Phylogenetic relationship of Salmonella enterica strains in Tehran, Iran, using 16S rRNA and gyrB gene sequences

Mercedeh Tajbakhsh¹, Babak Noory Nayer¹, Kamyar Motavaze², Pedram Kharaziha¹, Mohsen Chiani¹, Mohammad Reza Zali¹, John D. Klena³

¹Research Center for Gastroenterology and Liver Disease, Shahid Beheshti University, Medical Science, Tehran, Iran
²Department of Microbiology, Islamic Azad University, North Tehran Branch, Tehran, Iran
³School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

Abstract

Introduction: We assessed whether 16S rDNA and gyrB gene sequences, alone or combined, were suitable for determining the phylogenetic relationship among Salmonella enterica strains isolated from Tehran, Iran. Patients over five years of age enrolled in an acute diarrheal surveillance project in Tehran province between May 2004 and October 2006 were selected as our study group.

Methodology: 16S ribosomal DNA (rDNA) and gyrB genes from 40 Salmonella isolates obtained from patients with acute diarrhea were sequenced and the data was used to generate phylogenetic trees that facilitated isolate comparison.

Results: Salmonella strains clustered into five to seven phylogenetic groups, dependent on analysis of 16S rDNA (1546 bp), gyrB (1256 bp) or a combination of the two genes. By 16S rDNA sequence analysis, only strains of Salmonella enterica serovar Typhi (S. Typhi) clustered exclusively together. gyrB sequences permitted clustering of all the S. Typhi and S. Paratyphi A isolates, and clustering of S. Enteritidis into two separate but exclusive groups. Concatenation of the two data sets did not significantly improve the resolution of the strains compared to the gyrB gene. None of the analyses completely resolved S. enterica Paratyphi B and C into mutually exclusive groups.

Conclusion: Sequencing of gyrB represents a potentially useful tool for determining the phylogenetic relationship of S. enterica strains in Tehran, Iran. Genetic analysis of the 16S rRNA gene alone or in combination with gyrB did not increase the resolution between serotypes of S. enterica. We speculate that inclusion of additional genetic markers would improve the sensitivity of the analysis.

Key words: Salmonella; 16S rDNA; gyrB; phylogeny; Tehran


(Received 16 August 2010 – Accepted 14 November 2010)

Introduction

The genus Salmonella comprises a group of Gram-negative bacilli belonging to the family Enterobacteriaceae and contains a number of closely related organisms. The genus consists of two species, S. enterica and S. bongeri [1]. S. enterica has been further subdivided into over 2500 serotypes that can be differentiated by the Kauffman-White scheme, which is based on the serologic identification of O (somatic) and H (flagellar) antigens.[2]. Approximately 1400 of the serotypes have been reported to cause gastroenteritis in humans, while only a handful are capable of causing typhoid [3], a potentially fatal systemic infection.

S. enterica includes six subspecies on the basis of chromosomal DNA hybridization and multilocus enzyme electrophoresis as follows: S. enterica subsp. enterica subspecies (I), S. enterica subsp. salmonae (II), S. enterica subsp. arizonae (IIIa), S. enterica subsp. diarizona (IIIb), S. enterica subsp. indica (VI), S. enterica subsp. houtenae (IV) [2,4,5]. The majority of these serotypes belong to S. enterica subspecies I; these serotypes also cause most infections in humans and warm-blooded animals [6]. Whole genome comparative analyses of Salmonella serotypes [7-9] have shown that approximately 90% of the genome is conserved, and there is on average about 97% sequence identity when comparing homologous genes [3]. However, for most laboratories performing routine public health surveillance for detection and identification of Salmonella, whole genome methods are not possible due to resource and time constraints.
Molecular techniques utilizing polymerase chain reaction (PCR), alone or in combination with DNA sequence analysis, have become increasingly popular in determining the evolutionary relationships of bacteria [10-12]. With respect to *Salmonella* phylogenetics, these methods include sequence determination of the small subunit ribosomal RNA (rRNA or 16S rRNA) [13-16], large subunit (23S rRNA [17]) and various housekeeping (*atpD* [18], *gapA* [19], *gyrB* [20], *mdh* [21]) or virulence (*sefA* [22], *sopE* [23]), *inv-spa* region [24, 25] genes. Several multilocus sequence typing methods have also been developed [26-29]. The current database (http://mlst.ucc.ie/mlst/dbs/Senterica) contains over 3,047 isolates and 735 sequence types. Phylogenetic relationships are subsequently inferred after DNA sequencing comparison; this comparison is possible using a number of commercially available computer programs.

