Asia Pacific Journal of Molecular Biology and Biotechnology, 2000
Vol. 8 (2) : 155-160

Genomic analysis of *Salmonella* species based on 16S rRNA gene sequences

G. Subramanian[1], S. Puthucheary[1], R. Yassin[2], T. Pang[3] and K. L. Thong[4*]

[1]Dept. of Medical Microbiology, Faculty of Medicine, [4]Inst. of Postgraduate Studies and Research, Faculty of Science, University of Malaya, [2]Inst. of Medical Research, Kuala Lumpur, Malaysia, [3]WHO, Geneva.

Received 15 August 2000/ Accepted 15 November 2000

Abstract: The complete 16S rRNA gene (rDNA) sequences of 15 strains of *Salmonella* species have been analyzed using direct sequencing of PCR amplicons generated using universal primers. The percentage similarity between the *Salmonella* spp. studied was very high (96.4 to 99.2%) indicating a high degree of sequence conservation within this gene. Among the serovars studied, *S. bovismorbificans* was found to be the most dissimilar compared to the others. The base substitutions detected occurred throughout the gene with two apparent 'hot-spots' or signature regions, one of which corresponded to the V6 hypervariable region of the 16SrRNA gene. In addition, a 9-base pair insertion and a 5-base pair deletion were discovered in *S. waycross* and *S. matopeni*, respectively. We have also detected the presence of putative *Salmonella*-specific sequences within the 16S rRNA gene which can be used in the design of a species-specific probe. The *Salmonella* serovars studied appeared to belong to a single species that can be divided into four subclusters. Interestingly, the three human-adapted serovars associated with human enteric fever, that is, *S. typhi*, *S. paratyphi* A and *S. paratyphi* C, were grouped in the same cluster.

Key Words: 16S rRNA sequencing, *Salmonella* spp., genomic analysis

INTRODUCTION

Bacteria from the genus *Salmonella* belong to the family Enterobacteriaceae and many species are important human pathogens which cause gastroenteritis, septicemia and enteric fever. The identification and classification of the salmonellae has traditionally been based on phenotypic characteristics (eg cell surface antigens) (Kauffman, 1996) which subdivide the genus into groups based on O (somatic), Vi (capsular) and H (flagella) antigens. Subsequently, DNA reassociation analysis (Crosa *et al.*, 1973) demonstrated that all *Salmonella* serovars form a single DNA hybridization group which eventually led to the proposal that *Salmonella* consists of a single *Salmonella enterica* with six subspecies. Other typing methods have been used to characterize *Salmonella* including phage typing, multilocus enzyme electrophoresis (Selander *et al.*, 1991), ribotyping (Esteban *et al.*, 1993) and pulsed-field gel electrophoresis (Thong *et al.*, 1996). Despite all these efforts, the true phylogenetic relations among the 2435 *Salmonella* serovars remain poorly understood. More recently, the comparison of rRNA gene sequences has been widely applied as a means to assess bacterial phylogeny (Olsen *et al.*, 1993). rRNA or ribosomal DNA (rDNA) sequence comparisons has provided a useful approach for studying phylogenetic relationships among microorganisms due to its high information content, highly-conserved nature and universal distribution. The approach is particularly useful for organism with similar phenotypes like the salmonellae. However, to date, there has been very few systematic comparative analysis of the salmonellae based on rRNA gene sequences. In this study, we report the complete rRNA gene sequences from 15 species of *Salmonella* and compared the sequences to assess the putative phylogenetic relationships among these species.

MATERIALS AND METHODS

Bacterial strains. A total of 15 *Salmonella* spp were used in this study (Table 1). These were all human isolates

*Author for correspondence. Mailing address: Institute of Postgraduate Studies and Research, University of Malaya, 50603, Kuala Lumpur. Tel: 603-79594437; Fax: 603-79568940; E-mail: q5thong@umcsd.um.edu.my

Table 1. *Salmonella* species used in this study and GenBank accession numbers of 16S rRNA gene sequences.

