WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 7.523
Volume 6, Issue 13, 506-522. Research Article ISSN 2277– 7105

COMPUTATIONAL SEROTYPING OF *STREPTOCOCCUS PNEUMONIAE* STRAINS BASED ON CAPSULAR GENETIC LOCI AND POLYMERASE GENE THROUGH PHYLOGENETIC ANALYSIS

Jothi Ramalingam[1]* and Balamurugan Sivaprakasam[2]

[1]Department of Advanced Zoology and Biotechnology, Government Arts College (Autonomous), Nandanam, Chennai 600 035, Tamil Nadu, India.

[2]Department of Computer Science, Vels University, Pallavaram, Chennai 600 117, Tamil Nadu, India.

Article Received on
23 August 2017,

Revised on 13 Sep. 2017,
Accepted on 04 Oct. 2017

DOI: 10.20959/wjpr201713-9805

***Corresponding Author**
Jothi Ramalingam
Department of Advanced Zoology and Biotechnology, Government Arts College (Autonomous), Nandanam, Chennai 600 035, Tamil Nadu, India.

ABSTRACT

Streptococcus pneumoniae has about ninety serotypes, each serotype specific capsule is made up of a unique polysaccharide and nearly 20 genes in a genetic locus are involved in its synthesis. Despite the diversity of capsular polysaccharides, the genetic loci of all serotypes exhibit a similar organization, but few genes including polymerase are unique to given serotypes. Generally, one of the pneumococcal vaccine is derived from capsular polysaccharide. But the wide variation in capsular types and discrepancies in the conventional serotyping make most of the available vaccines ineffective. Typing of pneumococcal strains into effective and accurate serotypes without discrimination is inevitable in developing effective pneumococcal vaccines. The present study considered capsular genetic loci and polymerase genes of ninety serotypes for sequence based computational serotyping through phylogenetic analysis. The study reveals that serotypes from the serogroups 7,9,15,16,19,23,33,35,41 and 47 have genetic differences which are very different from the results of conventional serotyping methods. Similarly, the identical serotypes among serogroups 40/7F/7A; 38/25F/25A and 37/33F/33A are in contrast to the differences revealed by conventional serotyping methods. It provides an understanding about the diversification of capsular polysaccharides and the importance of developing effective vaccines.

KEYWORDS *Streptococcus pneumoniae*, serotyping, capsular genetic loci, polymerase, phylogenetic analysis.

1. INTRODUCTION

Capsule is a surface exposed component of *Streptococcus pneumoniae*, which acts as its major virulence factor and it is one of the causes for the pneumococcal diseases to human beings.[1] Serologically distinct capsules were recognized in 1900s and the polysaccharide nature of the capsule was revealed by studies starting from 1925, which include the landmark works of Griffith et al. and Avery et al. (1944).[2] The diversity of pneumococcal isolates has arisen by capsular polysaccharide antigens imposed by the host immune system. Based on these capsular types and their patterns of reactivity with the factor sera, about ninety serotypes were recognized.[3-5] In earlier days, the different serotypes were identified by agglutination of the cocci with the type-specific anti serum, precipitation of the specific soluble substances with the specific serum and the capsule swelling reaction.[6] Later, many modern molecular typing methods of pneumococcal capsules have come into practice with special reference to serotype specific sequences.[7-12]

Each serotype specific capsule is made up of a unique polysaccharide and nearly 20 genes are involved in its synthesis as a gene cluster.[13] Among 90 serotypes, the serotypes include 1, 2, 4-36 and 38-48 usually follows Block-type (wzy-dependent) capsular.[14,15] Other two serotypes, 3 and 37 are synthesized through synthase dependent pathway.[16-18] Despite the diversity of the capsular polysaccharides, the genetic loci of all serotypes exhibit a similar organization in which genes required for synthesis of a specific capsule type are flanked by genes common to all types.[14,15,19,20] But few genes contained within the type-specific regions are unique to given serotype and serogroup. The proteins encoded by these regions consist of glycosyl transferases, flippases, polymerases, transporters and enzymes necessary for the synthesis of nucleotide sugars unique to a given capsule.[15]

Expression of a specific type of capsule in *S. pneumoniae* is important for their survival in the host blood and is strongly associated with the ability to cause invasive disease. Nowadays vaccines like seven-valent[21,22] and 23-valent polysaccharide[23] are available for pneumococcal diseases; still, there are some drawbacks with the available vaccines. Wide variations in the capsule types make most of the available vaccines ineffective. Discrepancies can occur in the conventional serotyping methods[24-27] and molecular typing has the potential to improve discrimination and provide additional information on pneumococcal

serotyping.[28] Several methods, based on molecular typing have been developed[10,11] to reduce discrepancies in pneumococcal serotyping. Therefore, it is necessary to understand the complex mechanism by which antigenic diversity arises to develop effective methods of serotyping. Effective and accurate serotyping will lead to the discovery of effective vaccines. Analyzing the sequences of the complete set of pneumococcal CPS loci is the only means by which the above goal can be achieved.[14,29]

In the post genomic revolution, the availability of all the 90 genetic loci and their individual gene sequences allow us to carry out sequence based serotyping. Already Bentley et al. (2006)[15] have analyzed all the 90 genetic loci and Tomita et al. (2011)[30] have analyzed the specific gene of glycosyl transferase through sequence based capsular typing of the pneumococcal strains. Many studies carried out from the beginning of the present century have lead to the development of simple methods of accurate serotyping utilizing the sequences of CPS locus.[10,11,15,30] In this background, the present study has been undertaken with an aim to serotype the 90 pneumococcal strains using the already available sequences of their genetic loci and polymerase gene through phylogenetic analysis.

