Infection, Genetics and Evolution 83 (2020) 104326

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Genetics and physiology of *Salmonella houtenae* isolated from a river in Mexico provides insight into the aquatic habitat influence on its adaptation and pathogenesis

Gloria M. Castañeda-Ruelas, Areli Burgeño-Román, Maribel Jiménez-Edeza*

Laboratorio de Investigación y Diagnóstico Microbiológico, Programa Regional de Posgrado en Biotecnología, Facultad de Ciencias Químico-Biológicas, Universidad Autónoma de Sinaloa, Culiacán, Sinaloa, Mexico

ARTICLE INFO

Keywords:
Genome
Pathogenicity
Resistance
River
Salmonella enterica houtenae

ABSTRACT

Salmonella enterica subsp. *houtenae* (IV) is a non-enteric subspecies of the genus *Salmonella* and has recently been implicated in extraintestinal diseases in humans. In Mexico, its reported that rivers are a reservoir of *Salmonella houtenae*, however, detailed information about the virulence and infective capacity of this bacterium are limited. Here, we present the high-quality genome draft of *Salmonella houtenae* str. CFSAN039533 isolated from a river in Culiacan, Mexico. In addition, its virulence, antimicrobial resistance profile and the use of carbon sources for its primary metabolism are analyzed. The genome sequence of CFSAN03953 strain comprises 4.74-Mb which contains 4266 protein coding genes and 77 tRNA genes. The strain belongs to the Marine serotype (48: g, z51: -), and showed several virulence genes related to the SPI-1, SPI-2, SPI-3, SPI-5 and fimbriae genes. The strain exhibited a susceptible phenotype, which contrasted with the resistance genes of aminoglycosides family identified into the genome. Carbon source degradation evaluation along with the genomic analysis, indicated that the strain can use a relatively broad spectrum of compounds related to the central metabolic pathways. This study provides information on the genetics and physiology of *Salmonella houtenae* that confirms its ability to survive and adapt to the environment. In addition, its pathogenic potential and infective capacity towards warm-blooded hosts are demonstrated.

1. Background

The genus *Salmonella* belongs to the bacterial family *Enterobacteriaceae* which comprises gram-negative, mobile, non-sporulated and facultative anaerobes bacilli (Pui et al., 2011). Based on the 16S rRNA sequence and biochemical analysis, *Salmonella* is divided into two species with 2579 serotypes: 2557 and 22 serotypes for *S. enterica* and *S. bongori*, respectively (Lamas et al., 2018). Similarly, *S. enterica* is further divided into six subspecies denoted by Roman numerals: *S. enterica enterica* (I), *S. enterica salamae* (II), *S. enterica arizonae* (IIIa), *S. enterica diarizonae* (IIIb), *S. enterica houtenae* (IV) and *S. enterica indica* (VI) (Su and Chiu, 2007).

Salmonella causes an annual estimate of salmonellosis cases of > 179 million, including diarrhea and invasive enteric diseases; In addition, it causes about 300 thousand deaths worldwide, which is why it is considered a major public health problem (Kirk et al., 2015). Most *Salmonella* diseases are linked to a wide variety of serotypes of *S. enterica* subsp. I, being its main route of dissemination contaminated food and water (Chen et al., 2013). However, the participation of *Salmonella* subsp. II-VI in cases of atypical diseases in humans, has recently been described (Editorial team collective et al., 2008; Abbott et al., 2012).

Likewise, the emergence of multiple antimicrobial resistant *Salmonella* strains used clinically for prophylaxis is frequently reported, therefore current prevention and treatment strategies have become a priority (Chen et al., 2013). In Mexico, serotypes of non-typhoid *Salmonella* have been isolated from rivers, which exhibited phenotypes and genotypes of antimicrobial resistance against b-lactams, tetracyclines, sulfonamides, trimethoprim, quinolones, aminoglycosides, cephalosporins, among other drugs (Castañeda-Ruelas and Jiménez-Edeza, 2018; Burgueño-Roman et al., 2019). The antimicrobial resistance is a common phenomenon of *Salmonella* conferred by the presence and exchange of plasmids and mobile genes. The presence of genomics

* Corresponding author at: Laboratorio de Investigación y Diagnóstico Microbiológico, Programa Regional de Posgrado en Biotecnología, Facultad de Ciencias Químico-Biológicas, Universidad Autónoma de Sinaloa, Blvd. de las Américas and Josefa Ortiz de Domínguez S/N, Ciudad Universitaria, 80013 Culiacán, Sinaloa, Mexico.

E-mail address: mjimeneze@uas.edu.mx (M. Jiménez-Edeza).

https://doi.org/10.1016/j.meegid.2020.104326
Received 17 January 2020; Received in revised form 11 April 2020
Available online 19 April 2020

islands (SGI1) has been related to multidrug resistance profile among *Salmonella* strains (Gerlach and Hensel, 2007). The indiscriminate use of antibiotics in human activities also favors this property among bacteria (Castañeda-Ruelas and Jiménez-Edeza, 2018).

The etiology of salmonellosis is reported as a complex pathogenesis that involves multiple virulence genes grouped within the *Salmonell*a chromosome, and is called SPI "pathogenic islands of *Salmonella*" (Gerlach and Hensel, 2007). Plasmid, phage and fimbriae genes have also been identified as virulence factors for the pathogenesis of salmonellosis (Sabbagh et al., 2010; Hansmeier et al., 2017). The position and composition of the SPIs are associated with a possible acquisition by horizontal transfer. In addition, a pathogenic *Salmonella* clone could share and preserve virulence genes throughout the vertical transfer (Salama and Falkow, 1999). Currently, there are 21 SPIs, which differ in characteristics but encode the determinants responsible for the establishment of specific host interactions (Sabbagh et al., 2010); SPI-1 and SPI-2 have been easily studied and determined as the main SPI for host infection. The concentration of virulence genes within the SPI contributes to the particular virulence phenotype between subspecies, serotypes and strains (Gerlach and Hensel, 2007). Currently, these SPIs are being described mainly in strains belonged to *S. enterica* (I).

Salmonella houtenae inhabits the intestinal tract of reptiles and establishes an endophytic relationship with the natural environment (Zhu et al., 2013). Today, 73 serotypes of *Salmonella houtenae* have been described (Lamas et al., 2018), and some of them have been implicated as a source of serious and life-threatening diseases, such as sepsis, endocarditis and urinary tract infections that primarily affect vulnerable people (Editorial team collective et al., 2008; Abbott et al., 2012; Chen et al., 2013). However, the pathogenic potential of *Salmonella houtenae* has been underestimated due to the limited genetic information published to describe the route of infection, the adaptation process and the virulence factors that this bacterium can harbor. As far as we know, there is only one published report of a genome of *Salmonella houtenae* (Zhu et al., 2013).

Mexico is considered an endemic *Salmonella* country with a high incidence of salmonellosis cases every year (Dirección General de Epidmeiología (DGE), 2018). And, water resources have been identified as an important source of multiple non-typhoid *Salmonella* serotypes, and *Salmonella houtenae* is not an exception (Jiménez et al., 2014). Traditionally, it is denoted that *Salmonella* does not use water as its primary habitat; therefore, its persistence is related to the continuous fecal contamination of the environment (Pui et al., 2011). However, river water in regions with tropical climates is characterized by optimal conditions for the survival and persistence of *Salmonella* (Jiménez et al., 2014). Medrano-Félix et al. (2017) found that the physicochemical characteristics and carbohydrate sources of river water allowed the adaptation of non-typhoid *Salmonella* serotypes.

