International Journal of Infectious Diseases 49 (2016) 134–140

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Genotyping of *Salmonella* with lineage-specific genes: correlation with serotyping

Qing-Hua Zou [a], Ren-Qing Li [b], Gui-Rong Liu [c,d], Shu-Lin Liu [c,d,e,*]

[a] Department of Microbiology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China
[b] Institute of Immunology, Beijing Center for Disease Control and Prevention, Beijing, China
[c] Genomics Research Center, Harbin Medical University, 157 Baojian Road, Harbin, 150081, China
[d] HMU-UCFM Centre for Infection and Genomics, Harbin Medical University, Harbin, China
[e] Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form 25 May 2016
Accepted 27 May 2016

Corresponding Editor: Eskild Petersen, Aarhus, Denmark.

Keywords:
Salmonella
Serotyping
rpoB
MLST
Serotype-specific gene

SUMMARY

Background: The bacterial genus *Salmonella* encompasses a large number of serotypes that are genetically very similar but biologically quite different, especially in pathogenic properties and host specificity. Serotyping has been used for the classification, identification, and epidemiological investigation due to its excellent discriminating power, but it cannot distinguish the different pathogenic lineages within a polyphyletic serotype. Additionally, very few institutions have the comprehensive set of antisera for typing. Therefore various studies have been performed to explore alternative assays to differentiate *Salmonella* isolates, such as the search for genes that can be used as potential molecular substitutes for serotyping. However, the genes tested so far have often given inconsistent results.

Methods: In this study, the discriminating power of seven genes to differentiate 309 *Salmonella* strains representing 26 serotypes was evaluated and the results were compared with those of other methods.

Results: The seven newly selected genes have a good power to differentiate different serovars. The tree based on the concatenated sequences of these genes revealed phylogenetic relationships of the bacteria consistent with that of the whole genome tree.

Conclusion: Individual *Salmonella* lineages each have specific genes that can be used to differentiate *Salmonella* isolates on a phylogenetic basis.

1. Introduction

Since the divergence from a common ancestor with *Escherichia coli* more than 100 million years ago,[1] *Salmonella* have developed into more than 2600 serotypes that cause a variety of illnesses in humans and other animals. All *Salmonella* lineages are closely related, as revealed initially by DNA–DNA re-association assays[2] and then by physical mapping[3] and genomic sequencing.[4–9] Notwithstanding the observed genetic relatedness among the *Salmonella* lineages, they differ profoundly in host range and pathogenic features,[10,11] causing clinical consequences ranging from no obvious disease to mild gastroenteritis to potentially fatal systemic infections such as typhoid fever in humans. Therefore, the timely and accurate identification of *Salmonella* isolates is of great clinical significance.

Various typing methods, such as multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and phage typing, have been developed to discriminate these closely related bacteria,[12–16] of which serotyping has been the most widely used assay owing to its excellent discriminating power. However, serotyping has multiple disadvantages, such as low throughput, high expense, the need for expertise, and the requirement of a comprehensive set of antisera, which is not available to many institutions. Most importantly, however, is the fact that many *Salmonella* serotypes are polyphyletic, containing more than one phylogenetic (and pathogenic) lineage. As a result, alternative methods have been attempted for discriminating *Salmonella*, such as PFGE, ribotyping,[17] sequencing of H antigen genes (*fliC* and *fljB*),[18] and 16S–23S rRNA spacer restriction fragment length polymorphism (RFLP).[19] These methods may yield important information for epidemic analysis, but their discriminating ability is usually insufficient for accurate identification. MLST is an excellent method to discriminate strains based on their sequence

* Corresponding author. Tel.: +86 45186669150; fax: +86 45187502720.
E-mail address: slliu@ucalgary.ca (S.-L. Liu).

http://dx.doi.org/10.1016/j.ijid.2016.05.029

differences at seven house-keeping loci and has been used for the molecular typing of many bacteria.[20,21] Recently, MLST has been proposed for *Salmonella* typing due to its high resolution performance in delineating the serotypes.[12] Other genes such as *rpoB* have also demonstrated utility in *Salmonella* identification.[22,23] However, no sets of genes hitherto reported have shown a discriminating power similar to that of serotyping.

In a previous study by the present study group, 27 *Salmonella* genomes (which were completed before March 2012) were compared and the polymorphisms of a selected set of genes were analyzed, including some conserved genes (i.e., genes common to all compared genomes) and some from genomic islands, among the different serotypes.[24] It was found that 10 of the analyzed genes were polymorphic among most of the serotypes compared and thus it was considered that they may be useful in delineating *Salmonella*. In the present study, a phylogenetic analysis of seven selected genes was conducted and comparisons were made with the *rpoB* gene and the seven MLST genes to examine their discriminating power among the different *Salmonella* serotypes. A series of additional genes present only in individual serotypes were also evaluated and it was found that the combined use of the genes could significantly improve gene-based *Salmonella* typing.

