



Comprehensive Analysis Reveals Two Distinct Evolution Patterns of *Salmonella* Flagellin Gene Clusters

Yue Liu[†], Dao-Feng Zhang[†], Xiujuan Zhou, Li Xu, Lida Zhang and Xianming Shi*

MOST-USDA Joint Research Center for Food Safety, School of Agriculture and Biology, and State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:

John R. Battista,
Louisiana State University,
United States

Reviewed by:

Xiaohui Zhou,
University of Connecticut,
United States
Sunil D. Saroj,
Symbiosis International University,
India

*Correspondence:

Xianming Shi
xmshi@sjtu.edu.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 01 September 2017

Accepted: 14 December 2017

Published: 22 December 2017

Citation:

Liu Y, Zhang D-F, Zhou X, Xu L,
Zhang L and Shi X (2017)
Comprehensive Analysis Reveals Two
Distinct Evolution Patterns of
Salmonella Flagellin Gene Clusters.
Front. Microbiol. 8:2604.
doi: 10.3389/fmicb.2017.02604

Salmonella is one of the primary causes of foodborne disease, especially *Salmonella enterica* subsp. *enterica* (I) which has caused ~99% of clinical salmonellosis cases for humans and domestic mammals. The flagella genes, *fliC* and *fliB*, which encode the *Salmonella* phase 1 and phase 2 antigens respectively, are considered as the *Salmonella* serotype determinant genes, and contribute to the virulence of *Salmonella*. However, the evolution of the two flagellin genes is still not well-understood. In this study, the *fliC* and *fliB* gene clusters were analyzed among 205 *S. enterica* subspecies I genomes. The dataset covered 87 different serovars of *S. enterica* subsp. *enterica* and included 9 genomes (six serovars) of four other *Salmonella* subspecies. Based on a pan-genome definition and flanked gene linkages, the *fliC* and *fliB* gene clusters were identified in 207 (91 serovars) and 138 (61 serovars) genomes, respectively. A phylogenetic tree constructed based on SNPs (Single Nucleotide Polymorphisms) of core genes were used to reflect the essential evolutionary relationships among various serovars. Congruence analysis was performed among the core genome and each gene of *fliC* and *fliB* gene clusters, with only *fliA* and *fliS* showing congruence to *Salmonella* core genome. Congruence was also observed among *fliB*, *fliC/fliB*, and *fliD* genes, and their phylogeny revealed a division into two major groups, which strongly corresponded to monophasic and biphasic serovars. Besides, homologous recombination events referring *fliB*, *fliC*, and *fliD* were found to have mainly occurred within each group. These results suggested two distinct evolutionary patterns of *Salmonella* flagellin gene clusters. Further insight on the evolutionary implication of the two patterns and a framework for phase variation mechanism are needed to be further processed.

Keywords: *Salmonella enterica*, flagellin genes, *fliC*, *fliB*, recombination

INTRODUCTION

Salmonella is one of the primary causes of foodborne illness around the world, leading to a huge burden of illness calculated at nearly 93.8 million global cases annually (Majowicz et al., 2010). For understanding the epidemiology of these microorganisms and the investigation of outbreaks caused by *Salmonella*, serotyping has been greatly facilitated by the characterization of isolates. The *Salmonella* genus is composed of two species, *Salmonella bongori* and *Salmonella enterica*, the latter being further classified into six subspecies, *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI) (Grimont and Weill, 2007). *Salmonella* is an

extremely diverse species, of which more than 2,500 serovars has been reported. *S. enterica* subsp. *enterica* (I) strains, the vast majority of *Salmonella* strains, could be isolated from humans and warm-blooded animals, while all the other subspecies and *S. bongori* are more frequently isolated from cold-blooded animals (Nataro et al., 2011). Importantly, *S. enterica* subsp. *enterica* (I) has caused ~99% of diagnosed salmonellosis cases for humans and domestic mammals (Centers for Disease Control and Prevention (CDC), 2011). Strains with different serotypes show different levels of variation, a single host (e.g., *S. Typhi*) being adapted by monomorphic serotypes comprise a single strain in a single lineage, whilst polymorphic serotypes comprise multiple lineages of strains which may or may not be host adapted (e.g., *S. Choleraesuis*) (Wain and O'Grady, 2017).

Based on White-Kauffmann scheme, the *Salmonella* strains have been identified by their surface antigens, on the basis of somatic (O), flagellar (H), and capsular (Vi) antigens (Brenner et al., 2000) and their reactivity to antisera (Schrader et al., 2008). Flagella (*fliC* or *fljB*) protein is the primary globular protein forming the filament of flagellum in the majority of *Salmonella* strains, which functions as a main structural subunit (Haiko and Westerlund-Wikstrom, 2013). These two flagellin genes in most *Salmonella* strains, *fliC* and *fljB*, encode the phase 1 and phase 2 flagellins, respectively. The *fliC* gene is existent in all *Salmonella* cells and has homologs in other enteric bacteria, located in one of the flagellar biosynthesis operons (Macnab, 1992). The *fljB* gene is unique to *S. enterica* and is present in four of the *S. enterica* subspecies (subspecies I, II, IIIb, and VI), separately located on the chromosome from the *fliC* gene (McQuiston et al., 2004). However, these two flagellin genes cannot be expressed simultaneously. The second flagellin locus is considered as a genetic "spare tyre" availed in certain environmental circumstances (McQuiston et al., 2008). *Salmonella* isolates are indicated biphasic, monophasic, or non-motile according to the "phase variation," expressing both, one, or no flagellar phases, which is regulated by a switch mechanism, an invertible element *hin* (Bonifield and Hughes, 2003). The *fljB* is considered as a duplication of *fliC* but is flanked with the genes, *fljA* and *hin* (Szekely and Simon, 1983), involving in the switch mechanism, which appears to be specific to *Salmonella* (Zieg et al., 1978). Recombination of these genes is the basic source of flagella antigenic variation (Smith et al., 1990) and may occur between *fliC* of co-infecting organisms or *fliC* and *fljB* of the same isolate (Okazaki et al., 1993). The function of flagellar genes have been studied well; the *fliA* existing upstream of the coding region is specific to promoters for flagellar operons (Ohnishi et al., 1990), *fliB* is responsible for flagellar methylation (Parkhill et al., 2001), *fliC* encodes the flagellar filament protein (Ogushi et al., 2001), and *fliD* encodes the filament cap protein of the flagellar apparatus (Yokoseki et al., 1995). The expression of *fliS* facilitates the export of flagellin through the flagellum-specific export pathway (Yokoseki et al., 1995), whereas the *fljB* gene, which expresses phase II flagellin protein, constitutes an operon with the *fljA* gene, which encodes a negative regulator for *fliC* expression (Yamamoto and Kutsukake, 2006).

The flagellin genes are useful as a model system for studying genetic differentiation (Okazaki et al., 1993). Previous study has

revealed that the 5' and 3' ends of the *fliC* and *fljB* gene sequences from different antigenic types are highly conserved, but the central region sequences can be highly variable (Okazaki et al., 1993). A molecular serotyping for the H antigens by targeting the genes encoding the immunologically recognized antigen could be linked to traditional serotyping methods (McQuiston et al., 2011). Based on phylogenetic relationships adapted from previous studies, the acquisition of phase variation, the biphasic *Salmonella* flourished (McQuiston et al., 2008). However, this conclusion was drawn from analysis of limited numbers of samples and the basic evolutionary process of this apparently high level of surface-antigen diversity in the *Salmonella* is not well-understood (McQuiston et al., 2008). There is also a lack of data on the detailed phylogenetic description of the flagellar gene relationship among the *Salmonella*, as a framework for evolutionary studies. Therefore, it is worthwhile to seek full evolutionary insights and the evolutionary relationships of the *Salmonella* flagellin alleles to estimate the recombination within these loci. So, in this study, comparative genomics was performed to obtain additional evolutionary information regarding flagellin genes, which showed that there are two distinct evolutionary patterns of *fliC* and *fljB* clusters during the speciation and serovar diversification of *Salmonella*.

