

## **Epidemiological and ecological modelling reveal diversity in upper respiratory tract microbial population structures from a cross-sectional community swabbing study**

Abigail L. Coughtrie<sup>1</sup>; Denise E. Morris<sup>1</sup>; Rebecca Anderson<sup>1</sup>; Nelupha Begum<sup>1</sup>; David W. Cleary<sup>1</sup>; Saul N. Faust<sup>1,3,4</sup>; Johanna M. Jefferies<sup>1,4</sup>; Alex R. Kraaijeveld<sup>2</sup>; Michael V. Moore<sup>1</sup>; Mark A. Mullee<sup>1,5</sup>; Paul J. Roderick<sup>1</sup>; Andrew Tuck<sup>1</sup>; Robert N. Whittaker<sup>1</sup>; Ho Ming Yuen<sup>1</sup>; C. Patrick Doncaster<sup>2</sup>; Stuart C. Clarke<sup>1,4,6</sup>

<sup>1</sup>Faculty of Medicine and Institute for Life Sciences, University of Southampton, Southampton, UK;

<sup>2</sup>Biological Sciences, University of Southampton, Southampton, UK;

<sup>3</sup>Southampton NIHR Wellcome Trust Clinical Research Facility, University Hospital Southampton Foundation NHS Trust, Southampton, UK;

<sup>4</sup>Southampton NIHR Respiratory Biomedical Research Unit, University Hospital Southampton Foundation NHS Trust, Southampton, UK;

<sup>5</sup>NIHR Research Design Service South Central, University Hospital Southampton Foundation NHS Trust, Southampton, UK;

<sup>6</sup>Global Health Research Institute, University of Southampton, Southampton, UK

### **Key words**

Ecology, Epidemiology, Respiratory Infection

### **Running Title**

Application of epidemiological and ecological modelling to microbial populations

### **Corresponding Author**

Dr Stuart C. Clarke

Infectious Disease Epidemiology Group

Academic Unit of Clinical and Experimental Sciences

Mailpoint 814, Level C

Faculty of Medicine, University of Southampton

University Hospital Southampton Foundation NHS Trust

Southampton, SO16 6YD

United Kingdom

S.C.Clarke@southampton.ac.uk

## ABSTRACT

Respiratory tract infections (RTI) are responsible for over 4 million deaths per year worldwide with pathobiont carriage a required precursor to infection. Through a cross-sectional community-based nasal self-swabbing study we sought to determine carriage epidemiology for respiratory pathogens amongst bacteria (*Streptococcus pneumoniae*, *Haemophilus influenza*, *Moraxella catarrhalis*, *Staphylococcus aureus*, *Pseudomonas aeruginosa* and *Neisseria meningitidis*) and viruses (RSV, Influenza viruses A and B, Rhinovirus/Enterovirus, Coronavirus, Parainfluenza viruses 1-3 and Adenovirus (ADV)). Carriage of bacterial and viral species was shown to vary with participant age, recent RTI and the presence of other species. The spatial structure of microbial respiratory communities was less nested (more disordered) in the young (0-4 years) and those with recent RTI. Species frequency distributions were flatter than random expectation in young individuals ( $X^2 = 20.42$ ,  $p = 0.002$ ), indicating spatial clumping of species consistent with facilitative relationships amongst them. Deviations from a neutral model of ecological niches were observed for samples collected in the summer and from older individuals (those aged 5-17, 18-64 and  $\geq 65$  years) but not in samples collected from winter, younger individuals (those aged 0-4 years), individuals with recent RTI and individuals without recent RTI, demonstrating the importance of both neutral and niche processes in respiratory community assembly. The application of epidemiological methods and ecological theory to sets of respiratory tract samples has yielded novel insights into the factors that drive microbial community composition, such as seasonality and age, as well as species patterns and interactions within the nose.

## INTRODUCTION

Respiratory tract infections (RTI) remain a significant cause of mortality worldwide, with 2.8 million deaths, from an estimated four million per annum, resulting from lower RTI and a further 3,000 deaths resulting from upper RTI in 2010 (1). An estimated six million individuals in the UK visit their general practitioner each year regarding a RTI, posing a substantial burden on the health service (2).

