a testimony to the conservative nature of evolution (Lapierre and Gogarten, 2009). Genes related to the basic biology of *Salmonella* such transport and metabolism of carbohydrates, cell wall and envelope biogenesis, DNA process (replication, transcription and translation), intracellular trafficking, secretion and vesicular transport were found as the main component of core genome (Anjum et al., 2005).

Environmental stressful conditions may trigger, in shell and cloud

genomes, a high level of gene duplication, which might relate to the evolutionary process in many prokaryotic organisms. Therefore, the faster the bacteria settle in the environment, the lesser need for duplication of genes is generated (Bratlie et al., 2010). Furthermore, a small number of protein families (~ 5021 groups) detected in the serovars under study, reflects the flexibility of those genes in their ability to adapt new function indicating that the preferred mode of adaptation consists in the exploration of new strategies by gene duplication or mutation (Lapierre and Gogarten, 2009). Singletons of pseudogenes were detected for all three serovars, supporting the notion that the evolution of *S. enterica* serovars towards becoming host-adapted and host-restricted is characterized by the accumulation of pseudogenes (Waldner et al., 2012). The superoxide dismutase gene is a genome-constituent of *Salmonella*. In aquatic environments, the gene expression results in the prevention of the oxidation process set by phagocytic superoxide (Tidhar et al., 2015). Moreover, environmental isolates must survive predation from bacteriophage and predatory bacteria, under this context we detected biofilm formation genes which have been reported as a defense mechanism by a membrane rearrangement to avoid being ingested (Erken et al., 2013). Also, T3SS has been described as an *E. coli* and *V. parahaemolyticus* strategy to survive inside trophozoites of *A. castellanii* and *C. roenbergensis* which are predominant protozoa in natural environments such river and ocean water (Matz et al., 2011; Siddiqui et al., 2011), these strategies may additionally help *Salmonella* to predation evasion and survival in non-host environments.

D-galactose operon, present in all environmental isolate genomes, plays a crucial role in the degradation process of the D-galactonate. Humans are not the principal source of D-galactose contrary only a mere fraction is produced. Most production comes from nature, mainly prokaryotic cells (Singh et al., 2019). The metabolic flexibility by *Salmonella* during D- galactonate degradation (Selzer, 2016) makes the bacteria survive, establish, and persist in river water with high possibility to reach back human hosts by means of consumption of contaminated food or water.

Nevertheless, type II and IV secretory pathways genes, toxins, pilus assembly, and mandatory genes detected mainly in clinical isolates, contribute to bacterial colonization by the invasion of phagocytic and non-phagocytic cells (Ibarra and Steele-Mortimer, 2009). T3SS is an essential secretion system for actin rearrangements required for

Salmonella internalization, pathogenicity, and survival inside the host. Moreover, this system has been found as crucial for other pathogenic bacteria as *Yersinia* and *Shigella* (Green and Mecsas, 2016). Additionally, Salmonella eases the cellular internalization with T1SS, fimbriae adherence, and, flagella. For example, *lpf* operon detected for serovars Saintpaul and Thompson plays an important role in *Salmonella* adhesion to Peyer's patches, the portal of entry for all non-typhoidal *Salmonella* infections (Wagner and Hensel, 2011). The presence/absence of virulence operons found in this work suggest that fimbriae represent a source of diversity among *Salmonella* serovars as previously reported for several authors (Dufresne et al., 2018). Even though T5SS is not described as an essential system for *Salmonella*, it may contribute to bacterial virulence. The secretion of substrates may serve as toxins and receptor binding proteins (Wells and Henderson, 2013).

Bacteria infected with lysogenic phages produce a more diversified progeny including increases in virulence factors (Gupta et al., 2019). Moreover, Gifsy-1 and Gifsy-2 like phage are the most common phages found in *Salmonella* genomes and had been reported in other *Salmonella* serovars as Typhimurium playing an important role in the contribution of virulence (Ho and Schlauch, 2001). *SsrB* and *gogB* were identified as important cargo genes of Gifsy-1 and Gifsy-2 which have a principal role in activating global regulon of acquired genes (Worley et al., 2000) and increasing *Salmonella* survival after phagocytosis (Kutter and Sulakvelidze, 2004), respectively. Our study supports the notion that phages have an important role in the pathogenicity in *Salmonella* serovars.

CRISPR-Cas is an immunity strategy to avoid phage attack based in a set of Cas proteins with functional domains related to nucleases, helicases, and polynucleotide-binding proteins (Medina-Aparicio et al., 2018), due to the high death rate (4–50% killed bacteria) caused by phage lysis. Here we report that *Salmonella* genomes have the CRISPR array 1 (CRISPR1) associated with the Type I-E set of cas genes, which had been exclusively reported for *Salmonella* (Shariat et al., 2015) and other Enterobacteriaceae members as *E. coli*. Spacer contained in the CRISPR arrays provide a memory of the foreign elements that have invaded *Salmonella*. The 38% of spacers found in this study belong to phages, representing higher values than reported for other studies with 10% associated values. Non identified spacer sequences are believed to correspond to plasmids and their own genome sequence as shown in other studies (Medina-Aparicio et al., 2018). Differences in spacer numbers were detected among isolation sources, for example, clinical isolates belonging to serovar Thompson displayed higher numbers associated with spacers. This could be explained under a microbial adaptation context considering that environmental isolates rely on HGT mechanism to fit the non-host environment and the acquisition of CRIPR elements had been reported as a barrier for HGT, the reason why bacteria may eliminate or reduce CRISPR elements (Westra et al., 2016). Nevertheless, further studies may be conducted.

By the employment of AMRfinder all studied serovars concurred with the same pattern of antibiotic resistance and stress response. Interestingly, the metallochaperone *arsD* which is a constituent gene in *ars* operon providing arsenic resistance in natural environments and industrial contaminated sites (Jackson and Dugas, 2003) was the only difference detected in the environmental Thompson and Weltevreden isolates respecting the clinical ones. This adaptive signature gives *Salmonella* ways to survive and establish themselves in non-host environments such as high arsenic river water (Joerger et al., 2010). The *parC* mutation observed in both isolates (environmental and clinical) analyzed in the present study is the same that *Salmonella* Typhimurium and *E. coli* poses. This mutation provides resistance to fluoroquinolones due to amino acid change in the *ParC* subunit of topoisomerase IV, which is the second target of quinolones, and a growing worldwide problem (Baucheron et al., 2004; Dalhoff, 2012).

5. Conclusion

In the spite of the limited number of analyzed genomes in this study,

the application of WGS and bioinformatics approaches allowed a comprehensive analysis and full genomic comparison between *Salmonella* isolates from environmental and clinical sources. Genes associated with stress resistance and metabolic pathways were identified as environmental isolates signature that may represent an adaptation and survival advantage in non-host environments as river water, suggesting that the isolation source plays an important role in genome plasticity. In addition to HGT we identified gene duplication and pseudogenes accumulation as environmental survival strategies characterized by genome rearrangements. According to the accessory genome of clinical isolates, secretion system, flagellum, and virulence genes seem to be preserved among all genomes as essential elements to invade, internalize, and cause pathogenesis. The most common phages sequences as Gifsy_1 and Gifsy_2 reported for *Salmonella* were detected, these held virulence cargo genes contributing of bacterial virulence, supporting the idea that horizontal gene transfer enhanced bacterial fitness.

Declarations of interest

None

Credit author statement

J. R. Aguirre-Sanchez: Formal analysis, Writing – original draft and Investigation. J. Martínez-Urtaza: Writing – review & editing. J. R. Ibarra-Rodriguez: Investigation. I. F. Vega-Lopez: Investigation. C. Chaidez-Quiroz: Writing – review & editing, Supervision.

Acknowledgements

We thank you the National Laboratory for Food Safety and Research (LANIA) at Centro de Investigación y Desarrollo A. C. (CIAD) at Culiacan for sharing *Salmonella enterica* serovars Saintpaul, Thompson, and Weltevreden strains.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2021.104771>.

References

- Abby, S.S., Néron, B., Ménager, H., Touchon, M., Rocha, E.P.C., 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* 9.
- Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., Rocha, E.P.C., 2016. Identification of protein secretion systems in bacterial genomes. *Sci. Rep.* 6, 1–14.
- Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krawland, M.G., Hale, J. L., Harbottle, H., Uesbeck, A., 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 8, e1002776.
- Anjum, M.F., Marooney, C., Fookes, M., Baker, S., Dougan, G., Ivens, A., Woodward, M. J., 2005. Identification of core and variable components of the *Salmonella enterica* sub-species I genome by microarray. *Infect. Immun.* 73 <https://doi.org/10.1128/IAI73.12.7894-7905.2005>, 7894 LP – 7905.
- Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21.
- Baloux, F., Brynildsrud, O.B., Van Dorp, L., Shaw, L.P., Chen, H., Harris, K.A., Wang, H., Eldholm, V., 2018. From theory to practice: translating whole-genome sequencing (WGS) into the clinic. *Trends Microbiol.* 26, 1035–1048.
- Bankovich, A., Nuzik, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Pjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477.
- Baucheron, S., Chaslus-Dancla, E., Cloeckaert, A., 2004. Role of *TolC* and *parC* mutation in high-level fluoroquinolone resistance in *Salmonella enterica* serotype Typhimurium DT204. *J. Antimicrob. Chemother.* 53, 657–659.
- Becerra, A., Tomás, M., Bou, G., 2013. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* 26, 185–230.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.

