# *Genomic epidemiology: from transmission to the evolution of pathogenic microorganisms*


Trabajo realizado por

**María Lorena Mejía Castañeda**


Para optar al título de Doctor por la Universitat de València

Programa de doctorado en "Biodiversidad y Biología Evolutiva"


Director:

**Fernando González Candelas**



VNIVERSITAT DE VALÈNCIA

**Marzo, 2022**

D. Fernando González Candelas, Doctor en Ciencias Biológicas y Catedrático del Departamento de Genética de la Universitat de València.

CERTIFICA

Que Dª. María Lorena Mejía Castañeda, Licenciada en Biotecnología por la Universidad San Francisco de Quito, ha realizado bajo mi dirección el trabajo titulado:

"*Genomic epidemiology: from transmission to the evolution of pathogenic microorganisms*", para optar al Grado de Doctor por la Universitat de València.

Y para que conste, en el cumplimiento de la legislación vigente, firmo el presente certificado en Valencia, a 23 de marzo de 2022.

Fdo.: Dr. Fernando González Candelas

A Iván y a Arturo, porque la vida les quedó corta.

A Damien, porque hay todo un camino por recorrer.

A mi bebé, porque hay toda una vida para caminar juntos.

# DECLARACIÓN

El trabajo presentado en esta tesis ha sido desarrollado en la Unidad Mixta de Investigación "Infección y Salud Pública" FISABIO-Universitat de València, Instituto de Biología Integrativa de Sistemas (CSIC-UV), ubicada en el área Genómica y Salud de FISABIO, en Valencia, España, y en el Instituto de Microbiología de la Universidad San Francisco de Quito, en Quito, Ecuador.

Este trabajo no hubiera sido posible si no hubiera sido beneficiada con una beca de la Fundación Carolina, en convenio con la Universidad San Francisco de Quito y la Universitat de València, en la convocatoria 2016-2017.

Los fondos para la investigación proceden del proyecto BFU2017-89594R (MICIN, Gobierno de España), PROMETEO 2016-0122 (Generalitat Valenciana, España), y de la Unión Europea a través del Programa Operativo del Fondo Europeo de Desarrollo Regional (FEDER) de la Comunidad Valenciana (España) 2014-2020. Además, parte de esta investigación fue financiada por el Grupo Asesor de la OMS sobre Subvención de Vigilancia Integrada de la RAM (AGISAR) AGISAR-JM/lg. Apoyo adicional fue brindado por el Centro de Investigación en Red en Epidemiología y Salud Pública (CIBERESP).

# AGRADECIMIENTOS

Es increíble todo lo que siento al terminar esta tesis. Hace algunos años empecé este capítulo con emoción, miedo, determinación y un deseo enorme de aprender. Me apasioné con la epidemiología molecular cuando cursaba mi maestría en microbiología. Por eso mi primer agradecimiento es para Gabriel Trueba, la persona que me permitió conocer y enamorarme de la microbiología y luego de la epidemiología molecular. Sin él, sería otro tipo de científica y muy probablemente estaría contando otra historia.

Ya desde la maestría tenía claro que mi deseo para el doctorado era trabajar con una persona que sienta pasión por la epidemiología molecular/genómica. Para él, mi segundo agradecimiento. Conocí a Fernando en Ecuador en un congreso organizado por mi Universidad. Fue invitado precisamente por su trabajo en epidemiología genómica. Cuando lo escuché, supe que eso quería hacer por el resto de mi vida académica. Gracias, Fernando, por el tiempo, la paciencia, las enseñanzas, la experiencia, la puerta de la oficina siempre abierta.

En España, hay muchas personas a quien también deseo agradecer. Carlos, Iván, Paula, Neris, Bea y Martha, miembros del grupo Epimol, que me mostraron desde cómo abrir una terminal hasta la satisfacción de sentirme querida en una tierra lejana.

Siento un agradecimiento muy especial por Ximena y Alberto. Gracias por darme un hogar en Valencia.

Todo este trabajo no habría sido posible sin el gran apoyo de Sonia Zapata. Su absoluta confianza en mí fue clave para esta oportunidad y todo el camino recorrido, porque tal vez sin darse cuenta me apoyó también para convertirme en una mejor versión de mí mismo. Su mano extendida e incondicional fueron fundamentales en esta última etapa de mi tesis. Gracias.

Gracias también a mis colegas en el Instituto de Microbiología de la USFQ, que duplicaron su trabajo para que yo pueda estar ausente

Gracias a mis hermanas y hermano de otros papás y mamás, que siempre estuvieron pendientes de cada paso dado y meta alcanzada. Gracias porque el mínimo logro era digno de celebrar. Gracias por su paciencia durante mi ausencia. Gracias por ser parte de cada capítulo de mi vida.

Gracias tambień a los fiesteros de FISABIO por las salidas, encuentros, comidas y aventuras que me hicieron mejor persona al compartir juntos la vida del estudiante de doctorado, especialmente a Bob, Pau, Nico, Bea.

Gracias a mis abuelas Piedad y Yolita, porque son ejemplo de vida, de constancia, de amor incondicional y de hacer sentir familia hasta los que no son.

Gracias infinitas a mi familia, a mi mamá, papá y hermanos. Gracias por ser mi apoyo eterno. Sin su cariño y ejemplo, mi vida no sería la misma. Gracias por su amor, su locura, su pasión, su constancia y su paciencia. Gracias por su trabajo durante tantos años para inculcar en mí el deseo de siempre dar lo mejor, de volar alto, de no ser indiferente, de aceptarme y amarme y amar y respetar a los demás. Todo lo que he logrado también es su logro. Mi hermano Jorge merece una mención especial, gracias por siempre preguntar: ¿Cómo te puedo ayudar, bolita?, ¿Qué hago?, y hacerlo. Gracias por ser el mejor significado de apoyo.

Gracias a mi compañero de vida. Empecé y terminé esta etapa contigo. Gracias por siempre inspirarme a seguir. No fue fácil y el sentirte siempre cerca, a pesar de los 9000 kilómetros que nos separaban, fue esencial para recorrer este camino con la suficiente determinación para llegar. Gracias por caminar junto a mi. Sigamos alcanzando mis sueños, tus sueños y los nuestros.

Gracias a la vida por este capítulo que culmina con esta tesis y también por el nuevo capítulo que empieza con intensidad, alegría, emoción, amor, en el que estoy dispuesta a entregar todo y a enseñar a mi bebé a amar la vida en todas sus formas y con todas sus fuerzas.

**Table of Contents**

## Abstract

Genomic epidemiology has revolutionized the microbial world. Its applications allow to know and better understand the behavior, transmission and evolution of microorganisms. Despite its great and recognized potential, its use is still limited, especially in developing regions.

The main objective of this thesis was to study and analyze the complete genomes of three pathogenic and zoonotic microorganisms to better understand their epidemiology, pathogenic potential and evolution.

In the first chapter, we characterized at the genomic level *Salmonella enterica* isolated from food, farms, and clinical samples in Ecuador. In addition, we presented evidence of the presence of virulence factors and a megaplasmid with genes encoding antimicrobial resistance and tolerance to environmental stress. Due to the phylogenetic relationship between isolates, we conclude the transmission of *S. enterica* through the consumption of chicken in Ecuador, and the urgent need for better control of the use of antimicrobials in poultry farms.

In the second chapter, genomic epidemiology was applied to isolates of *Listeria monocytogenes*, a microorganism that has received little attention in Ecuador despite its high rate of hospitalization and mortality worldwide. We compared the genomes of clinical and food isolates and inferred the epidemiological relationship through the food chain to humans in this country. We characterized isolates that, despite their susceptibility to antimicrobials, have significant pathogenic potential due to the presence of pathogenicity islands associated with severe disease. In addition, we established the need to implement routine surveillance of this pathogen and to search for other foods as possible sources of transmission of listeriosis in Ecuador.

The third chapter focused on the evaluation of genetic recombination among pathogenic species of *Leptospira*. The analysis was performed amid species, intra *L. interrogans* and among genes of the LPS locus, important for the diagnosis and development of vaccines for leptospirosis. With the results of this section, we suggest the existence of a selective pressure promoting genetic variation of this locus and of several genes, making of Leptospira a successful zoonotic pathogen.

# Resumen

# Introducción

## Enfermedades infecciosas

Las enfermedades infecciosas se pueden clasificar según su modo de transmisión. La transmisión de patógenos microbianos incluye tres modos principales: de persona a persona (contacto directo, indirecto o por gotículas en el aire), por vehículos (transmitidos por el agua, los alimentos, el aire, el suelo) o por vectores, como artrópodos/insectos. En este trabajo nos hemos centrado en dos patógenos transmitidos por alimentos y uno zoonótico.

Las enfermedades infecciosas y parasíticas causaron el 9.2% de las muertes a nivel mundial en el 2019. De 1,415 especies patógenas conocidas, 538 son bacterias que representan el 38% del total. Estas cifras son subestimadas porque en varias regiones del mundo no siempre se identifica la causa exacta de la muerte.

## Patógenos transmitidos por alimentos

A pesar de los avances tecnológicos en la investigación y prevención de enfermedades transmitidas por alimentos (ETAs), éstas siguen siendo un problema de salud pública mundial, especialmente en los países en desarrollo. Sus síntomas van desde leves y autolimitados, como náuseas, fiebre o diarrea, hasta debilitantes y graves, como insuficiencia hepática, o incluso aborto o muerte fetal en mujeres embarazadas. La intensidad y los impactos de estas enfermedades varían mucho entre regiones y también entre diferentes grupos de personas. Los bebés, niños pequeños, mujeres embarazadas, adultos mayores y personas con un sistema inmunitario debilitado tienen un mayor riesgo de desarrollar complicaciones o morir a causa de ETAs.

Existe un esfuerzo continuo de los organismos internacionales y nacionales para estimar el impacto de las enfermedades transmitidas por los alimentos en todo el mundo, no solo en la salud sino también en el desarrollo socioeconómico, y para comprender cómo los patógenos evolucionan y se diseminan de la granja a la mesa. Debido a la falta de datos sobre la prevalencia y la epidemiología de las enfermedades transmitidas por los alimentos en los países en desarrollo, la carga de enfermedades transmitidas por los alimentos no se puede estimar con precisión y, por lo general, los brotes de enfermedades transmitidas por los alimentos en estos países no se detectan o no se notifican. Esta falta de información puede deberse a dificultades para la detección temprana y la cantidad insuficiente de recursos en las áreas de investigación.

A pesar de estas limitaciones, la OMS estimó la carga de las enfermedades transmitidas por los alimentos en 2010, determinando que una de cada diez personas enferma tras el consumo de alimentos inseguros, provocando 600 millones de casos de enfermedad cada año y 420,000 muertes.

*Salmonella* y *Listeria* son los dos patógenos bacterianos transmitidos por los alimentos que se asocian con mayor frecuencia con productos agrícolas, porque estos productos generalmente se consumen crudos o con un procesamiento mínimo. La contaminación cruzada o la recontaminación son dos eventos importantes que contribuyen a estas ETAs y muestran lo difícil que es erradicar estos patógenos en las instalaciones de procesamiento de alimentos. La información precisa sobre la prevalencia de *Salmonella* y *Listeria* en los alimentos y el medio ambiente puede informar con bastante precisión a los responsables de la formulación de políticas para el control de la inocuidad de los alimentos y los esfuerzos de prevención.

**Patógenos zoonóticos**

Las zoonosis son un importante problema de salud pública como resultado de nuestros estrechos contactos y relaciones con los animales. Una infección zoonótica es el resultado del salto de un patógeno desde un animal a un humano por contacto directo con animales o su lana, o piel contaminada, por mordeduras de animales, o vectores como mosquitos, piojos o garrapatas, pero también por consumo de agua o alimentos contaminados. El 61% de los agentes infecciosos (868/1415) son zoonóticos, y el 75% de los patógenos emergentes (132/175) causan zoonosis.

Aunque muchas zoonosis implican un gran riesgo para la salud pública, algunas se consideran desatendidas, especialmente en regiones en desarrollo. Es fundamental unir esfuerzos entre países afectados por la misma enfermedad y coordinar medidas para su control. Dado que las estrategias de control requieren un apoyo financiero que generalmente no está disponible en los países en desarrollo, los países desarrollados deben involucrarse para lograr un control eficiente de las zoonosis.

**Epidemiología genómica**

La secuenciación genómica ha revolucionado nuestro conocimiento del mundo microbiano gracias a la disponibilidad y menores costos de la secuenciación. La secuenciación de genomas completos de patógenos bacterianos permite la evaluación de la información genética total de los aislados hasta el nivel de un solo nucleótido. El primer genoma bacteriano estuvo disponible en 1995, y actualmente tenemos acceso a secuencias de ADN y ARN de más de 125,000

especies de bacterias y virus en distintas bases de datos de acceso libre. Las aplicaciones de los análisis genómicos permiten conocer y entender mejor el comportamiento, transmisión y evolución (resistencia a antimicrobianos, tolerancia a estrés ambiental, emergencia de nuevas características) de los microorganismos.

A pesar de su gran y reconocido potencial, su uso todavía es limitado, especialmente en regiones en desarrollo por acceso a la tecnología y competencia en genómica y bioinformática. Por otro lado, continuamente se desarrollan nuevos programas, lo que dificulta la estandarización de las herramientas utilizadas en todo el mundo. Además, a pesar de que la gran mayoría de la información y los programas para los análisis genómicos son de acceso abierto, existen grupos de investigación que limitan este acceso a sus herramientas bioinformáticas.

## Objetivo

El objetivo principal de esta tesis fue estudiar y analizar los genomas completos de tres microorganismos patógenos y zoonóticos (*Salmonella enterica, Listeria monocytogenes,* y *Leptospira*) para comprender mejor su epidemiología, potencial patogénico y evolución.

## Capítulo 1. Epidemiología genómica de *Salmonella* Infantis en Ecuador: de las granjas avícolas a las infecciones humanas

*Salmonella* es un patógeno Gram negativo, transmitido por alimentos, que pertenece al orden Enterobacterales y la familia Enterobacteriacea. *Salmonella enterica* representa una de las cuatro principales causas mundiales de enfermedades diarreicas, aunque algunos serotipos pueden causar fiebre tifoidea. No sólo es importante por su implicación en un elevado número de infecciones humanas sino también por su diversidad (más de 2,500 serotipos) y resistencia a varios antibióticos diferentes. Algunos de los alimentos implicados en las infecciones por *Salmonella* son los huevos, las aves y otros productos de origen animal, así como las verduras.

Las infecciones transmitidas por los alimentos causadas por *S. enterica* son de importancia primordial en todo el mundo. La OMS estima que *Salmonella* causa más de 153 millones de casos de enfermedad y 120,281 muertes. Dado que los alimentos pueden estar contaminados en varias partes de la cadena alimentaria, es necesario un enfoque "de la granja a la mesa" para comprender la epidemiología de *Salmonella*. Aunque *Salmonella* puede contaminar verduras, los animales, especialmente las aves de corral, se consideran fuentes importantes de infecciones humanas. En los países latinoamericanos, la avicultura es una de las principales fuentes de

proteínas de origen animal. Este es el caso de Ecuador, donde la carne de ave es el producto básico más consumido, con un consumo anual per cápita de 30,4 Kg.

En el primer capítulo, caracterizamos a nivel genómico *Salmonella enterica* aisladas de muestras de alimentos, granjas y clínicas de Quito, capital de Ecuador. Además, evidenciamos la presencia de factores de virulencia y de un megaplásmido con genes que codifican resistencia a antimicrobianos y tolerancia a estrés ambiental. Por la relación filogenética entre aislados, confirmamos la importancia de la transmisión de *S. enterica* por el consumo de pollo en el Ecuador y concluimos la urgente necesidad de un mejor control del uso de antimicrobianos en granjas avícolas.

## Metodología

Las muestras para el aislamiento de *Salmonella* se recolectaron semanalmente desde noviembre de 2017 hasta noviembre de 2018 siguiendo los lineamientos del documento: "*Integrated Surveillance of Antimicrobial Resistance in Foodborne Bacteria*" de la Organización Mundial de la Salud.

Granjas avícolas: 133 parvadas de 69 granjas fueron muestreadas durante el período de estudio. Por cada parvada muestreada, se recolectaron al azar 25 ciegos de 25 pollos a nivel del matadero y se transportaron al laboratorio en una hielera dentro de las siguientes dos horas. Las muestras de heces son recomendadas por la OMS porque brindan una mayor recuperación de aislados y representan mejor la contaminación de animales individuales a nivel de granja. En el laboratorio, se procesó como un grupo de muestras de 25 g para el aislamiento bacteriológico.

Canales de pollo: Se recolectaron 335 canales en tres tipos de mercados de la siguiente manera: 125 muestras de supermercados, 126 muestras de pequeñas tiendas y 84 muestras de mercados abiertos. El muestreo de las canales de pollo se realizó de manera alterna entre el norte y el sur de la ciudad. Cada canal fue recolectada en su bolsa original y transportada al laboratorio en una hielera dentro de las siguientes dos horas. En el laboratorio, se recolectaron asépticamente 25 g de piel de pechuga de cada canal para análisis bacteriológico.

Muestras de heces humanas: se recolectaron 302 muestras en dos centros de salud ubicados en la periferia urbana de Quito (centro de salud Guamaní, al sur, y centro de salud Calderón, al norte) de pacientes con dos o más episodios de diarrea o vómitos en las últimas 24 horas. Las muestras de heces humanas se transportaron al laboratorio en una hielera en las siguientes dos horas. Se recolectaron aproximadamente 25 g de heces para análisis bacteriológico.

De acuerdo con la legislación nacional, no se requería aprobación ética para las granjas avícolas y el muestreo de canales de pollo, ya que no se sacrificaron animales durante este estudio. Para el componente humano, el proyecto fue aprobado por el comité de bioética del Instituto Nacional de Salud Pública "Leopoldo Izquieta Pérez" (Protocolo ID:CEISH-INSPI-005). Los participantes fueron informados sobre el objetivo del estudio. Todos los voluntarios dieron su consentimiento por escrito. Toda la información personal fue anonimizada.

**Aislamiento e identificación de *Salmonella enterica***

El aislamiento de *Salmonella* se realizó mediante un método basado en el protocolo ISO 6579-1:2007. Brevemente, se añadieron 225 mL de agua de peptona tamponada (BPW; Difco, BD, Sparks, MD) a cada muestra, se homogeneizó a mano durante 1 minuto y se incubó a 37 °C durante 20 h. Luego, 100 µL de cada enriquecimiento se inocularon en agar Semisólido Rappaport-Vassiliadis Modificado (MSRV; Oxoid, Basingstoke, UK) en tres puntos equidistantes y se incubaron a 42 °C durante 24 h. Posteriormente, se examinaron las placas para detectar la presencia de un halo blanco alrededor de al menos un punto de inoculación. Una porción tomada del borde del halo blanco se sembró en agar Xilosa Lisina Desoxicolato (XLD, Difco) y se incubó a 37 °C durante 24 h. Después de la incubación, una colonia sospechosa de *Salmonella* fue confirmada bioquímicamente mediante las pruebas *Triple Sugar Iron* (TSI, Difco, BD), *Iron Lysine* (LIA, BBL, BD), Urea (BBL, BD) y *Sulphur Indole Motility* (SIM, BBL, BD). Las colonias aisladas se confirmaron por PCR. Se calculó el intervalo de confianza del 95 % ($IC_{95\%}$) para la prevalencia de *Salmonella* en cada componente.

**Pruebas de susceptibilidad antimicrobiana**

Todos los aislados de *Salmonella* fueron examinados por el método de difusión en disco de Kirby-Bauer con los siguientes antibióticos: sulfametoxazol + trimetoprima (25 µg), gentamicina (10 µg), ciprofloxacina (5 µg), cefotaxima (30 µg), tetraciclina (30 µg), estreptomicina (10 µg), cloranfenicol (30 µg), cefoxitina (30 µg), amikacina (30 µg), nitrofurantoína (300 µg), azitromicina (15 µg), fosfomicina (200 µg), ertapenem (10 µg), amoxicilina + ácido clavulánico (30 µg). Se utilizó la cepa *E. coli* ATCC 25922 como control de calidad. Los resultados y métodos se interpretaron de acuerdo con los criterios CLSI (*Clinical and Laboratory Standards Institute*) de 2019, considerando todos los fenotipos intermedios como resistentes.

**Detección de genes de betalactamasas de espectro extendido (BLEE)**

Los aislados de *Salmonella* que presentaban fenotipos resistentes a los antibióticos betalactámicos se analizaron adicionalmente mediante PCR para la identificación de genes BLEE: *bla*$_{\text{CTX-M}}$ (*bla*$_{\text{CTX-M-1}}$, *bla*$_{\text{CTX-M-2}}$, *bla*$_{\text{CTX-M-8}}$, *bla*$_{\text{CTX-M-9}}$, *bla*$_{\text{CTX-M-14}}$), *bla*$_{\text{TEM}}$, *bla*$_{\text{CMY}}$ y *bla*$_{\text{SHV}}$. Todos los productos de PCR se purificaron y secuenciaron en Macrogen Inc (Seúl-Corea del Sur). Las secuencias obtenidas se alinearon con las secuencias de referencia con la herramienta en línea ResFinder v3.2.

**Secuenciación del genoma completo**

Para la secuenciación del genoma completo, se realizó una selección de aislados de *Salmonella* a partir de componentes animales y alimentarios. Al seleccionar aislados de granjas avícolas se consideró la primera muestra positiva de cada granja. Para las canales de pollo se seleccionó la primera muestra positiva de cada semana de muestreo en cada tipo de mercado. Todos los aislados que no fueron *S.* Infantis y todos los aislados del componente humano fueron seleccionados.

El ADN genómico se extrajo con el kit Invitrogen *PureLink Genomic DNA* (Thermo Fisher Scientific, Walthman, MA, EE.UU.) siguiendo las recomendaciones del fabricante para lisados de células bacterianas Gram negativas. El ADN se cuantificó con el fluorómetro Invitrogen Qubit 3.0 (Thermo Fisher Scientific, Walthman, MA, EE.UU.) y se secuenció con la plataforma Illumina NextSeq con el kit de preparación de bibliotecas Nextera XT, obteniendo secuencias de extremos pareados de 150 x 2 pb (Illumina, San Diego, CA, EE.UU.). Se utilizaron parámetros predeterminados para todas las herramientas y programas bioinformáticos a menos que se especifique lo contrario. Las lecturas se limpiaron con Trimmomatic para eliminar los nucleótidos ambiguos y aquellos con valores de puntuación de calidad <20. Para la evaluación de la calidad se utilizaron los programas Fastqc y Multiqc.

**Identificación de serotipos**

Los serotipos de *Salmonella* fueron identificados por PCR como se describe por Akiba *et al*. Además, los serotipos de los aislados sometidos a WGS se confirmaron mediante el análisis de sus lecturas crudas utilizando el programa SeqSero.

**Análisis MLST, genes de resistencia antimicrobiana y detección de plásmidos**

Para identificar las secuencias tipo (ST) de MLST, los genes de resistencia a los antimicrobianos y las secuencias de plásmidos, se utilizó ARIBA con las bases de datos PubMLST, ResFinder v3.2 y PlasmidFinder 2.1, respectivamente. El fenotipo de resistencia de

cada muestra se comparó con la presencia de genes de resistencia encontrados por WGS. Adicionalmente, realizamos un mapeo contra el megaplásmido p-F219 descrito por Vallejos-Sánchez *et al*. usando Burrows-Wheeler Aligner con el algoritmo BWA-MEM. Se usaron BCFtools y vcfutils de SAMtools para obtener los archivos fastq de archivos SAM, y las secuencias fasta se transformaron de fastq con Seqtk. Las secuencias se concatenaron y se obtuvo un árbol filogenético de máxima verosimilitud con IQ-TREE 2. Como obtuvimos dos grupos bien definidos para *S*. Infantis, anotamos un representante de cada grupo de plásmidos con Prokka y realizamos un análisis de genes ortólogos (cobertura del 90%, similitud en secuencias proteicas del 80%) con Proteinortho5. Se realizó una comparación manual de todos los genes presentes en los plásmidos. Realizamos el mismo análisis para aislados no Infantis.

**Análisis de megaplásmidos**

Se analizaron dos megaplásmidos (pESI y p-F219) comúnmente asociados a cepas patógenas y multirresistentes de *S*. Infantis para identificar su relación. Utilizamos D-Genies para obtener un gráfico de puntos de comparación del genoma, una alineación del genoma con progressiveMauve para identificar bloques localmente colineales y un cálculo ANI para estimar la identidad de nucleótidos promedio en secuencias compartidas por ambos plásmidos.

**Análisis de metadatos y genoma "core"**

Una cepa peruana de *S*. Infantis, FARPER-219, y dos aislados ecuatorianos (SRR4019589 y SRR4019602) analizados por el Centro de Control y Prevención de Enfermedades de EE.UU. de dos pacientes que desarrollaron salmonelosis después de viajar a Ecuador se agregaron como referencias en el análisis filogenético de todos los aislados de *S*. Infantis. A partir de lecturas limpias, se utilizó Spades para generar ensamblados. Posteriormente, se realizó la anotación del genoma con Prokka. Se realizó un análisis de genes ortólogos con las mismas condiciones que para la detección de plásmidos con Proteinortho5. Los genes del "*core*" estricto, presentes en todos los aislados, se extrajeron con la herramienta Proteinortho: grab_proteins.pl. Mafft y un script propio se utilizaron para el alineamiento múltiple de cada gen y la concatenación posterior, respectivamente. El árbol filogenético del alineamiento del genoma "*core*" se obtuvo utilizando IQ-TREE 2 con 1000 réplicas de *bootstrap*. Los metadatos para el origen de la muestra, los patrones fenotípicos de resistencia a los antibióticos y la detección de plásmidos *in silico* se agregaron al árbol final en iTOL.

## Resultados

### Prevalencia de *Salmonella* e identificación de serotipos

*Salmonella* estuvo presente en 41,4% (55/133; IC95%:33-49,7), 55,5% (186/335; IC95%:50,2-60,8) y 1,98% (6/302; IC95%:0,4-3,6) en granjas de aves de corral, canales de pollo y muestras de heces humanas, respectivamente. *S.* Infantis representó el 98,2% (*n*=54) de los aislados de granjas avícolas, el 97,8% (*n*=182) de los aislados de canales de pollo y la mitad (*n*=3) de los aislados de muestras humanas. Además, un aislado se tipificó como *S.* Enteritidis en parvadas de pollos de engorde; a nivel minorista uno y tres aislados fueron tipificados como *S.* Typhimurium y *S.* Enteritidis, respectivamente, mientras que en las muestras de heces humanas dos aislados fueron tipificados como *S.* Enteritidis y un aislado correspondió a *S.* Typhimurium monofásico 4,[5], 12:i:- (Tabla complementaria 1 del texto principal).

### Resistencia a los antimicrobianos

Para los aislados de *S.* Infantis, las tasas de resistencia antimicrobiana a nitrofurantoína, tetraciclina, sulfametoxazol + trimetoprima, estreptomicina, gentamicina, cefotaxima, ciprofloxacina y cloranfenicol oscilaron entre 64,8% y 100%. Por otro lado, las tasas de resistencia a fosfomicina y azitromicina fueron más bajas, con un rango de 0% a 42,6%. Solo un aislado de una muestra de heces presentó resistencia fenotípica a amikacina, mientras que ninguno de los aislados de *Salmonella* en este estudio fue resistente a ertapenem (Tabla 1).

Considerando las clases de antimicrobianos que se probaron, los aislados de *S.* Infantis presentaron 43 patrones de resistencia a antimicrobianos. Con la excepción de un aislado de una muestra de heces, todos los aislados mostraron fenotipos multirresistentes. Es importante destacar que el 87% y el 82% de los aislados de granjas avícolas y canales de pollo, respectivamente, presentaron resistencia de 6 a 9 clases de antimicrobianos (Tabla complementaria 2). Una cepa de *S.* Infantis aislada de canales de pollo no pudo recuperarse para este análisis.

Los serotipos de *Salmonella* distintos de *S.* Infantis también presentaron patrones multirresistentes, excepto 3 aislados de *S.* Enteritidis que solo fueron resistentes a un grupo de antimicrobianos. Para este conjunto de aislados, cada patrón resistente incluía aislados pertenecientes a un solo serotipo (Tabla 2).

Un aislado de *S.* Enteritidis, uno de *S.* Typhimurium y uno de *S.* Typhimurium monofásico 4,[5],12:i:-; y 205 aislados de *S.* Infantis fueron identificados como resistentes a betalactámicos.

Seis aislados de *S*. Infantis de canales de pollo y uno de granjas avícolas no presentaron ninguno de los genes BLEE investigados. Todos los demás *S*. Infantis y una *S*. Enteritidis aislada de una canal presentaron el gen *bla*$_{\text{CTX-M-65}}$.

**Análisis genómico**

Para el análisis genómico, se seleccionaron 144 aislados (40 del componente animal, 98 del componente alimentario y 6 del componente humano). Los datos de secuencias crudas están disponibles en el bioproyecto PRJEB37560. Las secuencias de tres muestras no fueron suficientes para realizar el análisis genómico. El número medio obtenido de lecturas por cepa fue de 1,356,678 (rango 247,022-14,106,025) y después de los pasos de control de calidad, el número medio fue 1,266,242 (rango 228,263-13,094 594) (Tabla complementaria 3). El puntaje promedio de *Phred* fue Q34.

La tipificación MLST mostró que todos los aislados de *S*. Infantis (*n*=137) pertenecían al ST32. Los cinco aislados de *S*. Enteritidis pertenecían al ST11. Además, los aislados individuales de *S*. Typhimurium y S. Typhimurium monofásico 4,[5],12:i:- pertenecían a los ST19 y ST2379, respectivamente (Tabla complementaria 1).

El genoma "*core*" estricto de todas las *S*. Infantis incluidas en el análisis correspondía a 3,552 genes y abarcaba 3,161,448 pb, 1,414 de los cuales eran variables (SNPs). El alineamiento de los genes concatenados presentes en este núcleo se utilizó para obtener un árbol de máxima verosimilitud utilizando FARPER 219 como grupo externo (Figura 3). Se eligió esta cepa porque fue aislada en Perú, país vecino de Ecuador. A pesar de su inclusión en ST32, FARPER-219 presentó divergencia genética con las cepas ecuatorianas. Los dos genomas de *Salmonella* de EE.UU. (SRR4019589, SRR4019602) se agruparon de manera indistinta con algunos de los genomas de este estudio. Cabe destacar que las cepas analizadas no se agruparon según su origen de muestreo o sus patrones de resistencia fenotípica.

Los genes de resistencia a los antimicrobianos también se confirmaron con datos de genomas completos (Archivo complementario 1). Para la mayoría de las clases de antimicrobianos (inhibición de la vía del folato, aminoglucósidos, betalactámicos, tetraciclinas, fosfomicina y fenicol), las tasas de correspondencia fueron superiores al 80%. Sin embargo, no se encontraron genes responsables de la resistencia fenotípica a quinolonas, nitrofuranos y macrólidos en los aislados secuenciados (Tabla 3).

En cuanto a los genes de virulencia, se encontraron *sifA, sseL, pipB, sopD2* y *srlP* (parte de SPI 2) en el genoma "*core*" de todas las cepas de *S*. Infantis. El operón *lpf*, que codifica las fimbrias

polares largas, el grupo de genes *fim*, que codifica las fimbrias tipo 1, y los operones *csg*, que codifica las fimbrias Tafi (fimbrias agregadas delgadas), también estaban presentes en estos genomas.

La presencia de plásmidos fue confirmada *in silico* por PlasmidFinder. Sin embargo, solo un aislado de *S*. Infantis (U2449s) presentó un determinante de plásmido (IncX1 e IncX1_1) a pesar de los patrones de resistencia a múltiples fármacos encontrados en nuestros aislados (Figura 1 y archivo complementario 1). Para analizar más a fondo la baja incidencia de plásmidos, mapeamos las lecturas crudas de nuestras muestras contra el megaplásmido p-F219 descrito por Vallejos-Sánchez *et al*. Encontramos dos plásmidos similares a p-F219. El primero, denominado plásmido A, estuvo presente en la mayoría de las cepas. Este plásmido contenía 338 genes compartidos con el megaplásmido p-F219. El segundo, denominado plásmido B, estaba presente en 7 cepas. El plásmido B carecía de 72 de los genes presentes en el plásmido A y presentaba 6 genes exclusivos (Tabla complementaria 4). Las cepas que presentaron el plásmido B pertenecían a un clado monofilético (indicado con una estrella roja en la Fig. 1). Estas cepas también comparten susceptibilidad a fosfomicinas, macrólidos, fenicoles y betalactámicos. Se aislaron cepas de *Salmonella* que albergaban el plásmido B de canales de pollo muestreadas durante diferentes semanas del año, diferentes partes de la ciudad y diferentes tipos de tiendas minoristas (datos no mostrados).

Se encontraron proteínas hipotéticas ($n$=147) en ambos plásmidos; 43 encontrados exclusivamente en el plásmido A y 5 en el plásmido B (Tabla complementaria 4).

La comparación del plásmido p-F219 con el megaplásmido pESI, que se encuentra comúnmente en cepas patógenas de *S*. Infantis, mostró que comparten más del 79% de sus secuencias (Figura 1 complementaria). Además, notamos una gran inversión genómica en el plásmido p-F219 en comparación con el plásmido pESI que también se observa en el alineamiento con progressiveMauve (Figura 2 complementaria). A partir del cálculo de ANI, se encontró un 99,41% de identidad en los genes ortólogos presentes en ambas muestras.

La detección de plásmidos *in silico* en aislados no Infantis mostró la presencia de dos plásmidos en todas las cepas de *S*. Enteritidis, mientras que *S*. Typhimurium 4,[5],12:i:- y *S*. Typhimurium monofásicos presentaron 1 y 5 plásmidos, respectivamente ( Tabla complementaria 1).

## Capítulo 2. *Listeria monocytogenes*, un patógeno silencioso transmitido por alimentos en Ecuador.

*Listeria monocytogenes* es un Gram-positivo del orden Bacillales y de la familia Listeriaceae, patógeno intracelular facultativo transmitido por los alimentos que puede causar enfermedades y afecciones graves como septicemia, encefalitis y meningitis, aborto e infecciones neonatales en humanos, así como enfermedades graves en otros mamíferos y aves. Aunque las infecciones por *L. monocytogenes* presentan incidencia generalmente baja, la tasa de letalidad es una de las más altas entre las infecciones transmitidas por alimentos y afecta principalmente a niños pequeños, personas inmunocomprometidas, mujeres embarazadas y ancianos.

Se puede encontrar *Listeria* en alimentos crudos y procesados, pero los vehículos más frecuentes son los productos alimenticios listos para el consumo (RTE) como quesos, carnes, mariscos y productos frescos. Entre estas matrices alimentarias, los quesos son muy importantes para transmisión de *Listeria* principalmente porque suelen almacenarse a temperaturas de refrigeración durante varios días permitiendo el crecimiento de este microorganismo psicrofílico y normalmente no reciben un tratamiento adicional antes de su consumo.

*L. monocytogenes* representa una amenaza notable en los alimentos, porque puede persistir en los entornos de procesamiento durante largos períodos de tiempo, debido en parte a su capacidad para sobrevivir en una gran diversidad de condiciones estresantes (desinfección, pH, actividad de agua y temperatura) y la capacidad de formar biopelículas en las superficies. Esta bacteria puede colonizar las instalaciones de procesamiento de alimentos cuando se introducen materias primas contaminadas y cuando no existen barreras físicas ni procedimientos de desinfección entre los alimentos procesados y las áreas de materias primas.

En el segundo capítulo, se aplicó la epidemiología genómica en aislados de *Listeria monocytogenes*, un microorganismo sin la suficiente atención en el Ecuador. Comparamos los genomas de aislados clínicos y de alimentos e inferimos la relación epidemiológica a través de la cadena alimenticia a humanos en este país. Caracterizamos aislados que, a pesar de su susceptibilidad a antimicrobianos, presentan potencial patogénico importante por la presencia de islas de patogenicidad asociadas con enfermedad severa. Además, establecimos la necesidad de implementar una vigilancia rutinaria de este patógeno y de buscar otros alimentos como posibles fuentes de transmisión de listeriosis en Ecuador.

## Metodología

### Cepas bacterianas

Desde el Instituto Nacional de Investigación en Salud Pública de Ecuador, tuvimos acceso a todos los aislados de *L. monocytogenes* de la colección de cultivos desde 2015 hasta 2018 (14 aislados de muestras de queso y 20 aislados clínicos de casos de listeriosis). También tuvimos acceso a 31 aislados de muestras de queso artesanal de 8 provincias de Ecuador de un estudio previo. En total se evaluaron 65 aislados de *Listeria monocytogenes*. A partir de estos aislados, se evaluó la prueba de susceptibilidad antimicrobiana mediante la técnica de microdilución, utilizando paneles de antibióticos liofilizados para bacterias Gram positivas (Sensititre GPALL1F, Thermo Scientific ®) y su serogrupo fue determinado por PCR.

### Secuenciación del genoma completo

El ADN genómico de todos los aislados se purificó utilizando el kit de purificación de ADN genómico Wizard (Promega Corporation, Madison, WI, EE.UU.) siguiendo las recomendaciones del fabricante para bacterias Gram positivas. A continuación, se utilizó un fluorómetro Invitrogen Qubit 3.0 (Thermo Fisher Scientific, Walthman, MA, EE.UU.) para la cuantificación del ADN. Las muestras de ADN se secuenciaron con la plataforma Illumina NextSeq utilizando el kit de preparación de bibliotecas Nextera XT y se obtuvieron secuencias de extremos emparejados de 150 x 2 pb. Se utilizó Trimmomatic para la limpieza de lecturas y Fastqc y Multiqc para la evaluación de la calidad. Se utilizaron los parámetros predeterminados para los siguientes programas a menos que se especifique lo contrario.

### Determinación del ST, genes de resistencia antimicrobiana, factores de virulencia y detección de plásmidos.

Se utilizó ARIBA versión 2.11.1 con las bases de datos públicas CARD, MEGAres y ResFinder para identificar genes que pueden conferir resistencias a antimicrobianos. Las bases de datos VirFinder para genes de virulencia, PlasmidFinder para determinantes de plásmidos y PubMLST del Institut Pasteur para *Listeria monocytogenes* para la determinación de ST también se utilizaron con el software ARIBA.

También se determinó la presencia de plásmidos previamente descritos en otras cepas de *L. monocytogenes*. Realizamos una búsqueda manual en bases de datos públicas y descargamos 29 plásmidos del NCBI. Para ver qué tan similares eran estos plásmidos, se usó Proteinortho5 para determinar las relaciones de ortología, donde se establecieron valores de 80% para la cobertura y 80% para la similitud en las secuencias de proteínas. Con estos resultados,

eliminamos manualmente los plásmidos con patrones genéticos similares. Luego, seleccionamos 13 plásmidos para el paso de mapeo. Mapeamos nuestros datos sin procesar contra los plásmidos siguiendo la estrategia descrita en Detección de megaplásmidos de *Salmonella,* del capítulo anterior. Con las secuencias fasta de cada cepa, realizamos un alineamiento con cada plásmido para evaluar su cobertura. Finalmente, clasificamos la presencia de plásmidos en rangos para cada cepa (20-49%, 50-79% y 80-100%). Las coberturas de plásmido inferiores al 20% se descartaron como ausentes. Después de analizar la gran longitud cubierta de siete plásmidos (pMF6172, pl2015TE24968, PAUSMDU00000224_01, pPIR00540, pLM58, pMF4545, pPIR00541), realizamos una comparación adicional basada en ortólogos y en la identificación promedio de nucleótidos (ANI). También usamos progressiveMauve para identificar bloques localmente colineales e inversiones.

**Análisis filogenético**

Se utilizó Spades para obtener ensamblados a partir de lecturas limpias. Se descargaron 26 genomas cerrados de *L. monocytogenes* como referencias. Luego, se utilizó Prokka para la anotación del genoma. A continuación, las determinaciones de ortólogos entre todos los aislados de *L. monocytogenes* y entre los aislados de ST2 se realizaron de forma independiente con Proteinortho5 (90% de cobertura y 80% de similitud en las secuencias de proteínas). Los genes presentes en todos los aislados se extrajeron con la herramienta grab_proteins.pl de Proteinortho5. Luego, todos los genes fueron alineados y concatenados en un archivo multifasta con Mafft y un script interno. El alineamiento del genoma "*core*" se utilizó para obtener un árbol filogenético de máxima verosimilitud con IQ-TREE 2 con 1000 réplicas de *bootstrap* para evaluar los soportes de las ramas. Se utilizó iTOL para visualizar los árboles finales y agregar los metadatos.

**Determinación de factores genéticos**

Para detectar genes que codifican los principales factores de virulencia de *Listeria* (LIPI-1, LIPI-2, LIPI-3, LIPI-4 e internalinas *inlA* e *inlB*), genes asociados a la formación de biopelículas (f*laA, luxS, cheY*) y genes de islotes de supervivencia al estrés (SSI-1, SSI-2, SSI-F2365), se ejecutó blastn con los genes de referencia como sujetos y los conjuntos de secuencias como consulta. El porcentaje de identificación se fijó en el 75%.

## Resultados

### Datos crudos

Los datos brutos de las lecturas de secuenciación están disponibles en el bioproyecto PRJEB48671. Obtuvimos datos de 65 aislados de *L. monocytogenes*, 1 *L. seeligeri*, 1 *L. welshimeri* y 5 *L. innocua*. En este capítulo, nos enfocamos principalmente en los aislados de *L. monocytogenes*. El número promedio obtenido de lecturas en millones por cepa fue 1.5 (rango 0.3-2.6) (Tabla complementaria 7). El puntaje promedio de *Phred* fue Q34. Se excluyó un aislado de *L. monocytogenes* (Lm73) de los análisis posteriores porque el genoma "*core*" de todos los aislados, incluido Lm73, estaba compuesto por 1,227 genes ortólogos, y sin Lm73 estaba compuesto por 2,112 genes. Este aislado carecía de 885 genes ortólogos que estaban presentes en el resto de cepas (datos no mostrados).

### Secuencias Tipo (ST) y origen de la muestra

Se analizaron un total de 64 aislados de *L. monocytogenes* (Archivo complementario 3). Fueron aislados de varios tipos de quesos, incluyendo queso blando (*n*=31), queso amasado (*n*=7), queso picado (*n*=3), queso sin sal (*n*=2), cuajada (*n*=1), pero también de casos de listeriosis, incluidas muestras de sangre (*n*=19) y líquido cefalorraquídeo (*n*=1). Estas muestras se recolectaron entre 2015 y 2018 y se detectaron 8 secuencias tipo (ST) diferentes. La Tabla 4 muestra los ST entre alimentos y muestras clínicas. El ST predominante fue el ST2 (84,62% 55/65), y se encontraron otros ST en menos del 4% de las muestras. ST2 fue el único ST encontrado en alimentos y muestras clínicas.

### Resistencia antimicrobiana

Todos los aislados se analizaron previamente y mostraron susceptibilidad a penicilina, ampicilina, eritromicina, trimetoprim/sulfametoxazol y meropenem. Los determinantes genéticos para AMR también se detectaron con datos genómicos con tres bases de datos como se muestra en la Tabla 5. Los determinantes *fosX* y *mprF_2* se encontraron solo por la base de datos CARD, mientras que GYRA_23 y TUFAB_7 solo fueron encontrados por Megares, y *norB* y *tetS* fueron detectados por 2 y 3 bases de datos, respectivamente.

Dos aislados (Lm03 y Lmo22) presentaron 9 y 3 otros determinantes de resistencia además de los detectados en la mayoría de las muestras, y carecen del gen *tetS* (archivo complementario 3).

**Análisis genómico**

A partir de las lecturas crudas, PlasmidFinder y VirulenceFinder no detectaron plásmidos ni determinantes de virulencia, respectivamente (datos no mostrados). El genoma "*core*" de todos los aislados de *L. monocytogenes* (*n*=64) y referencias (*n*=26) incluye 2,112 genes que abarca 1,896,388 nt con 175,308 sitios variables (9.3 % del alineamiento de genoma completo). Este alineamiento generó un árbol de máxima verosimilitud que se muestra en la Figura 4, donde los valores de *bootstrap* de 90% o superiores se muestran con un círculo negro. Los aislados se agruparon según los linajes evolutivos previamente determinados. La mayoría de nuestras muestras se agruparon en el linaje I. La fecha de recolección, el tipo de secuencia, la fuente del aislado, el linaje genético, la presencia de SSI-1, SSI-F2365, LIPI-1, LIPI-3 , LIPI-4 e internalinas también se representan en la Figura 4.

Con respecto a los islotes de supervivencia al estrés, las muestras presentaron SSI-1 o SSI-F2365, excepto el aislado Lm21 que presentó SSI-F2365 pero también un gen de SSI-1. Solo 3 de nuestros aislados (2 aislados clínicos ST3 y 1 aislado alimentario ST796) contenían SSI-1. Ninguna de las muestras presentó SSI-2, que es común en *L. innocua*.

