

Comparative genome analysis of *Salmonella enterica* serovar Gallinarum biovars Pullorum and Gallinarum decodes strain specific genes

Rajesh Kumar Vaid (✉ rk_vaid@yahoo.com)

National Research Centre on Equines <https://orcid.org/0000-0002-7554-5238>

Zoozeal Thakur

NRCE: National Research Centre on Equines

Taruna Anand

National Research Centre on Equines

Sanjay Kumar

National Research Centre on Equines

Bhupendra Nath Tripathi

National Research Centre on Equines

Research article

Keywords: Poultry, Salmonella Pathogenicity Island, Pan-genome, Prophage, Toxin-antitoxin systems, India

Posted Date: November 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-112825/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Salmonella enterica serovar Gallinarum biovar Pullorum (bvP) and biovar Gallinarum (bvG) are the etiological agents of pullorum disease (PD) and fowl typhoid (FT) respectively, which cause huge economic losses to poultry industry especially in developing countries including India. Vaccination and biosecurity measures are currently being employed to control and reduce the *S. Gallinarum* infections. High endemicity, poor implementation of hygiene and lack of effective vaccines pose challenges in prevention and control of disease in intensively maintained poultry flocks. Comparative genome analysis unravels similarities and dissimilarities thus facilitating identification of genomic features that aids in pathogenesis, niche adaptation and in tracing of evolutionary history. The present investigation was carried out to assess the genotypic differences amongst *S. enterica* serovar Gallinarum strains including Indian strain *S. Gallinarum* Sal40 VTCCBAA614 with a focus on identification of candidate virulence factors. The comparative genome analysis revealed an open pan-genome consisting of 5091 coding sequence (CDS) with 3270 CDS belonging to core-genome, 1254 CDS to dispensable genome and strain specific genes *i.e.* singletons ranging from 3 to 102 amongst the analyzed strains. Moreover, the investigated strains exhibited diversity in genomic features such as genomic islands, prophage regions, toxin-antitoxin cassettes, and acquired antimicrobial resistance genes. The identified genetic differences among the *S. enterica* serovar Gallinarum strains could be used for bacterial typing, design of rapid and reliable diagnostics, structure based inhibitor development and vaccine design for effective infection control as well as eradication and will also form basis for future experimental investigations.

1. Introduction

Pullorum disease (PD) and fowl typhoid (FT) are two distinct septicaemic diseases caused by non-motile *Salmonella enterica* subsp. *enterica* serovar Gallinarum (*S. Gallinarum*) biovar Pullorum (bvP) and biovar Gallinarum (bvG), respectively, which exhibit host-specificity towards poultry and aquatic birds (De Carli et al. 2017; Alves Batista et al. 2018; Wilson et al. 2000). The diseases caused by these invasive avian pathogens cause high morbidity and acute mortality in poultry in India and various countries of Asia, Africa and South America. Pullorum disease (PD) occurs in young birds and persists for long periods in spleen and reproductive tract which leads to high mortality (Wigley et al. 2001). The disease is characterised by white diarrhoea chaperoned with increased antimicrobial resistance (AMR) and high infection rates (Hu et al. 2019; Guo et al. 2019). On the other hand, FT can affect birds of all ages but primarily occurs in adult birds and results in variable morbidity depending on the age, species, and breed of the bird (Barrow & Freitas Neto 2011). *Salmonella* Enteritidis, a major food-borne pathogen causes infection which leads to enteritis in multiple hosts in addition to poultry (Langridge et al. 2015). Although *S. Gallinarum* has negligible importance in humans, *S. Enteritidis* is an important zoonoses (Moutaouakkil et al. 2017; Langridge et al. 2015) The *S. Gallinarum*, in addition has been found to be a recently evolved ancestors of *S. Enteritidis* (Thomson et al, 2008).

The virulence of *Salmonella* spp. is mediated by an arsenal of genes which are capable of invasion, replication, and colonization inside the host cells (Uzzau et al. 2001; Ilyas et al. 2017). The genetic factors involved in pathogenesis of FT and PD at molecular and cellular mechanisms are still under elucidation (Alves Batista et al. 2018; Xu et al. 2016). The genomic sources of virulence systems in *Salmonella enterica* serovars are mainly divided into two elements, one of which horizontally acquired chromosomally located mobile genetic elements known as *Salmonella* pathogenicity Islands (SPIs), and prophage elements, whereas other is plasmids associated with salmonellae (Moreno Switt et al. 2012). These genetic elements have played a seminal role in various ecological niche adaptations in different host and pathogenicity life-style in *Salmonella* by encoding for proteins carrying out cell-adherence, cellular invasion at host level, induction of innate immune and/or pathophysiological responses to infection (Haraga et al. 2008; Jantsch et al. 2011). Two pathogenicity islands, *Salmonella* pathogenicity island 1 (SPI-1) and SPI-2 have been described, which play roles in mediating disease by *Salmonella enterica* through their respective type III secretion systems (TTSS). SPI which encode T6SSs have been detected in SPI6, SPI19, 20, 21 in *S. Gallinarum* and *S. Enteritidis* (Blondel et al. 2009).

Additionally, prophage lysogeny is a rich contributor of *Salmonella* genome diversity, and their acquisition leads to enhanced virulence and pathogenicity (Figueroa-Bossi et al. 2001; Thomson et al. 2004). The prophage elements are being increasingly utilised as molecular markers for strain discrimination (Brüssow et al. 2004; Colavecchio et al. 2017). Another crucial genetic element that is involved in persistence, virulence and AMR includes toxin-antitoxin system (TA system) (Huguet et al. 2016; Di Cesare et al. 2016; VanDrissse et al. 2017). Furthermore, the *Salmonella* serovars including the Gallinarum and Enteritidis are increasingly reporting incidence of antimicrobial resistance (AMR) (Pan et al. 2009; Jain et al. 2020). The comparative genomic analysis of pathogens helps in elucidating the diversity and depth of functional set of genes in form of pan-genome and decoding of evolutionary history (Rouli et al. 2015; Sahl et al. 2013). Moreover, it aids in functional annotation, as well as decodes molecular mechanisms underlying pathogenesis and niche adaptation by precise measurement of genetic variation within and between pathogenic groups (Lu et al. 2019; Rasko et al. 2008; Armstrong et al. 2019).

Recently we determined the first draft genome sequence of an Indian strain of *S. Gallinarum* VTCCBAA614 isolated from diseased poultry (Vaid et al. 2015). The present investigation dealt with whole genome characterization and assessment of genotypic differences amongst bvG and bvP strains including *in house* *S. Gallinarum* Sal40 VTCCBAA614 strain (Table 1). The study unravelled pan-genome, core-genome, dispensable genome, and strain specific genes of the analyzed *S. Gallinarum* strains. Moreover, the investigation decoded diversity in genomic islands, prophage regions, TA systems, and acquired AMR genes in the investigated strains via employment of various computational tools. The genotypic differences revealed by the study will serve as the genomic resource for the identification and discrimination of biovars, development of rapid diagnostics, smart vaccines, and could form the basis of structure based inhibitors design and that would be beneficial to poultry industry. Although comparative genomics

analysis of host-adapted salmonellae genomes have elucidated various evolutionary and virulence themes (Langridge et al. 2015; Hu et al. 2019), this is the first report for the assessment of genotypic differences, which includes an Indian strain among the investigated strains from America and China.

Table 1

Genome statistical information of investigated *Salmonella* Enteritidis and *S. Gallinarum* strains obtained by Prokka pipeline and ANI calculator

| #Organism name | Geographic location | Accession number | Assembly status | Total length | Size (MB) | GC Content % | CDS | tRNA | tmRNA | rRNA | rr |
|--|---------------------|------------------|-----------------|--------------|-----------|--------------|------|------|-------|------|----|
| <i>S. enterica</i> Enteritidis str. P125109 | United Kingdom | NC_011294.1 | Complete | 4,685,848 | 4.68 | 52.17 | 4352 | 85 | 1 | 22 | 2 |
| <i>S. enterica</i> Gallinarum str. 287/91 | Brazil | NC_011274.1 | Complete | 4,658,697 | 4.65 | 52.20 | 4452 | 77 | 1 | 22 | 2 |
| <i>S. enterica</i> Gallinarum str. 9184 | Not available | NZ_CP019035.1 | Complete | 4,609,911 | 4.60 | 52.20 | 4399 | 78 | 1 | 22 | 2 |
| <i>S. enterica</i> Gallinarum str. Sal40 | India | JSWQ00000000.1 | Contig | 4,598,206 | 4.59 | 52.22 | 4455 | 68 | 1 | 4 | 2 |
| <i>S. enterica</i> Pullorum str. ATCC 9120 | USA | NZ_CP012347.1 | Complete | 4,694,842 | 4.69 | 52.19 | 4474 | 81 | 1 | 23 | 2 |
| <i>S. enterica</i> Pullorum str. S06004 | China | NC_021984.1 | Complete | 4,682,599 | 4.68 | 52.14 | 4635 | 80 | 1 | 22 | - |
| <i>S. enterica</i> Pullorum str. QJ-2D-Sal | China | NZ_CP022963.1 | Complete | 4,728,875 | 4.72 | 52.17 | 4548 | 75 | 1 | 22 | 2 |
| <i>S. enterica</i> Gallinarum/Pullorum str. CDC1983-67 | China | NC_022221.1 | Complete | 4,623,089 | 4.62 | 52.23 | 4397 | 78 | 1 | 22 | 2 |
| <i>S. enterica</i> Gallinarum/pullorum str. RKS5078 | Not available | NC_016831.1 | Complete | 4,637,962 | 4.63 | 52.21 | 4451 | 75 | 1 | 22 | 2 |

2. Material And Methods

2.1 Data collection of *Salmonella* genomes:

The complete genome sequences of eight *S. Gallinarum* strains *i.e.*, *S. Gallinarum* str. 287/91 (NC_011274.1), *S. Gallinarum* str. 9184 (NZ_CP019035.1), *S. Pullorum* str. ATCC 9120 (NZ_CP012347.1), *S. Pullorum* str. S06004 (NC_021984.1), *S. Pullorum* QJ-2D-Sal (NZ_CP022963.1), *S. Gallinarum/pullorum* str. CDC1983-67 (NC_022221.1), and *S. Gallinarum/pullorum* str. RKS5078 (NC_016831.1) including draft genome sequence of *S. Gallinarum* Strain VTCCBAA614 (isolated from chicken in India from FT inflicted broiler flock) were mined from NCBI database to decode genotypic differences among the *S. enterica* serovar *Gallinarum* strains. In addition the genome sequence of *S. Enteritidis* str. P125109 (NC_011294.1)(Thomson et al. 2008) a well studied pathogenic *Salmonella* strain was also retrieved from NCBI for employment as reference for comparative genome analysis (Table 1). The Indian strain *S. Gallinarum* VTCCBAA614 (Vaid et al. 2015) is available from our culture collection at NCVTC, NRCE Hisar. All the sequences were extracted in fasta and GenBank (gbk) format.

2.2 Genome statistics and visualization:

Genomic features such as number of CDS, tRNA, tmRNA, rRNA and repeat region (rr) were determined by running Prokka pipeline at UseGalaxy webinterface (Table 1) (Seemann 2014). The GC% of the genomes was determined by ANI Calculator which uses OrthoANIu algorithm (Yoon et al. 2017). In addition, circular plot visualization of CDS, GC content, and GC skew with *S. Enteritidis* str. P125109 as reference genome against other *Salmonella* genomes was generated by EDGAR (Blom et al. 2016).

