

RESEARCH ARTICLE

Open Access



# Genetic boundaries delineate the potential human pathogen *Salmonella bongori* into discrete lineages: divergence and speciation

Xiaoyu Wang<sup>1,2,3</sup>, Songling Zhu<sup>1,2,3</sup>, Jian-Hua Zhao<sup>1,2,3</sup>, Hong-Xia Bao<sup>1,2,3</sup>, Huidi Liu<sup>1,2,3</sup>, Tie-Min Ding<sup>4</sup>, Gui-Rong Liu<sup>1,2,3</sup>, Yong-Guo Li<sup>5</sup>, Randal N. Johnston<sup>6</sup>, Feng-Lin Cao<sup>1,7\*</sup>, Le Tang<sup>1,2,3,8\*</sup> and Shu-Lin Liu<sup>1,2,3,5,9\*</sup>

## Abstract

**Background:** *Salmonella bongori* infect mainly cold-blooded hosts, but infections by *S. bongori* in warm-blooded hosts have been reported. We hypothesized that *S. bongori* might have diverged into distinct phylogenetic lineages, with some being able to infect warm-blooded hosts.

**Results:** To inspect the divergence status of *S. bongori*, we first completely sequenced the parakeet isolate RKS3044 and compared it with other sequenced *S. bongori* strains. We found that RKS3044 contained a novel T6SS encoded in a pathogenicity island-like structure, in addition to a T6SS encoded in SPI-22, which is common to all *S. bongori* strains so far reported. This novel T6SS resembled the SPI-19 T6SS of the warm-blooded host infecting *Salmonella* Subgroup I lineages. Genomic sequence comparisons revealed different genomic sequence amelioration events among the *S. bongori* strains, including a unique CTAG tetranucleotide degeneration pattern in RKS3044, suggesting non-overlapping gene pools between RKS3044 and other *S. bongori* lineages/strains leading to their independent accumulation of genomic variations. We further proved the existence of a clear-cut genetic boundary between RKS3044 and the other *S. bongori* lineages/strains analyzed in this study.

**Conclusions:** The warm-blooded host-infecting *S. bongori* strain RKS3044 has diverged with distinct genomic features from other *S. bongori* strains, including a novel T6SS encoded in a previously not reported pathogenicity island-like structure and a unique genomic sequence degeneration pattern. These findings alert cautions about the emergence of new pathogens originating from non-pathogenic ancestors by acquiring specific pathogenic traits.

**Keywords:** *Salmonella*, Bacterial pathogens, Genomic divergence, Genetic boundary

## Background

*Salmonella* bacteria are ubiquitous pathogens, with over 2600 serotypes documented to date [1]. Based on different levels of relatedness, the *Salmonella* bacteria are categorized into eight subgroups, e.g., Subgroups I, II, IIIa, IIIb, IV, V, VI and VII [2, 3]. Human salmonellosis is primarily caused by *Salmonella* Subgroup I lineages but may occasionally be elicited also by bacteria of the other

subgroups, which usually infect cold-blooded hosts [4]. Bacteria of Subgroup V (also known as *Salmonella bongori*; see the dynamic changes of *Salmonella* taxonomy and nomenclature in previous publications [5, 6]) lineages are among the least likely human salmonellosis agents, partly due to their lack of *Salmonella* Pathogenicity Island 2, which is required for the bacteria to survive in phagocytes and invade the deep tissues of the host [7, 8]. However, *S. bongori* infections do occur in humans or other warm-blooded hosts [9–11], opening a question about whether certain *S. bongori* lineages with special genetic traits might have facilitated such infections. A first step toward answering this question is the

\* Correspondence: sliu@hrbmu.edu.cn; le.tang@ucalgary.ca; CFL023@163.com

<sup>1</sup>Systemomics Center, College of Pharmacy, and Genomics Research Center (State-Province Key Laboratories of Biomedicine-Pharmaceutics of China), Harbin Medical University, 157 Baojian Road, Harbin 150081, China  
Full list of author information is available at the end of the article



revelation of the genetic differences among the *S. bongori* bacteria isolated from different host species. We postulate that certain genomic characteristics might be owned by some but not other phylogenetic lineages of *S. bongori*, and such lineage-specific characteristics may be identified through comparative genomic analyses. To start with, the core question here is whether *S. bongori* consist of genetically distinct lineages.

In a previous study, we found that the physical structure of bacterial genomes could remain highly conserved for hundreds of millions of years in evolution [12, 13], but in the meantime some subtle genomic features can unambiguously reflect the phylogenetic distinction between even very closely related bacteria [14, 15]. As such, genome structure analysis may provide objective and reliable parameters for differentiating bacteria based on their evolutionary relationships rather than according to any arbitrary standards. To prove this postulation, we recently profiled genomic characteristics among representative human-infecting *Salmonella* pathogens and demonstrated the existence of genetic boundaries that can be used to divide the *Salmonella* bacteria into clear-cut phylogenetic clusters [16–21].

The *Salmonella* bacteria were initially treated as individual species beginning from early 1880s primarily due to their distinct pathogenic features, such as *Salmonella typhimurium* causing gastroenteritis or *Salmonella typhi* causing typhoid fever in humans, and were differentiated by serotyping based on their different combinations of O and H antigens [22, 23]. As a result, the *Salmonella* species were represented each by an antigenic formula, such as *S. typhimurium* by 1,4, [5],12:i:1,2 or *S. typhi* by 9,12, [vi]:d:-. However, in the mid-1980s, all *Salmonella* species were re-classified into a single new species under the specific name *enterica* based on their close genetic relatedness. Consequently, the previous species became serovars of the only *Salmonella* species, *Salmonella enterica* [24]. Later, Subgroup V regained the scientific name of *Salmonella bongori* based on its greater genetic divergence from all other subgroups [25]. In this report, we use the pre-1980 *Salmonella* nomenclature for reasons as detailed in a previous publication [6].

Different from Subgroup I lineages that are mostly monophyletic serotypes, each having had a Latinized scientific name prior to the reclassification of all *Salmonella* species into the single species *enterica*, Subgroup V lineages all have the same scientific name *Salmonella bongori*. To date, more than 20 serotypes have been documented in *S. bongori*, but the precise phylogenetic relationships of the bacteria among the serotypes and the phyletic status within each of the serotypes remain largely unclear.

In this study, we completely sequenced and annotated the genome of a parakeet isolate of *S. bongori*, strain

RKS3044, which is one of the *Salmonella* Reference Collection C strains kindly provided by Dr. Robert K. Selander ([www.ucalgary.ca/~kesander](http://www.ucalgary.ca/~kesander)), and compared it with other sequenced *S. bongori* strains that had different antigenic formulae or were isolated from cold-blooded hosts. Here we report our findings that the sequenced *S. bongori* strains have diverged into distinct lineages and, interestingly, strain RKS3044 had a novel cluster of T6SS-associated genes. Whereas we cannot conclude that the additional T6SS might be involved in the warm-blooded host invasion by RKS3044, our results show that *S. bongori* bacteria can be circumscribed into discrete phylogenetic clusters, each having a distinct set of genomic characteristics.

