

RESEARCH ARTICLE

Open Access



Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003–2013

Thabo Mohale^{1,2*}, Nicole Wolter^{1,2}, Mushal Allam¹, Kedibone Ndlangisa^{1,2}, Penny Crowther-Gibson³, Mignon du Plessis^{1,2} and Anne von Gottberg^{1,2}

Abstract

Background: The capsular polysaccharide is the principal virulence factor of *Streptococcus pneumoniae* and a target for current pneumococcal vaccines. However, some pathogenic pneumococci are serologically nontypeable [nontypeable pneumococci (NTPn)]. Due to their relative rarity, NTPn are poorly characterized, and, as such, limited data exist which describe these organisms. We aimed to describe disease and genotypically characterize NTPn causing invasive pneumococcal disease in South Africa.

Results: Isolates were detected through national, laboratory-based surveillance for invasive pneumococcal disease in South Africa and characterized by whole genome analysis. We predicted ancestral serotypes (serotypes from which NTPn may have originated) for Group I NTPn using multilocus sequence typing and capsular region sequence analyses. Antimicrobial resistance patterns and mutations potentially causing nontypeability were identified. From 2003–2013, 39 (0.1 %, 39/32,824) NTPn were reported. Twenty-two (56 %) had partial capsular genes (Group I) and 17 (44 %) had complete capsular deletion of which 15 had replacement by other genes (Group II). Seventy-nine percent (31/39) of our NTPn isolates were derived from encapsulated *S. pneumoniae*. Ancestral serotypes 1 (27 %, 6/22) and 8 (14 %, 3/22) were most prevalent, and 59 % (13/22) of ancestral serotypes were serotypes included in the 13-valent pneumococcal conjugate vaccine. We identified a variety of mutations within the capsular region of Group I NTPn, some of which may be responsible for the nontypeable phenotype. Nonsusceptibility to tetracycline and erythromycin was higher in NTPn than encapsulated *S. pneumoniae*.

Conclusions: NTPn are currently a rare cause of invasive pneumococcal disease in South Africa and represent a genetically diverse collection of isolates.

Keywords: South Africa, *Streptococcus pneumoniae*, Nontypeable, Invasive pneumococcal disease, Whole genome sequencing

* Correspondence: thabom@nicd.ac.za

¹Centre for Respiratory Diseases and Meningitis (CRDM), National Institute for Communicable Diseases (NICD), National Health Laboratory Services (NHLS), Johannesburg, South Africa

²School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Full list of author information is available at the end of the article



Background

Streptococcus pneumoniae frequently colonizes the nasopharynx asymptotically and is also a significant human pathogen causing diseases such as otitis media, pneumonia and meningitis [1]. The capsule is a major virulence factor of *S. pneumoniae*, protecting it from host cell-mediated phagocytosis [2]. The capsule induces protective antibodies which provide protection against pneumococcal disease, and serves as the basis for current pneumococcal vaccines [3]. Based on the structure and antigenicity of the capsule, more than 90 serotypes have been identified [4, 5].

Nontypeable pneumococci (NTPn) are not assigned a serotype when using the Quellung reaction [6]. The inability to do so may be due to low-level capsule expression or novel capsule types not detectable by the Quellung reaction, or absence of capsule due to genetic modifications in the capsular polysaccharide synthesis locus (*cps*) [7–11]. NTPn are predominantly detected in carriage or non-invasive disease episodes, and rarely cause invasive pneumococcal disease (IPD) [12–14]. In vitro studies have shown that NTPn display increased adherence to epithelial cells and are more easily transformable compared to encapsulated *S. pneumoniae* (Ec-*Sp*) isolates [15, 16].

NTPn have been categorized into two groups based on the contents of the *cps* locus [7]. Group I has at least partial *cps* genes, whereas in Group II the *cps* genes are completely deleted and may be replaced with other genes. Group II is further subdivided into four null capsule clades (NCC) [17]. NCC1 has the *psk* gene which has been shown to play a role in adherence to epithelial cells and colonisation [18]. NCC2 has *aliC* and *aliD* genes either with (NCC2a) or without (NCC2b) a putative toxin-antitoxin system (encoded by *ntaAB* genes). *AliC* and *aliD* genes facilitate upregulation of competence for genetic transformation and mediate colonization, respectively [19]. NCC3 has the *aliD* gene, and NCC4 contains only transposable elements in the *cps* locus.

There are limited studies describing invasive NTPn [10, 12], possibly due to the rarity of these isolates in IPD. The prevalence of NTPn as a cause of IPD has been reported to range from 0.6 to 3% [10, 12]. The majority (~90%) of invasive NTPn described in the literature is Group I and is thought to be derived from Ec-*Sp* [10, 12]. In contrast, almost all Group II isolates are related to established NTPn lineages [12]. Exploring the potential changes in *cps* structure and understanding the genomic diversity of NTPn is of importance, especially because current pneumococcal conjugate vaccines (PCV) are not effective against these isolates. We aimed to describe disease and genotypically characterize NTPn causing IPD in South Africa.

Results

Invasive pneumococcal disease surveillance in South Africa, 2003–2013

From 2003 through 2013, 46,485 cases of IPD were reported, of which 32,824 (71%) had viable isolates. Thirty-nine (0.1%) were nontypeable by the Quellung reaction and were thus classified as NTPn. Case characteristics are shown in Table 1. Forty-four percent (17/38) of cases were in children <5 years. Two patients with known vaccination status had received one dose of PCV13 and two doses of PCV7, respectively. Thirteen percent (2/15) of patients with known immune status were immunocompromised. Of ten patients with known HIV status, 50% (5/10) were seropositive. Twenty-two percent (4/18) of patients with known outcome data died.

Classification of NTPn and multi-locus sequence type

Fifty-six percent (22/39) of NTPn had at least partial *cps* genes and were thus categorized as Group I (Table 2). The remaining isolates (17/39, 44%) had complete deletion of *cps* genes ($n = 17$) and replacement by *pspK* ($n = 1$), *aliC*, *aliD* and *ntaAB* ($n = 12$) or *aliC* and *aliD* genes only ($n = 2$) and were thus categorized as Group II. Based on the classification proposed by Park et al. [12, 17], these isolates were defined as NCC1 ($n = 1$), NCC2a ($n = 12$), NCC2b ($n = 2$) and NCC4 ($n = 2$).

