Development and Evaluation of A High Throughput Automated DNA Microarray For Serotyping *S. Pneumoniae.*

Authors
Vandana Govindan, Feroze A Ganaie, Avid Hussain, Geeta Nagaraj, K L Ravi Kumar
Central Research Laboratory, Kempegowda Institute of Medical Sciences Hospital and Research Centre,
K R Road, V V Puram, Bangalore-560004
Corresponding Author
K L Ravi Kumar
Email: *nm.vandana@gmail.com*, *klravikumar@gmail.com*

ABSTRACT

Streptococcus pneumonia is a major human pathogen commonly responsible for pneumonia, bacteraemia, meningitis, and otitis media, especially among young children and older adults. Pneumococcal serotype identification is a prerequisite to understand pathogenesis and estimate the disease burden. It is also a basic feature needed to track changes in epidemiology as well as to optimize the vaccines. Conventional gold standard typing method (Quellung reaction) requires the use of a large panel of expensive anti-sera, is subjective with cross reactions and few strains are non type able.

Modern molecular methods have provided innovative opportunities to address the problem of microbial detection. The drawbacks of some of these currently used methods include inadequate discrimination between the isolates and poor reproducibility.

To contend with these issues a novel automated microarray assay based on Capsular polysaccharide genes was developed to track the 90 serotypes of S. pneumoniae. A set of 15,150 specific probes of 60bp length each, corresponding to 1680 genes was generated utilizing the sequence data source from Sanger institute/EMBL. The assay was validated on 90 standard strains sourced from Staten Serum Institute, Copenhagen. The results demonstrated discriminatory ability of the Microarray among 90 serovars. Overall 61 (67.7%) serotypes were uniquely identified and 29 (32.2%) were assigned with a combination of their homologous serotypes. Bioinformatics data analysis revealed that detection of homologous serotypes is due to homology in the CPS region..

INTRODUCTION

Streptococcus pneumoniae is a bacterium that causes infections, ranging from serious diseases such as meningitis, septicaemia and pneumonia to milder but commoner sinusitis and otitis media. Pneumococcal diseases are a cause of morbidity and mortality worldwide, with rates of disease and death higher in developing countries than in industrialized country settings. Mortality for bacteremic pneumococcal pneumonia ranges from 10 to 30% in adults and up to 3% in children.[1]

Capsule, a layer consisting of polysaccharide chain that surrounds pneumococci is the primary virulence factor by its ability, to inhibit phagocytosis and to protect from lytic enzymes. To date, 94 immunologically distinct capsular serotypes have been described. The serotypes are dictated by the order and type of the monosaccharide units within the polysaccharide chain containing multiple sugars, linkages and side chains[2]. Serotypes vary in their ability to cause invasive disease and loss of capsule by mutation renders the bacterium a virulent [3].

The genes responsible for capsular synthesis are located in cps gene cluster which are flanked upstream by dex B and downstream by ali A genes common to all types [4]. The genes of this cluster are organized as common and type specific regions. The common region contains genes cpsA (wzg)–cpsB (wzh)–cpsC (wzd)–cpsD (wze)and in most types cps E. These genes encode proteins that are involved in the modulation of capsule production and synthesis. (3,38,50,80 from pneumococcus capsule).Genes contained in the type specific regions are unique to a given serotype or serogroup. The proteins encoded by these regions consist of glycosyl transferase, polymerases, transporters and enzymes necessary for the synthesis of nucleotide sugars unique to a given capsule[6]. The sequence data of 90 *S.pneumoniae*cps gene clusters is available at Sanger website (http://www.sanger.ac.uk/ Projects /*S.pneumoniae*/CPS/).

