Hypothesis: AT Homopolymer Strings in *Salmonella enterica* subspecies I Contribute to Speciation and Serovar Diversity

Jean Guard [1],*, Adam R Rivers [2], Justin N Vaughn [1], Michael J Rothrock Jr [1], Adelumola Oladeinde [1] and Devendra H Shah [3]

[1] Affiliation 1; US Department of Agriculture, US National Poultry Research Center, Athens, GA 30605
[2] Affiliation 2; US Department of Agriculture, Genomics and Bioinformatics Research, Gainesville, GA 32608
[3] Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164
* Correspondence: jean.guard@usda.gov; Tel.: 1+706-546-3446

Abstract: Adenine and thymine homopolymer strings of at least 8 nucleotides (AT 8+mers) were characterized in *Salmonella enterica* subspecies I. The motif differed between cother taxonomic classes but not between *Salmonella enterica* serovars. The motif in plasmids was associated with serovar. Approximately 12.3% of the *S. enterica* motif loci had mutations. Mutability of AT 8+mers suggests that genomes undergo frequent repair to maintain optimal gene content, and that the motif facilitates self-recognition; in addition, serovar diversity is associated with plasmid content. A theory that genome regeneration accounts for both persistence of predominant *Salmonella* serovars and serovar diversity provides a new framework for investigating root causes of foodborne illness.

Keywords: keyword 1; *Salmonella enterica* 2; food safety 3; genome 4; theory 5; single nucleotide polymorphisms 6; recombination 7; serovar

1. Introduction

Approximately 30 of 1500 *Salmonella enterica* subspecies I (*S. enterica*) serovars are persistent agents of foodborne illness in people [1]. Despite improved biosecurity throughout the food production pipeline, reduction of salmonellosis has plateaued over 20 years [2]. The inability to reduce salmonellosis indicates new approaches to understanding the biology of this important pathogen are needed. Recently, the most commonly occurring single nucleotide polymorphism (SNP) that caused disruption of a gene in *S. enterica* serovar Enteritidis (Enteritidis) was identified, and it was deletion of a single adenine in a homopolymer string of 8 nucleotides (nt) within the fimbrial gene *sef*D [3]. Mutational analysis, phenotype microarray, and infection experiments in the egg-laying hen indicated that the *sef*D mutation increased organ invasion and mortality in hens, disturbed egg production, enhanced growth of the pathogen to high cell density, and otherwise behaved as a regulator of dimorphism of phenotype [4]. As a result of this finding the performance of a killed vaccine for hens was enhanced by increasing SefD in preparations [5]. The drastic change in biological phenotype imparted by the single base pair deletion suggested that characterization of purine homopolymer strings of adenine, AAAAAAAA, and its pyrimidine base pair (bp) of thymine, TTTTTTTT, in *S. enterica* should be explored. It is not evident that *S. enterica* of different serovars would have conserved AT 8+mer content, because it has a mosaic genome with frequent inversions of sections of the core genome as well as differences originating from mobile genetic elements such as bacteriophage, transposons, insertion elements, and plasmids.

Previous analysis of homopolymeric A and T strings across 81 bacterial and 18 archeal genomes showed that the motif in bacteria was hypermutable, occurred preferentially at the 5' end of genes, and was not biased for AA- or TT- encoded amino acids [6]. Furthermore, detailed analysis of these tracts in the *inlA* gene of *Listeria monocytogenes* supported the premise that AT homopolymeric tracts were part of a robust regulatory mechanism that facilitated a reversible, rapid evolutionary mechanism facilitating adaptation

and phase variation in gene regulation. Finding that *S. enterica* serovar Enteritidis, the world's leading cause of foodborne salmonellosis, had the hypermutable gene *sefD* linked to phenotypic phase variation suggested that review of the *S. enterica* genome for AT homopolymeric tracts was warranted.

In addition to research already cited, there is other evidence that AT homopolymer tracts, referred to in this manuscript AT 8+mers, constitute an important regulatory element in bacteria. It is a DNA motif suggested by conformational studies to bend DNA out of the Z-conformation [7]. Polyadenine regions can impact gene regulation in prokaryotes and can contribute to microsatellite instability in eukaryotes [8-11]. It has been shown that homopolymer nucleotide strings contribute to non-programmed slipped strand replication and the accumulation of errors in DNA [12-14]. Thus, the physicochemical impact of these strings was another reason to catalogue this motif in the genome of *S. enterica*.

The incidence of homopolymer strings in completed genomes of subspecies I of *S. enterica* and two different taxonomic classes, namely Gammaproteobacteria and Bacilli, were compared to better understand if *S. enterica* was at all unique. Reference genomes of *S. enterica* serovars Enteritidis and Typhimurium were analyzed to locate mutated AT 8+mers in genomes because they are from different genomic lineages and have been extensively sequenced and annotated. Together they cause approximately 40% of all foodborne salmonellosis in the US [1]. *S. enterica* serovar Gallinarum was included in the same analysis because it is a biological outlier within subspecies I that does not cause foodborne disease. Unlike either Enteritidis or Typhimurium serovars, Gallinarum causes disease in poultry resulting in high morbidity, mortality, and economic loss. A comparative approach is useful for linking phenotype to single nucleotide polymorphisms occurring in genomes **[15]**. In this study the three genomes were compared to better understand the association the motif might have with naturally occurring mutation that disrupts open reading frames of genes.

1.1. Additional Background on Poultry-associated S. enterica Serovars

S. enterica serovars Enteritidis and Typhimurium differ biologically **[16]**. Epidemiological patterns for the two predominant pathogens also differ. Enteritidis is an exceptional *Salmonella* pathogen in part because it efficiently contaminates the internal contents of eggs produced by otherwise healthy-appearing hens. It produces a high molecular mass (HMM) O-antigen, which not only protects killing of the pathogen by the host but also acts as a protective capsule in the hostile environment of the egg **[17-19]**. Typhimurium is also resistant to complement but it does not produce HMM O-antigen, and thus it does not survive in the internal contents of eggs to an extent that can be detected by epidemiological surveillance. Both Typhimurium and Enteritidis can contaminate a broad spectrum of other food sources such as poultry carcasses, other meats, and fresh vegetables. Both serovars can invade organs and survive, which contributes to systemic spread during infection **[20]**. Variation between strains within each serovar occurs but serovar characteristics and general genome organization are maintained **[21, 22]**. There are serovar-specific patterns in plasmid carriage and fimbrial genes. Comprehensive reviews are available on the similarities and differences between *Salmonella* serovars **[23-27]**.

