Phylogenetic tree constructed of *Salmonella enterica* subspecies *enterica* isolated from animals and humans in Basrah and Baghdad governorates, Iraq

M.S. Sadiq¹ and R.M. Othman²

¹Department of Anesthesia, College of Health and Medical Technology, Middle Technical University, Baghdad, ²Department of Microbiology, College of Veterinary Medicine, University of Basrah, Iraq

**Abstract**

The genetic relatedness of *Salmonella enterica sub sp. enterica* isolated from human and animal origin has more interest as its data possibly will offer an essential confirmation for the source of human infection. This study aimed to determine the genetic relationship of *S. enterica subspecies enterica* isolated from human and animal sources. A total of 300 samples were collected from two primary sources, human and animal, from two different regions, Baghdad and Basrah governorates. For constructing the phylogenetic Tree of *Salmonella enterica* subspecies *enterica*, the sequencing of PCR product (positive samples) for each target genes 16s rRNA, avrA, and spvC were analysed using BLAST analysis to determine the similarities and differences between the Iraqi strains and the existing global strains. The similarity rate in the first gene 97.77%, the second gene 98.29%, and the third 96.82%, respectively. The genetic Tree of each of the three genes was set up separately using two methods, Maximum Likelihood, and the second Minimum evolution. The phylogenetic analysis reveals that Iraqi strains of Salmonella are highly similar, and they share the same sequence of 16s rRNA gene with national Salmonella strains. However, their bases of avrA and spvC genes are not similar. This difference leads us to conclude that the Iraqi Salmonella evolution path was characterized by its path in developing global strains with some correlation in some samples; it may be linked with the same ancestors from which it emerged.

**Keywords**: Salmonella enterica, PCR, Phylogenetic Tree

**Introduction**

Salmonellosis is considered one of the most important zoonotic diseases (1-3). *Salmonella* is most likely to inhabit the abdominal lumen of wide-ranging vertebrates (4-6). Over 2,610 different serovars are recognized by a schematic pattern of Kaufmann-White-Le minor. *Salmonella Enteritidis* and *Salmonella typhimurium* are the worldwide distribution and the most common serotypes that cause Salmonellosis in animals and humans, as they are generally transmitted by infested animals and their productions to humans been (7-10). The investigated genes comprised 16s rRNA, avrA, and spvC, in which avrA genes that are heading *Salmonella* for inducing inflammation through inhibition of both pro-inflammatory and the antiapoptotic (NF Kappa B) pathway and as a genetic marker for the existence of SPIs, which is related with the increase invasion and intracellular persist in the phagocytic and non-phagocytic cells (11). While the other gene, *Salmonella* plasmid virulence gene C (*spvC*), this gene exists in a virulence plasmid; generally formed by infected of definite host and it always requires for complete expresses of
virulence in Salmonella (12), Moreover to promote the rapid growing and surveillance inside the same host. However, all these virulence determinants are broadly spreading among animals, humans, and the environment, with specific variation (13). Alternatively, the 16S rRNA gene analysis is the most accepting factor for molecular study (14). Genomes of bacterial strain seemed to contain several genes with a specific role. For instance, in bacteria, the 16S rRNA is recognized to have nine hypervariable areas (V1-V9), which are often used to identify the microorganisms at species level as they show a sequencing diversity in this region many species of bacteria.

Furthermore, many species of bacterial strains have a conserved region closest to these hypervariable parts, and this conserved region considers the quorum sensing sequences that could be amplified by using a universal primer (15,16). The hypervariable area of 16S rRNA sequences was used in many genetic studies to identify a particular bacterial species or recognize among a different bacterial species (17-19). Additionally, many universal primers, such as the primer used to amplify the variable regions 1,3, and 6 (V1, V3, V6), were previously mentioned in previous studies are wide-ranging based primers that were used to distinguish among the various species of Salmonella (20-22).

Phylogenetic trees were used almost limiting to describe relationships within the species in systematics and taxonomy before the beginning of DNA sequencing technologies. In addition to explaining relationships between species on the Tree of life, phylogenies represent relationships between paralogous in a gene family (23); to understand pathogens’ development and epidemiological activity (24,25). More recently, molecular phylogenetics has become an indispensable tool for genome comparisons. It is used to identify genes and classify metagenomic sequences (26,27). Also, to interpret modern and ancient individual genomes and reconstruct ancestral genomes (28-30). So, this study was achieved to determine the genetic relationship of S. enterica sub sp. enterica isolated from human and animal sources and to determine the similarities and differences between the Iraqi strains and the existing global strains.