Pneumococcal sero group and type identification is currently performed using large panels of expensive anti sera by various methods, including the capsular swelling (Quellung) reaction, latex agglutination, and co agglutination.[7] The gold standard, Quellung reaction is labor intensive, cross reactive with a significant number of isolates being non typeable, restricting its use to specialized research laboratories. [8]

A variety of molecular serotyping methods are now employed to overcome these difficulties. Some of these methods include multiplex real time PCR, DNA microarray, Pulse Field Gel Electrophoresis, Ribotyping, Amplified Fragment length polymorphism analysis, Multilocus sequence typing and pyrosequencing. Most DNA-based methods are prone to amplification biases, false positive detections, erroneous quantification of relative abundances and allow the identification of a limited number of serotypes or serogroups.[9]

DNA microarrays offer the latest technological advancement for multi-gene detection and diagnostics. In addition, their flexibility and high throughput capabilities hold tremendous potential for pathogen identification, quantification and multiple serotype detection in molecular diagnostic laboratories. Herein, we put for wardan innovative serotyping approach that relies on the positions of unique genes located in the capsular operon. The custom oligonucleotide microarray was experimentally validated using 57- to 60-mer hybridization probes targeting the genes. Due to advances in fabrication, robotics, and bioinformatics the future role of this robust technology in diagnostic microbiology is indisputable.

MATERIALS AND METHODS

Bacterial strain, growth conditions and genomic DNA extraction:

90 characterized strains of *S. pneumoniae* were obtained from the Statens Serum Institute, Copenhagen, Denmark. Strains were cultivated on brain-heart infusion broth, streaked onto 5% Sheep Blood agar plate and incubated for 24 h at 37^0C in 5% CO_2. Genomic DNA was extracted using Qiagen DNA blood mini kit (Qiagen, Germany).The standard manufacturer's extraction protocol was followed .The suspension was centrifuged to pellet the bacteria, re-suspended in 180µL of lysis buffer and incubated at 37ºC for 60 min. Proteinase K (20µL) and Qiagen buffer AL (200µL) were added to the mixture and incubated for another 60 min at 56ºC. 4µL RNase A was added and incubated at room temperature for 5 min before incubation at 70ºC for 10 min. For the study fluorometric (Qubit) and spectrophotometric (Nanodrop) methods were utilized to access

gDNA purity and concentration. Extracted DNA was stored at 4ºC for analysis.

Probe Designing

Oligonucleotide probes were designed using Primer 3 and sure design software.The sequences have been analyzed for their GC content. Ninety percent of these sequences have a GC content of 30-45%. The probes sequences are listed here. Probes were designed based on the targeted genes for the *S. pneumoniae* groups.

The custom microarray comprised 15167 marker gene probes with 87 internal controls, 72 reserve negative controls, 316 Biological negative controls, 140 structural negative controls and 49 dark corners. For each marker gene probe three technical replicates were randomly distributed on the 8x15K format chip (Agilent technologies, Santa Clara, USA)

DNA labelling

2.5 µL of random primer was added to each reaction tube containing 13 µL of gDNA to make a total volume of 15.5 µL. Sample tubes were transferred to a heat block at 95°C for 3 minutes, then moved to ice for 5minutes.Following the incubation the samples were centrifuged at 6000g for 1 min.DNA was then labelled by adding 9.5µL of labelling master mix including 2.5 µL of 10X dNTP Nucleotide Mix, 1.5µL Cy3 or Cy5-dUTP and 0.5µL of Exo-Klenow Fragment. Sample tubes were thermo cycled at 37^0C for 2 hours, 65^0C for 10 minutes and on hold at 4^0C.Labelled DNA was then purified to determine the yield and specific activity.

Probe hybridization and washing

29 µL of the hybridization master mix was added to probe mixture making it to a total volume of 45 µL. The mixture was incubated at 95^0C for 3 mins and 37^0C for 30 mins and loaded onto the microarray slide and hybridized at 65^0C for 24 hrs. Washed microarray slide was scanned.