A total of 28 STs were identified (Table 2). Among Group I isolates, 18 STs were identified: 6 were novel STs and 12 were STs usually associated with Ec-*Sp*. Among Group II isolates, 10 STs were identified: 7 were novel STs and 3 (NCC2a) were ST344 which is usually associated with carriage NTPn, and ST105 and ST218 (NCC4) which are associated with Ec-*Sp*. Two CCs were identified among Group I isolates; CC217 ($n = 6$) and CC53 ($n = 3$) (Fig. 1). The remainder ($n = 12$) of Group I isolates were not assigned to any CC, except for ST5604 which is a single-locus variant of a ST105 isolate from Group II. CC344 ($n = 6$) was identified among Group II isolates. The remainder ($n = 10$) of Group II isolates were not assigned to any CC.

Predicted ancestral serotypes of Group I isolates

The most common predicted ancestral serotypes were serotype 1 ($n = 6$), 8 ($n = 3$), 12F ($n = 2$), 6A ($n = 2$) and 14 ($n = 2$). Fifty-nine percent (13/22) of the predicted ancestral serotypes were PCV13 serotypes 1 ($n = 6$), 18C, 19A, 6A ($n = 2$), 14 ($n = 2$) and 23F (Table 2). Two of the NTPn isolates (NT224 and NT225) were detected in two patients that were co-infected with a serotype 1 and serotype 18C isolate, respectively.

cps mutations in Group I isolates

Comparisons between Group I isolates and their ancestral serotype *cps* regions revealed a variety of mutations

Table 1 Characteristics of patients with invasive disease due to nontypeable and encapsulated *Streptococcus pneumoniae*, South Africa, 2003–2013

Characteristic		Nontypeable <i>S. pneumoniae</i> N = 39 n (%)	Encapsulated <i>S. pneumoniae</i> N = 32785 n (%)	Total N = 32824 n (%)
Year	2003	2 (5)	2885 (9)	2887 (9)
	2004	4 (10)	3465 (11)	3469 (11)
	2005	5 (13)	3643 (11)	3648 (11)
	2006	5 (13)	3411 (10)	3416 (10)
	2007	2 (5)	3318 (10)	3320 (10)
	2008	6 (15)	3319 (10)	3325 (10)
	2009	7 (18)	3377 (10)	3384 (10)
	2010	3 (8)	2869 (9)	2872 (9)
	2011	0	2409 (7)	2409 (7)
	2012	3 (8)	2158 (7)	2161 (7)
Gender	Male	23 (59)	15842 (48)	15865 (48)
	Female	16 (41)	16291 (50)	16307 (50)
Age group (years)	Unknown	0	652 (2)	652 (2)
	<5	17 (44)	9446 (29)	9463 (29)
	5–14	3 (8)	2960 (9)	2963 (9)
	15–24	2 (5)	1918 (6)	1920 (6)
	25–44	10 (26)	11714 (36)	11724 (36)
	45–64	4 (10)	4355 (13)	4359 (13)
	>64	2 (5)	1137 (3)	1139 (3)
Specimen type	Unknown	1 (3)	1255 (4)	1256 (4)
	Cerebrospinal fluid	11 (28)	11486 (35)	11497 (35)
	Blood	22 (56)	19284 (59)	19306 (59)
PCV vaccination ^b	Other ^a	6 (16)	2015 (6)	2021 (6)
	Yes ^c	2 (100)	580 (77)	582 (77)
PCV13 serotypes ^d	No	0	131 (17)	131 (17)
	Unknown	0	46 (6)	46 (6)
	PCV13 serotypes	13 ^e (59)	23619 (72)	23632 (72)
HIV status	Non PCV13 serotypes	9 ^e (41)	9166 (28)	9175 (28)
	Positive	5 (13)	7718 (23)	7723 (23)
	Negative	5 (13)	2489 (8)	2494 (8)
Immuno compromising conditions ^f	Unknown	29 (74)	22578 (69)	22607 (69)
	Yes	2 (5)	3114 (10)	3116 (10)
	No	13 (33)	6974 (21)	6987 (21)
Comorbid conditions ^g	Unknown	24 (62)	22697 (69)	22721 (69)
	Yes	2 (5)	1496 (5)	1498 (5)
	No	13 (33)	8592 (26)	8605 (26)
Outcome	Unknown	24 (62)	22697 (69)	22721 (69)
	Died	4 (10)	4024 (12)	4028 (12)
	Recovered	14 (36)	9887 (30)	9901 (30)
	Unknown	21 (54)	18874 (58)	18895 (58)

^aOther specimen types: pleural, joint and vitreous fluid^bOnly children born after February 2009 were included^cReceived at least 1 dose of the vaccine^dPCV13 serotypes: 4, 6B, 9V, 14, 18C, 19F, 1, 5, 7F, 3, 6A and 19A^eNontypeable predicted ancestral serotypes for Group I isolates^fImmunocompromising conditions were defined as medical record-documented pre-existing history of head injury, connective tissue disease, asplenia, pregnancy, premature birth, malignancy, burns, gastric acid suppression, aplastic anemia, organ transplant, primary immunodeficiency conditions, chromosomal conditions, protein energy malnutrition, alcohol dependency, current smoking, or immunosuppressive therapy^gComorbid conditions were defined as medical record-documented pre-existing history of pulmonary disease, renal disease, cerebrovascular accident, hepatic disease, cardiac disease, or diabetes mellitus

Table 2 Classification, predicted ancestral serotypes and *cps* mutations of invasive nontypeable *Streptococcus pneumoniae* in South Africa, 2003–2013 (N = 39)