Materials and methods

Ethical approve

Ethical approve was taken from Research Ethical Committee of College (RECC). No:6,7/35/1910, Date:21/5/2017.

Samples collection and processing

A total of 300 samples were collected from two regions, Basrah and Baghdad governorates, involving the following samples: 50 stool samples were collected from healthy workers in the field who were in contact with the domestic animals, and another 50 samples were collected from the diarrheic patients. While the animal samples were involved, 50 cows’ fecal samples were taken directly from the bowel. Another 50 samples were taken from the rectum of the chickens directly. A total of 50 fresh eggs were gathered from a typical local house of chicken and included two forms of samples: The first was swab samples directly collected after taking out the shell of the eggs. At the same time, the second type of samples was included, taken the contents of whole eggs and directly put in a sterilized container. The final type of samples was involved 50 swab samples were accumulated from sewerage water and the ground area of domestic animals’ houses and slaughter tools. According to a previous study, all methods and processing samples collection were carried out (31).

Primary detection of Salmonella

The direct detection of Salmonella was mentioned in the previous study (31,32). Further confirmation was done by Polymerase Chain Reaction (PCR) and analysis of DNA sequencing.

DNA extraction

Bacterial DNA from suspected colonies was extracted by suspending colonies of bacteria growth in 200 µl of sterilized DWDW and then boiling at 100 °C for 15 min to lyse the bacteria. After heating, about 800 µl of Sterile DW was added to get 1ml and vortex until the solution was remixed. The purified DNA was then harvested by centrifuge at 12,000 rpm for 10min.

PCR technique

Three sets of oligonucleotide primers, 16s rRNA, avrA, and spvC, were used in PCR to detect Salmonella species and construct the phylogenetic tree. Oligonucleotide sequences of primers and amplification region are shown in table 1. While the reaction mixture was prepared in a total volume of 25µl for all genes as a following: one µl (10pmol/µl) of forward and reverse primer were added to 5 µl of master mix, then 10 µl of extracted DNA was added as a template the final volume was adjusted by adding 8 µl of Nuclease free water PCR was performed on a PCR thermocycler under the conditions: 95°C for 5min, 94°C for 10 sec, and for annealing (55°C for 30 sec for 16s rRNA gene and 58°C for avrA and spvC genes ), then 72°C (60 sec) for elongation. Cycles were repeated 35×. And the final extension 72 °C for 10 min. The PCR products were investigated by electrophoresis using 1 % agarose gel and staining with ethidium bromide the observed after under UV light (33).
Table 1: The Oligonucleotides Primers Details used under study

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5’ 3′</th>
<th>Amplicon size (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>16s rRNA</td>
<td>TGTTGTGTTAATAAACCGCA CACAAATCCATCTCCTGGA</td>
<td>572</td>
<td>31</td>
</tr>
<tr>
<td>avrA</td>
<td>CCT GTATTGGTGACGTCTGG AAGAAGCTTCGTTAATTGTCCTCC</td>
<td>422</td>
<td>12</td>
</tr>
<tr>
<td>spvC</td>
<td>ACC AGAGACATTGCTTGCAACC AGAGACATTGCTTGCA</td>
<td>467</td>
<td></td>
</tr>
</tbody>
</table>

DNA sequencing, phylogenetic data analysis, and tree drawing

The 20 µl of PCR product (positive samples) with 2 µl for each sample of 17 pica Mole of the F primer for each target genes 16s r RNA, avrA, and spvC were sent to MACROGEN "http://dna.macrogen.com" for getting the sequencing of the target genes. The crude sequencing is then analyzed using the Basic Local Alignment search tool (BLAST) of the Geneious software to search for a similar sequence previously recorded in the National Centre for Biotechnology Information database (NCBI) http://blast.ncbi.nlm.nih.gov. The phylogenetic tree was constructed using the Maximum Likelihood and Minimum Evolution method by MEGA 7 software (34).

Results

The suspected samples from primary identification were analyzed for the presence of 16s rRNA, avrA, and spvC genes using three pairs of primers for the amplification of 572 bp; 422bp and 467bp DNA fragments, respectively. PCR results of 16s rRNA, avrA, and spvC genes demonstrated that 39 of investigated samples showed this gene’s presence (Figures 1-3).

Figure 1: Conventional PCR amplification of 16rRNA gene (A) for salmonella strains. Lane M: molecular marker (100bp), lane 1 positive result for 16rRNA gene (572bp), lane 2 for control negative.

Figure 2: Conventional PCR amplification of avrA gene (B) for salmonella strains. Lane M: molecular marker (100bp), lane 1 to eight positive results for avrA gene (422bp).

