Molecular Phylogenetic Approach for Classification of
Salmonella typhi

1Ramaiah Arunachalam, 2Balakrishnan Senthilkumar, 3Duraisamy Nenbagam,
2Ponnusamy Selvamaleeswaran and 4Mosos Rajasekaranandan
1Environmental Nanotechnology Division, Sri Paramakalyani Centre for Environmental Sciences,
Manonmaniam Sundaranar University, Alwarkuruchi-627 412, Tamil Nadu, India
2Centre for Biotechnology, Muthayammal College of Arts and Science, Rasipuram-637 408, TN, India
3Department of Biotechnology, Vivekananda College of Engineering for Women, Tiruchengode-637 205,
TN, India
4Department of Zoology, Arignar Anna Govt. Arts College, Namakkal-637 001, TN, India

Corresponding Author: B. Senthilkumar, Centre for Biotechnology, Muthayammal College of Arts and Science,
Rasipuram-637 408, TN, India

ABSTRACT
Salmonella typhi causes a severe systemic illness through intake of contaminated food or water
containing bacterium. This kind of pathogen is still residing in asymptomatic carriers which may
be considered as a sole source for the outbreak. Thus, the aim of the present investigation was to
identify the Salmonella typhi strains isolated from asymptomatic typhoid carriers. The cultural
characteristics on BSA, Macconkey agar and XLD were used for S. typhi identification. The OmpR
gene regions of these strains were amplified and sequenced. The Maximum Parsimony algorithm
tree of OmpR was constructed using max-mini branch and bound model. The OmpR gene
sequences were used for finding similarities and genetic distances with other related bacterial
species. The results of similarity and genetic distance analyses were further used for statistical and
cluster analyses. The two strains were designated as SS3 and SS5. Both strains were identified as
S. typhi on the basis of biochemical, phylogenetic, comparative and statistical analyses. Thereby,
we conclude that gene such as OmpR might be more useful for identifying bacteria at even in
species level. However, use of entire genomic sequences is more accurate yet it is more expensive
and time consuming process.

Key words: Asymptomatic carrier, Salmonella typhi, OmpR, phylogeny, classification

INTRODUCTION
Every academicians and scientists gave the impression of creating definition for bioinformatics
through their field of interests and were framed various definitions. However, in our point of view,
we defined bioinformatics as a combination of biotechnology and information technology. Here,
bioengineer applying information sciences in molecular biology to solve the biological problems,
especially in molecules such as DNA, RNA and proteins. The growing quantities of biological data’s
have initiated the development of huge number of specialized systems in managing and mining
aforementioned biological molecules (Higashi et al., 2009). Now-a-days availability of enormous
number of genome, gene and protein sequences of various organisms in the public databases initiated the development of numerous statistical method. These methods used to analyze the biological molecules have facilitated quick improvements in phylogenetics to classify and discover evolution of various organisms (Dixit et al., 2010; Blanquer and Uriz, 2007). The molecular phylogenetic reconstruction could be given uniqueness to classify the living things even in using of genes (Dixit et al., 2010). Hence, bioinformatics is helping in classification of bacteria's efficiently aided by computational methods (Higashi et al., 2009). Bacterial analysis by 16S rRNA and OmpR has become popular because of these divisions of genes are easy to sequence (Fukushima et al., 2002; Chang et al., 1997; Malickbasha et al., 2010). It is noted that the species delineation based on the 16S rRNA gene sequence analysis is still widely preferred because, when this 16S rRNA gene sequence identity $\geq$97% might be specified a specific species. Furthermore, sequencing of this gene has become much cheaper and faster (Pruesse et al., 2007; Slabbinck et al., 2010).

The both basic and advance approaches such as phenotypic and genotypic characterization were used for bacterial classification. The presences of similarities between microorganisms have derived by numerical taxonomic methods based on a range of present-day observable characteristics (phonetics), usual morphological and physiological test characters. In addition, chemotaxonomic markers such as whole-cell protein profiles, mol% G+C content and DNA-DNA homologies were also used. However, for classification, the powerful technique such as phylogenetic tree reconstruction using nucleotide sequences of conserved genes might acts as molecular chronometers. Both phonetics and phylogenetics was referred to as polyphasic taxonomy. It is suggested that the strategy in report of new species and genus along with organisms taxonomy. The numerical analysis of ribosomal RNA genes leading to the construction of branching phylogeny representing the divergence distance from a common ancestor had provided the foundation of microbial phylogenetics (Owen, 2004).