MATERIALS AND METHODS

Data Collection

A total of 214 *Salmonella* complete genomes and genome assemblies of highest quality (Table S1) were retrieved from the NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>). All the *Salmonella* genomes with annotations of different serotypes were downloaded as of May, 2016, including most of complete genomes as well as draft genomes of type strains or strains of some serovars for which a complete genome was not available. Five subspecies of *S. enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), and *houtenae* (IV), and a total of 87 serovars of *enterica* (I), were under consideration. MLST (Multilocus Sequence Typing) (Achtman et al., 2012), together with checking the original studies was implemented to identify their serovars. Housekeeping genes, *thrA*, *purE*, *sucA*, *hisD*, *aroC*, *hemD*, and *dnaN* were targeted for the MLST scheme (Achtman et al., 2012). The nucleotide sequences were then aligned against seven housekeeping genes. Outcome of sequences were subsequently submitted to the MLST database (<http://mlst.warwick.ac.uk/mlst/dbs/Senterica>) and assigned existing or novel allele type numbers. The composite sequence STs (sequence types) were defined by the database based on the set of allelic profiles derived from each of the seven loci.

Determination of Homologous Groups

OrthoMCL (version 2.0.9) (Li et al., 2003) was used to determine homologous clusters of the pan-genome for subsequent analyses. All extracted protein sequences were adjusted to a prescribed format and were grouped into homologous clusters based on sequence similarity. The BLAST reciprocal best hit algorithm (Moreno-Hagelsieb and Latimer, 2008) was employed with 70%

match cutoff, and Markov Cluster Algorithms (MCL) (Enright et al., 2002) were applied with an inflation index of 1.5.

Phylogenetic Analysis

To determine the phylogenetic relationships among *Salmonella* subspecies and different serovars, a phylogenetic tree was inferred from single nucleotide polymorphisms (SNPs) of core genes. According to the identification of homologous clusters, a total of 127,544 genes presented in single-copy orthologous genes shared by all 214 *Salmonella* strains were selected for phylogenetic analysis. For each orthologous cluster, nucleotide sequences were aligned and then concatenated. Custom Perl scripts were used to parse the concatenated sequence and extract SNPs. The organismal phylogeny of SNPs and other genes were inferred using the maximum likelihood (ML) algorithm in RAxML version 8.2.9 (Stamatakis, 2014). The tree reconstruction is repeated 1,000 times, and the percentage of times each interior branch which is noted as bootstrap-value is given. Each tree of genes in the two flagellin gene clusters was constructed by MEGA (Kumar et al., 2016), using Neighbor-Joining method with 1,000 bootstraps.

Identification of *fliC* and *fljB*

To characterize the evolution pattern of flagellin genes, reference sequences of *fliC*, *fljB*, and the flanked genes, *fliA*, *fliB*, *fliD*, *fliS*, *fljA*, and *hin*, were extracted from the *S. Typhimurium* LT2 genome (GenBank accession number AE006468) (McClelland et al., 2001). The relevant homologous clusters in the pan-genome were determined by BLAST with 70% match cutoff and 50% coverage cutoff. In cases of shared homologous genes related to flagellin genes, phylogenetic trees were inferred by ML algorithms in RAxML version 8.2.9 (Stamatakis, 2014). Together with the phylogenetic trees, the flanked gene contents were taken into account to confirm the gene loci of *fliA*, *fliB*, *fliC*, *fliD*, *fliS*, *fljA*, *fljB*, and *hin*.

Congruence Analysis

PAUP version 4.0a152 (Swofford, 1998) was used to conduct maximum-likelihood (ML) analysis of the congruence of gene trees, as described by Feil et al. (2001). The HKY85 model of DNA substitutions was used to estimate α parameter, and the transition/transversion (Ti/Tv) ratio, assuming gamma distribution. The difference between the log likelihood scores of the trees of two genes Δ -lnL was calculated, generating from the comparison of one gene tree against another based on the 99th percentile of the distribution of log likelihood scores for 200 trees from random topology. If the Δ -lnL values for the comparisons is less than that reference gene comparisons within the 99th percentile of this distribution, then these two gene trees should be considered significantly congruent (Octavia and Lan, 2006).

Regression Analysis for Evolutionary Synchronization

A key element of our study was to determine whether the evolution of flagella gene was associated with the evolutionary rate estimate. To test for this pattern we conducted a least squares linear regression for the genome-wide rate as a function

of sampling distance. The MEGA version 7 (Kumar et al., 2016) was used for pairwise distance calculating. A potential shortcoming of this analysis is that the data do not necessarily represent independent samples because some correspond to closely related lineages. This can be addressed by using phylogenetic independent contrasts, or phylogenetic generalized least squares. For the regression of the rate as a function of sampling evolutionary consensus the degree of association was estimated when fitting a regression for a ratio as a function of its denominator.

Recombination Analysis

Potential recombination within *fliB*, *fliC*, and *fliD* genes was screened using seven methods (RDP, GENECONV, MaxChi, Bootscan, Chimera, SiScan, and 3Seq) implemented in the Recombination Detection Program version 4.88 (RDP4) (Martin et al., 2015). The breakpoints were also defined by RDP4. Similarity between the recombinants and their possible major and minor parents was estimated using BootScan, embedded in RDP4 (Stamatakis, 2014). The recombination event evaluated by RDP4 was considered significant if it satisfied at least two criteria when the P -value (p) < 0.05 and the RDP recombination consensus score (RDPRCS) was >0.4. When p < 0.05 and the RDPRCS was between 0.4 and 0.6, the recombination event was possible. An RDPRCS < 0.4 with p < 0.05 indicated the rejection of the recombination event (Martin et al., 2015).

RESULTS

Core Genome Tree Revealed Detailed Phylogenetic Relationships among Different Serovars

A total of 214 *Salmonella* genomes were analyzed in this study, consisting of 205 *S. enterica* subsp. *enterica* genomes and 9 genomes of other four *S. enterica* subspecies, which covered 87 serovars. More than one genome (maximum twelve) was included for some famous serovars (e.g. *S. Enteritidis* and *S. Typhi*), while only one genome was considered for most other serovars because no more genomes were available. More detailed information on these genomes was summarized in **Table S1**. After homologous gene clustering, the total 947,793 proteins were grouped into 14,583 homologous clusters, including common genes represented by 596 single-copy groups of 712 core homologous clusters. A total of 57,946 SNPs in the “core” regions were identified that were present and aligned with high confidence in all taxa. These SNPs were used to construct the phylogeny (**Figure 1**).

The role of the phylogenetic tree based on SNPs in genome variation and evolution has been documented previously (Filliol et al., 2006; Holt et al., 2008). The core genome tree can support the major genetic group phylogeny and reveal much more phylogenetic detail, including the divergence of major phylogenetic groups. Moreover, the core genome tree also revealed a number of strong concordances with serotyping. From **Figure 1**, core genome tree was able to categorize *S.*