2.3 Pan- and core-genome calculations:

Pan-genome is the complete set of orthologous genes (OGGs) harboured within a collection of investigated genomes, whereas, the core-genome refers to the set of genes present in all the genomes of a collection. On the other hand, dispensable genome refers to the set of genes harboured by one or a subset of investigated genomes. Singleton genes are the unique genes that do not have any homologs in the investigated genomes (Tettelin et al. 2005; Carlos Guimaraes et al. 2015). EDGAR served as the resource for pan-genome, core-genome, and singleton genes calculations for the selected *Salmonella* strains (Blom et al. 2016). A customised project was set up by EDGAR to conduct all the calculations for pan-genome, core-genome, dispensable genome, and singletons with selection of *S. Enteritidis* str. P125109 as reference genome.

2.4 Functional annotation of pan- and core-genome

The functional annotation of core-genome and complete set of strain-specific genes via orthology assignment was carried by eggNOG-mapper v2 online portal (<http://eggno-mapper.embl.de>) (Huerta-Cepas et al. 2017, 2019). The tool employs precomputed clusters and phylogenies from its *in house* database and provides orthology assignment to a large set of sequences via fast orthology mapping. Out of the total 29,430 CDS that comprised core-genome of nine investigated *Salmonella* genomes, 29,125 (98.96%) CDS were queried by eggNOG-mapper v2 for orthology mapping. On the other side, 153 (48.11%) out of the total 318 strain specific CDS (singletons) were queried for orthology mapping.

2.5 Phylogenetic analysis of the strains

The identified core-genome of the investigated *Salmonella* strains was employed by EDGAR to decipher phylogenetic relationship among the strains. Alignment of each CDS set was obtained by MUSCLE (Edgar 2004), and then further joined to form one huge alignment. EDGAR employs neighbour joining algorithm, and FastTree software for phylogenetic tree construction. Conservation of gene order and genome rearrangements among *Salmonella* genomes were also explored by using EDGAR wherein *S. Enteritidis* P125109 was chosen as a reference to create synteny plots. The Indian strain, bvG VTCCBAA614, was not included, as the use of draft genomes is not recommended in synteny plot representation.

2.6 Detection of *Salmonella* pathogenicity islands (SPI)

Two approaches were employed to detect SPI's -the major determinant of *S. enterica* virulence via usage of web tool SPIFinder 1.0 available at (<https://cge.cbs.dtu.dk/services/SPIFinder>) (Roer et al. 2016) with default parameters, and NCBI BLAST search of known SPIs extracted from PAIDB v2.0 (<http://www.paidb.re.kr/>) against investigated genomes (Yoon et al. 2005, 2007).

2.7 Identification and *in silico* characterization of prophage sequences:

Potential prophage sequences within the *Salmonella* genomes were detected and annotated by employing PHASTER (Phage Search Tool Enhanced Release). PHASTER, an improved version of PHAST phage search tools detects candidate prophage regions in the bacterial genomes and then classifies the identified regions into three classes *i.e.* intact, incomplete and questionable on the basis score obtained (Arndt et al. 2016; Zhou et al. 2011).

2.8 Detection and analysis of Type II Toxin-Antitoxin (TA) gene cassettes

Type II TA loci were predicted in the *Salmonella* genomes by employing TAFinder (Makarova et al. 2009). The parameters used for the detection were BLAST e value: 0.01; HMMer E-value: 1; Maximum length for candidate toxin/antitoxin: 300; maximum overlap between candidate toxin and antitoxin: (-20–150).

2.9 Screening of acquired antimicrobial resistance (AMR) genes

Resfinder available at (<https://cge.cbs.dtu.dk/services/ResFinder/>) was employed to investigate the presence of acquired antimicrobial resistance (AMR) genes in the selected *Salmonella* strains (Zankari et al. 2012). Resfinder identifies acquired antibiotic resistance genes for aminoglycoside, β -lactam, colistin, fluoroquinolone, fosfomycin, fusidic acid, glycopeptide, MLS- macrolide, lincosamide and streptogramin B nitroimidazole, oxazolidinone, phenicol, rifampicin, sulphonamide, tetracycline and trimethoprim. Moreover, both known and unknown chromosomal point mutations were also detected in AMR resistance genes such as *gyrA*, *parE*, *pmrA*, *pmrB*, *pmrB*, *parC* and 16S_*rrsD*.

3. Results

3.1 Comparative genome statistics

Comparative genome analysis of selected nine strains of *Salmonella* species (Table 1) was carried out by employing various bioinformatics tools. The average genome size of the investigated *Salmonella* strains was 46,57,781 bp, ranging from 4,598,206 bp (*S. Gallinarum* Sal40 strain VTCCBAA614) to 47,28,875 bp (*S. Pullorum* QJ-2D-Sal). The average GC content of all the analyzed genomes was 52.19%, ranging from 52.14% (*S. Pullorum* str. S06004) to 52.23% (*S. Gallinarum/pullorum* str. CDC1983-67) (Table 1). Genomic features of the investigated *Salmonella* strains were determined by running Prokka pipeline at Use Galaxy web interface. The average number of CDS was observed to be 4463 with highest number detected in *S. Pullorum* str. S06004 (4635) and the lowest in *S. Enteritidis* str. P125109 (4352) (Table 1). The tRNA sequences identified by Prokka pipeline in the *S. Gallinarum* genomes fell in the range of 75–81 *i.e.* 75 in *S. Gallinarum/pullorum* str. RKS5078 and 81 in *S. Pullorum* ATCC 9120. It was 85 in *S. Enteritidis* str. P125109 (Table 1). In addition, all the investigated *Salmonella* genomes were detected to possess only one tmRNA. Notably, only four rRNA sequence were detected in *S. Gallinarum* Sal40 strain VTCCBAA614 in comparison to possession of 23 rRNA sequences in *S.*

Pullorum str. ATCC 9120 and presence of 22 rRNA sequences in rest of the genomes (Table 1). The circular plot visualization of investigated genomes with reference to *S. Enteritidis* str. P125109 depicts varied GC content and GC skew along with predominant similarity of core-regions of the investigated genomes (Fig. 1).

3.2 Pan and core-genome analysis unravels strain specific genes

The pan-genome calculation by customised set up in EDGAR led us to have discrete idea of the total genetic repository of the *Salmonella* genomes under study. The pan-genome of *S. Gallinarum* was found to be in open state with growth exponent value of 0.089 (95% confidence interval 0.084 to 0.093) (Suppl Table 1). The pan-genome development plot showed steady growth with addition of each new genome and reached 5091 on addition of ninth genome which is nearly 1.1 times the average number of genes of nine strains (Fig. 2a, Suppl Table 1). On the other hand, core-genome development plot became limited to 3270 genes, with shared genes decreasing with new genome addition. The extrapolated core-genome size was 2684 (95% confidence interval 2507.66 to 2861.11) (Fig. 2b, Suppl Table 1). In addition, singleton development plot also indicated *S. enterica* serovar Gallinarum to be in an open pan-genome state as 43 new genes were predicted to be found at every genome addition. (Fig. 2c, Suppl Table 1)

The pan-genome of investigated *Salmonella* strains with the reference to *S. Enteritidis* str. P125109 strain is composed of 5,091 coding sequences (CDS) which include a core-genome of 3,270 (64.2%) CDS, a dispensable genome of 1254 (24.6%) CDS, and 567 (11.1%) singletons. The complete list of CDS detected as part of pan-genome, core-genome, and singletons with their function are enlisted in Suppl Table 2, 3, 4 (a-i). Notably, a total of 318 CDS in the range of 3-102 (Table 2, Suppl Table 4) were found as strain specific genes aka singletons in the investigated genomes. Significantly, *S. Gallinarum* Sal40 strain VTCCBAA614 harboured the highest number (102) of singletons among the investigated *Salmonella* strains (Table 2, Suppl Table 4). The singletons detected were identified to be as hypothetical proteins (61), transposase (13), membrane proteins (8), ATP binding proteins (3) among others. Interestingly, genomes of *S. Enteritidis* str. P125109 and *S. Pullorum* QJ-2D-Sal possessed 96 and 85 singletons each, respectively.

Table 2
CDS identified in pangenome, core-genome and singletons of the investigated *Salmonella* genomes

| Organism and strain | Accession number | Pan genome | Core genome | No of singletons |
|--|------------------|------------|-------------|------------------|
| <i>S. enterica</i> Enteritidis str. P125109 | NC_011294.1 | 5091 | 3270 | 96 |
| <i>S. enterica</i> Gallinarum str. 287/91 | NC_011274.1 | | | 3 |
| <i>S. enterica</i> Gallinarum str. 9184 | NZ_CP019035.1 | | | 4 |
| <i>S. enterica</i> Gallinarum str. Sal40 strain | JSWQ00000000.1 | | | 102 |
| <i>S. enterica</i> Pullorum str. ATCC 9120 | NZ_CP012347.1 | | | 10 |
| <i>S. enterica</i> Pullorum str. S06004 | NC_021984.1 | | | 10 |
| <i>S. enterica</i> Pullorum str. QJ-2D-Sal | NZ_CP022963.1 | | | 85 |
| <i>S. enterica</i> Gallinarum/pullorum str. CDC1983-67 | NC_022221.1 | | | 3 |
| <i>S. enterica</i> Gallinarum/pullorum str. RKS5078 | NC_016831.1 | | | 5 |

On the other hand, *S. Pullorum* str. ATCC 9120 and *S. Pullorum* str. S06004 genomes harboured ten singletons each, whereas, the rest of the genomes of possessed ≤ 5 singletons each (Table 2, Suppl Table 4(a-i)).

Functional annotation of core-genome and strain specific CDS by assignment of orthology (COG) was performed by eggNOG-mapper v2. The annotation by eggNOG-mapper provided detailed description of GO term, EC number, annotation level, COG category and description (Suppl Table 5). 29,125 CDS (98.96%) out of the total 29,430 (core-genome sequences) were queried by eggNOG-mapper v2 for orthology mapping. The highest number among them belonged to the functional classes of function unknown (5923) transcription (2592), amino acid metabolism and transport (2196), energy production and conversion (2178), and cell wall/membrane/envelop biogenesis (2133) (Fig. 3a, Suppl Table 5a). On the other hand, 153 out of the 318 CDS (complete set of identified singleton sequences) were queried by eggNOG-mapper v2. The highest number amongst them belonged to the classes of function unknown (68), replication and repair (24) and defence mechanism (6) among others. Thirty five singletons were not provided any COG category by eggNOG-mapper (Fig. 3b, Suppl Table 5b).

3.3 Phylogenetic and synteny plot analysis

A phylogenetic tree was constructed on the basis of core-genome comprising of 9,83,299 aa-residues/bp per genome and 88,49,691 in total of nine *Salmonella* genomes by EDGAR (Fig. 3) using *S. enterica* serotype Enteritidis strain P125109 as a reference sequence to demonstrate the evolutionary relationship among the *S. Gallinarum* strains used in the study. The bvG and bvP formed two distinct and strongly supported clades in

the phylogeny (Fig. 4). In concordance with their taxonomic classification bvG strains clustered together *i.e.* *S. Gallinarum* str. 287/91, *S. Gallinarum* str. 9184 and *Gallinarum* Sal40 strain VTCCBAA614) and formed one clade. Whereas, the rest of the bvP strains bifurcated into 2 clades (Fig. 4).