Salmonella typhi synthesizes three major Outer Membrane Proteins (OMPs) which were greatly rich growth in standard laboratory media. There were three outer membrane proteins such as OmpC, OmpF porins and OmpA (a structural protein). Another major protein was PhoE, synthesized under phosphate limitation. In E. coli, the expression of OmpC and OmpF was under the control of EnvZ and OmpR, a two-component signal transduction system coded by the OmpB (OmpR-EnvZ) locus (Puente et al., 1987). There is an interest, in particularly on OmpR gene of S. typhi is used for finding genealogical relationship (Sam, 2008) with different bacterial species. Thereby, the aim of the present investigation is to classify the bacterial model organism Salmonella typhi strains by phylogenetic and comparative analyses.

MATERIALS AND METHODS
Screening of bacterial strains: To screen the Salmonella typhi from asymptomatic typhoid carriers, stool samples from fifty food handlers were collected. These samples were subsequently transported to the laboratory using the Cary-Blair transport medium in the screw capped tubes. Then, samples were transferred to the Selenite-F broth (Hi-Media, India) to enhance the growth of organisms. It was incubated at 37°C for 24 h. The colonies were inoculated into the Mac Conkey agar media and Bismuth Sulphite agar media (Hi-media, India). It was also incubated at 37°C for 24 h. Among fifty, the two isolates were identified as Salmonella typhi by Gram’s staining, motility,
catalase, oxidase, sugar fermentation, indole, methyl red, vogues-proskauer, citrate, triple sugar iron and urease test (Old, 1996; Malickbasha et al., 2010). Salmonella typhi SS3 and SS5 strains genomic DNA was isolated (Sambrook et al., 1989) and used for OmpR gene amplification described in our previous report (Malickbasha et al., 2010).

**Purification of amplified OmpR gene product and gel elution:** PCR products were separated on low melting agarose (Sigma) and purified using Eppendorf perfectprep gel cleanup kit (Sigma, India) for cycle sequencing process. Nucleotide sequencing was done for the OmpR gene of SS3 and SS5 isolates by automated ABI-3100 Genetic Analyser (GeNei, India). Nucleotide sequences were deposited in GenBank under the following accession numbers viz. EU834745 and EU849617, respectively.

**Phylogenetic analysis:** BLASTN (optimized for megablast) searches were manipulated with the OmpR gene sequences of S. typhi SS3 and SS5 to obtain a highly identical bacterial species. MEGA5 software (Tamura et al., 2011) programme was started with a set of aligned sequences of OmpR gene of S. typhi SS3 and SS5 with representative bacterial species (highly identical species) using Clustal W (Thompson et al., 1994). Subsequently search was made for reconstructing phylogenetic trees that are optimal according to Maximum Parsimony (MP) algorithm using Max-Mini Branch and Bound model (Tamura et al., 2011). The phylogenetic trees were reconstructed for each gene of SS3 and SS5 independently. The reliability of evolutionary trees was evaluated by the bootstrap method with 1000 replications. Each tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The sequence alignments were performed under complete deletion option. The gap open and gap extension penalties in the sequence alignments were 15 and 6.66, respectively.

In pairwise comparison, the OmpR sequence of SS3 and SS5 strains along with reference bacterial species were independently used to calculate the sequence similarity and Genetic Distance (GD) using Clustal W2 (Thompson et al., 1994) and MEGA 4.1 (Tamura et al., 2007), respectively.

**Statistical analysis:** The cluster analysis was carried out for both obtained variables such as sequence similarity and genetic distance of SS3 and SS5 strains using Single Linkage algorithm. This algorithm was optimized under robust and widely applicable Euclidean distance model by PAleontological STatistics (PAST) 2.05 software (Hammer et al., 2001).

**RESULTS AND DISCUSSION**

**Identification of bacterial strains:** Human’s poor personal hygiene and inadequate food handling could potentiate the transmission of S. typhi. Several food products kept at room temperature were found to favour the growth of Salmonella species. The food handlers prominently played a role in disseminating typhoid bacilli through different food products and water (Lin et al., 1988; Senthilkumar and Prabakaran, 2005). It was suggested that a periodic survey should be made on samples from food handlers and food stuffs. Furthermore, the proper sanitation methods should be followed in hotels and restaurants to avoid food contamination and spread of Salmonella sp (Sesikumar et al., 2005). Similarly, the present study was also carried out to screen the typhoid
asymptomatic carriers among the food handlers in Salem District, Tamilnadu, India. A total of two 
S. typhi strains were identified out of fifty stool samples of asymptomatic typhoid carriers. These 
two strains were identified as S. typhi by gram’s staining, biochemical characters, jet black colonies 
on BSA and colorless colonies on MacConkey agar medium that were reported by our group 
(Malickbasha et al., 2010). The PCR products of OmpR gene was analyzed by agarose gel 
electrophoresis for specific size and compared with standard DNA molecular marker. Both the 
representatives S. typhi isolates gave rise to 799 bp for OmpR. It is indicated the presence of the 
wild type gene which encoding for the virulence of the isolates (Fig. 1). The results obtained in this 
study are similar to those of Pickard et al. (1994).