En cuanto a la presencia de islas de patogenicidad y factores de virulencia, los aislados exhibieron patrones heterogéneos. Todos exhibieron al menos 5 de 6 genes de LIPI-1 (*prfA, plcA, hly, mpl, actA, plcB*), pero todas las referencias mostraron la isla completa. Solo dos aislados ecuatorianos (Lm07 y Lmo20) albergaron los 6 genes de LIPI-1. La mayoría de nuestros aislados no revelaron presencia del gen *actA* o el porcentaje de identidad fue inferior al 75% con respecto al gen de referencia. Los genes de LIPI-2, que generalmente se encuentran en *L. ivanovii*, estaban ausentes en todos los aislados. LIPI-3 solo estuvo presente en 13 aislados del linaje 1 (7 referencias de ST66, ST6, ST4 y ST1). Los aislados clínicos Lmo15 (ST3), Lmo10 (ST3), Lmo31 (ST4), Lmo16 (ST1) y el aislado alimentario Lmo20 (ST6) presentaron todos los genes de LIPI-3. LIPI-4 solo estuvo presente en 3 aislados (2 referencias con ST87 y ST4 y Lmo31). Las internalinas A y B estaban presentes en todos los aislados y referencias. Todos los aislados y referencias también mostraron la presencia de genes asociados con la formación de biopelículas (*flaA*, *luxS*, *cheY*) (datos no mostrados).

Dado que la mayoría de nuestras muestras pertenecen al ST2, obtuvimos un segundo árbol filogenético basado en el alineamiento de 57 cepas ST2, incluidas dos referencias, que se presenta en la Figura 5 (los valores de *Bootstrap* mayor o igual al 90% se muestran en ramas con un círculo negro). El alineamiento múltiple incluía 2,376,722 pb con solo 1,022 sitios variables (menos del 0,05 % del alineamiento completo). La fuente de aislamiento y la fecha

de recopilación se especifican como metadatos. Los clados (A-F) están coloreados para una mejor visualización. Las cepas en los clados B y E se aislaron solo de quesos, mientras que el resto de los clados incluye muestras humanas y de alimentos. La presencia de trece plásmidos se muestra en rangos (20-49 %, 50-79 % y 80-100 %) que representan una similitud global. Todos los aislados presentaron el NCRC7974_plasmid3 completo, mientras que el clado E, parte de F y Lm08_S6 (del clado B) mostraron el mismo patrón de plásmidos, una cobertura del 80-100 % de los plásmidos pLM58, pl2015TE24968 (p24968), pMF6172, pMF4545, pPIR00541, pPIR00540, PAUSMDU00000224_01 (p224_01); 50-79% de plásmidos pLM5578, N1-011A y J1776; y 20-49% de los plásmidos pOB050226 y pLmN1546.

Al analizar con más detalle el patrón de plásmidos presentes, nos dimos cuenta de que muchos genes estaban presentes en la mayoría de los plásmidos. Once genes que abarcan 10 kb están presentes en estos 7 plásmidos, mientras que siete genes (8 kb) son compartidos por 6 plásmidos y 20 genes (14,7 kb) están presentes en 5 plásmidos. Entre estos genes, encontramos seis transposasas, dos ADN invertasas, una ADN polimerasa IV, una transposón resolvasa, una NADH peroxidasa, una proteasa, un gen implicado en la regulación de la resistencia al cadmio, una ATPasa transportadora de cadmio y 24 proteínas hipotéticas (Archivo Suplementario 4). Después del cálculo de ANI entre los siete plásmidos, detectamos un alto porcentaje de similitud (99.51-100%) en diferentes tamaños de fragmentos compartidos (9743-50589 pb) por cada dos plásmidos (Tabla complementaria 8). El alineamiento de los siete plásmidos mostró grandes bloques localmente colineales que habían perdido sintenia a lo largo de los genomas, como se ve en la Figura 6. Dos plásmidos parecen haber sufrido inversiones.

## Capítulo 3. El impacto de la recombinación genética en *Leptospira* patógena

El tercer capítulo se centró en la evaluación de la recombinación genética entre especies patógenas de *Leptospira*. La recombinación en bacterias es la transferencia física de ADN entre dos moléculas. Este proceso genético ha jugado un importante rol en virulencia, evasión de respuesta inmune y tropismo al hospedador en algunas bacterias patógenas. Por otro lado, *Leptospira* es un género de bacterias ambientales que pueden clasificarse en comensales y patógenas. Las especies patógenas son causantes de la leptospirosis, una enfermedad tropical, considerada la zoonosis más común a nivel mundial. Su sintomatología puede variar de leve (síntomas similares a influenza) a severos (daño renal, hepático y muerte). La intensidad de la enfermedad depende del serovar que causa la infección, la edad del paciente y su sistema inmunológico. Se han logrado identificar más de 300 serovares en 35 especies. La clasificación en serovares se basa en la expresión de distintos epítopos del LPS (lipopolisacárido). El LPS

se ha asociado con inmunidad protectora, ya que los anticuerpos frente a un serovar protegen de todos los miembros de ese mismo serovar. Por esta razón, el LPS es una de las moléculas diana para el diagnóstico y candidata para el desarrollo de vacunas.

*Leptospira* tiene dos cromosomas. El cromosoma I varía en tamaño (3.55 a 4.37 Mb) y en su contenido de GC (35 a 44.9%). El cromosoma II habitualmente contiene menos del 9% del tamaño y número de genes del cromosoma I.

El análisis se realizó entre especies patógenas de *Leptospira.* Con los resultados de esta sección, sugerimos la existencia de una presión selectiva que promueve la variación genética de este locus y de varios genes, haciendo de *Leptospira* un patógeno zoonótico exitoso.

## Metodología

### Fuentes de datos, alineación del genoma central y árboles filogenéticos.

Los análisis de recombinación se evaluaron en los cromosomas I de cada cepa a tres niveles: interespecies, intraespecies entre cepas *interrogans* y en genes LPS. Los datos completos del cromosoma I de seis especies patógenas de *Leptospira* se descargaron de bases de datos públicas (el 7 de enero de 2019) para el análisis entre especies.

Para los análisis de primer y segundo nivel, los archivos fasta se anotaron con Prokka y se realizó una determinación de ortólogos (80% como similitud mínima en secuencias de proteínas y 90% como cobertura mínima de genes) entre todas las especies con Proteinortho5. El genoma “core” se obtuvo después de extraer todos los genes presentes en todas las cepas con el script Proteinortho: grab_proteins.pl, alineándose con Mafft y concatenando con un script propio.

Para el tercer análisis de recombinación, se extrajeron todos los genes (*n*=38) involucrados en la ruta de biosíntesis de lipopolisacáridos de la base de datos *KEGG Pathway* (https://www.genome.jp/kegg/pathway.html) de 29 cepas patógenas. El alineamiento del genoma de referencia correspondió a estos genes alineados y concatenados.

Para todos los análisis, el genoma de referencia y las filogenias de cada gen se generaron con 1000 réplicas de *bootstrap* usando IQ-TREE 2, con la opción ModelFinder para seleccionar el modelo de evolución más apropiado. Usando iTOL, los metadatos se agregaron al árbol filogenético.

### Detección de recombinación: método de incongruencia filogenética (PIM por sus siglas en inglés)

La detección de recombinación se realizó mediante una comparación de topologías gen a gen entre la filogenia de cada gen del "*core*" frente al árbol filogenético de referencia siguiendo el procedimiento PIM. Cada gen del "*core*" se evaluó en cuanto a su contenido filogenético utilizando IQ-TREE 2 con valor aleatorio de 10000 cuartetos. Las pruebas de topologías de los árboles también se calcularon con IQ-TREE 2. El árbol del genoma de referencia se rechazó si el peso de probabilidad esperado (c-ELW) y la prueba de Kishino-Hasegawa (valor *p*) eran inferiores a 0,05. Esto significa que las diferencias entre topologías de la filogenia del genoma de referencia y de la filogenia del gen fueron estadísticamente significativas. Se utilizó un script propio para clasificar cada gen recombinante según su categoría COG. Se usó Datamonkey 2.0 para caracterizar la selección diversificadora en cada gen recombinante con el análisis BUSTED para la prueba de todo el gen.

También analizamos 20 ORFs del LPS publicados previamente. Usamos BLASTn para cada ORF y analizamos secuencias que presentaban anomalías, como aislados de dos especies diferentes de *Leptospira* pertenecientes al mismo serovar, o secuencias pertenecientes a diferentes especies que mostraban similitudes inconsistentes, con el fin de comparar visualmente las topologías de los árboles.

## Resultados

### Recombinación entre especies patógenas de *Leptospira*

El alineamiento del genoma "*core*" se obtuvo con 2,095 genes concatenados que incluían 1,552,284 pb y 263,593 sitios informativos (17.6% del alineamiento completo). La filogenia de máxima verosimilitud construida a partir del genoma "*core*" se muestra en la Figura 7A. El genoma accesorio comprende 4,348 genes, que corresponden al 67.48% del pangenoma de estas seis especies.

Para evaluar el contenido filogenético de los genes "core" realizamos un análisis de mapeo de probabilidad. Encontramos 1,213 genes con más del 90% de árboles resueltos, lo que puede interpretarse como genes que presentan una alta señal filogenética. Los resultados de las pruebas de topología sugirieron que 34 (2.8 %) de esos 1,213 genes son recombinantes (Tabla complementaria 10). Nueve de estos genes fueron identificados como proteínas hipotéticas (26.5%). La proporción de genes recombinantes para 10 categorías de Grupos de genes ortólogos (COG) se muestra en la Tabla 9. La categoría COG más representada (25%) estaba

relacionada con funciones de traducción, estructura ribosomal y biogénesis. Cada una de las siguientes categorías de genes representó un 12% de los genes recombinantes: 1) pared celular, membrana, biogénesis de la envoltura, 2) transporte y metabolismo de aminoácidos, 3) producción de energía y metabolismo, y 4) transporte y metabolismo de carbohidratos (Figura 8).

Los resultados de la prueba de selección positiva de todo el gen con BUSTED mostraron que al menos un sitio en al menos una rama de cuatro genes (capreomicidina sintasa, alanina-tRNA ligasa, menaquinona reductasa, subunidad de unión a grupos de hierro-azufre y proteína ribosomal 50S L10) presenta selección diversificadora episódica (Tabla complementaria 10).

Se construyó un nuevo árbol filogenético del genoma "*core*" después de eliminar los genes recombinantes del alineamiento inicial. Este nuevo alineamiento, libre de recombinación, incluyó 1,507,524 nt y 255,546 sitios informativos (17.6 % del alineamiento completo) (Figura 7B). Si bien la recombinación en estas 6 especies de *Leptospira* se confirmó en el 2,8% de los genes con alto contenido filogenético y las topologías de los genes recombinantes difieren de la filogenia del genoma central, la resolución de la topología libre de recombinación no mejoró. Las relaciones filogenéticas de todas las especies son consistentes en ambos árboles. En la Tabla complementaria 10 mostramos los resultados de la cantidad de movimientos para que las topologías, gen versus genoma central, sean idénticas.

No se observa un patrón relevante en la distribución de los genes recombinantes entre las 6 especies. Los 34 genes no están ubicados en una región genómica particular, como se muestra en la Figura 9.

**Recombinación intraespecies entre miembros de *L. interrogans***

Para el alineamiento usamos 1,872 genes compartidos por todas las cepas, que comprenden 941,992 pb con 5,770 sitios informativos (0.6% del genoma completo) (Figura 10).

Después del análisis de señal filogenética, 798 genes presentaron una alta señal filogenética. Los resultados de las pruebas de topología sugirieron que 178 de esos genes son recombinantes, lo que representa un 22% (Tabla complementaria 11). De estos genes, 58 genes codifican proteínas con función desconocida (32%). La proporción de genes recombinantes para 18 categorías COG se muestra en la Tabla 10. Se encontró que un 12% de los genes recombinantes codifican proteínas implicadas en la biogénesis de la pared celular, la membrana o la envoltura. Todas las categorías de genes recombinantes entre especies también estaban presentes entre las categorías de genes recombinantes dentro de *L. interrogans*. Además, 5 genes fueron

recombinantes en ambos análisis: acetoína:2,6-diclorofenolindofenol oxidorreductasa subunidad alfa, alanina-tRNA ligasa, capreomicidina sintasa, proteína bifuncional de biosíntesis de coenzima A CoaBC y UDP-N-acetilglucosamina 1-carboxiviniltransferasa.

Al igual que con los genes recombinantes entre especies, no detectamos puntos críticos de recombinación dentro de los genomas de *L. interrogans*. La Figura 11 representa los genes recombinantes distribuidos a lo largo de los genomas. La sintenia de genes recombinantes se ha mantenido entre los dos genomas de *interrogans* serovar Lai.

**Recombinación entre genes del LPS**

Decidimos estudiar la recombinación de genes de lipopolisacáridos porque fueron los primeros genes que mostraron evidencia de recombinación. Treinta y ocho genes relacionados con LPS se concatenaron y se usaron como referencia. Estos genes se obtuvieron de 29 cepas de 10 especies (Tabla 8). El alineamiento del genoma incluyó 50,286 nt con 16,095 sitios informativos (32.7 %) y generó la filogenia que se muestra en la Figura 12.

El análisis de señal filogenética mostró que 36 genes presentaron una alta señal filogenética. De estos, 18 fueron recombinantes según los resultados de las pruebas de topología (Tabla 11). El análisis de selección mostró que solo un gen, *lptC* que codifica para la proteína del sistema de exportación de lipopolisacáridos, presentó selección positiva según el método BUSTED.

Además, diez de los 20 ORFs de LPS de cepas de *Leptospira* que mostraban anomalías en BLAST (dos aislados de *Leptospira* que pertenecían a dos especies diferentes pero que compartían los mismos antígenos de serovariedad o mostraban una mayor identidad con otras especies en lugar de con miembros de la misma especie), mostraron evidencia de incongruencia de topología (Figuras complementarias 3-12).

## Conclusiones

La epidemiología genómica permite una resolución de subtipificación sin precedentes a través del análisis del genoma "*core*" y accesorio para evaluar la diversidad genómica, los factores determinantes de la resistencia, los genotipos de virulencia, la resistencia al estrés y el potencial patogénico.

Los aislados de *Salmonella enterica* de Ecuador estudiados en este trabajo mostraron resistencia a múltiples fármacos, lo que señala la importancia de un mejor uso de los antimicrobianos en entornos de granjas avícolas intensivas en este país.

*Salmonella enterica* serovar Infantis aisladas de granjas, alimentos y muestras humanas presentaron una gran similitud genómica, lo que sugiere la diseminación de este patógeno transmitido por los alimentos a través de la cadena alimentaria a los humanos.

La detección *in silico* de determinantes presentó limitaciones porque no se encontraron plásmidos a pesar de los patrones de resistencia a múltiples fármacos en nuestros aislados de *Salmonella*.

Dada la relación filogenética de *Listeria monocytogenes* de quesos con muestras de origen clínico, se recomienda la genotipificación de alimentos y aislados clínicos para la vigilancia de rutina a fin de inferir vínculos epidemiológicos y la adopción de medidas de control que eviten un aumento de casos de listeriosis.

La alta similitud genética en las cepas ST2 de quesos y casos clínicos sugiere una alta probabilidad de asociación epidemiológica y propagación de *L. monocytogenes* a través de la cadena alimentaria, específicamente el consumo de queso, a humanos en Ecuador.

En Ecuador, la listeriosis no es una enfermedad de declaración obligatoria y no existe un programa nacional de vigilancia para este patógeno transmitido por los alimentos. Aunque este estudio proporciona valiosa información genómica y epidemiológica sobre *Listeria monocytogenes* en Ecuador, aún son necesarias más investigaciones de otras fuentes de contaminación además de los quesos.

La evidencia de recombinación en *Leptospira* sugiere una presión selectiva que promueve la variación en diferentes loci en este patógeno con posibles impactos en la identificación de serotipos y el desarrollo de vacunas.

El análisis genómico proporciona información sobre los genes recombinantes que están bajo selección, los genes que hacen de *Leptospira* un patógeno exitoso. Por lo tanto, este tipo de investigación también puede informar sobre nuevas formas de prevenir y tratar la leptospirosis.

## Abbreviations

AMR – Antimicrobial resistance

BLAST – Basic Local Alignment Search Tool

BWA – Burrows-Wheeler Aligner

CLSI – Clinical and Laboratory Standards Institute

DNA – Deoxyribonucleic acid

EUCAST – European Committee for Antimicrobial Susceptibility Testing

FBD – Foodborne diseases

GC – Guanine-Cytosine

HTS – High-throughput sequencing

LCB – Locally collinear blocks

LPS – Lipopolysaccharide

MLST – Multilocus Sequence Typing

NCBI – National Center for Biotechnology information

PCR – Polymerase Chain Reaction

PFGE – Pulse-field Gel Electroforesis

PIM – Phylogenetic Incongruence Method

RNA – Ribonucleic acid

RTE – Ready-to-eat

SAM – Sequence Alignment/Map

SNP – Single Nucleotide Polymorphism

ST – Sequence Types

TA – Toxin-antitoxin systems

WGS – Whole genome sequencing

WHO – World Health Organization

## List of figures

**Figure 7**. A) Maximum likelihood phylogeny of 2095 core genes from *L. borgpetersenii, L. interrogans, L. kmetyi, L. mayottensis, L. santarosai* and *L. tipperaryensis*. B) Maximum likelihood phylogeny of 2061 non-recombinant core genes from the same species as A. Branches with bootstrap values equal to 100 are displayed with a black circle.

**Figure 8**. Comparison of tree topologies of 34 leptospiral recombinant core genes. The strain names are specified in the legend. The genes are in this order from left to right: UDP-N-acetylglucosamine 1-carboxyvinyltransferase, hypothetical prot., hypothetical prot., Cysteine--tRNA ligase, hypothetical prot., hypothetical prot., Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha, hypothetical prot., Protein QmcA, Capreomycidine synthase, hypothetical prot., Dipeptide permease D, Elongation factor 4, Glyceraldehyde-3-phosphate dehydrogenase 1, hypothetical prot., N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase, Glucose-1-phosphate thymidylyltransferase 1, Putative teichuronic acid biosynthesis glycosyltransferase TuaG, Coenzyme A biosynthesis bifunctional protein CoaBC, NADH-quinone oxidoreductase subunit 1, Alanine--tRNA ligase, 50S ribosomal protein L10, DNA-directed RNA polymerase subunit beta', Threonine--tRNA ligase 2, hypothetical prot., GTPase Der, hypothetical prot., Menaquinone reductase, iron-sulfur cluster-binding subunit, Outer-membrane lipoprotein carrier protein, Elongation factor G 1, hypothetical prot. Sulfurtransferase, putative succinyl-diaminopimelate desuccinylase, 4,5-DOPA dioxygenase extradiol.

**Figure 9**. Distribution of recombinant genes among *Leptospira* species along their genomes. Displayed dots indicate the beginning of the gene position.

**Figure 10**. Maximum likelihood phylogeny of 1872 core genes from 7 *L. interrogans* strains. Branches with bootstrap support values equal to 100% are shown with a black circle.

**Figure 11**. Distribution of recombinant genes among *Leptospira interrogans* strains along their genomes. Depicted dots indicate the beginning of the gene position.

**Figure 12**. Maximum likelihood phylogeny of 38 genes involved in the LPS pathway. Branches with bootstrap values larger than 90% are shown with a black circle. Clades with more than one isolate are colored.

## List of tables

# Introduction

## 1. Bacteria as pathogens

Most microorganisms in nature and those that live in and on the human body are harmless. Nevertheless, a small proportion of those microbes has the ability to infect or cause disease. **Infection** refers to the growth of microorganisms in or on a host, and **disease** involves cell or tissue damage or impairment of any function of the host [1] (Figure 1).

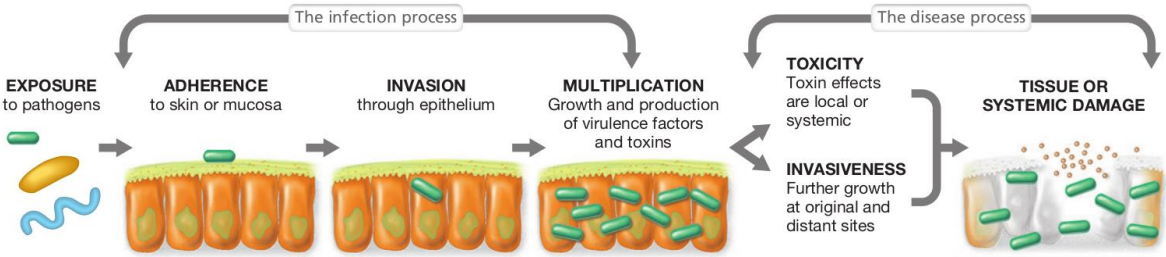



**Figure 1.** Microorganisms, from exposure to disease. Source: Madigan *et al*, 2022 [1]

For a microbe to cause disease, the first step is to adhere to host cells, by interactions between the microorganism and the host, by forming biofilm, or accessing through mucous membranes, skin surface or by cuts, wounds, or bites. Adherence can be achieved by adhesins, capsules, mucus layers, fimbriae, pili, or by surface proteins that specifically bind to cell proteins of certain regions of the host body [1]. Once attached to a surface, microorganisms need to grow and multiply to achieve infection. This growth may also involve the production of biofilms (of one or several species). After multiplying, the microbes need to invade cells or tissues at localized sites or they can enter the bloodstream (causing bacteremia or complicating to septicemia) to cause disease with mild or severe symptoms [1].

The virulence of a pathogen depends on its capability to adhere, colonize and invade, and on its virulence factors [1]. Bacterial virulence factors are classified as secreted (proteins that kill the host cells or help pathogens resist the host immune response), membrane associated (allowing adhesion or evasion of the host cell) or cytosolic (those that allow metabolic, physiological and morphological changes) [2]. The pathogens' capacity to cause disease may differ among strains of the same species [3], but it also depends on the genetic and physiological state of the infected host [4], or on food consumption and safety behaviors [5].

According to the Global Health Observatory, infectious and parasitic diseases were responsible for 9.2% of the global deaths in 2019 [6]. These figures must be underestimated because in various regions of the world, the exact cause of death is not always identified. Out of 1415 known pathogen species, 538 are bacteria accounting for 38% of the total [7].

## 1.1 Antimicrobial resistance

With the discovery of penicillin, quickly followed by other antibiotics, most bacterial infections were easily treated and controlled [8]. However, the ability of microorganisms to tolerate and resist antimicrobials represents now a serious public health problem [9]. Despite the increase in antimicrobial resistance in pathogenic bacteria, the investigation of new compounds for treatment is limited. It is essential to encourage the search for new natural or chemical products that have antimicrobial potential [9].

There are several mechanisms by which bacteria can acquire or develop resistance to antimicrobials [10,11]. In a bacterial population, a group of cells (usually less than 1%) remain dormant, so antimicrobial compounds may have no effect on them. This group of cells are known as persisters [12], and their antibiotic resistance or tolerance is non-genetic [10], being dependent on a physiological state of "dormancy" that prevents the action of antimicrobials.

Some microorganisms present a natural resistance (intrinsic when it is expressed in all the strains of a species, or induced when the resistance is expressed when exposed to the compound) to antimicrobials. Among the mechanisms that confer this type of resistance are efflux pumps or low permeability of the outer membrane [12].

Many microorganisms, in addition to their chromosome or chromosomes, harbor plasmids (extrachromosomal linear or circular DNA molecules) that usually contain genes that are not essential for the growth of the microbe, but that code for traits that provide a special advantage to the host cell. The most studied plasmids correspond to mobile genetic elements that confer resistance to antimicrobials [1]. Although plasmids are probably the most important genetic elements for distributing or disseminating antimicrobial resistance genes, they have a limitation. Since they imply an energetic load for the cell, bacteria tend to eliminate them when the selective pressure ceases to exist [13].

Another way of gaining resistance to antimicrobials is through point mutations, and less frequently deletions or insertions, in genes that encode targets, transporters or enzymes associated with antimicrobials [12].

Toxin-antitoxin systems (TAs) may have an important role in the maintenance of AMR. TAs were originally identified on plasmids and were thought to be of minor benefit to cells, although they were known to promote plasmid maintenance [13]. TAs are genetic complexes, that can be found on plasmids and on the bacterial chromosome and encode a toxin (when overexpressed it kills the cell or inhibits cell growth), and an antitoxin (which neutralizes the toxin) [13]. It is

thought that TAs are associated with AMR because they appear to stabilize not only resistance plasmids, but also genomic islands in the absence of the antimicrobials [13].

The formation of biofilms, extracellular matrices produced by bacteria adhered to a surface, allows greater tolerance to antimicrobials [1], due to lower penetration of the compounds, low growth rate, or certain genetic determinants typical of populations living in biofilms [10].

## 2. Bacterial genetics

Bacteria have the advantageous ability to adapt relatively easily to a changing environment. This capacity is mainly based on its genetic diversity generated from alterations in its genetic material (mutations) and the acquisition of genetic information, through horizontal gene transfer [1]. These genomic changes are balanced with genomic stability, which together enable microbial evolution [14].

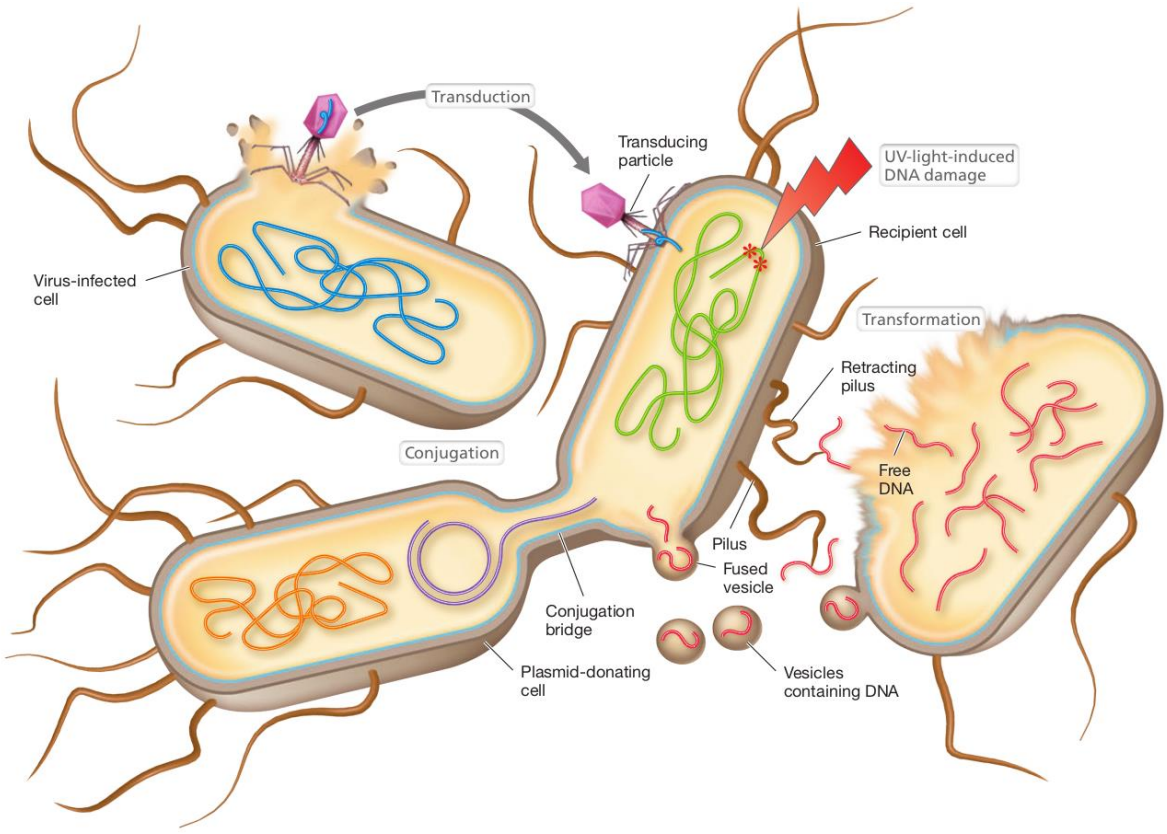



**Figure 2.** Bacterial (and archaeal) mechanisms of DNA transfer: transduction, conjugation and transformation. Source: Madigan *et al*, 2022 [1]

## 2.1 Mutation

A heritable alteration or change in the DNA sequence of an organism's genome is a mutation. A mutation can represent a **beneficial** effect, a **neutral** effect or a **harmful** one in an organism.

To better understand the effect of a mutation, it is essential to differentiate the **genotype** (the nucleotide sequence of a genome) from the **phenotype** (the observable characteristics of an organism). A mutation does not necessarily generate a different phenotype, and several different mutations (different genotypes) can generate the same phenotype [1].

During DNA replication, DNA polymerase may mistakenly place an incorrect nucleotide, generating a **spontaneous mutation**, while an environmental agent or chemicals (mutagen) change the structure or sequence of the DNA, producing an **induced mutation** [15].

## 2.2 Horizontal Gene Transfer

Horizontal or lateral gene transfer is the flow of genes between cells that are not descendants of one another. There are 3 main processes for this only one direction genetic movement in bacteria [16]:

1. Transformation, which is the acquisition of free DNA from the environment.
2. Transduction, in which the genetic information transfer is phage-mediated.
3. Conjugation, which implies the cell-to-cell contact to transfer a conjugative plasmid from a donor cell to a recipient.

Once inside the cell, the new DNA may be degraded by host destruction systems, replicate (if an origin of replication exists), or recombine with the recipient DNA [1]. Recombination after horizontal gene transfer allows the bacteria to acquire genetic material for new virulence, resistance or other metabolic traits that might improve better.

## 2.3 Genetic recombination

Recombination in bacteria is the physical transfer of DNA between two molecules, and it may involve legitimate (homologous) or illegitimate recombination [17]. Recombination is a frequent process and an important factor in the evolution of many bacterial species [17–23]. Recombination is mainly detected when recombinant genes cause phenotypic or genetic variation; consequently, recombination events that include genetic exchange of identical segments remain undetected [17].

Operational genes are more frequently and successfully transferred than informational genes [24]. The products of operational genes belong to simple and small assemblies, while the products of informational genes usually present complex interactions [24].

In pathogenic microorganisms, it has been shown that recombination plays a role in increasing virulence [18], evasion of the immune responses [22,25], host tropism [17,26]. At a different level,

recombination also implies that transferred genes have a different phylogenetic history to those strictly vertically inherited in the recipient genome [27]

## 3. Transmission of bacterial pathogens

Infectious diseases can be classified according to their mode of transmission. The transmission of microbial pathogens includes three major modes: person to person (direct, indirect contact or by airborne droplets), by vehicles (waterborne, foodborne, airborne, soilborne) or by vectors, such as arthropods/insects [1]. In this work, we have focused on two foodborne and one zoonotic pathogens.

### 3.1 Foodborne pathogens

Despite technological advances on research and prevention of foodborne diseases (FBD) they are still a global public health concern, especially in developing countries. The symptoms range from mild and self-limiting, such as nausea, fever or diarrhea, to weakening and severe, such as liver failure, brain disorders or even abortion or stillbirth in pregnant women [28]. The intensity and impacts of these illnesses vary greatly across regions and also among different groups of people. Infants, young children, pregnant women, older adults and people with weak immune system are at a higher risk of developing complications or dying from FBD [29].

There is a continuous effort from international and national organisms to estimate the impact of foodborne diseases worldwide not only on health but also on socio-economic development, and to understand how pathogens evolve and disseminate from farm to fork. Because of the data gap about prevalence and epidemiology of food-transmitted diseases in developing countries the foodborne disease burden cannot be accurately estimated and usually foodborne outbreaks in these countries go undetected, unreported and even unexplored [29]. This lack of information can be caused by difficulties for early detection and insufficient amount of resources on research areas. Despite these limitations, the WHO estimated the burden of foodborne diseases in 2010, determining that one in every ten people gets sick after consumption of unsafe food, causing 600 million of disease cases every year, and 420,000 deaths [29].

Changes in food production and diet have influenced FBD. The increasing habit of people consuming ready-to-eat and prepared food away from home, as well as the concentrated animal feeding operations and intense agriculture, might be contributing factors to an increased number of infections and outbreaks [30].

Foodborne diseases are preventable. It is necessary to carry out a better surveillance for the control of FBD. On the one hand, it is important to develop risk assessment strategies (prevalence analysis) to generate recommendations based on real information from each country. On the other hand, managing these risks would allow their reduction or mitigation, if possible, throughout the food production chain, and to take the appropriate legislative measures [31].

*Salmonella* and *Listeria* are the two bacterial foodborne pathogens most frequently associated with produce commodities because these products are usually consumed raw or with minimal processing [32]. Cross-contamination or recontamination are two major events that contribute to these FBD and show how difficult it is to eradicate these pathogens in food-processing facilities [33]. Accurate information on the burden of *Salmonella* and *Listeria* prevalence on food and the environment can quite precisely inform policy-makers for food safety control and prevention efforts.

### 3.1.1 *Salmonella enterica*

*Salmonella* is a Gram-negative foodborne pathogen belonging to the order Enterobacterales and family Enterobacteriacea. *Salmonella enterica* represents one of the four main global causes of diarrheal diseases, even though some serotypes can cause typhoid fever. It is not only important for its implication in a high number of human infections but also because of its diversity (more than 2500 serotypes) and resistance to several different antibiotics. Some of the foods involved in *Salmonella* infections are eggs, poultry and other products of animal origin as well vegetables [34].

Foodborne infections caused by *S. enterica* are of primary importance worldwide. The WHO estimates that *Salmonella* causes more than 153 million illnesses, 120,281 deaths, and 8.27 million disability-adjusted life years annually [29]. As foodstuffs can be contaminated in several parts of the food chain, a "from farm to table" approach is necessary to understand the epidemiology of *Salmonella*. Although *Salmonella* can contaminate vegetables, food-producing animals, especially poultry, are considered important sources for human infections [35,36]. In Latin American countries, poultry is one of the main sources of proteins of animal origin. This is the case of Ecuador, where poultry meat is the most consumed commodity with a yearly per capita consumption of 30.4 Kg [37].

Antimicrobials are commonly used in poultry production as both therapeutics and growth promoters. However, even when antimicrobials are used under technical criteria, they can select resistant strains of *Salmonella* that pose a public health problem when reaching consumers [38].

This is of special concern in developing countries where the misuse of antimicrobials and lack of control are issues to be addressed.

There is a wide diversity of virulence factors that are essential for pathogenicity of *Salmonella* in host cells. Among these factors, fimbriae, flagella, plasmids, pathogenicity islands, toxins and secretion systems are the more frequently associated to pathogenic strains of *Salmonella* [39].

### 3.1.2 *Listeria monocytogenes*

*Listeria monocytogenes* is a Gram-positive of the order Bacillales and family Listeriaceae, facultative intracellular foodborne pathogen that can cause severe illnesses and conditions such as septicemia, encephalitis and meningitis, abortion, stillbirths, and neonatal infections in humans [40] as well as severe disease in other mammals and birds [41,42]. Although infections of *L. monocytogenes* are rare and the incidence of listeriosis is generally low, the case-fatality rate is one of the highest among foodborne infections [40], affecting principally young children, immunocompromised individuals, pregnant women, and the elderly [43].

*Listeria* can be found in raw and processed foods, but the most frequent vehicles are ready-to-eat (RTE) food products such as cheeses, meat, seafood, and fresh products [44]. Among these food matrices, cheeses are very important for *Listeria* transmission mainly because they are usually stored at refrigeration temperatures [45] for several days allowing the growth of this psychrophilic microorganism [46], and normally they do not receive an additional treatment before consumption. In a review about food outbreaks caused by contaminated cheese (pasteurized and unpasteurized milk) in USA from 1998 to 2011, even though *Listeria* was not the most common pathogen, it was the cause of death of five out of six cases [47].

*L. monocytogenes* represents a remarkable threat in foods because it can persist in food-processing environments for long periods of time, due to in part to its ability to survive under a great diversity of stressing conditions (sanitizing, pH, water activity, and temperature), and the ability to form biofilms in surfaces [48,49]. This bacterium can colonize food-processing facilities when contaminated raw materials are introduced, and when physical barriers and disinfection procedures between processed food and raw materials areas do not exist [50].

Although isolates of *L. monocytogenes* are considered equally pathogenic by food regulatory authorities, certain serotypes have been shown to be more frequently associated with clinical cases, as most human listeriosis outbreaks and sporadic cases are related to serotype 4b, and some less frequently with serotypes 1/2b and 1/2a [43,51]. Not only does the presence of different

serotypes in food imply their association with clinical cases, it is now recognized that there are phenotypic differences between lineages and serotypes (presence of virulence factors, ability to survive stress conditions and/or produce biofilms, presence of stop codons in virulence genes, among other genetic traits) [51,52].

## 3.2 Zoonotic pathogens

Zoonoses are a major public health problem as a result of our close contacts and relationships with animals [53]. A zoonotic infection is the result of a pathogen that has jumped from an animal to a human by direct contact with animals or their contaminated wool, fur or hides, via animal bites, or vectors such as mosquitoes, lice or ticks, but also by consumption of contaminated water or food [31,53]. Taylor and collaborators prepared a literature review and found that 61% of infectious agents (868/1415) are zoonotic, and they also mentioned that 75% of the emerging pathogens (132/175) cause zoonoses [7].

Although many zoonoses imply a great risk to public health, some are considered neglected, especially in developing regions [31]. It is essential to join efforts between countries affected by the same disease and coordinate measures for their control. As control strategies require financial support that is generally not available in developing countries, developed countries need to get involved to achieve efficient control of zoonoses.

### 3.2.1 *Leptospira*.

Pathogenic species of *Leptospira* are the etiological agents of leptospirosis, the most common zoonotic disease worldwide causing more than 500.000 cases a year with a mortality rate of 10% [54]. Leptospirosis is a systemic disease that can present with mild (influenza-like) or severe (liver and kidney damage and death) symptoms. The intensity of the symptoms depends on the serovar causing the infection, and the age and immunological status of the patient. Not only does it affect humans, it can also cause illness in domestic animals [55].

The genus *Leptospira,* belonging to order Leptospirales and family Leptospiraceae, comprises 35 species [56] with more than 300 serovars [57]. Its classification in serogroups and serovars is based on the expression of distinct LPS antigens [58]. These LPS antigens are also associated with protective immunity because antibodies against a serovar protect from all the members of the same serovar [59]. Therefore, LPS is one the target for diagnosis and potential candidates for vaccines [60].

Isolates and species within the genus *Leptospira* show great differences in chromosome I size (3.55-4.37 Mb), GC content (35-44.9%) and habitat (environmental or restricted to host-to-

host transmission cycle) [61]. Leptospiral chromosomes II usually contain less than 9% of chromosome I size and CDSs [62–65]

Many wild and domestic mammals carry this bacterium in the kidneys and disseminate the pathogen through their urine [54]. Different species of mice and rats are chronic reservoirs for *Leptospira*, harboring the bacteria in their renal tubes with no signs of disease [66]. Urine-contaminated water or mud are also sources of infecting spirochetes which can enter the new host through lacerations or the mucosal epithelium [67].

Clinical serodiagnosis of leptospirosis is difficult and it requires microscopic agglutination test (MAT) that uses live leptospires to confirm the infection. The bacteria to be used must be from serovars that circulate in the region where the patient may have been infected [68].

## 4. Genomic epidemiology

Even though genotyping tools such as PFGE and MLST have been used successfully for typing isolates, they present the restriction of analyzing a limited portion of the genetic information of the microorganisms [69]. In some cases, the discriminatory capacity is not sufficient to differentiate isolates. This occurred with the epidemiological surveillance of *Listeria monocytogenes* in France where despite the fact that the evidence provided by PFGE analyzes has led to great advances in food safety, the use of WGS (whole genome sequencing) has shown a better discrimination of isolates to the point that the use of PFGE has been discontinued [70].

Molecular and genomic analyses are also important to understanding the biology of organisms, source attribution of clinical cases, and to understand genomic diversity and evolution and the emergence of newer traits [1].

Genomic epidemiology is undergoing a revolution as a result of the increasing availability and reduced costs of whole genome sequencing (WGS) [69]. The onset of cheaper and faster technologies, which allow accessing the complete sequences of bacterial genomes in short times, is putting an end to long-standing controversies about the most suited genetic markers for each organism.

### 4.1 Whole genome sequencing

The first bacterial genome was available in 1995 [71], and it changed dramatically the research on microbiology and molecular biology. Now, we have access to DNA and RNA sequences from more than 125000 prokaryotic and viral species [1]. Currently available bacterial genomes are in the order of tens of thousands, with a very biased distribution among species, and regions.

Specifically, on March 13th, 2017 there were 13,381 genome sequences for *S. enterica*, 4,175 for *L. monocytogenes,* and 1,506 for *Leptospira* species accessible at GenBank.

Complete genome sequences are now obtained by means of high-throughput sequencing (HTS) technologies, which produce large (in the order of millions) numbers of short sequences (typically between 100 and 300 bp) that have to be processed with bioinformatics tools to assemble the complete, original sequence of the genome they were derived from [1]. Although there are several competing methodologies [72,73], one of the most popular nowadays and the one that we used in this project is based on Illumina sequencing of paired-ends reads of 150 bp each.

The sequencing of whole genomes of bacterial pathogens allows the evaluation of the total genetic information of isolates to the single nucleotide level [69]. Genomic epidemiology has revolutionized the capacity of researchers and laboratories to deeply investigate transmission, virulence, antimicrobial resistance and evolution of microorganisms [69].

Despite the great advantages of whole genome sequencing, there are still several limitations of this methodology. On the one hand, there is an uneven distribution of access to technology and training in analysis and competence in genomics and bioinformatics [74]. On the other hand, new programs or approaches are continually being developed, making it difficult to standardize the tools used around the world [69]. In addition, despite the fact that the vast majority of genomic information and programs are open access, there are research groups that limit access to their bioinformatics tools.

## 4.2 Epidemiological surveillance

Epidemiological surveillance, which includes observation, recognition, registration, case reporting and data analysis, is an important part of disease control to identify trends or possible outbreaks (section 4.3). There are two fundamental concepts at this point. The **incidence** of a disease is the number of new cases in a population at a given time, while the **prevalence** is the total number of new and existing cases of a disease in a population at a given time. These two measurements are essential to predict the risk of a disease and its impact, respectively, and together, they show us the public health status of a population [1].

Infectious diseases affect everyone, they are global health problems. The distribution of a disease can change rapidly due to adaptive changes in pathogens, changes in the environment, or demographic or behavioral changes in the host, as well as climatic changes. Emerging

diseases are those that suddenly become prevalent, and re-emerging diseases are those diseases that were previously controlled but suddenly appear as a new epidemic [75].

When epidemiological surveillance is done correctly, it can lead to better control of infectious diseases and accurate measures for public health by increasing the knowledge of transmission, causes, disease epidemiology, reservoirs, and mortality and morbidity rates [76].

Regarding zoonotic diseases, surveillance is essential for the early detection of infectious pathogens in animals and humans, vectors, and reservoirs. For surveillance to be effective, it must be carried out at different levels: detection of pathogens, serological monitoring, surveillance of symptoms, and determination of transmission risk factors, and accompanied by vector control [31].

If this surveillance is globalized, there is the possibility of eradicating diseases. That is the goal of the drive to join efforts between different regions to control a disease more efficiently.

## 4.3 Outbreak investigation

Due to the COVID-19 pandemic, terms like epidemic, pandemic, or outbreak have become familiar. A disease is **endemic** when it has a relatively constant and low incidence in a population. A disease becomes **epidemic** if it affects unusually large numbers of people in a particular population simultaneously. An epidemic becomes a **pandemic** when it spreads globally [1]. A **sporadic** disease implies cases in distant times or areas (unrelated cases), while an **outbreak** corresponds to an increase in the number of cases in a relatively short time in a geographic area where that disease was endemic or sporadic. Cases need to be related [1,77].
The investigation of an outbreak starts with the identification of cases, where symptoms of the disease and common epidemiological characteristics among the people involved are considered. Once the number of cases of a given disease is greater than the expected number in that region, an outbreak is suspected. The research seeks to identify the causative agent of the symptoms, the size of the affected population, the expected incidence of the disease in that community, and the possible infection source (in case of foodborne or vector-borne diseases) [77]. For some microorganisms, such as *Clostridium botulinum* or Poliovirus, one single diseased case is considered an outbreak.

The identification of microorganisms at the laboratory level is essential for the detection of outbreaks, and the determination and elimination, when possible, of the sources of infection [76]. Depending on the disease, it is important to track and match microorganisms from human or animal samples with isolates from possible sources [69]. The main objective of outbreak

investigation is to prevent, through control measures, the appearance of more cases and new outbreaks [77].

Despite the availability of whole genome sequencing for molecular typing, methodologies that have shown limited discriminatory power are still being used in many countries [76]. In Ecuador, for example, the Institute of Public Health carries out epidemiological surveillance of a few bacterial pathogens, and if an outbreak is identified, molecular techniques of PFGE or MLST are used. At the research level, whole genome sequencing has already been used [78–80].

# Objectives

The main objective of this thesis was to use genomic epidemiology to investigate the transmission and evolution of 3 pathogenic microorganisms: *Salmonella enterica, Listeria monocytogenes, Leptospira*.