S. enterica serovars Gallinarum and Enteritidis are genetically closely related **[28]**. Serovar Gallinarum's antigenic formula is 1,9,12:-:-, which indicates it has the same lipopolysaccharide O-antigen epitopes as Enteritidis; however, it lacks both H1 and H2 flagellin proteins and is thus non-motile. Both Gallinarum and Enteritidis can contaminate the internal contents of eggs; however, Gallinarum has mutations and rearrangements that restrict its host range to the avian host, possibly by reducing immunological response to infection and thereby facilitating systemic infection **[20]**. Thus, the most striking differences between the foodborne pathogen Enteritidis and host-restricted Gallinarum is that the latter makes poultry sick, reduces egg production and causes mortality. In contrast hens infected with Enteritidis often appear healthy, remain in production, eggs become contaminated internally, and are a source of foodborne illness. The ability of Enteritidis to

spread through flocks that appear healthy was one of the contributing factors in its world-wide spread through the layer industry. The differences in the epidemiology, association with food, and virulence characteristics of the three pathogens, all of which occur in the poultry environment, supported comparing them to better understand the association between the AT 8+mer motif and naturally occurring mutation of an important food borne pathogen. Other pathogenic Salmonellae and bacteria with different taxonomy were also included in analysis to facilitate comparison to results obtained by others [6].

2. Materials and Methods

2.1. Salmonella enterica subspecies I Strains Analyzed for Strings of Homopolymers

A database of 49 completed genomes of *S. enterica* subspecies I (taxid:59201 was accessed at the National Center for Biotechnology Information (NCBI) **[29]** (Table S1), and the last accession date was August 20, 2021. S. *enterica* serovars Typhimurium, Typhi, and Enteritidis genomes were over-represented compared to other serovars, and together they comprised 39.4% of all completed genomes available at the time of analysis. Only 51.2% of *S. enterica* subspecies I genomes had a complete adenylate cyclase (*cyaA*) gene, which is required for virulence as a foodborne pathogen, thus analysis was restricted to completed genomes to avoid issues associated with incomplete assembly and annotation. The other sequences were plasmids, which were also evaluated for the AT 8+mer motif. Genome CP018657 was excluded from analyses due to errors. Genome databases at NCBI show homopolymer strings, as well as other combinations of low-complexity regions, in lower-case gray font because there is recognition that some sequence strings might be susceptible to alignment error; thus, they required masking during the alignment process. For the BLAST searches conducted here, each gene was observed for high fidelity of surrounding regions, therefore it is unlikely low complexity impacted observed alignments. For *S. enterica* subspecies I, 12 complete genomes were assessed for Typhimurium, Enteritidis, and Typhi, and another grouping included a mixture of 12 serovars often recovered from foods. It is important to note that the NCBI is in the process of consolidating redundancy in over-represented genomes, thus the total number of genomes represented by high quality and well-annotated accession numbers may be larger than it appears as redundant material is condensed and re-annotated (https://www.ncbi.nlm.nih.gov/refseq/about/prokaryotes/reannotation/).

2.2 Other Bacteria Analyzed for Comparison to S. enterica

To compare homopolymer strings in *S. enterica* subspecies I, AT and GC 8+mer homopolymers were tabulated from 42 strains encompassing 5 other genera in Phyla Proteobacteria (*Escherichia coli, Proteus mirabilis, Shigella sonnei, Yersinia pseudotuberculosis, Vibrio vulnificus* (chromosome I and II)), and 5 genera of Firmicutes *(Staphylococcus aureus, Streptococcus pyogenes, Enterococcus faecalis, Bacillus anthracis and Bacillus cereus)* (Table S1). *Salmonella, Shigella,* and *Escherichia* are all members of the Family Enterobacteriaceae, whereas *Staphylococcus, Streptococcus, Enterococcus*, and *Bacillus* are in the Family Bacilli. At least 3 complete genomes of the other genera were assessed, including 12 genomes from *Escherichia coli* (Table 1).

2.3. Statistical Analysis of kmer Content

Strings of homopolymers of different lengths were entered as text in the find function and the genomes were searched. Searches would count single locations multiple times if the length exceeded 8nt, e. g. kmers of 10nt would often be counted 3 times as 8, 9, and 10nts. Thus, each kmer was reviewed for length to make sure a single location was counted only once, and then it was catalogued according to total length. Review for undercounting homopolymer strings due to wrapping at end of lines in sequence reports was also done; thus, the Geneious or NCBI search programs were preferred. The genomes of interest were stored in SeqBuilder Pro, Lasergene V16.0.0 352 (DNASTAR, Madison, Wisconsin, USA) and in Geneious V2030.0.3 (Biomatters, Inc., San Diego, California, USA)

format ((https://www.geneious.com). Results were copied into an Excel ".csv" file as Unicode text (Microsoft Excel for Mac, V16.16.20 (200307). The text to column feature, and appropriate delimiters, were used to produce columns of data to calculate distance between nucleotide strings. The average, standard deviation, and median values between AT 8+mer homopolymers were then calculated. Homopolymer strings of all 4 nucleotides, ranging from 5–20 nucleotides, were counted in *S. enterica* serovars and other genera and results are shown in columns F and G of Table 1. To account for every AT kmer of at least 8 nucleotides, the longer motifs were added to 8mers in further analyses; thus, the term AT 8+mer is applied throughout to describe the motif. Ttest analysis was used to determine if differences in kmer counts between groupings were significant at $p < 0.01$.

S. enterica subspecies I serovar Typhimurium LT2 was the reference genome to produce a common denominator to normalize genomes of different sizes (Table 1). Thus, values greater than 1 indicated that more than the expected number of motifs were observed in comparison to *S. enterica* after normalizing for the size of the genome, and less than 1 indicated fewer motifs were observed.

AT 8+kmers for Typhimurium LT2 were classified as intergenic, intragenic, or regulatory using Genious Prime 2020.0.3 (Table S2). Another approach used to establish a baseline incidence of AT 8+mers occurring in genes was to generate a list of random numbers using the 4600 predicted genes of the reference genome. Two hundred random numbers were generated between 1 to 4600 corresponding to numbered genes, a FASTA file was then compiled, and the number of AT 8+mers within the randomly generated sets was determined.

