Complete genome sequence of *Salmonella enterica* subspecies *arizonae* str. RKS2983

Chun-Xiao Wang¹, Song-Ling Zhu¹, Xiao-Yu Wang¹, Ye Feng², Bailiang Li¹, Yong-Guo Li³, Randal N Johnston⁴, Gui-Rong Liu¹, Jin Zhou⁵* and Shu-Lin Liu¹,³,⁶*

**Abstract**

*Salmonella arizonae* (also called *Salmonella* subgroup IIIa) is a Gram-negative, non-spore-forming, motile, rod-shaped, facultatively anaerobic bacterium. *S. arizonae* strain RKS2983 was isolated from a human in California, USA. *S. arizonae* lies somewhere between *Salmonella* subgroups I (human pathogens) and V (also called *S. bongori*; usually non-pathogenic to humans) and so is an ideal model organism for studies of bacterial evolution from non-human pathogen to human pathogens. We hence sequenced the genome of RKS2983 for clues of genomic events that might have led to the divergence and speciation of *Salmonella* into distinct lineages with diverse host ranges and pathogenic features. The 4,574,836 bp complete genome contains 4,203 protein-coding genes, 82 tRNA genes and 7 rRNA operons. This genome contains several characteristics not reported to date in *Salmonella* subgroup I or V and may provide information about the genetic divergence of *Salmonella* pathogens.

**Keywords:** *S. enterica* subspecies *arizonae* RKS2983, Facultative anaerobe, Genomic evolution, Host-adapted, *Salmonella* pathogenicity islands

**Introduction**

*Salmonella* are Gram-negative facultative anaerobic bacteria of the family *Enterobacteriaceae* inhabiting the gastrointestinal tract of a wide variety of animals. There are currently over 2,600 serotypes (also called serovars) documented in the genus *Salmonella*. By chromosomal DNA hybridization experiments and MLSE, *Salmonella* currently are classified into two species, *S. enterica* and *S. bongori* (formerly subgroup V). The species *S. enterica* is further divided into six subspecies, including *S. enterica* subspecies *enterica*, *S. enterica* subspecies *salamae*, *S. enterica* subspecies *arizonae*, *S. enterica* subspecies *diarizonae*, *S. enterica* subspecies *houtenae*, and *S. enterica* subspecies *indica*, corresponding to the former subgroups I, II, IIIa, IIIb, IV and VI, respectively. Additionally, subgroup VII was described by Boyd et al. [1,2]. *Salmonella* taxonomy is a dynamic field of research and many issues remain unsolved, especially regarding species definition [3-5]. To avoid confusions, therefore, we use the traditional *Salmonella* classification system and the terms subgroup and serotype rather than subspecies or serovar (see more detailed explanation in [5]). Most of *Salmonella* infections in warm-blooded animals are caused by *Salmonella* subgroup I serotypes, and non-subgroup I serotypes are typically associated with cold-blooded vertebrates and rarely colonize the intestines of warm-blooded animals.

*Salmonella* evolved from a common ancestor with *Escherichia coli* about 120–150 million years ago [6,7]. During the evolutionary process, several key genomic events might have led bacteria to diverge, such as gene mutation and gene acquisition or loss [8]. Importantly, numerous lines of evidence have indicated that gene acquisition and loss are the major force driving the
evolution of virulence in *Salmonella* [9]. In fact, it has been postulated that the evolution of *Salmonella*-specific virulence can be divided into three phases. The first phase is the split of *Salmonella* and *E. coli* by the *Salmonella* acquisition of *Salmonella* pathogenicity island 1, which is present in all lineages of *Salmonella* but absent from *E. coli*. SPI-1 encodes virulence factors that strengthen the infection of *Salmonella* serotypes by different mechanisms, including the invasiveness of the bacteria into intestinal epithelial cells [10], induction of neutrophil recruitment, and secretion of intestinal fluid [11-13]. The second phase is the divergence of *Salmonella* into *S. bongori* and *S. enterica*; this pathogenic lineage acquired SPI-2 [14-17], which contains genes encoding a type III secretion system that is required for survival in macrophages [18]. The third phase is the adaptation of *Salmonella* subgroup I to warm-blooded animals, but the genetic events involved remain unknown.

