



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

**IDENTIFICACIÓN Y CARACTERIZACIÓN DE PSEUDOGENES
DE *Salmonella enterica* AISLADA DEL RIO CULIACÁN**

Por:

Karla María Sandoval Ramos

TESIS APROBADA POR LA

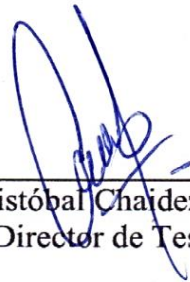
COORDINACIÓN EN CIENCIA Y TECNOLOGÍA PARA PRODUCTOS AGRÍCOLAS EN
ZONAS TROPICALES Y SUBTROPICALES

Como requisito parcial para obtener el grado de

MAESTRA EN CIENCIAS

APROBACIÓN

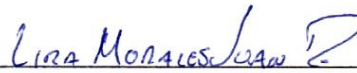
Los miembros del comité designado para la revisión de la tesis de Karla María Sandoval Ramos la han encontrado satisfactoria y recomiendan que sea aceptada como requisito parcial para obtener el grado de Maestra en Ciencias.



Dr. Cristóbal Chaidez Quiroz
Director de Tesis



Dra. Nohelia Castro del Campo
Integrante de comité de tesis



Dr. Juan Daniel Lira Morales
Integrante de comité de tesis



Dr. José Andrés Medrano Félix
Integrante de comité de tesis

DECLARACIÓN INSTITUCIONAL

La información generada en la tesis "Identificación y Caracterización de Pseudogenes de *Salmonella enterica* Aislada del Rio Culiacán" 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 de la autora Karla María Sandoval Ramos, 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.



CENTRO DE INVESTIGACIÓN EN
ALIMENTACIÓN Y DESARROLLO, A.C.
Coordinación de Programas Académicos

A handwritten signature in blue ink, appearing to read "Graciela Caire Juvera", written over a horizontal line.

Dra. Graciela Caire Juvera
Directora General

AGRADECIMIENTOS

Agradezco al Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) por su apoyo económico durante la realización de mi proyecto de maestría.

Al Centro de Investigación en Alimentación y Desarrollo A.C. (CIAD) por haber brindado la oportunidad de ser parte de esta prestigiosa institución y por facilitarme los espacios para llevar a cabo mi investigación de manera efectiva.

Al Dr. Cristóbal Chaidez por aceptar dirigir este trabajo de investigación y brindarme la oportunidad de desarrollarme profesionalmente. Gracias por su confianza, su conocimiento y experiencia; así como por su apoyo y orientación los cuales fueron indispensables en cada etapa de mi investigación.

A mis asesores de tesis Dra. Nohelia Castro del Campo, Dr. José Andrés Medrano Félix y Dr. Juan Daniel Lira Morales, muchas gracias por aceptar formar parte de este proyecto. Su dedicación, valiosos comentarios, aportes y revisiones contribuyeron a mejorar indudablemente mi trabajo de investigación.

A mi asesor extendido Dr. José Roberto Aguirre Sánchez, mi más sincero agradecimiento por su invaluable guía y apoyo incondicional durante la realización de esta tesis. Su mentoría, compromiso, experiencia y dedicación fueron una pieza fundamental para alcanzar una finalización exitosa. Gracias por siempre brindarme orientación y motivación.

A mis compañeros de generación y a todas las personas que conocí durante el curso de esta etapa, agradezco su apoyo y todos los recuerdos invaluable.

A mis amigas quienes me han acompañado a lo largo de este proceso y en diferentes momentos de mi vida. Su compañía siempre hace que todo sea más ameno y significativo. Agradezco su fe en mí y su apoyo incondicional. Gracias por su amistad.

A mi mamá y hermana, quienes han sido mi mayor fuente de motivación y ejemplo. Gracias por ser mi guía constante en todas las etapas de mi vida y por impulsarme a perseguir mis sueños. Su apoyo ha sido fundamental para mi desarrollo y crecimiento personal y siempre estaré agradecida por tenerlas a mi lado.

¡Muchas gracias!

CONTENIDO

APROBACIÓN	2
DECLARACIÓN INSTITUCIONAL	3
AGRADECIMIENTOS	4
CONTENIDO	6
LISTA DE FIGURAS	8
LISTA DE CUADROS	9
RESUMEN	10
ABSTRACT	11
1. INTRODUCCIÓN	12
2. ANTECEDENTES	14
2.1. Historia del Género <i>Salmonella</i>	14
2.2. Características Generales De <i>Salmonella</i> Spp.....	14
2.3. Metabolismo y Perfil Bioquímico de <i>Salmonella</i> Spp.	15
2.4. Antígenos de Superficie de <i>Salmonella</i> Spp.....	17
2.5. Taxonomía de <i>Salmonella</i> Spp.....	18
2.6. Epidemiología de <i>Salmonella enterica</i>	19
2.7. Especificidad de <i>Salmonella</i> por Hospederos	24
2.7.1. <i>Salmonella</i> en Ambientes Hospederos	25
2.7.2. <i>Salmonella</i> en Ambientes no Hospederos	26
2.8. El Genoma de <i>Salmonella enterica</i>	26
2.9. Descubrimiento de Pseudogenes	28
2.10. Definición y Clasificación de Pseudogenes.....	29
2.10.1. Pseudogenes no Procesados.....	30
2.10.2. Pseudogenes Procesados.....	30
2.10.3. Pseudogenes Unitarios.	31
2.11. Proceso de Pseudogenización	31
2.12. Evolución y Función de los Pseudogenes	32
2.13. Identificación de Pseudogenes.....	33
2.13.1. Exploración de los Pseudogenes en Genomas Bacterianos	34
2.14. Justificación.....	36
3. HIPÓTESIS	37
4. OBJETIVOS	38
4.1. Objetivo General.....	38
4.2. Objetivos Específicos	38
5. MATERIALES Y MÉTODOS	39
5.1. Obtención De Secuencias Objetivo	39
5.2. Obtención de Secuencias de Consulta.....	39

CONTENIDO (continuación)

5.3. Identificación De Pseudogenes.....	40
5.3.1. Prerrequisitos	40
5.3.1.1 Creación de pseudocromosoma.....	43
5.3.1.2 Anotación funcional de secuencias.....	43
5.3.1.3 Creación y alineación de bases de datos.....	41
5.4. Predicción de Pseudogenes.....	41
5.4.1. Primer Módulo De Psi-Phi.....	41
5.4.2. Segundo Modulo de Psi-Phi	42
5.4.3. Anotación COG	44
6. RESULTADOS Y DISCUSIÓN.....	45
6.1. Identificación de Pseudogenes	45
6.2. Categorías Funcionales COG Asociadas a Genes Asociados a Pseudogenes	47
6.3. Principales Genes Asociados a Pseudogenes	50
6.4. Descripción de Genes Asociados a Pseudogenes	54
6.4.1. Gen <i>lpp</i>	54
6.4.2. <i>fdnG</i>	55
6.4.3. <i>ccmH</i>	55
6.4.4. <i>tufB</i>	56
6.4.5. <i>queA</i>	56
7. CONCLUSIONES.....	57
8. RECOMENDACIONES.....	58
9. REFERENCIAS	59
10. APÉNDICES.....	64
10.1 Categoría Funcional de los Genes Asociados a Pseudogenes en <i>S. Weltevreden</i>	64
10.2. Categoría Funcional de los Genes Asociados a Pseudogenes en <i>S. enterica</i> Saintpaul.	65
10.3. Categoría Funcional de los Genes Asociados a Pseudogenes en <i>S. enterica</i> Oranienburg.	67

LISTA DE FIGURAS

Figura		Página
1	Estructura antigénica de <i>Salmonella</i>	16
2	Incidencia de fiebre tifoidea a nivel mundial.....	21
3	Distribución mundial de casos de <i>Salmonella</i> no tifoidea en el año 2010...	22
4	Ciclo de vida de <i>Salmonella</i> en ambientes externos y hospederos.....	24
5	Regiones de un gen y pseudogen.....	29
6	Principales clases de pseudogenes.....	31
7	Captura de pantalla del módulo 1 del algoritmo Psi-Phi.....	41
8	Captura de pantalla del módulo 2 del algoritmo Psi-Phi.....	43
9	Conteo de pseudogenes presentes en 84 secuencias de <i>Salmonella</i> enterica serotipos Oranienburg, Saintpaul y Weltevreden.....	46
10	Principales eventos que originan a los pseudogenes en los serotipos Oranienburg, Saintpaul y Weltevreden.....	47
11	Categorías COG presentes en los genes asociados a pseudogenes en las secuencias de los serotipos Saintpaul, Oranienburg y Weltevreden.....	49
12	Heatmap de presencia/ausencia de los principales genes homólogos a pseudogenes en <i>S. Saintpaul</i> , <i>S. Oranienburg</i> , <i>S. Weltevreden</i>	51

LISTA DE CUADROS

Cuadro		Página
1	Perfil bioquímico presente en las subespecies de <i>Salmonella</i>	16
2	Características bioquímicas de <i>Salmonella</i>	16
3	Clasificación, designación y ocurrencia de las seis subespecies de <i>Salmonella enterica</i>	19
4	Casos por entidad federativa de Fiebre Tifoidea, Fiebre Paratifoidea y Otras Salmonelosis hasta la semana 51 del 2021.....	23
5	Relación de secuencias objetivo que fueron empleadas durante la identificación y caracterización de pseudogenes en serotipos de <i>Salmonella enterica</i>	39
6	Genomas de referencia descargados de NCBI Pathogen Detection.....	39
7	Porcentaje de pseudogenes presentes en el genoma de los serotipos Oranienburg, Weltevreden y Saintpaul.....	43

RESUMEN

En bacterias, los pseudogenes son secuencias de ADN no funcional similares a genes funcionales, se originan durante el curso de la evolución mediante mutaciones puntuales en regiones reguladoras y marcos de lectura abiertos (ORF). A pesar de que los pseudogenes son escasos en genomas bacterianos, la pseudogenización se considera un proceso importante en el que ciertos genes se vuelven dispensables durante la colonización de un nuevo nicho ecológico. Este proceso puede considerarse como adaptativo para las bacterias y, además, puede fungir como un registro de la historia evolutiva del organismo. Por esta razón, es de interés determinar el papel que tienen los pseudogenes en la adaptación ambiental de este patógeno. Para lograrlo, 84 secuencias genómicas de *Salmonella enterica* pertenecientes a los serotipos Weltevreden (n=7), Saintpaul (n=37) y Oranienburg (n=40) aislados del río Culiacán fueron analizadas con el algoritmo Psi-Phi; comparando las secuencias de nucleótidos de cada genoma objetivo contra la secuencia de proteínas anotadas de un genoma de consulta para cada serotipo. Los criterios de evaluación fueron porcentaje de identidad (< 80%), *e*-value (< $10e^{-15}$) y la presencia de mutaciones en regiones reguladoras. Se lograron anotar un total de 94, 62 y 24 pseudogenes en los serotipos Oranienburg, Saintpaul y Weltevreden, respectivamente. De igual manera, se identificaron 25 categorías COG (Cluster of Orthologous Group) presentes en los genes asociados a pseudogenes en los cuales predominó la ausencia de categoría COG (31.56%), seguido de la categoría asociada a traducción, estructura ribosomal y biogénesis (10.30%), producción y conversión de energía (9.97%), transporte y metabolismo de carbohidratos (7.97%), modificaciones postraduccionales, modificaciones de proteínas y chaperonas (7.64%); las cuales se ven implicadas en el metabolismo y reparación de ADN, por lo que de acuerdo con los resultados obtenidos se puede inferir que los perfiles de pseudogenización identificados en los aislados ambientales de *Salmonella enterica* proporcionan una valiosa huella adaptativa y evolutiva. Estos pseudogenes, que se originan a través de mutaciones en regiones reguladoras y ORF, así como otros mecanismos, pueden representar una estrategia adaptativa para las bacterias en un nuevo nicho ecológico.

Palabras claves: Pseudogen, *Salmonella enterica*, identificación, ambientes no hospederos

ABSTRACT

In bacteria, pseudogenes are non-functional DNA sequences that are like functional genes. They originate during evolution through point mutations in regulatory regions and open reading frames (ORFs). Although pseudogenes are rare in bacterial genomes, pseudogenization is considered an important process in which certain genes become dispensable during the colonization of a new ecological niche. Thus, this process can be adaptive for bacteria and can also serve as a record of the organism's evolutionary history. Currently, there are few studies on the annotation of pseudogenes in *Salmonella enterica* genomes. Therefore, it is of interest to determine the role that pseudogenes play in the environmental adaptation of this pathogen. To achieve this, 84 genomic sequences of *Salmonella enterica* belonging to the Weltevreden (n=7), Saintpaul(n=37), and Oranienburg (n=40) serotypes isolated from the Culiacan River were analyzed using the Psi-Phi algorithm, comparing the nucleotide sequences of each target genome against the annotated protein sequence of a reference genome for each serotype. The evaluation criteria were identity percentage (<80%), e-value (<10e⁻¹⁵), and the presence of mutations in regulatory regions. As a result, a total of 94, 62, and 24 pseudogenes were annotated in the Oranienburg, Saintpaul and Weltevreden serotypes, respectively. Similarly, 25 COG categories were identified in the genes associated with pseudogenes, with the absence of the COG category (31.56%) being predominant, followed by the categories associated with translation, ribosomal structure, and biogenesis (10.30%), energy production and conversion (9.97%), carbohydrate transport and metabolism (7.97%), and post-translational modifications, protein modifications, and chaperones (7.64%). These categories are implicated in DNA metabolism and repair, so based on the results obtained, it can be inferred that the pseudogenization profiles identified in environmental isolates of *Salmonella enterica* provide valuable insights into adaptive and evolutionary patterns. These pseudogenes, originating from mutations in regulatory regions, ORFs, and other mechanisms, may represent an adaptive strategy for bacteria in a new ecological niche.

Keywords. Pseudogen, *Salmonella enterica*, identification, non-host environments.

1. INTRODUCCIÓN

Salmonella enterica es una bacteria Gram negativa perteneciente a la familia *Enterobacteriaceae*. Al ser una bacteria entérica, esta se encuentra principalmente colonizando los intestinos de un amplio rango de especies animales, de igual manera puede estar de forma ubicua en agua, suelo y vegetales (Eng *et al.*, 2015).

Salmonella enterica es un patógeno relevante para la salud pública, comúnmente se asocia a al consumo de agua o alimentos contaminados con materia fecal. Entre los principales alimentos que ocasionan brotes epidemiológicos se encuentran los de origen animal como: carne, huevo y leche. No obstante, también se han reportado brotes de *Salmonella* relacionados con el consumo de hortalizas y frutas frescas (Ehuwa *et al.*, 2021; Eng *et al.*, 2015).

La vía de transmisión de *Salmonella enterica* es mediante la ruta fecal-oral, permitiendo la diseminación de serotipos causantes de salmonelosis no tifoidea (serotipo Enteritidis y *Typhimurium*) y fiebre tifoidea (serotipo Typhi) en ambientes no hospederos donde puede ser prevalente durante largas temporadas (Bell y Kyriakides, 2002).

Salmonella tiene una amplia distribución en México y a nivel mundial, debido al rango amplio de las condiciones para su crecimiento, con valores, de temperatura que abarca desde los 7°C hasta los 48°C y un pH de 3.8 a 7.5, lo que le permite adaptarse en diversos ambientes (Ellermeier y Schlauch, 2006).

Para hacer frente a estas condiciones de estrés ambiental *Salmonella enterica* ha desarrollado diversos mecanismos genéticos que le confieren capacidad de adaptación al estrés ambiental, tal es el caso de la pseudogenización, que es un proceso evolutivo mediante el cual un gen funcional se convierte en pseudogenes (Emmanuelle Lerat y Howard Ochman, 2005).

Un pseudogen corresponde a una secuencia de ADN que es homologa a la secuencia de un gen funcional que ha perdido su capacidad de codificar una proteína funcional debido a mutaciones acumulativas en el genoma, como inserciones, deleciones o mutaciones puntuales, las cuales afectan negativamente la estructura del gen causando una alteración mayor del 20% en su estructura (Emmanuelle Lerat y Howard Ochman, 2005; Podlaha y Zhang, 2010).

El estudio de los pseudogenes en *Salmonella* es importante porque nos ayuda a comprender mejor su biología básica ya que nos brinda información sobre la estructura de su genoma, la evolución de

la bacteria y cómo se ha adaptado a diferentes entornos.

Además, el estudio de pseudogenes permite identificar genes que son esenciales para la supervivencia de la bacteria y aquellos que no lo son (Feng *et al.*, 2022; Li *et al.*, 2013).

Actualmente, son escasos los trabajos enfocados en la identificación de pseudogenes presentes en genomas de *Salmonella enterica* aislados de origen ambiental; por lo tanto, es necesario que se realice la identificación y anotación funcional de sus pseudogenes para conocer qué papel cumplen en su adaptación y colonización a nuevos ambientes.

Bajo este contexto a lo anterior, el objetivo del presente estudio es determinar qué papel tienen los pseudogenes en la adaptación de *Salmonella enterica* en ambientes no hospederos, por lo cual se analizaron 84 secuencias genómicas de *Salmonella enterica* pertenecientes a los serotipos Weltevreden (n=7), Saintpaul (n=37) y Oranienburg (n=40) aislados del río Culiacán.

2. ANTECEDENTES

2.1. Historia del Género *Salmonella*

En 1884, el epidemiólogo Theobald Smith se unió al Oficina de Industria Animal (BAI por sus siglas en inglés) de Estados Unidos como asistente de laboratorio asignado a investigaciones patológicas de enfermedades animales infecciosas bajo la dirección del Dr. Daniel Elmer Salmon (Andrews *et al.*, 2010)

Uno de sus primeros trabajos realizados para el BAI fue sobre el cólera porcino. Su investigación condujo al descubrimiento de una nueva especie de bacteria (*Salmonella enterica*, antes llamada *Salmonella choleraesuis*) en 1885, la cual aisló por primera vez de los intestinos de cerdos infectados con cólera porcina y por lo tanto se creía que era el agente causal de esta enfermedad (Eng *et al.*, 2015). Sin embargo, fue demostrado que el cólera porcino es una enfermedad producida por una infección viral y que el bacilo de *Salmonella* a menudo estaba presente como una infección secundaria (Andrews *et al.*, 2010).

