2017

Phylogenetic and Associated Phenotypic Analysis of Salmonella Enterica Serovar Mbandaka

Linto Antony
South Dakota State University

PHYLOGENETIC AND ASSOCIATED PHENOTYPIC ANALYSIS OF *SALMONELLA ENTERICA* SEROVAR MBANDAKA

BY

LINTO ANTONY

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Biological Sciences

Specialization in Microbiology

South Dakota State University

2017

PHYLOGENETIC AND ASSOCIATED PHENOTYPIC ANALYSIS OF *SALMONELLA ENTERICA* SEROVAR MBANDAKA

LINTO ANTONY

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Radhey Shyam Kaushik — Date
Thesis advisor

Joy Scaria — Date
Thesis co - advisor

Volker S. Brözel — Date
Head, Biology and Microbiology

Dean, Graduate School — Date

For my beloved family

ACKNOWLEDGEMENTS

I would first like to thank my advisor Prof. Dr. Radhey S. Kaushik, for giving me an opportunity to be his student under his mentorship. I can hardly find words to express my deep appreciation for his valuable guidance and support that always helped me whenever I had troubles and questions about my research or writing.

I also would like to express my sincere thanks to my co-advisor Dr. Joy Scaria, for the help and warm encouragement that I have received from him. Without his constant help, I couldn't have dreamt of completing this project successfully. It's my pleasure to convey my special thanks to Dr. Feng Li and Dr. Guanghui Hua for being my committee members. I would also extend my immense gratitude to all my colleagues for their assistance and advice throughout the advancement of this project.

Finally, I must express my very profound gratitude to my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. At the very outset, I express my thanks to the Almighty God for all the blessings endowed on me.

Thank you.

Linto Antony.

TABLE OF CONTENTS

ABBREVIATIONS

ADRDL : Animal disease research and diagnostic laboratory

AMR : Antimicrobial resistance

ANOVA : Analysis of variance

ATCC : American type cell culture

ATP : Adenosine triphosphate

ATR : Acid tolerance response

BLAST : Basic local alignment search tool

CDC : Center for disease control and prevention

CDS : Coding sequence

CDT : Cytolethal distending toxin

CFU : Colony forming unit

CGH : Comparative genomic hybridization

cgMLST : Core genome multilocus sequence typing

CLSI : Clinical and laboratory standards institute

CO_2 : Carbon dioxide

COG : Cluster of orthologous groups

DALY : Disability adjusted life years

DMEM : Dulbecco's modified eagle medium

DNA : Deoxyribose nucleic acid

DT : Definitive type

EDTA : Ethylene diamine tetra acetate

ELISA : Enzyme-linked immunosorbent assay

ESBL : Extended spectrum beta lactamase

FAD : Fimbrial adherence determinants

FBS : Fetal bovine serum

FSIS : Food safety and inspection service

Fig. : Figure

FDA : Food and drug administration

g : Gram

GIT : Gastro intestinal tract

h : Hour

HCL : Hydrochloric acid

HGT : Horizontal gene transfer

i.e. : That is

InDels : Insertions and deletions

kb : Kilo base pair

kDa : Kilo Dalton

LB : Luria Bertani/ Lysogeny broth/Luria broth

LGT : Lateral gene transfer

LPS : Lipopolysaccharide

MB : Mega bases

Mbp : Mega base pair

MCRA : Most common recent ancestor

MDR : Multi drug resistance

MFS : Major facilitator superfamily

Mg^{2+} : Magnesium ion

MIG : Macrophage inducible gene

ml : Milliliter

MLEE : Multilocus enzyme electrophoresis

MLST : Multilocus sequence typing

MLVA : Multiple locus variable number tandem repeat analysis

MNV : Multi nucleotide variation

MOI : Multiplicity of infection

MST : Minimum spanning tree

N : Normality

ng : Nanogram

nm : Nanometer

NARMS : National antimicrobial resistance monitoring system

NCBI : National center for biotechnology information

NGS : Next generation sequencing

NTS : Non typhoidal *Salmonella*

OD : Optical density

OD_{600} : Optical density at 600nm

ORF : Open reading frame

PBS : Phosphate buffered saline

PCB : Protein cluster bins

PCR : Polymerase chain reaction

PFGE : Pulse field gel electrophoresis

PMQR : Plasmid mediated quinolone resistance

QC : Quality control

RNA : Ribonucleic acid

rpm : Revolutions per minute

SGI : *Salmonella* genomic island

SNP : Single nucleotide polymorphism

SNV : Single nucleotide variation

SPI : *Salmonella* pathogenicity island

Spp. : Species

SPV : *Salmonella* plasmid virulence

SRA : Sequence read archive

ST : Sequence type

Sub sp. : Sub species

SCV : *Salmonella* containing vacuoles

tRNA : Transfer RNA

TSI	: Triple sugar iron
T1SS	: Type I secretion system
T3SS	: Type III secretion system
USA	: United States of America
UK	: United Kingdom
USDA	: United states department of agriculture
VFDB	: Virulence factor database
VNTR	: Variable number tandem repeats
v/v	: Volume by volume
WGS	: Whole genome sequencing
WHO	: World health organization
µ	: Micron
µl	: Microliter
µg	: Microgram
%	: Percentage
~	: Approximate
^{0}C	: Degree centigrade
≥	: Greater than or equal to
<	: Less than
>	: Greater than
≤	: Less than or equal to
=	: Equal to
±	: Plus, or minus

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

PHYLOGENETIC AND ASSOCIATED PHENOTYPIC ANALYSIS OF *SALMONELLA ENTERICA* SEROVAR MBANDAKA

LINTO ANTONY

2017

Food borne salmonellosis is a global public health concern caused by *Salmonella*, that causes enteric disease both in humans and animals. Most of the pathogenic *Salmonella* serovars fall under *Salmonella enterica subspecies enterica*, a major subspecies group that includes more than 50% of total identified *Salmonella* serovars. New serovars are identified each year and overall incidence of salmonellosis may mask the outbreak incidences caused by individual serovars. Infrequently reported serovar outbreaks can be a significant threat to public health.

Salmonella enterica serovar Mbandaka is one of the infrequently reported causative agents of non-typhoidal salmonellosis in USA. But it has been considered as one of the frequent human *Salmonella* serovar in other countries such as European countries, Israel, Africa as well as in New Zealand. Published researches about this serovar were very limited and no prior studies have been reported about S. Mbandaka isolates from USA especially at the genomic level. Knowledge about the population structure and intra serovar genetic diversity that exists within the *S*. Mbandaka isolates in a global context remains a mystery.

This research was aimed to understand the population structure of globally distributed *S*. Mbandaka isolates with a hypothesis that isolates of this serovar from different

geographical area may be genetically close and form specific clonal groups. Genome sequence data of 465 isolates from different parts of the world were collected from NCBI database and were used to analyze diversity at single nucleotide level. Phylogenetic tree, created based on SNP analysis, revealed partitioning of isolates into two major clusters and six sub clusters. Cluster formation was validated by further analysis with core genome MLST. This research was also aimed to understand the antimicrobial resistance gene pattern and distribution of virulence factors in *S.* Mbandaka isolates from different isolation sources. Analysis of ability to invade host cells and resistance to low pH environment in 76 USA isolates showed no major difference in these phenotypic properties irrespective of isolation source. Overall this research provides a solid platform for the epidemiological investigation of future Salmonella outbreaks caused by serovar *S.* Mbandaka.

Chapter 1. Literature review and Experimental Objectives

1.1 The genus *Salmonella*

Salmonella, a member of *Enterobacteriaceae* family [2], is a common foodborne pathogen with global public health concern [3]. It has a wide host range from cold blooded animals to warm blooded animals and causes infectious disease both in humans and animals [4-6]. It was originally discovered by Eberth and cultured by Gaffky in 1884 from a typhoid patient [7]. But the first successful isolation was done by Theobald Smith and Daniel Elder Salmon from infected pigs with hog cholera (swine fever) in 1885 and the name *Salmonella* was given in honor of D. E. Salmon [8, 9].

1.1.1 Morphology and biochemical properties

The members of *Salmonella* are morphologically categorized as Gram-negative, non-spore forming, rod shaped intracellular pathogens. These facultatively anaerobic bacilli are having a size of 0.7 to 1.5μm width and 2.0 to 5.0 μm length [4, 10, 11]. Because of the presence of higher lipid content and thinner cell wall they are Gram-negative in nature while the presence of peritrichous flagella made most of them motile except *Salmonella enterica* serovars Gallinarum and Pullorum.

The genus *Salmonella* can grow both in the presence and absence of oxygen [12] and they are oxidase negative and catalase positive. Biochemically they can grow on simple media containing minimal requirements such as glucose for carbon energy source and ammonium ion for nitrogen source. Enrichment broths can be used for the lab scale identification of the culturing samples and can be isolated on selective media. Normally the bacterium ferments glucose with the production of gas [13]. The biochemical

identification of the colonies is possible by the method of Triple Sugar Iron (TSI) agar slopes and can be confirmed by slide agglutination of *Salmonella* somatic antigen groups with antisera. Some of the biochemical properties of most *Salmonella* spp. which help in their identification and serovar differentiation are the utilization of citrate, production of hydrogen sulphide from inorganic sulphur, decarboxylation of ornithine and lysine. Inability to deaminate tryptophan and phenyl alanine and failure to produce indole and beta galactosidase also help in this context.

Table 1. Different biochemical characters of *Salmonella* species and sub species

Species	***S. enterica***						***S. bongori***
Subspecies	*enterica*	*salamae*	*arizonae*	*diarizonae*	*houtenae*	*indica*	
Biochemistry							
Dulcitol	+	+	-	-	-	d	+
ONPG(2h)	-	-	+	+	-	d	+
Malonate	-	+	+	+	-	-	-
Gelatinase	-	+	+	+	+	+	-
Sorbitol	+	+	+	+	+	-	+
Growth with KCN	-	-	-	-	+	-	+
L(+)-tartrate(a)	+	-	-	-	-	-	-
Galacturonate	-	+	-	+	+	+	+

γ-glutamil-transferace	+(*)	+	-	+	+	+	+
β-glucuronidace	d	d	-	+	-	d	-
Mucate	+	+	+	- (70%)	-	+	+
Salicine	-	-	-	-	+	-	-
Lactose	-	-	- (75%)	+ (75%)	-	d	-
Lysed by phage O1	+	+	-	+	-	+	d

(a) Dextro rotatory (L-) tartrate; (*) *S. enterica* serovars Typhimurium (d), Dublin -; (+) 90% or more positive reactions; (-) 90% or more negative reactions; (d) different reactions by different serovars; Adopted from Patrick et al,[14].

1.1.2 Nomenclature and classification

Taxonomy of genus *Salmonella* is complex and has changed many times. Formerly, this genus was considered as a single species known as *Salmonella choleraesuis* [15]. Later, in 1987 Le Minor and Popoff suggested the novel name "*Salmonella enterica*", coined by Kauffman and Edwards in 1952 [5], instead of *Salmonella choleraesuis* to avoid confusion with serovar choleraesuis [15, 16]. In 1961, Fritz Kauffman used the somatic and flagella antigen analysis method developed by Bruce (in 1926) to distinguish more than 2000 serovars. Based on this method Kauffman - White in 1980 proposed current *Salmonella* nomenclature [14] and is currently maintained by World Health Organization (WHO) Collaborating Centre for Reference and Research on *Salmonella* at the Pasteur Institute, Paris, France (WHO collaborating Center) [5, 14].

Genus *Salmonella* consists of two species, *Salmonella enterica* and *Salmonella bongori* [17]. The major difference between these two species is that *S. enterica* has got *Salmonella* pathogenicity Island-2 (SPI-2) while *S. bongori* does not have this in its genome [18]. Presently, as per supplement 2008-2010 White-Kauffman-Le minor scheme (No.48), species *Salmonella enterica* is divided in to six subspecies [5, 19], the details of which are given in table.2. Out of 2659 serovars of genus *Salmonella*, 1586 of them are grouped into subspecies *enterica* [19] which is more than 59% of total isolated *Salmonella* strains [14].

Table 2. Present number of serovars in each species and subspecies of *Salmonella* [19]

Species and subspecies	Serovars(n)
S. enterica	2637
Subsp. enterica	1586
Subsp. salamae	522
Subsp. arizonae	102
Subsp. diarizonae	338
Subsp. houtenae	76
Subsp. indica	13
S. bongori	22
Total	2659

1.1.3 Typing of *Salmonella*

Bacterial typing is the analyzation of isolate specific phenotypic and/or genotypic characters, below the species/subspecies level that can be used to investigate bacterial

transmission pattern, and to find out source of infection. “Subtyping” is often used as synonym for typing. Typing helps in the surveillance of infectious disease, outbreak investigation, study of pathogenesis and course of infection, as well as in the study of bacterial population genetics [20]. There are several typing methods for *Salmonella*, some of them are discussed in the following section.

1.1.3.1 Bio typing

Bio typing helps to differentiate *Salmonella* spp. based on their biochemical characteristics and it has been used in several studies to discriminate strains of different *Salmonella* serovars [21, 22]. This typing method has excellent typeability, but variable discriminatory power. The advantages of this typing method are that it is technically easy, inexpensive and easy to perform in small laboratories [20].

1.1.3.2 Serotyping

The Kauffman and White scheme classified *Salmonella* serovars based on this traditional phenotypic typing method [19]. Serotyping differentiates *Salmonella* isolates based on the surface antigens – Somatic ‘O’ antigen (lipopolysachharide; encoded by *rfb* genes) and H antigen (phase 1 and phase 2 flagellar antigens; encoded by *fliC* and *fliB*). Isolates are assigned into serotypes or serovars depending on a bacterial agglutination test using a panel of antisera prepared against these antigens [23, 24]. Serotyping is the predominant method for laboratory based surveillance of *Salmonella* infections [25] and it has significant role in disease outbreak investigation [26]. However, multiple drawbacks of this method, including low throughput, high expense and need of high expertise, confined this labor intensive procedure to specialist reference laboratories [24].

1.1.3.3 Bacteriophage typing

Phage typing relies on the lytic pattern of bacterial isolates of interest that have been exposed to a defined panel of typing phages. This typing method has got variable discrimination, poor reproducibility and often partial typeability [20]. Limited discrimination of phage typing in some important serovars like *S.* Enteriditis due to phenomena like phage type conversion confounded the use of this typing method for epidemiological outbreak investigation [27, 28]. Along with other draw backs like requirement of production and continuous quality control of phages, and need of extensive expertise caused loss of its use as a reference typing method [20].

1.1.3.4 Plasmid profile typing

The basis of the plasmid typing profile in *Salmonella* is the difference in size, frequency and distribution of plasmids among various *Salmonella* serovars. Plasmids with size between 2- 150 kb can be taken by most of the *Salmonella* strains. The purpose of the plasmid profile typing in *Salmonella* is to inspect the clonality of strains, explain the transmission routes and in epidemiological surveillance [29, 30]. However, in case of strains without plasmid and the instability of plasmid content rendered it unsuitable as a reliable clonal marker in some bacterial studies. Moreover, plasmid typing alone is not strain discriminatory [20, 22, 31].

1.1.3.5 Pulsed - Field Gel Electrophoresis (PFGE)

This method is considered as the "gold standard" of molecular typing method few years before [32]. It is based on the electrophoretic mobility pattern of the bacterial chromosomal DNA fragments, formed by the digestion using restriction enzymes [20, 32, 33]. It has been widely used in the study of *Salmonella* organisms as an invaluable tool in

tracing outbreak strains as well as in the study of genetic relatedness. Potential draw backs of this technique such as low throughput, time consuming, labor-intensive and the requirement of high expertise for the reproducibility caused researchers to think about replacing this technique with other efficient typing methods.

1.1.3.6 Multilocus sequence typing

Maiden *et al,* who described MLST for the first time [34], proposed this technique as a “gold standard” for molecular typing and characterization of almost all pathogenic and non-pathogenic bacterial species as it avoids the drawbacks of traditional typing methods like ribotyping, pulsed field gel electrophoresis (PFGE), and PCR with repetitive element primers or arbitrary primers [34, 35]. It is an updated version of Multilocus Enzyme Electrophoresis (MLEE) [36], where the variation in the house keeping core genes is indexed indirectly based on the difference in the electrophoretic mobility of the gene products on starch gels [34, 37]. In contrast, MLST analyzes the nucleotide change in the internal fragment of selected housekeeping genes [35].

MLEE was never generally accepted by microbiologists as it included unrelated isolates to the main lineage [24]. One of the serious drawbacks of these gel based methods is the difficulty to compare the results achieved by different laboratories [35]. MLST identifies the variation in the nucleotide sequence of gene fragment present in the multiple housekeeping loci. For a given locus, all unique sequences are assigned an allele number. The allele numbers of all loci for a given isolate are combined in to allelic profile and assigned a Sequence Type (ST) designation. This ST can be compared with other isolates. Isolates with identical STs or STs that differ at a few loci are considered as closely related [38]. The general procedures in MLST analysis is shown in Figure 1.

MLST is one of the preferred *Salmonella* typing method. MLST scheme has got some advantages over conventional serotyping as it provides the true phylogeny of the analyzed isolates [39]. Previous studies have indicated that lateral gene transfer event results in the emergence of new serovars. Conventional serotyping thus fails to recognize the relationship difference between the new serovar and its predecessor [39]. Researchers recommend replacing conventional serotyping by MLST or its equivalents as these recognize evolutionary groupings [24]. A few years before PFGE was used for genotyping bacterial strains, but it has got limited firmness of purpose for genetic differentiation of highly clonal strains [40]. Since MLST involves direct comparison of nucleotide sequences, the interpretation and comparison of the data is easy and readily accomplished. This sequence based isolate characterization identify the genetic variation between isolates with high discrimination which is not possible in PFGE as the later requires additional data to understand the genetic reason for the pattern change [41]. Over and above, the portability of MLST data, convenience in the data storage and data comparison between laboratories gave this high-resolution technique an upper hand over other traditional *Salmonella* typing methods. Although MLST has high resolution for typing of any bacterial pathogen, use of this method is limited in case of monomorphic pathogens like *S.* typhi and Paratyphi A. These pathogens show very few polymorphic sites and hence, fine typing for epidemiological studies of these bacteria demands the use of other typing methods with MLST.

.

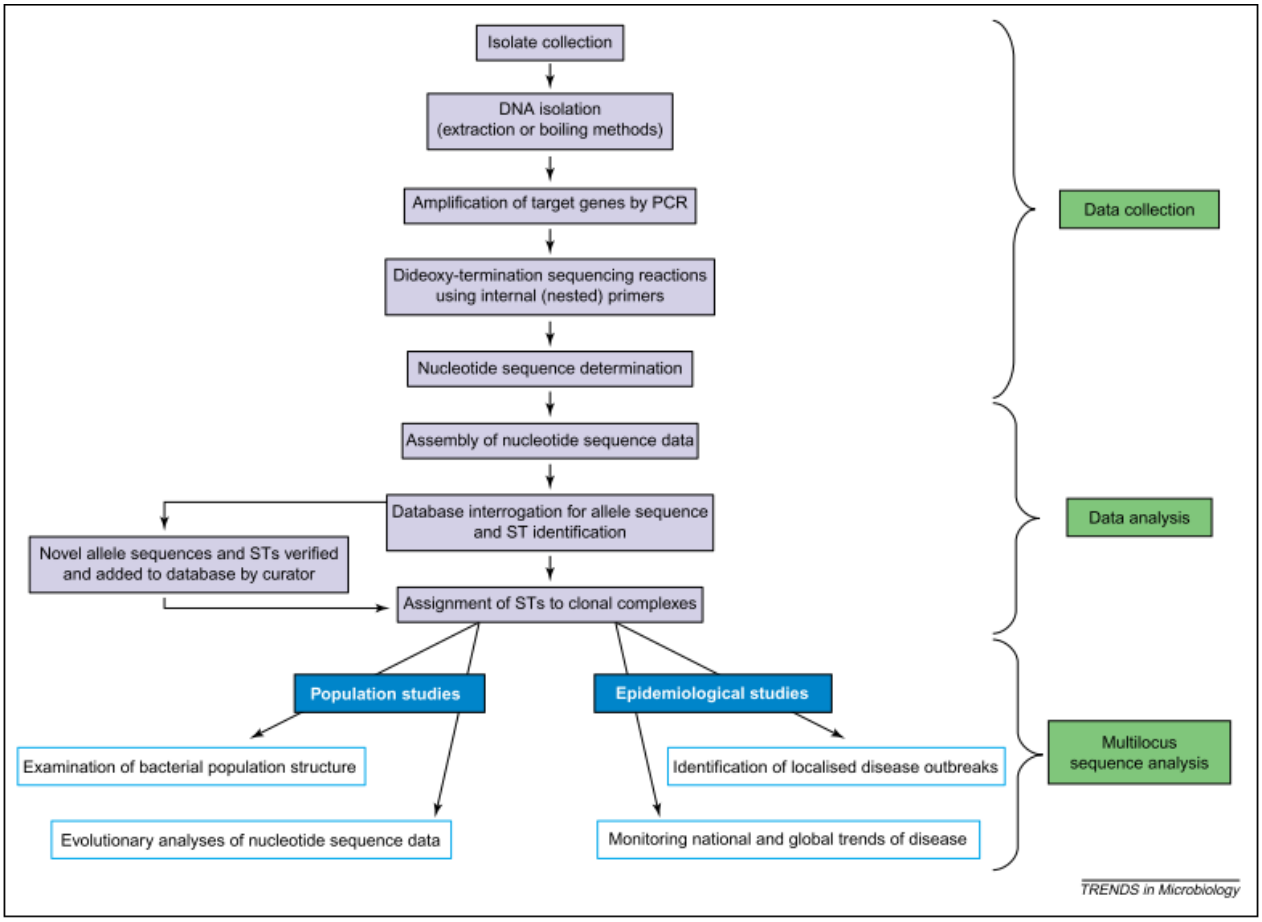


Fig.1. Multilocus Sequence Typing (MLST) and analysis. adopted from Urwin et al, 2003 [38]

1.1.3.7 Multiple-locus variable number tandem repeat analysis (MLVA)

Prokaryotic organisms have interspersed repetitive sequences that are widely distributed in their genome. These low copy number non-coding DNA repeats are called as Variable Number Tandem Repeats (VNTR) [42, 43]. MLVA is a PCR based bacterial typing which analyzes the multiple VNTR loci on the chromosome and assigns the profile based on the number of repeats in each locus [44]. Studies have shown that MLVA is faster to perform, easy to analyze and more discriminatory than PFGE that it

can be used for surveillance and outbreak analysis of pathogens like *S.* Typhimurium [45].

1.1.4 Typhoidal *Salmonella*

Salmonella serovars which cause disease in humans and animals can be grouped into Typhoidal and Non Typhoidal *Salmonella*. Despite the fact that most of these serovars belong to subsp. *enterica* and they share genetic similarity, these two groups are characterized by host specificity and clinical syndrome. Serovars *S.* Typhi, Paratyphi A, B or C and Sendai are referred as Typhoidal *Salmonella* serovars which are highly adapted to human and cause disease 'enteric fever' otherwise known as 'Typhoid or Paratyphoid fever' [46-48]. Typhoid fever (caused by *S.* Typhi) and Paratyphoid fever caused by (*S.* Paratyphi) are acute, life threatening febrile illnesses resulted by fatal systemic infection of these pathogens [49].

It has been estimated that in the year 2000 Typhoid fever caused more than 2.7 million illnesses and 0.2 million deaths worldwide. Global estimation of illnesses caused by paratyphoid fever was 5.4 million [48, 50]. A revised estimate by Buckle *et al* in 2012 suggested that in 2010, there were an estimated 13.5 million typhoid fever episodes globally [49].

1.1.5 Non Typhoidal *Salmonella* (NTS)

Salmonella serovars coming under this group are considered as generalist pathogens with broad specificity. Only a small number of serovars are commonly responsible for human non typhoidal *Salmonella* infection. Epidemiologically important serovars included in this category are *S.* Typhimurium, and *S.* Enteritidis [51, 52]. The global burden caused by NTS has been estimated to be 93.8 million illnesses and 155,000 deaths

each year [53]. Gastroenteritis, bacteremia and subsequent focal infections are the major disease manifestation of these food borne pathogens [15]. Non typhoidal *S*almonella are considered as important zoonotic pathogens as they can easily transmit from animals to human [54].

1.1.6 Host Range

The species *S. enterica* is found in reptiles and warm blooded animals [55] while the species *S. bongori* is predominantly associated with cold-blooded animals [56]. *S. enterica* subsp. *enterica* can infect broad range of hosts. Generally, based on hosts infected, we can divide members of subsp. *enterica* as host adapted and host ubiquitous (non-adapted) [55]. Host adapted serovars can again be divided into host restricted and host specific serovars. Host restricted serovars can infect only a narrow selection of hosts while host specific are those which can infect only one specific host. *S.* Typhi, Paratyphi A, Paratyphi B, and Paratyphi C are human specific *S. enterica* serovars predominantly causing Typhoid fever [49, 57]. Examples for another host specific serovars are *S.* Sendai, *S.* Gallinarum and Abortusovis, these are exclusively associated with systemic illnesses in human, fowl [58] and ovines respectively [55, 59, 60]. *S.* Dublin and *S.* Choleraesuis are generally associated with systemic disease in ruminants and pigs but may also cause infection in other mammalian hosts infrequently [55, 61, 62]. Host non-adapted serovars induce a self-limiting gastroenteritis in a broad range of unrelated host species. The important pathogenic serovars of this ubiquitous group are *S.* Typhimurium and Enteritidis which are also referred to as un-restricted serotypes [55, 63]. *S.* Typhimurium is a broad host range serotype frequently associated with disease in numerous species including human, livestock, rodents and birds [64]. Domestic animals

are the major reservoir for the food borne spread of these unrestricted serovars, which account for the high incidence of global non-typhoid *Salmonella* infections. Based on host specificity we can also classify *Salmonella* serovars as host specialist (example- *S.* Typhi) and host generalist (example – *S.* Typhimurium) in a broader sense [63].

1.1.7 Important genetic features

Approximately 100 million years ago, *Salmonella* diverged from *Escherichia coli*. Presence of large number of additional virulence genes distinguished *Salmonella* from this closely related species [65, 66]. With an approximate size of 4.5 – 5 Mbp *Salmonella* chromosome contained nearly 4500 genes, of which nearly 80% genes were shared with genetically similar sister species *E.Coli*. Albeit, study of gene content of well characterized *Salmonella* serovars showed the presence of SPI1-5 with approximately same distribution similarities between these serovars. Absence of these SPI loci in *E. coli* indicates that they were acquired soon after the divergence of *Salmonella* from *E. coli* [67, 68].

Pangenome analysis of *S. enterica* genomes suggested that any *Salmonella* strain has a stable large core genome and abundance of accessory genes composed of SPI, transposible elements, phages and plasmid DNA [65]. Pairwise comparison of the *Salmonella* serovars reveal that 10-12% of the genome is unique to each serovar. Genes in these unique region contribute to different abilities of the serovars to infect a variety of hosts [67]. These unique genes may be acquired through a process called Lateral or horizontal Gene Transfer (LGT). Conjugation, Transduction and transformation are the three mechanisms involved in the transfer of genetic material between bacterial cells.

Horizontal gene transfer is considered as the major contributor to *Salmonella* evolution, adaptation as well as emergence of novel serovars [39, 69-71].

1.1.8 *Salmonella* virulence

Pathogenicity of *Salmonella* depends on the virulence of the strain which in turn is determined by so-called virulence factors. These genes are grouped into two classes. The first class includes the genes that are involved in physiological process like nutrient uptake, synthesis of proteins and factors that protect the bacteria from stressful environment like acidic environment in the stomach and anaerobic environment in the GIT, cell maintenance and damage repair. These genes are generally found in both pathogenic and non-pathogenic bacteria. Examples for these genes in *Salmonella* are regulatory gene phoP/phoQ, and biosynthetic genes such as aroA. Second class of virulence genes are unique to pathogens and rarely seen in non-pathogenic organisms. This class of genes include classic virulence factors such as endotoxin, exotoxin, fimbriae, flagella and virulence plasmids [72]. Majority of virulence genes are clustered in regions called SPI distributed over the chromosome. These gene clusters are thought to have been acquired by horizontal gene transfer [72-74].

Currently 22 SPIs have been Identified from the genomes of *S. enterica* and *S. bongori* [56, 75]. Distribution of these 22 known SPI are shown in the Table 3. There are some distinguishing differences between SPIs of *S. enterica* from *S. bongori*. SPI-2, encodes Type III Secretion System -2 (T3SS-2), is absent in *S. bongori* [56, 76, 77]. In *S. enterica* SPI-3 occupies as a single element but it exists as two independent elements in *S. Bongori*, SPI – 3A and SPI – 3B. Instead of two regions of SPI-5 in *S. enterica*, SPI-5 in *S. bongori* composed of only one region [56].

The most prominent SPIs are SPI-1 to SPI-5 and SPI7. A brief description of these SPIs is given in the following section and distribution of SPI in *Salmonella Spp.* are shown in Table 3.

SPI-1 has a size of 40-kb and encodes a T3SS which helps in uptake of bacteria by the host cell. This T3SS transport effector proteins like actin binding proteins SptP and SopE into the cytoplasm of the eukaryotic target cell leading to the rearrangement of the cytoskeleton and thereby induce the uptake of bacterium [78]. SPI-2 encodes a second T3SS and the genes present in this island help the bacteria in intracellular replication, survival and persistence within *Salmonella* Containing Vacuoles (SCV) by evading the oxidative damage. Effector proteins of SPI-2 also take part in modulation of the vesicular transporting of the host cell [79-81]. Important effectors encoded by SPI-2 are – *spi*C, *sse*F, and *sse*G. SpiC prevent fusion of lysosomes with SCV [82]. SseF and SseG alter exocytosolic transport processes and recruit exocytic transport vesicles to SCV [83]. Both SPI-1 and SPI-2 gene expressions are regulated by global regulatory systems such as OmpR/EnvZ and PhoP/PhoQ [84, 85]. These regulators modulate the gene expression by activation of *hilE* or *hilD* two important mediators of SPI-1 and SPI-2.

The 17-kb large SPI-3 haboring 10 Open Reading Frames (ORF) has been identified by Blanc-potard and Groisman at the *selC* tRNA locus in *S*. Typhimurium [74, 86]. These ORFs are organized in 6 transcriptional units out of which one is an important virulence factor called *mgt*CB operon. Presence of this operon is essential for intramacrophage survival and growth under low Mg2+ environment [86].

SPI-4 was first described by Wong et al in 1998 [87]. This 27-kb pathogenicity island located at centisome 92 contains 6 open reading frames. SPI-4 is likely to carry a

Type 1 Secretion System (T1SS) involved in secretion of a large (595-kDa) novel protein SiiE encoded by *siiC, siiD,* and *siiF*. SiiE is non-fimbrial adhesion protein that helps *Salmonella* for epithelial cell surface adhesion [88, 89].

SPI-5 appears to be conserved in *Salmonella* Spp. Genes encoded in SPI-5 are necessary for enteropathogenic phenotype. Out of 6 novel genes present in this SPI, *sop*B plays a vital role in cytoskeleton rearrangement and bacterial entry into host cell [90]. These genes have been shown to be involved with fluid secretion and inflammatory responses [91].

SPI-7 is a 134-kb size mosaic pathogenicity island that has been found to be specific in serovars Typhi, Paratyphi C and Dublin. This SPI encodes virulence factors like SopE and ViaB [92].

Other factors responsible for *Salmonella* virulence:

Salmonella virulence plasmids : They vary in size between 30 and 100-kb and 1-2 copies per cell are present in serovar Typhimurium, Enteritidis, Choleraesuis, Dublin, Gallinarum and Typhi, but not all the isolates of these serotypes carry the virulence plasmid [93]. All plasmids contain a virulence locus called *spv* locus. Expression of genes in *spv* locus thought to be play roles in intra cellular multiplication of *Salmonellae* and destabilization of eukaryotic cytoskeleton [94, 95].

Toxins: *Salmonella* produce both endotoxin and exotoxin. Exotoxin is again divided into cytotoxins and the enterotoxins. Endotoxin is the lipid portion of outer membrane lipopolysaccharide (LPS). Examples for exotoxins of *Salmonella* include Shigella dysenteriae-1 like cytotoxin, salmolysin and heat labile *Salmonella* enterotoxin [74].

Table 3. Distribution of known SPIs in the *Salmonellae*. Adopted and modified from [56]

Genomic Island	*S. bongori 12419*	S. Typhi CT18	S. Typhimurium LT2	S. Enteritidis P125109	S. Gallinarum 287/91	S. Cholerasuis SC-B67
SPI-1	+	+	+	+	+	+
SPI-2	–	+	+	+	+	+
SPI-3	%	+	+	+	+	+
SPI-4	+	+	+	+	+	+
SPI-5	£	+	+	+	+	+
SPI-6	–	+	+	+Ψ	+Ψ	+
SPI-7	–	+	–	–	–	–
SPI-8	–	+	–	–	–	–
SPI-9	+	+	+	+	+	+Ψ
SPI-10	–	+	–	+Ψ	+Ψ	–
SPI-11	V	V	V	V	V	+
SPI-12	–	+	+	+	+	+
SPI-13	–	–	+	+	+	+
SPI-14	–	–	+	+	+	+
SPI-15	–	+	–	–	–	–
SPI-16	–	+	+	+	+	V
SPI-17	–	+	–	+	+	–
SPI-18	–	+	–	–	–	–
SPI-19	–	–	–	+Ψ	+	–
SPI-20	–	–	–	–	–	–
SPI-21	–	–	–	–	–	–
SPI-22	+	–	–	–	–	–

+, all genes are conserved; -, entire SPI is missing; V, partially present; Ψ, pseudogenes are contained on SPI; %, SPI-3 present as two independent genomic islands in *S. bongori*: SPI-3A and SPI-3B; £, only half of this island is represented in *S. bongori*.

Fimbrial clusters: Fimbriae are the bacterial surface protein structures made up of helically arranged protein fimbrin. Fimbriae are important in virulence as these structures help bacteria to cell adhesion [96]. Proteins for the biosynthesis, structure and assembly of fimbriae are encoded by 8-11 genes clustered in a 7-9 kb large operon [97].

1.1.9 Antimicrobial Resistance in *Salmonella*

Indiscriminate use of antimicrobial drugs in animal feed to promote the growth of food animals is one of the key reason for the development of Antimicrobial Resistance (AMR) in organisms like *Salmonella* [98]. Acquisition of these resistant *Salmonella* organisms through food chain confound with the treatment of both typhoidal and non-typhoidal salmonellosis using conventional antibiotics. This has been associated with increased morbidity, mortality, hospitalization as well as high economic lost [99]. Hence the emergence of antimicrobial resistance strains has become a serious threat to public health. Studies in the previous years have shown that there is world wide spread, in both of developed and developing countries, of *Salmonella* strains belonging to different serovars and showing multiple antimicrobial resistance. Most of the strains are zoonotically important. They acquire resistance from food animals and cause infection in human through food chain [54, 100]. There are different mechanisms for resistance development in a bacterium. Previously susceptible bacteria become resistant to a particular antimicrobial agent by one of the following methods: mutation, selection and/or resistance gene acquisition from other bacteria. These processes will attribute to expression of enzymes that destroy the antimicrobial agent, activation of efflux pumps that extrude the antibacterial agent, alteration of cell permeability to these agents and modification of drug targets [101]. Plasmids, transposons, gene cassettes and mobile genetic elements are the mediators of resistance gene transfer between bacteria [102].

Horizontal transfer of plasmid containing resistance genes happens via conjugation. Acquisition of several resistance genes may happen by this method [103]. In plasmids, resistance genes are located in genetic elements called transposons. Gene expression

element called integrons are responsible for the recruitment and expression of resistance genes. Integrons are located within transposons and encodes a recombinase called integrase. Integrase mediated site-specific recombination mechanism is an efficient genetic mechanism that helps bacteria to acquire resistance genes [54]. Critical relevance of plasmid mediated antimicrobial resistance in *Salmonella* is explained by following reports. In 1986 incidence report of both Apramycin and Gentamicin resistant *S.* Typhimurium Definitive Type 204c (DT204c) from calves as well as in human provided the evidence that use of antibiotic (Apramycin) in animal husbandry gave rise to gentamicin resistance, an antibiotic used in human as well [54, 104, 105]. In 1990s, several reports pointed out the increased frequency of *Salmonella* with plasmid conferring resistance to Apramycin and Gentamicin, as well as β-lactam resistance *Salmonella* in European countries [106, 107]. Broad spectrum Cephalosporin resistant *Salmonella* emergence was reported by National Antimicrobial Resistance Monitoring System (NARMS) in 1999. Several *Salmonella* isolates with plasmid mediated cephalosporin resistance encoded by CMY-2 AmpC β-lactamase were isolated from animals and humans in different states of United states in 1990s [108, 109]. This list does not end here, but even this much implies that plasmid mediated antimicrobial resistance in *Salmonella* is a global problem irrespective of place, host and serotype.

In several *Salmonella* serovars, genes conferring antibiotic resistance are also present in an integrative mobilizable element of *Salmonella* chromosome known as *Salmonella* Genomic Island (SGI). 42.4-kb size SGI 1 from *Salmonella* serovar Typhimurium DT104 is the first completely sequenced and well-studied genomic island which carries five resistance genes conferring resistance to seven antibiotics; ampicillin,

chloramphenicol, florfenicol, streptomycin, spectinomycin, sulfamethoxazole, and tetracyclin [110, 111]. SGI1 structure involves 27.4-kb backbone and a 15kb complex class1 integron. SGI1 is found integrated in the end of *thdF* gene with integron located adjacent to resolvase encoding *res* gene. SGI2 is a SGI1 related SGI with same SGI1 structure but the integron is in the S023 gene. SGI2 has been identified in serovar Emek and Virchow isolates [112].

1.1.9.1 Some important antimicrobial genes in *Salmonella*

Out of five Tetracycline resistance gene identified so far, *tet* (A), *tet* (B), *tet* (C), *tet* (D), *tet* (E), two of them (*tet* (A) and *tet* (B) are found in transposons. *Tet* (G) was found in *Salmonella* Genomic Isalnd 1(SGI1) or SGI2. Phenicol resistance in *Salmonella* is offered by type A (*cat A1* and *cat A2*) and type B (*cat B2*, *cat B3* or *catB8*) chloramphenicol acetyl transferase genes, chloramphenicol exporter genes (*cmlA1* and *cmlA4*) as well as chloramphenicol/florfenicol exporter genes *cmlA9* and *floR* [113].

There are 10 different genes known for aminoglycoside resistance in *Salmonella*. These genes encode for aminoglycoside-3”-O-adenyl transferases and are part of SGI1 or SGI2 MDR (Multi Drug Resistance) gene clusters or located on gene cassettes in class 1or class2 integrons. They - *aadA1, aadA2, aadA5, aadA6, aadA7, aadA12, aadA21, aadA22, aadA23, aadA24, aadA26, and aadA2* - confer resistance to streptomycin and spectinomycin. Other aminoglycoside resistance genes known to found in *Salmonella* include Streptomycin phosphotransferase genes - *str*A and *str*B – streptomycin resistance, aminoglycoside-2”-O-adenyltransferase gene *aadB* – gentamicin, Kanamycin and tobramycin resistance, aminoglycoside-N-acetyltransferase genes - *aacC* and *aacA* - gentamicin resistance [113-116].

β-lactam resistance in *Salmonella* is offered by a broad range of genes which were known to code for at least 13 different types of β lactamases [113, 117, 118]. There are 17 different *dfrA* genes and one *dfrB* have been reported in *Salmonella* for trimethoprim resistance [115]. *sul1, sul2, and sul3* are the three sulfonamide resistance genes reported in *Salmonella* so far [113]. There are some plasmid mediated genes, which code for DNA topoisomerase protecting proteins are responsible for quinolone resistance in *Salmonella*. These PMQR (Plasmid Mediated Quinolone Resistance) genes include *qnrD, qnrA, qnrB* and *qnrS* variants. qepA coding for a quinolone-specific efflux pump, gene *aac(60)-Ib-cr* nad mutations in the *gyrA, gyrB, parC* and/or *parE* are the other reported quinolone resistance genes in *Salmonella* [118, 119].

1.1.9.2 Antimicrobial resistance in Non-typhoidal *Salmonella* - Trend in USA

Based on the data available from NARMS for a period of 2004-2013, Micheal *et al*, summarized the trend of antimicrobial resistance in NTS (Fig.2) [113]. According to this data the percentage of fully susceptible NTS isolates in USA varied between 79.9% (2004) and 80.8% (2013). As per latest NARMS report antimicrobial resistance in *Salmonella* varies by serovar.

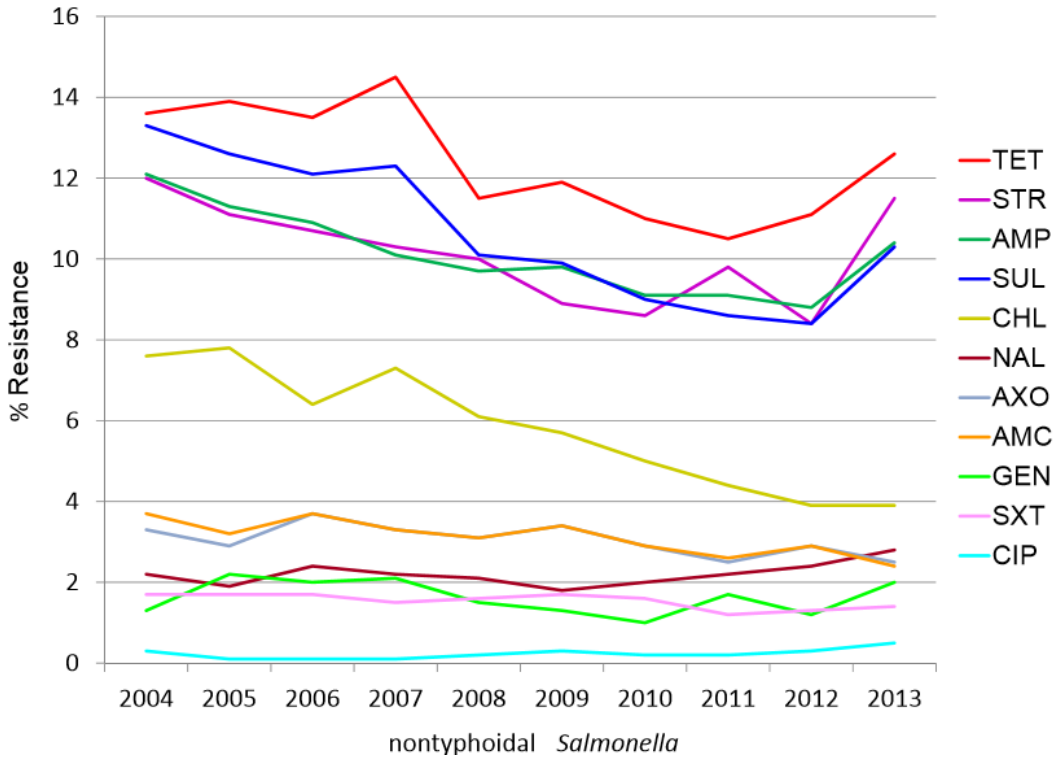


Fig.2. Summary of trends in resistance to various antimicrobial agents or combinations of agents detected among nontyphoidal *Salmonella* isolates in the USA during 2004-2013. Summarized data from NARMS. AMC, amoxicillin/clavulanic acid; AMP, ampicillin; AXO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; GEN, gentamicin; NAL, nalidixic acid; STR, streptomycin; SUL, sulphonamides; SXT, sulfamethoxazole / trime - thoprim; TET, tetracycline. Adopted and Modified from [113].

1.1.10 Host cell invasiveness

Pathogenicity of *Salmonella* depends upon the ability to invade host intestinal epithelial cells [120-122]. *Salmonella* can invade both phagocytic and non-phagocytic cells. Studies have shown that upon infection in non-phagocytic host cells, invasive *Salmonella* induce membranous ruffles formation at the site of bacterium-host cell interaction. This will facilitate the uptake of both invasive and non-invasive *Salmonella* by a mechanism that is distinct from receptor mediated phagocytosis [123]. This *Salmonella* host cell invasion process via trigger mechanism depends upon T3SS-1

encoded by SPI-1. The survival inside the host cell depend on T3SS-2 encoded by SPI-2 [80, 124, 125]. Few years before in vitro studies have shown that *Salmonella* can invade non-phagocytic cells via another T3SS-1 independent mechanism called Zipper mechanism. *Salmonella* thus became the first bacteria shown to be able to invade host cells via both Trigger and Zipper mechanism. Rck invasin, a protein encoded by the *rck* gene located on the large virulence plasmid, is the key protein in the zipper mechanism of *Salmonella* cell invasion [126, 127]. This protein is highly conserved in *S.* Enteritidis and *S.* Typhimurium and could be found in *S.* Dublin but never found in other serovars [127, 128]. Membrane ruffling in trigger entry mechanism is mediated by a cocktail of effector proteins called T3SS-1 effectors (SipA, SipC, SopB, SopE, SopE2). These effector proteins directly or indirectly modulate actin activity and cause cytoskeletal rearrangement [129-131].

1.1.11 Tolerance to Acidic environment

Salmonella spp. have well developed regulatory networks that deploy mechanisms to sense and respond to environmental changes and thereby protect the organism from a range of biotic and abiotic stresses. Food processing and storage, defense mechanisms of the hosts like extreme stomach pH, presence of bile salts in the gut, and low oxygen tension in the intestine are examples of these stress conditions that these organism constantly face during their life cycles [132]. Acid Tolerance Response (ATR) is an adaptive mechanism elicited by *Salmonella* spp. to survive in a low pH environment. ATR was first described and had been well studied in *S.* Typhimurium by Foster et al. in 1990 [133-135]. ATR is a phenomenon where, if adapted to a low pH (5.5 to 6.0) *S.* Typhimurium has been found to survive extreme low pH (pH 3.0 to 4.0) [133, 136].

Several studies suggested relationship between ATR and pathogenicity of bacterial strain. Wilmes - Riesenberg et al. explained the role of acid tolerance response in virulence of *S.* Typhimurium[137]. In another study, de Jonge et al. reported greater ATR for a group of highly virulent *S.* Typhimurium DT104 strains [138]. ATR thus has significant importance in case of foodborne pathogens like non-typhoidal *Salmonella* as these organisms should survive low pH environment in stomach as well as in intracellular environment to make infection in host.

1.1.12 Pathogenesis

Enteric fever caused by typhoidal *Salmonella* is an endemic disease in developing countries. Lack of sanitation and hygiene facilitate the fecal - oral route transmission of these pathogens [46, 139]. NTS infections is a worldwide concern. Person to person contact, contact with animals as well as consumption of contaminated animal and plant food products are the important transmission route of NTS pathogens [140-144]. Both groups of bacteria, after reaching the intestine, adhere to the host intestinal epithelial cells, which is then followed by cell invasion. As explained before, interaction of *Salmonella* organism with non-phagocytic host cells induces its own entry to the cells either by trigger entry or by zipper entry mechanism [123, 126]. SPI-2 encoded T3SS-2 effector proteins help the bacteria to survive inside the *Salmonella* Containing Vacuoles (SCV) [80, 81]. NTS serovars cause profound intestinal inflammation and induce massive neutrophil infiltration [145]. In contrast, Typhoidal *Salmonella* serovars induce a low intestinal inflammatory response characterized by negligible neutrophil recruitment during the initial invasion stage [146, 147]. This low inflammatory response will facilitate the bacteria to invade deeper tissue of intestine and infection become systemic

with bacterial dissemination to mesenteric lymph node, liver, spleen and gall bladder [46, 148, 149].

1.1.13 Clinical manifestation of salmonellosis

The status of the host determined by age, genetic and environmental factors and status of the bacterium determined by virulence factors influence the outcome of *Salmonella* infection [74]. Four different clinical manifestations of human salmonellosis are enteric fever, gastroenteritis, bacteremia and other extra intestinal complications, and chronic carrier state [150]. Since the symptoms of typhoid fever and paratyphoid fever are indistinguishable, the term “Enteric Fever” is used collectively for both fevers [151, 152]. The average incubation period for typhoidal serovars is 14 days with symptoms, such as headache, abdominal pain, diarrhea or constipation and fever, persisting for up to 3 weeks [153, 154]. Other symptoms are nausea, anorexia, hepatosplenomegaly, chills, rose spots, and dry cough [155]. In contrast to enteric fever, NTS infections have a shorter incubation period of 6-12 h and symptoms last less than 10 days [156]. NTS infection cause self-limiting acute gastroenteritis and watery diarrhea accompanied by vomiting, nausea, abdominal cramps, and fever [157]. Cholecystitis, pancreatitis and appendicitis are the other complications of enteritis caused by NTS infection [15]. In case of typhoidal salmonellosis after the acute stage of infection the infected person may act as a chronic carrier and shed the bacteria through stool [154, 158]. Carrier state is less frequent in NTS infection as the primary reservoir of NTS is animals, instead of humans [15].