Unfortunately, it is documented that rivers in Culiacan Mexico are often used as a water source for agricultural and recreational activities, and the safety of this source has been questioned due to reports of non-typhoid *Salmonella* outbreaks transmitted by food that are linked to irrigation water of crops (Behravesh et al., 2008; Centers for Disease Control and Prevention (CDC), 2012; Centers for Disease Control and Prevention (CDC), 2013; Centers for Disease Control and Prevention (CDC), 2015). Therefore, it is necessary to broaden the knowledge of the influence of the aquatic habitat on the transmission and pathogenesis of *Salmonella*, especially in subspecies or poorly reported subspecies, one of them being *Salmonella houtenae*.

2. Methods

2.1. Growth bacterial conditions

Salmonella houtenae str. CFSAN039533 was isolated from river at northwestern region of Mexico in 2009 (Jiménez et al., 2014). The strain was obtained from the collection of the National Food Safety Laboratory Research (LANNIA). One colony of CFSAN039533 strain cultured on Hektoen enteric agar (Oxoid, USA) was inoculated in tryptic soy broth (Oxoid, USA) for 24 h yielding a final density of 1×10^8 CFU/mL (OD = 0.1). The bacteria culture was employed for genotypic and phenotypic analysis.

2.2. Physiological profile

The physiological profile was determined using Biolog-Ecoplates® (Biolog Inc., USA) according to the manual guide provided by Biolog Inc. The Biolog-Ecoplates® consist of 96-wells that contain 31 different substrates (per triplicate) including carbohydrates, amino acids, polymers, amines/amides, carboxylic acids and acetic acids. An aliquot of 150-μL of standardized inoculum was poured into the Biolog-Ecoplates® and incubated at 37 °C for 72 h. The utilization of a particular substrate was indicated by the transformation of the indicator dye (tetrazolium salt) to purple. The data was measured with a spectrophotometer (λ = 590 nm) each 15 min, and analyzed with the Microlog Data Collection software 1.2 (Biolog Inc., USA).

2.3. Antimicrobial resistance analysis

The antimicrobial resistance phenotype of the CFSAN039533 strain was determined using the disk diffusion method via the standard procedure of the Clinical and Laboratory Standards Institute (CLSI) (Clinical and Laboratory Standards Institute (CLSI), 2010). The antimicrobial drugs tested were amikacin, amoxicillin-clavulanic acid, ampicillin, apramycin, cefoperazone, ceftazidime, chloramphenicol, colistin sulfate, furazolidone, gentamicin, nalidixic acid, neomycin, streptomycin, sulfonamides, tetracycline, and trimethoprim-sulfamethoxazole. The results were classified as susceptible, intermediate, or resistant strain according to the interpretation of the zone diameter standards recommended by CLSI. *Escherichia coli* ATCC 25922 strain was used as control for antimicrobial determinations.

2.4. DNA extraction and whole-genome sequencing

Chromosomal DNA was extracted from bacterial culture using DNeasy Blood & Tissue Kit (Qiagen, USA), following manufacturer's instructions. The concentration and purity of genomic DNA concentration was tested by Qubit 2.0 fluorometer (Life Technologies, USA) and NanoDrop 2000 UV-VIS spectrophotometer (Thermo Scientific, USA), respectively. The sequencing libraries were prepared using 2 ng of DNA employing Nextera XT kit (2 × 251 bp) (Illumina, USA). Whole genomes were sequenced on an Illumina MiSeq platform (Illumina, USA). The quality metrics of reads was performed by FastQC, PEAT-V1–2.1.4 (Li et al., 2015) and Trimmomatic version 0.36 (Bolger et al., 2014) platforms.

2.5. Genome assembly and annotation

De novo assembling analysis was performed with the pipeline A5-miseq (Coil et al., 2015). Genome annotation was performed via NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016). The rRNA and tRNA prediction was performed by PGAP. Signal peptides were identified by Signalp v4.1 (Petersen et al., 2011) and clustered regularly interspaced short palindromic repeats (CRISPR) by CRISPRs Finder online (Grissa et al., 2007). Analyses to identify genes that were assigned to transmembrane domains were performed by using TMHMM server v.2.0 (Krogh et al., 2001). Also, the protein domains and clusters of orthologous groups (COG) assignation of the protein sequences and metabolic pathways (KEEG) were performed in WebMGA (Wu et al., 2011). MLST finder 1.8 (Larsen et al., 2012), SeqSero software (Zhang et al., 2015) and Srst2 software (Inouye et al., 2014) were used to identify the sequence type, serotype and antimicrobial resistance genes in the draft genome of CFSAN039533 strain,

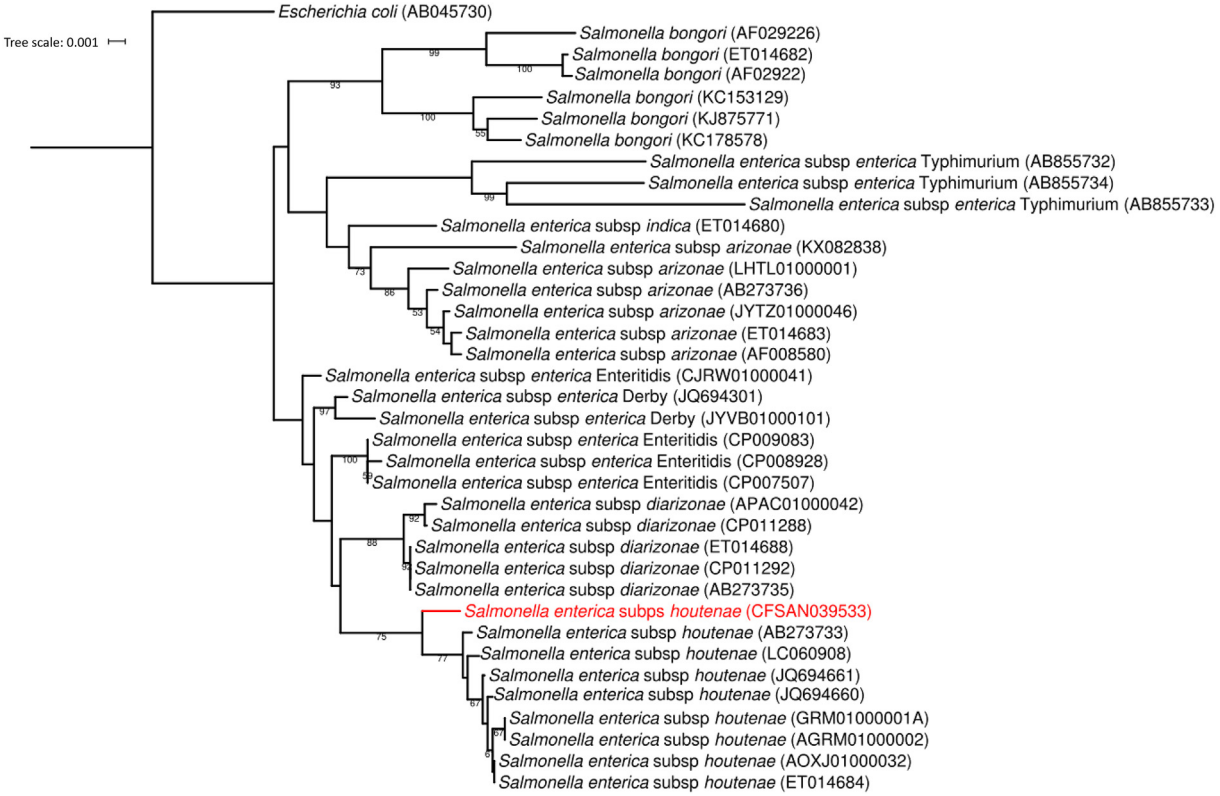


Fig. 1. Phylogenetic tree highlighting the position of *Salmonella houtenae* str. CFSAN03953 relative to other *Salmonella* species and subspecies. Phylogenetic tree was obtained by the Maximum Likelihood method within MEGA7 software using JC + G for the 36 sequences of 16 s RNA gene representatives of all *Salmonella* species. GenBank accession numbers are presented in parentheses. The position of the isolate CFSAN330395 is indicated in red. *Escherichia coli* ABDO45730 served as the outgroup. The scale bar represents 0.01 substitutions per nucleotide position. Support value of nodes was estimated by bootstrap (1000 replicates using neighbor-joining). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively.