Figure 3: Conventional PCR amplification of spvC gene (C) for salmonella strains. Lane M: molecular marker (100bp), lane 1 to 6 negative result, lane 8and nine positive results for spvC gene (467bp), lane 7 for control negative.

Sequence Analysis and Phylogenetic tree construction

For a complete detection for Salmonella spp, the alignment information for each gene was analyzed to see the distribution of the similarity with NCBI information as shown in table 2, which gives the nearest national strains to the Iraqi strains and its identity level. Additionally, the
phylogenetic tree was constructed, and all positions containing gaps and missing data were eliminated automatically by MEGA 7 software. Two methods of phylogenetic tree were used. The first method was (Maximum Likelihood), and the second method was (Minimum Evolution) (Figures 4-6). For recognizing between the three genes, it was written in one latter (A, B, and C) for 16s rRNA, avrA, and spvC respectively, and to recognize between Iraqi strains and national strains from NCBI information, a pink color sign was put before national sample with the recording code and the country where it was isolated. On the other hand, 10 of the local Iraqi strains were registries in NCBI, and the accession number was the following: MK968117, MK968116, MK972334, MK940757, MK940799, MK941605, MK920189, MK920200, MK940815, MK940590.

Table 2: Level of Identity for the Iraqi strains with GenBank and giving the accession number for the nearest national strains

<table>
<thead>
<tr>
<th>Sample No.</th>
<th>Species</th>
<th>Identity (%)</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>97</td>
<td>LR590082.1</td>
</tr>
<tr>
<td>2</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>97</td>
<td>CP034230.1</td>
</tr>
<tr>
<td>3</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP034233.1</td>
</tr>
<tr>
<td>4</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>MK215845.1</td>
</tr>
<tr>
<td>6</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP029852.1</td>
</tr>
<tr>
<td>7</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP029914.1</td>
</tr>
<tr>
<td>8</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP029900.1</td>
</tr>
<tr>
<td>11</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>95</td>
<td>AF057363.1</td>
</tr>
<tr>
<td>14</td>
<td>Salmonella enterica sub sp. Enterica</td>
<td>98</td>
<td>CP007323.2</td>
</tr>
<tr>
<td>21</td>
<td>Salmonella enterica sub sp. Enterica</td>
<td>99</td>
<td>KJ672306.1</td>
</tr>
<tr>
<td>22</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>KU 641446.1</td>
</tr>
<tr>
<td>31</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>MK972334.1</td>
</tr>
<tr>
<td>32</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP007323.2</td>
</tr>
<tr>
<td>33</td>
<td>Salmonella enterica sub sp. Enterica</td>
<td>98</td>
<td>CP007344.2</td>
</tr>
<tr>
<td>34</td>
<td>Salmonella enterica sub sp. Enterica</td>
<td>99</td>
<td>MK941605.1</td>
</tr>
<tr>
<td>35</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>CP014358.1</td>
</tr>
<tr>
<td>36</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>100</td>
<td>CP014536.1</td>
</tr>
<tr>
<td>37</td>
<td>Salmonella enterica sub sp. Enterica</td>
<td>98</td>
<td>CP007354.2</td>
</tr>
<tr>
<td>38</td>
<td>Salmonella enterica sub sp. Enterica</td>
<td>98</td>
<td>CP007381.2</td>
</tr>
<tr>
<td>39</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>86</td>
<td>CP007377.2</td>
</tr>
<tr>
<td>40</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>CP014358.1</td>
</tr>
<tr>
<td>41</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>CP007354.2</td>
</tr>
<tr>
<td>42</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP007309.2</td>
</tr>
<tr>
<td>43</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>97</td>
<td>CP007423.2</td>
</tr>
<tr>
<td>51</td>
<td>Salmonella enterica</td>
<td>98</td>
<td>LR134233.1</td>
</tr>
<tr>
<td>52</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>97</td>
<td>KT036673.1</td>
</tr>
<tr>
<td>53</td>
<td>Salmonella enterica</td>
<td>99</td>
<td>MK920189.1</td>
</tr>
<tr>
<td>54</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>CP029958.1</td>
</tr>
<tr>
<td>55</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>MK215094.1</td>
</tr>
<tr>
<td>56</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP029960.1</td>
</tr>
<tr>
<td>57</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>LN999997.1</td>
</tr>
<tr>
<td>58</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>99</td>
<td>CP029952.1</td>
</tr>
<tr>
<td>60</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>97</td>
<td>KY766065.1</td>
</tr>
<tr>
<td>61</td>
<td>Salmonella enterica sub sp. enterica</td>
<td>98</td>
<td>KX146473.1</td>
</tr>
</tbody>
</table>
Figure 4: Molecular Phylogenetic Tree of \textit{Salmonella enterica sub sp. enterica} based on 16s r RNA gene A, illustration by using Maximum Likelihood method (a) and Minimum evolution method (b), in which Iraqi strains have the following number: (1-10) Human; (11-20) Chicken; (21-30) Cows; (31-50) Egg and (51-70) other samples.