Genome sequencing efforts in *Salmonella* have mostly focused on *Salmonella* subgroup I serotypes, largely due to their pathogenicity in humans. In this study, we sequenced the genome of a strain from *Salmonella* subgroup IIIa (also known as *Salmonella arizonae*), which lies somewhere between *Salmonella* subgroups I and V in evolution. Based on the important evolutionary position of *Salmonella* subgroup IIIa, we anticipated that its genomic comparisons with other *Salmonella* subgroups, especially subgroups I and V, may provide novel insights into the evolutionary transition of *Salmonella* adaptation from cold- to warm-blooded hosts.

### Organism information

**Classification and features**

*S. arizonae* is classified to Class *Gammaproteobacteria*, Order *Enterobacteriales*, Family *Enterobacteriaceae* and Subfamily *Salmonellaceae*.

| MIGS ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
| RKS2983 | Strain classification | Domain *Bacteria* | TAS [34] |
|         |           | Phylum *Proteobacteria* | TAS [35] |
|         |           | Class *Gammaproteobacteria* | TAS [35,36] |
|         |           | Order *Enterobacteriales* | TAS [37-39] |
|         |           | Family *Enterobacteriaceae* | TAS [39,40] |
|         |           | Genus *Salmonella* | TAS [40-41] |
|         |           | Species *Salmonella enterica* | TAS [41,42] |
|         |           | Subspecies *Salmonella enterica* subsp. *arizonae* | TAS [42] |
|         |           | Strain RKS2983 | TAS [42] |
|         |           | Serovar 62:z36:- | TAS [42] |

**Gram stain**: Negative

**Cell shape**: Rod-shaped

**Motility**: Motile

**Sporulation**: Non-sporulating

**Temperature range**: 35°C–37°C

**pH**: 7.2–7.6

**Carbon source**: Glucose

**Habitat**: Human

**Salinity**: Medium

**Oxygen requirement**: Facultative anaerobes

**Biotic relationship**: Endophyte

**Pathogenicity**: Pathogenic

**Geographic location**: California, USA

**Sample collection time**: 1985

**Latitude**: Not report

**Longitude**: Not report

**Depth**: Not report

**Altitude**: Not report

### Project information

| MIGS ID | Property | Term |
|---------|----------|------|
| RKS2983 | Finishing quality | Finished |
|         | Libraries used | Illumina Paired-End library and SOLID mate_pair library (2 x 50 bp) |
|         | Sequencing platforms | Illumina HiSeq 2000 and SOLID 3.0 |
|         | Fold coverage | 100 x |
|         | Assemblers | SOAPdenovo v1.05 |
|         | Gene calling method | Glimmer software that used in the RAST pipeline |
|         | Genbank ID | CP006693.1 |
|         | Genbank date of release | September 22, 2014 |
|         | GOLD ID | GJ686507741 |
|         | BIOPROJECT | PRJNA215272 |
|         | Source material identifier | CDC 409–85 |

**Project relevance**: Evolution in bacteria
Genus *Salmonella* (Table 1). *S. arizonae* was first described in 1939 by the name *Salmonella dar es salaam* and was categorized as *Salmonella* subgroup IIIa, later named *S. arizonae* [19]. *S. arizonae* is a rare cause of human infection and is naturally found in reptiles.

We obtained RKS2983 from the *Salmonella* Genetic Stock Center (SGSC) as one of the strains in the set of *Salmonella* Reference Collection C strain (SARC6) [2]; it was initially isolated from a human of California in 1985. It is, like other *Salmonella* bacteria, Gram-negative with diameters around 0.7 to 1.5 μm and lengths of 2 to 5 μm, facultatively anaerobic, non-spore-forming, and predominantly motile with peritrichous flagella. The bacteria were grown at 37°C in Luria broth with pH of 7.2-7.6. Detailed information on the strain can be found at SGSC [20].

**Genome sequencing information**

**Genome project history**

This complete genome project was deposited in the Genomes On-Line Database (GOLD) and the complete genome sequence of strain RKS2983 was deposited at DDBJ/EMBL/GenBank under the accession CP006693.1. Table 2 presents the project information and its association with MIGS version 2.0 [21].