2.6. Phylogenetic analyses

The phylogenetic tree was obtained by the maximum likelihood method within MEGA7 software (Kumar et al., 2016) using C + G (%) for the 36 sequences of 16 s RNA gene representatives of all *Salmonella* species. The GenBank accession numbers of the strains are presented in Fig. 1. Phylogenetic distances were calculated with Kimura's 2-parameter model and the support value of nodes was estimated by bootstrap analysis based on 1000 replicates. *Escherichia coli* ABD45730 strain served as an out-group.

2.7. Analysis of pangenome

The genomes of *S.* Typhimurium LT2 and *S.* Enteritidis 771,427 strains were included to construct the pangenome of *S. houtenae* str. CFSAN039533. DNA sequences and protein-coding sequence (CDS) from all strains identified by annotation via NCBI prokaryotic genome (version 3.3) were clustered and visualized using the Gview server (version 3.0) (Petkau et al., 2010). The default parameters included for the analysis were BLASTn with an expected value cut-off (1×10^{-10}), genetic code (bacteria), identity (80%), and alignment length (100 bp).

2.8. Virulotyping

Selected protein sequences annotated via RAST server related to pathogenic and adaptation properties (subsystems of virulence and defense, secretion systems, iron metabolism, regulation, and stress response) were analyzed and clustered using multivariate analysis and the VennPainter tool (version 1.2.0) (Lin et al., 2016). *S.* Typhimurium LT2 and *S.* Enteritidis 77–1427 strains were used as reference strains. For defining the type or virulence of CFSAN039533 strain, we performed a virulence gene database constituted with 62 representatives' genes of SPIs (SPI-1, SPI-2, SPI-3, SPI-4, SPI-5 and SPI-7) and fimbriae genes of *Salmonella* within the CLC Genomics Workbench program. Nucleotide sequences of the pathogenic islands genes and virulence-relegated genes were obtained from Virulence Factors Database (VFD) and Protein Data Bank (PDB). The Basic Location Alignment Search Tool (BLAST) was used for mapping the current identity and coverage of virulence gene database. A threshold of 99% of similarity was considered to determine the identity of a gene.

3. Results

3.1. Physiological profile

The physiological profile of *S. houtenae* str. CFSAN039533 is showed in Table 1. The CFSAN039533 strain metabolized an average of 45% out of the 31 carbon sources of the Biolog-Ecoplates® tasted; polymers, carbohydrates, amino acids and miscellaneous compounds were identified as mainly chemical groups assimilated by the strain. The amino and amides groups were not used by the strain during growth condition.

3.2. Antimicrobial resistance profile

S. houtenae str. CFSAN039533 was identified as susceptible to all antimicrobials drugs of the panel tested.

3.3. Genome properties

The draft genome sequence of *S. houtenae* str. CFSAN039533 is part

Table 1
Metabolic profile of *Salmonella houtenae* str. CFSAN039533 from river in Mexico.

Type	Carbon sources	Utilization	Chemical guild rate
Amino acids	L-Arginine	−	50%
	L-Asparagine	+	
	Ác. Glycil L-Glutamic acid	+	
	L-Phnylalanine	−	
	L-Serine	+	
	L-Threonine	−	
Carbohydrates	D-Cellobiose	−	43%
	i-Erythritol	−	
	N-Acetyl D-Glucosamine	+	
	β-Methyl-D-Glucoside	+	
	α-D-Lactose	−	
	D-Mannitol	+	
	D-Xylose	−	
Carboxylic acids	γ-Hidroxy butyric acid	−	33%
	α-Keto butyric acid	−	
	2-Hidroxy benzoic acid	−	
	4-Hidroxy benzoic acid	−	
	D-Galactonic acid	−	
	D-Glucosaminic acid	+	
	γ-Lactone D-Galacturonic acid	+	
	Itaconic acid	+	
	D-Malic acid	−	
Amines/amides	Phenylethylamine	−	0%
	Putrescine	−	
Polymers	Tween 40	+	50%
	Tween 80	+	
	Glycogen	−	
	α-Ciclodextrina	−	
Miscellaneous	α-D,L-Glycerol phosphate	+	100%
	Glucose-1-phosphate	+	
	Pyruvic acid Methyl ester	+	

+: Positive result, −: Negative result.

Table 2
Genome statistics *Salmonella houtenae* str. CFSAN039533 from river in Mexico.

Attribute	Value	(%)[a]
Genome size (bp)	4,743,243	100.0
DNA coding (bp)	4,151,358	87.52
DNA G + C (bp)	2,447,638	51.60
DNA scaffolds	0	100.0
Total genes	4553	100.0
Protein coding genes	4266	93.69
RNA genes	107	2.35
Pseudo genes	181	3.97
Genes in internal clusters	NA	NA
Genes with function prediction	3338	73.31
Genes assigned to COGs	3561	78.21
Genes with Pfam domains	3823	83.97
Genes with signal peptides	1073	25.15
Genes with transmembrane helices	1036	22.75
CRISPR repeats	1	0.02

[a] The total is based on either the size of the genome in base pairs or on the total number of protein coding genes in the annotated genome.

of GenBank BioProject PRJNA186035 under the accession no. MEKC00000000. The summary of features for the draft genome is shown in Table 2. The genome assembly of CFSAN039533 strain consists of 62 DNA contigs amounting to ~4.74 Mb with a G + C content of 51.6%. Of the 4553 total genes, 4266 were protein-coding sequences, 106 RNAs related genes (including 77 tRNAs and 17 rRNAs), and 181 pseudogenes were also identified. The genome coverage was 76×. The genes assigned to COG functional categories are listed in Table 3, revealing that the most abundant COG category was protein related to transport and metabolism of carbohydrates (373 proteins) and amino acids (363 proteins). Also, KEGG tool was used to investigate the

Table 3
Number of genes associated with general COG functional categories.