2. Materials and methods

2.1. Retrieval of gene sequences

Among the 10 highly conserved genes identified in the previous study that were polymorphic across most of the 15 serotypes analyzed,[24] three, i.e., *STM2379*, *cpsG*, and *STM4261*, were not included in this current study, because *STM2379* is a pseudogene in *Salmonella* Heidelberg B182, *cpsG* has duplicates in several genomes, and *STM4261* is 16 680 bp, which is too long to be of practical use. To find *Salmonella* strains with all of the seven selected genes (*nuoG*, *srfC*, *napA*, *yhgE*, *priA*, *cpdB*, and *entF*), *rpoB*, and the seven MLST genes (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) being sequenced, a BLAST search of the nucleotide sequences of each of these genes in *Salmonella* Typhimurium LT2 was performed against the NCBI non-redundant and whole-genome shotgun contigs database, and only strains that had all of these genes sequenced were picked up for further analysis.

2.2. Construction of phylogenetic trees

Nucleotide sequences of the seven selected genes, *rpoB*, and the seven MLST genes were aligned using ClustalW in BioEdit software with default parameters. Phylogenetic trees were constructed by neighbor-joining method with MEGA software (version 5.0). The reliability of the neighbor-joining trees was estimated by bootstrap analysis with 1000 replicate datasets. The substitution model was 'maximum composite likelihood', substitutions included 'transitions + transversions', rates among sites were set as 'uniform', the pattern among lineages was set as 'same' (homogeneous), and the gaps/missing data treatment was 'complete deletion'.

The core genome tree of 32 *Salmonella* strains representing 18 serotypes (Supplementary Material Table S1) was constructed on coding sequences common to all compared genomes using the all-blast-all program in the NCBI Basic Local Alignment Search Tool (BLAST), with the criteria set at identity >75% and e-value <1e−10. For each query sequence, only the highest-scoring match above the defined identity and e-value cut-off in the 32 genomes was retained. Matched genes were then made into clusters using a Perl script. Genes present in all 32 genomes were aligned using ClustalW 2.1 and were concatenated to construct the core genome for each strain. A phylogenetic tree based on the core genome was constructed, as described above.

2.3. Lineage-specific genes

Genes present only in all analyzed strains (Supplementary Material Table S1) of one particular *Salmonella* lineage but absent in all the other lineages were considered lineage-specific. For further confirmation, all identified lineage-specific genes were searched against the NCBI non-redundant database using BLAST to exclude genes that had homologues in other *Salmonella* lineages.

2.4. Identification of clinical strains

Clinical strains of *Salmonella* were single-colony isolated and cultured in Luria-Bertani (LB) broth; DNA was extracted with a DNA extraction kit (Sangon Biotech, China). Primers for amplifying the segments of the selected genes were synthesized by Sangon Biotech, China. PCR fragments were sequenced using the Sanger AB3130 platform and the phylogenetic trees were constructed using MEGA 5.0.

3. Results

A total of 309 *Salmonella* strains representing 26 serotypes and having all the seven selected genes, *rpoB*, and the seven MLST genes being completely sequenced in the NCBI database were included in this study. In this analyzed collection, *Salmonella* Enteritidis was the predominant serotype (n = 86), followed by *Salmonella* Agona (n = 66) and *Salmonella* Montevideo (n = 37). Other serotypes were represented by 1 to 32 strains (Table 1).

3.1. The performance of the rpoB gene, the individual MLST genes, and the newly selected genes in distinguishing different Salmonella serotypes

The strains representing 20 serotypes formed serotype-specific branches on the *rpoB* gene tree, with the remaining six serotypes not well discriminated (Figure 1A). While it was not surprising to see *Salmonella* Enteritidis and *Salmonella* Gallinarum clustered together and *Salmonella* 4,[5],12:i:- being mixed with *Salmonella* Typhimurium, as judged by their phylogenetic relationships, it was unexpectedly found that strains of *Salmonella* Newport formed two separate clusters and the two strains of *Salmonella* Saintpaul did not cluster together.

Among the seven MLST genes, *sucA* distinguished strains representing 19 serotypes, with the remaining six genes, *aroC, thrA, hemD*, *dnaN*, *hisD* and *purE*, distinguishing strains representing 16, 18, 17, 16, 16, and 15 serotypes, respectively (Supplementary Material Figure S1, Table 1).

With the exception of *nuoG*, which only distinguished strains representing 13 serotypes, the other six genes selected in this study discriminated the 309 strains very clearly. The clustering correlated well with serotyping (Supplementary Material Figure S1, Table 1). Eighteen to 20 serotypes could be discriminated successfully by these genes.

Salmonella Saintpaul, *Salmonella* 4,[5],12:i:-, *Salmonella* Newport, and *Salmonella* Typhimurium were the least discriminated serotypes for *rpoB*, the MLST genes, and the seven newly selected genes, although the seven newly selected genes could clearly discriminate strains of *Salmonella* Dublin, *Salmonella* Heidelberg, and *Salmonella* Gallinarum/Pullorum (Supplementary Material Figure S1, Table 1).