Posteriormente, el género *Salmonella* fue descrito y definido por Daniel Salmon, cuyo apellido sirvió como base para nombrar al género en su honor. Su taxonomía comprende dos especies: *Salmonella bongori* y *Salmonella enterica*. Esta última se clasifica en seis subespecies: *S. enterica* subs. *enterica* (I), *S. enterica* subs. *salamae* (II), *S. enterica* subs. *arizonae* (IIIa), *S. enterica* subs. *diarizonae* (IIIb), *S. enterica* subs. *houtenae* (IV) y *S. enterica* subs. *indica* (VI), que en conjunto engloban a más de 2,600 serotipos (Bell y Kyriakides, 2002; Contreras-Soto *et al.*, 2018; Eng *et al.*, 2015).

2.2. Características Generales De *Salmonella* Spp.

El género *Salmonella* pertenece a la familia *Enterobacteriaceae*, que incluye a otros géneros como *Escherichia*, *Klebsiella* y *Proteus*. Estos se encuentran ampliamente distribuidos en plantas, agua,

suelo e intestinos de animales y el ser humano (Eng *et al.*, 2015).

El género *Salmonella* está compuesto por bacilos Gram negativos, generalmente motiles por flagelos peritricos (excepto *Salmonella enterica* Pullorum y *S. enterica* Gallinarum), anaerobios facultativos y no esporulados. Su tamaño promedio oscila entre 2 a 5 µm de largo y de 0.8 a 1.5 µm de ancho (Percival y Williams, 2014).

Los miembros de este género utilizan al citrato como única fuente de carbono y poseen metabolismo de tipo oxidativo y fermentativo (ICMSF, 1996).

El crecimiento y sobrevivencia de *Salmonella* se ve influenciado principalmente por tres condiciones: temperatura, pH y actividad de agua (Aw). Este microorganismo puede crecer en un amplio rango de temperatura que va desde los 5.2°C a 46.2°C, por lo que es considerado como un microorganismo mesófilo. Su temperatura óptima de crecimiento entre los 35°C y 43°C (ICMSF, 1996).

En el caso del pH, este microorganismo crece en un pH que oscila entre 3.8 a 9.5, con un crecimiento optimo en un pH de 7-7.5, por lo que es considerado como un organismo neutrófilo. Adicionalmente, los niveles de actividad de agua (Aw) pueden influir en la multiplicación de *Salmonella* en alimentos, la cual se da un rango de 0.99 a 0.94 (Santillana *et al.*, 2013; Spector y Kenyon, 2012).

2.3. Metabolismo y Perfil Bioquímico de *Salmonella* Spp.

La determinación del perfil bioquímico (Cuadro 1) de *Salmonella* ha permitido identificar a las especies y subespecies del género. Sin embargo, para una identificación más detallada se sugiere realizar una serotipación a los aislados con la técnica Kauffmann-White, la cual clasifica al género *Salmonella* en serotipos basado en la presencia de los antígenos somáticos (O), flagelares (H) y capsulares (K) (Jajere *et al.*, 2006).

Salmonella se caracteriza bioquímicamente por utilizar al citrato como única fuente de carbono, con excepción de *S. Typhi*. Es positiva a la reacción de catalasa, lisina, ornitina y arginina, pero, son indol, urea y oxidasa negativos (Cuadro 2). Fermenta glucosa y otros carbohidratos mediante la producción de gas y ácido sulfhídrico (excepto subespecies *S. arizonae* y *S. diarizonae*) (Bumann

y Schothorst, 2017).

Existen serotipos que presentan excepciones al perfil bioquímico como *S. Typhi*, *S. Gallinarum*, *S. Pullorum* y *S. Paratyphi A*.

Cuadro 1. Perfil bioquímico presente en las subespecies de *Salmonella*. Modificado de (Bell,2002)

Reacciones bioquímicas	<i>S. enterica enterica</i>	<i>S. enterica salamae</i>	<i>S. enterica arizonae</i>	<i>S. enterica diarizonae</i>	<i>S. enterica houtenae</i>	<i>S. enterica indica</i>	<i>S. enterica bongori</i>
Dulcitol	+	+	-	-	-	d	+
Lactosa	-	-	-	+	-	d	-
O.N.P. G	-	-	+	+	-	d	+
Salicina	-	-	-	-	+	-	-
Sorbitol	+	+	+	+	+	-	+
Malonato	-	-	-	-	-	-	-
<i>d</i> -tartrato	+	-	-	-	-	-	-
Mucato	+	+	+	d	-	+	+
Gelatinasa	-	+	+	+	+	+	-
β -glucuronidasa	d	d	-	+	-	d	-
Galacturonato	-	+	-	+	+	+	+
Crecimiento en KCN	-	-	-	-	+	-	+

+ = > 90% de los serotipos positivos

- = > 90% de los serotipos negativos

d= reacciones diferentes.

Cuadro 2. Características bioquímicas de *Salmonella*. Modificado de (Bell,2002)

Característica	Reacción usual
Catalasa	+
Oxidasa	-
Acido producido a partir de lactosa	-
Gas producido a partir de glucosa *	+
Indol	-
Ureasa producida	-
Sulfuro de hidrógeno producido a partir de agar de hierro con triple azúcar	+
Citrato utilizado como única fuente de carbono*	+
Rojo de Metilo	+
Voges-Proskauer	-
Lisina descarboxilasa	+
Ornitina descarboxilasa	+

+ = reacción positiva; - = reacción negativa. * Una excepción importante es *S. Typhi*, la cual tiene reacción negativa a estas pruebas.

2.4. Antígenos de Superficie de *Salmonella* Spp

Salmonella se caracteriza por presentar tres antígenos de superficie en su estructura externa: antígeno somático (O), antígeno capsular (K) y antígeno flagelar (H) (Figura 1) (Andrews-Polymeris, Baumber, McCormick, y Fang, 2010). La presencia de estos antígenos es utilizada para clasificar a los serotipos de *Salmonella* de acuerdo con el esquema propuesto por Kauffman y White (Grimont y Weill, 2007).

El antígeno somático (O) está conformado por una cadena repetida de polisacáridos componente del lipopolisacárido (LPS) bacteriano que se encuentra en la membrana externa de la envoltura celular bacteriana. Su función es actuar como una barrera de protección contra los agentes externos e interactuar contra la defensa humoral del hospedero (Crawford *et al.*, 2013).

Generalmente, el antígeno O no es único, si no que este puede expresar diversos factores antigénicos, que pueden ser comunes entre otros serotipos, lo que ocasiona que *Salmonella* se puede dividir en distintos grupos O, por ejemplo: O:9, O:12.

El antígeno flagelar (H) es una proteína que es componente del flagelo bacteriano, juega un papel importante en la activación de respuesta contra la respuesta del hospedero. (Figura 1)

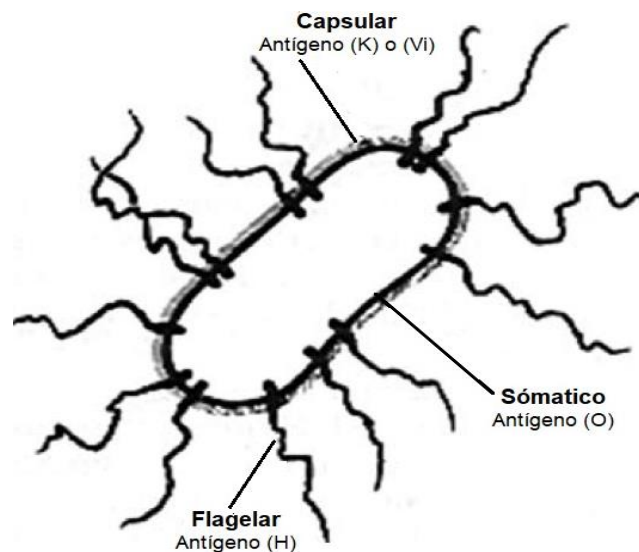


Figura 1. Estructura antigénica de *Salmonella* tomada de (Zúñiga y Abraham, 2006)

La mayoría de los serotipos de *Salmonella* pueden contener en su genoma a dos genes distintos que pueden codificar a dos proteínas flagelares (fase I y fase II) (Grimont y Weill, 2007).

Cada serotipo expresa una fase I específica en su antígeno flagelar (H) y la cual es responsable de su identidad inmunológica. En algunos casos, un serotipo puede solo expresar su fase I y es conocido como serotipo monofásico. Sin embargo, existen otros serotipos que se conocen como difásicos porque no contienen antígenos específicos para un serotipo, pueden ser común entre varios serotipos, debido a que alternan entre su fase I y su fase II (Zúñiga y Abraham, 2006).

El antígeno capsular (K) está compuesto por polisacáridos sensibles al calor que se localizan en la superficie de la capsula bacteriana y cuenta con un subtipo conocido como antígeno de virulencia (Vi), el cual solo se localiza en serotipos específicos de *Salmonella* como Paratyphi C, Dublín y Typhi. Este antígeno brinda resistencia contra la respuesta humoral del hospedero y, por lo tanto, es determinante en la virulencia de *Salmonella* Typhi (Eng *et al.*, 2015).

2.5. Taxonomía de *Salmonella* Spp

La nomenclatura de *Salmonella* se encuentra en constante evolución y por lo tanto puede llegar a ser confusa. Debido a esto, el Centro para Control y Prevención de Enfermedades (CDC) recomienda utilizar el sistema de nomenclatura para *Salmonella* empleado por la Organización Mundial de la Salud (WHO) y que se basa en la filogenia de las especies (Chalker y Blaser, 1988). De acuerdo con este sistema de clasificación, el género *Salmonella* cuenta con dos especies, *Salmonella enterica* y *Salmonella bongori*, basado en sus diferencias dentro de su secuencia 16S de ARN ribosomal (Coburn *et al.*, 2007).

Salmonella enterica se puede clasificar en seis subespecies (subgéneros), cada una de estas antecedida por un número romano (Eng *et al.*, 2015) (Cuadro 3); Los cuales comprenden a los subgéneros I (*S. enterica* subs. *enterica*), II (*S. enterica* subs. *salamae*), IIIa (*S. enterica* subs. *arizonae*), IIIb (*S. enterica* subsp. *diarizonae*), IV (*S. enterica* subsp. *houtenae*) y VI (*S. enterica* subsp. *indica*) (Cuadro 3). Generalmente, el nombre de los serotipos de la subsp. *enterica* (I) se establecen de acuerdo con la región geográfica donde fueron aislados por primera vez (Eng *et al.*, 2015).

La subespecie *S. enterica* subsp. *enterica* (I) cuenta con más de 2,500 serotipos descritos, por lo que es considerada como la subespecie más diversa. Los miembros de esta subespecie se encuentran predominantemente en mamíferos y se estima que contribuye al 99% de las infecciones causadas por *Salmonella* en humanos y en animales de sangre caliente.

Por lo contrario, el resto de las subespecies de *Salmonella enterica* y la especie *Salmonella bongori* han sido aisladas principalmente del ambiente y de animales de sangre fría.

Nomenclatura de <i>Salmonella</i>	<i>S. enterica enterica</i>	<i>S. enterica salamae</i>	<i>S. enterica arizonae</i>	<i>S. enterica diarizonae</i>	<i>S. enterica houtenae</i>	<i>S. enterica indica</i>	<i>S. enterica bongori</i>
Subgrupo de hibridación de ADN	I	II	IIIa Serotipos monofásicos	IIIb Serotipos difásicos	IV	VI	V
Subgéneros correspondientes de Kauffmann	I	II	III	III	IV		
Designación	La mayoría de los serotipos están en esta subespecie.	Designados por su estructura antigénica.	Anteriormente conocida como <i>Arizona</i>		Designados por su estructura antigénica. Solo los – no son nombrados.	Serotipos previamente incluidos en el subgrupo I o II	Anteriormente considerados serotipos atípicos del subgrupo I
Ocurrencia	Ocurrencia en humanos y animales.	Comúnmente en reptiles, rara vez en humanos.			Escasa; principalmente en reptiles.		

Cuadro 3. Clasificación, Designación y Ocurrencia de las Seis Subespecies de *Salmonella Enterica*. Modificado de (Bell,2002)

2.6. Epidemiología de *Salmonella enterica*

Los datos de vigilancia epidemiológica publicados por la CDC en los años 2006 – 2013 sugieren que *Salmonella* es el principal microorganismo transmitido por el consumo de alimentos; tan solo en Estados Unidos ocurren anualmente 1.35 millones de casos y 420 muertes por este microorganismo (Crump *et al.*, 2015).

A nivel mundial se estima que la fiebre tifoidea causa alrededor de 200,000 muertes y 22 millones de casos anuales al año (Figura 2), siendo los países en vía de desarrollo los más afectados debido

a la falta de saneamiento en agua utilizada para el consumo humano e irrigación de cultivos (Mogasale *et al.*, 2014).

Los serotipos no tifoideos de *Salmonella* tienen un amplio rango de hospederos; se estima que causan alrededor de 93 millones de infecciones entéricas y 155,000 muertes anualmente (Figura 3). Su prevalencia en África es mayor que en otras regiones geográficas del mundo (Balasubramanian *et al.*, 2019).

En México, su presencia se ha documentado desde hace décadas en muestras ambientales, clínicas y de alimentos siendo los principales serotipos reportados *S. Enteritidis*, *S. Typhimurium*, *S. Anatum*, *S. Agona*, *S. Meleagridis*, *S. Oranienburg*, *S. Derby*, *S. Infantis*, *S. Ohio* y *S. Havana* (Contreras-Soto *et al.*, 2018; Godínez *et al.*, 2020; Gutiérrez *et al.*, 2000; Jiménez *et al.*, 2014; Quiroz-Santiago *et al.*, 2009).

Salmonella es el patógeno transmitido por alimentos más reportado en México, según los datos publicados en el boletín epidemiológico del Sistema Nacional de Vigilancia Epidemiológica de México (SINAVE). En el año 2021 fue reportado un total de 65,085 casos, donde, un total de 43,054 de los casos asociados están asociados a serotipos no tifoideos, 16,530 casos de *Salmonella Typhi* y 5,501 casos de *Salmonella Paratyphi A* (SINAVE,2021) (Cuadro 4).

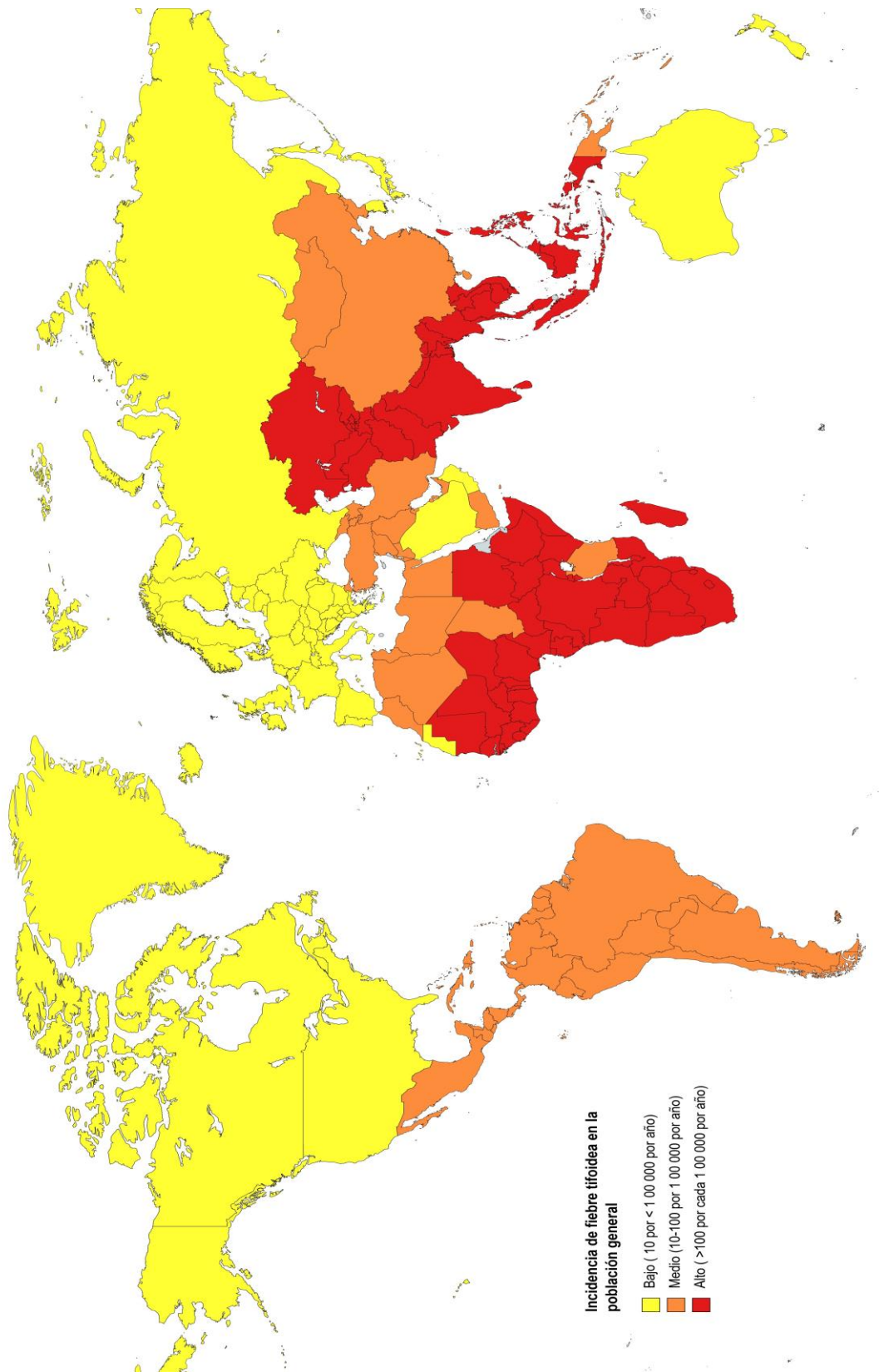


Figura 2. Incidencia de fiebre tifoidea a nivel mundial. Modificado de (Mogasale *et al.*, 2014).