Group	Isolate number	Sequence type	Sequence type associated serotype(s)/serogroup ^a	<i>cps</i> locus BLAST	Ancestral serotype	Mutations in the <i>cps</i> locus when compared to ancestral serotypes	Comments
I	NT224	217	1	1	1	} <i>rmlA, rmlB, rmlC, rmlD</i> and transposons present, all other <i>cps</i> genes deleted	} SLV of ST217
	NT12	217	1	1	1		
	NT17	217	1	1	1		
	NT18	217	1	1	1		
	NT45	8314	1	1	1		
	NT11	612	1	1	1	} <i>glf, wzd, wze</i> and transposons present. All other <i>cps</i> genes deleted	} SLV of ST217
	NT5	5604	25F	25F	25F		
	NT1	53	8	8	8		
	NT3	53	8	8	8		
	NT36	3847	8	8	8	SNP in <i>wchA</i>	} SLV of ST53
	NT225	9817	None	18C	18C	Deletion of <i>wzg</i> , partial deletion of <i>wzh</i> and <i>wciS</i> . SNPs in <i>wciS</i> . SNP in <i>wze</i> and insertion in <i>wchA</i>	
	NT6	9813	None	35B	35B	SNP in <i>wciB</i>	} Novel ST is SLV of different STs mostly associated with 18C
	NT40	989	12	12F	12F	2 bp deletion in <i>wciJ</i>	
	NT34	2416	12F & others ^b	12F	12F	Partial deletions of <i>wciJ, mnaA, wzx</i> and insertion in <i>fnlA</i>	} Novel ST has no SLV or DLV
NT13	9818	None	19A	19A	SNPs in <i>wchA, wchO, wchQ, wzy</i> and <i>mnaA</i> . Partial deletion in <i>wchQ, wzy</i> and <i>wzx</i>		
NT10	6373	6A	6A	6A	SNPs in <i>wchA</i> and <i>wciP</i> . Partial deletion in <i>wciN</i>	} Novel ST SLV of ST4870,4873 (associated with serotype 19A)	
NT20	9814	None	6A	6A	SNPs in <i>wze, wchA, HG262, wciP, wzx</i> . Partial deletion of <i>wzx</i>		
NT48	2208	24F	24F	24F	Deletion in <i>wzxK, rbsF</i> and <i>glf</i> . SNPs in <i>rbsF</i> and <i>wzx</i> .	} Novel ST SLV of ST8718 (associated with serotype 6A)	
NT14	2652	14	14	14	Deletion of <i>wciY</i> . SNPs in <i>wchA, wchJ, wchK, wchL, wchM, and wchN</i> and <i>Irp</i> .		
NT32	230	14 & others ^b	14	14	SNP in <i>wze</i> . Deletion in <i>wchK, wzy, wciY</i> and <i>Irp</i> .	} ST also associated with 19A, 19F, 20, 23F, 24 & 24F	
NT49	9816	None	23F	23F	Insertion in <i>wzg, wzy, wzx, gtp3, rmlB</i> and <i>wchF</i>		
NT30	9815	None	16F	16F	SNP in <i>wzy</i> . SNPs and deletions in <i>wzg</i>	} Novel ST DLV of ST353 (associated with serotype 23F)	
NT38	9812	None	ND	N/A	Replacement of <i>cps</i> genes with <i>pspK</i> gene		
II-NCC 1	NT27	344	NT	ND	N/A	Replacement of <i>cps</i> genes (except <i>glf</i> ^c with <i>aliC, aliD, ntaA</i> & 2 <i>ntaB</i> genes	} Novel ST SLV of ST9519 (associated with serotype 3)
II-NCC 2a	NT2	344	NT	ND	N/A	Replacement of <i>cps</i> gene (except <i>glf</i> ^c s with <i>aliC, aliD, ntaA</i> & <i>ntaB</i> genes	
II-NCC 2a	NT39	344	NT	ND	N/A	Replacement of <i>cps</i> gene (except <i>glf</i> ^c s with <i>aliC, aliD, ntaA</i> & <i>ntaB</i> genes	} Novel STs SLV of ST344 (associated with nonencapsulated <i>S. pneumoniae</i> strains)
II-NCC 2a	NT16	9809	None	ND	N/A	Replacement of <i>cps</i> gene (except <i>glf</i> ^c s with <i>aliC, aliD, ntaA</i> & <i>ntaB</i> genes	
II-NCC 2a	NT44	10239	None	ND	N/A	} Replacement of <i>cps</i> genes (except <i>glf</i> ^c with <i>aliC</i> and <i>aliD</i> genes	} Novel ST SLV of ST448 (associated with nonencapsulated <i>S. pneumoniae</i> strains)
II-NCC 2b	NT47	10241	None	ND	N/A		
II-NCC 2a	NT46	9810	None	ND	N/A	} Replacement of <i>cps</i> genes (except <i>glf</i> ^c with <i>aliC, aliD, ntaA</i> & <i>ntaB</i> genes	} Novel ST DLV of ST7252 (associated with serotype 38)
II-NCC 2a	NT50	9810	None	ND	N/A		
II-NCC 2b	NT4	10240	None	ND	N/A	} Replacement of <i>cps</i> genes (except <i>glf</i> ^c with <i>aliC</i> and <i>aliD</i> genes	} Novel STs DLV of ST7252 (associated with serotype 38)
II-NCC 2a	NT15	9811	None	ND	N/A		
II-NCC 2a	NT31	9811	None	ND	N/A	} Replacement of <i>cps</i> genes(except <i>glf</i> ^c with <i>aliC, aliD, ntaA</i> & <i>ntaB</i> genes	
II-NCC 2a	NT42	9811	None	ND	N/A		
II-NCC 2a	NT43	9811	None	ND	N/A		
II-NCC 2a	NT29	9811	None	ND	N/A		
II-NCC 4	NT28	218	12F,7F	ND	NP	Deletion of all <i>cps</i> genes	} Deletion of all <i>cps</i> genes
	NT33	105	25,38,NT	ND	NP		

Bold type represents serotypes included in PCV13

SLV single-locus variant, differs at one of the seven alleles. DLV Double-locus variant, differs at two of the seven alleles

None Novel ST and therefore no serotype/ST association exists

ND not done, isolates have no *cps* genes

N/A Not applicable, resembles nontypeable pneumococci from carriage (no *cps* genes, but instead have typical nontypeable genes)

NP not predicted. Isolates have no *cps* genes and/or are associated with more than 1 serotype

^aAs per *Streptococcus pneumoniae* multilocus sequence typing database (<http://pubmlst.org/spneumoniae/>, accessed March 2015)