The synteny plot analysis performed by EDGAR of the investigated *Salmonella* genomes in reference to *S. Enteritidis* str. P125109 depicted large scale genomic rearrangements which included relocations, inversions, duplications and deletions (Fig. 5). It was observed that all bvP strains genomes showed high degree of conservation of gene order among genomes, The highly similar genomic rearrangement within bvP comprised of inversion and duplication. However, the bvG strain 9184 was characterized by large region of inversion, which was not observed in bvG 287/91 in reference to *S. Enteritidis* str. P125109 strain. Synteny plot of individual investigated strains against the reference genome of *S. Enteritidis* str. P125109 is shown in Suppl Fig. 1.

3.4 Detection of Genomic islands (GIs):

A total of 113 homologs of SPIs with an average of 14.77 SPIs were detected in the investigated *Salmonella* genomes by the combined usage of web tool SPIFinder 1.0 and NCBI BLAST search. The integrated approach revealed presence of high sequence identity SP-1, SP-2, SP-3, SP-4, SPI-5, SPI-12, SPI-13, SP-14, C63PI and SPCS54 islands in their genomes (Suppl Table 6a-b). The detected SPIs varied in the range of 0.3 to 41.8 kb in terms of size. Interestingly, 3 SPI-13 (AY956834, AY956833, AY956832) and 2 SPI-14 (AY956835, AY956836) were observed in each of the investigated genome with the exception of *S. Gallinarum* str. 9184 genome, wherein SPI-14 (AY956836) (0.4 kb) was not detected. Moreover, SPI-12 was also not observed in *S. Gallinarum* Sal40 strain VTCCBAA614. Notably, various other SPIs and resistance islands reported in *S. enterica* serovar such as SESS LEE, SGI-1, SPI-10, SPI-11, SPI-2, SPI-6, SPI-7, and HP1 whose sequence were extracted from PAIDB database were not found to be present by both the approaches. The features of detected SPI homologs in their respective genomes such as starting position, end position, size, and external annotation if any are detailed in (Suppl. Table 6a-b).

3.5 Identification and analysis of prophage and prophage remnant regions:

Each of the analyzed *Salmonella* genome was detected to possess at least one candidate prophage region by PHASTER (Suppl. Figure 2, Suppl Table 7). In total, 23 prophage regions were identified in which fifteen were classified as putative intact regions and eight as incomplete prophage regions by PHASTER (Suppl. Figure 2, Suppl Table 7a). The detailed information of the identified prophage regions *i.e.*, region name, region length, completeness, score, total protein, region position and GC percentage in the respective *Salmonella* genomes as determined by the web server are enlisted in Suppl Table 7a. The highest number of prophage region were detected in *S. Pullorum* QJ-2D-Sal and *S. Pullorum* str. S06004 with each possessing three complete prophages and one incomplete prophage region. In total, there were 5 different prophage elements with Gifsy_2 being the most common, which was detected in all genomes analysed. On the other hand, all 3 bvG possessed only a single (Gifsy_2) intact prophage element. Whereas bvP harboured a variety of phage elements including Gifsy_2. The average size of identified prophage regions was 36.31 kb in which largest (64.3 kb) and smallest (8.5 kb) candidate prophage regions were observed in *S. Enteritidis* str. P125109. On the other hand, the average size of identified complete prophage sequences in the analyzed *Salmonella* genomes was found to be 44.98 kb with lowest size of 29.2 kb belonging to *S. Pullorum* str. ATCC 9120. (Suppl Table 7a, b). Whereas, largest intact prophage region of 64.3 kb was observed in *S. Enteritidis* str. P125109 genome. The GC percentage in identified prophage region varied from 45.59–53.21% (Suppl Table 7a). On the other hand GC% of complete prophage regions varied from 47.23–53.21% (Suppl Table 7b).

3.6 Presence of Toxin antitoxin cassettes:

TA loci were identified in all the investigated *Salmonella* genomes. A total of 149 Type II TA loci were identified, ranging from 16 to 18 in each of the genome analysed. (Fig. 6, Suppl Table 8). The length of the identified toxin and antitoxin proteins varied from 73–475 and 55–319 aa, respectively each. Majority of the identified candidate toxin proteins (58) harboured relElike_domain in their structure. While, other candidate toxin proteins showed the presence of GNATlike_domain (24), MazFlike_domain (9), COG2929like_domain (9), yeeU (8), doc (5), PINlike_domain (3), and NULL (24) in their structures respectively (Fig. 6, Suppl Table 8). On the other hand, identified candidate antitoxin proteins majorly possessed RHHlike_domain (60). Whilst, other domains that were observed included Xrelike_domain (27), COG5606like_domain (9), YhfGlike_domain (9), yeeU (8), PHDlike_domain (5), doc (4), AbrBlike_domain (3) and NULL (24) in rest of the investigated candidate antitoxins (Fig. 6, Suppl Table 8). The detailed description of identified TA loci in *Salmonella* genomes which includes location in the genome, length, strand, family and domain are summarised in Suppl Table 8.

3.7 Detection of acquired antibiotic resistance genes/chromosomal point mutations:

Interestingly, all the analyzed *Salmonella* genomes harboured acquired aminoglycoside resistance gene (aac(6')-Iaa) (Suppl Fig. 3, Suppl Table 9) as determined by Resfinder. The information related to identity, contig and position of the identified aminoglycoside gene in their respective genomes is enlisted in Suppl Table 9. On the other hand, no acquired resistance gene for beta-lactam, colistin, fluoroquinolone, fosfomycin, fusidic acid, glycopeptide, MLS- macrolide, lincosamide and streptogramin B nitroimidazole, oxazolidinone, phenicol, rifampicin, sulphonamide, tetracycline, and trimethoprim was detected in the analyzed *Salmonella* genomes by Resfinder. Strikingly, known chromosomal point mutation in *gyrA* was detected in 2 bvG and 2 bvP genomes out of the nine analysed *i.e.*, *S. Gallinarum* str. 287/91, *S. Gallinarum* Sal40 strain VTCCBAA614, *S. Pullorum* str. S06004

and *S. Pullorum* QJ-2D-Sal (Suppl Table 9). In addition, unknown mutations were detected in 16S_rrsD gene of all the analysed genomes (Suppl Table 9). Moreover, unknown mutations in *parE* were detected by Resfinder in *S. Gallinarum* Sal40 strain VTCCBAA614 and *S. Pullorum* str. S06004. A total of six antimicrobial resistance genes (acquired, known and unknown chromosomal point mutations) were identified among the nine investigated strains.

4. Discussion

Salmonella enterica subsp. *enterica* comprises of both host-adapted and host-promiscuous pathotypes that causes a spectrum of diseases depending on the serovar or host (Jajere 2019; Cheng et al. 2019). Pullorum Disease (PD) and Fowl Typhoid (FT) caused by *S. Gallinarum* biovars Pullorum (bvP) and Gallinarum (bvG) respectively are endemic in countries of Asia and South America leading to huge economic losses to poultry industry (Shivaprasad 2000; De Carli et al. 2017; Hu et al. 2019). It is interesting to note that as compared to FT, the PD reports have been low in India (Rajagopal and Mini, 2013). On the other hand, PD is frequent in China, being prevalent in every province (Hu et al. 2019). Biosecurity and vaccinations applied in combination are important strategies to prevent and control these diseases in commercial and breeding flocks. Most commercially available vaccines are killed vaccines, although vaccination strategies for effective *Salmonella* control mainly include use of live attenuated strains capable of inducing a cellular immune response (Mastroeni et al. 2001) still concerns about poor protection, lack of understanding of genetic basis of attenuation and residual pathogenicity remain (Silva et al. 1981; Kwon & Cho 2011; Koerich et al. 2018). A progress can be made by a thorough understanding of the genetic virulence makeup of prevalent strains. Recently, we had obtained the first draft genome sequence of an Indian strain of *S. Gallinarum* VTCCBAA614 isolated from diseased poultry (Vaid et al. 2015). In the present investigation, we performed comparative genome analysis of eight strains of *S. Gallinarum* bvG and bvP originating from different countries (Brazil, USA, China and India) including *in house* strain with host-promiscuous *S. Enteritidis* str. P125109 taken as reference genome to characterize their genomes and gain insights into pan genome, pathogenesis, mobilome, resistome, and taxonomy.

For a comprehensive analysis in an open pan-genome, determination of minimum number of necessary strains is difficult (Rouli et al. 2015). The present analysis entails sufficient genotypic variability, as our selected eight genomes represent a single serovar Gallinarum of *Salmonella enterica* ssp. *enterica*, and the total SNP count in each genome varied from 7424 to 12,726 (data not shown). The pan-genome of investigated strains of *S. Gallinarum* comprises of 5091 CDS with a core-genome of 3270 CDS and an indispensable genome of 1254 CDS (Table 2). With a R_{cp} of 64%, the genomes depict a fairly high degree of genomic diversity and asymmetry (Ghatak et al. 2016).

The pan-genome and core-genome development plot analysis in the present investigation revealed *S. Gallinarum* to possess an open pan-genome state (Fig. 2a, b). A steady growth with addition of new genomes is depicted in the pan-genome development plot, which indicates the capacity of this sympatric species to rapidly acquire exogenous DNA (Medini et al. 2005). The core development plot becomes limited to about 2600 genes (Fig. 2b). The acquisition of exogenous DNA poses obvious challenge for control strategies against FT/PD, however core-genome also offers insight for vaccine candidates. It is important to examine these core genes for reverse vaccinology application as their presence in all strains and high degree of conservation makes them effective source of potentially universal antigens (Rinaudo et al. 2009). Although the dispensable genes are strain-restricted, they can also be explored and exploited for immune-antigens (Rinaudo et al. 2009) (Fig. 3, Suppl Table 5b). Interestingly, strain specific genes were also detected in all the investigated genomes in the range of 3-102 (Suppl Table 4). Furthermore, singleton development plot demonstrated the possibility of finding 43 new genes with each newly sequenced genome (Fig. 2c). Indian strain *S. Gallinarum* Sal40 strain VTCCBAA614 harboured the highest number of singletons *i.e.* 102 whilst *S. Enteritidis* str. P125109 possessed 96 singletons (Suppl Table 4). Whether the presence of large number of singletons in *S. Gallinarum* Sal40 strain VTCCBAA614 and *S. Pullorum* QJ-2D-Sal, in comparison to other investigated strains is related to higher acquisition of foreign genes requires to be investigated.

Lower core-genome to pan-genome ratio and presence of large number of singletons in *S. Enteritidis* str. P125109 and other investigated strains suggests the dynamic state of *S. Gallinarum* genome. *Salmonella* Gallinarum has undergone extensive degradation and also acquired specific genes related to virulence, non-virulence, metabolism, information storage by horizontal gene transfer which may contribute to its avian host specificity (Bacciu et al. 2004; Thomson et al. 2008; Rouli et al. 2015).

Studies on very large number of *Salmonella enterica* genomes such as one employing 4893 genomes detected a pan-genome of 25.3 Mbp, a strict core of 1.5 Mbp present in all genomes, and a conserved core of 3.2 Mbp found in at least 96% of these genomes (Laing et al. 2017). This makes for a strong case for creating a database of serovar specific genome components such as for *S. Gallinarum*, a host specific pathogen.

Following core-genome identification, its functional analysis by employment of eggNOG-mapper v2 revealed that large number of the CDS (20%) belonged to function unknown category (5923), which indicates that our understanding of genetic repertoire of *Salmonella* genome is limited. It is probable that a sizeable number of such CDS may be mutant phenotypes for protein-coding genes (Price et al. 2018). Otherwise the maximum genes were related to transcription, cell-wall and membrane biogenesis, energy metabolism, amino acid metabolism and transportation among others (Fig. 3, Suppl Table 5a). The identified core-genome is also dominated by ATP-binding cassette (ABC) exporters which could be explored as novel drug targets (Rempel et al. 2020). On the other hand, the functional annotation of complete set of strain specific CDS of investigated strains fell in the categories of function unknown, replication and repair, and defence mechanism among others (Fig. 3, Suppl Table 5b).

