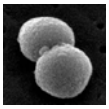
Molecular serotyping of *Streptococcus pneumoniae*: a microarray-based tool with enhanced utility.

Hinds, J.[1], Gould, K.A.[1], Witney, A.A.[1], Baldry, S.J.[2], Lambertsen, L.[3], Hanage, W.P.[4], Antonio, M.[5], Turner, P.[6,7], Newton R.H.[8], Wernisch, L.[8], Aanensen, D.M.[4] and Bentley, S.D.[2]

[1]St. George's, University of London, UK. [2]Wellcome Trust Sanger Institute, Cambridge, UK. [3]Statens Serum Institut, Copenhagen, Denmark. [4]Imperial College, London, UK. [5]Medical Research Council Laboratories, Fajara, The Gambia. [6]Shoklo Malaria Research Unit, Mae Sot, Thailand. [7]Centre for Tropical Medicine, University of Oxford, UK. [8]MRC BioStatistics Unit, Cambridge, UK.

Introduction



Streptococcus pneumoniae is an important bacterial pathogen and is a major cause of morbidity and mortality worldwide, particularly in the young and elderly. The pneumococcus causes a range of diseases from acute otitis media and sinusitis to more severe pneumonia, septicaemia and meningitis.

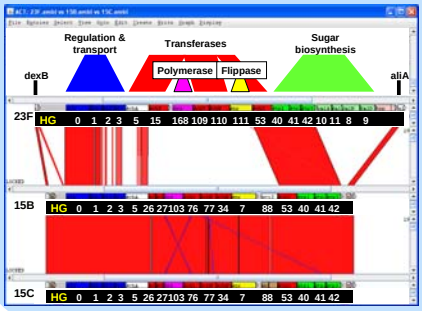
To date, 91 different capsular serotypes have been identified with each serotype expressing a serologically distinct polysaccharide capsule. This capsule represents the interface between bacterium and host and so is associated with immunity and invasive disease and forms the basis of polyvalent vaccines currently available.

Identifying the serotype of individual strains provides an invaluable tool to monitor the impact of vaccine introduction, by providing surveillance for serotype replacement, as well as revealing the association of particular serotypes with carriage or invasive disease.

Basis of molecular serotyping by microarray

- Biosynthesis and assembly of the capsular polysaccharide (CPS) is achieved by a repertoire of proteins encoded by *cps* genes clustered in a defined genomic locus.
- Complete sequencing of the *cps* loci of the reference strains for 90 serotypes has been completed and the *cps* gene content for each serotype revealed. (Figure 1)
- Encoded proteins functionally classified into 249 homology groups (HG) based on amino acid similarity and predicted function (Bentley *et al,* 2006, PLoS Genetics).

Figure 1: Comparison of the cps locus for 23F, 15B and 15C



Sequencing of the *cps* gene locus for each serotype revealed the complement of genes required to biosynthesize the pneumococcal capsular polysaccharide.

Representative *cps* gene clusters are shown for serotypes 23F, 15B and 15C together with their corresponding HG profile.

Red blocks indicate regions of sequence similarity between gene clusters, demonstrating that some serotypes are more similar than others. The gene content of 15B and 15C is identical, reflected in the HG profile, and differ only in a small mutation in *wciZ* (HG88).

BµG@S SP-CPS Microarray

- Robust design with multiple 60mer oligonucleotide reporters per HG or *cps* gene.
- Design concept to determine serotype from combination of HGs present in DNA.
- Agilent 8×15K array format with various components for multiple analyses:
  - **HGID:** serotype determination by HG profile of *cps* genes present.
  - **STID:** targeted discrimination of serotypes with identical HG profiles.
  - **PathID:** pathogen identification for co-colonisation of nasopharynx.
  - **AbRID:** detection of antibiotic resistance gene presence is isolates.
  - **SpTIGR4+R6:** genome backbone for strain association by arrayCGH.

Enhanced utility and application of approach

- Comprehensive molecular serotyping for any one of the known 91 serotypes.
- Indication of novel serotypes from unexpected combinations of genes. (Figure 2)
- Surveillance of *cps* genes in non-typeables or related *Streptococcus* species.
- Detection of multiple serotype carriage from COBA plate sweeps. (Figure 3)
- Relative abundance of multiple serotypes determined down to ~1% levels.
- Culture-free serotyping achieved directly from nasopharyngeal swabs. (Table 1)
- Simultaneous determination of *S.pneumoniae* serotype, co-colonisation by other pathogens, antibiotic resistance gene presence and strain relatedness by CGH.
- Automated web-based tool developed to analyse microarray data and report reliable calls for multiple analyses based on robust Bayesian statistical methods.