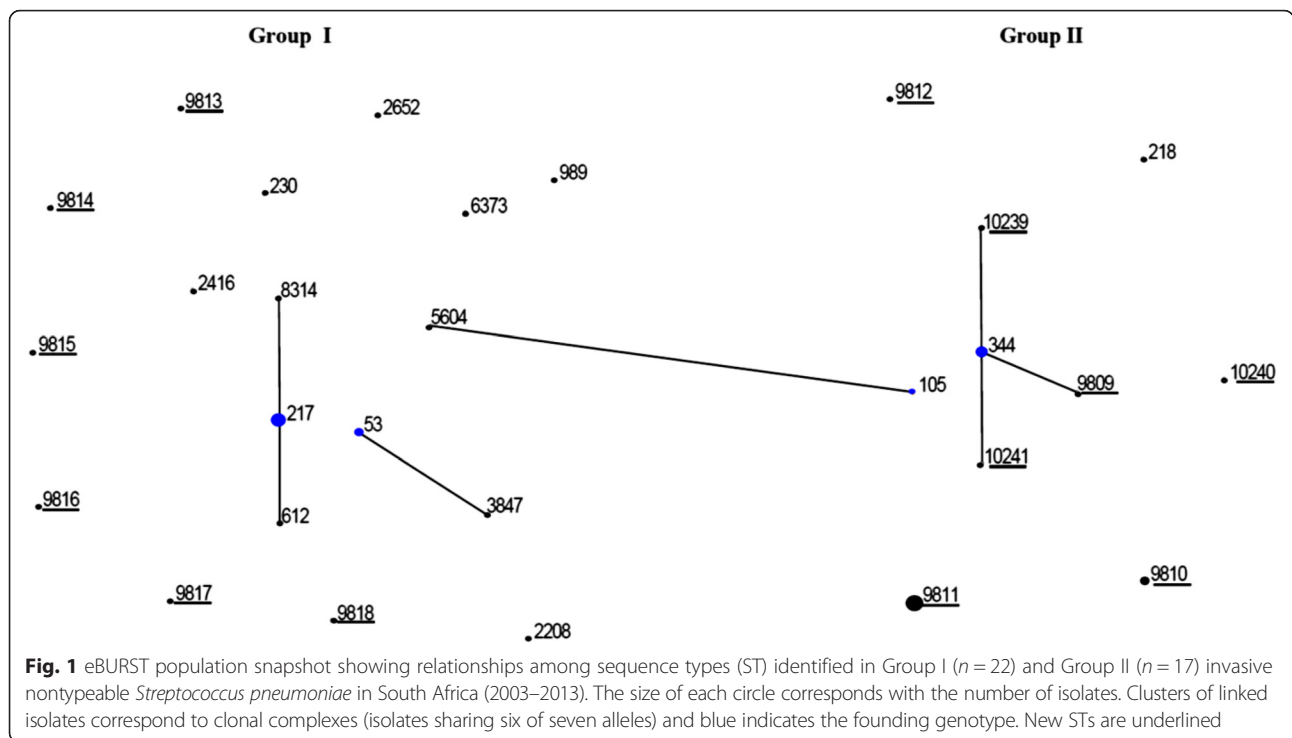
^bOther serotype/s associated with that ST. See comments

^c*glf* pseudo gene present within the *cps* locus

within the *cps* region of Group I isolates. These mutations are summarized in Table 2, described in detail in Additional file 1 and sketched in Additional file 2. Briefly, predicted ancestral serotype 1 isolates had identical partial deletions of *cps* genes (NT11, NT12, NT17, NT18, NT45 and NT224). Isolate NT5 also had partial deletions. The remaining isolates had almost all *cps* genes present, as in their ancestral serotypes; however, we identified a variety of mutations [single nucleotide polymorphisms (SNPs), insertions and deletions] in some of the *cps* genes. Most isolates had a combination of mutations. The vast majority of SNPs resulted in amino acid changes and a small number in stop codons. Base deletions and insertions were not common as SNPs.

Phylogenomic analysis

Ancestral serotype prediction was further confirmed by phylogenomic clustering of Group I isolates with representative genomes of their predicted ancestral serotypes (Fig. 2). An exception was NT30 which was predicted to be derived from serotype 16F but was more closely related to serotype 11A. For some isolates, there was more divergence between the NTPn and their ancestral serotypes as indicated by the long branches. Group II NCC1 (NT38), NCC4 (NT28 and NT33) and two carriage NTPn (MNZ37 and MNZ11b) from other studies [20] also clustered with invasive Ec-*Sp* and appear to be related to serotypes 3, 7E, 25A and 15A, respectively. Group II NCC2 NTPn belonging to ST9811 (*n* = 5) and ST10240 formed a distinct clade separate from other



Group II NCC2 isolates and seem to be more closely related to invasive *Ec-Sp*. The remaining Group II NCC2 NTPn ($n = 8$) formed a distinct clade with carriage NTPn from other studies ($n = 4$) [20, 21]. Overall, 31/39 (79 %) of NTPn from this study ([22] Group I, Group II NCC1 ($n = 1$), NCC4 ($n = 2$) and six NCC2 isolates belonging to ST9811 ($n = 5$) and ST10240) appear to have been derived from *Ec-Sp*.

Antimicrobial resistance patterns

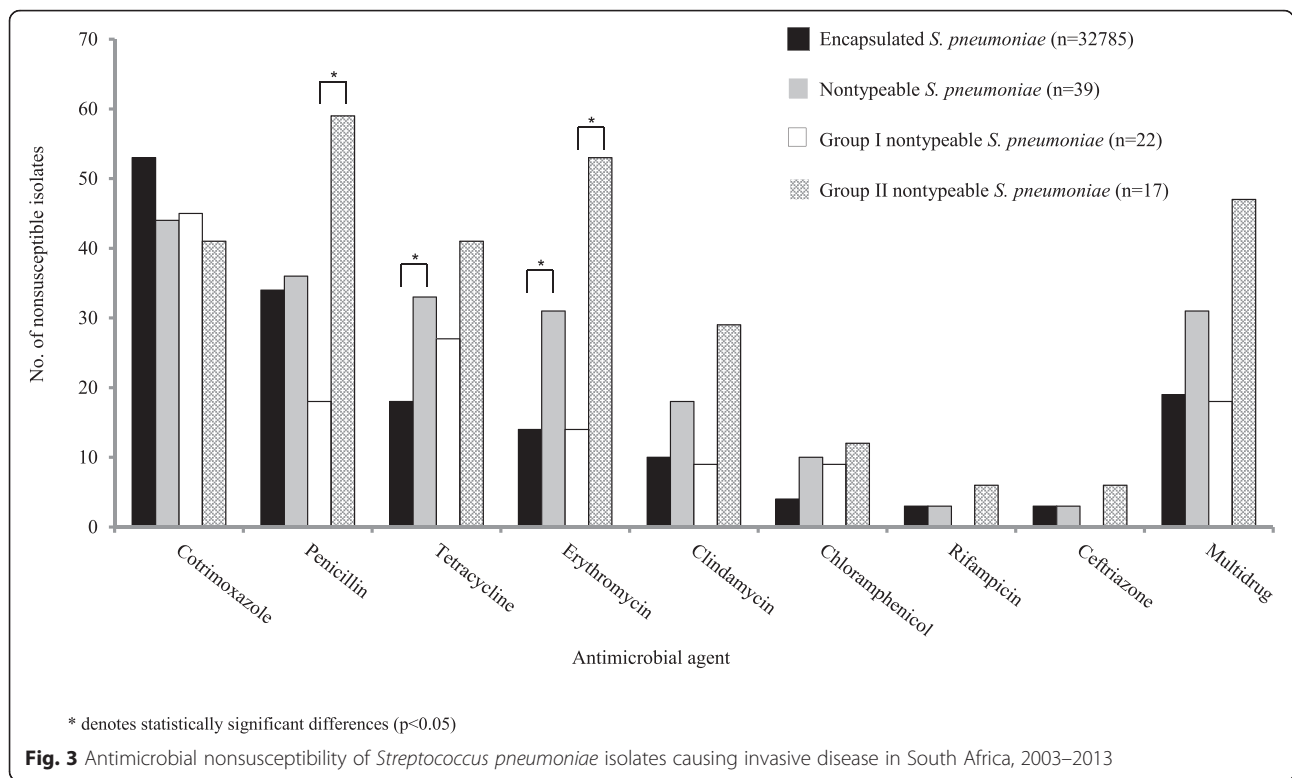
Nonsusceptibility among NTPn was highest for cotrimoxazole (17/39, 44 %), penicillin (14/39, 36 %), tetracycline (13/39, 33 %) and erythromycin (12/39, 31 %) (Fig. 3). Prevalence of nonsusceptibility was higher among NTPn than *Ec-Sp* for most antimicrobials; however this difference was only significant for tetracycline [33 % (13/39) versus 18 %, (5902/32785), $P = 0.02$] and erythromycin [31 % (12/39) versus 14 % (4589/32785), $P = 0.005$]. In addition, 31 % (12/39) of NTPn were multidrug resistant [of which 21 % (8/39) included penicillin] compared to 19 % (6230/32785) of *Ec-Sp* however this difference was not statistically significant ($P = 0.09$). With the exception of cotrimoxazole and chloramphenicol, Group II NTPn had a higher prevalence of nonsusceptibility than Group I, although this difference was only statistically significant for penicillin [59 % (10/17) versus 18 % (4/22), $P = 0.02$] and erythromycin [53 % (9/17) versus 14 % (3/22), $P = 0.01$]. Multidrug resistance was also higher in Group II NTPn (47 %, 8/17) compared to Group I