- Bratlie, M.S., Johansen, J., Sherman, B.T., Lempicki, R.A., Drablos, F., 2010. Gene duplications in prokaryotes can be associated with environmental adaptation. *BMC Genomics* 11, 588.
- Caratoli, A., Zankari, E., Garcia-Fernandez, A., Larsen, M.V., Lund, O., Villa, L., Aarestrup, F.M., Hasman, H., 2014. PlasmidFinder and pMLST: in silico detection and typing of plasmids. *Antimicrob. Agents Chemother.* 58 (7), 3895–3903.
- CDC (Center for Disease Control and Prevention), 2018. *Salmonella* Homepage. <http://www.cdc.gov/salmonella/index.html>.
- Chen, S.H., Byrne, R.T., Wood, E.A., Cox, M.M., 2015. E. coli rad D (yejH) gene: a novel function involved in radiation resistance and double-strand break repair. *Mol. Microbiol.* 95, 754–768.
- Coil, D., Jospin, G., Darling, A.E., 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31, 587–589.
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P.C., Vergnaud, G., Gautheret, D., Pourcel, C., 2018. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251.
- Dalhoff, A., 2012. Global fluoroquinolone resistance epidemiology and implications for clinical use. *Interdiscip. Perspect. Infect. Dis.* 2012.
- Drago-Serrano, M.E., 2009. Sistemas de adquisición de hierro en *Salmonella enterica*. *Rev. Biomed.* 20, 41–54.
- Dufresne, K., Saulnier-Bellemare, J., Daigle, F., 2018. Functional analysis of the chaperoneusher fimbrial gene clusters of *Salmonella enterica* serovar Typh. *Front. Cell. Infect. Microbiol.* 8, 26.
- Eren, A.M., Eren, Ö.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* 3, e1319.
- Erken, M., Lutz, C., McDougald, D., 2013. The rise of pathogens: predation as a factor driving the evolution of human pathogens in the environment. *Microb. Ecol.* 65, 860–868.
- Estrada-Acosta, M.D., Ramirez, K., Medrano-Felix, J.A., Castro-Del Campo, N., López-Moreno, H.S., Jiménez-Edeza, M., Martínez-Urtaza, J., Chaldez, C., 2018. Effect of river water exposition on adhesion and invasion abilities of *Salmonella Oranienburg* and *Saintpaul*. *Int. J. Environ. Health Res.* 28, 43–54.
- Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.-H., McDermott, P.F., 2019. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* 63, e00483–19.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., Milo, R., 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci.* 110, 10039–10044.
- Garai, P., Gnanadhas, D.P., Chakravorty, D., 2012. *Salmonella enterica* serovars Typhimurium and Typh. as model organisms: revealing paradigm of host-pathogen interactions. *Virulence* 3, 377–388.
- González-Escalona, N., Aguirre-Sánchez, J.R., Ibarra-Rodríguez, J.R., Chaldez-Quiróz, C., Martínez-Urtaza, J., 2019. Closed genome sequences of three *Salmonella enterica* strains belonging to serovars Saintpaul, Weltevreden, and Thompson, isolated from Mexico. *Microbiol. Resour. Annu.* 8, e00656–19.
- Gorski, L., Parker, C.T., Liang, A., Cooley, M.B., Jay-Russell, M.T., Gordus, A.G., Atwill, E.R., Mandrell, R.E., 2011. Prevalence, distribution, and diversity of *Salmonella enterica* in a major produce region of California. *Appl. Environ. Microbiol.* 77, 2734–2748.
- Green, E.R., Meccia, J., 2016. Bacterial secretion systems: an overview. *Virulence Mech. Bact. Pathog.* 213–239.
- Greene, S.K., Daly, E.R., Talbot, E.A., Demma, L.J., Holzbauer, S., Patel, N.J., Hill, T.A., Walderhaug, M.O., Hoekstra, R.M., Lynch, M.F., 2008. Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. *Epidemiol. Infect.* 136, 157–165.
- Guimarães, L.C., Florczak-Wyspińska, J., de Jesus, L.B., Viana, M.V.C., Silva, A., Ramos, R.T.J., de Soares, S.C., de Soares, S.C., 2015. Inside the Pan-genome - methods and software overview. *Curr. Genomics* 16, 245–252. <https://doi.org/10.2174/1389202916666150423002311>.
- Gupta, S.K., Sharma, P., McMillan, E.A., Jackson, C.R., Hiott, L.M., Woodley, T., Humayoun, S.B., Barrett, J.B., Frye, J.G., McClelland, M., 2019. Genomic comparison of diverse *Salmonella* serovars isolated from swine. *PLoS One* 14.
- Hanning, I.B., Nutt, J.D., Ricker, S.C., 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. *Foodborne Pathog. Dis.* 6, 635–648.
- Ho, T.D., Slauch, J.M., 2001. OmpC is the receptor for Gifx-1 and Gifx-2 bacteriophages of *Salmonella*. *J. Bacteriol.* 183, 1495–1498. <https://doi.org/10.1128/JB.183.4.1495-1498.2001>.
- Huang, Y.-H., Huang, C.-Y., 2018. SAAV2152 is a single-stranded DNA binding protein: the third SSB in *Staphylococcus aureus*. *Oncotarget* 9, 20239–20254. <https://doi.org/10.18632/oncotarget.24427>.
- Hunt, M., Mather, A.E., Sánchez-Busó, L., Page, A.J., Parkhill, J., Keane, J.A., Harris, S.R., 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genom.* 3.
- Ibarra, J.A., Steele-Mortimer, O., 2009. *Salmonella*—the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. *Cell. Microbiol.* 11, 1579–1586.
- Jackson, C.R., Dugas, S.L., 2003. Phylogenetic analysis of bacterial and archaeal *arsC* gene sequences suggests an ancient, common origin for arsenate reductase. *BMC Evol. Biol.* 3, 18.
- Jiménez, M., Martínez-Urtaza, J., Chaldez, C., 2011. Geographical and temporal dissemination of salmonellae isolated from domestic animal hosts in the Culiacan Valley, Mexico. *Microb. Ecol.* 61, 811–820.
- Jiménez, M., Martínez-Urtaza, J., Rodríguez-Alvarez, M.X., León-Félix, J., Chaldez, C., 2014. Prevalence and genetic diversity of *Salmonella* spp. in a river in a tropical environment in Mexico. *J. Water Health* 12, 874–884.
- Joerger, R.D., Hanning, I.B., Ricker, S.C., 2010. Presence of arsenic resistance in *Salmonella enterica* serovar Kentucky and other serovars isolated from poultry. *Avian Dis.* 54, 1178–1182.
- Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3.
- Kas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., Lund, O., 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One* 9.
- Kanehisa, M., Sato, Y., 2020. KEGG mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35.
- Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731.
- Kenyon, W.J., Sayers, D.G., Humphreys, S., Roberts, M., Spector, M.P., 2002. The starvation-stress response of *Salmonella enterica* serovar Typhimurium requires σ^E , but not CpxR-regulated extracytoplasmic functions. *Microbiology* 148, 113–122.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736.
- Kuramoto, N., Baba, K., Gion, K., Sugiyama, C., Taniura, H., Yoneda, Y., 2003. Xenobiotic response element binding enriched in both nuclear and microsomal fractions of rat cerebellum. *J. Neurochem.* 85, 264–273. <https://doi.org/10.1046/j.1471-4159.2003.01679.x>.
- Kutter, E., Sulakvelidze, A., 2004. Bacteriophages: Biology and Applications. CRC Press.
- Kwong, J.C., McCallum, N., Sintchenko, V., Howden, B.P., 2015. Whole genome sequencing in clinical and public health microbiology. *Pathology* 47, 199–210.
- Lapierre, P., Gogarten, J.P., 2009. Estimating the size of the bacterial pan-genome. *Trends Genet.* 25, 107–110.
- Letunic, I., Bork, P., 2007. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128.
- Levantesi, C., Bonadonna, L., Briancisco, R., Grohmann, E., Tote, S., Tundo, V., 2012. *Salmonella* in surface and drinking water: occurrence and water-mediated transmission. *Food Res. Int.* 45, 587–602.
- Liu, B., Zheng, D., Jin, Q., Chen, L., Yang, J., 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692.
- Liu, H., Whitehouse, C.A., Li, B., 2018. Presence and persistence of *Salmonella* in water: the impact on microbial quality of water and food safety. *Front. Public Health* 6, 159.
- Makendi, C., Page, A.J., Wren, B.W., Le Thi Phuong, T., Clare, S., Hale, C., Goulding, D., Klemm, E.J., Pichard, D., Okoro, C., 2016. A phylogenetic and phenotypic analysis of *Salmonella enterica* serovar Weltevreden, an emerging agent of diarrheal disease in tropical regions. *PLoS Negl. Trop. Dis.* 10, e0004446.
- Martínez-Urtaza, J., Liebana, E., García-Migura, L., Pérez-Piñero, P., Saco, M., 2004. Characterization of *Salmonella enterica* serovar Typhimurium from marine environments in coastal waters of Galicia (Spain). *Appl. Environ. Microbiol.* 70, 4030–4034.
- Matz, C., Nouri, B., McCarter, L., Martínez-Urtaza, J., 2011. Acquired type III secretion system determines environmental fitness of epidemic *Vibrio parahaemolyticus* in the interaction with bacterivorous protists. *PLoS One* 6, e20275.
- Medina-Aparicio, L., Davila, S., Rebollar-Flores, J.E., Calva, E., Hernández-Lucas, I., 2018. The CRISPR-Cas system in Enterobacteriaceae. *Pathog. Dis.* 76, fty002.
- Medrano-Felix, J.A., Castro-del Campo, N., de Peraza Garay, F.J., Martínez-Rodríguez, C., I., Chaldez, C., 2018. Carbon source utilization-based metabolic activity of *Salmonella Oranienburg* and *Salmonella Saintpaul* in river water. *Water Environ. J.* 32, 118–124.
- Morishige, Y., Koike, A., Tamura-Ueyama, A., Amano, F., 2017. Induction of viable but nonculturable *Salmonella* in exponentially grown cells by exposure to a low-humidity environment and their resuscitation by catalase. *J. Food Prot.* 80, 288–294.
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693.
- Page, A.J., Alilhan, 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. *Microb. Genom.* 3.
- Pons, B.J., Vignard, J., Mirey, G., 2019. Cytolethal distending toxin subunit b: a review of structure-function relationship. *Toxins (Basel)* 11, 595.
- Porwollik, S., McClelland, M., 2003. Lateral gene transfer in *Salmonella*. *Microbes Infect.* 5, 977–989.
- Ranjbar, R., Rahmati, H., Shokouhizadeh, L., 2018. Detection of Common Clones of *Salmonella enterica* Serotype Infantis from Human Sources in Tehran Hospitals Gastroenterology and Hepatology from Bed to Bench.
- Raudales, R.E., Parke, J.L., Guy, C.L., Fisher, P.R., 2014. Control of waterborne microbes in irrigation: a review. *Agric. Water Manag.* 143, 9–28.
- Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069.
- Selzer, P.M., 2016. Host-Pathogen Interaction: Microbial Metabolism, Pathogenicity and Antimetabolites. John Wiley & Sons.