1. To determine *Salmonella* prevalence in poultry farms, chicken carcasses and human stool samples from Quito, Ecuador.
2. To describe by phenotypic methods and by whole genome sequencing the antimicrobial resistance of *Salmonella enterica* isolated in Quito, Ecuador.
3. To perform a genomic analysis to determine the relatedness of *Salmonella* strains isolated from poultry, food and human samples from Quito, Ecuador.
4. To perform a genomic characterization of *Listeria monocytogenes* strains isolated from cheeses and clinical samples in Ecuador.
5. To compare the presence of plasmids by *in silico* determinants search and by mapping our raw reads to plasmids previously described in *Listeria monocytogenes*.
6. To evaluate the presence of genes encoding major *Listeria* virulence factors in Ecuadorian isolates.
7. To identify genetic recombination in core genes among 6 pathogenic species of *Leptospira*.
8. To identify genetic recombination in core genes among 7 *Leptospira interrogans* strains.
9. To detect recombinant genes involved in leptospiral LPS pathway among pathogenic *Leptospira* strains.

# Chapter 1

# Genomic Epidemiology of *Salmonella* Infantis in Ecuador: From Poultry Farms to Human Infections

This chapter has been published as:

Mejía L, Medina JL, Bayas R, Salazar CS, Villavicencio F, Zapata S, Matheu J, Wagenaar JA, González-Candelas F and Vinueza-Burgos C (2020) Genomic Epidemiology of *Salmonella* Infantis in Ecuador: From Poultry Farms to Human Infections. Front. Vet. Sci. 7:547891. doi: 10.3389/fvets.2020.547891

# 1. Background

Previous investigations in Latin America showed that *Salmonella enterica* serovar Infantis is an increasingly important serotype on poultry farms [81–85]. Moreover, this serotype has also been reported to cause infections in local inhabitants and travelers that have visited Latin American countries [86–88]. However, no genomic data considering isolates of *Salmonella* originated from animals, foodstuff, and humans have been released in Latin America so far.

This research was aimed at describing by phenotyping methods and whole-genome sequencing, the antimicrobial resistance (AMR) characteristics and genetic profiles of *Salmonella* isolates obtained from broiler farms, broiler carcasses and humans in Quito – Ecuador.

# 2. Materials and methods

## 2.1. Study design and sampling

Samples for *Salmonella* isolation were collected weekly from November 2017 to November 2018 following the guidelines of the document: "Integrated Surveillance of Antimicrobial Resistance in Foodborne Bacteria" by the World Health Organization [89].

**Poultry Farms**: 133 flocks from 69 farms were sampled during the study period. For every sampled flock, 25 caeca from 25 chicken were randomly collected at the slaughterhouse level and transported to the laboratory in an icebox within the next two hours. Caecal samples are recommended by the WHO because they provide a higher recovery of isolates and better represent contamination of individual animals at the farm level (15). At the laboratory, a sample pool of 25 g was obtained for bacteriological isolation as previously described [84].

**Chicken carcasses**: 335 carcasses were collected in three kinds of markets as follows: 125 samples from supermarkets, 126 samples from small shops, and 84 samples from open markets. Sampling of chicken carcasses was performed alternately between the north and south of the city. Each carcass was collected in its original bag and transported to the laboratory in an ice-box within the next two hours. At the laboratory, 25 g of breast skin of every carcass were aseptically collected for bacteriological analysis.

**Human stool samples**: 302 samples were evenly collected in two health care centers located in the urban periphery of Quito (Guamani health care center at the south and Calderon health care center at the north) from patients with two or more episodes of diarrhea or vomiting in the last 24 hours. Human stool samples were transported to the laboratory in an icebox within the next two hours. Approximately 25 g of feces were collected for bacteriological analysis.

According to national legislation, ethics approval was not required for poultry farms and chicken carcasses sampling since no animals were sacrificed during this study. For the human component, the project was approved by the bioethics committee from the National Institute of Public Health "Leopoldo Izquieta Pérez" (Protocol ID:CEISH-INSPI-005). The participants were informed about the objective of the study. All volunteers provided a written consent. All personal information was anonymized.

## 2.2. Isolation and Identification of *Salmonella enterica*

*Salmonella* isolation was performed by a method based in the ISO 6579-1:2007 protocol. Briefly, 225 mL of Buffered Peptone Water (BPW; Difco, BD, Sparks, MD) was added to every sample, homogenized by hand for 1 minute and incubated at 37 °C for 20 h. Then, 100 µL of each enrichment was inoculated onto Modified Semi-solid Rappaport-Vassiliadis agar (MSRV; Oxoid, Basingstoke, UK) in three equidistant points and incubated at 42 °C for 24 h. Afterwards, plates were examined for the presence of a white halo around of at least one inoculation point. A loopful taken from the edge of the white halo was streaked on a Xylose Lysine Deoxycholate agar (XLD, Difco) and incubated at 37 °C for 24 h. After incubation, one suspect colony of *Salmonella* was biochemically confirmed by Triple Sugar Iron (TSI, Difco, BD), Iron Lysine (LIA, BBL, BD), Urea (BBL, BD) and Sulfur Indole Motility tests (SIM, BBL, BD). Isolated colonies were confirmed by PCR as previously described [90]. The 95% confidence interval ($CI_{95\%}$) for the prevalence of *Salmonella* at each component was calculated.

## 2.3. Antimicrobial Susceptibility Testing

All *Salmonella* isolates were examined by the Kirby-Bauer disk diffusion method with the following antibiotics: sulfamethoxazole + trimethoprim (25 µg), gentamicin (10 µg), ciprofloxacin (5 µg), cefotaxime (30 µg), tetracycline (30 µg), streptomycin (10 µg), chloramphenicol (30 µg), cefoxitin (30 µg), amikacin (30 µg), nitrofurantoin (300 µg), azithromycin (15 µg), fosfomycin (200 µg), ertapenem (10 µg), amoxicillin + clavulanic acid (30 µg). *E. coli* ATCC 25922 strain was used as quality control. Results and methods were interpreted according to CLSI (Clinical and Laboratory Standards Institute) 2019 criteria considering all intermediate phenotypes as resistant for further analysis [91].

## 2.4. Detection of Extended Spectrum Beta-Lactamase (ESBL) Genes

*Salmonella* isolates that presented resistant phenotypes to beta-lactam antibiotics were further tested by PCR for the identification of ESBL genes. PCR conditions and primers were the ones described by Hasman et al. [92] for $bla_{CTX\text{-}M}$, Olesen et al. [93] for $bla_{TEM}$, Kruger et al. [94] for $bla_{CMY}$

and Arlet et al. [95] for *bla*$_{SHV}$. Sub-families of *bla*$_{CTX-M}$ genes were identified with PCR protocols described by Carattoli et al. [96] for *bla*$_{CTX-M-1}$, Jiang et al. [97] for *bla*$_{CTX-M-2}$, Hopkins et al. [98] for *bla*$_{CTX-M-8}$, Paauw et al. [99] for *bla*$_{CTX-M-9}$ and Dierikx et al. [100] for *bla*$_{CTX-M-14}$. Amplification products were confirmed by gel electrophoresis using a 2% agarose gel. All PCR products were purified and sequenced at Macrogen Inc (Seul-South Korea). Obtained sequences were aligned against reference sequences with the online tool ResFinder v3.2 [101].

## 2.5. Whole Genome Sequencing

For whole genome sequencing (WGS), a selection of *Salmonella* isolates was made from the animal and food components. When selecting isolates from poultry farms the first positive sample of each farm was considered. For chicken carcasses, the first positive sample of every sampling week in each kind of market was selected. All non-*S.* Infantis isolates and all isolates from the human component were selected for WGS.

Genomic DNA was extracted using Invitrogen PureLink Genomic DNA Kit (Thermo Fisher Scientific, Walthman, MA, USA) following manufacturer's recommendations for Gram-negative bacterial cell lysates. DNA was quantified using Invitrogen Qubit 3.0 fluorometer (Thermo Fisher Scientific, Walthman, MA, USA), and sequenced with the Illumina NextSeq platform using Nextera XT Library Preparation Kit obtaining 150 x 2 bp paired-ends sequences (Illumina, San Diego, CA, USQ). Default parameters were used for all bioinformatic tools and programs unless otherwise specified. Reads were trimmed with Trimmomatic [102] to remove ambiguous nucleotides and those with quality score values <20. The programs Fastqc [103] and Multiqc [104] were used for quality assessment.

### 2.5.1 Serotype Identification

*Salmonella* serotypes were identified by PCR as described by Akiba et al. [90]. Additionally, serotypes of isolates subjected to WGS were further confirmed by the analysis of their raw sequencing reads using the SeqSero pipeline [105].

### 2.5.2 MLST Analysis, Antimicrobial Resistance Genes, and Plasmid Detection

In order to identify MLST sequence types (ST), antimicrobial resistance genes and plasmid sequences, ARIBA [106] was used with PubMLST [107], ResFinder v3.2 [101] and PlasmidFinder 2.1 [108] databases, respectively. Phenotype resistance was compared with the presence of resistance genes found by WGS. Additionally, we performed a mapping against the megaplasmid p-F219

described by Vallejos-Sánchez *et al*. [109] using Burrows-Wheeler Aligner with BWA-MEM algorithm [110]. BCFtools and vcfutils from SAMtools [111] were used to obtain the fastq files from SAM files, and the fasta sequences were transformed from fastq with Seqtk [112]. The sequences were concatenated and a maximum likelihood phylogenetic tree was obtained with IQ-TREE 2 software [113]. As we obtained two well-defined clusters for *S.* Infantis, we annotated one representative of each plasmid cluster with Prokka [114] and performed an orthologous genes analysis (coverage of 90%, similarity on protein sequences of 80%) with Proteinortho5 [115]. A manual comparison of all genes present in the plasmids was carried out. We performed the same analysis for non-Infantis isolates.

### 2.5.3 Megaplasmid Analysis

Two megaplasmids (pESI and p-F219) commonly associated to pathogenic and MDR strains of *S.* Infantis were analyzed in order to identify their relatedness. We used D-Genies [116] to obtain a dot plot of genome comparison, a genome alignment with progressiveMauve [117] in order to identify locally collinear blocks, and an ANI calculation [118] for computing average nucleotide identity in sequences shared by both plasmids.

### 2.5.4 Core Genome and Metadata Analysis

A Peruvian *S.* Infantis strain, FARPER-219 [109], and two Ecuadorian isolates (SRR4019589 and SRR4019602) analyzed by the US Centers for Disease Control and Prevention from two patients that developed salmonellosis after traveling to Ecuador [86] were added as references in the phylogenetic analysis of all *S.* Infantis isolates. From trimmed reads, Spades [119] was used to generate assemblies. Later, genome annotation was performed with Prokka [114]. An orthologous genes analysis with the same conditions as for plasmid detection was performed with Proteinortho5 [115]. The strict core genes, those present in all the isolates, were extracted with the Proteinortho tool: grab_proteins.pl. Mafft [120] and an in-house script were used for multiple alignment of every gene and subsequent concatenation in a single multiple alignment, respectively. The phylogenetic tree from the core genome alignment was obtained using IQ-TREE 2 [113] with 1000 bootstrap replicates. The metadata for sample origin, phenotypic antibiotic resistance patterns, and plasmid *in silico* detection was added to the final tree with iTOL tools [121].

# 3. Results

## 3.1. *Salmonella* Prevalence and Serotype Identification

*Salmonella* was present in 41.4% (55/133; $CI_{95\%}$:33-49.7), 55.5% (186/335; $CI_{95\%}$:50.2-60.8) and 1.98% (6/302; $CI_{95\%}$:0.4-3.6) in poultry farms, chicken carcasses and human stool samples, respectively. *S.* Infantis accounted for 98.2% (*n*=54) of isolates from poultry farms, 97.8% (*n*=182) of isolates from chicken carcasses, and one half (*n*=3) of human sample isolates. Additionally, one isolate was typed as *S.* Enteritidis in broiler flocks; at the retail level one and three isolates were typed as *S.* Typhimurium and *S.* Enteritidis, respectively, while in the human stool samples two isolates were typed as *S.* Enteritidis and one isolate corresponded to monophasic *S.* Typhimurium 4,[5],12:i:- (Supplementary Table 1).

## 3.2 Antimicrobial Resistance

For *S.* Infantis isolates, antimicrobial resistance rates to nitrofurantoin, tetracycline, sulfamethoxazole + trimethoprim, streptomycin, gentamicin, cefotaxime, ciprofloxacin and chloramphenicol ranged from 64.8% to 100%. On the other hand, fosfomycin and azithromycin resistance rates were lower, ranging from 0% to 42.6%. Only one isolate from a stool sample presented phenotypic resistance to amikacin while none of the *Salmonella* isolates in this study was resistant to ertapenem (Table 1).

Considering antimicrobial classes that were tested, *S.* Infantis isolates presented 43 antimicrobial resistance patterns. With the exception of one isolate from a stool sample, all isolates showed multidrug-resistant phenotypes. Importantly, 87% and 82% of isolates from poultry farms and chicken carcasses, respectively, presented resistance from 6 up to 9 classes of antimicrobials (Supplementary Table 2). One *S.* Infantis isolated from chicken carcasses could not be recuperated for this analysis.

*Salmonella* serotypes other than *S.* Infantis also presented multiresistant patterns, except for 3 *S.* Enteritidis isolates that were only resistant to one group of antimicrobials. For this set of isolates, every resistant pattern included isolates belonging to only one serotype (Table 2).

One isolate of *S.* Enteritidis, one of *S.* Typhimurium and one of Monophasic *S.* Typhimurium 4,[5],12:i:-; and 205 isolates of *S.* Infantis were identified as resistant to beta-lactam antibioticss.

**Table 1**. Number of S. Infantis isolates resistant to each tested antimicrobial.

| Antimicrobial | Number (%) of resistant isolates | | |
|---|---|---|---|
| | **Poultry farms (farm)** | **Chicken carcasses (food)** | **Stool samples (human)** |
| Nitrofurantoin | 54 (100) | 180 (99,4) | 2 (66,7) |
| Tetracycline | 54 (100) | 176 (97,2) | 3 (100) |
| Sulfamethoxazole + trimethoprim | 44 (81,5) | 158 (87,3) | 1 (33,3) |
| Streptomycin | 46 (85,2) | 154 (85,1) | 3 (100) |
| Gentamicin | 45 (83,3) | 155 (85,6) | 2 (66,7) |
| Cefotaxime | 51 (94,4) | 150 (82,9) | 1 (33,3) |
| Chloramphenicol | 45 (83,3) | 149 (82,3) | 2 (66,7) |
| Ciprofloxacin | 35 (64,8) | 116 (64,1) | 1 (33,3) |
| Fosfomycin | 23 (42,6) | 68 (37,6) | 1 (33,3) |
| Azithromycin | 10 (18,5) | 31 (17,1) | 0 (0) |
| Cefoxitin | 7 (13) | 11 (6,1) | 0 (0) |
| Amoxicillin + clavulanic acid | 7 (13) | 7 (3,9) | 0 (0) |
| Amikacin | 0 (0) | 0 (0) | 1 (33,3) |
| Ertapenem | 0 (0) | 0 (0) | 0 (0) |

**Table 2**. Antimicrobial resistance patterns of S. Enteritidis, S. Typhimurium, and monophasic S. Typhimurium 4,[5],12:i:-.

| Resistant pattern | No. Antimicrobial classes | *S*. Enteritidis | *S*. Typhimurium | Monophasic S. Typhimurium 4,[5],12:i:- |
|---|---|---|---|---|
| SAQBTFN | 7 | | 1 | |
| SAQTFNM | 7 | 1 | | |
| SABTFNP | 7 | 1 | | |
| SABTFN | 6 | | | 1 |
| QFM | 3 | 1 | | |
| N | 1 | 3* | | |

* 1 isolate obtained from poultry farms and 2 isolates from human stool samples.

Sulfonamide (S), aminoglycosides (A), quinolones (Q), Beta-lactams (B), tetracyclines (T), phenicol (F), nitrofuran (N), macrolides (M), Fosfomycin (P)

Six *S*. Infantis isolates from chicken carcasses and one from poultry farms did not present any of the investigated ESBL genes. All other *S*. Infantis and one *S*. Enteritidis isolated from a carcass presented the $bla_{\text{CTX-M-65}}$ gene.

## 3.3 Genomic Analysis

For WGS analysis, 144 isolates (40 from the animal component, 98 from the food component and 6 from the human component) were selected. Raw sequence data is available under bioproject PRJEB37560. The sequences from three samples were not enough to perform genomic analysis. The obtained average number of reads per strain was 1,356,678 (range 247,022-14,106,025) and after the quality control steps, the average number was 1,266,242 (range 228,263-13,094,594) (Supplementary Table 3). Average Phred Score was Q34.

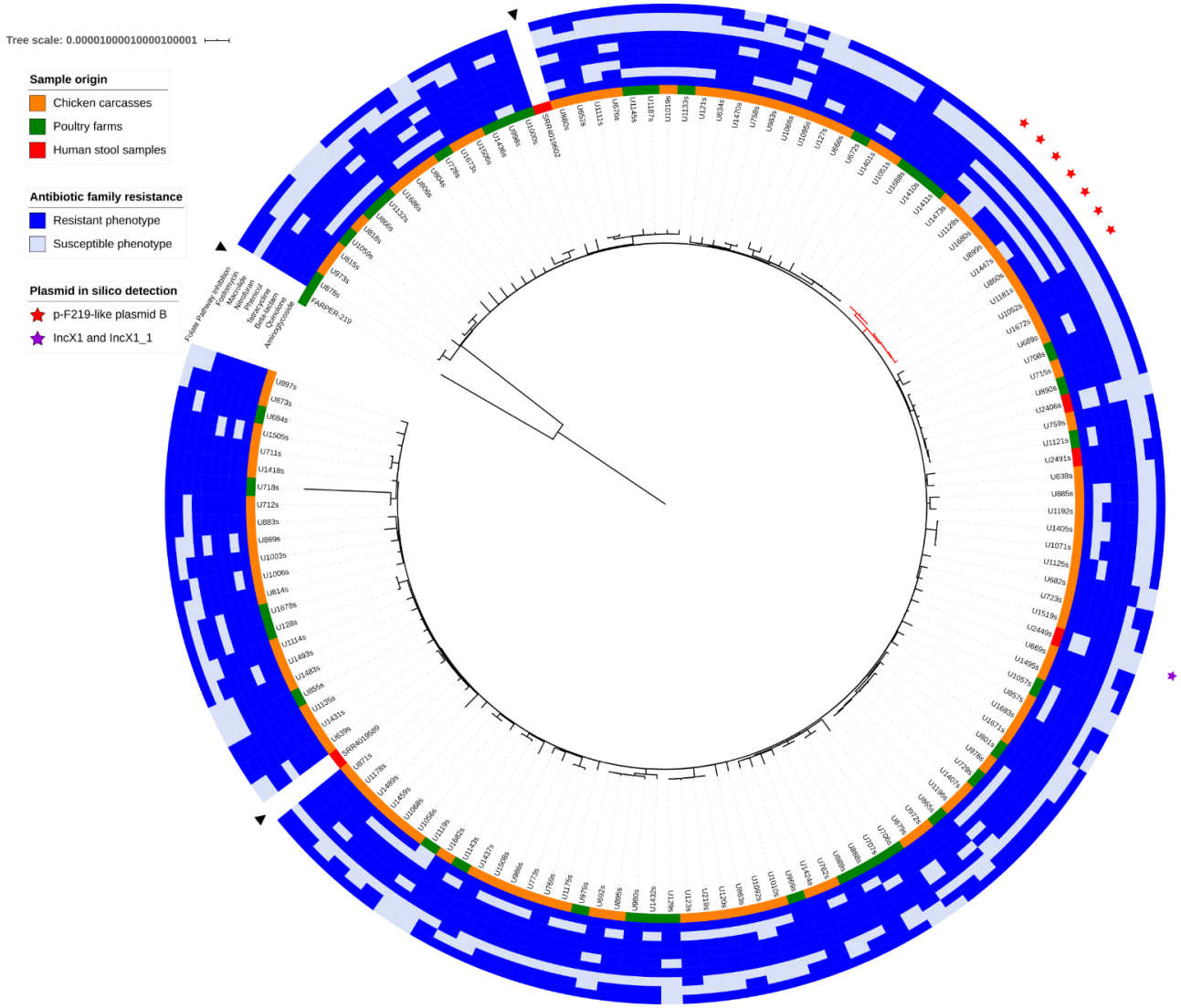



**Figure 3**. Maximum likelihood phylogenetic tree of core genome alignment of all 137 S. Infantis isolates based on 3,552 genes. Two S. Infantis genomes from Ecuador detected in USA (SRR4019589, SRR4019602) and a Peruvian strain (FARPER-219) were included in the analysis and are indicated with a black triangle. The origin of each sample is colored in red for human stool samples, in green for poultry farms isolates and in orange for chicken carcasses strains. The phenotypic resistance for nine antibiotic families is marked with a blue box. Strains with p-F219-like plasmid B are marked with a red star. The rest of the samples harbor the p-F219-like plasmid A. The purple star indicates the presence of IncX1 and IncX1_1 plasmids in one of the strains. Digital version of the phylogenetic tree is available with iTOL login LMejia at https://itol.embl.de/shared_projects.cgi.

MLST typing showed that all *S*. Infantis isolates (*n*=137) belonged to ST32. The five *S*. Enteritidis isolates belonged to ST11. Additionally, the single isolates of *S*. Typhimurium and monophasic *S*. Typhimurium 4,[5],12:i:- belonged to ST19 and ST2379, respectively (Supplementary Table 1).

The strict core genome of all *S*. Infantis included in the analysis corresponded to 3552 genes and spanned 3,161,448 bp, 1,414 of which were variable (SNPs). The alignment of the concatenated genes present in this core was used to obtain a maximum-likelihood tree using FARPER 219 as outgroup (Figure 3). This strain was chosen because it was isolated in Peru, a neighbor country to Ecuador. Despite its inclusion in ST32, FARPER-219 presented genetic divergence with the Ecuadorian strains. The two *Salmonella* genomes from the USA (SRR4019589, SRR4019602) grouped indistinctive with some of the genomes of this study. Notably, the analyzed strains did not group according to their sampling origin or their phenotypic resistance patterns.

Genes of antimicrobial resistance were also confirmed with WGS data (Supplementary File 1). For most of the antimicrobial classes (folate pathway inhibition, aminoglycoside, beta-lactams, tetracyclines, fosfomycin and phenicol) correspondence rates were higher than 80%. However, no genes responsible for the phenotypic resistance to quinolones, nitrofurans and macrolides were found in sequenced isolates (Table 3).

**Table 3**. Comparison of phenotypic AMR with AMR genes obtained from WGS data.

| **Antibiotic family** | **Phenotype (%)**[a] | **Phenotype + AMR gene (%)**[b] | **No phenotype + AMR gene (%)**[c] |
|---|---|---|---|
| Folate pathway inhibition | 84,56 | 92,17 | 7,83 |
| Aminoglycoside | 97,06 | 100 | 3,03 |
| Quinolone | 58,09 | 1,27 | 0 |
| Beta-lactams | 84,56 | 94,87 | 4,27 |
| Tetracycline | 97,06 | 99,24 | 0,76 |
| Phenicol | 80,88 | 80,91 | 6,36 |
| Nitrofuran | 98,53 | 0 | 0 |
| Macrolide | 17,65 | 0 | 0 |
| Fosfomycin | 37,5 | 90,2 | 41,18 |

Regarding virulence genes, *sifA, sseL, pipB, sopD2* and *srlP* (part of SPI 2) were found in the core genome of all *S*. Infantis strains. The *lpf* operon that encodes the long polar fimbriae (LPF), the *fim* gene cluster that encodes type 1 fimbriae, and the *csg* operons that encodes the Tafi fimbriae (Thin aggregative fimbriae) were also present in these genomes.

The presence of plasmids was confirmed *in silico* by PlasmidFinder. Only one *S*. Infantis isolate (U2449s) presented one plasmid determinant (IncX1 and IncX1_1) despite the multidrug resistance patterns found in our isolates (Fig. 1 and Supplementary File 1). To further analyze the low incidence of plasmids, we mapped the raw reads from our samples against the megaplasmid p-F219 described by Vallejos-Sánchez et al. [109]. We found two p-F219-like plasmids. The first one, denoted as plasmid A, was present in most of the strains. This plasmid contained 338 genes shared with the p-F219 megaplasmid. The second one, denoted as plasmid B, was present in the remaining 7 strains. Plasmid B lacked 72 of the genes present in plasmid A and presented 6 exclusive genes (Supplementary Table 4). The strains that presented plasmid B belonged to a monophyletic clade (denoted with a red star in Fig. 1). These strains also share susceptibility to fosfomycins, macrolides, phenicols and beta-lactams. *Salmonella* strains harboring plasmid B were isolated from chicken carcasses sampled during different weeks of the year, different parts of the city and different types of retail stores (data not shown).

New hypothetical proteins (n=147) were found in both plasmids; 43 exclusively found in plasmid A and 5 in plasmid B (Supplementary Table 4).

The comparison of the p-F219 plasmid with another megaplasmid, commonly found in pathogenic *S.* Infantis strains, pESI plasmid, showed that they share more than 79% of their sequences (>75% of identity) (Supplementary Figure 1). Besides, we noticed a large genomic inversion in plasmid p-F219 when compared with the pESI plasmid that is also observable in progressiveMauve genome alignment (Supplementary Figure 2). From the ANI calculation, 99.41% of identity was found in the ortholog genes present in both samples.

*In silico* plasmid detection in non-Infantis isolates showed the presence of two plasmids in all *S*. Enteritidis strains, while monophasic *S*. Typhimurium 4,[5],12:i:- and *S*. Typhimurium presented 1 and 5 plasmids, respectively (Supplementary Table 1).

## 4. Discussion

In the last decades, there has been a clear rise in the prevalence of multidrug resistant *Salmonella enterica* worldwide, especially of serovar Infantis [87,122–124].

To better explain the epidemiology of *Salmonella* in Ecuador, we studied *S. enterica* isolated from poultry farms, chicken carcasses and human stool samples in Quito. The prevalence of *Salmonella* was high in poultry farms and chicken carcasses, similarly to other studies in the region [83,84,125]. Although this research did not look for *Salmonella* in earlier stages of the broiler production chain, previous studies in Ecuador have reported the importance of compound feed,

one-day-old chicks and broiler pens in the *Salmonella* exposure of broilers [83,126]. These results highlight the necessity to improve broilers production systems in Ecuador towards a better control of *Salmonella* in the food chain.

In this study, it was also seen that samples from different kinds of markets delivered similar rates of *Salmonella* isolates, denoting that the geographical distribution of retailers does not influence the presence of *Salmonella* in carcasses (Supplementary Table 5). On the other hand, all human isolates originated in the southern health care center. However, it must to be considered that the low *Salmonella* prevalence in human samples could be influenced by the fact that other pathogens might be the main causes of diarrhea. This fact has been studied in Ecuador were other viruses, parasites and bacteria are the main cause of gastroenteritis cases [127–130]. Additionally, the state of health carriers should be considered when accessing the real prevalence of *Salmonella* in humans [131]. These circumstances represent limitations of this study and should be considered in future research.

The predominance of *S.* Infantis in this study is in accordance with other reports in the world that show that this serotype is becoming an emergent pathogen. For example, in Europe, *S.* Infantis has been reported to be one of the most common serovars in poultry and in humans [132]. The same tendency has been reported in the neighboring country of Peru, where *S.* Infantis is the most prevalent serotype in broilers [82]. However, a wider variety of serotypes has been reported in other Latin American countries [125,133,134]. This could be explained by the fact that the poultry industry of Peru and Ecuador have close commercial interactions which could determine a common epidemiology of this pathogen. Nevertheless, a recent publication from Chile reported that 24% of broiler meat samples (n=361) were positive to the isolation of *S.* Infantis [135]. This data highlights the necessity of more research in the field to better understand the epidemiology of *Salmonella* in the region.

Several studies have shown that control programs of targeted *Salmonella* serotypes could have favored the occurrence of other serotypes [35,136]. This kind of shifts might explain to some extent the low presence of *S.* Enteritidis and Typhimurium among our samples but further research is needed to prove this hypothesis.

Most *Salmonella* isolates in this study presented multidrug resistance (MDR) phenotypes. This issue is especially evident in *S.* Infantis, as already been seen in Ecuador [83,84] and other countries of the region [137–139]. Although in a lower extent, this feature has also been reported in Europe where high rates of MDR (up to 85%) are reported in *S.* Infantis isolated from poultry [140]. The high levels of resistance in *Salmonella* isolates in Latin American could be related to

the intensive use of antimicrobials in poultry production as prophylactics, therapeutics and growth promotors [141,142].

*S*. Infantis isolates belonged to ST32, that is among those more frequently identified within this serovar [107,143]. All Ecuadorian *S*. Infantis isolates showed a high genomic similarity with an apparently common origin. However, in order to verify that all in fact share a common ancestor, a larger analysis including isolates from countries around the world is necessary for identifying the origin of this serovar in Ecuador. The close similarity found between isolates from farms, animals and humans show that this pathogen may be responsible for human infections through the food chain.

We did not observe any clear clustering of *S*. Infantis isolates and antibiotic phenotypic resistance patterns, what makes sense since the phylogenetic tree was obtained from the core genome and most of the antimicrobial resistance determinants are expected to be part of the accessory genome.

The resistance genes analysis carried out here could not successfully explain resistance to macrolides, nitrofurans and quinolones in our isolates. In fact, phenotypic patterns and genetic detection correlation differences have been described previously [144–146], but point mutations not considered in this study may explain these phenotypes. Additional studies are still needed in order to identify the genes or mutations responsible for antimicrobial resistance in *S*. Infantis.

We found two p-F219-like plasmids (named A and B) present in all the analyzed isolates. Our samples share a significant number of genes with the p-F219 plasmid. The similarity between both p-F219-like plasmids requires a recent common ancestor from which they may have evolved. These plasmids may provide fitness and pathogenic advantages to the strains since they contain several genes for antimicrobial resistance, fimbriae, transposases and environmental stress tolerance (Supplementary Table 4). Other *S*. Infantis isolates around the world have been shown to also harbor pESI-related megaplasmids that appear to be the difference between non-pathogenic and pathogenic isolates since it also provides antimicrobial resistance, oxidative stress tolerance, pathogenicity traits and mercury environmental tolerance [143,144,147,148]. Besides, pESI has shown more pathogenicity and increased intestinal inflammation in experimental mice infections when compared to the plasmid-free isolates [147].

Both plasmids (p-F219 and pESI) share regions of very close similarity with an extensive genomic inversion. Genes present in both plasmids share more than 99% average nucleotide identity suggesting that p-F219 is actually a pESI-like plasmid and the variants found in our samples may also be cataloged as such. A genomic comparison between *S*. Infantis strains

showed that large plasmids from multiple isolates actually share a pESI backbone with some genetic plasticity to add different mobile genetic elements due to insertion sequences [149]. As pESI-like plasmids confer a MDR phenotype but also several virulence factors and tolerance to environmental stress, their acquisition may have been involved for making *S.* Infantis a successful emerging pathogen worldwide.

The negative results from the unmapped reads from *S.* Enteritidis and *S.* Typhimurium isolates against the p-F219 plasmid suggest that this plasmid may not be widespread among other *Salmonella* serovars, but more isolates need to be analyzed to confirm this hypothesis.

Other virulence genes in *Salmonella* are essential for pathogenicity and infection. We looked for some of the more relevant virulence and invasion determinants in the core genome of *S.* Infantis as determined by ProteinOrtho. Some genes that are usually identified as part of the *Salmonella* pathogenicity island 2 (SPI-2) were found: *sifA, sseL, pipB, sopD2, srlP* [149]. The long polar fimbriae (LfP), encoded by the *lpf* operon, was found in all samples as putative proteins in all isolates; it is believed that this protein is involved in adhesion and growth on the small intestine mucosa [150], evasion of the host immune system [151]. Moreover, variations in these genes may also impact the host range of the bacteria [152]. We also found the *fim* gene cluster that encodes type 1 fimbriae and it is involved in the initiation of biofilm formation [153]. Another determinant found in our strains was the *csg* operon that encodes the Tafi fimbriae (Thin aggregative fimbriae). Tafi is responsible of adhesive activities and biofilm formation [150]. The presence of these genes in *S.* Infantis may suggest a pathogenic character of this serotype that, together with the multidrug resistant profiles, represent a potential public health concern.

To the best of our knowledge, this is the first report based on *Salmonella enterica* in Ecuador that uses phenotypic and WGS information to analyze the relatedness of strains isolated from poultry, food and human samples. Isolates from this study show multidrug resistance patterns highlighting the importance of a reduced and better usage of antimicrobials in intensive poultry farms settings. The presence of related megaplasmids together with the high similarity of the core genome may suggest the dissemination of *S.* Infantis through the food chain to humans. The data presented here has shown the importance of *Salmonella enterica* serovar Infantis as a foodborne pathogen in Ecuador and provide critical information about its clonality and circulating strains.

# Chapter 2

## *Listeria monocytogenes,* a silent foodborne pathogen in Ecuador.

# 1. Background

Listeriosis is a notifiable disease in North America and in many European countries, but not in Ecuador. Considering the serious consequences of listeriosis, its widespread distribution, and the absence of data from Ecuador, we used WGS to genotype and characterize *L. monocytogenes* isolates recovered from cheeses and clinical samples to assess the potential risk of this pathogen to cause disease in Ecuador.

# 2. Material and methods

## 2.1. Bacterial strains

Sixty-five *Listeria monocytogenes* isolates were evaluated in this study. From the National Institute of Public Health Research in Ecuador, we had access to all *L. monocytogenes* isolates from the culture collection from 2015 until 2018 (14 isolates from cheese samples and 20 clinical isolates from listeriosis cases). We also had access to 31 isolates from artisanal cheese samples from 8 provinces in Ecuador from a previous study [154]. From these isolates, the antimicrobial susceptibility testing was assessed by micro dilution technique, using panels of lyophilized antibiotics for Gram-positive bacteria (Sensititre GPALL1F, Thermo Scientific ®) and their serogroup was determined by PCR as described by Tao and colleagues [155].

## 2.2. Whole Genome Sequencing

Genomic DNA from all the isolates was purified using Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) following manufacturer's guidelines for Gram positive bacteria. Next, an Invitrogen Qubit 3.0 fluorometer (Thermo Fisher Scientific, Walthman, MA, USA) was used for DNA quantification. DNA samples were sequenced with the Illumina NextSeq platform using Nextera XT Library Preparation Kit, and 150 x 2 bp paired-ends sequences were obtained. Trimmomatic [102] was used for reads trimming, and Fastqc [103] and Multiqc [104] were employed for quality assessment. Default parameters were used for the following software unless otherwise specified.

## 2.3. Sequence type determination, antimicrobial resistance genes, virulence factors and plasmid detection.

The software ARIBA version 2.11.1 [106] was used with public databases CARD, MEGAres and ResFinder to identify genes that may confer antimicrobial resistance (AMR). The databases VirFinder for virulence genes, PlasmidFinder for plasmid determinants, and the PubMLST

from Institut Pasteur [156] for *Listeria monocytogenes* for ST determination were also used with the ARIBA software.

We also wanted to determine the presence of plasmids previously described in other *L. monocytogenes* strains. We performed a manual search in public databases and downloaded 29 plasmids from NCBI. In order to see how similar these plasmids were, Proteinortho5 [115] was used to determine orthology relationships, where 80% for coverage and 80% for similarity on protein sequences were set. With these results, we manually eliminated plasmids with similar gene patterns. Then, we selected 13 plasmids for the mapping step. We mapped our raw data against the plasmids following the strategy described in *Salmonella* megaplasmids detection [78]. With the fasta sequences from each strain, we performed an alignment with every plasmid in order to evaluate their coverage. Finally, we classified the plasmid presence in ranges for each strain (20-49%, 50-79%, and 80-100%). Plasmid coverages lower than 20% were discarded as absent. After analyzing the high length covered of seven plasmids (pMF6172, pl2015TE24968, PAUSMDU00000224_01, pPIR00540, pLM58, pMF4545, pPIR00541), we performed an additional comparison based on orthologs and on average nucleotide identify [118]. We also used progressiveMauve [117] in order to identify locally collinear (LCB) blocks and inversions.

## 2.4. Phylogenetic analysis

Spades [119] was used to obtain the sequence assemblies from trimmed reads. Twenty-six *L. monocytogenes* closed genomes were downloaded as references (Supplementary File 2). Then, Prokka [114] was used for genome annotation. Next, determinations of orthologs among all *L. monocytogenes* isolates and among ST2 isolates was performed independently with Proteinortho5 [115] (90% for coverage and 80% for similarity on protein sequences). Genes in strict core for all isolates were extracted with grab_proteins.pl script from Proteinortho5. Then, all the genes were aligned and concatenated in a multifasta file with Mafft [120] and an in-house script. The core genome alignment was used for obtaining a maximum-likelihood phylogenetic tree with IQ-TREE 2 [113] with 1000 ultrafast bootstrap replicates to evaluate branch supports[157]. iTOL [121] was used to visualize the final trees and to add the metadata.

## 2.5 Genetic factors determination

In order to detect genes encoding major *Listeria* virulence factors (LIPI-1, LIPI-2, LIPI-3, LIPI-4 and internalins *inlA* and *inlB*), genes associated to biofilm formation (*flaA, luxS, cheY*) and genes from the stress survival islets (SSI-1, SSI-2, SSI-F2365), blastn was run with the

reference genes (Supplementary Table 6) as subjects and the sequence assemblies as query. The percentage identify was set to 75%.

# 3. Results

## 3.1. Raw data

Raw sequence data is available under bioproject PRJEB48671. We obtained whole genome sequence data from 65 *L. monocytogenes,* 1 *L. seeligeri,* 1 *L. welshimeri* and 5 *L. innocua* isolates. In this article, we focused mainly on *L. monocytogenes* isolates. The obtained average number of reads in millions per strain was 1.5 (range 0.3-2.6) (Supplementary Table 7). Average Phred Score was Q34. One *L. monocytogenes* isolate (Lm73) was excluded from subsequent analyses because the core genome of all isolates including Lm73 was composed of 1227 orthologous genes, and without Lm73 it was composed of 2112 genes. This isolate lacked 885 orthologous genes that were present in the rest of strains (data not shown).

## 3.2. Sequence Types and sample origin

A total of 64 *L. monocytogenes* isolates were analyzed (Supplementary File 3). They were isolated from various types of cheeses, including soft cheese (*n*=31), kneaded cheese (*n*=7), chopped cheese (*n*=3), unsalted cheese (*n*=2), curd (*n*=1), but also from listeriosis cases, including blood (*n*=19), and cerebrospinal fluid (*n*=1) samples. These samples were collected from 2015 to 2018 and 8 Sequence Types (STs) were detected. Table 4 shows the STs among food and clinical samples. The predominant ST was ST2 (84.62% 55/65), and other STs were found in less than 4% of the samples. ST2 was the only ST found in food and clinical samples.

**Table 4.** *Listeria monocytogenes* sequence types (STs) among sample origin.

| Sample origin | ST | *n* (%) |
|---|---|---|
| Food | ST2 | 42 (93.33) |
| | ST6 | 1 (2.22) |
| | ST796 | 2 (4.44) |
| Clinical | ST2 | 13 (65.00) |
| | ST1 | 1 (5.00) |
| | ST3 | 2 (10.00) |
| | ST4 | 1 (5.00) |
| | ST378 | 2 (10.00) |
| | ST392 | 1 (5.00) |

## 3.3. Antimicrobial resistance

All isolates were previously tested and showed susceptibility to penicillin, ampicillin, erythromycin, trimethoprim/sulfamethoxazole, and meropenem [154]. Genetic determinants for AMR were also detected with WGS data with three databases as shown in Table 5. The determinants *fosX* and *mprF_2* were found only by CARD database, while GYRA_23 and TUFAB_7 were only found by Megares, and *norB* and *tetS* were detected by 2 and 3 databases, respectively.

**Table 5.** *Listeria monocytogenes* genetic determinants for antimicrobial resistance.

| Determinant | CARD | Resfinder | Megares | *n* |
|---|---|---|---|---|
| FosX (fosfomycin) | 1 | 0 | 0 | 65 |
| GYRA_23[1] | 0 | 0 | 1 | 63 |
| MprF_2[2] | 1 | 0 | 0 | 62 |
| norB[3] | 1 | 0 | 1 | 63 |
| tetS (tetracycline) | 1 | 1 | 1 | 4 |
| TUFAB_7[4] | 0 | 0 | 1 | 65 |

[1] GYRA is a component of the bacterial DNA gyrase. Mutations in this protein may cause fluoroquinolones resistance [158].

[2] MprF confers resistance to cell membrane disrupting cationic peptides (defensins) [159].

[3] NorB confers resistance to fluoroquinolones and tetracycline [159].

[4] Mutations in TUFAB (components of EF-Tu) may cause resistance to elfamycins [158].

Two isolates (Lm03 and Lmo22) presented 9 and 3 other AMR determinants apart from the ones detected in most samples, and they lack *tetS* gene (Supplementary File 3).

## 3.4 Genomic analysis

From the raw data, no plasmids or virulence determinants were found by PlasmidFinder and VirulenceFinder, respectively (data not shown). The core genome of all *L. monocytogenes* (*n*=64) and references (*n*=26) encompasses 2112 genes spanning 1,896,388 nt with 175,308 variable sites (9.3% of complete sequence alignment). This alignment generated a maximum-likelihood tree shown in Figure 4, where bootstrapping values of 90 or higher are displayed with a black circle. Isolates grouped according to the evolutionary lineages previously determined [51]. Most of our samples clustered into lineage I. The collection date, sequence type, isolate source, genetic lineage, presence of SSI-1, SSI-F2365, LIPI-1, LIPI-3, LIPI-4 and internalins are also depicted.

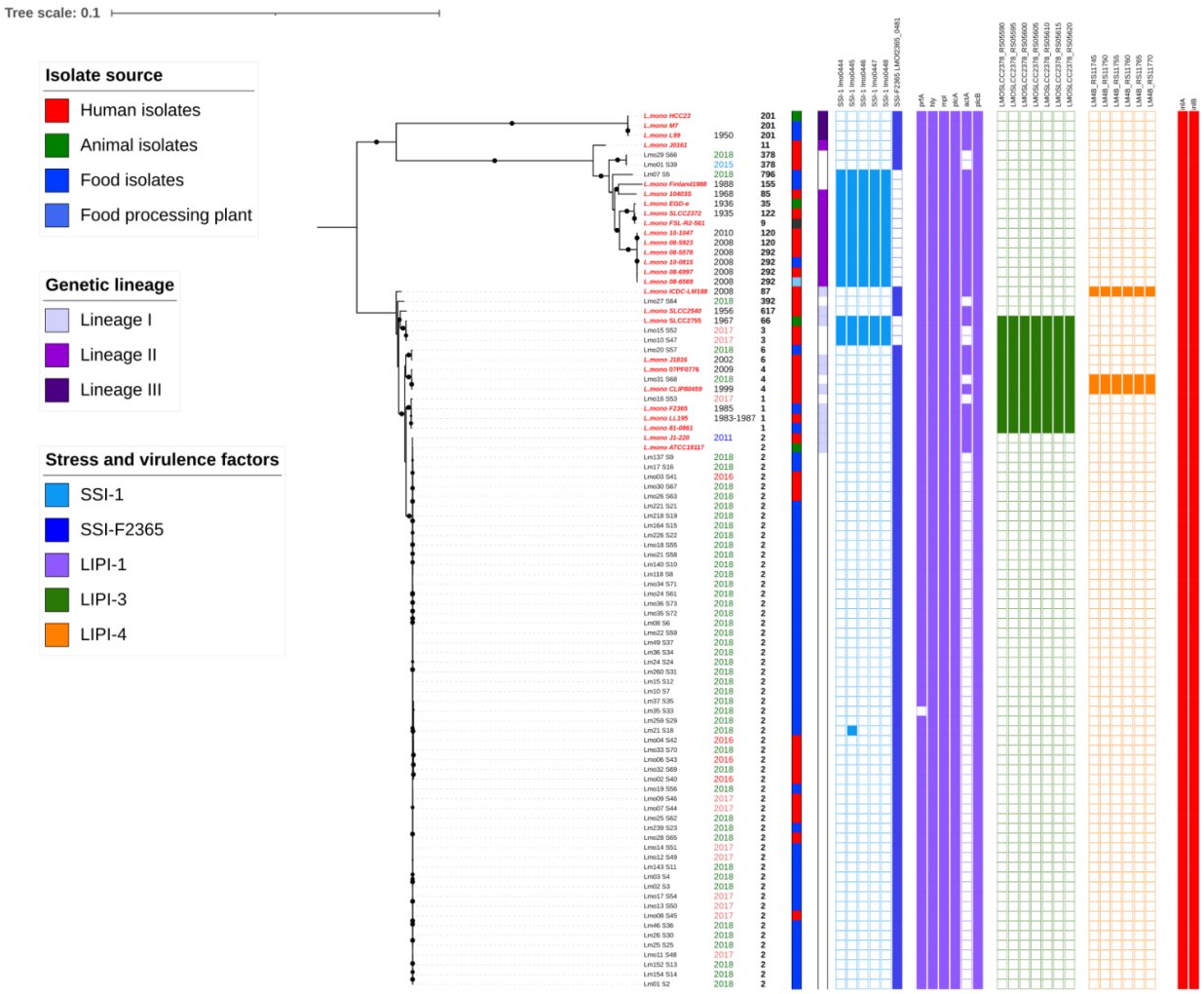



**Figure 4.** Phylogenetic tree from core genome alignment of 64 *L. monocytogenes* isolates and 26 references (2112 genes are included). Reference sequences names are labeled red and in italics. The isolation date and ST are indicated next to isolates names. The isolate source, genetic lineage, presence of SSI-1, SSI-F2365, LIPI-1, LIPI-3, LIPI-4 are represented as binary data. Colored squares indicate presence. Digital version of the phylogenetic tree is available on iTOL: https://itol.embl.de/shared/epimol

With respect to the stress survival islets, samples presented either SSI-1 or SSI-F2365, except for isolate Lm21 that presented SSI-F2365 but also one gene from SSI-1. Only 3 of our isolates (2 ST3 clinical isolates and 1 ST796 food isolate) contained SSI-1. None of the samples presented SSI-2, which is common in *L. innocua*.