1.2 *S.* Mbandaka - An Overview

S. Mbandaka is a non-typhoidal food borne pathogen causing diarrheal disease with diverse range of host including human. It was first isolated from Belgian, Congo in 1948. It was later classified as one of the top ten serovars causing human salmonellosis in European Union [159]. Studies have shown that *S.* Mbandaka (clone ST413) could be able to survive more than 15 years in Poland and successfully spread between poultry, feeds, food and finally human. *S.* Mbandaka was represented as one out of the 10 most predominant serovars in Poland after its occurrence in humans in1997 [159]. In Australia, first identification of this serovar as a cause of human infection was in 1978. Later several outbreaks were reported during the period 1985 to 1996 [160]. In USA, reports of human outbreak caused by this serovar were very rare. In1999, one of the major outbreak caused by *S.* Mbandaka happened in Oregon, Idaho, Washington and California. Upon investigation the source for this multistate outbreak was identified as Alfalfa sprouts and ungerminated seeds from a farm in southern California [161]. Data analysis of food borne disease outbreak caused by *Salmonella* enterica serovars and associated food commodities in United States during the period 1998 to 2008 by Jackson et al., pointed out that sprouts were the common food commodity for *S.* Mbandaka outbreak. Recent multistate human *Salmonella* outbreaks (2013 and 2016 outbreaks) gained some epidemiological importance to this serovar. Upon investigation it was found out that sources of these 2013 and 2016 outbreaks were tahini sesame paste and live backyard poultry respectively [162, 163].

This serovar was isolated with similar frequency both in USA and UK [75]. Cattle, chicken and pigs are the major isolation source for *S.* Mbandaka in USA with an isolation frequency of 27%, 25% and 14% respectively. While in UK, 65% incidence are from

chicken and 20% from cattle [164]. In UK, *S.* Mbandaka was the most frequently isolated *Salmonella* serovar from animal feed during the year 2013 [165].

In an epidemiological study of *S.* Mbandaka strains isolated from animals, feed and human sources in Poland, Hoszowski et al suggested that *S.* Mbandaka got into chicken through animal feed and then to human via food chain [166]. Later in an article, Hayward et al. showed that isolates of *S.* Mbandaka in UK comprised of a single clonal lineage. This clonal lineage has the ability to form biofilm at 25^{o}C and to utilize the metabolites of the soya bean. Based on these findings they suggested that this serovar might be well adapted to survival *ex vivo* like in animal feed [165].

Not many genomic studies or reports are available specifically about serovar *S.* Mbandaka and its strains. First sequencing of *S.* Mbandaka strains (two strains- M1 and M2) was achieved by Hayward et al in 2013. These were UK strains isolated from cattle in 2008 and 2009. According to this study, total length of both M1 and M2 strains genome were 4.72Mb nucleotides with a GC skew of 51.91% and 52.01% respectively. Number of genes predicted were 4616 and 4619 genes respectively. The important feature they observed in the genome of these strains was the presence of 860kb sequence inversion region, that codes for 909 genes, located between base 1086415 and 1947250 of M1, and 1132370 and 1992477 of M2. They reported a less diverse intra serovar difference between these serovars which was attributed by phage associated genes [75].

Large scale phylogenetic analysis of 78 serovars in *Salmonella enterica* subsp. *enterica* by Timme et al. showed two major sister lineages Clade A and Clade B. There were two sub lineages in Clade A: A1 and A2. *S.* Mbandaka was grouped into sublineage A2 close to sector Typhi, a strong cluster of isolates of serovar *S.* Typhi and *S.* Paratyphi

[167]. In another study involving UK isolates of *S*. Mbandaka reported that those isolates were comprised of one clonal lineage [165]. No studies have been found in our search regarding phylogenetic analysis of USA strains of *S*. Mbandaka. As mentioned before advent of next generation sequencing made possible the global availability of sequence data of *S*. Mbandaka strains. But so far, no phylogenetic studies have been reported in large scale including strains of *S*. Mbandaka from different parts of the world.

1.3 Concepts of Genomics approaches

1.3.1 Whole Genome Sequencing (WGS)

The era of genome sequencing began with the sequencing of phi X 174 bacteriophage genome in 1977 by Sanger et al. Approximately 5375 nucleotide size, single stranded circular DNA genome was sequenced by Sanger's "plus and Minus" method [168]. Genome sequencing methods then evolved through W.M. Barnes ribo substitution method, Maxam Gilberts sequencing method based on chemical degradation of DNA and through Sanger's chain termination method [169] and it reached present era of NGS. In 1995 sequencing of *Haemophilus influenza* genome, the first bacterial genome sequenced, opened the gates of microbial genomics world. It was sequenced using random shotgun sequencing approach based on Sanger's chain termination method [170]. In a review article by Ronholm *et al* the microbial WGS has been categorized into three generations (Fig.3) [171]. Automated Sanger sequencing introduced by ABI replaced the earlier Sanger method of using radiolabeled dideoxy nucleotides with fluorescent labelled bases. This relatively high through put sequencing platform formed the first-generation sequencing and was used to sequence first finished human genome

[171]. This was later followed by second and third generation sequencing platforms each with slightly different approaches in DNA preparation, sequencing and analysis.

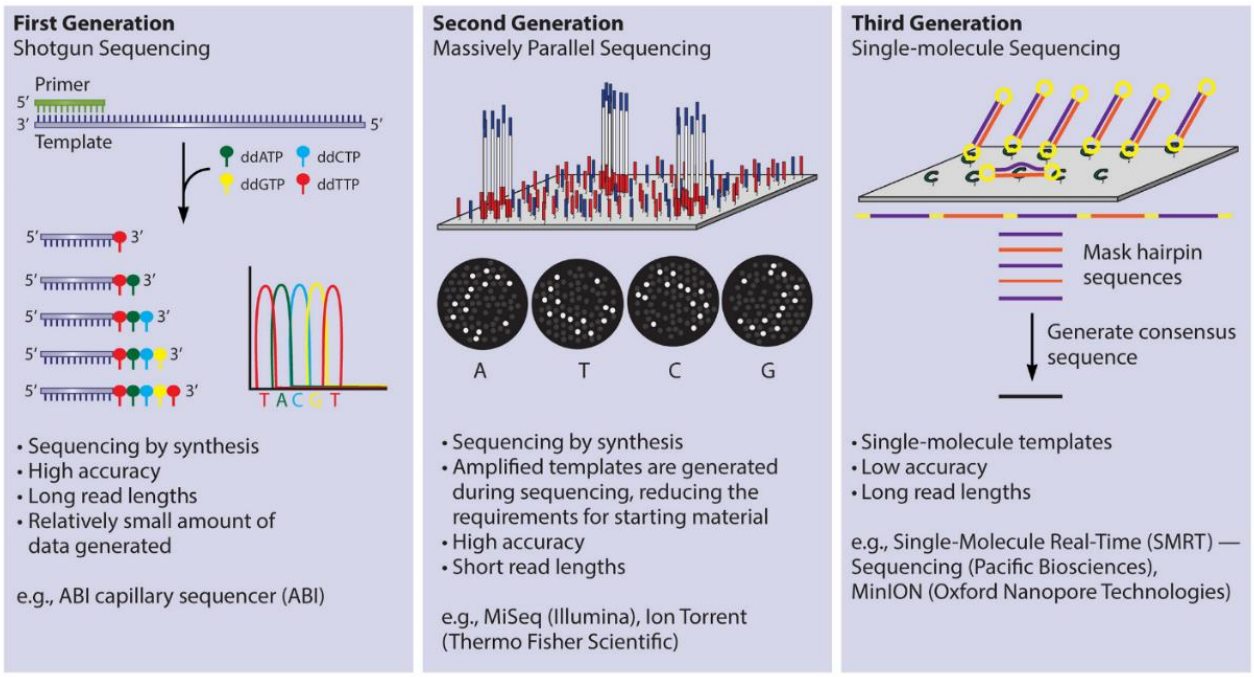


Fig. 3. Methods of Microbial whole genome sequencing methods described by Ronholm et al. *[171]*

The advent of WGS revolutionized the world of bacterial typing with high discriminatory power and resolution. WGS facilitates the genetic comparison between organisms down to the resolution of a single base pair [172]. The resolution achieved by WGS broadens the understanding of the evolution of pathogen and their population structure. This enabled the accurate study of outbreak epidemiology, pattern of pathogen spread, globally and locally, as well as the host adaptation of emerging pathogens. High discriminatory power of WGS allowed genome wide analysis of monomorphic pathogens with very little nucleotide diversity such as *Yersinia pestis* [173], *Bordetella pertussis* [174], and *Salmonella* Typhi [175]. WGS has also showed remarkable differentiation ability in case of highly clonal *Salmonella* serovars like *S.* Enteritidis [176], *S.*

Typhimurium and *S*. Montevideo [40, 171] where traditional typing methods like PFGE, failed to achieve this level of differentiation. Advancement of sequencing technologies extend the WGS implications into the field of epidemiologic investigation and surveillance of bacterial pathogen. First application of WGS for food borne outbreak investigation was reported in 2010 related to 2008 listeriosis outbreak in Canada. Two outbreak associated isolates, which showed similar but distinct PFGE pattern, when analyzed with WGS provided the clue that multiple distinct but highly related strains might have been responsible for the outbreak [177, 178]. WGS has also been used to link historical cases of salmonellosis to current out breaks [179] as well as for the identification and source attribution of laboratory – acquired salmonellosis [171, 180]. These suggest the integration of WGS application in the public health microbiology. WGS technology has proven its potential to identify outbreak source, pathogen virulence determinants, mechanism of pathogenicity, drug resistance and spreads [181] and also has improved trace back epidemiological investigation studies [182].

1.3.2 Next Generation Sequencing (NGS)

Automated Sanger sequencing remained as choice of genome sequencing for almost twenty years [171]. In the last five years, tremendous transformation happened in the sequence technology. 'High throughput sequencing' or 'NGS platforms now became the dominants in the realm of both bacterial and eukaryotic genome sequencing. Roche introduced first of this kind massively parallel sequencing platform in 2005 in the form of 454 GS20 [171]. By 2010, several other sequencing platforms came into industry each with different strength and weaknesses. These NGS platforms can generate enormous volume of sequence data with less cost and time than with Sangers chain terminating

sequence method. Even though all these platforms are based on conceptually similar approach on sequencing methods, each one used slightly different strategies for template preparation, sequencing, imaging and analysis [183]. NGS platforms are categorized into two categories, high end instruments and bench-top instruments. High end instruments are usually suitable for large research facilities because of high set up cost. These instruments provide high through put sequence data and/or long reads compared to bench-top instruments with low set up cost and short read lengths [1]. A general work flow and sequencing chemistry of these sequencing platforms are illustrated in figure 4. Comparison of different next generation sequencing instruments is shown in table 4.

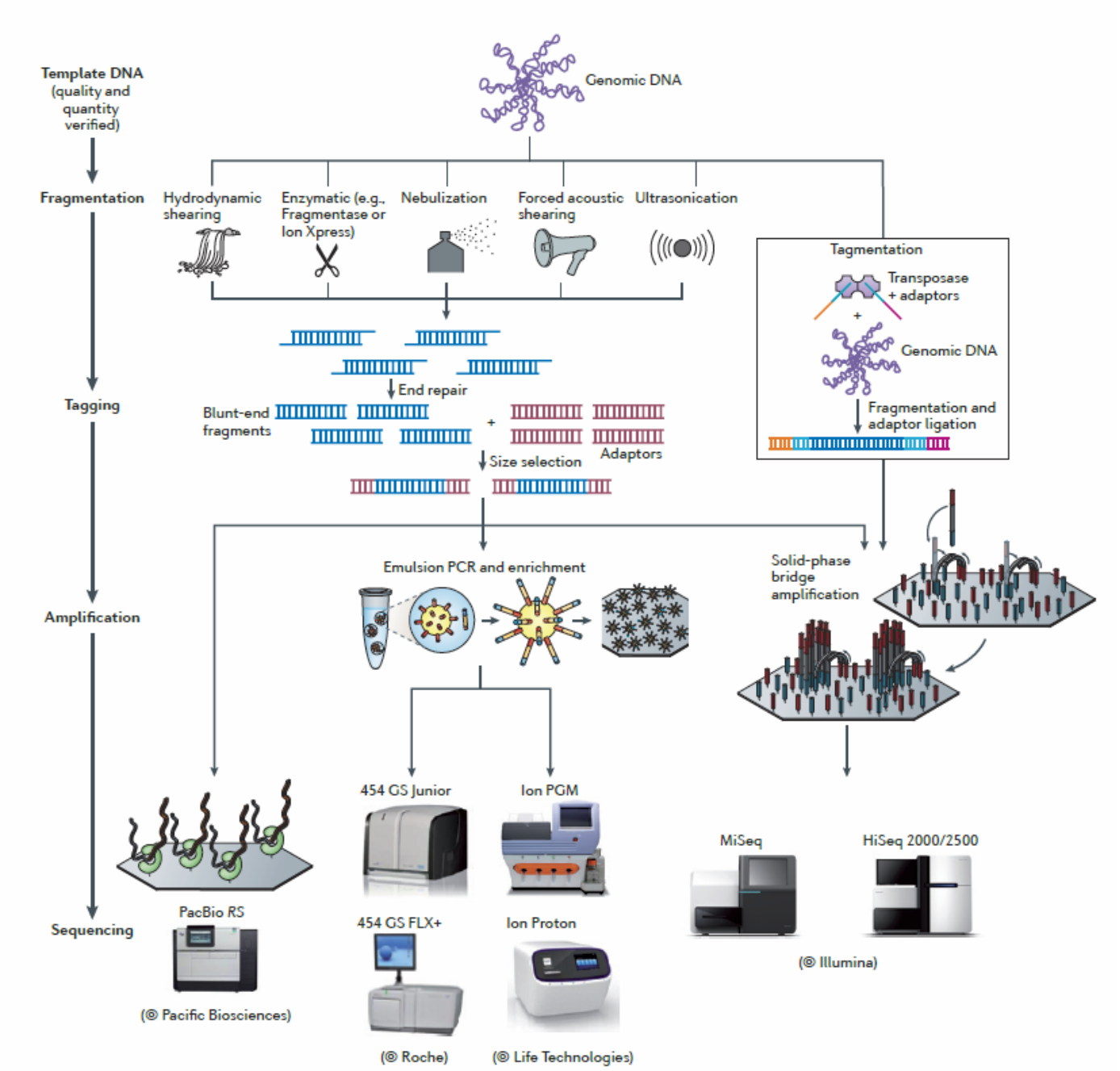


Fig.4. High-throughput sequencing platforms. The schematic shows the main high-throughput sequencing platforms, the associated sample preparation and template amplification. Adopted from review article by Loman et al, 2012 [1].

Table 4. Comparison of next-generation sequencing platforms. Adopted from review article by Loman et al, 2012 [1].

Machine (manufacturer)	Chemistry	Modal read length* (bases)	Run time	Gb per run	Current, approximate cost (US$)‡	Advantages	Disadvantages
High-end instruments							
454 GS FLX+ (Roche)	Pyrosequencing	700–800	23 hours	0.7	500,000	• Long read lengths	• Appreciable hands-on time • High reagent costs • High error rate in homopolymers
HiSeq 2000/2500 (Illumina)	Reversible terminator	2×100	11 days (regular mode) or 2 days (rapid run mode)§	600 (regular mode) or 120 (rapid run mode)§	750,000	• Cost-effectiveness • Steadily improving read lengths • Massive throughput • Minimal hands-on time	• Long run time • Short read lengths • HiSeq 2500 instrument upgrade not available at time of writing (available end 2012)
5500xl SOLiD (Life Technologies)	Ligation	75 + 35	8 days	150	350,000	• Low error rate • Massive throughput	• Very short read lengths • Long run times
PacBio *RS* (Pacific Biosciences)	Real-time sequencing	3,000 (maximum 15,000)	20 minutes	3 per day	750,000	• Simple sample preparation • Low reagent costs • Very long read lengths	• High error rate • Expensive system • Difficult installation
Bench-top instruments							
454 GS Junior (Roche)	Pyrosequencing	500	8 hours	0.035	100,000	• Long read lengths	• Appreciable hands-on time • High reagent costs • High error rate in homopolymers
Ion Personal Genome Machine (Life Technologies)	Proton detection	100 or 200	3 hours	0.01–0.1 (314 chip), 0.1–0.5 (316 chip) or up to 1 (318 chip)	80,000 (including OneTouch and server)	• Short run times • Appropriate throughput for microbial applications	• Appreciable hands-on time • High error rate in homopolymers
Ion Proton (Life Technologies)	Proton detection	Up to 200	2 hours	Up to 10 (Proton I chip) or up to 100 (Proton II chip)	145,000 + 75,000 for compulsory server	• Short run times • Flexible chip reagents	• Instrument not available at time of writing
MiSeq (Illumina)	Reversible terminator	2×150	27 hours	1.5	125,000	• Cost-effectiveness • Short run times • Appropriate throughput for microbial applications • Minimal hands-on time	• Read lengths too short for efficient assembly

1.3.3 Comparative bacterial genomics

Thousands of bacterial genome sequencing data generated by high throughput sequencing is available in the public domain like Genbank. Major part of these bacterial genome assemblies is in ‘draft form’ (as a set of sequence fragments instead of a single sequence representing whole genome). short gun sequencing methods, Short read length and assembly using bench top sequencing platform often resulted in fragmented assemblies [184, 185]. The general work flow of genome comparison studies using these raw sequence data involves assembly of overlapping raw sequence reads to contiguous sequences (contigs), ordering of contigs, annotation, genome comparison and typing [185]. But the core process of comparative genomics is the alignment of DNA sequences, which involves mapping of the nucleotides in one sequence on to the other sequence [186]. Comparative genomic analysis provided insights into bacterial evolution, genetic differences between closely related as well as distantly related species, horizontal gene transfer, gene acquisition and loss, evolution of bacterial pathogens, as well as differences in the gene content between members of similar species.

1.3.3.1 SNP (Single Nucleotide Polymorphism) based analysis

This is a comparative genomic method to determine the phylogenetic relationship of two or more organisms. SNP is the single nucleotide change in the genome sequence. It may occur in coding sequence or non-coding sequence. SNP in a gene may be synonymous if it does not change the amino acid sequence or non-synonymous if it alters the amino acid sequence [171]. Comparison of whole genome sequences forms one of the way to detect SNPs. Analysis of SNPs between genomes give insights in to the relatedness of strains by

comparison of isolates on phylogenetic basis. Detection of SNP can be done either by reference-guided assembly where reads are mapped to reference genome or by analyzing the de novo- assembled genome where assembled contigs are annotated and then ORFs are compared to reference ORFs [171].

1.4 Objectives of this Study

i. Elucidate the intra serovar genetic diversity within the *Salmonella enter*ica serovar Mbandaka in global context by involving publicly available sequence data of many isolates from different parts of the world

ii. Understand the evolutionary structure of those isolates by using phylogenetic analysis based on SNP as well as cgMLST which may provide help in future for epidemiological investigation of any outbreaks caused by this serovar

iii. Find out the existence of any genetic relationships between these isolates with regard to isolation source and its geographical origin which may give clues about the host specificity, host adaptation and geographical isolation

iv. Explore the virulence and antimicrobial resistance gene profile in *S.* Mbandaka isolate's genome to understand the potential pathogenicity of this serovar as well as prevalence of antimicrobial resistant strains

v. Functional analysis of characters such as host cell invasion, resistance to low pH and antimicrobial susceptibility to assess the difference among *S.* Mbandaka isolates at the phenotypic level.

Chapter 2: Whole genome sequencing based phylogenetic analysis, virulence gene mapping and antimicrobial resistance gene profiling of *Salmonella enterica* serovar Mbandaka

2.1 Introduction

*S*almonella Mbandaka, a non-typhoidal *Salmonella* pathogen, has been classified as one of the top ten *Salmonella* serovar causing human food borne illnesses in European union. Clones of this serovar have been shown to be capable of surviving for many years and to spread between different hosts including feed, animals, food and human [159]. Irrespective of geographical location this serovar has been identified as a cause of human salmonellosis in many countries [159, 160] that make this serovar as a global concern. After its first isolation from Belgian Congo in 1948, this serovar has been reported in several countries despite the continental boundaries [159, 166]. Human food borne illnesses caused by this serovar were rare in USA, but recent multistate out breaks reported in 2013 and 2016 raise concerns about this potential enteric pathogen since the sources of these outbreaks were identified as imported food commodity and backyard poultry [162, 163]. Based on annually compiled data from public health laboratory information system, Hayward et al showed that cattle, chicken and pigs are the major sources in USA from which this serovar was mainly isolated [75]. Prevalence in common food animals, ability to transmit through food commodities potential to cause illnesses in

human and worldwide incidence make this serovar as a serious risk to public health safety.

Very few attempts have been made to understand the genomic features as well as the population structure of *S*. Mbandaka, a *Salmonella* serovar which also contributes to the overall foodborne illnesses caused by *Salmonella*. Those studies were limited to small number of isolates from a specific geographical region. There were no prior studies conducted to analyze the population structure of *S*. Mbandaka isolates from USA. More than that, despite the advancement of next generation sequencing and public availability of worldwide distributed isolate's sequence data, no studies ever conducted before in a global context to infer the evolutionary history of this serovar. Since there was lack of reliable information of any isolate included in this study in relation to outbreak occurrence this study does not really investigate epidemiological evidence of any outbreak caused by this serovar. However, understanding the phylogenetic structure of genetic diversity between the isolates within serovar will promote immediate traceback of future outbreaks caused by this serovar.

In this study, we sought to determine the intra serovar genetic diversity of *S*. Mbandaka isolates using a whole genome single nucleotide polymorphism based approach to answer several questions regarding the evolution, host adaptation and geographical distribution of this serovar. Specifically, we aimed to answer following questions, how well isolates of this serovar are genetically diversified, how closely related these isolates from different geographical origin are, whether they show any host dependent discrimination in genetic relatedness, and if present, genetically at which level they show the relatedness or diversification. Apart from this we attempted to elucidate

virulence gene profile and antimicrobial resistance of this serovar to explore the potential of this serovar to pose public health concerns.

2.2 Materials and Methods

2.2.1 Bacterial culture and genomic DNA isolation

Seventy-Six *S*. Mbandaka isolates contributed by different centers (Animal disease research and diagnostic laboratory (ADRDL) (3), Kansas state university (50), Oklahoma state university (9), Michigan state university (13), Immunology and Microbiology Laboratory, PUCRS (1)) were used in this study. Details including their origin, bio sample and SRA accession numbers are described in Table 5. One ml of overnight culture of all isolates was used for genomic DNA extraction after pelleting the bacteria by spinning down at 8000 rpm for 5 minutes. DNA was extracted using Qiagen's DNeasy blood and tissue kit (Qiagen, Inc., Valencia, CA) according to manufacturer's protocol. Quality of DNA was assessed using Nanodrop ^{TM} one (Thermo scientific^{TM}, DE). Quantity of DNA was measured using Qubit® 3.0 (Thermo Fisher Scientific Inc., MA) fluorometer and stored at -20^0 C until further use.

2.2.2 Library preparation, sequencing and assembling

For whole genome sequencing (WGS), all 76 DNA samples were processed using Nextera XT DNA sample prep kit (Illumina inc. San Diego, CA) after adjusting the DNA concentration to 0.3ng/µl. DNA libraries thus prepared then normalized using bead based procedure and pooled together at equal volume. Pooled libararies were then sequenced using Miseq reagent (version 2.0) (Illumina Inc., CA) on Illumina Miseq platform using 2X 250 paired end V2 chemistry.

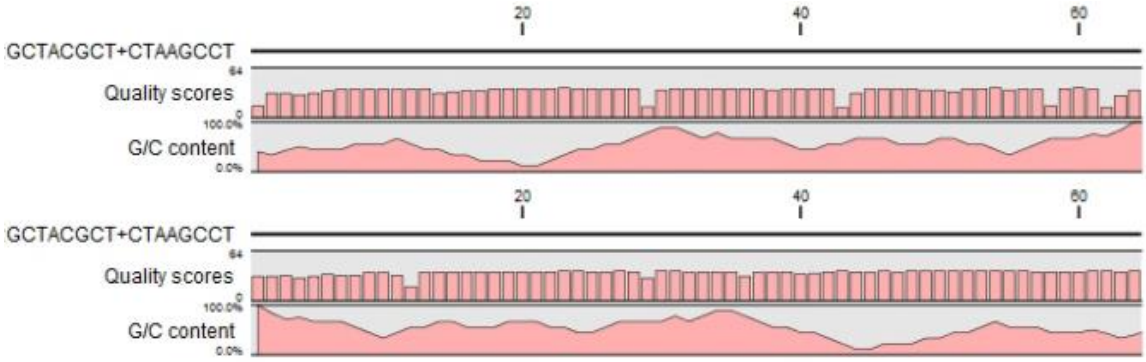


Fig. 5. Sequence read pairing by CLC genomics: Forward and reverse sequence data imported into CLC Genomics workbench as raw sequence reads in Fast Q format. These forward and reverse raw reads are paired before further analysis. Figure shows paired sequences along with their quality score as bar plot and G/C content as Quality scores and GC content as line graph. Adopted from CLC Genomics Workbench manual

2.2.3 Phylogenetic Analysis

After sequencing by Illumina Miseq next generation sequencing (NGS) platform, raw read sequence data of forward and reverse sequences were kept in the form of separate Fastq format files. That means there were two files for a specific sample. For newly sequenced 76 *S.* Mbandaka samples, sequence data (raw reads) were obtained from Illumina Mi seq platform. Sequence data of remaining 389 isolates was downloaded from NCBI - SRA (National Center for Biotechnology Information – Sequence Read Archive) database using sra tool kit (version 2.8.1-2). Details of 465 isolates used in this study were shown in table 5. Sequence data of all 465 samples was then imported in to CLC Genomics Workbench (version 9.5.3) by Qiagen bioinformatics. Sequence data were imported as a single file of paired reads by selecting paired end (forward –reverse) method while importing. Reads having a base length of 200 to 500 bases were imported. Those reads with length below 200 and above 500 bases as well as failed reads were discarded while importing. Fig. 5 shows schematic representation of one out of 1,988,838

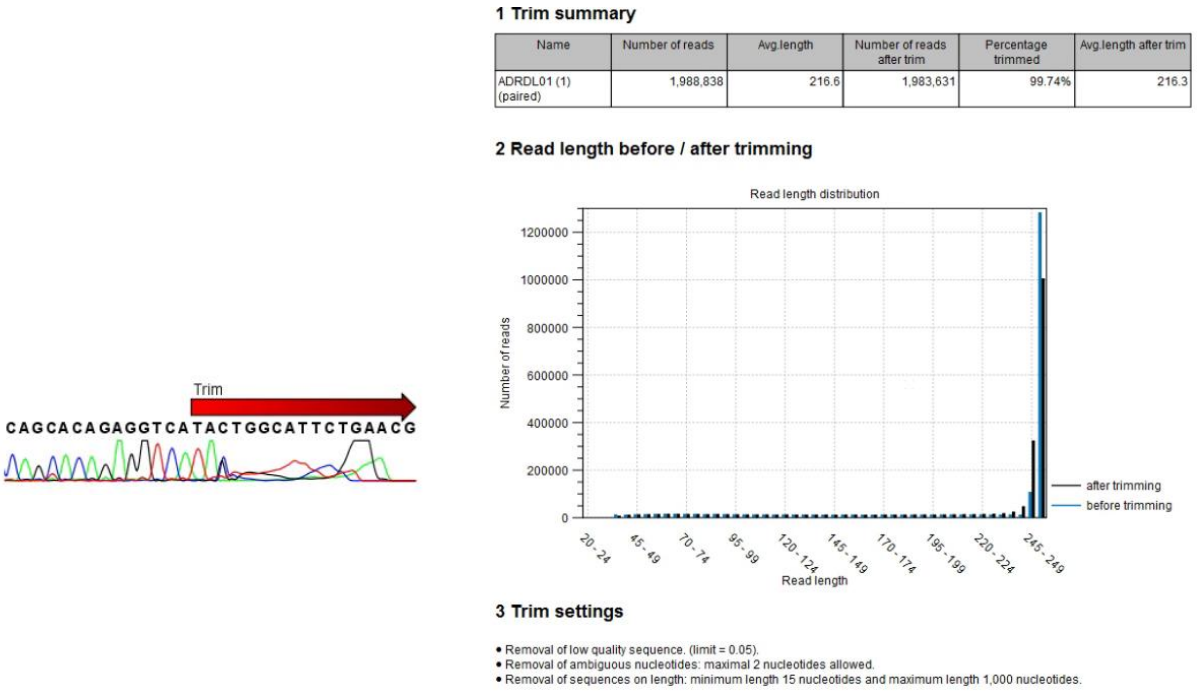


Fig.6. Trimming sequence reads: A) Schematic representation of creating annotation on region that is ignored in the further processing. B) Trimming report of sample ADRDL01 that shows trim summary, read length after trimming and settings used for trimming.

paired reads of isolate ADRDL01 sequence data after importing to CLC Genomics workbench.

Trimming sequence reads: Before doing any analysis imported reads were trimmed to remove any ambiguous nucleotides and to remove sequences with low quality scores. We performed quality trimming based on quality scores, ambiguous trimming to avoid ambiguous nucleotides and length trimming to remove reads shorter or longer than a specified threshold. Trimming here does not mean it will remove any data instead, software tool will create annotation in the region that needs to be trimmed (Fig. 6).

For trimming based on quality scores the program uses the modified-Mott trimming algorithm. Quality score limit that we used in this study is 0.05. Ambiguous nucleotides (for example N) are incorporated by the sequencer program which were removed before any analysis or assembling of sequence data. Maximum allowed ambiguities we selected was 2, that means this is the maximal number of ambiguous nucleotides allowed in the sequence data after trimming. Length trimming is a process of filtering the reads based on their length. For this study, we discarded those reads having a length below 15 bases and above 1000.

Assembling of Sequence reads to contigs:

De novo assembling of sequence reads is required for some analysis such as AMR gene mapping, virulence gene profile and MLST analysis etc. Assembling was carried out in the same CLC genomics work bench using an algorithm that works by using de Bruijn graphs. This creates stretches of contiguous sequences known as contigs. A minimum contig length of 200 bases was used as the assembling parameter and assembling was done without any scaffolding. Assembled sequences were then used for further analysis in CLC genomics. Assembled data exported in FASTA file format were used for MLST analysis as well as for protein clustering analysis.

2.2.3.1 SNP Analysis and Tree construction

Before doing actual SNP tree construction, paired, trimmed reads were mapped to contigs using reference genome ATCC51958 (NZ_CP019183.1). Variant detection was done using fixed ploidy variant detection using parameter ploidy 1 (for bacteria) and variant probability percentage of 90. Variant calls and read mappings results were used to determine the SNP positions. Read mappings were used to estimate the consensus

sequence. SNP tree was constructed using neighbor joining method. Distance between samples computed as “Number of input positions used where the consensus sequence is different / number of input positions used”. Branch length was based on the distance between the samples.

Representative isolates from each cluster were selected to do comparative SNP detection between cluster 1 and other sub clusters of cluster 2 to elucidate number of unique SNVs that differentiate each cluster. Unique SNPs present in representative isolates of each sub cluster were also identified. Details of genes/coding sequences from which the unique SNPs that characterize each cluster were described in table 7. Additional data filtering was carried out to extract the specific SNVs present in selected isolate of each cluster by removing SNVs which were present both in reference genome and selected representative sample. Only SNVs, that meet the criteria of minimum coverage ≥ 50x and with a minimum frequency of 100 were selected avoiding InDels (Insertions and Deletions) and MNVs to identify SNVs unique to a specific cluster.

2.2.3.2 Core genome multilocus sequence typing (cgMLST):

Cluster formation in SNP analysis was validated by gneration of Minimum Spanning Tree (MST) based on core genome MLST. Instead of using few (usually 7) house keeping genes as in traditional multi locus sequence typing, core genome MLST make use of a fixed set of conserved genome wide genes (core genes) for sequence comparison and genotyping providing higher resolution and accurate strain typing [187, 188]. De novo assembled sequences in CLC genomics workbench were used in another bioinformatics software Ridome seq sphere+ (Ridom©GmbH, Germany) for cg MLST. 2502 core genes identified common to both reference genome (ATCC51958, accession :

NZ_CP019183.1) and study isolates were used for sequence typing. Those isolates which were having any missing genes were not used in this study which reduced the study isolates number from 465 to 399. Sequence typing was performed by the program based on the single nucleotide variation in the selected target alleles. Sequence type generated for each isolate was then used to generate the MST (fig.9). MST was generated using a modified version of Kruskal's algorithm [189, 190]. Based on genotype multiple samples can be represented by a single node. Distance between genotypes were represented as links between the nodes.

2.2.3.3 Pangenome analysis and protein clustering

A high-resolution genome content variation between isolates was identified by analyzing the presence and absence of genes in 12 selected isolate genomes and reference genome (*S.* Mbandaka strain ATCC 51958). An analysis and visualization platform called Anvi'o (Meren lab) [191]. FASTA file format of de nova assembled selected genomes were used for gene annotation, protein identification and amino acid sequence alignment in Anvi'o. Platform Anvi'o uses other programs such NCBI's blastp for protein search, muscle for amino acid sequence alignment and MCL for clustering.

2.2.4 Virulence and antimicrobial resistance mapping

Assembled genomes of study isolates were used for virulence gene and resistance mapping. Virulence gene sequences downloaded from Virulence Factor Database (VFDB) [192, 193] and AMR gene sequences downloaded from ResFinder database were used to do the analysis. A BLAST (Basic Local Alignment Search Tool, NCBI) search was done between the isolate genome and gene sequences from the database to identify the presence and absence of respective genes. Minimum sequence length identity of 90%

for virulence gene profiling and 85% for AMR gene mapping was used with minimum sequence length criteria of 50%.

Table 5.**Metadata of *S*. Mbandaka isolates analyzed in this study.** Sample ID was given corresponding to sequence center. Name of the center collected the isolate was given under the title center name. First part of source location indicates the country and second part represents the state if given. Any unavailable information was left blank.

Sample ID	Biosample accession	SRA accession	Center Name	Collection date (Year)	Source location	Isolation_Source	Isolation Source Group
ADRDL-01	SAMN04993480	SRR4446618	CFSAN	2016	USA:SD	GROUND PORK	Food
ADRDL-02	SAMN05366692	SRR5380863	CFSAN	2003	USA:OK	Canine feces	Canine
ADRDL-03	SAMN05366694	SRR5380864	CFSAN	2006	USA:OK	Equine feces	Equine
ADRDL-04	SAMN05366696	SRR4256594	CFSAN	2007	USA:OK	Porcine intestine	Porcine
ADRDL-05	SAMN05366697	SRR4256593	CFSAN	2008	USA:AR	Kangaroo intestine	Others
ADRDL-06	SAMN05366698	SRR4256087	CFSAN	2009	USA:OK	Caprine intestine	Caprine
ADRDL-07	SAMN06669733	SRR5418728	CFSAN	2010	USA:MI	Chicken drag swab	Environmental
ADRDL-08	SAMN06669732	SRR5418726	CFSAN	2010	USA:MI	Chicken feed	Animal Feed
ADRDL-09	SAMN06669731	SRR5418724	CFSAN	2010	USA:MI	Chick box paper	Environmental
ADRDL-10	SAMN06669730	SRR5418725	CFSAN	2010	USA:FL	Bovine feces (Bos taurus)	Bovine
ADRDL-11	SAMN06669729	SRR5418727	CFSAN	2010	USA:MI	Chick box paper	Environmental
ADRDL-12	SAMN06669728	SRR5418517	CFSAN	2010	USA:FL	Bovine feces	Bovine
ADRDL-13	SAMN06669727	SRR5418515	CFSAN	2011	USA:SC	Feline Small Intestine	Feline
ADRDL-14	SAMN06669739	SRR5418516	CFSAN	2012	USA:MI	Equine feces	Equine
ADRDL-15	SAMN06669738	SRR5418512	CFSAN	2013	USA:MI	Chicken tissue pool	Avian
ADRDL-16	SAMN06669737	SRR5418508	CFSAN	2013	USA:MI	Chicken drag swab	Environmental
ADRDL-17	SAMN06669736	SRR5418514	CFSAN	2016	USA:MI	Chicken feed	Animal Feed
ADRDL-18	SAMN06669735	SRR5418510	CFSAN	2016	USA:MI	Bovine feces	Bovine
ADRDL-19	SAMN06669734	SRR5418509	CFSAN	2012	USA:MI	Chicken drag swab	Environmental
ADRDL-20	SAMN03734321	SRR5380867	CFSAN	2014	USA:MN	bovine colon	Bovine

ADRDL-21	SAMN04240663	SRR2971412	CFSAN	2015	USA:NE	bovine intestin	Bovine
ADRDL-22	SAMN06113938	SRR5173676	CFSAN	2008	Brazil	Meat and bones Meal	Animal Feed
ADRDL-23	SAMN06113971	SRR5182179	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-24	SAMN06113970	SRR5182188	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-25	SAMN06113969	SRR5182189	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-26	SAMN06113968	SRR5182178	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-27	SAMN06113967	SRR5182191	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-28	SAMN06113966	SRR5182192	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-29	SAMN06113965	SRR5182186	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-30	SAMN06114003	SRR5182180	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-31	SAMN06114002	SRR5185864	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-32	SAMN06114001	SRR5185869	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-33	SAMN06114022	SRR5185871	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-34	SAMN06114021	SRR5185863	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-35	SAMN06114020	SRR5185866	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-36	SAMN06114019	SRR5185862	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-37	SAMN06114018	SRR5185860	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-38	SAMN06114017	SRR5185737	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-39	SAMN06114016	SRR5185861	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-40	SAMN06114015	SRR5185739	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-41	SAMN06114014	SRR5185735	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-42	SAMN06114013	SRR5185738	CFSAN	2006	USA:TX	bovine feces	Bovine
ADRDL-43	SAMN06114012	SRR5380960	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-44	SAMN06114011	SRR5185726	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-45	SAMN06114010	SRR5182185	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-46	SAMN06114009	SRR5182181	CFSAN	2007	USA:TX	bovine feces	Bovine

ADRDL-47	SAMN06114008	SRR5292186	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-48	SAMN06114007	SRR5292183	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-49	SAMN06114006	SRR5292182	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-50	SAMN06114005	SRR5292180	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-51	SAMN06114040	SRR5292177	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-52	SAMN06114039	SRR5292172	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-53	SAMN06114038	SRR5292170	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-54	SAMN06114037	SRR5292169	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-55	SAMN06114036	SRR5292171	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-56	SAMN06114035	SRR5292166	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-57	SAMN06114034	SRR5292168	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-58	SAMN06114033	SRR5292167	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-59	SAMN06114032	SRR5292163	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-60	SAMN06114031	SRR5292165	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-61	SAMN06114030	SRR5292161	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-62	SAMN06114029	SRR5292156	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-63	SAMN06114028	SRR5291672	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-64	SAMN06114027	SRR5291669	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-65	SAMN06114026	SRR5291673	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-66	SAMN06114025	SRR5291670	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-67	SAMN06114024	SRR5291660	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-68	SAMN06114023	SRR5291663	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-69	SAMN06114044	SRR5291659	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-70	SAMN06114043	SRR5291668	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-71	SAMN06114042	SRR5291662	CFSAN	2007	USA:TX	bovine feces	Bovine
ADRDL-72	SAMN06114041	SRR5291667	CFSAN	2007	USA:TX	bovine feces	Bovine

ADRDL-73	SAMN05366685	SRR4256089	CFSAN	1988	USA:OK	Bovine fecal swab	Bovine
ADRDL-74	SAMN05366686	SRR4256088	CFSAN	1989	USA:OK	Bovine feces	Bovine
ADRDL-75	SAMN05366690	SRR4256085	CFSAN	1992	USA:OK	Bovine feces	Bovine
ADRDL-76	SAMN05366691	SRR5386735	CFSAN	1998	USA:OK	Porcine intestine	Porcine
CDC001	SAMN03469748	SRR1970582	EDLB-CDC		USA	Stool	Human
CFSAN001	SAMN05195908	SRR5120746	CFSAN	2014	USA:GA	Environmental (Pond)	Environmental
CFSAN002	SAMN05195909	SRR5120748	CFSAN	2014	USA:GA	Environmental (Pond)	Environmental
CFSAN003	SAMN05195910	SRR5120749	CFSAN	2014	USA:GA	Environmental (Pond)	Environmental
CFSAN004	SAMN05195911	SRR5120750	CFSAN	2014	USA:GA	Environmental (Pond)	Environmental
EUR001	SAMN02368725	SRR1106374	UCD-100K				Others
EUR002	SAMN02368608	SRR1106454	UCD-100K	2010	Denmark	feed	Animal Feed
EUR003	SAMN02368607	SRR1106455	UCD-100K	2010	Denmark	broiler	Avian
EUR004	SAMN02368606	SRR1106456	UCD-100K	2009	Nigeria	Chicken	Avian
EUR005	SAMN02368605	SRR1106457	UCD-100K	2002	Taiwan	Human	Human
EUR006	SAMN03168592	SRR1645208	GBRU	2013	United Kingdom: North of England	Human	Human
EUR007	SAMN03169272	SRR1645959	GBRU	2013	United Kingdom: London	Human	Human
EUR008	SAMN03465654	SRR1957759	GBRU	2014	United Kingdom: London	Human	Human
EUR009	SAMN03465692	SRR1957797	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human
EUR010	SAMN03465905	SRR1958003	GBRU	2014	United Kingdom: London	Food	Food
EUR011	SAMN03466060	SRR1958160	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human
EUR012	SAMN03466356	SRR1958463	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human
EUR013	SAMN03466359	SRR1958466	GBRU	2014	United Kingdom: London	Food	Food
EUR014	SAMN03466390	SRR1958496	GBRU	2014	United Kingdom: North of England	Human	Human

EUR015	SAMN03466483	SRR1958590	GBRU	2014	United Kingdom: North of England	Human	Human
EUR016	SAMN03466562	SRR1958669	GBRU	2014	United Kingdom: North of England	Human	Human
EUR017	SAMN03468476	SRR1959271	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human
EUR018	SAMN03468502	SRR1959297	GBRU	2015	United Kingdom	Food	Food
EUR019	SAMN03468506	SRR1959301	GBRU	2014	United Kingdom: London	Human	Human
EUR020	SAMN03468513	SRR1959308	GBRU	2014	United Kingdom: North of England	Human	Human
EUR021	SAMN03468528	SRR1959359	GBRU	2014	United Kingdom: London	Human	Human
EUR022	SAMN03468627	SRR1959446	GBRU	2014	United Kingdom: North of England	Human	Human
EUR023	SAMN03468723	SRR1959508	GBRU	2015	United Kingdom	Food	Food
EUR024	SAMN03469065	SRR1960295	GBRU	2014	United Kingdom: North of England	Human	Human
EUR025	SAMN03469072	SRR1960313	GBRU	2014	United Kingdom: London	Human	Human
EUR026	SAMN03469127	SRR1960594	GBRU	2015	United Kingdom	Food	Food
EUR027	SAMN03469153	SRR1961019	GBRU	2015	United Kingdom	Food	Food
EUR028	SAMN03469202	SRR1962339	GBRU	2014	United Kingdom: North of England	Human	Human
EUR029	SAMN03469950	SRR1963497	GBRU	2014	United Kingdom	Food	Food
EUR030	SAMN03475134	SRR1965197	GBRU	2014	United Kingdom: London	Human	Human
EUR031	SAMN03475839	SRR1965734	GBRU	2014	United Kingdom: London	Human	Human
EUR032	SAMN03476011	SRR1965919	GBRU	2014	United Kingdom: South of England	Human	Human
EUR033	SAMN03476017	SRR1965925	GBRU	2015	United Kingdom: South of England	Human	Human
EUR034	SAMN03476320	SRR1966237	GBRU	2015	United Kingdom: London	Human	Human

EUR035	SAMN03476477	SRR1966394	GBRU	2014	United Kingdom: South of England	Human	Human
EUR036	SAMN03476573	SRR1966490	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human
EUR037	SAMN03477023	SRR1966945	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human
EUR038	SAMN03477095	SRR1967017	GBRU	2014	United Kingdom: London	Human	Human
EUR039	SAMN03477383	SRR1967293	GBRU	2014	United Kingdom: North of England	Human	Human
EUR040	SAMN03477432	SRR1967342	GBRU	2014	United Kingdom: South of England	Human	Human
EUR041	SAMN03477597	SRR1967507	GBRU	2014	United Kingdom: North of England	Human	Human
EUR042	SAMN03477643	SRR1967553	GBRU	2014	United Kingdom: London	Human	Human
EUR043	SAMN03477732	SRR1967642	GBRU	2014	United Kingdom: South of England	Human	Human
EUR044	SAMN03478181	SRR1968090	GBRU	2014	United Kingdom	Food	Food
EUR045	SAMN03478321	SRR1968230	GBRU	2014	United Kingdom: London	Human	Human
EUR046	SAMN03478626	SRR1968533	GBRU	2014	United Kingdom: South of England	Human	Human
EUR048	SAMN03478819	SRR1968721	GBRU	2014	United Kingdom: London	Human	Human
EUR049	SAMN03478825	SRR1968727	GBRU	2014	United Kingdom: North of England	Human	Human
EUR050	SAMN03478828	SRR1968730	GBRU	2014	United Kingdom	Food	Food
EUR051	SAMN03479022	SRR1968924	GBRU	2014	United Kingdom: South of England	Human	Human
EUR052	SAMN03479182	SRR1969105	GBRU	2014	United Kingdom: London	Human	Human
EUR053	SAMN03479325	SRR1969248	GBRU	2014	United Kingdom: North of England	Human	Human
EUR054	SAMN03479539	SRR1969462	GBRU	2014	United Kingdom: Midlands and East of England	Human	Human

EUR055	SAMN03480046	SRR1969968	GBRU	2014	United Kingdom: North of England	Human	Human
EUR056	SAMN03480253	SRR1970177	GBRU	2014	United Kingdom: South of England	Human	Human
EUR057	SAMN03480286	SRR1970210	GBRU	2015	United Kingdom: South of England	Human	Human
EUR058	SAMN03480306	SRR1970230	GBRU	2014	United Kingdom: North of England	Human	Human
EUR059	SAMN03480345	SRR1970269	GBRU	2014	United Kingdom: London	Human	Human
EUR060	SAMN04362942	SRR3048574	PHE		United Kingdom	Human	Human
EUR061	SAMN04363133	SRR3048881	PHE		United Kingdom	Human	Human
EUR062	SAMN04363136	SRR3048884	PHE		United Kingdom	Human	Human
EUR063	SAMN04363203	SRR3048963	PHE		United Kingdom	Human	Human
EUR064	SAMN04363401	SRR3049163	PHE		United Kingdom	Human	Human
EUR065	SAMN04363579	SRR3049284	PHE		United Kingdom	Human	Human
EUR066	SAMN04363778	SRR3049483	PHE		United Kingdom	Human	Human
EUR067	SAMN04600314	SRR3321900	PHE		United Kingdom	Food	Food
EUR068	SAMN04600345	SRR3322012	PHE	2015	United Kingdom	Environmental	Environmental
EUR069	SAMN04600394	SRR3322073	PHE	2015	United Kingdom	Human	Human
EUR070	SAMN04600544	SRR3322144	PHE	2015	United Kingdom	Human	Human
EUR071	SAMN04600657	SRR3322361	PHE	2015	United Kingdom	Human	Human
EUR072	SAMN04600934	SRR3322687	PHE		United Kingdom	Human	Human
EUR073	SAMN04601072	SRR3322985	PHE	2015	United Kingdom	Environmental	Environmental
EUR074	SAMN04601151	SRR3323062	PHE		United Kingdom	Human	Human
EUR075	SAMN06247684	SRR5193613	PHE		United Kingdom	Human	Human
EUR076	SAMN06247764	SRR5193691	PHE		United Kingdom	Human	Human
FDA001	SAMN02367951	SRR1122614	UCD-100K	2011		Chicken Breast	Food
FDA002	SAMN02698272	SRR1220767	CFSAN	2013	Turkey	tahini	Food

FDA003	SAMN02678886	SRR1264976	CFSAN	2013	Turkey	ground cumin	Food
FDA004	SAMN02640947	SRR1272810	CFSAN	2003	USA:GA	Chicken breasts	Food
FDA005	SAMN02698203	SRR1287031	CFSAN	2011	India	manchow soup powder	Food
FDA006	SAMN02698218	SRR1288381	CFSAN	2011	India	tilapia	Food
FDA007	SAMN02678733	SRR1292256	CFSAN	2008	Mexico	chipotle chili powder	Food
FDA008	SAMN02640920	SRR1299303	CFSAN	2002	USA:GA	Chicken breasts	Food
FDA009	SAMN02766991	SRR1300634	CFSAN	2004	USA:MN	Chicken Breast	Food
FDA010	SAMN02766992	SRR1300670	CFSAN	2004	USA:MN	Chicken Breast	Food
FDA011	SAMN02678490	SRR1300724	CFSAN	2010	Canada	sesame seed	Food
FDA012	SAMN02698309	SRR1425261	CFSAN	2013	Turkey	tahini	Food
FDA013	SAMN02698311	SRR1425281	CFSAN	2013	India	spice mix	Food
FDA015	SAMN02698295	SRR1481937	CFSAN	2013	Bangladesh	chili powder	Food
FDA016	SAMN02849739	SRR1501485	CFSAN	2007	USA:WA	Bovine (feed)	Animal Feed
FDA017	SAMN02849762	SRR1501639	CFSAN	2007	USA:ID	Bovine	Bovine
FDA018	SAMN02849784	SRR1501668	CFSAN	2007	USA:ID	Bovine (feces)	Bovine
FDA019	SAMN02698446	SRR1503322	CFSAN	2010	USA:NE	meat and bone meal	Animal Feed
FDA020	SAMN02678763	SRR1509602	CFSAN	2008	El Salvador	iguana meat, frz	Food
FDA021	SAMN02678567	SRR1511519	CFSAN	2010	USA:IN	swab	Environmental
FDA022	SAMN02678569	SRR1511540	CFSAN	2010	Mexico	sesame seed	Food
FDA023	SAMN02678566	SRR1511552	CFSAN	2010	USA:IN	swab	Environmental
FDA024	SAMN02849949	SRR1515029	CFSAN	2007	USA:WA	Avian	Avian
FDA025	SAMN02849964	SRR1515964	CFSAN	2007	USA:UT	Bovine (feces)	Bovine
FDA026	SAMN02698164	SRR1553806	CFSAN	2010	China	isolated soy protein	Food
FDA027	SAMN02782539	SRR1553817	CFSAN	2004	USA:GA	Chicken Breast	Food
FDA028	SAMN02782559	SRR1556105	CFSAN	2004	USA:GA	Chicken Breast	Food
FDA029	SAMN02698313	SRR1560594	CFSAN	2013	Syria	anise seeds	Food