The core-genome based phylogeny (Fig. 4) of *S. Gallinarum* biovars. from different geographical origins has clearly divided strains into divergent clades, and Indian strain was grouped externally in the *S. Gallinarum* clade, as in previous study (De Carli et al. 2017). Similarly, the phylogenetic relationship between *S. Gallinarum* biovars and *S. Enteritidis* has been genomically analysed, which highlighted the shared ancestry of bvP and bvG with *S. Enteritidis* descendent originating in a 'second' clade (Langridge et al. 2015). In a recent study, the core-genome SNP phylogeny of only bvP strains from China has divided it into four lineages (Hu et al. 2019), which suggests that the study of within biovar genomic epidemiology of Indian bvG strains will similarly need inclusion of more number of local strains. The genomic diversity within the investigated *S. Gallinarum* strains is supported by large scale genomic rearrangement events including inversions, deletions, relocations and duplications detected in synteny analysis. (Fig. 5, Suppl Fig. 1).

The evolution of host-adapted serovars of salmonellae has traversed a path of periodic gene acquisition and gene disruption as a mechanism of niche adaptation away from enteric location to systemic one (Langridge et al. 2015). These genetic elements contribute to the bacterial virulence, pathogenesis, fitness in specific niche, and evolution (Langridge et al. 2015; Ilyas et al. 2017). So, we next focussed our investigation on identification and analysis of candidate virulence factors that were previously not identified in the genomes of the analyzed strains which included genomic islands, prophages, TA cassettes and AMR genes. The acquisition of virulence capability enhancement en masse by acquisition of SPI has been hall mark of *S. Gallinarum* (Groisman & Ochman 1996; Blondel et al. 2009). A total of 113 SPI homologs with nearly 15 SPIs per genome were detected by combinatorial usage of SPIfinder and BLAST searching (Suppl Table 6a-b), which uncovered the presence of homologs of SP-1, SP-2, SP-3, SP-4, SPI-5, SPI-12, SPI-13, SP-14, C63PI and SPCS54 island in the investigated genomes (Suppl Table 6a-b). The presence of particular SPIs in each of the genomes indicates their specificity towards *S. Gallinarum* and *S. Enteritidis* and thus corroborates with previously published reports of SPIs showing serovar specificity (Roer et al. 2016; Hensel 2004; Gerlach & Hensel 2007). Notably, the role of SPI's other than SP1 and SP2 in virulence and pathology of *S. Gallinarum* infections has not been analyzed in detail and thus requires further experimental investigations with functional knockout mutations (Jones et al. 2001; Wigley et al. 2002). Notably, SPI-12 (NC_006905_P4, 11.1 kb), and SPI-14 (AY956836, 0.4 kb) were not detected in *S. Gallinarum* Sal40 strain VTCCBAA614S and *Gallinarum* str. 9184, respectively. SPI-12 has been shown to be crucial for bacterial survival in the host (Tomljenovic-Berube et al. 2013). SPI-14 has been recently identified as being involved in *Salmonella* intestinal survival and invasion systemic via activation of SPI-1 genes (Jiang et al. 2017).

Prophages are reported to encode for virulence proteins, toxins, and antimicrobial resistance genes (Brüssow et al. 2004; Colavecchio et al. 2017) all of which contribute to diversity of strains. In our study, the intact prophages were detected in the range of 1–3 with the highest number in bvP QJ-2D-Sal and bvP str. S06004. As in our study, a recent study also detected only a single phage in all of their bvG genomes, which is less than 5 ± 3 prophage regions per genome detected in 1,760 *S. enterica* genomes, using Phaster (Mottawea et al. 2018). However the observed prophage diversity in our study is in contrast with lower diversity reported by (Matthews et al. 2015), using 2 strains of *S. Gallinarum*. The gifsy_2 prophage element which was detected in all genomes analysed in this study was earlier detected predominantly in *S. Typhimurium* genome (Mottawea et al. 2018). Significantly prophage element ST104 detected in bvP QJ-2D-Sal has been previously reported in MDR *S. Typhimurium* DT104 as phage ST104. Notably, prophage and plasmids acquisition has been demonstrated in *S. Pullorum* leading to emergence of multi-drug resistant (MDR) strains (Tanaka et al. 2004; Hu et al. 2019). Our study also highlights variable number of propahges detected in bvP as compared to single prophage element detected in bvG. Hu et al, 2019 also detected multiple prophage elements in their bvP genomes, which leads to high genetic diversity besides AMR (Hu et al. 2019). The prophage diversity revealed in the study can be employed for *Salmonella* typing and epidemiology as additional biomarkers (Mottawea et al. 2018). However, experimental investigation is required to check the induction and characterization of identified prophages.

Toxin-antitoxin operons which are implicated in pathogenicity, virulence, persistence, stress endurance, and antibiotic resistance in bacteria and archae were searched for in the investigated *Salmonella* genomes (Chimal-Cázares et al. 2020; Thakur et al. 2018; Yamaguchi et al. 2011). The computational search revealed a total of 149 TA cassettes in the range of 16–18 in the investigated *S. Gallinarum* genomes (Suppl Table 8). Significantly, a majority of the toxins of identified TA systems belonged to relElike family and highest number of identified antitoxins possessed RHH like domain (Suppl Table 8, Fig. 7). The *in silico* identified TA systems also requires to be experimentally characterized before considering them as bonafide TA operons by designing assays to test whether toxicity of over-expressed toxin protein of TA pair is abrogated by over-expression of putative antitoxin protein in heterologous systems (Ramage et al. 2009; Lobato-Márquez et al. 2015; Chan et al. 2014). Moreover, experimental investigation is required to decode their role in pathogenesis. The *Salmonella* TA systems can be considered as potent targets for structure based inhibitor design as has been previously employed against various other bacteria (Chan et al. 2015; Równicki et al. 2018).

Although plasmids are important carriers of AMR genes in *Salmonella* serovars, (Feng et al, 2012), we detected the presence of acquired aminoglycoside resistance gene (aac(6')-Iaa) in all the investigated *Salmonella* genomes (Suppl. Figure 3, Suppl Table 9a). The observation is in concordance with previously published reports stating that aminoglycosides exhibit weak bactericidal activity against intracellular *S. enterica* serovars (Menashe et al. 2008). Strikingly, known mutation in *gyrA* were also observed in four genomes out of the nine investigated i.e., *S. Gallinarum* str. 287/91, *S. Gallinarum* Sal40 strain VTCCBAA614, *S. Pullorum* str. S06004 and *S. Pullorum* QJ-2D-Sal (Suppl Table 9b). Koerich et al (2018) also reported high resistance to drugs from macrolide and quinolone groups in *Salmonella* *Gallinarum* field isolates. Literature mining has revealed that mutations in *gyrA* are associated with resistance to fluoroquinolones in eight species of *Enterobacteriaceae*. (Weigel et al. 1998). Antimicrobial resistance in *Salmonella* serovars is a serious poultry husbandry and public health problem all over the world including Asian, and

South American region (Taddele et al. 2012; Parveen et al. 2007; Li et al. 2015; Koerich et al. 2018) and NGS based mapping of AMR genes in *S. Gallinarum* serovars is an important step in this direction for their control (Hendriksen et al. 2019).

A number of investigators have previously carried out comparative genome analysis of *S. Gallinarum* with several other serovars which includes *S. Enteritidis*, *S. Typhimurium* to decode the genomic differences amongst them. The comparative genome analysis of *S. Enteritidis* PT4, *S. Typhimurium* LT2 and *S. Gallinarum* 287/91 was carried by Thompson *et al.*, (2008) wherein they reported predominant similarity and synteny in their core genomes and overrepresentation of putative pseudogenes in *S. Gallinarum* 287/91 in comparison to *S. Enteritidis* str. P125109. Similarly, a genomic comparison between *S. Gallinarum* and *S. Pullorum* was carried out by Feng and associates and they included *S. Gallinarum/pullorum* str. CDC1983-67, *S. Gallinarum/pullorum* str. RKS5078, *S. Gallinarum* str. 287/91, and *S. Enteritidis* str. P125109 and reported a high number and differential distribution of pseudogenes in bvG and bvP strains in reference to *S. Enteritidis* strain (Feng et al. 2013). On the other hand, genome comparison between closely related *S. enterica* serovars Enteritidis, Dublin, and Gallinarum revealed differential distribution of prophages and pseudogenes in their genomes along besides SNPs, insertions, and deletions amongst the investigated strains (Matthews et al. 2015). In a similar manner, to decode different host-specificities, genes related to SPI of *S. Enteritidis* PT4 (NCTC 13349) and *S. Gallinarum* 287/91 NCTC 13346 were compared by (Eswarappa et al. 2009) and detected 24 positively selected genes that included SPI-2 TTSS and effector proteins of SPI-1 TTSS.

The present study suggests that the host restricted *S. Gallinarum* strains harbour strain specific genes and they exhibit differential distribution of putative virulence factors such as genomic islands, prophage regions, TA cassettes, and acquired AMR genes in their genomes. This study also highlights the need to analyse more number of *S. Gallinarum* genomes to understand the phylogeny and better capture of its pan-genome biodiversity. This comparative genomic analysis of *S. Gallinarum* serovar has provided a valuable insight and laid foundation for future experimental studies to be carried out to decipher the underlying mechanisms driving the pathogenesis and virulence of this avian restricted pathogen.

5. Conclusion

Salmonella enterica serovar Gallinarum biovar Pullorum (bvP) and biovar Gallinarum (bvG), are the causative agents of pullorum disease (PD) and fowl typhoid (FT) respectively, which causes considerable economic losses to poultry industry worldwide especially in developing countries including India. Comprehensive comparative genome analysis of nine *S. enterica* serovar Gallinarum strains originating from different geographical regions including Indian strain *S. Gallinarum* Sa140 VTCCBAA614 was carried out to decode the genotypic differences amongst them with a focus on detection and analysis of candidate virulence factors. The investigation revealed an open pan-genome for *S. Gallinarum* encompassing 5091 coding sequence (CDS) with 3270 CDS belonging to core-genome, 1254 CDS to dispensable genome and strain specific genes amongst the analyzed strains. Furthermore, analysis of distribution of candidate virulence factors in the investigated strains revealed diversity and differential distribution of genomic features such as genomic islands, prophage regions, toxin-antitoxin operons, and acquired antimicrobial resistance genes. The computational search unravelled the existence of high sequence identity genomic islands SP-1, SP-2, SP-3, SP-4, SPI-5, SPI-12, SPI-13, SP-14, C63PI and SPCS54 in their genomes. Additionally, 23 prophage regions and 149 Type II TA loci were also identified and characterized in the investigated genomes. Moreover, antimicrobial resistance genes were also detected in the studied genomes. The genomic variability detected among the *S. enterica* serovar Gallinarum strains will form the basis for future experimental investigations and could be used for bacterial typing, design of diagnostics, as well as drug and vaccine target prediction for effective infection control and prevention.

Declarations

Ethics approval and consent to participate

As our manuscript does not report on or involve any animals, humans, human data, human tissue or plants, therefore we declare this section as “Not applicable”.

Consent for publication

Our manuscript does not contain data from any individual person, therefore this section is “Not applicable”.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files. The genome data is publicly available in NCBI.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RKV was involved in conceptualization, methodology, Project administration, supervision, draft reviewing, re-writing and editing, resources, funding acquisition; ZT contributed in Data curation, methodology, formal analysis, writing- original draft preparation. TA was into investigation, reviewing

and editing; SK helped in funding acquisition; reviewing and editing; and BNT assisted in providing resources, supervision, reviewing and editing.