Code	Description	Proteins	
		No.	%
J	Translation, ribosomal structure and biogenesis	180	4.22
A	RNA processing and modification	1	0.02
K	Transcription	319	7.48
L	Replication, recombination and repair	181	4.24
B	Chromatin structure and dynamics	0	0
D	Cell cycle control, Cell division, chromosome partitioning	37	0.87
V	Defense mechanisms	56	1.31
T	Signal transduction mechanisms	198	4.64
M	Cell wall/membrane biogenesis	245	5.74
N	Cell motility	101	2.37
U	Intracellular trafficking and secretion	122	2.86
O	Posttranslational modification, protein turnover, chaperones	146	3.42
C	Energy production and conversion	282	6.61
G	Carbohydrate transport and metabolism	373	8.74
E	Amino acid transport and metabolism	363	8.51
F	Nucleotide transport and metabolism	86	2.02
H	Coenzyme transport and metabolism	157	3.68
I	Lipid transport and metabolism	96	2.25
P	Inorganic ion transport and metabolism	210	4.92
Q	Secondary metabolites biosynthesis, transport and catabolism	65	1.52
R	General function prediction only	401	9.40
S	Function unknown	354	8.30
–	Not in COGs	705	16.53

metabolic pathways of the CFSAN039533 strain showing genes to metabolize amino acids, carbohydrates, glycans, lipids, polyketides, nucleotides, cofactors and vitamins. Additionally, many genes related to ABC transporters, bacterial secretion and phosphotransferase systems were identified as its main membrane transport to uptake nutrients, which its related with its carbon source phenotype (Table 1). MLST finder 1.8 identified the sequence type of CFSAN039533 strain as ST-454, while SeqSero software determined its antigenic formula as 48:g,z51 corresponding to Marina serotype. Regarding SRST2 platform, it was identified that the CFSAN039533 strain displayed resistance genes belonging to aminoglycosides drugs family (*Aac6-Iaa_AGly*).

3.4. Phylogenetic tree

S. houtenae str. CFSAN039533 is classified within the order *Enterobacteriales* of the class *Enterobacteriaceae* and genus *Salmonella*. Based on the phylogenetic tree constructed with 16S rRNA, the CFSAN039533 strain is placed in the distinct cluster formed by the *Salmonella houtenae* (IV) strains as shown in Fig. 1. The 16S rRNA phylogenetic tree also revealed that the CFSAN039533 strain was closely related to *Salmonella diarizonae* (IIIb) and to *Salmonella enterica* (I) strains, but distant from *Salmonella bongori*.

3.5. Analysis of pangenome

A pangenome of *S. houtenae* str. CFSAN039533 was constructed based on its DNA sequence and CDS, and taking as a reference the *S.* Typhimurium LT2 and *S.* Enteritidis 77–1427 strains (Fig. 2). The strains have an open pangenome with a conserved core genome based on the DNA sequence (Fig. 2A). Contributions of unique DNA regions of CFSAN039533 strain were observed. Interestingly, *S. houtenae* str. CFSAN039533 differed markedly in the CDSs content regarding to reference strains (Fig. 2B). A total of 5615 CDSs were identified among the strains, approximately sharing a 62.2% (n = 3491) of CDSs of the represented core genome (Fig. 2C). Particularly, *S. houtenae* str. CFSAN039533 had a higher content of unique CDSs (n = 846) in the constructed pangenome when compared to serotypes of *Salmonella enterica*.

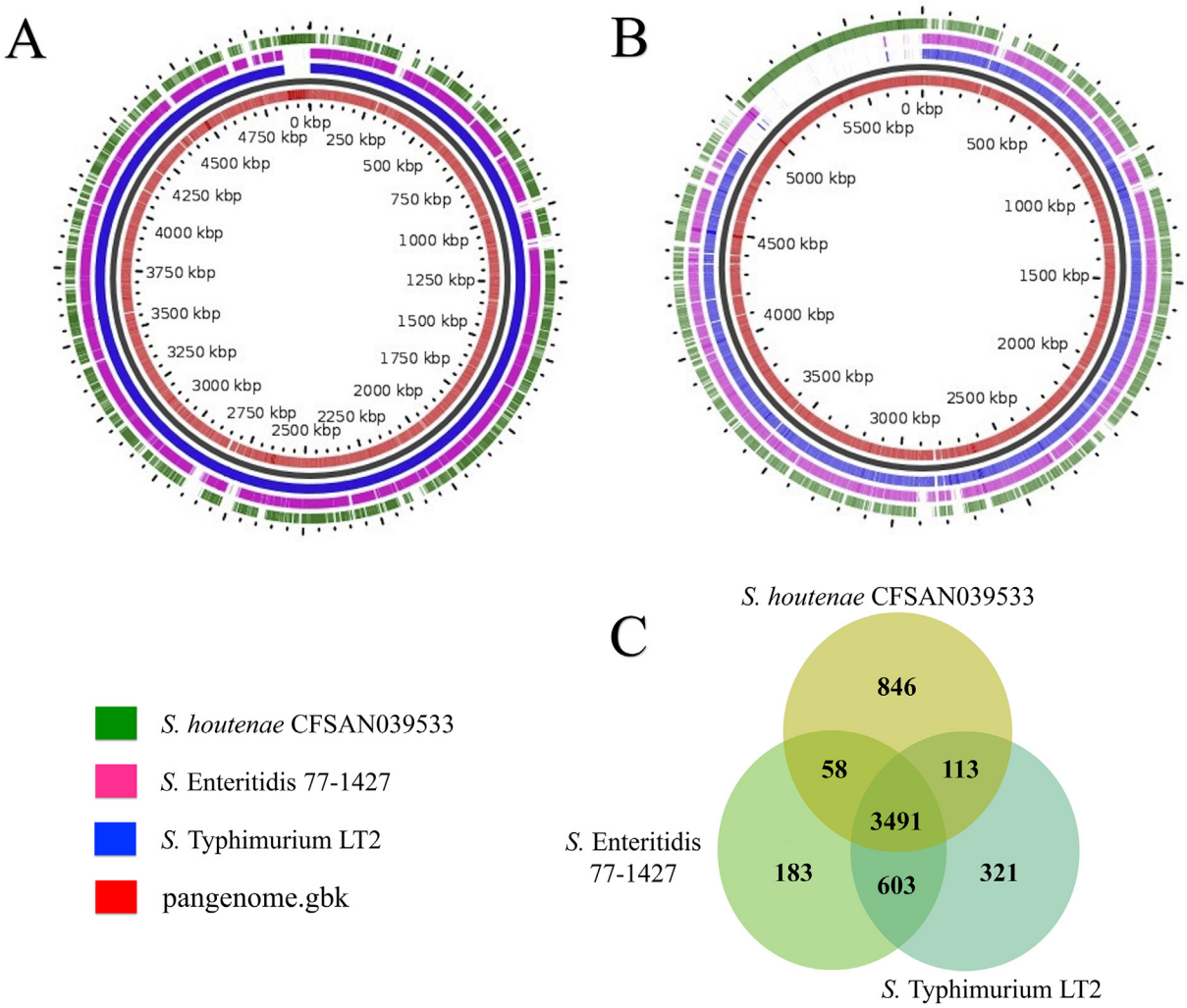


Fig. 2. Circular plots of pangenomes of *Salmonella houtenae* str. CFSAN039533. The genome sequences were subjected to pangenome analysis using the GView server based in A) sequences and B) CDSs. The innermost slot shows the constructed pangenome and the outer slots indicate the specific strain genomes. The white space indicates a region missing in the specified genome. C) Venn diagram illustrating the number of unique CDSs in each strain, core genome (in the middle) and accessory CDSs.