2. MATERIALS AND METHODS

Sequences of capsular biosynthetic genetic loci of 90 serotypes were retrieved from CPS database (ftp://ftp.sanger.ac.uk/pub/pathogens/spn/cps-clusters/) in EMBL format. The above mentioned 90 sequences of genetic loci were converted into FASTA format using the tool ReadSeq and made all the sequences as a single FASTA file. Likewise, the polymerase gene sequences of 89 serotypes (except serotype 3, whose doesn't have polymerase gene) were retrieved from their corresponding genetic loci with reference to their coordinate position using the tool BioEdit. The obtained polymerase sequences were also converted into FASTA format. Then sequence composition of each genetic loci and polymerase gene were obtained using Bioedit.

In order to investigate the genetic diversity of pneumococcal strains, the above prepared sequence files of genetic loci and polymerase genes were considered for phylogenetic tree construction using MEGA4 with the Neighbor-Joining (NJ) method, choosing default parameters. Bootstrapping was done for the phylogenetic tree construction of both genetic loci and polymerase gene.

3. RESULTS AND DISCUSSION

As provided in Supplementary data 1, the experimentally determined 90 serotypes were classified through conventional serotyping into 25 individual serotypes and 21 serogroups consisting of 65 serotypes,[5] which are shown in Supplementary data 2. But discrepancies in the conventional serotyping was noticed and the molecular typing methods have been developed in order to reduce the discrepancies in serotyping.[7,9,10,15,27,30] The availability of all the annotated CPS sequences in CPS database has given room for the comparison of modern sequence based serotyping and the conventional serology based studies.[14,15,30] Therefore, the present study considered genetic loci and polymerase gene for serotyping through phylogenetic analysis.

3.1. Serogrouping from the phylogenetic analysis of genetic loci

In order to show the commonality between the CPS of different serotypes and serogroups, initially comparison was made among capsular polysaccharide genetic loci of 90 serotypes. The result shows that genes in between dexB and aliA of a genetic locus ranged in size from 10337 bp (serotype 3) to 30298 bp (serotype 38) with an average of 20714 bp with low G+C content as provided in the Supplementary data 1.

Further, in order to show the pneumococcal diversity, phylogenetic tree was created for the sequences of 90 CPS genetic loci using MEGA4 with bootstrapping and it is shown in Figure.1. From the phylogenetic tree, it is shown that the genetic loci of a few serotypes are similar and many serotypes are variable. The serogrouping of the serotypes noted from the phylogenetic tree constructed are provided in Table 1. It shows that, unlike the conventional serotyping, there are 34 (instead of 25) individual serotypes and 26 (instead of 21) serogroups consisting of 56 serotypes (instead of 65). It is due to the varied arrangement of gene clusters in gene loci of different serotypes. But the phylogenetic tree of genetic loci constructed does not provide a clear picture of the expected serotyping. Therefore, in order to group serotypes more specifically, in the present study, yet another phylogenetic analysis has been carried out using polymerase genes.