Species[a]	Abbreviation	GenBank accession no.
Salmonella typhi	St	U88545
Salmonella paratyphi A	SpA	U88546
Salmonella paratyphi B	SpB	U88547
Salmonella paratyphi C	SpC	U88548
Salmonella blockey	Sbl	U90314
Salmonella matopeni	Sm	U90315
Salmonella typhimurium	Stm	U90316
Salmonella weltevreden	Sw	U90317
Salmonella enteritidis	Se	U90318
Salmonella chingola	Sc	U92192
Salmonella waycross	Swy	U92194
Salmonella bovismorbificans	Sbm	U92193
Salmonella houten	Sh	U92195
Salmonella bareilly	Sb	U92196
Salmonella agona	Sa	U92197
Salmonella sp.	Ssp	X80676.1
Salmonella sofia	Ss	X80677.1
Salmonella shomron	Ssh	X80678.1
Salmonella give	Sg	X80676.1

[a]All *Salmonella* spp. studied belonged to subspecies 1 (*S. enterica* subsp. *enterica*) except for *S. houten* which belongs to subspecies 4 (*S. enterica* subsp. *houtenae*).

obtained from the University Hospital, Kuala Lumpur and the Institute of Medical Research. All the isolates were identified, grown and maintained according to standard procedures (Cowan and Steel, 1974).

DNA extraction and PCR. Total genomic DNA was extracted from 5 ml LB culture using a modified method of Saito and Miura (1963) and was used in the amplification of the 16S rRNA gene. Polymerase chain reaction (PCR) amplification was based on the modified method of Dorsch and Stackerbrandt (1992). Briefly, approximately 50 ng of DNA was used per 100 ml PCR reaction mix containing 0.7 mM primer, 400 mM of each dNTP (dATP, dCTP, dGTP and dTTP), 50 mM KCl, 10 mM Tris-HCl (pH8.4), and 2.5 mM $MgCl_2$. Initial denaturation was done for 3 min at 95°C and the samples were chilled on ice. *Taq* DNA polymerase (1.5U) was then added. The thermal profile included 35 cycles of denaturation at 93°C for 1 min, reannealing at 55°C for 1 min and extension at 72°C for 2 min. After the 35 cycles were completed, a final extension was done at 72°C for 10 min. The primers used in both PCR and sequencing reactions were based on the published sequence of the 16S rRNA gene of the *Escherichia coli* (Brosius *et al.*, 1978) and a total of eight internal primers (Dorsch and Stackerbrandt, 1992) in both forward and reverse directions were used to complete the sequencing.

DNA sequencing. The PCR-amplified 16S rRNA genes were purified using the Glass Max DNA isolation matrix (Gibco BRL, Gaithersburg, MD, USA). Direct sequencing of the PCR products was performed using the Sequenase Kit, Version 2.0 (USB, Cleveland) with the incorporation of radiolabeled a-^{32}P-dATP (Amersham International, Buckinghamshire, UK). In order to validate the sequencing results, the amplification and sequencing reactions were performed at least twice for each *Salmonella* spp studied.

Alignment and analysis of nucleotide sequences. The entire DNA sequences of the 16S rRNA gene (approximately 1500 bp) of the *Salmonella* serovars were aligned using a multiple-sequence alignment programme (Geneworks, Intelligenetics, CA, USA). The alignments were then manually edited and checked to minimise the possibility of multistate positions. No gaps were found among the aligned sequences with the exception of *S. matopeni* (5 bp deletion) and *S. waycross* (9 base pair insertion). These gaps were treated as a single, difference regardless of length. The phylogram was analyzed and the distances between sequences were calculated (data not shown) using the Jukes-Cantor method and the tree was then constructed by the neighbor joining tree method. Sequences of 16S rRNA genes from other organisms were obtained through standard BLAST searching of deposited gene sequences.

Nucleotide accession numbers. The accession of the 16S rRNA gene sequences from the 15 *Salmonella* spp. studied is given in Table 1.

RESULTS

Based on the entire sequence of the 16S rRNA genes the percentage similarity in nucleotide sequences among the 15 *Salmonella* serovars was calculated (Table 2). Overall, the sequences of the 16S rRNA genes among these serovars were highly conserved with similarities ranging from 97% to 99% (Table 2). The comparative results showed that *S. bovismorbificans* was the most dissimilar of the 15 serovars studied with 8 out of 14 comparisons giving less than 98% similarity (Table 2). Alignment of these sequences showed that the base substitution occurred throughout the known hypervariable regions of the 16 S rRNA gene with two apparent 'hot-spots' regions,

Table 2. Matrix of similarity between 16S rRNA genes of *Salmonella* spp. abbreviation as in Table 1.