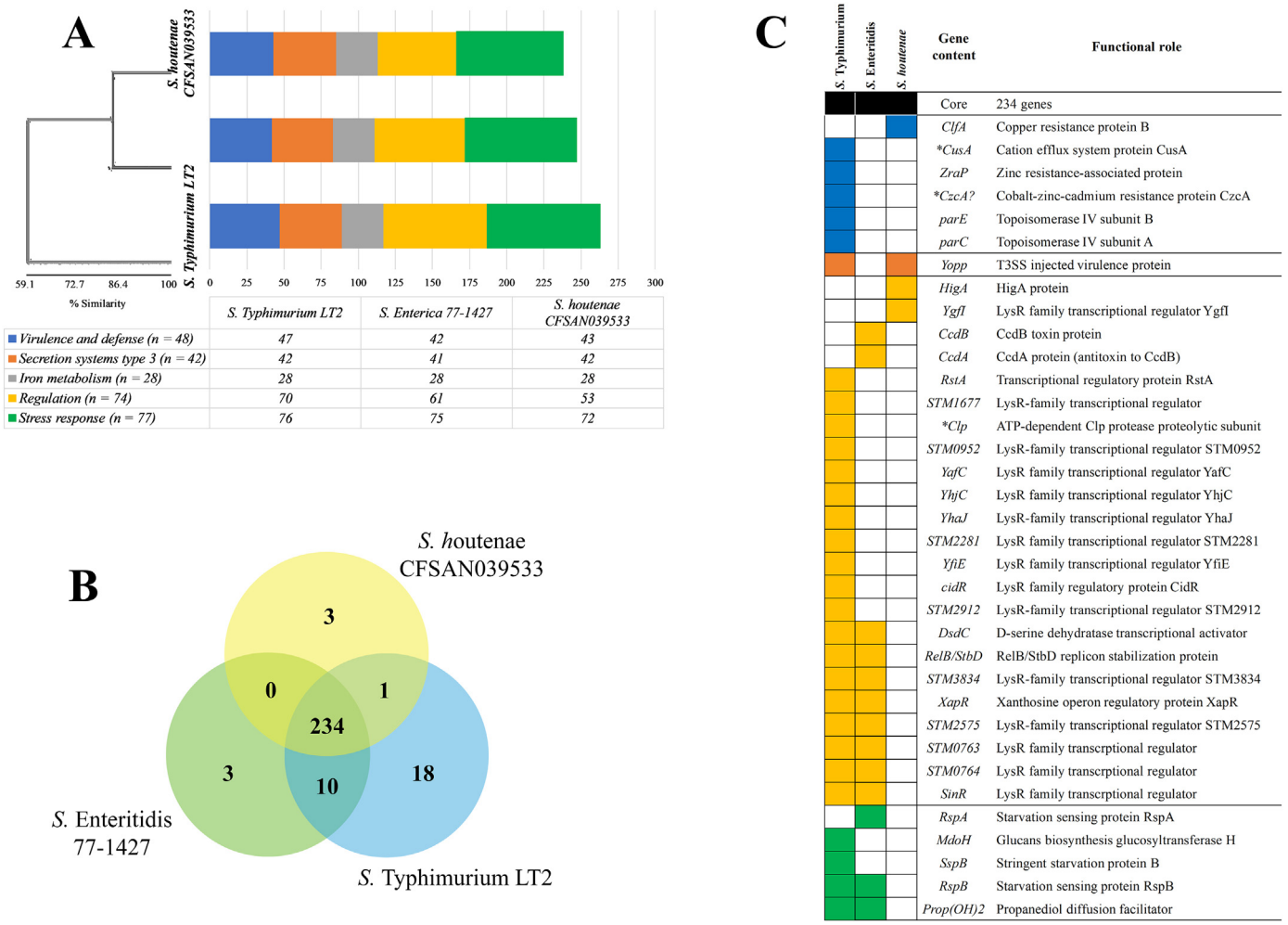
3.6. Virulotyping

The comparison of *S. houtenae* str. CFSAN039533 with clinical *Salmonella* strains is presented in Fig. 3. It showed that these *Salmonella* strains shared a similar content of the most selected proteins annotated via RAST (Fig. 3A), which playing roles as pathogenicity factors. It was observed that *S. houtenae* str. CFSAN039533 shared a genetic homology of ≈86% and ≈60% with *S.* Typhimurium LT2 and *S.* Enteritidis 77–1427, respectively. The three strains of *Salmonella* shared 234 (86%) of selected proteins (Fig. 3B); 3 genes were classified as specific for CFSAN039533 strain related to virulence functions (*ClfA*) and regulation systems (*HigA, Ygfl*) (Fig. 3C). The comparison of the virulence database with CFSAN39533 strain revealed the presence of 41 (66.0%) virulence genes mainly located in pathogenicity islands (SPIs) and related with fimbriae genes (Fig. 4). The concentration of virulence genes located in SPIs of the CFSAN039533 strain plays an essential role for pathogenic mechanisms (adhesion, invasion, intracellular survival, toxin production, fimbriae expression, resistance), and define a particular virulence type of *Salmonella houtenae*.

4. Discussion

The physiological characterization and De novo sequencing of the genome of *S. houtenae* str. CFSAN039533 were conceived in this study, in order to clarify its abilities to persist at the aquatic environment, and determine its role as a potential waterborne pathogen. Furthermore, currently unreported genetic information is provided for *S. houtenae* serotype Marina, pointing out key properties for colonizing warm-blooded host. The complexity and diversity of aquatic ecosystem at the northwestern of Mexico, which is consequence of agricultural and domestic activities (Jiménez et al., 2014; Arellano-Aguilar et al., 2017), could be an important influence over the physiology and prevalence of *Salmonella spp*.

The physiological fingerprint observed in the CFSAN039533 strain, by using Biolog-Ecoplates®, showed the versatile metabolic capacity of this bacteria to use substrates from different chemical groups (Table 1). This functional microbial behavior, could give *Salmonella houtenae* the ability to inhabit as a possible symbiont in the aquatic ecosystem. Several studies have employed the Biolog-Ecoplates® for describing the microbial ecology from environmental samples likewise soil, water, sea, land, and wastewater, providing an identifiable metabolic fingerprint (Preston-Mafham et al., 2002; Xu and Ge, 2015). In order to make conclusive asseverations further assays must be conducted with *Salmonella* subspecies strains. This study reports a concordance between the genomic analysis of the metabolic pathways of *S. houtenae* str. CFSAN039533 and the expressed metabolic phenotype (Table 1).



	S. Typhimurium LT2	S. Enterica 77-1427	S. houtenae CFSAN039533
Virulence and defense (n = 48)	47	42	43
Secretion systems type 3 (n = 42)	42	41	42
Iron metabolism (n = 28)	28	28	28
Regulation (n = 74)	70	61	53
Stress response (n = 77)	76	75	72

Fig. 3. The virulence core genome of *Salmonella houtenae* str. CFSAN039533. A) The total protein belonging to the RAST subsystems related to pathogenicity and adaptation of *Salmonella*. B) Cluster analysis of the selected protein; black, dark-gray and gray represented the core, unique and accessories genes, respectively. The tree on the upper side of the figure shows a hierarchical complete-linkage clustering of the profiles based on the Euclidean distance.