Regarding the presence of pathogenicity islands and virulence factors, the isolates exhibited heterogeneous patterns. They all exhibited at least 5 genes from LIPI-1 (*prfA, plcA, hly, mpl, actA, plcB*), but all references showed the complete island. Only two Ecuadorian isolates (Lm07 and Lmo20) harbored all genes from LIPI-1. Most of our isolates did not reveal presence of gene *actA* or the percentage identities were lower than 75% with respect to the reference gene. Genes from LIPI-2, which are usually found in *L. ivanovii,* were absent in all isolates.

LIPI-3 was only present in 13 isolates from lineage 1 (7 references of ST66, ST6, ST4 and ST1). Clinical isolates Lmo15 (ST3), Lmo10 (ST3), Lmo31 (ST4), Lmo16 (ST1) and food isolate Lmo20 (ST6) presented all genes from LIPI-3. LIPI-4 was only present in 3 isolates (2 references with ST87 and ST4 and Lmo31). Internalins A and B were present in all isolates and references. All isolates and references also showed the presence of genes associated with biofilm formation (*flaA, luxS, cheY*) (data not shown).

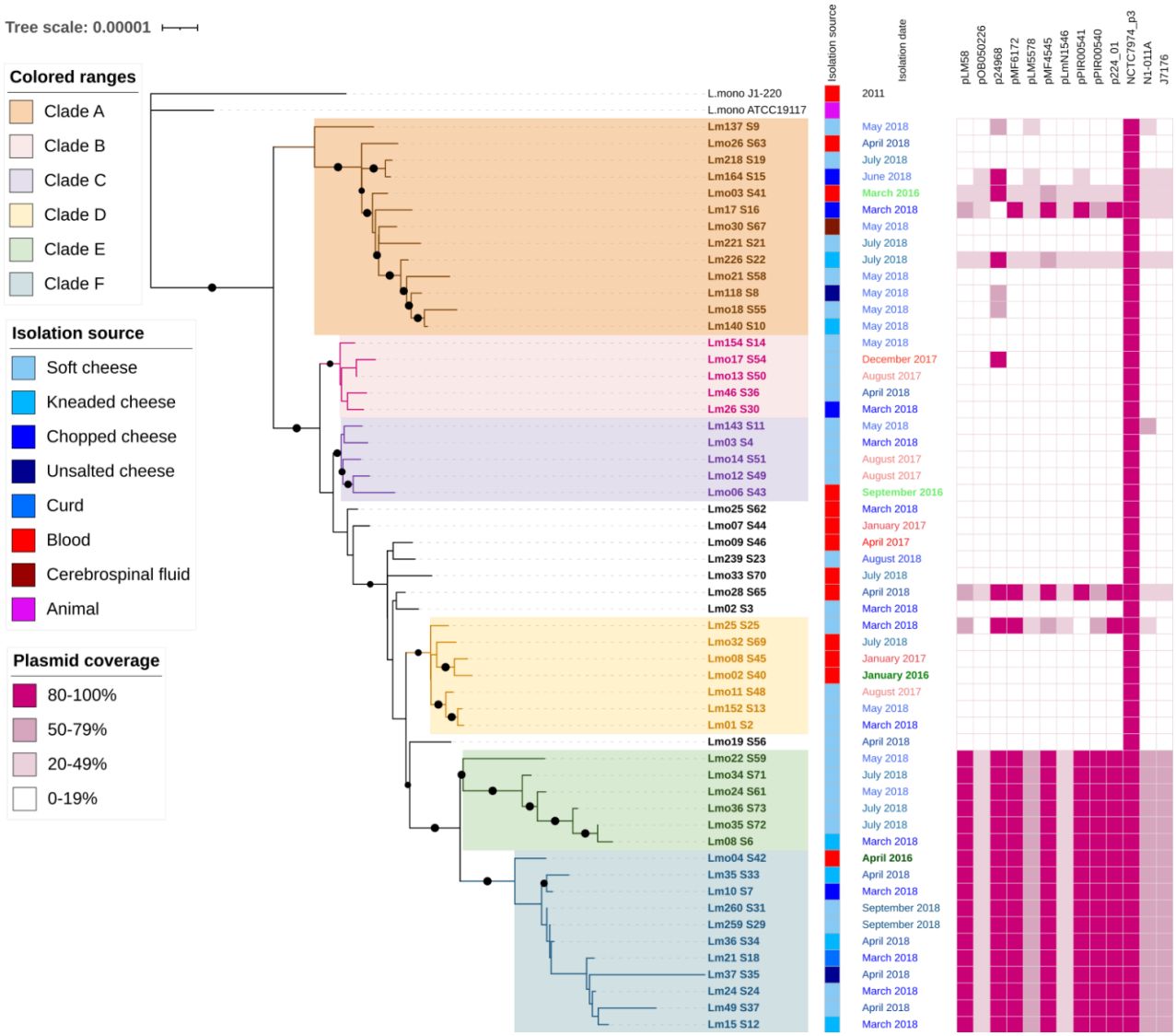



**Figure 5**. Phylogenetic tree from core genome alignment of 55 *L. monocytogenes* isolates and 2 references (2112 genes are included) that belong to ST2. The isolation source, date and place are indicated next to isolates names. The plasmids coverage is presented in ranges (20-49%, 50-79%, and 80-100%). Colored squares indicate presence. P24968 corresponds to pl2015TE24968. P224_01 corresponds to pAUSMDU00000224. Digital version of the phylogenetic tree is available on iTOL: https://itol.embl.de/shared/epimol

Since most of our samples belong to ST2, we obtained a second phylogenetic tree based on the alignment of 57 ST2 strains, including two references, which is presented in Figure 5

(Bootstrap values BS>=90% are shown in branches with a black circle). The multiple alignment included 2376722 bp with only 1022 variable sites (less than 0.05% of complete alignment). The isolation source and collection date are specified as metadata. Clades (A-F) are colored for better display. Strains in clade B and E were isolated only from cheeses, while the rest clades include human and food samples. The presence of thirteen plasmids is shown in ranges (20-49%, 50-79%, and 80-100%) representing global similarity. All isolates presented the whole NCRC7974_plasmid3, while clade E, part of F and Lm08_S6 (from clade B) showed the same plasmid pattern, a 80-100% coverage of plasmids pLM58, pl2015TE24968 (p24968), pMF6172, pMF4545, pPIR00541, pPIR00540, PAUSMDU00000224_01 (p224_01); 50-79% of plasmids pLM5578, N1-011A and J1776; and 20-49% of plasmids pOB050226 and pLmN1546.

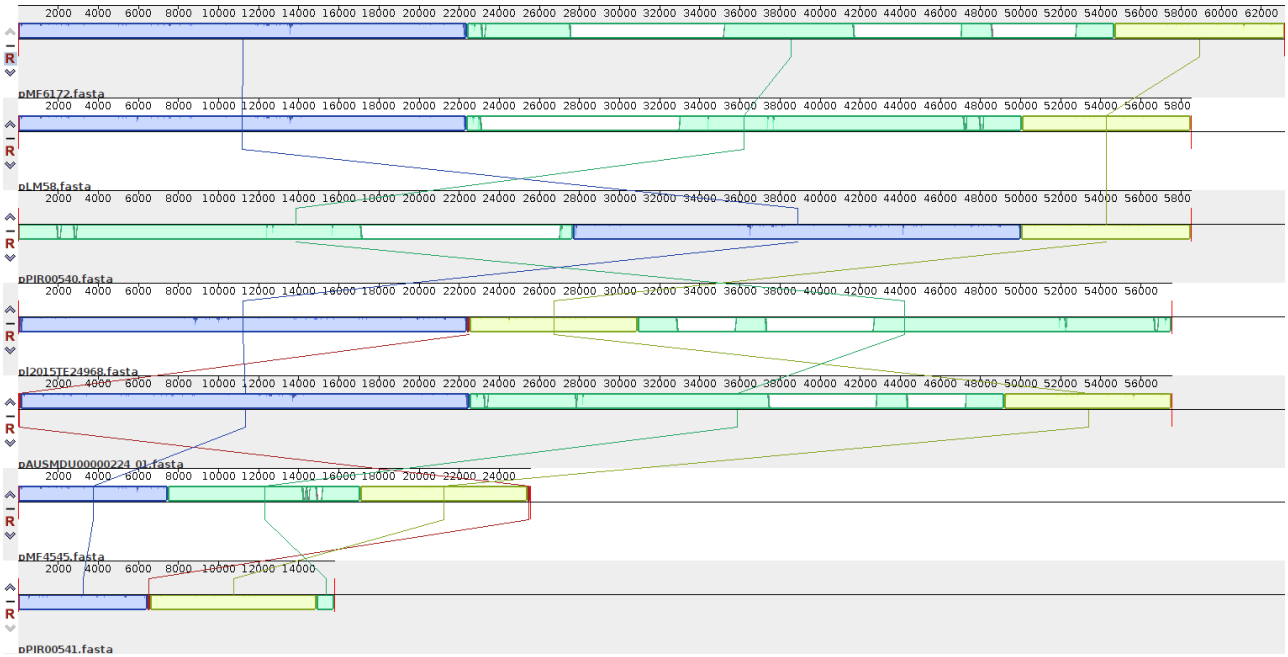



**Figure 6.** Locally collinear blocks (LCB) among the plasmids detected in Ecuadorian *L. monocytogenes* isolates. LBCs are colored according to the conserved fragments that are possibly homologous and free of internal genome rearrangements. LCBs with inverse orientation appear in the bottom of the center line.

When analyzing with more detail the pattern of plasmids present, we realized that many genes were present in most plasmids. Eleven genes spanning 10 kb are present in these 7 plasmids, while seven genes (8 kb) are shared by 6 plasmids, and 20 genes (14.7 kb) are present in 5 plasmids. Among these genes, we found six transposases, two DNA invertases, one DNA polymerase IV, one transposon resolvase, one NADH peroxidase, one protease, one gene involved in regulation of Cadmium resistance, one Cadmium-transporting ATPase, and 24 hypothetical proteins (Supplementary File 4). After the ANI calculation among the seven

plasmids, we detected a high percentage of similarity (99.51-100%) in different sizes of fragments shared (9743-50589 bp) by every two plasmids (Supplementary Table 8). The alignment of the seven plasmids showed large locally collinear blocks (LCBs) that have lost synteny along the genomes as seen in Figure 6. Two plasmids seem to have undergone inversions.

## 4. Discussion

This study provides important epidemiological information on *Listeria monocytogenes* in Ecuador. We have obtained complete genome sequences of *L. monocytogenes* isolates from various types of cheese from different provinces, and from human clinical cases. Of the total analyzed strains, 93.33% of cheese isolates and 65% of clinical isolates were ST2 (from clonal complex 2, CC2). This sequence type seems to be an important etiological agent of listeriosis in this country because more than half of the clinical cases from 2015 to 2018 from the National Institute of Public Health of Ecuador were caused by ST2 isolates, according to our data. Further, other Ecuadorian *L. monocytogenes* isolated from illegally imported foods into the European Union were also identified as ST2 (3/3 isolates from dairy and meat products) [160]. ST2 strains have been also isolated from dairy, meat and fish products in Albania, Argentina, Georgia, Republic of Moldova, Peru and Ukraine [160], from ready-to-eat fish, meat, sausage and chicken products in other European countries [161], from meat products in South Africa [162], and from food and clinical cases in Chile [163], in USA, and from food in China [164]. Isolates from CC2 have been previously associated with listeriosis cases with few or no immunosuppressive comorbidities [165], and with sporadic but no outbreak-related cases [161]. CC2 strains have also been isolated from a cheese plant as the second most prevalent CC, and the isolates obtained from the plant environment grouped in different clusters than the cheese isolates [161], suggesting a high genetic diversity among environmental isolates.

Apart from ST2, other STs were recovered from food samples: ST6 (lineage I, soft cheese, 2008) and ST796 (lineage II, kneaded cheese, 2018). A ST6 outbreak strain caused 1060 listeriosis cases in 2017-2018 in South Africa, with a case-fatality rate of 27%. The implicated strain was found in polony samples, a ready-to-eat processed meat product, but also in environmental swabs of the producing plant [166]. In Europe, another ST6 strain was implicated in an outbreak with 21 cases (3 deaths and 1 miscarriage) in the Netherlands and Belgium during 2017-2019 for the consumption of ready-to-eat meat products [167]. In Chile, a ST6 was isolated from the environment in a food plant [163], and in France a ST6 was recovered from a

turkey deli meat [164]. On the other, there are no records on ST796 strains at https://bigsdb.pasteur.fr/listeria/.

Strains from ST1, ST3, ST4, ST378 and ST392 were recovered only from clinical cases. ST1 strains have also been isolated from meat products in South Africa [162], from food and clinical samples from Chile [163], from food causing outbreaks in China, USA and Austria [164]. ST3 strains have been recovered from food in Chile [163], from food in France and China, and from clinical samples in China [164]. Strains from food in Italy, sporadic cases in USA, and food in France have been identified as ST4 [164], but also from clinical cases in Ireland [168]. ST392 isolates have been obtained from clinical cases in Chile [163]. This information, with the fact that we only analyzed cheese samples, suggests that the possible source of infection in clinical cases of these STs may be other food products.

The AMR analyses showed phenotypic susceptibility patterns to all antibiotics tested in all the Ecuadorian isolates. With the bioinformatic prediction of AMR, we found some determinants in different databases. With CARD, we found the presence of *fosX*, that is explained because *L. monocytogenes* are naturally resistant to fosfomycin [169], and *mprF*. This latter determinant allows the bacteria to resist CAMPs (cationic antimicrobial peptides), components of the host immune response [170]. When using Megares as reference database, we detected the presence of Gyra_23 and TUFAB-7, which may confer resistance to fluoroquinolones and efamycins, respectively. However, in a previous study, mutations on *gyrA* were obtained with no effect in the transformed *L. monocytogenes* strains in their resistance to nalidixic acid [171]. While working on this article, we did not find any previous publications on *Listeria* and efamycin resistance. We found *norB* and *tetS* in several databases. *L. monocytogenes* strains harboring *norB*, *fosX*, *lin, tetS* and *mprF* have been identified in South Africa [162].

The number of antimicrobial resistance determinants identified varies among the 3 chosen databases. CARD is based on BLAST (Basic Local Alignment Search Tool) against its own database. ResFinder is based on BLAST and KMA (k-mer alignment) against its own database. MEGARes is based on BWA (Burrows-Wheeler Aligner) against a database derived from ARG-ANNOT, CARD, NCBI, Lahey Clinic beta-lactamase archive, and ResFinder [172]. In a study comparing CARD, ARG-ANNOT and Resfinder with different antimicrobial resistance identifiers, the largest number of genes was obtained when using CARD as the reference database [173]. Differences on the outcome of AMR prediction may be due to the default settings for percentage of identity and gene length on the BLAST search of each program, or the

curation of the databases [172]. With the differences obtained depending on the chosen database, there exists a need to standardize the pipelines and the databases.

Among stress factors, we found SSI-1 in 3 isolates, Lm07 (ST796) from lineage II, isolated from food, Lmo10 (ST3) and Lmo15 (ST3) from lineage I, isolated from human cases. This five-gene islet was also present in 11 references from lineage II, isolated from human, animal, food and food processing plants samples, and in one reference from lineage I, isolated from an animal sample. These isolates belong to ST9, 35, 85, 66, 120, 122, 155 and 292. Deletion of the islet has showed decreased growth at suboptimal conditions, such as high salt concentrations and low pH [174,175], and bile and gastric stresses [176]. SSI-1 has also been associated with robustness and adherence of biofilm at 30ºC [177]. The majority of our isolates lacked SSI-1 but contained SSI-F2365 instead, an ORF that is transcribed in the opposite direction in the same location [175], with unknown function [176].

We also looked for SSI-2, two homologues of *lin0464* and *lin0465* present in *L. innocua* that are also transcribed in opposite direction [176]. None of our *L. monocytogenes* isolates contained SSI-2. Since these two genes are usually present in ST121 strains [174] that persist well in food processing plants for months and even years, it is suggested that the presence of SSI-2 in *L. monocytogenes* provides tolerance to alkaline and oxidative stress [176], different to the ones SSI-1 provides. An earlier investigation reported the prevalences of these 3 inserts. SSI-1 (33.3%), SSI-2 (11.8%), SSI-F2365 (54.8%) were evaluated in 476 *L. monocytogenes* isolates from foods, food-processing environments and humans from 37 countries from America, Europe, Asia, and Africa [176].

We also evaluated these 3 insertions in non-*monocytogenes Listeria* isolates. SSI-1 was found in *L. welshimeri* and SSI-2 was present in all five *L. innocua* isolates. All *L. innocua, L. seeligeri, L. welshimeri* isolates lacked SSI-F2365, but it was present in most of our *L. monocytogenes* isolates, suggesting importance to this pathogenic species.

The complete LIPI-1 island was detected in all *L. monocytogenes* references. These six physically linked genes are responsible for crucial steps in the intracellular life cycle of this species [178]. Nevertheless, *actA* gene was absent (or presented premature stop codons) in all Ecuadorian ST2, ST3, ST4 and ST392 isolates, but present in the references from the same STs. The a*ctA* gene is responsible for actin polymerization and motility within the cytoplasm of the host cell, for biofilm formation [179], and for cell-to-cell spread [178]. This gene was also missing from ST2 strains isolated from the food chain in South Africa [162]. Other study that characterized 100 *L. monocytogenes* isolates from environment and food samples in 3 food

plants over a 4-year period found that 33% of the isolates contained a deletion or truncation of *actA*, suggesting that these isolates may present reduced intracellular mobility [52].

The complete LIPI-3 was found in 5 Ecuadorian isolates (food and clinical) and 7 references (animal, food and clinical) with no monophyletic clustering which suggests independent gains in different isolates. LIPI-3 includes an inducible haemolytic and cytotoxic peptide and virulence enhancer, which is known as a listeriolysin S [180]. Strains harboring LIPI-3 belong to ST1, 3, 4, 6 and 66, from lineage I. In South Africa, 21.7% isolates from meat products (from lineage I and II) harbored LIPI-3 [181].

The only isolate containing LIPI-4 was a ST4 strain isolated from blood from a 4-day-old baby. This pathogenicity island in ST4 strains is recognized as the first virulence factor specifically associated with central nervous system and maternal-neonatal listeriosis since the presence of LIPI-4 enhances bacterial tropism to placentas and fetuses of pregnant humanized mice, and also contributes to neuroinvasion [165]. LIPI-4 has been usually associated with ST4 and CC4 strains [168]. Nevertheless, strains from CC2 and CC87 presented this island in China and South Africa [181], and ST1 and ST204 strains also harbored LIPI-4 also in South Africa [162].

In all Ecuadorian *L. monocytogenes* isolates, complete *inlA* and *inlB* genes were detected. Some studies have showed that the presence of premature stop codons in *inlA* are associated with reduced ability to invade human intestinal epithelial cells *in vitro* [182], like in 31% of *L. monocytogenes* isolates from food processing-plants in Ireland [52].

All of our isolates harbored *flaA*, *luxS* and *cheY* genes. The *flaA* gene, coding for flagellin A, has been identified as important for biofilm formation since motility may be required for surface attachment to biotic and abiotic surfaces [183]. The gene *luxS* has been also associated with biofilm formation because it encodes S-ribosylhomocysteine. This enzyme is a precursor of autoinducer-2 in pathogenic bacteria and *luxS* mutants have formed biofilms with changed architecture or did not form any at all [184]. *cheY* is part of a chemotactic system with two functions: it regulates the response to external signals, and it is also responsible of the flagellum rotor regulation. A *cheY* mutant presented a loss on motility activity and the biofilm formation was lower than the one formed by the wild type strain [185]. The fact that our results showed the presence of these 3 genes in all Ecuadorian isolates suggest a high capacity of these *L. monocytogenes* strains to form biofilm.

Among the ST2 isolates, we determined sublineages with wide distribution in time (at least from two different years of isolation) and types of samples (food and clinical isolates). Clades

B and E included only cheese isolates, while clades A, C, D and F grouped isolates from food and clinical samples.

Isolates from clade A showed the greatest diversity on sample origin (blood and cerebrospinal fluid for clinical samples and 4 different types of cheeses). All isolates were collected in 2018, except for a clinical case from 2016. Clade F showed the greatest nucleotide divergence along the core genome of 15 isolates, as seen in Figure 5. When compared these 15 genomes, 565 variables sites were found. Out of those, 430 SNPs were found in the segment 478144-479843 of the core genome. There are two clear subclades. The first group includes the early-branching isolates. They were all isolated from cheeses in 2018, and only 13 variable sites were obtained when analyzing their genomes. The rest of the isolates harbored 124 variable sites, and they were isolated from clinical (2016 and 2018) and cheese samples (2018).

Clade E and F include isolates harboring the same plasmid pattern, suggesting a plasmid gain in their common ancestor and a following plasmid loss in the 4 clinical isolates at the bottom of the tree (Lmo33, Lmo06, Lmo32 and Lmo02). Food isolates from these clades were collected only in 2018, while clinical isolates in 2016 and 2018. There is only one isolate from another clade (B) that presented the same plasmid pattern, Lm08, collected in 2018, suggesting an independent horizontal gene transfer event.

In South Africa, ST2 strains with N1-011A and J1776 have been identified, and the presence of pLM5578 was found only in ST121 and ST321 isolates [162]. We did not detect ST121 and ST321 strains.

Although it would seem that these strains contain several plasmids, the results of the comparative analysis suggest that the plasmids share the same backbone and have great plasticity of gene gain and loss.

Almost the complete NCTC7974_plasmid3 was present in all Ecuadorian isolates (more than 91% of the plasmid length). The length is 593685 nucleotides with 568 CDSs, from which 164 are hypothetical proteins. There is no published information about this plasmid. We also found the presence of this plasmid, with different coverage lengths, in non-*monocytogenes* isolates: 66.58% in the only *L. seeligeri* strain, from 82.98 to 95.07% in 5 *L. innocua* strains and 95.09% in the only *L. welshimeri* isolate (data not shown). After this plasmid annotation with Prokka [114], we found the presence of housekeeping and accessory genes (Supplementary Table 9). Some noteworthy genes to mention are tRNAs, 5S, 16S and 23S ribosomal RNA, 30S and 50S ribosomal proteins, ATP synthase subunits, DNA gyrase, DNA topoisomerase, cell division proteins, chromosomal replication initiator protein and partition protein, phosphoglycerate

kinase, CRISPR with 32 repeat units and Cas1, Cas2, Cas9 and Csn2, but also internalins J. This latter has been associated with virulence in *L. monocytogenes* [186]. In *Rhizobium* species, there has been identified a plasmid with housekeeping genes important for pantothenate synthesis. The authors suggested an intragenomic transference from the rhizobia chromosome to the plasmid [187]. And this might also be the case for this plasmid and *L. monocytogenes.*

The high genetic similarity on ST2 isolates from cheeses and clinical cases suggest the spread of *L. monocytogenes* through the food chain (in this case, cheese consumption) to humans in Ecuador. We also found a few isolates with virulence factors and stress survival islets, which in general means that Ecuadorian isolates show lower pathogenic potential.

In Ecuador, listeriosis is not a notifiable disease and there is not a National Surveillance Program for this foodborne pathogen. Although this study provides valuable genomic and epidemiological information on *L. monocytogenes*, further investigations on other contamination sources apart from cheeses and plant production facilities are still necessary. Genotyping of food, plants and clinical isolates is recommended for routine surveillance in order to infer epidemiological links and the adoption of control measures to prevent listeriosis.

# Chapter 3

## The impact of genetic recombination on pathogenic *Leptospira*

# 1. Background

Evidence of recombination events has been documented in some *Leptospira* species as possible explanation for serotype conversion of Copenhageni to serovar Hardjo [188].

Signatures of natural selection in microbial genomes may contribute to elucidate the crucial mechanisms of pathogenesis [189]. Here, we aimed at identifying recombination events on genes from chromosome I (chromosomes II harbor usually less than 9% of number of genes from chromosomes I [62–65]) whose phylogenetic history is different from the one typically associated within and among species along the core genome of *Leptospira*. First, the analyses were performed at two levels: amid pathogenic *Leptospira* species and intraspecies within *L. interrogans*. Then, another analysis was done among genes involved in *Leptospira* LPS pathway due to the importance of LPS antigens in serodiagnosis and immunity. This study generates evidence of recombination within and between *Leptospira* strains, and on LPS locus increasing genetic and phenotypic diversity with potentially important evolutionary, clinical and epidemiological implications.

# 2. Materials and methods

## 2.1 Data sources, core genome alignment and phylogenetic trees.

The recombination analyses were assessed in chromosomes I of each strain at three levels: interspecies, intraspecies among *interrogans* strains and on LPS genes. Complete chromosome I data from six pathogenic *Leptospira* species was downloaded from public databases (on January 7th, 2019) for the interspecies analysis. Accession numbers and genome data are listed in Table 6.

**Table 6.** Information on closed genomes of *Leptospira* used in recombination analysis among pathogenic species.

| Strain name | RefSeq/INSDC | Chromosome 1 size (Mb) | GC% |
|---|---|---|---|
| *Leptospira santarosai* serovar Shermani str. LT 821 | NZ_CP006694.1 | 3,66 | 41,8 |
| *Leptospira mayottensis* 200901116 | NZ_CP024871.1 | 3,82 | 39,7 |
| *Leptospira kmetyi* strain LS 001/16 | NZ_CP033614.1 | 3,93 | 44,9 |
| *Leptospira interrogans* serovar Bratislava str. PigK151 | NZ_CP011410.1 | 4,37 | 35 |
| *Leptospira borgpetersenii* serovar Ballum strain 56604 | CP012029.1 | 3,55 | 40,2 |
| *Leptospira tipperaryensis* strain GWTS#1 | NZ_CP015217.1 | 4,11 | 42,4 |

Since most closed genomes belonged to *L. interrogans*, we performed a recombination analysis at this species level (Table 7). The third recombination analysis was based on genes from the LPS pathway among ten species (Table 8).

**Table 7.** Information on closed *Leptospira interrogans* genomes used in intraspecies recombination analysis

| Strain name | RefSeq | Chromosome 1 size (Mb) | GC% |
|---|---|---|---|
| *Leptospira interrogans* serovar Copenhageni strain FDAARGOS_203 | NZ_CP020414.2 | 4,28 | 35 |
| *Leptospira interrogans* serovar Manilae strain UP-MMC-NIID LP | NZ_CP011931.1 | 4,24 | 35 |
| *Leptospira interrogans* serovar Bratislava str. PigK151 | NZ_CP011410.1 | 4,37 | 35 |
| *Leptospira interrogans* serovar Linhai str. 56609 | NZ_CP006723.1 | 4,33 | 35 |
| *Leptospira interrogans* serovar Lai str. 56601 | NC_004342.2 | 4,34 | 35 |
| *Leptospira interrogans* serovar Lai str. IPAV | NC_017551.1 | 4,35 | 35 |
| *Leptospira interrogans* serovar Hardjo str. Norma | NZ_CP012603.1 | 4,41 | 35 |

**Table 8.** Information on 29 pathogenic *Leptospira* strains used in LPS genes recombination analysis

| Strain name | RefSeq/GenBank | Chromosome 1 size (Mb) | GC% |
|---|---|---|---|
| *Leptospira alexanderi* serovar Manhao 3 str. L 60 | NZ_AHMT00000000.2 | 4,22 | 40,2 |
| *Leptospira alstonii* serovar Pingchang str. 80-412 | NZ_AOHD02000040.1 | 4,44 | 42,5 |
| *Leptospira alstonii* serovar Sichuan str. 79601 | NZ_ANIK00000000.1 | 4,44 | 42,5 |
| *Leptospira borgpetersenii* serovar Ballum strain 56604 | CP012029.1 | 3,55 | 40,2 |
| *Leptospira borgpetersenii* serovar Ceylonica strain Piyasena | NZ_CP026671.1 | 3,67 | 40,1 |
| *Leptospira borgpetersenii* serovar Hardjo-bovis strain 203 | CP021412.1 | 3,59 | 40,2 |
| *Leptospira borgpetersenii* serovar Hardjo-bovis str. JB197 | CP000350.1 | 3,58 | 40,2 |
| *Leptospira borgpetersenii* serovar Hardjo-bovis strain L49 | CP033440.1 | 3,62 | 40,2 |
| *Leptospira borgpetersenii* serovar Hardjo-bovis str. L550 | CP000348.1 | 3,61 | 40,2 |
| *Leptospira interrogans* serovar Bratislava str. PigK151 | NZ_CP011410.1 | 4,37 | 35 |
| *Leptospira interrogans* serovar Copenhageni strain FDAARGOS_203 | NZ_CP020414.2 | 4,28 | 35 |
| *Leptospira interrogans* serovar Copenhageni str. Fiocruz L1-130 | NC_005823.1 | 4,28 | 35 |
| *Leptospira interrogans* serovar Hardjo str. Norma | NZ_CP012603.1 | 4,41 | 35 |
| *Leptospira interrogans* serovar Lai str. 56601 | NC_004342.2 | 4,34 | 35 |
| *Leptospira interrogans* serovar Lai str. IPAV | NC_017551.1 | 4,35 | 35 |
| *Leptospira interrogans* serovar Linhai str. 56609 | NZ_CP006723.1 | 4,33 | 35 |
| *Leptospira interrogans* serovar Manilae strain UP-MMC-NIID LP | NZ_CP011931.1 | 4,24 | 35 |
| *Leptospira interrogans* serovar Manilae strain UP-MMC-NIID | NZ_CP011934.1 | 4,24 | 35 |
| *Leptospira kirschneri* serovar Grippotyphosa str. Duyster-Boelhouwer | NZ_AHPZ00000000.1 | 4,27 | 35,9 |
| *Leptospira kirschneri* serovar Mozdok strain:61H | NZ_JSVJ00000000.2 | 4,48 | 35,9 |
| *Leptospira kmetyi* strain LS 001/16 | NZ_CP033614.1 | 3,93 | 44,9 |
| *Leptospira mayottensis* 200901116 | NZ_CP024871.1 | 3,82 | 39,7 |
| *Leptospira mayottensis* strain VS2413 | NZ_CP030142.1 | 3,8 | 39,5 |
| *Leptospira noguchii* serogroup Panama strain U73 | NZ_LHQT00000000.1 | 4,76 | 35,7 |
| *Leptospira noguchii* serovar Panama str. CZ214 | NZ_AKWY00000000.2 | 4,71 | 35,5 |
| *Leptospira santarosai* strain U23 | NZ_CP028377.1 | 3,61 | 42 |
| *Leptospira santarosai* serovar Shermani str. LT 821 | NZ_CP006694.1 | 3,66 | 41,8 |
| *Leptospira weilii* str. LNT 1194 | NZ_AHQX00000000.1 | 4,25 | 40,7 |
| *Leptospira weilii* str. UI 14631 | NZ_AHQY00000000.1 | 4,31 | 40,7 |

For the first and second level analyses, fasta files were annotated with Prokka [114], and a determination of orthologs (80% as minimum similarity on protein sequences and 90% as minimum gene coverage) among all species was performed with Proteinortho5 [115]. The core genome was obtained after extracting all genes present in all strains with Proteinortho script: grab_proteins.pl, aligning with Mafft [120], and concatenating with an in-house script.

For the third recombination analysis, all the genes (*n*=38) involved in the lipopolysaccharide biosynthesis pathway from the KEGG Pathway Database (https://www.genome.jp/kegg/pathway.html) were extracted from 29 pathogenic strains (Table 8). The reference genome alignment corresponded to these aligned and concatenated genes.

For all analyses, the reference genome and each gene phylogenies were generated with 1000 bootstrap replicates using IQ-TREE2 [113], with the ModelFinder option [190] to select the most appropriate model of evolution. Using iTOL [121], the metadata was added to the phylogenetic tree.

## 2.2 Recombination detection: Phylogenetic incongruence method (PIM)

The recombination detection was performed by a gene-by-gene topology comparison between each core gene phylogeny versus the reference phylogenetic tree following the PIM procedure [27]. Every core gene was evaluated for its phylogenetic content using IQ-TREE 2 [113] with 10000 as the random number of quartets. The tree topology tests were computed also with IQ-TREE 2 [113]. The reference genome tree was rejected if the expected likelihood weight (c-ELW) and the Kishino-Hasegawa test (*p*-value) were less than 0.05. This means that the differences between topologies the reference genome phylogeny and the gene phylogeny were statistically different. An in-house script was used to classify each recombinant gene according to its COG category. Datamonkey 2.0 [191] was used for characterizing diversifying selection on each recombinant gene with the BUSTED analysis [192] for gene-wide testing.

We also analyzed 20 previously LPS published ORFs [61]. We used BLASTn for each ORF and analyzed sequences that presented anomalies, such as isolates of two different leptospiral species belonging to the same serovar, or sequences belonging to different species showing inconsistent similarities, in order to visually compare tree topologies.

# 3. Results

## 3.1. Recombination among pathogenic *Leptospira* species

For the inter species analysis, we used chromosome I of closed genomes from *L. borgpetersenii, L. interrogans, L. kmetyi, L. mayottensis, L. santarosai* and *L. tipperaryensis*. We performed a comparison of gene-by-gene tree topologies against the core genome phylogeny. The core genome alignment was obtained with 2,095 concatenated core genes that included 1,552,284 bp and 263,593 informative sites (17.6% of the complete alignment). The

maximum likelihood phylogeny built from the core genome is shown in Figure 7A. The accessory genome comprises 4,348 genes, which correspond to 67.48% of pangenome of these six species.

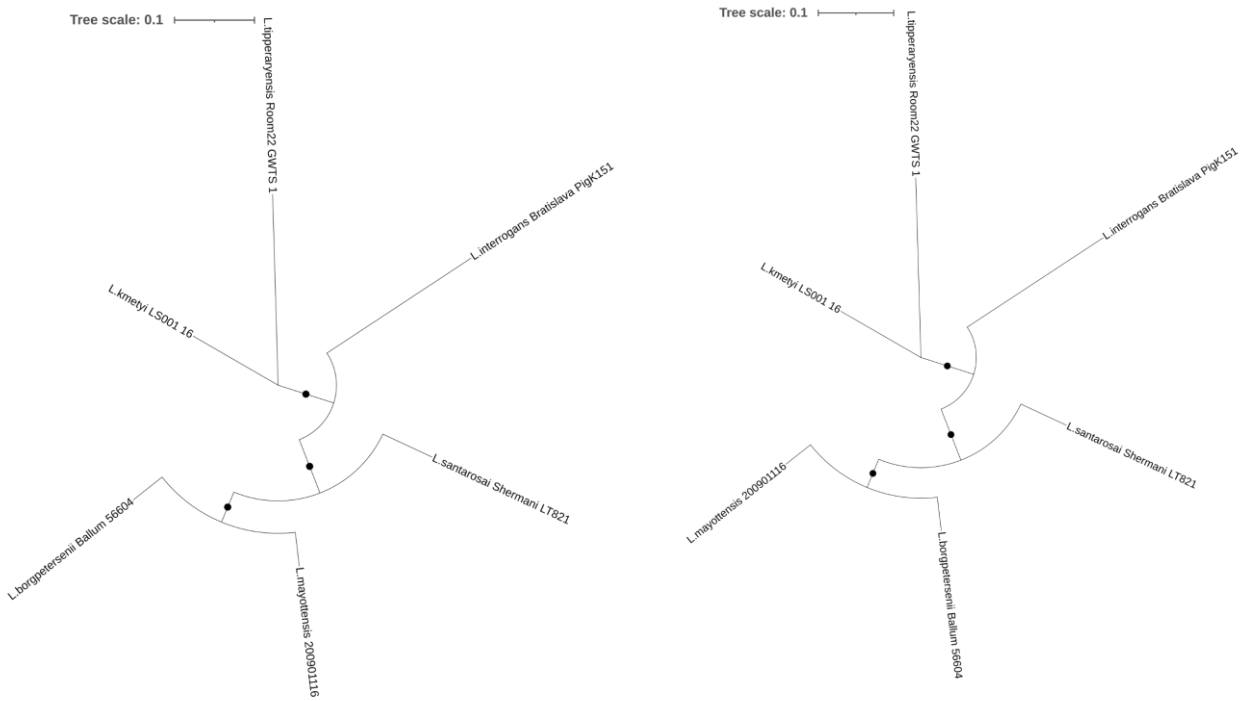



**Figure 7.** A) Maximum likelihood phylogeny of 2095 core genes from *L. borgpetersenii, L. interrogans, L. kmetyi, L. mayottensis, L. santarosai* and *L. tipperaryensis*. B) Maximum likelihood phylogeny of 2061 non-recombinant core genes from the same species as A. Branches with bootstrap values equal to 100 are displayed with a black circle.

In order to evaluate the phylogenetic content of the core genes, we performed a likelihood mapping analysis. We found 1,213 genes with more than 90% of resolved trees, which can be interpreted as genes presenting high phylogenetic signal. The topology tests results suggested that 34 (2.8%) out of those 1,213 genes are recombinant (Supplementary Table 10). Nine of these genes were identified as hypothetical proteins (26.5%). The proportion of the recombinant genes for 10 Cluster of Orthologous Groups (COGs) categories is shown in Table 9. The most represented COG category (25%) was related to translation, ribosomal structure, and biogenesis functions. Each of the following gene categories represented 12% of the recombinant genes: 1) cell wall, membrane, envelope biogenesis, 2) amino acid transport, and metabolism, 3) energy production and metabolism, and 4) carbohydrate transport, metabolism (Figure 8).

**Table 9.** Proportion of inter *Leptospira* species recombinant genes according to COG categories.

| COG categories | Proportion of genes |
|---|---|
| Translation, ribosomal structure and biogenesis | 24,00 |
| Cell wall, membrane, envelope biogenesis | 12,00 |
| Amino acid transport and metabolism | 12,00 |
| Energy production and conversion | 12,00 |
| Carbohydrate transport and metabolism | 12,00 |
| Coenzyme transport and metabolism | 8,00 |
| General function prediction only | 4,00 |
| Inorganic ion transport and metabolism | 4,00 |
| Posttranslational modification, protein turnover, chaperones | 4,00 |
| Secondary metabolites biosynthesis, transport and catabolism | 4,00 |
| Transcription | 4,00 |

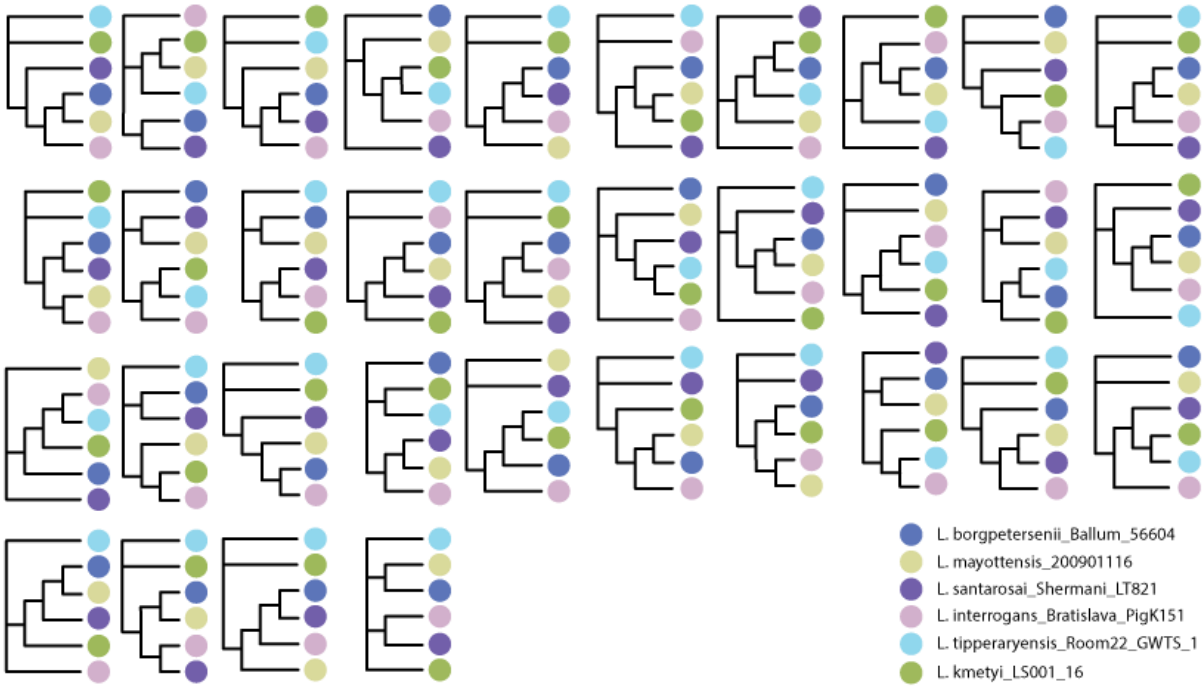



**Figure 8**. Comparison of tree topologies of 34 leptospiral recombinant core genes. The strain names are specified in the legend. The genes are in this order from left to right: UDP-N-acetylglucosamine 1-carboxyvinyltransferase, hypothetical prot., hypothetical prot., Cysteine--tRNA ligase, hypothetical prot., hypothetical prot., Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha, hypothetical prot., Protein QmcA, Capreomycidine synthase, hypothetical prot., Dipeptide permease D, Elongation factor 4, Glyceraldehyde-3-phosphate dehydrogenase 1, hypothetical prot., N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase, Glucose-1-phosphate thymidylyltransferase 1, Putative teichuronic acid biosynthesis glycosyltransferase TuaG, Coenzyme A biosynthesis bifunctional protein CoaBC, NADH-quinone oxidoreductase subunit 1, Alanine--tRNA ligase, 50S ribosomal protein L10, DNA-directed RNA polymerase subunit beta', Threonine--tRNA ligase 2, hypothetical prot., GTPase Der, hypothetical prot., Menaquinone reductase, iron-sulfur cluster-binding subunit,

Outer-membrane lipoprotein carrier protein, Elongation factor G 1, hypothetical prot. Sulfurtransferase, putative succinyl-diaminopimelate desuccinylase, 4,5-DOPA dioxygenase extradiol.

The results of the gene-wide positive selection test with BUSTED showed that at least one site on at least one branch of four genes (capreomycidine synthase, alanine-tRNA ligase, menaquinone reductase, iron-sulfur cluster-binding subunit, and 50S ribosomal protein L10) present episodic diversifying selection (Supplementary Table 10).

A new core genome phylogeny was constructed after removing the recombinant genes from the core alignment. This new recombinant-free alignment included 1,507,524 nt and 255,546 informative sites (17.6% of complete alignment) (Figure 7B). While recombination in these 6 *Leptospira* species was confirmed in 2.8% of genes with high phylogenetic content and the recombinant genes topologies differ from the core genome phylogeny, the resolution of the recombinant-free topology did not improve. The phylogenetic relationships of all the species is consistent in both trees. In Supplementary Table 10, we show the results of the matching split analysis (the number of movements for the topologies, gene versus core genome, to be identical).

No hotspots of recombination were detected along these genomes. The 34 recombinant genes are not located in a particular genomic region, as shown in Figure 9. No relevant pattern is observed in the distribution of genes among the 6 species.

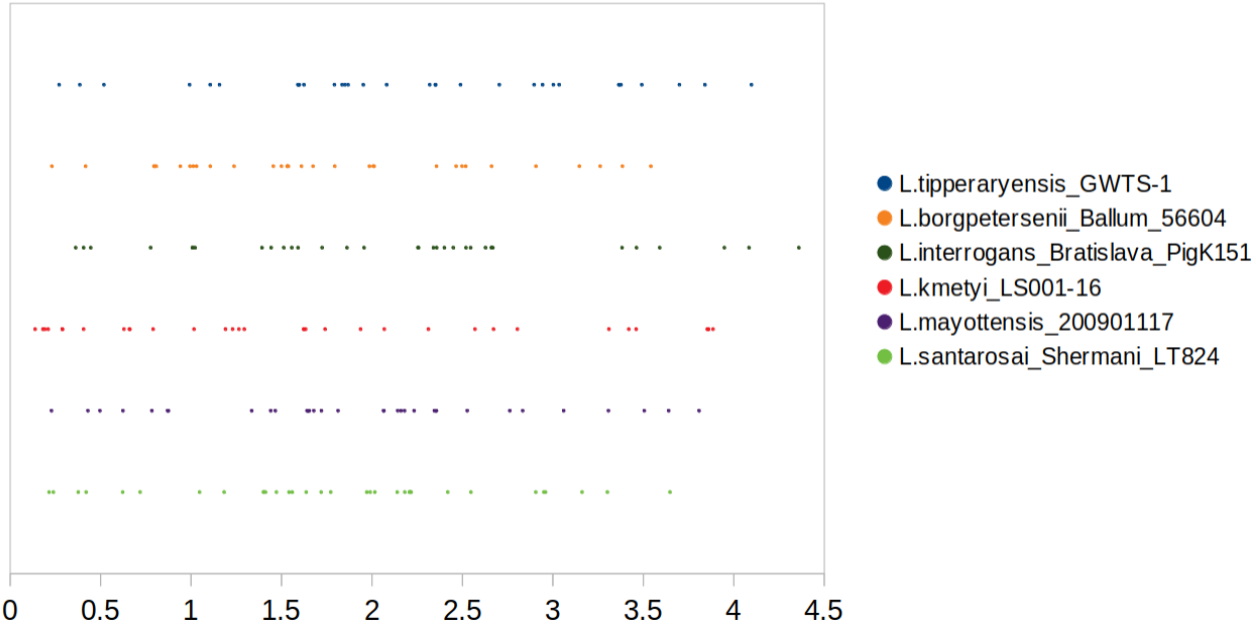



**Figure 9.** Distribution of recombinant genes among *Leptospira* species along their genomes. Displayed dots indicate the beginning of the gene position.