	St	SpA	SpB	SpC	Sbl	Sm	Stm	Sw	Se	Sc	Swy	Sbm	Sh	Sb	Sa
St	-	98.9	98.9	99.2	99.0	98.8	98.4	98.6	98.9	98.6	99.0	98.4	98.6	99.2	98.8
SpA		-	98.6	98.4	99.1	98.1	98.8	98.7	98.6	98.5	98.4	98.4	98.6	99.0	98.6
SpB			-	98.7	98.7	97.8	98.2	98.4	98.0	97.8	97.9	97.6	97.9	98.4	98.1
SpC				-	98.2	97.9	98.1	98.0	98.4	97.7	98.1	97.9	97.7	98.4	98.2
Sbl					-	98.6	98.8	99.1	98.6	98.8	98.5	98.1	98.6	99.2	98.9
Sm						-	98.4	98.8	98.2	98.3	98.5	98.0	97.8	98.3	98.2
Stm							-	99.1	98.3	98.6	98.2	97.8	98.0	98.5	98.2
Sw								-	98.4	99.0	98.7	97.8	98.4	98.8	98.5
Se									-	98.4	98.6	97.9	98.1	98.8	98.4
Sc										-	98.5	97.7	98.3	99.1	98.2
Swy											-	98.2	97.9	98.8	98.3
Sbm												-	97.6	98.2	97.9
Sh													-	98.9	98.4
Sb														-	98.9
Sa															-

the first between nucleotides 820 to 884 and the second between positions 1001 to 1042. In addition, a five base pair deletion (positions 167-171) was found in *S. matopeni* and a nine base pair insertion (positions 396 to 404) in *S. waycross*. Further analysis of the sequences also showed the presence of putative *Salmonella*-specific sequences (GGUGUUCUGUG and ACCGC) at positions 451 to 500 when these serovars were aligned with other Gram negative bacteria (Fig 1). The comparative sequence analysis depicted in Figure 1 also included alignment with four other *Salmonella* sequences (*S. shomron, S, give, S. sofia, Salmonella* spp) obtained independently in another laboratory and procured through a BLAST search of deposited sequences, the data showed 100% concordance of the sequence in this region. A phylogenetic cluster analysis was also performed by the neighbor-joining tree method to assess the degree of relatedness among the 15 serovars (Figure 2). The 15 *Salmonella* spp. studied were grouped into 4 subclusters within the *Salmonella* genus (Figure 2). It was noted that the four major serovars associated with enteric fever in humans (S. *typhi*, S. *paratyphi* A, B, C) were in a single cluster (Figure 2).

DISCUSSION

The widespread use of rRNA sequences as a means to study phylogenetic relationships (Olsen and Woese, 1974) was considered a major advance in the field as prior cellular and physiological approaches were inadequate for inferring such relationships between microorganisms. In contrast, rRNA sequences provided a natural, phylogenetically valid system of classification (Olsen *et al.*, 1994). It has been suggested that rRNA sequences provide a particularly useful approach for studying the phylogeny of microorganisms with similar phenotypes, as has been reported for *Chlamydia* spp. (Pudjiatmoko *et al.*, 1997) and *Vibrio* spp. (Dorsch *et al.*, 1992). Although ribotyping, a method comparing bacterial rRNA gene restriction patterns, has been used in studying *Salmonella* spp, there have been very few studies to compare the 16S rRNA gene sequences among these organisms. In this study, we have performed comparative sequence analysis of the entire 16S rRNA gene of selected *Salmonella* spp. Of the 15 serovars studied, fourteen belonged to subspesies 1 (*S. enterica* subsp. *enterica*) where 99% of clinical isolates are found, and one (*S. houten*) belonged to subspecies 4 (*S. enterica* subsp. *houtenae*). As such, it would be logical to expect that they would show a high degree of 16S rRNA sequence homology and conservation. This was found to be true in the present study which also confirmed that the base substitutions occurred within known hypervariable regions of the 16S rRNA gene, known as the V6 region (Barry *et al.*, 1990). Studies with other organisms (Pudjiatmoko *et al.*, 1997) have similarly found changes occurring in one of the variable regions detected in this study (between positions 1001 to 1042).