Microbial carriage of pathogens in the upper respiratory tract is a precursor to RTI, meningitis and sepsis and can also facilitate the transmission of microbes between individuals. Carriage is a dynamic process involving fluctuations of species and strains (3). The interactions and interplay between bacterial and viral species within the respiratory tract is thought to affect the shift from asymptomatic carriage to serious invasive illness (4). Such interactions will be driven by species presence and diversity of the resident communities of the respiratory tract. Here diversity in species richness and abundance is thought to offer protection from invading pathogens and overgrowth of a single species (5, 6). It is likely that the dynamics of community structure may impact on infection risk. Species diversity can be measured in the presence of different species within the respiratory tract (inter-species diversity) or the presence of different bacterial strains and types of a single species (intra-species). The former can be assessed using conventional culture methods or molecular techniques such as real-time polymerase chain reaction (PCR) and 16S rRNA gene sequencing (7, 8). Intra-species analysis gives insight into the distribution of different types of a species that may display differential capacity for disease. It requires techniques such as serotyping and multi-locus sequence typing (MLST) (9-12), or expansions of the MLST methodology such as whole-genome MLST (wgMLST) (13) and ribosomal MLST (rMLST) (14).

The processes that sustain biodiversity may involve species occupying more or less distinct ecological niches, or they may involve neutral processes of stochastic zero-sum dynamics amongst more or less identical species (15). Niche dynamics describe the distinct nutritional and environmental influences that determine the occurrences and abundances of different species (16, 17). Neutral dynamics describe stochastic influences on species abundance and diversity, such as birth, death, immigration, dispersal and speciation (5, 18). These alternative dynamics have been much studied in terms of species distribution and community assembly patterns for plant-insect (19), plant-fungi (20) and parasite-fish (21) interactions. They have not, to our knowledge, been applied to respiratory tract microbial ecology.

Here, we aim to understand how bacterial and viral species maintain diversity within the respiratory tract, the nature of their interactions and the factors influencing community assembly and colonisation. To our best knowledge this is the first study that aims to discern multi-species epidemiology using a broad cross-sectional community cohort with the added dimension of seasonality. This was achieved using culture-based techniques, molecular methods, mathematical modelling and the application of ecological models.

## **METHODS**

### **Sample Collection and Analysis**

Nose swab samples were collected as part of a large population-based cross-sectional carriage study undertaken with two separate time-points, May to August 2012 (late spring/summer) and February to April 2013 (winter/early spring), following protocols described in reference (22). Briefly this involved the collection of nose swabs from the two time-points were collected from different individuals from twenty general practitioners (GP) practices from the South East hub of the Wessex Primary Care Research Network. Figure 1 depicts the structure of the study. Swabs were immersed into skimmed milk, tryptone, glucose and glycerine (STGG) storage media and analysed for the presence of six bacterial species: *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, *S. aureus*, *P. aeruginosa* and *N. meningitidis* using standard microbiological identification techniques (methodology described in Supplementary Methods 1).

### **Statistical Analysis**

Binary logistic regression (SPSS, version 20) determined the effect of patient characteristics and demographic factors on carriage of the target bacterial species. Univariate linear regression was used to pre-select variables for inclusion in the final model (at Type-I threshold  $\alpha = 0.10$ ). Multivariable binomial logistic regression was then performed using these pre-selected variables, followed by backward elimination of those not meeting a Type-I threshold  $\alpha = 0.05$  within the final model. Crude and adjusted odds ratios (OR), 95% confidence intervals and *p*-values were recorded. Final model fit was assessed using Nagelkerke pseudo- $R^2$  values and Receiver Operating Characteristic (ROC) analysis for calculation of the area under the curve (AUC). High pseudo- $R^2$  values and an AUC significantly different from 0.5 indicate a model that predicts carriage better than chance variation.

Variables assessed for inclusion by univariate analyses included age (in years), recent RTI (self-reported, RTI within a month prior to swabbing), geographical location (practices were grouped into six geographical areas, recent use of antibiotics (self-reported, antibiotics taken within a month prior to swabbing), immunisation status (self-reported, immunisations in-line with the national schedule), season (winter or summer swabbing) and presence of other microbial species within the same sample.

## Phenotypic and Epidemiological Typing

Phenotypic and epidemiology typing of *S. pneumoniae*, *H. influenzae* and *S. aureus* was performed in order to determine serotype, sequence type, antibiotic resistance gene and vaccine candidate gene distribution among isolates. PCR serotyping was used to determine serotypes of *S. pneumoniae* and *H. influenzae* identified within summer samples only. *In silico* serotyping was used to determine serotypes of *S. pneumoniae* and *H. influenzae* identified within summer and winter samples. Methicillin resistance typing was done on *S. aureus* isolates identified within both summer and winter samples to determine the prevalence of methicillin-resistant *S. aureus* (MRSA).