## 3.2. Intraspecies recombination among members of *L. interrogans*

As most of the closed genomes belong to *L. interrogans*, we also wanted to detect recombination among these strains (Table 7). For the alignment we used 1,872 genes shared by all strains, that comprises 941,992 bp with 5,770 informative sites (0.6% of the complete genome) (Figure 10).

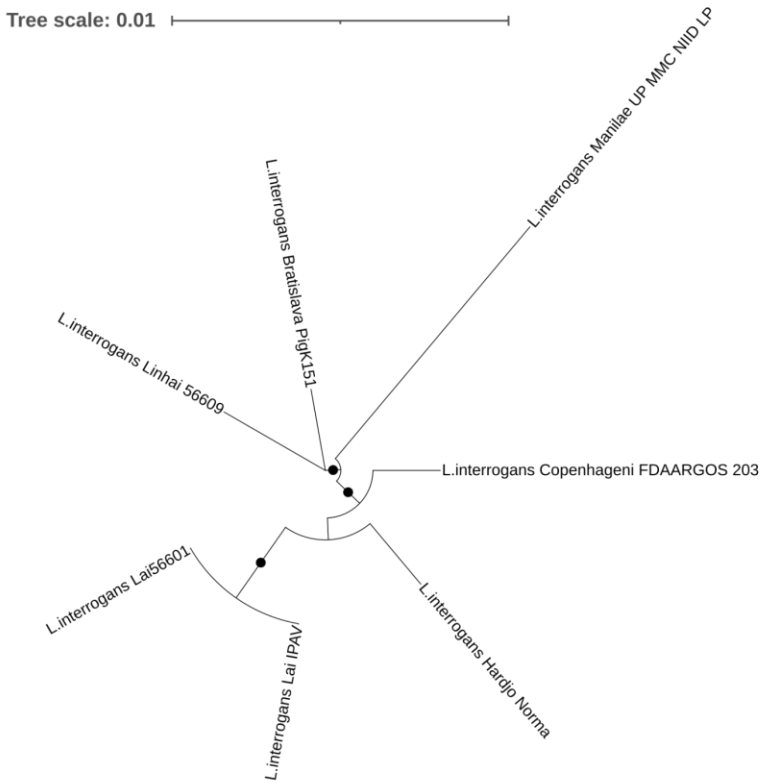



**Figure 10.** Maximum likelihood phylogeny of 1872 core genes from 7 *L. interrogans* strains. Branches with bootstrap support values equal to 100% are shown with a black circle.

After the likelihood mapping analysis, 798 genes presented high phylogenetic signal. The topology tests results suggested that 178 out of those genes are recombinant, which represents a 22% (Supplementary Table 11). From these genes, 58 genes code for proteins with unknown function (32%). The proportion of the recombinant genes for 18 COG categories is shown in Table 10. A 12% of recombinant genes was found to code for proteins involved in cell wall, membrane or envelope biogenesis. All the categories of recombinant genes between species were also present among the categories of recombinant genes within *L. interrogans*. Besides, 5 genes were recombinant in both analysis: acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha, alanine--tRNA ligase, capreomycidine synthase, coenzyme A

biosynthesis bifunctional protein CoaBC, and UDP-N-acetylglucosamine 1-carboxyvinyltransferase.

**Table 10.** Proportion of intra *Leptospira interrogans* recombinant genes according to COG categories.

| COG categories | Proportion of genes |
|---|---|
| Cell wall, membrane, envelope biogenesis | 12,21 |
| Cell cycle control, cell division, chromosome partitioning | 11,45 |
| Translation, ribosomal structure and biogenesis | 11,45 |
| Coenzyme transport and metabolism | 8,40 |
| Amino acid transport and metabolism | 6,87 |
| Energy production and conversion | 6,87 |
| Posttranslational modification, protein turnover, chaperones | 6,87 |
| Inorganic ion transport and metabolism | 5,34 |
| Replication, recombination and repair | 5,34 |
| General function prediction only | 4,58 |
| Cell motility | 3,82 |
| Carbohydrate transport and metabolism | 3,05 |
| Signal transduction mechanisms | 3,05 |
| Transcription | 3,05 |
| Secondary metabolites biosynthesis, transport and catabolism | 2,29 |
| Defense mechanisms | 2,29 |
| Lipid transport and metabolism | 1,53 |
| Nucleotide transport and metabolism | 1,53 |

As with the recombinant genes among species, we did not detect any hotspots of recombination within *L. interrogans* genomes. Figure 11 depicts the recombinant genes distributed along the genomes. The synteny for recombinant genes has been maintained between the two *interrogans* serovar Lai genomes (Figure 11).

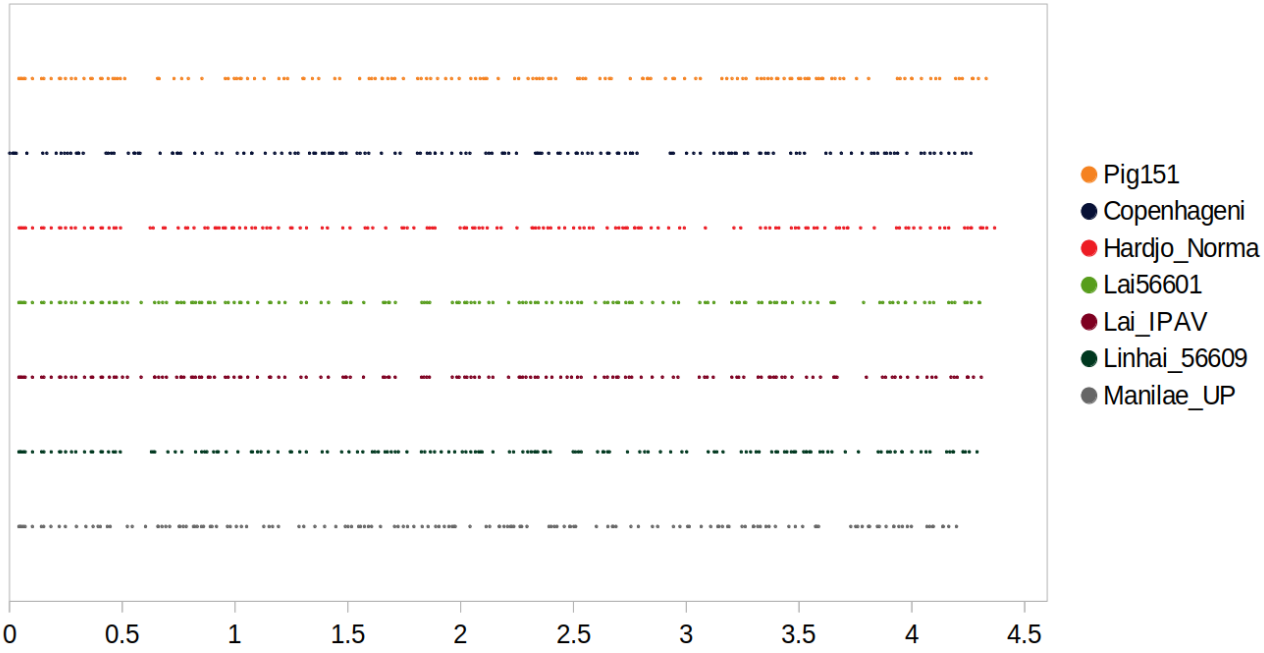



**Figure 11.** Distribution of recombinant genes among *Leptospira interrogans* strains along their genomes. Depicted dots indicate the beginning of the gene position.

## 3.3 Recombination among LPS genes

We decided to study the recombination of lipopolysaccharide genes because they were the first genes showing evidence of recombination 147. Thirty-eight genes related to LPS were concatenated and used as reference. These genes were obtained from 29 strains from 10 species (Table 8). A gene-by-gene comparison was done against this reference. The genome alignment included 50,286 nt with 16,095 informative sites (32.7%) and generated the phylogeny shown in Figure 12.

The likelihood mapping analysis showed that 36 genes presented high phylogenetic signal. Out of these, 18 were recombinant according to the topology tests results (Table 11). The results of the matching split suggest that the following number of movements are needed for the topologies of LPS genes versus reference genome to be the same: 14-20 movements in 3 genes; 21-30 movements in 10 genes; 31-40 movements in 4 genes; and 41-49 movements in 2 genes. Selection analysis showed that only one gene, l*ptC* that codes for lipopolysaccharide export system protein, presented positive selection according to BUSTED approach.

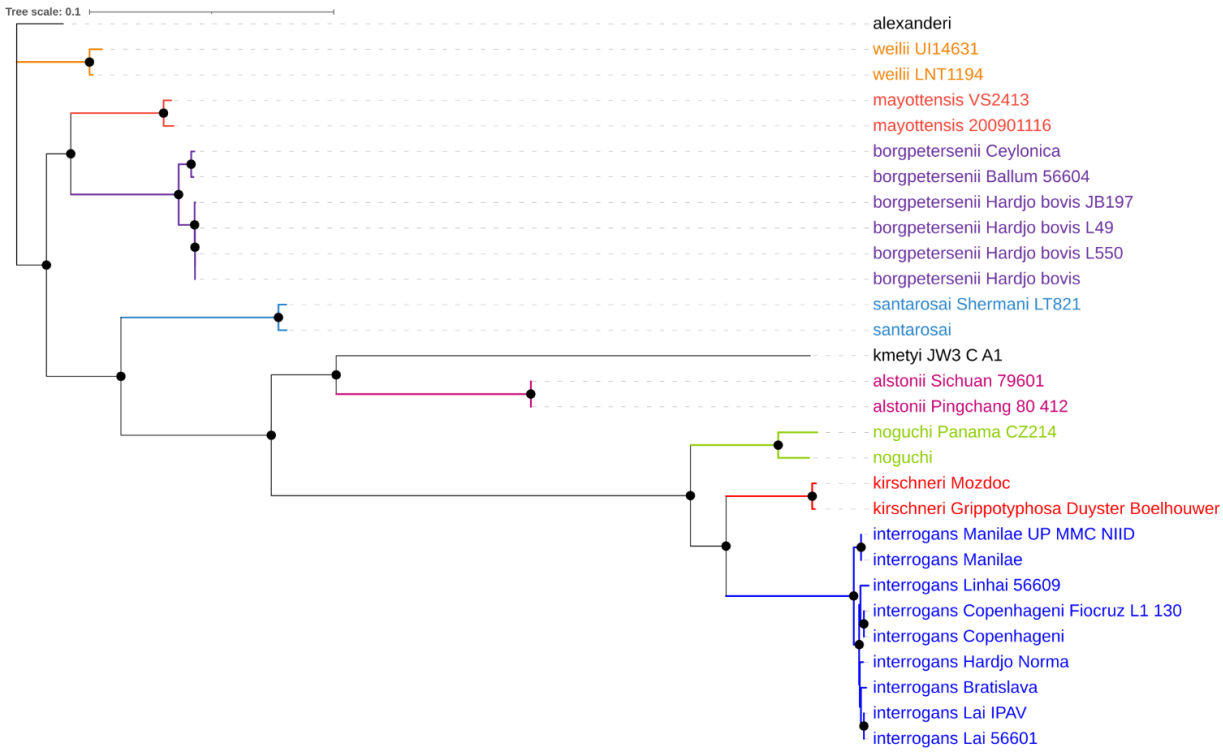



**Figure 12.** Maximum likelihood phylogeny of 38 genes involved in the LPS pathway. Nodes with bootstrap values larger than 90% are shown with a black circle. Clades with more than one isolate are colored.

**Table 11.** List of recombinant genes involved in the LPS pathway of *Leptospira*.

| Protein name | Gene | Matching Split to reference |
|---|---|---|
| Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase | lpxA | 31 |
| dTDP-4-dehydrorhamnose 3,5-epimerase | rfbC_2 | 35 |
| dTDP-glucose 4,6-dehydratase | rfbB_1 | 14 |
| dTDP-glucose 4,6-dehydratase 2 | rffG | 42 |
| Glucose-1-phosphate thymidylyltransferase 1 | rfbA | 49 |
| Lipid-A-disaccharide synthase | lpxB | 21 |
| Lipopolysaccharide assembly protein B x3 | lapB_2 | 24-28 |
| Lipopolysaccharide core heptosyltransferase RfaQ | rfaQ | 17 |
| Lipopolysaccharide export system ATP-binding protein LptB | lptB | 24 |
| Lipopolysaccharide export system permease protein LptG | lptG | 23 |
| **Lipopolysaccharide export system protein LptC** [1] | lptC | 35 |
| Mannose-1-phosphate guanylyltransferase RfbM | rfbM | 22 |
| Methyl-accepting chemotaxis protein McpA | mcpA_3 | 28 |
| Outer membrane protein assembly factor BamD | bamD_1 | 30 |
| Tetraacyldisaccharide 4'-kinase | lpxK | 23 |
| Undecaprenyl phosphate-alpha-4-amino-4-deoxy-L-arabinose arabinosyl transferase | arnT | 23 |

Additionally, ten out of 20 LPS ORFs from leptospiral isolates showing anomalies in BLAST (two leptospiral isolates belonging to two different species but sharing the same serovar

antigens or showing higher identity with other species rather than with members of the same species), showed evidence of topology incongruence (Supplementary Figures 3-12).

## 4. Discussion

We have found clear evidence of recombination in *Leptospira* genomes at different levels, within and among species of the same genus, as well as in LPS loci. These observations indicate that recombination may be an essential feature on this genus, still ongoing and not restricted to specific periods in its evolutionary past.

For recombination to occur, different leptospiral strains must be in close contact. No evidence of co-infections in humans or domestic animals has been identified to date. Nevertheless, the presence of more than one *Leptospira* species has been found in Mengyin County bats in China [193], Egyptian slit-faced bats and Egyptian fruit bats in Swaziland and in South Africa, and in small terrestrial mammals in Tanzania [194]. There must be a special attention in places with high *Leptospira* diversity, like Ecuador, where 6 species were found circulating among cows, pigs, rats and humans simultaneously in the same communities [195], or Malaysia, where 7 species were identified in rats, soil and water from an urban area [196]. These two facts suggest a temporal and ecological basis for recombination between species or strains of *Leptospira* to take place.

The identification of recombinant genes was based on incongruent topology when compared to the core genome phylogenetic tree, revealing a different evolutionary history for these genes, and the likelihood of having been horizontally transferred. Among the gene categories which were more associated with recombination among chromosome I from different leptospiral species and among leptospiral chromosome I belonging to the same species, some genes were swapped both inter-species and intraspecies: acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha, alanine--tRNA ligase, capreomycidine synthase, coenzyme A biosynthesis bifunctional protein CoaBC, UDP-N-acetylglucosamine 1-carboxyvinyltransferase. As expected, recombination was more frequent among members of the same species [17,197]. This phenomenon occurs because members of the same species are usually in closer contact due to shared ecological niches, and have high DNA similarity [22], which facilitates legitimate recombination.

Genomic regions with high levels of recombination have been shown evidence of positive selection in other microorganisms [21,23,27], and genes affected by natural selection are precisely the ones that are critical for microbial success, and may include functions such as host invasion, evasion of immune responses, replication within the hosts, etc. [22,26].

Informational genes involved in translation, ribosomal structure and biogenesis belong to the category with most evidence of recombination among members of leptospiral species. The interspecies recombination analysis indicated that only 4% of the recombinant genes are associated with transcription, posttranslational modification, protein turnover and chaperones; while for the intra-*interrogans* analysis, only 3% for transcription, and 6.87% for posttranslational modification, protein turnover and chaperones. This finding is intriguing because informational genes are usually less likely to be horizontally transferred [197] probably because the products of these genes must interact with other molecules encoded in the genome of the recipient, and are part of a complex systems (the complexity hypothesis) [24]. Accordingly, Riedl's burden hypothesis states that the donor and the recipient should be closely related for the informational genes to be successfully transferred [198]. Recombination of informational genes may cause unoptimized rates of growth, which seems paradoxical; however, other bacteria (such as *Mycobacterium tuberculosis*) have genes that reduce the rate of growth, protecting the host from a highly replicative pathogen causing a fatal disease [199]. For *M. tuberculosis* it seems more effective to keep a relatively healthy chronically infected host, which disseminates the bacterium and thus allows to infect larger number of new hosts [199]. Similarly, *Leptospira* colonizes the distal tubes of the kidney of animal reservoirs, causing an asymptomatic chronic infection [54]. A restricted growth rate may prevent pathological alterations of the kidney.

Leptospiral LPS is an important antigen for vaccine development, as protective immunity is based primarily on anti-LPS antibodies [59,200,201], and leptospiral LPS is also responsible for serovar identification [58]. Recombination in LPS genes was first described by de la Peña Moctezuma and colleagues, where the LPS biosynthetic loci (*rfb*) of two subtypes (Hardjoprajitno and Hardjobovis from two different species) showed such high similarity that resulted in two serologically indistinguishable serovar Hardjo subtypes [188]. Also, it has been observed that antibody pressure selects LPS mutants with different LPS antigenic reactivity [202,203] and may also select LPS-gene recombinants.

In this work we show evidence that the recombination of LPS genes is widespread in the different leptospiral genomes. Analysis of 29 strains, belonging to 10 species, showed that more than half of LPS genes were recombinant suggesting that there must be a selective pressure promoting variation on LPS. In addition, bioinformatic analysis indicated that genes coding for LPS component are under selective pressure (Table 11). We obtained additional evidence of recombination in 12 LPS genes of the LPS locus that showed serological inconsistencies in BLASTn analysis (Supplementary Figures 3-12). Recombination in LPS loci has been shown

to be beneficial for other bacteria preventing immune recognition and bacteriophage attachment [18,20–23,204,205]. Leptospiral LPS recombination has also implications in diagnosis because the gold standard for *Leptospira* detection is microscopic agglutination test (MAT), and it is also based on LPS recognition. Besides, recombination of LPS genes is probably the most important reason why serological classification does not correspond to genetic classification [188].

Similar selective pressures may be also associated with the high levels of recombination of outer components of the cell wall. In the interspecies analysis, some noteworthy genes are the *murA* (UDP-N-acetylglucosamine 1-carboxyvinyltransferase), *wbbL* (N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase), *tuaG* (Putative teichuronic acid biosynthesis glycosyltransferase), which are important for cell wall biogenesis; and *qmcA_1* (Protein QmcA), *dtpD* (Dipeptide permease D), *lolA_1* (Outer-membrane lipoprotein carrier protein), part of the outer membrane activities. Recombination of these genes may be beneficial because the outer membrane is the main communication structure between the bacteria and the environment/mammalian hosts [206], and also because leptospires are known to escape NOD1 and NOD2 sensing during cell wall biosynthesis [207]. We also observed evidence of recombination in the *LipL32* gene, another outer membrane component (Figure 8 and Supplementary Table 10).

Recombination also provides genomic diversity, which increases the possibility of discrimination between isolates. We observed that phylogenetic analysis of different alleles used in leptospiral MLST (multilocus sequence typing) schemes also showed evidence of recombination (data not shown).

Recombination requires that donor and recipient DNAs share the same environment and one potential environment could be fresh water or mud contaminated with urine from infected animals [67]. However, some species of *Leptospira* cannot survive in the environment such as *L. borgpetersenii* serovar Hardjo [208] and, interestingly, this was the first report or recombination reported (recombination between *L. interrogans* serovar Hardjo and *L. borgapetersenii* serovar Hardjo) [188]. We posit that this finding may suggest that recombination may occur in the kidney or urine of animals co-infected with different leptospiral species (both *L. interrogans* serovar Hardjo and *L. borgpetersenii* serovar Hardjo are frequently found in cattle) [188].

The approach used in our study detects only effective recombination events among the strains included in the analyses. The genes to be included in the analysis had to meet two requirements: present high phylogenetic content and their phylogenetic comparison with the reference genome had to have significant statistical incongruence in the topology tests. There are

programs that use different strategies for the recombination detection, like Gubbins [209], or ClonalFrameML [210]. In a previous work on *Treponema,* these 3 approaches were compared. The results showed that 57 recombinant genes were detected by Gubbins, while 44 were reported by ClonalFrameML, but only 10 of them fulfilled the criteria to be considered as recombinant. Most of the other genes did not present phylogenetic incongruence [27]. A high SNP concentration within a DNA fragment could be explained by the fact that spontaneous mutations accumulate due to diversifying selection [22], and not necessarily to recombination. One of the main advantages of our method of analysis is that recombinant detection was performed for each gene separately.

No effective recombination events were found in genes associated with RNA processing and modification, chromatin structure and dynamics, intracellular trafficking, secretion, and vesicular transport, extracellular structures, mobilome: prophages, transposons, or cytoskeleton. For most of these genes, there must be non-effective recombination events that occur when the transfer of DNA replaces an identical fragment in the recipient cell or is rapidly purged by purifying selection. According to Didelot and Maiden, the main role of homologous recombination is DNA repair [17]. It is also important to mention that the recombination analysis was performed on the core genome, so genes that make up the accessory genome were not included. Thus, it is important that the accessory genome is included in future analyses of recombination because these genes are usually involved in pathogenesis, virulence and antimicrobial resistance.

A better understanding of the impact of recombination in *Leptospira* not only enhance our knowledge on its evolution, but also provides evidence of microbial pathogenesis, possible impact on serovar identification and vaccine development.

# Discussion

Genomic epidemiology proved its usefulness in typing bacterial strains to the highest resolution, as we were able to detect the transmission of two bacterial pathogens, *S. enterica* and *L. monocytogenes*, through the food chain to humans in Ecuador.

In the first chapter of this thesis, we used genomic epidemiology to characterize *Salmonella enterica* isolated from three sources. *S. enterica* is considered one of the most important foodborne pathogens worldwide due to their high genetic diversity and increasing multidrug resistance that limits the available treatments when this pathogen causes systemic diseases. The implementation of a global surveillance program is essential for the adequate control of this preventable FBD, especially in developing countries, like Ecuador, where more efforts and resources need to be addressed on this matter.

When comparing the *in silico* resistance determinants with the phenotypic resistance of *Salmonella* isolates, we found different levels of correlation among the antibiotic families. This implies a limitation when databases do not include point mutations or new genes responsible for antimicrobial resistance. It is important to remember that genomic data found on databases come mainly from isolates from developed countries, which implies an important bias. For resistance breakpoints, for example, different values can be found for CLSI and EUCAST (European Committee for Antimicrobial Susceptibility Testing). One isolate becomes resistant or susceptible depending on the resistance breakpoints used.

*In silico* prediction of plasmid determinants did not work for our *Salmonella* and *Listeria* isolates. The fact that the *Salmonella* strains presented multi-resistance profiles and the apparent absence of plasmids prompted us to carry out a more exhaustive analysis in search of the reason for the high resistance to antimicrobials. When a mapping approach was performed after a literature revision on multidrug resistant *Salmonella*, the presence of a megaplasmid was suggested in our isolates. With the experience from the first chapter, a similar analysis was performed for *Listeria* despite the antimicrobial susceptibility. This is clearly another limitation if the analyses are based solely in the *in silico* predictions since the presence of some plasmids will not be identified. Nevertheless, in several publications, a good result has been found in the active search for plasmid determinants, in particular with species of clinical importance such as *E. coli*, *K. pneumoniae*, or *P. aeruginosa*.

Another obstacle that can be found in genomic analyses is the use of mapping to a reference versus *de novo* assembly when drafting a genome. At a computational level, mapping will always be faster and less expensive than assembly, but all the genetic information present in

the isolates but absent in the reference is lost. An approach that performs both analyses is recommended due to the type of information that we can obtain, especially important in the SNP calling.

Thanks to whole genome sequencing and subsequent bioinformatic analyses, we were able to work with the core and accessory genomes of *S.* Infantis. This allowed us to reveal the importance of this foodborne pathogen in Ecuador, where our analyses have shown its high clonality and presence of virulence factors as well as multidrug resistance patterns. Therefore, we recommend less and better antimicrobials use in poultry setting in this country.

For Chapter 2, we decided to work with complete genomes from another foodborne pathogen with little epidemiological information in Ecuador. In this study, isolates from cheeses and clinical samples were included in order to determine the relatedness among *L. monocytogenes* from food and human samples. Some important conclusions were derived from this research. Even though data on the prevalence or death toll from *L. monocytogenes* are unknown from Ecuador, after the performed analyses we concluded that this pathogen is causing disease in this country.

For the resistance determinants, three databases were compared and the results varied dramatically. The reasons for the number of determinants found may be related to the default parameters of percentage of identity or gene length, as well as the level of curation of the databases. With the increasing available data on a daily basis, it may be difficult to keep databases accurately updated. However, this limitation must be managed, either by making comparisons and correlations readily available or by pushing for databases to be kept up to date. Perhaps one option is to generate information control centers for different pathogens in different institutions. The fact that a single institution may have to work with massive information on several pathogens does not allow the curation and selection of data to be possible as fast as it is now required.

The *L. monocytogenes* isolates included in our study showed largely susceptible profiles to antimicrobial but also the presence of several virulence factors, with the potential of causing severe disease. From the analyses of clinical isolates, we also determined the presence of some strains not found in cheeses, suggesting that other types of food products are responsible for these clinical cases. This finding opened a new research project with other Ecuadorian institutions. We are currently working on the detection and isolation of *L. monocytogenes* in ready-to-eat foods from street vendors in the capital of Ecuador. This research is complemented

with training workshops to achieve better food practices and greater control of this pathogen in our country. There is an urgent need to include *L. monocytogenes* in a National Surveillance Program in clinical and food establishments.

In Chapter 3, we focused in the etiological agent for a neglected disease that is the most common zoonosis worldwide, leptospirosis. We found evidence of recombination in *Leptospira* at different levels suggesting that it is a major process contributing to the evolution and adaptation of *Leptospira* to new hosts, including the human species.

Recombination in leptospiral LPS genes was also evident. For this genus, LPS is important as antigen for vaccine target and diagnosis. The finding that more than half of the LPS genes are recombinant suggests a selective pressure for variation on this locus. In other pathogens, recombination on LPS has been confirmed as an advantage for avoiding immune recognition and phage infection.

We chose the PIM (phylogenetic incongruence method) approach for this research, despite the availability of bioinformatic programs that take less time and computational power to identify possible recombination events. As we presented earlier, our approach proved to be more precise to determine recombinant genes since our analysis includes several steps that discarded genes that did not meet the requirements of presenting high phylogenetic content and statistical significance in the phylogenetic inconsistency with the reference tree.

It is really interesting that no co-infections on humans or domestic animals have been found so far. When a more detailed literature review was carried out, we found a few publications showing the presence of more than one *Leptospira* serovar or species in bats and small mammals. With this in mind, we applied for a new funding call to investigate the presence of different species or isolates of *Leptospira* in animals that are normally considered reservoirs of this disease. We started working with dogs in poor urban areas of a province of Ecuador with a high incidence of leptospirosis. The aim is to follow these animals for several months to identify if there is a turnover of strains in persistent infections that may explain the ecological basis for recombination to occur.

Increasing the genomic information of isolates from different parts of the world is essential because microorganisms can be subjected to different types of stress or selective pressures in different habitats. For example, multiple studies have confirmed only one species of *Leptospira*

circulating in soils and waters in some regions and countries, but in other parts of the world, there is evidence of multiple species and serovars co-circulating in natural habitats.

In this work, we have demonstrated the great advantage of using WGS for epidemiological studies. The information we obtain from this type of technology has enormous potential to better understand the epidemiology, transmission and evolution of pathogens. We have also established the importance of open access databases. Chapter 3 focused on the analysis of closed complete genomes available in databases and it was possible to demonstrate a fundamental evolutionary process in a pathogen that causes the most common zoonosis worldwide.

We have also exposed the need to implement genomic epidemiology in national surveillance programs to help control infectious diseases. Particularly in this thesis, we have worked with isolates of two foodborne pathogens from a developing country, which without the financial support of a developed country would probably have taken much more time and human and financial resources, or maybe it would not have been possible at all. Thanks to the results of these investigations, new projects have begun to have more local information and help decision makers control these infections in Ecuador.

# Conclusions

1. Genomic epidemiology allows an unprecedented subtyping resolution through the analysis of the core and the accessory genome in order to assess the genomic diversity, resistance determinant factors, virulence genotypes, stress resistance and pathogenic potential.

2. *Salmonella enterica* isolates from Ecuador studied in this work showed multidrug resistance pointing out the importance of a better use of antimicrobials in intensive poultry farm settings in this country.

3. *Salmonella enterica* serovar Infantis isolated from farms, food and human samples presented high genomic similarity suggesting the dissemination of this foodborne pathogen through the food chain to humans.

4. The *in silico* detection of determinants presented limitations because no plasmids were found despite the multidrug resistance patterns on our *Salmonella* isolates.

5. Given the phylogenetic relatedness of *Listeria monocytogenes* from cheeses with samples of clinical origins, genotyping of food and clinical isolates is recommended for routine surveillance in order to infer epidemiological links and the adoption of control measures that prevent an increase in listeriosis cases.

6. The high genetic similarity on ST2 strains from cheeses and clinical cases suggests a high probability of epidemiological association and spread of *L. monocytogenes* through the food chain, specifically the consumption of cheese, to humans in Ecuador.

7. In Ecuador, listeriosis is not a notifiable disease and there is no national surveillance program for this foodborne pathogen. Although this study provides valuable genomic and epidemiological information on *Listeria monocytogenes* in Ecuador, further investigations of other contamination sources apart from cheeses are still necessary.

8. Evidence of recombination in *Leptospira* suggests a selective pressure promoting variation on different loci in this pathogen with possible impacts on serovar identification and vaccine development.

9. Genomic analysis provides information about recombinant genes that are under selection, the genes that make *Leptospira* a successful pathogen. Therefore, this type of research may also inform about new ways to prevent and treat leptospirosis.

# References

1. Madigan, M., Bender, K., Buckley, D., Sattley, M. & Stahl, D. *BROCK Biology of Microorganisms*. (Pearson Education, Inc., 2022).
2. Sharma, A. K. *et al.* Bacterial Virulence Factors: Secreted for Survival. *Indian J. Microbiol.* **57**, 1–10 (2017).
3. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond.)* **27**, 113–159 (1928).
4. Lund, B. M. & O'Brien, S. J. The Occurrence and Prevention of Foodborne Disease in Vulnerable People. *Foodborne Pathog. Dis.* **8**, 961–973 (2011).
5. Evans, E. W. & Redmond, E. C. Behavioral Risk Factors Associated with Listeriosis in the Home: A Review of Consumer Food Safety Studies. *J. Food Prot.* **77**, 510–521 (2014).
6. WHO. *Global health estimates: Leading causes of death*. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death (2022).
7. Taylor, L. H., Latham, S. M. & woolhouse, M. E. J. Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 983–989 (2001).
8. Vouga, M. & Greub, G. Emerging bacterial pathogens: the past and beyond. *Clin. Microbiol. Infect.* **22**, 12–21 (2016).
9. Maglangit, F., Yu, Y. & Deng, H. Bacterial pathogens: threat or treat (a review on bioactive natural products from bacterial pathogens). *Nat. Prod. Rep.* **38**, 782–821 (2021).
10. Banin, E., Hughes, D. & Kuipers, O. P. Editorial: Bacterial pathogens, antibiotics and antibiotic resistance. *FEMS Microbiol. Rev.* **41**, 450–452 (2017).
11. Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat. Rev. Genet.* (2019) doi:10.1038/s41576-019-0108-4.
12. C Reygaert, W. & Department of Biomedical Sciences, Oakland University William Beaumont School of Medicine, Rochester, MI, USA. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol.* **4**, 482–501 (2018).

13. Yang, Q. E. & Walsh, T. R. Toxin–antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. *FEMS Microbiol. Rev.* **41**, 343–353 (2017).
14. Darmon, E. & Leach, D. R. F. Bacterial Genome Instability. *Microbiol. Mol. Biol. Rev.* **78**, 1–39 (2014).
15. Watford, S. & Warrington, S. J. Bacterial DNA Mutations. in *StatPearls* (StatPearls Publishing, 2022).
16. Burmeister, A. R. Horizontal Gene Transfer. *Evol. Med. Public Health* **2015**, 193–194 (2015).
17. Didelot, X. & Maiden, M. C. J. Impact of recombination on bacterial evolution. *Trends Microbiol.* **18**, 315–322 (2010).
18. Blokesch, M. The Lifestyle of Vibrio cholerae Fosters Gene Transfers: Growing on chitinous surfaces helps these bacteria to initiate horizontal gene transfer and, perhaps, to swap pathogenic traits. *Microbe Mag.* **9**, 64–70 (2014).
19. Coscolla, M., Comas, I. & Gonzalez-Candelas, F. Quantifying Nonvertical Inheritance in the Evolution of Legionella pneumophila. *Mol. Biol. Evol.* **28**, 985–1001 (2011).
20. Davies, M. R., Broadbent, S. E., Harris, S. R., Thomson, N. R. & van der Woude, M. W. Horizontally Acquired Glycosyltransferase Operons Drive Salmonellae Lipopolysaccharide Diversity. *PLoS Genet.* **9**, e1003568 (2013).
21. Dettman, J. R., Rodrigue, N. & Kassen, R. Genome-Wide Patterns of Recombination in the Opportunistic Human Pathogen Pseudomonas aeruginosa. *Genome Biol. Evol.* **7**, 18–34 (2015).
22. Hester, S. E. *et al.* Horizontally acquired divergent O-antigen contributes to escape from cross-immunity in the classical bordetellae. *BMC Evol. Biol.* **13**, 209 (2013).
23. Milkman, R., Jaeger, E. & McBride, R. D. Molecular evolution of the Escherichia coli chromosome. VI. Two regions of high effective recombination. *Genetics* **163**, 475–483 (2003).
24. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci.* **96**, 3801–3806 (1999).

25. Pollett, S. *et al.* A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *Sci. Rep.* **11**, 17365 (2021).
26. Didelot, X., Achtman, M., Parkhill, J., Thomson, N. R. & Falush, D. A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: Convergence or divergence by homologous recombination? *Genome Res.* **17**, 61–68 (2006).
27. Pla-Díaz, M. *et al.* Evolutionary Processes in the Emergence and Recent Spread of the Syphilis Agent, *Treponema pallidum*. *Mol. Biol. Evol.* msab318 (2021) doi:10.1093/molbev/msab318.
28. NIH. Food Poisining. (No date).
29. WHO. *WHO ESTIMATES OF THE GLOBAL BURDEN OF FOODBORNE DISEASES*. 1–15 https://apps.who.int/iris/bitstream/handle/10665/200046/WHO_FOS_15.02_eng.pdf?sequence=1 (2015).
30. CIFOR. GUIDELINES FOR FOODBORNE DISEASE OUTBREAK RESPONSE. (2022).
31. Rahman, Md. T. *et al.* Zoonotic Diseases: Etiology, Impact, and Control. *Microorganisms* **8**, 1405 (2020).
32. Strawn, L. K. *et al.* Risk Factors Associated with Salmonella and Listeria monocytogenes Contamination of Produce Fields. *Appl. Environ. Microbiol.* **79**, 7618–7627 (2013).
33. Carrasco, E., Morales-Rueda, A. & García-Gimeno, R. M. Cross-contamination and recontamination by Salmonella in foods: A review. *Food Res. Int.* **45**, 545–556 (2012).
34. WHO. Salmonella (non-typhoidal). https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal) (2018).
35. Antunes, P., Mourão, J., Campos, J. & Peixe, L. Salmonellosis: the role of poultry meat. *Clin. Microbiol. Infect.* **22**, 110–121 (2016).
36. EFSA Panel on Biological Hazards (EFSA BIOHAZ Panel) *et al.* Salmonella control in poultry flocks and its public health impact. *EFSA J.* **17**, (2019).
37. Corporación Nacional de Avicultores del Ecuador. Estadísticas del Sector Avícola. (2020).

38. McDermott, P. F., Zhao, S. & Tate, H. Antimicrobial Resistance in Nontyphoidal *Salmonella*. in *Antimicrobial Resistance in Bacteria from Livestock and Companion Animals* (eds. Schwarz, S., Cavaco, L. M. & Shen, J.) 261–287 (ASM Press, 2018). doi:10.1128/9781555819804.ch12.
39. Ibarra, J. A. & Steele-Mortimer, O. *Salmonella* - the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. *Cell. Microbiol.* **11**, 1579–1586 (2009).
40. Seeliger, H. P. R. Listeriosis — History and actual developments. *Infection* **16**, S80–S84 (1988).
41. Jeffers, G. T. *et al.* Comparative genetic characterization of Listeria monocytogenes isolates from human and animal listeriosis cases. *Microbiology* **147**, 1095–1104 (2001).
42. Murray, E. G. D., Webb, R. A. & Swann, M. B. R. A disease of rabbits characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillusBacterium monocytogenes (n.sp.). *J. Pathol. Bacteriol.* **29**, 407–439 (1926).
43. Bergholz, T. M., Shah, M. K., Burall, L. S., Rakic-Martinez, M. & Datta, A. R. Genomic and phenotypic diversity of Listeria monocytogenes clonal complexes associated with human listeriosis. *Appl. Microbiol. Biotechnol.* **102**, 3475–3485 (2018).
44. Jordan, K. & McAuliffe, O. Listeria monocytogenes in Foods. in *Advances in Food and Nutrition Research* vol. 86 181–213 (Elsevier, 2018).
45. Walker, S. J., Archer, P. & Banks, J. G. Growth of *Listeria monocytogenes* at refrigeration temperatures. *J. Appl. Bacteriol.* **68**, 157–162 (1990).
46. Melo, J., Andrew, P. W. & Faleiro, M. L. Listeria monocytogenes in cheese and the dairy environment remains a food safety challenge: The role of stress responses. *Food Res. Int.* **67**, 75–90 (2015).
47. Gould, L. H., Mungai, E. & Barton Behravesh, C. Outbreaks Attributed to Cheese: Differences Between Outbreaks Caused by Unpasteurized and Pasteurized Dairy Products, United States, 1998–2011. *Foodborne Pathog. Dis.* **11**, 545–551 (2014).

48. Ferreira, V., Wiedmann, M., Teixeira, P. & Stasiewicz, M. J. Listeria monocytogenes Persistence in Food-Associated Environments: Epidemiology, Strain Characteristics, and Implications for Public Health. *J. Food Prot.* **77**, 150–170 (2014).
49. Jordan, K., Hunt, K., Lourenco, A. & Pennone, V. Listeria monocytogenes in the Food Processing Environment. *Curr. Clin. Microbiol. Rep.* **5**, 106–119 (2018).
50. Bolocan, A. S. *et al.* Dynamics of Listeria monocytogenes colonisation in a newly-opened meat processing facility. *Meat Sci.* **113**, 26–34 (2016).
51. Orsi, R. H., Bakker, H. C. den & Wiedmann, M. Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics. *Int. J. Med. Microbiol.* **301**, 79–96 (2011).
52. Hurley, D. *et al.* Whole-Genome Sequencing-Based Characterization of 100 Listeria monocytogenes Isolates Collected from Food Processing Environments over a Four-Year Period. *mSphere* **4**, (2019).
53. WHO. Zoonoses. (2020).
54. Ko, A. I., Goarant, C. & Picardeau, M. Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nat. Rev. Microbiol.* **7**, 736–747 (2009).
55. Mohammed, H., Nozha, C., Hakim, K., Abdelaziz, F. & Rekia, B. Leptospirosis: Epidemiology and Usuel Manifestations. *Bacteriol. J.* **1**, 1–7 (2010).
56. Vincent, A. T. *et al.* Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. *PLoS Negl. Trop. Dis.* **13**, e0007270 (2019).
57. Cerqueira, G. M. & Picardeau, M. A century of Leptospira strain typing. *Infect. Genet. Evol.* **9**, 760–768 (2009).
58. Cinco, M., Banfi, E. & Panfili, E. Leptospiral lipopolysaccharide presence in the outer envelope: Electrophoretic evidence and immunological specificity. *Zentralblatt Für Bakteriol. Mikrobiol. Hyg. Ser. Med. Microbiol. Infect. Dis. Virol. Parasitol.* **269**, 277–283 (1988).
59. Adler, B. Vaccines Against Leptospirosis. in *Leptospira and Leptospirosis* (ed. Adler, B.) vol. 387 251–272 (Springer Berlin Heidelberg, 2015).

60. Lauretti-Ferreira, F. *et al.* New strategies for Leptospira vaccine development based on LPS removal. *PLOS ONE* **15**, e0230460 (2020).
61. Adler, B. & de la Peña Moctezuma, A. Leptospira and leptospirosis. *Vet. Microbiol.* **140**, 287–296 (2010).
62. Llanes, A., Restrepo, C. M. & Rajeev, S. Whole Genome Sequencing Allows Better Understanding of the Evolutionary History of Leptospira interrogans Serovar Hardjo. *PLOS ONE* **11**, e0159387 (2016).
63. Moreno, L. Z. *et al.* Genomic characterization and comparative analysis of Leptospira interrogans serogroup Australis isolated from swine. *Pathog. Dis.* **75**, (2017).
64. Nascimento, A. L. T. O. *et al.* Comparative Genomics of Two *Leptospira interrogans* Serovars Reveals Novel Insights into Physiology and Pathogenesis. *J. Bacteriol.* **186**, 2164–2172 (2004).
65. Picardeau, M. *et al.* Genome Sequence of the Saprophyte Leptospira biflexa Provides Insights into the Evolution of Leptospira and the Pathogenesis of Leptospirosis. *PLoS ONE* **3**, e1607 (2008).
66. Boey, K., Shiokawa, K. & Rajeev, S. Leptospira infection in rats: A literature review of global prevalence and distribution. *PLoS Negl. Trop. Dis.* **13**, e0007499 (2019).
67. Schmalzle, S. A. *et al.* Recreational 'mud fever': Leptospira interrogans induced diffuse alveolar hemorrhage and severe acute respiratory distress syndrome in a U.S. Navy seaman following 'mud-run' in Hawaii. *IDCases* **15**, e00529 (2019).
68. Goris, M. G. A. & Hartskeerl, R. A. Leptospirosis Serodiagnosis by the Microscopic Agglutination Test. *Curr. Protoc. Microbiol.* **32**, (2014).
69. Gilmour, M. W., Graham, M., Reimer, A. & Van Domselaar, G. Public Health Genomics and the New Molecular Epidemiology of Bacterial Pathogens. *Public Health Genomics* **16**, 25–30 (2013).
70. Moura, A. *et al.* Real-Time Whole-Genome Sequencing for Surveillance of *Listeria monocytogenes* , France. *Emerg. Infect. Dis.* **23**, 1462–1470 (2017).
71. Fleischmann, R. D. *et al.* Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).

72. Levy, S. E. & Myers, R. M. Advancements in Next-Generation Sequencing. *Annu. Rev. Genomics Hum. Genet.* **17**, 95–115 (2016).
73. Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* **19**, 9–20 (2018).
74. Rabinowicz, P. D. Genomics in Latin America: Reaching the Frontiers. *Genome Res.* **11**, 319–322 (2001).
75. Snowden, F. M. Emerging and reemerging diseases: a historical perspective. *Immunol. Rev.* **225**, 9–26 (2008).
76. Deng, X., den Bakker, H. C. & Hendriksen, R. S. Genomic Epidemiology: Whole-Genome-Sequencing–Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. *Annu. Rev. Food Sci. Technol.* **7**, 353–374 (2016).
77. MacDonald, P. *Methods in Field Epidemiology*. (Jones & Bartlett Learning, 2012).
78. Mejía, L. *et al.* Genomic Epidemiology of Salmonella Infantis in Ecuador: From Poultry Farms to Human Infections. *Front. Vet. Sci.* **7**, 547891 (2020).
79. Gutierrez, B. *et al. Genomic epidemiology of SARS-CoV-2 transmission lineages in Ecuador*. http://medrxiv.org/lookup/doi/10.1101/2021.03.31.21254685 (2021) doi:10.1101/2021.03.31.21254685.
80. Garzon-Chavez, D. *et al.* Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. *Sci. Rep.* **10**, 6237 (2020).
81. Donado-Godoy, P. *et al.* Prevalence, Resistance Patterns, and Risk Factors for Antimicrobial Resistance in Bacteria from Retail Chicken Meat in Colombia. *J. Food Prot.* **78**, 751–759 (2015).
82. Valderrama, W., Pastor, J., Mantilla Salazar & Ortiz, M. Estudio de prevalencia de serotipos de salmonella en granjas avícolas tecnificadas en el perú. (2014).
83. Vinueza-Burgos, C., Baquero, M., Medina, J. & De Zutter, L. Occurrence, genotypes and antimicrobial susceptibility of Salmonella collected from the broiler production chain within an integrated poultry company. *Int. J. Food Microbiol.* **299**, 1–7 (2019).