In relation to the antimicrobial susceptibility of *Salmonella*, several studies reported the increase of antimicrobial resistance in non-typhoid *Salmonella* strains (Liang et al., 2015), but, available information about *Salmonella houtenae* is scared. Previously, Franco et al. (2011) showed a wild-type susceptible phenotype of *Salmonella houtenae* towards the major classes of antimicrobials used in human therapy. Nowadays, it has been observed that *Salmonella houtenae* strains have acquired resistance against to ampicillin, tetracycline (Singh et al., 2013), and streptomycin (Barazorda et al., 2015). In this study, the CFSAN039533 strain isolated from river water, displayed a susceptible phenotype to different antimicrobial families such as β-lactams, aminoglycosides, tetracyclines, cephalosporin, trimethoprim, and quinolones. But, in silico analysis of its antimicrobial resistance, showed putative genes associated with aminoglycoside resistance, which it could be under-expressed. However, aminoglycosides family is not recommended to salmonellosis therapy due to limited biocide activity. The bacterial resistance of aminoglycosides drugs family is due to aminoglycoside 6'-*N*-acetyltransferase (*Aac6-Iaa_AGly*) enzymes mainly encoded into chromosomal structure, which phenotype has been correlated with the weakly expressed or unexpressed *Aac6* family genes (Magnet et al., 1999). Even though, a susceptible strain was observed, the interchange of antimicrobial resistance genes should not be surprised, considering that the rivers at the northwestern of Mexico have being identified as a harboring source of multidrug resistance non-typhoid *Salmonella* serotypes (Castañeda-Ruelas and Jiménez-Edeza, 2018).

For the 16S rRNA phylogenetic tree, CFSAN039533 strain and reference strains of *Salmonella houtenae* were tightly clustered. Zhu et al. (2013) showed the phylogenetic relation of *Salmonella houtenae* (RKS3027 str.) with *Salmonella species* placed it between *S. enterica* (I) and *S. bongri*. A key factor of divergence of *Salmonella species* (*bongori* and *eneterica)* is the acquisition of pathogenic island-2 (SPI-2) that contributes its ability to colonize a host cell type. It was defined that *Salmonella houtenae* diverged from subspecies *Salmonella enterica, Salmonella indica* and *Salmonella salamae* by the content of virulence and metabolic factors (Pui et al., 2011; Zhu et al., 2013). A total of 73 serotypes of *Salmonella houtenae* have been described (Pui et al., 2011); CFSAN039533 strain was typed as Marina serotype, which it is considered an opportunistic pathogen related with cold-blood animal transmission via (Editorial team collective et al., 2008; Pui et al., 2011; Abbott et al., 2012).

The DNA sequencing methods together with bioinformatics tools have allowed to describe the biology of several microorganisms. A great number of genomes sequences of *Salmonella enterica* subsp. *enterica* have been published at genomics databases (NCBI, RAST, IMG), but there is scared data concerning of *Salmonella houtenae*. To best our knowledge, it was reported a total of 41 assembly genomes at National Center of Biotechnology Information, and only one research publication (Zhu et al., 2013) was done for this non-enterica subspecies of *Salmonella*. The pangenome performed in this study showed the genetic relationship grade within the selected strains with respect to DNA

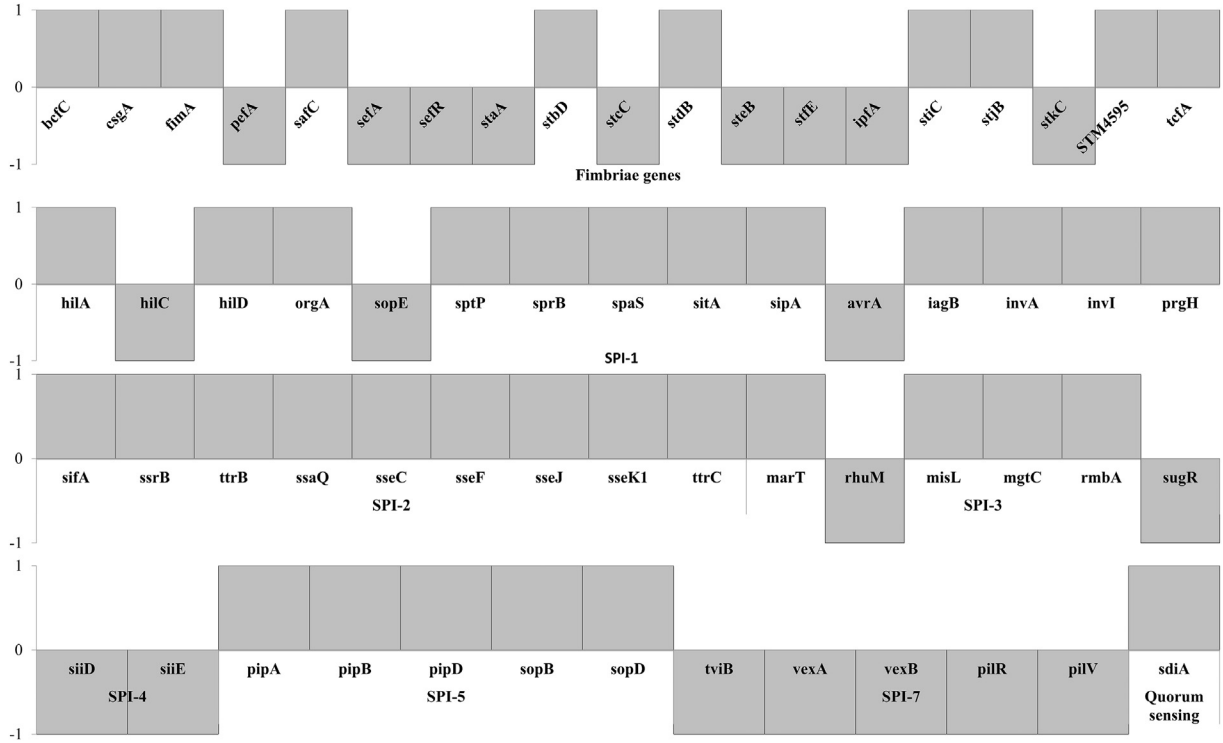


Fig. 4. Virulotyping of *Salmonella houtenae*str. CFSAN03953 isolated from river water in Northwestern Mexico. The bioinformatics analysis of virulence genes was denoted with values of 1 (present) and − 1 (absent). The virulence gene database was constructed including SPIs (I, II, III, IV, V and VII), and fimbriae and quorum sensing genes.

sequencing, where the main differences could be associated by different host-adapted pathogenic lifestyles. This assumption also can be related with the discrepancy in the CDS content between the strains (Fig. 2). The comparison between the two most recognized human *Salmonella* pathogens and an "environmental" *Salmonella* serovar, allows us to hypothesize about the distinctive genes of each strain can be associated with the virulence in cold-blooded host (CFSAN039533 strain) and/or warm-blooded host (*S*. Typhimurium LT2 and *S*. Enteritidis 77–1427 strains). Likewise, CFSAN039533 strain shared an important number of CDS with reference strains, which could support its ability to colonizes warm-blooded host as observed with the virulotyping analysis done (Fig. 3 and Fig. 4).

The genome annotation of the CFSAN039533 strain helps to identify the factors that allow it to colonize a host or the environment, and infers its implication in human health as a pathogen. The CFSAN039533 strain displayed putative metabolic genes related with survival strategies at a non-host source (Table 3 and Fig. 3), such as water. Also, a closely relatedness (60–86%) of these genetic properties was observed between CFSAN039533 strain and the clinical reference strains (Fig. 3). The relationship between the genomic content with its pathogenic potential in vitro, has been evidenced in non-typhoid *Salmonella* strains from aquatic origin (Burgueño-Roman et al., 2019). In this sense, water is identified as reservoir and/or potential transmission vehicle of pathogenic *Salmonella* strains, such *Salmonella houtenae.*

One of the main goals of this study was the contribution to the knowledge of the pathogenesis of *Salmonella houtenae* since its epidemic interest has become relevant. Nowadays, *Salmonella houtenae* have been reported as an opportunistic pathogen, but its genetic information and interpretation, is still scare. The CFSAN039533 strain displayed virulence genes belonging to SPI-1, SPI-2, SPI-3 and SPI-5, which allow an efficient pathogenic life cycle into wide range of hosts (Fig. 3 and Fig. 4). Rychlik et al. (2009) have described the SPI 1–5 as main SPIs into *Salmonella* genome. Many studies have highlighted that SPI-1, SPI-2 and SPI-4 are conserved genetic islands, while SPI-3 and SPI-5 displayed a variable genetic information (Gerlach and Hensel, 2007; Sabbagh et al., 2010), which could favor the ability to colonize the host.