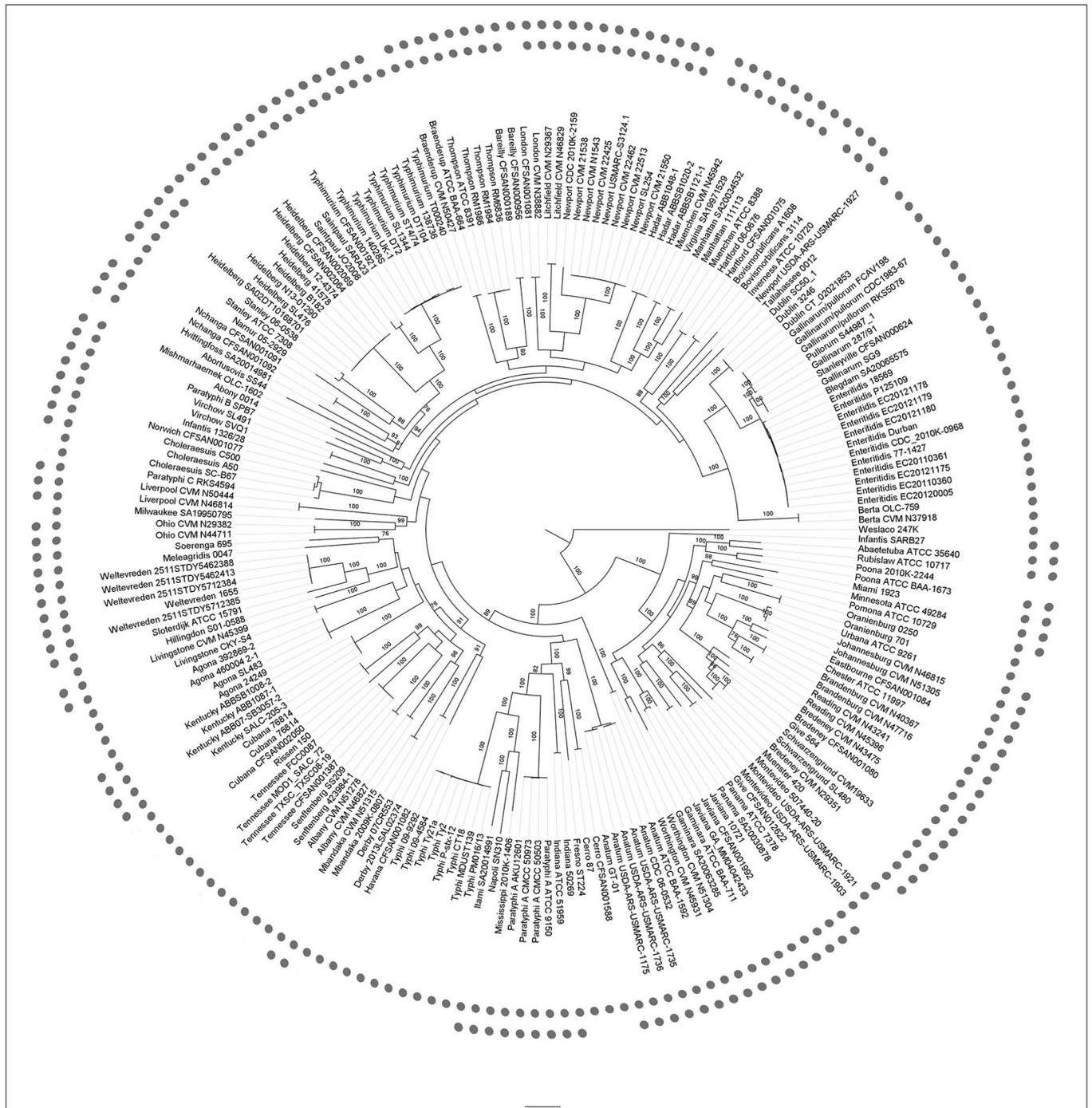


FIGURE 1 | Phylogenetic tree of *Salmonella enterica*. Maximum Likelihood phylogenetic tree based on 57,946 SNPs of 596 single copy core genes of 214 genomes. Bootstrap values (>70%) are presented on the branches. The inner dots indicated the presence of *fljC* gene clusters, while the outer dots indicated *fljB* gene clusters. The bootstrap values near the tips of the trees were omitted. The other four subspecies of *S. salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), and *houtenae* (IV) were far distant from subspecies *enterica* (I) and omitted.

Typhimurium isolates but unable to ultimately separate the *S. Stanleyville* from the *S. Gallinarum* with high resolution. In most cases, the genomes within a cluster belonged to the same serovar, as for *S. Typhi* or *Enteritidis*. Although some isolates

were mixed up with each other, for example, it was hard to differentiate *S. Give*, *S. Schwarzengrund*, and *S. Bredney*, the core genome tree generally separated the serovars from each other.

The Monophasic and Biphasic Strains Were Less Grouped into the Same Cluster on Core Genome Tree

The prevalence of phase I and phase II related flagellin gene locus of the 214 genomes was also indicated in **Figure 1**. In total, *fliC* gene cluster was found in 207 genomes (96.7%) of 91 serovars (97.8%), including subspecies, while *fljB* gene cluster was found in 138 genomes (64.5%) of 61 serovars (65.6%). The detailed list (**Table S2**) was dominated by *fliC* and *fljB* genes and included their flanked genes and *hin*. Some well-studied serovars (Enteritidis, Dublin, Typhi etc.) are monophasic the same as the previous study (Selander et al., 1992; Baker et al., 2007). Nevertheless, the *fliC* gene was not found present in the genomes of strain *S. Poona* 2010K-2244, *S. Heidelberg* B182, *S. Abortusovis* SS44, *S. enterica* subsp. *salamae* 3588/07, *S. Bovismorbificans* A1608, *S. Hadar* ABBSB1020-2, and *S. London* CFSAN001081, of which only a draft genome was obtained for these strains. Duplicated annotations and misprediction of some genes that span contig boundaries may occur in draft genomes. The homologs (**Table S2**) were able to mitigate but not eliminate these limitations. We further checked the contigs and related annotation of seven genomes of *fliC* gene cluster absent, and found some residue of the gene cluster at some contigs' ends in each genome. On the other hand, there was no report on the absence of *fliC* in these serovars, for example, *fliC* should be present in all *S. Heidelberg* (Macnab, 1992). Therefore, the absence of *fliC* gene in these genomes should be a result of misassembly or the related fragment not fully sequenced.

Interestingly, the monophasic and biphasic being linked to the evolution of *S. enterica* subsp. *enterica* showed discrete phylogenetic clusters on the core genome tree (**Figure 1**). A few serovars with identical phase, clustered together in the core genome tree. In one larger monophasic clade, with 25 genomes, containing *S. Dublin*, *Gallinarum*, *Pullorum*, *Stanleyville*, *Blegdam*, *Enteritidis*, and *Berta*, all the genomes of each serovar were located together in one distinct cluster in the

core genome tree. For the other monophasic serovars, the 45 genomes were identified at different locations, of which another larger clade of 12 genomes, containing *S. enterica* subsp. *enterica* serovars *Cubana*, *Rissen*, *Tennessee*, *Senftenberg*, and *Albany* were located adjacent to *Derby*, *Havana*, and *Agona*, but were disrupted by *Mbandaka* and *Kentucky* of biphasic.

Evolutionary Congruence and Phylogenetic Analysis Implied Two Distinct Evolution Patterns of the Flagellin Gene Clusters

Evolutionary analysis of flagellin gene *fliC* and *fljB* in *Salmonella* has been performed previously (Mortimer et al., 2007). However, the co-evolution of these two genes and their flanking genes has never been studied. To evaluate evolutionary congruence between and within the *fliC* and *fljB* clusters, ML analysis was carried out, and the results were summarized in **Table 1**. As is shown, none of the genes in *fliC* gene cluster were congruent to all the genes in *fljB* gene cluster, and none of the genes in *fljB* gene cluster were congruent to all the genes in *fliC* gene cluster. The gene trees with the largest number of congruencies were those of *fliC* and *hin*, which were congruent to four other gene trees. More concretely, *fliC* gene tree was similar to those of *fliB*, *fliD*, *fljA*, and *fljB* gene, and *hin* gene tree was similar to those of *fliB*, *fliD*, *fljA*, and *fljB* gene. As for *fljB* (equaling to *fljA* and *fljB* gene trees) and *fljB* (equaling to *fliB* and *fljA* trees), they were congruent to two gene trees. And *fliA* and *fliD* trees were congruent to *fljA* and *fliB* respectively. Overall, only 29.2% (14/48) of the flagellar gene tree comparisons were congruent among these *S. enterica* strains. As for the *fliC* and its neighbored genes, *fliB* and *fliD*, they were closely related, while for *fljA* and *fljB* genes, their relationship was more distant. Moreover, there was generally higher level of congruence between *fliC* and *fljB/fliD* genes than between *fliA* and *fliB* genes and between *fliD* and *fliS* genes. However, correlation calculations were not available between *fliC* and *fliS* genes, because approximation limit dynamically readjusted too low.

TABLE 1 | ML analysis for congruence between each gene tree of the flagella genes analyzed in this study.