### *Nucleic acid extraction*

For serotyping and whole genome sequencing, genomic DNA was extracted from bacterial colonies using the QIAamp DNA Mini Kit Blood (Qiagen, Germany) according to the manufacturer's protocol. For real-time PCR, nucleic acid extraction (DNA and RNA) from STGG samples was undertaken using the QIAextractor automated nucleic acid purification instrument and QIAamp Viral RNA Mini kits (Qiagen, Germany) according to the manufacturer's vaccum protocol.

### *PCR Serotyping*

*S. pneumoniae* isolates were serotyped using the Centre for Disease Control and Prevention (CDC) protocol (9). This involved eight separate PCR reaction pools covering 40 of the most common pneumococcal serotypes (Supplementary Table 1). Each PCR was undertaken in a 12.5µl volume containing 0.5µl 50mM MgCl<sub>2</sub>, 6.25µl 2x Red PCR, 0.0125µl 100mM *cpsA* (capsular polysaccharide biosynthesis gene for identification of *S. pneumoniae*) forward and reverse primers, 100mM serotype forward and reverse primers (volumes in Supplementary Table 1), Ultra-High Quality (UHQ) water and 1µl extracted DNA (or control). Thermal cycling was performed under the following conditions: 94°C for 4 minutes followed by 30

cycles of 94°C for 45 seconds, 54°C for 45 seconds and 72°C for 2 minutes 30 seconds. Serotyping of *H. influenzae* was undertaken using previously published methods (23, 24). Speciation PCR was undertaken for detection of the *ompP2* gene. Each PCR was undertaken in a 12.5µl volume containing 0.625µl 10mM forward and reverse primers, 6.25µl 2x Red PCR Mix, 4µl UHQ water and 1µl extracted DNA (or control). Thermal cycling was performed under the following conditions: 95°C for 2 minutes then 25 cycles of 94°C for 1 minute, 52°C for 1 minute and 72°C for 1 minute and then a final step of 72°C for 8 minutes. Encapsulation PCR was undertaken for detection of the *bexB* gene. Each PCR was undertaken in a 20µl volume containing 1µl 10mM forward and reverse primers, 10µl 2x Red PCR Mix, 6.4µl UHQ water and 1.6µl extracted DNA (or control). Thermal cycling was performed under the following conditions: 95°C for 4 minutes then 30 cycles of 95°C for 30 seconds, 54°C for 30 seconds and 72°C for 45 seconds and then a final step of 72°C for 8 minutes. Serotyping PCR was undertaken for detection of serotype a-f genes. Each PCR was undertaken in a 12.5µl volume containing 0.625µl 10mM forward and reverse primers for each serotype, 6.25µl 2x Red PCR Mix, 4µl UHQ water and 1µl extracted DNA (or control). Thermal cycling was performed under the following conditions: 95°C for 2 minutes then 25 cycles of 94°C for 1 minute, 52°C for 1 minute and 72°C for 1 minute and then a final step of 72°C for 8 minutes. All PCR products were visualised using a 2% high-resolution agarose gel.

#### *Whole Genome Sequencing*

Sequencing libraries were prepared using Nextera XT (Illumina, USA). Whole genome sequencing was performed using the Illumina MiSeq paired-end chemistry (2x150bp or 2x250bp). Adapter sequences and low quality regions were removed using Trimmomatic (Usadel Lab, 0.32), quality control of reads was performed using FastQC (Babraham Bioinformatics, version 0.11.2) and then assembled *de novo* using Velvet (Victorian Bioinformatics Consortium, version 2.2.5) (25). Assemblies were then improved using AssemblyImprovement (Wellcome Trust Sanger Institute, version 1.140300).

#### *In silico serotyping*

*In silico* PCR analysis was then undertaken using iPCRess (European Bioinformatics Institute, Exonerate package, version 1.2.0). Primer sequences for *in silico* serotyping used in iPCRess are shown in Supplementary Table 2.