## Results

### Genomic comparisons of the completely sequenced *S. bongori* strains: RKS3044 representing a warm-blooded host pathogen

For overall comparisons, we first completely sequenced *S. bongori* strain RKS3044. Its genome consists of a single chromosome of 4,394,500 bp (51.4% GC content), which has a perfectly balanced physical structure between *oriC* and *terC* like most other reported *Salmonella* genomes [13, 26, 27] (Additional file 1: Figure S1); detailed information on the genome is summarized in Table 1 and the distribution of genes into COGs functional categories is presented in Table 2. We compared RKS3044 with three previously sequenced *S. bongori* strains, including N268–08 [28], NCTC12419 [29] and SA19883065 ([www.ucalgary.ca/~kesander](http://www.ucalgary.ca/~kesander)), and draft genomic sequences of several other *S. bongori* strains (Additional file 3: Table S1). When the four complete genomes were aligned, we found

**Table 1** General characteristics of the complete *S. bongori* RKS3044 genome

Attribute	Value	% of Total
Genome size (bp)	4,394,500	100% of total genome
DNA coding (bp)	3,859,991	87.84% of total genome
DNA G + C (bp)	2,258,648	51.40% of total genome
Total predicted genes	4121	87.84% of total genome
Protein coding genes	4017	97.48% of total genes
RNA genes	22	0.53% of total genes
Pseudo genes	77	1.87% of total genes
Genes in internal clusters	2847	69.09% of total genes
Genes with function prediction	3563	86.46% of total genes
Genes assigned to COGs	3066	74.40% of total genes
Genes with Pfam domains	2449	68.54% of total genes
Genes with signal peptides	378	9.17% of total genes
Genes with transmembrane helices	1023	24.82% of total genes
CRISPR repeats	2	

**Table 2** Categories of genes associated with 25 general COG functions

Code	Value	% of total <sup>a</sup>	Description
J	174	0.57	Translation
A	1	0.03	RNA processing and modification
K	257	7.49	Transcription
L	132	3.85	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	31	0.90	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	46	1.34	Defense mechanisms
T	152	4.43	Signal transduction mechanisms
M	215	6.27	Cell wall/membrane biogenesis
N	103	3.00	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	113	3.29	Intracellular trafficking and secretion
O	143	4.17	Posttranslational modification, protein turnover, chaperones
C	236	6.88	Energy production and conversion
G	324	9.45	Carbohydrate transport and metabolism
E	318	9.27	Amino acid transport and metabolism
F	81	2.36	Nucleotide transport and metabolism
H	144	4.20	Coenzyme transport and metabolism
I	82	2.39	Lipid transport and metabolism
P	182	5.31	Inorganic ion transport and metabolism
Q	47	1.37	Secondary metabolites biosynthesis, transport and catabolism
R	333	9.17	General function prediction only
S	316	9.12	Function unknown
-	1055	25.60	Not in COGs

<sup>a</sup>The total is based on the total number of protein coding genes in the annotated genome

remarkable differences among them, particularly in different sets of insertions such as *Salmonella* Pathogenicity Islands (SPIs), which make the genome sizes considerably different among them (Fig. 1 and Additional file 4: Table S2). Whereas strains NCTC12419 and SA19983065, both being 66:z41:-, are highly similar, RKS3044 (48:z41:-) and N268-08 (antigenic formula unknown) are very different from each other and both are different from the pair of NCTC12419/SA19983065 strains. Such fundamental genomic differences suggest phylogenetic divergence over long evolutionary times among these bacteria as discussed previously [17, 18, 20].

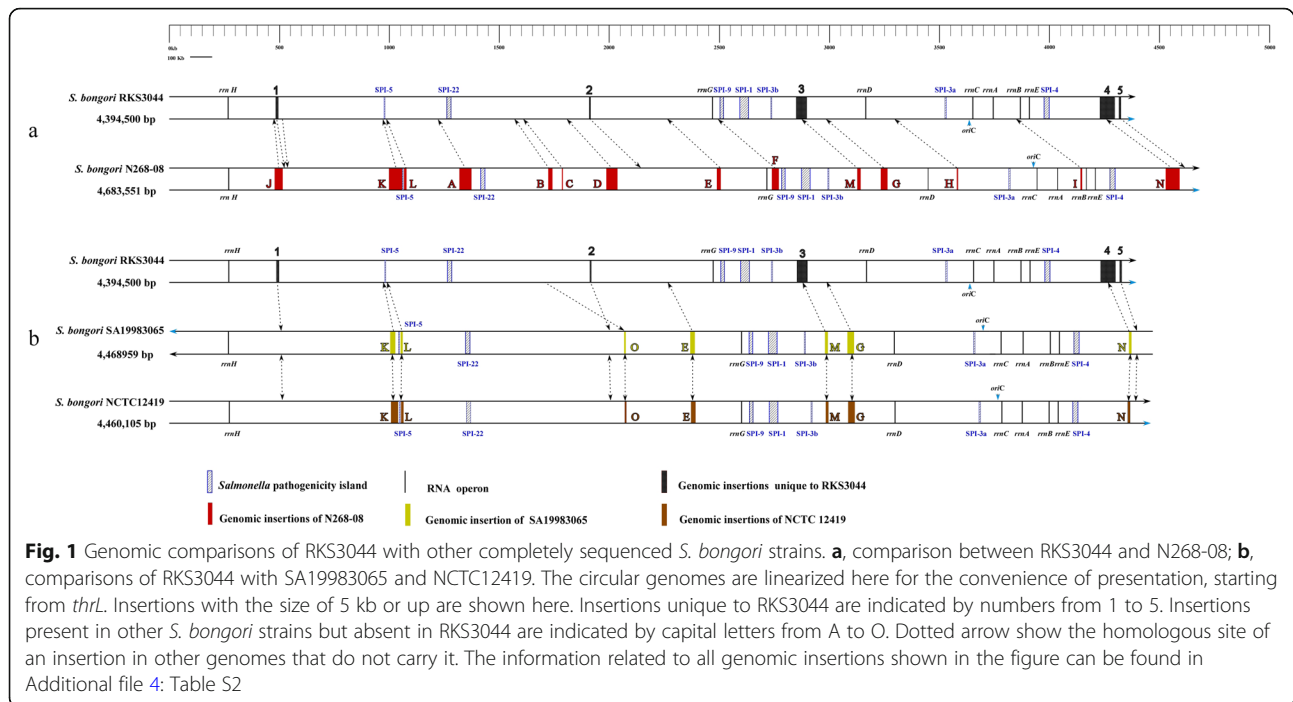
#### Phylogenetic relationships among the *S. bongori* bacteria

To estimate the population phylogenetic structure of *S. bongori*, we analyzed the sequenced *S. bongori* strains (see Additional file 3: Table S1) in comparison with representative strains of the other *Salmonella* subgroups (Additional file 5: Table S3). We concatenated 2804 genes

common to the 26 analyzed strains and constructed a phylogenetic tree. We found that in general the *S. bongori* strains had similar genetic distances from one another as those seen among the Subgroup I strains (Fig. 2). For example, *S. bongori* strains RKS3044 and N268-08 had a genetic distance between them similar to that as between *S. typhi* and *S. typhimurium*, two serotypes that are currently categorized into the same taxonomic species but are in fact vastly different pathogens. As we have recently resolved *S. typhi* and *S. typhimurium* into different natural species based on the clear-cut genetic boundary (i.e., allelic distance) between them [6, 19], we wondered whether a genetic boundary may also have been formed between *S. bongori* strains RKS3044 and N268-08.

