Genomic differences among carriage and invasive nontypeable pneumococci circulating in South Africa

Thabo Mohale1,2,*, Nicole Wolter1,2, Mushal Allam3, Susan A. Nzenze4,5, Shabir A. Madhi4,5, Mignon du Plessis1,2 and Anne von Gottberg1,2

Abstract
Most pneumococci express a polysaccharide capsule, a key virulence factor and target for pneumococcal vaccines. However, pneumococci showing no serological evidence of capsule expression [nontypeable pneumococci (NTPn)] are more frequently isolated from carriage studies than in invasive disease. Limited data exist about the population structure of carriage NTPn from the African continent. We aimed to characterize carriage NTPn and compare them to previously described invasive NTPn. Carriage and invasive NTPn isolates were obtained from South African cross-sectional studies (2009 and 2012) and laboratory-based surveillance for invasive pneumococcal disease (2003–2013), respectively. Isolates were characterized by capsular locus sequence analysis, multilocus sequence typing, antimicrobial non-susceptibility patterns and phylogenetic analysis. NTPn represented 3.7% (137/3721) of carriage isolates compared to 0.1% (39/32824) of invasive isolates (P<0.001), and 24% (33/137) of individuals were co-colonized with encapsulated pneumococci. Non-susceptibility to cotrimoxazole [84% (112/133) vs 44% (17/39)], penicillin [77% (102/133) vs 36% (14/39)], erythromycin [53% (70/133) vs 31% (12/39)] and clindamycin [36% (48/133) vs 18% (7/39)] was higher (P=0.03) among carriage than invasive NTPn. Ninety-one per cent (124/137) of carriage NTPn had complete deletion of the capsular locus and 9% (13/137) had capsule genes, compared to 44% (17/39) and 56% (22/39) of invasive NTPn, respectively. Carriage NTPn were slightly less diverse [Simpson’s diversity index (D)=0.92] compared to invasive NTPn [D=0.97]. Sixty-seven per cent (92/137) of carriage NTPn belonged to a lineage exclusive to NTPn strains compared to 23% (9/39) of invasive NTPn. We identified 293 and 275 genes that were significantly associated with carriage and invasive NTPn, respectively. NTPn isolates detected in carriage differed from those causing invasive disease, which may explain their success in colonisation or in causing invasive disease.

DATA SUMMARY
The genome assemblies of all 137 carriage nontypeable pneumococcal isolates sequenced in this study have been deposited in GenBank under BioProject ID PRJNA289243. Individual sample accession numbers can be found in Table S1 (available with the online version of this article).

INTRODUCTION
Streptococcus pneumoniae is a respiratory pathogen that frequently colonizes the nasopharynx, and can cause significant morbidity and mortality [1]. Most pneumococci are surrounded by a capsular polysaccharide, a key virulence factor [2]. The capsule induces protective antibodies and forms the basis for current pneumococcal vaccines [3]. More than
100 capsular types can be identified by the Quellung reaction using serotype-specific anti-sera [4–7]. Some pneumococci show no serological evidence of capsule expression and these nontypeable pneumococci (NTPn) are predominately carried in the nasopharynx [8–11]. The mechanisms underlying their nontypeability are diverse [8, 12–16]. NTPn have been associated with conjunctivitis outbreaks and non-invasive disease episodes, but rarely with invasive disease [17–19].

NTPn can be classified into two groups based on the gene content of the capsular polysaccharide synthesis (cps) locus [8]. In group I there are at least partial cps genes, whereas in group II the cps genes are completely deleted and may be replaced with non-cps genes. Group II is subdivided into four null capsule clades (NCCs) [19, 20]. NCC1 has the psk gene, NCC2 has aliC and aliD genes either with (NCC2a) or without (NCC2b) a putative toxin–antitoxin system (encoded by ntaAB genes). NCC3 has the aliD gene, and NCC4 contains only transposable elements in the cps locus.

Pneumococcal carriage studies from Europe reported specific lineages associated with NTPn strains [8, 11, 21]. A lineage of exclusively NTPn strains (classic lineage), comprising mainly sequence types (STs) 344, 448 and 449, and a lineage that appears to be related to encapsulated pneumococci (EcPn) strains (sporadic lineage) have been described [8, 21]. This was recently confirmed by whole-genome sequencing of 131 carriage NTPn isolates from 17 countries [22]. Carriage NTPn strains from Africa were not represented in these studies, and little is known about their population structure.