- Shariat, N., Timme, R.E., Pettengill, J.B., Barrangou, R., Dudley, E.G., 2015. Characterization and evolution of *Salmonella* CRISPR-Cas systems. *Microbiology* 161, 374–386.
- Sherwood, K.E., Cano, D.J., Maupin-Furlow, J.A., 2009. Glycerol-mediated repression of glucose metabolism and glycerol kinase as the sole route of glycerol catabolism in the haloarchaeon *Haloferax volcanii*. *J. Bacteriol.* 191, 4307–4315.
- Siddiqui, R., Malik, H., Sagheer, M., Jung, S.-Y., Khan, N.A., 2011. The type III secretion system is involved in *Escherichia coli* K1 interactions with *Acanthamoeba*. *Exp. Parasitol.* 128, 409–413.
- Singh, B., Arya, G., Kundu, N., Sangwan, A., Nongthombam, S., Chaba, R., 2019. Molecular and functional insights into the regulation of D-galactonate metabolism by the transcriptional regulator DgoR in *Escherichia coli*. *J. Bacteriol.* 201 e00281–18.
- Smollett, K.L., Smith, K.M., Kahramanoglou, C., Amvig, K.B., Buxton, R.S., Davis, E.O., 2012. Global analysis of the regulon of the transcriptional repressor LexA, a key component of SOS response in *Mycobacterium tuberculosis*. *J. Biol. Chem.* 287, 22004–22014.
- Spector, M.P., 1998. The starvation-stress response (SSR) of *Salmonella*. In: *Advances in Microbial Physiology*. Elsevier, pp. 233–279.
- Spector, M.P., Cubitt, C.L., 1992. Starvation-inducible loci of *Salmonella typhimurium*: regulation and roles in starvation-survival. *Mol. Microbiol.* 6, 1467–1476.
- Spector, M.P., Kenyon, W.J., 2012. Resistance and survival strategies of *Salmonella enterica* to environmental stresses. *Food Res. Int.* 45, 455–481.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Steele, M., Odumeru, J., 2004. Irrigation water as source of foodborne pathogens on fruit and vegetables. *J. Food Prot.* 67, 2839–2849.
- Stevens, M.P., Humphrey, T.J., Maskell, D.J., 2009. Molecular insights into farm animal and zoonotic salmonella infections. *Philos. Trans. R. Soc. B Biol. Sci.* <https://doi.org/10.1098/rstb.2009.0094>.
- Thomsen, L.E., Olsen, J.E., Foster, J.W., Ingmer, H., 2002. ClpP is involved in the stress response and degradation of misfolded proteins in *Salmonella enterica* serovar Typhimurium. *Microbiology* 148, 2727–2733.
- Tidhar, A., Rushing, M.D., Kim, B., Slauch, J.M., 2015. Periplasmic superoxide dismutase SodCI of *Salmonella* binds peptidoglycan to remain tethered within the periplasm. *Mol. Microbiol.* 97, 832–843.
- Treangen, T.J., Ondov, B.D., Koren, S., Phillippy, A.M., 2014. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15, 524.
- Velge, P., Wiedemann, A., Roszelin, M., Abed, N., Boumart, Z., Chaussé, A.M., Grépinet, O., Namdari, F., Roche, S.M., Rossignol, A., Virlogeux-Payant, I., 2012. Multiplicity of *Salmonella* entry mechanisms, a new paradigm for *Salmonella* pathogenesis. *Microbiologyopen* 1, 243–258. <https://doi.org/10.1002/mbo3.28>.
- Wagner, C., Hensel, M., 2011. Adhesive mechanisms of *Salmonella enterica*. In: *Bacterial Adhesion*. Springer, pp. 17–34.
- Waldner, L., MacKenzie, K., Köster, W., White, A., 2012. From exit to entry: long-term survival and transmission of *Salmonella*. *Pathogens* 1, 128–155. <https://doi.org/10.3390/pathogens1020128>.
- Wells, T.J., Henderson, I.R., 2013. Type 1 and 5 secretion systems and associated toxins. In: *Escherichia Coli*. Elsevier, pp. 499–532.
- Westra, E.R., Dowling, A.J., Broniewski, J.M., van Houte, S., 2016. Evolution and ecology of CRISPR. *Annu. Rev. Ecol. Evol. Syst.* 47, 307–331.
- Winfield, M.D., Groisman, E.A., 2003. Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl. Environ. Microbiol.* 69, 3687–3694.
- Worley, M.J., Ching, K.H.L., Heffron, F., 2000. *Salmonella* SsrB activates a global regulon of horizontally acquired genes. *Mol. Microbiol.* 36, 749–761.
- Wray, C., Wray, A., 2000. *Salmonella* in Domestic Animals (Cabi).
- Yao, K., Muruvanda, T., Allard, M.W., Hoffmann, M., 2017. Complete genome sequences of three *Salmonella enterica* subsp. *enterica* serovar Saintpaul isolates associated with a 2013 multistate outbreak in the United States. *Genome Announc.* 5.
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644.

3. GENOMIC-WIDE ANALYSIS APPROACH REVEALED GENOMIC SIMILARITY FOR ENVIRONMENTAL MEXICAN *S. Oranienburg* GENOMES

J. R. Aguirre-Sanchez¹, I. F. Vega-Lopez², N. Castro del Campo¹, J. A. Medrano-Felix³, J. Martínez-Urtaza⁴, C. Chaidez-Quiroz¹

¹ Centro de Investigación en Alimentación y Desarrollo, Culiacán, México.

² Parque de Innovación Tecnológica de la Universidad Autónoma de Sinaloa, Culiacán, México.

³ Investigadoras e Investigadores por México-Centro de Investigación en Alimentación y Desarrollo A.C., Culiacán, México.

⁴ Departement of Genetics and Microbiology, Universitat Autònoma de Barcelona, Barcelona, Spain.

*Correspondence:

Dr. Cristobal Chaidez Quiroz

CIAD, Culiacán

Carretera El Dorado Km 5.5

Campo el Diez, 80110

Culiacán, Sinaloa, México

email: chaqui@ciad.mx

Artículo enviado a la revista International Journal of Environment Health

Abstract

As the human population grows, an increase in food trade is needed. This elevates the risk of contamination and epidemiological outbreaks. One of the prevalent bacteria in Mexico has been *Salmonella* Oranienburg. This work describes the genetic composition of 162 Mexican *S. Oranienburg* genomes. Bioinformatic analysis was used to explore genomic relatedness. A major clonal group formed by ST23 was identified comprising 4 STs with one allelic profile difference. Genomic similarity based on SNPs was identified among isolates with a maximum difference of 202 SNPs; in particular clades, up to 50 SNPs were detected from different isolation sources. Typical virulence genes for *Salmonella* host invasion and colonization as *RpoS*, fimbria type 1, and, T3SS were found common for all isolates, showing the presence of genetic machinery to cause diseases. This study suggests that Mexican *S. Oranienburg* strains are potential pathogens circulating continuously in the region between hosts and the environment.

Keywords: *S. Oranienburg*, genome comparison, SNP, Sequence Type (ST), epidemiology

Introduction

Since the earliest 80's, there has been a dramatic increase in international fresh produce trading by exportation or importation (Shimizu, 2022). 40 years later, developing countries rely on these economic activities as one of the major sources of richness for them (Lengai et al., 2022). This has achieved greater diversification and access to a vast product diversity which otherwise we could not have (FAO, 2020). Mexico has not been the exception to this. According to the Banco de México (2022), this country is in the 12th place as an exporter and importer with a generated sum of 465 billion dollars, equivalent to 49% of the gross national product (GNP). In addition to this, Mexican exportations represented 37.9% of the total LATAM (Latin America) exportations getting first place for this activity (Pacífico, 2009).

Nevertheless, fresh produce production and animal food trading have an important role in the spread of disease-causing organisms among people by the contaminated food intake or contact which can occur during the food production chain (primary production, slaughterhouse, processing, storage, and distribution) (Lipkin, 2013; Savelli et al., 2019). The World Health Organization (WHO) estimated that 600 million people (1 in 10 people in the world) get sick from the consumption of contaminated food causing an economic burden of 110 billion dollars per year associated with productivity and medical expenses. For its part, the European Food Safety Authority (EFSA) and The Centers for Disease Control and Prevention (CDC), reported *Salmonella* as one of the most frequent agents in food outbreaks.

Multiple recalls have been made by the Food and Drug Administration (FDA) related to Mexican products, associating with *S. enterica* serovar Oranienburg. For example, in 2021 an outbreak associated with onions from the State of Chihuahua caused 1,040 illnesses and 260 hospitalizations in the District of Columbia and Puerto Rico (CDC, 2022). In addition to this, *S. Oranienburg* has

been reported as one of the most common isolated serotypes in water, soil, and packing surfaces in Mexico (Contreras-Soto et al., 2018). Also has been responsible for the outbreak of 150 inmates in a Mexican prison (Vázquez-Garcidueñas et al., 2014), prevalent isolated from captive animal feces (Silva-Hidalgo et al., 2012) and found in an integrative food chain surveillance system (Zaidi et al., 2008).

Due to the epidemiological importance for a fast and accurate etiological agent identification, a broad array of techniques have been used for food-borne outbreaks investigation (Sandt et al., 2006). Pulsed-field gel electrophoresis (PFGE) was considered as the gold standard for bacteria identification by the generation of DNA fingerprints (Neoh et al., 2019). However, technique limitations such as time consuming, low resolution for genetic relatedness strains, and some strains cannot be typed by this technique (Hedberg et al., 2001). These issues have resulted in the migration toward more sensitive techniques as whole genome sequencing.

With the big break of modern sequencing technologies and the optimization of computational algorithms, the analysis of bacterial genomes represents a big chance to introduce higher level of resolution to identify etiological agents from a new perspective in this open and trading world. The Single Nucleotide Polymorphism (SNP) approach has been widely used for outbreak identification and surveillance purposes (Coipan et al., 2020; Connor et al., 2015; Gymoese et al., 2017). The exploration of genome information could lead to the creation of genomic catalogs to know bacterial population structure and their main characteristics, resulting in valuable information to better understand bacteria biology and establish containment measures to prevent their spread (Connor et al., 2015).

In this regard, the aim of this work is to explore and describe the entire genetic atlas of the *S. Oranienburg* in Mexican populations due to the national and international importance of this

Salmonella serovar in food and fresh producers related infections. These approaches can enhance the management of the infections by improving the knowledge of this bacteria and also can have a significative role as food defense evidence. Furthermore, by understanding the genetic relationships among *S. Oranienburg* populations from the environment will enable us, to detect and respond to the emergence of strains with pathogenic potential.

Methods

DNA extraction and genome sequencing

A total of 53 *Salmonella enterica* serovar Oranienburg strains were isolated from 3 different environmental sources (23 from artificial waterways, 16 from river water, and 14 from asymptomatic domestic animals). The strains were obtained from the collection of the National Food Safety Laboratory (LANIIA) at the Centro de Investigación en Alimentación y Desarrollo (CIAD), Culiacan, Mexico. DNA extraction was performed using the DNeasy Blood & Tissue Kit (QIAGEN, Mexico City, Mexico) according to the manufacturer's instructions. The concentration of the extracted DNA was determined using the Qubit dsDNA Broad Range Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). For sequencing, genomic DNA from each *S. Oranienburg* strain was adjusted to a 0.2 ng/μL concentration and aliquoted at a final amount of 1 ng for preparing genomic DNA libraries with the Nextera XT DNA Library Preparation Kit (Illumina Inc., San Diego, CA, USA). Subsequently, the prepared genomic libraries were sequenced using a MiSeq™ Reagent Kit v2 (300-cycle format) to obtain a 2×150 bp paired-end read output with a MiSeq™ System (Illumina, Inc.) at the Earlham Institute (Norwich Research Park, Norwich, United Kingdom). In addition, 109 Mexican environmental genomes belonging to *S. Oranienburg*

were downloaded (Supplementary Table 1) in FASTA format from the National Center for Biotechnology Information (NCBI).