3.2. Performance of the concatenated sequences of the selected genes to reflect phylogenetic relationships of the Salmonella serotypes

While examining the combined performance of the concatenated sequences, trees were constructed and it was found that the

Table 1
The discriminative power of the individual and concatenated MLST genes, the seven newly selected genes, and the *rpoB* gene in distinguishing 309 strains representing 26 serotypes[a]

Salmonella serotype	Strains	Seven MLST gene								Seven newly selected gene								*rpoB*
		Con_m	*sucA*	*aroC*	*thrA*	*hemD*	*dnaN*	*hisD*	*purE*	Con_c	*srfC*	*yhgE*	*napA*	*entF*	*priA*	*cpdB*	*nuoG*	
Enteritidis	86	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Agona	66	√	√	×	×	√	√	√	√	√	√	√	√	×	√	√	√	√
Montevideo	37	×	×	×	×	×	√	×	×	√	×	×	√	×	×	×	×	√
Newport	32	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Typhimurium	25	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Heidelberg	12	√	√	√	√	√	×	×	×	√	√	√	×	√	√	√	√	√
Gallinarum/ Pullorum	7	√	×	×	×	×	×	×	√	√	√	√	√	√	×	×	×	√
Kentucky	6	√	√	√	√	√	√	√	×	√	√	√	√	√	√	√	√	√
Paratyphi A	6	√	√	√	√	×	√	√	√	√	√	√	√	√	√	√	×	√
4,[5],12:i:-	3	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Dublin	3	√	√	×	√	×	×	×	×	√	√	√	√	√	√	√	√	√
Javiana	3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Typhi	3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	×	√
Weltevreden	2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	×	√
Cubana	2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Hadar	2	√	√	√	√	√	√	×	×	√	√	√	√	√	√	√	√	√
Schwarzengrund	2	√	√	×	√	√	√	√	√	√	√	√	√	√	√	×	×	√
Thompson	2	√	√	√	√	√	×	√	×	√	√	√	√	√	√	√	×	√
Virchow	2	√	√	√	√	√	×	√	√	√	√	√	×	√	√	√	×	√
Saintpaul	2	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Choleraesuis	1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Infantis	1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Paratyphi B	1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Paratyphi C	1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
62:z4,z23:-	1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Houtenae	1	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Discriminated		20	19	16	18	17	16	16	15	21	20	20	19	19	19	18	13	21

MLST, multilocus sequence typing; Con_c: concatenated seven newly selected genes; Con_m: concatenated MLST genes.
[a] √ all strains in the serotype formed one distinct cluster from other serotypes; × strains in one serotype were not clustered together.

rpoB gene and the seven newly selected genes could delineate strains of 21 serotypes with better resolution than the MLST genes, which could delineate strains in 20 serotypes (Figure 1, Table 1). Strains in five serotypes – *Salmonella* Saintpaul, *Salmonella* 4,[5],12:i:-, *Salmonella* Typhimurium, *Salmonella* Enteritidis, and *Salmonella* Newport – were not delineated into distinct clusters by any of the three trees, a fact that might reflect the phylogenetic complexity of bacteria in these serotypes rather than insufficient resolution power of any the three trees. On the tree of the seven newly selected genes, the 85 *Salmonella* Enteritidis strains formed a single cluster, with only one strain, SARB17, not clustered with it; conversely, on the *rpoB* gene tree and the MLST genes tree, the 85 *Salmonella* Enteritidis strains were mixed with the *Salmonella* Gallinarum/Pullorum strains (Figure 1), which does however reflect differences in resolution power among the three trees, as *Salmonella* Enteritidis and *Salmonella* Gallinarum/Pullorum, although very closely related, are unambiguously distinct lineages.[25] Additionally, the seven newly selected genes also had a better resolution for *Salmonella* Gallinarum and *Salmonella* Montevideo than the *rpoB* gene and the MLST genes.

The trees were then evaluated for their reliability to reveal the phylogenetic relationships among the bacteria by comparing them with the tree constructed on the core genome (genes common to all compared strains). A total of 2304 core gene clusters were found in the 32 *Salmonella* genomes. It was found that the topology of the tree based on the seven newly selected genes was much more similar to the core genome tree than the *rpoB* and the MLST gene trees. In the tree of the seven newly selected genes, the majority of the serotypes showed similar evolutionary relationships to that of the core genome tree; for example, *Salmonella* Typhimurium, *Salmonella* Heidelberg, *Salmonella* Dublin, *Salmonella* Enteritidis, *Salmonella* Gallinarum/Pullorum, *Salmonella* Paratyphi C, *Salmonella* Choleraesuis, *Salmonella* Agona, *Salmonella* Typhi, and *Salmonella* Paratyphi A (Figure 1C, D). However, in the other two trees, the relationships were thoroughly confused. For example, the closely related *Salmonella* Paratyphi A and *Salmonella* Typhi were clustered as close neighbors in the tree of the seven newly selected genes in this study (Figure 1C), consistent with whole genome analysis (Figure 1D);[9] however, they were split up by other serotypes in the MLST genes tree (Figure 1B). Similar situations were seen with *Salmonella* Gallinarum/Pullorum, *Salmonella* Enteritidis, and *Salmonella* Dublin in comparison with *Salmonella* Agona and the *Salmonella* Heidelberg/*Salmonella* Typhimurium branches (Figure 1), in which only the tree of the seven newly selected genes was consistent with whole genome analysis.