Data Analysis

The hybridized microarray was scanned with a laser at 635 nm using the G4900 DA scanner. The signal and background intensities of each spot were quantified using Agilent Feature extraction software. An in-house model was developed with Microsoft excel for data analysis taking into consideration the dependencies between serotypes induced by genes they have in common, and by homologous genes. Empirical Bayesian model explained by Richard Newton et al [10] was adopted for analyzing serotyping results and to calculate the probabilities of combinations of serotypes.

RESULTS

A total of 15,150 oligonucleotide probes were designed for this analysis including positive and negative controls. The entire array of 2000 genes present in the CPS locus was taken into account. 80-300, oligonucleotide probes were devised from published sequences for identification of all serotypes. For each genotyped position, probes with the highest signal to background ratio was considered. Positivity of the 528 internal control probes for all the strains tested confirmed the test validity.

In this experiment, 61 serotypes of S. pneumoniae were specifically and distinctively identified. 13 types present in the conjugate vaccine were identified, 11 to the type level, 6A and 6B to the group level. All 23 types in the polysaccharide vaccine were identified to the type level. 29 homologous types constituting 10 serogroups couldn't be distinctly identified due to high sequence similarity. The concordance between the known standard strain and Microarray results was 100%.

DISCUSSION

Identifying the serotype of *S.pneumoniae* provides an invaluable tool to monitor the impact of vaccine introduction, by revealing the association of particular serotype with carriage or invasive disease besides providing surveillance for

serotype replacement. Conventionally, serotyping of pneumococciis performed by culture followed by Quellung reaction which is not widely feasible due to the high cost of antisera, difficulties in interpretation, and requirement of technical expertise. Various molecular techniques are employed to overcome the drawback of culture-based procedures. Microarray assays have the unparalleled ability to simultaneously identify and serotype the pneumococciin a rapid and sensitive manner. We put forward a novel microarray assay that relies on a combination of genetic properties to identify the serotypes.

Previously to type *S.pneumoniae*, diagnostic array of oligonucleotide probes targeting species specific variable regions were utilized. The 16S rRNA marker has a long history in DNA based microbial identification, but its use on microarray adds the complexity. First, the specific region of 16S rDNA may be prone to frequent mutations, and a false negative result will be produced by a strain with random mutations in this region. In addition, the strains of different species may cross-hybridize, as the 16S sequence is, by definition a ''universal'' marker and is quite well conserved (11).Roth et al employed gyrB and parE genes on microarray platform only to identify *S.pneumoniae*.

CPS gene cluster encodes for polysaccharides and proteins specific for synthesis, transfer and assembly of capsule that define serotypes. These genes are highly variable and type specific. The capsular serotype specific genes are targeted for identification and intra-species differentiation of small groups or individual pneumococcal sero-types. Wang et al [12] targeted wzy or capA genes which proved to be specific for detection and partial serotyping of 43 serotypes. Tomita et al in the year 2011[13] described the use of GT genes for serotype recognition. In their study of the 23 strains representing 23 valent vaccine serotype, 4 strains hybridized to almost all the specific set of probes. The presence of certain GT genes (for example, *wciB*, *wcrC*, *wciF*, *wcrD*, and *wciE*) and the commonalities in structure at the 3′ end of

the *cps* loci of these serotypes account for their non specific hybridization. In a study by Raymond et al [13] 12 polymorphisms located in the capsular operon were considered allowing the identification of 22 serotypes at the serotype level and the assignation of 24 other serotypes to subgroup level. The multi carriage study by Satzke et al utilized the BµG@S SP-CPS v1.4.0 microarray[14] as one of the twenty different pneumococcal serotyping methods. The microarray was able to detect all known serotypes and accurately measure their percent relative abundance apart from determining the genetic relatedness of isolates.