Quality control evaluation and reads assembly

Reads initial quality was evaluated using FASTQC (Andrews, 2017). Next, the first 20 bp were removed from each read and a 4-base wide sliding window was used to cut when the average Phred quality score per base was below 15 using Trimmomatic v0.32 (Bolger et al., 2014). Reads less than 50 bp were removed from the dataset. Draft genomes were assembled *de novo* with the pipeline A5-miseq v20160825 (Coil et al., 2015). Resulted assemblies were selected according to genome size, N50, and scaffolds number.

Taxonomy and ST assignation

Salmonella in silico typing resource (SISTR) stand-alone version (Yoshida et al., 2016) was used to corroborate the taxonomy identity in the previously assembled and downloaded *S. Oranienburg* genomes. MLST program (Jolley et al., 2018) was used to scan the assemblies' files against PubMLST *Salmonella* typing schemes (Achtman et al., 2012). Only genomes belonging to *S. Oranienburg* were included in the following analysis. A spanning tree was created to show ST diversity and locus variants using PHYLOViZ online version (Francisco et al., 2012).

Core SNP phylogenetic tree construction

A SNP core genome alignment from 165 *S. Oranienburg* assemblies was performed using Snippy (Torsten Seemann, 2018a) with the option snippy-multi to simplify the running of isolates against the closed *S. Oranienburg* strain CFSAN076211 (GCA_003710145.1). Ambiguous characters were removed from the alignment and replaced with N using the snippy-clean_full_aln. A maximum-likelihood tree was constructed using the resulting core alignment considering 1000 bootstrap

replicates with IQ-TREE (Trifinopoulos et al., 2016) with the automatic model selection option. Tree file was visualized and edited with the Interactive Tree Of Life (iTOL) (Letunic & Bork, 2019).

Genomic analysis

To explore genomic similarity in terms of SNPs among all *S. Oranienburg* isolates an SNP matrix was constructed using the program snp-dists 0.8.2 (Torsten Seemann, 2018b) converting the multi-FASTA core alignment to a numerical matrix. The versatile matrix visualization and analysis software MORPHEUS (Morpheus, n.d.) was used to create a heatmap. In addition, a cladogram and a Pearson correlation were performed to explore the relationship strength between the isolates. Virulence genes were explored by ABRicate v 1.0.1 (T Seemann, 2018) considering the Virulence Factor Database (VFDB) with a coverage of >90% and an identity >95%.

Results

MLST *in silico* analysis revealed 9 different ST profiles for the 165 Mexican *S. Oranienburg* genomes (Figure 1). The most prevalent group was ST23 (86%, n=142), serving as founder for the other eight STs. In addition, ST2683, ST2654, ST3111 and ST2751 were found as single locus variants (SLV) regard to ST23. For example, ST2654 only differs in *aroC* allele, ST2683, and ST3111 in *hemD* allele, and finally ST2751 in *hisD* allele. On the other hand, four STs groups (ST50, ST 1510, ST174, ST32) with 7 alleles differences (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) regard to founder ST were found.

Genomic similarity among *S. Oranienburg* genomes was tested by a maximum likelihood phylogenetic tree construction (Figure 2). Seven clades designated from A to G were found and colored according to observed aggrupation. All isolation sources were classified into 3 general

categories: food, water, and animal. Nevertheless, no specific isolation source clade was found, on the contrary, all 53 genomes from this study were grouped based on genetic similarity in different clades. A high genetic relatedness between the public repositories genomes and those obtained in this research was observed despite isolation source. For example, clade D (Figure 2) groups local isolates from river and animal sources that showed close relationship with previous reported genomes from the 3 general categories. In addition to this, clade A and B were the most abundant clades respecting the different isolation sources and the number of genomes included on each clade. Genomes from this study and the previously reported genomes were found homogeneously among clades suggesting a close phylogenetic relationship despite collection date or isolation source.

To quantify differences in core genome in terms of SNPs a SNP distance matrix was constructed (Figure 3). The distance matrix revealed a variation ranging from 1 as a minimum to 202 SNPs as a maximum among isolates. Additionally, a Pearson correlation was performed in order to show the measure of linear correlation strength among isolates (Figure 4). Genomes included in each of the high similarity blocks were found as close independent clades. For instance, first red block isolates form part of the clade D (Figure 2). In the other hand, blue blocks are forming an independent clade but the main difference between red blocks is the length of the branches that represent genomic relatedness among isolates. Under this context, clade A and B are formed (Figure 2).

Regarding the virulence profile, an average of 146 virulence genes were found as a conservative collection for all strains (Supplementary Figure 1). Major regulator genes *Fur* and *RpoS* for ferric uptake and the production of fimbriae were found. The *fim* cluster *fimA*, *I*, *C*, *D*, *H*, *F*, *Z*, *Y*, and *W*, and the single operon were found for all isolates. The curli fimbriae operon *csgBAC* and the master regulatory gene for adhesive curli fimbriae expression *csgD* were consistent for all strains.

Virulence cassette *ssaA* belonging to *Salmonella* pathogenicity island (SPI-2) was found. Its main role is the secretion of effector proteins to facilitate the replication of intracellular bacteria. The genes *sopAD*, *AvrA*, *SipABC* related to the TTSS-1 secreted effectors were constant for the genome collection. The magnesium uptake genes *MgtBC* necessary for intracellular survival and virulence were also found. The gene duo *entAB* associated with the iron acquisition was found. Interestingly, in the sporadic episodes of diarrhea in humans and animals the EAST1 toxin was found for only a strain that was isolated from serrano pepper (SORA2008JA02). In addition to this, the precursor of an enterobactin pathway 2,3-dihydrobenzoic acid was found for a single strain isolated from river water (C-ORA09001). The only not consistent gene among all isolates was *ratB* which has an intestinal persistence role.

Discussion

Even though *S. Oranienburg* is one of the top 10 serovars reported in Mexico of the most prevalent pathogens *S. serovars*, it is responsible for worldwide outbreaks and food recalls in developed/developing countries such as Mexico. In addition to this, *S. Oranienburg* is included on the list of most frequently isolated *Salmonella* serovars in humans and the agri-food sector over a period of 10 years (Cherchame, Ilango, et al., 2022). Despite the epidemiological importance of *S. Oranienburg*, a limited number of genomic studies have been performed. This work represents an effort in the elucidation of genomic diversity and genomic structure by genome-wide analysis approach for this serovar.

In the analyzed samples, the most common isolation source for this serovar was the water (river and canal), which coincides with the data reported by U. S. Department of Agriculture (USDA)

which found a high prevalence of *S. Oranienburg* on surface water in the agricultural region of the central coastal regions in California during a five-year survey (Lisa et al., 2022).

Genomic comparisons suggest that Mexican *S. Oranienburg* population display low genetic diversity, based on the limited number of STs, SNPs difference and virulence profile found among genomes even when strains were isolated from different sources and years. These compared with other studies. For example, a high difference of 870 SNPs were observed among 10 *S. enterica* genomes (Rahman et al., 2022b). In addition to this, genetic studies focused on four *S. Montevideo* lineages have shown differences of up to 17,600 SNPs among clades (Rahman et al., 2022b). For its part, Rahman et al., reported strongly geographically clustered populations of *S. Typhi* from 49 clades, he suggests that subclades can be identified from 68 SNPs thorough a genotyping approach (Rahman et al., 2022b).

Despite the low number of studies worldwide focused on *S. Oranienburg*, previous works had observed low homogeneity genomic patterns for this serovar on isolated captive reptiles in the Czech Republic (Tomastikova et al., 2017). Although this research does not conclude the directionality of *S. Oranienburg* transmission, the genomic similarity exhibited and the indifferent grouping to the isolation site in the phylogenetic tree suggest that *S. Oranienburg* strains are in constant circulation among animals and transmission to food by the fecal-oral via. These findings coincide with Toro et al. (2016) and Bloomfield et al (2017) studies focused on different serovars that suggest *S. enterica* is being transmitted bi-directionally (Bloomfield et al., 2017; Toro et al., 2016).

S. Oranienburg has been previously identified from environmental, food, and animal isolation sources (Contreras-Soto et al., 2018) in Mexico. Nevertheless, no genomic studies concerning genomic structure were performed. Under this context, ST23 was shown by MLST as the

predominant ST in Mexico with the presence of 86% from all strains, displaying high clonal similarity among all strains assigned to the same ST group. Moreover, *S. Oranienburg* population is integrated by a unique clonal complex formed by 5 STs (ST23, ST2654, ST2683, ST2751, and ST3111) sharing six of the seven identical alleles (Enright et al., 2002).

This supports the idea that *S. Oranienburg* population structure is monophyletic for sharing a most common ancestor (MRCA) among isolates. Nevertheless, global *S. Oranienburg* population should be considered. This finding differs from other studies where *S. Oranienburg* was found scattered across different lineages (Cherchame, Ilango, et al., 2022). Nevertheless, STs found in the previous work as part of lineage one (ST23 and ST174), coincide with the STs found in this study forming part of the same Mexican lineage, maybe this could explain the close aggrupation among Mexican isolates.

Owing to the genetic diversity nature of the *Salmonella* genus the use of genomic analysis plays an important role in their differentiation. The identification of core SNPs has been one of the main criteria for determining relatedness among isolates (Q. Wang et al., 2015). This strategy is today a key element in outbreaks investigation and traceback procedures (Le & Diep, 2013). Several works rely on SNPs as a crucial criterion in the establishment of presumptive or responsible outbreaks strains (Cherchame, Guillier, et al., 2022; den Bakker et al., 2011; Rahman et al., 2022a) For example, 250 confirmed polymorphisms in *S. enteritidis* have been used to differentiate subpopulations (Guard et al., 2011). Moreover, SNPs differentiation has generated more epidemiological information than even gold-standard techniques such as pulsed-field gel electrophoresis (Scaltriti et al., 2015).

In our study, a maximum total of 202 SNPs were found for the strains. The small difference demonstrates high genetic similarity among isolates despite isolation source and even isolation

year. The evidence suggested that *S. Oranienburg* has been in circulation in Mexico through a recurrent environmental and host exchange, for example, Clade D integrated by strains isolated from water and animals were quite similar with no more than 20 SNPs. This means that clonal descendent of the same strain were persistent among environmental and animal reservoir in the region of Sinaloa over a period of two years. Previous information can help us to investigate outbreaks in clonally related strains (differing by <100 SNPs). Against this background, the SNP number, phylogenetic inference, and isolation source of outbreaks producers' strains could be potentially used to put one's finger on outbreaks' responsible strains.