3.3. Concatenated partial sequences of the seven newly selected genes: similar discriminating power to the concatenated complete genes and capable of discriminating clinical strains

To save costs on the assays, it was attempted to determine whether parts instead of the full length of these genes could be used to discriminate the *Salmonella* lineages. For the seven MLST genes, segments of 399–501 bp in size were used, as indicated in the MLST databases (http://mlst.ucc.ie/mlst/dbs/Senterica/documents/primersEnterica_html). The locations of the fragments on the chromosome can be found in Table 2. Since the selection of the seven genes in the previous study was based on the polymorphisms of nucleotides to each serotype and it was found that the polymorphisms were distributed randomly in the genes, the seven genes were arbitrarily divided into segments consecutively with lengths of about 400–600 bp. The phylogenetic trees were constructed based on each of the segments and their discriminating powers were compared, which differed among the segments (data not shown). The one that had the highest discriminating power was selected for further analyses (see Table 2 for locations of the fragments on the chromosome). For

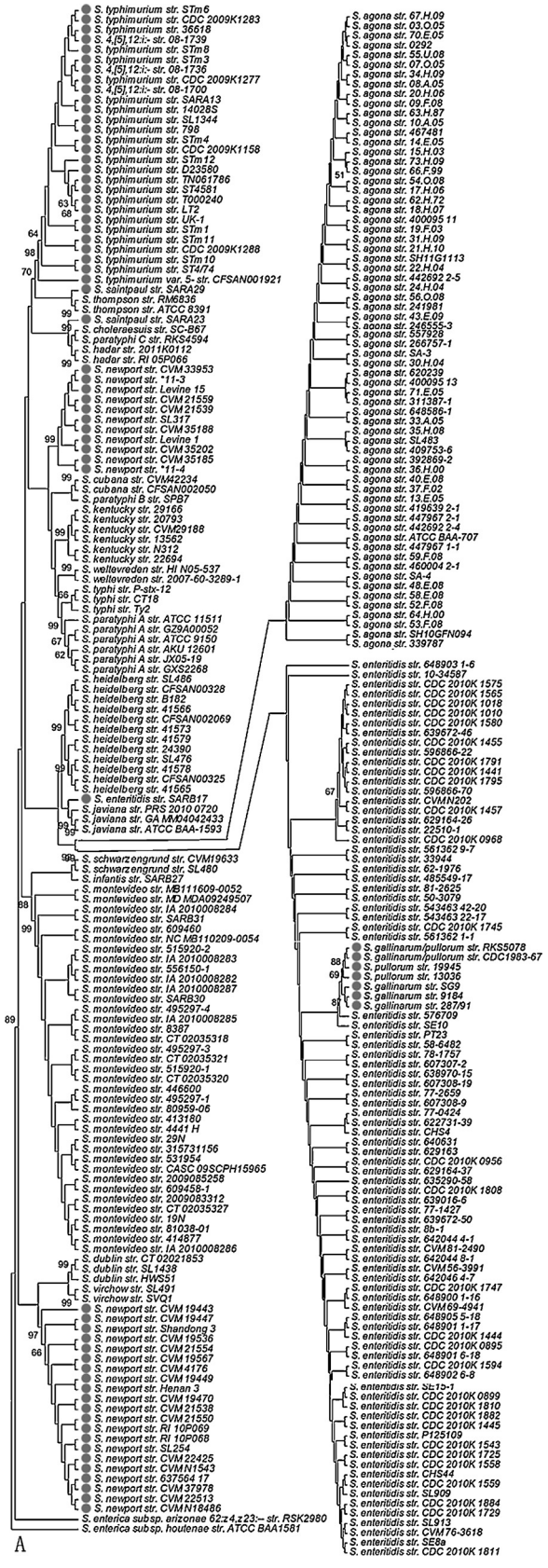

Figure 1. Phylogenetic trees of the *Salmonella* strains based on the *rpoB* gene (A), the concatenated MLST genes (B), the concatenated seven newly selected genes (C), and the core genome tree (D). Strains marked with a circle failed to cluster with the strains in corresponding serotypes.