2.4. Locating AT 8+mers within mutated genes

S. enterica serovar Typhimurium LT2 (NC_003197.2) was used as the primary reference sequence to name genes and gene functions **[30]**. The two other reference genomes were Enteritidis strain P125109 and Gallinarum strain 9184, with respective NCBI accession numbers of NC_011294.1 and CP019035.1 **[31, 32]**. Classes and functions of proteins were assessed at https://uniprot.org. Determining impact on open reading frames (ORFs) within annotated genes was done with Genious V2030.0.3. In addition, online software available at https://web.expasy.org was used to translate proteins and align amino acids with nucleotides. BLAST analyses and generation of reverse complemented sequences used the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

3. Results

3.1. Homopolymer Strings of at Least 8 Nucleotides were Dispersed in the Genome of S. enterica

The AT 8+mers were dispersed throughout the entire genome of serovar Typhimurium LT2 (Figure 1). The genome of reference strain *S. enterica* serovar Typhimurium LT2 is 52.2% GC. When data were expressed as ratios of AT:GC homopolymer strings, the AT 8mer homopolymers (e. g. AAAAAAAA and TTTTTTTT) were much more prevalent than GC 8mers in the reference genome (Figure 2). In total there were 294 AT 8mers and 11 GC 8mers in the reference serovar, which is a ratio of 27 AT 8mers to every GC 8mer. AT strings longer than 8bp were less frequently observed (Figure 2). On average the motif occurred every 16,450nt (Table S2). The range of AT 8+kmer distance was 11 to 117,141nt, and the median was 11,498nt (Table S2). Distances of 51,975nt or greater between motifs were over 3 standard deviations and were thus possibly deficient in AT 8+mers. Of 13 putatively deficient regions, the 4 longest regions were assessed for phage genes, pseudo genes, insertion elements, transposases, ribosome binding sites and regulons. The 4 regions were located between nucleotides i) 1368633-1444823 (76,198nt), ii) 2612956-2730097 (117,148nt), iii) 4124625-4209022 (84,404nt), and iv) 4342879-4418289 ((75,418). At this time, no feature could be found that differentiated AT 8+mer deficient regions from regions separated by a shorter distance.
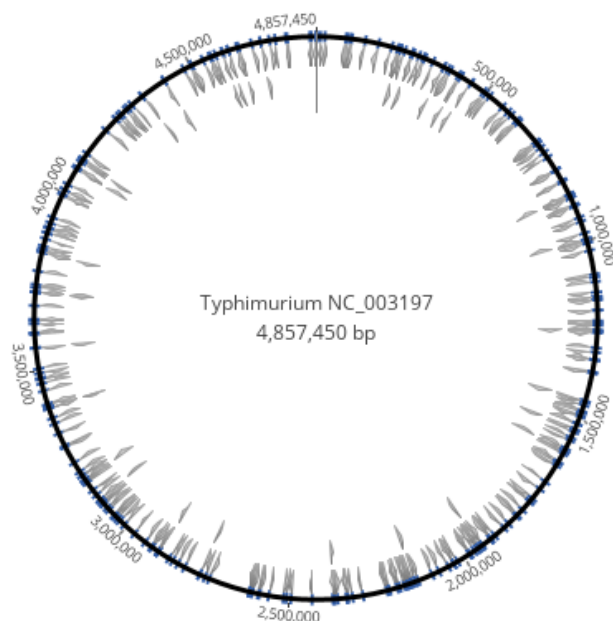
Figure 1. Locations of AT 8+mers in the genome of *Salmonella enterica* serovar Typhimurium LT2 NC_003197.2.

3.2. The AT 8+mer Motif in Bacteria was Specific to Genus and Species

Results of comparisons between Bacteria are shown in Table 1. Details include: 1) AT 8+mers in *S. enterica* groups were significantly more frequent than what was observed for E. coli ($p < 0.005$); 2) across bacterial classes *Vibrio vulnificus* cII had a minimum of 90.0 AT 8+mers and *Proteus mirabilis* had a 712.7; 3) standard deviations between strains in each Genera ranged from 2.3 for *Yersinia pseudotuberculosis* to 84.1 for *Enterococcus faecalis*; 4) all the genera examined, including those that differed by Phylum and Class, had a relative paucity of GC 8+mers as compared to AT 8+mers; thus, it appears there is a bias for Bacteria maintaining AT 8+mers in genomes, or inversely, selecting against GC 8+mers; 5) each species appeared distinctly different from others; 6) *Vibrio vulnificus* had 180 and 90 AT 8+mers in chromosomes cI and cII respectively; thus, AT 8+mer content might be a chromosomal characteristic that maintains the separation of the two chromosomes.

Results based upon taxonomic class were an average of 379.57 and 430.8 AT 8+mers, respectively, for the Gammaproteobacteria and Bacilli. In addition, standard deviations were large for both classes, namely 174.51 and 251.24 respectively. The range reflected by standard deviations masks the individuality of each genus and species. For example, *Bacillus anthracis* had an average of 432 AT 8+mers, whereas *Bacillus cereus* had an average of 700.3. Results from *S. enterica* subspecies I had a relatively small overall standard deviation of 12.83 AT 8+mers across 24 serovars, which agrees with the classification of *S. enterica* subspecies I as a single genomic grouping.
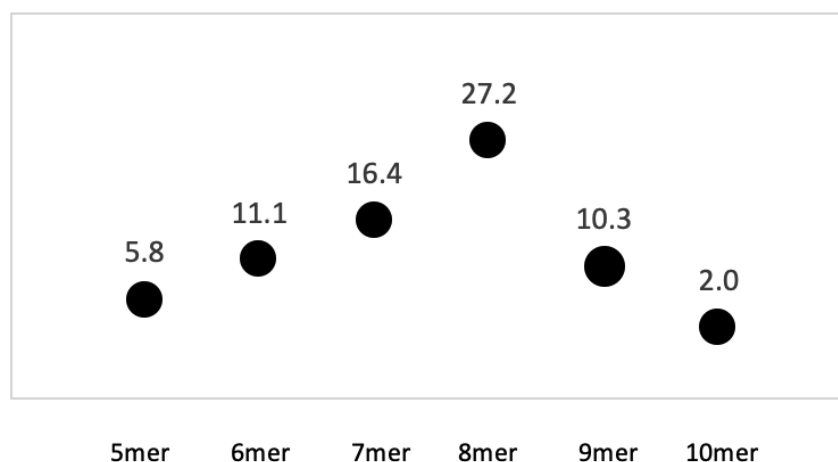
Figure 2. Ratios of AT homopolymers from 5 to 10 nucleotides in *Salmonella enterica* **serovar Typhimurium LT2 NC_003197.2.** The ratio of AT homopolymer kmers, either adenine or thymine but not mixed, to GC homopolymers was determined using Geneious software as described in text. The range in number of nucleotides per kmer searched was 5 to 10 (see legend label). Results show that a nucleotide motif of 8 was the most commonly encountered, and that approximately 27 AT homopolymers were found for every 1 GC AT homopolymer in the reference sequence of S. enterica LT2 NC_003197.2.