The fimbriae genes displayed in CFSAN039533 strain ensure the first step of *Salmonella* pathogenesis because are involved in the cell host adhesion process, and further colonization and dissemination (Hansmeier et al., 2017). Interestingly, this strain presents 10 of the approximately 20 adhesion genes (fimbriae and non-fimbriae genes) described for *Salmonella* (Wagner and Hensel, 2011; Hansmeier et al., 2017). Further, this strain presents the *tcfA* gene, which is a fimbriae operon previously reported as specific for *S.* Typhi and absent in *S*. Typhimuirum (Bishop et al., 2008). The detection of the SPI-1 and SPI-2 into the genome of CFSAN039533 strain, could suggest the capacity of this strain to use the secretion system III and to encode genes that guarantee the adherence, invasion and intracellular replication into phagocytic and non-phagocytic cells. Also, the genes of SPI-3 and SPI-5 have been associated with survival into macrophages and enteropathogenesis, respectively (Gerlach and Hensel, 2007; Rychlik et al., 2009; Sabbagh et al., 2010). The arsenal of molecular factors described in this draft-genome define *S. houtenae* str. CFSAN039533 as a potential pathogen, and requires the exploration of genomic bases into diverse strains of this subspecies.

5. Conclusions

High-draft genome sequence described here is the first published document providing an insight into genomic basis and metabolic characteristics of *S. houtenae* serotype Marina (strain CFSAN039533). The CFSAN039533 strain shows the ability to colonize warm-blooded host and display antimicrobial resistance defenses. This genome sequence has been deposited in GenBank, contributing to garner information of *Salmonella houtenae* genome, and stimulate further epidemiologic investigations into this *Salmonella* subspecies. This information analyzes in detail the *Salmonella houtenae* chromosome and its implication in public health problems.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the National Council of Science and Technology (CONACyT) grant CB-2014/235989.

Availability of data and material

All data generated during this study are included in this article.

Consent for publication

Not applicable.

Author contribution

Gloria M. Castañeda-Ruelas drafted the manuscript. Gloria M. Castañeda-Ruelas and Areli Román-Burgueño performed the laboratory experiments and carried out the bioinformatics analysis. All authors discussed the data. Maribel Jiménez-Edeza designed the study and approved the final manuscript.

Declaration of Competing Interest

The authors have no declarations or conflicts of interest associated with this work.

Acknowledgements

The authors thank the National Food Safety Laboratory Research (LANNIA) at Centro de Investigación en Alimentación y Desarrollo, A.C (CIAD) in Culiacán, México for sharing the strain and permitting it use in this research.

References

Abbott, S.L., Ni, F.C.Y., Janda, J.M., 2012. Increase in extraintestinal infections caused by *Salmonella enterica* subspecies II–IV. Emerg. Infect. Dis. https://doi.org/10.3201/eid1804.111386.

Arellano-Aguilar, O., Betancourt-Lozano, M., Aguilar-Zárate, G., Ponce de Leon-Hill, C., 2017. Agrochemical loading in drains and rivers and its connection with pollution in coastal lagoons of the Mexican Pacific. Environ. Monit. Assess. https://doi.org/10.1007/s10661-017-5981-8.

Barazorda, S., Knotek, Z., Čížek, A., Masaříková, M., Myšková, P., 2015. The incidence and antibiotic resistance of *Salmonella species* isolated from cloacae of captive veiled chameleons. Acta Vet. Brno. https://doi.org/10.2754/avb201584030209.

Behravesh, C.B., Mody, R.K., Jungk, J., Gaul, L., Redd, J.T., Chen, S., Cosgrove, S., Hedican, E., Sweat, E., Chávez-Hauser, L., Snow, S.L., Hanson, H., Nguyen, T., Sodha, S., Boore, A.L., Russo, E., Mikoleit, M., Theobald, L., Gerner-Smidt, P., Hoekstra, R.M., Angulo, F.J., Swerdlow, D.L., Tauxe, R.V., Griffin, P.M., Williams, I., T., 2008. Outbreak of *Salmonella* Saintpaul infections associated with raw produce. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa1005741.

Bishop, A., House, D., Perkins, T., Baker, S., Kingsley, R.A., Dougan, G., 2008. Interaction of *Salmonella enterica* serovar Typhi with cultured epithelial cells: roles of surface structures in adhesion and invasion. Microbiology. https://doi.org/10.1099/mic.0.2008/016998-0.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu170.

Burgueño-Roman, A., Castañeda-Ruelas, G.M., Pacheco, R., Jiménez, M., 2019. Pathogenic potential of non-typhoidal Salmonella serovars isolated from aquatic environment in Mexico. Genes & Genomics. https://doi.org/10.1007/s13258-019-00798-7.

Castañeda-Ruelas, G.M., Jiménez-Edeza, M., 2018. Evaluación de ríos del noroeste de México como reservorios de serotipos de *Salmonella* resistentes a antibióticos. Revista Internacional de Contaminación Ambiental 34, 191–201. http://dx.doi.org/10.20937/RICA.2018.34.02.01.

Centers for Disease Control and Prevention (CDC), 2012. Multistate Outbreak of *Salmonella* Braenderup Infections Associated to Mangoes. https://www.cdc.gov/salmonella/braenderup-08-12/index.html.

Centers for Disease Control and Prevention (CDC), 2013. Multistate Outbreak of *Salmonella* Saintpaul Infections Linked to Imported Cucumbers. http://www.cdc.gov/*Salmonella*/saintpaul-04-13/index.html.

Centers for Disease Control and Prevention (CDC), 2015. Multistate Outbreak of *Salmonella* Poona Infections Linked to Imported Cucumbers (Final Update). https://www.cdc.gov/salmonella/poona-09-15/index.html.

Chen, H.M., Wang, Y., Su, L.H., Chiu, C.H., 2013. Nontyphoid *Salmonella* infection: microbiology, clinical features, and antimicrobial therapy. Pediatr. Neonatol. https://doi.org/10.1016/j.pedneo.2013.01.010.

Clinical and Laboratory Standards Institute (CLSI), 2010. Performance Standards for Antimicrobial Susceptibility Testing; Twentieth Informational Supplement. Wayne, PA.

Coil, D., Jospin, G., Darling, A.E., 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu661.

Dirección General de Epidmeiología (DGE), 2018. Anuarios de Morbilidad 1984–2016. https://www.gob.mx/salud/acciones-y-programas/anuarios-de-morbilidad-1984-2016.

Editorial team collective, Bertrand, S., Rimhanen-Finne, R., Weill, F.X., Rabsch, W., Thornton, L., Perevoscikovs, J., van Pelt, W., Heck, M., 2008. *Salmonella* infections associated with reptiles: the current situation in Europe. Euro Surveill. https://doi.org/10.2807/ese.13.24.18902-en.

Franco, A., Hendriksen, R.S., Lorenzetti, S., Onorati, R., Gentile, G., Dell'Omo, G., Aarestrup, F.M., Battisti, A., 2011. Characterization of *Salmonella* occurring at high prevalence in a population of the land Iguana Conolophus subcristatus in Galápagos Islands, Ecuador. PLoS One. https://doi.org/10.1371/journal.pone.0023147.