**Growth conditions and DNA isolation**

*S. arizonae* RKS2983 was cultured to mid-logarithmic phase in 50 ml of Luria Broth on a gyratory shaker at

---

**Table 2 Nucleotide content and gene count levels of the genome**

| Attribute               | Value     | % of totalb |
|-------------------------|-----------|-------------|
| Genome Size (bp)        | 4,574,836 |             |
| G + C content (bp)      | 2,356,040 | 51.50       |
| Coding region (bp)      | 3,924,843 | 85.79       |
| Total genesb            | 4,390     |             |
| rRNA genes              | 22        | 0.50        |
| tRNA genes              | 82        | 1.87        |
| Protein-coding genes    | 4,203     | 95.70       |
| Pseudogenes             | 98        | 2.23        |
| Frameshifted Genes      | 78        | 1.78        |
| Genes assigned to COGs  | 3,383     | 77.06       |

b) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.
37°C. DNA was isolated from the cells using a CTAB bacterial genomic DNA isolation method [22].

**Genome sequencing and assembly**

The genome of *S. arizonae* RKS2983 was sequenced by use of two sequencing platforms, SOLiD 3.0 and Illumina HiSeq 2000. First, genomic DNA was sequenced with the Illumina sequencing platform by the paired-end strategy (2×100 bp) and the details of library construction and sequencing can be found at the Illumina web site [23]. The sequence data from Illumina HiSeq 2000 were assembled by SOAPdenovo v1.05 and the assembly contained 103 scaffolds with a genome size of 4.5 Mb. Then, the genomic DNA was sheared into 3 kb fragments by the Hydrosheer instrument and was sequenced on a SOLiD sequencer by the mate-pair strategy (2 × 50 bp) according to the manual for the instrument (Applied Biosystems). The two sets of data from different methods were assembled by the Velvet v1.2.09 software. The final assembly contained 20 scaffolds. Gaps between contigs were closed by PCR amplification using ABI3730 sequencer.

**Genome annotation**

Genes were predicted by Rapid Annotation using Subsystem Technology [24] with Glimmer 3 [25] followed by manual curation. The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information non-redundant database and Clusters of Orthologous Groups databases. These data sources were combined to assert a product description for each predicted protein. Then, we compared them with the annotated genes from four available *Salmonella* genomes, including *S. typhi* Ty2, *S. typhimurium* LT2 (AE006468) [26], *S. arizonae* RKS2980 (CP000880) [27] and *S. bongori* NCTC12419 (NC_015761) [17]. Non-coding genes and miscellaneous features were predicted using tRNAscanSE [28], RNAMMer [29], Rfam [30] and TMHMM [31].

**Genome properties**

The genome (Figure 1) consists of a chromosome of 4,574,836 bp (51.5% GC content) with 4,390 genes predicted, including 4,203 protein-coding genes, 22 rRNA genes, 82 tRNA genes and 98 pseudogenes. The properties and the statistics of the genome are summarized in Tables 3 and 4.

**Insights from the genome sequence**

We first looked into the genetic relatedness of *Salmonella* and *E. coli*. For this, we concatenated the 945 genes common to the 25 sequenced strains analyzed in this study and conducted comparisons using BLAST with the parameters set at >70% DNA identity and >0.7 gene length ratio to categorize genes into common genes. The multiple sequence alignment program MAFFT program [32] was used to align the gene sequences of the *Salmonella* and *E. coli* strains. Phylogenetic trees were constructed with the aligned gene sequences using the Neighbor-Joining methods based on 1,000 randomly selected bootstrap replicates by MEGA 4.0 software [33]. The tree showed that *S. bongori* positioned between *Salmonella* subgroup I and *E. coli*, *S. arizonae* RKS2983 positioned between *Salmonella* subgroup I and *S. bongori*, and all *Salmonella* subgroup I strains were clustered together (Figure 2).