#### Genetic boundaries delineating the *S. bongori* strains into clear-cut phylogenetic clusters

To evaluate the phylogenetic significance of the demonstrated genetic separation among the *S. bongori* strains,



i.e., to determine whether the genetic divergence might be clear-cut (i.e., making the lineages clearly discrete) or continuous (i.e., showing a spectrum of gradual or continuous genetic differences without clear cut-offs among the “lineages”), we detected the percentages of the homologous genes that share identical nucleotide sequences (zero nucleotide degeneracy) between the bacteria (Additional file 2: Figure S2 and Additional file 6: Table S4). Based on the percentages, we categorized the *S. bongori* strains into three groups: high, from 85 to 100% (e.g., between N268-08 and BCW\_1557 or CATO-2016, between NCTC12419 and SA19983065 or BCW\_1556, etc.); medium, 20–35% (e.g., between RKS3044 and BCW\_1554, between SA19983065 and CATO-2016, between 1308 and 83 and BCW\_1552); low, lower than 20% and mostly around 10% (the majority of the analyzed *S. bongori* strains; Additional file 2: Figure S2). The “high” group resembles strains within *S. typhi*, which are highly cohesive as members of the same phylogenetic cluster [19]; the “medium” group resembles bacteria between *S. gallinarum* and *S. pullorum*, which have diverged into separate lineages for not long and hence remain highly related though with obvious genetic distinction between them [16]; the “low” group resembles bacteria between most *Salmonella* Subgroup I serotypes like *S. typhi* and *S. typhimurium*, which are circumscribed into discrete phylogenetic clusters of bacteria equivalent to natural species [6, 16, 19]. In this study, we did not find *S. bongori* strains that had the percentages between 35 and 85%, probably because the number of *S. bongori* strains analyzed here was not large enough to

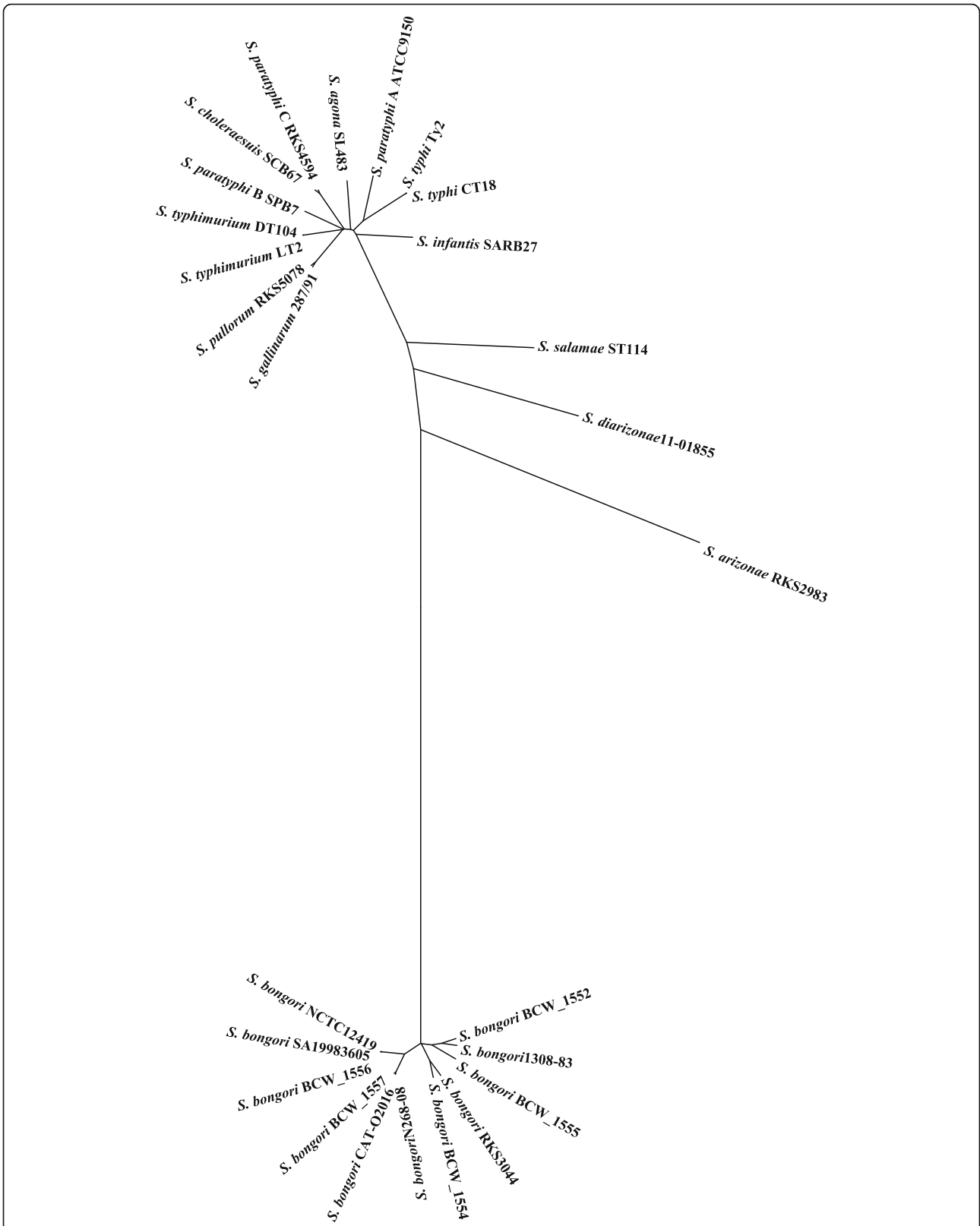
cover sufficient ranges of the genetic divergence, although our previous work has demonstrated that the “intermediate percentages”, such as those around 40% between *S. gallinarum* and *S. pullorum* [16], is indicative of phylogenetic divergence of the compared bacteria into distinct species.

#### Genes involved in pathogenesis and virulence: different levels of evolutionary conservation between *S. bongori* and other *Salmonella* subgroups

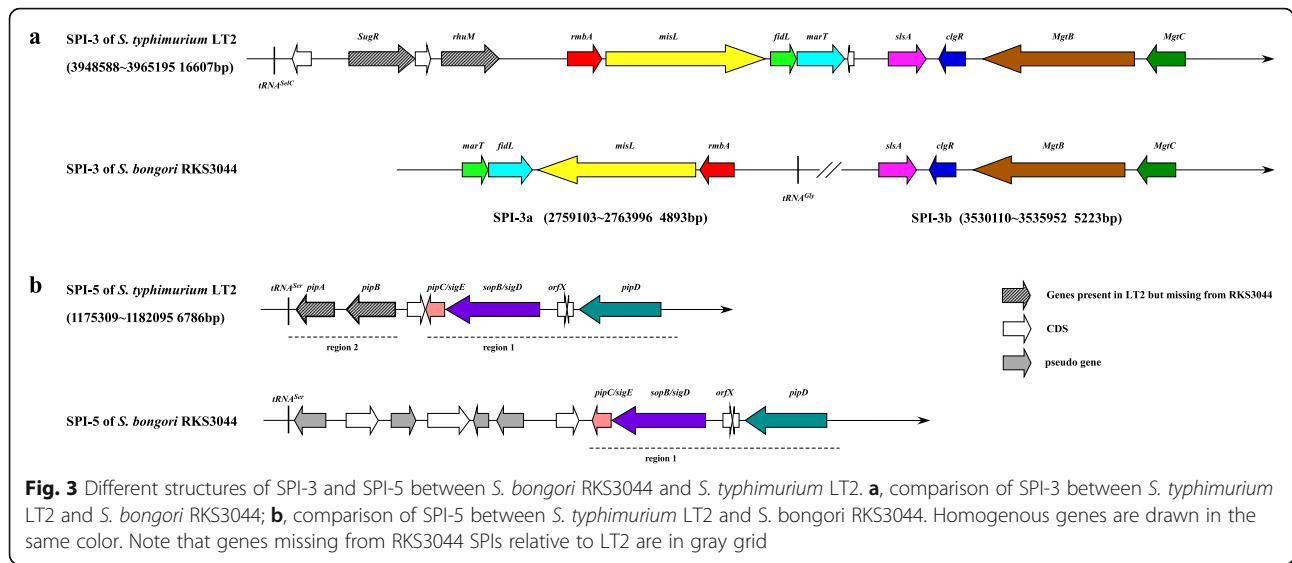
RKS3044 possesses all SPIs previously reported in *S. bongori*, including SPIs-1, 3, 4, 5, 9 and 22, which have various levels of structural or sequence similarities as exemplified by SPIs-3 and 5 between *S. bongori* and *S. typhimurium* LT2 (Fig. 3). By search against VFDB (Virulence Factors Pathogenic Bacteria, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), we identified most of the previously reported *S. bongori* virulence genes in RKS3044 strain, except some of the genes encoding T3SS effector proteins (Additional file 7: Table S5). In addition to the known SPIs, we identified a novel genomic island in RKS3044, which we temporarily designate as SPI-RKS3044 (Insertion 3 in Fig. 1). This 48 kb SPI was RKS3044-specific, flanked at the upstream end by a tRNA<sup>Phe</sup> gene and having fluctuated GC contents, an indication of mosaic DNA segments of different evolutionary origins.