FDA030	SAMN02678668	SRR1586544	CFSAN	2013	Lebanon	halawa candy	Food
FDA031	SAMN02847332	SRR1593394	CFSAN	2011	Mexico	papaya	Food
FDA032	SAMN02847333	SRR1593396	CFSAN	2011	Mexico	papaya	Food
FDA033	SAMN02847334	SRR1593414	CFSAN	2011	Mexico	papayas	Food
FDA034	SAMN02847337	SRR1596368	CFSAN	2011	Mexico	papayas	Food
FDA035	SAMN02847338	SRR1596369	CFSAN	2011	Mexico	papaya	Food
FDA036	SAMN02847339	SRR1596372	CFSAN	2011	Mexico	papaya	Food
FDA037	SAMN02847349	SRR1596381	CFSAN	2011	Mexico	papaya	Food
FDA038	SAMN02847354	SRR1596382	CFSAN	2011	Mexico	papaya	Food
FDA039	SAMN02847348	SRR1596388	CFSAN	2011	Mexico	papaya	Food
FDA040	SAMN02847346	SRR1596394	CFSAN	2011	Mexico	papaya	Food
FDA041	SAMN02847347	SRR1596395	CFSAN	2011	Mexico	papaya	Food
FDA042	SAMN02847336	SRR1596402	CFSAN	2011	Mexico	papaya	Food
FDA043	SAMN02847353	SRR1596408	CFSAN	2011	Mexico	papaya	Food
FDA044	SAMN02847340	SRR1596413	CFSAN	2011	Mexico	papayas	Food
FDA045	SAMN02847352	SRR1596414	CFSAN	2011	Mexico	papaya	Food
FDA046	SAMN02678608	SRR1613889	CFSAN	2008	Singapore	sultana biscuits	Food
FDA047	SAMN02678613	SRR1613890	CFSAN	2008	Mexico	serrano pepper	Food
FDA048	SAMN02678609	SRR1613910	CFSAN	2008	India	cashew snack	Food
FDA049	SAMN02678788	SRR1615081	CFSAN	2009	Mexico	senna tea	Food
FDA050	SAMN02844023	SRR1619530	CFSAN	2004	India	sesame seeds	Food
FDA051	SAMN02843931	SRR1619598	CFSAN	2004	USA:GA	animal feed blend	Animal Feed
FDA052	SAMN02918992	SRR1623027	CFSAN	2012	Mexico	animal feed, wheat millrun	Animal Feed
FDA053	SAMN02777685	SRR1637012	CFSAN	2003	USA:MN	Chicken Breast	Food
FDA054	SAMN02777687	SRR1637016	CFSAN	2003	USA:MN	Chicken Breast	Food
FDA055	SAMN02777684	SRR1637044	CFSAN	2003	USA:MN	Chicken Breast	Food

FDA056	SAMN02777688	SRR1637049	CFSAN	2003	USA:MN	Chicken Breast	Food
FDA057	SAMN02777686	SRR1637054	CFSAN	2003	USA:MN	Chicken Breast	Food
FDA058	SAMN02847395	SRR1638895	CFSAN	2012	Mexico	peanut, shelled	Food
FDA059	SAMN02846970	SRR1646539	CFSAN	2010	USA:CA	pistachio, in shell	Food
FDA060	SAMN02846971	SRR1646581	CFSAN	2010	USA:CA	pistachio, in shell	Food
FDA061	SAMN02846533	SRR1658072	CFSAN	2010	USA:CA	environmental swabs	Environmental
FDA062	SAMN02846531	SRR1658077	CFSAN	2010	USA:CA	environmental swabs	Environmental
FDA063	SAMN02846532	SRR1658078	CFSAN	2010	USA:CA	environmental swabs	Environmental
FDA064	SAMN02846534	SRR1658083	CFSAN	2010	USA:CA	environmental swabs	Environmental
FDA065	SAMN02846687	SRR1685379	CFSAN	2009	India	rte snack food - cornflakes & potato bits	Food
FDA066	SAMN02843821	SRR1687189	CFSAN	2003	USA	animal feed	Animal Feed
FDA067	SAMN02918862	SRR1705573	CFSAN	2012	USA:CO	environmental swab	Environmental
FDA068	SAMN02846688	SRR1720474	CFSAN	2009	India	rte snack food - cornflakes & potato bits	Food
FDA069	SAMN02843438	SRR1730384	CFSAN	2001	India	celery seed	Food
FDA070	SAMN02847149	SRR1732582	CFSAN	2010	Canada	sesame seed	Food
FDA071	SAMN02918827	SRR1745632	CFSAN	2011	USA:AR	animal feed, poultry feed	Animal Feed
FDA072	SAMN02777760	SRR1778004	CFSAN	2003	USA:GA	Chicken Breast	Food
FDA073	SAMN03285412	SRR1778061	CFSAN	2014	Mexico	arbol peppers	Food
FDA074	SAMN03269470	SRR1783178	CFSAN	2005	USA:GA	chicken breast	Food
FDA075	SAMN03269469	SRR1783205	CFSAN	2005	USA:GA	chicken breast	Food
FDA076	SAMN02846555	SRR1802945	CFSAN	2009	Japan	curry powder	Food
FDA077	SAMN02844670	SRR1803114	CFSAN	2009	Mexico	cucumber	Food
FDA078	SAMN02844630	SRR1805606	CFSAN	2008	Mexico	serrano peppers	Food
FDA079	SAMN03344531	SRR1810505	CFSAN	2015	USA:CO	beef tracheas	Food

FDA080	SAMN03269471	SRR1812835	CFSAN	2006	USA:CO	chicken breast	Food
FDA081	SAMN02846488	SRR1821290	CFSAN	2009	Taiwan	garlic powder	Food
FDA082	SAMN02846487	SRR1821304	CFSAN	2009	Taiwan	garlic powder	Food
FDA083	SAMN03269472	SRR1823694	CFSAN	2006	USA:CT	ground beef	Food
FDA084	SAMN02367998	SRR1840623	UCD-100K	2011	USA:Maryland	Chicken Breast	Food
FDA085	SAMN02368017	SRR1840628	UCD-100K	2011	USA:New York	Chicken Breast	Food
FDA086	SAMN02846612	SRR1916078	CFSAN	2009	Viet Nam	curry powder	Food
FDA087	SAMN02846606	SRR1916123	CFSAN	2009	Indonesia	cheese roll cake	Food
FDA088	SAMN02844503	SRR1946898	CFSAN	2007	USA	parsley powder	Food
FDA089	SAMN02846047	SRR1952752	CFSAN	2008	USA:CA	alfalfa seed	Food
FDA090	SAMN02846046	SRR1974111	CFSAN	2008	USA:CA	alfalfa seed	Food
FDA091	SAMN02845928	SRR1980615	CFSAN	2007	USA:AR	peanut butter	Food
FDA092	SAMN03276007	SRR1980731	CFSAN	2007	USA:CA	pork chop	Food
FDA093	SAMN02844258	SRR1982186	CFSAN	2005	China	spice powder	Food
FDA094	SAMN02846007	SRR2005920	CFSAN	2007	India	coconut shredded	Food
FDA095	SAMN02844040	SRR2025278	CFSAN	2004	China	dog chew	Animal Feed
FDA096	SAMN02844751	SRR2075974	CFSAN	2009	India	frz grated coconut	Food
FDA097	SAMN02844770	SRR2076020	CFSAN	2009	India	moth beans crisp disc	Food
FDA098	SAMN02845156	SRR2078204	CFSAN	2002	USA:CA	meat and bone meal	Animal Feed
FDA099	SAMN02844799	SRR2085686	CFSAN	2010	India	sesame seeds, indian hulled	Food
FDA100	SAMN02844596	SRR2086584	CFSAN	2008	Syria	cumin, ground	Food
FDA101	SAMN03276097	SRR2086981	CFSAN	2007	USA:GA	pork chop	Food
FDA102	SAMN02844635	SRR2087270	CFSAN	2008	Taiwan	tilapia	Food
FDA103	SAMN02844236	SRR2096626	CFSAN	2005	USA:IN	custom feed mix	Animal Feed
FDA104	SAMN03842241	SRR2143502	CFSAN	2007	USA:CA	chicken breast	Food
FDA105	SAMN02845473	SRR2156506	CFSAN	2004	India	hulled sesame seeds	Food

FDA106	SAMN02845474	SRR2156565	CFSAN	2004	India	hulled sesame seeds	Food
FDA107	SAMN02845472	SRR2173998	CFSAN	2004	India	hulled sesame seeds	Food
FDA108	SAMN02845471	SRR2174059	CFSAN	2004	India	hulled sesame seeds	Food
FDA109	SAMN03988234	SRR2407554	FDA	2014	USA:MO	Chicken Wings	Food
FDA110	SAMN03988418	SRR2407738	FDA	2014	USA:NM	Chicken Breast	Food
FDA111	SAMN03988424	SRR2407744	FDA	2014	USA:NM	Chicken Breast	Food
FDA112	SAMN03842301	SRR2532630	FDA	2007	USA:NM	chicken breast	Food
FDA113	SAMN03291697	SRR2533345	FDA	2006	USA:OR	ground turkey	Food
FDA114	SAMN03291698	SRR2533346	FDA	2006	USA:OR	ground beef	Food
FDA115	SAMN03894188	SRR2566990	FDA	2013	USA:TN	Chicken Wings	Food
FDA116	SAMN03894384	SRR2567186	FDA	2013	USA: NM	Ground Beef	Food
FDA117	SAMN03894393	SRR2567195	FDA	2013	USA: NY	Chicken Breast	Food
FDA118	SAMN03894406	SRR2567208	FDA	2013	USA:WA	Chicken Breast	Food
FDA119	SAMN03842397	SRR2670701	CFSAN	2008	USA:NY	chicken breast	Food
FDA120	SAMN02918825	SRR2728278	CFSAN	2011	USA:NY	animal feed, dairy cattle feed	Animal Feed
FDA121	SAMN02918913	SRR2728282	CFSAN	2012	Canada	animal feed, canola meal	Animal Feed
FDA122	SAMN03938825	SRR2939007	CFSAN	2008	USA:MD	chicken breast	Food
FDA123	SAMN02845059	SRR3038212	CFSAN	2002	USA:AZ	cottonseed	Animal Feed
FDA124	SAMN04224256	SRR3055310	CFSAN	2008	USA:MD	Chicken Breast	Food
FDA125	SAMN02843508	SRR3057159	CFSAN	2002	USA:AZ	blood meal	Animal Feed
FDA126	SAMN04256097	SRR3098665	CFSAN	2010	USA:NY	Chicken Breast	Food
FDA127	SAMN04396122	SRR3110510	CFSAN	2015	USA:MN	Raw Cashews Pieces	Food
FDA128	SAMN04224384	SRR3115191	CFSAN	2009	USA:CT	Chicken Breast	Food
FDA129	SAMN02844212	SRR3156748	CFSAN	2005	Canada	saw palmetto seed powder	Food
FDA130	SAMN02846022	SRR3173513	CFSAN	2009	USA:GA	environmental swab	Environmental

FDA131	SAMN04255401	SRR3173557	CFSAN	2009	USA:PA	Chicken Breast	Food
FDA132	SAMN02918555	SRR3191380	CFSAN	2009	USA:OH	poultry feed	Animal Feed
FDA133	SAMN04218119	SRR3194547	CFSAN	2008	USA:TN	Chicken Breast	Food
FDA134	SAMN04218120	SRR3194548	CFSAN	2008	USA:TN	Chicken Breast	Food
FDA135	SAMN04218126	SRR3199879	CFSAN	2008	USA:TN	Chicken Breast	Food
FDA136	SAMN04218123	SRR3199882	CFSAN	2008	USA:TN	Chicken Breast	Food
FDA137	SAMN02847454	SRR3205946	CFSAN	2012	Mexico	papaya	Food
FDA138	SAMN04218138	SRR3210390	CFSAN	2008	USA:TN	Ground Beef	Food
FDA139	SAMN04218140	SRR3210391	CFSAN	2008	USA:TN	Ground Beef	Food
FDA140	SAMN04218142	SRR3210393	CFSAN	2008	USA:TN	Ground Beef	Food
FDA141	SAMN04218145	SRR3210530	CFSAN	2008	USA:TN	Ground Beef	Food
FDA142	SAMN04218146	SRR3210531	CFSAN	2008	USA:TN	Ground Beef	Food
FDA143	SAMN04218147	SRR3210532	CFSAN	2008	USA:TN	Ground Beef	Food
FDA144	SAMN04218150	SRR3210535	CFSAN	2008	USA:TN	Pork Chop	Food
FDA145	SAMN04218151	SRR3210536	CFSAN	2008	USA:TN	Pork Chop	Food
FDA146	SAMN04218152	SRR3210537	CFSAN	2008	USA:TN	Pork Chop	Food
FDA147	SAMN04218153	SRR3210538	CFSAN	2008	USA:TN	Pork Chop	Food
FDA148	SAMN04218155	SRR3210540	CFSAN	2008	USA:TN	Pork Chop	Food
FDA149	SAMN04218157	SRR3210542	CFSAN	2008	USA:TN	Pork Chop	Food
FDA150	SAMN04217422	SRR3217374	CFSAN	2009	USA:NY	Chicken Breast	Food
FDA151	SAMN02846382	SRR3219072	CFSAN	2009	Egypt	artificially flavored beef bouillon	Food
FDA152	SAMN02918837	SRR3223732	CFSAN	2011	USA:PA	animal feed, chicken meal	Animal Feed
FDA153	SAMN02843451	SRR3242144	CFSAN	2001	Lebanon	halva	Food
FDA154	SAMN04577302	SRR3295755	FDA	2014	USA:FL	Dairy Cattle	Bovine
FDA155	SAMN04577361	SRR3295814	FDA	2014	USA:PA	Hogs	Porcine

FDA156	SAMN04577366	SRR3295819	FDA	2014	USA:FL	Dairy Cattle	Bovine
FDA157	SAMN04577373	SRR3295826	FDA	2014	USA:FL	Beef Cattle	Bovine
FDA158	SAMN02847343	SRR3379320	CFSAN	2011	Mexico	papayas	Food
FDA159	SAMN02845941	SRR3453119	CFSAN	2008	USA	swab	Environmental
FDA160	SAMN02845942	SRR3453120	CFSAN	2008	USA	swab	Environmental
FDA161	SAMN02845945	SRR3453123	CFSAN	2008	USA	swab	Environmental
FDA162	SAMN02699294	SRR3664643	FDA	2012	USA:NM	chicken breast	Food
FDA163	SAMN02699344	SRR3664883	FDA	2012	USA:CA	chicken breast	Food
FDA164	SAMN05201964	SRR3664884	FDA	2012	USA:CA	Chicken Breast	Food
FDA165	SAMN02699356	SRR3664919	FDA	2012	USA:NM	chicken breast	Food
FDA166	SAMN05201503	SRR3664987	FDA	2011	USA:OR	Chicken Breast	Food
FDA167	SAMN05201613	SRR3665138	FDA	2011	USA:MD	Chicken Breast	Food
FDA168	SAMN02699252	SRR3665190	FDA	2011	USA:NY	chicken breast	Food
FDA169	SAMN05201660	SRR3665205	FDA	2011	USA:GA	Ground Beef	Food
FDA170	SAMN02846717	SRR3721579	CFSAN	2012	Mexico	papaya	Food
FDA171	SAMN05416490	SRR3932996	FDA	2015	USA:GA	Chicken Breast	Food
FDA172	SAMN05417586	SRR3933129	FDA	2013	USA:NJ	fecal	Canine
FDA173	SAMN05417587	SRR3933130	FDA	2013	USA:PA	fecal	Canine
FDA174	SAMN02918523	SRR3938672	CFSAN	2008	China	soybean meal, animal feed	Animal Feed
FDA175	SAMN02918591	SRR3952228	CFSAN	2009	Mexico	animal feed, fish meal	Animal Feed
FDA176	SAMN04962024	SRR4012815	CFSAN	2013	USA:GA	Environmental (Pond)	Environmental
FDA177	SAMN04962031	SRR4012822	CFSAN	2013	USA:GA	Environmental (Pond)	Environmental
FDA178	SAMN02918607	SRR4014994	CFSAN	2010	USA:ID	dog food	Animal Feed
FDA179	SAMN02847019	SRR4015002	CFSAN	2010	Mexico	barley drink	Food
FDA180	SAMN02844839	SRR4119781	CFSAN	2010	India	sesame seeds	Food
FDA181	SAMN02847372	SRR4124944	CFSAN	2012	USA:NM	environmental sample	Environmental

FDA182	SAMN02847371	SRR4124946	CFSAN	2012	USA:NM	environmental sample	Environmental
FDA183	SAMN02847344	SRR4279933	CFSAN	2011	Mexico	papaya	Food
FDA184	SAMN02847319	SRR4292702	CFSAN	2011	Mexico	papaya	Food
FDA185	SAMN02918989	SRR4301128	CFSAN	2012	Mexico	animal feed, wheat millrun	Animal Feed
FDA186	SAMN02918994	SRR4301131	CFSAN	2012	Mexico	animal feed, wheat millrun	Animal Feed
FDA187	SAMN02843824	SRR4427087	CFSAN	2003	Taiwan	black pepper	Food
FDA188	SAMN05897854	SRR4733929	CFSAN	2016	India	Mix Spice Powder	Food
FDA189	SAMN02919045	SRR4733957	CFSAN	2014	USA:MI	animal feed, dry cat food	Animal Feed
FDA190	SAMN02845884	SRR5105537	CFSAN	2007	Mexico	dried pepper flakes	Food
FDA191	SAMN06213901	SRR5195799	CFSAN	2016	United Kingdom	Pistachio Kernels	Food
FMA001	SAMN02345360	SRR1139518	CFSAN	2011	USA:NM	peanut butter	Food
FMA002	SAMN02345449	SRR1158009	CFSAN	2012	USA:NM	environmental sample	Environmental
FMA003	SAMN02345424	SRR1175791	CFSAN	2011	USA:NM	environmental sample	Environmental
FMA004	SAMN02345578	SRR1198887	FSIS SRCAMB	2012	Mexico	papaya	Food
FMA005	SAMN02345577	SRR1198926	FSIS SRCAMB	2012	Mexico	papaya	Food
FNE001	SAMN02345251	SRR1097807	CFSAN	2011	Turkey	halva cocoa	Food
FNE002	SAMN02345073	SRR1185831	FSIS SRCAMB	2011	Hong Kong	dried gecko	Food
FNW001	SAMN02344770	SRR1170779	CFSAN	2010	Mexico	barley drink	Food
FNW002	SAMN02344928	SRR1177185	CFSAN	2010	Mexico	papaya	Food
FNW003	SAMN02344972	SRR1177233	CFSAN	2010	Mexico	frozen octopus	Food
FNW004	SAMN02344962	SRR1177258	CFSAN	2010	Mexico	papayas	Food
FNW005	SAMN02344963	SRR1177558	CFSAN	2010	Mexico	papayas	Food
FNW006	SAMN02345274	SRR1203043	CFSAN	2011	Mexico	papayas	Food
FSW001	SAMN02344988	SRR1068334	CFSAN	2010	Mexico	papaya	Food
FSW002	SAMN02344985	SRR1068343	CFSAN	2010	Mexico	papaya	Food
MDH001	SAMN02378204	SRR1029568	CFSAN	2013	USA:MN	tahini	Food

MDH002	SAMN02646878	SRR1265009	CFSAN	2004	USA:MN	Chicken Breast	Food
MDH003	SAMN02646867	SRR1269360	CFSAN	2003	USA:NC	Porcine Tissue	Porcine
MDH004	SAMN02646849	SRR1272512	CFSAN	2003	USA:NC	Porcine Tissue	Porcine
MDH005	SAMN02646792	SRR1299349	CFSAN	2003	USA:WI	Bovine Feces	Bovine
MDH006	SAMN02699424	SRR1300702	CFSAN	2000		Chicken Thigh	Food
MDH007	SAMN02699477	SRR1425269	CFSAN	2002	USA:MN	Canine Feces	Canine
MDH008	SAMN02699511	SRR1461784	CFSAN	2002	USA:MN	Porcine Tissue	Porcine
MDH009	SAMN02699522	SRR1501631	CFSAN	2002	USA:MN	Canine Feces	Canine
MDH010	SAMN02699596	SRR1553727	CFSAN	2003	USA:MN	Porcine Tissue	Porcine
NYSDH001	SAMN02222898	SRR1030357	CFSAN	2007	USA:CT	raw meat for pet food	Animal Feed
NYSDH002	SAMN02222934	SRR1067637	CFSAN	2013	USA:CT	Milk, Raw	Food
NYSDH003	SAMN01902330	SRR1106268	CFSAN	2009	USA:NY	DOG FOOD	Animal Feed
NYSDH004	SAMN01902413	SRR1107476	CFSAN	2012	USA:NY	GROUND BEEF	Food
NYSDH005	SAMN01902362	SRR1157834	CFSAN	2010	USA:NY	CHICKEN BREAST	Food
NYSDH006	SAMN01902345	SRR1272891	CFSAN	2009	USA:NY	CHICKEN BREAST	Food
NYSDH007	SAMN03795424	SRR2847927	CFSAN	2017	USA:NY	stool	Human
ODA001	SAMN05721600	SRR4237694	CFSAN	2012	USA:OH	Animal Feed	Animal Feed
ODA002	SAMN05577716	SRR4237742	CFSAN	2010	USA:OH	Feed	Animal Feed
ODA003	SAMN05721585	SRR5040864	CFSAN	2013	USA:OH	Dog Food	Animal Feed
ODA004	SAMN05721573	SRR5104745	CFSAN	2015	USA:OH	Animal Feed	Animal Feed
OSU001	SAMN03577516	SRR3392184	CFSAN	2008	USA:NC	floor swab	Environmental
OSU002	SAMN03577713	SRR3554452	CFSAN	2014	USA:OH	poultry env.	Environmental
OSU003	SAMN03577719	SRR3554464	CFSAN	2013	USA:OH	poultry env.	Environmental
OSU004	SAMN03577717	SRR3554467	CFSAN	2013	USA:OH	poultry env.	Environmental
OTH001	SAMN01924646	SRR1030358	CFSAN	2011		Water GA Pond- VH1	Environmental
OTH002	SAMN02345576	SRR1033524	CFSAN	2012	USA:MD	organic anise seed, whole	Food

OTH003	SAMN02344971	SRR1041533	CFSAN	2010	Mexico	papayas	Food
OTH004	SAMN02183004	SRR1106152	CFSAN	2011	USA:WA	chicken	Avian
OTH005	SAMN02182958	SRR1106269	CFSAN	2012	USA:WA	giblets combo	Food
OTH006	SAMN02368517	SRR1118716	UCD-100K		USA		Others
OTH007	SAMN02345100	SRR1175817	CFSAN	2011	Mexico	peanut, shelled	Food
OTH008	SAMN02403245	SRR1188475	FSIS SRCAMB	2005	USA:VA	Chicken	Avian
OTH009	SAMN02403247	SRR1207489	CFSAN	2005	USA:VA	Chicken	Avian
OTH010	SAMN02344968	SRR1212275	CFSAN	2010	Mexico	papaya	Food
OTH011	SAMN02403373	SRR1212353	CFSAN	2006	USA:VA	Horse	Equine
OTH012	SAMN02645801	SRR1548424	CFSAN	2009	USA:Dauphin PA	Chicken Breast	Food
OTH013	SAMN02345456	SRR1615964	CFSAN	2012	USA:NM	peanut butter	Food
OTH014	SAMN02345467	SRR1615965	CFSAN	2012	USA:NM	environmental swab	Environmental
OTH015	SAMN03112866	SRR1618705	CFSAN	2011	USA:VA	seagull	Avian
OTH016	SAMN03112867	SRR1618710	CFSAN	2011	USA:VA	seagull	Avian
OTH017	SAMN02908542	SRR1686510	USDA	2004	USA	Market hog swab	Porcine
OTH018	SAMN02908629	SRR1686597	USDA	2006	USA	Ground chicken	Food
OTH019	SAMN02908630	SRR1686598	USDA	2010	USA	Market hog	Porcine
OTH020	SAMN03255329	SRR1745631	CFSAN	2012	Turkey	egg	Food
OTH021	SAMN03702749	SRR2054190	CFSAN	2015	USA:TX	raw ground beef	Food
OTH022	SAMN03702738	SRR2096542	CFSAN	2014	USA:TX	ground hamburger	Food
OTH023	SAMN03702740	SRR2096596	CFSAN	2014	USA:TX	hamburger	Food
OTH024	SAMN04506151	SRR3185059	EDLB-CDC		USA	Urine	Environmental
OTH025	SAMN04528237	SRR3205869	CFSAN	2013	USA:NC	Soil	Environmental
OTH026	SAMN04528240	SRR3205872	CFSAN	2013	USA:NC	Soil	Environmental
OTH027	SAMN04544918	SRR3278367	CFSAN	2013	USA:NC	Lagoon	Environmental
OTH028	SAMN03800261	SRR3394981	CFSAN	2012	Chile:Rancagua	feces	Environmental

OTH029	SAMN03218269	SRR4011083	CFSAN	2008	United Kingdom	retail meat	Food
OTH030	SAMN05560723	SRR4069542	CFSAN	2015	USA:IA	Raw Beef materials	Food
OTH031	SAMN05829080	SRR4299026	CFSAN	2015	USA:NY	canis rufus feces	Canine
OTH032	SAMN05948800	SRR4473729	CFSAN	2016	USA:IA	Beef Trim	Food
OTH033	SAMN05203386	SRR5010543	CFSAN		USA:WA	orange juice	Food
OTH034	SAMN05203387	SRR5010546	CFSAN	2013	USA:MN	tahini	Food
OTH035	SAMN03743881	SRR5057141	CFSAN	2005	Argentina:Buenos Aires	factory swab	Environmental
OTH036	SAMN03743905	SRR5057145	CFSAN	2005	Argentina:Buenos Aires	flour	Environmental
OTH037	SAMN03743880	SRR5057147	CFSAN	2005	Argentina:Buenos Aires	environmental sample	Environmental
USDA001	SAMN03218359	SRR1720462	CFSAN	2012	USA:GA	whole eggs	Food
USDA002	SAMN03218356	SRR1720470	CFSAN	2012	USA:GA	egg yolks	Food
USDA003	SAMN03218378	SRR1745534	CFSAN	2012	USA:GA	egg whites	Food
USDA004	SAMN03218236	SRR1745544	CFSAN	2012	USA:GA	egg whites	Food
USDA005	SAMN03218249	SRR1745558	CFSAN	2012	USA:GA	whole eggs	Food
USDA006	SAMN03218253	SRR1745623	CFSAN	2012	USA:AR	egg whites	Food
USDA007	SAMN03285092	SRR1774090	CFSAN	2012	USA:GA	product-eggs-raw-whole	Food
USDA008	SAMN03776990	SRR2068068	USDA FSIS	2015	USA:NC	NRTE Comminuted Poultry Exploratory Sampling	Food
USDA009	SAMN03649173	SRR2075049	CFSAN	2012	USA:AR	product-eggs-raw-whole	Food
USDA010	SAMN03763518	SRR2078900	CFSAN	2012	USA:AL	product-eggs-raw-whole	Food
USDA011	SAMN03464558	SRR2124289	CFSAN	2012	USA:NY	product-eggs-raw-whites	Food
USDA012	SAMN03464560	SRR2124492	CFSAN	2012	USA:IA	product-eggs-raw-yolks	Food
USDA013	SAMN03921964	SRR2125864	USDA-FSIS	2015	USA:GA	NRTE Comminuted Poultry Exploratory Sampling	Food
USDA014	SAMN03838236	SRR2152998	CFSAN	2012	USA:GA	product-eggs-raw-yolks	Food
USDA015	SAMN03649168	SRR2156500	CFSAN	2012	USA:TX	product-eggs-raw-whole	Food

USDA016	SAMN03838261	SRR2174022	CFSAN	2012	USA:MN	product-eggs-raw-whole	Food
USDA017	SAMN04014662	SRR2188119	USDA-FSIS	2015	USA:PA	Animal-Swine-Sow	Porcine
USDA018	SAMN04027097	SRR2239755	USDA-FSIS	2015	USA:NC	Animal-Swine-Market Swine	Porcine
USDA019	SAMN04090025	SRR2421589	USDA-FSIS	2015	USA:IN	Raw Intact Chicken	Avian
USDA020	SAMN03218369	SRR2890151	CFSAN	2012	USA:NY	whole eggs	Food
USDA021	SAMN03218366	SRR2890153	CFSAN	2012	USA:NY	egg yolks	Food
USDA022	SAMN04260528	SRR2920118	USDA-FSIS	2015	USA:WI	Animal-Cattle-Beef Cow	Bovine
USDA023	SAMN04260961	SRR2920120	USDA-FSIS	2015	USA:ID	Animal-Cattle-Dairy Cow	Bovine
USDA024	SAMN04376256	SRR3062684	USDA-FSIS	2015	USA:IA	Animal-Swine-Market Swine	Porcine
USDA025	SAMN04437782	SRR3115432	USDA-FSIS	2015	USA:TX	Animal-Swine-Market Swine	Porcine
USDA026	SAMN04481314	SRR3156224	USDA-FSIS	2016	USA:TX	Animal-Cattle-Steer	Bovine
USDA027	SAMN04575036	SRR3284587	USDA-FSIS	2016	USA:TX	Animal-Cattle-Steer	Bovine
USDA028	SAMN03922164	SRR3323110	USDA-FSIS	2015	USA:GA	NRTE Comminuted Poultry Exploratory Sampling	Food
USDA029	SAMN04942710	SRR3475887	USDA-FSIS	2016	USA:CA	Animal-Cattle-Dairy Cow	Bovine
USDA030	SAMN04942595	SRR3476439	USDA-FSIS	2015	USA:AL	Chicken Carcass	Avian
USDA031	SAMN05150605	SRR3555083	USDA-FSIS	2016	USA:KS	Animal-Cattle-Beef Cow	Bovine
USDA032	SAMN05150600	SRR3555189	USDA-FSIS	2016	USA:MI	Animal-Swine-Sow	Porcine
USDA033	SAMN05720506	SRR4106448	USDA-FSIS	2016	USA:MA	Animal-Swine-Market Swine	Porcine
USDA034	SAMN05900914	SRR4418021	USDA-FSIS	2016	USA:GA	Animal-Cattle-Dairy Cow	Bovine
USDA035	SAMN05900903	SRR4419039	USDA-FSIS	2016	USA:WA	Product-Raw-Intact-Beef	Food
USDA036	SAMN05945097	SRR4453696	USDA-FSIS	2016	USA:TN	Animal-Chicken-Young Chicken	Avian
USDA037	SAMN06015764	SRR5019458	USDA-FSIS	2016	USA:MD	Animal-Chicken-Young Chicken	Avian
USDA038	SAMN06015786	SRR5019515	USDA-FSIS	2016	USA:PA	Animal-Swine-Sow	Porcine
USDA039	SAMN06046041	SRR5043210	USDA-FSIS	2016	USA:OK	Raw-Ground, Comminuted Pork	Food

USDA040	SAMN06048913	SRR5045370	USDA-FSIS	2016	USA:FL	Comminuted Chicken	Food
USDA041	SAMN06187008	SRR5132770	USDA-FSIS	2016	USA:FL	Raw Intact Chicken	Avian
USDA042	SAMN06235216	SRR5182328	USDA-FSIS	2016	USA:TN	Animal-Swine-Market Swine	Porcine

*EUR047 and FDA014 were deliberately removed as those isolates were not included in this study

2.3 Results and Discussion

2.3.1 Phylogenetic analysis of *S*. Mbandaka

Currently, no information is available regarding phylogenetic organization within single serovar *S*. Mbandaka. A phylogenetic tree was constructed using next-generation sequence data of 465 *S*. Mbandaka isolates to explore the evolutionary genetic diversity of *Salmonella* serovar Mbandaka. Sequences for newly sequenced 76 *S*. Mbandaka isolates from 9 different North American States and one South American state were phylogenetically analyzed along with sequences for 388 isolates from different parts of the world. Metadata including biosample accession number and NCBI SRA accession number were given in table 5. A comprehensive phylogenetic characterization, using next-generation sequencing based SNP analysis and cgMLST methods, was performed to elucidate the evolutionary relationship of these isolates in a global context. Resultant phylogenetic trees were used to generate various hypothesis related to the evolution, host distribution and ability to cause human outbreaks from various sources.

2.3.1.1 SNP based Analysis

A total of 87,089 genome positions with SNPs were detected by pairwise comparison of all *S*. Mbandaka isolate genomes to the reference genome *S*. Mbandaka str. ATCC 51958 (NCBI Reference Sequence accession: NZ_CP019183.1). Similar nucleotide changes happened at 1974 genome positions in all 465 isolates. Remaining

SNVs at 85115 genome positions with varying nucleotide changes were distributed between study isolates. Phylogenetic analysis of above SNP data generated an evolutionary tree comprised of two primary clusters, ‘Cluster 1’ and ‘Cluster 2’. For the ease of understanding, we identified six sub clusters in ‘Cluster 2’ based on the close relationship of isolates in the context of isolation source group and geographical origin (fig.7). Number of unique SNPs that differentiate each cluster from any other cluster were identified by comparing representative isolates from each cluster. The details of identified SNPs were given in table 7.

Cluster 1

A cluster consists of minimum number of genomic positions with SNPs from the reference genome. Cluster1 included 13 isolates, all of which were collected from same geographical area, Texas USA (Fig.8). This outlying group was again characterized by the presence of isolates collected from same isolation source, cattle, except for one isolate (USDA015), which was isolated from raw whole egg. Interestingly, 35 isolates collected from the same location (Texas, USA) and from same isolation source (Cattle) were clustered together in another place of the phylogenetic tree forming a sub cluster (sub cluster 2B). Out of 85115 only 9 genomic position with SNVs distributed between the members of this cluster. Remaining 85106 SNVs distributed in cluster 2. When compared with sub cluster 2B (representative sample ADRDL27 of cluster 1and FDA083 of subcluster2B) 14 SNPs present in coding sequences (CDS) were unique to cluster 1 that differentiated this primary cluster from sub cluster 2B. Eight of these SNPs were non-synonymous resulting in an amino acid change in respective proteins while six of them were synonymous. Important ones include Non-Synonymous SNP in Gene *pyrG*

encoding CTP synthatse and in CDS for protein magnesium translocating P-type ATPase resulted in amino acid change asparagine (Asn) to (Tyr) tyrosine and (Val) Valine to Leucine (Leu) respectively. Poultry isolate USDA015, on the other hand carried one unique SNP differentially from other bovine isolates of the cluster 1. This unique SNP occurred in CDS for type II secretion system protein GSpE resulting in an amino acid change Glutamine(Glu) to Aspartate (Asp).

Cluster 2

Remaining 452 *S.* Mbandaka isolates constituted cluster 2. These isolates were again sub clustered into 6 sub clusters in correlation to isolation source and geographical area of isolate collection.

Sub cluster 2A: A cluster consisting of 40 isolates differentiated by SNPs in 89 genomic positions. Correlating factor that outline this cluster is the isolation source especially for 29 *S.* Mbandaka isolates from 14 different states of USA collected over a period of 2002 to 2015. All these 29 isolates were obtained from poultry and/or poultry related products such as chicken breast, egg and chicken wings. Isolation source of other closely related isolates in this cluster consisting of equine (1), porcine (2), human (1), animal feed (2) and environmental sources (4). Two non-synonymous SNPs by frame shift mutation in the coding sequences of phage tail proteins were found unique to this cluster (represent-ative sample USDA005) compared to cluster1. A total of 21 unique SNPs that includes 10 non-synonymous SNPs resided in this cluster separated sub cluster 2A from its sister lineage sub cluster 2B containing mainly bovine isolates.

Sub cluster 2B: Fifty isolates, out of seventy-six newly sequenced *S*. Mbandaka isolates here in our lab, were bovine isolates collected from Texas, USA during the period 2006-2007. Out of these fifty isolates, 12 were clustered in cluster 1, while 35 were clustered together along with 18 other isolates forming a sub cluster (2B) in cluster 2. There were 39 unique SNPs including 27 non - synonymous SNPs differentiated this sub cluster (representative sample FDA083 for sub cluster 2B) from cluster 1 bovine isolates. There was a total of 53 isolates in sub cluster 2B. 51 isolates shared a common characteristic of isolation source, that they were isolated either from cattle, beef or beef products. Two exceptions (OTH019 and FDA092), were from pig or pig products. There were 268 genomic positions with SNPs distributed between members of this sub cluster. Out of 39 unique SNPs, 15 SNPs were in phage elements and 9 were in hypothetical proteins. Nine non-synonymous SNPs were identified in major bacterial protein coding sequences that include sequences encoding flagellin FliC, SPI-1 effector StpP.

Analysis of SNPs between representative isolates from cluster1 (ADRDL27) and sub cluster2B (ADRDL 45) collected from same isolation source (cattle) and location (Texas:USA) revealed 82 unique SNPs that contain 61 non-synonymous SNPs in sub cluster 2B isolate (ADRDL 45) which is nearly five times greater than what had for the cluster1. Out of 61, two of the non- synonymous SNPs occurred in *ligA* (SEEM1958 _RS07470) and *thiP* (SEEM1958_RS19865) genes that encode for DNA ligase and thiamine/thiamine pyrophosphate ABC transporter permease ThiP respectively. ligA protein is essential for DNA replication and repair. Protein encoded by *thiP* functions in the transport of thiamine in to the cell. Non-synonymous SNPs in coding sequences of two integrases (SEEM1958_RS06125 and SEEM1958_RS10835), MFS transporter

(SEEM1958_RS17580), CesD/SycD/LcrH family type III secretion system chaperone (SEEM1958_RS13550) might be playing important roles as the first two are related to antimicrobial resistance and later one to pathogenicity. Nearly 22 SNPs occured in phage related protein coding sequences (CDS) and 14 were in CDS for hypothetical proteins.

Moreover, 36 SNPs were identified as unique to this cluster when did a comparison between sub cluster 2A poultry isolate (USDA005) with sub cluster 2B isolate (FDA 083). Twenty-four SNPs were non-synonymous SNPs that involve SNPs in CDS of diamino- pimelate decarboxylase, flagellin FliC, tRNA (guanosine(18)-2'-O)-methyl transferase TrmH, Integrase, and AraC family transcriptional regulator. Remaining ones occurred mainly in phage related elements.

Sub cluster 2C: Twenty-eight isolates obtained from two different isolation source group constitute this cluster. 71% of isolates in this cluster (20 isolates- including two isolates FDA046 and -048 in the sister lineage) were obtained from food commodities of Asia, North America and Europe. Remaining 29% (8 isolates) were human isolates from Europe alone. SNPs at 141 genomic positions were distributed in this cluster. Fifteen unique SNPs identified in the representative sample FDA015, differentiated this sub cluster from cluster 1bovine isolates. Eleven SNPs were non-synonymous including SNP in a gene *srlA* (SEEM1958_RS05760) that encodes a protein PTS sorbitol transporter subunit IIC. Sub unit IIC domain forms the PTS system translocation channel required for the translocation of sugar substrate across the cell membrane. Irrespective of geographical location these closely related *S.* Mbandaka strains shared a common ancestry. Apart from that, clustering of these genetically similar isolates from food

commodities and human source may indicate the transmission capabilities and potential strength of food related *S*. Mbandaka strains to cause food borne salmonellosis in human.

Sub cluster 2D: Sixteen isolates collected from Tennessee state of USA, clustered together forming a sister cluster of sub cluster 2C and sub cluster 2E. All isolates in this cluster isolated from meat products that includes chicken breast, ground beef, and pork chop collected during the year 2008. There were no SNPs unique to this cluster at ≥50x coverage compared with cluster 1. Since this cluster shared a common ancestry with sub cluster 2C as well as with sub cluster 2E, an analysis was done to find out how many unique SNPs differentiate this sub cluster from other two sub clusters. There were no SNPs found at coverage of ≥50x. When we reduce the criteria to ≥ 20x, for the representative sample FDA133, there were 11 SNPs in comparison with sub cluster 2C (FDA015) and 14 with sub cluster 2E (EUR069). Ten were common in comparison with subcluster2C and 2E that included SNPs in two genes flil (codes for flagellum specific ATP synthase) and *artM* (codes for arginine transporter permease subunit ArtM). Non-synonymous SNP that causes an aminoacid change Valine (Val) to isoleucine (Ile) happened in *artM* gene while synonymous SNP occurred in flil gene. Out of ten, six were non-synonymous unique SNPs.

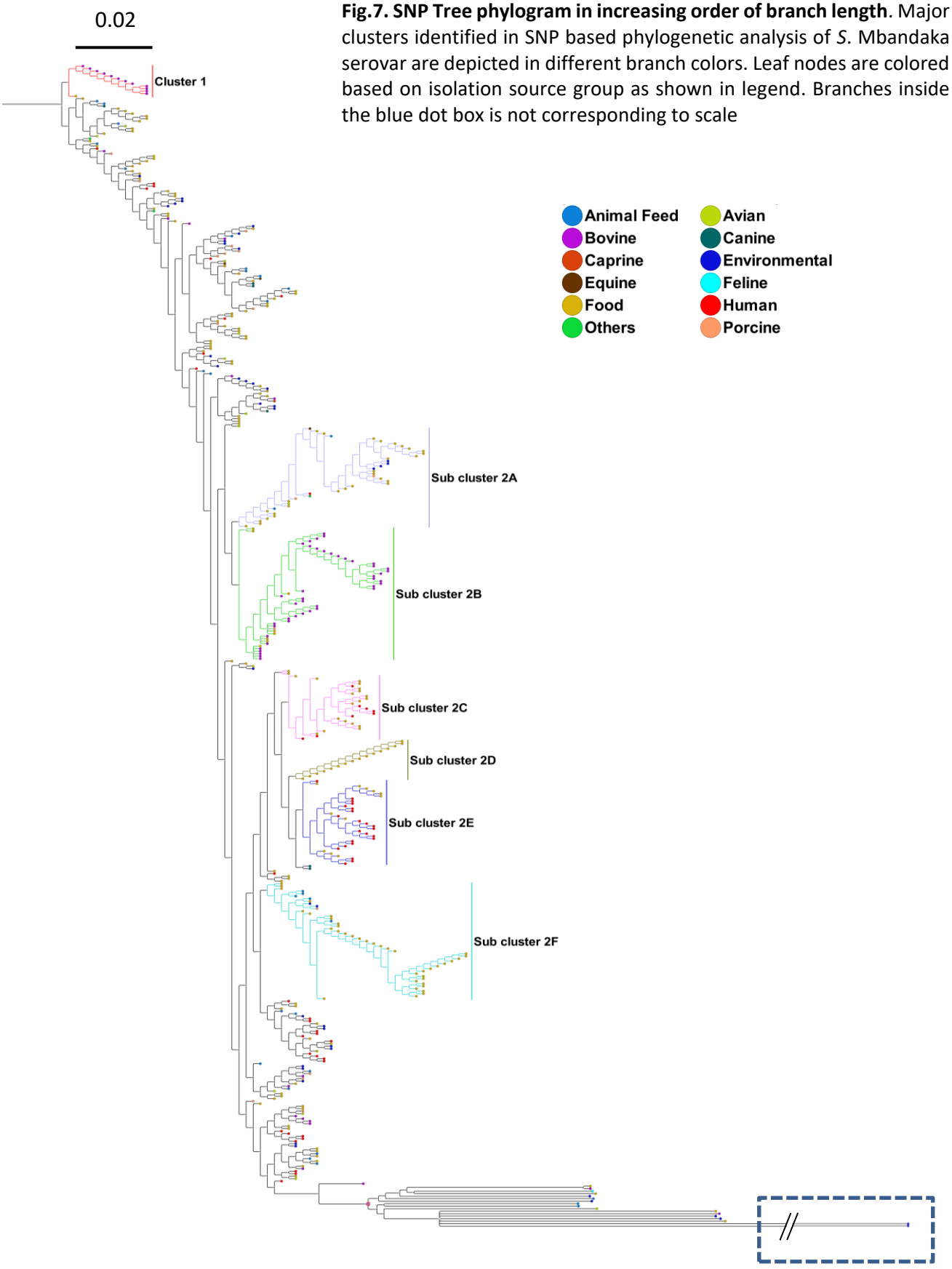
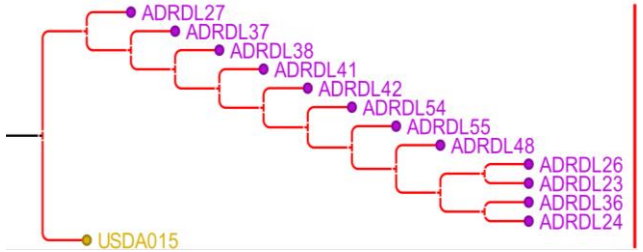
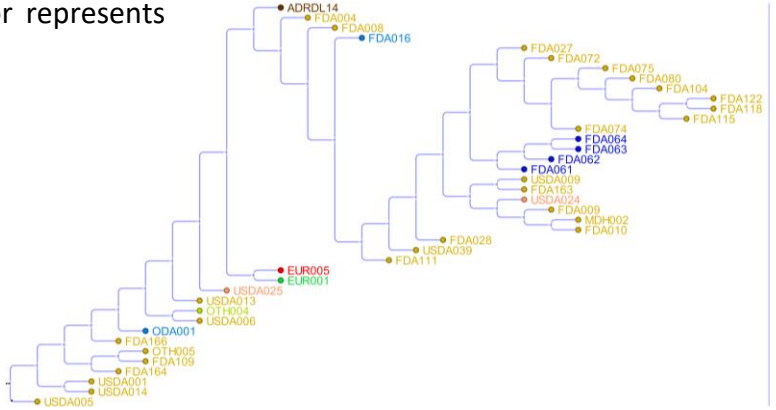


Fig.7. SNP Tree phylogram in increasing order of branch length. Major clusters identified in SNP based phylogenetic analysis of *S*. Mbandaka serovar are depicted in different branch colors. Leaf nodes are colored based on isolation source group as shown in legend. Branches inside the blue dot box is not corresponding to scale

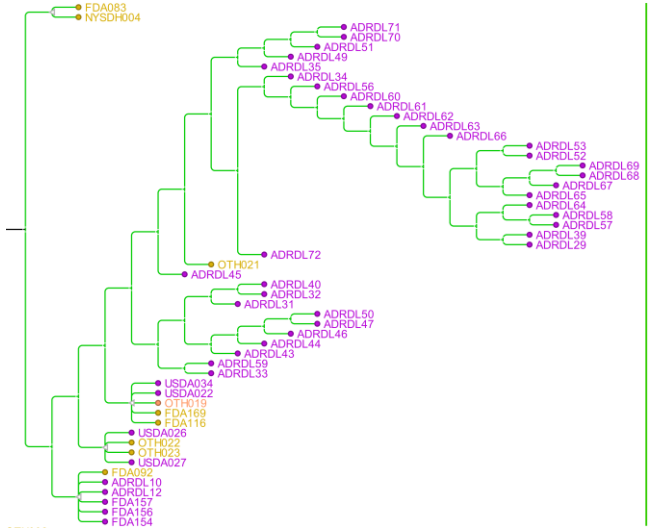
Fig.8. Major clusters in SNP tree phylogram in increasing order of branch length. Major clusters identified in phylogenetic analysis of *S*. Mbandaka serovar are shown in different branch colors. Cluster 1 and six sub clusters of primary cluster 2 are shown individually. To make leaf label visible magnified images shown in this cartoon are not corresponding to any scale. Leaf label represents sample ID and color represents isolation source group