We previously characterized NTPn causing invasive pneumococcal disease (IPD) in South Africa [13]. These NTPn isolates represented 0.1% (39/32 824) of IPD cases and 5.1% (2/39) of IPD individuals were co-infected with EcPn. They were predominantly group I (56.4%) isolates that harboured a variety of mutations within their cps locus and had higher antimicrobial non-susceptibilities than EcPn. Invasive NTPn isolates were diverse [Simpson’s diversity index (D) of 0.97; 95% confidence interval 0.95–0.99] by multilocus sequence typing (MLST) analysis and the majority (79.4%) belonged to sporadic lineages. In this study, we characterized carriage NTPn and compared their population structure to the previously described invasive NTPn from South Africa.

**METHODS**

**NTPn isolates**

Invasive NTPn were obtained from active, national laboratory-based surveillance for IPD in South Africa from 2003 to 2013 and have been previously described [13]. Carriage NTPn were obtained from cross-sectional community carriage studies assessing the impact of seven-valent pneumococcal conjugate vaccine (PCV7) on pneumococcal nasopharyngeal colonization in South Africa. These cross-sectional carriage studies were conducted among randomly selected households with at least one child<2 years of age in a rural community (2009 and 2011) [23], and among human immunodeficiency virus (HIV)-infected and HIV-uninfected mother–child pairs in an urban community (2010 and 2012) [24]. PCV7 was introduced into the national immunization programme of South Africa in April 2009 and was replaced with PCV13 in July 2011.

Nasopharyngeal swabs (Medical Wire and Equipment) were collected, and placed in skimmed milk, tryptose, glycerol and glucose (STGG) broth transport media and stored at −70 °C. Swabs were cultured on 5% horse blood agar with 5 mg gentamicin sulphate ml⁻¹ (Diagnostic Media Products, National Health Laboratory Service, Johannesburg, South Africa) and incubated at 37°C in 5% CO₂ for 48 h. Pneumococcal isolates were identified by susceptibility to ethylhydrocupreine hydrochloride (5 µg optochin; Becton Dickinson Microbiology Systems), bile solubility as well as real-time PCR detecting lytA [25]. Nontypeable pneumococcal isolates were detected during routine serotyping using the Quellung reaction (Statens Serum Institut, Copenhagen, Denmark) [26]. Co-colonization was defined as simultaneous isolation of NTPn with EcPn from the same nasopharyngeal swab.

**Antimicrobial-susceptibility testing**

Minimum inhibitory concentrations (MICs) for all antibiotics were determined by the broth microdilution method and interpreted using the Clinical and Laboratory Standards Institute 2014 MIC breakpoints [27]. Isolates defined as either intermediately resistant or resistant to any of the antibiotics were regarded as non-susceptible. For penicillin and ceftriaxone, the more conservative meningitis MIC breakpoints...
were used as we wanted more resolution and as this analysis was not to guide clinical therapy. Multidrug resistance was defined as non-susceptibility to three or more classes of antibiotics.

**Nucleic acid extraction, genome sequencing and assembly**

DNA extraction, whole-genome sequencing and genome analysis were carried out as previously described for invasive NTPn [13]. Briefly, genomic DNA was extracted from overnight broth cultures using a QIAamp DNA mini kit (Qiagen). Paired-end libraries (2×300 bp) were prepared using the Nextera XT DNA sample preparation kit (Illumina) and sequencing was performed on an Illumina MiSeq. Reads were imported into CLC Genomics Workbench v8 (CLC Bio) and were quality trimmed using the Trim sequences module in the software with the limit set at 0.05 (e.g. sequences with phred score of <30 were trimmed). The trimmed reads were assembled using the de novo assembly algorithm of CLC Genomics Workbench v8. The word size was set at 21, minimum contig length at 500 and bubble size at 80% of mean read length. The assembled genomes were ordered relative to *S. pneumoniae* ATCC 700669 using Mauve [28] and annotated using Prokka v1.11 [29] (see Table S1 for genome characteristics).

**Characterization of the cps locus and STs**

The annotated genomes were imported into CLC Genomics Workbench v8. Using the ‘Find annotation’ tool, we searched for *dexB* and *aliA* genes, and extracted sequence data between these two genes (e.g. the *cps* locus) using the ‘Extract Sequences Tool’. STs were derived from genome data using the Bio-MLST-Check module (https://metacpan.org/release/AJPAGE/Bio-MLST-Check-1.133090). The eBURST v3 algorithm (http://eburst.mlst.net) was used to examine genetic relationships between carriage and invasive NTPn. A clonal complex (CC) was defined as isolates sharing 6/7 MLST alleles [30].