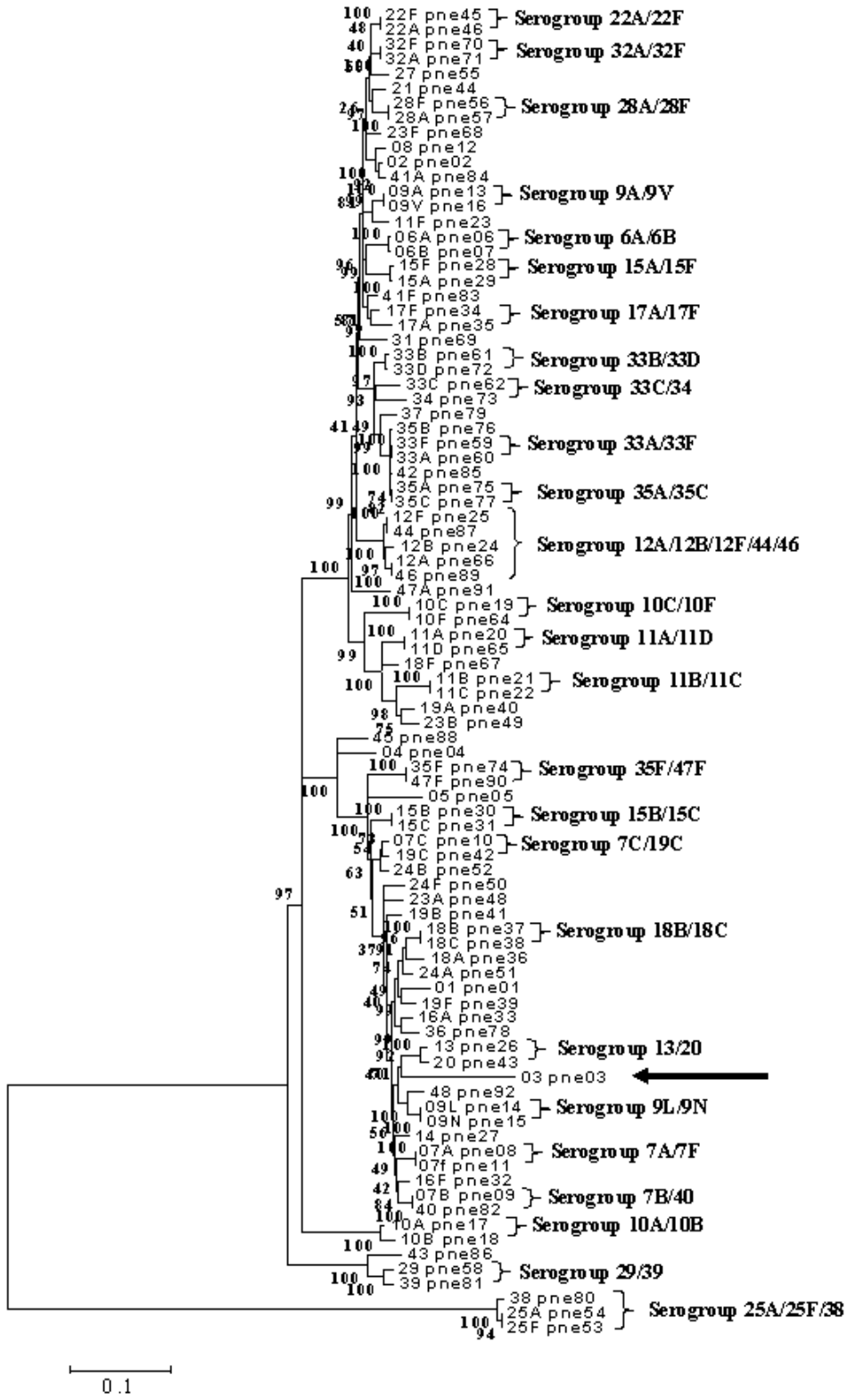

Figure 1: Phylogenetic tree of capsular polysaccharide synthesizing (CPS) genetic loci of 90 different serotypes including, 03 pne03 of *S. pneumoniae* (shown with arrow ←).

Table 1: Sixty serogroups of 90 serotypes from the CPS sequences of genetic loci through phylogenetic analysis. It includes 34 individual serotypes and 26 serogroups consisting of 56 serotypes.

Sl. No.	Individual Serotypes	Sl. No.	Serotypes in Serogroup
1	1	1	6A/6B
2	2	2	7F/7A
3	3	3	7B/40
4	4	4	7C/19C
5	5	5	9A/9V
6	8	6	9L/9N
7	11F	7	10A/10B
8	14	8	10C/10F
9	16A	9	11A/11D
10	16F	10	11B/11C
11	18A	11	12A/12B/12F/44/46
12	18F	12	13/20
13	19A	13	15B/15C
14	19F	14	15A/15F
15	19B	15	17A/17F
16	21	16	18B/18C
17	23A	17	22A/22F
18	23B	18	25A/25F/38
19	23F	19	28A/28F
20	24A	20	29/39
21	24B	21	32A/32F
22	24F	22	33B/33D
23	27	23	33A/33F
24	31	24	33C/34
25	35B	25	35A/35C
26	36	26	35F/47F
27	37		
28	41A		
29	41F		
30	42		
31	43		
32	45		
33	47A		
34	48		

3.2. Serogrouping from the phylogenetic analysis of polymerase gene (wzy)

Among all the genes in a gene cluster, the genes for glycosyl transferase, flippase and polymerase are very much capsular specific.[14] Recently Tomita et al. (2011)[30] have grouped the 90 serotypes using the highly variable gene of glycosyl transferase and have revealed that the ninety serotypes are grouped into 64 serogroups. Likewise, in order to group the pneumococcal serotypes, the present study has been carried out considering wzy, yet

another gene polymerase of 89 serotypes (excluding serotype 3, which does not have any polymerase gene) for phylogenetic analysis.

Initially, the polymerase genes of the pneumococcal serotypes were compared for their gene length and G+C content which are provided in the Supplementary data 1. It reveals that the length of the polymerase genes ranges from 1068 bp (serotype 20) to 1452 bp (serotype 7A). Comparison of G+C percentage content shows the range from 23.95% (serotype 43) to 36.40% (serotype 5). Generally, the serotype specific gene, wzy tend to have A+T rich sequences.

The bootstrapped phylogenetic tree of 89 polymerase gene sequences was constructed using MEGA4[31] with Neighbor-Joining method[32] and is shown in Figure 2. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary relationships of the taxa analyzed.[33] Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches.[33] The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method[34] and are in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+ Noncoding. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 343 positions in the final dataset.

Based on the sequence identity, the serotypes are grouped manually and are labeled. The obtained bootstrap consensus tree with clusters is shown in Figure 2 with their corresponding percentage of replicate trees encircled. Through visual inspection, the phylogenetic tree shows that polymerase genes are also suitable targets for serotype / serogroup identification. Comparison of polymerase gene sequences of different serotypes reveals that very few of them show close similarity while the others show variations.