| Code | Value | % of total | Description                                      |
|------|-------|------------|-------------------------------------------------|
| J    | 168   | 3.83       | Translation, ribosomal structure and biogenesis |
| A    | 1     | 0.02       | RNA processing and modification                 |
| K    | 0     | 0.00       | Transcription                                   |
| L    | 216   | 4.92       | Replication, recombination and repair           |
| B    | 0     | 0.00       | Chromatin structure and dynamics                |
| D    | 32    | 0.73       | Cell cycle control, mitosis and meiosis         |
| Y    | 0     | 0.00       | Nuclear structure                               |
| V    | 44    | 1.00       | Defense mechanisms                              |
| T    | 106   | 2.41       | Signal transduction mechanisms                  |
| M    | 223   | 5.08       | Cell wall/membrane biogenesis                   |
| N    | 89    | 2.03       | Cell motility                                   |
| Z    | 0     | 0.00       | Cytoskeleton                                    |
| W    | 0     | 0.00       | Extracellular structures                        |
| U    | 44    | 1.00       | Intracellular trafficking and secretion         |
| O    | 137   | 3.12       | Posttranslational modification, protein turnover, chaperones |
| C    | 240   | 5.47       | Energy production and conversion                |
| G    | 307   | 6.99       | Carbohydrate transport and metabolism           |
| E    | 314   | 7.15       | Amino acid transport and metabolism             |
| F    | 76    | 1.73       | Nucleotide transport and metabolism             |
| H    | 142   | 3.23       | Coenzyme transport and metabolism               |
| I    | 88    | 2.00       | Lipid transport and metabolism                  |
| P    | 182   | 4.15       | Inorganic ion transport and metabolism          |
| Q    | 53    | 1.21       | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 311   | 7.08       | General function prediction only                |
| S    | 348   | 7.93       | Function unknown                                |
|     | 1007  | 22.94      | Not in COGs                                    |

a.) The total is based on the total number of protein coding genes in the annotated genome.
Figure 2. Phylogenetic tree highlighting the position of *S. arizonae* RKS2983 (shown in bold) relative to strains of other *Salmonella* lineages. The corresponding GenBank accession numbers are displayed in parentheses. The tree was built based on the comparison of concatenated nucleotide sequences of 945 conserved genes in all strains. Individual orthologous sequences were aligned by the MAFFT program [32] and concatenated. The phylogenetic tree was constructed by using the MEGA 4.0 software [33] with Neighbor-Joining method. The bootstrap values are shown at branch points.

Figure 3. Venn diagram showing the core genes in *S. arizonae* RKS2983, *S. bongori* NCTC 12419 and *S. typhimurium* LT2. The core genes conducted using BLAST with the parameters set at “>70% DNA identity and >0.7 gene length ratio.”
The core gene data of *S. arizonae* RKS2983, *S. bongori* NCTC 12419 and *S. typhimurium* LT2 (representing *Salmonella* subgroup I) were presented in Figure 3. There are 2823 genes common to all three genomes and 926 genes specific in RKS2983. SPI-2 is in the set of 516 genes common to RKS2983 and LT2 and is absent in *S. bongori* NCTC 12419. As many as 1017 genes are in LT2 but not in the other two genomes; we postulate that some of these genes may be associated with virulence to warm-blooded hosts.

We compared these genomes for presence or absence of *Salmonella* pathogenicity islands (SPIs) and found that *S. arizonae* RKS2983 shared some of the SPIs with *S. bongori* NCTC 12419 and others with *S. typhimurium* LT2 or *S. typhi* Ty2 (Table 5), providing opportunities of evolutionary studies about acquisition of SPIs during transition of *Salmonella* from cold- to warm-blooded animal pathogens.

### Table 5 Distribution of known SPIs in four representation genomes of *Salmonella* genus

| Genomic Island | *S. bongori* 12419 | *S. arizonae* RKS2983 | *S. typhimurium* LT2 | *S. typhi* Ty2 |
|---------------|-----------------|-----------------|-----------------|----------------|
| SPI-1         | +               | +               | +               | +              |
| SPI-2         | -               | +               | +               | +              |
| SPI-3         | +               | -               | +               | +              |
| SPI-4         | +               | +               | +               | +              |
| SPI-5         | +               | +               | +               | +              |
| SPI-6         | -               | -               | +               | +              |
| SPI-7         | -               | -               | -               | +              |
| SPI-8         | -               | -               | -               | +              |
| SPI-9         | +               | +               | +               | +              |
| SPI-10        | -               | -               | -               | +              |
| SPI-11        | +               | +               | +               | -              |
| SPI-12        | -               | -               | +               | +              |
| SPI-13        | -               | +               | -               | -              |
| SPI-14        | -               | +               | +               | -              |
| SPI-15        | -               | -               | -               | -              |
| SPI-16        | -               | -               | +               | -              |
| SPI-17        | -               | -               | -               | +              |
| SPI-18        | -               | -               | -               | -              |
| SPI-19        | -               | -               | -               | -              |
| SPI-20        | +               | +               | -               | -              |
| SPI-21        | +               | +               | -               | -              |
| SPI-22        | -               | -               | -               | -              |

+ means SPI is present in the serotype.
- means SPI is absent in the serotype.