Gerlach, G.C., Hensel, M., 2007. *Salmonella* Pathogenicity Islands in host specificity, host pathogen-interactions and antibiotics resistance of *Salmonella enterica*. Berl. Munch. Tierarztl. Wochenschr. https://doi.org/10.2376/0005-9366-120-317.

Grissa, I., Vergnaud, G., Pourcel, C., 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. https://doi.org/10.1093/nar/gkm360.

Hansmeier, N., Miskiewicz, K., Elpers, L., Liss, V., Hensel, M., Sterzenbac, T., 2017. Functional expression of the entire adhesiome of *Salmonella enterica* serotype typhimurium. Nature. https://doi.org/10.1038/s41598-017-10598-2.

Inouye, M., Dashnow, H., Raven, L.A., Schultz, M.B., Pope, B.J., Tomita, T., Zobel, J., Holt, K.E., 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genom. Med. https://doi.org/10.1186/s13073-014-0090-6.

Jiménez, M., Martinez-Urtaza, J., Rodriguez-Alvarez, M.X., Leon-Felix, J., Chaidez, C., 2014. Prevalence and genetic diversity of *Salmonella* spp. in a river in a tropical environment in Mexico. J. Water Health 12, 874–884. https://doi.org/10.2166/wh.2014.051.

Kirk, M.D., Pires, S.M., Black, R.E., Caipo, M., Crump, J.A., Devleesschauwer, B., Döpfer, D., Fazil, A., Fischer-Walker, C.L., Hald, T., Hall, A.J., Keddy, K.H., Lake, R.J., Lanata, C.F., Torgerson, P.R., Havelaar, A.H., Angulo, F.J., 2015. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. PLoS Med. https://doi.org/10.1371/journal.pmed.1001921.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

Lamas, A., Miranda, J.M., Regal, P., Vázquez, B., Franco, C.M., Cepeda, A., 2018. A comprehensive review of non-enterica subspecies of *Salmonella enterica*. Microbiol. Res. https://doi.org/10.1016/j.micres.2017.09.010.

Larsen, M.V., Consentino, S., Rasmussen, S., Friis, C., Hasman, H., Marving, R.L., Jelsbak, L., Sicheritz-Potén, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total genome sequenced bacteria. J. Clin. Microbiol. 50, 1355–1361.

Li, Y.L., Weng, J.C., Hsiao, C.C., Chou, M.T., Tseng, C.W., Hung, J.H., 2015. PEAT: an intelligent and efficient paired-end sequencing adapter trimming algorithm. BMC Bioinforma. https://doi.org/10.1186/1471-2105-16-S1-S2.

Liang, Z., Ke, B., Deng, X., Liang, J., Ran, L., Lu, L., He, D., Huang, Q., Ke, C., Li, Z., Yu, H., Klena, J.D., Wu, S., 2015. Serotypes, seasonal trends, and antibiotic resistance of non-typhoidal *Salmonella* from human patients in Guangdong Province, China, 2009–2012. BMC Infect. Dis. 15, 53. https://doi.org/10.1186/s12879-015-0784-4.

Lin, G., Chai, J., Yuan, S., Mai, C., Cai, L., Murphy, R., Zhou, W., Luo, J., 2016. VennPainter: a tool for the comparison and identification of candidate genes based on Venn diagrams. PLoS One. https://doi.org/10.1371/journal.pone.0154315.

Magnet, S., Courvalin, P., Lambert, T., 1999. Activation of the cryptic *aac(69)-Iy* aminoglycoside resistance gene of *Salmonella* by a chromosomal deletion generating a transcriptional fusion. J. Bacteriol. 181, 6650–6655.

Medrano-Félix, J.A., Castro-del Campo, N., Peraza, F., Martínez-Rodríguez, De J., Chaidez, C., 2017. Carbon source utilization-based metabolic activity of *Salmonella* Oranienburg and *Salmonella* Saintpaul in river water. Water Environ. J. https://doi.org/10.1111/wej.12306.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods. https://doi.org/10.1038/nmeth.1701.

Petkau, A., Stuart-Edwards, M., Stothard, P., Van Domselaar, G., 2010. Interactive microbial genome visualization with GView. Bioinformatics. https://doi.org/10.1093/bioinformatics/btq588.

Preston-Mafham, J., Boddy, L., Randerson, P.F., 2002. Analysis of microbial community functional diversity using sole-carbon-source utilization profiles - a critique. FEMS Microbiol. Ecol. https://doi.org/10.1111/j.1574-6941.2002.tb00990.x.

Pui, C.F., Wong, W.C., Chai, L.C., Tunung, R., Jeyaletchumi, P., Noor, H., Ubong, A., Farinazleen, M.G., Cheah, Y.K., Son, R., 2011. Review article *Salmonella*: a foodborne pathogen. Int. Food Res. J. 18, 465–473.

Rychlik, I., Karasova, D., Sebkova, A., Volf, J., Sisak, F., Havlickova, H., Kummer, V., Imre, A., Szmolka, A., Nagy, B., 2009. Virulence potential of five major pathogenicity islands (SPI-1 to SPI-5) of *Salmonella enterica* serovar Enteritidis for chickens. BMC Microbiol. https://doi.org/10.1186/1471-2180-9-268.

Sabbagh, S.C., Forest, C.G., Lepage, C., Leclerc, J.M., Daigle, F., 2010. So similar, yet so diferent: uncovering distinctive features in the genomes of *Salmonella enterica* serovars typhimurium and Typhi. FEMS Microbiol. Lett. 305, 1–13.

Salama, N.R., Falkow, S., 1999. Genomic clues for defining bacterial pathogenicity. Microbes Infect. 1, 615–619.

Singh, B.R., Singh, V., Ebibeni, N., Singh, R.K., 2013. Antimicrobial and herbal drug resistance in enteric bacteria isolated from faecal droppings of common house Lizard/Gecko (*Hemidactylus frenatus*). Int. J. Microbiol. 2013, 1–8. https://doi.org/10.1155/2013/340848.

Su, L.H., Chiu, C.H., 2007. *Salmonella*: clinical importance and evolution of nomenclature. Chang Gung Med. J. 30, 210–219.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw569.

Wagner, C., Hensel, M., 2011. Adhesive mechanisms of *Salmonella enterica*. Adv. Exp. Med. Biol. https://doi.org/10.1007/978-94-007-0940-9_2.

Wu, S., Zhu, Z., Fu, L., Niu, B., Li, W., 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. https://doi.org/10.1186/1471-2164-12-444.

Xu, W., Ge, Z., 2015. Application and optimization of biolog EcoPlates in functional diversity studies of soil microbial communities. In: MATEC Web of Conferences, https://doi.org/10.1051/matecconf/20152204015.

Zhang, S., Yin, Y., Jones, M.B., Zhang, Z., Kaiser, B.L.D., Dinsmore, B.A., Fitzgerald, C., Fields, P.I., Deng, X., 2015. *Salmonella* serotype determination utilizing hihg-throughput genome sequencing data. J. Clin. Microbiol. https://doi.org/10.1128/JCM.00323-15.

Zhu, S., Wang, H., Wang, C., Tang, L., Wang, X., Yu, K., Liu, S., 2013. Non-contiguous finished genome sequence and description of *Salmonella enterica subsp.* houtenae str. RKS3027. Stand. Genomic Sci. https://doi.org/10.4056/sigs.3767427.