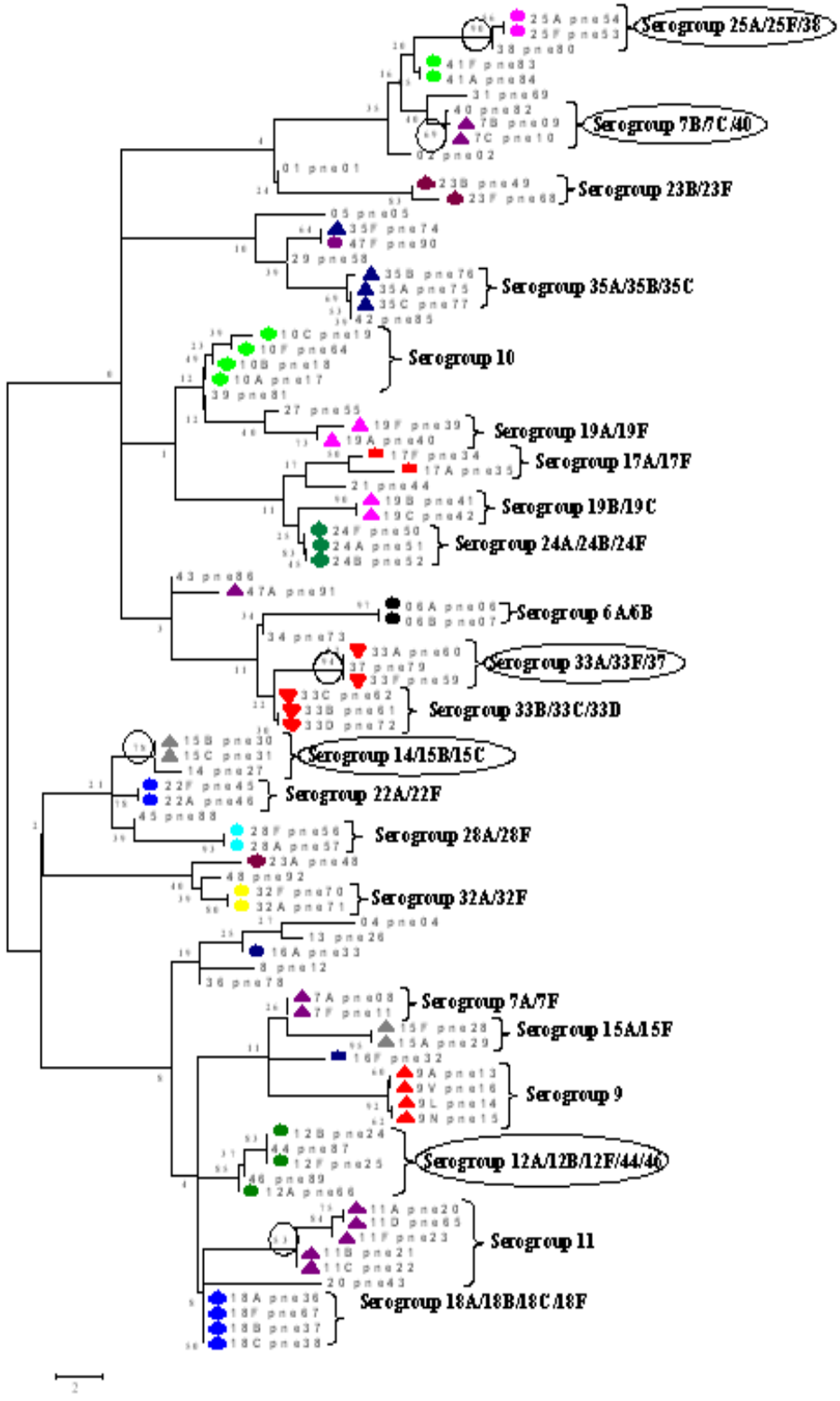

Figure 2: Phylogenetic tree of gene encoding polymerase (wzy) of 89 serotypes of *S. pneumoniae* excluding 03 pne03 (Note: Serotype 3 (03 pne03) does not have polymerase gene in its genetic loci) was inferred with NJ method (Bootstrap consensus tree inferred from 500 replicates) using MEGA4. Clusters were made with more than 50% bootstrap replicates (encircled) between sequences.

Table 2: Forty six serogroups of 89 serotypes from the sequences of polymerase (wzy) gene through phylogenetic analysis. It includes 27 individual serotypes and 22 serogroups consisting of 62 serotypes.

Sl. No.	Individual Serotypes	Sl. No.	Serotypes in Serogroup
1	1	1	6A/6B
2	2	2	7F/7[a]
3	4	3	7B/7C/40
4	5	4	9A/9L/9N/9V
5	8	5	10A/10B/10C/10F
6	13	6	11A/11B/11C/11D/11F
7	16[a]	7	12A/12B/12F/44/46
8	16F	8	14/15B/15C
9	20	9	15A/15F
10	21	10	17A/17F
11	23[a]	11	18A/18B/18C/18F
12	27	12	19A/19F
13	29	13	19B/19C
14	31	14	22A/22F
15	34	15	23B/23F
16	35B	16	24A/24B/24F
17	35F	17	25A/25F/38
18	36	18	28A/28F
19	39	19	32A/32F
20	41[a]	20	33A/33F/37
21	41F	21	33B/33C/33D
22	42	22	35A/35C
23	43		
24	45		
25	47[a]		
26	47F		
27	48		

The phylogenetic tree of polymerase gene sequences constructed in the present study reveals a few discrepancies when compared with the results of conventional serotyping. Conventional serotyping revealed 90 serotypes with 46 serogroups among which 25 are individual serotypes and the other 65 serotypes grouped into 21 serogroups [Supplementary data 2]. The present study shows that the 89 serotypes can be categorized into 27 (instead of 25) individual serotypes and 22 (instead of 21) serogroups consisting of 62 (instead of 65) serotypes (Table 2). These discrepancies highlight the variations exhibited by the serotypes of the same serogroup. The study observed the presence of some genetic differences within the serogroups. The present study also revealed that unlike conventional method, serotype 14 is grouped with the serogroup 15B/15C. As the serotype 14 is considered for 7 valent vaccines, it can also be useful for treating infection of the serotypes 15B/15C.