### Conclusions

*S. arizonae* is phylogenetically positioned between *S. bongori* and *Salmonella* subgroup I and shares some pathogenicity-associated genes with *S. bongori* and some others with *Salmonella* subgroup I lineages. Therefore *S. arizonae* genome analyses may provide important clues to key genomic events that might have facilitated the evolution of warm-blooded animal pathogens from cold-blooded parasites.

### Abbreviations

SARC: *Salmonella* reference collection; CTAB: Cetyl trimethyl ammonium bromide; MLEE: Multilocus enzyme electrophoresis.

### Competing interests

The authors declare that they have no competing interests.

### Authors’ contributions

**CW** carried out the genome sequence analysis and drafted the manuscript. **SLZ** and **BL** participated in genome sequence analysis. **XYW** participated in PCR amplification and sequencing of the PCR products by ABI3130 sequencer. **RJ**, **JZ** and **GRL** participated in the study design and provided reagents for the project. **YGL** and **JZ** provided the SOLIO and ABI Sequencing platform. **SLL** conceived the study and finalized the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

This work was supported by a Heilongjiang Innovation Endowment Award for graduate studies to **CW** (YJSCX2012-23SHLJ) and to **XYW** (YJSCX2012-198HLJ); National Natural Science Foundation of China (NSFC30970078 and NSFC81201248) and a grant of Natural Science Foundation of Heilongjiang Province of China to **GRL**; and grants of the National Natural Science Foundation of China (NSFC30970119, 81030029, 81271786, NSFC-NIH 81161120416) to **SLL**.

### Author details

1. Genomics Research Center, Harbin Medical University, Harbin, China.
2. Institute for Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China.
3. Department of Infectious Diseases, The First Affiliated Hospital, Harbin Medical University, Harbin, China.
4. Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Canada.
5. Department of Hematology, The First Affiliated Hospital, Harbin Medical University, Harbin, China.
6. Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Canada.

**Received:** 29 September 2014 **Accepted:** 21 April 2015

**Published online:** 03 June 2015

**References**

1. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. *Salmonella* nomenclature. J Clin Microbiol. 2000;38(7):2465–6.
2. Boyd EF, Wang FS, Whittam TS, Selander RK. Molecular genetic relationships of the salmonellae. Appl Environ Microbiol. 1996;62(3):804–8.
3. Tang L, Wang CX, Zhu SL, Li Y, Deng X, Johnston RN, et al. Genetic boundaries to delineate the typhoid agent and other *Salmonella* serotypes into distinct natural lineages. Genomics. 2013;102(4):331–7.
4. 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;14:89.
5. 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.
6. Baumler AJ, Tsolis RM, Ficht TA, Adams LG. Evolution of host adaptation in *Salmonella* enterica. Infect Immun. 1998;66(10):4579–87.
7. Doulliette RF, Feng DF, Tsang S, Cho G, Little E. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science (New York, NY). 1996;271(5248):470–7.
8. Lawrence JG, Ochman H. Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci U S A. 1998;95(16):9413–7.
9. Ponwillok S, McClelland M. Lateral gene transfer in *Salmonella*. Microbes and Infection / Institut Pasteur. 2003;5(11):977–89.
Salmonella typhimurium – Salmonella – Salmonella – pathogenicity island 2 and the –

74. 6. 9. 23. – Escherichia coli 41. 64. – survival in host cells. Proc Natl 500.

sp. nov., nom. rev., as the type and only species of the genus

8. 9. 86. 9. pathogenicity island 2 are required for bacterial virulence and

2. Shigella 4. provides insights into the evolution of the

Salmonella enterica – 4. Enterobacteriaceae – – serovar

initiates murine 7. 66. Salmonella bongori – – – 15.

K-12. J Bacteriol. 2007;189(11):5733–40.

10. Jones BD, Ghoori N, Falkow S. Salmonella typhimurium initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer’s patches. J Exp Med. 1994;180(1):15–23.