Molecular phylogenetic analysis: A 538 (SS3) and 518 (SS5) bp sequences were amplified from 
the genomic DNA with MGR 06 (F) 5'AGG GCCGT TTC ATC TGC-3' and MGR 07 (R) 5'-ACC 
AGG CTG ACG AACAG-3' primers (Pickard et al., 1994). The OmpR genes showed high similarity 
with OmpR genes deposited in the GenBank (Table 1, 2). In this study, OmpR gene of different 
species (different strains of a species) were obtained by BLASTN search. However, 10 species were 
selected on the basis of high % of sequence identity with good E value for phylogenetic analysis 
(Table 1, 2). The cumulative results from a limited number of studies to date suggest that 16S rRNA 
gene sequencing provides genus identification in most cases (>90%) but less so with regard to 
species (65-83%). With regard to species, 1-14% of the isolates were remaining unidentified even 
after testing. Fukushima et al. (2002) reported that gyrB gene based phylogenetic trees could be 
able to classify some bacteria that could not be classified by their 16S rRNA sequences. Michael and 
Sharon (2007) reported that, the minimum 500-525 bp essential for phylogenetic analysis and also 
for species identification, minimum >99% similarity and ideal >99.5% similarity should be desirable.
Table 1: Result of identical sequence searches in GenBank by BLASTN Algorithm (optimized for megablast) using OmpR gene of SS3 isolated in the present investigation.

<table>
<thead>
<tr>
<th>GenBank accession No. SS3</th>
<th>Highest identical species</th>
<th>Matched sequence accession No.</th>
<th>Sequence identity (%)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU884746</td>
<td><em>Salmonella typhi</em></td>
<td>EU8840017</td>
<td>94</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Klebsiella pneumoniae</em></td>
<td>CP000647</td>
<td>92</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Escherichia coli</em></td>
<td>CP0001509</td>
<td>92</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella flexneri</em></td>
<td>CP000296</td>
<td>91</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella boydii</em></td>
<td>CP000396</td>
<td>91</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella sonnei</em></td>
<td>CP000308</td>
<td>91</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella dysenteriae</em></td>
<td>CP000394</td>
<td>91</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Cronobacter turicensis</em></td>
<td>FN548393</td>
<td>87</td>
<td>3e-170</td>
</tr>
<tr>
<td></td>
<td><em>Eradinia tasmaniensis</em></td>
<td>CU468135</td>
<td>86</td>
<td>7e-162</td>
</tr>
<tr>
<td></td>
<td><em>Eradinia caradocensis</em></td>
<td>BX390851</td>
<td>85</td>
<td>4e-154</td>
</tr>
</tbody>
</table>

Table 2: Result of identical sequence searches in GenBank by BLASTN Algorithm (optimized for megablast) using CmpR gene of SS5 isolated in the present investigation.

<table>
<thead>
<tr>
<th>GenBank accession No. SS5</th>
<th>Highest identical species</th>
<th>Matched sequence accession No.</th>
<th>Sequence identity (%)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU849617</td>
<td><em>Escherichia coli</em></td>
<td>CP0001509</td>
<td>97</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella flexneri</em></td>
<td>CP000206</td>
<td>97</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella boydii</em></td>
<td>CP000306</td>
<td>97</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella sonnei</em></td>
<td>CP000308</td>
<td>97</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Shigella dysenteriae</em></td>
<td>CP000394</td>
<td>96</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Escherichia fergusonii</em></td>
<td>CU923158</td>
<td>96</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Salmonella typhi</em></td>
<td>EU884745</td>
<td>94</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td><em>Klebsiella pneumoniae</em></td>
<td>CP000647</td>
<td>87</td>
<td>8e-171</td>
</tr>
<tr>
<td></td>
<td><em>Salmonella enterica</em></td>
<td>CP000826</td>
<td>86</td>
<td>2e-157</td>
</tr>
<tr>
<td></td>
<td><em>paratyphi A</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Eradinia tasmaniensis</em></td>
<td>CU468135</td>
<td>84</td>
<td>2e-142</td>
</tr>
</tbody>
</table>

Less than 0.5% similarity and other properties such as phenotype should be considered as final species identification and also, gyrB sequence method might be more useful for identifying bacteria to the species level (Fukushima et al., 2002). E-value should be below 0.05 would be considered significant; at least they might be worth considering (Lesk, 2005).