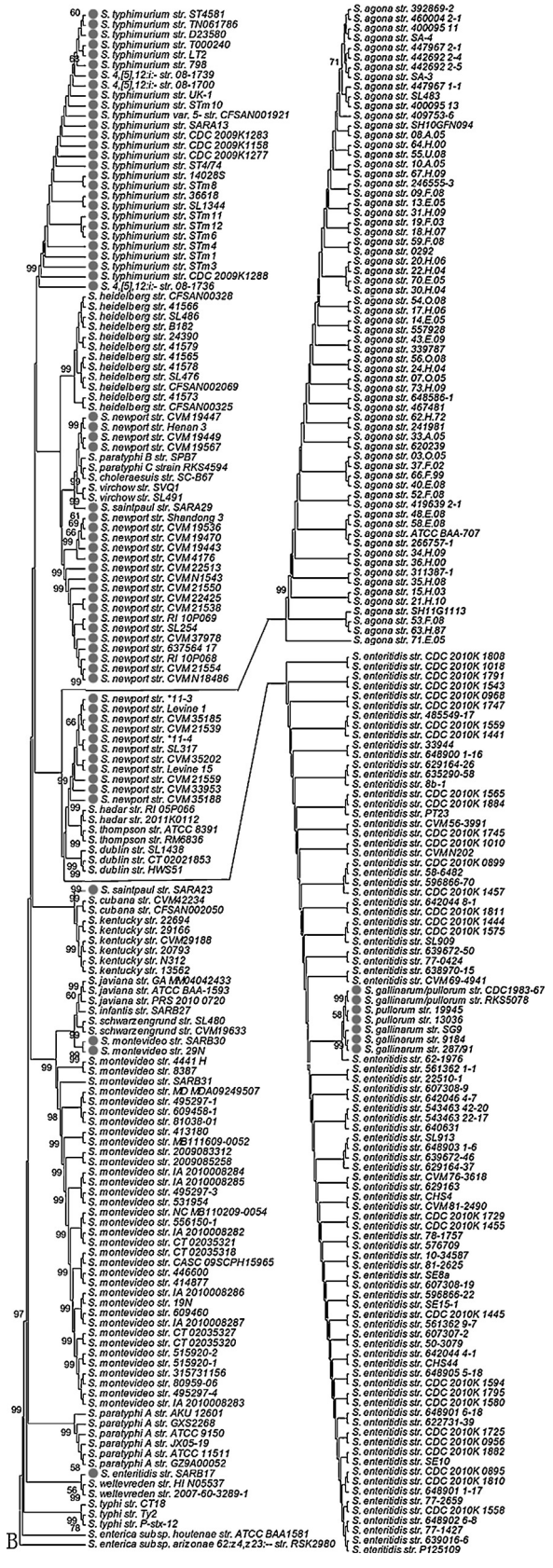

Figure 1. (*Continued*).

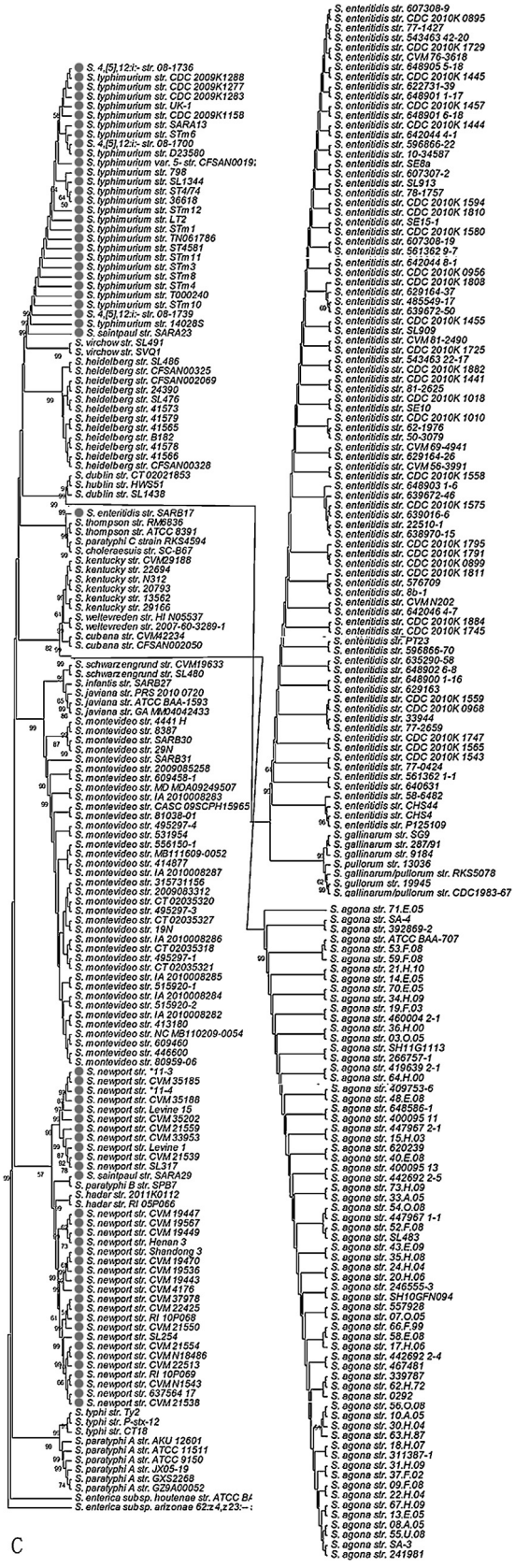

Figure 1. (*Continued*).