The genotype of *S. Oranienburg* population presents a conserved virulence profile. Genes involved in Type III secretion system (T3SS) assembly, magnesium uptake, adhesion, and stress response were consistently found for all genomes. This category has been observed in well-known pathogenic *Salmonella* serovar as Typhi (Al-Khafaji et al., 2021; Kaur & Jain, 2012) and Typhimurium (Dos Santos et al., 2019; Johnson et al., 2018) as an essential gene repertory to invade, survive, reproduce and cause a gastrointestinal clinical profile (Jajere, 2019). For example, the sigma factor RpoS, has been described as in control of the adaptative response to environmental stress, virulence, and biofilm formation (Lago et al., 2017). The presence of type 1 fimbria plays an important role in the initial attachment to host intestinal mucosa causing persistence and pathogenesis (Kolenda et al., 2019). The punctual identification of previously mentioned virulence markers suggests that this serotype has the potential to infect a host and trigger a pathogenesis process in it.

Conclusion

The genomic structure of the genome collection of *S. Oranienburg* from Mexico exhibited low genomic diversity regarding STs, SNPs, and virulence profile. ST23 was found as prevalent forming a unique clonal complex with ST2654, ST2683, 2751, and ST3111, differing in one allele according to the MLST profile. Phylogenetic inference revealed an independent aggragation pattern respecting strain isolation source, which could be represented by a constant introduction and circulation of this serovar from water, asymptomatic domestic animals, and food. SNP analysis revealed genetic similarity among strains. A maximum difference of 204 SNPs was detected. Nevertheless, specific clades with less than 50 SNPs were found. The phylogenetic core genome could have potential use for epidemiological investigations due to the identification of emergency pathogens that group or share few SNPs. Moreover, all strains have pathogenic potential according to the punctual identification of virulence markers related to cell invasion, effector proteins, curli fibers, siderophores, and macrophage survival, suggesting that this serotype has the potential to infect, and host trigger a pathogenesis process. We suggest taking this information into account for the outbreak's future identification.

Acknowledgment

This genomic material was provided by The National Laboratory for Food Safety and Research (LANIIA) at Centro de Investigación y Desarrollo A.C (CIAD) at Culiacán, Sinaloa, México.

Disclosure statement

No potential conflict of interest was reported by the authors

Authors contribution

Formal analysis and investigation: José Roberto Aguirre-Sánchez and Inés Fernando Vega-López.

Writing – original draft: José Roberto Aguirre-Sánchez, José Andrés Medrano-Félix and Jaime Martínez-Urtaza. Writing – review and editing: Nohelia Castro del Campo. Resources: Jaime Martínez-Urtaza and Cristóbal Chaidez-Quiroz. Supervision: Cristobal Chaidez-Quiroz.

References

- Aanensen, D. M., & Spratt, B. G. (2005). The multilocus sequence typing network: mlst.net. *Nucleic Acids Research*, 33(suppl_2), W728–W733. <https://doi.org/10.1093/nar/gki415>
- Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krauland, M. G., Hale, J. L., Harbottle, H., & Uesbeck, A. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog*, 8(6), e1002776.
- Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*, 37(2), 125–132.
- Al-Khafaji, N. S. K., Al-Bayati, A. M. K., & Al-Dahmoshi, H. O. M. (2021). Virulence Factors of *Salmonella Typhi*. In *Salmonella spp.-A Global Challenge*. IntechOpen.
- Alvarado, O., & Isac, W. (2022). *Caracterización del agente causal de Salmonellosis aviar” mediante revisión bibliográfica*. BABAHOYO: UTB, 2022.
- Andino, A., & Hanning, I. (2015). *Salmonella enterica*: survival, colonization, and virulence differences among serovars. *The Scientific World Journal*, 2015.
- Andrews, S. (2017). *FastQC: a quality control tool for high throughput sequence data*. 2010.
- Anjum, M. F., Marooney, C., Fookes, M., Baker, S., Dougan, G., Ivens, A., & Woodward, M. J. (2005). Identification of Core and Variable Components of the Salmonella enterica Subspecies I Genome by Microarray. *Infection and Immunity*, 73(12), 7894 LP – 7905. <https://doi.org/10.1128/IAI.73.12.7894-7905.2005>

- Antony, L., Behr, M., Sockett, D., Miskimins, D., Aulik, N., Christopher-Hennings, J., Nelson, E., Allard, M. W., & Scaria, J. (2018). Genome divergence and increased virulence of outbreak associated *Salmonella enterica* subspecies *enterica* serovar Heidelberg. *Gut Pathogens*, 10(1), 1–9.
- Baker, S., & Dougan, G. (2007). The genome of *Salmonella enterica* serovar Typhi. *Clinical Infectious Diseases*, 45(Supplement_1), S29–S33.
- Baudart, J., Lemarchand, K., Brisabois, A., & Lebaron, P. (2000). Diversity of *Salmonella* strains isolated from the aquatic environment as determined by serotyping and amplification of the ribosomal DNA spacer regions. *Applied and Environmental Microbiology*, 66(4), 1544–1552.
- Baum, D. A., & Smith, S. D. (2013). Tree thinking. *An Introduction to Phylogenetic Biology*. Roberts and Company Publishers.
- 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.
- Blair, J., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13(1), 42–51.
- Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P., & Allaoui, A. (2001). Structure and composition of the *Shigella flexneri* ‘needle complex’, a part of its type III secreton. *Molecular Microbiology*, 39(3), 652–663.
- Bloomfield, S. J., Benschop, J., Biggs, P. J., Marshall, J. C., Hayman, D. T. S., Carter, P. E., Midwinter, A. C., Mather, A. E., & French, N. P. (2017). Genomic analysis of *Salmonella enterica* serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand, 1998–2012. *Emerging Infectious Diseases*, 23(6), 906.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., & Huang, X. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology*, 26(10),

1146–1153.

- Brocchieri, L. (2001). Phylogenetic inferences from molecular sequences: review and critique. *Theoretical Population Biology*, 59(1), 27–40.
- Caballero, E. J. L., & Suárez, G. P. (1999). Métodos de análisis en la Reconstrucción Filogenética. *Bol. SEA*, 26, 45–56.
- Cabañas-Romero, P. V., & Huerta-Saquero, A. (2014). Nanomáquinas biológicas: los sistemas de secreción bacterianos. *Mundo Nano. Revista Interdisciplinaria En Nanociencias y Nanotecnología*, 7(13), 28–36.
- Carattoli, A. (2003). Plasmid-mediated antimicrobial resistance in *Salmonella enterica*. *Current Issues in Molecular Biology*, 5(4), 113–122.
- CDC. (2022). *Salmonella Outbreak Linked to Onions*. <https://www.cdc.gov/salmonella/oranienburg-09-21/index.html#:~:text=Multiple companies recalled onions,information about each company's recall>.
- Chao, W. L., Ding, R. J., & Chen, R. S. (1987). Survival of pathogenic bacteria in environmental microcosms. *Zhonghua Minguo Wei Sheng Wu Ji Mian Yi Xue Za Zhi= Chinese Journal of Microbiology and Immunology*, 20(4), 339–348.
- Chen, L., Xiong, Z., Sun, L., Yang, J., & Jin, Q. (2012). VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research*, 40(D1), D641–D645.
- Cherchame, E., Guillier, L., Lailler, R., Vignaud, M.-L., Silva, J.-D., Le Hello, S., Weill, F.-X., & Cadel-Six, S. (2022). *Salmonella enterica* subsp. *enterica* Welikade: guideline for phylogenetic analysis of serovars rarely involved in foodborne outbreaks. *BMC Genomics*, 23(1), 1–13.
- Cherchame, E., Ilango, G., Noël, V., & Cadel Six, S. (2022). Polyphyly in widespread *Salmonella enterica* serovars and using genomic proximity to choose the best reference genome for bioinformatics analyses. *Frontiers in Public Health*, 2820.
- Coil, D., Jospin, G., & Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, 31(4), 587–589.

- Coipan, C. E., Dallman, T. J., Brown, D., Hartman, H., van der Voort, M., van den Berg, R. R., Palm, D., Kotila, S., van Wijk, T., & Franz, E. (2020). Concordance of SNP-and allele-based typing workflows in the context of a large-scale international *Salmonella* Enteritidis outbreak investigation. *Microbial Genomics*, 6(3).
- Connor, T. R., Barker, C. R., Baker, K. S., Weill, F.-X., Talukder, K. A., Smith, A. M., Baker, S., Gouali, M., Thanh, D. P., & Azmi, I. J. (2015). Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *Elife*, 4, e07335.
- Contreras-Soto, M. B., Medrano-Félix, J. A., Chaidez, C., Ibarra-Rodríguez, J. R., Martínez-Urtaza, J., & Castro-del Campo, N. (2018). Los últimos 50 años de *Salmonella* en México: Fuentes de aislamiento y factores que influyen en su prevalencia y diversidad. *Revista Bio Ciencias*, 6, 26.
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518–524.
- Delmont, T. O., & Eren, A. M. (2018). Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*, 6, e4320.
- den Bakker, H. C., Moreno Switt, A. I., Cummings, C. A., Hoelzer, K., Degoricija, L., Rodriguez-Rivera, L. D., Wright, E. M., Fang, R., Davis, M., & Root, T. (2011). A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type. *Applied and Environmental Microbiology*, 77(24), 8648–8655.
- Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, 13(9), 601–612.
- Domingues, S., da Silva, G. J., & Nielsen, K. M. (2012). Integrons: vehicles and pathways for horizontal dissemination in bacteria. *Mobile Genetic Elements*, 2(5), 211–223.
- Dos Santos, A. M. P., Ferrari, R. G., & Conte-Junior, C. A. (2019). Virulence factors in *Salmonella* Typhimurium: the sagacity of a bacterium. *Current Microbiology*, 76(6), 762–773.
- Duncan, M. C., Linington, R. G., & Auerbuch, V. (2012). Chemical inhibitors of the type three secretion system:

- disarming bacterial pathogens. *Antimicrobial Agents and Chemotherapy*, 56(11), 5433–5441.
- Eklblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7(9), 1026–1042.
- Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G., & Lee, L.-H. (2015). Salmonella: a review on pathogenesis, epidemiology and antibiotic resistance. *Frontiers in Life Science*, 8(3), 284–293.
- Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., & Spratt, B. G. (2002). The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proceedings of the National Academy of Sciences*, 99(11), 7687–7692.
- FAO. (2020). *Agricultural production statistics 2000–2020*. Agricultural Production Statistics 2000–2020. <https://www.fao.org/3/cb9180en/cb9180en.pdf>
- Fatica, M. K., & Schneider, K. R. (2011). Salmonella and produce: survival in the plant environment and implications in food safety. *Virulence*, 2(6), 573–579.
- Ferrari, R. G., Panzenhagen, P. H. N., & Conte-Junior, C. A. (2017). Phenotypic and genotypic eligible methods for Salmonella Typhimurium source tracking. *Frontiers in Microbiology*, 8, 2587.
- Figueira, R., & Holden, D. W. (2012). Functions of the Salmonella pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiology*, 158(5), 1147–1161.
- Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., & Carriço, J. A. (2012). PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, 13(1), 1–10.
- Galán, J. E. (2001). Salmonella interactions with host cells: type III secretion at work. *Annual Review of Cell and Developmental Biology*, 17(1), 53–86.
- Gerlach, R. G., Jäckel, D., Stecher, B., Wagner, C., Lupas, A., Hardt, W., & Hensel, M. (2007). Salmonella Pathogenicity Island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. *Cellular Microbiology*, 9(7), 1834–1850.