#### *Gene Detection*

SRST2 (Short Read Sequence Typing, version 0.1.4) (26) was used to screen the whole genome sequences of the selected *S. pneumoniae* and *H. influenzae* isolates for defined

sets of genes: housekeeping genes for determination of sequence type (ST), antibiotic resistance genes (ResFinder/Comprehensive Antibiotic Resistance Database (27, 28)) and genes associated with virulence (Pathogenic Bacteria Database (VFBD) (29, 30)). Antigens that are or have been considered as vaccine candidates were of particular interest. *S. pneumoniae* vaccine candidates included *piaA*, *pspA*, *pspC/cbpA*, *psaA*, *nanA*, *lytA* and *ply* (31-34). *H. influenzae* vaccine candidates included *hpd*, *ompP2*, *ompP5*, *tbpA/B*, *hmw1/2* and *Hia* (35, 36). iPCReSS was also used to identify penicillin binding protein genes *pbp1a*, *pbp2x* and *pbp2b* genes within *S. pneumoniae* isolates (37) and the *H. influenzae* protein D (*hdp*) gene (38).

Simpson's Index of Diversity,  $1 - D$ , was used to quantify the diversity of species' serotypes and STs. The value of  $D$  gives the probability that two randomly-sampled individuals belong to different species. The index was calculated from the sum across all species of the number of individuals of each species:

$$1 - D = 1 - \left( \frac{\sum n(n-1)}{N(N-1)} \right)$$

$n$  = Total number of individuals of a particular species

$N$  = total number of individuals of all species

### *MRSA typing*

*S. aureus* isolates were tested for methicillin resistance by plating out 10 $\mu$ l of pure culture stored in STGG onto Brilliance MRSA Agar (Oxoid, PO1162) and incubating at 37°C for 18-20 hours. The presence of denim blue colonies of the plates was indicative of MRSA.

### **Real-time PCR**

A subset of nose swabs ( $N=190$ ) was analysed by real-time PCR to detect the presence of *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, *S. aureus*, *P. aeruginosa*, *Neisseria meningitidis*, Respiratory Syncytial Virus, Influenza viruses A and B, Rhinovirus/Enterovirus, Coronavirus, Parainfluenza viruses 1-3 and Adenovirus. Details of primers and probes used in the real-time PCR reaction are listed in Supplementary Table 3. Argene PCR kits were used for the detection of Coronavirus and Parainfluenza Virus (71-045) as well as Enterovirus/Rhinovirus (71-042). Bacterial real-time PCR reactions were performed in monoplex 20 $\mu$ l reactions, consisting of 5 $\mu$ l of the primer/probe working stock solution, 10 $\mu$ l Roche Lightcycler 480 Probes Master and 5 $\mu$ l extracted DNA. Viral real-time PCR reactions were done in multiplex 25 $\mu$ l reactions. The Influenza A-B and BMV reaction consisted of 12.5 $\mu$ l Superscript III One-Step RT-PCR System 2x Reaction Mix, 0.8 $\mu$ l Invitrogen

SuperScript III RT/Platinum Taq Enzyme Mix, 6.7µl of the primer/probe working stock solution and 5µl extracted RNA. The RSVA-B and MPV reaction consisted of 12.5µl Superscript III One-Step RT-PCR System 2x Reaction Mix, 0.5µl SuperScript III RT/Platinum Taq Enzyme Mix, 6.7µl of the primer/probe working stock solution and 5µl extracted RNA. The Adenovirus reaction consisted of 12.5µl Applied Biosystems TaqMan 2x Universal PCR Mastermix, 7.5 µl of the primer/probe working stock solution and 5µl extracted DNA. For the commercial kits, a 1 in 10 dilution of the reverse transcriptase (RT) was made up in UHQ water. For each sample, 0.15µl of diluted RT was added to 15µl pre-amplification mix. 15µl of the amplification mix was then aliquoted into the reaction tube along with 10µl of extracted RNA. Each multiplex bacterial real-time PCR was run on the Rotor-Gene Q (Qiagen, Germany) with the following thermal cycling conditions: 95°C for 5 minutes and 50 cycles of 95°C for 15 seconds followed by 60°C for 45 seconds. Each multiplex viral real-time PCR was run on the ABI 7500 (Applied Biosystems, USA) with the following thermal cycling conditions: 50°C for 30 minutes, 95°C for 2 minutes and 50 cycles of 95°C for 15 seconds followed by 60°C for 60 seconds. Both commercial PCR kits were run on the Rotor-Gene Q with the following thermal cycling conditions: 50°C for 5 minutes, 95°C for 15 minutes and 45 cycles of 95°C for 10 seconds, 60°C for 40 seconds and 72°C for 25 seconds. Detection of bacterial and viral species was determined using Cycle Threshold ( $C_T$ ) values reported by the Rotor-Gene Q and ABI 7500 system.