Acknowledgements and Funding

This work was supported by Department of Biotechnology, Govt. of India, New Delhi (BT/BI/25/07/2012-BIF) and Director, National Research Centre on Equines, National Centre for Veterinary Type Cultures, Hisar, Haryana, India. ZT acknowledges DBT “Bioinformatics infrastructure facility (BIF)” for providing Research Associateship. Funding source(s) had no involvement if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

References

- Alves Batista DF et al. 2018. Evaluation of pathogenicity of *Salmonella* Gallinarum strains harbouring deletions in genes whose orthologues are conserved pseudogenes in *S. Pullorum*. PLoS One. 13(1). doi: 10.1371/journal.pone.0200585.
- Armstrong J, Fiddes IT, Diekhans M, Paten B. 2019. Whole-Genome Alignment and Comparative Annotation. Annu. Rev. Anim. Biosci. 7:41–64. doi: 10.1146/annurev-animal-020518-115005.
- Arndt D et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44:W16–W21. doi: 10.1093/nar/gkw387.
- Bacciu D et al. 2004. Transposition of the heat-stable toxin astA gene into a Gifsy-2-related prophage of *Salmonella enterica* serovar abortusovis. J. Bacteriol. 186:4568–4574. doi: 10.1128/JB.186.14.4568-4574.2004.
- Barrow PA, Freitas Neto OC. 2011. Pullorum disease and fowl typhoid-new thoughts on old diseases: A review. Avian Pathol. 40:1–13. doi: 10.1080/03079457.2010.542575.
- Blom J et al. 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res. 44:W22–W28. doi: 10.1093/nar/gkw255.
- Blondel CJ, Jiménez JC, Contreras I, Santiviago CA. 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. BMC Genomics. 10. doi: 10.1186/1471-2164-10-354.
- Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. Microbiol. Mol. Biol. Rev. 68:560–602. doi: 10.1128/mmr.68.3.560-602.2004.
- De Carli S et al. 2017. Molecular and phylogenetic analyses of *Salmonella* Gallinarum trace the origin and diversification of recent outbreaks of fowl typhoid in poultry farms. Vet. Microbiol. 212:80–86. doi: 10.1016/j.vetmic.2017.11.001.
- Carlos Guimaraes L et al. 2015. Inside the Pan-genome - Methods and Software Overview. Curr. Genomics. 16:245–252. doi: 10.2174/1389202916666150423002311.
- Di Cesare A et al. 2016. Diverse distribution of Toxin-Antitoxin II systems in *Salmonella enterica* serovars. Sci. Rep. 6. doi: 10.1038/srep28759.
- Chan WT, Balsa D, Espinosa M. 2015. One cannot rule them all: Are bacterial toxins-antitoxins druggable? FEMS Microbiol. Rev. 39:522–540. doi: 10.1093/femsre/fuv002.
- Chan WT, Yeo CC, Sadowy E, Espinosa M. 2014. Functional validation of putative toxin-antitoxin genes from the Gram-positive pathogen *Streptococcus pneumoniae*: Phd-doc is the fourth bona-fide operon. Front. Microbiol. 5. doi: 10.3389/fmicb.2014.00677.
- Cheng RA, Eade CR, Wiedmann M. 2019. Embracing diversity: Differences in virulence mechanisms, disease severity, and host adaptations contribute to the success of nontyphoidal salmonella as a foodborne pathogen. Front. Microbiol. 10. doi: 10.3389/fmicb.2019.01368.
- Chimal-Cázares F et al. 2020. Molecular Characterization of SehB, a Type II Antitoxin of *Salmonella enterica* Serotype Typhimurium: Amino Acid Residues Involved in DNA-Binding, Homodimerization, Toxin Interaction, and Virulence. Front. Microbiol. 11. doi: 10.3389/fmicb.2020.00614.
- Colavecchio A, Cadieux B, Lo A, Goodridge LD. 2017. Bacteriophages contribute to the spread of antibiotic resistance genes among foodborne pathogens of the Enterobacteriaceae family - A review. Front. Microbiol. 8. doi: 10.3389/fmicb.2017.01108.
- Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.
- Eswarappa SM, Janice J, Balasundaram S V., Dixit NM, Chakravorty D. 2009. Host-specificity of *Salmonella enterica* serovar Gallinarum: Insights from comparative genomics. Infect. Genet. Evol. 9:468–473. doi: 10.1016/j.meegid.2009.01.004.

- Feng Y et al. 2012. Inheritance of the *Salmonella* virulence plasmids: Mostly vertical and rarely horizontal. *Infection, Genetics and Evolution*. 12:1058–1063. doi: 10.1016/j.meegid.2012.03.004.
- Feng Y, Johnston RN, Liu GR, Liu SL. 2013. Genomic Comparison between *Salmonella* Gallinarum and Pullorum: Differential Pseudogene Formation under Common Host Restriction. *PLoS One*. 8. doi: 10.1371/journal.pone.0059427.
- Figuroa-Bossi N, Uzzau S, Maloriol D, Bossi L. 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol. Microbiol.* 39:260–272. doi: 10.1046/j.1365-2958.2001.02234.x.
- Gerlach RG, Hensel M. 2007. *Salmonella* Pathogenicity Islands in host specificity, host pathogen-interactions and antibiotics resistance of *Salmonella enterica*. *Berl. Munch. Tierarztl. Wochenschr.* 120:317–327. doi: 10.2376/0005-9366-120-317.
- Ghatak S et al. 2016. Pan-genome analysis of *Aeromonas hydrophila*, *Aeromonas veronii* and *Aeromonas caviae* indicates phylogenomic diversity and greater pathogenic potential for *Aeromonas hydrophila*. *Antonie van Leeuwenhoek*, 109:945–956. doi: 10.1007/s10482-016-0693-6.
- Groisman EA, Ochman H. 1996. Pathogenicity islands: Bacterial evolution in quantum leaps. *Cell*. 87:791–794. doi: 10.1016/S0092-8674(00)81985-6.
- Guo R et al. 2019. Induction of arthritis in chickens by infection with novel virulent *Salmonella* Pullorum strains. *Vet. Microbiol.* 228:165–172. doi: 10.1016/j.vetmic.2018.11.032.
- Haraga A, Ohlson MB, Miller SI. 2008. Salmonellae interplay with host cells. *Nat. Rev. Microbiol.* 6:53–66. doi: 10.1038/nrmicro1788.
- Hendriksen RS et al. 2019. Using Genomics to Track Global Antimicrobial Resistance. *Front. Public Heal.* 7. doi: 10.3389/fpubh.2019.00242.
- Hensel M. 2004. Evolution of pathogenicity islands of *Salmonella enterica*. *Int. J. Med. Microbiol.* 294:95–102. doi: 10.1016/j.ijmm.2004.06.025.
- Hu Y et al. 2019. Loss and Gain in the Evolution of the *Salmonella enterica* Serovar Gallinarum Biovar Pullorum Genome. *mSphere*. 4. doi: 10.1128/msphere.00627-18.
- Huerta-Cepas J et al. 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47:D309–D314. doi: 10.1093/nar/gky1085.
- Huerta-Cepas J et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34:2115–2122. doi: 10.1093/molbev/msx148.
- Huguet KT, Gonnet M, Doublet B, Cloeckert A. 2016. A toxin antitoxin system promotes the maintenance of the IncA/C-mobilizable *Salmonella* Genomic Island 1. *Sci. Rep.* 6. doi: 10.1038/srep32285.
- Ilyas B, Tsai CN, Coombes BK. 2017. Evolution of *Salmonella*-host cell interactions through a dynamic bacterial genome. *Front. Cell. Infect. Microbiol.* 7. doi: 10.3389/fcimb.2017.00428.
- Jain P et al. 2020. Characterization of non-typhoidal *Salmonella* isolates from children with acute gastroenteritis, Kolkata, India, during 2000–2016. *Brazilian J. Microbiol.* 51:613–627. doi: 10.1007/s42770-019-00213-z.
- Jajere SM. 2019. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and adaptation and antimicrobial resistance including multidrug resistance. *Vet. World.* 12:504–521. doi: 10.14202/vetworld.2019.504-521.
- Jantsch J, Chikkaballi D, Hensel M. 2011. Cellular aspects of immunity to intracellular *Salmonella enterica*. *Immunol. Rev.* 240:185–195. doi: 10.1111/j.1600-065X.2010.00981.x.
- Jiang L et al. 2017. Signal transduction pathway mediated by the novel regulator LoiA for low oxygen tension induced *Salmonella* Typhimurium invasion. *PLoS Pathog.* 13. doi: 10.1371/journal.ppat.1006429.
- Jones MA, Wigley P, Page KL, Hulme SD, Barrow PA. 2001. *Salmonella enterica* serovar Gallinarum requires the *Salmonella* pathogenicity island 2 type III secretion system but not the *Salmonella* pathogenicity island 1 type III secretion system for virulence in chickens. *Infect. Immun.* 69:5471–5476. doi: 10.1128/IAI.69.9.5471-5476.2001.
- Koerich PKV et al. 2018. *Salmonella* Gallinarum field isolates and its relationship to vaccine strain SG9R. *Br. Poult. Sci.* 59:154–159. doi: 10.1080/00071668.2017.1406062.
- Kwon HJ, Cho SH. 2011. Pathogenicity of SG 9R, a rough vaccine strain against fowl typhoid. *Vaccine*. 29:1311–1318. doi: 10.1016/j.vaccine.2010.11.067.