We would, however, be a little cautious in making definitive conclusions about the true phylogenetic relationships between the *Salmonella* serovars studied. Although a high degree of sequence conservation was noted, at the same time, they would also be expected to show an unpredictable relationship to each as judged by cluster analysis. We have found this to be true in the present study and it

```
          451                                                                500
St        AGGUGUUGUGGUUAAUAACCGCAGCAAUUGACGUUACCCGCAGAAGAAGC
Shs       .A.G.AGUAAAG.U.AU...UU..CA.UG.....
Shd       UUUGC.CA.UGACGU..C.....A.GAA.CACCGG.UAA.UCCGUCCCA
Ec        ...GAG.AAA.....CCUUUGCUC.....
Sem       .....G..A.C.....CGCU..U.....U.....
Kpn       ...C...AA.....UUG..GA.....
Pv        .....A.AAA.....C.UUUGU.....
Cf        GAAG.nGUU.AGGUUA.UAAC.UCAGCnAUUGACGUUA.U.GC.GA.GAA
Ap        GUG.UAAUA.CAC.U.GCAUUG.CGUUACUCGCAG.AGAAGCACC..CUA
Ha        .UACC..UGCUCAUUG.CGUUACC.GCAGACUU.C.G.G.GGAGGA.G.G
Era       ....UAAUAACC.UU.CCAUUG.CGUUACCCGCAG.AGAAGCACC.GCUA
Ewa       ....UAAUAACC.U.GUGAUUG.CGUUACUCGCAG.AGAAGCACC.GCUA
Yen       UU.ACG.UACUCGCAG..GAAGCA.CGGCU.AC.CCGUGC...C..CC..
Pls       UA..UAAUACC.AGUGGCAUU.ACGUUACUCGCAG.AGAAGCACC.CGUA
Ch        .A.GUGC..AG.....C.U..GCA.....UA.....
Sui       .A.GU..A.....C.U.UGAAU.....A...G.UAA.....
Dn        .C.GUGCA.....C..AUGCA.....A...G.UAAG...A....
SpA       .....
SpB       .....
SpC       .....
Sbl       .....
Sm        .....G.....
Stm       .....
Sw        .....
Se        .....
Sc        .....GCUC.....
Sbn       .....
Sh        .....GCU.....
Sb        .....
Sa        .....
Swy       .....
Ssh       .....
Sg        .....
Ss        .....
Ssp       .....
```

Figure 1. Alignment of 16S rRNA sequences from position 451 to 500. Shs = *Shigella sonnei*, Shd = *Shigella dysenteriae*, Ec = *Escherichia coli*, Sem = *Serratia marcessens*, Kpn = *Klebsiella pneumoniae*, Pv = *Proteus vulgaris*, Cf = *Citrobacter freundii*, Ap = *Aranicola proteolyticus*, Ha = *Hafnia alvei*, Era = *Erwinia amylovora*, Ewa = *Ewingella* americana, Yen = *Yersinia enterocolitica*, Pls = *Plesiomonas shigelloides*, Ch = *Cardiobacterium hominis*, Sui = *Suttonella indologenes*, Dn = *Dichelobacter nodosus*, Ssh = *S. shomron*, Sg = *S. give*, Ss = *S. sofia*, Ssp = *Salmonella* spp. All other abbreviations are as in Table 1. Bold, underlined sequences indicate putative *Salmonella*-specific sequences.

is relevant and important to point out the observation of others that rRNA gene sequence data are probably not ideally suited for deducing phylogenetic relationships between closely related or very recently diverged strains below the genus level (Fox *et al.*, 1992; Cilia *et al.*, 1996; Van De Peer *et al.*, 1996). On the basis of these considerations, it would perhaps be premature, in the absence of more sequence information on more *Salmonella* serovars, to make definite conclusions as to the true phylogenetic relationship between the serovars analysed in the present study.