#### A new type VI secretion system in SPI-RKS3044

Within SPI-RKS3044, we annotated a new Type VI secretion system (T6SS) at genomic location spanning nucleotides 2,861,308~2,886,179. Like other *S. bongori*



**Fig. 2** Phylogenetic tree of the *Salmonella* strains. The phylogenetic tree was constructed by genes common to the 26 *Salmonella* strains representing *Salmonella* subgroups I, II, IIIa, IIIb, and V

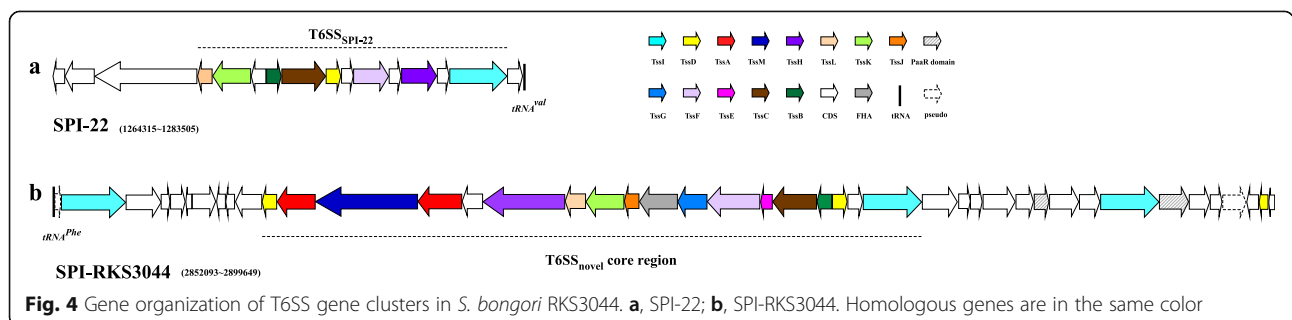


strains, RKS3044 also has a T6SS in SPI-22. Therefore RKS3044 carries two T6SSs (Additional file 8: Table S6), but the two T6SS clusters have very different structures (Fig. 4), suggesting distinct evolutionary histories and different roles in bacterial pathogenesis between the two T6SSs. The T6SS in SPI-22 (Fig. 4a) has homologs encoding an Hcp-like protein (N643\_RS05930) and a VgrG protein (N643\_RS05960), which are both required for T6SS apparatus functionality; cytoplasmic proteins VipA (N643\_RS05920) and VipB (N643\_RS05925), which interact directly and are required for intracellular growth in macrophages; and other essential components such as DotU (N643\_RS05890), ImpA (N643\_RS05895), gp25-like protein (N643\_RS05935) and VasA (N643\_RS05940). However, in the SPI-22 T6SS we did not find homologs encoding COG0542 (ClpV), COG3521 (SciN) or COG3523 (IcmF) components that were present in previously reported SPI-22 T6SSs [29], which are usually required for the T6SS functionality. The novel T6SS in the 48 kb SPI-RKS3044 (designated T6SS<sub>novel</sub>) carries many more genes than the SPI-22 T6SS, covering nucleotides 2,852,331–2,899,649 (Fig. 4b), with structural similarity to the T6SS in SPI-19 (T6SS<sub>SPI-19</sub>) of

*Salmonella* Subgroup I serotypes, such as *S. dublin*, *S. weltevreden*, *S. agona* and *S. gallinarum* [30].

**Phylogenetic analyses of T6SS gene clusters**

To assess the evolutionary relationships of the T6SSs in RKS3044 with those in closely related bacteria, we constructed a phylogenetic tree on concatenated TssB and TssC protein sequences of T6SS<sub>novel</sub> and T6SS<sub>SPI-22</sub> of RKS3044 and those from selected *E. coli* and *Salmonella* lineages. Previously, the *E. coli* T6SS gene clusters had been categorized into three distinct phylogenetic groups based on different levels of structural and sequence similarities: T6SS-1, T6SS-2 and T6SS-3, corresponding to Type i2, i1 and i4b, respectively [31–33]. On the phylogenetic tree, the *E. coli*-associated T6SSs and *Salmonella*-associated SPI-T6SSs were mixed together among five main branches, with the RKS3044 T6SS<sub>novel</sub> appearing on the type i1 branch (Fig. 5a). This finding proves the horizontal acquisition nature of T6SSs. When we focused the phylogenetic analysis on T6SS<sub>novel</sub> and other type i1 T6SSs in bacteria of 30 Proteobacteria, using the concatenated protein sequences of TssB, C, E, F, G, H, J, K and M, we found that RKS3044 T6SS<sub>novel</sub> was clustered