84. Vinueza-Burgos, C., Cevallos, M., Ron-Garrido, L., Bertrand, S. & De Zutter, L. Prevalence and Diversity of Salmonella Serotypes in Ecuadorian Broilers at Slaughter Age. *PLOS ONE* **11**, e0159567 (2016).
85. Voss-Rech, D. *et al.* A temporal study of Salmonella enterica serotypes from broiler farms in Brazil. *Poult. Sci.* **94**, 433–441 (2015).
86. Brown, A. C. *et al.* CTX-M-65 Extended-Spectrum β-Lactamase–Producing *Salmonella enterica* Serotype Infantis, United States1. *Emerg. Infect. Dis.* **24**, 2284–2291 (2018).
87. Cartelle Gestal, M., Zurita, J., Paz y Mino, A., Ortega-Paredes, D. & Alcocer, I. Characterization of a small outbreak of Salmonella enterica serovar Infantis that harbour CTX-M-65 in Ecuador. *Braz. J. Infect. Dis.* **20**, 406–407 (2016).
88. Tate, H. *et al.* Comparative Analysis of Extended-Spectrum-β-Lactamase CTX-M-65-Producing Salmonella enterica Serovar Infantis Isolates from Humans, Food Animals, and Retail Chickens in the United States. *Antimicrob. Agents Chemother.* **61**, (2017).
89. World Health Organization. *Integrated surveillance of antimicrobial resistance in foodborne bacteria: application of a one health approach: guidance from the WHO Advisory Group on Integrated Surveillanec of Antimicrobial Resistance (AGISAR)*. (World Health Organization, 2017).
90. Akiba, M., Kusumoto, M. & Iwata, T. Rapid identification of Salmonella enterica serovars, Typhimurium, Choleraesuis, Infantis, Hadar, Enteritidis, Dublin and Gallinarum, by multiplex PCR. *J. Microbiol. Methods* **85**, 9–15 (2011).
91. CLSI. PERFORMANCE STANDARDS FOR ANTIMICROBIAL SUSCEPTIBILITY TESTING. (2020).
92. Hasman, H., Mevius, D., Veldman, K., Olesen, I. & Aarestrup, F. M. β-Lactamases among extended-spectrum β-lactamase (ESBL)-resistant Salmonella from poultry, poultry products and human patients in The Netherlands. *J. Antimicrob. Chemother.* **56**, 115–121 (2005).
93. Olesen, I., Hasman, H. & Møller Aarestrup, F. Prevalence of β-Lactamases among Ampicillin-Resistant *Escherichia coli* and *Salmonella* Isolated from Food Animals in Denmark. *Microb. Drug Resist.* **10**, 334–340 (2004).

94. Kruger, T. *et al.* Infections with Nontyphoidal *Salmonella* Species Producing TEM-63 or a Novel TEM Enzyme, TEM-131, in South Africa. *Antimicrob. Agents Chemother.* **48**, 4263–4270 (2004).

95. Arlet, G. Substitution of alanine for aspartate at position 179 in the SHV-6 extended-spectrum β-lactamase. *FEMS Microbiol. Lett.* **152**, 163–167 (1997).

96. Carattoli, A. *et al.* Molecular Epidemiology of *Escherichia coli* Producing Extended-Spectrum β-Lactamases Isolated in Rome, Italy. *J. Clin. Microbiol.* **46**, 103–108 (2008).

97. Jiang, X. *et al.* Detection of Extended-Spectrum β-Lactamases in Clinical Isolates of *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **50**, 2990–2995 (2006).

98. Hopkins, K. L., Batchelor, M. J., Liebana, E., Deheer-Graham, A. P. & Threlfall, E. J. Characterisation of CTX-M and AmpC genes in human isolates of Escherichia coli identified between 1995 and 2003 in England and Wales. *Int. J. Antimicrob. Agents* **28**, 180–192 (2006).

99. Paauw, A., Fluit, A. C., Verhoef, J. & Leverstein-van Hall, M. A. *Enterobacter cloacae* Outbreak and Emergence of Quinolone Resistance Gene in Dutch Hospital. *Emerg. Infect. Dis.* **12**, 807–812 (2006).

100. Dierikx, C. M. *et al.* Occurrence and characteristics of extended-spectrum-β-lactamase- and AmpC-producing clinical isolates derived from companion animals and horses. *J. Antimicrob. Chemother.* **67**, 1368–1374 (2012).

101. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).

102. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

103. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).

104. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

105. Zhang, S. *et al.* Salmonella Serotype Determination Utilizing High-Throughput Genome Sequencing Data. *J. Clin. Microbiol.* **53**, 1685–1692 (2015).

106. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genomics* **3**, (2017).

107. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 124 (2018).

108. Carattoli, A. *et al. In Silico* Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).

109. Vallejos-Sánchez, K. *et al.* Whole-Genome Sequencing of a Salmonella enterica subsp. enterica Serovar Infantis Strain Isolated from Broiler Chicken in Peru. *Microbiol. Resour. Announc.* **8**, (2019).

110. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

111. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

112. H. Li. *Seqtk: Toolkit for processing sequences in FASTA/q formats.* (2018).

113. Minh, B. Q. *et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era*. http://biorxiv.org/lookup/doi/10.1101/849372 (2019) doi:10.1101/849372.

114. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

115. Lechner, M. *et al.* Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).

116. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).

117. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **5**, e11147 (2010).

118. Yoon, S.-H., Ha, S., Lim, J., Kwon, S. & Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, 1281–1286 (2017).

119. Nurk, S. *et al.* Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. in *Research in Computational Molecular Biology* (eds. Deng, M., Jiang, R., Sun, F. & Zhang, X.) vol. 7821 158–170 (Springer Berlin Heidelberg, 2013).

120. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

121. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

122. Hendriksen, R. S. *et al.* Global Monitoring of *Salmonella* Serovar Distribution from the World Health Organization Global Foodborne Infections Network Country Data Bank: Results of Quality Assured Laboratories from 2001 to 2007. *Foodborne Pathog. Dis.* **8**, 887–900 (2011).

123. European Food Safety Authority & European Centre for Disease Prevention and Control. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012. *EFSA J.* **12**, (2014).

124. Gymoese, P. *et al.* WGS based study of the population structure of Salmonella enterica serovar Infantis. *BMC Genomics* **20**, 870 (2019).

125. Donado-Godoy, P. *et al.* Prevalence, Risk Factors, and Antimicrobial Resistance Profiles of Salmonella from Commercial Broiler Farms in Two Important Poultry-Producing Regions of Colombia. *J. Food Prot.* **75**, 874–883 (2012).

126. Vinueza-Burgos, C. & Villagómez Estrada. Aislamiento y serotipificación de salmonella enteritidis, typhimurium, e infantis en carcasas de pollo destinadas para consumo humano en un camal industrializado de la Provincia de Pichincha. (2015).

127. Vasco, G. *et al.* Identifying Etiological Agents Causing Diarrhea in Low Income Ecuadorian Communities. *Am. J. Trop. Med. Hyg.* **91**, 563–569 (2014).

128. Naranjo, A., Cedeño, C., Teran, E., Castello, A., & and the CASERO Research Team. Prevalence of VP4 and VP7 genotypes of human rotavirus in Ecuadorian children with acute diarrhea. *J. Med. Virol.* **80**, 1106–1111 (2008).

129. Guderian, R. H., Ordóñez, G. & Bossano, R. [Acute diarrhea associated with a Campylobacter and other pathogenic agents in Quito, Ecuador]. *Boletin Oficina Sanit. Panam. Pan Am. Sanit. Bur.* **102**, 333–339 (1987).

130. Bhavnani, D., Goldstick, J. E., Cevallos, W., Trueba, G. & Eisenberg, J. N. S. Synergistic Effects Between Rotavirus and Coinfecting Pathogens on Diarrheal Disease: Evidence from a Community-based Study in Northwestern Ecuador. *Am. J. Epidemiol.* **176**, 387–395 (2012).

131. Gunn, J. S. *et al.* Salmonella chronic carriage: epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol.* **22**, 648–655 (2014).

132. European Food Safety Authority and European Centre for Disease Prevention and Control (EFSA and ECDC). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. *EFSA J.* **16**, (2018).

133. Boscán-Duque, L. A. *et al.* Reduced Susceptibility to Quinolones among Salmonella Serotypes Isolated from Poultry at Slaughter in Venezuela†. *J. Food Prot.* **70**, 2030–2035 (2007).

134. Medeiros, M. A. N., Oliveira, D. C. N. de, Rodrigues, D. dos P. & Freitas, D. R. C. de. Prevalence and antimicrobial resistance of Salmonella in chicken carcasses at retail in 15 Brazilian cities. *Rev. Panam. Salud Pública* **30**, 555–560 (2011).

135. Lapierre, L. *et al.* Phenotypic and Genotypic Characterization of Virulence Factors and Susceptibility to Antibiotics in Salmonella Infantis Strains Isolated from Chicken Meat: First Findings in Chile. *Animals* **10**, 1049 (2020).

136. Foley, S. L. *et al.* Population Dynamics of Salmonella enterica Serotypes in Commercial Egg and Poultry Production. *Appl. Environ. Microbiol.* **77**, 4273–4279 (2011).

137. Cunha-Neto, A. da *et al.* Salmonella isolated from chicken carcasses from a slaughterhouse in the state of Mato Grosso, Brazil: antibiotic resistance profile,

serotyping, and characterization by repetitive sequence-based PCR system. *Poult. Sci.* **97**, 1373–1381 (2018).

138. Quesada, A., Reginatto, G. A., Ruiz Español, A., Colantonio, L. D. & Burrone, M. S. Resistencia antimicrobiana de Salmonella spp aislada de alimentos de origen animal para consumo humano. *Rev. Peru. Med. Exp. Salud Pública* **33**, 32 (2016).

139. Rodríguez, R., Fandiño, C., Donado, P., Guzmán, L. & Verjan, N. Characterization of *Salmonella* from Commercial Egg-Laying Hen Farms in a Central Region of Colombia. *Avian Dis.* **59**, 57–63 (2015).

140. European Food Safety Authority & European Centre for Disease Prevention and Control. The European Union Summary Report on Antimicrobial Resistance in zoonotic and indicator bacteria from humans, animals and food in 2017/2018. *EFSA J.* **18**, (2020).

141. Vinueza-Burgos, C. Salmonella and Campylobacter in Broilers at Slaughter Age : A Possible Source for Carcasses Contamination in Ecuador. (2017).

142. Van Boeckel, T. P. *et al.* Global trends in antimicrobial use in food animals. *Proc. Natl. Acad. Sci.* **112**, 5649–5654 (2015).

143. Hindermann, D. *et al.* Salmonella enterica serovar Infantis from Food and Human Infections, Switzerland, 2010–2015: Poultry-Related Multidrug Resistant Clones and an Emerging ESBL Producing Clonal Lineage. *Front. Microbiol.* **8**, 1322 (2017).

144. Bogomazova, A. N. *et al.* Mega-plasmid found worldwide confers multiple antimicrobial resistance in Salmonella Infantis of broiler origin in Russia. *Int. J. Food Microbiol.* **319**, 108497 (2020).

145. Wajid, M., Saleemi, M. K., Sarwar, Y. & Ali, A. Detection and characterization of multidrug-resistant *Salmonella enterica* serovar Infantis as an emerging threat in poultry farms of Faisalabad, Pakistan. *J. Appl. Microbiol.* **127**, 248–261 (2019).

146. Vuthy, Y., Lay, K. S., Seiha, H., Kerleguer, A. & Aidara-Kane, A. Antibiotic susceptibility and molecular characterization of resistance genes among Escherichia coli and among Salmonella subsp. in chicken food chains. *Asian Pac. J. Trop. Biomed.* **7**, 670–674 (2017).

147. Franco, A. *et al.* Emergence of a Clonal Lineage of Multidrug-Resistant ESBL-Producing Salmonella Infantis Transmitted from Broilers and Broiler Meat to Humans in Italy between 2011 and 2014. *PLOS ONE* **10**, e0144802 (2015).

148. Aviv, G. *et al.* A unique megaplasmid contributes to stress tolerance and pathogenicity of an emergent *S almonella enterica* serovar Infantis strain: The contribution of a megaplasmid to the emergence of *S* . Infantis. *Environ. Microbiol.* **16**, 977–994 (2014).

149. Cohen, E., Rahav, G. & Gal-Mor, O. Genome Sequence of an Emerging Salmonella enterica Serovar Infantis and Genomic Comparison with Other S. Infantis Strains. *Genome Biol. Evol.* **12**, 223–228 (2020).

150. Ledeboer, N. A., Frye, J. G., McClelland, M. & Jones, B. D. *Salmonella enterica* Serovar Typhimurium Requires the Lpf, Pef, and Tafi Fimbriae for Biofilm Formation on HEp-2 Tissue Culture Cells and Chicken Intestinal Epithelium. *Infect. Immun.* **74**, 3156–3169 (2006).

151. Kingsley, R. A., Weening, E. H., Keestra, A. M. & Bäumler, A. J. Population Heterogeneity of *Salmonella enterica* Serotype Typhimurium Resulting from Phase Variation of the *lpf* Operon In Vitro and In Vivo. *J. Bacteriol.* **184**, 2352–2359 (2002).

152. Gupta, S. K. *et al.* Genomic comparison of diverse Salmonella serovars isolated from swine. *PLOS ONE* **14**, e0224518 (2019).

153. Boddicker, J. D., Ledeboer, N. A., Jagnow, J., Jones, B. D. & Clegg, S. Differential binding to and biofilm formation on, HEp-2 cells by Salmonella enterica Serovar Typhimurium is dependent upon allelic variation in the fimH gene of the fim gene cluster: Allelic variation in the Salmonella fimH gene. *Mol. Microbiol.* **45**, 1255–1265 (2002).

154. Espinosa-Mata, E., Mejía, L., Villacís, J. E., Alban, V. & Zapata, S. Detection and genotyping of Listeria monocytogenes in artisanal soft cheeses from Ecuador. *Rev. Argent. Microbiol.* S0325754121000493 (2021) doi:10.1016/j.ram.2021.02.013.

155. Tao, T., Chen, Q., Bie, X., Lu, F. & Lu, Z. Investigation on prevalence of Listeria spp. and Listeria monocytogenes in animal-derived foods by multiplex PCR assay targeting novel genes. *Food Control* **73**, 704–711 (2017).

156. Moura, A. *et al.* Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. *Nat. Microbiol.* **2**, 16185 (2017).

157. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

158. Doster, E. *et al.* MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* **48**, D561–D569 (2020).

159. Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* gkz935 (2019) doi:10.1093/nar/gkz935.

160. Rychli, K. *et al.* Listeria monocytogenes Isolated from Illegally Imported Food Products into the European Union Harbor Different Virulence Factor Variants. *Genes* **9**, 428 (2018).

161. Nielsen, E. M. *et al.* Closing gaps for performing a risk assessment on Listeria monocytogenes in ready-to-eat (RTE) foods: activity 3, the comparison of isolates from different compartments along the food chain, and from humans using whole genome sequencing (WGS) analysis. *EFSA Support. Publ.* **14**, (2017).

162. Mafuna, T., Matle, I., Magwedere, K., Pierneef, R. E. & Reva, O. N. Whole Genome-Based Characterization of Listeria monocytogenes Isolates Recovered From the Food Chain in South Africa. *Front. Microbiol.* **12**, 669287 (2021).

163. Toledo, V. *et al.* Genomic Diversity of Listeria monocytogenes Isolated from Clinical and Non-Clinical Samples in Chile. *Genes* **9**, 396 (2018).

164. Yin, Y. *et al.* Geographical and longitudinal analysis of Listeria monocytogenes genetic diversity reveals its correlation with virulence and unique evolution. *Microbiol. Res.* **175**, 84–92 (2015).

165. Maury, M. M. *et al.* Uncovering Listeria monocytogenes hypervirulence by harnessing its biodiversity. *Nat. Genet.* **48**, 308–313 (2016).

166. Smith, A. M. *et al.* Outbreak of *Listeria monocytogenes* in South Africa, 2017–2018: Laboratory Activities and Experiences Associated with Whole-Genome Sequencing Analysis of Isolates. *Foodborne Pathog. Dis.* **16**, 524–530 (2019).

167. ECDC & EFSA. Multi-country outbreak of Listeria monocytogenes sequence type 6 infections linked to ready-to-eatmeat products – 25 November 2019. *EFSA Support. Publ.* **16**, (2019).

168. Hilliard, A. *et al.* Genomic Characterization of Listeria monocytogenes Isolates Associated with Clinical Listeriosis and the Food Production Environment in Ireland. *Genes* **9**, 171 (2018).

169. Luque-Sastre, L. *et al.* Antimicrobial Resistance in *Listeria* Species. *Microbiol. Spectr.* **6**, (2018).

170. Thedieck, K. *et al.* The MprF protein is required for lysinylation of phospholipids in listerial membranes and confers resistance to cationic antimicrobial peptides (CAMPs) on Listeria monocytogenes. *Mol. Microbiol.* **62**, 1325–1339 (2006).

171. Lampidis, R. Molecular characterization of the genes encoding DNA gyrase and topoisomerase IV of Listeria monocytogenes. *J. Antimicrob. Chemother.* **49**, 917–924 (2002).

172. Hendriksen, R. S. *et al.* Using Genomics to Track Global Antimicrobial Resistance. *Front. Public Health* **7**, 242 (2019).

173. Doyle, R. M. *et al.* Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. *Microb. Genomics* **6**, (2020).

174. Hein, I. *et al.* Stress Survival Islet 1 (SSI-1) Survey in *Listeria monocytogenes* Reveals an Insert Common to *Listeria innocua* in Sequence Type 121 *L. monocytogenes* Strains. *Appl. Environ. Microbiol.* **77**, 2169–2173 (2011).

175. Ryan, S., Begley, M., Hill, C. & Gahan, C. G. M. A five-gene stress survival islet (SSI-1) that contributes to the growth of Listeria monocytogenes in suboptimal

conditions: Stress survival islet in L. monocytogenes. *J. Appl. Microbiol.* **109**, 984–995 (2010).

176. Harter, E. *et al.* Stress Survival Islet 2, Predominantly Present in Listeria monocytogenes Strains of Sequence Type 121, Is Involved in the Alkaline and Oxidative Stress Responses. *Appl. Environ. Microbiol.* **83**, (2017).

177. Keeney, K., Trmcic, A., Zhu, Z., Delaquis, P. & Wang, S. Stress survival islet 1 contributes to serotype-specific differences in biofilm formation in *Listeria monocytogenes*. *Lett. Appl. Microbiol.* **67**, 530–536 (2018).

178. Vázquez-Boland, J. A., Domínguez-Bernal, G., González-Zorn, B., Kreft, J. & Goebel, W. Pathogenicity islands and virulence evolution in Listeria. *Microbes Infect.* **3**, 571–584 (2001).

179. Travier, L. *et al.* ActA Promotes Listeria monocytogenes Aggregation, Intestinal Colonization and Carriage. *PLoS Pathog.* **9**, e1003131 (2013).

180. Cotter, P. D. *et al.* Listeriolysin S, a Novel Peptide Haemolysin Associated with a Subset of Lineage I Listeria monocytogenes. *PLoS Pathog.* **4**, e1000144 (2008).

181. Matle, I. *et al.* Population Structure of Non-ST6 Listeria monocytogenes Isolated in the Red Meat and Poultry Value Chain in South Africa. *Microorganisms* **8**, 1152 (2020).

182. Su, X. *et al.* Characterization of internalin genes in Listeria monocytogenes from food and humans, and their association with the invasion of Caco-2 cells. *Gut Pathog.* **11**, 30 (2019).

183. Warke, S. R. *et al.* Biofilm Formation and Associated Genes in Listeria Monocytogenes. *INDIAN J. Vet. Sci. Biotechnol.* **12**, (2017).

184. Vendeville, A., Winzer, K., Heurlier, K., Tang, C. M. & Hardie, K. R. Making 'sense' of metabolism: autoinducer-2, LUXS and pathogenic bacteria. *Nat. Rev. Microbiol.* **3**, 383–396 (2005).

185. Fan, Y. *et al.* Influence of different factors on biofilm formation of Listeria monocytogenes and the regulation of cheY gene. *Food Res. Int.* **137**, 109405 (2020).

186. Sabet, C., Lecuit, M., Cabanes, D., Cossart, P. & Bierne, H. LPXTG Protein InlJ, a Newly Identified Internalin Involved in *Listeria monocytogenes* Virulence. *Infect. Immun.* **73**, 6912–6922 (2005).

187. Villaseñor, T. *et al.* Housekeeping genes essential for pantothenate biosynthesis are plasmid-encoded in Rhizobium etli and Rhizobium leguminosarum. *BMC Microbiol.* **11**, 66 (2011).

188. de la Peña-Moctezuma, A., Bulach, D. M., Kalambaheti, T. & Adler, B. Comparative analysis of the LPS biosynthetic loci of the genetic subtypes of serovar Hardjo: Leptospira interrogans subtype Hardjoprajitno and Leptospira borgpetersenii subtype Hardjobovis. *FEMS Microbiol. Lett.* **177**, 319–326 (1999).

189. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* **15**, 379–393 (2014).

190. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

191. Weaver, S. *et al.* Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **35**, 773–777 (2018).

192. Murrell, B. *et al.* Gene-Wide Identification of Episodic Selection. *Mol. Biol. Evol.* **32**, 1365–1371 (2015).

193. Han, H.-J. *et al.* Pathogenic *Leptospira* Species in Insectivorous Bats, China, 2015. *Emerg. Infect. Dis.* **24**, 1123–1126 (2018).

194. Dietrich, M. *et al.* Biogeography of *Leptospira* in wild animal communities inhabiting the insular ecosystem of the western Indian Ocean islands and neighboring Africa. *Emerg. Microbes Infect.* **7**, 1–12 (2018).

195. Barragan, V. *et al.* High Leptospira Diversity in Animals and Humans Complicates the Search for Common Reservoirs of Human Disease in Rural Ecuador. *PLoS Negl. Trop. Dis.* **10**, e0004990 (2016).

196. Pui, C. F., Bilung, L. M., Apun, K. & Su'ut, L. Diversity of *Leptospira* spp. in Rats and Environment from Urban Areas of Sarawak, Malaysia. *J. Trop. Med.* **2017**, 1–8 (2017).

197. Wiedenbeck, J. & Cohan, F. M. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **35**, 957–976 (2011).

198. Riedl, R. *Order in living organisms: a systems analysis of evolution*. (Wiley, 1978).

199. Shimono, N. *et al.* Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mce1 operon. *Proc. Natl. Acad. Sci.* **100**, 15918–15923 (2003).

200. Jost, B. H., Adler, B. & Faine, S. Experimental immunisation of hamsters with lipopolysaccharide antigens of Leptospira interrogans. *J. Med. Microbiol.* **29**, 115–120 (1989).

201. Patra, K. P. *et al.* Comparative analysis of lipopolysaccharides of pathogenic and intermediately pathogenic Leptospira species. *BMC Microbiol.* **15**, 244 (2015).

202. Zapata, S. *et al.* Characterization of a lipopolysaccharide mutant of Leptospira derived by growth in the presence of an anti-lipopolysaccharide monoclonal antibody: Leptospira lipopolysaccharide. *FEMS Microbiol. Lett.* no-no (2010) doi:10.1111/j.1574-6968.2010.02029.x.

203. Zuerner, R. L. & Trueba, G. A. Characterization of IS 1501 mutants of Leptospira interrogans serovar pomona. *FEMS Microbiol. Lett.* **248**, 199–205 (2005).

204. Dettman, J. R., Rodrigue, N., Aaron, S. D. & Kassen, R. Evolutionary genomics of epidemic and nonepidemic strains of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci.* **110**, 21065–21070 (2013).

205. Liu, B. *et al.* Structural diversity in Salmonella O antigens and its genetic basis. *FEMS Microbiol. Rev.* **38**, 56–89 (2014).

206. Haake, D. A. & Zückert, W. R. The Leptospiral Outer Membrane. in *Leptospira and Leptospirosis* (ed. Adler, B.) vol. 387 187–221 (Springer Berlin Heidelberg, 2015).

207. Santecchia, I., Ferrer, M. F., Vieira, M. L., Gómez, R. M. & Werts, C. Phagocyte Escape of Leptospira: The Role of TLRs and NLRs. *Front. Immunol.* **11**, 571816 (2020).

208. Bulach, D. M. *et al.* Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. *Proc. Natl. Acad. Sci.* **103**, 14560–14565 (2006).

209. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15–e15 (2015).

210. Didelot, X. & Wilson, D. J. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Comput. Biol.* **11**, e1004041 (2015).

# Appendices

## Supplementary material for Chapter 1

**Supplementary Table 1.** Sequence Type and Plasmid determinants present in non-*S.* Infantis isolates. Detection was done *in silico* with PubMLST and PlasmidFinder.

| Sample ID | Serovar | ST | Col156 | Col8282 | ColRNAI | IncFIB | IncFII_S | IncQ1 |
|---|---|---|---|---|---|---|---|---|
| U875s | Enteritidis | 11 | 0 | 0 | 0 | 1 | 1 | 0 |
| U114s | Enteritidis | 11 | 0 | 0 | 0 | 1 | 1 | 0 |
| U1193s | Enteritidis | 11 | 0 | 0 | 0 | 1 | 1 | 0 |
| U2506s | Enteritidis | 11 | 0 | 0 | 0 | 1 | 1 | 0 |
| U2526s | Enteritidis | 11 | 0 | 0 | 0 | 1 | 1 | 0 |
| U2451s | Monophasic Typhimurium 4,[5],12:i:-. | 2379 | 1 | 0 | 0 | 0 | 0 | 0 |
| U113s | Typhimurium | 19 | 0 | 1 | 1 | 1 | 1 | 1 |

**Supplementary Table 2.** Antibiotic resistance patterns of *S.* Infantis.

| Resistant pattern | No. Antimicrobial classes | Sample origin | | | |
|---|---|---|---|---|---|
| | | Poultry farms (farm) | Chicken carcasses (food) | Stool samples (human) | Total |
| SAQBTFNMP | 9 | 3 | 10 | | 13 |
| SAQBTFNP | 8 | 10 | 26 | 1 | 37 |
| SAQBTFNM | 8 | 2 | 10 | | 12 |
| SABTFNMP | 8 | 1 | 1 | | 2 |
| AQBTFNMP | 8 | | 1 | | 1 |
| SAQBTFN | 7 | 9 | 28 | | 37 |
| SAQTFNM | 7 | | 1 | | 1 |
| SABTFNP | 7 | 6 | 19 | | 25 |
| SABTFNM | 7 | | 1 | | 1 |
| SAQBTNM | 7 | 2 | | | 2 |
| AQBTFNM | 7 | 1 | 2 | | 3 |
| SAQBTNP | 7 | | 1 | | 1 |
| SABTNMP | 7 | 1 | | | 1 |
| AQBTFNP | 7 | 1 | 3 | | 4 |
| SQBTFNP | 7 | 1 | | | 1 |
| SABFNMP | 7 | | 1 | | 1 |
| SABTFN | 6 | 5 | 22 | | 27 |
| AQBTNM | 6 | | 2 | | 2 |
| SAQTFN | 6 | 2 | 6 | | 8 |
| SAQBTN | 6 | 1 | 4 | | 5 |
| SABFNP | 6 | | 1 | | 1 |
| AQBTFN | 6 | 2 | 6 | | 8 |
| ABTFNP | 6 | | 1 | | 1 |
| AQBFNP | 6 | | 1 | | 1 |
| AQBTNP | 6 | | 1 | | 1 |
| SAQTNM | 6 | | 1 | | 1 |
| AQBTN | 5 | 1 | 1 | | 2 |
| SAQTN | 5 | | 8 | | 8 |
| AQBNP | 5 | | 1 | | 1 |
| ABTFN | 5 | 2 | 2 | | 4 |
| SABTN | 5 | 1 | 5 | | 6 |
| SATFN | 5 | | 6 | | 6 |
| SATNM | 5 | | 1 | | 1 |
| SABFP | 5 | | 1 | | 1 |
| AQTN | 4 | | 1 | | 1 |
| ABTN | 4 | 1 | | | 1 |
| SATN | 4 | | 1 | | 1 |
| SQTN | 4 | | 2 | | 2 |
| ATFN | 4 | | | 1 | 1 |
| BTN | 3 | 2 | 1 | | 3 |

| | | | | | |
|---|---|---|---|---|---|
| STN | 3 | | 2 | | 2 |
| AT | 2 | | | 1 | 1 |
| **Total** | | **54** | **181** | **3** | **239** |

Sulfonamide (S), Aminoglycosides (A), Quinolones (Q), Beta-lactams (B), Tetracyclines (T), Phenicol (F), Nitrofuran (N), Macrolides (M), Fosfomycin (P).

## **Supplementary Table 3.** Raw read statistics of *Salmonella enterica* dataset

| Sample name | Sequence ID | Number of reads before QC | Number of reads after QC | % Duplicate Reads | Average % GC Content | Average Sequence Length (bp) |
|---|---|---|---|---|---|---|
| CSG217 | Se_Q_001_1 | 1139333 | 1067280 | 19.24 | 51.00 | 139.64 |
| | Se_Q_001_2 | 1139333 | 1067280 | 18.38 | 51.00 | 137.37 |
| CSG232 | Se_Q_002_1 | 1023542 | 958762 | 18.37 | 51.00 | 138.34 |
| | Se_Q_002_2 | 1023542 | 958762 | 17.64 | 51.00 | 136.56 |
| CSG252 | Se_Q_003_1 | 755563 | 707087 | 14.63 | 51.00 | 138.92 |
| | Se_Q_003_2 | 755563 | 707087 | 13.99 | 51.00 | 136.98 |
| CSG177 | Se_Q_004_1 | 634856 | 593892 | 15.56 | 51.00 | 140.72 |
| | Se_Q_004_2 | 634856 | 593892 | 14.73 | 51.00 | 137.89 |
| CSG175 | Se_Q_005_1 | 648678 | 608515 | 12.83 | 51.00 | 139.27 |
| | Se_Q_005_2 | 648678 | 608515 | 12.39 | 51.00 | 137.10 |
| CSG132 | Se_Q_006_1 | 694442 | 647666 | 13.68 | 51.00 | 138.26 |
| | Se_Q_006_2 | 694442 | 647666 | 12.71 | 51.00 | 136.10 |
| U113s | Se_Q_007_1 | 618997 | 577466 | 13.77 | 51.00 | 140.03 |
| | Se_Q_007_2 | 618997 | 577466 | 13.16 | 51.00 | 137.34 |
| U114s | Se_Q_008_1 | 605178 | 566546 | 13.02 | 51.00 | 140.20 |
| | Se_Q_008_2 | 605178 | 566546 | 12.35 | 51.00 | 137.66 |
| U120s | Se_Q_009_1 | 1310684 | 1227772 | 23.40 | 50.00 | 136.86 |
| | Se_Q_009_2 | 1310684 | 1227772 | 22.24 | 50.00 | 134.78 |
| U121s | Se_Q_010_1 | 590923 | 552241 | 12.42 | 51.00 | 140.28 |
| | Se_Q_010_2 | 590923 | 552241 | 11.72 | 51.00 | 137.80 |
| U123s | Se_Q_011_1 | 537690 | 503581 | 12.82 | 50.00 | 140.51 |
| | Se_Q_011_2 | 537690 | 503581 | 12.20 | 50.00 | 137.57 |
| U127s | Se_Q_012_1 | 591806 | 552445 | 12.22 | 51.00 | 140.82 |
| | Se_Q_012_2 | 591806 | 552445 | 11.60 | 51.00 | 137.72 |
| U128s | Se_Q_013_1 | 842436 | 783719 | 15.62 | 51.00 | 141.08 |
| | Se_Q_013_2 | 842436 | 783719 | 14.33 | 51.00 | 137.19 |
| U129s | Se_Q_014_1 | 1043638 | 981104 | 20.35 | 49.00 | 139.05 |
| | Se_Q_014_2 | 1043638 | 981104 | 19.79 | 49.00 | 137.28 |
| U219s | Se_Q_015_1 | 840919 | 789739 | 16.27 | 50.00 | 138.85 |
| | Se_Q_015_2 | 840919 | 789739 | 15.59 | 50.00 | 137.14 |
| U634s | Se_Q_016_1 | 530082 | 496633 | 11.35 | 51.00 | 140.95 |
| | Se_Q_016_2 | 530082 | 496633 | 10.65 | 51.00 | 138.53 |
| U638s | Se_Q_017_1 | 801194 | 753077 | 15.02 | 51.00 | 139.37 |
| | Se_Q_017_2 | 801194 | 753077 | 14.74 | 51.00 | 137.65 |
| U639s | Se_Q_018_1 | 808578 | 758555 | 15.46 | 51.00 | 138.87 |
| | Se_Q_018_2 | 808578 | 758555 | 14.85 | 51.00 | 137.17 |
| U652s | Se_Q_019_1 | 802129 | 753111 | 16.27 | 50.00 | 139.44 |
| | Se_Q_019_2 | 802129 | 753111 | 15.53 | 50.00 | 137.48 |
| U666s | Se_Q_022_1 | 669033 | 627332 | 13.64 | 51.00 | 140.03 |
| | Se_Q_022_2 | 669033 | 627332 | 13.10 | 51.00 | 138.13 |
| U669s | Se_Q_023_1 | 697412 | 654117 | 13.88 | 51.00 | 139.70 |
| | Se_Q_023_2 | 697412 | 654117 | 13.29 | 51.00 | 137.64 |
| U672s | Se_Q_024_1 | 792573 | 742819 | 15.16 | 51.00 | 139.01 |
| | Se_Q_024_2 | 792573 | 742819 | 14.55 | 51.00 | 137.02 |
| U676s | Se_Q_025_1 | 637730 | 599491 | 13.72 | 50.00 | 140.95 |
| | Se_Q_025_2 | 637730 | 599491 | 13.05 | 50.00 | 138.81 |
| U679s | Se_Q_026_1 | 1168723 | 1094358 | 19.26 | 51.00 | 139.07 |
| | Se_Q_026_2 | 1168723 | 1094358 | 18.33 | 51.00 | 137.13 |
| U682s | Se_Q_027_1 | 977262 | 913901 | 16.61 | 51.00 | 136.88 |
| | Se_Q_027_2 | 977262 | 913901 | 16.02 | 51.00 | 135.24 |
| U684s | Se_Q_028_1 | 793817 | 742994 | 15.24 | 51.00 | 138.31 |
| | Se_Q_028_2 | 793817 | 742994 | 14.53 | 51.00 | 136.32 |
| U689s | Se_Q_029_1 | 807907 | 756384 | 15.17 | 51.00 | 138.29 |
| | Se_Q_029_2 | 807907 | 756384 | 14.43 | 51.00 | 136.37 |
| U692s | Se_Q_030_1 | 890987 | 831701 | 15.92 | 51.00 | 136.27 |
| | Se_Q_030_2 | 890987 | 831701 | 15.37 | 51.00 | 134.61 |

| Sample name | Sequence ID | Number of reads before QC | Number of reads after QC | % Duplicate Reads | Average % GC Content | Average Sequence Length (bp) |
|---|---|---|---|---|---|---|
| U706s | Se_Q_031_1 | 708538 | 662819 | 13.95 | 51.00 | 138.93 |
| | Se_Q_031_2 | 708538 | 662819 | 13.21 | 51.00 | 136.80 |
| U707s | Se_Q_032_1 | 853711 | 800421 | 17.95 | 50.00 | 139.40 |
| | Se_Q_032_2 | 853711 | 800421 | 16.97 | 50.00 | 137.22 |
| U708s | Se_Q_033_1 | 806660 | 755989 | 15.36 | 51.00 | 138.08 |
| | Se_Q_033_2 | 806660 | 755989 | 14.56 | 51.00 | 136.17 |
| U711s | Se_Q_034_1 | 589871 | 551231 | 12.17 | 51.00 | 140.59 |
| | Se_Q_034_2 | 589871 | 551231 | 11.73 | 51.00 | 138.17 |
| U712s | Se_Q_035_1 | 1063865 | 995970 | 18.45 | 51.00 | 138.62 |
| | Se_Q_035_2 | 1063865 | 995970 | 17.76 | 51.00 | 136.57 |
| U715s | Se_Q_036_1 | 1562915 | 1460637 | 23.78 | 51.00 | 139.04 |
| | Se_Q_036_2 | 1562915 | 1460637 | 22.51 | 51.00 | 136.53 |
| U718s | Se_Q_037_1 | 247022 | 228263 | 6.00 | 51.00 | 140.18 |
| | Se_Q_037_2 | 247022 | 228263 | 5.54 | 51.00 | 136.15 |
| U719s | Se_Q_038_1 | 2297867 | 2133371 | 28.80 | 51.00 | 132.61 |
| | Se_Q_038_2 | 2297867 | 2133371 | 27.90 | 51.00 | 131.13 |
| U723s | Se_Q_039_1 | 1384495 | 1277976 | 19.83 | 51.00 | 130.47 |
| | Se_Q_039_2 | 1384495 | 1277976 | 19.78 | 51.00 | 127.87 |
| U728s | Se_Q_041_1 | 2070236 | 1926880 | 27.28 | 51.00 | 132.93 |
| | Se_Q_041_2 | 2070236 | 1926880 | 26.66 | 51.00 | 131.70 |
| U729s | Se_Q_042_1 | 1144418 | 1070009 | 18.98 | 51.00 | 139.18 |
| | Se_Q_042_2 | 1144418 | 1070009 | 17.95 | 51.00 | 137.15 |
| U758s | Se_Q_043_1 | 1025769 | 957227 | 17.58 | 51.00 | 137.01 |
| | Se_Q_043_2 | 1025769 | 957227 | 16.83 | 51.00 | 135.00 |
| U759s | Se_Q_044_1 | 1097155 | 1022428 | 18.15 | 51.00 | 135.68 |
| | Se_Q_044_2 | 1097155 | 1022428 | 17.37 | 51.00 | 133.85 |
| U762s | Se_Q_045_1 | 1329339 | 1237204 | 20.84 | 51.00 | 133.53 |
| | Se_Q_045_2 | 1329339 | 1237204 | 20.13 | 51.00 | 131.94 |
| U769s | Se_Q_046_1 | 1029748 | 963248 | 18.32 | 51.00 | 138.92 |
| | Se_Q_046_2 | 1029748 | 963248 | 17.39 | 51.00 | 136.96 |
| U773s | Se_Q_047_1 | 977079 | 913903 | 17.51 | 51.00 | 139.11 |
| | Se_Q_047_2 | 977079 | 913903 | 16.59 | 51.00 | 136.87 |
| U801s | Se_Q_048_1 | 1853982 | 1719878 | 25.90 | 51.00 | 134.05 |
| | Se_Q_048_2 | 1853982 | 1719878 | 24.99 | 51.00 | 132.24 |
| U804s | Se_Q_049_1 | 1591766 | 1489774 | 23.53 | 51.00 | 136.92 |
| | Se_Q_049_2 | 1591766 | 1489774 | 22.70 | 51.00 | 135.36 |
| U806s | Se_Q_050_1 | 1101621 | 1030533 | 18.96 | 51.00 | 139.10 |
| | Se_Q_050_2 | 1101621 | 1030533 | 18.45 | 51.00 | 137.24 |
| U814s | Se_Q_051_1 | 2110096 | 1976981 | 28.99 | 51.00 | 139.79 |
| | Se_Q_051_2 | 2110096 | 1976981 | 27.92 | 51.00 | 137.80 |
| U815s | Se_Q_052_1 | 1024911 | 956914 | 19.02 | 51.00 | 140.23 |
| | Se_Q_052_2 | 1024911 | 956914 | 18.05 | 51.00 | 137.82 |
| U818s | Se_Q_053_1 | 1253886 | 1178226 | 21.12 | 51.00 | 139.17 |
| | Se_Q_053_2 | 1253886 | 1178226 | 20.46 | 51.00 | 137.65 |
| U855s | Se_Q_054_1 | 1128597 | 1058909 | 19.64 | 51.00 | 140.22 |
| | Se_Q_054_2 | 1128597 | 1058909 | 19.01 | 51.00 | 138.15 |
| U857s | Se_Q_055_1 | 1106425 | 1034670 | 19.35 | 51.00 | 139.22 |
| | Se_Q_055_2 | 1106425 | 1034670 | 18.43 | 51.00 | 137.29 |
| U860s | Se_Q_056_1 | 805401 | 756052 | 15.97 | 51.00 | 140.31 |
| | Se_Q_056_2 | 805401 | 756052 | 15.52 | 51.00 | 138.34 |
| U863s | Se_Q_057_1 | 14007207 | 13042665 | 63.40 | 52.00 | 133.58 |
| | Se_Q_057_2 | 14007207 | 13042665 | 62.41 | 52.00 | 132.31 |
| U865s | Se_Q_058_1 | 1614863 | 1511753 | 24.88 | 52.00 | 140.04 |
| | Se_Q_058_2 | 1614863 | 1511753 | 24.22 | 52.00 | 138.10 |
| U866s | Se_Q_059_1 | 818649 | 767177 | 16.47 | 51.00 | 140.10 |
| | Se_Q_059_2 | 818649 | 767177 | 15.67 | 51.00 | 137.89 |
| U869s | Se_Q_060_1 | 1607408 | 1506509 | 25.23 | 51.00 | 138.39 |
| | Se_Q_060_2 | 1607408 | 1506509 | 24.38 | 51.00 | 136.67 |

| Sample name | Sequence ID | Number of reads before QC | Number of reads after QC | % Duplicate Reads | Average % GC Content | Average Sequence Length (bp) |
|---|---|---|---|---|---|---|
| U871s | Se_Q_061_1 | 767191 | 719460 | 15.38 | 51.00 | 140.48 |
| | Se_Q_061_2 | 767191 | 719460 | 14.74 | 51.00 | 138.41 |
| U873s | Se_Q_062_1 | 1209880 | 1128762 | 21.70 | 51.00 | 140.91 |
| | Se_Q_062_2 | 1209880 | 1128762 | 20.55 | 51.00 | 137.98 |
| U875s | Se_Q_063_1 | 1600958 | 1496497 | 25.63 | 51.00 | 141.21 |
| | Se_Q_063_2 | 1600958 | 1496497 | 24.32 | 51.00 | 138.38 |
| U878s | Se_Q_064_1 | 893764 | 832305 | 16.62 | 51.00 | 140.30 |
| | Se_Q_064_2 | 893764 | 832305 | 15.51 | 51.00 | 137.61 |
| U880s | Se_Q_065_1 | 1138215 | 1068562 | 21.64 | 50.00 | 140.65 |
| | Se_Q_065_2 | 1138215 | 1068562 | 20.72 | 50.00 | 138.34 |
| U883s | Se_Q_066_1 | 885052 | 831626 | 16.74 | 51.00 | 139.27 |
| | Se_Q_066_2 | 885052 | 831626 | 16.07 | 51.00 | 137.34 |
| U885s | Se_Q_067_1 | 848237 | 792512 | 16.71 | 51.00 | 140.43 |
| | Se_Q_067_2 | 848237 | 792512 | 15.46 | 51.00 | 138.09 |
| U888s | Se_Q_068_1 | 2258769 | 2120434 | 30.02 | 51.00 | 139.23 |
| | Se_Q_068_2 | 2258769 | 2120434 | 29.03 | 51.00 | 137.43 |
| U889s | Se_Q_069_1 | 636071 | 595535 | 13.28 | 51.00 | 140.72 |
| | Se_Q_069_2 | 636071 | 595535 | 12.38 | 51.00 | 138.17 |
| U890s | Se_Q_070_1 | 870242 | 813915 | 16.90 | 51.00 | 139.02 |
| | Se_Q_070_2 | 870242 | 813915 | 15.96 | 51.00 | 137.06 |
| U895s | Se_Q_071_1 | 660463 | 618513 | 13.78 | 51.00 | 140.71 |
| | Se_Q_071_2 | 660463 | 618513 | 12.79 | 51.00 | 138.13 |
| U897s | Se_Q_072_1 | 597370 | 559621 | 12.60 | 51.00 | 140.76 |
| | Se_Q_072_2 | 597370 | 559621 | 12.09 | 51.00 | 138.43 |
| U899s | Se_Q_073_1 | 802986 | 750837 | 16.06 | 51.00 | 139.27 |
| | Se_Q_073_2 | 802986 | 750837 | 15.27 | 51.00 | 137.09 |
| U969s | Se_Q_074_1 | 745383 | 695067 | 14.67 | 51.00 | 141.33 |
| | Se_Q_074_2 | 745383 | 695067 | 13.74 | 51.00 | 137.99 |
| U972s | Se_Q_075_1 | 959647 | 898702 | 18.04 | 51.00 | 140.91 |
| | Se_Q_075_2 | 959647 | 898702 | 17.07 | 51.00 | 138.33 |
| U973s | Se_Q_076_1 | 710660 | 663834 | 14.66 | 51.00 | 140.79 |
| | Se_Q_076_2 | 710660 | 663834 | 13.84 | 51.00 | 138.08 |
| U976s | Se_Q_077_1 | 456188 | 427583 | 10.45 | 51.00 | 140.83 |
| | Se_Q_077_2 | 456188 | 427583 | 9.95 | 51.00 | 138.55 |
| U979s | Se_Q_078_1 | 786237 | 739350 | 15.63 | 51.00 | 140.63 |
| | Se_Q_078_2 | 786237 | 739350 | 15.02 | 51.00 | 138.49 |
| U980s | Se_Q_079_1 | 780165 | 729389 | 15.74 | 51.00 | 140.26 |
| | Se_Q_079_2 | 780165 | 729389 | 14.78 | 51.00 | 137.87 |
| U983s | Se_Q_080_1 | 948614 | 891439 | 17.75 | 51.00 | 139.86 |
| | Se_Q_080_2 | 948614 | 891439 | 17.07 | 51.00 | 138.02 |
| U986s | Se_Q_081_1 | 990612 | 926751 | 18.37 | 51.00 | 138.22 |
| | Se_Q_081_2 | 990612 | 926751 | 17.44 | 51.00 | 136.15 |
| U998s | Se_Q_082_1 | 620190 | 579474 | 12.96 | 51.00 | 141.13 |
| | Se_Q_082_2 | 620190 | 579474 | 12.18 | 51.00 | 138.27 |
| U1000s | Se_Q_083_1 | 1661763 | 1558740 | 25.52 | 51.00 | 140.21 |
| | Se_Q_083_2 | 1661763 | 1558740 | 24.39 | 51.00 | 137.75 |
| U1003s | Se_Q_084_1 | 698834 | 655006 | 14.29 | 51.00 | 140.78 |
| | Se_Q_084_2 | 698834 | 655006 | 13.47 | 51.00 | 138.27 |
| U1006s | Se_Q_085_1 | 646727 | 605550 | 13.54 | 51.00 | 141.24 |
| | Se_Q_085_2 | 646727 | 605550 | 12.77 | 51.00 | 138.49 |
| U1010s | Se_Q_086_1 | 2430323 | 2252194 | 30.97 | 51.00 | 137.32 |
| | Se_Q_086_2 | 2430323 | 2252194 | 29.04 | 51.00 | 134.57 |
| U1019s | Se_Q_087_1 | 1920388 | 1785515 | 29.36 | 50.00 | 140.74 |
| | Se_Q_087_2 | 1920388 | 1785515 | 27.34 | 50.00 | 137.14 |
| U1051s | Se_Q_088_1 | 1542875 | 1429365 | 24.61 | 51.00 | 140.26 |
| | Se_Q_088_2 | 1542875 | 1429365 | 22.64 | 51.00 | 136.60 |
| U1052s | Se_Q_089_1 | 1636264 | 1519817 | 24.54 | 51.00 | 136.91 |
| | Se_Q_089_2 | 1636264 | 1519817 | 23.20 | 51.00 | 134.43 |