In this pilot study, we analyzed the 16S rRNA and *gyrB* DNA sequences of 40 isolates of *S. enterica* isolated over a two year time period in Tehran, Iran. Our goal was to establish a simple and rapid sequence-based method for molecular identification and phylogenetic characterization of *Salmonella enterica* serotypes Enteritidis, Paratyphi A, Paratyphi B, Paratyphi C, and Typhi. This is the first report of the application of *gyrB* typing of *Salmonella* species isolated from clinical samples in Iran.

**Methodology**

**Bacterial isolates**

A total of 40 *Salmonella enterica* isolates from a total collection of 54 clinical isolates collected as part of routine treatment for diarrheal diseases were tested in this study. Samples were selected from patients over five years of age with acute diarrhea living in Tehran province, Iran. Bacterial isolates were acquired between May 2004 and October 2006 by the Research Center of Gastroenterology and Liver Disease, Food-borne Department, Shaheed Beheshti University. Patient histories including questions about the presence of fever, abdominal pain, and vomiting were taken at the time of fecal sample submission. Fecal specimens were cultured directly on MacConkey agar (Merck, KGaA, Darmstadt, Germany) and *Salmonella-Shigella* agar (Pronadisa; Hispanlab, Madrid, Spain), without enrichment, for the isolation and identification of *Salmonella*. Suspected *Salmonella* isolates were tested using standard microbiological biochemical assays and serological tests using O and H *Salmonella* antisera (MAST Group Ltd., Merseyside, UK) for confirmation. Most of the isolates were from patients between 15 and 60 years of age. The strains used in this study were *S. Enteritidis* (n = 7), *S. Paratyphi A* (n = 6), *S. Paratyphi B* (n = 8), *S. Paratyphi C* (n = 10) and, *S. Typhi* (n = 9).

**Preparation of chromosomal DNA**

Bacteria were cultured in MacConkey agar at 37ºC between 18 and 24 hours prior to extraction of total DNA using the phenol-chloroform-isooamyl alcohol procedure described by Sambrook *et al.* [30].

**Primer design**

PCR primers for amplification and sequencing of 16S rRNA were designed based on the sequences of *S. Typhi* TY2, Genbank accession number NC-004631, using the Gene Runner software program, version 3.05. Amplification of *gyrB* used primers described previously [31]. Sequencing primers for *gyrB* based on the published *S. Typhi* Ty2 sequence as described for the 16S rDNA gene were developed.

**PCR amplification**

To generate complete nucleotide sequences (1546 bp) for the 16S rDNA gene, primers were designed 80 bp upstream and and 44 bp downstream of the 5’ and 3’ ends of the 16S rDNA gene. Two PCR amplicons were generated, each covering approximately half of the 16S rDNA gene (880 bp and 850 bp), with 75 bp of overlapping sequence between them.

The 16S rDNA and of the majority of the *gyrB* gene (1256 bp) were amplified using an Eppendorf AG 22331 thermal cycler (Eppendorf, Hamburg, Germany). PCR was performed in 25 μl final volumes containing 10 mM Tris-HCl (pH: 9.0), 50 mM KCl, 1.5 mM MgCl2, 0.1% Triton X-100, 200 μl of each deoxynucleotide triphosphate, two picom primers (Table 1), and one unit *Taq* DNA polymerase (Gen Fanavaran Co, Tehran, Iran). A total of forty nanograms of bacterial DNA were used as template for each PCR.