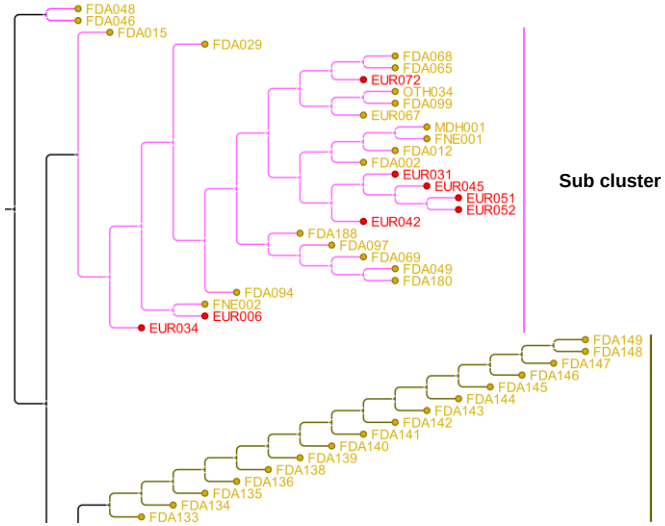
Cluster 1



Sub cluster 2A

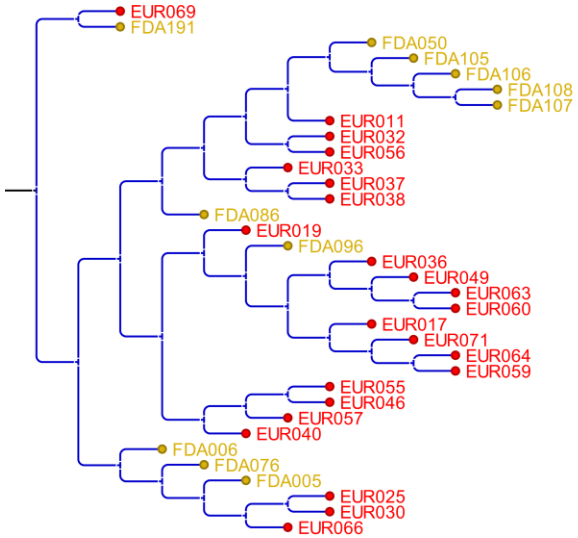


Sub cluster 2B

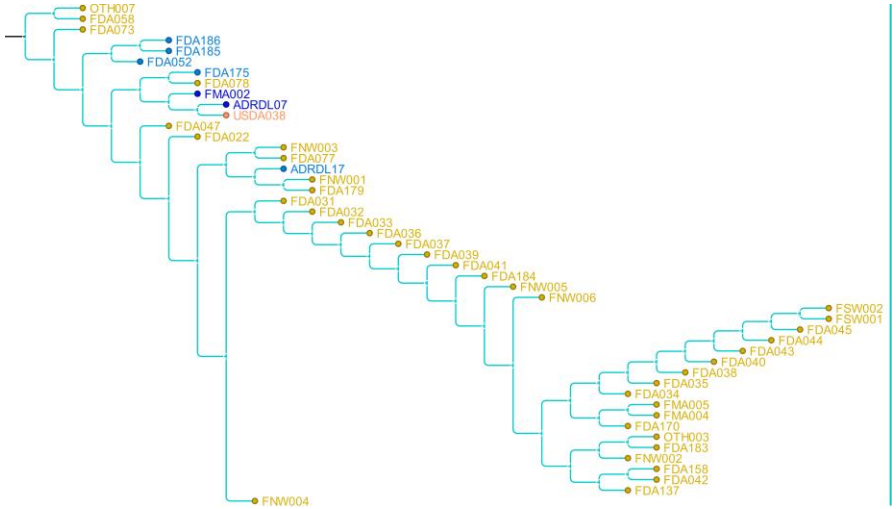


Sub cluster 2C

Sub cluster 2D



Sub cluster 2E



Sub cluster 2F

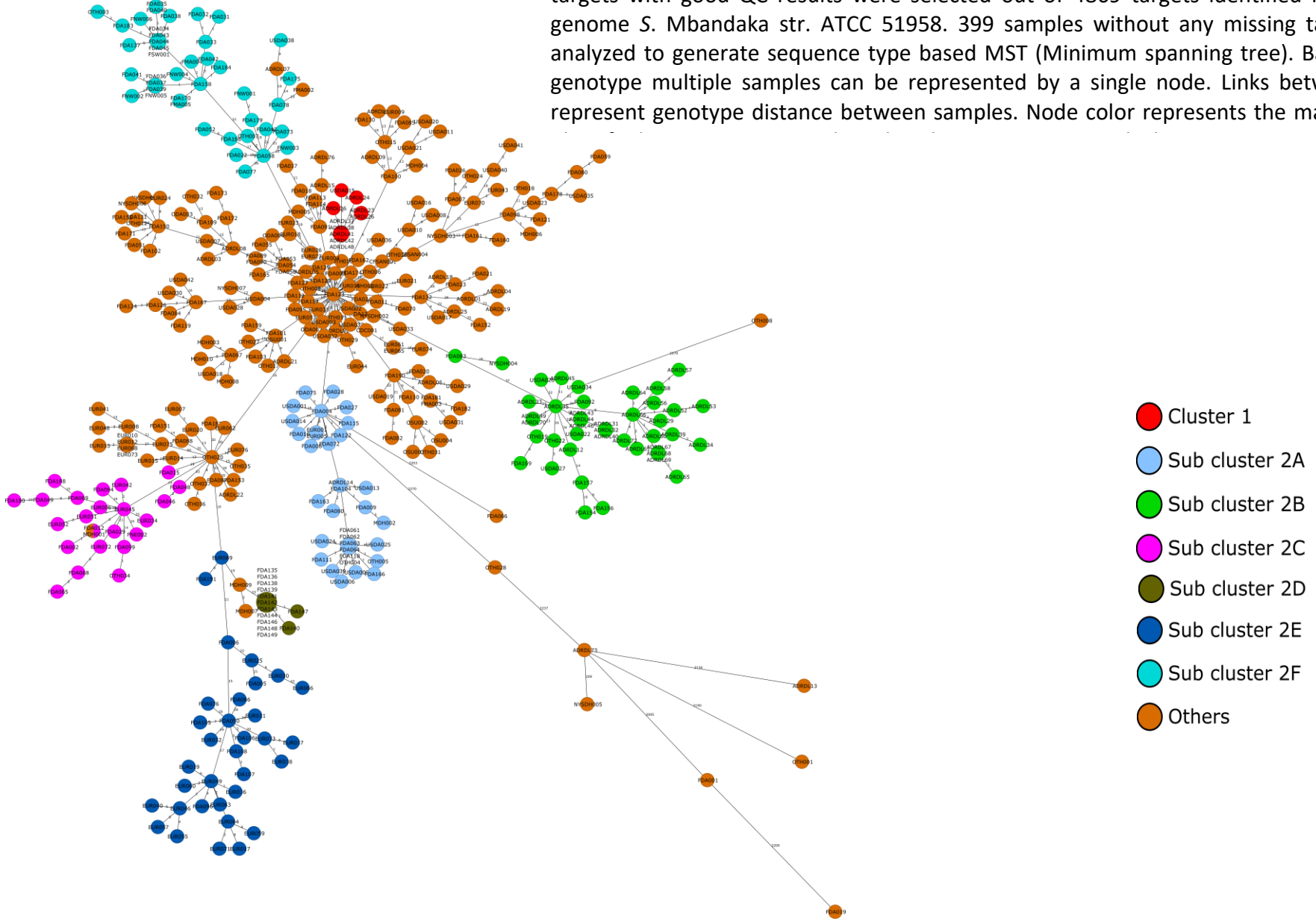


Fig. 9. Minimum spanning tree (MST) created by cgMLST. For distance calculation, 2502 targets with good QC results were selected out of 4869 targets identified in reference genome *S*. Mbandaka str. ATCC 51958. 399 samples without any missing targets were analyzed to generate sequence type based MST (Minimum spanning tree). Based on the genotype multiple samples can be represented by a single node. Links between nodes represent genotype distance between samples. Node color represents the major clusters

Sub cluster 2E: Comprised of 34 total isolates this cluster is defined by 157 SNPs. Isolates from UK formed 70.5% (24 isolates) of this cluster of which 23 were human isolates and one from food commodity. Isolates from Asian continent constituted remaining 29.5% all of which were obtained from food commodities. A close genetic relationship of Asian isolates from food commodities and human isolates from different geographical locations of UK form highlighting feature of this cluster formation. There were 15 unique SNPs identified for this cluster compared to cluster 1 including synonymous SNP in *carB* gene that codes for large subunit of carbamoyl phosphate synthase. Eight SNPs were non-synonymous SNPs, the important ones include SNPs in the coding sequences for MFS transporter (locus tag: SEEM1958_RS18305) that confers resistance to antibiotic fosfomycin, integrase (SEEM1958_RS17410), PhoPQ regulated protein (SEEM1958 _RS12735) and sugar efflux transporter (SEEM1958_RS12855). Out of these 15 SNPs, 13 SNPs made this cluster different from its closely related cluster sub cluster 2C, which contained isolates from similar sources, though they shared a common ancestor. Non-synonymous SNP in a CDS (locus tag SEEM1958_RS13070) that codes for major facilitator superfamily (MFS) transporter and a synonymous SNP at SEEM1958_RS02400 that codes for tRNA dihydrouridine synthase DusB were the two additional unique SNPs identified from cluster 1. The number of SNPs unique to sub cluster 2E from sub cluster 2D is similar as that from cluster 1, but three SNPs occurred at different locus. One non-synonymous SNP was in a hypothetical protein (SEEM1958_RS18715) while other two were synonymous ones in CDS for Rhs family protein (SEEM1958_RS18815) and flagellin FliC (SEEM1958_RS06100). Basically,

twelve SNPs unique to sub cluster 2E were similar in comparison with cluster1, sub cluster 2C and 2D.

Sub cluster 2F: This is a geographically well-defined cluster comprising 43 out of 47 Mexican isolates contained in this study. A total of 47 isolates sharing SNPs in 136 genomic positions were included in this cluster. Two isolates from Michigan, one from Pennsylvania and one from New Mexico were the remaining 4 isolates in the cluster. Out of 43 Mexican isolates, twenty-nine isolates were clustered together correlating with their common isolation source papaya. In contradiction to this unique clustering, one Isolate (OTH010), which was also a Mexican isolate from papaya located phylogenetically away from this cluster. Remaining fourteen isolates in this cluster were isolated either from a food commodity or from Animal feed. A total of 33 unique SNPs that differentiate this cluster from cluster 1. Fourteen were non-synonymous and 19 were synonymous substitutions within protein coding sequences. More than half of the non-synonymous substitutions happened in phage elements.

Further evaluation with cgMLST was performed to validate the consistencies of the clusters formed in SNP analysis.

2.3.1.2 Sequence type (ST) based Analysis

In contrast to SNP based neighbor joining tree method, where sequence reads were used to analyze single nucleotide variation, de novo assembled sequence contigs were used for MLST analysis. Instead of looking for nucleotide variations in the whole genome, variations in selected target genes common to reference as well as sample genomes were identified in MLST analysis. Distance calculation calculated using different algorithm based on genotype makes this analysis as an entirely different approach for phylogenetic

analysis. Genetic relatedness of isolates in different clusters in SNP analysis was reaffirmed by this different approach indicating that those isolates are truly remain close phylogenetically.

For MLST distance calculation 2502 targets with good QC results were selected out of 4869 targets identified in reference genome. Three hundred and ninety nine samples without any missing targets were filtered to generate sequence type based MST (Fig. 9). There were 14 SNV positions in 14 target sequences identified between 10 isolates of cluster1. SNVs were synonymous at two places while at 12 targets at least one sample showed non-synonymous SNVs. Including SNVs from reference sequences a total of 21115 non InDel SNV positions identified from 2300 targets identified between cluster1 isolates. While for sub cluster 2B total filtered SNVs without InDels (including SNVs from reference sequences) were 21409 from 2313 targets (Data not shown).

2.3.1.3 Pangenome analysis and protein clustering of *S*. Mbandaka

A much more resolution in the analysis of intra serovar genetic diversity within the serovar *S*. Mbandaka was accomplished by pangenome and protein clustering analysis based on distribution of presence and absence of genes between isolate's genome. In contrast to previous analysis methods that were based on nucleotide variations either in the genome sequences as in SNP analysis or in selected core genes as in cgMLST, pangenome analysis elucidate the gene content variation between isolates. SNP and cgMLST analysis provide information regarding genes or sequences in comparison with reference genome and does not give information about any additional genes present or absent in study isolates. Pangenome analysis can provide information about variable genes present in each isolate.

Salmonella core genome and pangenome were estimated to be around 2800 and 10000 gene families respectively [65]. Pangenome of a *Salmonella* strain is the total of stable core genome as well as abundance of accessory genome, including SPIs, plasmids, phage and transposable elements [65]. Pangenome has also been described as addition of core genome (orthologs shared among multiple genomes) and variable genome which includes gene families shared by two or more organism and singletons (referred to as strain specific genes having no orthologs in corresponding genomic strains) [194]. Orthologs are the genes in different species that evolved by speciation from an ancestor gene but maintaining similar function.

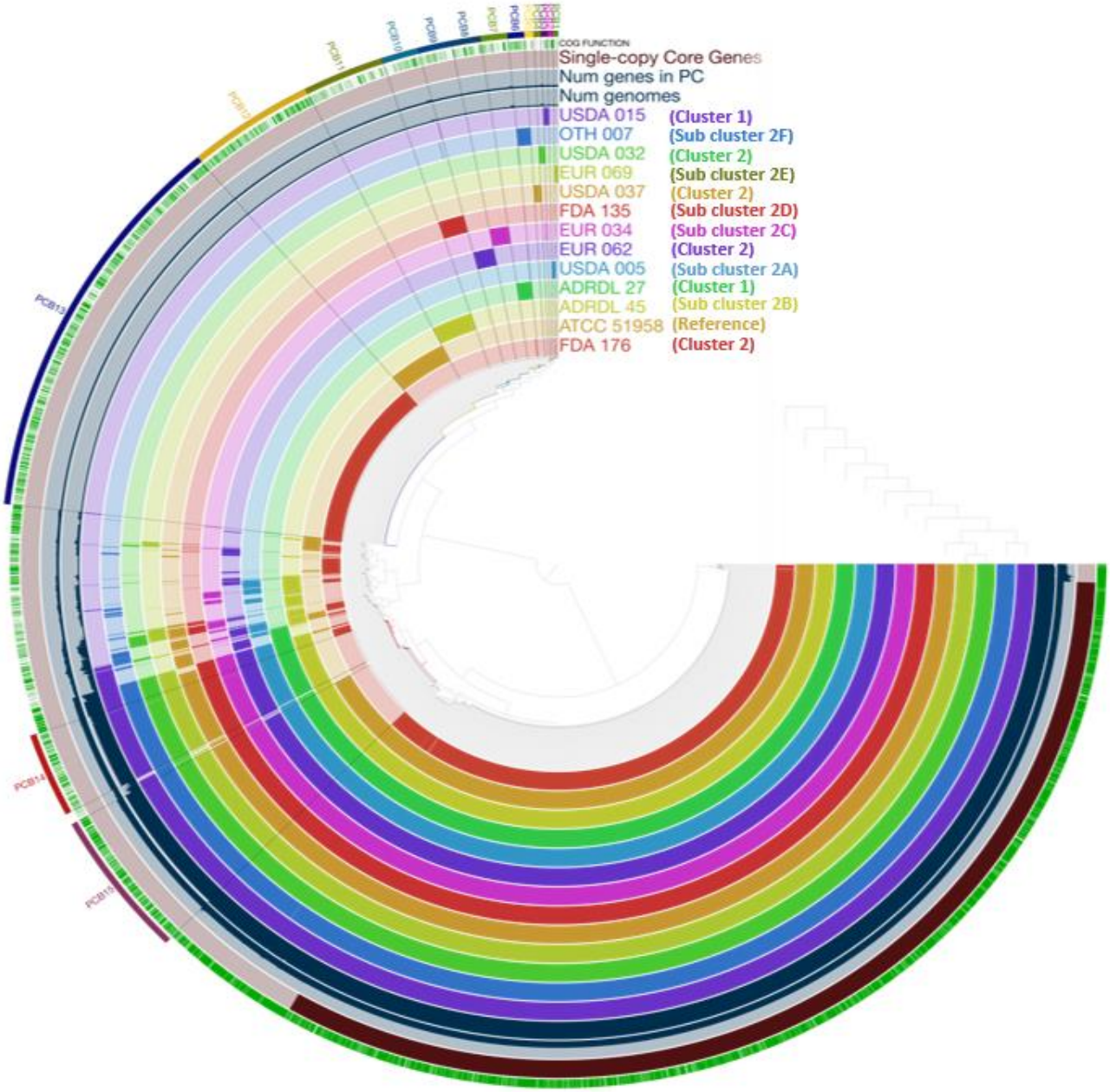
Twelve Selected isolate genomes representing different clusters and positions in SNP phylogenetic tree were selected for pangenome analysis along with ATCC 51958 reference strain genome. De novo assembled genomes in FASTA file format were analysed in Anvi'o (Meren lab,) a software for analysis and visualization of 'omics data [191]. Exclusive ortholog proteins were identified in all thirteen isolates, which were clustered into 13 PC Bins (Protein cluster bins – PCB) (Fig.10). Two other PC bins (PCB14 and PCB15) were created to analyze proteins that were present in all strains except either in isolate FDA176 alone or in both FDA176 and reference strain. A detailed information regarding exclusive COG (Cluster of orthologous groups) proteins identified in pan genome analysis in all the selected bins are given in table 8. Proteins without any known function are not mentioned in the table. A summary of number of exclusive COGs present in each cluster are shown in table 6.

Table 6. Summary of COGs exclusively present in representative isolate of each cluster

PC Bin number	Cluster/sub cluster name	Sample ID	Exclusive orthologs	Un-characterized	predicted	Phage elements	Other Known function	Unknown function
1	Sub cluster 2E	EUR069	13	1			1	11
2	Sub cluster 2A	USDA005	16				7	9
3	Cluster 1 (poultry)	USDA015	17	1			4	12
4	Cluster 2	USDA032	20					20
5	Cluster 2	USDA037	26	1			2	23
6	Sub cluster 2F	OTH007	44	3		1	15	25
7	Cluster 1(Bovine)	ADRDL27	64	1	1	1	12	49
8	Sub cluster 2C	EUR034	79	1	1	11	12	54
9	Cluster 2	EUR062	90		1	5	15	69
10	Sub cluster 2D	FDA135	106	1	2	2	27	74
11	Sub cluster 2B	ADRDL045	219	3	4	9	33	170
12	Reference	ATCC_51958	344	10	19	10	124	181
13	Cluster 2	FDA176	1117	37	32	41	404	603

PC bins from 1 to 13 were selected to pick proteins that were uniquely acquired by the representative isolates. In case of PC bin 4, 7 and 10 some of the proteins were also present in 2 or 3 other isolates, which were not included in finding exclusive orthologs for a selected bin. A higher degree of accessory gene presence as well as absence were found in FDA176. From PCB13 we could identified presence of additional 1117 proteins in this pond isolate from USA. In PC bin 14, there were some proteins that were present in all isolates except ATCC51958 and FDA176. Over all, this analysis revealed the higher degree acquisition of variable genes in the genome of salmonella isolate irrespective of geographical origin and isolation source which may attribute to the overall gene diversity within the serovar.

Fig. 10. Pan-genome analysis and visualization of thirteen *S*. Mbandaka isolates genome content. Outer most circle depicts selected protein cluster bins (PCB) based on the presence and absence of protein clusters. Inside this circle is a green circle that indicates COG (Cluster of orthologous groups) functions. Green color in this circle represents known function and white indicates unknown function. Brown circle Inner to green one shows single copy core genes. Next two circles show histogram of number of genes in protein cluster and number of genome. Remaining circles from outer to inner indicates genome of 13 *S*. Mbandaka isolate.



2.3.2 High-resolution phylogenetic analysis reveals genetic diversity and partitioning of *S*. Mbandaka isolates from similar isolation source

By the application of NGS method, this study presents a highly-resolved phylogeny of the *Salmonella* serovar Mbandaka isolates. SNP analysis of 465 *S*. Mbandaka isolates was validated by cgMLST to explore the genomic diversity within the serovar. Complete genome of *S*. Mbandaka str. ATCC 51958 (NCBI accession: NZ_CP019183.1) with a total length of 4.9 Mb was used as a reference sequence for the comparative analysis. Comparative analysis using WGS allowed us to identify two primary clusters and six sub clusters within the population structure of this serovar. Cluster 1 remained as an outlying group distinguished by 9 SNPs from cluster 2. Minimum number of nucleotide variations of cluster 1 isolates from the reference genome as well as homogeneity of the isolates with respect to the isolation source and geographical origin may shed light to the possibility of these isolates being closest, compared to all other isolates in this study, to the evolutionary origin of the *Salmonella* serovar Mbandaka. All isolates in cluster 1, twelve isolates from cattle and one from raw whole egg, were obtained from Texas state of USA. At the same time, 37 isolates from the same geographical location and isolation source (cattle) were clustered separately in cluster2 forming a sub cluster denoted as sub cluster 2B together with 16 other *S*. Mbandaka isolates. This represents the genetically diverse nature of *S*. Mbandaka isolates from the same host and geographical location.

Closely related 16 other isolates in sub cluster 2B varied in their isolation source and location. Three isolates (OTH021,022, and 023) were from Texas itself but from different sources (hamburger :2, ground beef:1) while remaining 13 isolates were from 7 different

states of USA except one isolate for which there was no record from which state it was collected. However, out of these 13, seven isolates were from cattle. Isolation source for remaining isolates were; four from ground beef, one from hog and one from pork chop.

Interesting observation is that out of 400 isolates, other than those from Texas USA, no strain from any other place showed any close relationship to cluster1 strains. Same thing applies to sub cluster 2B, where all the isolates were from North America and no isolates from any other countries clustered with cluster 2B isolates. This bring forth the assumption that in USA there is prevalence of two groups of *S.* Mbandaka strains, one containing geographically isolated and evolutionarily less diverse strains and second one accommodating more genetic diversity. Presence of a minimum number of unique SNPs provided by the comparative analysis between representative strains of these clusters may give substantial evidence for this assumption. Above all in our analysis we identified that number of unique SNPs that resides in two bovine isolates from Texas, one representing cluster 1 (ADRDL27: 14 SNPs) and the other representing sub cluster 2B (ADRDL 45: 82 SNPs), differed by five times. Further substantial evidence provided by MLST based MST that verified this phylogenetic separation of isolates from similar source and location. Sequence typing based on SNV analysis on 2502 targets showed a clear separation of cluster 1 isolates from sub cluster 2B irrespective of the isolation source and location.

Sub cluster 2B, which contains more bovine isolates is a sister clade of sub cluster 2A, one which contains more poultry associated isolates, without any overlap of isolation sources between clusters. That is, there was no bovine isolates in sub cluster 2A and no poultry associated isolates in sub cluster 2B. This is somewhat similar what we found in

cluster1 where poultry isolate and bovine isolates formed two separate lineages inside the cluster from their MCRA. Similar pattern exists here between sub cluster 2A and 2B as they also formed two separate lineages from their MCRA with a clear discrimination by accommodating 21 and 36 unique SNPs respectively. Interestingly we found that all the poultry related isolates in sub cluster 2A were obtained from USA. Based on this inference we suggest that in USA, cattle and poultry isolates of serovar Mbandaka undergone evolutionary changes maintaining their population structure. However, both theses clusters shared genetic similarity with porcine isolates. Sub cluster 2A isolates shared similarity also with equine, human, animal feed and environmental isolates indicating wide host adaptation of these strains. Poultry isolate USDA 015 in cluster 1 carried one unique SNP in a coding sequence of a type II secretion system protein GspE (SEEM1958_RS19665) in comparison with ADRDL027 bovine isolate in the same cluster. In sub cluster 2A isolate USDA005, we identified two unique SNPs (frame shift mutation) in CDS for phage tail proteins in comparison with cluster1 bovine isolate. However as mentioned earlier the number of unique SNPs difference between subcluster 2A and 2B was much more higher and most of them happened in phage related elements. This may suggest that changes in gene content happened in *S.* Mbandaka isolates as shown by pangenome analysis, but acquisition of genetic elements happened maintaining the specificity of the host especially in case of isolates from bovine and poultry, two major isolation sources of *S.* Mbandaka isolates in USA.

Sub cluster 2C and 2E shared their lineage with a small cluster of isolates from Tennessee state of USA i.e., sub cluster 2D. 2C and 2E sub clusters have a common feature, that they both contain human isolates from Europe and Food commodity isolates

from Asian countries. Majority of food commodity sources were plant commodities. Interestingly, isolates from tahini collected in 2013 by FDA and MDA, were clustered together in this sub cluster. Based on the collection year and collection center that matches with CDC investigation reports, we assume these isolates were the ones related to multistate outbreak caused by *S*. Mbandaka in United States in 2013. As per investigation report the outbreak was caused by *Salmonella* from contaminated tahini sesame paste imported from Turkey. If that is true, inferred from the close genetic relationship between isolates, our study reveals that the contaminated food commodities exported from Asian countries might be the source of human isolates in UK.

Isolates from Tennessee were collected from chicken breast, ground beef, and pork chop during the year 2008. These isolates were clustered together forming a sister clade to sub cluster 2C and 2E. These three clusters have a most common recent ancestor. The genomic evidence here in thus gives us clues that Tennessee isolates have the potential to adapt to different host environments. This well nested clustering also put forward the possibility of contamination meat products in a common place like a food processing facility. Tennessee isolates clustered together as clones from different isolation source. This cluster carried more than ten unique SNPs from other two nearby sub clusters 2C and 2E indicating these isolates are unique from other two clusters. There were 6 non-synonymous SNPs unique to this cluster from other two sub clusters, the important one is a SNP occured in a gene called *artM* that codes a protein required for the arginine transport across the inner membrane.

In general, a clear demarcation of clusters in the context of origin of isolates could not be seen in this study but clusters were selected based on the preference of more

isolates from similar isolation source to be grouped together. One exception is sub cluster 2F which was clearly defined by the geographical location of isolates. Forty three Mexican isolates clustered together in this cluster. Four more Mexican isolates contained in this study located some other positions in the phylogenetic tree were the exceptions.

2.3.3 Virulence gene mapping

De novo assembled 465 *S*. Mbandaka isolate genomes were analyzed for virulence determinants by BLAST search against virulence gene sequences available from virulence factor database [192, 193]. Virulence factors identification was performed in CLC Genomics workbench (Version 9.5.3, Qiagen) based on the parameters of minimum sequence identity of 90% and minimum sequence length of 50%.

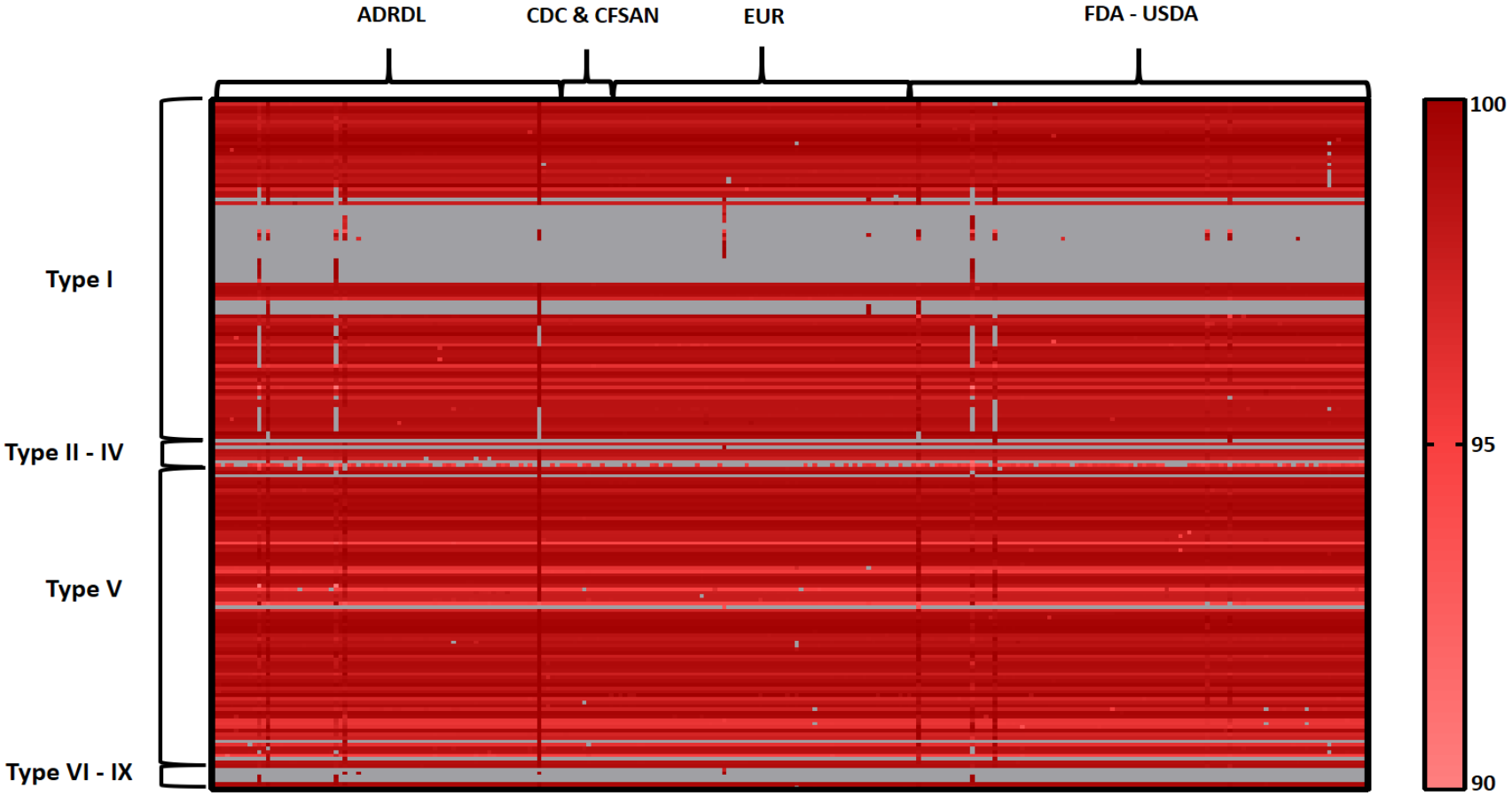
Overall, 195 virulence genes grouped into 9 categories were identified in our *S*. Mbandaka study strains. Results were shown as heat map in Fig.11. Fimbrial adherence determinanats, non-fimbrial adherence determinants, Macrophage inducible genes, Magnesium uptake genes, Genes involved with secretion system, serum resistance genes, stress protein genes, Toxin factors, and two component system genes were the 9 categories under which all 196 virulence factors available from VFDB were grouped for better understanding. 49% of the identified virulence factors belong to the group Fimbrial Adherence Determinant (FAD). FADs play important roles in host cell adherence and invasion. 42.6% genes were secretory system related genes. Nine genes (4.5%) were plasmid encoded genes.

Out of 196 genes represented by database 63 genes were present in all 465 isolates. Presence of 93 genes varied between 97.4% to 99.8% of isolates. One gene was not identified in any isolate while remaining 39 genes were highly varied in their presence.

Adherence determinants:

Ninety-seven fimbrial adherence determinants and 4 non- fimbrial adherence determinants were identified in total. Twenty fimbrial adherence genes were present in all 465 isolates. These twenty fimbrial genes include 6 bcf, 4 csg, 3fim, 5stb and 2 std genes. More than 99.5% isolates carried

Fig. 11. Heat map showing identified virulence factors in *S*. Mbandaka serovar isolates. *De novo* assembled sequences of 465 *S*.Mbandaka isolates under this study were analyzed for known virulence factors by performing a BLAST search against 196 virulence gene sequences available from Virulence factor database (VFDB) using CLC Genomics workbench. Minimum identity of 90% and minimum sequence length of 50 % were the criteria used for the identification of each virulence factor. For the better understanding, Virulence factors were categorized in to nine different types depicted as Type I to Type IX arranged in a serial order from top to bottom on Y axis. All 465 samples were arranged in alphabetical order from left to right on X axis. Type I – Fimbrial adherence determinants, Type II – Macrophage inducible gene, Type III – Magnsium uptake, Type IV – Non fimbrial adherence determinants, Type V – Secretion system, Type VI – Serum resistance, Type VII – stress protein, Type VIII – Toxin, Type IX – Two component system.



virulence factor Fim encoded by 9 genes (*fimA,C,D,F,H,L,W,Y,*and *Z*). Other virulence factors and genes that were present in more than 97.4% of isolates include Factors Stj, Ste, Stf, Sth, Sti, Stk and genes *bcfG, csgD,F,G, lpfA,B,C,E,stdA,tcfB* and *tcfC*.

Unique presence of 8 fimbrial genes (*pefB, pefC, pefD, sefC, spvA,spvB,spvC and spvR*) were found in a human isolate (EUR033) from United Kingdom. Isolate EUR033 carried maximum number 83 out of 97 fimbrial adherence genes. Minimum number of FADs were identified in two USA isolates FDA 176 and FDA 177 isolated from pond samples.

Virulance factor *sta* was present in only 5 isolates. All these five contained 7 sta (*staA-staG*) determinants in their genome. Similarly, virulence factor stc was identified in 8 isolates. Genes *stcA,B C* and *D* were present in 6 isolates. One human isolate EUR066 carried *stcB, C,*and *D*. Only *stcB* was present in MDH010 (Porcine isolate from USA). In case of another virulence factor Tcf only three isolates (FDA019, FDA071, OTH026) were identified with *tcfD,* but they were also present with *tcfB* and *tcfC*. Gene *tcfB* and *C* were present in a total of 456 samples. All these three determinants were absent in 7 isolates.

Virulence factors associated with Secretory system

48.8% of identified genes (41 out of 84) in this category were found present in all isolates. 39 out of 84 genes were found present in ≥ 98% of isolates analyzed suggesting that most of the genes in this category are highly conserved in *S.* Mbandaka isolates. Only four genes (*sspH2, sopE, srfH,* and *gog B*) encoding for Type three secretion effector proteins showed major variation in their presence. They were found only in < 10

isolates with gene *gogB* as the least one identified only in two isolates (ADRDL73 and FDA013). The virulence genes in this category are mainly from SPI1 or SPI-2. There were 30 SPI-1 genes encoding Type III secretion system proteins that includes *hil* and *inv* genes along with ten T3SS-1 effector genes. From SPI-2 there were 28 genes for T3SS and 15 for T3SS-2 effectors. Gene *slrP* encoding effectors that are translocated via both system were also identified.

Other virulence genes identified in *S*. Mbandaka serovar:

There were two macrophage inducible genes *mig14* and *mig5* identified in our study isolates. Virulence gene Mig14 that encodes for transcriptional regulator was identified in 463 *S*. Mbandaka isolates. Mig5 is a plasmid encoded gene was uniquely identified in human isolate EUR033.

Two Magnesium uptake genes *mgtB* and *mgtC* were identified in 464 and 4654 isolate respectively. *mgtB* was not identified in FDA168 for the set criteria of ≥ 90% of sequence identity. Plasmid encoded serum resistance gene *rck* was unique to EUR033, human isolate from UK. Stress protein virulence factor was present in 5 isolates ADRDL30, 33, 73, EUR033, and NYSDH005. Gene *sodCl* encodes for a superoxide dismutase (Cu-Zn) precursor was identified in these 5 isolates.

Two genes encoding for typhoid toxin subunits were also identified in *S*. Mbandaka isolates. Eight isolates (ADRDL11, 28, FDA013, 176, 177, OTH001, 10, 26) were identified with virulence gene *cdtB* that has limited amino acid sequence similarity with DNAse 1 family of proteins [195] and encodes a subunit of typhoid toxin. All but three

isolates FDA176, 177 and OTH001 in this group, were identified with another typhoid toxin subunit gene *pltA*.

Two *PhoPQ* virulence factors were found universally in all isolates except one EUR050 in which one gene *phoQ* was not identified.

2.3.4 Plasmid encoded as well as typhoid associated virulence factors were identified in *S*. Mbandaka isolates

A set of Five Fimbrial operons (*bcf, csg, fim, stb, and std*) were identified as common virulent factors present in our study set of 465 *S*. Mbandaka isolates irrespective of isolation source and geographical location. Some of these operons (*Bcf, stb, std* and *stj* fimbrial operons) have been found to contribute intestinal persistence and long term carriage of *S*. enterica serotype Typhimurium in genetically resistant mice [196, 197]. Curli fibers, thin aggregative fibers were found to be present and expressed in many enteropathogens including *Salmonella* and *E*. coli. *Two* operons, *csgBA (C)* and *csgDEFG,* involved with curli biogenesis [198] have been identified in *S*. Typhimurium[199]. All seven ORFs in these operons have been identified in >97% of *S*. Mbandaka isolates.

Presence of virulence plasmid, not in all isolates, has been reported in several *Salmonella* serovars such as serovar Abortusovis, Choleraesuis, Dublin, Enteritidis, Gallinarum/Pullorum, Paratyphi C, Sendai and Typhimurium [200-205]. All virulence plasmids contain *Salmonella* plasmid virulence locus, highly conserved 8-kb region. Presence of virulence plasmid with this spv operon, consisting of regulatory *spvR* locus

and structural *spvABCD* genes, was reported to sufficient to enable systemic infection in animal models [206, 207]. Out of five plasmid virulence genes spv R is a transcriptional activator that encodes for a positive regulatory protein required for the expression of *spvABCD* [206, 208]. Out of five genes in *spv* locus, four of them, *spvRABC,* were found unique to a human isolate from UK (EUR033). But *spvD* was not identified in any isolates of *S.* Mbandaka in our study dataset. Six more plasmid encoded genes (*pefB, pefC, pefD, sefC,* (FADs), *mig-5* (MIG) and *rck* (serum resistance gene)) were also found unique to this UK isolate obtained from human. According to Feng et al, *Salmonella* serovars acquire virulence plasmids mainly through vertical transmission, although exceptions are there as in case of Enteritidis where the acquisition was found via horizontal transfer [205]. If vertical transmission was the way of plasmid acquisition in this isolate, comparing virulence plasmid between this and other serovars may give insights to the source of this virulence plasmid acquisition and much more evolutionary information about this serovar. Since presence or absence of plasmid plays some role in host adaptation a comparison between isolates in this aspect may reveal the original causes of outbreaks as well.

Baumler et al, described scattered phylogenetic distribution of certain fimbrial operons, either due to evolutionary loss as in *lpf* (long polar fimbriae) operon, or by horizontal gene transfer as in *sef* (*S.* enteritidis fimbriae) and *pef* (plasmid encode fimbriae) operons, and their limited presence in small number of serovars [209, 210] . They also suggested with several examples that presence or absence of these fimbriae operons might have played role in host adaptation of certain *Salmonella* serovars. Our result showed that serovar Mbandaka carry lpf operon with the presence of *lpf A,B,C,*and

E genes in more than 96% isolates and *lpf D* gene in around 2.3% isolates. Interestingly one human isolate EUR033 showed unique presence of three *pef* genes (*pef B, C* and *D*) and one *sef* gene (*sef C*) suggesting the horizontal acquisition of fimbrial operons in *S.* Mbandaka isolates.

Presence of typhoid associated virulence factor genes in *S*.Mbandaka:

Certain virulence factors (*cdtB, pltA* and *pltB*) which were originally characterized in *S.* Typhi have already been identified in isolates of NTS serovars Montevideo, Schwarzengrund, Bredeney and 9,12:l,v:- by Comparative Genomic Hybridization (CGH) technique [211]. Cytolethal distending toxin subunit B (CdtB) and Putative pertussis-like toxin subunit (PltA) virulence factor genes (*cdtB* and *pltA*) were identified in eight and five isolates respectively in our study isolates. It has been shown that intracellular *S.* Typhi express a protein homologous to active enzymatic subunit of Cytolethal Distending Toxin (CDT) CdtB that cause cell cycle arrest, host cell distension and nuclear enlargement [195]. *cdtB* and *pltA* are encoded within SPI-11 together with *pltB*. These three genes were shown to produce a tripartite exotoxin intracellularly by *S.* Typhi, which was then transported extracellularly via vesicular mechanism [211, 212]. All eight isolates (ADRDL11, 28, FDA013, 176, 177, OTH001, 10, 26) with *cdtB* gene were isolated mainly from environmental (chicken box paper, pond, soil, and spice mix) samples except two isolates OTH001 (papaya) and ADRDL 28 (Bovine). This suggests the prevalence of Typhoid toxin producing *S.* Mbandaka isolates in environment and the transmission through food chain to hosts.

Typhi colonizing factor (*tcf*) operon was described as a putative fimbrial operon encoded in SPI-6, with four reading frames (*tcf*A, B, C, and D), that plays important role

in host specificity of *S*. typhi[209, 211]. This operon was reported to found in Salmonella serovars, Choleraesuis, Schwarzengrund and Heidelberg, Virchow and Montevideo in addition to Typhi and Paratyphi [213-215]. ORFs *tcfB* and *C* were identified in 98% of our isolates under study. In our result *tcfD* was found only in three USA isolates FDA019, FDA071, OTH026. First two isolates were collected from animal feed while third one from soil indicate the prevalence of *S*. Mbandaka strains in environment, which has the gene repertoire to cause serious Salmonella infection.

2.3.5 Antimicrobial Resistance pattern

All 465 isolates in this study were subjected to whole genome short gun sequencing. Using CLC genomics workbench (version 9.5.3 Qiagen) *de novo* Assembled sequences were used to identify resistance genes by BLAST search against 2156 resistance genes data set available from ResFinder database (Center for Genomic Epidemiology). Minimum percentage sequence identity of ≥ 85% and sequence length identity of ≥ 50% were the criteria used to identify resistance determinants. A total of 376 (17.4%) resistance genes were identified in 125 genomes (26.9% of total genomes analyzed). We identified genes that confer resistance to 9 different class of antimicrobial agents. Out of newly sequenced 76 isolates only 11 isolates were identified with resistance genes. Fig.12 shows heat map of identified genes against isolates showed at least one resistant gene.

Most common gene identified was for tetracycline resistance (*tetB*) followed by streptomycin resistance genes (*strA and strB*). More than hundred determinants were identified in 5 isolates (FDA003, ADRDL7, MDH004, EUR059 and EUR044) with one Isolate (FDA 003) (turkey isolate collected from ground cumin) contained highest number, 246 determinants, in our strain set. Genes that confer resistance to all 9 class of

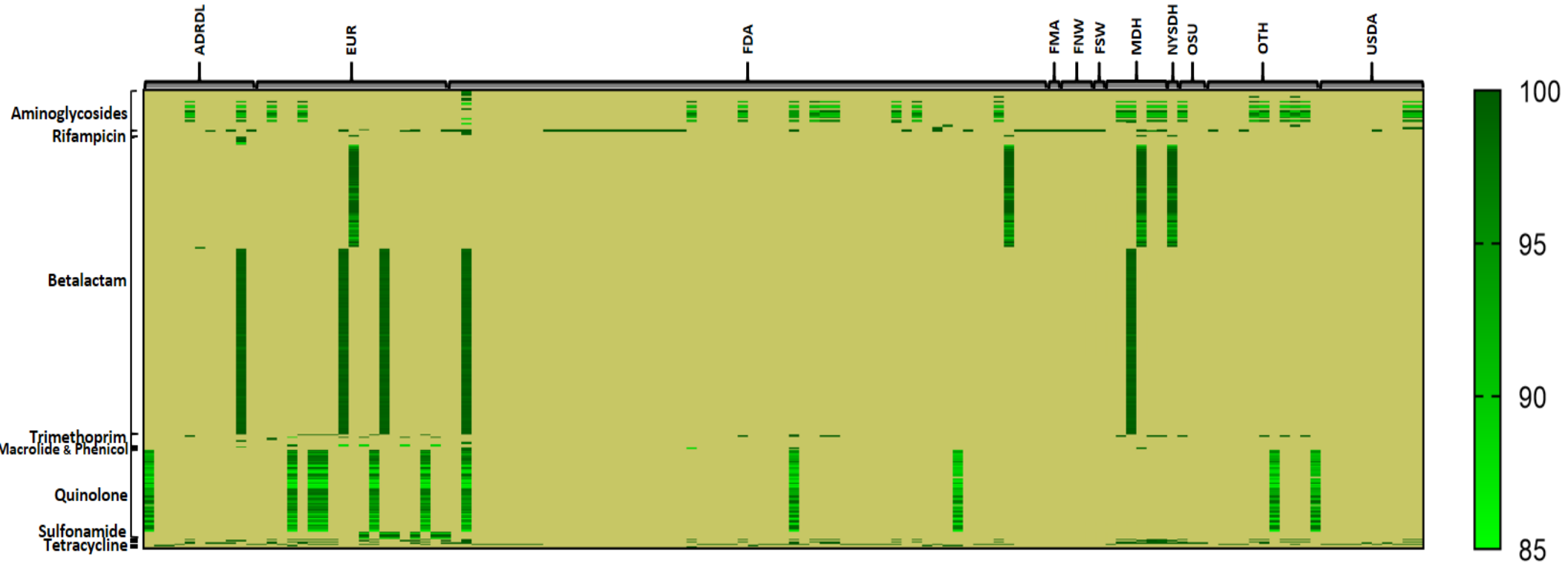
antimicrobial agents were identified in isolate FDA003. Forty-three isolates contained only one AMR determinants, the most common was *tetB* gene. No AMR determinants were identified in cluster 1, and subcluster 2C isolates but were identified in 36, 11, 10 and 28 isolates of Sub cluster2A, 2B, 2D and 2E. There were 5 resistant strains from Texas (Bovine-4, ground beef-1) in sub cluster 2B. Out of 5, one bovine isolate (ADRDL 45) carried aminoglycoside (*strA* and *strB*), sulphonamide (*sul2*) and tetracycline (*tetA*)AMR genes. One isolate (OTH021) from ground beef carried 67 quinolone resistance genes. All 5 bovine isolates from Florida (ADRDL10, 12, FDA 154, 156 and 157) carried genes for tetracycline resistance (*tetC*). Along with this gene, FDA156 contained two additional aminoglycoside genes (*aph (3')-Ia* and *aph(3')-Ic*) and FDA 157 carried 65 quinolone resistance genes.

Out of 60 human isolates in our study, 58 were from United Kingdom, one from Taiwan (EUR005) and one from USA (NYSDH005) Taiwan isolate (EUR005) identified with the presence of only one tetracycline resistant *tetB* gene. Isolate from USA (NYSDH005) was found to carry tetracycline (*tetA*), Sulfonamide (*sul2*) and beta lactam (*bla*CFE1, *bla* LAT-1, *bla*CMY and *bla*BIL) resistance genes. Fifteen UK isolates showed resistance determinants in their genome that include resistance against six group of antimicrobials but not for rifampicin, macrolides and phenicols. Eleven were identified either with *qnrB* or *qnrS* quinolone resistance genes. No human isolates with both *qnrB* and *qnrS* were found in our sample set.

In sub cluster 2E, 26 out of 28 resistant strains genome carried resistance genes *strA* and *strB* only. All those 26 strains were from papaya in Mexico. AMR determinants for aminoglycoside (13 genes), phenicol (1 gene), sulfonamide (2) and tetracycline (1) were

identified in one resistant strain (FDA047) from serrano pepper in Mexico. Another isolate in this cluster collected from chicken drag swab in USA:MI contained 65 genes for quinolone resistance.

Fig. 12. Heat map showing presence of antimicrobial resistance genes in *S*. Mbandaka serovar isolates. For the identification of AMR genes, de novo assembled genome sequences of 465 *S*. Mbandaka isolates were BLAST searched against 2156 AMR gene sequences available from ResFinder database. Minimum sequence identity of 85% and sequence length identity of 50% were parameters used for the identification of AMR gene presence. A total of 376 resistance genes belong to 8 classes of antimicrobial agents were identified in overall 125 isolates. Legend: heat color based on percentage identity of each gene ranging from 85 to 100.



Tetracycline resistance:

Genes *tetA, tetB, tetC* and *tetG* were the only four tetracycline genes identified out of 104 total genes available in the ResFinder reference data set. These genes were involved with efflux pump mechanism that confers drug resistance to bacteria [216]. A total of 82 isolates out of 125 resistant strains in our strain set showed tetracycline resistance determinants in their genome. The *tetB* gene was the commonest in our strain set identified in 50 isolates (60.9%). Prevalence of *tetA,* most frequently occurring tetracycline gene in Gram-negative bacteria [216], was 30.5% occurring in 25 isolates. Genes *tetC* and *tetG* were identified in 7 and 1 isolates respectively. All except one (FDA 079) contained no more than one *tet* genes. Both *tetA* and *tetB* were identified in FDA 079 along with 89 other AMR determinants. Gene *tetB* was the common gene identified in most of the isolates with only a single AMR determinant. One human isolate from Taiwan (EUR005) belongs to this category which carries only one determinant in its genome ie, *tetB*.

Aminoglycoside resistance:

In total 34 resistance genes were detected in 64 resistant strains. Genes *strA* and *strB* were the most common resistance genes present in >60% of resistant strains followed by genes *aadA1, aadA2, aadA3, aadA12, aadA13, aadA15, aadA17, aadA21, aadA22, aadA23, aadA24, aadA8,* and *aadA8b* (37.9%). Thirteen resistant genes (*ant* (3")-*Ih- aac* (6')-*IId, aadA6, aadA10, aadA11, aadA16, aac (3)-Ib-aac (6')-Ib, aac (3)-IIa, aac (3)-IIc, aac (3)-IId, aac (3)-IIe, aac (6')-Ib, aac (6') Ib-cr*,and *aacA4*) which form 38.2% of total aminoglycoside resistance genes identified were unique to Turkey isolate (FDA003) obtained from ground cumin.