- Laing CR, Whiteside MD, Gannon VPJ. 2017. Pan-genome analyses of the species *Salmonella enterica*, and identification of genomic markers predictive for species, subspecies, and serovar. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.01345.
- Langridge GC et al. 2015. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc. Natl. Acad. Sci. U. S. A.* 112:863–868. doi: 10.1073/pnas.1416707112.
- Li Q et al. 2015. Complete genome sequence of *Salmonella enterica* serovar pullorum multidrug resistance strain S06004 from China. *J. Microbiol. Biotechnol.* 25:606–611. doi: 10.4014/jmb.1406.06031.
- Lobato-Márquez D, Moreno-Córdoba I, Figueroa V, Díaz-Orejas R, García-Del Portillo F. 2015. Distinct type I and type II toxin-antitoxin modules control *Salmonella* lifestyle inside eukaryotic cells. *Sci. Rep.* 5. doi: 10.1038/srep09374.
- Lu QF et al. 2019. Genus-wide comparative genomics analysis of *Neisseria* to identify new genes associated with pathogenicity and niche adaptation of *Neisseria* pathogens. *Int. J. Genomics.* 2019. doi: 10.1155/2019/6015730.
- Makarova KS, Wolf YI, Koonin E V. 2009. Comprehensive comparative-genomic analysis of Type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct.* 4. doi: 10.1186/1745-6150-4-19.
- Mastroeni P, Chabalgoity JA, Dunstan SJ, Maskell DJ, Dougan G. 2001. *Salmonella*: Immune responses and vaccines. *Vet. J.* 161:132–164. doi: 10.1053/tvjl.2000.0502.
- Matthews TD et al. 2015. Genomic comparison of the closely-related *Salmonella enterica* serovars Enteritidis, Dublin and Gallinarum. *PLoS One.* 10. doi: 10.1371/journal.pone.0126883.
- Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15:589–594. doi: 10.1016/j.gde.2005.09.006.
- Menashe O, Kaganskaya E, Baasov T, Yaron S. 2008. Aminoglycosides affect intracellular *Salmonella enterica* serovars typhimurium and virchow. *Antimicrob. Agents Chemother.* 52:920–926. doi: 10.1128/AAC.00382-07.
- Moreno Switt AI et al. 2012. Identification and characterization of novel *Salmonella* mobile elements involved in the dissemination of genes linked to virulence and transmission. *PLoS One.* 7. doi: 10.1371/journal.pone.0041247.
- Mottawea W et al. 2018. *Salmonella enterica* prophage sequence profiles reflect genome diversity and can be used for high discrimination subtyping. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.00836.
- Moutaouakkil Y, Frikh M, Maleb A, Lemnouer A, Elouennass M. 2017. Infection caused by *Salmonella* Gallinarum emergent bacteria. *Saudi J. Heal. Sci.* 6:68–70. doi: 10.4103/sjhs.sjhs_81_16.
- Pan Z et al. 2009. Changes in antimicrobial resistance among *Salmonella enterica* subspecies enterica serovar Pullorum isolates in China from 1962 to 2007. *Vet. Microbiol.* 136:387–392. doi: 10.1016/j.vetmic.2008.11.015.
- Parveen S et al. 2007. Prevalence and antimicrobial resistance of *Salmonella* recovered from processed poultry. *J. Food Prot.* 70:2466–2472. doi: 10.4315/0362-028X-70.11.2466.
- Price MN et al. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature.* 557:503–509. doi: 10.1038/s41586-018-0124-0.
- Rajagopal R, Mini M. 2013. Outbreaks of salmonellosis in three different poultry farms of Kerala, India. *Asian Pacific Journal of Tropical Biomedicine.* 3:496–500. doi: 10.1016/S2221-1691(13)60103-3
- Ramage HR, Connolly LE, Cox JS. 2009. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* 5. doi: 10.1371/journal.pgen.1000767.
- Rasko DA et al. 2008. The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190:6881–6893. doi: 10.1128/JB.00619-08.
- Rempel S et al. 2020. A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature.* 580:409–412. doi: 10.1038/s41586-020-2072-8.
- Rinaudo CD, Telford JL, Rappuoli R, Seib KL. 2009. Vaccinology in the genome era. *J. Clin. Invest.* 119:2515–2525. doi: 10.1172/JCI38330.

- Roer L et al. 2016. Is the Evolution of *Salmonella enterica* subsp. *enterica* Linked to Restriction-Modification Systems? *mSystems*. 1. doi: 10.1128/msystems.00009-16.
- Rouli L, Merhej V, Fournier PE, Raoult D. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7:72–85. doi: 10.1016/j.nmni.2015.06.005.
- Równicki M et al. 2018. Artificial activation of *Escherichia coli* mazEF and hipBA toxin–antitoxin systems by antisense peptide nucleic acids as an antibacterial strategy. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.02870.
- Sahl JW et al. 2013. Evolution of a Pathogen: A Comparative Genomics Analysis Identifies a Genetic Pathway to Pathogenesis in *Acinetobacter*. *PLoS One*. 8. doi: 10.1371/journal.pone.0054287.
- Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 30:2068–2069. doi: 10.1093/bioinformatics/btu153.
- Shivaprasad HL. 2000. Fowl typhoid and pullorum disease. *OIE Rev. Sci. Tech.* 19:405–424. doi: 10.20506/rst.19.2.1222.
- Silva EN, Snoeyenbos GH, Weinack OM, Smyser CF. 1981. Studies on the use of 9R strain of *Salmonella Gallinarum* as a vaccine in chickens. *Avian Dis.* 25:38–52. doi: 10.2307/1589825.
- Taddele MH, Rathore R, Dhama K. 2012. Antibigram assay of *Salmonella Gallinarum* and other *Salmonella enterica* serovars of poultry origin in India. *Asian J. Anim. Vet. Adv.* 7:309–317. doi: 10.3923/ajava.2012.309.317.
- Tanaka K et al. 2004. Molecular Characterization of a Prophage of *Salmonella enterica* Serotype Typhimurium DT104. *J. Clin. Microbiol.* 42:1807–1812. doi: 10.1128/JCM.42.4.1807-1812.2004.
- Tettelin H et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial ‘pan-genome’. *Proc. Natl. Acad. Sci. U. S. A.* doi: 10.1073/pnas.0506758102.
- Thakur Z, Saini V, Arya P, Kumar A, Mehta PK. 2018. Computational insights into promoter architecture of toxin-antitoxin systems of *Mycobacterium tuberculosis*. *Gene*. 641:161-171 doi: 10.1016/j.gene.2017.10.054.
- Thomson N et al. 2004. The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. *J. Mol. Biol.* 339:279–300. doi: 10.1016/j.jmb.2004.03.058.
- Thomson NR et al. 2008. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella Gallinarum* 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res.* 18:1624–1637. doi: 10.1101/gr.077404.108.
- Tomljenovic-Berube AM et al. 2013. Mapping and regulation of genes within *Salmonella* pathogenicity island 12 that contribute to in vivo fitness of *Salmonella enterica* serovar typhimurium. *Infect. Immun.* 81:2394–2404. doi: 10.1128/IAI.00067-13.
- Uzzau S et al. 2001. *Salmonella enterica* serovar-host specificity does not correlate with the magnitude of intestinal invasion in sheep. *Infect. Immun.* 69:3092–3099. doi: 10.1128/IAI.69.5.3092-3099.2001.
- Vaid RK et al. 2015. First draft genome sequence of *Salmonella enterica* serovar Gallinarum strain VTCCBAA614, isolated from chicken in India. *Genome Announc.* 3. doi: 10.1128/genomeA.01221-15.
- VanDrisse CM, Parks AR, Escalante-Semerena JC. 2017. A toxin involved in *Salmonella* persistence regulates its activity by acetylating its cognate antitoxin, a modification reversed by CobB sirtuin deacetylase. *MBio*. 8. doi: 10.1128/mBio.00708-17.
- Weigel LM, Steward CD, Tenover FC. 1998. gyrA mutations associated with fluoroquinolone resistance in eight species of Enterobacteriaceae. *Antimicrob. Agents Chemother.* 42:2661–2667. doi: 10.1128/aac.42.10.2661.
- Wigley P, Berchieri A. J, Page KL, Smith AL, Barrow PA. 2001. *Salmonella enterica* serovar pullorum persists in splenic macrophages and in the reproductive tract during persistent, disease-free carriage in chickens. *Infect. Immun.* 69:7873–7879. doi: 10.1128/IAI.69.12.7873-7879.2001.
- Wigley P, Jones MA, Barrow PA. 2002. *Salmonella enterica* serovar Pullorum requires the *Salmonella* pathogenicity island 2 type III secretion system for virulence and carriage in the chicken. *Avian Pathol.* 31:501–506. doi: 10.1080/0307945021000005879.
- Wilson RL, Elthon J, Clegg S, Jones BD. 2000. *Salmonella enterica* serovars gallinarum and pullorum expressing *Salmonella enterica* serovar typhimurium type 1 fimbriae exhibit increased invasiveness for mammalian cells. *Infect. Immun.* 68:4782–4785. doi: 10.1128/IAI.68.8.4782-4785.2000.

- Xu Z et al. 2016. A critical role of bacterioferritin in *Salmonella* pullorum-induced IFN- β expression in DF-1 cells. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.00020.
- Yamaguchi Y, Park J-H, Inouye M. 2011. Toxin-Antitoxin Systems in Bacteria and Archaea. *Annu. Rev. Genet.* doi: 10.1146/annurev-genet-110410-132412.
- Yoon SH et al. 2005. A computational approach for identifying pathogenicity islands in prokaryotic genomes. *BMC Bioinformatics.* 6. doi: 10.1186/1471-2105-6-184.
- Yoon SH et al. 2007. Towards pathogenomics: A web-based resource for pathogenicity islands. *Nucleic Acids Res.* 35. doi: 10.1093/nar/gkl790.
- Yoon SH, Ha S min, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*, 110:1281–1286. doi: 10.1007/s10482-017-0844-4.
- Zankari E et al. 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67:2640–2644. doi: 10.1093/jac/dks261.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* 39. doi: 10.1093/nar/gkr485.

Additional Files/materials

| File name (e.g. Additional file 1) | File format including the correct file extension | Title of data | Description of data |
|------------------------------------|---|---|--|
| Supplementary Figures | | | |
| Supplementary Figure 1(a-h) | .pdf | Synteny plot analysis of investigated <i>Salmonella</i> strains with reference to <i>S. Enteritidis</i> str. P125109. | Depicts synteny plot analysis of investigated <i>Salmonella</i> strains |
| Supplementary Figure 2 | .tiff | Prophages detected by Phaster in analyzed <i>Salmonella</i> strains | Details of prophages detected |
| Supplementary Figure 3 | .tiff | Acquired antimicrobial resistance genes, known chromosomal mutations, and unknown chromosomal mutations detected in investigated <i>Salmonella</i> strains by Resfinder. | Detects acquired antimicrobial resistance genes |
| Supplementary Tables | | | |
| Supplementary Table 1: | .docx | Pan-genome and core genome development projections for investigated nine <i>Salmonella</i> strains | Pan-genome and core genome development projections |
| Supplementary Table 2: | .xlsx | List of CDS that are part of pan genome of investigated nine <i>Salmonella</i> strains with their name and function. | Pan genome CDS list of genomes |
| Supplementary Table 3: | .xlsx | List of CDS that are identified to be part of the core genome of analyzed nine <i>Salmonella</i> strains. | Core genome CDS list of genomes |
| Supplementary Table 4 (a to i): | .xlsx | Description of singletons detected in investigated <i>Salmonella</i> Gallinarum (a) <i>S. Enteritidis</i> str. P125109 (b) <i>S. Gallinarum/pullorum</i> str. CDC1983-67 (c) <i>S. Gallinarum/pullorum</i> str. RKS5078 (d) <i>S. Gallinarum</i> str. 287/91 (e) <i>S. Gallinarum</i> str. 9184 (f) <i>S. Pullorum</i> str. ATCC 9120 (g) <i>S. Pullorum</i> str. S06004 (h) <i>S. Pullorum</i> QJ-2D-Sal (i) <i>S. Gallinarum</i> Sal40 strains by EDGAR with their name and function with reference to <i>S. Enteritidis</i> str. P125109 | Detected Singletons listing of all genomes |
| Supplementary Table 5: (a-b) | .xlsx | The functional annotation of (a) core genome of analyzed <i>Salmonella</i> strains and (b) set of strain specific CDS performed by eggNOG-mapper v2 and enlists query, seed ortholog, e-value, score, best taxonomy level, preferred name, GO terms, EC number, annotation level, COG category and description. | functional annotation of CDS of analysed genomes |
| Supplementary Table 6(a)-(b) | .xlsx | The datasheet enlists homologs of <i>Salmonella</i> pathogenic islands (SPIs) detected by (a) SPIFinder 1.0 and (b) BLAST search, in the investigated <i>Salmonella</i> genomes | Listing of <i>Salmonella</i> pathogenicity island detected by SPI Finder and by BLAST search |
| Supplementary Table 7: (a-b) | .xlsx | The identified (a) complete list of prophage regions and (b) intact prophages, in the genomes of nine <i>Salmonella</i> strains that were investigated in the study with description of a) region number b) region length c) completeness d) score e) total proteins f) region position g) most common phage and h) GC%. The colour coding of the rows represents the classification of identified prophage regions by PHASTER, green- intact prophage region, red- incomplete prophage region and blue- questionable prophage region. | Description of prophages detected in all genomes with details |

| | | | |
|-------------------------------------|-------|---|--|
| Supplementary Table 8: | .xlsx | This worksheet contains 149 TA gene cassettes identified in nine <i>Salmonella</i> genomes by TA finder. | TA gene listing detected in all genomes |
| Supplementary Table 9: (a-c) | .xlsx | This datasheet details (a) acquired antimicrobial resistance genes, (b) known chromosomal mutations, and (c) unknown chromosomal mutations detected in investigated <i>Salmonella</i> strains by Resfinder. | Details of AMR genes detected in genomes |