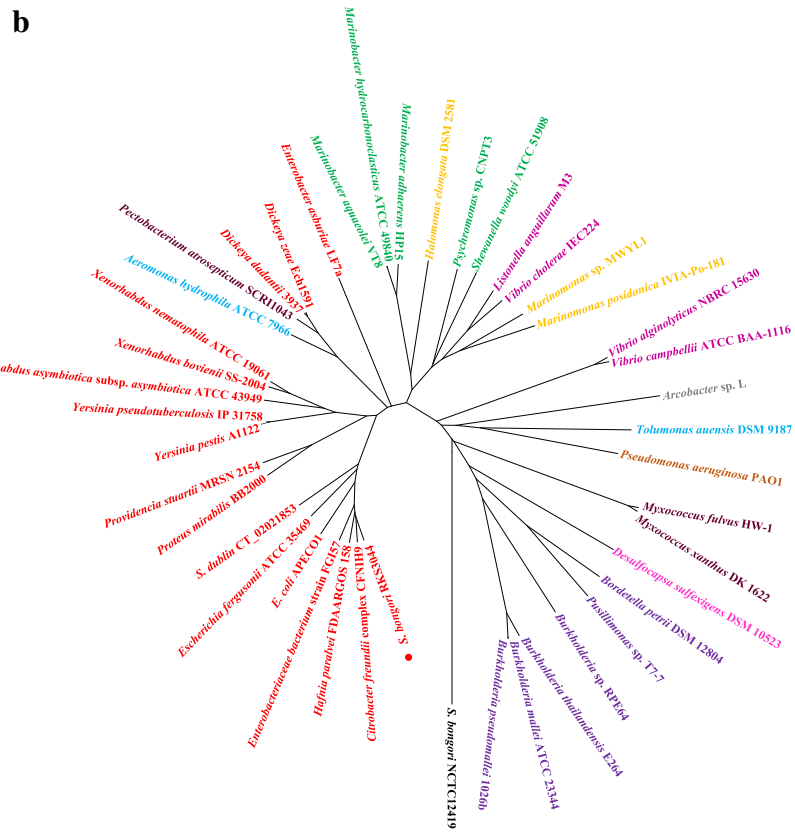
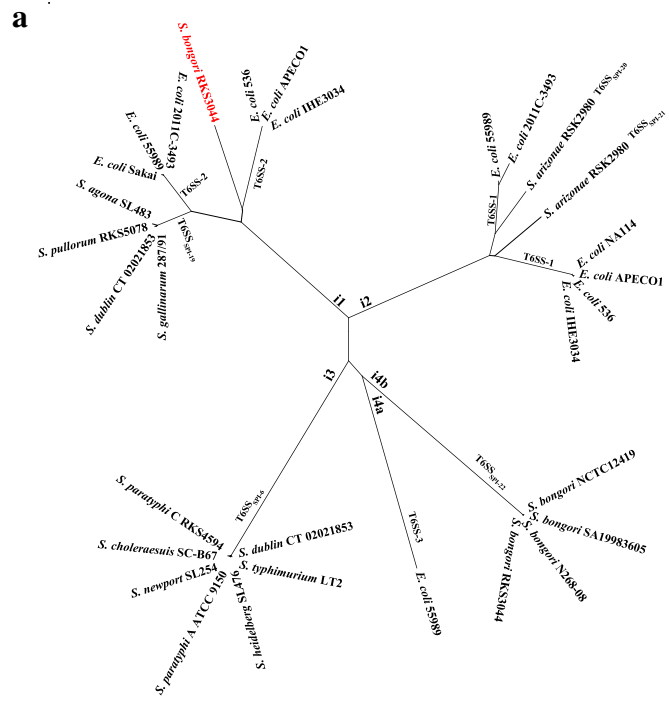


Fig. 5 (See legend on next page.)

(See figure on previous page.)

**Fig. 5** Evolutionary analysis of *S. bongori* RKS3044 T6SS<sub>novel</sub>. **a** comparison between T6SS<sub>novel</sub> and T6SSs in selected *Salmonella* and *E. coli* lineages; the neighbour-joining tree was constructed from concatenated TssB and TssC protein sequences. **b** comparison between T6SS<sub>novel</sub> and Type I T6SSs in selected Proteobacteria strains; the neighbour-joining tree was calculated from concatenated protein sequences of TssB, C, E, F, G, H, J, K and M, with the *S. bongori* NCTC12419 T6SS<sub>SPI-22</sub> as outgroup

with T6SSs of other *Enterobacteriaceae* bacteria and was, interestingly, more closely related to that of *Citrobacter freundii* than to the *E. coli* and *Salmonella*- associated T6SSs (Fig. 5b and Additional file 9: Table S7).

#### Genomic divergence by DNA sequence amelioration during evolution for adaptation as probed by unique CTAG degeneration patterns

RKS3044 and the other *S. bongori* strains form distinct phylogenetic lineages. We thus anticipated independent accumulation of genomic variations in these compared bacteria, including not only distinct sets of genomic insertions such as SPI-RKS3044 or T6SS<sub>novel</sub> encoded within it, but also independent nucleotide substitutions such as the degeneration of some highly conserved short nucleotide sequences [17, 18, 20]. Our hypothesis was that, following the acquisition of certain peculiar traits such as the ability to expand host range, genomic DNA sequence amelioration may ensue to become better adapted to the new selection pressures, as seen in the divergence between the closely related *Salmonella* lineages *S. paratyphi* C and *S. choleraesuis* [34]. To probe such hypothesized genomic sequence amelioration events, we profiled the tetranucleotide sequence CTAG in the completely sequenced *S. bongori* strains and conducted systematic comparisons among the bacteria, focusing on its genomic location and degeneration patterns. We found that the CTAG sequence had different genomic distributions among the *S. bongori* lineages and the divergence patterns were consistent with the phylogenetic clustering of the bacteria, suggesting that the lineages have diverged into different phylogenetic and ecological positions for long evolutionary times, so they do not have much chances to freely exchange their genetic materials (Additional file 10: Table S8).

#### Discussion

Among all documented bacteria to date, pathogenic species are a tiny portion but some of them are deadly, such as the typhoid agent *S. typhi*. Although genetic events involved in transforming a benign bacterial ancestor to a pathogen have been richly documented, a general picture about the origin of pathogenic bacterial species is lacking. Based on our previous findings that the ubiquitous *Salmonella* lineages, which are descendants of a common ancestor giving rise to *Escherichia* and *Salmonella* about 120–160 million years ago [35–37], all have their own unique sets of laterally acquired genes and

genomic sequence amelioration patterns [12, 17, 38–40], we proposed the Adopt-Adapt model of bacterial speciation, with the “adopted” lateral genes diverting the direction of evolution and the ensuing genomic sequence amelioration for “adaptation” to accept the adopted genes and become increasingly fit to the new niche, e.g., a new host or environment [6, 26, 41]. The newly speciated bacterial lineage would continue accumulating further genomic variations independently and eventually become established in a separate gene pool isolated by a genetic boundary from others [16]. To prove whether the warm-blooded host isolate *S. bongori* RKS3044 might represent a novel natural species, we compared it with previously sequenced *S. bongori* strains to identify the hypothesized unique set of laterally adopted genes and adaptive sequence amelioration events in RKS3044 and to determine whether genetic boundaries might have been formed between RKS3044 and other *S. bongori* strains.

*S. bongori* bacteria have diverged from the other *Salmonella* lineages for ca. 40–63 million years [42] and they still form a tight phylogenetic cluster with obvious genetic distances from all other *Salmonella* subgroups [13, 25, 29]. As the extant *Salmonella* representative that first diverged from the common ancestor with *E. coli*, *S. bongori* may provide information about the evolutionary processes from a benign bacterial ancestor to the diverse *Salmonella* pathogens. The emergence of warm-blooded host infecting agent from the well-established cold-blooded host pathogen alerts the potential of *S. bongori* to become a common human pathogen.