Beta lactam resistance:

Beta lactam resistant genes *blaCARB1, blaCARB2, blaCARB3, blaCARB4, blaCARB6, blaCARB11* and *blaCARB12* were unique to a bovine isolate from Oklahoma, United states (ADRDL73) collected in the year 1988. Another bovine isolate from Texas (ADRDL31) contained only one resistant gene that is a beta lactam resistant determinant blaSED1. This gene was identified as unique to ADRDL 31. 84 *blaCMY* genes were identified only in 4 isolates (EUR46, FDA166, MDH005, NYSDH005). 153 *blaTEM* genes were identified in only 5 isolates (ADRDL73, EUR044, EUR059, FDA003, MDH004).

Quinolone resistance:

73 quinolone resistant determinants were identified in our study strain set. Eight isolates were identified with 67 genes (EUR 011, 025, 030, 056, 066, FDA 003, 079, OTH021), one with 66 genes (ADRDL 07) followed by two with 65 genes (FDA157 and OTH 031). All of them were *qnrB* genes. Isolation sources of these above isolates were distributed widely including human, bovine canine and environmental isolates. Six genes (*QnrS1 – S6*) were uniquely present in 6 human isolates (EUR 54, 59, 61, 65, 71 and 74) from United Kingdom. *qnr* genes are plasmid mediated quinolone resistance (PMQR) determinants that have been identified in several enterobacterial organisms including Salmonella [217]. *Aac (6') lb-cr* determinant unique to isolate FDA 003 was the only fluoroquinolone resistant gene identified in our study. Aminoglycoside acetyltransferase *aac (6') lb-cr* determinant acetylates several fluoroquinolones and offer resistance to agents such as norfloxacin and ciprofloxacin [218]. Eleven human isolates from UK were

identified with quinolone resistant genes. Five of them carried 67 *Qnr*B genes (EUR 011, 025, 030, 056, 066) and six of them carried 6 *Qnr*S genes.

Sulfonamide and Trimethoprim resistance:

Three sulfonamide resistance determinants, *sul1, sul2* and *sul3* were identified in a total of 39 isolates. *Sul2* alone was present in 8 isolates (EUR 054,62, 65, 74, MDH 005, NYSDH 005, ADRDL 45, and ADRDL 73) from Human, Bovine and chicken sources, while three isolates FDA003, MDH 008, MDH 010 from ground cumin and porcine contained all three determinants. *Sul1* and *sul3* were identified together in 28 isolates from widely different isolation sources. But no isolates were identified with either *sul1* or *sul3* alone. *Dfr12* was the common determinant identified in 12 of our study strains. 11 isolates out of 12 contained only this allele. One isolate (FDA 0079) contained both *dfrA12* and *dfrA1*. Including this one there were total 7 isolates identified with dfrA1 out of which remaining 6 of them contained only *dfrA1* trimethoprim resistant allele. Determinants *dfrA16* (ADRDL 73), *dfrA15* and *dfrA15b* (EUR 004) *dfrA27* and *dfrA28* (FDA 003) were identified in one isolate each. *dfrA14, dfrA30* and *dfrA5,* all three occurred in 5 isolates in which no other trimethoprim resistant genes were identified. Four of them were human isolates while one (EUR044) was food isolate.

Macrolide, phenicol and rifampicin resistance:

Macrolide resistant determinant *ereA* was identified in bovine isolate from Oklahoma and *mphA* was identified in FDA003, Turkey isolate from ground cumin. Phenicol resistant determinant gene *floR* was identified in 4 isolates. Three rifampicin resistant genes were identified in this study which were all present in FDA 003.

2.3.6 Resistance to multiple antimicrobial agents was identified in S. Mbandaka isolates

Potential of WGS to predict antimicrobial susceptibility of bacteria has already been described by several research studies [182, 219, 220]. Here, we report antimicrobial resistance gene pattern in 465 *S.* Mbandaka isolates by analyzing publicly available sequence data using WGS. Overall, 376 genes that confer resistance to 9 classes of antimicrobial agents were identified from 125 *S.* Mbandaka isolates. 65.6% of isolates showed resistance to tetracycline followed by resistance to aminoglycosides (52.8% isolates). Our analysis revealed resistance of *S.* Mbandaka isolates to more than three classes of anti-microbial agents. Fifteen isolates showed presence of resistance genes against more than three classes of antimicrobial agents. Twenty-two isolates showed resistance against more than three classes of antimicrobial agents. One isolate from Turkey carried 246 resistant determinants involved with resistance to all 9 classes of antimicrobial agents. This MDR isolate obtained from ground cumin was the only one in our study strains that harbored 3 rifampicin resistant determinants as well as one macrolide resistant gene *mph* (A).

Extended Spectrum Beta Lactamase (ESBLs) CTX-2 has been identified in a *S.* Mbandaka strain, isolated from a Algerian infant, with reduced susceptibility to cefotaxime [221]. In another study *bla SHV-12*, a ESBL gene that confers resistance to Extended spectrum cephalosporins has been identified in a *S.* Mbandaka isolate [222]. But both these were not identified in our study strains. However, *blaTEM* genes, that have been reported to code for ESBLs [223], were identified in five isolates

(ADRDL73,EUR44, 59, FDA003, and MDH004). A total of 153 *blaTEM* genes were found uniquely present in these five isolates.

Ceftriaxone resistance is a serious concern because of its use in the treatment of salmonellosis in children. In Salmonella, resistance to ceftriaxone is mainly due to AmpC β-lactamase (*bla*$_{\text{CMY-2}}$) gene[224, 225]. There were four *S*. Mbandaka isolates (EUR046, FDA166, MDH005, and NYSDH005) in our study harbored *bla CMY-2* genes. EUR046 was a human isolate while others were from poultry product (FDA166 and NYSDH005) and cattle (MDH005). This explains the prevalence of ceftriaxone resistant *S*. Mbandaka isolates in human as well as in food animals or animal products. These four isolates harbored a total of 82 *blaCMY* genes which were found unique to these four in our study. CMY- β-lactamases are cephalosporinases that can hydrolyze all β-lactams except carbapenems [223].

Majority of human isolates in our study were from UK. Out of total 60 human isolates 58 were from UK. Of these 11 isolates carried quinolone resistant genes, five isolates with *qnrB* and six with *qnrS*. These are PMQR genes that confer resistance by protecting DNA gyrase and topoisomerase IV from quinolone inhibition [226]. First report of PMQR was in 1998 in *Klebsiella pneumoniae* isolate from USA [227]. Qnr proteins, aminoglycoside acetyl transferase AAC (6')-lb-cr and the efflux pump protein QepA were the different kinds of identified PMQR mechanisms [218, 228, 229]. Epidemiology of three types of *qnr* determinants has been very well explained in a review article by Cattoir et al, in 2009 [218]. Quinolones are widely used antimicrobial agents in both human and veterinary clinical practices. Development of quinolone resistance in Gram- negative and Gram-positive bacteria limits the use of quinolones for

various clinical applications [230]. Indiscriminate use of antibiotics in food animals is considered to be one of the major reasons for development of quinolone resistant strains of bacteria that may gain access to human through food chain. There were six strains from other isolation source also showed quinolone resistance in our study. These isolates were from chicken drag swab (1), cattle (1), beef products (2), ground cumin (1), and canine feces (1).

2.4 Conclusions

This study explores the genetic and evolutionary diversity of *Salmonella enterica* serovar Mbandaka iterating the power of massively parallel sequencing technologies to assess the diversity within a serovar at single base resolution. For the first time, we have studied the phylogenetic structure of *S.* Mbandaka incorporating sequence data of massive number isolates from different parts of the world. We also aimed to define the virulence and antimicrobial resistance gene repertoire of *S.* Mbandaka to reveal the capability of this serovar as a potential threat to public health. Our data show a split in the cluster formation of *S.* Mbandaka isolates from similar isolation source and origin.

Although acquisition of accessory genes may result in the genetic diversity within this serovar, this acquired diversity does not appears to alter host adaptiveness at least in poultry and bovine isolates which needs to be further substantiated by analyzing more number of host specific isolates. Moreover, we could show that irrespective of isolation source most of the isolates in our study carried a similar virulence repertoire. Additionally, existence of isolates with resistance to more than three classes of antimicrobial agents was also revealed by this analysis. To the best of our knowledge

this study forms the first high resolution genetic analysis of *S*. Mbandaka which may support future outbreak investigations and researches on this serovar.

Chapter 3: Analysis of host cell invasion and acid resistance phenotypic characters of *S.* Mbandaka

3.1 Introduction

Food borne infections caused by *Salmonellae* are an important worldwide concern [53, 231]. As per the latest reports based on CDC *Salmonella* outbreak analysis data, the overall incidence of *Salmonella* outbreak remains same for the last few years in USA but incidence scenario due to different serovars has changed [232]. Some of the serovars which were not a concern previously, started to became a concerned source of recently occured food borne outbreaks. *S.* Mbandaka is one of such *Salmonella* serovar which caused multistate human *Salmonella* outbreaks in USA in 2013 and 2016 [162, 163]. Cattle, chicken and pigs are the major host of *S.* Mbandaka in USA, and many of these act as asymptomatic carriers [75]. *S.* Mbandaka has also been isolated from a variety of sources including animal feed, plant and meat food products, as well as from environmental samples.

Comparative genomic analysis of global collection of 465 *S.* Mbandaka sequence data elucidate the population structure, intra serovar distribution of virulence factor and antimicrobial resistance genes of this serovar (chapter2). Since the genotype may not always reflect the phenotypic behavior of the bacteria, analysis of phenotypic properties is well demanded to understand functional relationship between strains from different isolation source especially clinical and non-clinical strains. For this purpose, seventy-six *S.* Mbandaka strains collected from different places of USA has been used for cell invasion assay, acid tolerance assay as well as for antimicrobial resistance study. Data

comparison between strains of bacteria was carried out to see whether any intra serovar differences exist between strains based on their source of isolation, geographical area of collection, clinical and non-clinical properties.

3.2 Materials

- LB (Luria Broth/Lysogeny broth/Luria bertani) Broth and LB plates

  For one liter of broth, 25g LB broth powder (LB Broth, Fisher Bio reagent; BP1426-2) was added to 1000ml of Milli Q water. When completely dissolved, autoclaved the media at 121^{o}C, at 15 psi pressure for 20 minutes. For LB plates, after mixing LB broth powder 1.5% bacteriological agar (Agar granulated, Fisher Bioreagent; BP9744-5) was added. Agar containing media was then autoclaved, cooled down 40-50^{o} C and then immediately poured out into petridishes.

- pH adjusted LB broth

  Normal LB broth had a pH ranges from 6.8 to 7.0. Broth with lower pH was prepared by using 12 N Hydrochloric Acid (HCL). LB broth with pH of 4.0, 5.0, 6.0, and 7.0 were prepared.

- Cell culture media

  Complete media: Dulbeccos Modified Eagle Medium (DMEM(1X) + Glutamax-I, Gibco; 10569-010) supplemented with 10% (v/v) FBS (Fetal Bovine Serum) and 1% Antibiotic and Antimycotic solution

  Plain media: contains no supplements

- Antibiotic (Gentamicin) containing Cell culture media

Added Antibiotic Gentamicin sulfate (Gentamicin sulfate, Across organics; 455310050) at a concentration of 100µg/Liter to Plain DMEM + Glutamax media. Filter sterilized the media with 0.2µ filter and stored at 4ºC.

- 1% Triton X

  Prepared 1% Triton X in Phosphate Buffered Saline (PBS) from 100x. To make 100ml of 1% Triton X, 1ml of 100x Triton X was mixed to 99ml of PBS (1x). After Vortexing well, the mixture was kept at room temperature for overnight. Filter sterilizing was carried out using 0.2µ filter before use.

- 0.125% Trypsin EDTA (1X): Gibco by life technologies, Cat# 25200-056
- Fetal bovine serum (heat inactivated), premium: Atlanta biologicals, Cat # S11150H
- T75 tissue culture flask: Fisherbrand, Cat # 353136

3.3 Methods

3.3.1 Cell invasion Assay

3.3.1.1 Bacterial culture and preparation

All 76 *S.* Mbandaka strains under study were cultured in 3ml LB broth. 3ml of LB broth without bacterial inoculation was used as a control to rule out the media contamination. After overnight incubation at 37ºC, OD (Optical Density) at 600nm was measured keeping the normal uninoculated LB broth as blank. An OD_{600} of 1 was considered as 10^9 Colony Forming Units (CFU)/ml of culture. Based on the number of epithelial cells per well in tissue culture plate, we calculated number of bacteria required to infect cells with a MOI (Multiplicity of Infection) of 100, that is 100 bacteria per cell. Once the number of bacteria required was determined, we found out the volume of each

bacterial strain culture based on OD_{600} value to get the number of bacteria to infect the epithelial cells. Respective volume of each bacterial culture was then aliquoted into 1.5ml Eppendorff's tube (Autoclaved). Duplicate tubes were prepared for each bacterial strain. Spun down the bacteria at 8000 rpm for 5 minutes and discarded the supernatant media. Re-suspended the bacterial pellet in 500µl sterile PBS (1X) by vortexing and then pelleted again the bacteria at 8000 rpm for 5 minutes. Repeated the PBS wash one more time and then discarded the supernatant after the final spin. Two times washed (with PBS) bacterial pellet was then re-suspended in 600µl of plain DMEM cell culture media.

3.3.1.2 Caco2 cell culture and preparation

Human Colorectal Adenocarcinoma cells (Caco2 cells) were obtained from ATCC. Cells culture Passage number ranged from 53 to 70 during the entire experiment period. Cells were grown in DMEM medium containing Glutamine (DMEM(1x) + Glutamine; Gibco) supplemented with 10% (v/v) FBS and 1% Antibioic and antimycotic solution at 37^{o}C in 5% CO_2 (v/v) in T75 Tissue culture flask. Once the cell monolayer was 70 - 80% confluent, cells were trypsinized with 0.125% Trypsin EDTA for 5-10 minutes. Trypsinization was stopped by adding FBS supplemented DMEM media. The cell suspension was centrifuged at 1200 rpm for 5 minutes and the supernatant medium was discarded to wash off the trypsin EDTA. The cells were re-suspended in complete DMEM + Glutamax media and uniform cell suspension was made by gentle pipetting.

Cell count per ml of cell suspension was determined using Neubours hemocyto-meter. Volume of cell suspension required for a seeding density of $0.3x\ 10^5$ per well of a 24 well plate was determined. A total volume of 1.5 ml media per well was used for cell culture in 24 well plate. Plate was then incubated at 37^{o}C in 5% CO_2 for 48 hours. After

incubation trypsinized cells in two wells of the plate and determined average number of cells per well based on which number of bacteria required for each well was calculated for a MOI of 100.

3.3.1.3 Cell invasion assay protocol

Gentamicin protection assay as described by Lee et al. was used with slight modification [233]. Cell monolayers in the 24 well plate were washed two times with sterile 1x PBS. Inoculated the Caco2 cells in each well with 600µl of bacterial suspension already prepared. In each experiment two wells for each bacterial strain were used. These wells were inoculated with duplicate tubes of each bacterial suspension. Incubated the plate at 37^{o}C in 5% CO_2 for 2 hours. After incubation, we removed the media containing bacteria from each well and then washed one time with 1ml of sterile PBS (1x). Cells were then treated with 400µl of Gentamicin containing (100 µg/ml) DMEM media. Gentamicin treatment kills extracellular non-invading bacteria while intracellular bacteria remain viable [233] . Incubated the plate next 1 hour at 37^{o}C in 5% CO_2. Washed the cells two times with sterile PBS (1x) followed by lysis of the cells with PBS containing 1% Triton X-100 (Sigma) for 10 minutes to release intracellular bacteria. 100 µl of lysed cell suspension from each well was then serially diluted using sterile PBS (1x). Dilutions of 10^{-1} to 10^{-5} were prepared, 100 µl of which were then used for spread plating on LB agar plates. Plate count was then taken after overnight incubation. Average CFU/ml was then calculated for each bacterial strain. The experiment was done two times with duplicate wells for each bacterial strain.

3.3.2 Acid Tolerance Assay

Bacterial strains were cultured overnight as mentioned before. Two hundred and forty microliter culture was then aliquoted into a non-tissue culture treated flat bottom 96 well plate. OD_{600} was measured using an ELISA plate reader. Plain LB broth was used as blank. Subtracted the blank reading from the sample OD_{600} reading to get the actual OD_{600} of the bacterial culture. Calculated the volume of LB broth needed to add in each sample to adjust the OD_{600} to 0.4. Set up flat bottom, non-tissue culture treated 96 well plates with LB broth, the pH of which was adjusted to 4.0. Triplicate wells for each bacterial sample were arranged with 180µl of sterile LB broth (pH=4.0). Aliquoted 20µl from the OD adjusted (OD=0.4) bacterial culture and mixed with respective wells (triplicate) of each sample with pH 4.0 media to make the total volume 200µl per well. Immediately after mixing OD_{600} was measured using the ELISA reader plate. This was considered as 0^{th} hour measurement. Plates were then incubated at $37^{o}C$ in 5% CO_2. OD_{600} was measured at time points 3hours and 6 hours.

3.3.3 Antibiotic Sensitivity Assay

Susceptibility to 14 antimicrobial agents were determined for 76 *S.* Mbandaka isolates using the Sensititre NARMS Gram Negative Plate (CMV3AGNF, Thermofisher). Resistance to Antimicrobial agents was determined as per Clinical and Laboratory Standards Institute (CLSI) guidelines. Five beta lactams (Amoxicillin/ Clavulanic acid, Ampicillin, Cefoxitin, Ceftiofur, and Ceftriaxone), two quinolones (Ciprofloxacin and Nalidixic acid), two aminoglycosides (Gentamicin and Streptomycin), azithromycin tetracycline, Sulfa trimethoprim (Trimethoprim/ sulfamethoxazole), sulfisoxazole and Chloramphenicol were the antibiotics used for susceptibility testing. Test results acquired

for antibiotics Azithromycin and sulfisoxazole were indeterminate and were excluded in further analysis.

3.3.4 Statistical analysis

Single factor ANOVA (one way ANOVA) and Tukey multiple comparison test were performed in Prism 7 (GraphPad software, Inc.), where a p-value less than 0.05 was considered as significant [165]. Significant difference in Tukey multiple comparison test was determined by pair wise comparison sample means. Difference between two means was computed at confidence interval of 95%. For each comparison, critical value q was calculated using the equation q = q=sqrt (2) *D/SED, where, D is the difference between two means and SED is the standard error of that difference (computed from all data).

3.4 Results and discussion

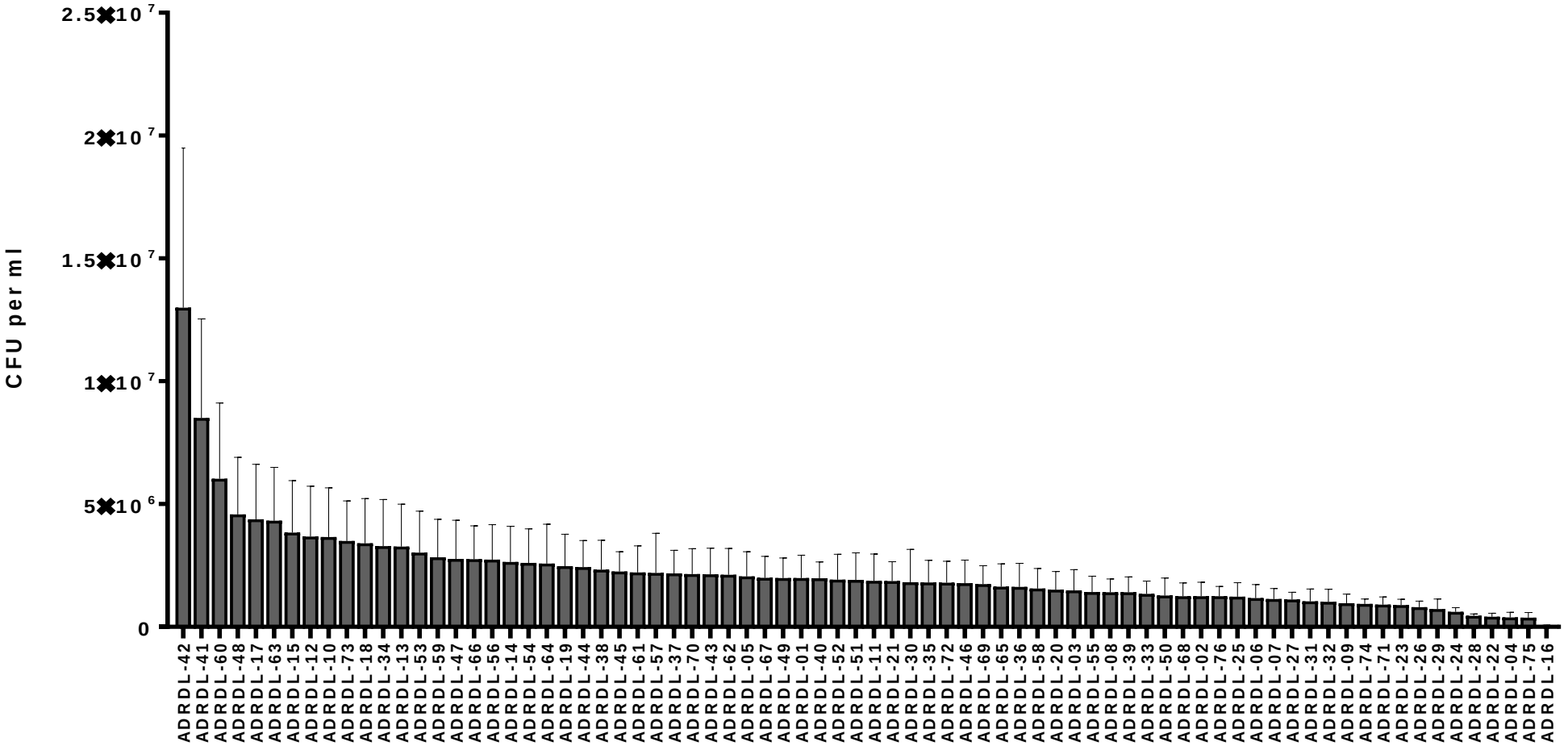
3.4.1 Analysis of host cell invasiveness of *S*. Mbandaka on Caco2 cells

Gentamicin protection assay [233] with slight modification was used for invasion assays. The results are shown in Fig.13. A MOI of 100:1 (100 bacteria per cell) was used to infect Caco2 cells as in vitro cell culture model. Cells were incubated with bacteria for 2 hours. Non-invading extra cellular bacteria were killed by gentamicin treatment at a concentration of 100µg/ml. Lysis of the cells with 1% triton X exposed the live intra cellular bacteria which were then quantified by plating on LB agar plates for CFU.

All strains have shown the property of cell invasion on Caco2 cells. Comparative analysis of the data between strains shows ADRDL-42, ADRDL-41 and ADRDL-60 invaded Caco2 cells with higher number compared to other strains. This increased cell invasion property does not correlate with virulence factors present in their genome as

analyzed in this study (not shown). These strains have similar virulence factors as most of other strains in this study. One way ANOVA showed an P value of 0.0063 which is lower than 0.05 indicating that there is significant difference among means of isolates. A multiple comparison analysis (Tukeys test) was performed to identify actual isolates which were showing the significant difference. Only one isolate, ADRDL 42, showed significant difference from others at a confidence interval of 95%. ADRDL-16 an isolate from chicken drag swab showed least number of bacteria invaded the cells. But this strain also has similar virulence factors as those other strains which showed the higher invasiveness. We expected a higher degree of cell invasiveness in case of eighteen isolates in our study group that had a background as clinical isolates from cattle. But in contrary, it was a non-clinical bovine isolate from Texas, ADRDL42, showed more invasiveness than any other isolates in this study. ADRDL16 was an environmental sample. Considering these two samples we may suggest that host associated strains may be more invasive than environmental sample but to confirm this further studies based on more host specific isolates is required.

Fig. 13. Result of *S*. Mbandaka invasion assay using Caco2 cells. Invasion assay was performed using all 76 strains under this study. Graph represents one way ANOVA of data from 2 experiments, each with 2 replicates of bacterial treatment. Data is expressed as number of bacteria in CFU per ml that invaded Caco2 cells.



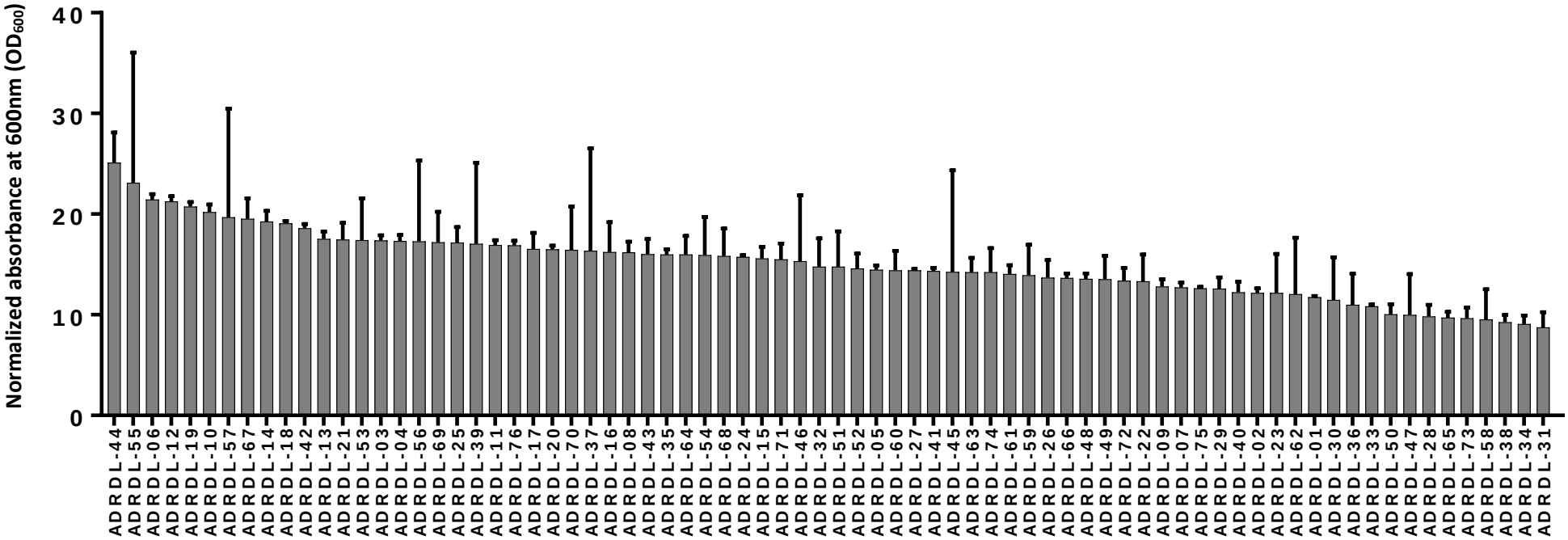
3.4.2 Tolerance to low pH varied between *S*. Mbandaka strains

To study the tolerance ability of different *S*. Mbandaka strains in low pH, overnight culture of bacteria in LB broth (pH 7.0 ± 0.2) were shifted to a low pH (pH 4.0 ± 0.1) LB media. The adaptation and survival of each strain to this low pH environment were assessed by overtime growth measured at different time points. Overnight cultures, OD_{600} of which was adjusted to 0.4 ± 0.03, of all strains were used to inoculate LB media with a pH of 4.0 ± 0.1. An initial OD_{600} was measured immediately after inoculation (0^{th} hour) which was later used to normalize the final reading taken after 6 hours of incubation. All 76 *S*. Mbandaka strains in this study survived the treatment with LB broth (pH 4.0 ± 0.1) without any acid adaptation treatment. An increase in OD_{600} could be observed in all the samples indicating an increase in the number of bacteria after incubation at 37^{o}C for a period of 6 hours. Results were shown in Fig 14.

Salmonella are neutrophilic organisms capable of growing in a pH range of 5 to 9. Acid tolerance response of Nontyphoidal *Salmonella* such as *S*. Typhimurium has been well studied previously. In pioneering studies of Foster et al., it has been shown that *S*. Typhimurium can adapt to survive conditions of severe low pH (pH 3.3) by a phenomenon called Acid Tolerance Response (ATR). An adaptation treatment with low pH media (acid shock) for a short time was shown to be required for a maximum induction of ATR in *S*. Typhimurium. But acid shock (at pH 4.3) exposure of >30min does not induce this response [133]. Here, in this study we were trying to find out how well the bacterial strains adapt to a low pH and is there any difference in the adaptation performance between strains. Results showed all the strains under this study could

survive an immediate exposure to a low pH environment. Single factor ANOVA of the present data has shown that there is significant difference in the OD_{600} value between different strains with a p value < 0.05. The difference was irrespective of the isolation source, geographical location of strain collection and clinical significance (clinical or non-clinical strains). Upon multiple comparison test (Tukeys test) only one isolate ADRDL44 showed significant difference in mean OD_{600} with other four isolate ADRDL73, 58, 38, 34, and 31. ADRDL 44 a non-clinical bovine isolate from Texas showed faster adaptance to low pH than isolates from similar location and source. Except ADRDL73 ,which was from Oklahoma, all other four were from Texas. Growth phase of the bacteria used for the treatment, the type of acidulant used for adjusting the pH of the media, composition of the adaptation media, and growth temperature were reported to be the factors which influence the ATR of *Salmonella* spp.[132]. Use of complex LB media, an incubation temperature of 37ºC, and overnight grown bacteria culture might have favored the development of this resistance to low pH in strains used in this study. But the explanation for the difference in resistance level, reflected by a relative difference in the growth, between different strains remains questionable. Wide variation in acid tolerance response among *Salmonella* serovars and even between strains of a given serovar has been reported previously [132], which may not be applicable here since we were looking the adaptation of strains when immediate exposure to low pH without any low pH adaptation treatment. Even then our data shows ability of *S.* Mbandaka isolates from different isolation source to be capable of adjusting to a low pH without any prior treatment.

Fig. 14. Growth of S. Mbandaka strains at pH 4.0. Overnight culture of all 76 *S*. Mbandaka strains (OD adjusted to 0.4) were inoculated to a low pH (pH 4.0± 0.1) LB media and incubated for a period of 6 hours at 37°C. Bacterial growth was assessed by measuring absorbance at 600 (OD_{600}) after 6 hours of incubation. Normalized data with 0^{th} hour OD_{600} was presented as bar graph.



3.4.3 Antimicrobial sensitivity

Antimicrobial resistance was determined by using Sensititre NARMS Gram Negative Plate. Out of 76 *S.* Mbandaka strains tested, only 48 isolate showed resistance against 12 antibiotics tested (fig.7). Forty-six isolates were resistant to streptomycin. Seventy-eight percentage of them (36 isolates) were resistant to only streptomycin. Remaining ones showed resistance to other antibiotics also. Beta lactam resistance was shown by only four isolates. Out of four ADRDL- 48, ADRDL-58 and ADRDL-09 showed only intermediate resistance to antibiotic Cefoxitin. The remaining one, ADRDL-73 showed resistance to Ampicillin (intermediate) and Amoxicillin/Clavulanic acid. Except ADRDL – 09, which showed intermediate resistance to Chloramphenicol as well, all beta lactam resistant strains were also resistant to streptomycin.

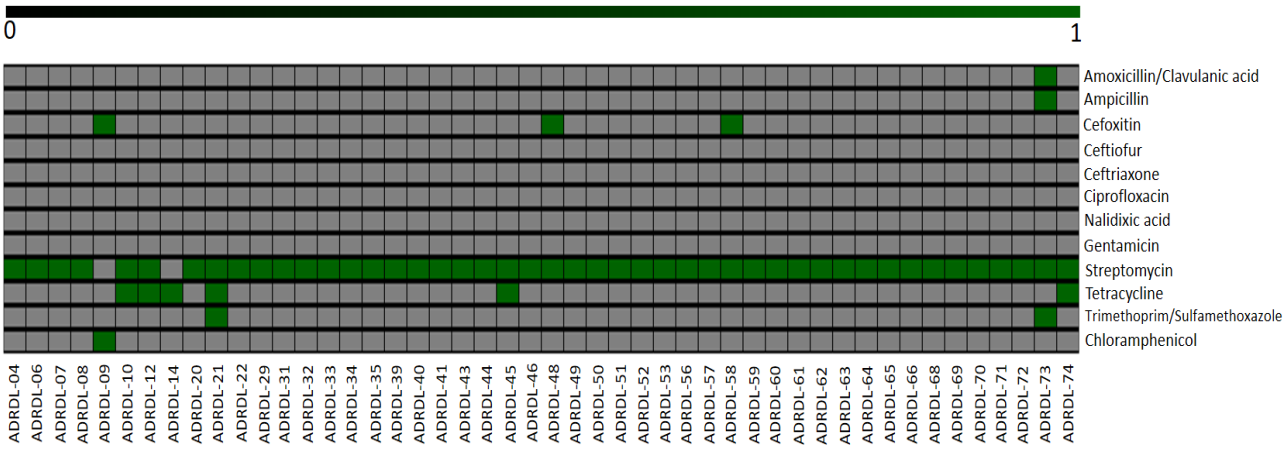


Fig.15. Heat map showing resistance to Antimicrobial agents. Resistance against 12 antimicrobial agents were determined using Sensititre NARMS Gram Negative Plate based on CLSI standards. Out of 76 *Salmonella* isolates, only 48 isolates showed resistance to at least one antibiotic. Only 10 isolates showed resistance to more than one antibiotic. Legend description: 0 = Susceptible, 1 = Resistant. (Intermediate resistance phenotypes were

Five isolates showed resistance to both streptomycin and tetracycline. One isolate, ADRDL-14, was resistant to only tetracycline. Only two streptomycin resistant isolates

also showed resistance against trimethoprim/sulfamethoxazole. Isolate ADRDL-73 was the one with resistant to more number of antibiotics (4) followed by ADRDL – 21 which was resistant to 3 antibiotics.

Theoretically, phenotypic features can be explained by the underlined genetic features. But in this study certain discrepancies exist between genotypic and phenotypic features of antimicrobial resistance in our study isolates. This may be explained by the currently unknown resistance mechanisms and related genes that were not included in the database. Although high degree of congruencies was expected between genotypic and phenotypic results, disagreements were also noticed.

In genotypic method, only 11 newly sequenced isolates were identified with resistance genes. Streptomycin resistant genes (*strA* and *strB*) were identified in three isolates ADRDL33, ADRDL45 and ADRDL 74. All three showed concordant results in phenotypic identification of streptomycin resistance. ADRDL 21 and ADRDL 73 that contained 13 aadA genes (aminoglycoside resistance genes) also showed resistant to streptomycin phenotypically. Remaining 41 isolates that were phenotypically resistant to streptomycin, showed disagreement with genotypic methods. Concordant results were obtained for tetracycline resistance in case of 6 isolates (ADRDL10, 12, 14, 21, 45, and 74), but isolates ADRDL 33 and 39 were showed disagreement to phenotypic methods since they were identified with tetracycline resistant genes (*tetA*). Genes for betalactam resistance were identified in ADRDL 31 (*blaSED*) and ADRDL 73 (*blaTEM*-153 genes, *blaCARB* – 7 genes), but only later one showed resistance phenotypically to Amoxicillin/clavulanic acid and ampicillin. There were other isolates (ADRDL 09, 48

and 58) that showed resistance to beta lactams (Cefoxitin) in Antibiotic sensitivity assay even though no resistance genes were identified.

In case of sulfonamide trimethoprim resistance, ADRDL21 (*sul1, sul3,* and *dfrA12*), and ADRDL73 (*sul2*) showed concordant result with Antibiotic sensitivity assay using trimethoprim/sulfamethoxazole. Gene *sul2* was present in ADRDL 45 but no resistance showed in phenotypic assay. ADRDL7 was identified with 66 *qnrB* quinolone resistant genes but no resistance showed in sensitivity assay against ciprofloxacin and nalidixic acid. No phenicol resistant genes identified in ADRDL9, but showed resistance against chloramphenicol.

3.5 Conclusions

In this study, we demonstrate that *S*. Mbandaka isolates could invade human colon carcinoma cells (Caco2 cells) without much significant difference among selected isolates. This clearly explains the ability of this serovar to cause significant human disease. However, a comparative study with other invasive non-typhoidal *Salmonella* may be needed to explain how well they differ from other serovars in the expression of this phenotypic character. Additionally, we showed the potential of this serovar to adapt to a low pH environment and thereby the capability of this serovar to survive in food commodities as well as to overcome host defense barriers like acidic environment in the stomach. Moreover, existence of isolates with multi drug resistance revealed by this study raises a serious concern as this serovar is more prevalent in food animals.

Table 7. Unique SNPs that define *S.* Mbandaka clusters in SNP tree

Sample ID	Chromosome Region	Reference	Allele	Gene	Locus_tag (NZ_CP019183 (CDS))	Note	Product	Protein ID	Coding region change	Amino acid change	Non-synonymous
Unique SNPs of cluster 1 bovine isolates from Texas in comparison with sub cluster 2B bovine isolates from Texas											
ADRDL27 (Cluster 1)	320976	G	A		SEEM1958_RS01505		branched-chain amino acid ABC transporter permease	WP_000003007.1			
	2212564	C	T		SEEM1958_RS11240		flagellar brake protein YcgR	WP_000017418.1	G>A	Gly-Asp	Yes
	4223531	G	A		SEEM1958_RS21295		peptide-methionine (S)-S-oxide reductase	WP_000051467.1	G>A	Gly-Asp	Yes
	82888	G	T		SEEM1958_RS00410		magnesium-translocating P-type ATPase	WP_000131288.1	G>T	Val-Leu	Yes
	1003669	C	T	pyrG	SEEM1958_RS05115	CTP synthase; cytidine triphosphate synthetase; catalyzes the ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia as the source of nitrogen; in Escherichia coli this enzyme forms a homotetramer	CTP synthetase	WP_000210863.1	A>T	Asn-Tyr	Yes

	2586466	A	T	SEEM1958_RS13125		amino acid transporter	WP_000412353.1	A>T	Glu-Val	Yes
	3434830	T	A	SEEM1958_RS17535		hypothetical protein	WP_000495330.1			
	2858628	G	T	SEEM1958_RS14605		cell division protein YceG	WP_000736459.1			
	904588	G	A	SEEM1958_RS04660		fimbrial protein	WP_001244646.1	G>A	Gly-Glu	Yes
	398475	T	C	SEEM1958_RS01835		carboxypeptidase/ penicillin-binding protein 1A	WP_001663100.1			
	144061	T	A	SEEM1958_RS00695		glycosyltransferase	WP_023217918.1	T>A	Asp-Glu	Yes
	1598613	C	T	SEEM1958_RS08030		citrate transporter	WP_023218370.1			
	3358397	G	T	SEEM1958_RS17120		oxidoreductase	WP_023218841.1	C>A	Ser	Yes
	2350269	G	T	SEEM1958_RS11960		two-partner secretion translocator ZirT	WP_023227594.1			
Unique SNPs of cluster1 poultry isolates in comparison with cluster1 bovine isolates										
USDA015 (Cluster 1)	3874840	A	C	SEEM1958_RS19665		type II secretion system protein GspE	WP_076031835.1	A>C	Glu-Asp	Yes
Unique SNPs of sub clusters in comparison with cluster1										
USDA005 (Sub cluster 2A)	1224826	A	C	SEEM1958_RS06300	frameshifted	phage tail protein		A>C	Lys-Thr	Yes
	1225044	A	C	SEEM1958_RS06305	frameshifted	phage tail protein		A>C	His-Pro	Yes

FDA083 (Sub cluster 2B)	3415090	A	T	SEEM1958_RS17400	incomplete; partial in the middle of a contig; missing stop	hypothetical protein		A>T	Thr-Ser	Yes
	1223450	A	G	SEEM1958_RS06300	frameshifted	phage tail protein		G>A	Gly-Asp	Yes
	1224913	C	G	SEEM1958_RS06305	frameshifted	phage tail protein		T>C	Cys-Arg	Yes
	1227644	C	T	SEEM1958_RS06325		tail sheath protein	WP_000046109.1			
	1200943	A	G	SEEM1958_RS06130		phage repressor protein	WP_000052560.1	G>A	Asp-Asn	Yes
	922239	A	C	SEEM1958_RS04755		diaminopimelate decarboxylase	WP_000056587.1	A>C	Ile-Leu	Yes
	2057250	G	A	SEEM1958_RS10400	structural flagella protein; individual Salmonella serotypes usually alternate between the production of 2 antigenic forms of flagella, termed phase 1 and phase 2, each specified by separate structural genes	flagellin FliC	WP_000079802.1	G>A	Ala-Thr	Yes
	1201562	T	A	SEEM1958_RS06135		hypothetical protein	WP_000102104.1	T>A	Lys	Yes
	1217272	A	G	SEEM1958_RS06245		hypothetical protein	WP_000171565.1	G>C	Val-Leu	Yes
	1221428	A	G	SEEM1958_RS06285		baseplate assembly protein	WP_000177484.1	A>G	Ile-Val	Yes
	488911	C	T	SEEM1958_RS02390		adenine-specific DNA-methyltransferase	WP_000642611.1			

	3847405	G	A	SEEM1958_RS19550		DeoR family transcriptional regulator	WP_000678254.1			
	1214990	C	T	SEEM1958_RS06225		phage capsid protein	WP_000730757.1	A>G	Thr-Ala	Yes
	1229644	C	T	SEEM1958_RS06340		GpE family phage tail protein	WP_000763316.1			
	33143	G	T	SEEM1958_RS00145		hypothetical protein	WP_000809951.1			
	4544964	G	A	SEEM1958_RS22885		isocitrate lyase	WP_000857884.1			
	1217113	A	G	SEEM1958_RS06240		tail protein X	WP_000868184.1	C>A	Gln-Lys	Yes
	4566649	C	A	SEEM1958_RS22985	binds specifically to the major sigma factor sigma 70; active in stationary phase	sigma D regulator	WP_000934317.1	C>A	Ser-Tyr	Yes
	1202399	G	A	SEEM1958_RS06145		hypothetical protein	WP_000956168.1			
	1202521	T	G	SEEM1958_RS06150		hypothetical protein	WP_000963480.1	A>T	Leu-Phe	Yes
	1225847	C	T	SEEM1958_RS06310		tail fiber assembly protein	WP_000972188.1	G>A	Val-Ile	Yes
	1219184	A	C	SEEM1958_RS06265		tail protein	WP_001039961.1	A>C	Met-Leu	Yes
	109552	T	A	SEEM1958_RS00510		tRNA (guanosine(18)-2'-O)-methyltransferase TrmH	WP_001068433.1	A>T	Glu-Val	Yes
	1217549	T	G	SEEM1958_RS06250		glycoside hydrolase	WP_001069919.1			

	1207353	G	T	SEEM1958_RS06180		hypothetical protein	WP_001154444.1	A>C	Ile-Leu	Yes
	1229006	C	T	SEEM1958_RS06330		major tail tube protein	WP_001207652.1	G>A	Ser-Asn	Yes
	1207577	C	T	SEEM1958_RS06185		DinI family protein	WP_001217581.1			
	1229315	C	T	SEEM1958_RS06335		phage tail assembly protein	WP_001280962.1	T>A	Leu-Gln	Yes
	1200547	A	T	SEEM1958_RS06125		integrase	WP_001536726.1			
	1221029	A	G	SEEM1958_RS06280		phage baseplate assembly protein V	WP_001556169.1	C>A	Leu-Met	Yes
	1205863	G	A	SEEM1958_RS06175		replication endonuclease	WP_023218233.1	A>G	Glu-Gly	Yes
	1211708	G	C	SEEM1958_RS06210		phage capsid portal protein	WP_023218236.1	G>T	Ala-Ser	Yes
	1212256	C	G	SEEM1958_RS06215		terminase subunit	WP_023218237.1	C>G	Pro-Ala	Yes
	1214089	G	C	SEEM1958_RS06220		phage capsid scaffolding protein	WP_023218238.1	G>C	Glu-Asp	Yes
	1216763	T	A	SEEM1958_RS06235		hypothetical protein	WP_023218239.1			
	1222504	T	C	SEEM1958_RS06295		phage tail protein I	WP_023218240.1			
	2143283	A	G	SEEM1958_RS10880		hypothetical protein	WP_023218578.1	A>G	Gln-Arg	Yes
	3616455	C	T	SEEM1958_RS18405		AraC family transcriptional regulator	WP_023219097.1	G>A	Ala-Thr	Yes
	1087107	A	G	SEEM1958_RS05520	GTP-activating protein/tyrosine phosphatase;	pathogenicity island 1 effector protein StpP	WP_076031708.1	A>G	Glu-Gly	Yes

						facilitates bacterial survival in host cells					
FDA015 (Sub cluster 2C)	1226005	C	T		SEEM1958_RS06315	frameshifted	phage tail protein		A>G	Lys-Arg	Yes
	4448414	A	T		SEEM1958_RS22520		acetyl-coenzyme A synthetase	WP_000083883.1			
	3728054	A	T		SEEM1958_RS18975		hypothetical protein	WP_000145239.1	A>T	His-Leu	Yes
	2444591	A	G		SEEM1958_RS12420		acyltransferase	WP_000155373.1			
	937099	G	A		SEEM1958_RS04810		prolipoprotein diacylglyceryl transferase	WP_000204647.1	G>A	Met-Ile	Yes
	1132162	T	C	srlA	SEEM1958_RS05760	catalyzes the phosphorylation of incoming sugar substrates along with their translocation across the cell membrane; the IIC domain forms the PTS system translocation channel and contains the specific substrate-binding site	PTS sorbitol transporter subunit IIC	WP_000573333.1	A>G	Asn-Asp	Yes
	4577367	G	A		SEEM1958_RS23040		DNA-directed RNA polymerase subunit beta'	WP_000653965.1	C>T	Ala-Val	Yes
	4202506	G	A		SEEM1958_RS21195		2-dehydro-3-deoxyphosphooctonate aldolase	WP_000779253.1			

	1225552	T	C	SEEM1958_RS06310		tail fiber assembly protein	WP_000972188.1	A>G	Gln-Arg	Yes
	489693	C	T	SEEM1958_RS02400		tRNA dihydrouridine synthase DusB	WP_001219664.1			
	304734	C	A	SEEM1958_RS01425		hypothetical protein	WP_021000870.1	G>T	Gly-Val	Yes
	1240217	C	T	SEEM1958_RS06385		VCBS repeat-containing protein	WP_023218145.1	G>A	Val-Ile	Yes
	2666853	A	G	SEEM1958_RS13600		pathogenicity island protein	WP_023218254.1	A>G	His-Arg	Yes
	3959663	G	A	SEEM1958_RS20055		L-carnitine CoA-transferase	WP_023218424.1	G>A	Arg-His	Yes
	623204	T	A	SEEM1958_RS03095		glycerate 2-kinase	WP_023219238.1	T>A	Asp-Glu	Yes
EUR069 (Sub Cluster 2E)	2576538	C	A	SEEM1958_RS13070		MFS transporter	WP_000091796.1	C>A	Tyr	Yes
	2534501	C	T	SEEM1958_RS12855		sugar efflux transporter	WP_000154617.1	C>T	Gln	Yes
	1149094	G	A	SEEM1958_RS05870	DNA-binding transcriptional repressor of microcin B17 synthesis and multidrug efflux; negative regulator of the multidrug operon emrAB	transcriptional regulator	WP_000378431.1			
	3592002	T	A	SEEM1958_RS18305	Confers resistance to fosfomycin and deoxycholate;	MFS transporter	WP_000446768.1	A>T	Asn-Ile	Yes
	3094297	G	A	SEEM1958_RS15795		membrane protein	WP_000505788.1			

	497337	A	T		SEEM1958_RS02435		mononuclear molybdenum enzyme YedY	WP_000723876.1			
	342809	C	A		SEEM1958_RS01620		aspartate-semialdehyde dehydrogenase	WP_000799940.1			
	926076	C	T		SEEM1958_RS04770		bifunctional 2-acylglycerophosphoethanolamine acyltransferase/acyl-ACP synthetase	WP_000896101.1	C>T	Pro-Leu	Yes
	1603990	G	A		SEEM1958_RS24795		hypothetical protein	WP_001121022.1	G>A	Met-Ile	Yes
	489693	C	T		SEEM1958_RS02400		tRNA dihydrouridine synthase DusB	WP_001219664.1			
	3964150	C	T	carB	SEEM1958_RS20080	four CarB-CarA dimers form the carbamoyl phosphate synthetase holoenzyme that catalyzes the production of carbamoyl phosphate; CarB is responsible for the amidotransferase activity	carbamoyl phosphate synthase large subunit	WP_023218427.1			
	2786092	G	A		SEEM1958_RS14210		hypothetical protein	WP_023219026.1			
	2350181	C	T		SEEM1958_RS11960		two-partner secretion translocator ZirT	WP_023227594.1	C>T	Thr-Ile	Yes
	3415888	G	T		SEEM1958_RS17410		integrase	WP_023237847.1	G>T	Met-Ile	Yes
	2512377	C	T		SEEM1958_RS12735		PhoPQ-regulated protein	WP_076031773.1	C>T	Thr-Ile	Yes