- González-Pedrajo, B., & Dreyfus, G. (2003). Sistemas de secreción de proteínas en las bacterias Gram negativas: Biogénesis flagelar y translocación de factores de virulencia. *Mensaje Bioquímico*, 27, 45–63.
- Gonzalez Pedraza, J., Pereira Sanandres, N., Soto Varela, Z., Hernández Aguirre, E., & Villarreal Camacho, J. (2014). Aislamiento microbiológico de *Salmonella* spp. y herramientas moleculares para su detección. *Revista Salud Uninorte*, 30(1), 73–94.
- Grimont, P. A. D., & Weill, F.-X. (2007). Antigenic formulae of the *Salmonella* serovars. *WHO Collaborating Centre for Reference and Research on Salmonella*, 9, 1–166.
- Guard, J., Morales, C. A., Fedorka-Cray, P., & Gast, R. K. (2011). Single nucleotide polymorphisms that differentiate two subpopulations of *Salmonella enteritidis* within phage type. *BMC Research Notes*, 4(1), 1–15.
- Gymoese, P., Sørensen, G., Litrup, E., Olsen, J. E., Nielsen, E. M., & Torpdahl, M. (2017). Investigation of outbreaks of *Salmonella enterica* serovar Typhimurium and its monophasic variants using whole-genome sequencing, Denmark. *Emerging Infectious Diseases*, 23(10), 1631.
- Hacker, J., & Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Reports*, 2(5), 376–381.
- Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., & Lindsay, J. A. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science*, 327(5964), 469–474.
- Hedberg, C. W., Smith, K. E., Besser, J. M., Boxrud, D. J., Hennessy, T. W., Bender, J. B., Anderson, F. A., & Osterholm, M. T. (2001). Limitations of pulsed-field gel electrophoresis for the routine surveillance of *Campylobacter* infections. *The Journal of Infectious Diseases*, 184(2), 242–243.
- Hensel, M. (2004). Evolution of pathogenicity islands of *Salmonella enterica*. *International Journal of Medical Microbiology*, 294(2–3), 95–102.
- Herschleb, J., Ananiev, G., & Schwartz, D. C. (2007). Pulsed-field gel electrophoresis. *Nature Protocols*, 2(3), 677–684.
- Ho, S. (2008). The molecular clock and estimating species divergence. *Nature Education*, 1(1), 168.

- Højby, N., Bjarnsholt, T., Givskov, M., Molin, S., & Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *International Journal of Antimicrobial Agents*, 35(4), 322–332.
- Homøe, P., Bjarnsholt, T., Wessman, M., Sørensen, H. C. F., & Johansen, H. K. (2009). Morphological evidence of biofilm formation in Greenlanders with chronic suppurative otitis media. *European Archives of Oto-Rhino-Laryngology*, 266(10), 1533–1538.
- Jacobsen, C. S., & Bech, T. B. (2012). Soil survival of *Salmonella* and transfer to freshwater and fresh produce. *Food Research International*, 45(2), 557–566.
- Jajere, S. M. (2019). A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Veterinary World*, 12(4), 504.
- Johnson, R., Mylona, E., & Frankel, G. (2018). Typhoidal *Salmonella*: Distinctive virulence factors and pathogenesis. *Cellular Microbiology*, 20(9), e12939.
- Jolley, K. A., Bray, J. E., & Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Research*, 3.
- Jones, B. D. (1997). Host responses to pathogenic *Salmonella* infection. *Genes & Development*, 11(6), 679–687.
- Kaur, J., & Jain, S. K. (2012). Role of antigens and virulence factors of *Salmonella enterica* serovar Typhi in its pathogenesis. *Microbiological Research*, 167(4), 199–210.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624–626.
- Kolenda, R., Ugorski, M., & Grzymajlo, K. (2019). Everything you always wanted to know about *Salmonella* type 1 fimbriae, but were afraid to ask. *Frontiers in Microbiology*, 10, 1017.
- Koonin, E., & Galperin, M. Y. (2002). *Sequence—evolution—function: computational approaches in comparative genomics*.
- Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, 6(8), 654–662.
- Lago, M., Monteil, V., Douche, T., Guglielmini, J., Criscuolo, A., Maufrais, C., Matondo, M., & Norel, F. (2017).

- Proteome remodelling by the stress sigma factor RpoS/ σ S in Salmonella: identification of small proteins and evidence for post-transcriptional regulation. *Scientific Reports*, 7(1), 1–15.
- Laing, C. R., Whiteside, M. D., & Gannon, V. P. J. (2017). Pan-genome analyses of the species *Salmonella enterica*, and identification of genomic markers predictive for species, subspecies, and serovar. *Frontiers in Microbiology*, 8, 1345.
- Le Minor, L., & Popoff, M. Y. (1987). Designation of *Salmonella enterica* sp. nov., nom. rev., as the Type and Only Species of the Genus *Salmonella*: Request for an Opinion. *International Journal of Systematic and Evolutionary Microbiology*, 37(4), 465–468.
- Le, V. T. M., & Diep, B. A. (2013). Selected insights from application of whole genome sequencing for outbreak investigations. *Current Opinion in Critical Care*, 19(5), 432.
- Lee, P.-C., Zmina, S. E., Stopford, C. M., Toska, J., & Rietsch, A. (2014). Control of type III secretion activity and substrate specificity by the cytoplasmic regulator PcrG. *Proceedings of the National Academy of Sciences*, 111(19), E2027–E2036.
- Lemey, P., Salemi, M., & Vandamme, A.-M. (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press.
- Lengai, G. M. W., Fulano, A. M., & Muthomi, J. W. (2022). Improving Access to Export Market for Fresh Vegetables through Reduction of Phytosanitary and Pesticide Residue Constraints. *Sustainability*, 14(13), 8183.
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Lipkin, W. I. (2013). The changing face of pathogen discovery and surveillance. *Nature Reviews Microbiology*, 11(2), 133–141.
- Lisa, G., Samarpita, W., Diana, C., & Ashley, A. N. (2022). *Salmonella enterica* serovar diversity, distribution, and prevalence in public access waters from a central California coastal leafy green growing region during 2011–2016.

- Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Research*, 47(D1), D687–D692.
- Liu, H., Whitehouse, C. A., & Li, B. (2018). Presence and persistence of Salmonella in water: the impact on microbial quality of water and food safety. *Frontiers in Public Health*, 6, 159.
- Loman, N. J., & Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nature Reviews Microbiology*, 13(12), 787–794.
- Long, S. G., DuPont, H. L., Gaul, L., Arafat, R. R., Selwyn, B. J., Rogers, J., & Casey, E. (2010). Pulsed-field gel electrophoresis for Salmonella infection surveillance, Texas, USA, 2007. *Emerging Infectious Diseases*, 16(6), 983.
- Lostroh, C. P., & Lee, C. A. (2001). The Salmonella pathogenicity island-1 type III secretion system. *Microbes and Infection*, 3(14–15), 1281–1291.
- Marcus, S. L., Brumell, J. H., Pfeifer, C. G., & Finlay, B. B. (2000). Salmonella pathogenicity islands: big virulence in small packages. *Microbes and Infection*, 2(2), 145–156.
- Martinez, J. L., & Baquero, F. (2000). Mutation frequencies and antibiotic resistance. *Antimicrobial Agents and Chemotherapy*, 44(7), 1771–1777.
- Mazurkiewicz, P., Thomas, J., Thompson, J. A., Liu, M., Arbibe, L., Sansonetti, P., & Holden, D. W. (2008). SpvC is a Salmonella effector with phosphothreonine lyase activity on host mitogen-activated protein kinases. *Molecular Microbiology*, 67(6), 1371–1383.
- McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., Ayers, S. L., Lam, C., Tate, H. P., & Zhao, S. (2016). Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal Salmonella. *Antimicrobial Agents and Chemotherapy*, 60(9), 5515–5520.
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31–46.
- Morpheus. (n.d.). *MORPHEUS: Versatile matrix visualization and analysis software*.
<https://software.broadinstitute.org/morpheus/>

- Mouslim, C., Hilbert, F., Huang, H., & Groisman, E. A. (2002). Conflicting needs for a *Salmonella* hypervirulence gene in host and non-host environments. *Molecular Microbiology*, 45(4), 1019–1027.
- Neoh, H., Tan, X.-E., Sapri, H. F., & Tan, T. L. (2019). Pulsed-field gel electrophoresis (PFGE): A review of the “gold standard” for bacteria typing and current alternatives. *Infection, Genetics and Evolution*, 74, 103935.
- Ochoa, I. M. F., & Rodríguez, A. V. (2005). Mecanismos moleculares de patogenicidad de *Salmonella* sp. *Revista Latinoamericana de Microbiología*, 47(1–2), 25–42.
- Oladeinde, A., Cook, K., Lakin, S. M., Woyda, R., Abdo, Z., Looft, T., Herrington, K., Zock, G., Lawrence, J. P., & Thomas IV, J. C. (2019). Horizontal gene transfer and acquired antibiotic resistance in *Salmonella enterica* serovar Heidelberg following in vitro incubation in broiler ceca. *Applied and Environmental Microbiology*, 85(22), e01903-19.
- Pacífico, A. del. (2009). *Alianza del Pacífico*.
- 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).
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22), 3691–3693.
- Palmer, A. D., & Slauch, J. M. (2017). Mechanisms of *Salmonella* pathogenesis in animal models. *Human and Ecological Risk Assessment: An International Journal*, 23(8), 1877–1892.
- Pfeifer, S. P. (2017). From next-generation resequencing reads to a high-quality variant data set. *Heredity*, 118(2), 111–124.
- Prevention, C. for D. C. and. (2015). *Multistate outbreak of Salmonella Poona infections linked to imported cucumbers*.
- Rabsch, W. (2007). *Salmonella typhimurium* phage typing for pathogens. In *Salmonella* (pp. 177–211). Springer.
- Rahman, M.-M., Lim, S.-J., & Park, Y.-C. (2022a). Development of Single Nucleotide Polymorphism (SNP)-Based