	α^a	Ti/Tv ratio ^b	-lnL	Δ -lnL (99th) ^c	<i>fliA</i>	<i>fliB</i>	<i>fliC</i>	<i>fliD</i>	<i>fliS</i>	<i>fljA</i>	<i>fljB</i>	<i>hin</i>
	Δ -lnL (99 percentile)											
<i>fliA</i>	0.02	7.59	1,442.96	1,333.38	–	1,446.74	1,855.03	1,588.38	1,336.41	1,332.98	1,559.4	1,392.09
<i>fliB</i>	0.2	3.37	7,850.6	24,309.07	28,488.81	–	26,254.26	24,838.09	29,386.81	24,142.47	24,295.95	27,444.31
<i>fliC</i>	0.52	1.29	25,042.12	127,229.25	152,120.99	124,348.02	–	125,530.74	–	125,113.23	124,009.55	145,249.22
<i>fliD</i>	0.29	3.2	7,535.46	35,561.8	43,616.44	35,031.96	40,180.51	–	44,377.23	36,789.97	36,477.65	43,028.63
<i>fliS</i>	0.01	8.72	877.71	802.22	815.62	850.98	1,059.74	902.58	–	809.76	900.25	842.57
<i>fljA</i>	0.15	3.2	1,649.33	2,071.13	2,224.6	2,100.08	2,443.48	2,172.55	2,236.68	–	2,143.27	2,236.19
<i>fljB</i>	0.25	2.27	8,965.19	24,268.16	28,402.47	23,710.94	25,860.96	24,279.29	28,701.65	24,179.59	–	27,639.31
<i>hin</i>	0.02	2.67	1,674.34	2,095.84	2,101.85	1,832.47	2,264.01	1,947.05	2,097.23	1,795.33	1,955.46	–

^aNucleotide substitution rate variation between sites, with gamma distribution as the parameter. ^bEstimated transition/transversion ratio. ^cDifference in -lnL score from -lnL column (reference data) and the 99th percentile from random topology. ^dDifference in -lnL score from reference data to each calculated score from other genes. Trees are deemed to be congruent if Δ -lnL is equal to or lower than the 99th percentile of random trees, when compared with reference data (Feil et al., 2001; Octavia and Lan, 2006). Bold character in the table symbolizes congruent tree between these two genes.

Besides the relationship among the genes in the *fliC* and *fljB* gene clusters, we further studied their congruence to the core genome using linear regression analysis instead of ML analysis, because datasets of the core genome were too large to calculate -lnL score. The mutual information analysis indicated a weak correlation between each flagellin gene and core genome. The low values of both trees for each linear regression indicate a high level of incongruence between individual tree of each gene and the core genome tree, as shown in **Figures 2, 3**. The topology of *fljB*, *fliC*, and *fliD* gene trees was much distinct from that of the core genome tree, while those of the *fliA* and *fliS* genes were more similar to the core genome tree than those of *fljB*, *fliC*, and *fliD* genes. However, the *fljA* tree for serovars displayed a similar topology to that of the core genome tree, while *fljB* did not.

To infer evolutionary relationships among different serovars, based upon similarities and differences in their genetic characteristics, we constructed phylogenetic trees of *fliA*, *fliB*, *fliC*, *fliD*, *fliS*, and *fljA*, *fljB*, *hin*, and *fliC/fljB* (**Figure 4**). Diagrammatic representation of *fliB*, *fliC*, and *fliD* indicated bifurcation to form two distinct lineages, where other *Salmonella enteric* subspecies joined together with subspecies enteric serovars. The other four subspecies could be easily recognized from *fliA* and *fliS* gene tree because of their distant branches, which was more consistent with core genome tree. Subspecies *salamae* was close to three neighbors, serovar Weltevreden, Sloterdijk and Kentucky, in *fljB* tree, while these serovars were scattered to different branches in *fljA* tree. The results of phylogenetic trees were consistent to the congruence analysis (**Table 1, Figure 2, 3**). Dramatically, the individual gene trees of *fliB*, *fliC* (*fliC* of *S. Typhi* was an exception) and *fliD* were consistent with a clustering dividing monophasic and biphasic serovars. This seemed to imply two distinct evolutionary paths of *fliC/fljB* corresponding to monophasic and biphasic serovars.

Phylogeny of *fliC* and *fljB* Showed Complex Relationships

Because different genes of the flagellin diverge under different evolutionary pressures, we manually curated the phylogenetic pattern of genes in flagellin gene clusters. For all flagellin and their flanked genes, the individual gene trees (**Figure 4**) can be used to demonstrate the evolutionary story on individual gene basis. To study the genetic evolution of the 207 *fliC* and 138 *fljB* allelic variants, phylogenetic tree based on nucleotide sequences was constructed (**Figure 4I**). There is a clear separation of monophasic *Salmonella* groups from the biphasic groups based on the *fliC* gene, while all the *fljB* alleles clustered with the biphasic *fliC* alleles. This may either reflect that monophasic were acting as incipient serotypes with a barrier to recombination with biphasic serotypes, or indicate that there were insufficient incidents in which recombination could have taken place between monophasic and biphasic serotypes. However, *S. Typhi* was not clustered into other monophasic genomes as one branch together. Similarly, previous study had also identified that monophasic *S. Typhimurium* existed, of which the serotype was always biphasic (Laorden et al., 2010).

Furthermore, all biphasic *fliC* and *fljB* alleles were analyzed based on the amino acid sequences (**Figure S1**). The genetic distance between *fliC* gene and *fljB* gene sequences showed no relationship within/among serovars, which was reflected in the phylogenetic tree. For example, serovar Schwarzengrund, Liverpool, and Livingstone were closely related based on *fliC* gene while they were distant based on *fljB*. *S. Heidelberg* and *S. Newport* were clustered together according to the *fljB* gene, while they were clustered separately according to the *fliC* gene. It suggested that the evolutionary histories were independent between *fliC* and *fljB* genes and frequent homologous recombination referring *fliC* and *fljB* occurred between biphasic serovars. The phylogenetic tree of 5' and 3' conserved segment of *fliC* and *fljB* genes was also constructed (data not shown), showing the similar division.

Less Recombination Events Were Found between Monophasic and Biphasic Serovars

Recombination is a significant source of genetic diversity in *Salmonella* (Silverman et al., 1979). This phenomenon, which occurs during co-infection with different strains, is usually observed at the *fliC* and its flanked genes. It is apparent that there were different levels of recombination within flagellar genes of *S. enterica*. Comparison of recombination of monophasic genes with biphasic genes, which represents the unique ability to change flagellar composition, by both ML (**Table 1**) and recombination detective analysis (**Figure 5**), showed that recombination occurred far more frequently within monophasic or biphasic gene clusters than between them. For example, the serovar Heidelberg, whose major parent (a sequence closely related to that from which the greater part of the recombinant's sequence may have been derived, Martin et al., 2015) was Virchow, and minor parent (a sequence closely related to that from which sequences in the proposed recombinant region may have been derived, Martin et al., 2015) was Newport, did not recombine to any monophasic genes. It was identical to Hartford, whose major parent was Pomona, and minor parent was Anatum. Even though it was found that Senftenberg (monophase) was minor parent for Newport (biphase), and other similar recombinant events of referring small segments were also found, recombination was believed mainly to have occurred within the same phase. Moreover, there were lots of breakpoints and the breakpoints tended to cluster far more around the edges of genes than they did within the central parts of genes (**Figure 5**).

DISCUSSION

With the cost of whole genome sequencing (WGS) decreasing dramatically (Didelot et al., 2012) and the technology becoming available for routine use, WGS is very useful and practical in various bacterial infectious studies including diagnostic and public health microbiology (Koser et al., 2012). Many *Salmonella* serovars are human-specific, which may cause gastroenteritis, a self-limiting illness. A few *Salmonella* serovars, especially *S. Typhi*, may elicit typhoid-like diseases and serious systemic