#### *Quantification of bacterial species*

Standard curve methodology was used for quantification of bacterial species. Reference strains of *S. pneumoniae* (ATCC 49619), *S. aureus* (clinical isolate), *H. influenzae* (NCTC 11931), *M. catarrhalis* (clinical isolate), *P. aeruginosa* (NCTC 10662) and *N. meningitidis* (MC58) were used to make 1-2 McFarland solutions in 1ml UHQ water. Ten-fold series dilution of each bacterial solution were prepared up to  $10^{-9}$  and plated out in triplicate 10µl spots onto CBA (or chocolate CBA for *H. influenzae*). Following incubation at 37°C for 24 hours (with 5% CO<sub>2</sub> for *S. pneumoniae*, *H. influenzae*, *M. catarrhalis* and *N. meningitidis* plates were counted and CFU/ml values calculated for each dilution. All dilutions were also extracted using the QIAextractor nucleic extraction and amplified in duplicate on the Rotor-Gene Q. The CFU/ml for each dilution was then correlated to a given  $C_T$  value to create a standard curve for each species, which was then used for quantification.

#### **Ecological Analysis**

All ecological analyses used real-time PCR results for bacterial species from nose swabs ( $N = 190$  species).

### *Nestedness*

Nestedness is a measure of organisation within an ecological system that quantifies the ordering of species incidence across space or time. A strongly nested system has a highly predictable sequence of species incidence amongst samples, while a weakly nested system has a more random turnover of species incidence. A nestedness tool, PlotTemp.R, was used within the R graphical user interface (CRAN, version 3.0.1) to determine spatial ordering amongst nose-swab samples analysed with real-time PCR. Real-time PCR datasets were converted to incidence matrices of the presence/absence of each species in each sample. The plot\_temp script shuffles each matrix to rank its rows of samples by their richness (most species rich at top) and its columns of species by their incidence (most frequently present to left). Within the final shuffled matrix, presences are represented by red squares and absences by white squares. Their locations in the matrix are compared to a line of perfect fill that would encompass all presences if they were perfectly packed towards the top-left of the matrix. This line then reveals the locations of 'surprise' presences lying to its right and 'surprise' absences lying to its left. The degree of disorder in the matrix is reported as a matrix 'temperature', ranging from 0-100°, calculated as a function of the sum across all surprises of squared deviations from the line of perfect fill. Low temperatures reveal strong nesting, characteristic of highly ordered systems, and high temperatures reveal weak nesting, characteristic of disordered systems. Nestedness was compared between groups of individuals partitioned by age, recent RTI and season.

### *Species distributions*

The distributions of the different bacterial and viral species across swab samples were analysed by plotting the observed frequencies of 1, 2, 3, ... species within swab samples. This observed distribution was compared to the Poisson distribution as a null model of random species affiliations. Deviations of observed frequencies from Poisson expectation were detected with  $p < 0.05$  in a  $\chi^2$  goodness-of-fit test. Narrower and more peaked distributions than Poisson expectation indicated a more regular than random distribution, with many swab samples containing a similar number of species, providing indirect evidence of mutual repulsion in competition. Wider and flatter distributions than Poisson expectation indicated a more clumped than random distribution, with many swab samples containing few or numerous species, providing indirect evidence of mutual attraction in facilitation. Frequency distributions were compared between groups of individuals partitioned by age, recent RTI and season.

### *Community assembly patterns*

Plots of ranked species abundances, measured as bacterial concentrations in CFU/ml, were evaluated against McArthur's Broken Stick model of neutral community assembly. This model compares an ecological environment to a stick that is broken at random points to produce fragments of lengths that represent the abundances of different species (39, 40). Deviations from neutral expectation were detected with  $p < 0.05$  in a  $\chi^2$  goodness-of-fit test. Observations of rare species below neutral expectation indicate niche characteristics of a dominance hierarchy, while observations of rare species above neutral expectation indicate niche characteristics of resource segregation (15).

## RESULTS

### Participants

A total of 2,103 patients from 20 GP practices across the South East hub of the Wessex Primary Care Research Network participated in self-swabbing across the two time-points. Different individuals were swabbed within each time-point. The characteristics of the study participants are shown in Table 1. Study participants were characterised as having a mean age of 38.7 years (SD 29.4), 8.6% with recent antibiotic use, 34.8% with recent RTI and 80.9% with up-to-date vaccination status. Participation was 23.4% ( $n = 1,260$ , 95% CI 22.27% to 24.53%) in summer and 18.9% ( $n = 843$ , 95% CI 17.7% to 20.0%) in winter. Individual GP practice participation varied from 9.3% ( $n = 27$ ) to 33.1% ( $n = 96$ ) in the summer self-swabbing group and 9.1% ( $n = 21$ ) to 30.4% ( $n = 80$ ) in the winter self-swabbing group.