- Triplex PCR Marker for Serotype-specific *Escherichia coli* Detection. *Pathogens*, 11(2), 115.
- Rahman, M.-M., Lim, S.-J., & Park, Y.-C. (2022b). Genome-Wide Searching Single Nucleotide-Polymorphisms (SNPs) and SNPs-Targeting a Multiplex Primer for Identification of Common *Salmonella* Serotypes. *Pathogens*, 11(10), 1075.
- Ryan, J. (2016). *Biosecurity and bioterrorism: containing and preventing biological threats*. Butterworth-Heinemann.
- Sandt, C. H., Krouse, D. A., Cook, C. R., Hackman, A. L., Chmielecki, W. A., & Warren, N. G. (2006). The key role of pulsed-field gel electrophoresis in investigation of a large multiserotype and multistate food-borne outbreak of *Salmonella* infections centered in Pennsylvania. *Journal of Clinical Microbiology*, 44(9), 3208–3212.
- Savelli, C. J., Bradshaw, A., Ben Embarek, P., & Mateus, C. (2019). The FAO/WHO international food safety authorities network in review, 2004–2018: learning from the past and looking to the future. *Foodborne Pathogens and Disease*, 16(7), 480–488.
- Scaltriti, E., Sassera, D., Comandatore, F., Morganti, M., Mandalari, C., Gaiarsa, S., Bandi, C., Zehender, G., Bolzoni, L., & Casadei, G. (2015). Differential single nucleotide polymorphism-based analysis of an outbreak caused by *Salmonella enterica* serovar Manhattan reveals epidemiological details missed by standard pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*, 53(4), 1227–1238.
- Schmieger, H. (1999). Molecular survey of the *Salmonella* phage typing system of Anderson. *Journal of Bacteriology*, 181(5), 1630–1635.
- Schrader, K. N., Fernandez-Castro, A., Cheung, W. K. W., Crandall, C. M., & Abbott, S. L. (2008). Evaluation of commercial antisera for *Salmonella* serotyping. *Journal of Clinical Microbiology*, 46(2), 685–688.
- Seemann, T. (2018). ABRicate: mass screening of contigs for antimicrobial and virulence genes. *Department of Microbiology and Immunology, The University of Melbourne, Melbourne, Australia*. Available Online: <https://Github.Com/Tseemann/Abriicate> (Accessed on 28 February 2019).
- Seemann, Torsten. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069.
- Seemann, Torsten. (2018a). *Snippy: fast bacterial variant calling from NGS reads*.

<https://github.com/tseemann/snippy>

Seemann, Torsten. (2018b). *Snp-dists: Convert a FASTA alignment to SNP distance matrix*.

<https://github.com/tseemann/snp-dists>

Selander, R. K., Li, J., & Nelson, K. (1996). Evolutionary genetics of *Salmonella enterica*. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, 2, 2691–2707.

Serra Valdés, M. Á. (2017). La resistencia microbiana en el contexto actual y la importancia del conocimiento y aplicación en la política antimicrobiana. *Revista Habanera de Ciencias Médicas*, 16(3), 402–419.

Shi, C., Singh, P., Ranieri, M. L., Wiedmann, M., & Moreno Switt, A. I. (2015). Molecular methods for serovar determination of *Salmonella*. *Critical Reviews in Microbiology*, 41(3), 309–325.

Shimizu, T. (2022). The Growth of the Fruit and Vegetable Export Industry in Peru. *SpringerBriefs in Economics*.

Silva-Hidalgo, G., Ortiz-Navarrete, V. F., Alpuche-Aranda, C. M., Rendón-Maldonado, J. G., López-Valenzuela, M., Juárez-Barranco, F., & López-Moreno, H. S. (2012). Non-typhi *Salmonella* serovars found in Mexican zoo animals. *Research in Veterinary Science*, 93(3), 1132–1135.

Simpson, G. G. (1964). Organisms and Molecules in Evolution: Studies of evolution at the molecular level lead to greater understanding and a balancing of viewpoints. *Science*, 146(3651), 1535–1538.

Smith, K. E., Medus, C., Meyer, S. D., Boxrud, D. J., Leano, F. E., Hedberg, C. W., Elfering, K., Braymen, C., Bender, J. B., & Danila, R. N. (2008). Outbreaks of salmonellosis in Minnesota (1998 through 2006) associated with frozen, microwaveable, breaded, stuffed chicken products. *Journal of Food Protection*, 71(10), 2153–2160.

Soto-Estrada, G., Moreno-Altamirano, L., & Pádua Díaz, D. (2016). Panorama epidemiológico de México, principales causas de morbilidad y mortalidad. *Revista de La Facultad de Medicina (México)*, 59(6), 8–22.

Steele, M., & Odumeru, J. (2004). Irrigation water as source of foodborne pathogens on fruit and vegetables. *Journal of Food Protection*, 67(12), 2839–2849.

Steemers, F. J., & Gunderson1 Kevin L, 2. (2005). *Illumina, Inc*.

- Stevens, M. P., Humphrey, T. J., & Maskell, D. J. (2009). Molecular insights into farm animal and zoonotic salmonella infections. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 364, Issue 1530, pp. 2708–2723). Royal Society. <https://doi.org/10.1098/rstb.2009.0094>
- Subedi, D., Kohli, G. S., Vijay, A. K., Willcox, M. D. P., & Rice, S. A. (2018). Accessory genome contributes to the virulence and resistance of the ocular isolate of *Pseudomonas aeruginosa*: A complete genome analysis. *BioRxiv*, 497974.
- Tomastikova, Z., Romero, S. B., Knotek, Z., & Karpiskova, R. (2017). Prevalence and characteristics of *Salmonella* species isolated from captive reptiles in the Czech Republic. *Veterinární Medicína*, 62(8), 456–469.
- Toro, M., Retamal, P., Ayers, S., Barreto, M., Allard, M., Brown, E. W., & Gonzalez-Escalona, N. (2016). Whole-genome sequencing analysis of *Salmonella enterica* serovar Enteritidis isolates in Chile provides insights into possible transmission between gulls, poultry, and humans. *Applied and Environmental Microbiology*, 82(20), 6223–6232.
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(W1), W232–W235.
- Tritt, A., Eisen, J. A., Facciotti, M. T., & Darling, A. E. (2012). An Integrated Pipeline for de Novo Assembly of Microbial Genomes. *PLOS ONE*, 7(9), e42304. <https://doi.org/10.1371/journal.pone.0042304>
- Trost, B., Walker, S., Wang, Z., Thiruvahindrapuram, B., MacDonald, J. R., Sung, W. W. L., Pereira, S. L., Whitney, J., Chan, A. J. S., & Pellicchia, G. (2018). A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *The American Journal of Human Genetics*, 102(1), 142–155.
- Urwin, R., & Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends in Microbiology*, 11(10), 479–487.
- Van Belkum, A. (2007). Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunology & Medical Microbiology*, 49(1), 22–27.
- Van Der Wal, C., & Ho, S. Y. W. (2019). *Molecular Clock* (S. Ranganathan, M. Gribskov, K. Nakai, & C. B. T.-E.

of B. and C. B. Schönbach (eds.); pp. 719–726). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-809633-8.20257-4>

- Vázquez-Garcidueñas, M. S., Romero-Pérez, N. L., Figueroa-Aguilar, G. A., Jaime-Sánchez, J. L., & Vázquez-Marrufo, G. (2014). Investigation of a food-borne *Salmonella* Oranienburg outbreak in a Mexican prison. *The Journal of Infection in Developing Countries*, 8(02), 143–153.
- Velge, P., Wiedemann, A., Rosselin, M., Abed, N., Boumart, Z., Chaussé, A. M., Grépinet, O., Namdari, F., Roche, S. M., Rossignol, A., & Virlogeux-Payant, I. (2012). Multiplicity of *Salmonella* entry mechanisms, a new paradigm for *Salmonella* pathogenesis. *MicrobiologyOpen*, 1(3), 243–258. <https://doi.org/10.1002/mbo3.28>
- Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics*, 40(4), 277.
- Vernikos, G., Medini, D., Riley, D. R., & Tettelin, H. (2015). Ten years of pan-genome analyses. *Current Opinion in Microbiology*, 23, 148–154.
- Voetsch, A. C., Van Gilder, T. J., Angulo, F. J., Farley, M. M., Shallow, S., Marcus, R., Cieslak, P. R., Deneen, V. C., Tauxe, R. V., & Group, E. I. P. F. W. (2004). FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. *Clinical Infectious Diseases*, 38(Supplement_3), S127–S134.
- Wain, J., House, D., Zafar, A., Baker, S., Nair, S., Kidgell, C., Bhutta, Z., Dougan, G., & Hasan, R. (2005). Vi antigen expression in *Salmonella enterica* serovar Typhi clinical isolates from Pakistan. *Journal of Clinical Microbiology*, 43(3), 1158–1165.
- Wang, M., & Caetano-Anollés, G. (2013). *Molecular Clock* (S. Maloy & K. B. T.-B. E. of G. (Second E. Hughes (eds.); pp. 461–462). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374984-0.00964-5>
- Wang, Q., Holmes, N., Martinez, E., Howard, P., Hill-Cawthorne, G., & Sintchenko, V. (2015). It is not all about single nucleotide polymorphisms: comparison of mobile genetic elements and deletions in *Listeria monocytogenes* genomes links cases of hospital-acquired listeriosis to the environmental source. *Journal of Clinical Microbiology*, 53(11), 3492–3500.

- White, N. J. (2002). *Salmonella typhi* (typhoid fever) and *S. paratyphi* (paratyphoid fever). In *Antimicrobial Therapy and Vaccines* (pp. 583–603). Apple Tree Productions, New York.
- Winfield, M. D., & Groisman, E. A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl. Environ. Microbiol.*, 69(7), 3687–3694.
- Wu, H. J., Wang, A. H. J., & Jennings, M. P. (2008). Discovery of virulence factors of pathogenic bacteria. In *Current Opinion in Chemical Biology* (Vol. 12, Issue 1, pp. 93–101). Elsevier Current Trends.
<https://doi.org/10.1016/j.cbpa.2008.01.023>
- Yoshida, C. E., Kruczkiewicz, P., Laing, C. R., Lingohr, E. J., Gannon, V. P. J., Nash, J. H. E., & Taboada, E. N. (2016). The *Salmonella* in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PloS One*, 11(1), e0147101.
- Zaidi, M. B., Calva, J. J., Estrada-Garcia, M. T., Leon, V., Vazquez, G., Figueroa, G., Lopez, E., Contreras, J., Abbott, J., & Zhao, S. (2008). Integrated food chain surveillance system for *Salmonella* spp. in Mexico. *Emerging Infectious Diseases*, 14(3), 429.
- Zaidi, M. B., López, C., & Calva, E. (2006). Estudios mexicanos sobre *Salmonella*: epidemiología, vacunas y biología molecular. *Rev Latinoam Microbiol*, 48(2), 121–125.
- Zhang, S., Kingsley, R. A., Santos, R. L., Andrews-Polymenis, H., Raffatellu, M., Figueiredo, J., Nunes, J., Tsolis, R. M., Adams, L. G., & Bäumler, A. J. (2003). Molecular pathogenesis of *Salmonella enterica* serotype Typhimurium-induced diarrhea. *Infection and Immunity*, 71(1), 1–12.
- Zuckerkindl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In V. Bryson & H. J. B. T.-E. G. and P. Vogel (Eds.), *Evolving Genes and Proteins* (pp. 97–166). Academic Press.
<https://doi.org/10.1016/b978-1-4832-2734-4.50017-6>

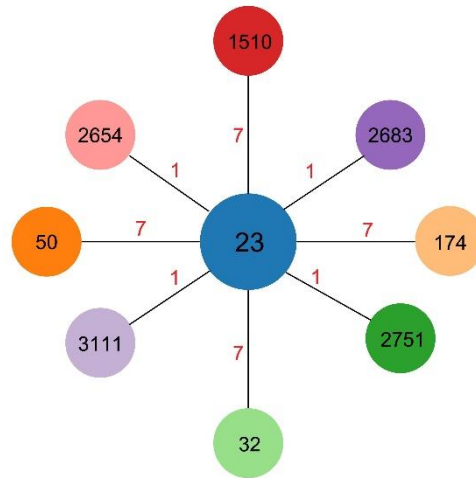


Figure 1. Spanning tree for *S. Oranienburg* genomes. The found 9 STs are represented in different color circles. Red numbers indicate the number of allele variants against the founder ST23. Circles size do not represent the proportion of genomes associated to each ST.