| Sample name | Sequence ID | Number of reads before QC | Number of reads after QC | % Duplicate Reads | Average % GC Content | Average Sequence Length (bp) |
|---|---|---|---|---|---|---|
| U1056s | Se_Q_090_1 | 1516150 | 1414930 | 23.86 | 51.00 | 139.88 |
| | Se_Q_090_2 | 1516150 | 1414930 | 22.44 | 51.00 | 137.05 |
| U1057s | Se_Q_091_1 | 14106025 | 13094594 | 63.58 | 51.00 | 139.33 |
| | Se_Q_091_2 | 14106025 | 13094594 | 61.06 | 51.00 | 136.38 |
| U1059s | Se_Q_092_1 | 2048875 | 1907887 | 28.33 | 51.00 | 139.46 |
| | Se_Q_092_2 | 2048875 | 1907887 | 26.82 | 51.00 | 136.49 |
| U1066s | Se_Q_093_1 | 2705722 | 2502027 | 32.64 | 51.00 | 134.17 |
| | Se_Q_093_2 | 2705722 | 2502027 | 30.89 | 51.00 | 131.89 |
| U1068s | Se_Q_094_1 | 1648740 | 1514273 | 23.03 | 52.00 | 130.20 |
| | Se_Q_094_2 | 1648740 | 1514273 | 22.11 | 52.00 | 129.12 |
| U1071s | Se_Q_095_1 | 2703682 | 2514344 | 32.79 | 52.00 | 140.55 |
| | Se_Q_095_2 | 2703682 | 2514344 | 30.63 | 52.00 | 137.09 |
| U1092s | Se_Q_096_1 | 3097842 | 2884752 | 35.79 | 51.00 | 139.76 |
| | Se_Q_096_2 | 3097842 | 2884752 | 33.56 | 52.00 | 136.85 |
| U1095s | Se_Q_097_1 | 3366537 | 3134041 | 36.73 | 51.00 | 138.49 |
| | Se_Q_097_2 | 3366537 | 3134041 | 34.72 | 51.00 | 135.91 |
| U1111s | Se_Q_098_1 | 1840857 | 1718591 | 27.13 | 51.00 | 140.96 |
| | Se_Q_098_2 | 1840857 | 1718591 | 25.62 | 51.00 | 137.97 |
| U1114s | Se_Q_099_1 | 2265612 | 2119611 | 30.61 | 51.00 | 140.69 |
| | Se_Q_099_2 | 2265612 | 2119611 | 29.04 | 51.00 | 138.11 |
| U1119s | Se_Q_100_1 | 1018113 | 945878 | 18.54 | 51.00 | 141.06 |
| | Se_Q_100_2 | 1018113 | 945878 | 17.31 | 51.00 | 137.46 |
| U1121s | Se_Q_101_1 | 1661754 | 1556908 | 25.42 | 51.00 | 140.81 |
| | Se_Q_101_2 | 1661754 | 1556908 | 24.51 | 51.00 | 138.38 |
| U1125s | Se_Q_102_1 | 1891940 | 1777370 | 27.65 | 51.00 | 140.91 |
| | Se_Q_102_2 | 1891940 | 1777370 | 26.47 | 51.00 | 138.60 |
| U1128s | Se_Q_103_1 | 1460645 | 1364446 | 23.64 | 51.00 | 140.64 |
| | Se_Q_103_2 | 1460645 | 1364446 | 22.31 | 51.00 | 138.14 |
| U1132s | Se_Q_104_1 | 1451757 | 1356201 | 22.71 | 51.00 | 141.12 |
| | Se_Q_104_2 | 1451757 | 1356201 | 21.87 | 51.00 | 138.24 |
| U1133s | Se_Q_105_1 | 1601472 | 1493529 | 24.76 | 51.00 | 140.89 |
| | Se_Q_105_2 | 1601472 | 1493529 | 23.41 | 51.00 | 137.79 |
| U1135s | Se_Q_106_1 | 11132882 | 10398816 | 59.45 | 51.00 | 140.49 |
| | Se_Q_106_2 | 11132882 | 10398816 | 57.89 | 51.00 | 137.90 |
| U1143s | Se_Q_107_1 | 989447 | 926374 | 19.01 | 51.00 | 140.55 |
| | Se_Q_107_2 | 989447 | 926374 | 17.95 | 51.00 | 137.96 |
| U1145s | Se_Q_108_1 | 935982 | 877638 | 17.98 | 51.00 | 140.31 |
| | Se_Q_108_2 | 935982 | 877638 | 17.05 | 51.00 | 138.05 |
| U1175s | Se_Q_109_1 | 712225 | 666827 | 14.40 | 51.00 | 140.83 |
| | Se_Q_109_2 | 712225 | 666827 | 13.65 | 51.00 | 138.31 |
| U1178s | Se_Q_110_1 | 1636297 | 1524812 | 25.13 | 51.00 | 141.05 |
| | Se_Q_110_2 | 1636297 | 1524812 | 23.56 | 51.00 | 137.74 |
| U1181s | Se_Q_111_1 | 931427 | 866838 | 17.82 | 51.00 | 141.54 |
| | Se_Q_111_2 | 931427 | 866838 | 16.58 | 51.00 | 137.70 |
| U1187s | Se_Q_112_1 | 783399 | 730173 | 16.79 | 50.00 | 140.58 |
| | Se_Q_112_2 | 783399 | 730173 | 15.76 | 50.00 | 137.40 |
| U1192s | Se_Q_113_1 | 790171 | 738417 | 15.81 | 51.00 | 140.61 |
| | Se_Q_113_2 | 790171 | 738417 | 14.73 | 51.00 | 137.75 |
| U1193s | Se_Q_114_1 | 959943 | 900076 | 18.48 | 51.00 | 140.55 |
| | Se_Q_114_2 | 959943 | 900076 | 17.45 | 51.00 | 138.04 |
| U1196s | Se_Q_115_1 | 1448712 | 1350952 | 23.13 | 51.00 | 140.54 |
| | Se_Q_115_2 | 1448712 | 1350952 | 21.83 | 51.00 | 137.65 |
| U1401s | Se_Q_116_1 | 737674 | 690562 | 14.56 | 51.00 | 139.70 |
| | Se_Q_116_2 | 737674 | 690562 | 13.93 | 51.00 | 137.46 |
| U1405s | Se_Q_117_1 | 964380 | 898048 | 17.86 | 51.00 | 141.28 |
| | Se_Q_117_2 | 964380 | 898048 | 16.62 | 51.00 | 137.63 |
| U1407s | Se_Q_118_1 | 1520020 | 1417727 | 23.71 | 51.00 | 140.54 |
| | Se_Q_118_2 | 1520020 | 1417727 | 22.32 | 51.00 | 137.67 |

| Sample name | Sequence ID | Number of reads before QC | Number of reads after QC | % Duplicate Reads | Average % GC Content | Average Sequence Length (bp) |
|---|---|---|---|---|---|---|
| U1410s | Se_Q_119_1 | 609264 | 568543 | 12.85 | 51.00 | 140.98 |
| | Se_Q_119_2 | 609264 | 568543 | 11.90 | 51.00 | 137.69 |
| U1411s | Se_Q_120_1 | 710193 | 663114 | 14.73 | 51.00 | 141.04 |
| | Se_Q_120_2 | 710193 | 663114 | 13.77 | 51.00 | 137.89 |
| U1418s | Se_Q_122_1 | 1050959 | 980550 | 18.47 | 51.00 | 140.41 |
| | Se_Q_122_2 | 1050959 | 980550 | 17.44 | 51.00 | 137.73 |
| U1424s | Se_Q_123_1 | 860486 | 801822 | 16.77 | 51.00 | 140.56 |
| | Se_Q_123_2 | 860486 | 801822 | 15.40 | 51.00 | 137.40 |
| U1431s | Se_Q_124_1 | 631376 | 588241 | 13.40 | 51.00 | 140.70 |
| | Se_Q_124_2 | 631376 | 588241 | 12.69 | 51.00 | 137.81 |
| U1432s | Se_Q_125_1 | 615936 | 574971 | 13.48 | 51.00 | 140.51 |
| | Se_Q_125_2 | 615936 | 574971 | 12.50 | 51.00 | 137.49 |
| U1436s | Se_Q_126_1 | 615389 | 576677 | 13.03 | 51.00 | 140.68 |
| | Se_Q_126_2 | 615389 | 576677 | 12.25 | 51.00 | 137.96 |
| U1437s | Se_Q_127_1 | 920726 | 856144 | 17.17 | 51.00 | 140.31 |
| | Se_Q_127_2 | 920726 | 856144 | 15.99 | 51.00 | 137.33 |
| U1447s | Se_Q_128_1 | 921906 | 861438 | 17.10 | 51.00 | 140.59 |
| | Se_Q_128_2 | 921906 | 861438 | 16.19 | 51.00 | 137.71 |
| U1459s | Se_Q_129_1 | 1018612 | 950675 | 18.82 | 51.00 | 138.16 |
| | Se_Q_129_2 | 1018612 | 950675 | 17.47 | 51.00 | 135.84 |
| U1470s | Se_Q_131_1 | 628796 | 585700 | 12.99 | 51.00 | 140.86 |
| | Se_Q_131_2 | 628796 | 585700 | 12.00 | 51.00 | 137.46 |
| U1473s | Se_Q_132_1 | 499772 | 465596 | 10.85 | 51.00 | 140.52 |
| | Se_Q_132_2 | 499772 | 465596 | 10.03 | 51.00 | 137.59 |
| U1483s | Se_Q_133_1 | 1075321 | 1004574 | 19.01 | 51.00 | 139.36 |
| | Se_Q_133_2 | 1075321 | 1004574 | 17.80 | 51.00 | 136.90 |
| U1489s | Se_Q_134_1 | 351084 | 323623 | 7.89 | 51.00 | 140.90 |
| | Se_Q_134_2 | 351084 | 323623 | 7.28 | 51.00 | 136.37 |
| U1493s | Se_Q_135_1 | 759308 | 706967 | 15.97 | 51.00 | 141.04 |
| | Se_Q_135_2 | 759308 | 706967 | 14.73 | 51.00 | 137.59 |
| U1495s | Se_Q_136_1 | 916561 | 855410 | 18.30 | 51.00 | 138.96 |
| | Se_Q_136_2 | 916561 | 855410 | 17.32 | 51.00 | 136.64 |
| U1505s | Se_Q_138_1 | 960003 | 900506 | 17.64 | 51.00 | 140.70 |
| | Se_Q_138_2 | 960003 | 900506 | 16.70 | 51.00 | 138.31 |
| U1506s | Se_Q_139_1 | 658877 | 609825 | 12.95 | 51.00 | 139.79 |
| | Se_Q_139_2 | 658877 | 609825 | 12.18 | 51.00 | 136.48 |
| U1508s | Se_Q_140_1 | 1605003 | 1502344 | 24.76 | 51.00 | 138.87 |
| | Se_Q_140_2 | 1605003 | 1502344 | 23.64 | 51.00 | 136.80 |
| U1519s | Se_Q_141_1 | 2738120 | 2558365 | 34.33 | 52.00 | 137.83 |
| | Se_Q_141_2 | 2738120 | 2558365 | 32.94 | 52.00 | 135.72 |
| U1671s | Se_Q_143_1 | 1507090 | 1408319 | 23.92 | 51.00 | 139.64 |
| | Se_Q_143_2 | 1507090 | 1408319 | 22.40 | 51.00 | 137.12 |
| U1672s | Se_Q_144_1 | 1351547 | 1263663 | 22.42 | 51.00 | 139.34 |
| | Se_Q_144_2 | 1351547 | 1263663 | 21.67 | 51.00 | 137.27 |
| U1673s | Se_Q_145_1 | 861065 | 804006 | 17.37 | 50.00 | 141.25 |
| | Se_Q_145_2 | 861065 | 804006 | 16.23 | 50.00 | 138.15 |
| U1678s | Se_Q_146_1 | 969535 | 908717 | 17.57 | 51.00 | 140.16 |
| | Se_Q_146_2 | 969535 | 908717 | 17.00 | 51.00 | 138.17 |
| U1680s | Se_Q_147_1 | 993689 | 922055 | 16.71 | 51.00 | 135.63 |
| | Se_Q_147_2 | 993689 | 922055 | 16.21 | 51.00 | 134.11 |
| U1682s | Se_Q_148_1 | 697221 | 652327 | 14.23 | 51.00 | 140.80 |
| | Se_Q_148_2 | 697221 | 652327 | 13.63 | 51.00 | 138.41 |
| U1683s | Se_Q_149_1 | 1299705 | 1217310 | 20.51 | 51.00 | 140.25 |
| | Se_Q_149_2 | 1299705 | 1217310 | 19.88 | 51.00 | 138.20 |
| U1686s | Se_Q_150_1 | 1526404 | 1430986 | 24.60 | 50.00 | 140.03 |
| | Se_Q_150_2 | 1526404 | 1430986 | 23.88 | 50.00 | 138.07 |
| U1688s | Se_Q_151_1 | 1355301 | 1267491 | 21.47 | 51.00 | 140.13 |
| | Se_Q_151_2 | 1355301 | 1267491 | 20.67 | 51.00 | 137.95 |

## Supplementary File 1. Metadata for *Salmonella enterica* dataset

Available at:

https://docs.google.com/spreadsheets/d/1Hmwt4rKcJsXQ9rnICcVt3H1YT9UDt8MH/edit?usp=sharing&ouid=101340286042764398769&rtpof=true&sd=true

**Supplementary Table 4.** List of proteins present in the p-F219-like plasmids found in all samples.

**Proteins present in both plasmids**

2,3-dihydroxybenzoate-AMP ligase
Adenosine monophosphate-protein transferase SoFic
Anaerobic nitric oxide reductase transcription regulator NorR
Antirestriction protein KlcA
Antitoxin CcdA
Arginine/agmatine antiporter
ATP-dependent RNA helicase DbpA
Biodegradative arginine decarboxylase
Chaperone protein FaeE
Cyclic di-GMP phosphodiesterase PdeL
D-alanine--poly(phosphoribitol) ligase subunit 1
Dihydropteroate synthase
DNA adenine methyltransferase YhdJ
Endoribonuclease PemK
Inner membrane transport protein YhjV
IS110 family transposase ISEc21
IS110 family transposase ISSfl8
IS200/IS605 family transposase IS609
IS21 family transposase IS1326
IS21 family transposase ISEc10
IS256 family transposase IS285
IS256 family transposase ISEc39 (n=2)
IS256 family transposase ISSod4
IS3 family transposase ISEam1
IS3 family transposase ISIba2
IS3 family transposase ISKpn40 (n=2)
IS3 family transposase ISPa74
IS3 family transposase ISYpe1
IS4 family transposase ISSfl1 (n=2)
IS481 family transposase ISErsp1
IS630 family transposase ISEc40
ISL3 family transposase ISStma11
ISNCY family transposase ISRor2 (n=3)
K88 fimbrial protein AC
L-methionine gamma-lyase

Lactococcin-G-processing and transport ATP-binding protein LagD

Leader peptidase PppA

Lipoprotein signal peptidase

Low affinity potassium transport system protein kup

Major fimbrial subunit SMF-1

Mercuric reductase

Multidrug transporter EmrE

N-acetylmuramoyl-L-alanine amidase AmiD

Nucleoid occlusion protein

Outer membrane usher protein HtrE

Pesticin receptor

Phthiocerol synthesis polyketide synthase type I PpsE

Plasmid-derived single-stranded DNA-binding protein

Protein AmpG

Protein PsiB

Protein SopB

Protein UmuD

Protein YgiW

putative fimbrial chaperone YadV

putative protein YggR

putative protein YjiK

putative signaling protein

putative transport protein HsrA

RepFIB replication protein A

S6 family transposase IS26 (n=6)

Salicylate synthase

Streptomycin 3"-adenylyltransferase

Succinate-acetate/proton symporter SatP

Surfactin synthase thioesterase subunit

Tellurite resistance protein TehA

Tetracycline repressor protein class A from transposon 1721

Tetracycline resistance protein, class C

Tn3 family transposase TnAs1

Toxin CcdB

Transcription antitermination protein RfaH (n=2)

tRNA(fMet)-specific endonuclease VapC

Tyrosine recombinase XerC (n=6)

Ubiquinone/menaquinone biosynthesis C-methyltransferase UbiE

Vitamin B12 import ATP-binding protein BtuD (n=2)

Hypothetical proteins (174 different ORFs)

**Proteins exclusive to plasmid A**

Antitoxin PemI

Arsenical pump-driving ATPase

Arsenical resistance operon trans-acting repressor ArsD

Beta-lactamase Toho-1

Bicyclomycin resistance protein

ECF RNA polymerase sigma factor SigL

F1 capsule-anchoring protein

Glutathione transferase FosA

HTH-type transcriptional regulator HdfR

Hygromycin-B 4-O-kinase

IS200/IS605 family transposase ISShwo2

IS21 family transposase ISEc57

IS5 family transposase IS903

IS6 family transposase ISRle7

Lipopolysaccharide core heptose(II)-phosphate phosphatase

Mercuric transport protein MerT

Mercuric transport protein periplasmic component

NADPH-dependent FMN reductase ArsH

Plasmid segregation protein ParM

Protein PndA

Protein UmuC

putative outer membrane usher protein LpfC

Signal recognition particle 54 kDa protein

SPBc2 prophage-derived aminoglycoside N(3')-acetyltransferase-like protein YokD

Tn3 family transposase

Transcriptional repressor PifC

Tyrosine recombinase XerD

Vitamin B12 transporter BtuB

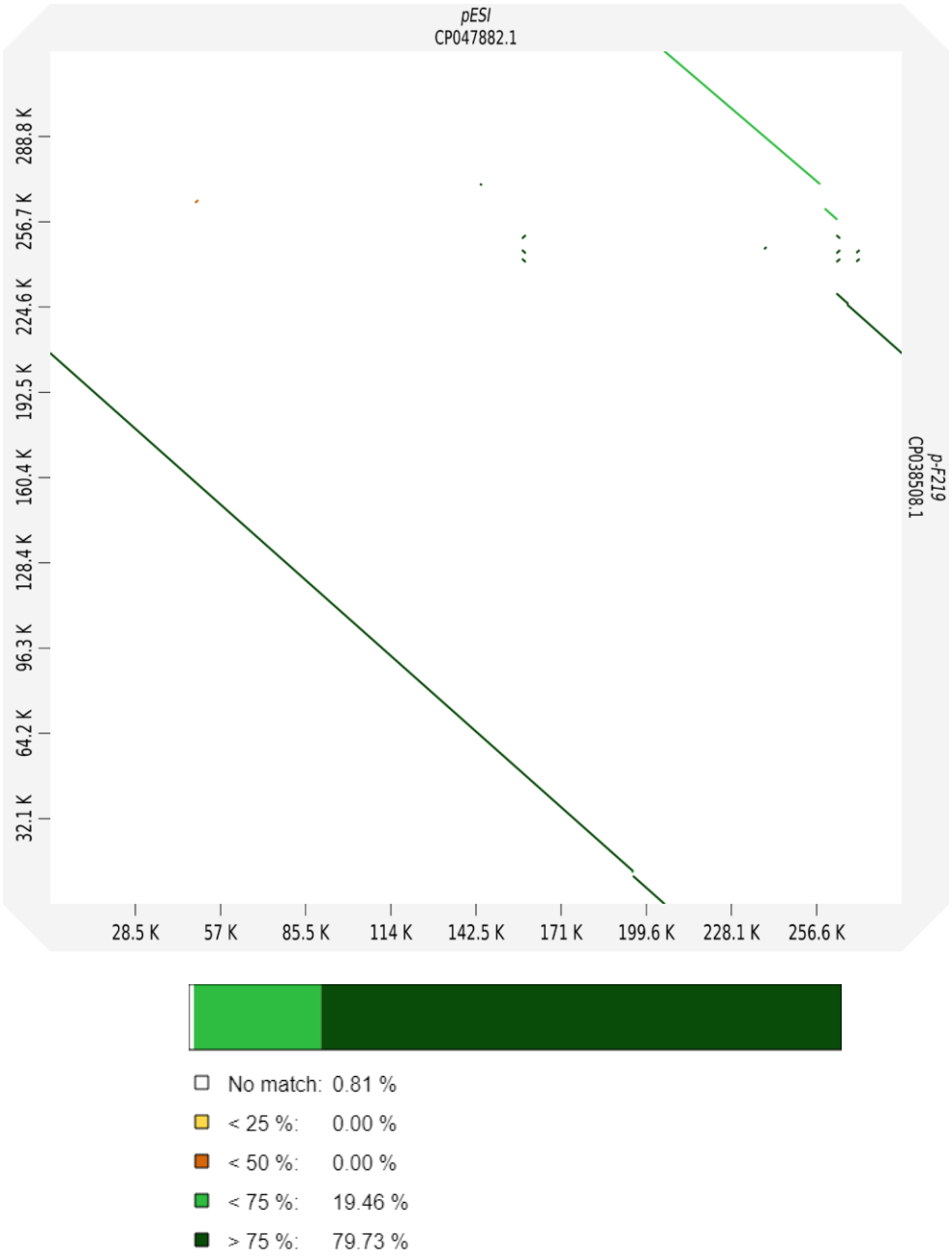
Hypothetical proteins (43 different ORFs)

**Proteins exclusive to plasmid B**

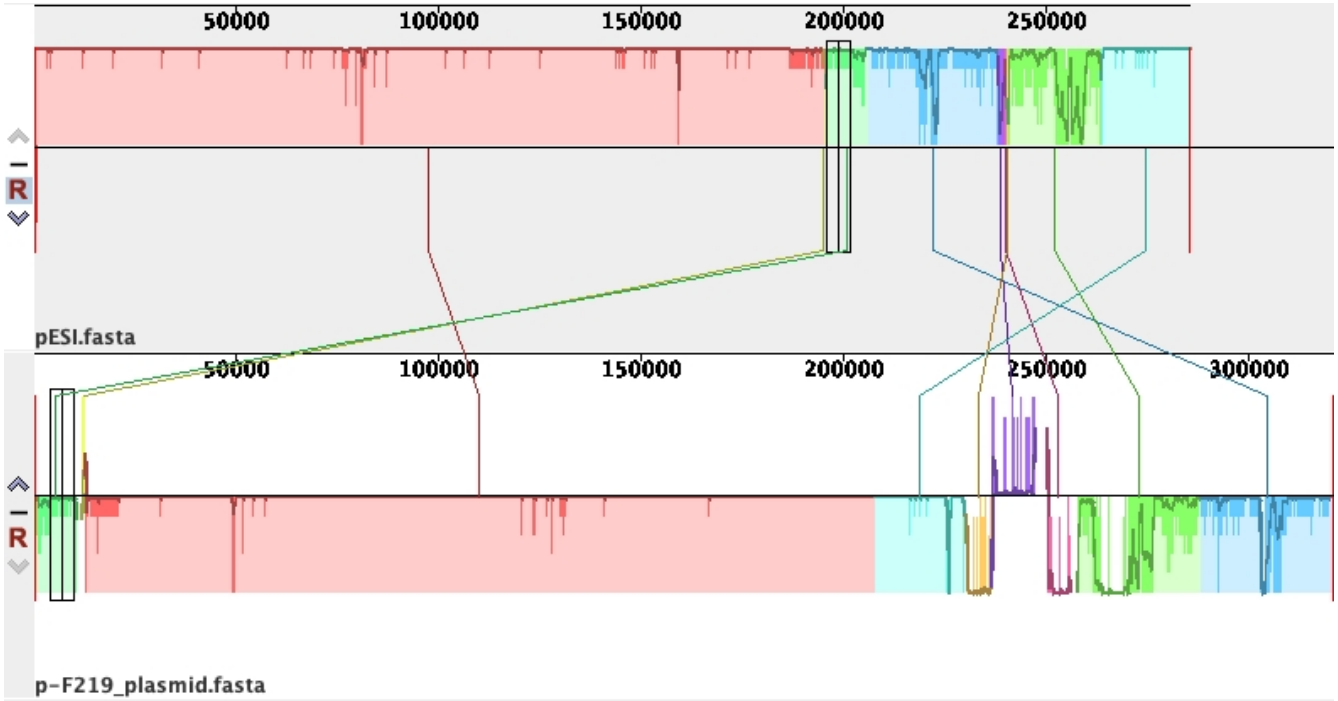
IS3 family transposase ISYps8

Hypothetical proteins (5 different ORFs)

**Supplementary Figure 1.** Dot Plot of genome comparison of pESI and p-F219 plasmids obtained with D-Genies aligner.



**Supplementary Figure 2.** Locally collinear blocks (LCBs) found in the genome alignment of pESI and p-F219 plasmids obtained with progressiveMauve.




LCBs are colored in order to identify the conserved segments that appear to be internally free from genome rearrangements and presumably homologous. The LCBs that appear below the center line indicate regions with inverse orientation.

**Supplementary Table 5.** *Salmonella* isolates in retailers and health care centers located in two different zones of Quito.

| **Origin of sample** | ***Salmonella* isolates (Number of samples)** | | |
|---|---|---|---|
| | **North Zone** | **South Zone** | **Total** |
| Open markets | 25 (42) | 23 (42) | 48 (84) |
| Supermarkets | 31 (61) | 33 (64) | 64 (125) |
| Small shops | 38 (63) | 35 (63) | 73 (126) |
| Total | 94 (166) | 92 (169) | 186 (335) |
| Health Care Centers | 0 (151) | 6(151) | 6 (302) |

## Supplementary material for Chapter 2

## **Supplementary File 2.** Metadata for *Listeria monocytogenes* references dataset.

Available at:

https://docs.google.com/spreadsheets/d/17OHRlEL8sk12Qs2AFkVMweQBL1uh7Dgf7HdpH6AJXM8/edit?usp=sharing

**Supplementary Table 6.** Accession number and gene location for reference genes of major *Listeria monocytogenes* virulence factors.

| Pathogenicity island | Gene | Accession number and gene location | Strain |
|---|---|---|---|
| **LIPI-1** | prfA | NC_003210.1:c204353-203640 | *Listeria monocytogenes* EGD-e |
| | plcA | NC_003210.1:c205577-204624 | |
| | hly | NC_003210.1:205819-207408 | |
| | mpl | NC_003210.1:207739-209271 | |
| | actA | NC_003210.1:209470-211389 | |
| | plcB | NC_003210.1:211425-212294 | |
| **LIPI-3** | LMOSLCC2378_RS05590 | NC_018585.1:1127442-1128329 | *Listeria monocytogenes* SLCC2378 |
| | LMOSLCC2378_RS05595 | NC_018585.1:1128329-1129078 | |
| | LMOSLCC2378_RS05600 | NC_018585.1:1129119-1129436 | |
| | LMOSLCC2378_RS05605 | NC_018585.1:1129433-1130308 | |
| | LMOSLCC2378_RS05610 | NC_018585.1:1130316-1131233 | |
| | LMOSLCC2378_RS05615 | NC_018585.1:1131226-1132539 | |
| | LMOSLCC2378_RS05620 | NC_018585.1:1132532-1133119 | |
| **LIPI-4** | LM4b_RS11745 | NC_012488.1:c2393494-2392132 | *Listeria monocytogenes* serotype 4b str. CLIP 80459 |
| | LM4b_RS11750 | NC_012488.1:c2395429-2393501 | |
| | LM4b_RS11755 | NC_012488.1:c2396192-2395433 | |
| | LM4b_RS11760 | NC_012488.1:c2396506-2396192 | |
| | LM4b_RS11765 | NC_012488.1:c2396874-2396506 | |
| | LM4b_RS11770 | NC_012488.1:c2398198-2396882 | |
| **SSI-1** | lmo0444 | NC_003210.1:473936-476716 | *Listeria monocytogenes* EGD-e |
| | lmo0445 | NC_003210.1:476960-478447 | |
| | lmo0446 | NC_003210.1:478721-479710 | |
| | lmo0447 | NC_003210.1:479765-481153 | |
| | lmo0448 | NC_003210.1:481250-482701 | |
| **SSI-F2365** | LMOf2365_0481 | NC_002973.6:c497276-496728 | *Listeria monocytogenes* serotype 4b str. F2365 |
| **SSI-2** | lin0464 | NC_003212.1:c485702-484770 | *Listeria innocua* Clip11262 |
| | lin0465 | NC_003212.1:485841-486479 | |
| **internalins** | inlA | NC_003210.1:454534-456936 | *Listeria monocytogenes* EGD-e |
| | inlB | NC_003210.1:457021-458913 | |

**Supplementary Table 7.** Raw read statistics of *Listeria monocytogenes* dataset

| Sample Name | % Dups | % GC | Length (bp) | M Seqs | N50 | Assembled total length | Number of contigs |
|---|---|---|---|---|---|---|---|
| Lm01_S2 | 14.10 | 37.00 | 144 | 0.7 | 176005 | 3077918 | 65 |
| Lm02_S3 | 23.00 | 38.00 | 131 | 1.4 | 198288 | 3083756 | 61 |
| Lm03_S4 | 17.40 | 38.00 | 133 | 0.8 | 176761 | 3072971 | 72 |
| Lm07_S5 | 29.00 | 38.00 | 129 | 1.9 | 316022 | 2932747 | 35 |
| Lm08_S6 | 22.10 | 38.00 | 134 | 1.2 | 154847 | 3141278 | 149 |
| Lm10_S7 | 20.00 | 37.00 | 141 | 0.9 | 206397 | 3100229 | 43 |
| Lm118_S8 | 37.10 | 37.00 | 127 | 1.5 | 205814 | 3071316 | 80 |
| Lm137_S9 | 22.70 | 37.00 | 143 | 0.7 | 153514 | 3144063 | 154 |
| Lm140_S10 | 23.90 | 37.00 | 136 | 1.3 | 209235 | 3059984 | 52 |
| Lm143_S11 | 17.90 | 38.00 | 127 | 1.5 | 176090 | 3581798 | 520 |
| Lm15_S12 | 19.30 | 37.00 | 137 | 0.7 | 196611 | 3100498 | 51 |
| Lm152_S13 | 31.90 | 38.00 | 139 | 1.9 | 205922 | 3087191 | 78 |
| Lm154_S14 | 32.80 | 37.00 | 140 | 1.9 | 198060 | 3082730 | 68 |
| Lm164_S15 | 24.70 | 38.00 | 134 | 1.4 | 197475 | 3193056 | 107 |
| Lm17_S16 | 29.40 | 38.00 | 138 | 1.5 | 209279 | 3095283 | 60 |
| Lm21_S18 | 13.90 | 38.00 | 139 | 0.5 | 159732 | 3099356 | 48 |
| Lm218_S19 | 36.50 | 37.00 | 140 | 2.6 | 206795 | 3092532 | 89 |
| Lm221_S21 | 15.90 | 37.00 | 145 | 0.6 | 150064 | 2984813 | 59 |
| Lm226_S22 | 27.10 | 37.00 | 141 | 1.5 | 166211 | 3133944 | 81 |
| Lm239_S23 | 23.70 | 37.00 | 143 | 1.3 | 198783 | 3092072 | 71 |
| Lm24_S24 | 26.60 | 38.00 | 134 | 1.5 | 198773 | 3102979 | 47 |
| Lm25_S25 | 35.70 | 37.00 | 143 | 2.4 | 225112 | 3061587 | 82 |
| Lm259_S29 | 22.30 | 37.00 | 146 | 1.1 | 225112 | 3102744 | 41 |
| Lm26_S30 | 30.50 | 37.00 | 143 | 1.6 | 190687 | 3040874 | 56 |
| Lm260_S31 | 20.30 | 37.00 | 143 | 0.9 | 205815 | 3098731 | 44 |
| Lm35_S33 | 11.80 | 38.00 | 137 | 0.3 | 39001 | 3094393 | 135 |
| Lm36_S34 | 24.10 | 37.00 | 143 | 1.2 | 198080 | 3099680 | 41 |
| Lm37_S35 | 11.60 | 37.00 | 136 | 0.4 | 109483 | 3275960 | 282 |
| Lm46_S36 | 26.20 | 37.00 | 140 | 1.4 | 176091 | 3067819 | 56 |
| Lm49_S37 | 22.00 | 38.00 | 137 | 1 | 198676 | 3074786 | 57 |
| Lm73_S38 | 21.70 | 38.00 | 136 | 1.7 | 2135 | 4097116 | 2477 |
| Lmo01_S39 | 14.40 | 37.00 | 142 | 0.5 | 232746 | 2927533 | 30 |
| Lmo02_S40 | 28.50 | 38.00 | 143 | 1.6 | 225112 | 3068622 | 49 |
| Lmo03_S41 | 27.20 | 38.00 | 144 | 1.7 | 209298 | 3112795 | 82 |
| Lmo04_S42 | 26.30 | 37.00 | 144 | 1.4 | 225112 | 3015337 | 38 |
| Lmo06_S43 | 26.20 | 38.00 | 145 | 1.3 | 225111 | 3100889 | 47 |
| Lmo07_S44 | 27.00 | 37.00 | 133 | 2 | 197793 | 3079125 | 48 |
| Lmo08_S45 | 30.20 | 37.00 | 140 | 2 | 206912 | 3077970 | 57 |
| Lmo09_S46 | 26.50 | 37.00 | 143 | 1.6 | 213125 | 3134948 | 46 |
| Lmo10_S47 | 23.70 | 37.00 | 139 | 1.3 | 294218 | 3074501 | 64 |
| Lmo11_S48 | 28.40 | 37.00 | 138 | 1.8 | 209893 | 3036787 | 48 |
| Lmo12_S49 | 30.90 | 38.00 | 135 | 2.2 | 176504 | 3043946 | 38 |
| Lmo13_S50 | 29.60 | 38.00 | 147 | 1.5 | 225112 | 3034608 | 53 |
| Lmo14_S51 | 30.90 | 37.00 | 147 | 1.8 | 151921 | 3013153 | 40 |
| Lmo15_S52 | 11.40 | 37.00 | 143 | 0.5 | 125624 | 3085872 | 57 |
| Lmo16_S53 | 38.00 | 37.00 | 144 | 2.6 | 299861 | 3021184 | 27 |
| Lmo17_S54 | 33.80 | 37.00 | 144 | 2.1 | 209827 | 3033743 | 47 |

| Sample Name | % Dups | % GC | Length (bp) | M Seqs | N50 | Assembled total length | Number of contigs |
|---|---|---|---|---|---|---|---|
| Lmo18_S55 | 35.20 | 37.00 | 145 | 1.3 | 209162 | 3062985 | 60 |
| Lmo19_S56 | 35.00 | 37.00 | 142 | 2 | 198472 | 2999457 | 36 |
| Lmo20_S57 | 24.10 | 37.00 | 143 | 1.3 | 294313 | 2957225 | 28 |
| Lmo21_S58 | 25.70 | 37.00 | 144 | 1.3 | 211305 | 3034592 | 43 |
| Lmo22_S59 | 20.50 | 37.00 | 142 | 0.9 | 206131 | 3159827 | 41 |
| Lmo24_S61 | 22.10 | 39.00 | 131 | 1.3 | 225112 | 2989033 | 38 |
| Lmo25_S62 | 33.30 | 38.00 | 141 | 2.5 | 198614 | 3052780 | 51 |
| Lmo26_S63 | 39.90 | 37.00 | 144 | 1.9 | 213197 | 2947422 | 37 |
| Lmo27_S64 | 57.30 | 37.00 | 139 | 2.4 | 216541 | 2924009 | 33 |
| Lmo28_S65 | 35.00 | 37.00 | 142 | 2.1 | 176092 | 3127751 | 55 |
| Lmo29_S66 | 31.60 | 37.00 | 144 | 1.4 | 284077 | 2910179 | 26 |
| Lmo30_S67 | 29.20 | 38.00 | 139 | 1.9 | 225124 | 2966037 | 36 |
| Lmo31_S68 | 29.00 | 38.00 | 144 | 1.7 | 237768 | 2885709 | 30 |
| Lmo32_S69 | 25.50 | 37.00 | 144 | 1.2 | 225112 | 3031770 | 42 |
| Lmo33_S70 | 28.00 | 37.00 | 143 | 1.5 | 210040 | 3045758 | 47 |
| Lmo34_S71 | 28.10 | 37.00 | 143 | 1.5 | 271295 | 3022819 | 33 |
| Lmo35_S72 | 29.20 | 37.00 | 145 | 1.6 | 210689 | 3022028 | 35 |
| Lmo36_S73 | 25.70 | 38.00 | 134 | 1.4 | 211305 | 3061005 | 36 |
| Minimum | 11.4 | 37 | 127 | 0.3 | 2135 | 2885709 | 26 |
| Maximum | 57.3 | 39 | 147 | 2.6 | 316022 | 4097116 | 2477 |
| Average | 26.50 | 37.37 | 139.98 | 1.45 | 201282.89 | 3083163.12 | 103.68 |

## **Supplementary File 3.** Metadata for *Listeria monocytogenes* dataset

Available at:

https://docs.google.com/spreadsheets/d/1u2-J6ZCsZ0F-e2ELNkF7j1mAKPFRe8XB6KhdFOmfTGg/edit?usp=sharing

## **Supplementary File 4.** Genes present in reference plasmids

Available at:

https://docs.google.com/spreadsheets/d/1yfBA5RWtdaHdbUxeLurNfOXl84kMRbh5fmM7u06IGdQ/edit?usp=sharing

**Supplementary Table 8.** Similitude matrix of seven plasmids. The upper hemimatrix shows the ANI value as a percentage and the lower hemimatrix shows the size (in bp) of the fragment common to the two compared plasmids.

| **OrthoANIu value (%) - Average aligned length (bp)** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Plasmids**<br>**Size** | **pLM58**<br>58523 bp | **pl2015TE24968**<br>57530 bp | **pMF6172**<br>63182 bp | **pMF4545**<br>25550 bp | **pPIR00541**<br>15792 bp | **pPIR00540**<br>58517 bp | **pAUSMDU00000224_01**<br>57553 pb |
| **pLM58**<br>58523 bp | | 99.51% | 99.75% | 100.00% | 99.94% | 99.98% | 99.59% |
| **pl2015TE24968**<br>57530 bp | 40007 bp | | 99.85% | 99.81% | 99.86% | 99.55% | 99.99% |
| **pMF6172**<br>63182 bp | 34567 bp | 44890 bp | | 99.91% | 99.77% | 99.66% | 99.94% |
| **pMF4545**<br>25550 bp | 18819 bp | 19724 bp | 20596 bp | | 99.93% | 99.97% | 99.89% |
| **pPIR00541**<br>15792 bp | 9841 bp | 10002 bp | 10199 bp | 9743 bp | | 99.93% | 99.61% |
| **pPIR00540**<br>58517 bp | 50589 bp | 39865 bp | 37842 bp | 21094 bp | 10956 bp | | 99.60% |
| **pAUSMDU00000224_01**<br>57553 pb | 37998 bp | 34858 bp | 42765 bp | 17487 bp | 10307 bp | 34596 bp | |

**Supplementary Table 9.** List of CDS of plasmid 3 isolated from *L. monocytogenes* strain NCTC7971

| NCTC7974_plasmid3_CDS |
|---|
| 1,4-dihydroxy-2-naphthoate octaprenyltransferase |
| 1,4-dihydroxy-2-naphthoate octaprenyltransferase |
| 16S ribosomal RNA |
| 16S ribosomal RNA |
| 2-dehydro-3-deoxy-phosphogluconate aldolase |
| 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase |
| 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase |
| 2,3-bisphosphoglycerate-independent phosphoglycerate mutase |
| 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase |
| 23S ribosomal RNA |
| 23S ribosomal RNA |
| 3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ |
| 3-oxoacyl-[acyl-carrier-protein] reductase FabG |
| 3',5'-cyclic adenosine monophosphate phosphodiesterase CpdA |
| 30S ribosomal protein S10 |
| 30S ribosomal protein S11 |
| 30S ribosomal protein S12 |
| 30S ribosomal protein S13 |
| 30S ribosomal protein S14 |
| 30S ribosomal protein S17 |
| 30S ribosomal protein S19 |
| 30S ribosomal protein S3 |
| 30S ribosomal protein S5 |
| 30S ribosomal protein S7 |
| 30S ribosomal protein S8 |
| 30S ribosomal protein S9 |
| 5-keto-L-gluconate epimerase |
| 5'-deoxynucleotidase YfbR |
| 50S ribosomal protein L13 |
| 50S ribosomal protein L14 |
| 50S ribosomal protein L15 |
| 50S ribosomal protein L16 |
| 50S ribosomal protein L17 |
| 50S ribosomal protein L18 |

50S ribosomal protein L2
50S ribosomal protein L22
50S ribosomal protein L23
50S ribosomal protein L24
50S ribosomal protein L29
50S ribosomal protein L3
50S ribosomal protein L30
50S ribosomal protein L31 type B
50S ribosomal protein L34
50S ribosomal protein L36
50S ribosomal protein L4
50S ribosomal protein L5
50S ribosomal protein L6
5S ribosomal RNA
5S ribosomal RNA
6-phospho-beta-glucosidase BglA
ABC transporter ATP-binding protein YtrB
Acetyl-coenzyme A synthetase
Acid sugar phosphatase
Adaptive-response sensory-kinase SasA
Adenosyl-chloride synthase
Adenylate kinase
Aldose 1-epimerase
Alkaline phosphatase synthesis sensor protein PhoR
Alkaline phosphatase synthesis transcriptional regulatory
Alpha-galactosylglucosyldiacylglycerol synthase
Alpha-monoglucosyldiacylglycerol synthase
Aminopeptidase C
Anthranilate synthase component 2
Arginine decarboxylase
Arginine--tRNA ligase
Aryl-phospho-beta-D-glucosidase BglC
Aryl-phospho-beta-D-glucosidase BglH
Ascorbate-specific PTS system EIIA component
Ascorbate-specific PTS system EIIA component
Ascorbate-specific PTS system EIIB component
Ascorbate-specific PTS system EIIC component
Aspartokinase 3
ATP synthase epsilon chain