For instance, as provided in Table 2 and Figure 2, the present study reveals some discrepancy in serotyping from conventional serotyping and they are given as follows. Serogroup 7 is diversified into 7F/7A and 7B/7C instead of the conventional serotyping 7F/7A/7B/7C. Serogroup 15 is diversified into 15B/15C and 15A/15F instead of 15B/15C/15A/15F. Serogroup 16 is diversified into 16A and 16F instead of 16A/16F. Serogroup 19 is diversified into 19A/19F and 19B/19C instead of 19A/19F/19B/19C. Serogroup 23 is diversified into 23A/23B/23F instead of 23A and 23B/23F. Serogroup 33 is diversified into 33A/33F and 33B/33C/33D instead of 33A/33F/33B/33C/33D. Similarly, serogroup 35 is diversified into 35A/35B/35C and 35F instead of 35A/35B/35C/35F. Likewise, the present study also revealed the diversification of serogroup 47 into 47A and 47F instead of the conventional serotyping 47A/47F.

Similar discrepancies were also reported by various studies with regard to different genetic loci.[10,11,15,30] Serotypes from the same serogroups have subtle differences and provide an example of multiple serotypes arising by divergence from a single CPS locus. But, their CPS genes fall into two or three pairs, the pairs differing significantly in sequences, as seen in Figure 2. This suggests an initial divergence to form two ancestral serotypes as also discussed by Bentley et al. (2006).[15] The divergence of a serotype from its serogroup may be due to the accumulation of point mutations, insertions and deletions of a single gene resulting in slightly different CPS structures.

Another remarkable observation made during the present analysis is the commonality of the CPS of different serotypes and serogroups. Comparison studies reveal the existence of such strong commonality between the CPS of different serotypes and serogroups (Figure 2). In contrast to the results of the conventional serotype studies, showing wide variations among the different serotypes, the present study shows clear-cut similarities among them. Illustrative examples demonstrating the combination of sequences in order to analyze the CPS loci are shown in Figures 3, 4 and 5 and are discussed below.

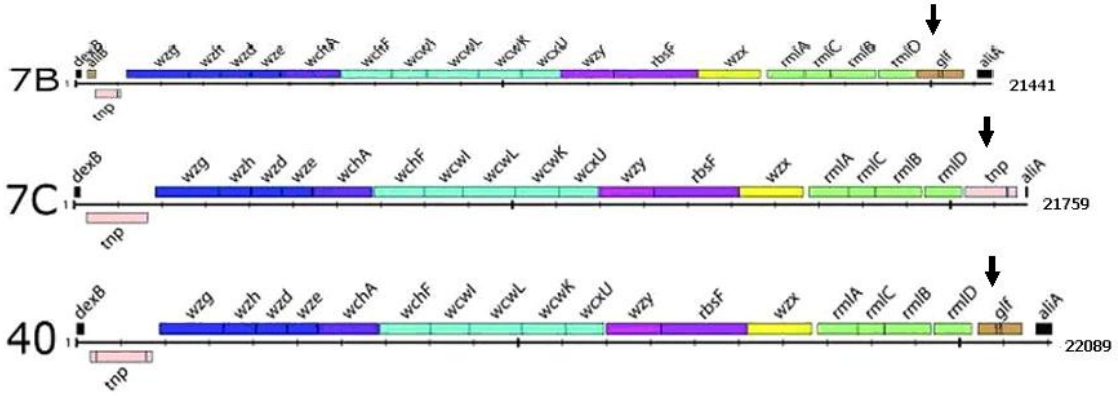

Figure 3:Comparison of the serotype 40 with the serogroup 7 (7B and 7C). The genes tnp and glf, which are differing among serotypes, are marked with arrows (↓). The figure is not to scale.

a. Serotype 40 is related to serogroup 7 (7B and 7C)

Present phylogenetic analysis of complete genetic loci and polymerase gene of 89 serotypes shows that the CPS gene clusters of serogroup 7 (7F/7A) and serotype 40 are almost identical, differing only in transposase (tnp) and glf genes (Figure 3). Figure 3 also shows that the sequence length of 7B, 7C and 40 are nearly same (21500-22000 bp) and the genes involved in their capsular synthesis also similar. The sequence identity of wzy found between serotypes 40 and 7B is 98%, serotype 40 and 7C is 98% and serotype 7B and 7C is 99% that reflects their close resemblance.