**In silico prediction of ancestral serotypes for group I carriage NTPn**

To predict ancestral serotypes of group I carriage NTPn (serotypes from which NTPn may have originated), three methods were used, namely serotype-ST associations, SeroBA and phylogenetic whole-genome analysis. In the first method, using the *S. pneumoniae* MLST database (http://pubmlst.org/spneumoniae/, accessed May 2016), we identified serotypes commonly associated with the ST of our group I carriage NTPn and, secondly, we used SeroBA v1.0.0, a rapid high-throughput *in silico* serotyping method for *S. pneumoniae* from whole-genome sequence data [31]. Lastly, we performed phylogenetic comparison (described below) of the carriage NTPn with invasive EcPn circulating in South Africa (Table S2). If group I NTPn clustered closely with a particular serotype in the phylogenetic tree, that serotype was assumed to be the likely ancestral serotype. If there were conflicting ancestral serotypes identified by the different
Table 1. STs, capsular loci (cps) and ancestral serotypes of group I carriage non-typeable *S. pneumoniae* (*n*=13) in South Africa, 2009–2012

| Isolate no. | ST | ST associated serotype | SeroBA serotype | Phylogenetic clustering | Predicted ancestral serotype | Mutations in the cps locus relative to ancestral serotypes | Comments |
|-------------|----|------------------------|-----------------|------------------------|-----------------------------|----------------------------------------------------------|----------|
| NT185       | 700| 3                      | 3               | 3                      | 3                           | Deletion of aliB, tnp, wgz, wzh. Partial deletion of wzd and wze. SNPs in wze, ugd, wchE and galU. Insertion in wze. |          |
| NT69        | 458| 3                      | 3               | 3                      | 3                           | Partial deletion of aliB and pgm. SNPs in wzh and aliB. |          |
| NT89        | 458| 3                      | 3               | 3                      | 3                           | Partial deletion of aliB, pgm and tnp. |          |
| NT205       | 458| 3                      | 3               | 3                      | 3                           | Partial deletion of pgm. |          |
| NT191       | 4872| 19A                   | 19A             | 19A                    | 19A                          | Deletion of IS1670. Partial deletion of IS1202. SNPs in IS1202, wgz, wxe, wzx, rmlD. Insertion in IS1202, wzx and rmlD. |          |
| NT103       | 5410| 4                     | 19A             | 4                      | 4                           | Deletion of IS1202. Partial deletion of IS1670 and rmlD. SNPs in wgz and IS1670. Insertion in rmlD. | Co-carried with 19A. Capsule compared to 19A*.
| NT244       | 6277| 19F                   | 19F             | 19F                    | 19F                          | Deletion of aliB and introns. Partial deletion and SNPs in tnp. Insertion in wxe. | Co-carried with 19F. |
| NT134       | 217| 1                      | 1               | 1                      | 1                           | Deletion of aliB. Partial deletion of tnp and wzx. SNPs in tnp, wzd, wze and gla. |          |
| NT207       | 2285| 6A                    | 6A              | 6A                     | 6A                           | SNPs in wzd. Partial deletion of wzx. |          |
| NT171       | 5647| 13                    | 13              | 13                     | 13                           | Deletion of tnp. Partial deletion of whatG and wzy. SNPs in wcIF. |          |
| NT177       | 393| 38                     | 38              | 38                     | 38                           | SNP in wcIF. Partial deletion of wzy, wcyD and gla. Insertion in wzy. |          |
| NT206       | 447| 37                     | 37              | 37                     | 37                           | SNPs in wgz, wzh, wzd, wxe, wcIB, wzy, wcIG, gfJ and wcIE. |          |
| NT66        | 10938| None                  | 10A             | 13                     | 13                           | SNP in aliB. Partial deletion of tnp. Capsule compared to 10A*. |          |

*Isolates had capsular genes not matching their predicted ancestral serotypes, maybe due to serotype switching (this is evident for isolate NT103, which was co-carried with serotype 19A and, therefore, likely took up the 19A capsule); hence, the capsule locus was compared to the switched capsule and not the ancestral serotype locus.

Methods, ancestral serotype assignment was based on the phylogenetic analysis result as this method accounted for the background genome of the isolate and was, therefore, assumed to represent the strain from which the NTPn isolate was potentially derived.

Molecular basis for the nontypeability of group I carriage NTPn

The ‘alignment tool’ in CLC Genomics Workbench v8 (default settings) was used to align the cps locus of each isolate,
Mohale et al., Microbial Genomics 2019;5

Fig. 2. eBURST comparison of nontypeable *S. pneumoniae* \((n=176)\) showing relationships between STs found in invasive \((n=39)\) [13] and carriage \((n=137)\) isolates from South Africa. STs shown in black were only found in invasive isolates, those shown in green were found in carriage isolates and those shown in pink were found in both. The size of each circle corresponds with the number of isolates. Clusters of linked isolates correspond to CCs (isolates sharing at least five of the seven alleles) and blue indicates the founding genotype. Group I carriage isolates are underlined with a dashed line. Asterisks denote new STs.

together with a number of its predicted ancestral serotype *cps* loci (Table S3). This was done to determine the genetic variation within the *cps* locus of group I carriage NTPn that may be responsible for the phenotypic nontypeability.