Figures

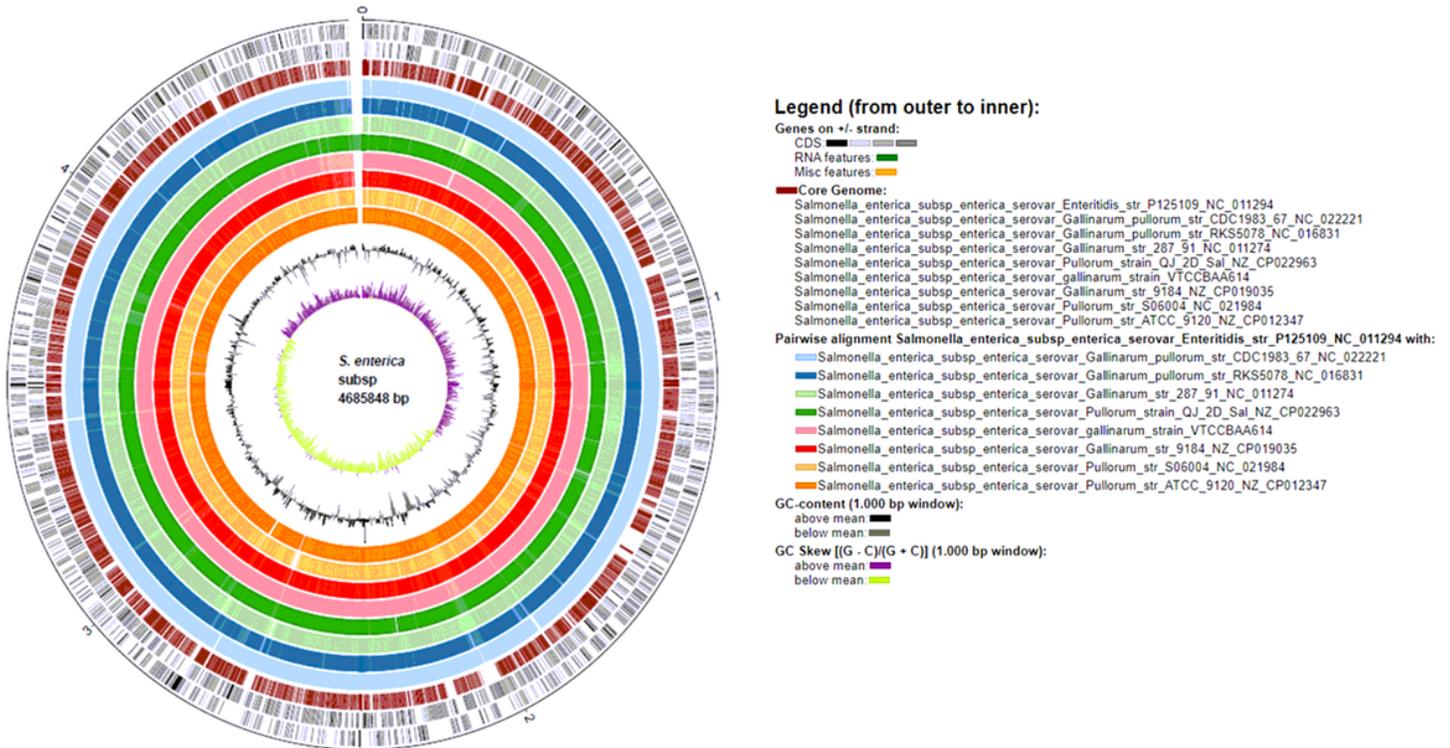


Figure 1
 Circular plot genome representation of nine investigated *Salmonella* strains i.e. a) *S. Gallinarum* str. 287/91 b) *S. Gallinarum* str. 9184 b) *S. Gallinarum* Sal40 strain d) *S. Pullorum* str. ATCC 9120 e) *S. Pullorum* str. S06004 f) *S. Pullorum* QJ-2D-Sal g) *S. Gallinarum*/Pullorum str. CDC1983-67 and h) *S. Gallinarum*/pullorum str. RKS5078 with demonstration of GC content, GC skew and CDS in reference to *S. Enteritidis* str. P125109.

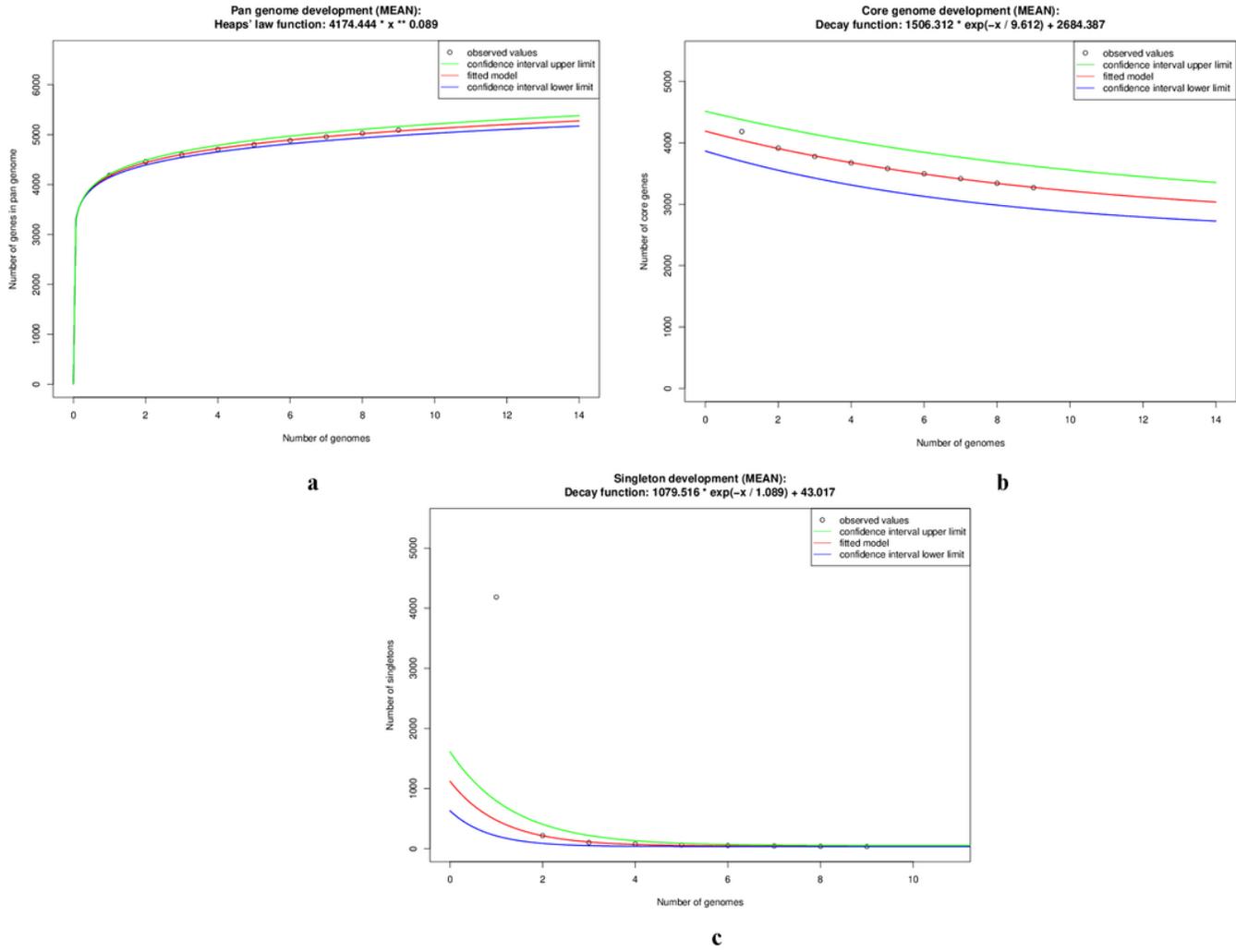
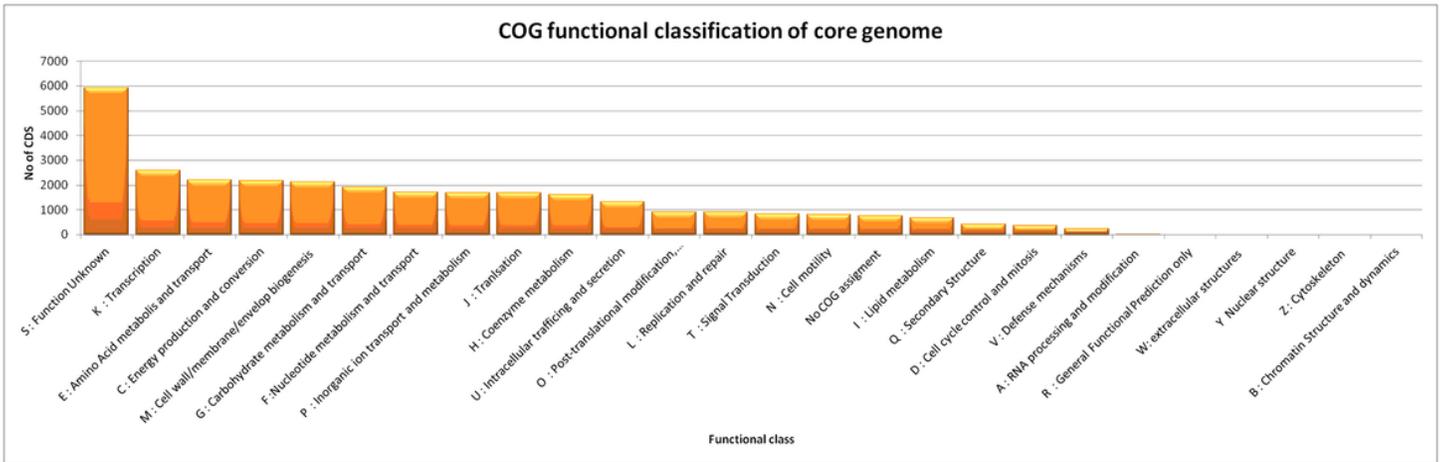
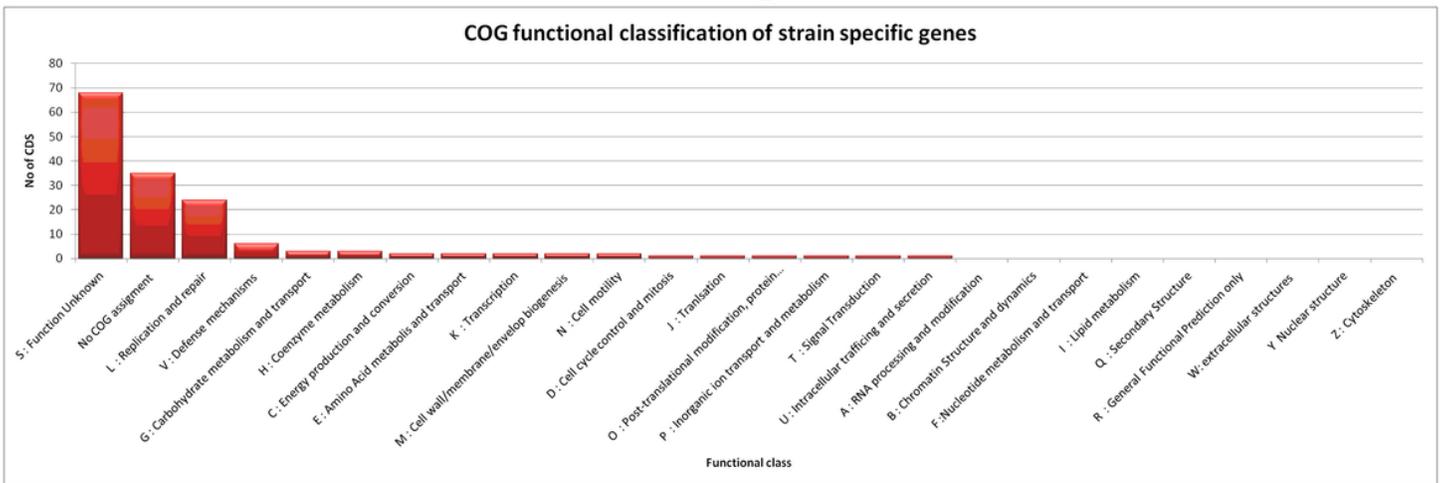