3.3. The AT 8+mer Motif in the chromosome is not Specific to Serovar

As mentioned in the previous section, AT 8+mers per *S. enterica* grouping was from 315.6 to 332.6, and the average was 322.2 +/- 12.83 AT 8+mers (Table 1). The length of the homopolymer is proposed to impact the physicochemical bending properties of DNA; thus, we wanted to account for every kmer of 8 nucleotides or more. Results from analysis of AT 8+mers between *S. enterica* serovars were: i) The incidence of AT 8+mers in the reference genome for serovar Typhimurium LT2 was the lowest of the 12 strains in the group, which suggests that using the serovar as a reference would not over-estimate the incidence of AT 8+mers for *S. enterica* or other genera; ii) the standard deviations for AT 8+mers in serovar Typhimurium and in the group of mixed serovars were, respectively, 13.0 and 13.9; iv) serovars Enteritidis and Typhi, with respective standard deviations of 10.5 and 5.9, appeared more clonal than Typhimurium, and this finding agrees with current knowledge; v) the foodborne serovars, namely Typhimurium, Enteritidis, and the group of mixed serovars, had a more variable motif content than human-restricted Typhi. Overall, the *S. enterica* serovar groups were not significantly different from each other although Typhi had the lowest standard deviation. There were not enough completed genomes of the poultry-restricted serovar Gallinarum to include it for this analysis; however, a reference genome of this important host-restricted poultry pathogen is evaluated in the following text.

Table 1. Expected versus observed occurrence of homopolymer strings of 8 and more nucleotides in genomes of Bacteria.

TABLE 1. Expected versus observed occurrence of homopolmer strings of 8 and more nucleotides in genomes of Eubacteria

Phylum	Genus species[1]	Other genome information	Number of genomes analysed		genome size (bp)	Common denominator (nt)[2]	Expected number of 8+k-mers	Observed AT 8-mers	Observed GC 8-mers	Observed vs expected AT 8-mers[3]	Observed vs expected GC 8-mers[3]
Proteobacteria	*Salmonella enterica*	Typhimurium	12	Average	4,890,448	16,299	300.0	332.6	17.2	1.11	0.06
				stdev	50,356	---	3.1	13.0	3.6	---	---
Proteobacteria	*Salmonella enterica*	Enteritidis	12	Average	4,686,462	16,299	287.5	323.7	21.5	1.13	0.07
				stdev	20,384	---	1.3	10.5	4.2	---	---
Proteobacteria	*Salmonella enterica*	Typhi	12	Average	4,770,414	16,299	292.7	316.9	29.5	1.08	0.10
				stdev	60,270	---	3.7	5.9	3.2	---	---
Proteobacteria	*Salmonella enterica*	mixed	12	Average	4,713,701	16,299	289.2	315.6	17.2	1.09	0.06
				stdev	80,652	---	4.9	13.9	4.4	---	---
Proteobacteria	*Escherichia coli*	---	12	Average	5,087,133	16,299	312.1	281.9	18.3	0.90	0.06
				stdev	262,098	---	16.1	30.2	6.5	---	---
Proteobacteria	*Proteus mirabilis*	---	3	Average	4,124,431	16,299	253.0	712.7	15.7	2.82	0.06
				stdev	83,305	---	5.1	42.0	2.1	---	---
Proteobacteria	*Shigella sonnei*	---	3	Average	4,929,599	16,299	302.4	261.3	11.7	0.86	0.04
				stdev	90,607	---	5.6	8.4	3.5	---	---
Proteobacteria	*Yersinia pseudotuberculosis*	---	3	Average	4,802,245	16,299	294.6	429.3	120.3	1.46	0.41
				stdev	118,706	---	7.3	2.3	3.8	---	---
Proteobacteria	*Vibrio vulnificus*	chromosome I	3	Average	3,330,104	16,299	204.3	180.0	10.3	0.88	0.05
				stdev	79,423	---	4.9	14.0	4.9	---	---
Proteobacteria	*Vibrio vulnificus*	chromosome II	3	Average	1,756,668	16,299	107.8	90.0	3.3	0.83	0.03
				stdev	87,177	---	5.3	7.9	3.1	---	---
Firmicutes	*Staphylococcus aureus*	---	3	Average	2948373	16,299	180.9	108.3	0.0	0.60	0.00
				stdev	114371	---	7.0	10.6	0.0	---	---
Firmicutes	*Streptococcus pyogenes*	---	3	Average	1,895,707	16,299	116.3	263.7	0.3	2.27	0.00
				stdev	42,370	---	2.6	15.4	0.6	---	---
Firmicutes	*Enterococcus faecalis*	---	3	Average	3,090,387	16,299	189.6	649.7	2.0	3.42	0.01
				stdev	117,259	---	7.2	84.1	3.5	---	---
Firmicutes	*Bacillus anthracis*	---	3	Average	5,228,732	16,299	320.8	432.0	1.3	1.35	0.00
				stdev	1,349	---	0.1	11.5	0.6	---	---
Firmicutes	*Bacillus cereus*	---	3	Average	5,406,060	16,299	331.7	700.3	13.0	2.11	0.04
				stdev	16615	---	1.0	53.7	6.1	---	---

[1] Genomes included in analysis are listed in supplementary Table S1 with NCBI accession numbers.

[2] The common denominater of 16,299 nucleotides (nt) used to normalize variation in geome size was obtained from *Salmonella enterica* subspecies I serotype Typhimurum LT2 (NC_003197.2) as described in text.

[3] Values greater than one indicate more motifs observed than expected, less than one indicates fewer were observed than expected.

3.4. Characterization of the AT 8+mer Motif in Poultry-Associated Serovars of Salmonella

The association of motif content with gene function was catalogued. Table S2 lists 294 intergenic, regulatory, and gene AT 8+mer loci in serovar Typhimurium LT2; of these, 131 were intergenic, 150 were in genes, and 13 were in regulatory regions. Table S3 lists 185 genes and regulatory regions with AT 8+mers, which were located by comparing reference genomes of serovars Typhimurium , Enteritidis, and Gallinarum; thus, 20 more genes and regulatory regions were identified using the comparative approach. Gene functions were assigned in the last column. Some genes in serovars Enteritidis and/or Gallinarum did not have homologs in the Typhimurium reference strain, and vice versa. Serovars Typhimurium, Gallinarum, and Enteritidis each had 3, 22, and 5 pseudogenes with AT

8+mers, respectively, while each genome had a total of 40, 287, and 96 pseudogenes. Thus respectively, 7.5%, 8.4%, and 5.2% of pseudogenes involved the motif.