**Phylogenetic analysis**

To determine genetic relationships among the South African and other globally available carriage NTPn genomes, a rapid large-scale prokaryote pan-genome analysis pipeline (Roary) [32] and randomized accelerated maximum-likelihood (RAxML) analysis [33] were used to reconstruct a maximum-likelihood phylogenetic tree based on core genome SNPs of carriage NTPn from this study \((n=137)\), and a previously described collection of carriage NTPn from other countries \((n=131)\) [22]. The core genome alignment module in Roary [32] was used to extract predicted coding regions from Prokka-annotated assemblies and convert them to protein sequences. All protein sequences were compared with each other using blastp. Proteins that had alignment similarity of \(\geq 70.0\%\) and were present in at least 90.0\% of the isolates were defined as the core genome. RAxML [33] was used to create a bootstrapped maximum-likelihood phylogenetic tree from the resulting core-genome alignment and visualized in Phan-dango v1.3.0 [34]. Invasive EcPn from South Africa \((n=48)\) (Table S2) and previously described invasive NTPn from our setting [13] were included for comparison.

**Genomic comparison of carriage and invasive NTPn**

The pan-genome of carriage and invasive NTPn was inferred with Roary [32]. Prokka annotations were provided to Roary as an input and, in turn, Roary produced a gene presence/absence matrix for the isolates. To identify genes that were associated with carriage or invasive NTPn, genes were classified as unique if they were present in at least one carriage NTPn isolate and absent in invasive NTPn isolates or vice versa. We also examined whether genes were significantly over-represented in carriage versus invasive isolates and vice versa, using the chi-square test with significance assessed at \(P<0.05\). wego software [35] (http://wego.genomics.org.cn/, accessed 1 February 2017) was used to perform gene ontology functional classification for the genes that were significantly
over-represented in invasive NTPn and present in the majority (>70%) of invasive NTPn isolates.

**Statistical analyses**

Proportional differences were analysed using the chi-square or Fisher’s exact test, where appropriate. Statistical analyses were performed with GraphPad InStat 3 (GraphPad Software) and Microsoft Excel. ST diversity was calculated using Simpson’s diversity index (D) (http://darwin.phyloviz.net/dokuwiki/doku.php – under the comparing partitions tab). D ranges from 0 to 1, and values closer to 1 indicate higher diversity. Statistical significance was assessed at P<0.05.

**RESULTS**

**Prevalence of carriage NTPn**

The overall prevalence of *S. pneumoniae* colonization was 31.7% (3721/11 739), for which 3.7% (137/3721) isolates were NTPn. Amongst participants colonized with NTPn and for whom demographic data were available, 60.6% (63/104) were female, 81.5% (110/135) were children aged 5 years and 43.6% (41/94) were HIV infected. A total of 24.1% (33/137) of individuals carrying NTPn were co-colonized with EcPn, and the most common co-colonizing serotypes were serotype 19F (5/33, 15.2%), 6B (4/33, 12.1%), 6A (3/33, 9.1%), 23F (3/33, 9.1%) and 34 (3/33, 9.1%).

**Antimicrobial non-susceptibility of carriage NTPn compared to invasive NTPn**

Among carriage NTPn, 84.2% (112/133) were non-susceptible to cotrimoxazole, 76.7% (102/133) to penicillin, 52.6% (70/133) to erythromycin, 51.1% (68/133) to tetracycline and 36.1% (48/133) to clindamycin (Fig. 1). Antimicrobial-susceptibility testing was not performed for four isolates due to loss of viability during storage. Overall, with the exception of chloramphenicol non-susceptibility, which was higher in invasive NTPn compared to carriage NTPn [10.3% (4/39) versus 0.8% (1/133), \(P=0.01\)], non-susceptibility was higher among carriage NTPn than invasive NTPn for cotrimoxazole [84.2% (112/133) versus 43.6% (17/39), \(P<0.01\)], penicillin [76.7% (102/133) versus 35.9% (14/39), \(P<0.01\)], erythromycin [52.6% (70/133) versus 30.8% (12/39), \(P=0.02\)] and clindamycin [36.1% (48/133) versus 17.9% (7/39), \(P=0.03\)]. Multidrug non-susceptibility was higher in carriage NTPn compared to invasive NTPn [58.6% (78/133) versus 30.8% (12/39), \(P<0.01\)].