(18 %, 4/22), although not statistically significant ($P = 0.08$).

Discussion

We characterized NTPn causing IPD in South Africa, a middle-income, high HIV prevalence country with PCV introduction in 2009. NTPn represented 0.1 % of IPD cases, much lower than other studies which have reported a prevalence of between 0.6 and 3 % [10, 12]. Our study showed that among patients with NTPn infection and for whom data were available, 44 % were children less than 5 years, proportionately more than *Ec-Sp* (29 %), 50 % were HIV positive versus 76 % for *Ec-Sp* and the case-fatality ratio (22 %) was similar to cases with *Ec-Sp* (29 %). Fifty-six percent of our NTPn were Group I, for which the most common predicted ancestral serotypes were 1 (27 %) and 8 (14 %). Fifty-nine percent of the predicted ancestral serotypes were PCV13 serotypes. The *cps* loci of the Group I isolates harbored a variety of mutations. Nonsusceptibility to tetracycline and erythromycin was significantly higher in NTPn than encapsulated *S. pneumoniae*. Nonsusceptibility to penicillin and erythromycin was also significantly higher in Group II NTPn than Group I.

The mechanisms by which NTPn are able to cause IPD without the capsule are not clear. Immune suppression of the host and/or other lesser known or novel virulence factors might possibly explain the success of NTPn in causing IPD. However, in this study the



majority of patients with NTPn infection (87 %, 13/15) were not immunocompromised and this was similar to patients with Ec-*Sp* strains (69 %). NTPn appear to be equally as virulent as Ec-*Sp* as the patients with NTPn infection did not seem to be more susceptible to infection.

We observed a higher prevalence (44 %) of Group II isolates amongst our NTPn that has not been described in other studies, where 4 to 10 % of NTPn were Group II isolates [10, 12]. The majority of Group II isolates (82 %) were NCC2 and no isolates belonging to NCC3 were identified, a finding similar to invasive NTPn collected through the Active Bacterial Core surveillance program in the United States during 2006–2009 [12]. In addition, two isolates were classified as NCC4, a newly defined clade in Group II, characterized by deletion of all *cps* genes [12]. Similar to what has been shown in the US Active Bacterial Core surveillance, NCC4 isolates in this study also appeared to be derived from Ec-*Sp* [12].

We were able to predict ancestral serotypes for all Group I isolates because they had partial *cps* genes, suggesting that these isolates were derived from Ec-*Sp* which, at some point, lost their capsule through mutations in the *cps* locus. Indeed we found a diverse range of mutations in the *cps* locus of Group I isolates, some of which may be responsible for the nontypeable phenotype. For predicted ancestral serotype 1 isolates, all *cps* mutations were identical. The deleted region is flanked

by identical IS1167 repeats and this, together with the fact that the predicted ancestral serotype 1 isolates were genotypically related, indicates that the *cps* mutations probably occurred as a result of a single deletion event in a particular strain rather than multiple independent deletions in different strains. The same mutations have been described for serotype 1-derived NTPn in other studies [9, 10].

Serotype predictions were confirmed by the phylogenomic analyses as these isolates clustered with encapsulated strains expressing the same serotype as the predicted ancestral serotypes. One isolate (NT30), however, was more closely related to serotype 11A than its predicted serotype 16F. This isolate has a novel ST, which shares three of seven alleles with the 11A isolate, suggesting that NT30 may have been derived from a common ancestor with serotype 11A at some point in time. For some isolates, there was more divergence between the NTPn and their ancestral serotypes. This could be that certain serotypes are inherently diverse while others are more stable.

Predicted ancestral serotypes 1 (27 %) and 8 (14 %) were most prevalent among our NTPn isolates. Analysis of invasive NTPn isolates from Native American communities collected from 1994–2007 showed that predicted ancestral serotype 1 (19 %) and 7 F (15 %) were most prevalent [10]. In addition, serotype 8 (25 %) was most prevalent among invasive NTPn from the US Active Bacterial Core surveillance [12]. These data suggest

that serotypes 1 and 8 may be more prone to mutations in their *cps* region than other serotypes.