The regulons with the motif encompassed mostly metabolic functions. Regulons associated with the genes STM2277, STM0664, and STM4585 have some association with antibiotics (references not shown). Another class of gene that could be of interest for targeting of antimicrobials include transporters, and there are 20 listed in Table S3 [33]. Cell surface molecules that include genes with the motif include colonic acid, lipopolysaccharide, flagella, and fimbria. A virulence factor with the motif is MviM.

3.5. Location of AT 8+mers within Mutated Open Reading Frames

Table 2 lists 22 genes that are a subset of those from Table S3. Serovar Gallinarum had 17 pseudogenes that were intact in serovars Enteritidis and Typhimurium. There were 4 pseudogenes in both serovars Enteritidis and Gallinarum that were intact in Typhimurium. STM1666 was unusual because it was a pseudogene in serovar Typhimurium, absent in Gallinarum, and intact in Enteritidis. STM1666 is thus one of the few genes that differentiates serovar Enteritidis from both Gallinarum and Typhimurium. The gene was present in 31 other Salmonellae genomes available at NCBI, but no serovar was reported for these and no other information was present. There was no homolog of STM1666 in other bacteria.

The location of the first adenine or thymine in homopolymer strings in the subset of mutated genes was located relative to start codons ATG or GTG. Results were that 8, 7, 6, and 1 gene(s) fell into the 1st, 2nd, 3rd, and 4th quartile of gene lengths. These findings thus agree in concept with others that homopolymer tracts are found closer to the 5' end of open reading frames [6]. However, we suggest that these data are more accurately summarized as AT 8+mers of *S. enterica* do not often occur near the 3' end of genes. An area of future research is to determine if taxonomic classes of bacteria differ in location of mutable homopolymer strings within open reading frames.

3.6. Plasmids of S. enterica subspecies I have serovar specific AT 8+mer motifs

Some of the most important foodborne *Salmonella* serovars harbor a large low-copy virulence plasmid [34]. *S. enterica* serovar Typhimurium, and related immunogenic variants with LPS O-antigens 1,4,[5],12, have one of the largest plasmids (pSLT) (NC_003277.2) [30]. At 93,933nt, pSLT would be expected to have 5 or 6 AT 8+mers if the average of motifs found within the chromosome, which was 16,450nt, applied to the extrachromosomal plasmid. Typhimurium pSLT had 6 AT 8+mer loci (Table 3). No evidence of hypermutability of these loci was found within serovar by BLAST analysis, thus AT 8+mers within serovar appeared conserved. The motif in plasmids of the other two serovars differed substantially from each other and from Typhimurium (Table 3). The serovar Enteritidis plasmid pSENV (CP063701.1) and Gallinarum SG9 strain 9 plasmid (CM001154.1) were 59,372 and 87,371 nucleotides, respectively, and each had 4 AT 8+mers. Overall, there was substantial variation between the AT 8+mer motif in each of the plasmids, thus AT 8+mer content was serovar specific.

Table 3. AT 8+mer motif variation in 3 serotypes of *Salmonella enterica* subspecies I

Serovar	Typhimurium	Enteritidis	Gallinarum	Other information		
Accession	NC_003277.2	NZ_CP063701.1	CM001154.1			
Other name	pSLT	pSENV	str. SG9			
Atmer size	*Description of loci*	*Variation from pSLT*	*Variation from pSLT*	*pSLT start*	*pSLT end*	*gene function*
9mer	Intergenic (PSLT039-PSLT038)	1nt deletion	1nt deletion	27961	27969	*spvB-spvC*;58nt upstream from spvC start
10mer	Intergenic (PSLT041-PSLT042)	1nt deletion	2nt deletion	32324	32333	*spvR*-PSLT041: 10nt upstream from *spvR* start
9mer	PSLT076	present	absent	62299	62307	*traY:* conjugative transfer: oriT nicking
8mer	PSLT088	absent	present	69093	69100	*traC*: conjugative transfer: assembly
8mer	PSLT102	absent	absent	82835	82842	*traS*: conjugative transfer: surface exclusion
8mer	PSLT111	truncated	truncated	93512	93519	*finO*: conjugative transfer: regulation

4. Discussion

The AT 8+mer motif was shown here to be associated with 1) *S. enterica* speciation, 2) serovar designation via details of the AT 8+mer in plasmids, 3) hypermutability in genes and regulatory regions that impact phenotype, growth potential, virulence and metabolism of *Salmonella enterica* subspecies I. AT 8+mers influence microbial biology. For example, A and T homopolymers impact transcription termination in Archea [35]. The canine herpesvirus thymidine kinase gene has mutational hotspots at stretches of 8 adenines [36]. T7 bacteriophage RNA polymerases undergo transcription slippage at A and T homopolymers [37]. As mentioned previously for *S. enterica* serovar Enteritidis, a mutational hotspot in 1 of 8 adenines increased virulence [3]. We did not conclude that mutation in the motif was primarily in the 5′ region of genes; however, it is a subject that requires development of better algorithms to address in detail [6]. These results suggest that the AT 8+mer motif facilitates maintenance of *S. enterica* as a persistent and important foodborne pathogen, as well as impacting epidemiological outcomes by contributing to ecological adaptation [38].

If AT 8+mers are mutational hotspots, then there must also exist a mechanism for repair. Otherwise, evolution of any one serovar of *S. enterica* would be unidirectional towards extinction. There are several examples of *Salmonella* serovars, e. g. Typhimurium, Enteritidis, Newport, Infantis, and Heidelberg, that persist over decades; however, the majority of the 1500 serovars within subspecies I cause illness inconsistently, rarely, or never [1, 16]. For this reason, we theorize there is another function for AT 8+mers. It is proposed that chromosomal AT 8+mers align sections of genomes during replication and DNA

repair processes. This function would result in repair of mutations occurring between stretches of wildtype AT 8+mers [39, 40]. It would also account for an inherent mechanism of self-recognition, which would facilitate preferential, but not exclusive, DNA exchange within the species. The pan-genome of *S. enterica* subspecies 1 has a mosaic structure, with frequent inversions, deletions, and insertions occurring between serovars; however, the chromosomal arrangement of many *Salmonella* lineages is comparatively stable [25, 32, 41, 42]. Thus the motif would account for i) the stability of some serovars with conserved genome features that are persistent, e. g. serovar Typhimurium [1], ii) the occasional emergence of a new serovar that happens to undergo clonal expansion in an environment favorable for growth, e. g. serovar Tennessee in peanut butter [43, 44], iii) the rare emergence of a hybrid strain following a major recombination event that results in rapid proliferation of a serovar with new biological properties, e. g. serovar Enteritidis [45], and iv) the periodic emergence and disappearance of serovars that are not optimized for the survival in the environment in which they are generated. Another finding that the AT 8+mer impacts speciation is that the two chromosomes of *Vibrio vulnificus* differ substantially and would thus never be predicted to coalesce.