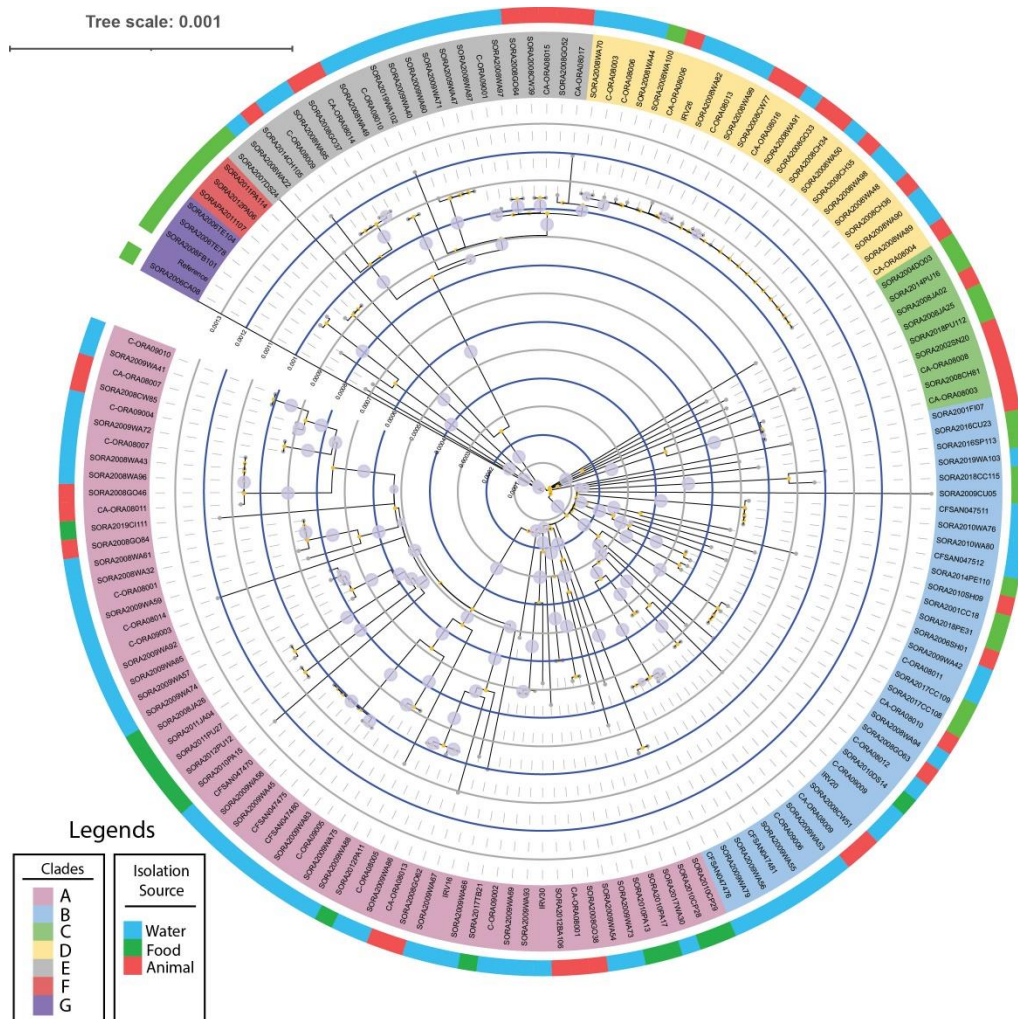


Figure 2. Core genome phylogenetic tree of 165 Mexican *S. Oranienburg* isolates. Specific clades are colored in different colors. Bootstraps higher than >85 are shown in blue circles. Outer ring represents isolation source. Tree scale bar indicate the genetic change amount of 0.001 expressed in substitutions per site.

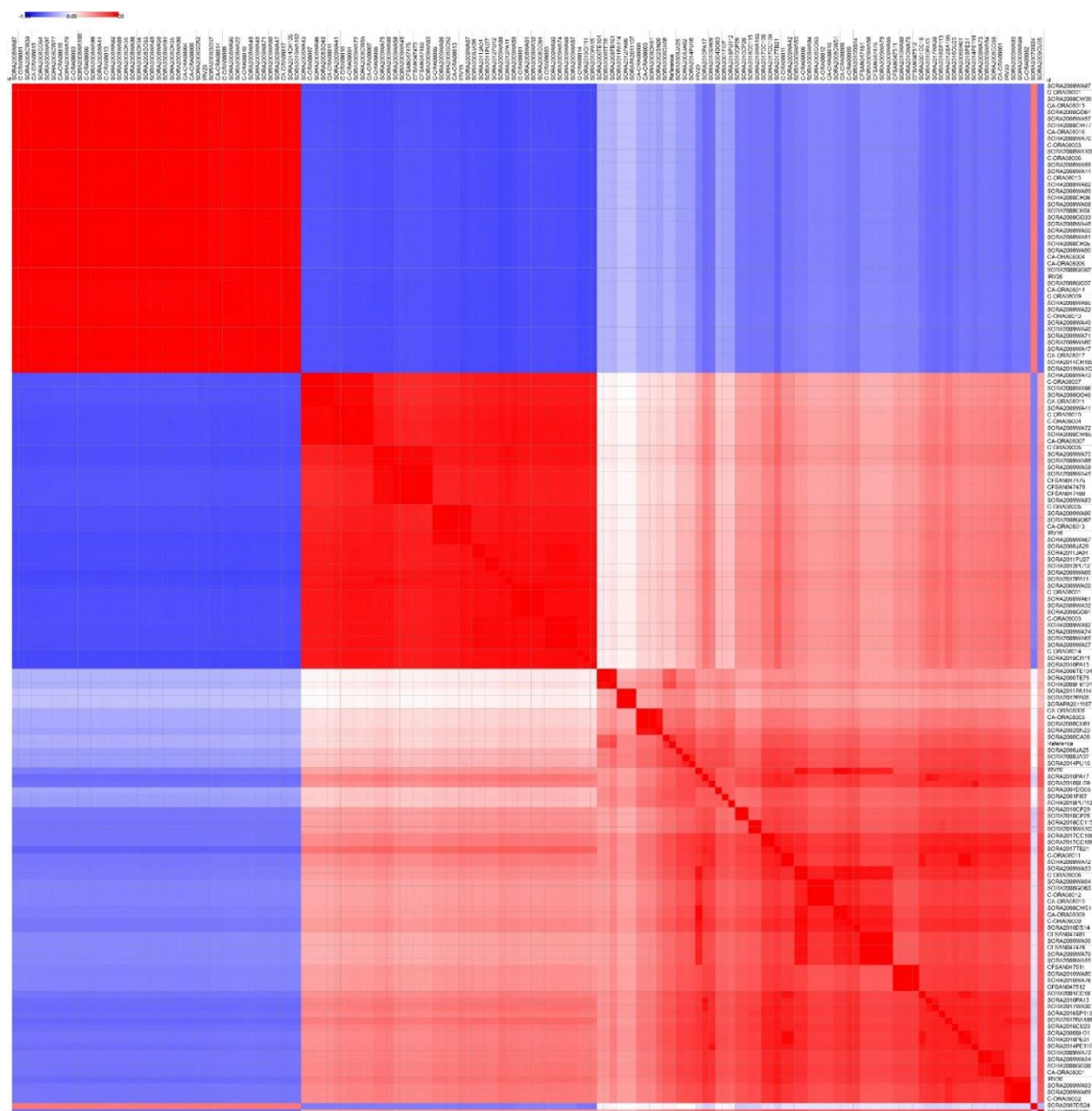


Figure 4. Heatmap based on SNP distance among *S. Oranienburg* core genomes. Colors indicate Pearson correlation coefficient. Red shows positive correlation (values near +1) while blue negative correlation (-1) and white no correlation (values near 0).

4. CONCLUSIONES GENERALES

Salmonella enterica de la región de Culiacán, Sinaloa de diversas fuentes ambientales presentaron alto grado de similitud genética de acuerdo al bajo número de SNPs detectados. Lo que sugiere que esta se encuentra recirculando en la región. En cuanto a su estructura poblacional, un grupo clonal identificado como ST23, correspondiente al serotipo Oranienburg fue encontrado. Genes asociados a la resistencia al estrés ambiental y de rutas metabólicas fueron encontrados en los aislados ambientales, lo que podría representar una ventaja adaptativa y de supervivencia en ambientes no hospederos, lo que sugiere que el punto de aislamiento juega un papel muy importante en la plasticidad genómica. Por el contrario, en los aislados clínicos se identificaron genes asociados a virulencia como flagelares y de sistemas de secreción, los cuales juegan un papel fundamental en la colonización, invasión y establecimiento en el hospedero. Adicionalmente, se encontró que entre los mecanismos que median la adaptación de *Salmonella* en el ambiente se encuentran la duplicación genética y acumulación de pseudogenes, como principales.

5. RECOMENDACIONES

Para futuros proyectos se recomienda aislar y secuenciar cepas de origen clínico para ampliar nuestro catálogo de genomas, los cuales pueden ser utilizados para tener una perspectiva más contextual de las diferencias genéticas encontradas entre los aislados ambientales y clínicos. De igual forma sería interesante explorar los genes de resistencia en dichas cepas, para proponer tratamientos adecuados de acuerdo a dichos genes, ya que este problema se acrecienta año con año afectando a millones de personas. Otro aspecto interesante es enriquecer la base genómica con aislados de un periodo de tiempo mayor, esto con la finalidad de mejorar la señal espacio temporal.