Amplification conditions for the 16S rRNA and *gyrB* gene targets were as follows: initial denaturation at 95ºC for four minutes, followed by 30 cycles of 94ºC for one minute, annealing for 40 seconds at either 60ºC (16S rDNA) or 64ºC (*gyrB*) and extension at 72ºC for one minute. A final extension step of 10 minutes at 72ºC was included to enable near 100% efficiency of the PCR. PCR products were analyzed by electrophoresis through
1.2% agarose gels (Merck, Tehran, Iran) stained with ethidium bromide and visualized on UV Gel Doc (BioRad, Hercules, CA, USA).

DNA sequencing

PCR products were purified using a QIAquick PCR purification kit (QIAGEN Tehran, Iran). Sequencing reactions were performed using a Big Dye Terminator Kit version 3.1 (Perkin Elmer, Foster City, CA, USA). Sequencing procedures were conducted using an Applied Biosystems 3130 Genetic Analyzer; data was collected and analyzed using data collection software version 2.0 and sequencing analysis software version 5.1.1 (Applied Biosystems). Additional internal primers were used to determine the complete coding sequences (Table 1). For quality assurance, a twofold sequencing redundancy of each PCR product was performed.

Phylogenetic analysis

DNA sequences were edited and assembled using the programs SeqMan and Edit Seq (DNA Star, Laser Gene 6, Madison, WI, USA). Sequences were aligned by using the CLUSTAL W program, version 1.81 [32]. Aligned sequences were analyzed using MEGA software version 4 [33]. Genetic distances were computed using the Kimura 2-parameter model [34] and evolutionary trees were constructed by the neighbour-joining method [35]. Percent divergence and similarity were calculated by comparing sequence pairs in relation by MegAlign DNASTAR.

Nucleotide sequence accessioning numbers

The nucleotide sequence data reported in this paper appear in the Gene Bank nucleotide sequence database with the following accession numbers: EU118076-EU118116 for 16S rRNA, EU146963, and EU146965- EU147003 for gyrB alleles.

Results

Phylogenetic structure on the basis of 16S rRNA gene sequences

A total of 40 strains of S. enterica were initially compared based on differences in 16S rDNA sequence (Figure 1). The 16S rDNA phylogenetic analysis organized the strains into five clusters. Overall genetic distances between the sequences within the same serovars ranged from 0.001 to 0.003; the degree of similarity within each group ranged from 99.6%-100%.

Analysis of the 16S rDNA gene grouped all nine S. Typhi strains together (cluster V). However, most of the serotypes of Salmonella were not able to be distinguished as unique by 16S rDNA analysis; all
the remaining clusters (I-IV) were populated by a variety of different serotypes. Cluster I was the most diverse, with nearly equal representation of S. Enteritidis, S. Paratyphi B and S. Paratyphi C strains. A single S. Paratyphi A strain (A1) was located in cluster I. Cluster II was dominated by S. Paratyphi A strains, but also contained a single S. Paratyphi C (C7) strain. Cluster III was mainly home to S. Paratyphi B strains, although a single representative S. Paratyphi C (C1) and S. Enteritidis (E3) strain also co-localized to this cluster. Our data indicates that the 16S rRNA gene sequence is not the most appropriate locus to definitively identify Salmonella serotypes, or to deduce phylogenetic relationships among bacteria in the same genus.

**Figure 1.** Dendrogram of Iran Salmonella strains based on 16S rDNA analysis.

Salmonella phylogenetic structure determined by gyrB gene sequence analysis

Several groups have used gyrB to identify bacterial species within a genus, and to determine the phylogenetic relationship of these organisms. We used a partial sequence of the gyrB gene (1256 bp) to determine the genetic relationship among the 40 Iran Salmonella strains (Figure 2).