Despite the possible limitations with regards to making conclusions about true phylogenetic relationships, the availability of 16S rRNA sequences is still of immense value. It allows, for example, an assessment to be made of the rate and type of mutation occurring within this gene. In this study, a five base pair deletion was found in *S. matopeni* and a nine base pair insertions in *S. waycross*. More studies with other *Salmonella* serovars, especially those which are human pathogens are indicated and may provide further evidence for the recent observations that *Salmonella* pathogens possess high mutations frequencies (Leclerc *et al.*, 1996). It has also been shown that the genome of some of these serovars have undergone chromosomal rearrangements involving the DNA fragments between the rRNA (*rrn*) operons

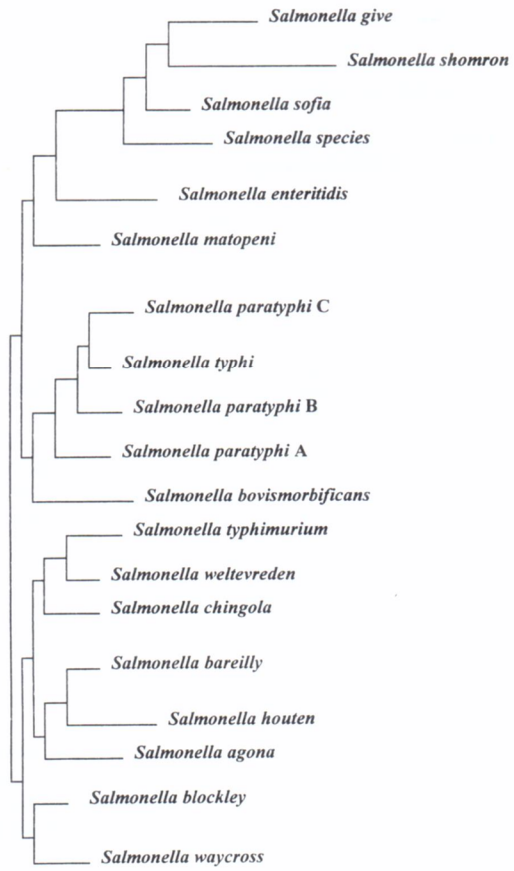

Figure 2. Phylogram based on the neighbour joining tree method depicting the relationships between *Salmonella* serovars based on the 16S rRNA gene sequences.

(Liu and Sanderson, 1995). The availability of sequence data may also have practical applications in the development of probes for diagnostic purposes. Our results suggest the possible presence of putative *Salmonella*-species sequences within the 16S rRNA gene in the region of nucleotide sequences 451 to 500. Alignment of this region with other *Salmonella* sequences obtained independently by another laboratory showed 100% concordance of nucleotide sequence in this region, thus attesting the accuracy and validity of the sequencing results. However, it is important to point out that although the sequence comparisons were made with many species of Gram-negative bacteria, only three belonged to the Enterobacteriaceae and a more extensive comparative study to confirm the specificity of the sequences is required. The general value of DNA sequences within the conserved regions of 16S rRNA gene of the development of specific DNA probes is well established (Barry *et al.*, 1990).

The cluster analysis performed on these strains showed the presence of four major subclusters in the *Salmonella* genus. Not surprisingly, the clustering based on 16S rRNA sequences does not show a good correlation with the serologic classification (Kauffmann, 1996) or that based on enzyme polymorphisms (Selander *et al.*, 1991), both of which are based on phenotypic, rather than genotypic, characteristics. Interestingly, the 4 major organisms most commonly associated with enteric fever in man (S. *typhi*, S. *paratyphi* A, B, C) were in a single cluster. It may be also significant that these four serovars are known to be primarily or exclusively limited in host range (host adapted) to humans (Selander *et al.*, 1991; Baumler, 1997). In contrast, previous studies utilizing multilocus enzyme electrophoresis found no close phylogenetic relationship among clones of different human-adapted serovars (Selander *et al.*, 1991). The data obtained provided an interesting correlation between molecular sequence data and biological behaviour of pathogenic *Salmonella* serovars. Ultimately, the type of analysis described may provide a systematic framework for analyzing the evolution of host adaptation and pathogenicity in this group of bacteria.