The above result perhaps is not surprising as serogroups may be defined by common epitopes. This is one of the examples of common ancestry that is not apparent from serology. But the conventional serotyping showed no factor serum cross reactions with these serotypes and the method does not bring out the commonality of these serotypes. Similar result was also obtained through sequential multiplex PCR approach.[10]

b. Serotype 38 is related to serogroup 25 (25F and 25A)

The present study shows that serotype 38 shares significant sequential similarity with the serogroup 25. But no significant serological cross reaction was found through conventional serotyping. All the genes including regulatory and transposase are very.

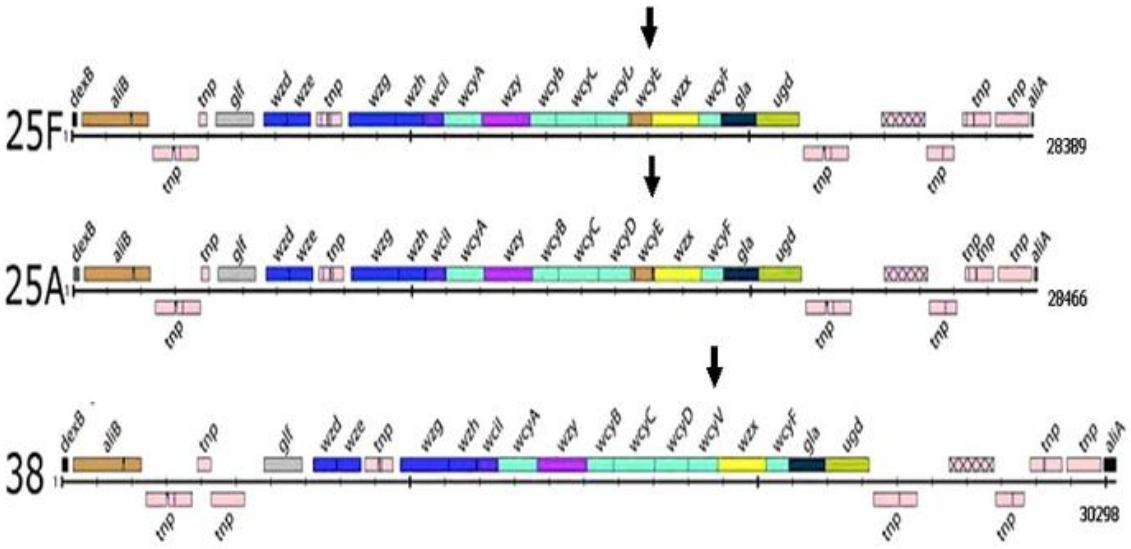

Figure 4: Comparison of the serotype 38 with the serogroup 25 (25F and 25A). The gene wcyV is differing among serotypes are marked with arrows (↓). The figure is not to scale.

Much similar in serotype 38 and serogroup 25 except the gene wcyV, which is present only in the serotype 38; but wcyE is absent in it (Figure 4). It is additionally supported by the sequence identity of polymerase gene of serotypes 25F and 25A being 100%, 25A and 38 being 99% and that of 25F and 38 being 99%, reflecting their close resemblance. Similar result was obtained by the sequential multiplex PCR analysis of capsular typing.[10]

c. Serotype 37 is related to serogroup 33 (33F and 33A)

Capsular typing of CPS polymerase gene sequences through phylogenetic analysis reflects the existence of homogenecity between serotype 37 and serogroup 33 (33A and 33F). The genetics of these three serotypes shows that all the genes are very much identical (Figure 5). Sequence identity of polymerase gene found between serotypes 33F and 33A is shown to be 100%, 33A and 37 is 99% and that between 33F and 37 is 99% indicating their close resemblance. Unlike earlier cases, genes in these serotypes are identical and there is no difference in gene arrangement (Figure 5). Similar results were obtained through the identification of serotyping using DNA microarray[11] and cpsB sequence polymorphism and serotype specific PCR.[7]

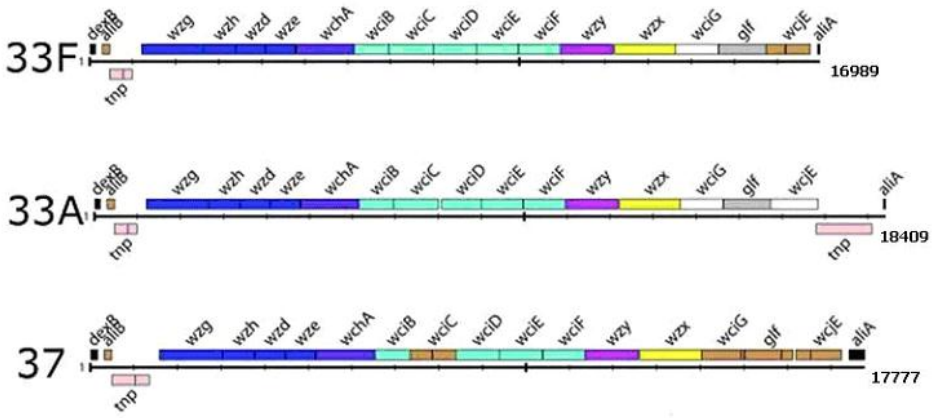

Figure 5: Comparison of serotype 37 with the serogroup 33 (33F and 33A). Genes among serotypes are identical. The figure is not to scale.