**Carriage NTPn cps loci**

Nine per cent (13/137) of carriage NTPn had conventional cps loci and were classified as group I (Table 1). For 84.6% (11/13) of group I carriage NTPn, all methods used to predict ancestral serotypes were in agreement. Two isolates (NT66 and NT103) had conflicting results where the ancestral serotype predicted by the capsular genes differed from that of the ST-associated serotype and phylogenetic clustering, possibly due to serotype switching. The most common predicted ancestral serotypes were serotype 3 (4/13, 30.8%) and 13 (2/13, 15.4%). Compared to their predicted ancestral serotypes (or switched serotype locus for NT66 and NT103), group I carriage NTPn had SNPs and deletions within their cps loci (Table 1). The remaining carriage NTPn (124/137, 90.5%) were classified as group II and had pspK (6/124, 4.8%), aliC, aliD and ntaAB (104/124, 83.9%), aliC and aliD (5/124, 4.0%), aliD (3/124, 2.4%), transposable elements (3/124, 2.4%) and aliC (3/124, 2.4%) in place of the cps locus. Based on the classification by Park et al. [20, 36], group II carriage NTPn were defined as NCC1, NCC2a, NCC2b, NCC3 and NCC4. Three group II carriage NTPn (3/124, 2.4%) contained only aliC and did not belong to any previously defined clade.

**ST analysis of carriage NTPn and comparison to invasive NTPn**

Among the 137 carriage NTPn, 44 STs were identified of which 24 were novel STs. The Simpson’s diversity index among the carriage NTPn was 0.92 (95% confidence interval: 0.90–0.95) (Fig. 2). Among the 13 group I carriage NTPn, only 3 isolates belonged to the same ST (ST458) (Fig. 2, Table 1). Among 124 group II carriage NTPn, 33 STs were identified of which 23 were novel STs. STs 9810 (25/124, 20.2%), 344 (21/124, 16.9%), 9809 (13/124, 10.5%), 9811 (13/124, 10.5%), 10972 (7/124, 5.6%), 9519 (6/124, 4.8%) and 7604 (5/124, 4.0%) represented 72.6% (90/124) of group II carriage NTPn. Four CCs namely CC344 (49/124, 39.5%), CC9810 (33/124, 26.6%), CC10972 (21/124, 16.9%) and CC9519 (8/124, 6.5%) accounted for 89.5% (111/124) of group II carriage NTPn. The majority (82/124, 66.1%) of group II carriage NTPn were related (sharing at least 5/7 MLST alleles) to international PMEN clones NorwayNT-ST344 and USAnt-ST448 (http://www.pneumogen.net/pmen/). Six STs (ST217, ST344, ST9809, ST9810, ST9811 and ST10241) were identified in both carriage (54.0%, 74/137) and invasive (41.0%, 16/39) NTPn.

**Phylogenetic analysis of carriage and invasive NTPn**

Carriage and invasive NTPn isolates clustered by ST, nontypeable lineage and nontypeable group (Fig. 3). Within the classic lineage, two CCs, namely CC344 and CC9810, were dominant. Within CC344, ST344 strains from South Africa clustered more closely with single-locus variant ST9809 South African isolates than with ST344 isolates from other countries. Overall, 67.2% (92/137) of carriage and 23.1% (9/39) (\(P<0.01\)) of invasive NTPn from our study belonged to the classic nontypeable lineage. The remaining carriage (45/137, 32.8%) and invasive (30/39, 76.9%) NTPn clustered with EcPn isolates in the sporadic nontypeable lineage. Group I carriage NTPn were closely related to their predicted ancestral serotypes (Table 1).

**Genomic comparison of carriage and invasive NTPn**

We identified a total of 11959 genes in the pan-genome of carriage and invasive NTPn, of which 39.9% (4777/11959) encoded hypothetical proteins. The core genome for all NTPn isolates was 1102 coding sequences, which translated to 55.1%...
Fig. 3. Maximum-likelihood phylogenetic tree of carriage (n=137) and invasive (n=39) nontypeable S. pneumoniae from South Africa, and a collection of carriage nontypeable isolates (n=131) from 17 countries (2 countries in Africa, 6 in Europe, 3 in North America, 1 in South America, 1 in Australasia and 4 in Asia). Encapsulated, invasive S. pneumoniae isolates representative of different serotypes from South Africa (n=47) were included for comparison. In CCs, other includes CCs: 74, 105, 353, 941, 1473, 4084, 5604 and 9817. The tree is based on core genome SNPs of the isolates. Bar, 0.04 nucleotide changes per 100 bp.