11. McCormick BA, Miller SJ, Carnes D, Madara J. Transepithelial signaling to neutrophils by salmonellae: a novel virulence mechanism for gastrointestinal infections. Infect Immun. 1995;63(6):3202–9.

12. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev. 2004;17(1):14–56.

13. Ochman H, Groisman EA. Distribution of pathogenicity islands in Salmonella spp. Infect. Immun. 1996;64(12):5410–2.

14. Hensel M, Shea JE, Baumler AJ, Gleeson C, Blattner F, Holden DW. Analysis of the boundaries of Salmonella pathogenicity island 2 and the corresponding chromosomal region of Escherichia coli K-12. J Bacteriol. 1997;179(4):1105–11.

15. Hensel M, Shea JE, Waterman SR, Mundy R, Nikolaus T, Banks G, et al. Genes encoding putative effector proteins of the type III secretion system of Salmonella pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. Mol Microbiol. 1998;30(1):163–74.

16. Cirillo DM, Valdivia RH, Monack DM, Falkow S. Macrophage-dependent induction of the Salmonella pathogenicity island 2 type III secretion system and its role in intracellular survival. Mol Microbiol. 1998;30(1):175–88.

17. Fooks A, Schroeder GN, Langridge GC, Blondel CJ, Marmion B, Connor TR, et al. Salmonella bongori provides insights into the evolution of the Salmonellae. PLoS Pathogens. 2011;7(8):e1002191.

18. Ochman H, Soncin FC, Solomon F, Groisman EA. Identification of a pathogenicity island required for Salmonella survival in host cells. Proc Natl Acad Sci U S A. 1996;93(15):7800–4.

19. Croso JH, Brenner DJ, Ewing WH, Falkow S. Molecular relationships among the Salmonellaeae. J Bacteriol. 1973;115(1):307–15.

20. SGSC web site. [www.ucalgary.ca/~kesander]

21. Field D, Garty G, Gray T, Morrison N, Selengut J, Stenk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26(5):541–7.

22. Doyle JD. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.

23. Illumina web site. [http://www.illumina.com/technology/sequencing_technology.Illumi]

24. Aszt RL, Bartels D, Best AA, Delong M, Ditz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

25. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics (Oxford, England). 2002;18(12):1744–5.

26. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature. 2001;413(6858):852–6.

27. Desai PT, Porwollik S, Long F, Cheng P, Wollam A, Bhonagiri-Palsikar V, Bhonsle R, et al. Predicting transmembrane protein topology with a hidden Markov model application to complete genomes. J Mol Biol. 2001;305(3):567–80.

28. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80.

29. Kato H, Matsuda 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):3905–9.

30. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 2007;24(5):1596–9.

31. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–9.

32. Ganry GM, Bell JA, Libburn T. Phylum XIV. Proteobacteria phyl. In Bergey’s manual of systematic bacteriology, volume 2, part B. 2nd ed. New York: Springer; 2005:1.

33. Ganry GM, Holt JG. Taxonomic outline of the archaea and bacteria. In Bergey’s manual of systematic bacteriology, volume 1. 2nd ed. New York: Springer; 2001. p. 155–66.

34. Rahn O. New principles for the classification of bacteria. Zentralbl Bakteriol Parasitenkd Infektionskr Hyg. 1937;6:273–86.

35. Commission J. Conservation of the family name Enterobacteriaceae, of the name of the type genus, and designation of the type species Opinion No. 15. Int Bull Bacteriol Nomencl Taxon. 1958;8:74.

36. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res. 2003;31(1):439–41.

37. Lagesen K, Hallin P, Rognes T, Staerfeldt HH, von Heijne G,见hea TK. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–4.

38. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–9.

39. Commission J. Conservation of the family name Enterobacteriaceae, of the name of the type genus, and designation of the type species Opinion No. 15. Int Bull Bacteriol Nomencl Taxon. 1958;8:74.

40. Le Minor L, Popoff MY. Request for an opinion. 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.

41. Shigella spp. Infect Immun. 1996;64(12):5410–2.

42. Ochman H, Soncin FC, Solomon F, Groisman EA. Identification of a pathogenicity island required for Salmonella survival in host cells. Proc Natl Acad Sci U S A. 1996;93(15):7800–4.

43. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567–80.

44. Kato H, Matsuda 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):3905–9.