According to Michael and Sharon (2007) and Lesk (2005) reports, the both strain SS3 and SS5 had 538 and 518 bp which were more than 500 bp that should be qualified for phylogenetic analysis. In BlastN analysis, isolate SS3 (538 bp) had significant (94%) >90% identity and E value (0.0) <0.05 with *S. typhi* (EU849617). Other species such as *K. pneumoniae* (CP000647) and *E. coli* (CP0001509) had 92% identity with 0.0 E value followed by *S. flexneri, S. boydii, S. sonnei* and *S. dysenteriae* had 91% similarities with 0.0 E values (Table 1). Isolate SS5 (518 bp) had significant (97%) >90% similarity and E value (0.0) with *E. coli* (CP001509), *S. flexneri* (CP000266), *S. boydii* (CP000036) and *S. sonnei* (CP000038) whereas *S. dysenteriae* (CP000034), *E. fergusonii* (CU923158) and *S. typhi* (EU884745) shows 96-94% identity with 0.0 E value (Table 2). It should be noted that SS3 has identical with *S. typhi* (EU849617) than other reference species whereas SS5 highly identical with *E. coli* (CP001509), *S. flexneri* (CP000266), *S. boydii* (CP000036) and *S. sonnei* (CP000038) when compare to *S. typhi* (EU884745) and other reference species. These results indicate that strain SS3 belongs to *S. typhi* (Table 1) whereas SS5 might be belongs to
S. typhi and other six species mentioned in Table 2. The BlastN result of SS5 illustrates that little perplexes to exactly confirm the species from the list of reference bacterial species. It is suggest, there is a need of improved analyses should be performed for bacterial confirmation and classification. For that, the phylogenetic and comparative analyses were carried out to identify the both strains accurately, even in species level.

As the number of sequences available for analysis continues to grow, the structure of phylogenetic trees derived from OmpR, gyrB and 16S rRNA sequences becomes both more elaborate and accurate. Chang et al. (1997) and Fukushima et al. (2002) reported that 16S rRNA and gyrB of various Salmonella, Shigella and Escherichia coli species were partially sequenced and used for defining all members of the genus, groups of species or individual species. The Fig. 2a had two branches. Among these two branches, one (bottom) divided into two clades contains SS3 along with S. typhi (EU849617) and Shigella dysenteriae (CP000034) whereas other reference species were distinct from SS3. These findings were supported by 100 bootstrap values. As shown in Fig. 2b it was clearly observed that the SS5 tree also had two branches and one (bottom) of these was divided into two clades which were supported by 100 bootstrap values. This clade had isolate SS5 along with Salmonella typhi (EU834745), Shigella dysenteriae (CP000034) and Klebsiella pneumoniae (CP000064). Both of these two trees suggest that isolate SS3 and SS5 are belongs to Salmonella typhi (Fig. 2). Cilia et al. (1996) have reported that 16S rRNA sequences cannot be used to derive phylogenetic tree analyses among closely related bacteria, for example, Shigella and E. coli, owing to the similarity in these gene regions. However, the present results shown the

![Fig. 2(a-b): Phylogenetic tree of SS3 and SS5 based on the nucleotide sequence of OmpR gene with reference sequences obtained from GenBank. The MP trees were constructed by MEGA5 using Max-Mini Branch and Bound model. The reliability of the tree was evaluated by the bootstrap method with 1000 replications. All position containing gaps and missing data were eliminated from the dataset (complete deletion option). (a) SS3 and (b) SS5](image)
possible classification of such closely related bacteria on the basis of OmpR gene. It might suggest that the OmpR gene is an alternative of 16S rRNA gene to determine the evolutionary relationships of bacteria.