Two of the NTPn isolates (NT224 and NT225) were detected in mixed infections in two patients with encapsulated serotype 1 and 18C isolates, respectively. For the serotype 1 and NT224 mixed infection, *S. pneumoniae* was identified from the same culture, with the NT isolate representing approximately 98 % of the culture (using the Quellung reaction). After several attempts to separate the two variants, only the NTPn isolate could be obtained, however, real-time PCR confirmed the presence of the serotype 1 in the mixed culture. NT224 was confirmed to be ST217 which is associated with serotype 1. For the serotype 18C and NT225 mixed infection, both isolates were available and NT225 had the same *cps* region, with the exception of a single SNP, compared to its encapsulated counterpart (serotype 18C), the same ST and their core genomes were identical (unpublished observations).

Pneumococcal serotypes targeted by PCV have the potential to switch from vaccine serotype to non-vaccine serotype or to nontypeable, thereby providing a mechanism whereby they may be able to evade vaccine pressure. However, no evidence of adaptation to PCV occurred among invasive NTPn from Native American communities, where 45 % of NTPn isolates had ancestral PCV7 serotypes pre-vaccine and none of their NTPn isolates were ancestral PCV7 serotypes post-vaccine [10]. Although 59 % of the predicted ancestral serotypes of NTPn in our study were PCV13 serotypes, there is no evidence to suggest that serotype switching occurred as a result of PCV pressure as the majority of NTPn were already present prior to PCV introduction.

Antimicrobial nonsusceptibility data for invasive NTPn are limited. In our study, there was a significantly higher prevalence of nonsusceptibility among NTPn than *Ec-Sp* for erythromycin and tetracycline. These higher nonsusceptibility rates coupled with higher transformation rates of NTPn [16] enable these strains to serve as a reservoir of antibiotic resistance genes. A serotype 19F clone from Switzerland became increasingly resistant to penicillin by acquisition of a *pbp2x* gene from NTPn [23]. Recently, a comparative genome analysis of over 3000 carriage pneumococci from a refugee camp in Thailand showed that the highest rates of receipt and donation of recombinant DNA occurred in NTPn and that the most commonly exchanged genes were those associated with antibiotic resistance and immune interactions [24].

It is possible that our NTPn isolates may have lost the ability to express capsule during sub-culturing in the laboratory, but we believe this was not the case as we would expect NTPn to be detected more frequently. We did not perform in vitro experiments to confirm whether the mutations identified in the *cps* locus of our isolates

were responsible for the lack of capsule expression. However, in some of these isolates, there were complete deletions of *cps* genes making expression of a capsule impossible. For *Ec-Sp* isolates different methods [agar dilution or Etest (2003–2008) or broth microdilution (2009–2013)] were used for MIC testing compared to NTPn (broth microdilution). This may have overestimated the prevalence of β -lactam resistance in NTPn for the 2003–2008 period as broth microdilution detects more resistance than agar dilution as was previously shown [25].

Conclusion

NTPn are currently a rare cause of IPD in South Africa and are genetically diverse but have a higher prevalence of antimicrobial resistance than *Ec-Sp*. The majority of NTPn were derived from *Ec-Sp*, of which a significant proportion was PCV13 serotypes. Further studies which explore capsule-independent virulence mechanisms of NTPn should be considered.

Methods

Bacterial isolates

Isolates were obtained through the GERMS-SA (Group for Enteric, Respiratory and Meningeal Disease Surveillance in South Africa) network, which conducts active, national, laboratory-based surveillance for IPD. Surveillance began in 1999 and was enhanced in 2003 [26, 27]. Over 200 diagnostic laboratories across South Africa routinely submit clinical isolates and basic patient demographic data to the network. A case of IPD was defined as the detection of *S. pneumoniae* from a normally sterile site (e.g., cerebrospinal fluid, blood) from January 2003 through December 2013. Isolates were cultured on 5 % horse blood agar (Diagnostic Media Products, Johannesburg, South Africa) and incubated at 37 °C in 5 % CO₂ for 18–24 h. Optochin sensitivity and bile solubility assays were performed to confirm *S. pneumoniae* identification. In addition, for NTPn isolates, real-time PCR detecting *lytA* was performed [28]. A mixed infection was defined as simultaneous isolation of at least two serotypes (including nontypeable isolates) either from the same normally-sterile site specimen or from two or more normally-sterile site specimens obtained from the same patient within 21 days of each other. Such cases were detected during routine serotyping by Quellung if an isolate reacted partially with a specific antiserum pool.

Serotyping and antimicrobial susceptibility testing

Serotyping was performed by the Quellung reaction using serotype-specific antisera (Statens Serum Institut, Copenhagen, Denmark). Minimum inhibitory concentrations (MIC) for NTPn isolates were determined by broth microdilution using commercially customized TREK panels (Trek Diagnostics Inc., Ohio, United States).

MICs for Ec-*Sp* isolates were determined as previously described [25, 27]. MIC breakpoints were interpreted using Clinical and Laboratory Standards Institute guidelines [29]. For penicillin, isolates with MICs of ≥ 0.12 mg/L were defined as nonsusceptible. Isolates that were intermediately resistant or resistant to any of the antibiotics were regarded as nonsusceptible. Multidrug resistance was defined as nonsusceptibility to three or more classes of antibiotics.

DNA preparations, genome sequencing and assembly

Brain heart infusion broth (Diagnostic Media Products) was inoculated with overnight cultures and incubated overnight at 37 °C in 5 % CO₂. A pre-lysis step was performed by suspending bacterial pellets in 200 μ l of 10 mg/ml lysozyme (Sigma-Aldrich, St. Louis, MO, USA) and incubating at 37 °C for 1 h. Genomic DNA was extracted using the QIAamp® DNA Mini kit (QIAGEN, Hilden, Germany). Paired-end libraries (2 x 300 bp) were prepared using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA) and sequencing was performed on an Illumina MiSeq. The resulting paired-end reads were quality trimmed and mapped to the reference genome of *S. pneumoniae* ATCC 700669 [GenBank accession no. FM211187, serotype 23 F, sequence type (ST) 81] using CLC Genomics Workbench v8 (CLC bio, Aarhus, Denmark), giving on average 121x depth of coverage and 88 % coverage of the reference genome. *De novo* assembly was performed for all isolates using CLC Genomics Workbench and ordered relative to *S. pneumoniae* ATCC 700669 using Mauve [30]. The ordered contigs were concatenated and annotated using Prokka v1.11 [31].

cps regions and multi-locus sequence typing (MLST)

The *cps* locus and seven MLST alleles were derived from whole genome data. The *cps* loci were used to categorize NTPn into different groups/clades [12, 17]. MLST allele numbers and sequence type (ST) were assigned using the Bio-MLST-Check module (<http://search.cpan.org/~ajpage/Bio-MLST-Check> 1.133090/lib/Bio/MLST/Check.pm). The eBURST v3 algorithm (<http://eburst.mlst.net>) was used to determine genetic relationships and grouped isolates into clonal complexes (CC) [32]. A CC was defined as a group of isolates sharing six of seven alleles with any other isolate in the group.