of the total genome. We found 5059 and 1298 genes that were exclusively present in carriage and invasive NTPn, respectively (Tables S4 and S5). Of these, 43.3% (2193/5059) and 41.7% (541/1298) encoded hypothetical proteins, in carriage and invasive NTPn, respectively. We also found 293 and 275 genes that were significantly associated with carriage and invasive NTPn, respectively (Tables S6 and S7). Of the 275 genes significantly associated with invasive NTPn, 59 genes (highlighted in red in Table S7) were present in at least 70% of invasive isolates, and 52 genes (88.1%) had BLAST matches to known proteins and were assigned to three gene ontology classes with 260 functional terms (Fig. 4). Assignments to biological processes, molecular function and cellular components were 48.8% (127/260), 34.2% (89/260) and 16.9% (44/260), respectively. Within biological processes, metabolic processes were highly represented (26.9%, 14/52); and within molecular function, catalytic activity (38.5%, 20/52) and binding (26.9%, 14/52) were highly represented.

**DISCUSSION**

We characterized and compared carriage and invasive NTPn circulating in South Africa. While NTPn caused IPD at a low prevalence (0.1%) [13], they were more frequently isolated in carriage (3.7%). Carriage NTPn were predominantly group II (90.5%) and classic NTPn lineage (67.1%), whereas invasive NTPn were predominantly group I (56.4%) and sporadic NTPn lineage (76.9%). Non-susceptibility to cotrimoxazole, penicillin, erythromycin and clindamycin was significantly higher in carriage NTPn compared to invasive NTPn. Carriage NTPn isolates were slightly less diverse than invasive NTPn (D=0.92 vs 0.97). We identified 293 and 275 genes that were exclusively found in carriage and invasive NTPn, respectively.

A small proportion (9.5%) of our carriage NTPn had partial cps genes compared to our invasive NTPn (56.4%) (P<0.01) [13]. This is similar to a pneumococcal carriage study conducted in Thailand from 2007 to 2010, where 9.2% (46/501) of carriage NTPn had partial cps genes [37]. For the two carriage group I isolates that had capsular genes not matching their predicted ancestral serotypes, we hypothesize that there may have been a serotype switch and that during or after this event the isolates lost the ability to express a capsule. In particular, isolate NT103 was co-carried with a serotype 19A isolate; therefore, we speculate that a serotype switch event resulted in the acquisition of 19A capsular genes.
As expected, we observed a higher prevalence (90.5%) of group II carriage NTPn than invasive NTPn (43.6%) \( (P<0.01) \) [13]. This was similar to the Thailand study, where group II represented 90.8% (455/501) of carriage NTPn [37]. Three of our group II carriage isolates (2.4%) contained only \( \text{allic} \) in the \( \text{cps} \) region and, thus, did not belong to any clade using the current classification [19, 20]. We propose that \( \text{cps} \) with \( \text{allic} \) be classified NCC5.

Twenty-four per cent of our carriage NTPn were co-carried with EcPn compared to invasive NTPn where 5.1% were co-infected with EcPn [13]. The EcPn co-colonizing with carriage NTPn were diverse, but serotypes 19F, 6B, 6A, 23F and 34 were most common. One group I carriage NTPn isolate (NT244), with a predicted 19F ancestral serotype, was co-colonized with a serotype 19F isolate. It is possible that NT244 is a variant of this 19F isolate; however, we did not assess the relationship between co-colonizing pairs. Schaffner et al. described a serotype 18C carriage isolate that lost its capsule, resulting in the co-existence of encapsulated and unencapsulated variants of the same strain [38]. We have also previously described two cases of IPD caused by co-infected encapsulated and unencapsulated variants of serotype 1 and 18C [14].

As co-colonization of NTPn with EcPn occurs frequently and provides an opportunity for exchange of genetic material [9, 11, 39], NTPn can serve as reservoir of resistance genes. Evidence for this was a study in Switzerland that found a serotype 19F clone became resistant to penicillin by acquisition of a gene from NTPn [40]. Recently, using whole-genome sequencing, a study in Thailand showed that higher rates of DNA exchange occurred in NTPn compared to EcPn, and the most commonly exchanged genes were those associated with antimicrobial resistance and immune interaction [37].

There were significantly higher rates of antimicrobial non-susceptibility among carriage NTPn than invasive NTPn for cotrimoxazole, penicillin, erythromycin and clindamycin. Since the majority of carriage NTPn (90.5%) lacked a capsule (true nontypeable), they are more likely to acquire resistance genes, compared to invasive NTPn where 56.4% were encapsulated but at some point sporadically lost the ability to express the capsule. Also, carriage NTPn have a greater risk of antimicrobial exposure than invasive NTPn and are, therefore, more likely to acquire resistance genes [41]. Lastly, this high resistance in carriage NTPn may be driven by global multidrug-resistant nontypeable clones Norway\textsuperscript{NT}-ST344 and USA\textsuperscript{NT}-ST448, as two-thirds of our carriage NTPn were related to these clones. High rates of antimicrobial non-susceptibility are common among carriage NTPn isolates [11, 37, 39].