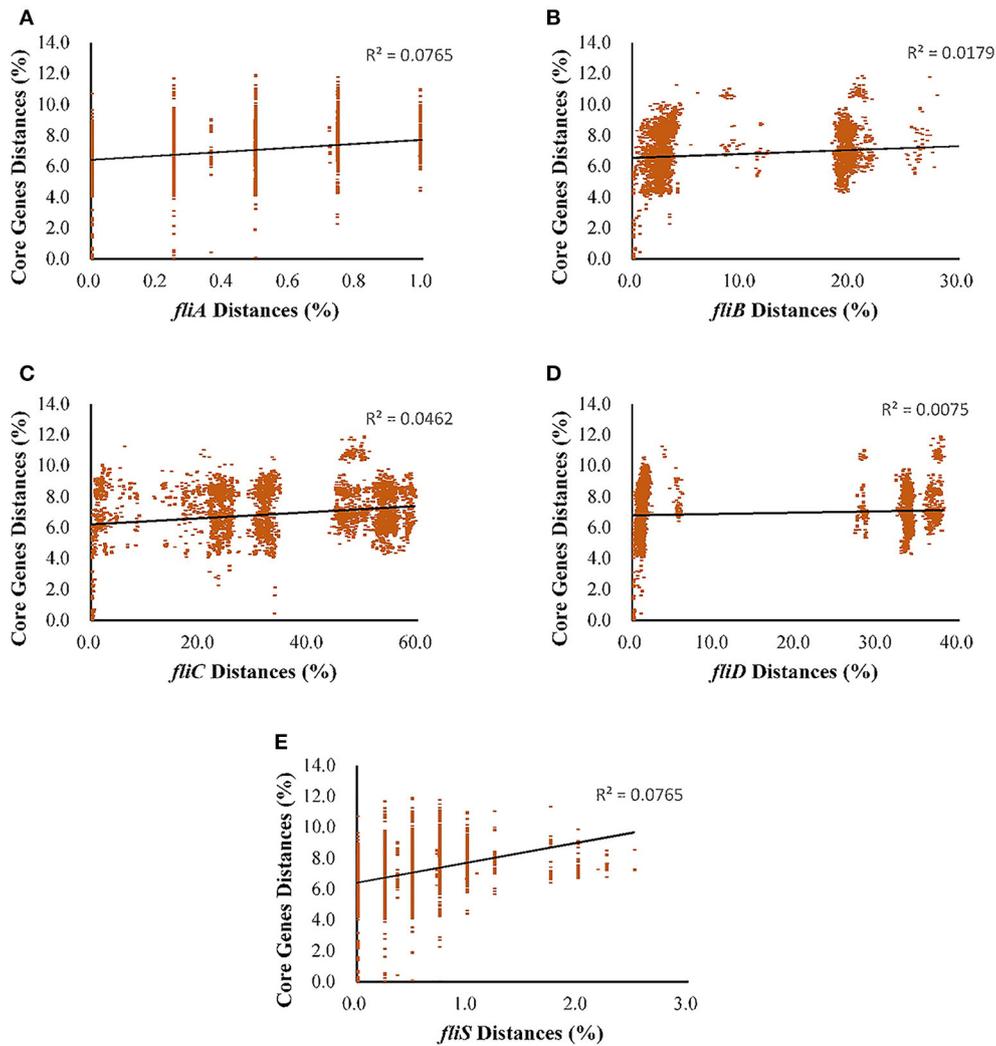


FIGURE 2 | Regressions between each gene of the *fljC* gene clusters and core genetic distance (expected nucleotide substitutions per site) as a function of sampling the evolutionary distance for *S. enterica* (I). **(A)**, *fljA*; **(B)**, *fljB*; **(C)**, *fljC*; **(D)**, *fljD*; and **(E)**, *fljS*. Each dot corresponds to a pair of sampled genomes, and the linear regression is shown as a solid line using least squares; the R^2 coefficients are also shown.

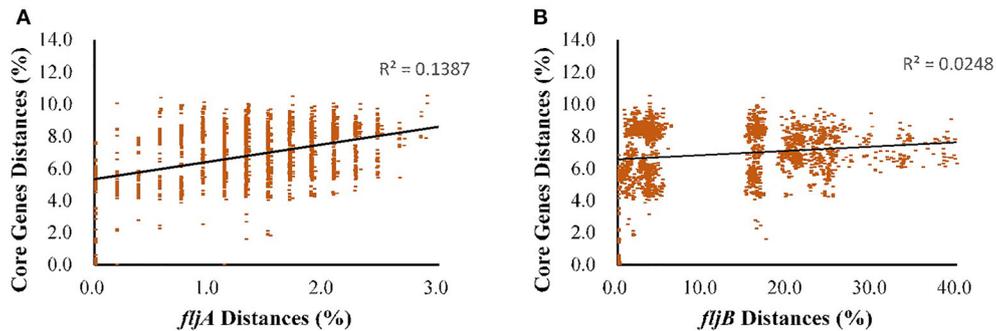
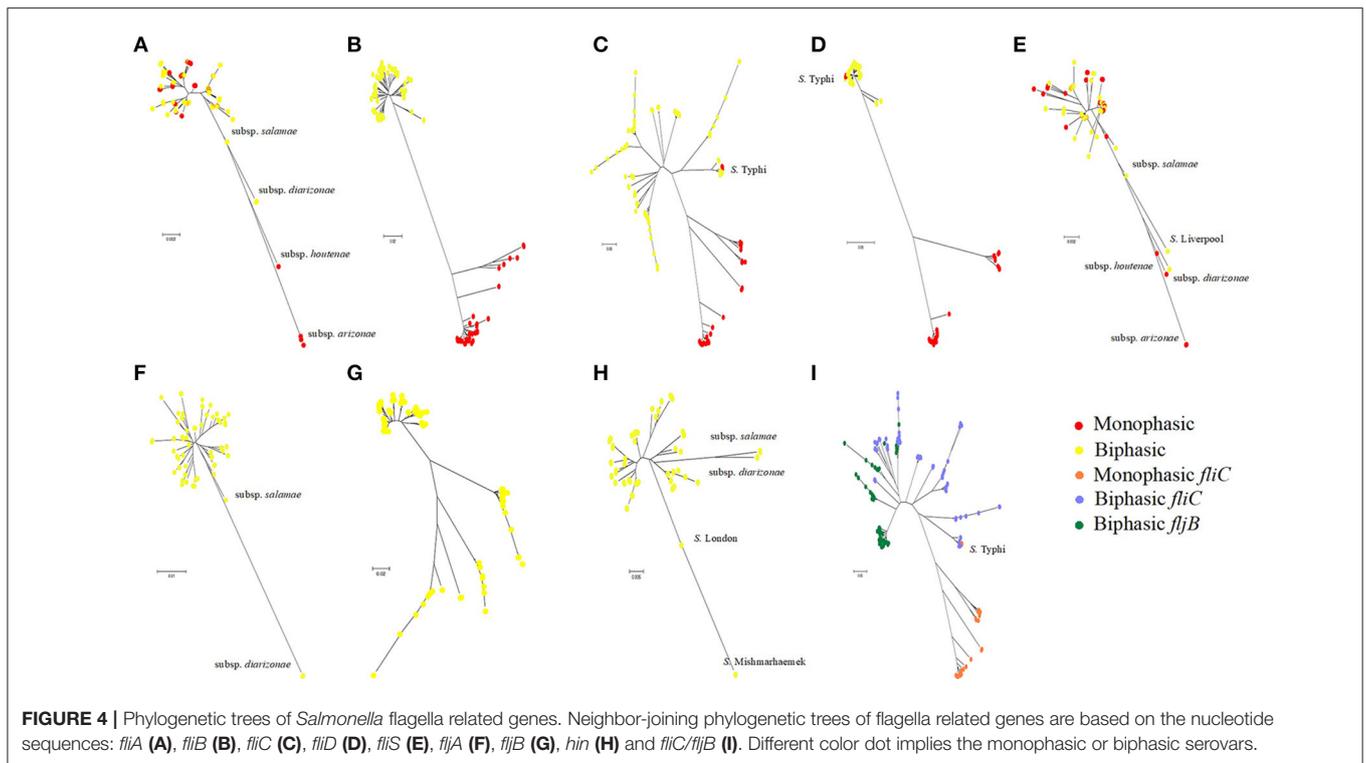


FIGURE 3 | Regressions between each gene of the *fljB* gene clusters and core genetic distance (expected nucleotide substitutions per site) as a function of sampling the evolutionary distance for *S. enterica* (I). **(A)**, *fljA*; and **(B)**, *fljB*. Each dot corresponds to a pair of sampled genomes, and the linear regression is shown as a solid line using least squares; the R^2 coefficients are also shown.