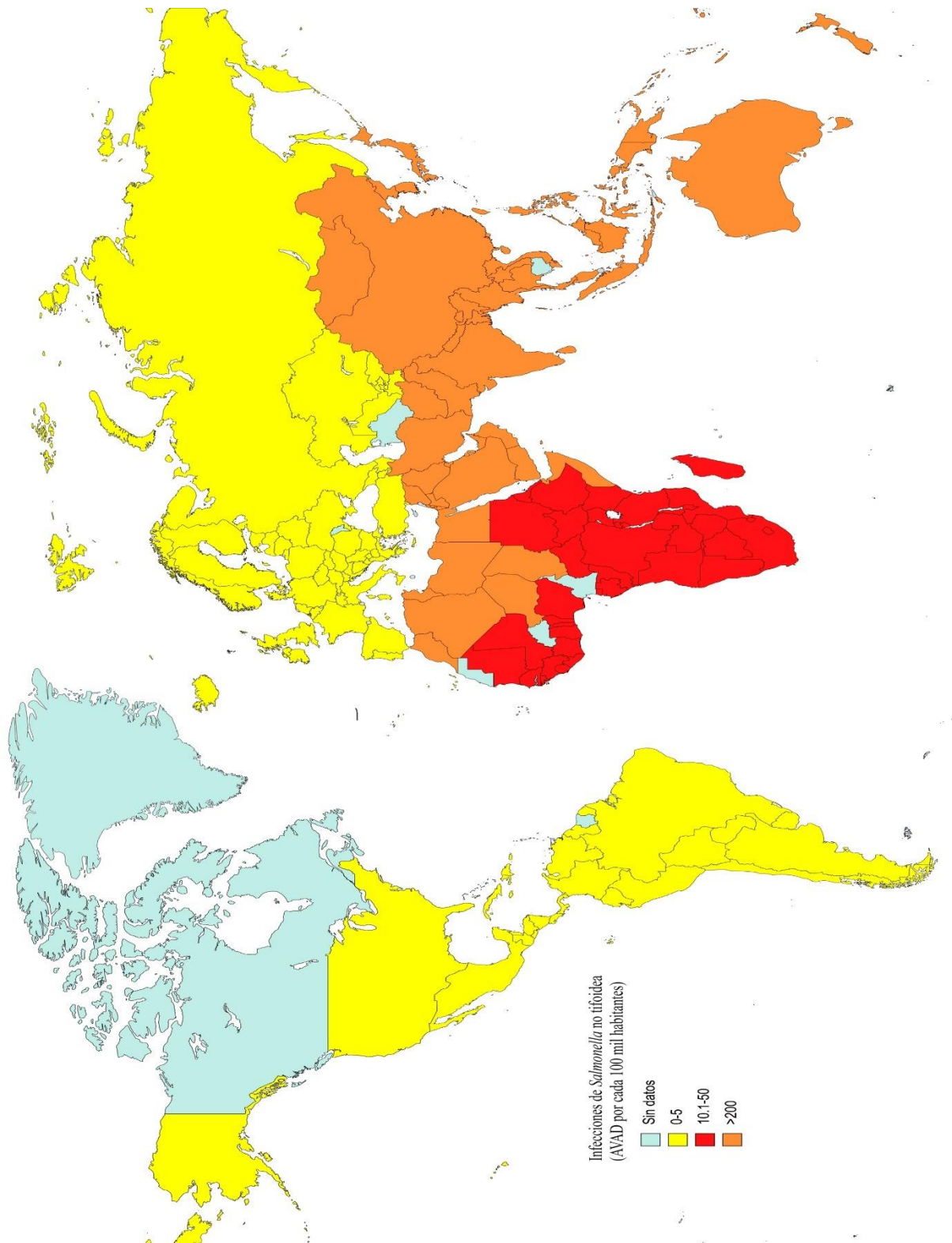


Figura 3. Distribución mundial de casos de *Salmonella* no tifoidea en el año 2010. Modificado de (Balasubramanian *et al.*, 2019).

Cuadro 4. Casos por entidad federativa de Fiebre Tifoidea, Fiebre Paratifoidea y Otras Salmonelosis hasta la semana 51 del 2021.

ENT. FED.	Fiebre Tifoidea CIE-10 ^a REV.				Paratifoidea CIE-10 ^a REV.				Otras Salmonelosis CIE-10 ^a REV.			
	Sem.	2021 Acum.		2020 Acum.	Sem.	2021 Acum.		2020 Acum.	Sem.	2021 Acum.		2020 Acum.
		M	F			M	F			M	F	
AGS.	-	25	24	85	-	-	-	7	-	66	104	187
B.C.	6	103	239	304	1	8	24	32	3	56	157	346
B.C.S.	1	6	15	13	-	6	3	6	-	48	66	67
CAMP.	-	24	48	135	-	6	24	46	2	64	160	64
COAH.	7	283	659	807	1	56	90	69	21	355	691	951
COL.	-	1	1	-	-	2	2	8	1	25	59	84
CHIS.	1	263	399	858	3	430	802	1,274	155	4,745	7,854	12,876
CHIH.	-	123	184	198	-	24	42	57	15	659	1,297	1,280
CDMX	-	5	5	154	-	3	5	17	4	123	147	303
DGO.	-	8	8	19	-	1	4	1	-	4	19	12
GTO.	1	100	172	311	-	31	69	63	2	182	293	534
GRO.	7	129	261	428	1	31	72	185	4	227	391	845
HGO.	-	6	7	11	-	-	5	1	1	39	57	136
JAL.	1	53	111	161	-	32	76	227	14	487	728	1,403
EDO.	-	64	83	83	7	9	27	40	23	376	640	784
MEX												
MICH.	6	260	442	519	-	87	152	259	13	333	580	971
MOR.	-	28	33	39	-	2	3	5	2	6	28	77
NAY.	-	2	5	6	1	21	32	44	14	465	911	2,132
N. L.	5	122	169	274	1	27	36	79	4	61	71	123
OAX.	-	78	90	248	-	33	75	138	7	330	476	947
PUEB.	-	27	35	66	-	18	31	70	42	1,103	1,405	2,549
QRO.	-	2	2	16	-	1	2	2	-	4	5	9
Q. ROO	1	8	6	2	-	42	8	47	11	259	529	661
S.L. P	-	21	67	148	-	32	39	30	4	53	104	287
SIN.	31	1,336	1,997	3,383	-	212	460	404	32	514	1,318	1,718
SON.	-	22	52	48	-	8	20	18	1	106	148	160
TAB.	18	477	732	1,457	2	229	364	568	18	1,090	1,810	3,108
TAMPS.	19	1,054	1,528	3,344	1	106	197	274	4	262	575	658
TLAX.	-	12	15	35	-	3	4	13	1	68	92	171
VER.	20	561	1,322	1,748	35	643	927	1,279	128	3,180	5,579	9,172
YUC.	-	3	8	14	-	1	12	39	1	50	99	184
ZAC.	55	920	1,436	1,616	4	20	27	199	3	126	252	255
TOTAL	179	6,126	10,155	16,530	57	2,124	3,634	5,501	530	15,466	26,645	43,054

Fuente: SINAVE/DGE/Salud 2021. Información preliminar, incluye casos probables

2.7. Especificidad de *Salmonella* por Hospederos

Si bien la mayoría de los serotipos de *Salmonella* son localizados en el ser humano, existen serotipos que son específicos de algunos hospederos y pueden permanecer sólo en una o en pocas especies animales.

En algunos casos, los serotipos de *Salmonella* tienen predilección por un solo hospedero por lo que se les conoce como serotipos con hospedero específico o restringido, tal es el caso de los serotipos Typhi, Paratyphi A, Paratyphi B y Paratyphi C, que infectan sólo al ser humano y se consideran como serotipos tifoideos. Estos organismos son el agente causal de la fiebre tifoidea y fiebre parafoidea, las cuales en su conjunto se conocen como fiebre enterica (Crump *et al.*, 2015). Otro ejemplo de serotipos restringidos a un solo hospedero son los serotipos Gallinarum y Pullorum que afectan únicamente a aves de corral y que pueden ocasionar pullorosis y fiebre tifoidea aviar (Hu *et al.*, 2019).

Los serotipos no tifoideos no presentan especificidad por algún hospedero por lo que son considerados como generalistas porque se encuentran colonizando un amplio rango de animales vertebrados (Crump *et al.*, 2015; Eng *et al.*, 2015).

Los principales serotipos no tifoideos son *S. Typhimurium* y *S. Enteritidis*, los cuales han sido encontrados tanto en muestras ambientales como en clínicas, sin embargo, se destaca también la alta presencia de otros serotipos no tifoideos como *S. Dublin*, *S. Heidelberg*, *S. Choleraesuis*, los cuales se estima que pueden llegar a causar infecciones muy parecidas e incluso más agudas que las que provoca *S. Typhimurium* en el ser humano (Crump *et al.*, 2015).



Figura 4. Ciclo de vida de *Salmonella* en ambientes externos y hospederos. Fuente: (Winfield y Groisman, 2003).

2.7.1. *Salmonella* en Ambientes Hospederos

Salmonella, un patógeno ampliamente estudiado, emplea estrategias hábiles para alcanzar e infectar a su hospedero. Su ruta hacia el interior del huésped involucra una serie de etapas cruciales que le permiten establecer una infección efectiva (Andrews *et al.*, 2010).

La principal fuente de infección por *Salmonella* en humanos es la ingesta de alimentos contaminados, como carne de res, cerdo, pollo, mariscos y huevos crudos o insuficientemente cocidos. Además, el consumo de vegetales y frutas frescas irrigadas con aguas contaminadas o tratadas con fertilizantes no seguros representa otro riesgo importante (Newell *et al.*, 2010).

Una vez que *Salmonella* ingresa al organismo a través de la ingesta de alimentos contaminados, este patógeno enfrenta una serie de desafíos para establecer una infección exitosa. En primer lugar, debe sobrevivir a la barrera gástrica, que presenta condiciones adversas como un bajo pH y ácidos gástricos que son letales para muchas bacterias. *Salmonella* puede resistir estas condiciones debido a su capacidad de neutralizar el ácido gástrico y mantenerse viable en el ambiente ácido del estómago (Barton *et al.*, 2021).

En los intestinos, *Salmonella* enfrenta condiciones aún más desafiantes. La falta de oxígeno, la presencia de sales biliares, péptidos antimicrobianos y la competencia por nutrientes y espacio con otros microorganismos residentes son obstáculos que debe superar (Barton *et al.*, 2021). *Salmonella* ha desarrollado mecanismos para resistir la acción de las sales biliares y evadir el sistema inmunológico en esta etapa (Spector y Kenyon, 2012).

Posteriormente, *Salmonella* comienza a invadir y a colonizar la mucosa intestinal del hospedero a través de células M, donde este patógeno es captado por una vacuola sin nutrientes. (Anderson y Kendall, 2017; Duprey *et al.*, 2014). Subsecuentemente, es liberado dentro de la submucosa intestinal donde es fagocitado y absorbido por macrófagos residentes de la mucosa, lo cual le permite escapar de las defensas humorales del hospedero (Spector y Kenyon, 2012).

Finalmente, *Salmonella* es absorbida por los macrófagos residentes de la mucosa, lo que le permite escapar de las defensas humorales del huésped y establecer una infección intracelular. Dentro de los macrófagos, *Salmonella* reside dentro de los fagosomas y utiliza una variedad de mecanismos de virulencia para evitar su degradación y multiplicarse intracelularmente, lo que le permite persistir y provocar la enfermedad (Anderson y Kendall, 2017).

2.7.2. *Salmonella* en Ambientes no Hospederos

Una vez que *Salmonella* es excretada del tracto gastrointestinal de sus hospederos por heces, este microorganismo es capaz de prosperar en ambientes no hospederos como suelo y agua superficial, donde puede permanecer durante largas temporadas enfrentándose constantemente a una serie de condiciones cambiantes en el ambiente como disminución de nutrientes, depredación, estrés osmótico, radiación solar, ultravioleta (UV), cambios en temperatura y pH (Pradhan *et al.*, 2003) (Figura 4)

Debido a lo anterior, los ambientes no hospederos son considerados como reservorios de *Salmonella* ya que facilitan su diseminación hacia nuevos hospederos. El papel que tienen en la transmisión de *Salmonella* se ha vuelto relevante para la salud pública debido al incremento de brotes asociados a reservorios no animales, por ejemplo, el consumo de agua contaminada y alimentos frescos mínimamente procesados como frutas y hortalizas (Liu *et al.*, 2018).

2.8. El Genoma de *Salmonella enterica*

El genoma de *Salmonella* tiene un contenido de G+C de ~ 52% mientras que el de AT es ~48%. (Baker y Dougan, 2007). Se encuentra constituido por un solo cromosoma de ADN de doble hélice circular con un tamaño aproximado de entre 4.5 a 5 Mb, el cual contiene aproximadamente entre 4000 a 5000 secuencias codificantes (CDSs), 75 a 85 ARNt y un número de plásmidos lineales o circulares variables (Alikhan *et al.*, 2018).

El cromosoma de todos los serotipos de *Salmonella* está compuesto por un genoma core (genes que están presentes en la mayoría de los serotipos) y un genoma accesorio (genes exclusivos y que están presentes sólo en algunos serotipos) (Brown *et al.*, 2021).

El genoma core abarca alrededor del 90% del genoma y está compuesto por genes ortólogos (genes que son similares entre si porque cuentan con un mismo origen evolutivo) los cuales son responsables de funciones esenciales para *Salmonella* como replicación transcripción, traducción, producción de energía y síntesis de moléculas esenciales como aminoácidos (Fu *et al.*, 2015).

Mientras que el genoma accesorio contiene a los genes altamente especializados que solo están en ciertos serotipos de *Salmonella*, estos se dividen en tres categorías: SPIs, profagos y operones fimbriales (Baker y Dougan, 2007).

Los SPIs (Islas de Patogenicidad de *Salmonella*) son secuencias de ADN que son contenidas generalmente en plásmidos y que son insertadas en el genoma de *Salmonella* durante la transferencia horizontal de genes vía ARNt. Algunas SPI codifican genes con funciones importantes como sistemas de secreción y otros juegan un papel importante durante la determinación del rango de hospederos y el grado de virulencia entre los diferentes serotipos (Amavisit *et al.*, 2003).

Los profagos son el genoma o una parte del genoma de un bacteriófago atenuado, el cual es insertado en el cromosoma de *Salmonella*, confiriéndole virulencia a la bacteria e incrementando su potencial patogénico hacia su hospedero (Duprey *et al.*, 2014).

Los operones fimbriales, los cuales determinan el tipo de fimbria en la superficie de la célula bacteriana y difieren ampliamente entre los serotipos de *Salmonella*. Cada serotipo tiene una combinación diferente de operones fimbriales y su distribución es un reflejo de la adaptación de los serotipos a diferentes nichos ecológicos (Dufresne *et al.*, 2018).

El genoma de *Salmonella* cuenta con una alta densidad de regiones codificantes o funcionales, las cuales reflejan una relación entre el tamaño del genoma y el número de genes, sin embargo, también cuenta con secuencias intergenicas y no funcionales las cuales contienen genes que se han silenciado por mutaciones y a pseudogenes (Abby y Daubin, 2007; Bobay y Ochman, 2017; Bolotin y Hershberg, 2015).

Los genomas de *Salmonella* a su vez albergan un amplio número de pseudogenes que se cree que están asociados a su adaptación a un nuevo nicho ecológico; se estima que abarcan aproximadamente del 3 al 4% de las secuencias codificantes en los genomas y se ha detectado su presencia en serotipos generalistas como en hospederos específicos, aunque su presencia es mayor en serotipos asociados a un hospedero como *Salmonella Gallinarum* o *Salmonella Pullorum* (Holt *et al.*, 2009).

2.9. Descubrimiento de Pseudogenes

En el año de 1977 Jacq et al. descubrieron en *Xenopus laevis* una copia trunca de la secuencia del gen 5S ARNr con una función aparente a la cual llamaron “pseudogen” (Jacq *et al.*, 1977). Posterior a su descubrimiento, los pseudogenes han sido identificados en los genomas de la mayoría de los organismos y se ha determinado que su número varía entre las especies y no presenta una correlación con la complejidad del organismo (Pink *et al.*, 2011).

Un pseudogen se define como una copia defectuosa de un gen funcional que ha perdido su capacidad de codificar un ARN o una proteína.

En el pasado, los pseudogenes eran considerados como ADN fósil o reliquias genómicas de genes antiguos los cuales no poseen ninguna función debido a que presentan mutaciones que no les permiten codificar proteínas. (Gil y Latorre, 2012) (Figura 5). Recientemente, se ha descubierto evidencia sobre una posible funcionalidad de algunos pseudogenes en la regulación de genes (Aguirre-Sanchez *et al.*, 2021; Dufresne *et al.*, 2018; Feng *et al.*, 2022; Kovalenko y Patrushev, 2018). Esto ha sido descrito en ciertos organismos eucariotas (Cheetham *et al.*, 2020; Pei *et al.*, 2012; Xu y Zhang, 2016) y últimamente en organismos procariotas (Feng *et al.*, 2022), por lo que la identificación de pseudogenes puede incrementar el conocimiento sobre la historia evolutiva de las especies y mejorar la anotación de genomas.

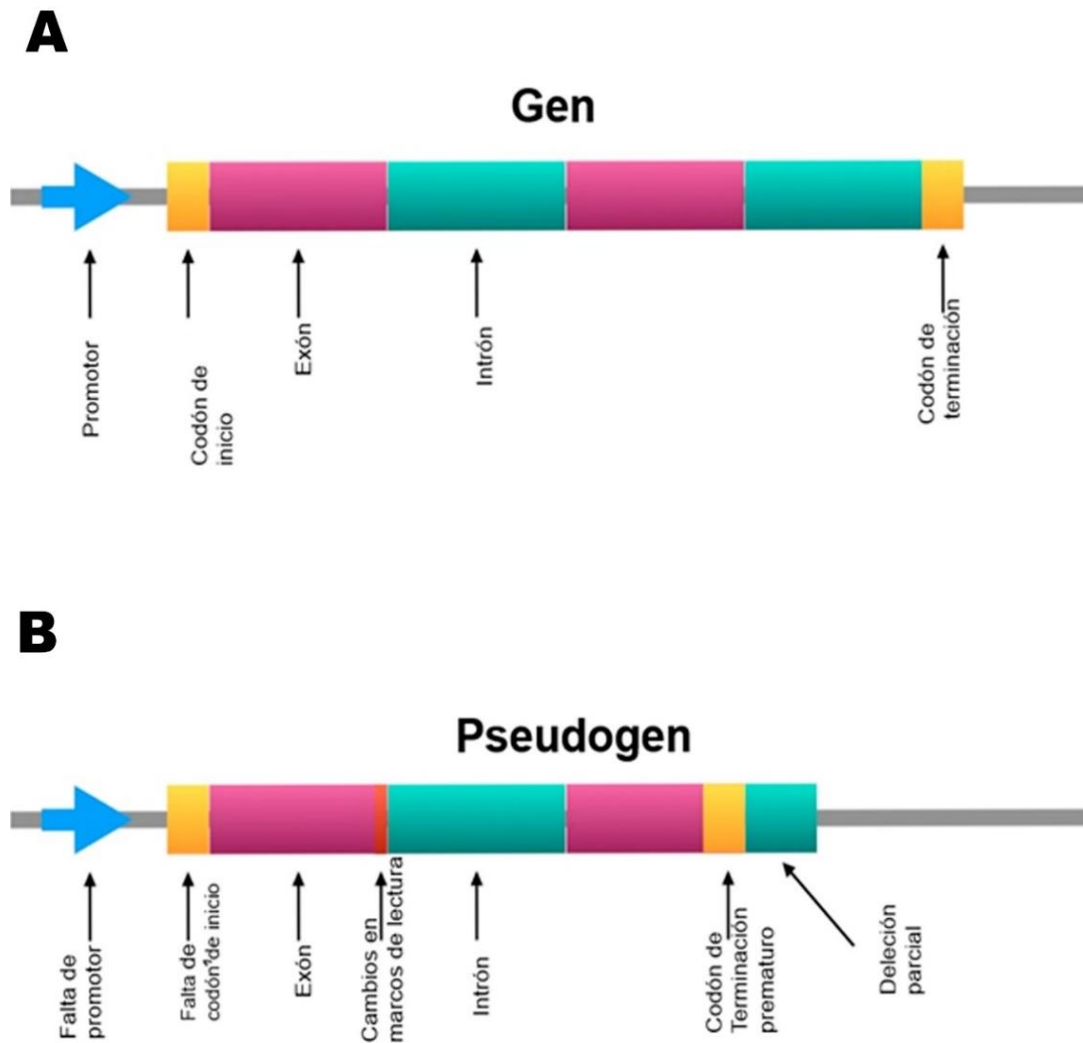


Figura 5. Regiones de un gen y pseudogen. A| Estructura y partes de un gen. B| Principales mutaciones que presenta un gen en sus regiones durante el proceso de pseudogenización.