FDA058 (Sub cluster 2F)	3968645	C	T	SEEM1958_RS25005	frameshifted	hypothetical protein				
	1223302	G	A	SEEM1958_RS06300	frameshifted	phage tail protein		G>A	Ser-Asn	Yes
	1224908	G	A	SEEM1958_RS06305	frameshifted	phage tail protein		G>A	Asp-Asn	Yes
	1227683	C	T	SEEM1958_RS06325		tail sheath protein	WP_000046109.1	C>A	Ala-Asp	Yes
	563931	G	A	SEEM1958_RS02780		LPS export ABC transporter periplasmic protein LptC	WP_000047845.1	C>T	Thr-Met	Yes
	1216391	A	G	SEEM1958_RS06230		terminase endonuclease subunit	WP_000059172.1			
	2576538	C	A	SEEM1958_RS13070		MFS transporter	WP_000091796.1	C>A	Tyr	Yes
	456794	T	G	SEEM1958_RS02190		50S ribosomal protein L6	WP_000091939.1			
	1201676	A	G	SEEM1958_RS06135		hypothetical protein	WP_000102104.1			
	1217302	T	C	SEEM1958_RS06245		hypothetical protein	WP_000171565.1			
	1221349	C	T	SEEM1958_RS06285		baseplate assembly protein	WP_000177484.1	G>A	Arg-Gln	Yes
	1221773	T	C	SEEM1958_RS06290		baseplate assembly protein	WP_000268273.1	A>G	Gln-Arg	Yes
	284614	G	A	SEEM1958_RS01330	involved in resistance to DNA-damaging agents;	universal stress protein A	WP_000323565.1			

3171839	C	T	SEEM1958_RS16160	permease	WP_000373611.1			
1215164	C	T	SEEM1958_RS06225	phage capsid protein	WP_000730757.1			
1217113	A	G	SEEM1958_RS06240	tail protein X	WP_000868184.1			
1232683	T	C	SEEM1958_RS06350	tail assembly protein	WP_000980409.1			
1218916	C	T	SEEM1958_RS06265	tail protein	WP_001039961.1	C>A	Leu-Met	Yes
1217489	C	T	SEEM1958_RS06250	glycoside hydrolase	WP_001069919.1			
1226936	T	C	SEEM1958_RS06320	multiple promoter invertase	WP_001165558.1			
1229243	C	T	SEEM1958_RS06330	major tail tube protein	WP_001207652.1			
489693	C	T	SEEM1958_RS02400	tRNA dihydrouridine synthase DusB	WP_001219664.1			
1199875	G	A	SEEM1958_RS06125	integrase	WP_001536726.1			
1229801	T	C	SEEM1958_RS06345	phage tail tape measure protein	WP_023218142.1	G>A	Arg-Lys	Yes
1233309	T	A	SEEM1958_RS06355	phage late control D family protein	WP_023218143.1	A>C	Lys-Gln	Yes
1205836	C	G	SEEM1958_RS06175	replication endonuclease	WP_023218233.1	C>A	Asp-Glu	Yes
1211306	C	T	SEEM1958_RS06210	phage capsid portal protein	WP_023218236.1	C>A	Leu-Met	Yes
1212037	G	A	SEEM1958_RS06215	terminase subunit	WP_023218237.1			

	1213954	C	T	SEEM1958_RS06220		phage capsid scaffolding protein	WP_023218238.1		
	1216877	T	C	SEEM1958_RS06235		hypothetical protein	WP_023218239.1		
	1222563	C	T	SEEM1958_RS06295		phage tail protein I	WP_023218240.1	C>T	Thr-Ile	Yes
	1287670	G	A	SEEM1958_RS06580	catalyzes the acetylation of lysine	CoA-binding domain/acetyltransferase	WP_023218612.1		
	1207865	T	A	SEEM1958_RS06190		hypothetical protein	WP_024147339.1	T>A	Ser-Thr	Yes
ADRDL45 (Sub cluster 2B)	3415090	A	T	SEEM1958_RS17400	incomplete; partial in the middle of a contig; missing stop	hypothetical protein		A>T	Thr-Ser	Yes
	1223100	G	A	SEEM1958_RS06300	frameshifted	phage tail protein		G>A	Ala-Thr	Yes
	1224913	C	G	SEEM1958_RS06305	frameshifted	phage tail protein		A>C	His-Pro	Yes
	1226005	C	T	SEEM1958_RS06315	frameshifted	phage tail protein		G>A	Cys-Tyr	Yes
	2137369	C	G	SEEM1958_RS24840	frameshifted; internal stop; incomplete; partial in the middle of a contig; missing start	phage tail protein		C>G	Pro-Ala	Yes
	3446149	T	G	SEEM1958_RS17580		MFS transporter	WP_000005070.1	A>C	Ile-Leu	Yes
	1227594	A	G	SEEM1958_RS06325		tail sheath protein	WP_000046109.1	A>G	Ile-Val	Yes
	1201345	G	A	SEEM1958_RS06130		phage repressor protein	WP_000052560.1			

	1215905	T	C	SEEM1958_RS06230		terminase endonuclease subunit	WP_000059172.1	C>T	Leu-Phe	Yes
	2746480	A	C	SEEM1958_RS13990	catalyzes the phosphorylation of incoming sugar substrates concomitant with their translocation across the cell membrane; involved in N,N'-diacetylchitobiose transport; protein IIA transfers a phosphoryl group to IIB which then transfers the phosphoryl group to the sugar; IIC forms the translocation channel for the sugar uptake;	PTS N,N'-diacetylchitobiose transporter subunit IIC	WP_000073063.1	T>G	Ser-Ala	Yes
	2575443	G	A	SEEM1958_RS13070		MFS transporter	WP_000091796.1			
	1201574	G	A	SEEM1958_RS06135		hypothetical protein	WP_000102104.1	T>A	Lys	Yes
	1508355	G	A	SEEM1958_RS07570		divalent metal cation transporter	WP_000131734.1			
	2769616	C	T	SEEM1958_RS14115		aldehyde dehydrogenase	WP_000153505.1			
	1217293	T	G	SEEM1958_RS06245		hypothetical protein	WP_000171565.1	G>A	Gly-Asp	Yes
	1221253	G	T	SEEM1958_RS06285		baseplate assembly protein	WP_000177484.1	G>T	Gln-His	Yes
	1425996	T	C	SEEM1958_RS07135		gluconate:proton symporter	WP_000200859.1	T>C	Val-Ala	Yes

	3919462	G	A	thiP	SEEM1958_RS19865	permease; with TbpA and ThiQ functions in transport of thiamine and thiamine pyrophosphate into the cell; repressed in presence of exogenous thiamine	thiamine/thiamine pyrophosphate ABC transporter permease ThiP	WP_000235635.1	G>A	Arg-Gln	Yes
	1607255	C	A		SEEM1958_RS08065	NuoCD; NDH-1 shuttles electrons from NADH, via FMN and iron-sulfur (Fe-S) centers, to quinones in the respiratory chain; subunits NuoCD, E, F, and G constitute the peripheral sector of the complex; in Escherichia coli this gene encodes a fusion protein of NuoC and NuoD that are found separate in other organisms	NADH-quinone oxidoreductase subunit C/D	WP_000247855.1	C>A	Pro-Gln	Yes
	1221611	C	G		SEEM1958_RS06290		baseplate assembly protein	WP_000268273.1	T>C	Tyr-His	Yes
	1219301	G	A		SEEM1958_RS06270		virion morphogenesis protein	WP_000343946.1	A>G	Asn-Ser	Yes
	3868235	T	C		SEEM1958_RS19625		aromatic amino acid transporter AroP	WP_000378222.1			

	1493178	G	T	ligA	SEEM1958_RS07470	this protein catalyzes the formation of phosphodiester linkages between 5'-phosphoryl and 3'-hydroxyl groups in double-stranded DNA using NAD as a coenzyme and as the energy source for the reaction; essential for DNA replication and repair of damaged DNA; similar to ligase LigB	DNA ligase (NAD(+)) LigA	WP_000433281.1	G>T	Asp-Tyr	Yes
	2788198	C	T		SEEM1958_RS14230		leucine efflux protein	WP_000457190.1	G>A	Gly-Arg	Yes
	1201813	T	C		SEEM1958_RS06140		hypothetical protein	WP_000460858.1	G>T	Glu-Asp	Yes
	4871730	A	G		SEEM1958_RS24525		tRNA uridine-5-carboxymethylaminomethyl(34) synthesis enzyme MnmG	WP_000499872.1	A>G	Glu-Gly	Yes
	2948786	C	A		SEEM1958_RS15135		3,4-dihydroxyphenylacetate 2,3-dioxygenase	WP_000517001.1			
	2105141	A	G		SEEM1958_RS10650		tRNA 5-methoxyuridine(34)/uridine 5-oxyacetic acid(34) synthase CmoB	WP_000569026.1			

	2770753	G	A	SEEM1958_RS14125		D-hexose-6-phosphate mutarotase	WP_000608657.1	G>A	Trp	Yes
	3319014	C	T	SEEM1958_RS16905		2-keto-3-deoxygluconate permease 2	WP_000694477.1	G>A	Ala-Thr	Yes
	2656377	C	T	SEEM1958_RS13550		CesD/SycD/LcrH family type III secretion system chaperone	WP_000711027.1	G>A	Arg-His	Yes
	1214783	A	G	SEEM1958_RS06225		phage capsid protein	WP_000730757.1	A>G	Ser-Gly	Yes
	1229629	G	A	SEEM1958_RS06340		GpE family phage tail protein	WP_000763316.1	T>C	Ser-Pro	Yes
	2238611	C	T	SEEM1958_RS11360		peptide chain release factor 1	WP_000804703.1	C>T	Pro-Ser	Yes
	33143	G	T	SEEM1958_RS00145		hypothetical protein	WP_000809951.1			
	1217113	A	G	SEEM1958_RS06240		tail protein X	WP_000868184.1	C>A	Gln-Lys	Yes
	4098702	T	C	SEEM1958_RS20705	serine sensor receptor	methyl-accepting chemotaxis protein	WP_000919519.1	A>G	Tyr-Cys	Yes
	4566649	C	A	SEEM1958_RS22985	binds specifically to the major sigma factor sigma 70; active in stationary phase	sigma D regulator	WP_000934317.1	C>A	Ser-Tyr	Yes
	56965	C	T	SEEM1958_RS00280	membrane protein regulates uhpT expression	MFS transporter family glucose-6-phosphate receptor UhpC	WP_000948174.1			
	1225461	A	C	SEEM1958_RS06310		tail fiber assembly protein	WP_000972188.1	T>G	His-Gln	Yes

	1232551	A	G	SEEM1958_RS06350		tail assembly protein	WP_000980409.1			
	4680436	A	G	SEEM1958_RS23545	response regulator in two-component regulatory system with CpxA; part of the envelope stress response system	DNA-binding response regulator	WP_001033731.1	A>G	Glu-Gly	Yes
	1218880	G	A	SEEM1958_RS06265		tail protein	WP_001039961.1	C>A	Asp-Glu	Yes
	2780356	G	A	SEEM1958_RS14160		GGDEF domain-containing protein	WP_001048657.1	G>A	Gly-Asp	Yes
	1217477	G	A	SEEM1958_RS06250		glycoside hydrolase	WP_001069919.1			
	1207342	C	T	SEEM1958_RS06180		hypothetical protein	WP_001154444.1	C>T	Leu-Phe	Yes
	2365060	G	A	SEEM1958_RS12040		tRNA 2-thiocytidine(32) synthetase TtcA	WP_001156210.1			
	1226945	A	G	SEEM1958_RS06320		multiple promoter invertase	WP_001165558.1			
	1218560	G	A	SEEM1958_RS06260		hypothetical protein	WP_001201940.1	G>A	Ala-Thr	Yes
	1228754	G	A	SEEM1958_RS06330		major tail tube protein	WP_001207652.1	G>T	Leu-Phe	Yes
	1207571	T	C	SEEM1958_RS06185		DinI family protein	WP_001217581.1			
	2491669	C	T	SEEM1958_RS12620		Na+/H+ antiporter NhaC	WP_001276631.1	G>A	Val-Met	Yes
	1229315	C	T	SEEM1958_RS06335		phage tail assembly protein	WP_001280962.1	T>A	Leu-Gln	Yes

	4097339	C	T	SEEM1958_RS20700	catalyzes the transfer of phosphoglycerol to the glucan backbone	phosphoglycerol transferase I	WP_001292726.1		
	1199788	G	A	SEEM1958_RS06125		integrase	WP_001536726.1	G>A	Val-Ile	Yes
	1220723	C	T	SEEM1958_RS06280		phage baseplate assembly protein V	WP_001556169.1	A>G	Asn-Asp	Yes
	1725878	C	T	SEEM1958_RS08590		hypothetical protein	WP_014344110.1	C>T	His-Tyr	Yes
	107659	C	T	SEEM1958_RS00505	catalyzes branch migration in Holliday junction intermediates	DNA helicase RecG	WP_017442022.1		
	1202900	C	G	SEEM1958_RS06155		hypothetical protein	WP_021293768.1	G>A	Met-Ile	Yes
	99368	T	C	SEEM1958_RS00480		alpha-xylosidase	WP_023217903.1	T>C	Ile-Thr	Yes
	3144512	A	G	SEEM1958_RS16025		mechanosensitive channel protein	WP_023218008.1	A>G	Met-Val	Yes
	1229744	C	T	SEEM1958_RS06345		phage tail tape measure protein	WP_023218142.1	C>A	Asp-Glu	Yes
	1233372	G	C	SEEM1958_RS06355		phage late control D family protein	WP_023218143.1	C>G	Asn-Lys	Yes
	2432965	G	A	SEEM1958_RS12370		D-alanyl-D-alanine dipeptidase	WP_023218164.1	G>A	Arg-Gln	Yes
	1204577	C	T	SEEM1958_RS06170		DNA adenine methylase	WP_023218232.1	C>T	Thr-Ile	Yes
	1204817	G	A	SEEM1958_RS06175		replication endonuclease	WP_023218233.1	G>A	Glu-Lys	Yes
	1211129	G	A	SEEM1958_RS06210		phage capsid portal protein	WP_023218236.1	C>G	Ile-Met	Yes
	1212037	G	A	SEEM1958_RS06215		terminase subunit	WP_023218237.1	G>T	Gln-His	Yes

	1213978	T	C	SEEM1958_RS06220		phage capsid scaffolding protein	WP_023218238.1	T>A	Val-Glu	Yes
	1216607	C	A	SEEM1958_RS06235		hypothetical protein	WP_023218239.1	A>G	Lys-Arg	Yes
	1222504	T	C	SEEM1958_RS06295		phage tail protein I	WP_023218240.1	A>G	Ser-Gly	Yes
	2089378	A	G	SEEM1958_RS10585		methyl-accepting chemotaxis protein II	WP_023218563.1	A>G	Ile-Val	Yes
	2143292	T	G	SEEM1958_RS10880		hypothetical protein	WP_023218578.1	T>G	Leu-Arg	Yes
	4106293	G	A	SEEM1958_RS20740		type I restriction-modification enzyme R subunit	WP_023218593.1			
	831161	A	G	SEEM1958_RS04235		IclR family transcriptional regulator	WP_023218731.1			
	2801682	G	A	SEEM1958_RS14315		hypothetical protein	WP_023218768.1			
	616093	C	T	SEEM1958_RS03070		tagatose-1,6-bisphosphate aldolase	WP_023219240.1			
	1087107	A	G	SEEM1958_RS05520	GTP-activating protein/tyrosine phosphatase; facilitates bacterial survival in host cells	pathogenicity island 1 effector protein StpP	WP_076031708.1	A>G	Glu-Gly	Yes
	1203119	T	G	SEEM1958_RS06160		hypothetical protein	WP_076031711.1	T>G	Ser-Ala	Yes
	2136830	C	G	SEEM1958_RS10835		integrase	WP_076031751.1	C>T	Leu-Phe	Yes

	2465723	G	A		SEEM1958_RS12505		MFS transporter	WP_076031765.1			
	3535700	T	C		SEEM1958_RS18015		hypothetical protein	WP_078054929.1	A>G	Tyr-Cys	Yes
Unique SNPs of Sub cluster 2A in comparison with Sub cluster 2B											
USDA005 (Sub cluster 2A)	1223166	G	A		SEEM1958_RS06300	frameshifted	phage tail protein		G>A	Val-Ile	Yes
	982690	G	A		SEEM1958_RS05030		MFS transporter	WP_000097015.1			
	1201601	G	T		SEEM1958_RS06135		hypothetical protein	WP_000102104.1			
	1217344	T	C		SEEM1958_RS06245		hypothetical protein	WP_000171565.1			
	1003669	C	T	pyrG	SEEM1958_RS05115	CTP synthase; cytidine triphosphate synthetase; catalyzes the ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia as the source of nitrogen; in Escherichia coli this enzyme forms a homotetramer;	CTP synthetase	WP_000210863.1			
	1221956	T	G		SEEM1958_RS06290		baseplate assembly protein	WP_000268273.1	T>G	Asp-Glu	Yes
	435944	T	G		SEEM1958_RS02005		ABC transporter ATP-binding protein	WP_000634765.1	A>C	Asp-Ala	Yes
	1215137	G	C		SEEM1958_RS06225		phage capsid protein	WP_000730757.1			

3044754	T	C	SEEM1958_RS15550	dimethyl sulfoxide reductase subunit A	WP_000850250.1	A>G	Asn-Asp	Yes
1232624	G	A	SEEM1958_RS06350	tail assembly protein	WP_000980409.1	G>A	Ala-Thr	Yes
1218916	C	T	SEEM1958_RS06265	tail protein	WP_001039961.1			
1207492	A	T	SEEM1958_RS06180	hypothetical protein	WP_001154444.1	A>T	Ser-Cys	Yes
1227392	A	G	SEEM1958_RS06320	multiple promoter invertase	WP_001165558.1			
1148368	C	A	SEEM1958_RS05865	EmrA/EmrK family multidrug efflux transporter periplasmic adaptor subunit	WP_001275597.1			
3511347	G	A	SEEM1958_RS17895	multidrug ABC transporter permease/ATP-binding protein	WP_023218106.1			
1230428	G	A	SEEM1958_RS06345	phage tail tape measure protein	WP_023218142.1	C>T	Arg-Trp	Yes
1234121	A	G	SEEM1958_RS06355	phage late control D family protein	WP_023218143.1			
1206907	T	C	SEEM1958_RS06175	replication endonuclease	WP_023218233.1	A>G	Gln-Arg	Yes
1212054	A	C	SEEM1958_RS06215	terminase subunit	WP_023218237.1	T>G	Ser-Ala	Yes
1222971	C	T	SEEM1958_RS06295	phage tail protein I	WP_023218240.1			

	4866604	C	T	SEEM1958_RS24495		ATPase RavA	WP_023218914.1	C>T	Pro-Leu	Yes
FDA083 (Sub cluster 2B)	1223450	A	G	SEEM1958_RS06300	frameshifted	phage tail protein				
	1227644	C	T	SEEM1958_RS06325		tail sheath protein	WP_000046109.1	T>A	Ser-Thr	Yes
	1200770	C	T	SEEM1958_RS06130		phage repressor protein	WP_000052560.1	G>A	Ser-Asn	Yes
	922239	A	C	SEEM1958_RS04755		diaminopimelate decarboxylase	WP_000056587.1	A>C	Ile-Leu	Yes
	2057250	G	A	SEEM1958_RS10400	structural flagella protein; individual Salmonella serotypes usually alternate between the production of 2 antigenic forms of flagella, termed phase 1 and phase 2, each specified by separate structural genes	flagellin FliC	WP_000079802.1	G>A	Ala-Thr	Yes
	1201592	T	C	SEEM1958_RS06135		hypothetical protein	WP_000102104.1	T>A	Lys	Yes
	1217272	A	G	SEEM1958_RS06245		hypothetical protein	WP_000171565.1	G>C	Val-Leu	Yes
	1221428	A	G	SEEM1958_RS06285		baseplate assembly protein	WP_000177484.1	A>G	Ile-Val	Yes
	488911	C	T	SEEM1958_RS02390		adenine-specific DNA-methyltransferase	WP_000642611.1			

	3847405	G	A	SEEM1958_RS19550		DeoR family transcriptional regulator	WP_000678254.1			
	1214990	C	T	SEEM1958_RS06225		phage capsid protein	WP_000730757.1	A>G	Thr-Ala	Yes
	1229644	C	T	SEEM1958_RS06340		GpE family phage tail protein	WP_000763316.1			
	33143	G	T	SEEM1958_RS00145		hypothetical protein	WP_000809951.1			
	4544964	G	A	SEEM1958_RS22885		isocitrate lyase	WP_000857884.1			
	1217131	T	G	SEEM1958_RS06240		tail protein X	WP_000868184.1	C>A	Gln-Lys	Yes
	4566649	C	A	SEEM1958_RS22985	binds specifically to the major sigma factor sigma 70; active in stationary phase	sigma D regulator	WP_000934317.1	C>A	Ser-Tyr	Yes
	1202399	G	A	SEEM1958_RS06145		hypothetical protein	WP_000956168.1			
	1202530	A	T	SEEM1958_RS06150		hypothetical protein	WP_000963480.1	A>T	Leu-Phe	Yes
	1219184	A	C	SEEM1958_RS06265		tail protein	WP_001039961.1	A>C	Met-Leu	Yes
	109552	T	A	SEEM1958_RS00510		tRNA (guanosine(18)-2'-O)-methyltransferase TrmH	WP_001068433.1	A>T	Glu-Val	Yes
	1217549	T	G	SEEM1958_RS06250		glycoside hydrolase	WP_001069919.1			
	1207353	G	T	SEEM1958_RS06180		hypothetical protein	WP_001154444.1	A>C	Ile-Leu	Yes

1229006	C	T	SEEM1958_RS06330	major tail tube protein	WP_001207652.1	G>A	Ser-Asn	Yes
1207577	C	T	SEEM1958_RS06185	DinI family protein	WP_001217581.1			
1229315	C	T	SEEM1958_RS06335	phage tail assembly protein	WP_001280962.1	T>A	Leu-Gln	Yes
1200532	C	G	SEEM1958_RS06125	integrase	WP_001536726.1	C>T	His-Tyr	Yes
1221041	C	A	SEEM1958_RS06280	phage baseplate assembly protein V	WP_001556169.1	C>A	Leu-Met	Yes
1202882	A	T	SEEM1958_RS06155	hypothetical protein	WP_021293768.1			
1205863	G	A	SEEM1958_RS06175	replication endonuclease	WP_023218233.1	A>G	Glu-Gly	Yes
1211719	C	A	SEEM1958_RS06210	phage capsid portal protein	WP_023218236.1	G>T	Ala-Ser	Yes
1212256	C	G	SEEM1958_RS06215	terminase subunit	WP_023218237.1	C>G	Pro-Ala	Yes
1214089	G	C	SEEM1958_RS06220	phage capsid scaffolding protein	WP_023218238.1	G>C	Glu-Asp	Yes
1216763	T	A	SEEM1958_RS06235	hypothetical protein	WP_023218239.1			
1222504	T	C	SEEM1958_RS06295	phage tail protein I	WP_023218240.1			
2143272	T	C	SEEM1958_RS10880	hypothetical protein	WP_023218578.1	A>G	Gln-Arg	Yes
3616455	C	T	SEEM1958_RS18405	AraC family transcriptional regulator	WP_023219097.1	G>A	Ala-Thr	Yes

Unique SNPs Sub cluster 2C Versus Sub cluster 2D

FDA015 (sub cluster 2C)	1226005	C	T		SEEM1958_RS06315	frameshifted	phage tail protein		A>G	Lys-Arg	Yes
	4448414	A	T		SEEM1958_RS22520		acetyl-coenzyme A synthetase	WP_000083883.1			
	3728054	A	T		SEEM1958_RS18975		hypothetical protein	WP_000145239.1	A>T	His-Leu	Yes
	937099	G	A		SEEM1958_RS04810		prolipoprotein diacylglyceryl transferase	WP_000204647.1	G>A	Met-Ile	Yes
	1132162	T	C	srlA	SEEM1958_RS05760	catalyzes the phosphorylation of incoming sugar substrates along with their translocation across the cell membrane; the IIC domain forms the PTS system translocation channel and contains the specific substrate-binding site	PTS sorbitol transporter subunit IIC	WP_000573333.1	A>G	Asn-Asp	Yes
	4577367	G	A		SEEM1958_RS23040		DNA-directed RNA polymerase subunit beta'	WP_000653965.1	C>T	Ala-Val	Yes
	4202506	G	A		SEEM1958_RS21195		2-dehydro-3-deoxyphosphooctonate aldolase	WP_000779253.1			
	4402875	C	A		SEEM1958_RS22315		AraC family transcriptional regulator	WP_000921676.1	C>A	Asn-Lys	Yes
	1225552	T	C		SEEM1958_RS06310		tail fiber assembly protein	WP_000972188.1	A>G	Gln-Arg	Yes

	304734	C	A		SEEM1958_RS01425		hypothetical protein	WP_021000870.1	G>T	Gly-Val	Yes
	1240217	C	T		SEEM1958_RS06385		VCBS repeat-containing protein	WP_023218145.1	G>A	Val-Ile	Yes
	2666853	A	G		SEEM1958_RS13600		pathogenicity island protein	WP_023218254.1	A>G	His-Arg	Yes
	3959663	G	A		SEEM1958_RS20055		L-carnitine CoA-transferase	WP_023218424.1	G>A	Arg-His	Yes
	623204	T	A		SEEM1958_RS03095		glycerate 2-kinase	WP_023219238.1	T>A	Asp-Glu	Yes
FDA133 (Sub cluster 2D)	303013	C	A		SEEM1958_RS01415		pheromone autoinducer 2 transporter	WP_000179588.1	G>T	Gly-Val	Yes
	2045544	G	A	fliI	SEEM1958_RS10330	involved in type III protein export during flagellum assembly	flagellum-specific ATP synthase	WP_000213257.1			
	2292255	T	A		SEEM1958_RS11650	bifunctional anthranilate synthase II/anthranilate phosphoribosyltransferase; TrpD; forms a heterotetramer with Trp E and the complex catalyzes the formation of anthranilate from chorismate and glutamine; also catalyzes the formation of N-(5-phospho-D-ribosyl)-anthranilate from athranilate and 5-phospho-alpha-D-ribose 1-diphosphate;	bifunctional glutamine amidotransferase/ anthranilate phosphoribosyltransferase	WP_000763492.1	A>T	Asn-Tyr	Yes

						functions in tryptophan biosynthesis					
	3088180	G	A	artM	SEEM1958_RS15760	with ArtPQJI acts to transport arginine across the inner membrane	arginine transporter permease subunit ArtM	WP_000895393.1	G>A	Val-Ile	Yes
	3012404	G	A		SEEM1958_RS15420		chromosome partition protein MukF	WP_001288828.1			
	3143561	A	T		SEEM1958_RS16025		mechanosensitive channel protein	WP_023218008.1	A>T	Ser-Cys	Yes
	3725451	A	G		SEEM1958_RS18970		hypothetical protein	WP_023218471.1	A>G	Gln-Arg	Yes
	4109610	C	T		SEEM1958_RS20745		type I restriction endonuclease subunit M	WP_023218594.1			
	1011720	G	A		SEEM1958_RS05145		assimilatory sulfite reductase (NADPH) flavoprotein subunit	WP_023218864.1			
	2786092	G	A		SEEM1958_RS14210		hypothetical protein	WP_023219026.1			
	2405913	T	C		SEEM1958_RS12235		carboxylesterase	WP_076031760.1	T>C	Ile-Thr	Yes

Unique SNPs Sub cluster 2C Versus Sub cluster 2E

EUR069 (Sub cluster 2E)	2534501	C	T		SEEM1958_RS12855		sugar efflux transporter	WP_000154617.1	C>T	Gln	Yes
	1149094	G	A		SEEM1958_RS05870	DNA-binding transcriptional repressor of microcin B17 synthesis and multidrug efflux; negative regulator of the multidrug operon emrAB	transcriptional regulator	WP_000378431.1			
	3592002	T	A		SEEM1958_RS18305	Confers resistance to fosfomycin and deoxycholate;	MFS transporter	WP_000446768.1	A>T	Asn-Ile	Yes
	3094297	G	A		SEEM1958_RS15795		membrane protein	WP_000505788.1			
	497337	A	T		SEEM1958_RS02435		mononuclear molybdenum enzyme YedY	WP_000723876.1			
	342809	C	A		SEEM1958_RS01620		aspartate-semialdehyde dehydrogenase	WP_000799940.1			
	926076	C	T		SEEM1958_RS04770		bifunctional 2-acylglycerophosphoethanolamine acyltransferase/acyl-ACP synthetase	WP_000896101.1	C>T	Pro-Leu	Yes
	1603990	G	A		SEEM1958_RS24795		hypothetical protein	WP_001121022.1	G>A	Met-Ile	Yes
	3964150	C	T	carB	SEEM1958_RS20080	four CarB-CarA dimers form the carbamoyl phosphate synthetase holoenzyme that catalyzes the production of carbamoyl phosphate; CarB is	carbamoyl phosphate synthase large subunit	WP_023218427.1			

						responsible for the amidotransferase activity				
	2786092	G	A		SEEM1958_RS14210		hypothetical protein	WP_023219026.1			
	2350181	C	T		SEEM1958_RS11960		two-partner secretion translocator ZirT	WP_023227594.1	C>T	Thr-Ile	Yes
	3415888	G	T		SEEM1958_RS17410		integrase	WP_023237847.1	G>T	Met-Ile	Yes
	2512377	C	T		SEEM1958_RS12735		PhoPQ-regulated protein	WP_076031773.1	C>T	Thr-Ile	Yes
FDA015 (Sub cluster 2C)	1226005	C	T		SEEM1958_RS06315	frameshifted	phage tail protein		A>G	Lys-Arg	Yes
	4448414	A	T		SEEM1958_RS22520		acetyl-coenzyme A synthetase	WP_000083883.1			
	3728054	A	T		SEEM1958_RS18975		hypothetical protein	WP_000145239.1	A>T	His-Leu	Yes
	937099	G	A		SEEM1958_RS04810		prolipoprotein diacylglyceryl transferase	WP_000204647.1	G>A	Met-Ile	Yes
	1132162	T	C	srlA	SEEM1958_RS05760	catalyzes the phosphorylation of incoming sugar substrates along with their translocation across the cell membrane; the IIC domain forms the PTS system	PTS sorbitol transporter subunit IIC	WP_000573333.1	A>G	Asn-Asp	Yes

					translocation channel and contains the specific substrate-binding site					
	4577367	G	A	SEEM1958_RS23040		DNA-directed RNA polymerase subunit beta'	WP_000653965.1	C>T	Ala-Val	Yes
	4202506	G	A	SEEM1958_RS21195		2-dehydro-3-deoxyphosphooctonate aldolase	WP_000779253.1			
	1225552	T	C	SEEM1958_RS06310		tail fiber assembly protein	WP_000972188.1	A>G	Gln-Arg	Yes
	304734	C	A	SEEM1958_RS01425		hypothetical protein	WP_021000870.1	G>T	Gly-Val	Yes
	1240217	C	T	SEEM1958_RS06385		VCBS repeat-containing protein	WP_023218145.1	G>A	Val-Ile	Yes
	2666853	A	G	SEEM1958_RS13600		pathogenicity island protein	WP_023218254.1	A>G	His-Arg	Yes
	3959663	G	A	SEEM1958_RS20055		L-carnitine CoA-transferase	WP_023218424.1	G>A	Arg-His	Yes
	623204	T	A	SEEM1958_RS03095		glycerate 2-kinase	WP_023219238.1	T>A	Asp-Glu	Yes
	1386368	G	C	SEEM1958_RS07025		hypothetical protein	WP_076031718.1	G>C	Ala-Pro	Yes
Unique SNPs Sub cluster 2D versus Sub cluster 2E										
EUR069 (Sub cluster 2E)	2534501	C	T	SEEM1958_RS12855		sugar efflux transporter	WP_000154617.1	C>T	Gln	Yes

	1149094	G	A		SEEM1958_RS05870	DNA-binding transcriptional repressor of microcin B17 synthesis and multidrug efflux; negative regulator of the multidrug operon emrAB	transcriptional regulator	WP_000378431.1			
	3592002	T	A		SEEM1958_RS18305	Confers resistance to fosfomycin and deoxycholate	MFS transporter	WP_000446768.1	A>T	Asn-Ile	Yes
	3094297	G	A		SEEM1958_RS15795		membrane protein	WP_000505788.1			
	497337	A	T		SEEM1958_RS02435		mononuclear molybdenum enzyme YedY	WP_000723876.1			
	342809	C	A		SEEM1958_RS01620		aspartate-semialdehyde dehydrogenase	WP_000799940.1			
	926076	C	T		SEEM1958_RS04770		bifunctional 2-acylglycerophosphoethanolamine acyltransferase/acyl-ACP synthetase	WP_000896101.1	C>T	Pro-Leu	Yes
	1603990	G	A		SEEM1958_RS24795		hypothetical protein	WP_001121022.1	G>A	Met-Ile	Yes
	3675600	A	T		SEEM1958_RS18715		hypothetical protein	WP_006499449.1	T>A	Val-Glu	Yes
	3964150	C	T	carB	SEEM1958_RS20080	four CarB-CarA dimers form the carbamoyl phosphate synthetase holoenzyme that catalyzes the production of carbamoyl phosphate; CarB is responsible for the amidotransferase activity	carbamoyl phosphate synthase large subunit	WP_023218427.1			

	3697903	G	T		SEEM1958_RS18815		Rhs family protein	WP_023218486.1			
	1197125	C	T		SEEM1958_RS06100	structural flagella protein; individual Salmonella serotypes usually alternate between the production of 2 antigenic forms of flagella, termed phase 1 and phase 2, each specified by separate structural genes	flagellin FliC	WP_023219298.1			
	2350181	C	T		SEEM1958_RS11960		two-partner secretion translocator ZirT	WP_023227594.1	C>T	Thr-Ile	Yes
	3415888	G	T		SEEM1958_RS17410		integrase	WP_023237847.1	G>T	Met-Ile	Yes
	2512377	C	T		SEEM1958_RS12735		PhoPQ-regulated protein	WP_076031773.1	C>T	Thr-Ile	Yes
FDA133 (Sub cluster 2D)	303013	C	A		SEEM1958_RS01415		pheromone autoinducer 2 transporter	WP_000179588.1	G>T	Gly-Val	Yes
	2045544	G	A	fliI	SEEM1958_RS10330	involved in type III protein export during flagellum assembly	flagellum-specific ATP synthase	WP_000213257.1			

	2292255	T	A		SEEM1958_RS11650	bifunctional anthranilate synthase II/anthranilate phosphoribosyltransferase; TrpD; forms a heterotetramer with Trp E and the complex catalyzes the formation of anthranilate from chorismate and glutamine; also catalyzes the formation of N-(5-phospho-D-ribosyl)-anthranilate from athranilate and 5-phospho-alpha-D-ribose 1-diphosphate; functions in tryptophan biosynthesis	bifunctional glutamine amidotransferase/ anthranilate phosphoribosyltransferase	WP_000763492.1	A>T	Asn-Tyr	Yes
	4572349	G	C		SEEM1958_RS23015		2-iminoacetate synthase ThiH	WP_000847486.1				
	3088180	G	A	artM	SEEM1958_RS15760	with ArtPQJI acts to transport arginine across the inner membrane	arginine transporter permease subunit ArtM	WP_000895393.1	G>A	Val-Ile	Yes
	3012404	G	A		SEEM1958_RS15420		chromosome partition protein MukF	WP_001288828.1				
	3143561	A	T		SEEM1958_RS16025		mechanosensitive channel protein	WP_023218008.1	A>T	Ser-Cys	Yes
	1382471	T	C		SEEM1958_RS07020		hypothetical protein	WP_023218393.1				
	3725451	A	G		SEEM1958_RS18970		hypothetical protein	WP_023218471.1	A>G	Gln-Arg	Yes

	4109610	C	T	SEEM1958_RS20745		type I restriction endonuclease subunit M	WP_023218594.1			
	1011720	G	A	SEEM1958_RS05145		assimilatory sulfite reductase (NADPH) flavoprotein subunit	WP_023218864.1			
	2589168	C	T	SEEM1958_RS13145	allows for ions and hydrophilic solutes to cross the outer membrane	phosphoporin PhoE	WP_023219003.1			
	1195940	T	C	SEEM1958_RS06100	structural flagella protein; individual Salmonella serotypes usually alternate between the production of 2 antigenic forms of flagella, termed phase 1 and phase 2, each specified by separate structural genes	flagellin FliC	WP_023219298.1			
	2405913	T	C	SEEM1958_RS12235		carboxylesterase	WP_076031760.1	T>C	Ile-Thr	Yes

Table 8. Details of proteins present or absent in *S*. Mbandaka isolates in selected protein cluster bins derived from pan genome analysis. Comparative genome content analysis revealed 4701 proteins with known COG function and 2275 with unknown function. Protein cluster bins (PCB) 1 to 13were selected to identify exclusive COGs present in each selected S.Mbandaka isolates. PCB 14 and 15 were selected to identify protein that were found absent in isolate FDA176. Table contains only proteins with known COG function after removing duplicates of proteins similar COG function access number.

Bin name	Protein cluster ID	Genome ID	COG function accession	COG function
PCB1	PC_00006821	EUR_069	COG1192	Cellulose biosynthesis protein BcsQ
	PC_00005656	EUR_069	COG4372	Uncharacterized conserved protein, contains DUF3084 domain
PCB2	PC_00005223	USDA_005	COG4591	ABC-type transport system, involved in lipoprotein release, permease component
	PC_00004853	USDA_005	COG3052	Citrate lyase, gamma subunit
	PC_00004858	USDA_005	COG2197	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains
	PC_00004857	USDA_005	COG2025	Electron transfer flavoprotein, alpha subunit
	PC_00005645	USDA_005	COG0046|COG0047	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain|Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain
	PC_00004864	USDA_005	COG1974	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)
	PC_00005014	USDA_005	COG4694	Wobble nucleotide-excising tRNase
PCB3	PC_00004914	USDA_015	COG1917|COG2207	Cupin domain protein related to quercetin dioxygenase|AraC-type DNA-binding domain and AraC-containing proteins

	PC_00006064	USDA_015	COG0543|COG0633	NAD(P)H-flavin reductase|Ferredoxin
	PC_00004920	USDA_015	COG1762|COG3711	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)|Transcriptional antiterminator
	PC_00005657	USDA_015	COG0515|COG0790|COG0790	Serine/threonine protein kinase|TPR repeat|TPR repeat
	PC_00005040	USDA_015	COG2929	Uncharacterized conserved protein, DUF497 family
PCB4				
PCB5	PC_00005450	USDA_037	COG3727	G:T-mismatch repair DNA endonuclease, very short patch repair protein
	PC_00006870	USDA_037	COG0270	Site-specific DNA-cytosine methylase
	PC_00005491	USDA_037	COG1479	Uncharacterized conserved protein, contains ParB-like and HNH nuclease domains
PCB6	PC_00005962	OTH_007	COG2217|COG3350	Cation transport ATPase|Uncharacterized conserved protein, YHS domain
	PC_00004899	OTH_007	COG2372	Copper-binding protein CopC (methionine-rich)
	PC_00004907	OTH_007	COG0419|COG3593	DNA repair exonuclease SbcCD ATPase subunit|Predicted ATP-dependent endonuclease of the OLD family, contains P-loop ATPase and TOPRIM domains
	PC_00006357	OTH_007	COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain
	PC_00006378	OTH_007	COG2205	K+-sensing histidine kinase KdpD
	PC_00005750	OTH_007	COG2132|COG2132	Multicopper oxidase with three cupredoxin domains (includes cell division protein FtsP and spore coat protein CotA)|Multicopper oxidase with three cupredoxin domains (includes cell division protein FtsP and spore coat protein CotA)
	PC_00006811	OTH_007	COG0845	Multidrug efflux pump subunit AcrA (membrane-fusion protein)

	PC_00006760	OTH_007	COG0739	Murein DD-endopeptidase MepM and murein hydrolase activator NlpD, contain LysM domain
	PC_00006008	OTH_007	COG1538	Outer membrane protein TolC
	PC_00006392	OTH_007	COG5569	Periplasmic Cu and Ag efflux protein CusF
	PC_00006425	OTH_007	COG5525	Phage terminase, large subunit GpA
	PC_00006551	OTH_007	COG3668	Plasmid stabilization system protein ParE
	PC_00005909	OTH_007	COG0270	Site-specific DNA-cytosine methylase
	PC_00005592	OTH_007	COG2801	Transposase InsO and inactivated derivatives
	PC_00005204	OTH_007	COG0286	Type I restriction-modification system, DNA methylase subunit
	PC_00006474	OTH_007	COG3704	Type IV secretory pathway, VirB6 components
	PC_00005324	OTH_007	COG3019	Uncharacterized conserved protein
	PC_00006615	OTH_007	COG3544	Uncharacterized conserved protein, DUF305 family
	PC_00005913	OTH_007	COG3667	Uncharacterized protein involved in copper resistance
PCB7	PC_00004854	ADRDL_27	COG0161	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
	PC_00004855	ADRDL_27	COG0132	Dethiobiotin synthetase
	PC_00004915	ADRDL_27	COG2086	Electron transfer flavoprotein, alpha and beta subunits
	PC_00006737	ADRDL_27	COG0366	Glycosidase
	PC_00006190	ADRDL_27	COG0582	Integrase
	PC_00006273	ADRDL_27	COG0477	MFS family permease
	PC_00004775	ADRDL_27	COG4384	Mu-like prophage protein gp45
	PC_00004836	ADRDL_27	COG4386	Mu-like prophage tail sheath protein gpL
	PC_00006807	ADRDL_27	COG4626	Phage terminase-like protein, large subunit, contains N-terminal HTH domain
	PC_00006305	ADRDL_27	COG5495	Predicted oxidoreductase, contains short-chain dehydrogenase (SDR) and DUF2520 domains
	PC_00006832	ADRDL_27	COG2271	Sugar phosphate permease

	PC_00006633	ADRDL_27	COG3793	Tellurite resistance protein
	PC_00005681	ADRDL_27	COG1063	Threonine dehydrogenase or related Zn-dependent dehydrogenase
	PC_00006492	ADRDL_27	COG3547	Transposase
	PC_00006929	ADRDL_27	COG3316	Transposase (or an inactivated derivative)
	PC_00004717	ADRDL_27	COG1662	Transposase and inactivated derivatives, IS1 family
	PC_00006417	ADRDL_27	COG3451	Type IV secretory pathway, VirB4 component
	PC_00004859	ADRDL_27	COG1881	Uncharacterized conserved protein, phosphatidylethanolamine-binding protein (PEBP) family
PCB8	PC_00006936	EUR_034	COG4227	Antirestriction protein ArdC
	PC_00005055	EUR_034	COG1106	ATPase/GTPase, AAA15 family
	PC_00006181	EUR_034	COG0740	ATP-dependent protease ClpP, protease subunit
	PC_00006971	EUR_034	COG5511	Bacteriophage capsid protein
	PC_00006709	EUR_034	COG4385	Bacteriophage P2-related tail formation protein
	PC_00005527	EUR_034	COG1192	Cellulose biosynthesis protein BcsQ
	PC_00005020	EUR_034	COG1475	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
	PC_00005979	EUR_034	COG0863|COG1475	DNA modification methylase|Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
	PC_00005575	EUR_034	COG0847	DNA polymerase III, epsilon subunit or related 3'-5' exonuclease
	PC_00006348	EUR_034	COG0358	DNA primase (bacterial type)
	PC_00005911	EUR_034	COG2944	DNA-binding transcriptional regulator YiaG, XRE-type HTH domain
	PC_00006344	EUR_034	COG5004	P2-like prophage tail protein X
	PC_00006226	EUR_034	COG4540	Phage P2 baseplate assembly protein gpV
	PC_00006915	EUR_034	COG3500	Phage protein D
	PC_00005927	EUR_034	COG3497	Phage tail sheath protein FI

	PC_00005815	EUR_034	COG3498	Phage tail tube protein FII
	PC_00006196	EUR_034	COG5525	Phage terminase, large subunit GpA
	PC_00006206	EUR_034	COG3948	Phage-related baseplate assembly protein
	PC_00006539	EUR_034	COG5301	Phage-related tail fibre protein
	PC_00005003	EUR_034	COG5283	Phage-related tail protein
	PC_00006829	EUR_034	COG3620	Predicted transcriptional regulator with C-terminal CBS domains
	PC_00006018	EUR_034	COG0270	Site-specific DNA-cytosine methylase
	PC_00005068	EUR_034	COG1974	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)
	PC_00006274	EUR_034	COG1479	Uncharacterized conserved protein, contains ParB-like and HNH nuclease domains
PCB9	PC_00006947	EUR_062	COG1051	ADP-ribose pyrophosphatase YjhB, NUDIX family
	PC_00006457	EUR_062	COG1192	Cellulose biosynthesis protein BcsQ
	PC_00006038	EUR_062	COG1475	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
	PC_00006744	EUR_062	COG2131	Deoxycytidylate deaminase
	PC_00006231	EUR_062	COG0262	Dihydrofolate reductase
	PC_00006252	EUR_062	COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain
	PC_00006630	EUR_062	COG0543|COG0633	NAD(P)H-flavin reductase|Ferredoxin
	PC_00005809	EUR_062	COG3081	Nucleoid-associated protein YejK (function unknown)
	PC_00006688	EUR_062	COG1835	Peptidoglycan/LPS O-acetylase OafA/YrhL, contains acyltransferase and SGNH-hydrolase domains
	PC_00005649	EUR_062	COG3740	Phage head maturation protease
	PC_00006516	EUR_062	COG4695	Phage portal protein BeeE
	PC_00006562	EUR_062	COG5525	Phage terminase, large subunit GpA
	PC_00005403	EUR_062	COG3728	Phage terminase, small subunit

	PC_00004921	EUR_062	COG3668	Plasmid stabilization system protein ParE
	PC_00006311	EUR_062	COG3905	Predicted transcriptional regulator
	PC_00006147	EUR_062	COG1715	Restriction endonuclease Mrr
	PC_00006184	EUR_062	COG1396	Transcriptional regulator, contains XRE-family HTH domain
	PC_00005421	EUR_062	COG3706	Two-component response regulator, PleD family, consists of two REC domains and a diguanylate cyclase (GGDEF) domain
PCB10	PC_00004607	ADRDL_45	COG1426	Cytoskeletal protein RodZ, contains Xre-like HTH and DUF4115 domains
	PC_00004608	ADRDL_45	COG1974	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)
	PC_00005720	FDA_135	COG4227	Antirestriction protein ArdC
	PC_00006864	FDA_135	COG2336	Antitoxin component of the MazEF toxin-antitoxin module
	PC_00004692	FDA_135	COG0507	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I
	PC_00006410	FDA_135	COG1192	Cellulose biosynthesis protein BcsQ
	PC_00005783	FDA_135	COG2916	DNA-binding protein H-NS
	PC_00005721	FDA_135	COG4197	DNA-binding transcriptional regulator YdaS, prophage-encoded, Cro superfamily
	PC_00006319	FDA_135	COG2944	DNA-binding transcriptional regulator YiaG, XRE-type HTH domain
	PC_00004863	FDA_135	COG1344|COG1344	Flagellin and related hook-associated protein FlgL|Flagellin and related hook-associated protein FlgL
	PC_00004934	FDA_135	COG0426	Flavorubredoxin
	PC_00005597	FDA_135	COG2337	mRNA-degrading endonuclease, toxin component of the MazEF toxin-antitoxin module
	PC_00004984	FDA_135	COG0845	Multidrug efflux pump subunit AcrA (membrane-fusion protein)

PC_00004666	FDA_135	COG0697	Permease of the drug/metabolite transporter (DMT) superfamily
PC_00006066	FDA_135	COG2932	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains
PC_00004868	FDA_135	COG5301	Phage-related tail fibre protein
PC_00005132	FDA_135	COG1502	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme
PC_00004872	FDA_135	COG4928	Predicted P-loop ATPase, KAP-like
PC_00005973	FDA_135	COG3091	Predicted Zn-dependent metalloprotease, SprT family
PC_00005469	FDA_135	COG1715	Restriction endonuclease Mrr
PC_00006225	FDA_135	COG1961	Site-specific DNA recombinase related to the DNA invertase Pin
PC_00004648	FDA_135	COG0741	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
PC_00004701	FDA_135	COG0286	Type I restriction-modification system, DNA methylase subunit
PC_00006088	FDA_135	COG0630	Type IV secretory pathway ATPase VirB11/Archaellum biosynthesis ATPase
PC_00005940	FDA_135	COG3736	Type IV secretory pathway, component VirB8
PC_00005181	FDA_135	COG2948	Type IV secretory pathway, VirB10 components
PC_00005454	FDA_135	COG3702	Type IV secretory pathway, VirB3 components
PC_00004905	FDA_135	COG3451	Type IV secretory pathway, VirB4 component
PC_00004753	FDA_135	COG3704	Type IV secretory pathway, VirB6 components
PC_00004679	FDA_135	COG3504	Type IV secretory pathway, VirB9 components
PC_00005667	FDA_135	COG3505	Type IV secretory pathway, VirD4 component, TraG/TraD family ATPase
PC_00006942	FDA_135	COG4737	Uncharacterized protein