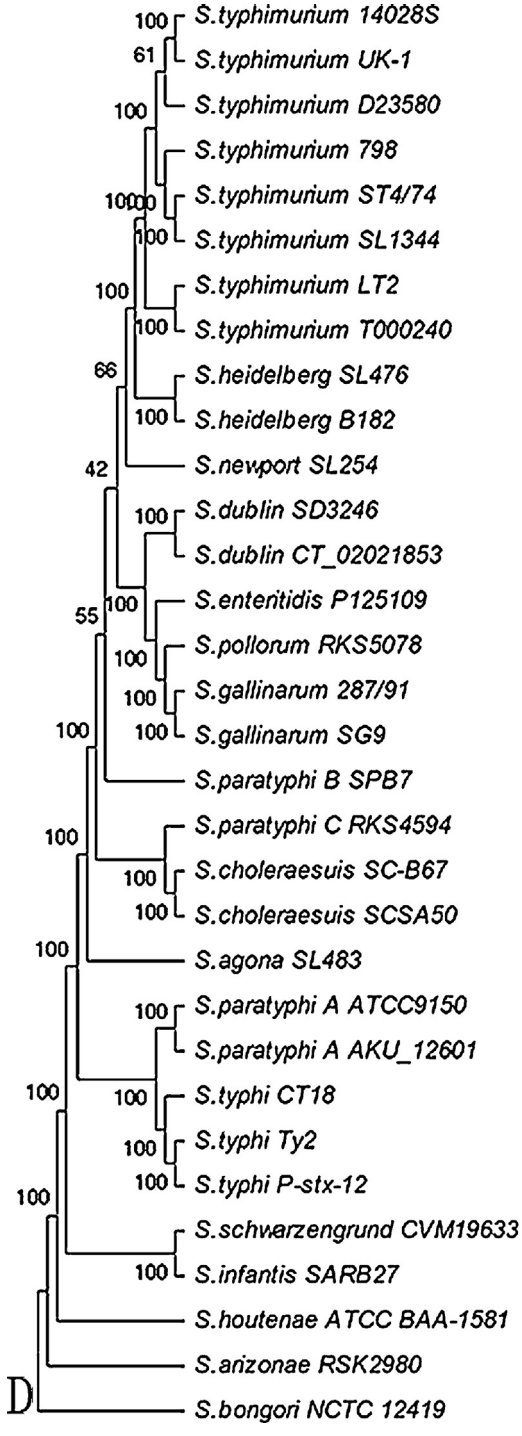

Figure 1. (*Continued*).

the seven newly selected genes, the concatenated segments had similar discriminating power as the complete genes, with rare exceptions in *Salmonella* Enteritidis (strain 58–6482) and *Salmonella* Montevideo (strains 4441H, 8387, SARB30 and 29N; Supplementary Material Figure S2A). For the seven MLST genes, the 32 *Salmonella* Newport strains formed two clusters (Supplementary Material Figure S2B).

The discriminating ability of the fragments of the seven newly selected genes and the seven MLST genes for clinical strains were then compared. The fragments of the 14 genes in five clinical strains were sequenced and compared with other strains. All five strains fell into corresponding places on the tree of the

Table 2
Genomic distribution of the seven newly selected genes and the seven MLST genes along with the segments used for phylogenetic analysis

Gene	Locus_tag in LT2	Size (bp)	Gene position in LT2	Fragment position in LT2
Seven newly selected genes				
nuoG	STM2323	2733	2431894–2434626	2431909–2432559
srfC	STM1595	2145	1684073–1686217	1685100–1685706
napA	STM2259	2487	2356418–2358904	2357020–2357673
yhgE	STM3499	1710	3654477–3656186	3654974–3655637
priA	STM4095	2199	4303361–4305559	4303825–4304472
cpdB	STM4403	1944	4639560–4641503	4639913–4640579
entF	STM0588	3885	645462–649346	647403–648102
Seven MLST genes				
aroC	STM2384	1086	2494541–2495626	2494719–2495369
dnaN	STM3837	1101	4042519–4043619	4042852–4043502
hemD	STM3937	741	4144312–4145052	4144333–4144953
hisD	STM2072	1305	2150617–2151921	2151127–2151586
purE	STM0534	510	597116–597625	597163–597617
sucA	STM0736	2802	801745–804546	802330–802922
thrA	STM0002	2463	337–2799	889–1511

MLST, multilocus sequence typing.

concatenated fragments of the seven newly selected genes (Supplementary Material Figure S2A), while in the MLST fragment tree, the *Salmonella* Enteritidis strain SE154 failed to cluster with other *Salmonella* Enteritidis strains and *Salmonella* Typhimurium strain DT104 showed a closer relationship with *Salmonella* Saintpaul strain SARA29 than with other *Salmonella* Typhimurium strains (Supplementary Material Figure S2B).