We focused on comparisons between *S. bongori* RKS3044 and other sequenced *S. bongori* strains to reveal their possible genomic differences. We found that RKS3044 shared most of the virulence genes with the previously sequenced *S. bongori* strains and also lacked SPI-2, but on the other hand contained a novel T6SS (T6SS<sub>novel</sub>) encoded in a new SPI (SPI-RKS3044) identified in this study. As T6SS<sub>novel</sub> shares significant structural and sequence similarity with SPI-19 T6SS, which is required for survival the macrophages and for efficient colonization into deep tissues of warm-blooded hosts [43–45], we postulated its evolutionary roles to divert a lineage of *S. bongori* toward infecting warm-blooded hosts. If so, according to the hypothesized Adopt-Adapt model of bacterial speciation, genomic analysis should reveal a unique nucleotide sequence amelioration pattern [17, 18, 20], distinct from those of all previously reported *S. bongori* strains. When we profiled the tetranucleotide sequence CTAG in the genome of



RKS3044 and conducted comparisons with the other sequenced *S. bongori* strains, we found that the patterns were unique for RKS3044 and strain N268–08; strains NCTC12419 and SA19983065 shared another special pattern (Additional file 10: Table S8). As NCTC12419 and SA19983065 belong to the same serotype (66:z41:-) that is different from the antigenic formula of RKS3044 (48:z41:-), it seemed that the serotype situation in *S. bongori* might be similar to that of *Salmonella* Subgroup I lineages, in which most serotypes are monophyletic and the monophyletic serotypes correspond to natural species [6]. To prove this, we needed to resolve genetic boundaries among them and eventually found that the *S. bongori* serotypes were indeed separated into discrete phylogenetic clusters by large allelic distances similar to those of *Salmonella* Subgroup I lineages (Additional file 6: Table S4 and Additional file 1: Figure S1). These findings are consistent with our previous reports that bacteria of a monophyletic *Salmonella* lineage, such as *S. typhimurium*, *S. typhi*, *S. gallinarum* or *S. pullorum*, collected from a wide range of geographical, temporal or spatial spans, have a common genome structure or share the same gene pool as reflected by the genetic boundary that circumscribe them together and separate the monophyletic *Salmonella* lineages into discrete phylogenetic clusters [17, 18, 20, 46].

The concept of natural species for bacteria is based on the notion that bacteria of a monophyletic *Salmonella* lineage collected from a wide range of geographical, temporal or spatial spans have high percentages of homologous genes with zero sequence degeneracy, reflecting dynamic processes of clonal expansion to eliminate less fit subpopulations. On the other hand, bacteria of different *Salmonella* lineages, no matter how closely related they might be from one another, such as between *S. gallinarum* or *S. pullorum*, have low percentages of homologous genes with zero sequence degeneracy, since they as distinct lineages have diverged for a long evolutionary time and in the process accumulated mutations independently. Notably, there are usually broad windows between the “high” and “low” percentages (higher than 70% vs lower than 20%, without intermediates between 70 and 20%, with rare exceptions) [17, 18, 20, 46], which would make the allelic distance an applicable parameter to define and delineate bacteria into natural species. The method we used can effectively profile the nucleotide sequence information and the results can be evaluated by investigators of different laboratories. The low percentages of homologous genes with zero sequence degeneracy between different *Salmonella* lineages can be interpreted to be the consequences of independent accumulation of nucleotide variations by the bacteria over a long period.

To summarize, our results in this research show that *S. bongori* consist of discrete phylogenetic groups corresponding to the individual serotypes, which are equivalent

to natural species. Whereas all sequenced *S. bongori* strains have a T6SS in SPI-22, strain RKS3044 has an additional T6SS in SPI- RKS3044. As the SPI-22 T6SS in RKS3044 is defective but the SPI-RKS3044 T6SS resembles the functional SPI-19 T6SS of the warm-blooded host infecting *Salmonella* Subgroup I lineages, it is possible that acquisition of the SPI-19-like T6SS might have facilitated the evolution of a branch of *S. bongori* represented by RKS3044 toward infecting warm-blooded hosts. A unique genomic sequence amelioration profile as reflected by the distinct CTAG tetranucleotide degeneration pattern from even very closely related *S. bongori* lineages suggests independent accumulation of genomic variations and the detection of a large allelic distance between RKS3044 and other *S. bongori* lineages/strains supported this postulation. Our study alerts cautions about the emergence of new pathogens originating from non-pathogenic ancestors, so preventive strategies are necessary to minimize such risks.

## Conclusions

The warm-blooded host-infecting *S. bongori* strain RKS3044 has distinct genomic features from other *S. bongori* strains, including a novel T6SS encoded in a previously not reported pathogenicity island-like structure and a unique genomic sequence degeneration pattern. These findings provide new support to the model of bacterial speciation to become pathogens.

## Methods

### Bacterial strains, growth conditions and DNA isolation

All bacterial strains used in this study were obtained from *Salmonella* Genetic Stock Center (SGSC; [www.ucalgary.ca/~kesander](http://www.ucalgary.ca/~kesander)) and cultured at 37 °C in LB broth or on LB agar plates. DNA was isolated from the bacterial cells by the CTAB (Cetyltrimethyl ammonium bromide) bacterial genomic DNA isolation method, purified and eluted by the QIAGEN DNA purified Kit (Qiagen Germany).

### Genome sequencing, assembly and annotation

The genome of *S. bongori* RKS3044 was sequenced by two sequencing platforms (SOLiD 3.0 and Illumina HiSeq 2000) according to the manual for the instrument. This work was initiated with SoLid time and we then added the HiSeq data to finish the sequencing project. Sequence data from the two methods were assembled by the velvet v1.2.09 software. We used the PFGE techniques to create a structural framework to align the contigs of SoLid and HiSeq short reads. The assembled scaffolds contained several gaps, which were closed by PCR amplification and ABI3730 sequencing. Eventually, we verified the correctness of the finished genome by PFGE based on cleavage data of XbaI, AvrII and SpeI.

Genes were annotated using NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) [47]. Additional analysis of gene prediction and annotation was supplemented using the IMG platform [48]. The graphical circular map of the *S. bongori* RKS3044 genome was generated with the CGviewer software.

### Comparative genome analysis

We concatenated the core genes common to the bacterial strains analyzed in this study and conducted comparisons using the Basic Local Alignment Search Tool (BLAST) with the parameters set at >70% DNA identity and >70% gene length to categorize genes into common genes. Individual orthologous sequences were aligned by the MAFFT program [49]. The phylogenetic trees of the genomes and T6SS loci were structured using the Neighbor-Joining method [50] in MEGA6 [51] by 1000 bootstrap replicates. The evolutionary distances were computed using the p-distance method [52]. To identify the virulence factor genes in RKS3044, we performed a BLAST search of whole RKS3044 ORFs against the virulence factor protein sequences of core dataset listed in VFDB [53] with an e-value of 1e-5. Genes encoding the T6SS were identified using the SecReT6 database [54].

For comparative analysis, genome sequences of representative *Salmonella* and *E. coli* strains were downloaded from the NCBI website (Additional file 5: Table S3).

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-019-6259-z>.

**Additional file 1: Figure S1.** Graphical map of the *S. bongori* RKS3044 genome. From the outside to the center: genes on forward strand (color by COG categories), genes on reverse strand (color by COG categories), GC content, and GC skew. The map was generated with the CGviewer software.

**Additional file 2: Figure S2.** Genomic comparison among the *Salmonella bongori* strains. Sequences common to eleven strains were concatenated and pair-wise aligned for the number of genes that have 100% sequence identity.

**Additional file 3: Table S1.** Sequenced *S. bongori* strains used for genomic comparisons in this study.

**Additional file 4: Table S2.** Profiles of genomic insertions in the completely sequenced *S. bongori* strains.

**Additional file 5: Table S3.** *Salmonella* and *E. coli* strains included in the comparative analysis.

**Additional file 6: Table S4.** Percentages of homologous genes that share identical nucleotide sequences between pairs of the bacteria compared.

**Additional file 7: Table S5.** Comparison of virulence gene profiles between *S. bongori* strains RKS3044 and NCTC12419.