ATP synthase gamma chain
ATP synthase subunit a
ATP synthase subunit alpha
ATP synthase subunit b
ATP synthase subunit beta
ATP synthase subunit c
ATP synthase subunit delta
ATP-binding/permease protein CydD
ATP-dependent Clp protease proteolytic subunit
ATP-dependent DNA helicase RecQ
Beta sliding clamp
Beta-glucoside kinase
Beta-glucoside kinase
Beta-phosphoglucomutase
Cadmium, cobalt and zinc/H(+)-K(+) antiporter
Carboxylesterase
Carboxylesterase
Catabolite control protein A
Catabolite control protein B
Catalase
Cell division ATP-binding protein FtsE
Cell division protein FtsX
Cell division suppressor protein YneA
Cell division topological determinant MinJ
Cell wall-binding protein YocH
Central glycolytic genes regulator
Chromosomal replication initiator protein DnaA
Chromosome partition protein Smc
Cocaine esterase
ComF operon protein 1
Copper homeostasis protein CutC
Creatinine amidohydrolase
CRISPR with 32 repeat units
CRISPR-associated endonuclease Cas1
CRISPR-associated endonuclease Cas9
CRISPR-associated endoribonuclease Cas2
CRISPR-associated protein Csn2
CTP synthase
Cystathionine beta-lyase PatB

Cysteine desulfurase SufS

Cytochrome bd ubiquinol oxidase subunit 1

Cytochrome bd-I ubiquinol oxidase subunit 2

D-alanyl-D-alanine carboxypeptidase DacA

D-alanyl-D-alanine carboxypeptidase DacA

D-allulose-6-phosphate 3-epimerase

D-arabitol-phosphate dehydrogenase

Deoxyguanosinetriphosphate triphosphohydrolase

Diacylglycerol kinase

Dipeptide-binding protein DppE

DNA gyrase subunit A

DNA gyrase subunit B

DNA replication and repair protein RecF

DNA topoisomerase 3

DNA-directed RNA polymerase subunit alpha

DNA-directed RNA polymerase subunit delta

Elongation factor G

Elongation factor Tu

Energy-coupling factor transporter ATP-binding protein EcfA1

Energy-coupling factor transporter ATP-binding protein EcfA2

Energy-coupling factor transporter transmembrane protein EcfT

Energy-coupling factor transporter transmembrane protein EcfT

Energy-dependent translational throttle protein EttA

Enolase

Erythronate-4-phosphate dehydrogenase

FAD:protein FMN transferase

Fe/S biogenesis protein NfuA

Ferredoxin--NADP reductase 2

Ferrous-iron efflux pump FieF

FMN reductase (NADPH)

Fructose-bisphosphate aldolase

Galactitol 1-phosphate 5-dehydrogenase

Galactokinase

General stress protein 13

General stress protein 26

GlcNAc-binding protein A

Glucitol operon repressor

Gluconeogenesis factor

Glucosamine-6-phosphate deaminase 1

Glucosamine-6-phosphate deaminase 1

Glucose-6-phosphate isomerase

Glucosylglycerate phosphorylase

Glucosylglycerate phosphorylase

Glutamate decarboxylase

Glutamate decarboxylase

Glutamate/gamma-aminobutyrate antiporter

Glutamine--fructose-6-phosphate aminotransferase [isomerizing]

Glutathione biosynthesis bifunctional protein GshAB

Glyceraldehyde-3-phosphate dehydrogenase

Glycerate 2-kinase

Glycerate 2-kinase

Glycine cleavage system H protein

Glyoxal reductase

Glyoxal reductase

Heme response regulator HssR

Heme sensor protein HssS

Heptaprenyl diphosphate synthase component 2

High-affinity heme uptake system protein IsdE

Holliday junction ATP-dependent DNA helicase RuvB

Homoserine dehydrogenase

Homoserine kinase

HPr kinase/phosphorylase

HTH-type transcriptional activator RhaR

HTH-type transcriptional activator RhaS

HTH-type transcriptional regulator AdhR

HTH-type transcriptional regulator AdhR

HTH-type transcriptional regulator ArgP

HTH-type transcriptional regulator GlvR

HTH-type transcriptional regulator HexR

HTH-type transcriptional repressor CzrA

HTH-type transcriptional repressor PurR

IMPACT family member YigZ

Inner membrane ABC transporter permease protein YcjO

Inner membrane ABC transporter permease protein YcjP

Inosine-5'-monophosphate dehydrogenase

Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase

Internalin-J

Internalin-J

Iron-sulfur cluster carrier protein

IS1595 family transposase ISCac2

Isochorismate synthase MenF

KDP operon transcriptional regulatory protein KdpE

Kojibiose phosphorylase

L-cystine import ATP-binding protein TcyN

L-cystine transport system permease protein TcyL

L-cystine transport system permease protein TcyM

L-cystine-binding protein TcyK

L-rhamnose isomerase

L-rhamnose mutarotase

L-Rhamnulokinase

L-threonine 3-dehydrogenase

Lactose phosphotransferase system repressor

Leucine-rich protein

Lichenan permease IIC component

Lichenan permease IIC component

Lichenan permease IIC component

Lichenan-specific phosphotransferase enzyme IIA component

Lichenan-specific phosphotransferase enzyme IIA component

Lichenan-specific phosphotransferase enzyme IIB component

Lipoyl-[GcvH]:protein N-lipoyltransferase

Magnesium-transporting ATPase, P-type 1

Major cardiolipin synthase ClsA

Major cardiolipin synthase ClsA

Mannitol-specific phosphotransferase enzyme IIA component

Mannosylglucosyl-3-phosphoglycerate phosphatase

Mannosylglycerate hydrolase

Membrane protein insertase MisCA

Membrane protein YdfJ

Methionine import ATP-binding protein MetN

Methionine import system permease protein MetP

Methionine-binding lipoprotein MetQ

MreB-like protein

Multidrug export protein EmrB

Multidrug export protein MepA

Multidrug resistance ABC transporter ATP-binding/permease protein BmrA

Multidrug resistance protein D

Myo-inositol 2-dehydrogenase

N-acetyldiaminopimelate deacetylase

N-acetylgalactosamine permease IIC component 1

N-acetylglucosaminyldiphosphoundecaprenol N-acetyl-beta-D-mannosaminyltransferase

N-acetylmannosamine kinase

N-substituted formamide deformylase

Na(+)/H(+) antiporter subunit A

Na(+)/H(+) antiporter subunit B

Na(+)/H(+) antiporter subunit C

Na(+)/H(+) antiporter subunit D1

Na(+)/H(+) antiporter subunit E

Na(+)/H(+) antiporter subunit F

NAD-dependent protein deacetylase

NADH dehydrogenase-like protein

NADPH dehydrogenase

NADPH dehydrogenase

Nucleoid occlusion protein

Nucleoid-associated protein

Nucleotide-binding protein YvcJ

O-acetyl-ADP-ribose deacetylase

o-succinylbenzoate synthase

Peptide chain release factor 1

Peptide chain release factor 2

Peptidoglycan glycosyltransferase RodA

Peptidoglycan glycosyltransferase RodA

Peptidoglycan glycosyltransferase RodA

Phosphatase YwpJ

Phosphate import ATP-binding protein PstB 3

Phosphate import ATP-binding protein PstB 3

Phosphate transport system permease protein PstA

Phosphate-binding protein PstS 1

Phosphate-specific transport system accessory protein PhoU

Phospho-furanose lactonase

Phosphoglucomutase

Phosphoglycerate kinase

Phosphoglycolate phosphatase

Phosphomevalonate kinase

Phosphoserine aminotransferase

Polyisoprenyl-teichoic acid--peptidoglycan teichoic acid

Potassium-transporting ATPase ATP-binding subunit

Potassium-transporting ATPase KdpC subunit

Potassium-transporting ATPase potassium-binding subunit

Prolipoprotein diacylglyceryl transferase

Protein CbbY

Protein DegV

Protein SapB

Protein translocase subunit SecA

Protein translocase subunit SecY

Protein UmuC

Protein-arginine-phosphatase

Pseudouridine kinase

Pseudouridine-5'-phosphate glycosidase

PTS system 2-O-alpha-mannosyl-D-glycerate-specific EIIABC component

PTS system beta-glucoside-specific EIIBCA component

PTS system beta-glucoside-specific EIIBCA component

PTS system beta-glucoside-specific EIIBCA component

PTS system cellobiose-specific EIIB component

PTS system fructose-specific EIIABC component

PTS system fructose-specific EIIB component

PTS system galactitol-specific EIIB component

PTS system galactitol-specific EIIC component

PTS system mannitol-specific EIICB component

PTS system mannose-specific EIID component

PTS system N,N'-diacetylchitobiose-specific EIIC componen

PTS system oligo-beta-mannoside-specific EIIA component

PTS system oligo-beta-mannoside-specific EIIB component

PTS system oligo-beta-mannoside-specific EIIB component

PTS system oligo-beta-mannoside-specific EIIC component

PTS system sorbose-specific EIIA component

PTS-dependent dihydroxyacetone kinase, ADP-binding subunit DhaL

PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit DhaK

PTS-dependent dihydroxyacetone kinase, phosphotransferase

putative ABC transporter ATP-binding protein

putative ABC transporter ATP-binding protein

putative ABC transporter ATP-binding protein

putative ABC transporter permease

putative ABC transporter permease

Putative agmatine deiminase

putative amino acid permease YhdG

Putative esterase

putative FMN/FAD exporter YeeO

Putative formate dehydrogenase

putative fructoselysine utilization operon transcriptional repressor

putative glutamate/gamma-aminobutyrate antiporter

putative glycosyltransferase

Putative hemin import ATP-binding protein HrtA

Putative hemin import ATP-binding protein HrtA

putative HTH-type transcriptional regulator YbbH

putative hydrolase YxeP

putative licABCH operon regulator

putative metabolite transport protein CsbC

Putative metabolite transport protein YjhB

Putative monooxygenase MoxC

Putative monooxygenase YcnE

putative murein peptide carboxypeptidase

Putative N-acetylmannosamine-6-phosphate 2-epimerase

putative N-acetyltransferase YtmI

putative peptidoglycan glycosyltransferase FtsW

Putative peptidyl-prolyl cis-trans isomerase

putative protein YaaQ

putative protein YpjQ

putative protein YutD

putative protein YwqG

putative protein-export membrane protein SecG

putative siderophore transport system ATP-binding protein YusV

putative siderophore transport system permease protein YfhA

Putative sporulation transcription regulator WhiA

putative sugar epimerase YhfK

Putative sugar phosphate isomerase YwlF

putative sulfoacetate transporter SauU

putative tautomerase.1

putative transport protein HsrA

putative undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate

Putative universal stress protein

Putrescine carbamoyltransferase

Pyridoxal phosphate phosphatase YbhA

Pyrophosphatase PpaX

Quinol oxidase subunit 1

Quinol oxidase subunit 3

Quinol oxidase subunit 4

Quinolone resistance protein NorB

Recombination protein RecR

Regulatory protein MgsR

Release factor glutamine methyltransferase

Rhamnulose-1-phosphate aldolase

Ribonuclease P protein component

Ribonuclease R

Ribose-5-phosphate isomerase B

Ribosomal RNA small subunit methyltransferase G

Ribosomal small subunit pseudouridine synthase A

Ribosome hibernation promotion factor

Ribosome-binding ATPase YchF

Ribulose-phosphate 3-epimerase

RNA polymerase sigma-54 factor

Sensor protein KdpD

Serine hydroxymethyltransferase

Serine--tRNA ligase

Single-stranded DNA-binding protein B

Spermidine N(1)-acetyltransferase

Spermine/spermidine acetyltransferase

Sporulation initiation inhibitor protein Soj

SsrA-binding protein

Stage 0 sporulation protein J

Sugar phosphatase YidA

Sugar phosphatase YidA

Sulfur carrier protein FdhD

Tagatose-6-phosphate kinase

Tetracycline resistance protein, class B

Thermostable beta-glucosidase B

Thioredoxin 1

Thioredoxin reductase

Threonine synthase

Threonylcarbamoyl-AMP synthase

Thymidine kinase

Thymidylate kinase

Transaldolase

Transcription antiterminator LicT

Transcription antiterminator LicT

Transcription antiterminator LicT

Transcription termination factor Rho

Transcriptional regulator ManR

Transcriptional regulator SlyA

Transcriptional regulator SlyA

Transcriptional regulatory protein DegU

Transcriptional regulatory protein SrrA

transfer-messenger RNA, SsrA

Transketolase

Translation initiation factor IF-1

Triosephosphate isomerase

tRNA modification GTPase MnmE

tRNA pseudouridine synthase A

tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG

tRNA-Ala(tgc)

tRNA-Ala(tgc)

tRNA-Arg(ccg)

tRNA-Asn(gtt)

tRNA-Asn(gtt)

tRNA-Asp(gtc)

tRNA-Cys(gca)

tRNA-Gln(ttg)

tRNA-Glu(ttc)

tRNA-Gly(gcc)

tRNA-His(gtg)

tRNA-Ile(gat)

tRNA-Ile(gat)

tRNA-Leu(caa)

tRNA-Met(cat)

tRNA-Phe(gaa)

tRNA-Ser(gga)

tRNA-specific adenosine deaminase

tRNA-Thr(ggt)

tRNA-Trp(cca)

tRNA-Tyr(gta)

tRNA-Val(gac)

tRNA-Val(tac)

UDP-glucose 4-epimerase

UDP-N-acetylglucosamine 1-carboxyvinyltransferase 1

UDP-N-acetylglucosamine 1-carboxyvinyltransferase 2

UDP-N-acetylglucosamine 2-epimerase

Uracil phosphoribosyltransferase

UvrABC system protein A

UvrABC system protein B

Vegetative protein 296

Xylulose kinase

Zinc-dependent sulfurtransferase SufU

Zinc-type alcohol dehydrogenase-like protein

## Supplementary material for Chapter 3

**Supplementary Table 10.** Recombinant genes among six pathogenic *Leptospira* species

| Protein name | Gene | EC number | COG | LRT, *p*-value | Matching Split |
|---|---|---|---|---|---|
| 4,5-DOPA dioxygenase extradiol | ygiD | 1.13.11.29 | COG3384 | | 6 |
| 50S ribosomal protein L10 | rplJ | | COG0244 | 0 | 10 |
| Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha | acoA | 1.1.1.- | COG1071 | | 12 |
| Alanine--tRNA ligase | alaS | 6.1.1.7 | COG0013 | 0.021 | 8 |
| Capreomycidine synthase | vioD | 4.2.1.145 | | 0.012 | 4 |
| Chondroitin synthase | | | | | 4 |
| Coenzyme A biosynthesis bifunctional protein CoaBC | coaBC_1 | | COG0452 | | 9 |
| Cysteine--tRNA ligase | cysS | 6.1.1.16 | COG0215 | | 4 |
| Dipeptide permease D | dtpD | | COG3104 | | 4 |
| DNA-directed RNA polymerase subunit beta' | rpoC | 2.7.7.6 | COG0086 | | 7 |
| Elongation factor 4 | lepA | 3.6.5.- | COG0481 | | 10 |
| Elongation factor G 1 | fusA_2 | | | | 4 |
| Glucose-1-phosphate thymidylyltransferase 1 | rfbA | 2.7.7.24 | COG1209 | | 4 |
| Glyceraldehyde-3-phosphate dehydrogenase 1 | gap1 | 1.2.1.12 | COG0057 | | 4 |
| GTPase Der | der | | COG1160 | | 7 |
| Hypothetical protein (*n*=9) | | | | | 4-13 |
| Menaquinone reductase, iron-sulfur cluster-binding subunit | qrcC | 1.97.-.- | COG0437 | 0.006 | 4 |
| N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase | wbbL | 2.4.1.289 | COG1216 | | 7 |
| NADH-quinone oxidoreductase subunit 1 | nqo1 | 1.6.5.11 | COG1894 | | 4 |
| Outer-membrane lipoprotein carrier protein | lolA_1 | | | | 7 |
| Protein QmcA | qmcA_1 | | COG0330 | | 4 |
| putative succinyl-diaminopimelate desuccinylase | dapE | 3.5.1.18 | | | 7 |
| Rhamnosyltransferase WbbL | WbbL | | | | 7 |
| Sulfurtransferase | | 2.8.1.- | COG0607 | | 6 |
| Threonine--tRNA ligase 2 | thrZ | 6.1.1.3 | COG0441 | | 7 |
| UDP-N-acetylglucosamine 1-carboxyvinyltransferase | murA | 2.5.1.7 | COG0766 | | 4 |

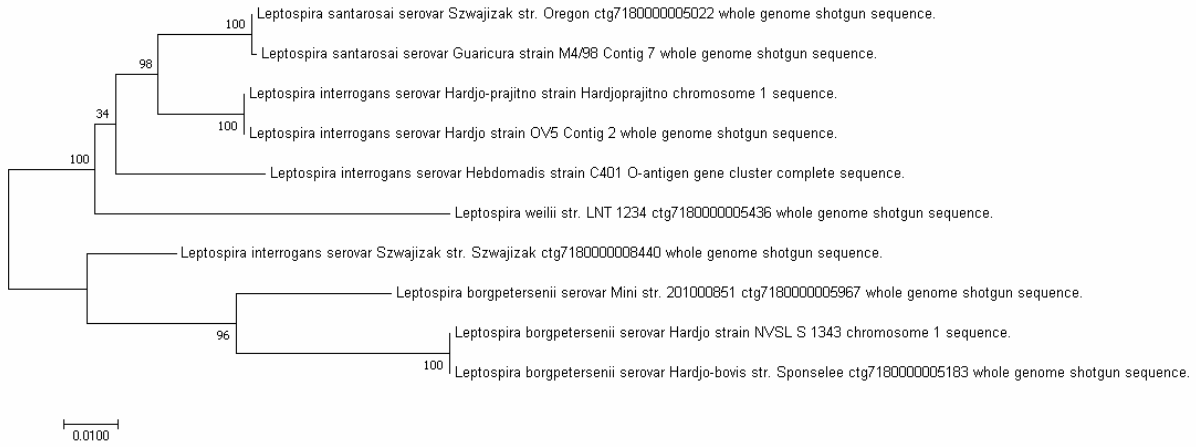
**Supplementary Table 11.** Recombinant genes among *L. interrogans*

| Protein name | Gene | EC number | COG |
|---|---|---|---|
| 1-deoxy-D-xylulose-5-phosphate synthase | dxs_3 | 2.2.1.7 | COG1154 |
| 2-dehydro-3-deoxy-L-rhamnonate dehydrogenase (NAD(+)) | | 1.1.1.401 | COG1028 |
| 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase | menH_4 | 4.2.99.20 | COG0596 |
| 23S rRNA (uracil-C(5))-methyltransferase RlmCD | rlmCD_1 | 2.1.1.189 | COG2265 |
| 30S ribosomal protein S11 | rpsK | | COG0100 |
| 30S ribosomal protein S9 | rpsI | | COG0103 |
| 4-hydroxybenzoate octaprenyltransferase | ubiA | 2.5.1.- | COG0382 |
| 50S ribosomal protein L4 | rplD | | COG0088 |
| 50S ribosomal protein L7/L12 | rplL | | COG0222 |
| 60 kDa chaperonin | groL | | COG0459 |
| Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha | acoA | 1.1.1.- | COG1071 |
| Acetyl-coenzyme A synthetase | acsA_2 | 6.2.1.1 | COG0365 |
| Acetylglutamate kinase | argB | 2.7.2.8 | COG0548 |
| ADP-heptose--LPS heptosyltransferase 2 | rfaF_2 | 2.-.-.- | COG0859 |
| Alanine--tRNA ligase | alaS | 6.1.1.7 | COG0013 |
| Alpha-maltose-1-phosphate synthase | glgM_2 | 2.4.1.342 | COG0438 |
| Anaerobic magnesium-protoporphyrin IX monomethyl ester cyclase | bchE | 1.21.98.3 | COG1032 |
| Arabinose 5-phosphate isomerase KdsD | kdsD_1 | 5.3.1.13 | COG0517 |
| Arginine--tRNA ligase | argS | 6.1.1.19 | COG0018 |
| Asparagine--tRNA ligase | asnS | 6.1.1.22 | COG0017 |
| Aspartate kinase Ask_LysC | lysC | 2.7.2.4 | COG0527 |
| ATP synthase subunit alpha | atpA | 3.6.3.14 | COG0056 |
| ATP-dependent zinc metalloprotease FtsH | ftsH | 3.4.24.- | COG0465 |
| Bifunctional folate synthesis protein | sulD | | COG0801 |
| Bifunctional methionine biosynthesis protein MetXA/MetW | metXA_2 | | COG0500 |
| Bifunctional protein HldE | hldE | | COG2870 |
| Bifunctional purine biosynthesis protein PurH | purH | | COG0138 |
| Capreomycidine synthase | vioD | 4.2.1.145 | |
| Carbonic anhydrase 2 | can | 4.2.1.1 | COG0288 |
| Chaperone protein ClpB | clpB | | COG0542 |
| Coenzyme A biosynthesis bifunctional protein CoaBC | coaBC_1 | | COG0452 |
| Cyclic di-GMP phosphodiesterase PdeL | pdeL | 3.1.4.52 | COG2200 |
| Cyclic-di-AMP phosphodiesterase PgpH | pgpH | 3.1.4.- | COG1480 |
| DEAD-box ATP-dependent RNA helicase CshE | cshE | 3.6.4.13 | COG0513 |
| Dihydrofolate synthase/folylpolyglutamate synthase | folC | 6.3.2.12 | COG0285 |
| Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex | pdhC | 2.3.1.12 | COG0508 |
| Dihydropteroate synthase | folP | 2.5.1.15 | COG0294 |
| DNA damage-responsive serine/threonine-protein kinase RqkA | rqkA | 2.7.11.1 | COG0515 |
| DNA polymerase I | polA | 2.7.7.7 | COG0258 |

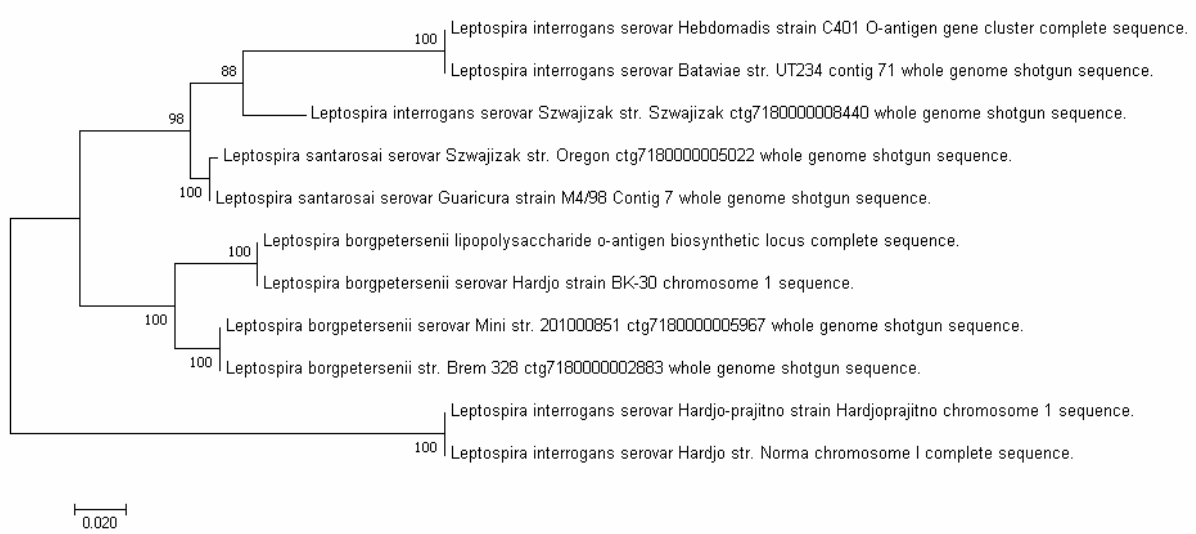
| Protein name | Gene | EC number | COG |
|---|---|---|---|
| DNA replication protein DnaC | dnaC_2 | | COG1484 |
| DNA topoisomerase 1 | topA | 5.99.1.2 | COG3569 |
| DNA-3-methyladenine glycosylase | alkA_1 | 3.2.2.21 | COG0122 |
| ECF RNA polymerase sigma-E factor | rpoE_2 | | COG1595 |
| Elongation factor Tu | tufA | | COG0050 |
| Exodeoxyribonuclease 7 large subunit | xseA | 3.1.11.6 | COG1570 |
| FK506-binding protein | fbp_1 | 5.2.1.8 | COG0545 |
| Flagellar filament 35 kDa core protein | flaB_4 | | COG1344 |
| Flagellar motor switch protein FliG | fliG_3 | | COG1536 |
| Formate hydrogenlyase subunit 7 | hycG | | COG3260 |
| Galactokinase | galK | 2.7.1.6 | COG0153 |
| Gamma-glutamyl phosphate reductase | proA | 1.2.1.41 | COG0014 |
| GDP-L-fucose synthase | fcl | 1.1.1.271 | COG0451 |
| Glycerol-3-phosphate dehydrogenase [NAD(P)+] | gpsA_3 | 1.1.1.94 | COG0240 |
| Glycine--tRNA ligase | glyQS | 6.1.1.14 | COG0423 |
| GTP cyclohydrolase 1 | folE | 3.5.4.16 | COG0302 |
| GTP pyrophosphokinase | relA | 2.7.6.5 | COG0317 |
| Histone deacetylase-like amidohydrolase | hdaH | 3.5.1.- | COG0123 |
| Homoserine O-acetyltransferase | metXA_1 | 2.3.1.31 | COG2021 |
| Hydroxyethylthiazole kinase | thiM | 2.7.1.50 | COG2145 |
| Hypothetical proteins (*n*=50) | | | |
| Inner membrane transport permease YadH | yadH_1 | | COG0842 |
| L-aspartate oxidase | nadB | 1.4.3.16 | COG0029 |
| Lon protease | lon_1 | 3.4.21.53 | COG0466 |
| Membrane protein insertase YidC | yidC | | COG0706 |
| Methyl-accepting chemotaxis protein McpU | mcpU_2 | | COG0840 |
| Multidrug resistance protein MdtC | mdtC_3 | | COG0841 |
| Multifunctional alkaline phosphatase superfamily protein PehA | pehA_2 | 3.1.4.- | |
| N-acetylgalactosamine-N,N'-diacetylbacillosaminyl-diphospho-undecaprenol 4-alpha-N-acetylgalactosaminyltransferase | pglJ | 2.4.1.291 | COG0438 |
| NAD(P) transhydrogenase subunit beta | pntB | 1.6.1.2 | COG1282 |
| NADH-quinone oxidoreductase subunit 4 | nqo4 | 1.6.5.11 | COG0649 |
| NADH-quinone oxidoreductase subunit M | nuoM | 1.6.5.11 | COG1008 |
| Oligopeptide transport ATP-binding protein OppD | oppD_1 | | COG0444 |
| Omega-amidase YafV | yafV | 3.5.1.3 | COG0388 |
| Orotate phosphoribosyltransferase | pyrE | 2.4.2.10 | COG0461 |
| Pantothenate synthetase | panC | 6.3.2.1 | COG0414 |
| Penicillin-binding protein 1A | mrcA_1 | | COG5009 |
| Penicillin-binding protein 1A | mrcA_3 | | COG5009 |
| Peptide transporter CstA | cstA | | COG1966 |
| Peptidoglycan glycosyltransferase MrdB | mrdB | 2.4.1.129 | COG0772 |
| Peptidoglycan O-acetyltransferase | patA_5 | 2.3.1.- | COG1696 |
| Phenylalanine--tRNA ligase beta subunit | pheT | 6.1.1.20 | COG0072 |
| Polyphosphate kinase | ppk | 2.7.4.1 | COG0855 |
| Propionyl-CoA carboxylase beta chain | pccB | 6.4.1.3 | COG4799 |
| Putative ABC transporter ATP-binding protein YadG | yadG_1 | | COG1131 |
| Putative CtpA-like serine protease | | 3.4.21.- | |
| Putative cysteine desulfurase | csd | 2.8.1.7 | COG1104 |

| Protein name | Gene | EC number | COG |
|---|---|---|---|
| Putative D,D-dipeptide-binding periplasmic protein DdpA | ddpA | | COG0747 |
| Putative iron-sulfur-binding oxidoreductase FadF | fadF | 1.-.-.- | COG0247 |
| Putative O-methyltransferase | | 2.1.1.- | COG4122 |
| putative peptidyl-prolyl cis-trans isomerase | ppiB | 5.2.1.8 | COG0652 |
| putative ribonucleotide transport ATP-binding protein mkl | mkl | | COG1127 |
| Putative serine protease HtrA | htrA | | COG0265 |
| Putative teichuronic acid biosynthesis glycosyltransferase TuaG | tuaG | 2.4.-.- | COG0463 |
| Putative transcriptional regulatory protein | | | COG0217 |
| Putative transport protein | | | COG0628 |
| Pyruvate kinase | pyk | 2.7.1.40 | COG0469 |
| RecBCD enzyme subunit RecB | recB | 3.1.11.5 | COG1074 |
| Ribitol-5-phosphate cytidylyltransferase | tarI | 2.7.7.40 | |
| Riboflavin biosynthesis protein RibBA | ribBA | | COG0108 |
| Ribonuclease BN | rbn_1 | 3.1.26.11 | COG1234 |
| Ribonuclease R | rnr | 3.1.13.1 | COG0557 |
| Ribosomal large subunit pseudouridine synthase C | rluC | 5.4.99.24 | COG0564 |
| RNA polymerase sigma factor RpoD | rpoD | | COG0568 |
| Shikimate dehydrogenase (NADP(+)) | aroE | 1.1.1.25 | COG0169 |
| Sigma factor SigB regulation protein RsbQ | rsbQ_1 | | |
| Small ribosomal subunit biogenesis GTPase RsgA | rsgA_1 | 3.6.1.- | COG1162 |
| Sporulation initiation inhibitor protein Soj | soj_1 | 3.6.-.- | COG1192 |
| SsrA-binding protein | smpB | | COG0691 |
| Sulfate transport system permease protein CysW | cysW | | COG4208 |
| Sulfate-binding protein | sbp_1 | | COG1613 |
| Sulfite reductase [ferredoxin] | sir | 1.8.7.1 | COG0155 |
| Transketolase 2 | tktB | 2.2.1.1 | COG0021 |
| Tryptophan synthase beta chain | trpB | 4.2.1.20 | COG0133 |
| Type II secretion system protein G | xcpT | | COG2165 |
| UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1,3-glucosyltransferase WfgD | wfgD | 2.4.1.305 | COG0463 |
| UDP-N-acetylglucosamine 1-carboxyvinyltransferase | murA | 2.5.1.7 | COG0766 |
| UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase | murE | 6.3.2.13 | COG0769 |
| Undecaprenyl phosphate-alpha-4-amino-4-deoxy-L-arabinose arabinosyl transferase | arnT | 2.4.2.43 | COG1807 |
| Vitamin B12-binding protein | btuF | | COG0614 |
| 2,3-bisphosphoglycerate-independent phosphoglycerate mutase | gpmI_1 | 5.4.2.12 | COG3635 |

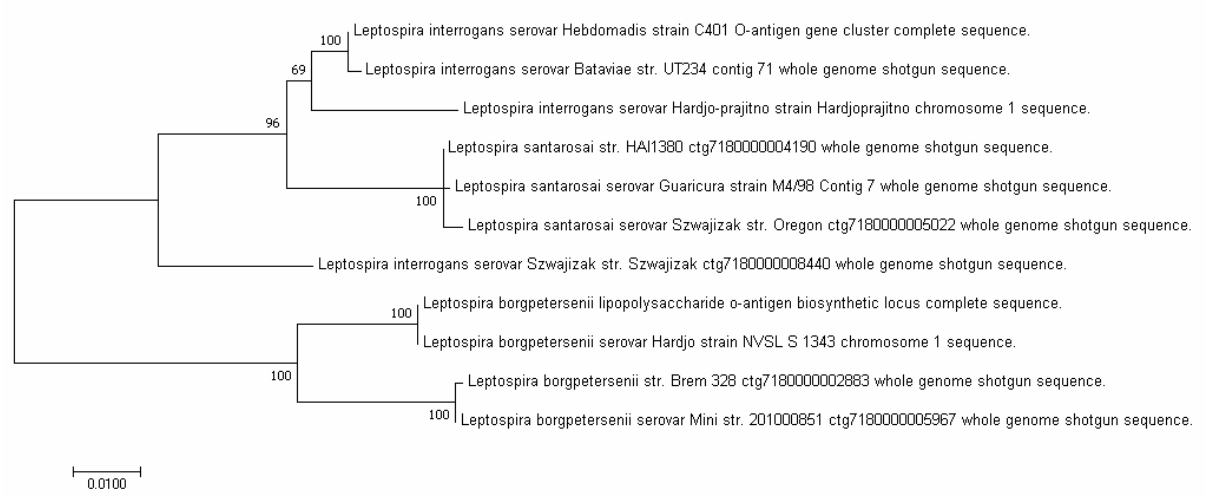
**Supplementary Figure 3.** Maximum-likelihood of ORF2 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans, santarosai,* and *weilii.*



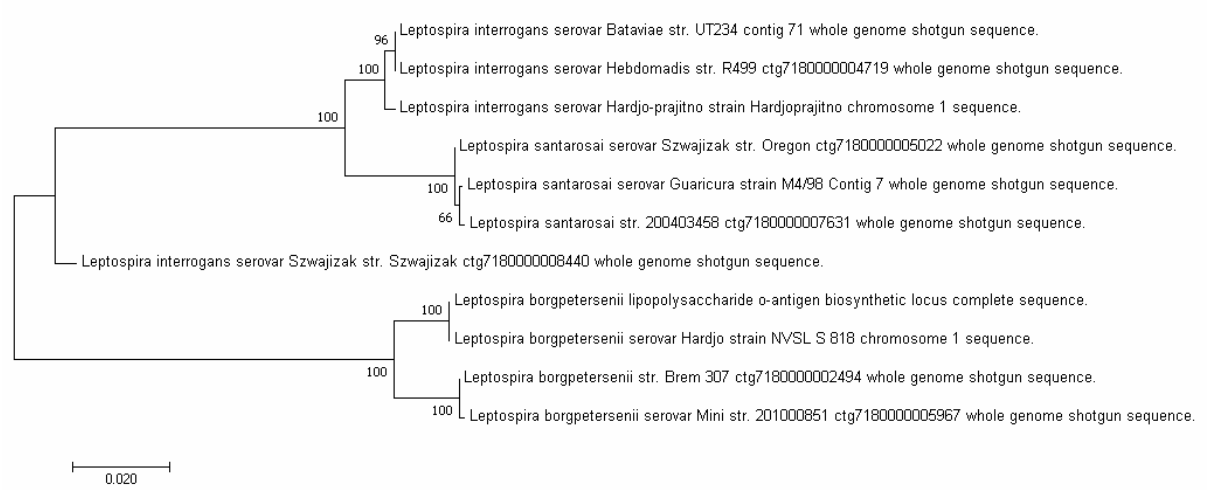
**Supplementary Figure 4.** Maximum-likelihood of ORF3 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



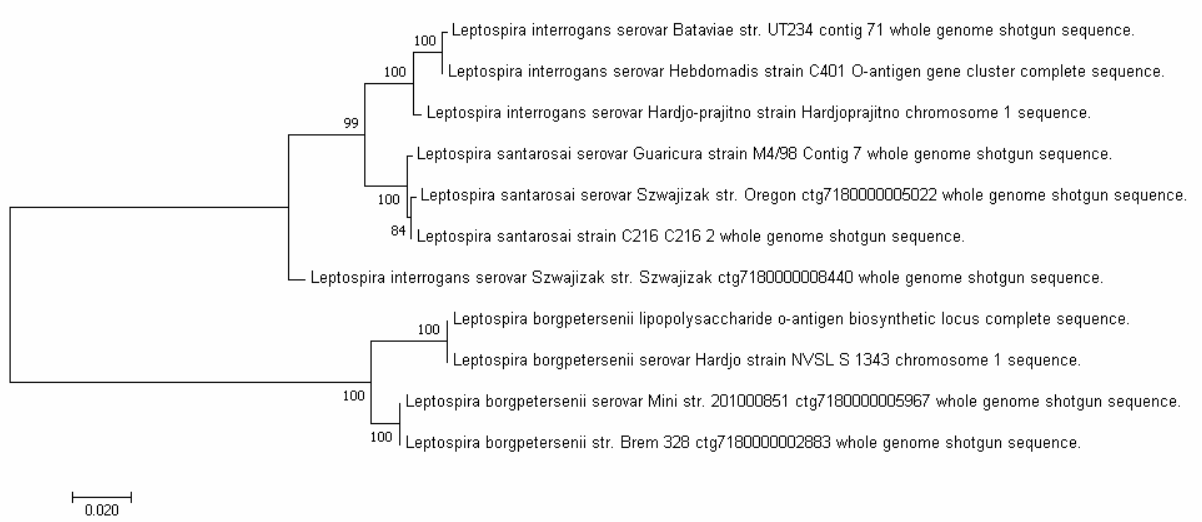
**Supplementary Figure 5.** Maximum-likelihood of ORF5 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



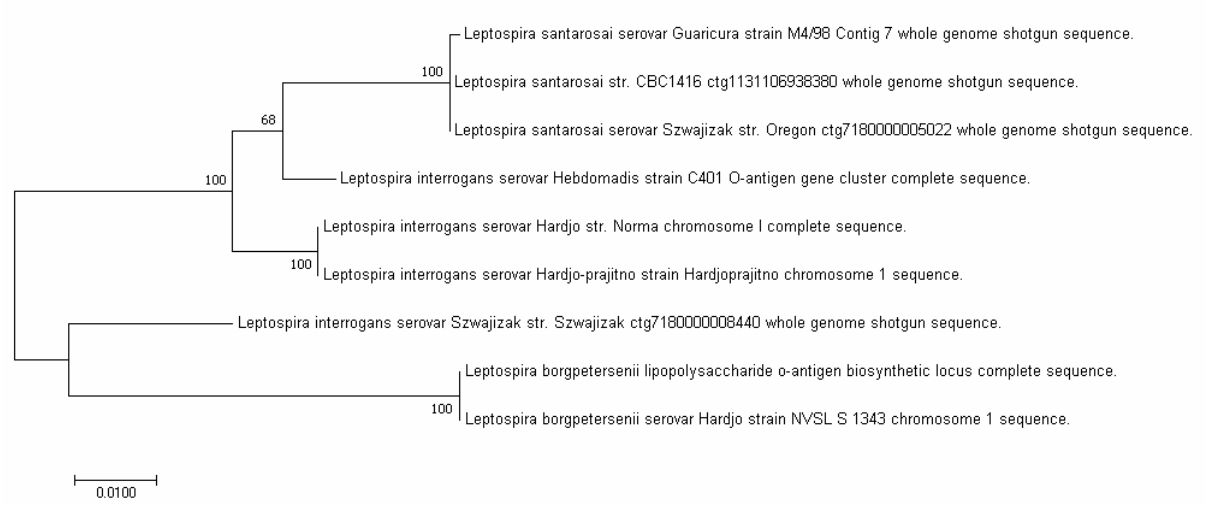
**Supplementary Figure 6.** Maximum-likelihood of ORF6 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



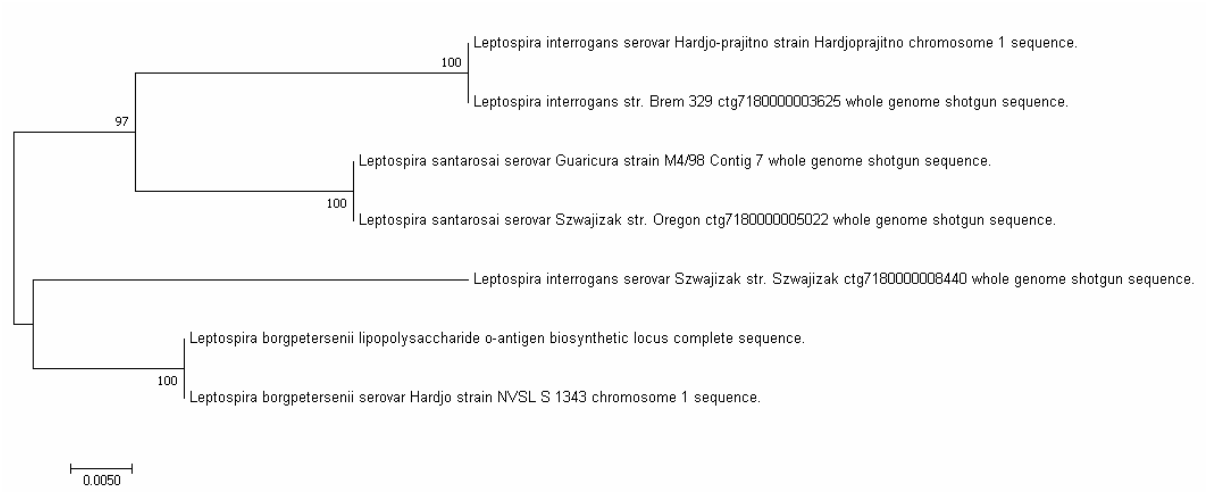
**Supplementary Figure 7.** Maximum-likelihood of ORF7 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



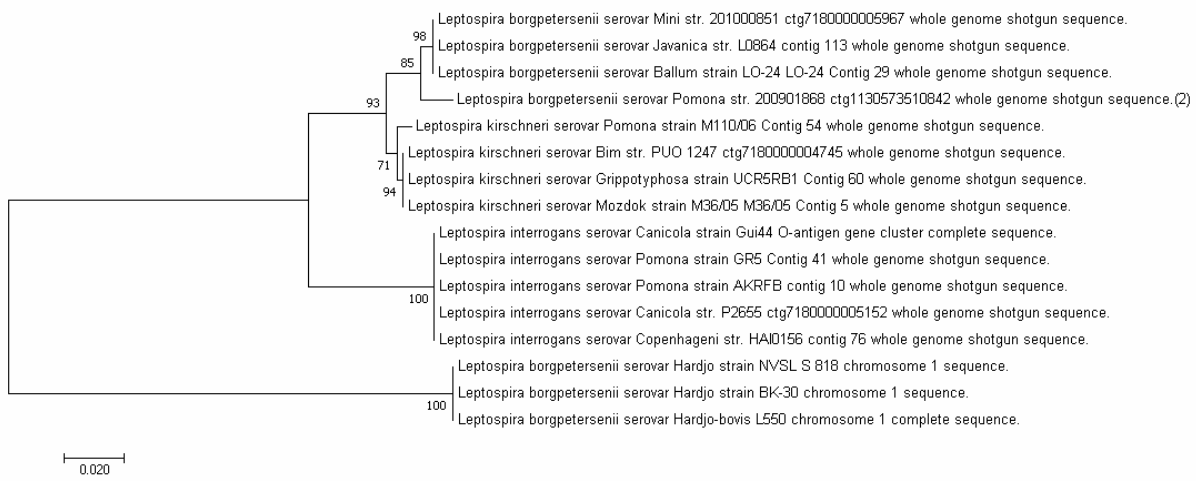
**Supplementary Figure 8.** Maximum-likelihood of ORF8 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



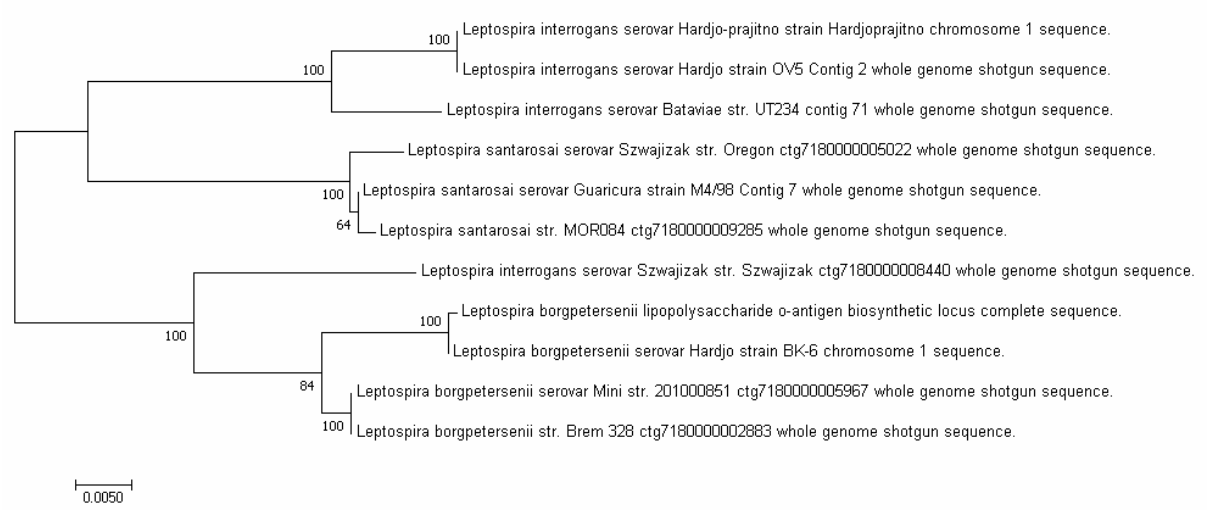
**Supplementary Figure 9.** Maximum-likelihood of ORF9 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



**Supplementary Figure 10.** Maximum-likelihood of ORF13 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *kirschneri.*



**Supplementary Figure 11.** Maximum-likelihood of ORF14 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *santarosai.*



**Supplementary Figure 12.** Maximum-likelihood of ORF19 from LPS locus of different serovars of *Leptospira borgpetersenii, interrogans,* and *kirschneri.*