Figure 5: Molecular Phylogenetic Tree of \textit{Salmonella enterica sub sp. enterica} based on \textit{avrA} gene (B), illustration by using Maximum Likelihood method (a) and Minimum evolution method (b), in which Iraqi strains have the following number: (1-10) Human; (11-20) Chicken; (21-30) Cows; (31-50) Egg and (51-70) other samples.
Figure 6: Molecular Phylogenetic Tree of *Salmonella enterica* sub sp. *enterica* based on *spvC* gene (C), illustration by using Maximum Likelihood method (a) and Minimum evolution method (b), in which Iraqi strains have the following number: (1-10) Human; (11-20) Chicken; (21-30) Cows; (31-50) Egg and (51-70) other samples.

Discussion

*Salmonella enterica* sub sp. *enterica* shows a significant challenge in public health as the total number of cases each year increases due to its ability to spread from animals to humans or by a food chain. A molecular technique that uses the highly conserved sequences such as housekeeping and virulence gene will offer an excellent and multipurpose approach for detecting *Salmonella enterica* sub sp. *enterica*. The molecular technique also helps to correctly identify the sources and routes of transmission of infection, besides the detection of the bacterial population. Additionally, since *S. enterica* sub sp. *enterica* has numerous sub ancestries and shows a complex molecular pathogenicity pathway, understanding the evolution pathway of *S. enterica* is more critical in detecting its sources routes of spreading besides predicting and preventing the occurrences (35,36). In our study, the positive strains of PCR products were sent to MACROGEN, Korea, to get the sequence results and construct the phylogenetic Tree of *Salmonella enterica*. Each sequence was BLAST in the NCBI to compare it with the national sequences of *Salmonella* from different sources and different sites. The comparison of strains was made with 100 national isolates to increase the validity and reliability of this study to find out the site of changes in the sequence to determine if there is a mutation or not between Iraqi strains and national reference. The acquired sequences to the BLAST analysis were done by Geneious software to identity between Iraqi *Salmonella* in this study and regional and national *Salmonella* from GenBank. It was done in two steps, the first one for each gene alone to see the distribution of Identity of this gene with GenBank to know the level of identity of all Iraqi strains for the *16s rRNA*, *avrA*, and *spvC* genes with national GenBank. The second one was done by comparing the three genes together to see if there are relationships in the level of identity of the three genes for the same strains to know the different strains if it is unique or there is an error in the isolation technique. The phylogenetic tree is essential to find the organism’s origin and a path of evolution, sequence similarity, and relationship (37). The phylogenetic tree for each gene was made by the maximum likelihood method and minimum evolution method to see the relationship of Iraqi isolates with the higher query cover above 90% of national isolates.

The distribution of the national strains among the Iraqi strains in the group 1 was content Canada, India, USA, Netherland, China, South Africa, Uganda, and Bangladesh and maximum similar of Iraqi strains, while group 2 content France and UK and similar Iraqi strains and group 4 content Pakistan, Turkey, Morocco, Mexico and Brazil with maximum similar Iraqi strains, except group 2 which missed national strains but it shares the same ancestor of
the group 1 which was a branch of the same ancestor of the group 3. On the other hand, the minimum evolution for 16s rRNA genes like in the maximum likelihood method the distribution of the Iraqi strains and national strains in the most groups of the tree found in the USA, Canada, and India in the second group and France and UK in the third group and Uganda, Bangladesh, South Africa, China and Netherland in 4 and Mexico, Brazil, Turkey, Morocco, and Pakistan in group 7. In contrast, groups 1, 5, and 6 arise from the same ancestor of groups 2, 4, and 7, respectively, but that had its evolution branch.

This means the Iraqi strains of Salmonella are highly similar to the national salmonella strains and share the same sequence of 16s rRNA gene because this gene is fixed and have a low mutation rate and need a very long time to change, so it is used in the classification of the bacteria in the world. Detection based on 16s rRNA is a reference method for bacterial identification and taxonomic studies (38).