5. Conclusion

These results support and add to the concept that the homopolymer AT motif is an important regulator of bacterial phase transition, ecological persistence, and association with disease [6]. In regards to *Salmonella enterica* subspecies I, the motif is proposed to be important for maintaining optimal gene content, with the coincidental impact of contributing to serovar diversity [46-49]. Future research on the AT 8+mer contribution to these proposed functions will require proof of concept experimentation. Biological experimentation will focus on finding environmental niches that facilitate genomic exchange and repair mechanisms. Analyzing the impact of the motif on the safety of the food supply may require methods with detection limits that are orders of magnitude lower than those used to currently detect bacteria. This is because a successful recombinant may at first be a rare cell type [50, 51]. Further analysis into the impact of AT 8+mers on the ability of *S. enterica* to survive and persist in environments associated with foodborne illness is thus warranted.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: List of bacterial genomes analyzed for AT 8+mer homopolymers, Table S2: Location and classification of all AT8+mers in Typhimurium LT2.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, Jean Guard and Adam Rivers; methodology, all authors; validation, Jean Guard, Adam Rivers, and Justin Vaughn; formal analysis, Adam Rivers; investigation, Jean Guard and Justin Vaughn; resources, Jean Guard.; data curation, Jean Guard; writing—original draft preparation, Jean Guard.; writing—review and editing, Jean Guard, Michael Rothrock, Adelumola Oladeinde, Devendra Shah.; visualization, Jean Guard, Adam Rivers, Justin Vaughn, Adelumola Oladeinde, Devendra Shah.; supervision, Jean Guard; project administration, Jean Guard, Adam Rivers, and Michael Rothrock; funding acquisition, Jean Guard, Adam Rivers, and Michael Rothrock. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The U.S. Department of Agriculture, Agricultural Research Service project plan number 6040-32000-012-00-D and by The National Institute of Food and Agriculture, Agriculture and Food Research Initiative Grant Number 2019-67021-29924.

Data Availability Statement: The database analyzed for this project can be found at the National Center for Biotechnology Institute (NCBI) at https://www.ncbi.nlm.nih.gov.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CDC. An Atlas of Salmonella in the United States, 1968-2011: Laboratory-based Enteric Disease Surveillance. Atlanta, Georgia: Centers for Disease Control and Prevention (CDC), 2013.

2. Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, et al. Preliminary Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food - Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2016-2019. MMWR Morb Mortal Wkly Rep. 2020;69(17):509-14. Epub 2020/05/01. doi: 10.15585/mmwr.mm6917a1. PubMed PMID: 32352955.

3. Guard J, Cao G, Luo Y, Baugher JD, Davison S, Yao K, et al. Genome sequence analysis of 91 Salmonella Enteritidis isolates from mice caught on poultry farms in the mid 1990s. Genomics. 2020;112(1):528-44. Epub 2019/04/12. doi: 10.1016/j.ygeno.2019.04.005. PubMed PMID: 30974149.

4. Morales CA, Guard J, Sanchez-Ingunza R, Shah DH, Harrison M. Virulence and metabolic characteristics of Salmonella enterica serovar enteritidis strains with different sefD variants in hens. Appl Environ Microbiol. 2012;78(18):6405-12. Epub 2012/06/26. doi: 10.1128/AEM.00852-12. PubMed PMID: 22729535; PubMed Central PMCID: PMCPMC3426706.

5. Sanchez-Ingunza R, Guard J, Morales CA, Icard AH. Reduction of Salmonella Enteritidis in the spleens of hens by bacterins that vary in fimbrial protein SefD. Foodborne Pathog Dis. 2015;12(10):836-43. Epub 2015/07/29. doi: 10.1089/fpd.2015.1971. PubMed PMID: 26218804; PubMed Central PMCID: PMCPMC4601671.

6. Orsi RH, Bowen BM, Wiedmann M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. BMC Genomics. 2010;11:102. Epub 2010/02/11. doi: 10.1186/1471-2164-11-102. PubMed PMID: 20144225; PubMed Central PMCID: PMCPMC2831843.

7. Reich Z, Friedman P, Levin-Zaidman S, Minsky A. Effects of adenine tracts on the B-Z transition. Fine tuning of DNA conformational transition processes. J Biol Chem. 1993;268(11):8261-6. Epub 1993/04/15. PubMed PMID: 8463336.

8. Hines ER, Kolek OI, Jones MD, Serey SH, Sirjani NB, Kiela PR, et al. 1,25-dihydroxyvitamin D3 down-regulation of PHEX gene expression is mediated by apparent repression of a 110 kDa transfactor that binds to a polyadenine element in the promoter. J Biol Chem. 2004;279(45):46406-14. Epub 2004/09/01. doi: 10.1074/jbc.M404278200. PubMed PMID: 15337762.

9. Lindemose S, Nielsen PE, Mollegaard NE. Polyamines preferentially interact with bent adenine tracts in double-stranded DNA. Nucleic Acids Res. 2005;33(6):1790-803. Epub 2005/03/25. doi: 10.1093/nar/gki319. PubMed PMID: 15788751; PubMed Central PMCID: PMCPMC1069516.

10. Jung BH, Beck SE, Cabral J, Chau E, Cabrera BL, Fiorino A, et al. Activin type 2 receptor restoration in MSI-H colon cancer suppresses growth and enhances migration with activin. Gastroenterology. 2007;132(2):633-44. Epub 2007/01/30. doi: 10.1053/j.gastro.2006.11.018. PubMed PMID: 17258738; PubMed Central PMCID: PMCPMC4154562.

11. Agnoli K, Haldipurkar SS, Tang Y, Butt AT, Thomas MS. Distinct Modes of Promoter Recognition by Two Iron Starvation sigma Factors with Overlapping Promoter Specificities. J Bacteriol. 2019;201(3). Epub 2018/11/21. doi: 10.1128/JB.00507-18. PubMed PMID: 30455278; PubMed Central PMCID: PMCPMC6349086.

12. Roberts JD, Nguyen D, Kunkel TA. Frameshift fidelity during replication of double-stranded DNA in HeLa cell extracts. Biochemistry. 1993;32(15):4083-9. Epub 1993/04/20. doi: 10.1021/bi00066a033. PubMed PMID: 8385995.