2.10. Definición y Clasificación de Pseudogenes

El término pseudogen se acuña a cualquier secuencia que es similar a un gen pero que presenta mutaciones en regiones codificantes o reguladoras que impiden su transcripción como codones de inicio/termino prematuros, truncamientos y cambios en marcos de lectura (Figura 5). Como resultado, el tamaño de la secuencia se ve afectado más del 80% y cuenta con un porcentaje de identidad mayor a 20% con su gen homólogo (Chu *et al.*, 2021; Lerat y Ochman, 2005).

Los pseudogenes son clasificados por su mecanismo de origen. Las categorías más comunes de pseudogenes son los pseudogenes procesados y los no procesados. Mientras que los pseudogenes unitarios representan a la minoría (Cheetham *et al.*, 2020).

2.10.1. Pseudogenes no Procesados

Los pseudogenes duplicados o mejor conocidos como pseudogenes no procesados, se originan como su nombre sugiere, de eventos parciales o completos de duplicación del ADN de genes (Pink *et al.*, 2011). Lo anterior ocurre durante el proceso de replicación del ADN donde existe un entrecruzamiento desigual entre dos cromosomas homólogos (Figura.6). Los eventos de duplicación genética han demostrado ser uno de los principales mecanismos precursores de familias de genes funcionales, por ejemplo, las subunidades α -globina and β -globina que codifican a la hemoglobina humana. Sin embargo, si la duplicación es truncada o si no es copiada fielmente se producen mutaciones que generan cambios en los marcos de lectura y otras anomalías que dan lugar a la formación de un pseudogen duplicado no funcional (Goodhead, 2015).

2.10.2. Pseudogenes Procesados

Los pseudogenes procesados se encuentran en largas proporciones en genomas eucariotas y se consideran el tipo de pseudogen más común en estos organismos (Podlaha y Zhang, 2010). Se caracterizan por tener secuencias sin intrones y muy similares a la secuencia de ARNm de su gen parental, además de contar con una poliadenilación en el extremo 3'. Se cree que su origen proviene de secuencias de ARNm que fueron retro transcriptas (transcripción inversa) en ADN complementario (ADNc) y que se insertan en diversas regiones del genoma.

2.10.3. Pseudogenes Unitarios.

A diferencia de los dos tipos anteriores, los pseudogenes unitarios no se derivan del resultado de un evento de inserción o duplicación, se originan cuando un gen original es inactivado a través de mutaciones espontaneas por lo que no se mantiene ninguna copia funcional del gen (Cheetham *et al.*, 2020). (Figura.6)

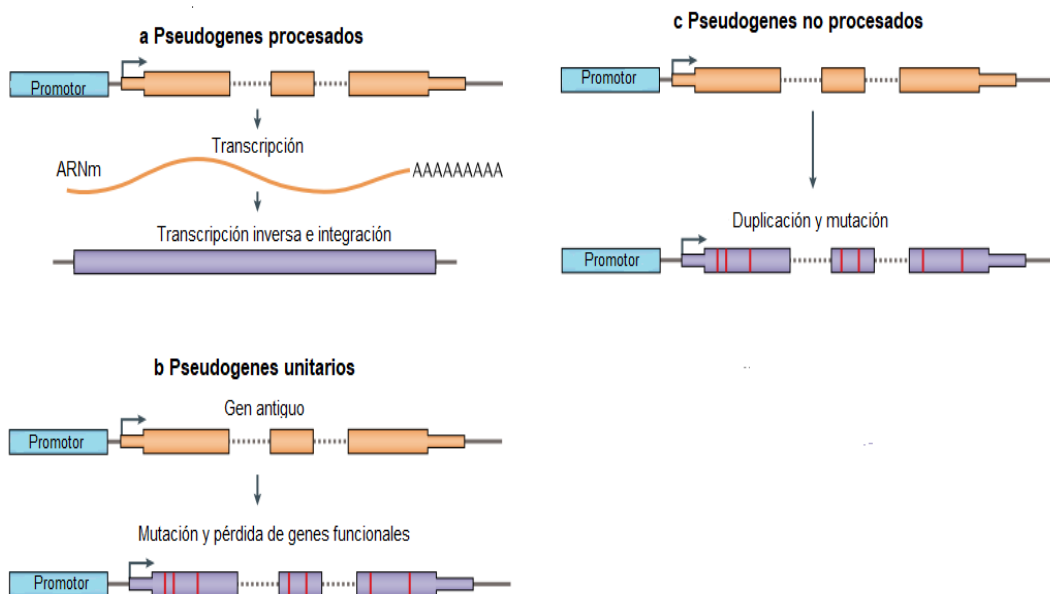


Figura.6. Principales clases de pseudogenes. a| Los pseudogenes procesados surgen de la transcripción inversa y la integración de un ARNm procesado. b| Los pseudogenes no procesados se originan a partir de duplicaciones de genes que acumulan mutaciones, lo que impide su traducción. c | Los pseudogenes unitarios se derivan sin duplicación de un gen que codifique proteínas, se caracteriza por ser un gen de origen ancestral que ha perdido la posibilidad de codificar proteínas. Fuente: (Cheetham *et al.*,2020).

2.11. Proceso de Pseudogenización

Los pseudogenes se originan a partir de cualquier secuencia en el genoma. Siendo los genes constitutivos, los genes expresados en la línea germinal y los genes que participan en regulaciones metabólicas básicas los que presentan mayor cantidad de eventos de pseudogenización (Li *et al.*,

2013). Esto es debido a que estos genes son altamente expresados por lo que es más probable que acumulen mutaciones o sean retro transpuestos en el genoma.

Un factor que influye en la generación de un pseudogen es el tamaño de un gen. Influye en la generación de un pseudogen, se ha demostrado que genes que codifican proteínas de cadena larga tienden a producir pseudogenes no procesados mientras que genes que codifican proteínas de cadena corta tienden a producir pseudogenes procesados (Li *et al.*, 2013).

Asimismo, es importante destacar que el número de pseudogenes es variable entre las especies y no necesariamente está correlacionado con el grado de complejidad del organismo (Tutar, 2012).

2.12. Evolución y Función de los Pseudogenes

La pseudogenización se refiere al proceso evolutivo en el cual un gen funcional se convierte en un pseudogen durante el curso de la evolución. Algunas veces puede llegar a ser adaptativa, a pesar de que la mayoría de los eventos de pseudogenización son neutrales y ocurren mediante la acumulación azarosa de mutaciones en genes con limitaciones funcionales. Este proceso ocurre cuando la ecología de una especie cambia repentinamente y una gran cantidad de genes funcionales se convierten en no funcionales, por ejemplo, una bacteria de vida libre se convierte en un organismo simbiótico (Chu *et al.*, 2021; Kuo y Ochman, 2010)

Las bacterias de vida libre tienen un genoma relativamente grande comparado con el de una bacteria intracelular. Sin embargo, cuando existe un cambio de nicho ecológico las bacterias de vida libre se ven obligadas a reducir el tamaño y la capacidad codificante de su genoma, pero conforme el organismo se adapta a un entorno intracelular, es posible que las restricciones selectivas en algunos genes se relajen debido a la protección o los nutrientes que brinda la célula hospedera (Goodhead, 2015) y puedan ser propensos al mecanismo de pseudogenización por mutación. Una mayor especialización por un hospedero da como resultado la eliminación de secuencias redundantes y a una reducción en el tamaño del genoma (Goodhead, 2015).

Lo anterior es posible ejemplificarlo en los serotipos de *Salmonella enterica*. Este microorganismo se caracteriza por tener aproximadamente a 2600 serotipos que son generalistas y un poco variedad de serotipos hospederos específicos. (Eng *et al.*, 2015)

La pseudogenización es considerada como una característica distintiva en los serotipos restringidos a un hospedero, esto se debe a que el número de pseudogenes es mayor en los serotipos restringidos a un hospedero que en los serotipos que son generalistas. Por ejemplo, en serotipos generalistas como *S. Enteritidis* y *S. Typhimurium* se ha reportado la presencia de 63 y 113 pseudogenes, respectivamente. Mientras que en serotipos restringidos a un solo hospedero como *S. Gallinarum* y *S. Typhi* se han encontrado 309 y 204 pseudogenes, respectivamente. Lo cual representa una mayor cantidad de pseudogenes comparado con los serotipos generalistas. (Goodhead, 2015; Kuo y Ochman, 2010).

2.13. Identificación de Pseudogenes

La identificación de pseudogenes puede realizarse mediante diversos métodos, como genómica comparativa, transcriptómica y proteómica.

El método más común para identificar pseudogenes es a través de la genómica comparativa. Es decir, comparar un genoma de interés contra un genoma objetivo con el objetivo de encontrar regiones que acumulen mutaciones que previenen la producción de una proteína funcional.

Generalmente, los pseudogenes presentan una serie de características que permiten distinguirlos de genes funcionales, tales como codones de terminación e inicio prematuros, cambios de marcos de lectura, deleciones que causan una pérdida mayor del 80% en la longitud de la secuencia proteica del gen o una alteración mayor del 20% en la secuencia en el caso de inserciones (Lerat y Ochman, 2005).

Los métodos de identificación de pseudogenes con transcriptómica son realizados mediante la búsqueda de transcritos de ARN que sean similares a genes funcionales pero que no tengan la habilidad de producir proteínas funcionales debido a mutaciones puntuales. Estos transcritos a menudo se encuentran truncados o contienen mutaciones que impiden que sean traducidos a proteínas funcionales. Estas proteínas son degradadas rápidamente por la célula y son difíciles de identificar, pero los avances en técnicas proteómicas han hecho que la identificación de pseudogenes derivados de proteínas sea más sencillo (Cheetham, 2020).

Los métodos proteómicos consisten en la identificación de proteínas que son similares a proteínas

pero que tienen una pérdida de dominios funcionales o motivos necesarios para que la proteína funcione correctamente.

Hoy en día la identificación de pseudogenes puede ser desafiante pero el desarrollo de la bioinformática y técnicas modernas en genómica y proteómica lo han hecho posible que el identificar pseudogenes sea más rápido y sencillo.

2.13.1. Exploración de los Pseudogenes en Genomas Bacterianos

Anteriormente se pensaba que los pseudogenes eran inusuales en los genomas bacterianos debido al costo metabólico que representa mantener secuencias no codificantes.

No obstante, el desarrollo de herramientas bioinformáticas ha permitido realizar análisis genomas bacterianos donde se ha identificado pseudogenes en la mayoría estos, los cuales representan al 1 al 5 % del tamaño total del genoma (Liu *et al.*,2004), con excepción de genomas pertenecientes a especies patógenas donde se han localizado en grandes cantidades de pseudogenes.

En el año 2004 Liu y colaboradores modificaron un pipeline de identificación de pseudogenes en eucariotas para realizarlo en genomas procariotas. Durante este trabajo, se realizó un análisis para determinar la proporción de pseudogenes presentes en 64 genomas que incluyeron a arqueas, bacterias patógenas y no patógenas. Entre las cepas analizadas se encontraba *Salmonella enterica* Typhi. Se encontró que alrededor de 7000 pseudogenes estaban relacionados con la transposición del ADN y a la respuesta ambiental de las bacterias. Finalmente, se concluyó que los pseudogenes no eran poco frecuentes como se pensaba y que abarcan aproximadamente del 1 al 5 % de los genomas analizados (Liu *et al.*,2004).

Un estudio conducido por Lerat y colaboradores en el año 2005 utilizando la herramienta Psi-Phi para la identificación de pseudogenes describió la presencia de estos en genomas de *Salmonella enterica*, así como también en las especies *Shigella flexneri*, *Yersinia pestis* y *Escherichia coli* (Lerat y Ochman, 2005). Como resultado, se encontró que el número de pseudogenes fue variable entre las especies, sin embargo, se destaca que las especies de origen clínico presentaban una mayor

cantidad de pseudogenes, lo cual se correlaciona con el estilo de vida de la bacteria y, además, el número de pseudogenes difiere considerablemente entre especies filogenéticamente relacionadas. Posteriormente, Lerat y Ochman en el año 2006 realizaron la identificación de pseudogenes con el programa Psi-Phi en 11 genomas pertenecientes a las especies *Staphylococcus*, *Streptococcus*, *Yersinia* y *Vibrio*, de los cuales se incluyó al menos una cepa de origen clínico con la hipótesis de que los pseudogenes de las especies clínicas acumulan más pseudogenes en comparación con las especies que se encuentran en el ambiente.

Entre los hallazgos de los autores, se encontró que los pseudogenes componen alrededor del 1 al 8% de los genomas analizados y que se encuentran relacionados con genes con funciones biológicas esenciales como el procesamiento, reparación y replicación de nucleótidos.

Además, debido a que en trabajos previos se identificaban a ORF cortos como pseudogenes y a inconsistencias en la definición de que es un pseudogen, los autores fijaron un criterio que considera a un gen como pseudogen si el 20% la secuencia del gen se ve afectada. Actualmente este criterio es la base para identificar a un pseudogen en procariontas

Liao y colaboradores en el año 2019 realizaron un análisis comparativo de 90 aislados de *Salmonella enterica* pertenecientes a los serotipos Dublin (n=21), Newport (n=32) y Typhimurium (n=18) obtenidos de heces de ganado bovino y de pacientes clínicos. Para llevar a cabo los autores realizaron una anotación del genoma con Prokaryotic Genome Annotation Pipeline de la NCBI y de igual manera se identificó el número de pseudogenes en cada serotipo. Se concluyó que los serotipos de *Salmonella enterica* pueden sobrevivir durante largos periodos en diferentes cuerpos de agua debido a la presencia de pseudogenes asociados con funciones celulares importantes para el metabolismo anaeróbico como utilización de etanolamina, precursores de la síntesis de vitB12 o actuando como donantes y aceptores de electrones (Liao *et al.*,2020).

En el año 2021, Chu y colaboradores realizaron un trabajo donde aislaron a bacterias pertenecientes al filo Roseobacter de tres nichos ecológicos distintos en las costas de Hong Kong. Los autores emplearon el algoritmo Psi-Phi modificado para la identificación de pseudogenes en los genomas, donde determinaron que la pseudogenización está correlacionada con el cambio de nicho en las bacterias y que la pérdida de genes por pseudogenización contribuye a la adaptación y a la diversificación ecológica de Roseobacter.

Los autores concluyeron que el mecanismo de pseudogenización puede activarse o no dependiendo del nicho donde se encuentre la bacteria y, además, puede permitir que bacterias en ambientes

acuáticos respondan rápidamente a los cambios ambientales y se adapten a nuevos hábitats por lo que los pseudogenes son un contribuyente importante a la variación genética y diversificación ecológica por lo que se espera observar este mismo mecanismo en bacterias generalistas (Chu *et al.*, 2021).

2.14. Justificación

Salmonella, es una bacteria patógena con importantes implicaciones tanto clínicas como ambientales, ya que puede originar infecciones graves en humanos como animales, representando así un riesgo para la salud pública.

Su capacidad para sobrevivir y adaptarse en ambientes no hospederos plantea varias incógnitas sobre que favorecen su supervivencia en estos entornos, y una de estas características son los pseudogenes encontrados en su estructura genómica.

Los pseudogenes son secuencias de ADN que han perdido su capacidad de codificar proteínas funcionales debido a mutaciones puntuales. Estas secuencias son producto de la evolución y ofrecen información valiosa sobre la historia evolutiva de *Salmonella*. Por lo tanto, identificar y analizar estos pseudogenes en cepas de *Salmonella* aisladas del ambiente puede contribuir a incrementar el conocimiento sobre la biología de esta bacteria, así como comprender mejor los eventos evolutivos que han facilitado su colonización en ambientes no hospederos. y

Además, la identificación pseudogenes en cepas ambientales proporciona información sobre las adaptaciones genéticas que han ocurrido en respuesta a diferentes condiciones del entorno. Esto nos permite identificar qué rutas metabólicas desempeñan un papel crucial en la supervivencia y propagación de *Salmonella* en el ambiente. esto puede ayudar a identificar que rutas metabólicas juegan un papel clave en la supervivencia y propagación de *Salmonella* en el ambiente.

3. HIPÓTESIS

1. Se encontrarán al menos 100 pseudogenes en los serotipos Weltevreden, Saintpaul y Oranienburg.
2. Las principales categorías de pseudogenes serán metabolismo, reparación de ADN, transporte de metabolitos, motilidad y señalización.
3. Los pseudogenes identificados se asociarán a la respuesta adaptativa de Weltevreden, Saintpaul y Oranienburg a ambientes no hospederos.

4. OBJETIVOS

4.1. Objetivo General

Determinar pseudogenes específicos para los serotipos Weltevreden, Saintpaul y Oranienburg que contribuyen a la adaptación ecológica de *Salmonella enterica* a un ambiente no hospedero.