Detailed genetic analysis and comparison of gene clusters of different serogroups involved in CPS synthesis have been done by Bentley et al. (2006).[15] They have discussed elaborately about the discrepancies and similarities found in serogrouping of 12A/12B/12F/44/46 and serogroup 14/15B/15C in comparison with conventional serotyping. The results and findings of the present study go hand in hand with that of Bentley et al. (2006).[15]

4. CONCLUSIONS

Accurate serotype determination of *S. pneumoniae* is of critical importance as vaccine development presently relies on serotype prevalence data.[35] Among ninety different serotypes, only 23 serotypes are responsible for 90% of all cases of pneumococcal diseases [36]. The emergence of antibiotic resistance and the spread of resistant strains have increased the importance of vaccines as a primary prevention. The drawback in developing very effective vaccine lies with the variation in the types of pneumococcal capsules. Therefore, typing of pneumococcal strains into correct serotypes and bringing them together in accurate serogroups barring discrepancies is an inevitable in developing effective vaccines.

Phylogenetic analyses present here on the complete genetic loci of pneumococcal serotypes shows variations in the arrangement of gene cluster in the gene loci of different serotypes. Phylogenetic analysis of polymerase gene provides few examples of common ancestry that are not revealed by conventional serological studies. As far as the serogroups 7, 9, 15, 16, 19, 23, 33, 35, 41 and 47 are concerned, the present serotyping reveals different results from that of conventional studies. These serogroups provide an example of multiple serotypes arising

by divergence from a single CPS locus. But within serogroups, their CPS genes fall into two or three pairs. For example, serotype 19A is highly similar to 19F; and 19B is highly similar to 19C, but with the two pairs differing significantly in sequence. It may be due to the accumulation of point mutations, insertions and deletions of a single gene within a CPS locus.

Similarly, the new serogroupings such as serogroup 7B/7C/40; serogroup 25A/25F/38, serogroup 33A/33F/37, 12A/12B/12F/44/46 and serogroup 14/15B/15C, obtained from the present study provide some examples of common ancestry that is also not apparent from serology. Few of the above results were supported through analysis of CPS genetic loci sequences and their known structures[15] and phylogenetic analysis of glycosyl transferase.[30] One reason for the capsular diversity is the transposition-like events, and it is evidenced by the fact that all the capsular gene clusters of *S. pneumoniae* are flanked by insertion sequence elements.[37] As the serotype 14 is grouped with the serogroup 15B/15C and serotype 14 is considered for 7 valent vaccines, it can also be useful for treating infection of the serotypes 15B/15C.

To conclude, the present phylogenetic analysis using the available sequences of the CPS loci and polymerase gene sequences of pneumococcal serotypes provides an understanding of spread of antigenic variants due to differences in the polysaccharides found at the cell surface. Further extensive study on the serotyping is indeed will be an interesting candidate for the design of a conjugate vaccine especially useful for DNA based vaccine design.

ACKNOWLEDGEMENT

One of the authors R. Jothi thank her research supervisor Dr. S. Parthasarathy, Associate Professor and Head, Department of Bioinformatics, Bharathidasan University, Tiruchirappalli – 600 024 for assigned the present work.

REFERENCES

1. Sheppard CL, Kapatai G, Broughton K, Schaefer U, Hannah M, Litt DJ and Fry NK, Clinical streptococcal isolates, distinct from Streptococcus pneumoniae, but containing the β-glucosyltransferase tts gene and expressing serotype 37 capsular polysaccharide. PeerJ, 2017; 5: e3571; DOI 10.7717/peerj.3571.
2. Avery OT, MacLeod CM, McCarty M Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a

deoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med, 1944; 79(2): 137-158.

3. Dhian RA, Camargo, Fabiano S Pais, Ângela C Volpini, Marluce AA Oliveira, Roney S Coimbra Revisiting molecular serotyping of Streptococcus pneumoniae, BMC Genomics, 2015; 16(Suppl 5): S1, 1-10.
4. Raymond F, Boucher N, Allary R, Robitaille L, Lefebvre B, et al. Serotyping of *Streptococcus pneumoniae* based on Capsular Genes Polymorphisms. PLoS ONE, 2013; 8(9): e76197. doi:10.1371/ journal.pone.0076197.
5. Henrichsen J, Six newly recognized types of *Streptococcus pneumoniae*. J Clin Microbiol, 1995; 33(10): 2759-2762.
6. Kellogg JA, Bankert DA, Elder CJ, Gibbs JL, Smith MC, Identification of *Streptococcus pneumoniae* Revisited. J Clin Microbiol, 2001; 39(9): 3373–3375.
7. Kong F, Gilbert GL, Using cpsA-cpsB sequence polymorphisms and serotype-/group-specific PCR to predict 51 *Streptococcus pneumoniae* capsular serotypes. J Med Microbiol, 2003; 52: 1047–1058.
8. Brito DA, Ramirez M, de Lencastre H. Serotyping *Streptococcus pneumoniae* by multiplex PCR. J Clin Microbiol, 2003; 41(6): 2378–2384.
9. Kong F, Wang W, Tao J, Wang L, Wang Q, Sabananthan A, Gilbert GL, A molecular-capsular-type prediction system for 90 *Streptococcus pneumoniae* serotypes using partial cpsA-cpsB sequencing and wzy or wzx-specific PCR. J Med Microbiol, 2005; 54(4): 351–356.
10. Pai R, Gertz RE, Beall B, Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. J Clin Microbiol, 2006; 44(1): 124-131.
11. Wang Q, Wang M, Kong F, Gilbert GL, Cao B, Wang L, Feng L Development of a DNA microarray to identify the *Streptococcus pneumoniae* serotypes contained in the 23-valent pneumococcal polysaccharide vaccine and closely related serotypes. J Microbiol Methods, 2007; 68(1): 128–136.
12. Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG, Predicted functions and linkage specificities of the products of the *Streptococcus pneumonia* capsular biosynthetic loci. J Bacteriol, 2007; 189(21): 7856–7876.
13. Kapatai et al. Kapatai G, Sheppard CL, Al-Shahib A, Litt DJ, Underwood AP, Harrison TG, Fry NK, Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. PeerJ, 2016; 4: e2477; DOI 10.7717/peerj.2477.