Unlike the 16S rDNA dendrogram, seven clusters of Salmonella spp. were delineated using the gyrB gene. The percentage similarity between each cluster was between 98.6% and 100% and the average genetic distance of isolates within a serovar was 0.005 for S. Typhi, 0.011 for S. Paratyphi A, 0.003 for S. Paratyphi B, 0.013 for S. Paratyphi C, and 0.01 for S. Enteritidis.

**Figure 2.** Phylogenetic tree of Iran Salmonella strains based on and gyrB DNA sequence analysis.

Full length gyrB gene sequences (1256 bp) were amplified by PCR and the nucleotide sequences determined. This is a neighbour-joining tree based on 40 Salmonella 16S rDNA sequences. The scale bar indicates two base substitution per 1000 nt position. The number shown next to each node indicates the bootstrap value (1000 replicates). The sequence from an E. coli O157:H7 strain (accession number FDL933) was used as an outgroup. Bootstrap values below 10% are not shown. Strains designated with the prefixes A, B, C, E and T indicate S. Paratyphi A, S. Paratyphi B, S. Paratyphi C, S. Enteritidis and S. Typhi serovars respectively.
neighbour-joining tree based on 40 Salmonella gyrB sequences. The scale bar indicates one base substitution per 100 nucleotide positions. The number shown next to each node indicates the bootstrap value (1000 replicates). The sequence from an E. coli O157:H7 strain (accession number FDL933) was used as an outgroup. Bootstrap values below 10% are not shown. Strains designated with the prefixes A, B, C, E and T indicate S. Paratyphi A, S. Paratyphi B, S. Paratyphi C, S. Enteritidis and S. Typhi serovars respectively.

**Figure 3.** Phylogenetic tree of Iran Salmonella strains based on the combined 16S rDNA and gyrB DNA sequence analysis.

Combined nucleotide sequences (2802 bp) were analyzed in this neighbour-joining tree based on 40 Iran Salmonella. The scale bar indicates one base substitution per 100 nucleotide positions. The number shown next to each node indicates the bootstrap value (1000 replicates). The sequence from an E. coli O157:H7 strain (accession number FDL933) was used as an outgroup. Bootstrap values below 70% are not shown. Strains designated with the prefixes A, B, C, E and T indicate S. Paratyphi A, S. Paratyphi B, S. Paratyphi C, S. Enteritidis and S. Typhi serovars respectively.

In contrast to the 16S rDNA dendrogram, the clusters that were organized based on gyrB sequence analysis were composed primarily of only a single serotype. All the S. Paratyphi B strains grouped to cluster I. Similarly, all S. Paratyphi A and S. Typhi strains grouped to clusters V and III, respectively. A total of two clusters, IV and V, consisted solely of S. Enteritidis strains and most of the S. Paratyphi C strains grouped to either cluster II or VI. Finally, three S. Paratyphi C strains, C1, C2 and C9, co-located to cluster I with the isolates of S. Paratyphi B strains.

**Comparison of 16S rDNA and gyrB gene sequences for phylogenetic analysis**

We combined all the sequence data collected for each strain in this study and performed a phylogenetic analysis (Figure 3). The overall topology of this tree was very similar to that of the gyrB tree. The main differences between the gyrB and combined analyses was the placement of strain C9, one of the three S. Paratyphi C strains that clustered with the S. Paratyphi B isolates using gyrB alone. In the combined tree, this isolate was an outlier within the S. Typhi cluster. All other major phylogenetic relationships identified using gyrB were preserved in the combined tree, confirming the utility of the gyrB marker over the 16S rDNA marker for at least these five serotypes of Salmonella.