4.2. Objetivos Específicos

Identificar y caracterizar a los pseudogenes presentes en las secuencias genómicas de los serotipos Weltevreden, Saintpaul y Oranienburg de *Salmonella enterica* aislados del río Culiacán utilizando herramientas bioinformáticas y genómica comparativa.

Determinar los pseudogenes que se asocian a la respuesta adaptativa de *Salmonella enterica* en ambientes no hospederos.

5. MATERIALES Y MÉTODOS

5.1. Obtención De Secuencias Objetivo

Se analizaron un total de 84 secuencias ensambladas y pertenecientes a tres serotipos de *Salmonella enterica* (Cuadro 5), las secuencias fueron aisladas del río Culiacán y fueron proporcionadas por el Laboratorio Nacional para la Investigación en Inocuidad Alimentaria (LANIA).

Cuadro 5. Relación de secuencias objetivo que fueron empleadas durante la identificación y caracterización de pseudogenes en serotipos de *Salmonella enterica*.

Serotipo	Numero de secuencias	Origen
Oranienburg	40	Ambiental
Saintpaul	37	
Weltevreden	7	

5.2. Obtención de Secuencias de Consulta

Las secuencias de referencia para el comparativo genómico se descargaron de la base de datos NCBI Pathogen Detection (<https://www.ncbi.nlm.nih.gov/pathogens/>) y los criterios de selección fue secuencias con origen clínico y ensamble completo (Cuadro 6).

Cuadro 6. Genomas de referencia descargados de NCBI Pathogen Detection (<https://www.ncbi.nlm.nih.gov/pathogens/>).

Serotipo	ID secuencia	NCBI BioSample	Origen	Ensamble
Saintpaul	FDAARGOS373	SAMN07312417	Clínico	Completo
Weltevreden	2511STDY5712388	SAMEA2336235	Clínico	Completo
Oranienburg	CFSAN001285	SAMN01805331	Clínico	Completo

5.3. Identificación De Pseudogenes

La identificación y filtrado de pseudogenes se llevó a cabo de acuerdo con la metodología propuesta por Lerat y colaboradores en el año 2005, la cual consiste en la utilización de Psi-Phi, una suite de programas diseñados para la identificación de pseudogenes a través de análisis comparativos con genomas relacionados. Los algoritmos se dividen en dos módulos escritos en el lenguaje de programación Pearl y bioperl que han sido probados previamente en los sistemas operativos Unix (SunOS 5.8) y GNU/Linux.

5.3.1. Prerrequisitos

5.3.1.1. Creación de pseudocromosoma. El programa Psi-Phi solo trabaja con un pseudo cromosoma. Sin embargo, actualmente la mayoría de las secuencias que se generan del genoma bacteriano no están ensambladas en un solo cromosoma, por lo que es necesario concatenar a todos los scaffolds en un solo pseudocromosoma.

5.3.1.2. Anotación funcional de secuencias. Es necesario que tanto como los genomas de secuencias objetivo como los de referencia sean anotados funcionalmente. Esto con el objetivo de obtener las secuencias proteicas y generar una base de datos Gen Bank para cada secuencia. Lo anterior fue realizado con el programa Prokka v5.

5.3.1.3. Creación y alineación de bases de datos. Para la creación y alienación de bases de datos de nucleótidos se utilizaron dos herramientas de BLAST: makeblastdb para la creación de una base de datos de los nucleótidos encontrados en la secuencia objetivo y tBLASTn para crear una alineación entre la base de nucleótidos construida con makeblastdb contra la secuencia de proteínas de consulta.

5.4. Predicción de Pseudogenes

5.4.1. Primer Módulo De Psi-Phi

El primer módulo consistió en el análisis del archivo de salida de tblastn (Figura 7), para lo cual se determinó la longitud de los CDS (regiones codificantes) de la secuencia de consulta mediante la búsqueda en el archivo Genbank del genoma de consulta. Como criterio, se fijó recuperar como candidatos a todos los genes que tuvieran una pérdida de más del 20% en su secuencia proteica y cuyo porcentaje de identidad con la secuencia de consulta sea mayor al 79%. En caso de que existieran dos CDS que se solaparan, se fijó el criterio de distancia de 300 nucleótidos para que se consideraran como regiones diferentes o como una sola.

Anotación funcional de secuencias.

```
(base) karla@karla-laptop:~/Documents/prueba10/Saintpaul/CFSAN04173/C-SAI08001$ perl ./Psi_Fi_module1_2.pl
***** Psi_Fi_module1_2 (version 0.2) *****
Enter the name of the tblastn file (format m=8)
tCSAI08001-q4173.tblastn
46593

Enter the name of the target species (as it appears in the tblastn file)
Salmonella_CSAT08001

Enter the name of the query genome file (genbank format)
CFSAN4173.gbk

!!! WARNING: the protein ids in the tblastn file (first column) must be the same as the protein ids in the genbank file (/protein_id=) !!!

Enter the name of the output file
sortie.out
***** Parameters *****
A    minimal percentage of identity: 79
B    maximal percentage of identity: 100
C    distance to merge: 300

Select the parameter you want to change or type Y

Y
running...
```

Figura 7. Captura de pantalla del módulo 1 del algoritmo Psi-Phi. En esta Figura se visualizan los criterios fijos que son necesarios para el funcionamiento del algoritmo, así como las etapas en las que son necesarios los archivos generados durante los prerequisites.

5.4.2. Segundo Modulo de Psi-Phi

Durante la ejecución del segundo modulo se realizó una comparación del archivo de salida del primer módulo de Psi-Phi contra el archivo de Genbank de la secuencia objetivo. Para lograr lo anterior, el algoritmo identifica y retiene aquellas secuencias cuyas características que cumplieron con los criterios seleccionados para asignar a pseudogenes potenciales (Figura 8).

Los criterios seleccionados fueron:

El valor E o valor esperado de BLAST: Parámetro que describe el número de veces en que se espera encontrar una secuencia por casualidad en una base de datos. Es significativo cuando su valor es igual o cercano a cero.

Para la utilización de este segundo modulo se fijó en BLAST un valor $E < 10^{-15}$ el cual permite identificar la presencia de codones prematuros como e identificar pseudogenes potenciales.

Longitud de secuencia: Perdida al menos 20% de la secuencia de aminoácidos originada por mutaciones puntuales.

Presencia de codones de terminación prematuros

Radio o distancia entre la longitud del gen objetivo y del gen de consulta (Mínimo 30% y máximo 80%)

```
karla@karla-laptop:~/Documents/prueba10/Saintpaul/CFSAN04173/C-SAI00001$ perl ./Psi_Fi_module2.pl
***** Psi_Fi_module2 *****
Enter the name of the file
sortie.out
Enter the name of the target genome file (genbank format)
CSAI00001.gbk
Enter the name of the out file
PseudoCSAI00001.txt
Enter the name of the genome bank (fasta format)
genome_CSAI00001.fasta
***** Parameters *****
A   maxinal e_value: 1e-15
B   maxinal percentage of the query size: 0.8
C   mininal percentage of the query size (match): 0.3
D   maxinal percentage of the query size (match): 0.8
Select the parameter you want to change or type Y
Y
running...
█
```

Figura 8. Captura de pantalla del módulo 2 del algoritmo Psi-Phi. En esta Figura se visualizan los criterios fijos que son necesarios para el funcionamiento del algoritmo, así como las etapas en las que son necesarios los archivos generados durante los prerrequisitos.

Finalmente, se obtuvo un archivo de salida que contiene 12 campos separados por tabulación:

Campo 1: e_value

Cadena en el genoma objetivo (+ y -)

Inicio de la coincidencia en el genoma objetivo.

Fin de la coincidencia en el genoma objetivo,

Inicio del gen en el genoma objetivo (si la coincidencia se superpone a un CDS ya anotado).

Fin del gen en el genoma objetivo (si la coincidencia se superpone a un CDS ya anotado)

Nombre del gen objetivo (si la coincidencia se superpone a un CDS ya anotado)

Función del gen objetivo (si la coincidencia se superpone a un CDS ya anotado).

Número de identificación de proteína del gen objetivo (si la coincidencia se superpone a un CDS ya anotado)

Categoría (Descritas abajo)

Relación entre la longitud del gen objetivo y la del gen de consulta (si la coincidencia se superpone a un CDS ya anotado)

A continuación, se desglosan 11 categorías de pseudogenes que son identificadas con el algoritmo Psi-Phi, las categorías con * no son pseudogenes:

Pseudogen identificado

Mala anotación: Presencia de un codón de terminación interno.

Mala anotación: Secuencia corta en el genoma objetivo, codón de terminación interno.

*En CDS, pero match muy corto.

Mala anotación: secuencia corta en el genoma objetivo y en el match.

*Mala anotación: Secuencia corta en el genoma objetivo e (pero $>X\%$ x consulta); codón de terminación interno

* Secuencia corta en el genoma objetivo, pero probablemente no es un pseudogen

Intergenico

* Intergenico, pero probablemente no es un pseudogen.

Intergenico, pero match corto (tamaño del match / tamaño del gen de consulta)

5.4.3. Anotación COG

Para clasificar a los genes asociados a pseudogenes según su función biológica se realizó una anotación funcional con ayuda de la base de datos Clusters of Orthologous Groups of Proteins (COG) de NCBI, para clasificar a las proteínas existentes en secuencias genómicas en base a su ortología y de acuerdo a los mecanismos en que están involucrados los productos de esos genes tales como almacenamiento y procesamiento de la información, procesos celulares y de señalización y metabolismo (Tatusov, 2000); estos mecanismos son clasificados mediante su función con una clave en letras mayúsculas que se ejemplifica a continuación:

6. RESULTADOS Y DISCUSIÓN

6.1. Identificación de Pseudogenes

84 secuencias de *Salmonella enterica* de origen ambiental y clínico pertenecientes a los serotipos Oranienburg, Saintpaul y Weltevreden (40, 37 y 7, correspondientemente) fueron analizadas en busca de pseudogenes con el objetivo de esclarecer y contribuir al entendimiento de los mecanismos de adaptación y evolución de este patógeno en ambientes no hospederos.

Se detectó a *Salmonella* Oranienburg como el principal serotipo que alberga el mayor número de pseudogenes con 169, seguido de Weltevreden con 70, y, por último, Saintpaul con 62 (Figura 9). En todos los casos la gran mayoría de los pseudogenes correspondieron a la categoría de proteínas hipotéticas y desconocidas. Adicionalmente, se localizó la presencia de secuencias de profagos, pero no fueron tomadas en cuenta para este estudio, ya que corresponde a material genético adquirido por mecanismos de transducción y no es propiamente del genoma bacteriano.

Considerando la relación entre número de pseudogenes con respecto al tamaño de genoma por serotipo (Cuadro 7), *S.* Oranienburg presentó mayor relación con un 3.63% del tamaño total de su genoma, seguido de Weltevreden con 1.47% y Saintpaul con 1.32%.

Asimismo, se identificaron a los principales eventos que originaron la presencia de pseudogenes (Figura 10). El cambio en los marcos de lectura abiertos fue el mecanismo que más generó pseudogenes entre los tres serotipos Oranienburg, Saintpaul y Weltevreden, produciéndose 34, 22 y 12 eventos, respectivamente.

En cuanto a la incidencia de un segundo mecanismo producto de pseudogenes, se encontraron diferencias entre los serotipos. Por ejemplo, para el caso de Oranienburg, la aparición de codones de terminación prematuros ocupó el segundo lugar (29 eventos). Mientras que para los serotipos Weltevreden y Saintpaul la aparición de genes trancos fue el segundo mecanismo de pseudogenización más importante con 15 y 7 eventos, respectivamente (Figura 10).

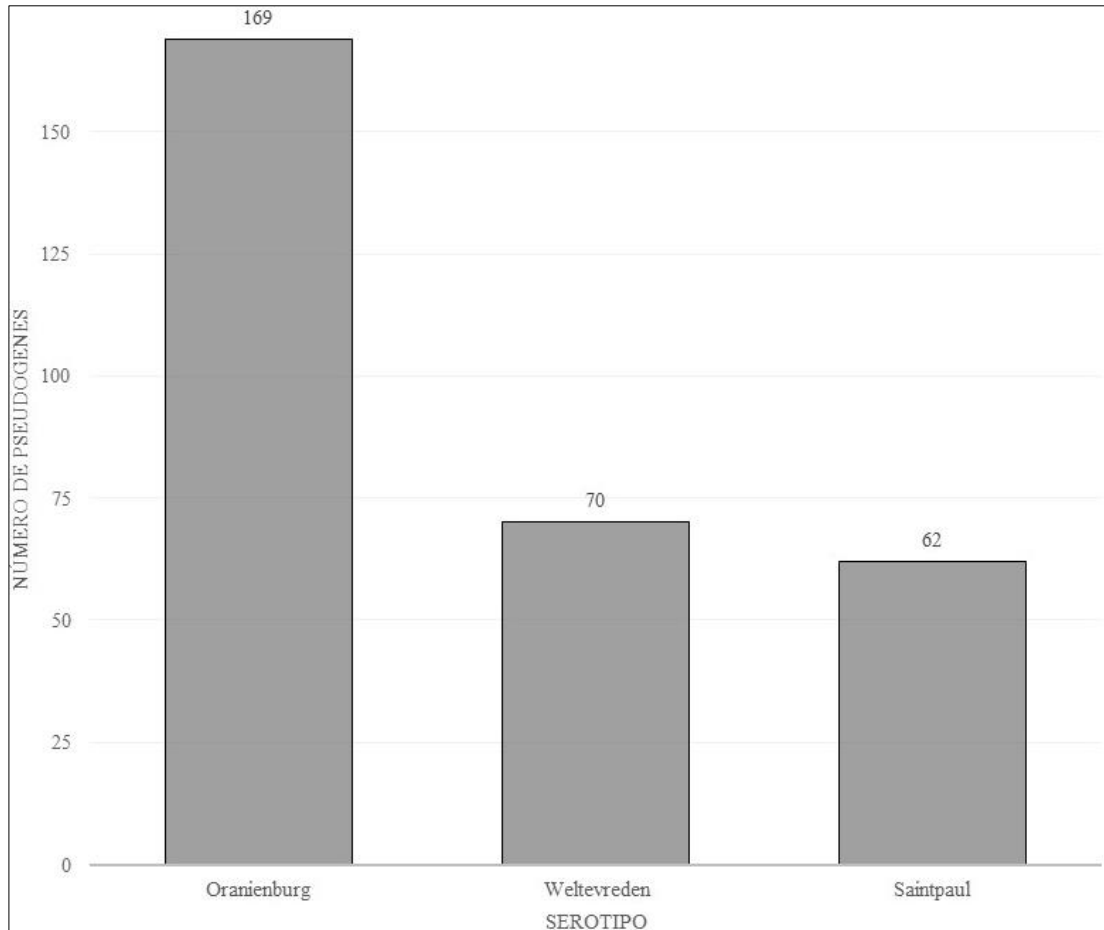


Figura 9. Conteo de pseudogenes presentes en 84 secuencias de *Salmonella enterica* serotipos Oranienburg, Saintpaul y Weltevreden (40,37 y 7), respectivamente.

Cuadro 7. Porcentaje de pseudogenes presentes en el genoma de los serotipos Oranienburg, Weltevreden y Saintpaul.

Serotipo	Genes	Porcentaje de pseudogenes
Oranienburg	4660	3.63%
Weltevreden	4777	1.47%
Saintpaul	4,710	1.32%

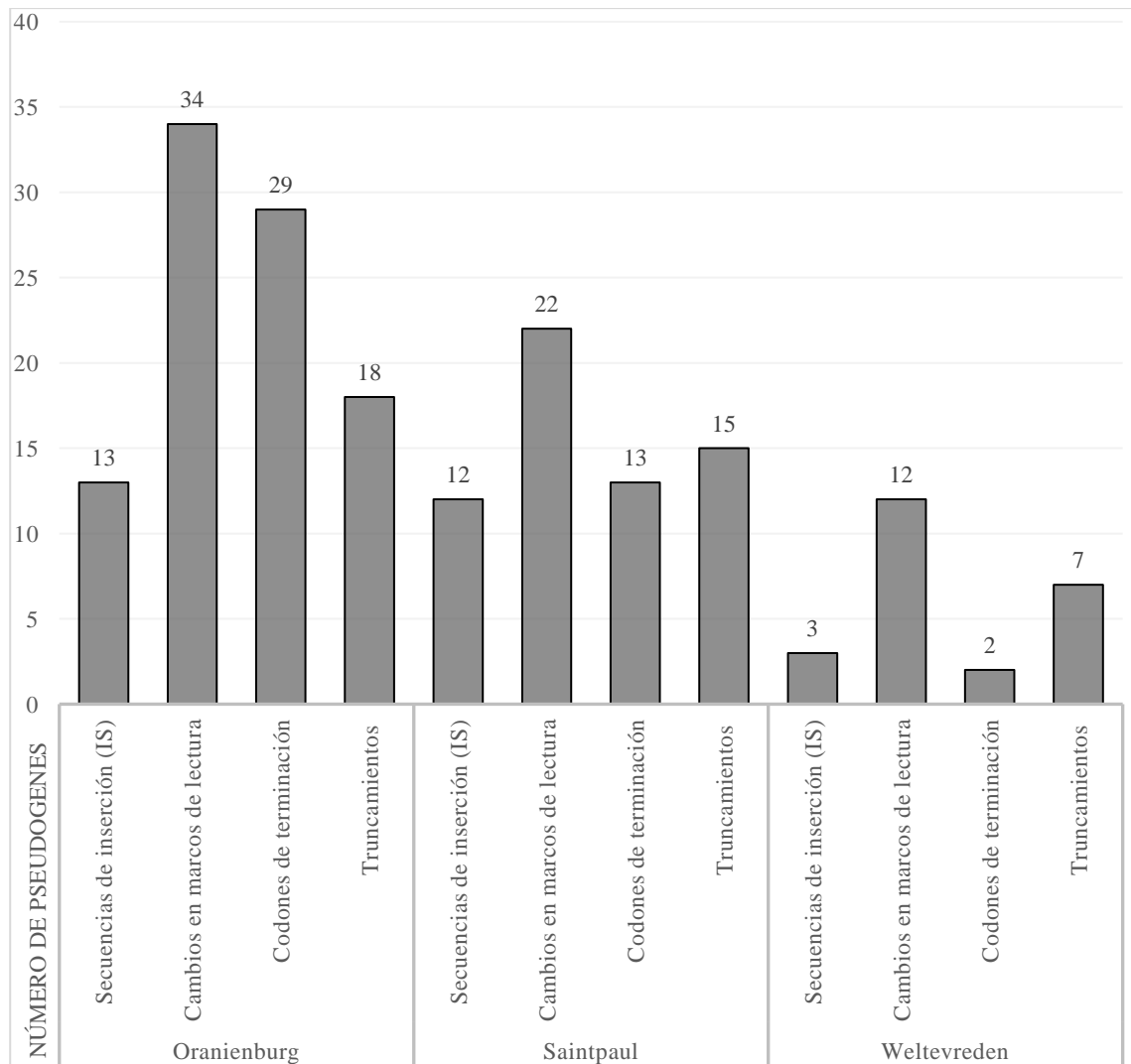


Figura 10. Principales eventos que originan a los pseudogenes en los serotipos Oranienburg, Saintpaul y Weltevreden.