**Additional file 8: Table S6.** Annotation of *S. bongori* RKS3044 SPI-22 and SPI-RKS3044 genes.

**Additional file 9: Table S7.** Bacterial strains used for comparative analysis of T6SS gene clusters.

**Additional file 10: Table S8.** CTAG profiles in completely sequenced *S. bongori* lineages/strains.

### Acknowledgements

Not applicable

### Authors' contributions

XYW coordinated the project and conducted data analysis; SZ, JHZ, HXB, HDL and TMD were involved in experiments and data analysis; GRL, YGL, RNJ, FLC and LT contributed reagents/ materials/ analysis tools; SLL conceived the study and finalized the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by Heilongjiang Innovation Endowment Awards for graduate studies (YJSCX2012-197HLJ, YJSCX2012-198HLJ), and grants of National Natural Science Foundation of China (NSFC30970078, 81271786, 81671980, 31600001, 31700126, 81871623). The funding bodies played no roles in the design of the study; the collection, analysis or interpretation of data; or in writing the manuscript.

### Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NCBI website repository, ([www.ncbi.nlm.nih.gov/ncbisearch/](http://www.ncbi.nlm.nih.gov/ncbisearch/)), with the accession number: *S. bongori* NCTC12419 (NC\_015761), *S. bongori* N268-08 (NC\_021870), *S. bongori* RKS3044 (NZ\_CP006692), *S. bongori* SA19983605 (NZ\_CP022120), *S. bongori* 1308-83 (NZ\_MXLD01000001), *S. bongori* BCW\_1557 (NZ\_MXOC01000001), *S. bongori* BCW\_1556 (NZ\_MXOD01000001), *S. bongori* BCW\_1555 (NZ\_MXOE01000001), *S. bongori* BCW\_1554 (NZ\_MXOF01000001), *S. bongori* BCW\_1552 (NZ\_MXOH01000001), *S. bongori* CATO-2016 (NZ\_NAPQ01000001), *S. typhi* Ty2 (NC\_004631), *S. typhi* CT18 (NC\_003198), *S. typhimurium* LT2 (NC\_003197), *S. typhimurium* DT104 (NC\_022569), *S. pullorum* RKS5078 (NC\_016831), *S. gallinarum* 287/91 (NC\_011274), *S. paratyphi* A ATCC 9150 (NC\_006511), *S. paratyphi* B SPB7 (NC\_010102), *S. paratyphi* C RKS4594 (NC\_012125), *S. choleraesuis* SC-B67 (NC\_006905), *S. infantis* SARB27(CM001274), *S. agona* L483 (NC\_011149), *S. arizonae* 62:z36- RSK2983 (NZ\_CP006693), *S. diarizonae* 11-01855 (NZ\_CP011288), *S. salamae* 57:z29:z42 St114 (NZ\_CP022467), *S. arizonae* 62:z4:z23- RSK2980 (NC\_010067), *S. heidelberg* SL476 (NC\_011083), *S. dublin* CT\_02021853 (NC\_011205), *E. coli* O157:H7 Sakai (NC\_002695), *E. coli* O104:H4 2011C-3493 (NC\_018658), *E. coli* 55989 (NC\_011748), *E. coli* 536 (NC\_008253), *E. coli* IHE3034 (NC\_017628), *E. coli* NA114 (NC\_017644). All data generated or analysed during this study are included in this published article and its supplementary information files.

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Systemomics Center, College of Pharmacy, and Genomics Research Center (State-Province Key Laboratories of Biomedicine-Pharmaceutics of China), Harbin Medical University, 157 Baojian Road, Harbin 150081, China. <sup>2</sup>HMU-UCCSM Centre for Infection and Genomics, Harbin Medical University, Harbin, China. <sup>3</sup>Translational Medicine Research and Cooperation Center of Northern China, Heilongjiang Academy of Medical Sciences, Harbin, China. <sup>4</sup>Department of Medicine and Food Engineering, Harbin Labor Technician College, Harbin, China. <sup>5</sup>Department of Infectious Diseases, The First Affiliated Hospital, Harbin Medical University, Harbin, China. <sup>6</sup>Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Canada. <sup>7</sup>Department of Hematology, The First Affiliated Hospital, Harbin Medical University, Harbin, China. <sup>8</sup>Department of Ecosystems and Public Health, University of Calgary, Calgary, Canada. <sup>9</sup>Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Canada.