PCB11	PC_00006742	ADRDL_45	COG1135	ABC-type methionine transport system, ATPase component
	PC_00006051	ADRDL_45	COG2189	Adenine specific DNA methylase Mod
	PC_00005544	ADRDL_45	COG0814	Amino acid permease
	PC_00006540	ADRDL_45	COG0507	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I
	PC_00005971	ADRDL_45	COG1887	CDP-glycerol glycerophosphotransferase, TagB/SpsB family
	PC_00006603	ADRDL_45	COG1192	Cellulose biosynthesis protein BcsQ
	PC_00005398	ADRDL_45	COG2303	Choline dehydrogenase or related flavoprotein
	PC_00005157	ADRDL_45	COG1196	Chromosome segregation ATPase
	PC_00006949	ADRDL_45	COG1475	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
	PC_00006236	ADRDL_45	COG1309	DNA-binding transcriptional regulator, AcrR family
	PC_00006498	ADRDL_45	COG3855	Fructose-1,6-bisphosphatase
	PC_00006097	ADRDL_45	COG0334	Glutamate dehydrogenase/leucine dehydrogenase
	PC_00006531	ADRDL_45	COG1391	Glutamine synthetase adenylyltransferase
	PC_00004940	ADRDL_45	COG0582|COG4974	Integrase|Site-specific recombinase XerD
	PC_00005482	ADRDL_45	COG1566	Multidrug resistance efflux pump
	PC_00006620	ADRDL_45	COG0036	Pentose-5-phosphate-3-epimerase
	PC_00005142	ADRDL_45	COG3645|COG3646	Phage antirepressor protein YoqD, KilAC domain|Phage regulatory protein Rha
	PC_00004976	ADRDL_45	COG2932	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains

PC_00006683	ADRDL_45	COG2932|COG2932|COG2932	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains|Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains|Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains
PC_00005425	ADRDL_45	COG3772	Phage-related lysozyme (muramidase), GH24 family
PC_00006975	ADRDL_45	COG5281|COG5281	Phage-related minor tail protein|Phage-related minor tail protein
PC_00004959	ADRDL_45	COG5301	Phage-related tail fibre protein
PC_00005416	ADRDL_45	COG2814	Predicted arabinose efflux permease, MFS family
PC_00005203	ADRDL_45	COG2072	Predicted flavoprotein CzcO associated with the cation diffusion facilitator CzcD
PC_00006681	ADRDL_45	COG1598	Predicted nuclease of the RNAse H fold, HicB family
PC_00006063	ADRDL_45	COG3654	Prophage maintenance system killer protein
PC_00004945	ADRDL_45	COG5527	Protein involved in initiation of plasmid replication
PC_00005271	ADRDL_45	COG3598	RecA-family ATPase
PC_00005214	ADRDL_45	COG3611	Replication initiation and membrane attachment protein DnaB
PC_00005516	ADRDL_45	COG3587	Restriction endonuclease
PC_00006574	ADRDL_45	COG1039	Ribonuclease HIII
PC_00005243	ADRDL_45	COG4585	Signal transduction histidine kinase
PC_00005793	ADRDL_45	COG0270	Site-specific DNA-cytosine methylase
PC_00005509	ADRDL_45	COG0741|COG5283	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)|Phage-related tail protein
PC_00006364	ADRDL_45	COG1974	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)
PC_00006304	ADRDL_45	COG3793	Tellurite resistance protein

	PC_00006325	ADRDL_45	COG0675	Transposase
	PC_00006745	ADRDL_45	COG2801	Transposase InsO and inactivated derivatives
	PC_00006806	ADRDL_45	COG3415|COG4373	Transposase|Mu-like prophage FluMu protein gp28
	PC_00006212	ADRDL_45	COG3843	Type IV secretory pathway, VirD2 components (relaxase)
	PC_00005547	ADRDL_45	COG3299	Uncharacterized phage protein gp47/JayE
	PC_00005463	ADRDL_45	COG4834	Uncharacterized protein
	PC_00005910	ADRDL_45	COG3567	Uncharacterized protein
PCB12	PC_00005322	ATCC_51958	COG0179	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)
	PC_00005771	ATCC_51958	COG3481	3'-5' exoribonuclease YhaM, can participate in 23S rRNA maturation, HD superfamily
	PC_00006666	ATCC_51958	COG0066	3-isopropylmalate dehydratase small subunit
	PC_00005079	ATCC_51958	COG0834	ABC-type amino acid transport/signal transduction system, periplasmic component/domain
	PC_00004968	ATCC_51958	COG1132	ABC-type multidrug transport system, ATPase and permease component
	PC_00006604	ATCC_51958	COG1129	ABC-type sugar transport system, ATPase component
	PC_00006461	ATCC_51958	COG1879	ABC-type sugar transport system, periplasmic component, contains N-terminal xre family HTH domain
	PC_00005067	ATCC_51958	COG4569	Acetaldehyde dehydrogenase (acetylating)
	PC_00006068	ATCC_51958	COG0827	Adenine-specific DNA methylase
	PC_00005066	ATCC_51958	COG2015	Alkyl sulfatase BDS1 and related hydrolases, metallo-beta-lactamase superfamily
	PC_00005049	ATCC_51958	COG2721	Altronate dehydratase

PC_00004950	ATCC_51958	COG2336	Antitoxin component of the MazEF toxin-antitoxin module
PC_00006012	ATCC_51958	COG2161	Antitoxin component YafN of the YafNO toxin-antitoxin module, PHD/YefM family
PC_00005855	ATCC_51958	COG2207	AraC-type DNA-binding domain and AraC-containing proteins
PC_00006172	ATCC_51958	COG0542	ATP-dependent Clp protease ATP-binding subunit ClpA
PC_00005148	ATCC_51958	COG0507|COG1112|COG1502	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I|Superfamily I DNA and/or RNA helicase|Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme
PC_00006490	ATCC_51958	COG5295	Autotransporter adhesin
PC_00006133	ATCC_51958	COG5295|COG5295|COG5295	Autotransporter adhesin|Autotransporter adhesin|Autotransporter adhesin
PC_00005012	ATCC_51958	COG1168	Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities
PC_00005061	ATCC_51958	COG0496	Broad specificity polyphosphatase and 5'/3'-nucleotidase SurE
PC_00004994	ATCC_51958	COG3266	Cell division protein DamX, binds to the septal ring, contains C-terminal SPOR domain
PC_00005114	ATCC_51958	COG1192	Cellulose biosynthesis protein BcsQ
PC_00004997	ATCC_51958	COG1196|COG3593	Chromosome segregation ATPase|Predicted ATP-dependent endonuclease of the OLD family, contains P-loop ATPase and TOPRIM domains

PC_00004929	ATCC_51958	COG1475	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
PC_00006682	ATCC_51958	COG2096|COG3193	Cob(I)alamin adenosyltransferase|Uncharacterized conserved protein GlcG, DUF336 family
PC_00005141	ATCC_51958	COG1203	CRISPR/Cas system-associated endonuclease/helicase Cas3
PC_00005461	ATCC_51958	COG1917	Cupin domain protein related to quercetin dioxygenase
PC_00005242	ATCC_51958	COG1064	D-arabinose 1-dehydrogenase, Zn-dependent alcohol dehydrogenase family
PC_00006792	ATCC_51958	COG0329	Dihydrodipicolinate synthase/N-acetylneuraminate lyase
PC_00006902	ATCC_51958	COG0847	DNA polymerase III, epsilon subunit or related 3'-5' exonuclease
PC_00005791	ATCC_51958	COG2003	DNA repair protein RadC, contains a helix-hairpin-helix DNA-binding motif
PC_00005198	ATCC_51958	COG3279	DNA-binding response regulator, LytR/AlgR family
PC_00006421	ATCC_51958	COG0745	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
PC_00005048	ATCC_51958	COG2390	DNA-binding transcriptional regulator LsrR, DeoR family
PC_00005369	ATCC_51958	COG1349	DNA-binding transcriptional regulator of sugar metabolism, DeoR/GlpR family
PC_00006553	ATCC_51958	COG2186	DNA-binding transcriptional regulator, FadR family
PC_00005437	ATCC_51958	COG1414	DNA-binding transcriptional regulator, IclR family
PC_00005554	ATCC_51958	COG0583	DNA-binding transcriptional regulator, LysR family

PC_00006000	ATCC_5195 8	COG1167	DNA-binding transcriptional regulator, MocR family, contains an aminotransferase domain
PC_00005789	ATCC_5195 8	COG0251	Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family
PC_00006381	ATCC_5195 8	COG0695	Glutaredoxin
PC_00006322	ATCC_5195 8	COG1100|COG4886	GTPase SAR1 family domain|Leucine-rich repeat (LRR) protein
PC_00005358	ATCC_5195 8	COG0561	Hydroxymethylpyrimidine pyrophosphatase and other HAD family phosphatases
PC_00005159	ATCC_5195 8	COG0582	Integrase
PC_00005080	ATCC_5195 8	COG0119	Isopropylmalate/homocitrate/citramalate synthases
PC_00005103	ATCC_5195 8	COG2205	K+-sensing histidine kinase KdpD
PC_00006146	ATCC_5195 8	COG4948	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily
PC_00005650	ATCC_5195 8	COG0662	Mannose-6-phosphate isomerase, cupin superfamily
PC_00005028	ATCC_5195 8	COG4373	Mu-like prophage FluMu protein gp28
PC_00006525	ATCC_5195 8	COG4388	Mu-like prophage I protein
PC_00005337	ATCC_5195 8	COG4382	Mu-like prophage protein gp16
PC_00005240	ATCC_5195 8	COG4383	Mu-like prophage protein gp29
PC_00006298	ATCC_5195 8	COG4387	Mu-like prophage protein gp36

PC_00006905	ATCC_51958	COG5005	Mu-like prophage protein gpG
PC_00005453	ATCC_51958	COG1566	Multidrug resistance efflux pump
PC_00005966	ATCC_51958	COG0739|COG1388|COG3773	Murein DD-endopeptidase MepM and murein hydrolase activator NlpD, contain LysM domain|LysM repeat|Cell wall hydrolase CwlJ, involved in spore germination
PC_00005784	ATCC_51958	COG1028	NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family
PC_00004866	ATCC_51958	COG2110	O-acetyl-ADP-ribose deacetylase (regulator of RNase III), contains Macro domain
PC_00006092	ATCC_51958	COG3121	P pilus assembly protein, chaperone PapD
PC_00006735	ATCC_51958	COG0036	Pentose-5-phosphate-3-epimerase
PC_00006271	ATCC_51958	COG0697	Permease of the drug/metabolite transporter (DMT) superfamily
PC_00006954	ATCC_51958	COG3499	Phage protein U
PC_00006156	ATCC_51958	COG3772	Phage-related lysozyme (muramidase), GH24 family
PC_00006962	ATCC_51958	COG5301	Phage-related tail fibre protein
PC_00005691	ATCC_51958	COG5283	Phage-related tail protein
PC_00005859	ATCC_51958	COG1502	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme
PC_00005590	ATCC_51958	COG1263	Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific

PC_00005187	ATCC_5195 8	COG1762	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
PC_00005186	ATCC_5195 8	COG3414	Phosphotransferase system, galactitol-specific IIB component
PC_00005231	ATCC_5195 8	COG3539	Pilin (type 1 fimbria component protein)
PC_00006589	ATCC_5195 8	COG0596	Pimeloyl-ACP methyl ester carboxylesterase
PC_00005019	ATCC_5195 8	COG3668	Plasmid stabilization system protein ParE
PC_00005069	ATCC_5195 8	COG4938	Predicted ATPase
PC_00004942	ATCC_5195 8	COG3522	Predicted component of the type VI protein secretion system
PC_00005351	ATCC_5195 8	COG3516	Predicted component of the type VI protein secretion system
PC_00005440	ATCC_5195 8	COG3518	Predicted component of the type VI protein secretion system
PC_00005636	ATCC_5195 8	COG3521	Predicted component of the type VI protein secretion system
PC_00005775	ATCC_5195 8	COG3515	Predicted component of the type VI protein secretion system
PC_00006106	ATCC_5195 8	COG3517	Predicted component of the type VI protein secretion system
PC_00006306	ATCC_5195 8	COG3520	Predicted component of the type VI protein secretion system
PC_00006728	ATCC_5195 8	COG3456	Predicted component of the type VI protein secretion system, contains a FHA domain

PC_00006167	ATCC_5195 8	COG3311	Predicted DNA-binding transcriptional regulator AlpA
PC_00006323	ATCC_5195 8	COG5635	Predicted NTPase, NACHT family domain
PC_00006229	ATCC_5195 8	COG1487	Predicted nucleic acid-binding protein, contains PIN domain
PC_00006621	ATCC_5195 8	COG0667	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
PC_00005180	ATCC_5195 8	COG3620	Predicted transcriptional regulator with C-terminal CBS domains
PC_00006708	ATCC_5195 8	COG2964	Predicted transcriptional regulator YheO, contains PAS and DNA-binding HTH domains
PC_00005006	ATCC_5195 8	COG2865	Predicted transcriptional regulator, contains HTH domain
PC_00005308	ATCC_5195 8	COG1359	Quinol monooxygenase YgiN
PC_00004923	ATCC_5195 8	COG0732|COG0732	Restriction endonuclease S subunit|Restriction endonuclease S subunit
PC_00006403	ATCC_5195 8	COG1172	Ribose/xylose/arabinose/galactoside ABC-type transport system, permease component
PC_00005654	ATCC_5195 8	COG0257	Ribosomal protein L36
PC_00005371	ATCC_5195 8	COG1734	RNA polymerase-binding transcription factor DksA
PC_00006314	ATCC_5195 8	COG3550	Serine/threonine protein kinase HipA, toxin component of the HipAB toxin-antitoxin module
PC_00004856	ATCC_5195 8	COG5002	Signal transduction histidine kinase
PC_00005269	ATCC_5195 8	COG0338	Site-specific DNA-adenine methylase

PC_00004896	ATCC_5195 8	COG4974	Site-specific recombinase XerD
PC_00005042	ATCC_5195 8	COG1070	Sugar (pentulose or hexulose) kinase
PC_00005150	ATCC_5195 8	COG2271	Sugar phosphate permease
PC_00005389	ATCC_5195 8	COG0553|COG0553	Superfamily II DNA or RNA helicase, SNF2 family|Superfamily II DNA or RNA helicase, SNF2 family
PC_00005151	ATCC_5195 8	COG1063	Threonine dehydrogenase or related Zn-dependent dehydrogenase
PC_00005872	ATCC_5195 8	COG4115	Toxin component of the Txe-Axe toxin-antitoxin module, Txe/YoeB family
PC_00005419	ATCC_5195 8	COG4584	Transposase
PC_00006074	ATCC_5195 8	COG2963	Transposase and inactivated derivatives
PC_00004900	ATCC_5195 8	COG2801	Transposase InsO and inactivated derivatives
PC_00006532	ATCC_5195 8	COG3706	Two-component response regulator, PleD family, consists of two REC domains and a diguanylate cyclase (GGDEF) domain
PC_00004922	ATCC_5195 8	COG0286	Type I restriction-modification system, DNA methylase subunit
PC_00005220	ATCC_5195 8	COG0610	Type I site-specific restriction-modification system, R (restriction) subunit and related helicases ...
PC_00006367	ATCC_5195 8	COG3267	Type II secretory pathway, component ExeA (predicted ATPase)
PC_00005466	ATCC_5195 8	COG3157	Type VI protein secretion system component Hcp (secreted cytotoxin)

	PC_00005704	ATCC_51958	COG3519	Type VI protein secretion system component VasA
	PC_00005339	ATCC_51958	COG3455	Type VI protein secretion system component VasF
	PC_00004882	ATCC_51958	COG3523	Type VI protein secretion system component VasK
	PC_00005406	ATCC_51958	COG3943	Uncharacterized conserved protein
	PC_00004860	ATCC_51958	COG3209|COG3209|COG4104	Uncharacterized conserved protein RhaS, contains 28 RHS repeats|Uncharacterized conserved protein RhaS, contains 28 RHS repeats|Zn-binding Pro-Ala-Ala-Arg (PAAR) domain, incolved in TypeVI secretion
	PC_00005707	ATCC_51958	COG1479	Uncharacterized conserved protein, contains ParB-like and HNH nuclease domains
	PC_00004998	ATCC_51958	COG2369	Uncharacterized conserved protein, contains phage Mu gpF-like domain
	PC_00005879	ATCC_51958	COG3501	Uncharacterized conserved protein, implicated in type VI secretion and phage assembly
	PC_00006749	ATCC_51958	COG4643|COG5519	Uncharacterized domain associated with phage/plasmid primase|Uncharcterized protein, DUF927 family
	PC_00005176	ATCC_51958	COG3477	Uncharacterized membrane protein YagU, involved in acid resistance, DUF1440 family
	PC_00006025	ATCC_51958	COG0730	Uncharacterized membrane protein YfcA
	PC_00006841	ATCC_51958	COG4705	Uncharacterized membrane-anchored protein
	PC_00005633	ATCC_51958	COG5351	Uncharacterized protein
	PC_00006019	ATCC_51958	COG4456	Virulence-associated protein VagC (function unknown)

PCB13	PC_00005790	FDA_176	COG3481	3'-5' exoribonuclease YhaM, can participate in 23S rRNA maturation, HD superfamily
	PC_00004893	FDA_176	COG1995	4-hydroxy-L-threonine phosphate dehydrogenase PdxA
	PC_00006892	FDA_176	COG2977	4'-phosphopantetheinyl transferase EntD (siderophore biosynthesis)
	PC_00005587	FDA_176	COG1896	5'-deoxynucleotidase YfbR and related HD superfamily hydrolases
	PC_00005630	FDA_176	COG1403	5-methylcytosine-specific restriction endonuclease McrA
	PC_00006772	FDA_176	COG0834	ABC-type amino acid transport/signal transduction system, periplasmic component/domain
	PC_00006637	FDA_176	COG2884	ABC-type ATPase involved in cell division
	PC_00004944	FDA_176	COG2274	ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain
	PC_00006414	FDA_176	COG3842	ABC-type Fe3+/spermidine/putrescine transport systems, ATPase components
	PC_00005581	FDA_176	COG1653	ABC-type glycerol-3-phosphate transport system, periplasmic component
	PC_00005525	FDA_176	COG0395	ABC-type glycerol-3-phosphate transport system, permease component
	PC_00006655	FDA_176	COG1121	ABC-type Mn2+/Zn2+ transport system, ATPase component
	PC_00005924	FDA_176	COG1108	ABC-type Mn2+/Zn2+ transport system, permease component
	PC_00005228	FDA_176	COG1132	ABC-type multidrug transport system, ATPase and permease component
	PC_00006812	FDA_176	COG1131	ABC-type multidrug transport system, ATPase component
	PC_00005899	FDA_176	COG3638	ABC-type phosphate/phosphonate transport system, ATPase component

PC_00006602	FDA_176	COG3221	ABC-type phosphate/phosphonate transport system, periplasmic component
PC_00005319	FDA_176	COG3639	ABC-type phosphate/phosphonate transport system, permease component
PC_00006900	FDA_176	COG4107	ABC-type phosphonate transport system, ATPase component
PC_00005517	FDA_176	COG1682	ABC-type polysaccharide/polyol phosphate export permease
PC_00005573	FDA_176	COG1134	ABC-type polysaccharide/polyol phosphate transport system, ATPase component
PC_00005695	FDA_176	COG4618	ABC-type protease/lipase transport system, ATPase and permease components
PC_00006071	FDA_176	COG1129	ABC-type sugar transport system, ATPase component
PC_00005708	FDA_176	COG1879	ABC-type sugar transport system, periplasmic component, contains N-terminal xre family HTH domain
PC_00005841	FDA_176	COG1175	ABC-type sugar transport system, permease component
PC_00006142	FDA_176	COG0803	ABC-type Zn uptake system ZnuABC, Zn-binding component ZnuA
PC_00006116	FDA_176	COG0028	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme
PC_00006194	FDA_176	COG0183	Acetyl-CoA acetyltransferase
PC_00005559	FDA_176	COG4799	Acetyl-CoA carboxylase, carboxyltransferase component
PC_00005352	FDA_176	COG0110	Acetyltransferase (isoleucine patch superfamily)

PC_00006943	FDA_176	COG0236|COG0500|COG1020|COG1020|COG1028|COG2226|COG3319|COG3321	Acyl carrier protein|SAM-dependent methyltransferase|Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family|Ubiquinone/menaquinone biosynthesis C-methylase UbiE|Thioesterase domain of type I polyketide synthase or non-ribosomal peptide synthetase|Acyl transferase domain in polyketide synthase (PKS) enzymes
PC_00005027	FDA_176	COG4670	Acyl CoA:acetate/3-ketoacid CoA transferase
PC_00006845	FDA_176	COG2030	Acyl dehydratase
PC_00004889	FDA_176	COG0427	Acyl-CoA hydrolase
PC_00006385	FDA_176	COG2189	Adenine specific DNA methylase Mod
PC_00006736	FDA_176	COG0827	Adenine-specific DNA methylase
PC_00005448	FDA_176	COG1051	ADP-ribose pyrophosphatase YjhB, NUDIX family
PC_00006611	FDA_176	COG1454	Alcohol dehydrogenase, class IV
PC_00005914	FDA_176	COG3627	Alpha-D-ribose 1-methylphosphonate 5-phosphate C-P lyase
PC_00006241	FDA_176	COG3454	Alpha-D-ribose 1-methylphosphonate 5-triphosphate diphosphatase PhnM
PC_00006163	FDA_176	COG3624	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnG
PC_00006542	FDA_176	COG3625	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnH
PC_00005345	FDA_176	COG3626	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnI
PC_00006042	FDA_176	COG4778	Alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL

PC_00006782	FDA_176	COG1501	Alpha-glucosidase, glycosyl hydrolase family GH31
PC_00006339	FDA_176	COG2721	Altronate dehydratase
PC_00006255	FDA_176	COG0531	Amino acid transporter
PC_00005986	FDA_176	COG0147	Anthranilate/para-aminobenzoate synthases component I
PC_00004906	FDA_176	COG4227	Antirestriction protein ArdC
PC_00005336	FDA_176	COG2207	AraC-type DNA-binding domain and AraC-containing proteins
PC_00004951	FDA_176	COG2207|COG4936	AraC-type DNA-binding domain and AraC-containing proteins|Ligand-binding sensor domain
PC_00005566	FDA_176	COG0433	Archaeal DNA helicase HerA or a related bacterial ATPase, contains HAS-barrel and ATPase domains
PC_00006873	FDA_176	COG1438	Arginine repressor
PC_00006473	FDA_176	COG0137	Argininosuccinate synthase
PC_00006138	FDA_176	COG3119	Arylsulfatase A or related enzyme
PC_00006635	FDA_176	COG5295|COG5295	Autotransporter adhesin|Autotransporter adhesin
PC_00005606	FDA_176	COG2911|COG3468	Autotransporter translocation and assembly factor TamB|Type V secretory pathway, adhesin AidA
PC_00006270	FDA_176	COG5614	Bacteriophage head-tail adaptor
PC_00006755	FDA_176	COG1874	Beta-galactosidase GanA
PC_00006434	FDA_176	COG3250	Beta-galactosidase/beta-glucuronidase
PC_00006411	FDA_176	COG0637	Beta-phosphoglucomutase or related phosphatase, HAD superfamily
PC_00006057	FDA_176	COG2931	Ca2+-binding protein, RTX toxin-related
PC_00005496	FDA_176	COG0664	cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein kinases
PC_00005302	FDA_176	COG0346	Catechol 2,3-dioxygenase or other lactoylglutathione lyase family enzyme

PC_00006526	FDA_176	COG3266	Cell division protein DamX, binds to the septal ring, contains C-terminal SPOR domain
PC_00005954	FDA_176	COG0791|COG1310	Cell wall-associated hydrolase, NlpC family|Proteasome lid subunit RPN8/RPN11, contains Jab1/MPN domain metalloenzyme (JAMM) motif
PC_00006058	FDA_176	COG1196	Chromosome segregation ATPase
PC_00005478	FDA_176	COG1196|COG5281|COG5281	Chromosome segregation ATPase|Phage-related minor tail protein|Phage-related minor tail protein
PC_00004909	FDA_176	COG1475	Chromosome segregation protein Spo0J, contains ParB-like nuclease domain
PC_00006592	FDA_176	COG3052	Citrate lyase, gamma subunit
PC_00005881	FDA_176	COG2032	Cu/Zn superoxide dismutase
PC_00006505	FDA_176	COG2124	Cytochrome P450
PC_00006707	FDA_176	COG1181	D-alanine-D-alanine ligase and related ATP-grasp enzymes
PC_00006031	FDA_176	COG0794	D-arabinose 5-phosphate isomerase GutQ
PC_00005144	FDA_176	COG0471|COG0471	Di- and tricarboxylate transporter|Di- and tricarboxylate transporter
PC_00005235	FDA_176	COG2376	Dihydroxyacetone kinase
PC_00006504	FDA_176	COG0847	DNA polymerase III, epsilon subunit or related 3'-5' exonuclease
PC_00005396	FDA_176	COG0358	DNA primase (bacterial type)
PC_00005113	FDA_176	COG2003	DNA repair protein RadC, contains a helix-hairpin-helix DNA-binding motif
PC_00004886	FDA_176	COG1484	DNA replication protein DnaC
PC_00005164	FDA_176	COG3636	DNA-binding prophage protein
PC_00005343	FDA_176	COG3279	DNA-binding response regulator, LytR/AlgR family
PC_00005043	FDA_176	COG2197	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains

PC_00004870	FDA_176	COG2390	DNA-binding transcriptional regulator LsrR, DeoR family
PC_00005439	FDA_176	COG1349	DNA-binding transcriptional regulator of sugar metabolism, DeoR/GlpR family
PC_00005056	FDA_176	COG4197	DNA-binding transcriptional regulator YdaS, prophage-encoded, Cro superfamily
PC_00005629	FDA_176	COG2944	DNA-binding transcriptional regulator YiaG, XRE-type HTH domain
PC_00005715	FDA_176	COG1309	DNA-binding transcriptional regulator, AcrR family
PC_00006891	FDA_176	COG0640	DNA-binding transcriptional regulator, ArsR family
PC_00006219	FDA_176	COG2186	DNA-binding transcriptional regulator, FadR family
PC_00004943	FDA_176	COG1802	DNA-binding transcriptional regulator, GntR family
PC_00005368	FDA_176	COG2188	DNA-binding transcriptional regulator, GntR family
PC_00005937	FDA_176	COG1959	DNA-binding transcriptional regulator, IscR family
PC_00004918	FDA_176	COG1609	DNA-binding transcriptional regulator, LacI/PurR family
PC_00005160	FDA_176	COG0583	DNA-binding transcriptional regulator, LysR family
PC_00006765	FDA_176	COG1846	DNA-binding transcriptional regulator, MarR family
PC_00004916	FDA_176	COG0789|COG1192	DNA-binding transcriptional regulator, MerR family|Cellulose biosynthesis protein BcsQ
PC_00005709	FDA_176	COG1167	DNA-binding transcriptional regulator, MocR family, contains an aminotransferase domain
PC_00005155	FDA_176	COG1737	DNA-binding transcriptional regulator, MurR/RpiR family, contains HTH and SIS domains
PC_00004898	FDA_176	COG1476	DNA-binding transcriptional regulator, XRE-family HTH domain
PC_00005557	FDA_176	COG3710	DNA-binding winged helix-turn-helix (wHTH) domain
PC_00005062	FDA_176	COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain

PC_00006409	FDA_176	COG2200	EAL domain, c-di-GMP-specific phosphodiesterase class I (or its enzymatically inactive variant)
PC_00006719	FDA_176	COG0251|COG0251|COG0251|COG0251	Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family|Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family|Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family|Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family
PC_00005301	FDA_176	COG0619	Energy-coupling factor transporter transmembrane protein EcfT
PC_00005224	FDA_176	COG0148	Enolase
PC_00006827	FDA_176	COG1024	Enoyl-CoA hydratase/carnithine racemase
PC_00005138	FDA_176	COG1073	Fermentation-respiration switch protein FrsA, has esterase activity, DUF1100 family
PC_00006214	FDA_176	COG1011	FMN phosphatase YigB, HAD superfamily
PC_00005117	FDA_176	COG2222	Fructoselysine-6-P-deglycase FrlB and related proteins with duplicated sugar isomerase (SIS) domain
PC_00005958	FDA_176	COG0076	Glutamate or tyrosine decarboxylase or a related PLP-dependent protein

PC_00005842	FDA_176	COG0001|COG0236|COG0236|COG0318|COG1020|COG1020|COG1020|COG1020|COG2141|COG3321	Glutamate-1-semialdehyde aminotransferase|Acyl carrier protein|Acyl carrier protein|Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II|Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Flavin-dependent oxidoreductase, luciferase family (includes alkanesulfonate monooxygenase SsuD and methylene tetrahydromethanopterin reductase)|Acyl transferase domain in polyketide synthase (PKS) enzymes
PC_00006838	FDA_176	COG0625	Glutathione S-transferase
PC_00005659	FDA_176	COG2379	Glycerate-2-kinase
PC_00005826	FDA_176	COG0366	Glycosidase
PC_00006168	FDA_176	COG0438	Glycosyltransferase involved in cell wall bisynthesis
PC_00006793	FDA_176	COG0463	Glycosyltransferase involved in cell wall bisynthesis
PC_00005551	FDA_176	COG1216	Glycosyltransferase, GT2 family
PC_00006819	FDA_176	COG2610	H+/gluconate symporter or related permease
PC_00005843	FDA_176	COG0672	High-affinity Fe2+/Pb2+ permease
PC_00006535	FDA_176	COG0561	Hydroxymethylpyrimidine pyrophosphatase and other HAD family phosphatases
PC_00005942	FDA_176	COG0134	Indole-3-glycerol phosphate synthase
PC_00004978	FDA_176	COG0582	Integrase
PC_00006768	FDA_176	COG1416	Intracellular sulfur oxidation protein, DsrE/DsrF family
PC_00005907	FDA_176	COG3385	IS4 transposase
PC_00005845	FDA_176	COG0473|COG1058	Isocitrate/isopropylmalate dehydrogenase|Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA

PC_00005916	FDA_176	COG0119	Isopropylmalate/homocitrate/citramalate synthases
PC_00005717	FDA_176	COG3210|COG5295	Large exoprotein involved in heme utilization or adhesion|Autotransporter adhesin
PC_00005376	FDA_176	COG3210|COG3468	Large exoprotein involved in heme utilization or adhesion|Type V secretory pathway, adhesin AidA
PC_00005540	FDA_176	COG3210|COG4932|COG4932	Large exoprotein involved in heme utilization or adhesion|Uncharacterized surface anchored protein|Uncharacterized surface anchored protein
PC_00005314	FDA_176	COG4886	Leucine-rich repeat (LRR) protein
PC_00006059	FDA_176	COG4886|COG4886	Leucine-rich repeat (LRR) protein|Leucine-rich repeat (LRR) protein
PC_00004904	FDA_176	COG1452	LPS assembly outer membrane protein LptD (organic solvent tolerance protein OstA)
PC_00005418	FDA_176	COG3486	Lysine/ornithine N-monooxygenase
PC_00005483	FDA_176	COG0331	Malonyl CoA-acyl carrier protein transacylase
PC_00005399	FDA_176	COG4580	Maltoporin (phage lambda and maltose receptor)
PC_00006448	FDA_176	COG4668	Mannitol/fructose-specific phosphotransferase system, IIA domain
PC_00005538	FDA_176	COG0246	Mannitol-1-phosphate/altronate dehydrogenases
PC_00006554	FDA_176	COG3064	Membrane protein involved in colicin uptake
PC_00006377	FDA_176	COG2244	Membrane protein involved in the export of O-antigen and teichoic acid
PC_00006259	FDA_176	COG1884|COG2185	Methylmalonyl-CoA mutase, N-terminal domain/subunit|Methylmalonyl-CoA mutase, C-terminal domain/subunit (cobalamin-binding)
PC_00004885	FDA_176	COG0477	MFS family permease
PC_00006007	FDA_176	COG0477|COG2814	MFS family permease|Predicted arabinose efflux permease, MFS family

PC_00006124	FDA_176	COG4675	Microcystin-dependent protein (function unknown)
PC_00006555	FDA_176	COG4381	Mu-like prophage protein gp46
PC_00005305	FDA_176	COG5005	Mu-like prophage protein gpG
PC_00004981	FDA_176	COG4379	Mu-like prophage tail protein gpP
PC_00005390	FDA_176	COG0845	Multidrug efflux pump subunit AcrA (membrane-fusion protein)
PC_00006550	FDA_176	COG0841	Multidrug efflux pump subunit AcrB
PC_00006121	FDA_176	COG1757	Na+/H+ antiporter NhaC
PC_00005109	FDA_176	COG2211	Na+/melibiose symporter or related transporter
PC_00005467	FDA_176	COG3023	N-acetyl-anhydromuramyl-L-alanine amidase AmpD
PC_00005730	FDA_176	COG1820	N-acetylglucosamine-6-phosphate deacetylase
PC_00005537	FDA_176	COG3055	N-acetylneuraminic acid mutarotase
PC_00005281	FDA_176	COG1028	NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family
PC_00005128	FDA_176	COG1252	NADH dehydrogenase, FAD-containing subunit
PC_00004939	FDA_176	COG1853	NADH-FMN oxidoreductase RutF, flavin reductase (DIM6/NTAB) family
PC_00005712	FDA_176	COG4221	NADP-dependent 3-hydroxy acid dehydrogenase YdfG
PC_00005767	FDA_176	COG2375	NADPH-dependent ferric siderophore reductase, contains FAD-binding and SIP domains
PC_00006078	FDA_176	COG1834	N-Dimethylarginine dimethylaminohydrolase
PC_00006351	FDA_176	COG1208	NDP-sugar pyrophosphorylase, includes eIF-2Bgamma, eIF-2Bepsilon, and LPS biosynthesis proteins
PC_00005626	FDA_176	COG4409	Neuraminidase (sialidase)
PC_00005508	FDA_176	COG1335	Nicotinamidase-related amidase
PC_00005601	FDA_176	COG2223	Nitrate/nitrite transporter NarK
PC_00005082	FDA_176	COG1021	Non-ribosomal peptide synthetase component E (peptide arylation enzyme)

PC_00006388	FDA_176	COG1020|COG1020|COG1020|COG2226|COG3433	Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Ubiquinone/menaquinone biosynthesis C-methylase UbiE|Aryl carrier domain
PC_00006712	FDA_176	COG1020|COG1020|COG2175	Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Taurine dioxygenase, alpha-ketoglutarate-dependent
PC_00005229	FDA_176	COG1020|COG1020|COG3320	Non-ribosomal peptide synthetase component F|Non-ribosomal peptide synthetase component F|Thioester reductase domain of alpha aminoadipate reductase Lys2 and NRPSs
PC_00006723	FDA_176	COG0451	Nucleoside-diphosphate-sugar epimerase
PC_00006940	FDA_176	COG0589	Nucleotide-binding universal stress protein, UspA family
PC_00005052	FDA_176	COG2823	Osmotically-inducible protein OsmY, contains BON domain
PC_00005769	FDA_176	COG4206	Outer membrane cobalamin receptor protein
PC_00005110	FDA_176	COG4571	Outer membrane protease
PC_00006170	FDA_176	COG3203	Outer membrane protein (porin)
PC_00006100	FDA_176	COG2885	Outer membrane protein OmpA and related peptidoglycan-associated (lipo)proteins
PC_00005465	FDA_176	COG2885|COG2913	Outer membrane protein OmpA and related peptidoglycan-associated (lipo)proteins|Outer membrane protein assembly factor BamE, lipoprotein component of the BamABCDE complex
PC_00005089	FDA_176	COG1538	Outer membrane protein TolC
PC_00005837	FDA_176	COG4771	Outer membrane receptor for ferrienterochelin and colicins

PC_00006910	FDA_176	COG4774	Outer membrane receptor for monomeric catechols
PC_00005991	FDA_176	COG0729	Outer membrane translocation and assembly module TamA
PC_00006032	FDA_176	COG4693	Oxidoreductase (NAD-binding), involved in siderophore biosynthesis
PC_00005833	FDA_176	COG3121	P pilus assembly protein, chaperone PapD
PC_00005362	FDA_176	COG0036	Pentose-5-phosphate-3-epimerase
PC_00006405	FDA_176	COG3409|COG3409	Peptidoglycan-binding (PGRP) domain of peptidoglycan hydrolases|Peptidoglycan-binding (PGRP) domain of peptidoglycan hydrolases
PC_00005099	FDA_176	COG0810	Periplasmic protein TonB, links inner and outer membranes
PC_00006897	FDA_176	COG0616	Periplasmic serine protease, ClpP class
PC_00006145	FDA_176	COG0265	Periplasmic serine protease, S1-C subfamily, contain C-terminal PDZ domain
PC_00006513	FDA_176	COG0697	Permease of the drug/metabolite transporter (DMT) superfamily
PC_00005868	FDA_176	COG3628	Phage baseplate assembly protein W
PC_00005091	FDA_176	COG3740	Phage head maturation protease
PC_00005085	FDA_176	COG4540	Phage P2 baseplate assembly protein gpV
PC_00004873	FDA_176	COG4695	Phage portal protein BeeE
PC_00005317	FDA_176	COG3500	Phage protein D
PC_00005177	FDA_176	COG3646	Phage regulatory protein Rha
PC_00005256	FDA_176	COG2932	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains
PC_00005898	FDA_176	COG1842	Phage shock protein A
PC_00005426	FDA_176	COG5362|COG5410	Phage terminase large subunit|Uncharacterized protein
PC_00006538	FDA_176	COG5525	Phage terminase, large subunit GpA

PC_00005268	FDA_176	COG3747	Phage terminase, small subunit
PC_00006320	FDA_176	COG4626	Phage terminase-like protein, large subunit, contains N-terminal HTH domain
PC_00006491	FDA_176	COG3772	Phage-related lysozyme (muramidase), GH24 family
PC_00006487	FDA_176	COG4679	Phage-related protein
PC_00006937	FDA_176	COG4733	Phage-related protein, tail component
PC_00005402	FDA_176	COG4733|COG4932	Phage-related protein, tail component|Uncharacterized surface anchored protein
PC_00005137	FDA_176	COG5301	Phage-related tail fibre protein
PC_00005253	FDA_176	COG5283	Phage-related tail protein
PC_00006232	FDA_176	COG5283|COG5412	Phage-related tail protein|Phage-related protein
PC_00006107	FDA_176	COG1942	Phenylpyruvate tautomerase PptA, 4-oxalocrotonate tautomerase family
PC_00006613	FDA_176	COG5083	Phosphatidate phosphatase PAH1, contains Lipin and LNS2 domains. can be involved in plasmid maintenance
PC_00004936	FDA_176	COG1080|COG1762|COG1925	Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)|Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)|Phosphotransferase system, HPr and related phosphotransfer proteins
PC_00005631	FDA_176	COG1235	Phosphoribosyl 1,2-cyclic phosphodiesterase
PC_00006333	FDA_176	COG1447	Phosphotransferase system cellobiose-specific component IIA
PC_00005153	FDA_176	COG1440	Phosphotransferase system cellobiose-specific component IIB
PC_00004878	FDA_176	COG1455	Phosphotransferase system cellobiose-specific component IIC

PC_00005328	FDA_176	COG1445	Phosphotransferase system fructose-specific component IIB
PC_00006815	FDA_176	COG1263|COG1264	Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific|Phosphotransferase system IIB components
PC_00006530	FDA_176	COG1762	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
PC_00005568	FDA_176	COG1299	Phosphotransferase system, fructose-specific IIC component
PC_00004874	FDA_176	COG1299|COG1445	Phosphotransferase system, fructose-specific IIC component|Phosphotransferase system fructose-specific component IIB
PC_00005323	FDA_176	COG1925|COG4668	Phosphotransferase system, HPr and related phosphotransfer proteins|Mannitol/fructose-specific phosphotransferase system, IIA domain
PC_00005693	FDA_176	COG2213	Phosphotransferase system, mannitol-specific IIBC component
PC_00005124	FDA_176	COG3444	Phosphotransferase system, mannose/fructose/N-acetylgalactosamine-specific component IIB
PC_00005292	FDA_176	COG2893	Phosphotransferase system, mannose/fructose-specific component IIA
PC_00004972	FDA_176	COG3539	Pilin (type 1 fimbria component protein)
PC_00005245	FDA_176	COG1752	Predicted acylesterase/phospholipase RssA, containd patatin domain
PC_00005577	FDA_176	COG2814	Predicted arabinose efflux permease, MFS family
PC_00005026	FDA_176	COG3179	Predicted chitinase
PC_00005046	FDA_176	COG3515	Predicted component of the type VI protein secretion system

PC_00005296	FDA_176	COG3520	Predicted component of the type VI protein secretion system
PC_00005864	FDA_176	COG3517	Predicted component of the type VI protein secretion system
PC_00005965	FDA_176	COG3522	Predicted component of the type VI protein secretion system
PC_00006528	FDA_176	COG3521	Predicted component of the type VI protein secretion system
PC_00005087	FDA_176	COG0673	Predicted dehydrogenase
PC_00006162	FDA_176	COG3311	Predicted DNA-binding transcriptional regulator AlpA
PC_00005861	FDA_176	COG2378	Predicted DNA-binding transcriptional regulator YafY, contains an HTH and WYL domains
PC_00005698	FDA_176	COG3576	Predicted flavin-nucleotide-binding protein, pyridoxine 5'-phosphate oxidase superfamily
PC_00006161	FDA_176	COG3596	Predicted GTPase
PC_00006371	FDA_176	COG1598	Predicted nuclease of the RNAse H fold, HicB family
PC_00006481	FDA_176	COG0667	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
PC_00005560	FDA_176	COG0679	Predicted permease
PC_00006150	FDA_176	COG4653	Predicted phage phi-C31 gp36 major capsid-like protein
PC_00005925	FDA_176	COG1724	Predicted RNA binding protein YcfA, dsRBD-like fold, HicA-like mRNA interferase family
PC_00006221	FDA_176	COG5472	Predicted small integral membrane protein
PC_00006021	FDA_176	COG5464	Predicted transposase YdaD
PC_00005534	FDA_176	COG0661	Predicted unusual protein kinase regulating ubiquinone biosynthesis, AarF/ABC1/UbiB family
PC_00006337	FDA_176	COG1989	Prepilin signal peptidase PulO (type II secretory pathway) or related peptidase

PC_00006054	FDA_176	COG3617	Prophage antirepressor
PC_00005102	FDA_176	COG3654	Prophage maintenance system killer protein
PC_00004988	FDA_176	COG1670	Protein N-acetyltransferase, RimJ/RimL family
PC_00005032	FDA_176	COG1651	Protein-disulfide isomerase
PC_00006326	FDA_176	COG1363	Putative aminopeptidase FrvX
PC_00005479	FDA_176	COG3657	Putative component of the toxin-antitoxin plasmid stabilization module
PC_00006933	FDA_176	COG1703	Putative periplasmic protein kinase ArgK or related GTPase of G3E family
PC_00005375	FDA_176	COG3723	Recombinational DNA repair protein RecT
PC_00005897	FDA_176	COG0684	Regulator of RNase E activity RraA
PC_00005543	FDA_176	COG0305	Replicative DNA helicase
PC_00005036	FDA_176	COG1715	Restriction endonuclease Mrr
PC_00005993	FDA_176	COG0732|COG0732	Restriction endonuclease S subunit|Restriction endonuclease S subunit
PC_00005309	FDA_176	COG3344	Retron-type reverse transcriptase
PC_00005405	FDA_176	COG3709	Ribose 1,5-bisphosphokinase PhnN
PC_00004961	FDA_176	COG0698	Ribose 5-phosphate isomerase RpiB
PC_00005571	FDA_176	COG1172	Ribose/xylose/arabinose/galactoside ABC-type transport system, permease component
PC_00005836	FDA_176	COG0499	S-adenosylhomocysteine hydrolase
PC_00005574	FDA_176	COG0515	Serine/threonine protein kinase
PC_00005262	FDA_176	COG0631	Serine/threonine protein phosphatase PrpC
PC_00005729	FDA_176	COG0172	Seryl-tRNA synthetase
PC_00004932	FDA_176	COG4264	Siderophore synthetase component
PC_00006442	FDA_176	COG5000	Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
PC_00005230	FDA_176	COG0629	Single-stranded DNA-binding protein

PC_00006726	FDA_176	COG1961	Site-specific DNA recombinase related to the DNA invertase Pin
PC_00005536	FDA_176	COG0338	Site-specific DNA-adenine methylase
PC_00005442	FDA_176	COG0270	Site-specific DNA-cytosine methylase
PC_00004913	FDA_176	COG4974	Site-specific recombinase XerD
PC_00005622	FDA_176	COG0741	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
PC_00005111	FDA_176	COG3109	sRNA-binding protein
PC_00004941	FDA_176	COG1082	Sugar phosphate isomerase/epimerase
PC_00005779	FDA_176	COG2271	Sugar phosphate permease
PC_00004966	FDA_176	COG0641	Sulfatase maturation enzyme AslB, radical SAM superfamily
PC_00005218	FDA_176	COG0210	Superfamily I DNA or RNA helicase
PC_00005121	FDA_176	COG3208	Surfactin synthase thioesterase subunit
PC_00006697	FDA_176	COG3005	Tetraheme cytochrome c subunit of nitrate or TMAO reductase
PC_00006299	FDA_176	COG0457	Tetratricopeptide (TPR) repeat
PC_00005064	FDA_176	COG1063	Threonine dehydrogenase or related Zn-dependent dehydrogenase
PC_00005839	FDA_176	COG3007	Trans-2-enoyl-CoA reductase
PC_00006465	FDA_176	COG3609	Transcriptional regulator, contains Arc/MetJ-type RHH (ribbon-helix-helix) DNA-binding domain
PC_00004877	FDA_176	COG1396	Transcriptional regulator, contains XRE-family HTH domain
PC_00005452	FDA_176	COG3335	Transposase
PC_00005906	FDA_176	COG3415	Transposase
PC_00006009	FDA_176	COG0675	Transposase
PC_00006247	FDA_176	COG3293	Transposase
PC_00006355	FDA_176	COG5421	Transposase

PC_00006960	FDA_176	COG4584	Transposase
PC_00005393	FDA_176	COG3328	Transposase (or an inactivated derivative)
PC_00004947	FDA_176	COG3328|COG3464	Transposase (or an inactivated derivative)|Transposase
PC_00005025	FDA_176	COG2963	Transposase and inactivated derivatives
PC_00004990	FDA_176	COG2801	Transposase InsO and inactivated derivatives
PC_00005891	FDA_176	COG3415|COG4373	Transposase|Mu-like prophage FluMu protein gp28
PC_00005303	FDA_176	COG3335|COG3415	Transposase|Transposase
PC_00005208	FDA_176	COG1638	TRAP-type C4-dicarboxylate transport system, periplasmic component
PC_00005226	FDA_176	COG1554	Trehalose and maltose hydrolase (possible phosphorylase)
PC_00005751	FDA_176	COG0149	Triosephosphate isomerase
PC_00006711	FDA_176	COG1767	Triphosphoribosyl-dephospho-CoA synthetase
PC_00005122	FDA_176	COG1746	tRNA nucleotidyltransferase (CCA-adding enzyme)
PC_00006282	FDA_176	COG0286|COG4227	Type I restriction-modification system, DNA methylase subunit|Antirestriction protein ArdC
PC_00006600	FDA_176	COG0610	Type I site-specific restriction-modification system, R (restriction) subunit and related helicases ...
PC_00005673	FDA_176	COG2804	Type II secretory pathway ATPase GspE/PulE or T4P pilus assembly pathway ATPase PilB
PC_00005998	FDA_176	COG1450	Type II secretory pathway component GspD/PulD (secretin)
PC_00004892	FDA_176	COG1459	Type II secretory pathway, component PulF
PC_00005780	FDA_176	COG3468	Type V secretory pathway, adhesin AidA
PC_00005818	FDA_176	COG3157	Type VI protein secretion system component Hcp (secreted cytotoxin)
PC_00006673	FDA_176	COG3455	Type VI protein secretion system component VasF
PC_00005949	FDA_176	COG3523	Type VI protein secretion system component VasK

PC_00005678	FDA_176	COG0562	UDP-galactopyranose mutase
PC_00005329	FDA_176	COG1422	Uncharacterized archaeal membrane protein, DUF106 family, distantly related to YidC/Oxa1
PC_00006503	FDA_176	COG3193	Uncharacterized conserved protein GlcG, DUF336 family
PC_00005685	FDA_176	COG3209	Uncharacterized conserved protein RhaS, contains 28 RHS repeats
PC_00006006	FDA_176	COG3251	Uncharacterized conserved protein YbdZ, MbtH family
PC_00005335	FDA_176	COG3110	Uncharacterized conserved protein YccT, UPF0319 family
PC_00006061	FDA_176	COG2841	Uncharacterized conserved protein YdcH, DUF465 family
PC_00006677	FDA_176	COG2373	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
PC_00006335	FDA_176	COG2373|COG2911|COG3210|COG4932	Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family|Autotransporter translocation and assembly factor TamB|Large exoprotein involved in heme utilization or adhesion|Uncharacterized surface anchored protein
PC_00005173	FDA_176	COG5492	Uncharacterized conserved protein YjdB, contains Ig-like domain
PC_00006101	FDA_176	COG4625	Uncharacterized conserved protein, contains a C-terminal beta-barrel porin domain
PC_00006632	FDA_176	COG4372|COG5280|COG5412	Uncharacterized conserved protein, contains DUF3084 domain|Phage-related minor tail protein|Phage-related protein
PC_00006932	FDA_176	COG5266	Uncharacterized conserved protein, contains GH25 family domain
PC_00005294	FDA_176	COG2369	Uncharacterized conserved protein, contains phage Mu gpF-like domain
PC_00005873	FDA_176	COG4453	Uncharacterized conserved protein, DUF1778 family

PC_00006920	FDA_176	COG3514	Uncharacterized conserved protein, DUF4415 family
PC_00006166	FDA_176	COG3501|COG4253	Uncharacterized conserved protein, implicated in type VI secretion and phage assembly|Uncharacterized conserved protein, DUF2345 family
PC_00005542	FDA_176	COG1361|COG1361|COG2373	Uncharacterized conserved protein|Uncharacterized conserved protein|Uncharacterized conserved protein YfaS, alpha-2-macroglobulin family
PC_00006402	FDA_176	COG0730	Uncharacterized membrane protein YfcA
PC_00005680	FDA_176	COG1285	Uncharacterized membrane protein YhiD, involved in acid resistance
PC_00005679	FDA_176	COG3299	Uncharacterized phage protein gp47/JayE
PC_00004879	FDA_176	COG4289	Uncharacterized protein
PC_00005047	FDA_176	COG4834	Uncharacterized protein
PC_00005794	FDA_176	COG3566	Uncharacterized protein
PC_00005892	FDA_176	COG3567	Uncharacterized protein
PC_00006109	FDA_176	COG4877	Uncharacterized protein
PC_00006642	FDA_176	COG4340	Uncharacterized protein
PC_00006678	FDA_176	COG5556	Uncharacterized protein
PC_00006727	FDA_176	COG3470	Uncharacterized protein probably involved in high-affinity Fe2+ transport
PC_00006043	FDA_176	COG3778	Uncharacterized protein YmfQ in lambdoid prophage, DUF2313 family
PC_00005585	FDA_176	COG0006	Xaa-Pro aminopeptidase
PC_00005468	FDA_176	COG4104	Zn-binding Pro-Ala-Ala-Arg (PAAR) domain, incolved in TypeVI secretion
PC_00006648	FDA_176	COG3227	Zn-dependent metalloprotease

4 References

1. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ: **High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity**. *Nat Rev Microbiol* 2012, **10**(9):599-606.
2. Darwin KH, Miller VL: **Molecular basis of the interaction of Salmonella with the intestinal mucosa**. *Clin Microbiol Rev* 1999, **12**(3):405-428.
3. Lan R, Reeves PR, Octavia S: **Population structure, origins and evolution of major Salmonella enterica clones**. *Infect Genet Evol* 2009, **9**(5):996-1005.
4. Foley SL, Nayak R, Hanning IB, Johnson TJ, Han J, Ricke SC: **Population Dynamics of Salmonella enterica Serotypes in Commercial Egg and Poultry Production**. *Appl Environ Microb* 2011, **77**(13):4273-4279.
5. F.W.Brenner: **Salmonella Nomenclature**. *Journal of Clinical Microbiology* 2000, **38**(7):2465-2467.
6. Coburn B, Grassl GA, Finlay BB: **Salmonella, the host and disease: a brief review**. *Immunol Cell Biol* 2007, **85**(2):112-118.
7. http://textbookofbacteriology.net/salmonella.html
8. D E: **Who named it, retrieved 12 November, 2007 from** http://www.whonamedit.com/Synd.cfm/402.html 2007.
9. Grimont PAD GF, Bouvet P: **Taxonomy of the Genus Salmonella, In:Salmonella in domestic animals. (Wray C, Wray A, editors). CABI Publishing , Wallingford, United Kingdom**. 2000:1-17.
10. Giannella RA: **Salmonella**. In: *Medical Microbiology.* Edited by Baron S, 4th edn. Galveston (TX); 1996.
11. Ohl ME, Miller SI: **Salmonella: a model for bacterial pathogenesis**. *Annu Rev Med* 2001, **52**:259-274.
12. Encheva V, Shah HN, Gharbia SE: **Proteomic analysis of the adaptive response of Salmonella enterica serovar Typhimurium to growth under anaerobic conditions**. *Microbiology* 2009, **155**(Pt 7):2429-2441.
13. Grimont P.A.D, Grimont F, Bouvet P. 2000 Taxonomy of the Genus Salmonella. In:Salmonella in domestic animals. (Wray C, Wray A, editors). CABI Publishing, Wallingford, United Kingdom, 1-17.
14. Patrick A.D. Grimont aF-XW: **Antigenic formulae of the Salmonella serovars**. *WHO Collaborating Centre for Reference and Research on Salmonella, Pasteur Institute, Paris, France* 2007, **9th ed.**
15. Hohmann EL: **Nontyphoidal salmonellosis**. *Clin Infect Dis* 2001, **32**(2):263-269.
16. Leon Le Minor MYP: **Designation of Salmonella enterica sp. nov., norn. rev., as the Type and Only Species of the Genus Salmonella** *International Journal of Systemic Bacteriology* 1987, **37**(4):465-468.
17. B.J.Tindall PADG, G.M.Garrity,and J. P. Euz´eby: **Nomenclature and taxonomy of the genus Salmonella**. *International Journal of Systematic and Evolutionary Microbiology* 2005, **55**(1):521–524.
18. Capra EJ, Laub MT: **Evolution of two-component signal transduction systems**. *Annu Rev Microbiol* 2012, **66**:325-347.