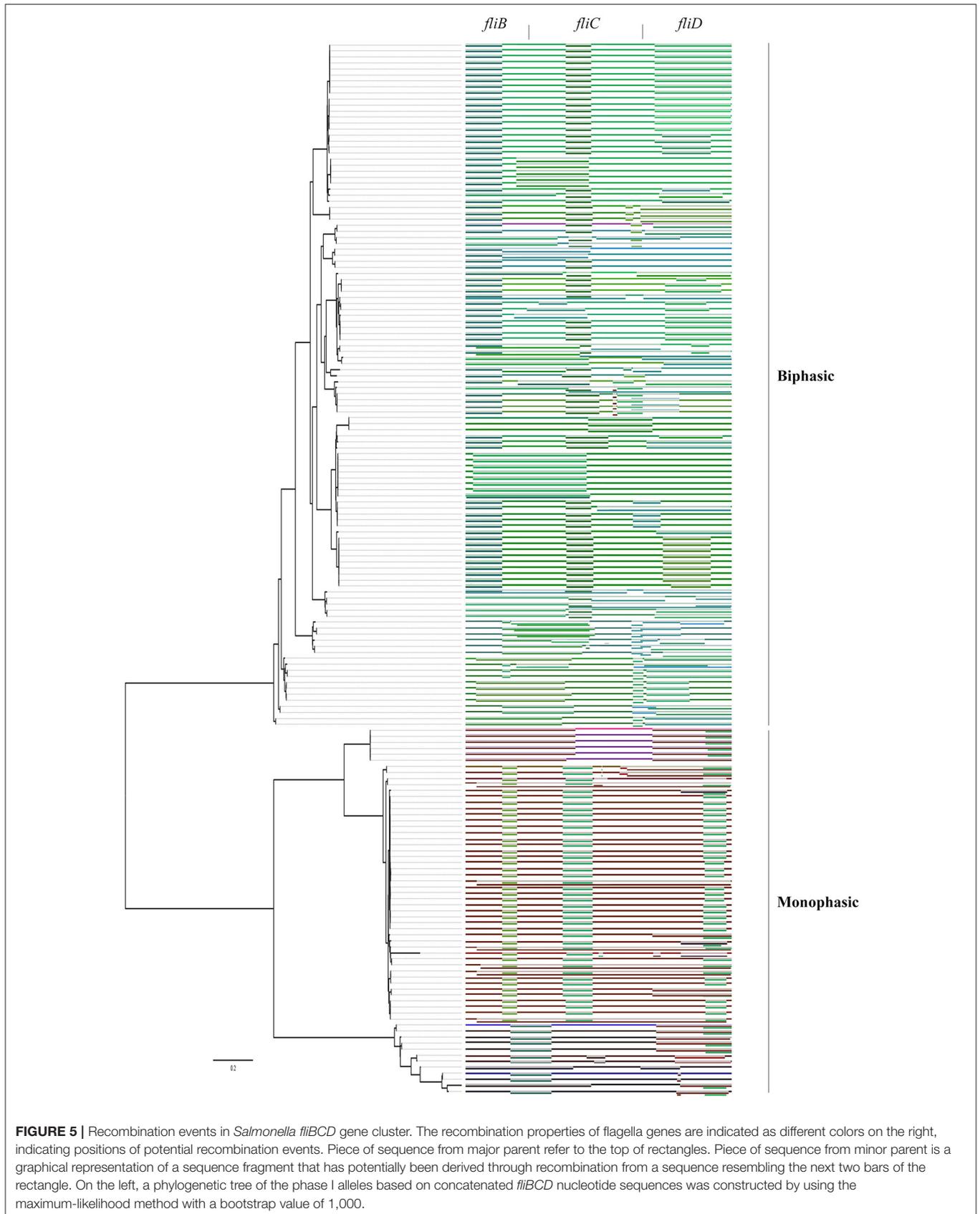


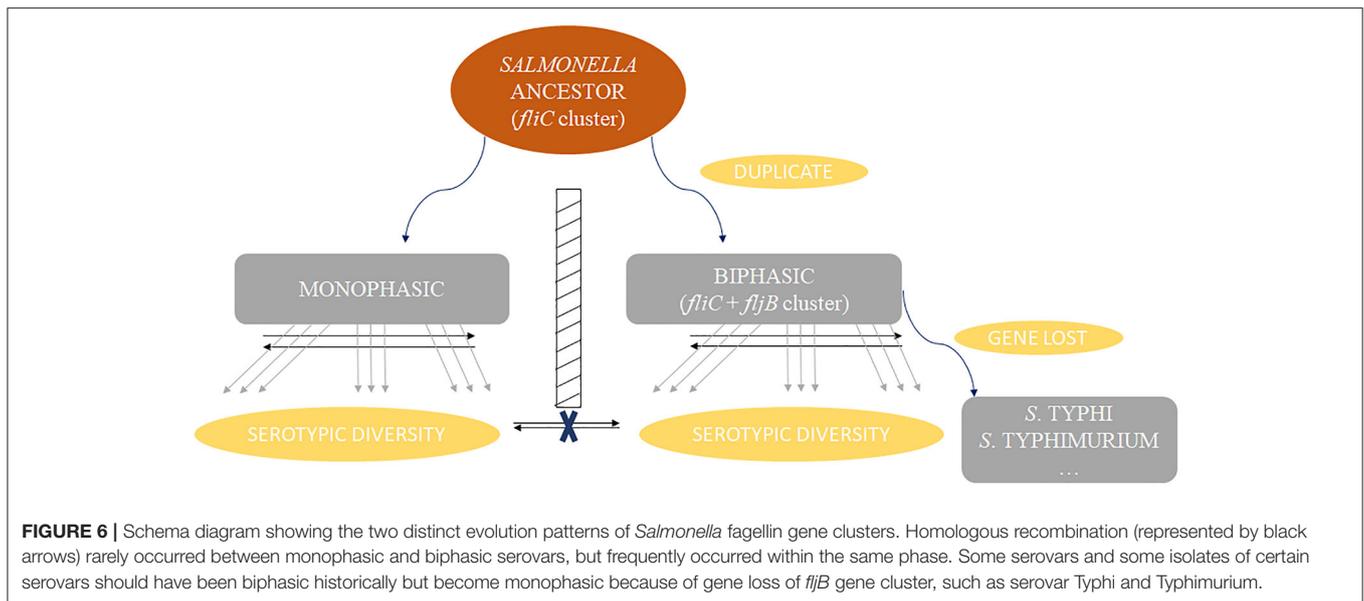
infections (Parkhill et al., 2001; Crump and Mintz, 2010). Recent studies support the phylogenetic tree based on SNPs as an outbreak surveillance tool (Allard et al., 2012; Hoffmann et al., 2013; Zhou et al., 2013). The performance of these 214 genomes based on SNPs was validated on the phylogenetic analysis. The fractions of the core genome tree (Figure 1) supported the partition of the taxa in previous study and reflected robust evolutionary relationship of *Salmonella* serovars (Chan et al., 2003; Jacobsen et al., 2011; Zou et al., 2014). Moreover, the rich information of *Salmonella* subspecies *enterica*, covering 87 serovars, provided a framework for determination of the evolutionary relationships of the flagellin alleles to assess the co-evolution and recombination events.

Flagellin genes are considered as a dominant antigen of the adaptive immune response (Haiko and Westerlund-Wikstrom, 2013). The *fliC* gene is present in most of *Salmonella* isolates, and has a homologous locus in *Escherichia coli* (Macnab, 1992). Selection within specific niches might have an impact in antigenic variation (McQuiston et al., 2008). For adapting to environmental changes, *Salmonella* adjust their expression, introducing a second set of antigens, to survive in the new niches. The complex branching division of diphasic serotypes, compared to monophasic serotypes, is suggestive of an expanded mutation rate or recombination in the flagellin when *Salmonella* acquired the second flagellin gene. Perhaps monophasic serotypes obtain enough antigenic diversity from SNPs, letting these serotypes to be successful as monophasic strains (Mortimer et al., 2007). However, the reason why *S. Typhi* is a monophasic serotype and cluster with biphasic serotypes in the phylogenetic trees of *fliBCD* is still unclear. We hypothesize that *S. Typhi* had a trend to be

diphasic, and the substitute of *fljB* was present in the history but got lost during speciation.