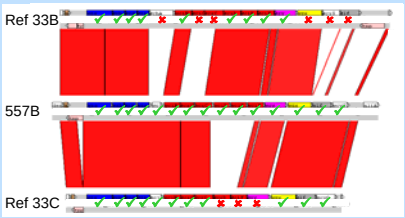
BµG@S Bacterial Microarray Group at St. George's

wellcome trust

Novel serotypes and non-typeables

- Surveillance for the presence of any known *cps* gene by the microarray enables rapid insights into the *cps* gene content of isolates, facilitating both novel serotype discovery and also characterisation of non-typeable isolates.
- Novel serotype discovery is achieved by detecting HG profiles that do not match the *cps* gene content of any of the currently known reference strain serotypes.
- Analysis of non-typeable isolates revealed they may either match the *cps* gene content of known serotypes, resemble remnants of serotypes, contain no known *cps* genes or identified as non-*pneumoniae Streptococcus spp.* by arrayCGH.

Figure 2: Potential novel serotype discovery for isolate 557B



Isolate 557B was serotyped as 33C by standard methods. However, microarray analysis did not match the results expected for a 33C.

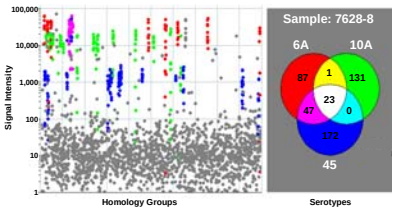
Whilst the majority of expected 33C genes were present (✓), others were absent/divergent (✗) and also genes found in other serotypes were unexpectedly present in 557B.

Sequencing of the *dexB-aliA* locus has confirmed the *cps* gene content and order for isolate 557B.

Multiple serotype carriage

- The microarray has provided the ability to detect and identify multiple serotype carriage and also determine the relative abundance of each serotype present.
- Results were obtained from DNA extracted from sweeps of primary plate cultures of *S. pneumoniae* on selective media, rather than single colony purified isolates, ensuring multiple bacterial colonies were included in the analysis.
- The sensitivity of the fluorescent detection has enabled serotypes in the order of 1% relative abundance to be detected in samples.

Figure 3: Detection and relative abundance of multiple carriage



The presence of multiple serotypes were detected in DNA extracted from a plate sweep of bacterial colonies grown from a nasopharyngeal swab.

The relative abundance of each serotype present is directly proportional to relative signal intensity for each serotype.

Statistical analysis reported that 6A(78%)+10A(28%)+45 (2%) were present in sample 7628-8.

Direct analysis from nasopharyngeal swabs

- Preliminary data has indicated that the microarray offers potential to directly serotype from nasopharyngeal swabs using DNA extracted from 200µl of STGG.
- Excellent concordance with established serotyping methods was shown with the additional advantage of detecting the presence of co-colonising pathogens.
- Sensitivity is greatly reduced when compared to plate sweeps due the lower bacterial load of *S. pneumoniae,* combined with the greater complexity of sample that includes DNA from other pathogens, commensals and host tissue.
- Whole genome amplification methods have been utilised but suggest that a more targeted enrichment or amplification strategy is required for future use.

Table 1: Molecular serotyping direct from nasopharyngeal swabs

Isolate	Serotype	*S.pneumoniae* cfu/ml	BµG@S SP-CPS Microarray Analysis		
			Plate Sweep	NPS	NPS+WGA
2651	34+19F	2.80×10^6	34(62%)+19F(38%)	34(75%)+19F(25%)	34(76%)+19F(24%)
2215	19B	1.80×10^6	19B(88%)+6A(12%)	19B	19B
1876	14+NT	1.40×10^6	NT(65%)+14(32%)+23A(3%)	NT+14	14
2184	19F+23F	8.00×10^5	23F(80%)+19F(20%)	23F	23F
2445	33C	2.20×10^5	33C(71%)+46(29%)	Fail	Fail
1617	14	1.60×10^5	14	14	14
3091	15B	1.50×10^5	15B(56%)+23B(44%)	Fail	15B
2443	28	1.40×10^5	28A(95%)+19F(5%)	28A	28A
1903	9A	4.00×10^4	9V	Fail	Fail
1657	10F	4.40×10^3	10F	Fail	Fail