3.4. Serotype-specific genes as potential markers for typing

Although the seven newly selected genes had great discriminating power for *Salmonella* strains, as shown above, strains of some serotypes still could not be well differentiated. It was then sought to determine whether there might be additional serotype-specific genes for use as potential markers of certain serotypes. By comprehensive analyses, a series of serotype-specific genes was identified in 13 serotypes (Supplementary Material Table S2). These genes were present in all of the analyzed strains of one given serotype, but not in any of the other serotypes in the NCBI non-redundant database. One such serotype-specific gene was identified in *Salmonella* Typhimurium and as many as 146 such serotype-specific genes in *Salmonella* Paratyphi B. Most of the serotype-specific genes are contiguous in the genome and are phage-related. For example, in *Salmonella* Typhi, from t1351 to t1397, there are 39 serotype-specific genes encoding bacteriophage-related proteins or hypothetical proteins with unknown functions. There are also some genes with evident functions like O-antigen formation, such as the SNSL254_A2006, SNSL254_A2266, and SNSL254_A2264 in *Salmonella* Newport, which encode O-acetyl transferase, O-antigen polymerase, and glycosyl transferase, respectively. Most of the serotype-specific genes that were identified in this study encode hypothetical proteins with unknown functions. Further studies on these genes may lead to new insights into the evolution of individual serotypes, especially their roles in pathogenesis.

4. Discussion

Various molecular methods have been applied in an attempt to replace the conventional serotyping method for *Salmonella*, but their discriminative ability is usually insufficient to distinguish different *Salmonella* serotypes. In this work, seven newly selected genes were assessed by comparison with the *rpoB* gene and the seven MLST genes for the differentiation of *Salmonella* lineages. The *rpoB* gene encodes the β subunit of DNA-dependent RNA polymerase. Multiple studies have shown that *rpoB*-based analysis could effectively overcome the intrinsic limitations of the intra-species heterogeneity of 16S rRNA[26,27] and could clearly differentiate among *Legionella* and *Salmonella* species.[23,28,29] In this study, the concatenated sequences of the seven newly selected genes showed higher resolution than *rpoB* and the MLST genes, especially for *Salmonella* Enteritidis, *Salmonella* Gallinarum/Pullorum, and *Salmonella* Newport. *Salmonella* Newport is known to be polyphyletic,[30] including at least two distinct sub-lineages; this was resolved in the present study only on the tree of the seven newly selected genes. These results demonstrate that the seven newly selected genes should be good candidates for *Salmonella* typing and their high discriminating power may be attributed to the polymorphic sites indicated in the previous study.[24]

Of great significance, among the three phylogenetic trees, only the one based on the seven newly selected genes showed similar evolutionary relationships of the bacteria with that revealed by the core genome tree. Strains of some serotypes such as *Salmonella* 4,[5],12:i:- could not be discriminated from another serotype (here, *Salmonella* Typhimurium) by any of the three phylogenetic trees, which is consistent with reports by other researchers using MLST, PFGE, and amplified fragment length polymorphism.[31] In the case of *Salmonella* 4,[5],12:i:-, an explanation could be that 4,[5],12:i:– may be polyphyletic and some of its members might be closely related to *Salmonella* Typhimurium but some others may be more distantly related.

Bacteria constantly need to adapt to the new host and external environment, often by obtaining novel genetic traits through accumulating mutations or acquiring laterally transferred genes (LTG). As the accumulation of mutations in core genes requires a long evolutionary time and the acquisition of LTG may take place at any time point, the use of core genes only may not precisely resolve newly developed sub-lineages of bacteria that have just acquired new LTGs to become a specific pathogen. For this reason, genes that are specific to only a single serotype or even a subset of a single serotype (especially while assuming that a given serotype might be polyphyletic) were sought. Such genes were identified in some serotypes, e.g., SNSL254_A2266 in *Salmonella* Newport, which encodes O-antigen polymerase. Many serotype-specific genes identified in this study were phage-related, reiterating the important roles of phage-mediated LTGs in bacterial evolution and their value in the identification of the *Salmonella* pathogens.

On the whole, in the current study, both the seven individual newly selected genes and their concatenated sequence showed a high power to discriminate different *Salmonella* serotypes. With the rapid developments made in genome sequencing, more and more *Salmonella* strains that have had all of the seven newly selected genes sequenced should be deposited in the database; the discriminatory power of the seven genes can then be tested further

with these strains. At the same time, since most of the seven newly selected genes show a good power to discriminate different serotypes, the discriminatory power of different combinations of these genes could be explored further.

In conclusion, seven newly selected genes that could differentiate 309 *Salmonella* strains into distinct clusters were identified in this study. Additionally, a set of serotype-specific genes was found, the combined use of which with the seven newly selected genes could type *Salmonella* isolates with high discriminating power.

Conflict of interest: None of the authors declare any conflict of interest regarding this manuscript.

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China to QHZ (NSFC 81572041, 31100134) and SLL (NSFC81030029, 81271786). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijid.2016.05.029.