As an important virulence factor in pathogenic bacteria, especially in *Salmonella* genus, flagellin might regulate their virulence and pathogenesis (Stecher et al., 2004). For example, it is beneficial to immunologic escape by turning on/off the expression of genes related to these structures, niche adaptation by expressing the second flagellin and avoiding predation by different motility or antigenic expression. All of these selective advantages are importantly involved in evolutionary step of having the second flagellin. The gain or loss of their unique ability to encode the major flagellin protein by switching expression can be discussed by an evolutionary model. Based on the phylogenetic tree of the *fliC* and *fljB* genes in the broad serovars involved, it is supposed that the phylogeny of flagellin genes illustrated a strong serovar-specific lineage, with strains of the same serovar clustering either together or in a few distinct branches (Figure 1), as described before (Yue et al., 2015). Clustering of the *fliC* and *fljB* genes showed that *fljB* arose from *fliC* gene duplication (Mortimer et al., 2007), which have highly conserved 5' and 3' ends and highly variable center regions (Masten and Joys, 1993; McQuiston et al., 2004). Most studies of flagellin genes focus either on its evolutionary relationships or as a dominant antigen of the sequence diversity (Mortimer et al., 2007) but do not undertake subsequent congruence analysis of their evolutionary process. The systematic approach, ML analysis and regression analysis (Table 1, Figures 2, 3), determined that *fljB*, *fliC*, and *fliD* gene trees were evolutionarily congruent to each other but not to the core genome. In the regression analysis of *fliA* and *fliS* against the core genome, low levels of correlation were found (low R^2





values). The *fliA* and *fliS* genes were highly conserved within subspecies *enterica* with limited variation (Figure 2), which will disturb the regression analysis, while their phylogenetic trees indicated a clear classification as that of core genome, especially other subspecies (Figure 4). These results suggested that *fliA* and *fliS* evolved congruently with the core genome. Notably, homologous recombination of *fliC* and flanked genes occurred frequently within monophasic and diphasic serovars (Figure 5). These results suggested that the level of clonality of flagellar genes within subspecies I was low.

Dramatically, the phylogenetic trees of *fliB*, *fliC*, and *fliD* revealed a similar division of the serovars into two distinct groups, corresponding to monophasic and biphasic serovars, respectively. This implied two distinct evolutionary patterns of *Salmonella fliC/fljB* gene cluster. This division was not pointed out previously, but similar phylogenetic structures of *fliC/fljB* were reported (Smith et al., 1990; Mortimer et al., 2007). It was divided into Non-G antigens and G antigens based on *fliC_{g,m,s}* encoding three factors. Through the phylogenetic tree in Mortimer (Mortimer et al., 2007) study, Non-G group and G group were clustered into different branches, which equaled to biphasic and monophasic groups respectively. According to these results, a flow diagram was constructed to demonstrate the evolutionary path of the monophasic and biphasic gene clusters (Figure 6). The *Salmonella* ancestor should just harbor a *fliC* gene cluster as most of *E. coli* isolates (McQuiston et al., 2008). Subsequently, some lineages suffered *fliC* duplication to assemble an *fljB* gene cluster under certain natural selection conditions (biphasic), while the others not (monophasic). Both monophasic and biphasic lineages evolved and adapted to different environments and occupied diverse ecological niches, resulting in H antigen (serovar) diversification. Ancestors of some biphasic serovars might have lost *fljB* gene cluster under natural selection because it was useless, becoming monophasic again (e.g., *S. Typhi* and some *S. Typhimurium* isolates). Frequent

homologous recombination occurred within monophasic and biphasic serovars. Finally, the current diversity of *Salmonella* was formed. This theory assumed that *fljB* gene cluster should perform a role together with *fliC* cluster in most biphasic serovars, but present knowledge suggested that just one of them was expressed in a cell (Imre et al., 2005). Therefore, the role of *fljB* needs to be further determined to clarify the implication of the two distinct evolutionary patterns.

Collectively, we provide the viewpoint that flagellin genes of *S. enterica* subsp. *enterica* represents two evolutionary patterns with a barrier between monophasic and biphasic serovars. The evolutionary barrier may inhibit an overlapping recombination. However, it is still unclear what the implication of the barriers on the immunological escape, niche adaptation, and avoiding predation is. That is an interesting area for further study. Besides, different serovar genome repertoires were included, suggesting that the core genome evolution of *S. enterica* was accompanied by *fliA/fliS* but distinct from those of *fliBCD* and *fljB* gene cluster. This study will provide a robust genetic framework for future studies on the evolution of the phase I and phase II flagellar characteristics in *S. enterica*.

AUTHOR CONTRIBUTIONS

YL and DFZ designed and initiated the study, interpreted the results, and wrote the manuscript; YL downloaded and analyzed the data; LZ, XZ, LX, and XS contributed to discussion of the results and improvement of the manuscript and designed the outline of this study and manuscript and provided laboratory equipment and space.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China [No. 31230058] and the National Key R&D

program of China [No. 2016YFE0106100]. The authors would like to thank Dr. Harold Corke, Dr. Hongyu Ou, Dr. Xiangyu Deng, and Dr. Arun Bhunia for their useful suggestions on this manuscript. In addition, we want to thank Mr. Ben Jia and Mr. Lincheng Shi who kindly provided help in custom script coding for this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02604/full#supplementary-material>

REFERENCES

- Achtman, M., Wain, J., Weill, F. X., Nair, S., Zhou, Z., Sangal, V., et al. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 8:e1002776. doi: 10.1371/journal.ppat.1002776
- Allard, M. W., Luo, Y., Strain, E., Li, C., Keys, C. E., Son, I., et al. (2012). High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. *BMC Genomics* 13:32. doi: 10.1186/1471-2164-13-32
- Baker, S., Hardy, J., Sanderson, K. E., Quail, M., Goodhead, I., Kingsley, R. A., et al. (2007). A novel linear plasmid mediates flagellar variation in *Salmonella* Typhi. *PLoS Pathog.* 3:e59. doi: 10.1371/journal.ppat.0030059
- Bonifield, H. R., and Hughes, K. T. (2003). Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *J. Bacteriol.* 185, 3567–3574. doi: 10.1128/JB.185.12.3567-3574.2003
- Brenner, F. W., Villar, R. G., Angulo, F. J., Tauxe, R., and Swaminathan, B. (2000). *Salmonella* nomenclature. *J. Clin. Microbiol.* 38, 2465–2467.
- Centers for Disease Control and Prevention (CDC) (2011). *National Salmonella Surveillance Annual Summary 2009*. Atlanta, GA: US Department of Health and Human Services, CDC.
- Chan, K., Baker, S., Kim, C. C., Detweiler, C. S., Dougan, G., and Falkow, S. (2003). Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar typhimurium DNA microarray. *J. Bacteriol.* 185, 553–563. doi: 10.1128/JB.185.2.553-563.2003
- Crump, J. A., and Mintz, E. D. (2010). Global trends in typhoid and paratyphoid Fever. *Clin. Infect. Dis.* 50, 241–246. doi: 10.1086/649541
- Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 13, 601–612. doi: 10.1038/nrg3226
- Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575
- Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P. J., Enright, M. C., et al. (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. U.S.A.* 98, 182–187. doi: 10.1073/pnas.98.1.182
- Filioli, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., Bobadilla del Valle, M., et al. (2006). Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* 188, 759–772. doi: 10.1128/JB.188.2.759-772.2006
- Grimont, P. A. D., and Weill, F. X. (2007). *Antigenic Formulae of the Salmonella serovars, 9th Edn*. Paris: WHO Collaborating Centre for Reference and Research on Salmonella.
- Haiko, J., and Westerlund-Wikström, B. (2013). The role of the bacterial flagellum in adhesion and virulence. *Biology* 2, 1242–1267. doi: 10.3390/biology2041242
- Hoffmann, M., Luo, Y., Lafon, P. C., Timme, R., Allard, M. W., McDermott, P. F., et al. (2013). Genome sequences of *Salmonella enterica* serovar Heidelberg isolates isolated in the United States from a multistate outbreak of human *Salmonella* infections. *Genome Announc.* 1:e00004-12. doi: 10.1128/genomeA.00004-12
- Holt, K. E., Parkhill, J., Mazzoni, C. J., Roumagnac, P., Weill, F. X., Goodhead, I., et al. (2008). High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat. Genet.* 40, 987–993. doi: 10.1038/ng.195
- Imre, A., Olasz, F., and Nagy, B. (2005). Development of a PCR system for the characterisation of *Salmonella* flagellin genes. *Acta Vet. Hung.* 53, 163–172. doi: 10.1556/AVet.53.2005.2.2
- Jacobsen, A., Hendriksen, R. S., Aarestrup, F. M., Ussery, D. W., and Friis, C. (2011). The *Salmonella enterica* pan-genome. *Microb. Ecol.* 62, 487–504. doi: 10.1007/s00248-011-9880-1
- Köser, C. U., Ellington, M. J., Cartwright, E. J., Gillespie, S. H., Brown, N. M., Farrington, M., et al. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* 8:e1002824. doi: 10.1371/journal.ppat.1002824
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Laorden, L., Herrera-Leon, S., Martinez, I., Sanchez, A., Kromidas, L., Bikandi, J., et al. (2010). Genetic evolution of the Spanish multidrug-resistant *Salmonella enterica* 4,5,12:i:- monophasic variant. *J. Clin. Microbiol.* 48, 4563–4566. doi: 10.1128/JCM.00337-10
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for Eukaryotic genomes. *Genome Res.* 13:2178. doi: 10.1101/gr.1224503
- Macnab, R. M. (1992). Genetics and biogenesis of bacterial flagella. *Annu. Rev. Genet.* 26, 131–158. doi: 10.1146/annurev.ge.26.1.120192.001023
- Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., et al. (2010). The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin. Infect. Dis.* 50, 882–889. doi: 10.1086/650733
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003
- Masten, B. J., and Joys, T. M. (1993). Molecular analyses of the *Salmonella* g... flagellar antigen complex. *J. Bacteriol.* 175, 5359–5365. doi: 10.1128/jb.175.17.5359-5365.1993
- McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., et al. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413, 852–856. doi: 10.1038/35101614
- McQuiston, J. R., Fields, P. I., Tauxe, R. V., and Logsdon, J. M. Jr. (2008). Do *Salmonella* carry spare tyres? *Trends Microbiol.* 16, 142–148. doi: 10.1016/j.tim.2008.01.009
- McQuiston, J. R., Parrenas, R., Ortiz-Rivera, M., Gheesling, L., Brenner, F., and Fields, P. I. (2004). Sequencing and comparative analysis of flagellin genes *fliC*, *fljB*, and *flpA* from *Salmonella*. *J. Clin. Microbiol.* 42, 1923–1932. doi: 10.1128/JCM.42.5.1923-1932.2004
- McQuiston, J. R., Waters, R. J., Dinsmore, B. A., Mikoleit, M. L., and Fields, P. I. (2011). Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. *J. Clin. Microbiol.* 49, 565–573. doi: 10.1128/JCM.01323-10