**Comparative analysis:** The sequences used in this study were considered for further comparative and statistical analyses to confirm the species. The isolates SS3 and SS5 were used as a query to estimate the pairwise GD and similarity with reference bacterial sequences. It shows that the mean of pairwise genetic distances for SS3-0.050-0.570 is obviously identical with SS5-0.034-0.571. The mean of pairwise genetic similarity (%) of SS3 shows 43-95% whereas SS5 had 43-96%, it suggest that interior of the SS3 and SS5 is obviously identical (Table 3). These findings were point out that query sequence of SS3 had very low GD and high similarity with S. typhi (EU849617) (0.050; 95%) and Shigella dysenteriae (CP000034) (0.076; 91%) and had moderate GD and less similarity with rest of the species. The strain SS5 had very low GD and high similarity with S. typhi (EU884745) (0.050; 95%), Shigella dysenteriae (CP000034) (0.034; 96%) and Klebsiella pneumoniae (CP000964) (0.122; 87%) and had moderate GD and less similarity with rest of the species (Table 3). These results indicate that the isolate SS3 and SS5 are belongs to S. typhi and had an evolutionary relationship with Shigella dysenteriae and Klebsiella pneumoniae.

**Cluster analysis:** The similarity and GD values of isolates SS3 and SS5 with reference bacterial species were used for cluster analysis. As shown in Fig. 3a and b, the clusters were joined based on the average distance between all members in the two groups (Hammer et al., 2001). It shows that the two rooted trees had two branches which might be typically a site for gene duplication. Bacterial strains from both branches have homologous genes in their genomes which might be reflecting the phylogenetic and statistical relationship of strains as well as generation of genes by duplication(s) and/or mutations. Moreover, isolate SS3 (S. No. 1) was cluster with S. typhi (EU849617) and Shigella dysenteriae (CP000034) which was represented by serial number 2 and 11. The isolate SS5 (S. No. 1) was cluster with S. typhi (EU884745), Shigella dysenteriae (CP000034) and Klebsiella pneumoniae (CP000964) were represented by serial number 2, 7 and 9 (Fig. 3, Table 3).

**Table 3:** The pairwise sequence similarity and genetic distance was calculated using Clustal W2 and MEGA4.1 for SS3 and SS5 strain with representative bacterial species.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>species with SS3</th>
<th>Similarity (%)</th>
<th>Genetic divergence</th>
<th>species with SS5</th>
<th>Similarity Genetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SS3</td>
<td>100</td>
<td>0.000</td>
<td>SS5</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>EU849617</td>
<td>95</td>
<td>0.050</td>
<td>EU884745</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td>CP061506</td>
<td>45</td>
<td>0.538</td>
<td>CP061506</td>
<td>44</td>
</tr>
<tr>
<td>4</td>
<td>CP000286</td>
<td>45</td>
<td>0.538</td>
<td>CP000286</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>CP000206</td>
<td>45</td>
<td>0.538</td>
<td>CP000206</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>CP000288</td>
<td>45</td>
<td>0.538</td>
<td>CP000288</td>
<td>45</td>
</tr>
<tr>
<td>7</td>
<td>CP000847</td>
<td>51</td>
<td>0.554</td>
<td>CP000847</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>FUS-43055</td>
<td>46</td>
<td>0.528</td>
<td>FUS-43055</td>
<td>46</td>
</tr>
<tr>
<td>9</td>
<td>CU468135</td>
<td>43</td>
<td>0.548</td>
<td>CU468135</td>
<td>43</td>
</tr>
<tr>
<td>10</td>
<td>BX960851</td>
<td>43</td>
<td>0.670</td>
<td>BX960851</td>
<td>43</td>
</tr>
<tr>
<td>11</td>
<td>CP000034</td>
<td>91</td>
<td>0.076</td>
<td>CP000034</td>
<td>47</td>
</tr>
</tbody>
</table>

The serial number represents the corresponding species and/or accession number of representative species with SS3 and SS5 isolated in the present study.
CONCLUSIONS

Both phylogenetic (Fig. 2) and cluster analyses (Fig. 3) confirm that the strains present in the lower branch might be evolutionarily related to the isolates used in the present study. There is a growing interest in the use of OmpR gene instead of 16S rRNA, gyrA and RNAase sequences. We believe that the OmpR region could have high reliability for identifying pathogenic bacteria and it is considerably new in phylogenetic analysis. The use of genome sequences will provide a rich source of data for future taxonomic analysis. However, there is a need of extensive study on OmpR gene sequences would give more detail for bacterial classification.
ACKNOWLEDGMENTS

Authors thank Muthayammal Educational and Charitable Trust, India for providing wet lab facilities to carryout preliminary work and RA thank Prof. G. Annadurai for providing facility for insilico analysis.

REFERENCES