Homologous genes, although not identical, are similar to each other in their DNA sequences, resulting in cross hybridisation. Evolutionarily some of the cps genes are homologous. Divergence of a single *cps* locus results in multiple serotypes with subtle differences. In microarray assay the intensities of the probes for a particular gene may be elevated not by the presence of the gene they were designed to target, but by the presence of a homologous gene.[15] In the present study, we validated the identification of 90 serotypes, 61 of which were precisely identified and 29 with homologous types. In the next generation chip we aim to add probes targeting serotype-specific sequences for the 29 types assigned to 10 homologous groups to accurately differentiate members of these cross-reacting groups. We further aim to add probes to determine antimicrobial resistance and to calculate the relative abundance.

Although the Quellung reaction is still considered the gold-standard, laboratories should take into account the number of pneumococcal isolates, the type of samples to be used for testing, the time frame for the results and the resources available in order to select the most appropriate method.

CONCLUSION

The customized oligonucleotide microarray was able to detect all these rotypes. However, for homologous serotypes auxiliary study to distinguish serotypes sharing similar sequences is

required and is currently under investigation. The specificity, reproducibility and flexibility of detection demonstrated in this study make this approach attractive for clinical diagnostics. Most likely, a combination of phenotypic and genotypic methods would be optimal to monitor and evaluate the impact of pneumococcal conjugate vaccines and to provide information for future vaccine formulations.

REFERENCES

1. http://www.medscape.com/viewarticle/702635_14

2. The Pneumococcus.Tuomanen E I, Mitchell T J, Morrison D A, Spratt B G Washington: American Society for Microbiology Press; 2004

3. Song JY, Nahm MH, Moseley MA. Clinical Implications of Pneumococcal Serotypes: Invasive Disease Potential, Clinical Presentations, and Antibiotic Resistance. J Korean Med Sci.2013; 28(1):4-15.

4. Morona JK, Paton JC, Miller DC, Morona R.Tyrosine phosphorylation of CpsD negatively regulates capsular polysacch-aride biosynthesis in *streptococcus pneumoniae*. Mol Microbiol. 2000;35 (6):1431-42.

5. Morona JK,Morona R and Paton JC. Characterization of the locus encoding the *S.pneumoniae*type 19F capsule biosynthe-tic pathway. Mol. Microbiol. 1997;23:751-763

6. Jannet Yother. Capsules. The Pneumococcus. American Society for Microbiology Press; 2004.

7. Lafong, A. C., and E. Crothers. Simple latex agglutination method fortyping pneumococci. J. Clin. Pathol. 1988; 41:230–231.

8. Brito D. A., Ramirez M and de LencastreH. Serotyping *Streptococcus pneumoniae* by multiplex PCR.J. Clin. Microbiol.2003; 41:2378–2384.

9. Ahn J, Yang L, Paster BJ, Ganly I, Morris L, et al. Oral microbiomeprofiles: 16S rRNA pyrosequencing and microarray assay comparison. PloS one. 2011; 6: e22788.

10. Newton R, Hinds J and Wernisch L. Empirical Bayesian models for analysing molecular serotyping microarrays.BMC Bioinformatics. 2011; 12: 88.

11. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences. 1985;82: 6955–6959.

12. WangQ ,Wang M, KongF, Gilbert G L, Cao B ,Wang L et al. 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; 128–136

13. Raymond F, Boucher N, Allary R, Robitaille L, Lefebvre B, Tremblay C et al. Serotyping of *Streptococcus pneumoniae* Based on Capsular Genes Polymorphisms. 2013;PLoS ONE 8: e76197.

14. Satzke C, Dunne EM, Porter BD, Klugman KP, Mulholland EK and Pneumo carriage project group. The Pneumo carriage project: A Multicenter Comparative Study to Identify the Best Serotyping Methods for Examining Pneumococcal Carriage in vaccine Evaluation Studies. PLoS Med. 2015;12(11): e1001903.

15. Newton R, Hinds J and Wernisch L. Empirical Bayesian models for analysing molecular serotyping microarrays.BMC Bioinformatics. 2011; 12: 88.