14. Yother J. Capsule. In: Tuomanen EI (ed) The Pneumococcus, ASM press, Washington, 2004; 30–48.
15. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, KaltoftMS, Barrell B, Reeves PR, Parkhill J, Spratt BG, Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet, 2006; 2(3): 262-269.
16. Dillard JP, Vandersea MW, Yother J, Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. J Exp Med, 1995; 181(3): 973–983.
17. Llull D, Muñoz R, López R, García E, A single gene (tts) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide type 37 Pneumococci are natural, genetically binary strains. J Exp Med, 1999; 190(2): 241-252.
18. Cartee RT, Forsee WT, Jensen JW, Yother J, Expression of the *Streptococcus pneumoniae* type 3 synthase in Escherichia coli. Assembly of type 3 polysaccharide on a lipid primer. J Biol Chem**,** 2001; 276(52): 48831–48839.
19. Kolkman MA, van der Zeijst BA, Nuijten PJ, Diversity of capsular polysaccharide synthesis gene clusters in *Streptococcus pneumoniae*. J Biochem, 1998; 123(5): 937–945.
20. García E, Llull D, Muñoz R, Mollerach M, López R Current trends in capsular polysaccharide biosynthesis of *Streptococcus pneumoniae*. Res Microbiol, 2000; 151(6): 429–435.
21. Black S, Shinefield H, Fireman B, Lewis E, Ray P, Hansen JR, Elvin L, Ensor KM, Hackell J, Siber G, Malinoski F, Madore D, Chang I, Kohberger R, Watson W, Austrian R, Edwards K, Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Pediatr Infect Dis J, 2000; 19(3): 187–195.
22. Reinert RR, Pneumococcal conjugate vaccines—a European perspective. Int J Med Microbiol, 2004; 294(5): 277–294.
23. Poolman JT, Pneumococcal vaccine development. Expert Rev Vaccines, 2004; 3(5): 597–604.
24. Heidelberger M, Precipitating cross-reactions among pneumococcal types. Infect Immun, 1983; 41(3): 1234–1244.
25. Barker JH, Musher DM, Silberman R, Phan HM, Watson DA, Genetic relatedness among nontypeable pneumococci implicated in sporadic cases of conjunctivitis. J Clin Microbiol, 1999; 37(12): 4039–4041.

26. Henrichsen J, Typing of *Streptococcus pneumoniae*: past, present, and future. Am J Med, 1999; 107(1): 50S–54S.
27. Lawrence ER, Arias CA, Duke B, Beste D, Broughton K, Efstratiou A, George RC, Hall LMC Evaluation of serotype prediction by cpsA-cpsB gene polymorphism in *Streptococcus pneumoniae*. J Clin Microbiol, 2000; 38(4): 1319–1323.
28. Hall LM, Application of molecular typing to the epidemiology of *Streptococcus pneumoniae*. J Clin Pathol, 1998; 51(4): 270–274.
29. Croucher NJ, Finkelstein JA, Pelton SI, Parkhill J, Bentley SD, Lipsitch M, Hanage WP, Population genomic datasets describing the post-vaccine evolutionary epidemiology of *Streptococcus pneumoniae*. Sci. Data, 2015; 2: 150058 doi: 038/sdata. 58.
30. Tomita Y, Okamoto A, Yamada K, Yagi T, Hasegawa Y, Ohta M, Capsular-type prediction by phylogenetic tree of glycosyltransferase gene polymorphism in *Streptococcus pneumoniae*. Open Access Bioinformatics, 2011; 3: 67–73.
31. Tamura K, Dudley J, Nei M, Kumar S MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol**,** 2007; 24(8): 1596-1599.
32. Saitou N, Nei M, The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol, 1987; 4(4): 406-425.
33. Felsenstein J, Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 1985; 39(4): 783-791.
34. Tamura K, Nei M, Kumar S, Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA, 2004; 101(30): 11030-11035.
35. Hausdorff WP, Bryant J, Paradis PR, Siber GR, Which pneumococcal serogroups cause the most invasive disease? Implications for conjugate vaccine formulation and use, part I. Clin Infect Dis**,** 2000; 30(1): 100-121.
36. Bridy-Pappas AE, Margolis MB, Center KJ, Isaacman DJ *Streptococcus pneumoniae*: description of the pathogen, disease epidemiology, treatment and prevention, Pharmacotherapy, 2005; 25(9): 1193–1212.
37. Lopez R, Pneumococcus: the sugar-coated bacteria. Int Microbiol, 2006; 9(3): 179-190.