Received: 15 March 2019 Accepted: 5 November 2019

Published online: 04 December 2019

## References

- Guibourdenche M, Roggentin P, Mikoleit M, Fields PI, Bockemuhl J, Grimont PA, Weill FX. Supplement 2003-2007 (No. 47) to the White-Kauffmann-Le Minor scheme. *Res Microbiol.* 2010;161(1):26–9.
- Selander RK, Beltran P, Smith NH, Helmuth R, Rubin FA, Kopecko DJ, Ferris K, Tall BD, Cravioto A, Musser JM. Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. *Infect Immun.* 1990;58(7):2262–75.
- Boyd EF, Wang FS, Whittam TS, Selander RK. Molecular genetic relationships of the salmonellae. *Appl Environ Microbiol.* 1996;62(3):804–8.
- Popoff MY, Le Minor LE. Genus XXXIII. *Salmonella*. In: Brenner DJ, Krieg NR, Stanley JT, editors. *Bergey's Manual of Systematic Bacteriology*, vol. 2. 2nd ed: Springer; 2005. p. 764–99.
- Agbaje M, Begum RH, Oyekunle MA, Ojo OE, Adenubi OT. Evolution of *Salmonella* nomenclature: a critical note. *Folia Microbiol (Praha).* 2011;56(6):497–503.
- Tang L, Liu SL. The 3Cs provide a novel concept of bacterial species: messages from the genome as illustrated by *Salmonella*. *Antonie Van Leeuwenhoek.* 2012;101(1):67–72.
- Groisman EA, Ochman H. How *Salmonella* became a pathogen. *Trends Microbiol.* 1997;5(9):343–9.
- Hensel M. *Salmonella* pathogenicity island 2. *Mol Microbiol.* 2000;36(5):1015–23.
- Giammanco GM, Pignato S, Mammina C, Grimont F, Grimont PA, Nastasi A, Giammanco G. Persistent endemicity of *Salmonella bongori* 48z(35):–in southern Italy: molecular characterization of human, animal, and environmental isolates. *J Clin Microbiol.* 2002;40(9):3502–5.
- Pignato S, Giammanco G, Santangelo C, Giammanco GM. Endemic presence of *Salmonella bongori* 48z35:– causing enteritis in children in Sicily. *Res Microbiol.* 1998;149(6):429–31.
- Marti R, Hagens S, Loessner MJ, Klumpp J. Genome Sequence of *Salmonella bongori* Strain N268–08 [corrected]. *Genome Announcements.* 2013;1(4):e00580-13.
- Liu SL, Hessel A, Sanderson KE. Genomic mapping with I-Ceu I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. *Proc Natl Acad Sci U S A.* 1993;90(14):6874–8.
- Liu SL, Schryvers AB, Sanderson KE, Johnston RN. Bacterial phylogenetic clusters revealed by genome structure. *J Bacteriol.* 1999;181(21):6747–55.
- Liu SL, Hessel A, Sanderson KE. The XbaI-BlnI-CeuI genomic cleavage map of *Salmonella enteritidis* shows an inversion relative to *Salmonella typhimurium* LT2. *Mol Microbiol.* 1993;10(3):655–64.
- Liu SL, Sanderson KE. Genomic cleavage map of *Salmonella typhi* Ty2. *J Bacteriol.* 1995;177(17):5099–107.
- Tang L, Li Y, Deng X, Johnston RN, Liu GR, Liu SL. Defining natural species of bacteria: clear-cut genomic boundaries revealed by a turning point in nucleotide sequence divergence. *BMC Genomics.* 2013;14:489.
- Tang L, Liu WQ, Fang X, Sun Q, Zhu SL, Wang CX, Wang XY, Li YG, Zhu DL, Sanderson KE, et al. CTAG-containing cleavage site profiling to delineate *Salmonella* into natural clusters. *PLoS One.* 2014;9(8):e103388.
- Tang L, Mastriani E, Zhou YJ, Zhu S, Fang X, Liu YP, Liu WQ, Li YG, Johnston RN, Guo Z, et al. Differential degeneration of the ACTAGT sequence among *Salmonella*: a reflection of distinct nucleotide amelioration patterns during bacterial divergence. *Sci Rep.* 2017;7(1):10985.
- Tang L, Wang CX, Zhu SL, Li Y, Deng X, Johnston RN, Liu GR, Liu SL. Genetic boundaries to delineate the typhoid agent and other *Salmonella* serotypes into distinct natural lineages. *Genomics.* 2013;102(4):331–7.
- Tang L, Zhu S, Mastriani E, Fang X, Zhou YJ, Li YG, Johnston RN, Guo Z, Liu GR, Liu SL. Conserved intergenic sequences revealed by CTAG-profiling in *Salmonella*: thermodynamic modeling for function prediction. *Sci Rep.* 2017;7:43565. <https://doi.org/10.41038/srep44356>.
- Bao HX, Tang L, Yu L, Wang XY, Li Y, Deng X, Li YG, Li A, Zhu DL, Johnston RN, et al. Differential efficiency in exogenous DNA acquisition among closely related *Salmonella* strains: implications in bacterial speciation. *BMC Microbiol.* 2014;14:157.
- Kauffmann F. On the serology of the *Salmonella* group. *Acta pathologica et microbiologica Scandinavica.* 1947;24(3–4):242–50.
- Kauffmann F, Edwards PR. A simplification of the serologic diagnosis of *Salmonella* cultures. *J Lab Clin Med.* 1947;32(5):548–53.
- Le Minor L, Popoff MY. Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*. *Int J Syst Bacteriol.* 1987;37:465–8.
- Reeves MW, Evins GM, Heiba AA, Pliakytis BD, Farmer JJ 3rd. Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. *J Clin Microbiol.* 1989;27(2):313–20.
- Liu SL, Sanderson KE. Rearrangements in the genome of the bacterium *Salmonella typhi*. *Proc Natl Acad Sci U S A.* 1995;92(4):1018–22.
- Liu GR, Liu WQ, Johnston RN, Sanderson KE, Li SX, Liu SL. Genome plasticity and ori-ter rebalancing in *Salmonella typhi*. *Mol Biol Evol.* 2006;23(2):365–71.
- Marti R, Hagens S, Loessner MJ, Klumpp J. Genome Sequence of *Salmonella bongori* Strain N268–08. *Genome Announcements.* 2013;1(6):e01018-13.
- Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, et al. *Salmonella bongori* provides insights into the evolution of the salmonellae. *PLoS Pathog.* 2011;7(8):e1002191.
- Blondel CJ, Jimenez JC, Contreras I, Santiviago CA. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics.* 2009;10:354.
- Barret M, Egan F, Fargier E, Morrissey JP, O'Gara F. Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered. *Microbiology.* 2011;157(Pt 6):1726–39.
- Barret M, Egan F, O'Gara F. Distribution and diversity of bacterial secretion systems across metagenomic datasets. *Environ Microbiol Rep.* 2013;5(1):117–26.
- Russell AB, Wexler AG, Harding BN, Whitney JC, Bohn AJ, Goo YA, Tran BQ, Barry NA, Zheng H, Peterson SB, et al. A type VI secretion-related pathway in *Bacteroidetes* mediates interbacterial antagonism. *Cell Host Microbe.* 2014;16(2):227–36.
- Liu WQ, Feng Y, Wang Y, Zou QH, Chen F, Guo JT, Peng YH, Jin Y, Li YG, Hu SN, et al. *Salmonella paratyphi C*: genetic divergence from *Salmonella choleraesuis* and pathogenic convergence with *Salmonella typhi*. *PLoS One.* 2009;4(2):e4510.
- Ochman H, Wilson AC. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol.* 1987;26(1–2):74–86.
- Doolittle RF, Feng DF, Tsang S, Cho G, Little E. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science.* 1996;271(5248):470–7.
- Feng DF, Cho G, Doolittle RF. Determining divergence times with a protein clock: update and reevaluation. *Proc Natl Acad Sci U S A.* 1997;94(24):13028–33.
- Liu SL, Hessel A, Cheng HY, Sanderson KE. The XbaI-BlnI-CeuI genomic cleavage map of *Salmonella paratyphi B*. *J Bacteriol.* 1994;176(4):1014–24.
- Liu SL, Hessel A, Sanderson KE. The XbaI-BlnI-CeuI genomic cleavage map of *Salmonella typhimurium* LT2 determined by double digestion, end labelling, and pulsed-field gel electrophoresis. *J Bacteriol.* 1993;175(13):4104–20.
- Liu SL, Sanderson KE. The chromosome of *Salmonella paratyphi a* is inverted by recombination between *rrnH* and *rrmG*. *J Bacteriol.* 1995;177(22):6585–92.
- Liu GR, Rahn A, Liu WQ, Sanderson KE, Johnston RN, Liu SL. The evolving genome of *Salmonella enterica* serovar Pullorum. *J Bacteriol.* 2002;184(10):2626–33.
- McQuiston JR, Fields PI, Tauxe RV, Logsdon JM Jr. Do *Salmonella* carry spare tyres? *Trends Microbiol.* 2008;16(4):142–8.
- Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, Seth-Smith HM, Barquist L, Stedman A, Humphrey T, et al. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc Natl Acad Sci U S A.* 2015;112(3):863–8.
- Blondel CJ, Jimenez JC, Leiva LE, Alvarez SA, Pinto BI, Contreras F, Pezoa D, Santiviago CA, Contreras I. The type VI secretion system encoded in *Salmonella* pathogenicity island 19 is required for *Salmonella enterica* serotype Gallinarum survival within infected macrophages. *Infect Immun.* 2013;81(4):1207–20.
- Blondel CJ, Yang HJ, Castro B, Chiang S, Toro CS, Zaldivar M, Contreras I, Andrews-Polymeris HL, Santiviago CA. Contribution of the type VI secretion system encoded in SPI-19 to chicken colonization by *Salmonella enterica* serotypes Gallinarum and Enteritidis. *PLoS One.* 2010;5(7):e11724.
- Liu SL, Sanderson KE. I-CeuI reveals conservation of the genome of independent strains of *Salmonella typhimurium*. *J Bacteriol.* 1995;177(11):3355–7.

47. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, et al. Toward an online repository of standard operating procedures (SOPs) for (meta) genomic annotation. *OMICS*. 2008;12(2):137–41.
48. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics*. 2009;25(17):2271–8.
49. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–66.
50. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406–25.
51. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
52. DeLong EF, Pace NR. Environmental diversity of bacteria and archaea. *Syst Biol*. 2001;50(4):470–8.
53. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res*. 2005;33(Database issue):D325–8.
54. Li J, Yao Y, Xu HH, Hao L, Deng Z, Rajakumar K, Ou HY. SecReT6: a web-based resource for type VI secretion systems found in bacteria. *Environ Microbiol*. 2015;17(7):2196–202.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