13. Traverse CC, Ochman H. Genome-Wide Spectra of Transcription Insertions and Deletions Reveal That Slippage Depends on RNA:DNA Hybrid Complementarity. mBio. 2017;8(4). Epub 2017/08/31. doi: 10.1128/mBio.01230-17. PubMed PMID: 28851848; PubMed Central PMCID: PMCPMC5574713.

14. Gragg H, Harfe BD, Jinks-Robertson S. Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in Saccharomyces cerevisiae. Mol Cell Biol. 2002;22(24):8756-62. Epub 2002/11/26. doi: 10.1128/mcb.22.24.8756-8762.2002. PubMed PMID: 12446792; PubMed Central PMCID: PMCPMC139878.

15. Guard J, Morales CA, Fedorka-Cray P, Gast RK. Single nucleotide polymorphisms that differentiate two subpopulations of Salmonella enteritidis within phage type. BMC Res Notes. 2011;4:369. Epub 2011/09/29. doi: 10.1186/1756-0500-4-369. PubMed PMID: 21942987; PubMed Central PMCID: PMCPMC3220660.

16.   Grimont P, Weill F-X. Antigenic formulae of the Salmonella serovars. 9th Edition. Paris, France: WHO Collaborating Centre for Reference and Research on Salmonella. Paris, France: World Health Organization, 2007.

17.   Guard-Bouldin J, Gast RK, Humphrey TJ, Henzler DJ, Morales C, Coles K. Subpopulation characteristics of egg-contaminating Salmonella enterica serovar Enteritidis as defined by the lipopolysaccharide O chain. Appl Environ Microbiol. 2004;70(5):2756-63. Epub 2004/05/07. doi: 10.1128/aem.70.5.2756-2763.2004. PubMed PMID: 15128529; PubMed Central PMCID: PMCPMC404386.

18.   Gantois I, Ducatelle R, Pasmans F, Haesebrouck F, Van Immerseel F. The Salmonella Enteritidis lipopolysaccharide biosynthesis gene rfbH is required for survival in egg albumen. Zoonoses Public Health. 2009;56(3):145-9. Epub 2008/11/08. doi: 10.1111/j.1863-2378.2008.01195.x. PubMed PMID: 18990194.

19.   Parker CT, Liebana E, Henzler DJ, Guard-Petter J. Lipopolysaccharide O-chain microheterogeneity of Salmonella serovars Enteritidis and Typhimurium. Environ Microbiol. 2001;3(5):332-42. Epub 2001/06/26. doi: 10.1046/j.1462-2920.2001.00200.x. PubMed PMID: 11422320.

20.   Huang K, Fresno AH, Skov S, Olsen JE. Dynamics and Outcome of Macrophage Interaction Between Salmonella Gallinarum, Salmonella Typhimurium, and Salmonella Dublin and Macrophages From Chicken and Cattle. Front Cell Infect Microbiol. 2019;9:420. Epub 2020/01/31. doi: 10.3389/fcimb.2019.00420. PubMed PMID: 31998655; PubMed Central PMCID: PMCPMC6966237.

21.   Foley SL, Nayak R, Hanning IB, Johnson TJ, Han J, Ricke SC. Population dynamics of Salmonella enterica serovars in commercial egg and poultry production. Appl Environ Microbiol. 2011;77(13):4273-9. Epub 2011/05/17. doi: 10.1128/AEM.00598-11. PubMed PMID: 21571882; PubMed Central PMCID: PMCPMC3127710.

22.   Branchu P, Bawn M, Kingsley RA. Genome Variation and Molecular Epidemiology of Salmonella enterica Serovar Typhimurium Pathovariants. Infect Immun. 2018;86(8). Epub 2018/05/23. doi: 10.1128/IAI.00079-18. PubMed PMID: 29784861; PubMed Central PMCID: PMCPMC6056856.

23.   McMillan EA, Gupta SK, Williams LE, Jove T, Hiott LM, Woodley TA, et al. Antimicrobial Resistance Genes, Cassettes, and Plasmids Present in Salmonella enterica Associated With United States Food Animals. Front Microbiol. 2019;10:832. Epub 2019/05/07. doi: 10.3389/fmicb.2019.00832. PubMed PMID: 31057528; PubMed Central PMCID: PMCPMC6479191.

24.   Desai PT, Porwollik S, Long F, Cheng P, Wollam A, Bhonagiri-Palsikar V, et al. Evolutionary Genomics of Salmonella enterica Subspecies. mBio. 2013;4(2). Epub 2013/03/07. doi: 10.1128/mBio.00579-12. PubMed PMID: 23462113; PubMed Central PMCID: PMCPMC3604774.

25.   Achtman M, Hale J, Murphy RA, Boyd EF, Porwollik S. Population structures in the SARA and SARB reference collections of Salmonella enterica according to MLST, MLEE and microarray hybridization. Infect Genet Evol. 2013;16:314-25. Epub 2013/03/20. doi: 10.1016/j.meegid.2013.03.003. PubMed PMID: 23507027.

26.   Turcotte C, Woodward MJ. Cloning, DNA nucleotide sequence and distribution of the gene encoding the SEF14 fimbrial antigen of Salmonella enteritidis. J Gen Microbiol. 1993;139(7):1477-85. Epub 1993/07/01. doi: 10.1099/00221287-139-7-1477. PubMed PMID: 8371111.

27.   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. Epub 1994/06/01. doi: 10.1111/j.1365-2958.1994.tb01077.x. PubMed PMID: 7934897.

28.   Matthews TD, Schmieder R, Silva GG, Busch J, Cassman N, Dutilh BE, et al. Genomic Comparison of the Closely-Related Salmonella enterica Serovars Enteritidis, Dublin and Gallinarum. PLoS One. 2015;10(6):e0126883. Epub 2015/06/04. doi: 10.1371/journal.pone.0126883. PubMed PMID: 26039056; PubMed Central PMCID: PMCPMC4454671.

29.   Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2019;47(D1):D23-D8. Epub 2018/11/06. doi: 10.1093/nar/gky1069. PubMed PMID: 30395293; PubMed Central PMCID: PMCPMC6323993.

30.   McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature. 2001;413(6858):852-6. Epub 2001/10/26. doi: 10.1038/35101614. PubMed PMID: 11677609.

31.   Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, et al. Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. 2008;18(10):1624-37. Epub 2008/06/28. doi: 10.1101/gr.077404.108. PubMed PMID: 18583645; PubMed Central PMCID: PMCPMC2556274.