6.2. Categorías Funcionales COG Asociadas a Genes Asociados a Pseudogenes

En relación con las categorías COG identificadas para los pseudogenes, la mayoría de los genomas analizados presentaron una incidencia alta de proteínas hipotéticas (31.56%) (Figura. 11). Al tratarse de proteínas muy degradadas, no fue posible asignarles una categoría estructural y funcional. Así mismo, se identificó un total de 25 categorías COG de genes funcionales asociados a pseudogenes (Figura 11).

La mayoría de los genomas analizados presentaron una gran cantidad de proteínas hipotéticas

(31.56%) por lo que no fue posible asignarles una categoría funcional ya que se trataba de proteínas muy degradadas. Sin embargo, la categoría con mayor prevalencia fue traducción, estructura ribosomal y biogénesis (Figura 11) (categoría COG J) asociada a pseudogenes que participan en la traducción, estructura ribosomal y biogénesis (10.7%), seguida de producción y conversión de energía (categoría COG C) (9.97%) que está relacionada con los procesos metabólicos que generan y convierten la energía necesaria para el funcionamiento celular la siguiente es transporte y metabolismo de carbohidratos (categoría COG G) (7.97%) dicha categoría se refiere a los procesos de transporte y metabolismo de carbohidratos, que son nutrientes esenciales para el crecimiento y la supervivencia bacteriana, modificaciones postraduccionales (categoría O) (7.64%) esta categoría está relacionada con los procesos metabólicos que generan y convierten la energía necesaria para el funcionamiento celular, conversión de energía y modificaciones postraduccionales (categorías COG O y C) (5.32%), metabolismo y transporte de aminoácidos (categoría COG E) (4.32%) la cual está relacionada con los procesos metabólicos que generan y convierten la energía necesaria para el funcionamiento celular, pared celular, membrana y biogénesis de envoltura (categoría COG M) (3.65%), transcripción (categoría COG K) (2.66%), transporte y metabolismo de iones inorgánicos (categoría COG P) (2.33%) , tráfico intracelular, secreción y transporte vesicular (categorías COG U, X, y R) (2.33%), transporte y metabolismo de nucleótidos (categoría COG F) (1.66%) , replicación, recombinación y reparación (categoría COG L) (1.66%), mecanismos de transducción de señales (categoría COG T) (1.66%), motilidad celular (categoría COG N) (1.3%) y metabolismo de lípidos (1%) (categoría COG I)

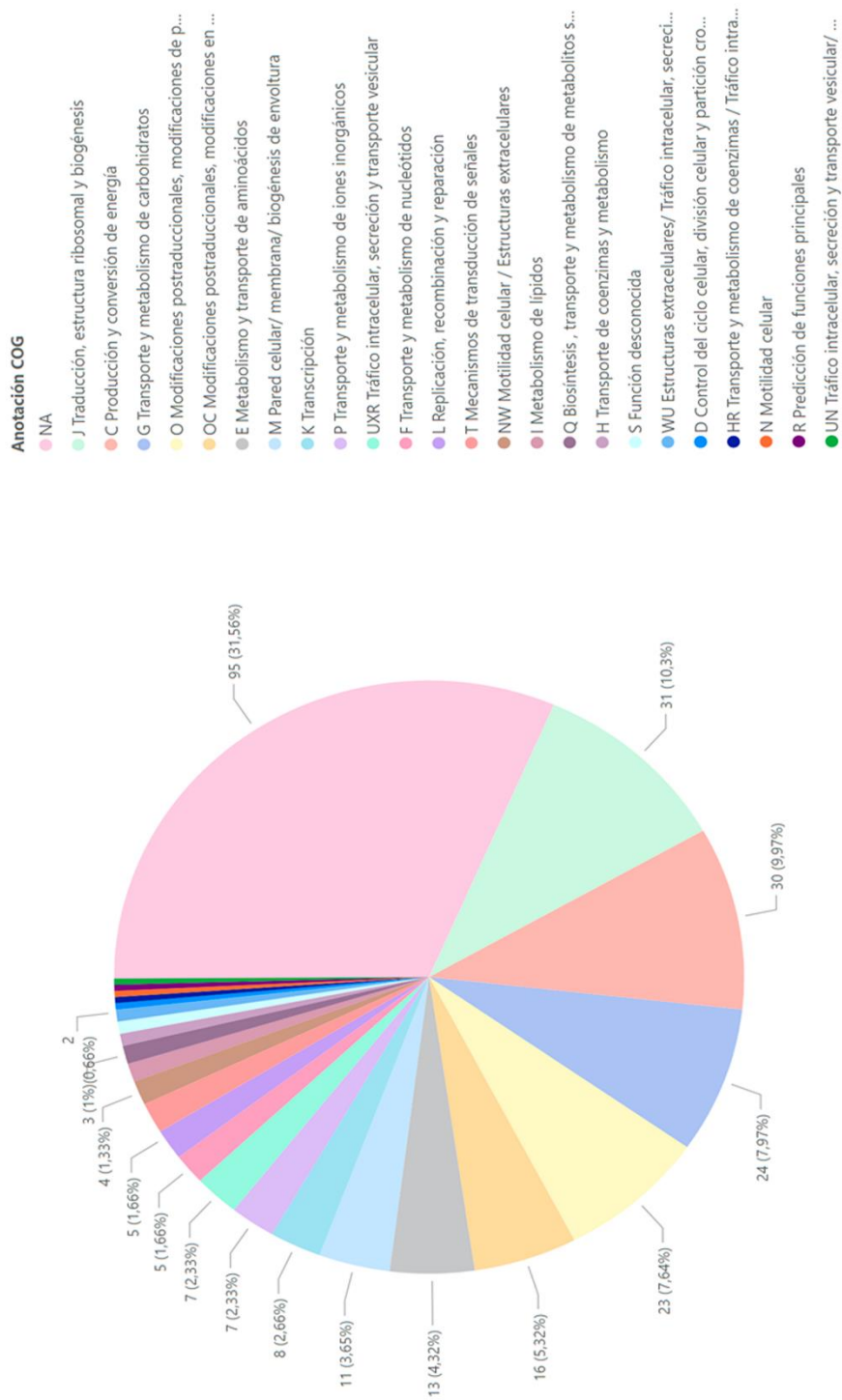


Figura 11. Categorías COG presentes en los genes asociados a pseudogenes en las secuencias de los serotipos Saintpaul, Oranienburg y Weltevreden.

6.3. Principales Genes Asociados a Pseudogenes

De acuerdo con un comparativo de homología de secuencias, se logró anotar funcional y estructuralmente los genes correspondientes a alguna categoría COG funcional. El principal gen de ocurrencia para los 3 serotipos es *lpp* (Figura 12), que codifica para una lipoproteína encargada de regular los mecanismos de la envoltura celular y se ha vinculado a aspectos de virulencia. Así mismo, se identificaron genes como *cmmA* y *ccmH* (categoría COG O) los cuales forman parte del sistema ABC involucrados en la biogénesis del citocromo tipo C. Por su parte, genes relacionados con procesos de respiración anaerobia como *fdnG* y *fdoG* (categoría COG C) fueron detectados(Figura 12).

Las secuencias parciales de los genes *fdhF* (categoría COG C), *sopA* (categoría COG S) y *oadA* (categoría COG I) se encontraron con alta incidencia para *S. Saintpaul* (78%, 78% y 67%, correspondientemente). El primero de ellos tiene una participación en la descomposición de ácido fórmico a hidrógeno y dióxido de carbono en condiciones de anaerobiosis. *sopA* codifica para una proteína efectora que altera la fisiología celular promoviendo su supervivencia en tejidos de hospederos. Por su parte, *oadA* (categoría COG I) consiste en una bomba presente en condiciones de anaerobiosis.

Adicionalmente, en la Figura 12 se muestran las secuencias parciales de genes identificados para los aislados. Se incluye un cladograma que muestra la agrupación de los aislados por ocurrencia de aparición de secuencias específicas. En los 3 cladogramas se puede apreciar agrupaciones, indicando similitud en el contenido de pseudogenes(Figura 12).

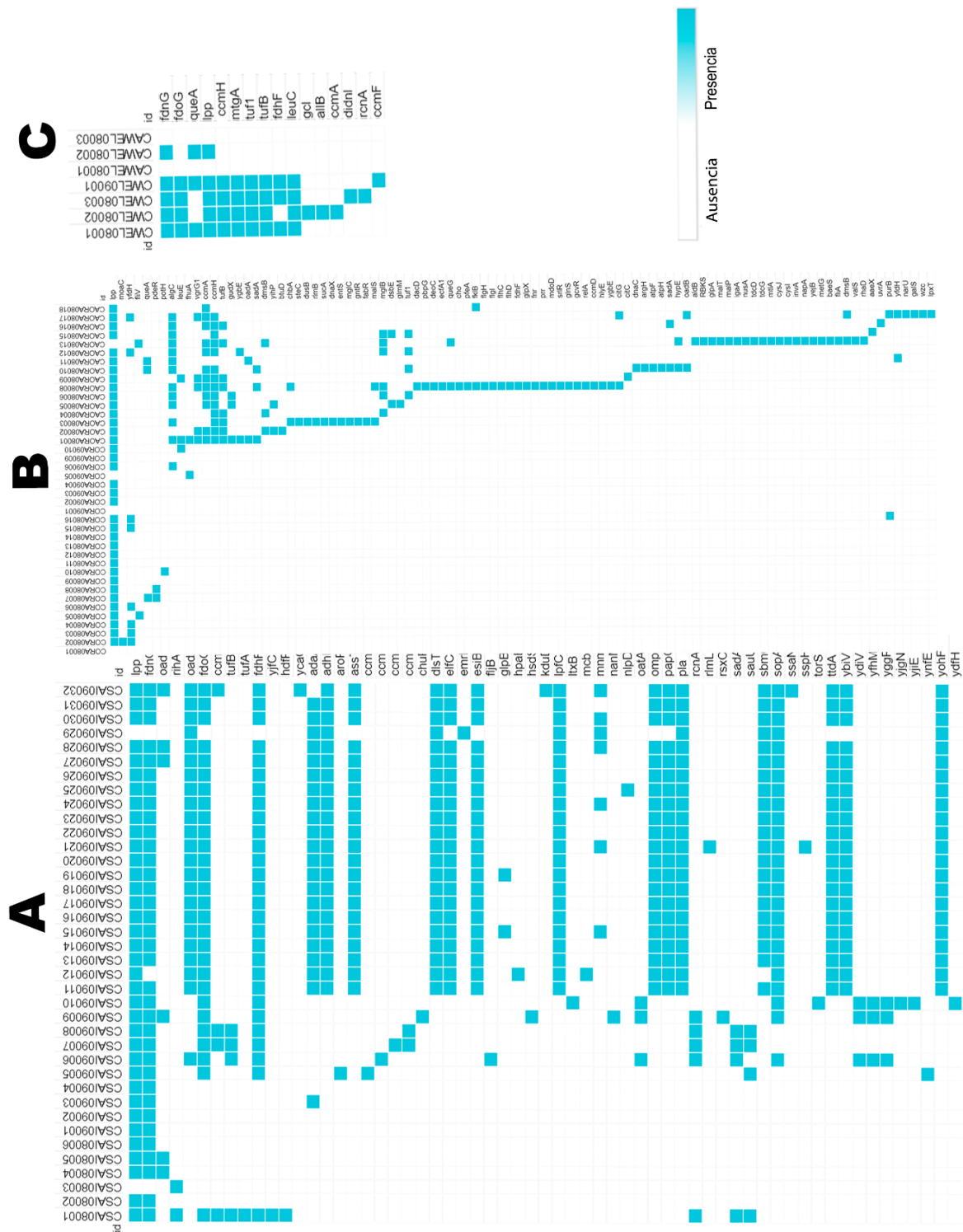


Figura 12. Heatmap de presencia/ausencia de los principales genes homólogos a pseudogenes en *S. Saintpaul*, *S. Oranienburg*, *S. Weltevreden*. Donde el color azul significa presencia y el color blanco ausencia. El panel A representa a *S. Saintpaul*, el B a Oranienburg y el C a *S. Weltevreden*.

En esta investigación, se analizaron un total de 84 secuencias de *Salmonella enterica* de los serotipos Oranienburg, Saintpaul y Weltevreden (40, 37 y 7 secuencias, respectivamente) con el objetivo de identificar pseudogenes, comprender su distribución y posibles funciones para estos serotipos específicos.

Los resultados obtenidos revelaron que la mayoría de los pseudogenes identificados en los tres serotipos correspondieron a proteínas hipotéticas y desconocidas, lo cual coincide con lo reportado en *E. coli* (Homma *et al.*, 2002).

En cuanto a la cantidad de pseudogenes identificados en cada serotipo, se observó que el *S. Oranienburg* presentó el mayor número de pseudogenes con un total de 94 pseudogenes (Cuadro 10), seguido por *S. Saintpaul* (Cuadro 9) con 62 pseudogenes. Por otro lado, *S. Weltevreden* fue el que presentó la menor cantidad de pseudogenes con un total de 24 (Cuadro 8).

Al analizar el porcentaje de pseudogenes en los genomas de cada serotipo, se encontró que hubo una variación considerable. *S. Oranienburg* fue el que mostró la mayor presencia de pseudogenes, representando el 3.63 % del tamaño total de su genoma. Sin embargo, los serotipos *S. Weltevreden* y *S. Saintpaul* presentaron un menor porcentaje con 1.47% y 1.32%, respectivamente (Cuadro 7).

Lo anterior concuerda con lo establecido por Liu y colaboradores (Liu *et al.*, 2004), esto se podría deber a que *S. Oranienburg* se encuentra en constante intercambio entre un ambiente hospedante y no hospedante (humano – medio ambiente), sin embargo, es importante considerar un periodo de tiempo prolongado e incorporar cepas de origen clínico quienes sugirieron un porcentaje de pseudogenes promedio en procariotas está dentro del rango del 1 al 5% del total de su genoma. Además, a pesar de que *Salmonella Oranienburg* se caracteriza por ser un microorganismo de vida libre, se encontró que su porcentaje de pseudogenes presenta una similitud con el porcentaje de pseudogenes de bacterias asociadas a un hospedero, tal es caso de *E. coli* MG1655 con 3.7%, *E. coli* O:157 con un porcentaje de 3.8% y *S. Typhimurium* LT2 con 4.1% (Liu *et al.*, 2004).

En cuanto a los principales mecanismos que originaron la presencia de pseudogenes, se identificó a los cambios en marcos de lectura como el evento más común en los tres serotipos Oranienburg, Saintpaul y Weltevreden (30, 22 y 12, respectivamente).

Esto coincide con lo reportado por Lerat y colaboradores (2005) quienes identificaron dicho mecanismo como el más frecuente en la generación de pseudogenes en los genomas procariotas.

Para *S. Oranienburg*, el segundo evento más frecuente que generó pseudogenes fue la presencia de codones de terminación prematuros, con un total de 29 eventos. En contraste, el serotipo Weltevreden presentó solo 2 eventos de este tipo, mientras que el serotipo Saintpaul mostró 13

eventos. Así mismo, el truncamiento se identificó como el segundo evento más frecuente en los serotipos Saintpaul y Weltevreden, con 15 y 7 eventos respectivamente. Aunque no se ha descrito la terminación prematura de codones como mecanismo de pseudogenización principal para *Salmonella* ya sean descrito como un evento importante en géneros como *S. aureus* y *Y. pestis*, eso podría sugerir que cada género bacteriano podría tener un mecanismo en particular o preferente para llevar a cabo la pseudogenización (Lerat y Ochman, 2005) , sin embargo es necesario identificar pseudogenes en diferentes géneros para determinar el mecanismo preferente de pseudogenización.

En relación con la funcionalidad de los pseudogenes identificados en *Salmonella enterica*, un porcentaje del 31.56 corresponde a proteínas hipotéticas. Este representa un numero alto en comparación con otros géneros bacterianos como *Pseudomonas* y *E. coli*, donde se detectó una presencia asociada a proteínas hipotéticas del 16.9 y 12% correspondientemente. Esta discrepancia asociada al incidencia de proteínas hipotéticas dentro del genoma bacteriano se puede explicar por la falta de estandarización en bases de datos o repositorios públicos y por la falta de flujos de trabajo que propician una notación más efectiva (Ijaq *et al.*, 2022).

Los pseudogenes identificados en este estudio mostraron una amplia diversidad en cuanto a las categorías de genes funcionales (COG) a las que estaban asociados. Se identificaron un total de 25 categorías, lo que indica una distribución heterogénea de los pseudogenes en términos de sus funciones biológicas.

La principal categoría asociada a pseudogenes traducción, estructura ribosomal y biogénesis (categoría COG J), que representó el 10.7% de los pseudogenes identificados. Esta categoría se relaciona con procesos de síntesis de proteínas y la formación de la maquinaria ribosomal.

Otra categoría detectada fue producción y conversión de energía (categoría COG C), que representó el 9.97% de los pseudogenes identificados.

Esta categoría está relacionada con los procesos metabólicos que generan y convierten la energía necesaria para el funcionamiento celular.

La categoría COG G transporte y metabolismo de carbohidratos también tuvo una proporción significativa de pseudogenes identificados, representando el 7.97%. La presencia de pseudogenes en esta categoría sugiere posibles alteraciones en los genes involucrados en la utilización y regulación de los carbohidratos.

Otras categorías relevantes incluyeron modificaciones postraduccionales (7.64%), OC Conversión de energía y modificaciones postraduccionales (5.32%), E Metabolismo y transporte de

aminoácidos (4.32%), y M Pared celular, membrana y biogénesis de envoltura (3.65%). Estas categorías representan procesos importantes en la biología bacteriana, como las modificaciones de proteínas, el metabolismo de aminoácidos y la formación de la pared celular y membranas.