In silico prediction of ancestral serotypes for Group I NTPn

To predict ancestral serotypes of Group I isolates (serotypes from which NTPn may have originated), two methods were used, namely, serotype-ST associations using the MLST database (<http://pubmlst.org/spneumoniae>), accessed March 2015, and blasting of the *cps* locus against reference *cps* loci for the 94 known serotypes (GenBank accession no's. CR93162 – CR931722, EF538714, HM171374,

GU074953 and JQ653094). The serotype that gave the highest BLAST bit score was assigned as the likely ancestral serotype.

cps mutations in Group I NTPn

To determine the genetic variations within the *cps* locus of Group I isolates that may be responsible for their phenotypic nontypeability, we used the alignment module in CLC to align the *cps* locus of each isolate together with a number of its predicted ancestral serotype *cps* loci shown in Additional file 3.

Phylogenomic tree construction

To confirm predicted ancestral serotypes for Group I isolates, a maximum likelihood phylogenetic tree was constructed based on core genome SNPs of NTPn from this study ($n = 39$), together with invasive Ec-*Sp* with different serotypes from South Africa ($n = 42$) Additional file 4 matching the predicted ancestral serotypes for the NTPn isolates, and carriage NTPn from other countries ($n = 6$) [20, 21]. The core genome alignment module in the rapid large-scale prokaryote pan genome analysis pipeline was used to extract predicted coding regions from annotated assemblies and convert them to protein sequences [33]. All protein sequences were compared with each other using BLASTP. Proteins that had alignment similarity of ≥ 70 % and were present in at least 90 % of the isolates were defined as the core genome. RAxML was used to create a phylogenetic tree from the resulting core genome alignment and this was visualized in Figtree v1.4.2 [34].

Statistical analysis

Differences between NTPn versus Ec-*Sp* antimicrobial nonsusceptibility and between Group I and Group II NTPn were analyzed using the chi-square test and Fisher's exact test, respectively, with statistical significance assessed at $P < 0.05$. Statistical analyses were performed with GraphPad InStat 3.

Additional files

Additional file 1: Description of the *cps* mutations of nontypeable *Streptococcus pneumoniae* ($n = 39$) from South Africa, 2003–2013. (XLS 44 kb)

Additional file 2: Diagrammatic representation of capsular locus of nontypeable *Streptococcus pneumoniae* ($n = 39$) from South Africa, 2003–2013. (DOCX 1547 kb)

Additional file 3: Accession numbers for encapsulated *S. pneumoniae* used for *cps* locus alignment (XLS 54 kb)

Additional file 4: Accession numbers for encapsulated *S. pneumoniae* isolates from South Africa and carriage nontypeable isolates from other studies used to construct the phylogenetic tree. (XLS 34 kb)

Abbreviations

CC, clonal complex; *cps*, capsular polysaccharide synthesis locus; Ec-*Sp*, encapsulated *Streptococcus pneumoniae*; GERMS-SA, Group for Enteric,

Respiratory and Meningeal Disease Surveillance in South Africa; IPD, invasive pneumococcal disease; MIC, Minimum inhibitory concentration; MLST, multi locus sequence typing; NCC, null capsule clade; NTPn, nontypeable pneumococci; PCV, pneumococcal conjugate vaccine; SNP, single nucleotide polymorphism; ST, sequence type.

Acknowledgements

This work was funded by the Medical Research Council of South Africa and an NHLS Research Trust Development grant (PI, Nicole Wolter). Surveillance was funded by the NHLS/NICD, and was supported, in part, by funds from the Centers for Disease Control, Atlanta, Georgia, USA. We thank all laboratories and clinical staff throughout South Africa who participated in GERMS-SA surveillance, all the staff of the CRDM bacteriology laboratory for technical assistance and expertise. We also thank the NICD core sequencing facility (Dr Arshad Ismail) for sequencing of isolates. The findings and conclusions in this paper are those of the authors and do not necessarily represent the official views of their affiliated institutions or the agencies funding the study.

Availability of data and materials

The draft genome sequences of the 39 NTPn isolates analysed in this study have been deposited in GenBank with accession numbers LHAF000000000 - LHAG000000000, LIWA000000000 - LIWT000000000 and LIVJ000000000 - LIVZ000000000.

Authors' contributions

The study was conceived by TM, NW, MdP and AvG. Whole genome analysis was carried out by MA and TM. PCG performed the statistical analysis. The manuscript was drafted by TM and critically revised by NW, MdP, KN and AvG. All authors have read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Ethical approval for this study and national IPD surveillance were obtained from the Human Research Ethics Committee (Medical), University of the Witwatersrand, Johannesburg, South Africa (protocol numbers: M140450 and M081117, respectively).

Author details

¹Centre for Respiratory Diseases and Meningitis (CRDM), National Institute for Communicable Diseases (NICD), National Health Laboratory Services (NHLS), Johannesburg, South Africa. ²School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ³Public Health Surveillance and Response, National Institute for Communicable Diseases, National Health Laboratory Services, Johannesburg, South Africa.