Compared to our invasive NTPn, which were slightly more heterogeneous [13], MLST analysis of carriage NTPn showed that this population is slightly less diverse. This may be due to the fact that the majority (two-thirds) of carriage NTPn isolates were of the same clones (Norway\textsuperscript{NT}-ST344 and USA\textsuperscript{NT}-ST448). Similarly, the population structure of NTPn isolates from carriage studies conducted in Portugal in children aged <5 years between 2001 and 2007 was also clonal, with 83% (20/24) of NTPn isolates belonging to global clones Norway\textsuperscript{NT}-ST344 and USA\textsuperscript{NT}-ST448 [11]. In Thailand, approximately 40% of carriage NTPn were represented by three STs: ST448, ST4133 and ST4136.

Immune deficiencies of the host and/or other known or novel virulence factors have been proposed as explanations for the ability of NTPn to cause IPD without the capsule [12]. However, in our previous study, the majority of patients with invasive NTPn infection (87.0%) were not immunocompromised, although the numbers were small [13]. The genes that
we found to be significantly associated with invasive NTPn may play an important role in aiding NTPn to cause IPD and should be further investigated.

Our study has some limitations. It is possible that our NTPn isolates may have lost the ability to express a capsule during sub-culturing in the laboratory, but we believe that if this were the case we would detect NTPn much more frequently. We did not perform in vitro experiments to confirm whether the mutations in the cps locus of our isolates were responsible for the lack of capsule expression. However, in the majority of these isolates, there were complete deletions of cps genes making expression of a capsule impossible. We did not perform gene knockout or mouse model studies for the genes that were exclusively present in invasive NTPn compared to carriage NTPn to determine their role in virulence.

In conclusion, our study showed differences between carriage and invasive NTPn related to prevalence, capsule locus, relatedness, virulence, antimicrobial resistance and genome content. The genes that were significantly associated with invasive NTPn in this study should be further investigated to determine their role in virulence. Continued surveillance of NTPn is important since these isolates represent a reservoir of antibiotic-resistance genes for EcPn.

References
1. Wahls B, O’Brien KL, Greenbaum A, Majumder A, Liu L et al. Burden of Streptococcus pneumoniae and Haemophilus influenzae type B disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. Lancet Glob Health 2018;6:e744-e757.
2. 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;78:704-715.
3. 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.
4. Bentley SD, Aanensen DM, Mavrodi A, Saunders D, Rabinowitsch E et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet 2006;2:e331.
5. Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J et al. 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;287:27885–27894.
6. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP et al. Pneumococcal capsules and their types: past, present, and future. Clin Microbiol Rev 2015;28:871–899.
7. Geno KA, Saad JS, Nahm MH. Discovery of novel pneumococcal serotype 35D, a natural Wc/G-deficient variant of serotype 35B. J Clin Microbiol 2017;55:1416–1425.
8. Hathaway LJ, Stutzmann Meier P, Bättig P, Aebi S, Mühlemann K. A homologue of aⅡd is found in the capsule region of nonencapsulated Streptococcus pneumoniae. J Bacteriol 2004;186:3721–3729.
9. Sá-Leão R, Simões AS, Nunes S, Sousa NG, Frazão N et al. Identification, prevalence and population structure of non-typable Streptococcus pneumoniae in carriage samples isolated from preschoolers attending day-care centres. Microbiology 2006;152:367–376.
10. Scott JR, Hinds J, Gould KA, Miller EV, Reid R et al. 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.
11. Simões AS, Valente C, de Lena Castre R, Sá-Leão R. Rapid identification of noncapsulated Streptococcus pneumoniae in nasopharyngeal samples allowing detection of colonization and reevaluation of prevalence. Diagn Microbiol Infect Dis 2011;71:208–216.
12. Melchiorre S, Camilli R, Pietrantoni A, Moschioni M, Berti F et al. Point mutations in wchA are responsible for the non-typability of two invasive Streptococcus pneumoniae isolates. Microbiology 2012;158:338–344.
13. Mohale T, Wolter N, Allam M, Ndlangisa K, Crowther-Gibson P et al. Genomic analysis of nontypeable pneumococci causing invasive pneumococcal disease in South Africa, 2003–2013. BMC Genomics 2016;17:470.
14. Ndlangisa KM, du Plessis M, Allam M, Wolter N, Mohale T et al. Two cases of serotypeable and non-serotypeable variants of Streptococcus pneumoniae detected simultaneously during invasive disease. BMC Microbiol 2016;16:126–745.
15. Saltor SJ, Hinds J, Gould KA, Lambertsen L, Hanage WP et al. Variation at the capsule locus, cps, of mistyped and non-typable Streptococcus pneumoniae isolates. Microbiology 2012;158:1560–1569.
16. Waite RD, Penfold DW, Struthers JK, Dowson CG. Spontaneous sequence duplications within capsule genes capE and its control phase variation in Streptococcus pneumoniae serotypes 8 and 37. Microbiology 2003;149:497–504.
17. Crum NF, Barrozo CP, Chapman FA, Ryan MAK, Russell KL. An outbreak of conjunctivitis due to a novel unencapsulated Streptococcus pneumoniae among military trainees. Clin Infect Dis 2004;39:1148–1154.