Moreover, the distribution of Iraqi strains, in strains B35 and B51 were significantly like to the national strains from Taiwan, UK, South Korea, Turkey, and Australia in group 2, and the strains B32, B33, B40, B41, B42 and B43 which were isolated from eggs was maximum similar to the strains from Canada and Mexico in group 3. In contrast, the other groups lacked the similarity to the national strains, and group 5 was unique with USA strains alone which is far away from Iraqi and other strains in this tree. Although the groups 1 and 2 come from the same ancestor and 3 and 4 share the same ancestor, it has a particular branch. Additionally, the evolution in the avrA gene, which is divided into two branches each one takes two directions of evolution one similar to national strains as in group 2 content the USA, Australia, Turkey, South Korea, Taiwan and UK and group 3 content Canada and Mexico, while the other direction was unique in evolution as in groups 1 and 4. The mutation effectiveness can explain this in avrA gene that led to some change in the Iraqi strains and plays an influential role in the distribution of the Iraqi strains in the tree led to a collection of the national isolates with Iraqi strains maximum likelihood in group 1, which contains France, China, Germany, Hungary, and South Korea and group 2 contains Australia, Brazil, Switzerland, USA, Canada, UK, and Taiwan. In contrast, groups 3 and 4 were lack of similarity to the national strains. Also, only group 2 shows the same evolution similarity with national strains which content France and China. In contrast, other groups had their evolution branch, except group 4, which contains only national isolates Switzerland, Canada, Taiwan, UK, USA, Brazil, Australia, Germany, South Korea, and Hungary and lacks any sharing with Iraqi strains, that may be explained as Iraqi strains had its evolution in spvC gene.

All above results were attributed to the evolutionary relationships between the ancestor and new generation in the tree is trunks ancestor and branches or new generation placed at the end of branches that have arisen from the ancestor. The distance between two groups is the indicator for the relationship degree; close branches mean closely related groups. Branch length as long as mean the number of changes (39).

Subsequently, the relationships between the differences and similarities of Iraqi strains and national strains may refer to mutation rate due to the environmental conditions, drugs use, and the transportation of bacteria by the animals, foods, and water pollution of rivers. Moreover, attributed to widespread use with over-the-counter accessibility and unrestricted obtaining to common antimicrobials drugs for human and animal usages, these factors create resistant strains (40).

In many countries, the infection rate by Salmonella is significantly increased among populations living in the waste water-spreading grounds than those living in the control areas that do not practice wastewater spreading. Furthermore, as a result of persistent occurrence of Salmonella in higher concentrations in effluent water and survival for an extended period in the moist soil, in which the cattle feeding on drain water irrigated pasture leads to be infected with Salmonella. Additionally, people may have the bacteria when consuming milk or ingestion meat from such infected cattle (41).

Conclusions

We concluded that Iraqi strains of Salmonella are highly similar to the national Salmonella strains and share the same sequence of 16s rRNA gene, but show some dissimilarity on the bases of avrA and spvC genes, which could be attributed to the mutation.

Acknowledgment

We are thankful to the College of Veterinary Medicine, University of Basrah, Iraq, for providing the laboratory facilities.

Conflict of interest

Conflict of Interest: The authors state that there is no conflict of interest.

References


الشجرة العرقية لسلالات السالمونيلا المعوية للت نوع
التحت المعوزة من الإنسان والحيوان في محافظتي البصرة وبغداد، العراق

ميثم صبح صادق و رشا منذر عثمان

قسم التخدير، كلية التقنيات الصحية والطبية، الجامعة التقنيّة الوسطى، بغداد، فرع الأحياء المجهرية، كلية الطب البيطري، جامعة البصرة، البصرة، العراق

الخلاصة

تعتبر القرابة الوراثية للأنماط المصلية للسالمونيلا المعوية للت نوع
التحت المعوزة من الإنسان والحيوان ذات أهمية لأن معلوماتها توفر دليلاً مهمًا لمعرفة مصدر العدوى البشرية. لذا هدفت
الدراسة الحالية إلى تحديد العلاقة الوراثية للسالمونيلا المعوية تحت
النوع المعوزة من الإنسان والحيوان. جمعت 300 عينة من
مصدرين رئيسيين بشري وحيواني في محافظتي البصرة وبغداد. لبناء
الشجرة الوراثية لجراثيم السالمونيلا المعوية تحت النوع المعوزي تم
تحليل نتائج تحديد التتابعات الجينات 16s r RNA و spvCavrA using MEGA software. نموذج النتائج مع النماذج العالمية تم
تحقيق تشابه وانسجام بين النماذج، حيث قدرت النماذج العالمية
على التشابه في سلاسل الجينات 16s r RNA و spvCavrA على التوالي
97.77% و 97.29% و 98.29%.