References

1. 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**:470–7.
2. Crosa JH, Brenner DJ, Ewing WH, Falkow S. Molecular relationships among the Salmonelleae. *J Bacteriol* 1973;**115**:307–15.
3. Liu SL, Schryvers AB, Sanderson KE, Johnston RN. Bacterial phylogenetic clusters revealed by genome structure. *J Bacteriol* 1999;**181**:6747–55.
4. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 2001;**413**:848–52.
5. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 2001;**413**:852–6.
6. Deng W, Liou SR, Plunkett G, Mayhew GF, Rose DJ, Burland V, et al. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J Bacteriol* 2003;**185**:2330–7.
7. McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S, Sabo A, et al. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nat Genet* 2004;**36**:1268–74.
8. Chiu CH, Tang P, Chu C, Hu S, Bao Q, Yu J, et al. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic Acids Res* 2005;**33**:1690–8.
9. Liu WQ, Feng Y, Wang Y, Zou QH, Chen F, Guo JT, et al. *Salmonella* Paratyphi C: genetic divergence from *Salmonella* Choleraesuis and pathogenic convergence with *Salmonella* Typhi. *PLoS One* 2009;**4**:e4510.
10. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid fever. *N Engl J Med* 2002;**347**:1770–82.
11. Barrow PA, Huggins MB, Lovell MA. Host specificity of Salmonella infection in chickens and mice is expressed in vivo primarily at the level of the reticuloendothelial system. *Infect Immun* 1994;**62**:4602–10.
12. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog* 2012;**8**:e1002776.
13. Ramisse V, Houssu P, Hernandez E, Denoeud F, Hilaire V, Lisanti O, et al. Variable number of tandem repeats in *Salmonella enterica* subsp. *enterica* for typing purposes. *J Clin Microbiol* 2004;**42**:5722–30.
14. Liu WB, Liu B, Zhu XN, Yu SJ, Shi XM. Diversity of Salmonella isolates using serotyping and multilocus sequence typing. *Food Microbiol* 2011;**28**:1182–9.
15. Harbottle H, White DG, McDermott PF, Walker RD, Zhao S. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. *J Clin Microbiol* 2006;**44**:2449–57.
16. Miller T, Braun PG, Fehlhaber K, Prager R, Pfeifer Y, Rabsch W. Typing of *Salmonella enterica* serovar Infantis isolates from 51 outbreaks in Germany between 1974 and 2009 by a novel phage-typing scheme. *Epidemiol Infect* 2014;**142**:75–83.
17. Oscar TP. Identification and characterization of Salmonella isolates by automated ribotyping. *J Food Prot* 1998;**61**:519–24.
18. Herrera-Leon S, Ramiro R, Arroyo M, Diez R, Usera MA, Echeita MA. Blind comparison of traditional serotyping with three multiplex PCRs for the identification of Salmonella serotypes. *Res Microbiol* 2007;**158**:122–7.
19. Christensen H, Moller PL, Vogensen FK, Olsen JE. 16S to 23S rRNA spacer fragment length polymorphism of *Salmonella enterica* at subspecies and serotype levels. *J Appl Microbiol* 2000;**89**:130–6.
20. Aanensen DM, Spratt BG. The multilocus sequence typing network: mlst.net. *Nucleic Acids Res* 2005;**33**(Web Server issue):W728–33.
21. Kotetishvili M, Stine OC, Kreger A, Morris Jr JG, Sulakvelidze A. Multilocus sequence typing for characterization of clinical and environmental Salmonella strains. *J Clin Microbiol* 2002;**40**:1626–35.
22. Kwon HJ, Park KY, Kim SJ, Yoo HS. Application of nucleotide sequence of RNA polymerase beta-subunit gene (rpoB) to molecular differentiation of serovars of *Salmonella enterica* subsp. *enterica*. *Vet Microbiol* 2001;**82**:121–9.
23. Seong WJ, Kwon HJ, Kim TE, Lee DY, Park MS, Kim JH. Molecular serotyping of *Salmonella enterica* by complete *rpoB* gene sequencing. *J Microbiol* 2012;**50**:962–9.
24. Zou QH, Li RQ, Wang YJ, Liu SL. Identification of genes to differentiate closely related Salmonella lineages. *PLoS One* 2013;**8**:e55988.
25. 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.
26. Mollet C, Drancourt M, Raoult D. *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* 1997;**26**:1005–11.
27. Dahllof I, Baillie H, Kjelleberg S. *rpoB*-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol* 2000;**66**:3376–80.
28. Ko KS, Lee HK, Park MY, Lee KH, Yun YJ, Woo SY, et al. Application of RNA polymerase beta-subunit gene (*rpoB*) sequences for the molecular differentiation of *Legionella* species. *J Clin Microbiol* 2002;**40**:2653–8.
29. Fresno M, Barreto M, Gutierrez S, Dougnac C, Abalos P, Retamal P. Serotype-associated polymorphisms in a partial *rpoB* gene sequence of *Salmonella enterica*. *Can J Microbiol* 2014;**60**:177–81.
30. Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, et al. Evolution and population structure of *Salmonella enterica* serovar Newport. *J Bacteriol* 2010;**192**:6465–76.
31. Ben Aissa R, Al-Gallas N, Troudi H, Belhadj N, Belhadj A. Trends in *Salmonella enterica* serotypes isolated from human, food, animal, and environment in Tunisia 1994-2004. *J Infect* 2007;**55**:324–39.