ACKNOWLEDGEMENT

The work described was funded by IRPA Grants 06-02-03-0750 from the Ministry of Science, Technology and Environment, Malaysia. We thanked Ms Goh Yee Ling for assistance in typing the manuscript.

REFERENCES

Barry, T., Powell, R. and Gannon, F. 1990. A general method to generate DNA probes for microorganisms. *Biotechnology* 8: 233-236.

Baumler, A. 1997. The record of horizontal gene transfer in *Salmonella*. *Trends in Microbiology* 5: 318-322.

Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. 1978. Complete nucleotide sequence of a 16S rRNA gene from *Escherichia coli*. *Proceeding of National Academy Science*, USA 75: 4901-4805.

Cilia, V., Lafay, B. and Christen, R. 1996. Sequence heterogenities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Molecular Biology Evolution* 13: 451-461.

Cowan, S.T. and Steel, J. 1974. Manual for identification of medical bacteria. 2nd ed. Cambridge: Cambridge University Press.

Crosa, J.H., Brenner, D.J., Ewing, W.H. and Falkow, S. 1973. Molecular relationship among the salmonellae. *Journal of Bacteriology* 115: 307-315.

Dorsch, M., Lane, D. and Stackerbrandt, E. 1992. Towards a phylogeny of the genus *Vibrio* based on 16S rRNA sequences. *Journal of Systematic Bacteriology* 42: 58-63.

Dorsch, M., Lane, D and Stackerbrandt, E. 1992. Some modification in the production of direct sequencing of PCR amplified 16S rDNA. *Journal of Microbiological Methods* 16: 271-279.

Esteban, E., Snipes, K., Hird, D, Kasren, R. and Kinde, H. 1993. Use of ribotyping for characterization of *Salmonella* serotypes. *Journal of Clinical Microbiology* 31: 233-237.

Fox, G.E., Wisotzkey, J.D. and Jurthsuk, P. 1992. How close is close- 16S rRNA sequences identity may not be sufficient to guarantee species identity. *Internationl Journal of Systematic Bacteriology* 42: 166-170.

Kauffmann, F. 1996. The bacteriology of Entero-bacteriaceae. pp. 56-57. Baltimore: Williams and Wilkins.

Leclerc, J.E., Li, B., Payne, W.L. and Cebula, T.A. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274: 1208-1211.

Liu, S.L. and Sanderson, K.E. 1995. Rearrangement in the genome of the bacterium *Salmoenlla typhi. Proceeding National Academy Science USA* 92: 1018-1022.

Liu, S.L. and K. E. Sanderson. 1996. Highly plastic chromosomal organization in *Salmonella typhi*. *Proceeding National Academy Science* USA 93: 10303-10308.

Olsen, G.J. and Woese, C.R. 1993. Ribosomal RNA – a key to phylogeny. *FASEB Journal* 7: 113-123.

Olsen, G.J., Woese, C.R. and Overbeek, R. 1994. The winds of (evolutionary) changing – breathing new life into microbiology. *Journal of Bacteriology* 175: 1-6.

Pudjiatmoko, Fukushi, H., Ochiai, Y., Yamaguchi, T. and Hirai, K. 1997. Phylogenetic analysis of the genus *Chlamydia* based on 16S rRNA gene sequences. *International Journal of Systematic Bacteriology* 47: 425-431.

Saito, H. and Miura, K.J. 1963. Preparation of transforming deoxyribonucleic acid direct sequencing of PCR amplified 16S rDNA. *Journal of Microbiological Method* 16: 271-279.

Selander, R.K., Beltran, P. and Smith, P.H. 1991. Evolutionary genetics of *Salmonella*. *In*: Evolution at the molecular level. eds. Selander, R. K, Clark, A.G. and Whitam, T.S., pp. 25-57. Sunderland: Sinauer Associates.

Thong, K.L., Passey, M., Clegg, A., Combs, B.G., Yassin, R. and Pang, T. 1996. Molecular analysis of isolates of *Salmonella typhi* obtained from patients with fatal and non fatal typhoid fever. *Journal of Clinical Microbiology* 43: 1029-1033.

Van De Peer, Y., Chapelle, S. and De Wachter, R. 1996. A quantitative map of nucleotide substitution rates on bacterial rRNA. *Nucleic Acids Research* 24: 3381-3391.