18. Onwubiko C, Shires C, Quin LR, Swiatlo E, McDaniel LS. Characterization of Streptococcus pneumoniae isolated from children with otitis media. FEMS Immunol Med Microbiol 2007;50:119–125.

19. 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.

20. Park IH, Kim K-H, Andrade AL, Brides DE, McDaniel LS et al. Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene cpsN. MBio 2012;3:e00035–12.

21. Hanage WP, Kajjalainen T, Saukkoriipi A, Rickcott JD, Spratt BG. A successful, diverse disease-associated lineage of nontypeable pneumococci that has lost the capsular biosynthesis locus. J Clin Microbiol 2006;44:743–749.

22. Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S et al. Global phylogenomic analysis of unencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol Evol 2014;6:3281–3294.

23. Nzenze SA, Shiri T, Nunes MC, Klugman KP, Kahn K et al. Temporal changes in pneumococcal colonization in a rural African community with high HIV prevalence following routine infant pneumococcal immunization. Pediatr Infect Dis J 2013;32:1270–1278.

24. Nzenze SA, van Gottberg A, Shiri T, van Nierkerk N, de Gouveia L et al. Temporal changes in pneumococcal colonization in HIV-infected and HIV-uninfected mother–child pairs following transition from 7-valent to 13-valent pneumococcal conjugate vaccine, Soweto, South Africa. J Infect Dis 2015;212:1082–1092.

25. Carvalho M daGS, Tondella ML, McCaustland K, Weidleich L, McGee L et al. Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. J Clin Microbiol 2007;45:2460–2466.

26. Austrian R. The Quellung reaction, a neglected microbiologic technique. Mt Sinai J Med 1976;43:699–709.

27. Clinical and Laboratory Standards Institute. Performance Standards for Antimicrobial Susceptibility Testing: Approved Standard, M100-52A. Wayne, PA: Clinical and Laboratory Standards Institute; 2014.

28. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–1403.

29. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2013;30:2098–2099.

30. 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;186:1518–1530.

31. Epping L, van Tonder AJ, Gladstone RA, Bentley SD, Page AJ et al. SeroBA: rapid high-throughput serotyping of Streptococcus pneumoniae from whole genome sequence data. Microb Genom 2018;4:mgen.000186.

32. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3693.

33. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–2690.

34. Hadfield J, Croucher NJ, Goater RJ, Abudakah K, Aanensen DM et al. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics 2018;34:292–293.

35. Ye J, Fang L, Zheng H, Zhang Y, Chen J et al. WEGO: a web tool for plotting go annotations. Nucleic Acids Res 2006;34:W293–W297.

36. Park IH, Pritchard DG, Cartee R, Brandaos A, Brandlione MCC et al. Discovery of a new capsular serotype (6C) within serogroup 6 of Streptococcus pneumoniae. J Clin Microbiol 2007;45:1225–1233.

37. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet 2011;44:305–309.

38. Schaffner TO, Hinds J, Gould KA, Wüthrich D, Bruggmann R et al. A point mutation in cpsE renders Streptococcus pneumoniae nonencapsulated and enhances its growth, adherence and competence. BMC Microbiol 2014;14:210.

39. Andrade ALS, Franco CM, Lamaro-Cardoso J, André MCDPB, Oliveira LLG et al. Non-typeable Streptococcus pneumoniae carriage isolates genetically similar to invasive and carriage isolates expressing capsular type 14 in Brazilian infants. J Infect 2010;61:314–322.

40. Hauser C, Aebi S, Mühlmann 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;48:3563–3566.

41. Lehtinen S, Blanquart F, Croucher NJ, Turner P, Lipsitch M et al. Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. Proc Natl Acad Sci USA 2017;114:1075–1080.