19. Issenhuth-Jeanjean S, Roggentin P, Mikoleit M, Guibourdenche M, de Pinna E, Nair S, Fields PI, Weill FX: **Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme**. *Res Microbiol* 2014, **165**(7):526-530.
20. van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P *et al*: **Guidelines for the validation and application of typing methods for use in bacterial epidemiology**. *Clin Microbiol Infect* 2007, **13 Suppl 3**:1-46.
21. Crichton PB, Old DC: **Salmonellae of serotypes gallinarum and pullorum grouped by biotyping and fimbrial-gene probing**. *J Med Microbiol* 1990, **32**(3):145-152.
22. Millemann Y, Lesage MC, Chaslus-Dancla E, Lafont JP: **Value of plasmid profiling, ribotyping, and detection of IS200 for tracing avian isolates of Salmonella typhimurium and S. enteritidis**. *J Clin Microbiol* 1995, **33**(1):173-179.
23. Borman EK, Wheeler KM, West DE, Mickle FL: **Salmonella Typing in a Public Health Laboratory**. *Am J Public Health Nations Health* 1943, **33**(2):127-134.
24. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A *et al*: **Multilocus sequence typing as a replacement for serotyping in Salmonella enterica**. *PLoS Pathog* 2012, **8**(6):e1002776.
25. Herikstad H, Motarjemi Y, Tauxe RV: **Salmonella surveillance: a global survey of public health serotyping**. *Epidemiol Infect* 2002, **129**(1):1-8.
26. Uesugi AR, Danyluk MD, Mandrell RE, Harris LJ: **Isolation of Salmonella Enteritidis phage type 30 from a single almond orchard over a 5-year period**. *J Food Prot* 2007, **70**(8):1784-1789.
27. Rankin S, Platt DJ: **Phage conversion in Salmonella enterica serotype Enteritidis: implications for epidemiology**. *Epidemiol Infect* 1995, **114**(2):227-236.
28. Ward LR, de Sa JD, Rowe B: **A phage-typing scheme for Salmonella enteritidis**. *Epidemiol Infect* 1987, **99**(2):291-294.
29. Malorny B, Schroeter A, Bunge C, Hoog B, Steinbeck A, Helmuth R: **Evaluation of molecular typing methods for Salmonella enterica serovar Typhimurium DT104 isolated in Germany from healthy pigs**. *Vet Res* 2001, **32**(2):119-129.
30. Holmberg SD, Wachsmuth IK, Hickman-Brenner FW, Cohen ML: **Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing Salmonella typhimurium isolates from outbreaks**. *J Clin Microbiol* 1984, **19**(2):100-104.
31. Dale JW: **Mobile genetic elements in mycobacteria**. *Eur Respir J Suppl* 1995, **20**:633s-648s.
32. Olive DM, Bean P: **Principles and applications of methods for DNA-based typing of microbial organisms**. *J Clin Microbiol* 1999, **37**(6):1661-1669.
33. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B: **Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing**. *J Clin Microbiol* 1995, **33**(9):2233-2239.
34. Enright MC, Spratt BG: **Multilocus sequence typing**. *Trends in Microbiology* 1999, **7**(12):482-487.
35. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou JJ, Zurth K, Caugant DA *et al*: **Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms**. *P Natl Acad Sci USA* 1998, **95**(6):3140-3145.
36. Cooper JE, Feil EJ: **Multilocus sequence typing - what is resolved?** *Trends in Microbiology* 2004, **12**(8):373-377.

37. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS: **Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics**. *Appl Environ Microbiol* 1986, **51**(5):873-884.
38. Urwin R, Maiden MC: **Multi-locus sequence typing: a tool for global epidemiology**. *Trends Microbiol* 2003, **11**(10):479-487.
39. Tankouo-Sandjong B, Sessitsch A, Liebana E, Kornschober C, Allerberger F, Hachler H, Bodrossy L: **MLST-v, multilocus sequence typing based on virulence genes, for molecular typing of Salmonella enterica subsp. enterica serovars**. *J Microbiol Methods* 2007, **69**(1):23-36.
40. Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW: **High resolution clustering of Salmonella enterica serovar Montevideo strains using a next-generation sequencing approach**. *BMC Genomics* 2012, **13**:32.
41. Feavers IM, Gray SJ, Urwin R, Russell JE, Bygraves JA, Kaczmarski EB, Maiden MCJ: **Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak**. *Journal of Clinical Microbiology* 1999, **37**(12):3883-3887.
42. Lin RJ, Capage M, Hill CW: **A repetitive DNA sequence, rhs, responsible for duplications within the Escherichia coli K-12 chromosome**. *J Mol Biol* 1984, **177**(1):1-18.
43. Lupski JR, Weinstock GM: **Short, interspersed repetitive DNA sequences in prokaryotic genomes**. *J Bacteriol* 1992, **174**(14):4525-4529.
44. Li W, Raoult D, Fournier PE: **Bacterial strain typing in the genomic era**. *FEMS Microbiol Rev* 2009, **33**(5):892-916.
45. Torpdahl M, Sorensen G, Lindstedt BA, Nielsen EM: **Tandem repeat analysis for surveillance of human Salmonella Typhimurium infections**. *Emerg Infect Dis* 2007, **13**(3):388-395.
46. Gal-Mor O, Boyle EC, Grassl GA: **Same species, different diseases: how and why typhoidal and non-typhoidal Salmonella enterica serovars differ**. *Front Microbiol* 2014, **5**:391.
47. Raffatellu M, Wilson RP, Winter SE, Baumler AJ: **Clinical pathogenesis of typhoid fever**. *J Infect Dev Ctries* 2008, **2**(4):260-266.
48. Crump JA, Luby SP, Mintz ED: **The global burden of typhoid fever**. *Bull World Health Organ* 2004, **82**(5):346-353.
49. Buckle GC, Walker CL, Black RE: **Typhoid fever and paratyphoid fever: Systematic review to estimate global morbidity and mortality for 2010**. *J Glob Health* 2012, **2**(1):010401.
50. Crump JA, Mintz ED: **Global trends in typhoid and paratyphoid Fever**. *Clin Infect Dis* 2010, **50**(2):241-246.
51. Galanis E, Lo Fo Wong DM, Patrick ME, Binsztein N, Cieslik A, Chalermchikit T, Aidara-Kane A, Ellis A, Angulo FJ, Wegener HC *et al*: **Web-based surveillance and global Salmonella distribution, 2000-2002**. *Emerg Infect Dis* 2006, **12**(3):381-388.
52. WHO: **WHO Global Salm-Surv: Progress Report 2000-2005. Available at** http://www.who.int/salmsurv/links/GSSProgressReport2005.pdf. 2006.
53. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM, International Collaboration on Enteric Disease 'Burden of Illness S: **The global burden of nontyphoidal Salmonella gastroenteritis**. *Clin Infect Dis* 2010, **50**(6):882-889.
54. Carattoli A: **Plasmid-mediated antimicrobial resistance in Salmonella enterica**. *Curr Issues Mol Biol* 2003, **5**(4):113-122.

55. Uzzau S, Brown DJ, Wallis T, Rubino S, Leori G, Bernard S, Casadesus J, Platt DJ, Olsen JE: **Host adapted serotypes of Salmonella enterica**. *Epidemiol Infect* 2000, **125**(2):229-255.
56. Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M *et al*: **Salmonella bongori provides insights into the evolution of the Salmonellae**. *PLoS Pathog* 2011, **7**(8):e1002191.
57. **Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old**.
58. Barrow PA, Huggins MB, Lovell MA: **Host specificity of Salmonella infection in chickens and mice is expressed in vivo primarily at the level of the reticuloendothelial system**. *Infect Immun* 1994, **62**(10):4602-4610.
59. **Ovine salmonellosis caused by Salmonella Abortusovis**.
60. Pardon P, Sanchis R, Marly J, Lantier F, Guilloteau L, Buzoni-Gatel D, Oswald IP, Pepin M, Kaeffer B, Berthon P *et al*: **Experimental ovine salmonellosis (Salmonella abortusovis): pathogenesis and vaccination**. *Res Microbiol* 1990, **141**(7-8):945-953.
61. Chu C, Hong SF, Tsai C, Lin WS, Liu TP, Ou JT: **Comparative physical and genetic maps of the virulence plasmids of Salmonella enterica serovars typhimurium, enteritidis, choleraesuis, and dublin**. *Infect Immun* 1999, **67**(5):2611-2614.
62. Gray JT, Stabel TJ, Fedorka-Cray PJ: **Effect of dose on the immune response and persistence of Salmonella choleraesuis infection in swine**. *Am J Vet Res* 1996, **57**(3):313-319.
63. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F *et al*: **Complete genome sequence of Salmonella enterica serovar Typhimurium LT2**. *Nature* 2001, **413**(6858):852-856.
64. Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschape H, Adams LG, Baumler AJ: **Salmonella enterica serotype Typhimurium and its host-adapted variants**. *Infect Immun* 2002, **70**(5):2249-2255.
65. Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, Friis C: **The Salmonella enterica pan-genome**. *Microb Ecol* 2011, **62**(3):487-504.
66. Lawrence JG, Ochman H: **Molecular archaeology of the Escherichia coli genome**. *Proc Natl Acad Sci U S A* 1998, **95**(16):9413-9417.
67. Edwards RA, Olsen GJ, Maloy SR: **Comparative genomics of closely related salmonellae**. *Trends Microbiol* 2002, **10**(2):94-99.
68. Ochman H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation**. *Nature* 2000, **405**(6784):299-304.
69. Porwollik S, McClelland M: **Lateral gene transfer in Salmonella**. *Microbes Infect* 2003, **5**(11):977-989.
70. Porwollik S, Wong RM, McClelland M: **Evolutionary genomics of Salmonella: gene acquisitions revealed by microarray analysis**. *Proc Natl Acad Sci U S A* 2002, **99**(13):8956-8961.
71. Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H: **Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution**. *Mol Microbiol* 1997, **23**(6):1089-1097.
72. Groisman EA, Ochman H: **Pathogenicity islands: bacterial evolution in quantum leaps**. *Cell* 1996, **87**(5):791-794.
73. Marcus SL, Brumell JH, Pfeifer CG, Finlay BB: **Salmonella pathogenicity islands: big virulence in small packages**. *Microbes Infect* 2000, **2**(2):145-156.
74. van Asten AJ, van Dijk JE: **Distribution of "classic" virulence factors among Salmonella spp**. *FEMS Immunol Med Microbiol* 2005, **44**(3):251-259.

75. Hayward MR, Jansen V, Woodward MJ: **Comparative genomics of Salmonella enterica serovars Derby and Mbandaka, two prevalent serovars associated with different livestock species in the UK**. *BMC Genomics* 2013, **14**:365.
76. Ochman H, Groisman EA: **Distribution of pathogenicity islands in Salmonella spp**. *Infect Immun* 1996, **64**(12):5410-5412.
77. Hensel M, Shea JE, Baumler AJ, Gleeson C, Blattner F, Holden DW: **Analysis of the boundaries of Salmonella pathogenicity island 2 and the corresponding chromosomal region of Escherichia coli K-12**. *J Bacteriol* 1997, **179**(4):1105-1111.
78. Knodler LA, Vallance BA, Hensel M, Jackel D, Finlay BB, Steele-Mortimer O: **Salmonella type III effectors PipB and PipB2 are targeted to detergent-resistant microdomains on internal host cell membranes**. *Mol Microbiol* 2003, **49**(3):685-704.
79. Chakravortty D, Hansen-Wester I, Hensel M: **Salmonella pathogenicity island 2 mediates protection of intracellular Salmonella from reactive nitrogen intermediates**. *J Exp Med* 2002, **195**(9):1155-1166.
80. Ochman H, Soncini FC, Solomon F, Groisman EA: **Identification of a pathogenicity island required for Salmonella survival in host cells**. *Proc Natl Acad Sci U S A* 1996, **93**(15):7800-7804.
81. Vazquez-Torres A, Xu Y, Jones-Carson J, Holden DW, Lucia SM, Dinauer MC, Mastroeni P, Fang FC: **Salmonella pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase**. *Science* 2000, **287**(5458):1655-1658.
82. Uchiya K, Barbieri MA, Funato K, Shah AH, Stahl PD, Groisman EA: **A Salmonella virulence protein that inhibits cellular trafficking**. *EMBO J* 1999, **18**(14):3924-3933.
83. Kuhle V, Abrahams GL, Hensel M: **Intracellular Salmonella enterica redirect exocytic transport processes in a Salmonella pathogenicity island 2-dependent manner**. *Traffic* 2006, **7**(6):716-730.
84. Garmendia J, Beuzon CR, Ruiz-Albert J, Holden DW: **The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the Salmonella typhimurium SPI-2 type III secretion system**. *Microbiology* 2003, **149**(Pt 9):2385-2396.
85. Bijlsma JJ, Groisman EA: **The PhoP/PhoQ system controls the intramacrophage type three secretion system of Salmonella enterica**. *Mol Microbiol* 2005, **57**(1):85-96.
86. Blanc-Potard AB, Groisman EA: **The Salmonella selC locus contains a pathogenicity island mediating intramacrophage survival**. *EMBO J* 1997, **16**(17):5376-5385.
87. Wong KK, McClelland M, Stillwell LC, Sisk EC, Thurston SJ, Saffer JD: **Identification and sequence analysis of a 27-kilobase chromosomal fragment containing a Salmonella pathogenicity island located at 92 minutes on the chromosome map of Salmonella enterica serovar typhimurium LT2**. *Infect Immun* 1998, **66**(7):3365-3371.
88. Gerlach RG, Jackel D, Stecher B, Wagner C, Lupas A, Hardt WD, Hensel M: **Salmonella Pathogenicity Island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system**. *Cell Microbiol* 2007, **9**(7):1834-1850.
89. Morgan E, Bowen AJ, Carnell SC, Wallis TS, Stevens MP: **SiiE is secreted by the Salmonella enterica serovar Typhimurium pathogenicity island 4-encoded secretion system and contributes to intestinal colonization in cattle**. *Infect Immun* 2007, **75**(3):1524-1533.
90. Zhou D, Chen LM, Hernandez L, Shears SB, Galan JE: **A Salmonella inositol polyphosphatase acts in conjunction with other bacterial effectors to promote host cell actin cytoskeleton rearrangements and bacterial internalization**. *Mol Microbiol* 2001, **39**(2):248-259.

91. Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE: **Identification of a pathogenicity island required for Salmonella enteropathogenicity**. *Mol Microbiol* 1998, **29**(3):883-891.
92. Pickard D, Wain J, Baker S, Line A, Chohan S, Fookes M, Barron A, Gaora PO, Chabalgoity JA, Thanky N *et al*: **Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding Salmonella enterica pathogenicity island SPI-7**. *J Bacteriol* 2003, **185**(17):5055-5065.
93. Gulig PA: **Virulence plasmids of Salmonella typhimurium and other salmonellae**. *Microb Pathog* 1990, **8**(1):3-11.
94. Rotger R, Casadesus J: **The virulence plasmids of Salmonella**. *Int Microbiol* 1999, **2**(3):177-184.
95. Marc L. Lesnick NER, Joshua Fierer,, Guiney aDG: **The Salmonella spvB virulence gene encodes an enzyme that ADPribosylates actin and destabilizes the cytoskeleton of eukaryotic cells**. 2001.
96. Collinson SK: **The location of four fimbrin-encoding genes, agfA, fimA, sefA and sefD, on the Salmonella enteritidis and/or S. typhimurium XbaI-BlnI genomic restriction maps**. 1996.
97. Clouthier SC, Collinson SK, Kay WW: **Unique fimbriae-like structures encoded by sefD of the SEF14 fimbrial gene cluster of Salmonella enteritidis**. *Mol Microbiol* 1994, **12**(6):893-901.
98. Hyeon JY, Chon JW, Hwang IG, Kwak HS, Kim MS, Kim SK, Choi IS, Song CS, Park C, Seo KH: **Prevalence, Antibiotic Resistance, and Molecular Characterization of Salmonella Serovars in Retail Meat Products**. *J Food Protect* 2011, **74**(1):161-166.
99. Travers K, Barza M: **Morbidity of infections caused by antimicrobial-resistant bacteria**. *Clin Infect Dis* 2002, **34 Suppl 3**:S131-134.
100. Threlfall EJ: **Antimicrobial drug resistance in Salmonella: problems and perspectives in food- and water-borne infections**. *FEMS Microbiol Rev* 2002, **26**(2):141-148.
101. Tenover FC: **Mechanisms of antimicrobial resistance in bacteria**. *Am J Med* 2006, **119**(6 Suppl 1):S3-10; discussion S62-70.
102. Hall RM, Collis CM: **Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination**. *Mol Microbiol* 1995, **15**(4):593-600.
103. Waters VL: **Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance**. *Front Biosci* 1999, **4**:D433-456.
104. Wray C, Hedges RW, Shannon KP, Bradley DE: **Apramycin and gentamicin resistance in Escherichia coli and salmonellas isolated from farm animals**. *J Hyg (Lond)* 1986, **97**(3):445-456.
105. Threlfall EJ, Rowe B, Ferguson JL, Ward LR: **Characterization of plasmids conferring resistance to gentamicin and apramycin in strains of Salmonella typhimurium phage type 204c isolated in Britain**. *J Hyg (Lond)* 1986, **97**(3):419-426.
106. Pohl P, Glupczynski Y, Marin M, Van Robaeys G, Lintermans P, Couturier M: **Replicon typing characterization of plasmids encoding resistance to gentamicin and apramycin in Escherichia coli and Salmonella typhimurium isolated from human and animal sources in Belgium**. *Epidemiol Infect* 1993, **111**(2):229-238.
107. Llanes C: **Propagation of TEM- and PSE-type β-lactamases among amoxicillin-resistanct Salmonella spp. isolated in France**. 1999.
108. **Emergence of domestically acquired ceftraxone-resistant Salmonella infections associated with AmpC beta-lactamase**.

109. Fey PD, Safranek TJ, Rupp ME, Dunne EF, Ribot E, Iwen PC, Bradford PA, Angulo FJ, Hinrichs SH: **Ceftriaxone-resistant salmonella infection acquired by a child from cattle**. *N Engl J Med* 2000, **342**(17):1242-1249.
110. Mulvey MR, Boyd DA, Olson AB, Doublet B, Cloeckaert A: **The genetics of Salmonella genomic island 1**. *Microbes Infect* 2006, **8**(7):1915-1922.
111. Boyd D, Peters GA, Cloeckaert A, Boumedine KS, Chaslus-Dancla E, Imberechts H, Mulvey MR: **Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of Salmonella enterica serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona**. *J Bacteriol* 2001, **183**(19):5725-5732.
112. Hall RM: **Salmonella genomic islands and antibiotic resistance in Salmonella enterica**. *Future Microbiol* 2010, **5**(10):1525-1538.
113. Michael GB, Schwarz S: **Antimicrobial resistance in zoonotic nontyphoidal Salmonella: an alarming trend?** *Clin Microbiol Infect* 2016, **22**(12):968-974.
114. Lopes GV, Michael GB, Cardoso M, Schwarz S: **Identification and characterization of Salmonella enterica subsp enterica serovar Derby isolates carrying a new aadA26 gene cassette in a class 1 integron obtained at pig slaughterhouses**. *Fems Microbiology Letters* 2014, **356**(1):71-78.
115. Lopes GV, Michael GB, Cardoso M, Schwarz S: **Antimicrobial resistance and class 1 integron-associated gene cassettes in Salmonella enterica serovar Typhimurium isolated from pigs at slaughter and abattoir environment**. *Vet Microbiol* 2016, **194**:84-92.
116. Ramirez MS, Tolmasky ME: **Aminoglycoside modifying enzymes**. *Drug Resist Update* 2010, **13**(6):151-171.
117. Falgenhauer L, Ghosh H, Guerra B, Yao YC, Fritzenwanker M, Fischer J, Helmuth R, Imirzalioglu C, Chakraborty T: **Comparative genome analysis of IncHI2 VIM-1 carbapenemase-encoding plasmids of Escherichia coli and Salmonella enterica isolated from a livestock farm in Germany**. *Vet Microbiol* 2017, **200**:114-117.
118. Le Hello S, Harrois D, Bouchrif B, Sontag L, Elhani D, Guibert V, Zerouali K, Weill FX: **Highly drug-resistant Salmonella enterica serotype Kentucky ST198-X1: a microbiological study**. *Lancet Infectious Diseases* 2013, **13**(8):672-679.
119. Piddock LJV: **Fluoroquinolone resistance in Salmonella serovars isolated from humans and food animals**. *Fems Microbiology Reviews* 2002, **26**(1):3-16.
120. Steele-Mortimer O: **The invasion-associated type III secretion system of Salmonella enterica serovar Typhimurium is necessary for intracellular proliferation and vacuole biogenesis in epithelial cells**. *Cellular microbiology* 2002.
121. Galan JE, Curtiss R, 3rd: **Cloning and molecular characterization of genes whose products allow Salmonella typhimurium to penetrate tissue culture cells**. *Proc Natl Acad Sci U S A* 1989, **86**(16):6383-6387.
122. Clark MA, Hirst BH, Jepson MA: **Inoculum composition and Salmonella pathogenicity island 1 regulate M-cell invasion and epithelial destruction by Salmonella typhimurium**. *Infect Immun* 1998, **66**(2):724-731.
123. Francis CL, Ryan TA, Jones BD, Smith SJ, Falkow S: **Ruffles induced by Salmonella and other stimuli direct macropinocytosis of bacteria**. *Nature* 1993, **364**(6438):639-642.
124. Cirillo DM, Valdivia RH, Monack DM, Falkow S: **Macrophage-dependent induction of the Salmonella pathogenicity island 2 type III secretion system and its role in intracellular survival**. *Mol Microbiol* 1998, **30**(1):175-188.

125. Hensel M, Shea JE, Waterman SR, Mundy R, Nikolaus T, Banks G, Vazquez-Torres A, Gleeson C, Fang FC, Holden DW: **Genes encoding putative effector proteins of the type III secretion system of Salmonella pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages**. *Mol Microbiol* 1998, **30**(1):163-174.
126. Rosselin M, Virlogeux-Payant I, Roy C, Bottreau E, Sizaret PY, Mijouin L, Germon P, Caron E, Velge P, Wiedemann A: **Rck of Salmonella enterica, subspecies enterica serovar enteritidis, mediates zipper-like internalization**. *Cell Res* 2010, **20**(6):647-664.
127. Boumart Z, Velge P, Wiedemann A: **Multiple invasion mechanisms and different intracellular Behaviors: a new vision of Salmonella-host cell interaction**. *FEMS Microbiol Lett* 2014, **361**(1):1-7.
128. Rychlik I, Gregorova D, Hradecka H: **Distribution and function of plasmids in Salmonella enterica**. *Vet Microbiol* 2006, **112**(1):1-10.
129. Schlumberger MC, Hardt WD: **Salmonella type III secretion effectors: pulling the host cell's strings**. *Curr Opin Microbiol* 2006, **9**(1):46-54.
130. McGhie EJ, Hayward RD, Koronakis V: **Cooperation between actin-binding proteins of invasive Salmonella: SipA potentiates SipC nucleation and bundling of actin**. *EMBO J* 2001, **20**(9):2131-2139.
131. hayward Rd: **Direct modulation of the host cell cytoskeleton by Salmonella actin-binding proteins**. 2002.
132. Álvarez-Ordóñez A, Prieto M, Bernardo A, Hill C, López M: **The Acid Tolerance Response of Salmonella spp.: An adaptive strategy to survive in stressful environments prevailing in foods and the host**. *Food Research International* 2012, **45**(2):482-492.
133. Foster JW, Hall HK: **Adaptive acidification tolerance response of Salmonella typhimurium**. *J Bacteriol* 1990, **172**(2):771-778.
134. Foster JW: **The acid tolerance response of Salmonella typhimurium involves transient synthesis of key acid shock proteins**. *J Bacteriol* 1993, **175**(7):1981-1987.
135. Foster JW, Hall HK: **Inducible pH homeostasis and the acid tolerance response of Salmonella typhimurium**. *J Bacteriol* 1991, **173**(16):5129-5135.
136. Foster JW: **Salmonella acid shock proteins are required for the adaptive acid tolerance response**. *J Bacteriol* 1991, **173**(21):6896-6902.
137. Riesenberg-Wilmes MR, Bearson B, Foster JW, Curtis R, 3rd: **Role of the acid tolerance response in virulence of Salmonella typhimurium**. *Infect Immun* 1996, **64**(4):1085-1092.
138. **<antibiotic resistance to mbandaka.pdf>**.
139. Meltzer E, Schwartz E: **Enteric fever: a travel medicine oriented view**. *Current Opinion in Infectious Diseases* 2010, **23**(5):432-437.
140. Mermin J, Hutwagner L, Vugia D, Shallow S, Daily P, Bender J, Koehler J, Marcus R, Angulo FJ, FoodNe EIP: **Reptiles, amphibians, and human Salmonella infection: A population-based, case-control study**. *Clin Infect Dis* 2004, **38**:S253-S261.
141. Braden CR: **Salmonella enterica serotype enteritidis and eggs: A national epidemic in the United States**. *Clin Infect Dis* 2006, **43**(4):512-517.
142. Haeusler GM, Curtis N: **Non-typhoidal Salmonella in Children: Microbiology, Epidemiology and Treatment**. *Hot Topics in Infection and Immunity in Children Ix* 2013, **764**:13-26.
143. Cavallaro E, Date K, Medus C, Meyer S, Miller B, Kim C, Nowicki S, Cosgrove S, Sweat D, Phan Q *et al*: **Salmonella Typhimurium Infections Associated with Peanut Products**. *New Engl J Med* 2011, **365**(7):601-610.
144. Bayer C, Bernard H, Prager R, Rabsch W, Hiller P, Malorny B, Pfefferkorn B, Frank C, de Jong A, Friesema I *et al*: **An outbreak of Salmonella Newport associated with mung**

bean sprouts in Germany and the Netherlands, October to November 2011. *Eurosurveillance* 2014, **19**(1):10-18.
145. Mccormick BA, Miller SI, Carnes D, Madara JL: **Transepithelial Signaling to Neutrophils by Salmonellae - a Novel Virulence Mechanism for Gastroenteritis**. *Infection and Immunity* 1995, **63**(6):2302-2309.
146. Sprinz H, Gangarosa EJ, Williams M, Hornick RB, Woodward TE: **Histopathology of the upper small intestines in typhoid fever. Biopsy study of experimental disease in man**. *Am J Dig Dis* 1966, **11**(8):615-624.
147. Kraus MD, Amatya B, Kimula Y: **Histopathology of typhoid enteritis: Morphologic and immunophenotypic findings**. *Modern Pathol* 1999, **12**(10):949-955.
148. Gordon MA: **Salmonella infections in immunocompromised adults**. *J Infection* 2008, **56**(6):413-422.
149. House D, Wain J, Ho VA, Diep TS, Chinh NT, Bay PV, Vinh H, Duc M, Parry CM, Dougan G *et al*: **Serology of typhoid fever in an area of endemicity and its relevance to diagnosis**. *Journal of Clinical Microbiology* 2001, **39**(3):1002-1007.
150. Darby J, Sheorey H: **Searching for Salmonella**. *Australian Family Physician* 2008, **37**(10):806-810.
151. Connor BA, Schwartz E: **Typhoid and paratyphoid fever in travellers**. *Lancet Infectious Diseases* 2005, **5**(10):623-628.
152. Meltzer E, Sadik C, Schwartz E: **Enteric fever in Israeli travelers: A nationwide study**. *Journal of Travel Medicine* 2005, **12**(5):275-281.
153. Olsen SJ, Bleasdale SC, Magnano AR, Landrigan C, Holland BH, Tauxe RV, Mintz ED, Luby S: **Outbreaks of typhoid fever in the United States, 1960-99**. *Epidemiology and Infection* 2003, **130**(1):13-21.
154. Bhan MK, Bahl R, Bhatnagar S: **Typhoid and paratyphoid fever**. *Lancet* 2005, **366**(9487):749-762.
155. Stuart BM, Pullen RL: **Typhoid; clinical analysis of 360 cases**. *Arch Intern Med (Chic)* 1946, **78**(6):629-661.
156. Glynn JR, Palmer SR: **Incubation Period, Severity of Disease, and Infecting Dose - Evidence from a Salmonella Outbreak**. *Am J Epidemiol* 1992, **136**(11):1369-1377.
157. McGovern VJ, Slavutin LJ: **Pathology of salmonella colitis**. *Am J Surg Pathol* 1979, **3**(6):483-490.
158. Gonzalez-Escobedo G, Marshall JM, Gunn JS: **Chronic and acute infection of the gall bladder by Salmonella Typhi: understanding the carrier state**. *Nat Rev Microbiol* 2011, **9**(1):9-14.
159. Hoszowski A, Zajac M, Lalak A, Przemyk P, Wasyl D: **Fifteen years of successful spread of Salmonella enterica serovar Mbandaka clone ST413 in Poland and its public health consequences**. *Ann Agric Environ Med* 2016, **23**(2):237-241.
160. Scheil W, Cameron S, Dalton C, Murray C, Wilson D: **A South Australian Salmonella Mbandaka outbreak investigation using a database to select controls**. *Aust N Z J Public Health* 1998, **22**(5):536-539.
161. division CfdpeOh: **Another outbreak of alfalfa sprout-associated salmonellosis**. *CD summary* 1999, **48**(5).
162. (CDC) CfDCaP: **Multistate Outbreak of Salmonella Montevideo and Salmonella Mbandaka Infections Linked to Tahini Sesame Paste (Final Update)**. 2013.
163. (CDC) CfDCaP: **Eight Multistate Outbreaks of Human Salmonella Infections Linked to Live Poultry in Backyard Flocks (Final Update)**. 2016.
164. CDC PHLIS: Salmonella annual survey, Volume 2010. Atlanta, USA: CDC; 2009.

165. Hayward MR, Petrovska L, Jansen VAA, Woodward MJ: **Population structure and associated phenotypes of Salmonella enterica serovars Derby and Mbandaka overlap with host range**. *BMC Microbiol* 2016, **16**:9.
166. Hoszowski A, Wasyl D: **Typing of Salmonella enterica subsp enterica serovar Mbandaka isolates**. *Vet Microbiol* 2001, **80**(2):139-148.
167. Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, Wehnes C, Van Kessel JS, Karns JS, Musser SM, Brown EW: **Phylogenetic diversity of the enteric pathogen Salmonella enterica subsp. enterica inferred from genome-wide reference-free SNP characters**. *Genome Biol Evol* 2013, **5**(11):2109-2123.
168. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M: **Nucleotide sequence of bacteriophage phi X174 DNA**. *Nature* 1977, **265**(5596):687-695.
169. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proc Natl Acad Sci U S A* 1977, **74**(12):5463-5467.
170. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM *et al*: **Whole-genome random sequencing and assembly of Haemophilus influenzae Rd**. *Science* 1995, **269**(5223):496-512.
171. Ronholm J, Nasheri N, Petronella N, Pagotto F: **Navigating Microbiological Food Safety in the Era of Whole-Genome Sequencing**. *Clin Microbiol Rev* 2016, **29**(4):837-857.
172. Price J, Gordon NC, Crook D, Llewelyn M, Paul J: **The usefulness of whole genome sequencing in the management of Staphylococcus aureus infections**. *Clin Microbiol Infect* 2013, **19**(9):784-789.
173. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL *et al*: **Genome sequence of Yersinia pestis, the causative agent of plague**. *Nature* 2001, **413**(6855):523-527.
174. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL *et al*: **Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica**. *Nat Genet* 2003, **35**(1):32-40.
175. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT *et al*: **Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18**. *Nature* 2001, **413**(6858):848-852.
176. Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R *et al*: **On the evolutionary history, population genetics and diversity among isolates of Salmonella Enteritidis PFGE pattern JEGX01.0004**. *PLoS One* 2013, **8**(1):e55254.
177. Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C: **High-throughput genome sequencing of two Listeria monocytogenes clinical isolates during a large foodborne outbreak**. *BMC Genomics* 2010, **11**:120.
178. Deng X, den Bakker HC, Hendriksen RS: **Genomic Epidemiology: Whole-Genome-Sequencing-Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens**. *Annu Rev Food Sci Technol* 2016, **7**:353-374.
179. Mohammed M, Delappe N, O'Connor J, McKeown P, Garvey P, Cormican M: **Whole genome sequencing provides an unambiguous link between Salmonella Dublin outbreak strain and a historical isolate**. *Epidemiol Infect* 2016, **144**(3):576-581.
180. Alexander DC, Fitzgerald SF, DePaulo R, Kitzul R, Daku D, Levett PN, Cameron AD: **Laboratory-Acquired Infection with Salmonella enterica Serovar Typhimurium Exposed by Whole-Genome Sequencing**. *J Clin Microbiol* 2016, **54**(1):190-193.

181. Punina NV, Makridakis NM, Remnev MA, Topunov AF: **Whole-genome sequencing targets drug-resistant bacterial infections**. *Hum Genomics* 2015, **9**:19.
182. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, Ayers SL, Lam C, Tate HP, Zhao S: **Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella**. *Antimicrob Agents Chemother* 2016, **60**(9):5515-5520.
183. Metzker ML: **APPLICATIONS OF NEXT-GENERATION SEQUENCING Sequencing technologies - the next generation**. *Nat Rev Genet* 2010, **11**(1):31-46.
184. Nagarajan N, Cook C, Di Bonaventura M, Ge H, Richards A, Bishop-Lilly KA, DeSalle R, Read TD, Pop M: **Finishing genomes with limited resources: lessons from an ensemble of microbial genomes**. *BMC Genomics* 2010, **11**:242.
185. Edwards DJ, Holt KE: **Beginner's guide to comparative bacterial genome analysis using next-generation sequence data**. *Microb Inform Exp* 2013, **3**(1):2.
186. Hardison RC: **Comparative genomics**. *PLoS Biol* 2003, **1**(2):E58.
187. Mellmann A, Andersen PS, Bletz S, Friedrich AW, Kohl TA, Lilje B, Niemann S, Prior K, Rossen JW, Harmsen D: **High Interlaboratory Reproducibility and Accuracy of Next-Generation-Sequencing-Based Bacterial Genotyping in a Ring Trial**. *J Clin Microbiol* 2017, **55**(3):908-913.
188. Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D, Mellmann A: **Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of Listeria monocytogenes**. *J Clin Microbiol* 2015, **53**(9):2869-2876.
189. JB K: **On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem**. *Proc of the Am Math Soc* 1956(7):48–50.
190. Francisco AP, Bugalho M, Ramirez M, Carrico JA: **Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach**. *BMC Bioinformatics* 2009, **10**:152.
191. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO: **Anvi'o: an advanced analysis and visualization platform for 'omics data**. *PeerJ* 2015, **3**:e1319.
192. Chen L, Zheng D, Liu B, Yang J, Jin Q: **VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on**. *Nucleic Acids Res* 2016, **44**(D1):D694-697.
193. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q: **VFDB: a reference database for bacterial virulence factors**. *Nucleic Acids Res* 2005, **33**(Database issue):D325-328.
194. Bhardwaj T, Somvanshi P: **Pan-genome analysis of Clostridium botulinum reveals unique targets for drug development**. *Gene* 2017, **623**:48-62.
195. Haghjoo E, Galan JE: **Salmonella typhi encodes a functional cytolethal distending toxin that is delivered into host cells by a bacterial-internalization pathway**. *Proc Natl Acad Sci U S A* 2004, **101**(13):4614-4619.
196. Weening EH, Barker JD, Laarakker MC, Humphries AD, Tsolis RM, Baumler AJ: **The Salmonella enterica serotype Typhimurium lpf, bcf, stb, stc, std, and sth fimbrial operons are required for intestinal persistence in mice**. *Infect Immun* 2005, **73**(6):3358-3366.
197. Nefise Akkoc BO, Beg¨um G. Tan, Mustafa Akc¸elik: **The role of stj fimbrial operon in the intestinal persistence of Salmonella Typhimurium in mice**. *Biologia* 2009:859-863.
198. Hammar M, Arnqvist A, Bian Z, Olsen A, Normark S: **Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12**. *Mol Microbiol* 1995, **18**(4):661-670.

199. Romling U, Bian Z, Hammar M, Sierralta WD, Normark S: **Curli fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation**. *J Bacteriol* 1998, **180**(3):722-731.
200. Barrow PA, Simpson JM, Lovell MA, Binns MM: **Contribution of Salmonella gallinarum large plasmid toward virulence in fowl typhoid**. *Infect Immun* 1987, **55**(2):388-392.
201. Chart H, Threlfall EJ, Rowe B: **Virulence of Salmonella enteritidis phage type 4 is related to the possession of a 38 MDa plasmid**. *FEMS Microbiol Lett* 1989, **49**(2-3):299-303.
202. Chikami GK, Fierer J, Guiney DG: **Plasmid-mediated virulence in Salmonella dublin demonstrated by use of a Tn5-oriT construct**. *Infect Immun* 1985, **50**(2):420-424.
203. Gulig PA, Curtiss R, 3rd: **Plasmid-associated virulence of Salmonella typhimurium**. *Infect Immun* 1987, **55**(12):2891-2901.
204. Ou JT, Baron LS, Dai XY, Life CA: **The virulence plasmids of Salmonella serovars typhimurium, choleraesuis, dublin, and enteritidis, and the cryptic plasmids of Salmonella serovars copenhagen and sendai belong to the same incompatibility group, but not those of Salmonella serovars durban, gallinarum, give, infantis and pullorum**. *Microb Pathog* 1990, **8**(2):101-107.
205. Feng Y, Liu J, Li YG, Cao FL, Johnston RN, Zhou J, Liu GR, Liu SL: **Inheritance of the Salmonella virulence plasmids: mostly vertical and rarely horizontal**. *Infect Genet Evol* 2012, **12**(5):1058-1063.
206. Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F, Rhen M: **Molecular analysis of spv virulence genes of the Salmonella virulence plasmids**. *Mol Microbiol* 1993, **7**(6):825-830.
207. Guiney DG, Fang FC, Krause M, Libby S: **Plasmid-mediated virulence genes in non-typhoid Salmonella serovars**. *FEMS Microbiol Lett* 1994, **124**(1):1-9.
208. Guiney DG, Fang FC, Krause M, Libby S, Buchmeier NA, Fierer J: **Biology and clinical significance of virulence plasmids in Salmonella serovars**. *Clin Infect Dis* 1995, **21 Suppl 2**:S146-151.
209. Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S: **Multiple insertions of fimbrial operons correlate with the evolution of Salmonella serovars responsible for human disease**. *Mol Microbiol* 1999, **33**(3):612-622.
210. Baumler AJ, Gilde AJ, Tsolis RM, van der Velden AW, Ahmer BM, Heffron F: **Contribution of horizontal gene transfer and deletion events to development of distinctive patterns of fimbrial operons during evolution of Salmonella serotypes**. *J Bacteriol* 1997, **179**(2):317-322.
211. Suez J, Porwollik S, Dagan A, Marzel A, Schorr YI, Desai PT, Agmon V, McClelland M, Rahav G, Gal-Mor O: **Virulence gene profiling and pathogenicity characterization of non-typhoidal Salmonella accounted for invasive disease in humans**. *PLoS One* 2013, **8**(3):e58449.
212. Spano S, Ugalde JE, Galan JE: **Delivery of a Salmonella Typhi exotoxin from a host intracellular compartment**. *Cell Host Microbe* 2008, **3**(1):30-38.
213. Townsend SM, Kramer NE, Edwards R, Baker S, Hamlin N, Simmonds M, Stevens K, Maloy S, Parkhill J, Dougan G *et al*: **Salmonella enterica serovar Typhi possesses a unique repertoire of fimbrial gene sequences**. *Infect Immun* 2001, **69**(5):2894-2901.
214. Bronowski C, Winstanley C: **Identification and distribution of accessory genome DNA sequences from an invasive African isolate of Salmonella Heidelberg**. *FEMS Microbiol Lett* 2009, **298**(1):29-36.
215. den Bakker HC, Moreno Switt AI, Govoni G, Cummings CA, Ranieri ML, Degoricija L, Hoelzer K, Rodriguez-Rivera LD, Brown S, Bolchacova E *et al*: **Genome sequencing**

reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of Salmonella enterica. *BMC Genomics* 2011, **12**:425.

216. Nguyen F, Starosta AL, Arenz S, Sohmen D, Donhofer A, Wilson DN: **Tetracycline antibiotics and resistance mechanisms**. *Biol Chem* 2014, **395**(5):559-575.
217. Hopkins KL, Day M, Threlfall EJ: **Plasmid-mediated quinolone resistance in Salmonella enterica, United Kingdom**. *Emerg Infect Dis* 2008, **14**(2):340-342.
218. Cattoir V, Nordmann P: **Plasmid-mediated quinolone resistance in gram-negative bacterial species: an update**. *Curr Med Chem* 2009, **16**(8):1028-1046.
219. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, Grundman H, Hasman H, Holden MT, Hopkins KL *et al*: **The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee**. *Clin Microbiol Infect* 2017, **23**(1):2-22.
220. Carroll LM, Wiedmann M, den Bakker H, Siler J, Warchocki S, Kent D, Lyalina S, Davis M, Sischo W, Besser T *et al*: **Whole-Genome Sequencing of Drug-Resistant Salmonella enterica Isolates from Dairy Cattle and Humans in New York and Washington States Reveals Source and Geographic Associations**. *Appl Environ Microbiol* 2017, **83**(12).
221. Poupart MC, Chanal C, Sirot D, Labia R, Sirot J: **Identification of CTX-2, a novel cefotaximase from a Salmonella mbandaka isolate**. *Antimicrob Agents Chemother* 1991, **35**(7):1498-1500.
222. Whichard JM, Gay K, Stevenson JE, Joyce KJ, Cooper KL, Omondi M, Medalla F, Jacoby GA, Barrett TJ: **Human salmonella and concurrent decreased susceptibility to Quinolones and extended-spectrum Cephalosporins**. *Emerg Infect Dis* 2007, **13**(11):1681-1688.
223. Michael GB, Butaye P, Cloeckaert A, Schwarz S: **Genes and mutations conferring antimicrobial resistance in Salmonella: an update**. *Microbes Infect* 2006, **8**(7):1898-1914.
224. Chen S, Zhao S, White DG, Schroeder CM, Lu R, Yang H, McDermott PF, Ayers S, Meng J: **Characterization of multiple-antimicrobial-resistant salmonella serovars isolated from retail meats**. *Appl Environ Microbiol* 2004, **70**(1):1-7.
225. Dunne EF, Fey PD, Kludt P, Reporter R, Mostashari F, Shillam P, Wicklund J, Miller C, Holland B, Stamey K *et al*: **Emergence of domestically acquired ceftriaxone-resistant Salmonella infections associated with AmpC beta-lactamase**. *JAMA* 2000, **284**(24):3151-3156.
226. Tran JH, Jacoby GA: **Mechanism of plasmid-mediated quinolone resistance**. *Proc Natl Acad Sci U S A* 2002, **99**(8):5638-5642.
227. Luis Martínez-Martínez AP, George A Jacob: **Quinolone resistance from a transferable plasmid**. *The Lancet* 1998, **351**:797 -799.
228. Robicsek A, Strahilevitz J, Jacoby GA, Macielag M, Abbanat D, Park CH, Bush K, Hooper DC: **Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase**. *Nat Med* 2006, **12**(1):83-88.
229. Perichon B, Courvalin P, Galimand M: **Transferable resistance to aminoglycosides by methylation of G1405 in 16S rRNA and to hydrophilic fluoroquinolones by QepA-mediated efflux in Escherichia coli**. *Antimicrob Agents Chemother* 2007, **51**(7):2464-2469.
230. Kim ES, Hooper DC: **Clinical importance and epidemiology of quinolone resistance**. *Infect Chemother* 2014, **46**(4):226-238.

231. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM: **Foodborne Illness Acquired in the United States-Major Pathogens**. *Emerg Infect Dis* 2011, **17**(1):7-15.
232. Stacy M. Crim PMG, Robert Tauxe, Ellyn P. Marder, Debra Gilliss, Alicia B. Cronquist, Matthew Cartter, Melissa Tobin-D'Angelo, David Blythe, Kirk Smith, Sarah Lathrop, Shelley Zansky, Paul R. Cieslak, John Dunn, Kristin G. Holt, Beverly Wolpert, Olga L. Henao: **Preliminary Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2006–2014**. *Morbidity and Mortality Weekly Report (MMWR), Centers for disease control and prevention (CDC)* 2015, **64**(18):5.
233. Lee CA, Falkow S: **The ability of Salmonella to enter mammalian cells is affected by bacterial growth state**. *Proc Natl Acad Sci U S A* 1990, **87**(11):4304-4308.