32.   Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, 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. Epub 2013/02/06. doi: 10.1371/journal.pone.0055254. PubMed PMID: 23383127; PubMed Central PMCID: PMCPMC3559427.

33.   Du D, van Veen HW, Murakami S, Pos KM, Luisi BF. Structure, mechanism and cooperation of bacterial multidrug transporters. Curr Opin Struct Biol. 2015;33:76-91. Epub 2015/08/19. doi: 10.1016/j.sbi.2015.07.015. PubMed PMID: 26282926.

34.   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. Epub 1994/11/15. doi: 10.1111/j.1574-6968.1994.tb07253.x. PubMed PMID: 8001760.

35.   Santangelo TJ, Cubonova L, Skinner KM, Reeve JN. Archaeal intrinsic transcription termination in vivo. J Bacteriol. 2009;191(22):7102-8. Epub 2009/09/15. doi: 10.1128/JB.00982-09. PubMed PMID: 19749050; PubMed Central PMCID: PMCPMC2772485.

36.   Yamada S, Matsumoto Y, Takashima Y, Otsuka H. Mutation hot spots in the canine herpesvirus thymidine kinase gene. Virus Genes. 2005;31(1):107-11. Epub 2005/06/21. doi: 10.1007/s11262-005-2206-y. PubMed PMID: 15965615.

37.   Koscielniak D, Wons E, Wilkowska K, Sektas M. Non-programmed transcriptional frameshifting is common and highly RNA polymerase type-dependent. Microb Cell Fact. 2018;17(1):184. Epub 2018/11/27. doi: 10.1186/s12934-018-1034-4. PubMed PMID: 30474557; PubMed Central PMCID: PMCPMC6260861.

38.   Silva C, Puente JL, Calva E. Salmonella virulence plasmid: pathogenesis and ecology. Pathog Dis. 2017. Epub 2017/06/25. doi: 10.1093/femspd/ftx070. PubMed PMID: 28645187.

39.   Brandis G, Cao S, Hughes D. Co-evolution with recombination affects the stability of mobile genetic element insertions within gene families of Salmonella. Mol Microbiol. 2018;108(6):697-710. Epub 2018/04/01. doi: 10.1111/mmi.13959. PubMed PMID: 29603442.

40.   Brandis G, Cao S, Hughes D. Measuring Homologous Recombination Rates between Chromosomal Locations in Salmonella. Bio Protoc. 2019;9(3):e3159. Epub 2019/02/05. doi: 10.21769/BioProtoc.3159. PubMed PMID: 33654967; PubMed Central PMCID: PMCPMC7854157.

41.   Achtman M, Zhou Z, Alikhan NF, Tyne W, Parkhill J, Cormican M, et al. Genomic diversity of Salmonella enterica -The UoWUCC 10K genomes project. Wellcome Open Res. 2020;5:223. Epub 2021/02/24. doi: 10.12688/wellcomeopenres.16291.2. PubMed PMID: 33614977; PubMed Central PMCID: PMCPMC7869069.

42.   Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of Salmonella. PLoS Genet. 2018;14(4):e1007261. Epub 2018/04/06. doi: 10.1371/journal.pgen.1007261. PubMed PMID: 29621240; PubMed Central PMCID: PMCPMC5886390.

43.   Dong HJ, Cho S, Boxrud D, Rankin S, Downe F, Lovchik J, et al. Single-nucleotide polymorphism typing analysis for molecular subtyping of Salmonella Tennessee isolates associated with the 2007 nationwide peanut butter outbreak in the United States. Gut Pathog. 2017;9:25. Epub 2017/05/05. doi: 10.1186/s13099-017-0176-y. PubMed PMID: 28469710; PubMed Central PMCID: PMCPMC5412032.

44.   Wilson MR, Brown E, Keys C, Strain E, Luo Y, Muruvanda T, et al. Whole Genome DNA Sequence Analysis of Salmonella subspecies enterica serovar Tennessee obtained from related peanut butter foodborne outbreaks. PLoS One. 2016;11(6):e0146929. Epub 2016/06/04. doi: 10.1371/journal.pone.0146929. PubMed PMID: 27258142; PubMed Central PMCID: PMCPMC4892500.

45.   J. G, D. S, C. M, D. C. Evolutionary trends associated with niche specialization as modeled by whole genome analysis of egg-contaminating *Salmonella enterica* serovar Enteritidis. In: Porwollik S, editor. Salmonella: From Genome to Function. San Diego: Caister Academic Press; 2011. p. 91 - 106.

46.   Zhou Z, McCann A, Litrup E, Murphy R, Cormican M, Fanning S, et al. Neutral genomic microevolution of a recently emerged pathogen, Salmonella enterica serovar Agona. PLoS Genet. 2013;9(4):e1003471. Epub 2013/05/03. doi: 10.1371/journal.pgen.1003471. PubMed PMID: 23637636; PubMed Central PMCID: PMCPMC3630104.

47.   Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, et al. Evolution and population structure of Salmonella enterica serovar Newport. J Bacteriol. 2010;192(24):6465-76. Epub 2010/10/12. doi: 10.1128/JB.00969-10. PubMed PMID: 20935094; PubMed Central PMCID: PMCPMC3008538.

48.   Park CJ, Andam CP. Distinct but Intertwined Evolutionary Histories of Multiple Salmonella enterica Subspecies. mSystems. 2020;5(1). Epub 2020/01/16. doi: 10.1128/mSystems.00515-19. PubMed PMID: 31937675; PubMed Central PMCID: PMCPMC6967386.

49.   Liu Y, Zhang DF, Zhou X, Xu L, Zhang L, Shi X. Comprehensive Analysis Reveals Two Distinct Evolution Patterns of Salmonella Flagellin Gene Clusters. Front Microbiol. 2017;8:2604. Epub 2018/01/10. doi: 10.3389/fmicb.2017.02604. PubMed PMID: 29312269; PubMed Central PMCID: PMCPMC5744181.

50.   Richards AK, Hopkins BA, Shariat NW. Conserved CRISPR arrays in Salmonella enterica serovar Infantis can serve as qPCR targets to detect Infantis in mixed serovar populations. Lett Appl Microbiol. 2020;71(2):138-45. Epub 2020/04/26. doi: 10.1111/lam.13296. PubMed PMID: 32333808.

51.   Shariat N, Dudley E. CRISPR Typing of Salmonella Isolates. Methods Mol Biol. 2021;2182:39-44. Epub 2020/09/08. doi: 10.1007/978-1-0716-0791-6_5. PubMed PMID: 32894485.