Es importante destacar que también se identificaron categorías menos frecuentes, como Q Biosíntesis, transporte y metabolismo de metabolitos secundarios, H Transporte de coenzimas y metabolismo, S Función desconocida, y D Control de ciclo celular, entre otras. Estas categorías indican posibles funciones y procesos biológicos en los que los pseudogenes que podrían estar involucrados, aunque su papel específico y relevancia aún requieren una investigación más detallada.

Los resultados revelan una amplia diversidad de categorías de genes funcionales asociadas a los pseudogenes identificados en *Salmonella enterica*. Estas categorías abarcan procesos biológicos fundamentales, como la traducción, el metabolismo energético, el transporte de nutrientes y la formación de estructuras celulares. Estos hallazgos proporcionan información valiosa sobre los posibles efectos de los pseudogenes en la biología y la adaptación de *Salmonella enterica*. Sin embargo, se requieren estudios futuros para comprender mejor las implicaciones funcionales y evolutivas de estos pseudogenes en esta bacteria patógena.

6.4. Descripción de Genes Asociados a Pseudogenes

Los serotipos comparten entre sí cinco pseudogenes (*lpp*, *fdnG*, *ccmH*, *tufB* y *queA*) los cuales están asociados a funciones biológicas importantes como metabolismo, transporte de metabolitos y modificaciones postraduccionales que tal vez permiten a los serotipos Weltevreden, Saintpaul y Oranienburg adaptarse fácilmente a ambientes no hospederos.

6.4.1. Gen *lpp*

El gen *lpp*, también conocido como lipoproteína de baja masa molecular, es un gen que codifica a la lipoproteína de Braun, la cual es una proteína que es abundante en la membrana externa de

bacterias Gram negativas como *E. coli*, *Salmonella* y *Pseudomonas*. Esta lipoproteína brinda estabilidad a la membrana externa, así como también proporciona resistencia a la lisis bacteriana y da protección contra factores ambientales negativos (Asmar y Collet, 2018). La implicación de un pseudogen de *lpp* podría influir en que *Salmonella* reduzca la producción en esta proteína que encuentra en su membrana externa. Esto podría alterar la interacción de la bacteria con el sistema inmunológico de su hospedero y disminuir su capacidad para reconocer y eliminar a *Salmonella*.

6.4.2. *fdnG*

El gen *fdnG* es un gen ancestral que se encuentra conservado en la mayoría de las bacterias entéricas como *S. Typhimurium* y *E. coli* codifica a la proteína formiato deshidrogenasa, nitrato inducible que participa en la respiración por nitrato y participa en la oxidación de compuestos orgánicos nitrogenados. La ausencia de este gen puede hacer que *Salmonella* no lleve a cabo de manera eficiente la respiración por nitrógeno por lo que este microorganismo no podrá utilizar compuestos nitrogenados como fuente de energía y podría dar lugar a que dependiera de otras rutas metabólicas alternativas para permitir su sobrevivencia en ausencia del gen *fdnG*.

6.4.3. *ccmH*

El gen *ccmH* es un gen que codifica una proteína involucrada en el ensamblaje de un complejo proteico llamado sistema de maduración de cito cromosomas (CCM). El CCM es esencial para la correcta incorporación del grupo hemo en las proteínas citocromo c, las cuales están implicadas en el transporte de electrones de muchas bacterias.

La presencia del pseudogen *ccmH* en *Salmonella* puede tener implicaciones en la capacidad de la bacteria para ensamblar correctamente el CCM y, por lo tanto, afectar la función de proteínas citocromo c y el transporte de electrones. Esto podría repercutir en la capacidad de la bacteria para generar energía y llevar procesos metabólicos importantes, por lo que es necesario realizar más

investigación para comprender las consecuencias que puede tener en la biología de la bacteria.

6.4.4. *tufB*

El gen *tufB* codifica a la proteína de factor de elongación Tu (EF-Tu), que es una proteína esencial para la síntesis de proteínas. La presencia de un pseudogen de *tufB* en *Salmonella* podría afectar la síntesis de proteínas y por lo tanto tener consecuencias en la viabilidad y función celular de la bacteria.

Dado que EF-Tu desempeña un papel crucial en la elongación de cada peptídica durante la traducción, un pseudogen de *tufB* podría alterar la velocidad y eficiencia de la síntesis de proteínas.

6.4.5. *queA*

El gen *queA* codifica una enzima que tiene un papel crucial en la modificación de la base de guanina en los tRNAs. Esta modificación es esencial para la síntesis precisa de proteínas y tiene implicaciones en la virulencia y patogenicidad de la bacteria.

Por lo que la presencia de un pseudogen de *queA* podría afectar la capacidad de *Salmonella* para infectar y colonizar a su hospedero.

7. CONCLUSIONES

Este estudio proporciona una visión inicial sobre la presencia y características de los pseudogenes presentes en tres serotipos de *Salmonella*: Oranienburg, Saintpaul y Weltevreden. Los resultados revelaron una considerable cantidad de pseudogenes en los genomas analizados, lo que sugiere que la pseudogenización es un fenómeno común entre las cepas de *Salmonella*.

Las diferentes clasificaciones de los pseudogenes detectados sugiere que estos podrían haber desempeñado roles importantes en la evolución y adaptación de *Salmonella* a diferentes ambientes. Si bien, este estudio brinda información valiosa sobre los pseudogenes de *Salmonella*, es necesario realizar un análisis funcional más detallado para evaluar la función de los pseudogenes identificados. Esto podría incluir análisis de la expresión génica donde se determine si los pseudogenes están siendo transcritos y si tienen algún efecto en la virulencia y resistencia a antibióticos de *Salmonella*.

Además, es necesario realizar estudios comparativos entre los pseudogenes de distintos serotipos ambientales de *Salmonella*. Esto permitirá identificar pseudogenes específicos para cada serotipo y podría incrementar la información sobre su adaptación y evolución a distintos entornos.

En conjunto, estos enfoques podrían brindar una comprensión más amplia de la importancia funcional de los pseudogenes de *Salmonella* y su contribución a la biología de la bacteria.

8. RECOMENDACIONES

Es necesario realizar un estudio de pseudogenes en otros serotipos de origen ambiental de *Salmonella* dado que este estudio se centró en los serotipos Oranienburg, Weltevreden y Saintpaul, esto con el fin de ampliar un poco más la visión de la distribución y función de los pseudogenes en diferentes cepas.

Se sugiere realizar experimentos de sobreexpresión e inhibición de pseudogenes para evaluar su efecto en la virulencia y resistencia en *Salmonella*.

9. REFERENCIAS

- Abby, S., y Daubin, V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends Microbiol*, 15(3), 135-141. doi:10.1016/j.tim.2007.01.007
- Aguirre-Sanchez, J. R., Ibarra-Rodriguez, J. R., Vega-Lopez, I. F., Martínez-Urtaza, J., y Chaidez-Quiroz, C. (2021). Genomic signatures of adaptation to natural settings in non-typhoidal *Salmonella enterica* Serovars Saintpaul, Thompson and Weltevreden. *Infect Genet Evol*, 104771. doi:10.1016/j.meegid.2021.104771
- Alikhan, N.-F., Zhou, Z., Sergeant, M. J., y Achtman, M. (2018). A genomic overview of the population structure of *Salmonella*. *Plos Genet*, 14(4), e1007261. doi:10.1371/journal.pgen.1007261
- Amavisit, P., Lightfoot, D., Browning, G. F., y Markham, P. F. (2003). Variation between pathogenic serovars within *Salmonella* pathogenicity islands. *J Bacteriol*, 185(12), 3624-3635. doi:10.1128/jb.185.12.3624-3635.2003
- Anderson, C. J., y Kendall, M. M. (2017). *Salmonella enterica* serovar typhimurium strategies for host adaptation. *Front Microbiol*, 8(OCT). doi:10.3389/fmicb.2017.01983
- Andrews-Polymenis, H. L., Baumler, A. J., McCormick, B. A., y Fang, F. C. (2010). Taming the Elephant: *Salmonella* Biology, Pathogenesis, and Prevention. *Infect Immun*, 78(6), 2356-2369. doi:10.1128/iai.00096-10
- Asmar, A. T., y Collet, J.-F. (2018). Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. *Fems Microbiol Lett*, 365(18). doi:10.1093/femsle/fny199
- Baker, S., y Dougan, G. (2007). The Genome of *Salmonella enterica* Serovar Typhi. *Clin Infect Dis*, 45(Supplement_1), S29-S33. doi:10.1086/518143
- Balasubramanian, R., Im, J., Lee, J.-S., Jeon, H. J., Mogeni, O. D., Kim, J. H., . . . Marks, F. (2019). The global burden and epidemiology of invasive non-typhoidal *Salmonella* infections. *Hum Vaccin Immunother*, 15(6), 1421-1426. doi:10.1080/21645515.2018.1504717
- Bell, C., y Kyriakides, A. (2002). *Salmonella: A practical approach to the organism and its control in foods* (1 Ed. Vol. 1): Blackwell Science Ltd.
- Bobay, L. M., y Ochman, H. (2017). The Evolution of Bacterial Genome Architecture. *Front Genet*, 8, 6. doi:10.3389/fgene.2017.00072
- Bolotin, E., y Hershberg, R. (2015). Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. *Genome Biol*, 7(8), 2173-2187. doi:10.1093/gbe/evv135
- Brown, E. W., Bell, R., Zhang, G., Timme, R., Zheng, J., Hammack, T. S., y Allard, M. W. (2021). *Salmonella* Genomics in Public Health and Food Safety. 9(2), eESP-0008-2020. doi:doi:10.1128/ecosalplus.ESP-0008-2020
- Bumann, D., y Schothorst, J. (2017). Intracellular *Salmonella* metabolism. *Cell Microbiol*, 19(10). doi:10.1111/cmi.12766

- Chalker, R. B., y Blaser, M. J. J. R. o. i. d. (1988). A review of human salmonellosis: III. Magnitude of Salmonella infection in the United States. *IO(1)*, 111-124.
- Cheetham, S. W., Faulkner, G. J., y Dinger, M. E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Rev Genet*, *21(3)*, 191-201. doi:10.1038/s41576-019-0196-1
- Chu, X., Li, S., Wang, S., Luo, D., y Luo, H. (2021). Gene loss through pseudogenization contributes to the ecological diversification of a generalist Roseobacter lineage. *ISME Journal*, *15(2)*, 489-502. doi:10.1038/s41396-020-00790-0
- Coburn, B., Grassl, G. A. y Finlay, B. J. I., (2007). Salmonella, the host and disease: a brief review. *Immunol Cell Biol*. 2007 Feb-Mar;85(2):112-8. doi: 10.1038/sj.icb.7100007
- Contreras-Soto, M., Medrano-Félix, J., Chaidez, C., Ibarra-Rodríguez, J., Martínez-Urtaza, J., y Castro-del Campo, N. J. R. (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)* . doi: 10.15741/revbio.06.nesp.e540
- Crawford, R. W., Wangdi, T., Spees, A. M., Xavier, M. N., Tsohis, R. M., y Bäumlér, A. J. (2013). Loss of very-long O-antigen chains optimizes capsule-mediated immune evasion by Salmonella enterica serovar Typhi. *Mbio*, *4(4)*. doi:10.1128/mBio.00232-13
- Crump, J. A., Sjölund-Karlsson, M., Gordon, M. A., y Parry, C. M. (2015). Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive Salmonella infections. *Clin Microbiol Rev*, *28(4)*, 901-937. doi:10.1128/CMR.00002-15
- Dufresne, K., Saulnier-Bellemare, J., y Daigle, F. (2018). Functional analysis of the chaperone-usher fimbrial gene clusters of Salmonella enterica serovar Typhi. *Front Cell Infect Microbiol*, *8(2)*. doi:10.3389/fcimb.2018.00026
- Duprey, A., Reverchon, S., y Nasser, W. (2014). Bacterial virulence and Fis: Adapting regulatory networks to the host environment. *Trends Microbiol*, *22(2)*, 92-99. doi:10.1016/j.tim.2013.11.008
- Ehuwa, O., Jaiswal, A. K., y Jaiswal, S. (2021). Salmonella, Food Safety and Food Handling Practices. *Foods*, *10(5)*, 907. doi:10.3390/foods10050907
- Ellermeier, C. D., y Slauch, J. M. (2006). The Genus Salmonella. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, y E. Stackebrandt (Eds.), *The Prokaryotes: A Handbook on the Biology of Bacteria Volume 6: Proteobacteria: Gamma Subclass* (pp. 123-158). New York, NY: Springer New York.
- Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G., y Lee, L.-H. (2015). Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance. *Front Life Sci*, *8(3)*, 284-293. doi:10.1080/21553769.2015.1051243
- Feng, Y., Wang, Z., Chien, K.-Y., Chen, H.-L., Liang, Y.-H., Hua, X., y Chiu, C.-H. (2022). “Pseudo-pseudogenes” in bacterial genomes: Proteogenomics reveals a wide but low protein expression of pseudogenes in Salmonella enterica. *Nucleic Acids Res*, *50(9)*, 5158-5170. doi:10.1093/nar/gkac302 %J Nucleic Acids Research
- Fu, S., Octavia, S., Tanaka, M. M., Sintchenko, V., y Lan, R. (2015). Defining the Core Genome

- of *Salmonella enterica* Serovar Typhimurium for Genomic Surveillance and Epidemiological Typing. *J Clin Microbiol*, 53(8), 2530-2538. doi:10.1128/jcm.03407-14
- Gil, R., y Latorre, A. (2012). Factors behind junk DNA in bacteria *Genes*, 3(4), 634-650. doi: 10.3390/genes3040634
- Godínez-Oviedo, A., Tamplin, M. L., Bowman, J. P., y Hernández-Iturriaga, M. (2020). *Salmonella enterica* in Mexico 2000-2017: Epidemiology, Antimicrobial Resistance, and Prevalence in Food. *Foodborne Pathog Dis*, 17(2), 98-118. doi:10.1089/fpd.2019.2627
- Goodhead, I., Alistair C. (2015). Taking the pseudo out of pseudogenes. *Curr Opin Microbiol*, 23, 102-109. doi:10.1016/j.mib.2014.11.012
- Grimont, P., y Weill, F.-X. (2007). Antigenic Formulae of the *Salmonella* serovars, (9th ed.) Paris: WHO Collaborating Centre for Reference and Research on *Salmonella*. *Institute Pasteur.*, 1-166.
- Gutiérrez-Cogco, L., Montiel-Vázquez, E., Aguilera-Pérez, P., y González-Andrade, M. (2000). Serotipos de *Salmonella* identificados en los servicios de salud de México. *Salud pública de México*, 42(6), 490-495.
- Holt, K. E., Thomson, N. R., Wain, J., Langridge, G. C., Hasan, R., Bhutta, Z. A., . . . Parkhill, J. (2009). Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *Bmc Genomics*, 10. doi:10.1186/1471-2164-10-36
- Homma, K., Fukuchi, S., Kawabata, T., Ota, M., y Nishikawa, K. (2002). A systematic investigation identifies a significant number of probable pseudogenes in the *Escherichia coli* genome. *Gene*, 294(1), 25-33. doi:https://doi.org/10.1016/S0378-1119(02)00794-1
- Hu, Y., Wang, Z., Qiang, B., Xu, Y., Chen, X., Li, Q., y Jiao, X. (2019). Loss and Gain in the Evolution of the *Salmonella enterica* Serovar Gallinarum Biovar Pullorum Genome. *mSphere*, 4(2). doi:10.1128/mSphere.00627-18
- ICMSF. (1996). *Microorganisms in Food 5: Microbiological Specifications of Food Pathogens*. London: Springer New York, NY.
- Ijaq, J., Chandra, D., Ray, M. K., y Jagannadham, M. V. (2022). Investigating the Functional Role of Hypothetical Proteins From an Antarctic Bacterium *Pseudomonas* sp. Lz4W: Emphasis on Identifying Proteins Involved in Cold Adaptation. *Front Genet*, 13, 825269. doi:10.3389/fgene.2022.825269
- Barton, J., Hill, A. (2021). Host restriction, pathogenesis and chronic carriage of typhoidal *Salmonella*. *Fems Microbiol Rev*, 45(5). doi:10.1093/femsre/fuab014
- Jacq, C., Miller, J. R., y Brownlee, G. G. (1977). A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell*, 12(1), 109-120. doi:https://doi.org/10.1016/0092-8674(77)90189-1
- 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), 504-521. doi:10.14202/vetworld.2019.504-521
- Jiménez, M., Martínez-Urtaza, J., Rodríguez-Alvarez, M. X., Leon-Felix, J., y Chaidez, C. (2014). Prevalence and genetic diversity of *Salmonella* spp. in a river in a tropical environment in Mexico. *J Water Health*, 12(4), 874-884. doi:10.2166/wh.2014.051 %J Journal of Water and Health

- Kovalenko, T. F., y Patrushev, L. I. (2018). Pseudogenes as Functionally Significant Elements of the Genome. *Biochemistry*, 83(11), 1332-1349. doi:10.1134/s0006297918110044
- Kuo, C. H., y Ochman, H. (2010). The Extinction Dynamics of Bacterial Pseudogenes. *Plos Genetics*, 6(8). doi:10.1371/journal.pgen.1001050
- Lerat, E., y Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Res*, 33(10). doi:10.1093/nar/gki631
- Li, W., Yang, W., y Wang, X.-J. (2013). Pseudogenes: Pseudo or Real Functional Elements? *J Genet Genomics*, 40(4), 171-177. doi:https://doi.org/10.1016/j.jgg.2013.03.003
- Liu, H., Whitehouse, C. A., y 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). doi:10.3389/fpubh.2018.00159
- Liu, Y., Harrison, P. M., Kunin, V., y Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol*, 5(9), 11. doi:10.1186/gb-2004-5-9-r64
- Mogasale, V., Maskery, B., Ochiai, R. L., Lee, J. S., Mogasale, V. (2014). Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. *Lancet Glob Health*, 2(10). doi: 10.1016/S2214-109X(14)70301-8
- Newell, D. G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., . . . Kruse, H. (2010). Food-borne diseases - the challenges of 20 years ago still persist while new ones continue to emerge. *Int J Food Microbiol*, 139 Suppl 1, S3-15. doi:10.1016/j.ijfoodmicro.2010.01.021
- Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X. J., Gerstein, M. B. (2012). The GENCODE pseudogene resource. *Genome Biol*, 13(9), R51. doi:10.1186/gb-2012-13-9-r51
- Percival, S. L., y Williams, D. W. (2014). Chapter Ten - Salmonella. In S. L. Percival, M. V. Yates, D. W. Williams, R. M. Chalmers, y N. F. Gray (Eds.), *Microbiology of Waterborne Diseases (Second Edition)* (pp. 209-222). London: Academic Press.
- Pink, R. C., Wicks, K., Caley, D. P., Punch, E. K., Jacobs, L., y Carter, D. R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA*, 17(5), 792-798. doi:10.1261/rna.2658311
- Podlaha, O., y Zhang, J. (2010). Pseudogenes and Their Evolution. *eLS*. doi: 10.1002/9780470015902.a0005118.pub2
- Pradhan, D., y Devi Negi, V. (2019). Stress-induced adaptations in Salmonella: A ground for shaping its pathogenesis. *Microbiol Res*, 229. doi:10.1016/j.micres.2019.126311
- Quiroz-Santiago, C., Rodas-Suárez, O. R., Vázquez Q, C. R., Fernández, F. J., Quiñones-Ramírez, E. I., y Vázquez-Salinas, C. (2009). Prevalence of Salmonella in Vegetables from Mexico. *J Food Prot*, 72(6), 1279-1282. doi:10.4315/0362-028X-72.6.1279
- Santillana Farakos, S. M., Frank, J. F., y Schaffner, D. W. (2013). Modeling the influence of temperature, water activity and water mobility on the persistence of Salmonella in low-moisture foods. *Int J Food Microbiol*, 166(2), 280-293. doi:https://doi.org/10.1016/j.ijfoodmicro.2013.07.007

- Spector, M. P., y Kenyon, W. J. (2012). Resistance and survival strategies of *Salmonella enterica* to environmental stresses. *Food Res Int*, 45(2), 455-481. doi:<https://doi.org/10.1016/j.foodres.2011.06.056>
- Syberg-Olsen, M. J., Garber, A. I., Keeling, P. J., McCutcheon, J. P., y Husnik, F. (2022). Pseudofinder: Detection of Pseudogenes in Prokaryotic Genomes. *Mol Biol Evol*, 39(7). doi:10.1093/molbev/msac153
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., . . . Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 28(5), 511-515. doi:10.1038/nbt.1621
- Tutar, Y. (2012). Pseudogenes. *Comp Funct Genomics*, 2012 (1). doi:10.1155/2012/424526
- Winfield, M. D., y Groisman, E. A. (2003). Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl Environ Microbiol*, 69(7), 3687-3694. doi:10.1128/aem.69.7.3687-3694.2003
- Xu, J., y Zhang, J. (2016). Are Human Translated Pseudogenes Functional? *Mol Biol Evol*, 33(3), 755-760. doi:10.1093/molbev/msv268
- Zúñiga, K., y Abraham, E. J. (2006). Reacción de Widal: interpretación clínica. *Rev. Panam. Infectol*, 14(2) ,40-44.