- Moreno-Hagelsieb, G., and Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24, 319–324. doi: 10.1093/bioinformatics/btm585
- Mortimer, C. K., Gharbia, S. E., Logan, J. M., Peters, T. M., and Arnold, C. (2007). Flagellin gene sequence evolution in *Salmonella*. *Infect. Genet. Evol.* 7, 411–415. doi: 10.1016/j.meegid.2006.12.001
- Nataro, J. P., Bopp, C. A., Fields, P. I., Kaper, J. B., Strockbine, N. A., Murray, P. R., et al. (2011). “*Escherichia*, *Shigella*, and *Salmonella*,” in *Manual of Clinical Microbiology, 10th Edn*, eds P. R. Murray, E. J. Tenover, J. H. Tenover, M. A. Tenover, and R. H. Tenover (Washington, DC: ASM Press), 603–626.
- Okazaki, N., Matsuo, S., Saito, A., and Tominaga, M., Enomoto M. (1993). Conversion of the *Salmonella* phase 1 flagellin gene *fliC* to the phase 2 gene *fljB* on the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* 175, 758–766. doi: 10.1128/jb.175.3.758-766.1993
- Octavia, S., and Lan, R. T. (2006). Frequent recombination and low level of clonality within *Salmonella enterica* subspecies I. *Microbiology Sgm* 152, 1099–1108. doi: 10.1099/mic.0.28486-0
- Ogushi, K., Wada, A., Niidome, T., Mori, N., Oishi, K., Nagatake, T., et al. (2001). *Salmonella enteritidis* FliC (flagella filament protein) induces human beta-defensin-2 mRNA production by Caco-2 cells. *J. Biol. Chem.* 276, 30521–30526. doi: 10.1074/jbc.M011618200
- Ohnishi, K., Kutsukake, K., Suzuki, H., and Iino, T. (1990). Gene *fliA* encodes an alternative sigma factor specific for flagellar operons in *Salmonella typhimurium*. *Mol. Gen. Genet.* 221, 139–147. doi: 10.1007/BF00261713
- Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D., Wain, J., et al. (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413, 848–852. doi: 10.1038/35101607
- Schrader, K. N., Fernandez-Castro, A., Cheung, W. K., Crandall, C. M., and Abbott, S. L. (2008). Evaluation of commercial antisera for *Salmonella* serotyping. *J. Clin. Microbiol.* 46, 685–688. doi: 10.1128/JCM.01808-07
- Selander, R. K., Smith, N. H., Li, J., Beltran, P., Ferris, K. E., Kopecko, D. J., et al. (1992). Molecular evolutionary genetics of the cattle-adapted serovar *Salmonella dublin*. *J. Bacteriol.* 174, 3587–3592. doi: 10.1128/jb.174.11.3587-3592.1992
- Silverman, M., Zieg, J., Hilmen, M., and Simon, M. (1979). Phase variation in *Salmonella*: genetic analysis of a recombinational switch. *Proc. Natl. Acad. Sci. U.S.A.* 76, 391–395. doi: 10.1073/pnas.76.1.391
- Smith, N. H., Beltran, P., and Selander, R. K. (1990). Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. *J. Bacteriol.* 172, 2209–2216. doi: 10.1128/jb.172.5.2209-2216.1990
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stecher, B., Hapfelmeier, S., Müller, C., Kremer, M., Stallmach, T., and Hardt, W. D. (2004). Flagella and chemotaxis are required for efficient induction of *Salmonella enterica* serovar Typhimurium colitis in streptomycin-pretreated mice. *Infect. Immun.* 72, 4138–4150. doi: 10.1128/IAI.72.7.4138-4150.2004
- Swofford, D. L. (1998). *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Sunderland, MA: Sinauer Associates.
- Szekely, E., and Simon, M. (1983). DNA sequence adjacent to flagellar genes and evolution of flagellar-phase variation. *J. Bacteriol.* 155, 74–81.
- Wain J., and O’Grady J. (2017). “Genomic diversity in *Salmonella enterica*,” in *Applied Genomics of Foodborne Pathogens*, eds X. Deng, H. C. den Bakker, and R. S. Hendriksen (Cham: Springer), 91–107.
- Yamamoto, S., and Kutsukake, K. (2006). FljA-mediated posttranscriptional control of phase 1 flagellin expression in flagellar phase variation of *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 188, 958–967. doi: 10.1128/JB.188.3.958-967.2006
- Yokoseki, T., Kutsukake, K., Ohnishi, K., and Iino, T. (1995). Functional analysis of the flagellar genes in the *fliD* operon of *Salmonella typhimurium*. *Microbiology* 141 (Pt 7), 1715–1722. doi: 10.1099/13500872-141-7-1715
- Yue, M., Han, X., De Masi, L., Zhu, C., Ma, X., Zhang, J., et al. (2015). Allelic variation contributes to bacterial host specificity. *Nat. Commun.* 6:8754. doi: 10.1038/ncomms9754
- Zhou, Z., McCann, A., Littrup, E., Murphy, R., Cormican, M., Fanning, S., et al. (2013). Neutral genomic microevolution of a recently emerged pathogen, *Salmonella enterica* serovar Agona. *PLoS Genet.* 9:e1003471. doi: 10.1371/journal.pgen.1003471
- Zieg, J., Hilmen, M., and Simon, M. (1978). Regulation of gene expression by site-specific inversion. *Cell* 15, 237–244. doi: 10.1016/0092-8674(78)90098-3
- Zou, Q. H., Li, R. Q., Liu, G. R., and Liu, S. L. (2014). Comparative genomic analysis between typhoidal and non-typhoidal *Salmonella* serovars reveals typhoid-specific protein families. *Infect. Genet. Evol.* 26, 295–302. doi: 10.1016/j.meegid.2014.06.008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Liu, Zhang, Zhou, Xu, Zhang and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.