**Discussion**

Diarrheal diseases remain a leading cause of morbidity and mortality in Iran [36]. A recent cross-sectional study of pediatric diarrhea in Iran identified a pathogen in 55% of 1,087 cases. The leading cause of bacterial diarrhea was Shigella spp.; Salmonella spp. was recorded as the sixth most common bacterial pathogen (3.9%). Only a limited number of studies are available from Iran assessing the incidence of Salmonella-induced diarrhea. In contrast, the public health situation with respect to typhoid fever, primarily caused by S. Typhi and S. Paratyphi, appears much more severe. Recent epidemiological studies in Asia indicate variable incidences of typhoid, ranging from 15.3 to 151.7 cases per 100,000 person years in China and Pakistan,
respectively [37]. The increase in cases of Paratyphi A and C as primary causes of typhoid fever is also of significant concern [38]. Phylogenetic analyses have been used to understand the emergence and spread of pathogenic organisms, including Salmonella, within human populations [1] and in epidemiological investigations of outbreaks [39].

In developing countries, it is a challenge for reference and hospital laboratories to routinely type isolates of Salmonella species using the Kaufmann-White serotyping scheme because of the costs and resources involved in attaining reagents. Reference laboratories in these countries require a rapid and cost-effective system that will provide accurate identification of isolates with readily available reagents and equipment. The analyses of two housekeeping genes and the probability of identifying Salmonella serotypes based on DNA sequence data that we describe in this paper is the first step in addressing this situation in Iran.

DNA-based techniques have been used for nearly two decades to identify serotypes of Salmonella and subsequently characterize strains within each serotype. These methods evolved from single gene [21] to multiple gene analyses [28]. No universally accepted sequencing method exists for Salmonella. Currently, several promising new methods for identifying serotypes are emerging, based on single nucleotide polymorphisms [1]. However, with respect to this study, these methods were unavailable for use.

Analysis of 16S rDNA has been used for several decades to characterize bacterial strains [15,16]. However, because 16S rDNA typing methods may not be sensitive enough to distinguish phylogenetic differences at the species level [40], housekeeping genes have been used for the study of phylogenetic and taxonomic relationships at this level [20,41].

A number of studies have demonstrated that the gyrB gene could be used as a suitable marker for the classification of some bacterial species [20,31,42-45]. The gyrB gene is present within almost all bacterial species as a single copy gene, and it encodes the ATPase domain of DNA gyrase, which is necessary for replication [46]. Phylogenetic analysis has suggested that the gyrB gene is evolving at a faster rate than the rRNA gene loci. Because of this, phylogenetic analysis using gyrB sequences is expected to provide higher resolution than 16S rRNA gene sequences [47,48].

Direct comparisons of the genetic distance and phylogenetic relationships between 16S rRNA and gyrB gene sequences are not possible since the rate at which the gyrB and 16S rDNA genes evolve is different [20]. In other studies, the gyrB gene has shown a mean substitution rate approximately six times higher than the 16S rRNA gene [49]. Similarly, we found that the phylogenetic tree based on 16S rRNA was unable to group similar serotypes together in discrete clusters in contrast with the gyrB gene analysis.

One of the main limitations of this study was the inability of trees constructed with either gene alone, or in combination with each other, to cluster all members of the same serotype in one group. Although gyrB was able to form discrete clusters of each serotype, with the exception of S. Typhi and S. Paratyphi A, strains were present in unrelated clusters. To achieve our objective of a rapid and simple assay for identification and characterization of Salmonella strains in Iran, additional genes must be analyzed and new methods, based on single nucleotide polymorphisms, must be tested.

Acknowledgements

The authors are grateful for the assistance provided by Fereshteh Jafari, Zahra Nochi, Nazanin Hossein Khan and Hasan Salimi, Research Center for Gastroenterology and Liver Disease, Shahid Beheshti University, Medical Science, Tehran, Iran, and for the support of Danilo Lo-Fo-Wong, Department of Food Safety and Zoonoses, World Health Organization. This work was financially supported by internal funding from the Research Center for Gastroenterology and Liver Disease, Shaheed Beheshti University of Medical Sciences, Tehran, Iran.

References


Corresponding author
John D Klena
Unit 7300, Box 060, DPO, AP, 96521-0050
School of Biological Sciences, University of Canterbury
Christchurch, New Zealand
Phone +66 139 1101 2902
Email: john.klena@gmail.com

Conflict of interests: No conflict of interests is declared.