10. APÉNDICES

10.1 Categoría Funcional de los Genes Asociados a Pseudogenes en *S. Weltevreden*.

Serotipo	Gen	Descripción	COG
	<i>allB</i>	Alantinasas	COG0044
	<i>ccmA</i>	Proteína de exportación de unión a ATP de biogénesis de citocromo c	COG4133
	<i>ccmF</i>	Proteína de biogénesis de citocromo tipo c	COG1138
	<i>ccmH</i>	Proteína de biogénesis de citocromo c tipo	COG3088
	<i>dinI</i>	Proteína I inducible por daño en el ADN	NA
	<i>fdhF</i>	Formato deshidrogenasa H	COG0243
	<i>fdnG</i>	Formato deshidrogenasa, subunidad mayor inducible por nitrato	COG0243
	<i>fdoG</i>	Formato deshidrogenasa-O, subunidad mayor	COG0243
	<i>gcl</i>	Carboligasa de glicolato	COG3960
Weltevreden	<i>leuC</i>	Subunidad grande de la 3-isopropilmalato deshidratasa	COG0065
	<i>lppI</i>	Proteína lipoproteica de la membrana externa principal LppI	Sin categoría
	<i>mtgA</i>	Transglicosilasa biosintética de peptidoglicano	COG0744
	<i>queA</i>	Isomerasa ribosiltransferasa de S-adenosilmetionina:tRNA	COG0809
	<i>rcnA</i>	Sistema de eflujo de níquel/cobalto RcnA	COG2215
	Sin categoría	Transposasa de la familia IS3 ISEae2	Sin categoría
	Sin categoría	Transposasa de la familia IS3 ISSen1	Sin categoría
	Sin categoría	Transposasa de la familia IS3 ISSty	Sin categoría
	Sin categoría	Transposasa de la familia IS3 IS1230B	Sin categoría
	Sin categoría	Transposasa de la familia IS256 ISSod4	Sin categoría
	<i>tufI</i>	Factor de elongación Tu 1	COG0050
	<i>tufB</i>	Factor de elongación Tu 2	COG0050
	<i>tufB</i>	Factor de elongación Tu	COG0050

10.2. Categoría Funcional de los Genes Asociados a Pseudogenes en *S. enterica* Saintpaul.

Serotipo	Gen	Descripción	COG
Saintpaul	<i>Lpp</i>	proteína lipoproteica de la membrana externa principal Lpp1	Sin categoría
	<i>fdnG</i>	Formato deshidrogenasa, subunidad mayor inducible por nitrato	COG0243
	<i>oadB</i>	Cadena beta de la oxaloacetato descarboxilasa	COG1883
	<i>rihA</i>	Ribonucleósido hidrolasa específica de pirimidina	COG1957
	<i>fbaB</i>	Aldolasa de fructosa-1,6-bifosfato clase 1	COG1830
	<i>oadA</i>	Cadena alfa de la oxaloacetato descarboxilasa	COG0511
	<i>fdoG</i>	Formato deshidrogenasa-O, subunidad mayor	COG0243
	<i>fliC</i>	Flagelina de fase 2	
	<i>ccmH</i>	Proteína de biogénesis de citocromo c tipo H	COG3088
	<i>tufB</i>	Factor de elongación Tu 2	
	<i>tufA</i>	Factor de elongación Tu 1	
	<i>fdhF</i>	Formato deshidrogenasa H	COG0243
	<i>yjfC</i>	Ligasa ácido-amina putativa	COG0754
	<i>hdfR</i>	Regulador transcripcional tipo HTH	Sin categoría
	<i>adaA</i>	Activador transcripcional bifuncional/Enzima de reparación de ADN	COG2169
	<i>adhE</i>	Aldehído-alcohol deshidrogenasa	COG1012
	<i>aroP</i>	Proteína de transporte de aminoácidos aromáticos	COG1113
	<i>assT</i>	Sulfotransferasa de arilsulfato AssT	Sin categoría
	<i>ccmA</i>	Proteína de exportación de unión a ATP de biogénesis de citocromo c	COG4133
	<i>ccmB</i>	Proteína exportadora de hemo B	COG2386
	<i>ccmC</i>	Proteína exportadora de hemo C	COG0755
	<i>ccmE</i>	Proteína de biogénesis de citocromo tipo C	COG2332
	<i>chuR</i>	Enzima de maduración de sulfatasa anaeróbica	COG0641
<i>dlsT</i>	Transportador de serina putativo	COG0814	
<i>elfC</i>	Proteína de guía de membrana externa putativa	COG3188	
<i>emrD</i>	Proteína de resistencia a múltiples fármacos D	Sin categoría	

<i>esiB</i>	Proteína de unión a inmunoglobulina A secretora	Sin categoría
<i>fljB</i>	Flagelina de fase 2	COG1344
<i>glpB</i>	Subunidad B de la deshidrogenasa de glicerol-3-fosfato anaeróbica	COG3075
<i>hpaB</i>	Componente oxigenasa de la monooxigenasa 4-hidroxifenilacetato 3	COG2368
<i>hsdS</i>	Proteína de especificidad de la enzima de restricción tipo 1	Sin categoría
<i>kduD</i>	2-deshidro-3-desoxi-D-Gluconato 5-deshidrogenasa	Sin categoría
<i>lpfC</i>	Proteína de guía de membrana externa putativa	COG3188
<i>ltxB</i>	Proteína de unión a ATP de exportación de leucotoxina	
<i>mcbR</i>	Regulador transcripcional tipo HTH McbR	COG1802
<i>mnmC</i>	Proteína bifuncional de biosíntesis de tRNA 5-metilaminometil-2-tiouridina	COG0665
<i>nanM</i>	Epimerasa de N-acetilneuraminato	Sin categoría
<i>nlpD</i>	Activador de hidrolasa de mureína	Sin categoría
<i>oatA</i>	O-acetiltransferasa	Sin categoría
<i>ompX</i>	Proteína de membrana externa	COG3637
<i>papC</i>	Proteína de guía de membrana externa	COG3188
<i>pla</i>	Coagulasa/fibrinolisina	Sin categoría
<i>rcnA</i>	Sistema de eflujo de níquel/cobalto	COG2215
<i>rlmL</i>	Metiltransferasa de la subunidad grande del ARNr ribosomal	COG0116
<i>rsxC</i>	Subunidad C del complejo de oxidoreductasa que transloca iones	COG4656
<i>sadA</i>	Adhesina autotransportadora SadA	COG5295
<i>sauU</i>	Transportador de sulfoacetato putativo	Sin categoría
<i>sbmC</i>	Inhibidor de la girasa de ADN	Sin categoría
<i>sopA</i>	Ligasa de ubiquitina E3	COG1357
<i>ssaN</i>	Sintasa de ATP del aparato del sistema de secreción putativo	COG1157

<i>sspH</i> 2	Ligasa de ubiquitina E3	COG4886
<i>torS</i>	Proteína sensora	COG0784
<i>ttdA</i>	Subunidad alfa de la L (+) -tartrato deshidratasa	COG1951
<i>tuf</i>	Factor de elongación Tu	Sin categoría
<i>ybiV</i>	Fosfatasa de azúcar	COG0561
<i>ydiV</i>	Factor anti-FlhC (2) FlhD (4)	COG2200
<i>yfhM</i>	Alfa-2-macroglobulina	COG2373
<i>yggR</i>	Proteína putativa	COG2805
<i>yjgN</i>	Proteína de la membrana interna	COG4269
<i>yjiE</i>	Regulador transcripcional tipo HTH	Sin categoría
<i>ynfE</i>	Cadena de la reductasa de dimetil sulfoxido putativa	COG0243
<i>yohF</i>	Oxidorreductasa putativa	COG1028

10.3. Categoría Funcional de los Genes Asociados a Pseudogenes en *S. enterica* Oranienburg.

Serotipo	Gene	Descripción	COG
	<i>vgrG1</i>	Toxina VgrG1 que une la actina	COG3501
	<i>moaC</i>	Sintasa de monofosfato de piranopterina cíclica	COG0315
	<i>ccmA_2</i>	Proteína de exportación de ATP de biogénesis de citocromo c CcmA	COG4133
	<i>ccmH</i>	Proteína de biogénesis de citocromo c tipo CcmH	COG3088
	<i>Sin categoría</i>	Flavoproteína de transferencia de electrones-ubiquinona oxidoreductasa	COG0644
Oranienburg	<i>tufB</i>	Factor de elongación Tu 2	COG0050
	<i>gudX</i>	Dehidratasa de glucarato	COG4948
	<i>Lpp</i>	Proteína lipoproteica principal de la membrana externa Lpp 1	Sin categoría
	<i>algC</i>	Fosfomanomutasa/fosfoglucomutasa	COG1109
	<i>potH</i>	Proteína de permeasa del sistema de transporte de putrescina PotH	COG1176
	<i>queA</i>	Isomerasa ribosiltransferasa de S-adenosilmetionina:tRNA	COG0809
	<i>leuE</i>	Proteína de efusión de leucina	COG1280

<i>fhuA</i>	Transportador de membrana externa de ferricromo/receptor de fagos	COG1629
<i>fliV</i>	Proteína biosintética de flagelos FliV	Sin categoría
<i>pdeR</i>	Fosfodiesterasa de ciclo di-GMP PdeR	COG2199
<i>dmsB</i>	Cadena B de la reductasa de dimetilsulfóxido anaeróbica	COG0437
<i>purB</i>	Liasa de adenylosuccinato	COG0015
<i>yfdH</i>	Transferasa de bactoprenol glucosilo de profago	COG0463
<i>citG</i>	Sintasa de 2-(5"-trifosforribosil) -3'-desfosfocoenzima-A	COG1767
<i>narU</i>	Transportador de nitrato/nitrito NarU	COG2223
<i>oadB</i>	Cadena beta de la oxaloacetato descarboxilasa	COG1883
<i>galS</i>	Regulador transcripcional tipo HTH GalS	COG1609
<i>wzc</i>	Quinasa de tirosina-proteína wzc	COG0489
<i>lpxT</i>	Sintasa de lipid A 1-difosfato	COG0671
<i>sadA</i>	Adhesina autotransportadora SadA	COG5295
<i>mglB</i>	Proteína periplasmática de unión a D-galactosa	COG1879
<i>uvrA</i>	Proteína A del sistema UvrABC	COG0178
<i>tufI</i>	Factor de elongación Tu 1	Sin categoría
<i>dsbE</i>	Proteína de intercambio tiol-disulfuro DsbE	COG0526
<i>aldB</i>	Aldehído deshidrogenasa B	COG1012
<i>RBKS</i>	Ribocinasa	Sin categoría
<i>glpA</i>	Subunidad A de la deshidrogenasa de glicerol-3-fosfato anaeróbica	COG0578
<i>malT</i>	Regulador transcripcional tipo HTH MalT	Sin categoría
<i>malP</i>	Fosforilasa de maltodextrina	COG0058
<i>igaA</i>	Proteína atenuadora del crecimiento intracelular igaA	Sin categoría
<i>nusA</i>	Proteína de terminación/antiterminación de la transcripción NusA	COG0195
<i>tdcD</i>	Quinasa de propionato	COG0282
<i>tdcG</i>	Dehidratasa de L-serina TdcG	COG1760
<i>mtlA</i>	Componente EIICB específico para manitol del sistema PTS	Sin categoría
<i>cysJ</i>	Componente alfa de la flavoproteína de la reductasa de sulfito [NADPH]	COG0369
<i>cysI</i>	Componente beta de la flavoproteína de la reductasa de sulfito [NADPH]	COG0155
<i>invA</i>	Proteína de invasión InvA	COG4789
<i>hypE</i>	Dehidratasa de carbamoilo HypE	COG0309

<i>napA</i>	Reductasa de nitrato periplasmática	COG0243
<i>yejB</i>	Proteína de permeasa de transportador ABC de membrana interna YejB	COG4174
<i>metG</i>	Ligasa de metionina-tRNA	COG0073
<i>baeS</i>	Quinasa de histidina de señalización de transducción	COG0642
<i>fliA</i>	Factor sigma de la polimerasa de ARN FliA	COG1191
<i>citC</i>	Ligasa de [citrato [pro-3S] -liaça]	COG3053
<i>da</i>	Proteína de replicación de ADN DnaC	COG1484
<i>queA</i>	Isomerasa ribosiltransferasa de S-adenosilmetionina:tRNA	COG0809
<i>dacD</i>	Carboxipeptidasa D-alanil-D-alanina DacD	COG1686
<i>pbpG</i>	Endopeptidasa D-alanil-D-alanina	COG1686
<i>deoC</i>	Aldolasa de desoxirribosa-fosfato	COG0274
<i>yihP</i>	Exportador putativo de 2,3-dihidroxiopropano-1-sulfonato	COG2211
<i>glmM</i>	Mutasa de fosfoglucosamina	COG1109
<i>dmsB</i>	Cadena B de la reductasa de dimetilsulfóxido anaeróbica	COG0437
<i>btuD</i>	Proteína de unión a ATP para la importación de vitamina B12 BtuD	Sin categoría
<i>malS</i>	Alfa-amilasa periplasmática	COG0366
<i>chbA</i>	Componente EIIA específico para N,N'-diacetilquitobiosa del sistema PTS	COG1447
<i>steC</i>	Quinasa de efector secretada SteC	Sin categoría
<i>dusB</i>	Sintasa de tRNA-dihidrouridina B	COG0042
<i>ygbE</i>	Proteína de membrana interna YgbE	Sin categoría
<i>oadA</i>	Cadena alfa de la oxaloacetato descarboxilasa	COG0511
<i>rlmB</i>	Metiltransferasa de 23S rRNA (guanósina-2'-O-) RlmB	COG0566
<i>sucA</i>	Componente E1 de la 2-oxoglutarato deshidrogenasa	COG0567
<i>dnaX</i>	Subunidad tau de la polimerasa III de ADN	COG2812
<i>entS</i>	Exportador de enterobactina EntS	COG0477
<i>mglC</i>	Proteína de permeasa del sistema de transporte de galactósidos MglC	COG4211
<i>gntR</i>	Regulador transcripcional tipo HTH GntR	Sin categoría
<i>fabR</i>	Represor transcripcional tipo HTH FabR	COG1309
<i>atpH</i>	Subunidad b de la sintasa de ATP	COG0711
<i>atpF</i>	Subunidad delta de la sintasa de ATP	COG0712
<i>ispU</i>	Sintasa de ditrans,policis-undecaprenil-difosfato específica para (2E,6E)-farnesil-difosfato	COG0020

<i>ecfA1</i>	Proteína de unión a ATP del transportador de factor de acoplamiento de energía EcfA1	COG1122
<i>queG</i>	Reductasa de epoxyqueuosina	COG1600
<i>cho</i>	Excinucleasa cho	COG0322
<i>pfeA</i>	Receptor de enterobactina férrica	COG4771
<i>fkfB</i>	Peptidil-prolil cis-trans-isomerasa de 22 kDa tipo FKBP	COG0545
<i>flgH</i>	Proteína L de flagelo	Sin categoría
<i>flgI</i>	Proteína P de flagelo	Sin categoría
<i>flhC</i>	Regulador transcripcional de flagelo FlhC	Sin categoría
<i>flhD</i>	Regulador transcripcional de flagelo FlhD	Sin categoría
<i>fdhF</i>	Formato deshidrogenasa H	COG0243
<i>glpX</i>	Fructosa-1,6-bisfosfatasa 1 clase 2	COG1494
<i>fnr</i>	Proteína reguladora de la reducción de fumarato y nitrato	COG0664
<i>prf</i>	Dehidrogenasa de gamma-aminobutiraldehído	COG1012
<i>srlR</i>	Represor del operón de glucitol	COG1349
<i>glnS</i>	Ligasa de glutamina-tRNA	COG0008
<i>gcvR</i>	Represor transcripcional del sistema de clivaje de glicina	COG2716
<i>relA</i>	Pirorfosfoquinasa de GTP	COG0317
<i>ccmD</i>	Proteína exportadora de hemo D	COG3114