Received: 11 December 2015 Accepted: 27 May 2016

Published online: 22 June 2016

References

- O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*. 2009;374:893–902.
- Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun*. 2010;78:704–15.
- Milligan GN, Fairchild RL, Sterner KE, Braley-Mullen H. Type 6 and 19 pneumococcal polysaccharides coupled to erythrocytes elicit pneumococcal cell wall-specific primary IgM responses and capsular polysaccharide-specific secondary IgG responses. *Eur J Immunol*. 1990;20:595–603.
- Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet*. 2006;2:e31.
- Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, Nahm MH. Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* Serotype 20 strains: discovery of a new pneumococcal serotype. *J Biol Chem*. 2012;287:27885–94.
- Austrian R. The quellung reaction, a neglected microbiologic technique. *Mt Sinai J Med*. 1976;43:699–709.
- Hathaway LJ, Stutzmann MP, Battig P, Aebi S, Muhlemann K. A homologue of *alib* is found in the capsule region of nonencapsulated *Streptococcus pneumoniae*. *J Bacteriol*. 2004;186:3721–9.
- Melchiorre S, Camilli R, Pietrantonio A, Moschioni M, Berti F, Del Grosso M, Superti F, Barocchi MA, Pantosti A. Point mutations in *wchA* are responsible for the non-typability of two invasive *Streptococcus pneumoniae* isolates. *Microbiology*. 2012;158:338–44.
- Salter SJ, Hinds J, Gould KA, Lamberts L, Hanage WP, Antonio M, Turner P, Hermans PW, Bootsma HJ, O'Brien KL, Bentley SD. Variation at the capsule locus, *cps*, of mistyped and non-typable *Streptococcus pneumoniae* isolates. *Microbiology*. 2012;158:1560–9.
- Scott JR, Hinds J, Gould KA, Millar EV, Reid R, Santosham M, O'Brien KL, Hanage WP. Nontypeable pneumococcal isolates among navajo and white mountain apache communities: are these really a cause of invasive disease? *J Infect Dis*. 2012;206:73–80.
- Waite RD, Penfold DW, Struthers JK, Dowson CG. Spontaneous sequence duplications within capsule genes *cap8E* and *tts* control phase variation in *Streptococcus pneumoniae* serotypes 8 and 37. *Microbiology*. 2003;149:497–504.
- Park IH, Geno KA, Sherwood LK, Nahm MH, Beall B. Population-based analysis of invasive nontypeable pneumococci reveals that most have defective capsule synthesis genes. *PLoS One*. 2014;9:e97825.
- Sa-Leao R, Simoes AS, Nunes S, Sousa NG, Frazao N, de Lencastre H. Identification, prevalence and population structure of non-typable *Streptococcus pneumoniae* in carriage samples isolated from preschoolers attending day-care centres. *Microbiology*. 2006;152:367–76.
- Xu Q, Kaur R, Casey JR, Sabharwal V, Pelton S, Pichichero ME. Nontypeable *Streptococcus pneumoniae* as an otopathogen. *Diagn Microbiol Infect Dis*. 2011;69:200–4.
- Adamou JE, Wizemann TM, Barren P, Langermann S. Adherence of *Streptococcus pneumoniae* to human bronchial epithelial cells (BEAS-2B). *Infect Immun*. 1998;66:820–2.
- Pearce BJ, Iannelli F, Pozzi G. Construction of new unencapsulated (rough) strains of *Streptococcus pneumoniae*. *Res Microbiol*. 2002;153:243–7.
- Park IH, Kim KH, Andrade AL, Briles DE, McDaniel LS, Nahm MH. Nontypeable pneumococci can be divided into multiple *cps* types, including one type expressing the novel gene *pspK*. *MBio*. 2012;3:e00035–12.
- Keller LE, Jones CV, Thornton JA, Sanders ME, Swiatlo E, Nahm MH, Park IH, McDaniel LS. *PspK* of *Streptococcus pneumoniae* increases adherence to epithelial cells and enhances nasopharyngeal colonization. *Infect Immun*. 2013;81:173–81.
- Hathaway LJ, Battig P, Reber S, Rotzetter JU, Aebi S, Hauser C, Heller M, Kadioglu A, Muhlemann K. *Streptococcus pneumoniae* detects and responds to foreign bacterial peptide fragments in its environment. *Open Biol*. 2014;4:130224. doi:10.1098/rsob.130224:130224.
- Keller LE, Thomas JC, Luo X, Nahm MH, McDaniel LS, Robinson DA. Draft Genome Sequences of Five Multilocus Sequence Types of Nonencapsulated *Streptococcus pneumoniae*. *Genome Announc*. 2013;1:e00520–13.
- Hilty M, Wuthrich D, Salter SJ, Engel H, Campbell S, Sa-Leao R, de Lencastre H, Hermans P, Sadowy E, Turner P, Chewapreecha C, Diggle M, Pluschke G, McGee L, Koseoglu EO, Low DE, Smith-Vaughan H, Endimiani A, Kuffer M, Dupasquier M, Beaudoin E, Weber J, Bruggmann R, Hanage WP, Parkhill J, Hathaway LJ, Muhlemann K, Bentley SD. Global phylogenomic analysis of nonencapsulated *Streptococcus pneumoniae* reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. *Genome Biol Evol*. 2014;6:3281–94.
- Schaffner TO, Hinds J, Gould KA, Wuthrich D, Bruggmann R, Kuffer M, Muhlemann K, Hilty M, Hathaway LJ. A point mutation in *cpsE* renders *Streptococcus pneumoniae* nonencapsulated and enhances its growth, adherence and competence. *BMC Microbiol*. 2014;14:210. doi:10.1186/s12866-014-0210-x:210-0210.
- Hauser C, Aebi S, Muhlemann K. An internationally spread clone of *Streptococcus pneumoniae* evolves from low-level to higher-level penicillin resistance by uptake of penicillin-binding protein gene fragments from nonencapsulated pneumococci. *Antimicrob Agents Chemother*. 2004;48:3563–6.
- Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ, Salter SJ, Harris D, Nosten F, Goldblatt D, Corander J, Parkhill J, Turner P, Bentley SD. Dense genomic

- sampling identifies highways of pneumococcal recombination. *Nat Genet.* 2014;46:305–9.
25. von Mollendorf C, Cohen C, de Gouveia L, Quan V, Meiring S, Feldman C, Klugman KP, von Gottberg A. Factors associated with ceftriaxone nonsusceptibility of *Streptococcus pneumoniae*: analysis of South African national surveillance data, 2003 to 2010. *Antimicrob Agents Chemother.* 2014;58:3293–305.
 26. Huebner RE, Klugman KP, Matai U, Eggers R, Hussey G. Laboratory surveillance for *Haemophilus influenzae* type B meningococcal, and pneumococcal disease. Haemophilus Surveillance Working Group. *S Afr Med J.* 1999;89:924–5.
 27. von Gottberg A, de Gouveia L, Tempia S, Quan V, Meiring S, von Mollendorf C, Madhi SA, Zell ER, Verani JR, O'Brien KL, Whitney CG, Klugman KP, Cohen C. Effects of vaccination on invasive pneumococcal disease in South Africa. *N Engl J Med.* 2014;371:1889–99.
 28. Carvalho MG, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, Steigerwalt A, Whaley M, Facklam RR, Fields B, Carlone G, Ades EW, Dagan R, Sampson JS. Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA. *J Clin Microbiol.* 2007;45:2460–6.
 29. Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing; approved standard M100-S24. 34th ed. Wayne: Clinical and Laboratory Standards Institute; 2014.
 30. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14:1394–403.
 31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
 32. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol.* 2004;186:1518–30.
 33. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015, %20.
 34. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