Figure 2

Core genome development plot for nine *Salmonella* genomes. The red curve shows the fitted model. Green represents confidence interval upper limit and blue demonstrates confidence interval lower limit



a



b

Figure 3

a) Distribution of COG functional classes in core genome of investigated Salmonella strains by eggnog mapper b) Distribution of COG functional categories in strain specific genes by eggnog mapper

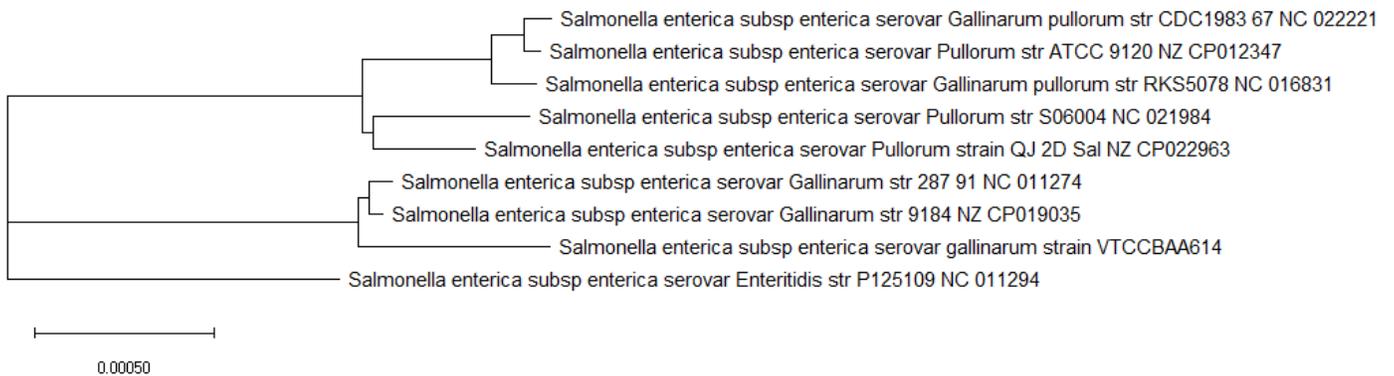


Figure 4

Unrooted phylogenetic tree showing taxonomic positions of the investigated Salmonella strains constructed by EDGAR on the basis of core genome set

Synteny plot

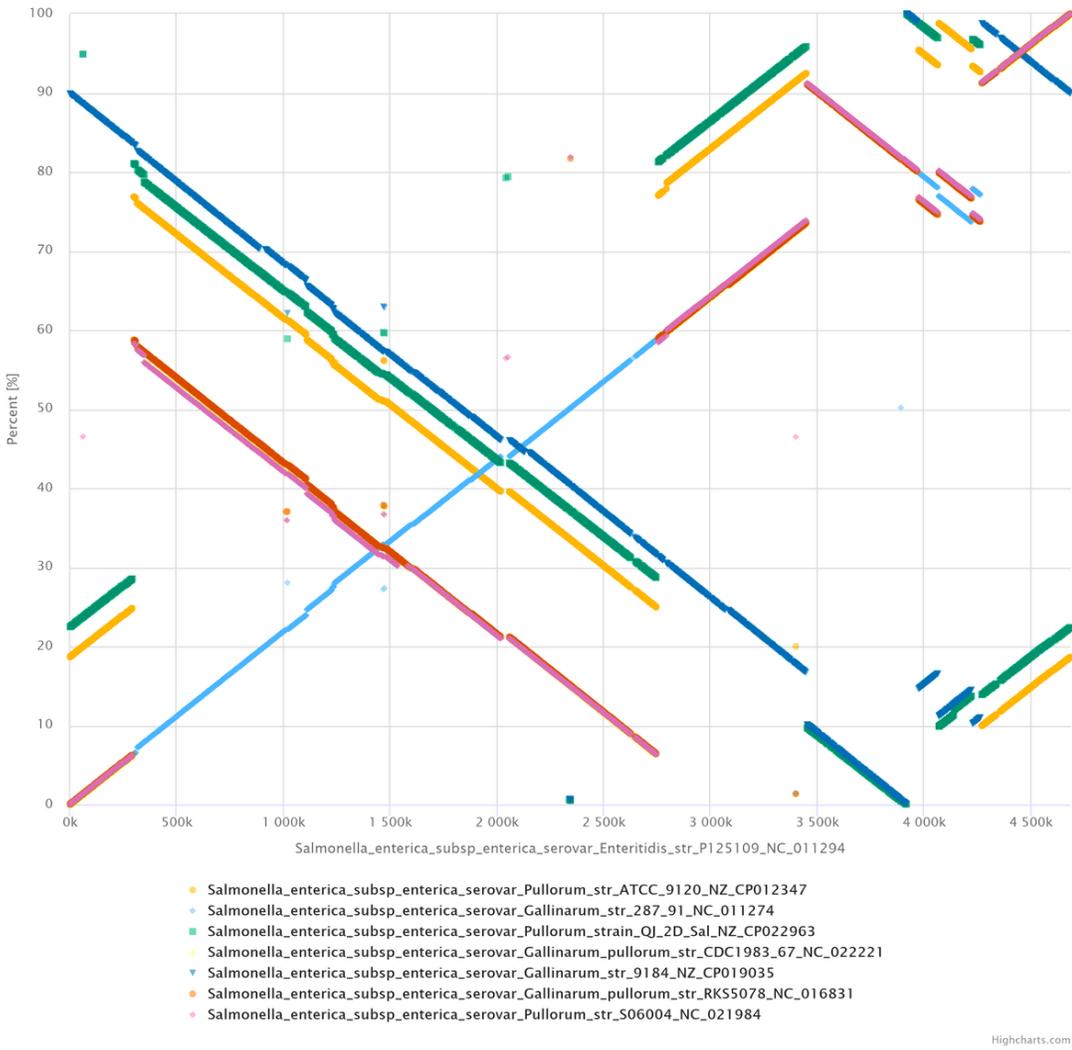


Figure 5

Synteny plot constructed by EDGAR of eight *S. Gallinarum* strains i.e. a) *S. Pullorum* str. ATCC 9120 (b) *S. Gallinarum* str. 287/91 c) *S. Pullorum* QJ-2D-Sal (d) *S. Gallinarum/Pullorum* str. CDC1983-67 (e) *S. Gallinarum* str. 9184, (f) *S. Gallinarum/Pullorum* str. RKS5078 (g) *S. Pullorum* str. S06004 against the reference of *S. Enteritidis* str. P125109

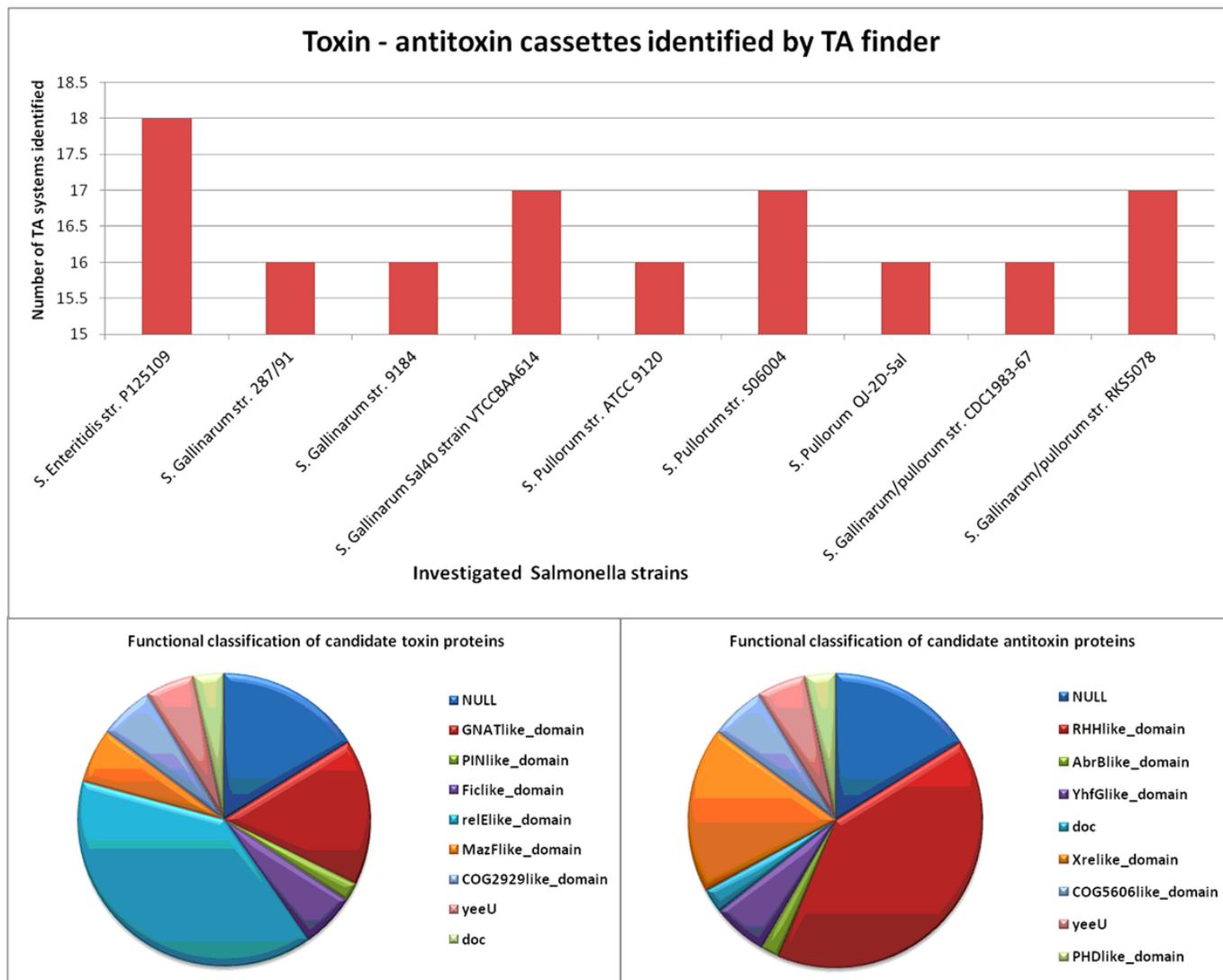


Figure 6

Toxin-antitoxin cassettes identified by TA finder in investigated Salmonella strains with functional classification of candidate toxin and antitoxin proteins

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFiguresandtables.zip](#)