### Carriage of Bacterial and Viral Species

Unadjusted and age group-specific prevalence of bacterial and viral species carriage within nose swabs detected by culture and real-time PCR are shown in Tables 2-4. All 2,103 nose samples were analysed by culture and a subset of 190 nose samples were analysed by real-time PCR. Carriage was higher in younger age-groups (0-4 and 5-17) for *S. pneumoniae*, *M. catarrhalis* and *H. influenzae*. Conversely higher carriage levels were observed for *S. aureus* in the older age groups. Comparatively fewer incidences of *P. aeruginosa* or *N. meningitidis* carriage were observed (Tables 2 and 3). Viral carriage was shown to be dominated by Rhinovirus/Enterovirus (13.2%), Coronavirus (10%) and RSV (6.8%) (Table 4). Carriage of bacterial and viral species within the respiratory tract was shown to vary with participant age (*S. pneumoniae* aOR=0.943; *H. influenzae* aOR=0.948; *M. catarrhalis* aOR=0.981, RV/EV FET  $p < 0.001$ ), recent RTI (*S. pneumoniae* aOR=2.071; RV/EV FET  $p < 0.001$ ; COV FET

$p=0.013$ ) and the presence of other species (Table 5). Modelling of viral species prevalence was not possible due to small numbers of viruses detected within this sample set, hence only descriptive statistics are shown in Table 6 for RSV, adenovirus, rhinovirus/enterovirus and coronavirus.

## Phenotypic and Epidemiological Typing

### *Serotyping*

A total of 209 *S. pneumoniae* isolates were serotyped, 10.5% ( $n = 22$ ) of which were PCV7/PCV13 vaccine types, 85.2% ( $n = 178$ ) non-vaccine types and 4.3% ( $n = 9$ ) non-typeable/unknowns. The most common serotypes were 11A/D (11.5%,  $n = 24$ ), 6C (8.6%,  $n = 18$ ), 15B/C (8.6%,  $n = 18$ ), 24A/B/F (6.7%,  $n = 14$ ), 23B (6.7%,  $n = 14$ ), 19A (6.2%,  $n = 13$ ) and 35F/47F (6.2%,  $n = 13$ ). Seventy-four *H. influenzae* isolates were serotyped, all of which were identified as non-typeable *H. influenzae* (NTHi).

### *Sequence Typing*

Sixty-three distinct sequence types (ST) of *S. pneumoniae* were identified within the set of isolates. There were also 11 isolates of previously unknown ST. The most common STs were 62 (10.5%,  $n = 22$ ), 199 (10.0%,  $n = 21$ ), 162 (5.7%,  $n = 12$ ), 439 (5.3%,  $n = 11$ ) and 1635 (5.3%,  $n = 11$ ). ST199 was frequently associated with serotypes 19A (9/13, 69.2%) and 15B/C (11/18, 61.1%). ST162 was frequently associated with serotype 24A/B/F (10/14, 71.4%). ST439 was frequently associated with serotype 23B (11/14, 78.6%) ST1635 was frequently associated with serotype 35F/47F (11/13, 84.6%). ST62 was frequently associated with serotype 11A/D (22/24, 91.7%). The set of *S. pneumoniae* STs gave a Simpson's Index of Diversity (1-D) of 0.96.

Forty-two distinct sequence types of *H. influenzae* were identified within the set of isolates. The *H. influenzae* isolates were diverse with only one or two isolates of each ST identified. The most common STs were 474 (6.8%,  $n = 5$ ), 57 (4.1%,  $n = 3$ ), 348 (4.1%,  $n = 3$ ), 569 (4.1%,  $n = 3$ ) and 1215 (4.1%,  $n = 3$ ). Thirteen isolates were of previously unknown sequence type. The set of *H. influenzae* isolates gave a Simpson's Index of Diversity (1-D) of 0.99.

### *Antibiotic Resistance*

Genes associated with tetracycline resistance, *tetM* ( $n = 9$ , 4.3%), *tetS* ( $n = 2$ , 1.0%) and *tet32* ( $n = 1$ , 0.5%); macrolide resistance *mefA* ( $n = 4$ , 1.9%) and *msrD* ( $n = 4$ , 1.9%); erythromycin resistance *ermB* ( $n = 6$ , 2.9%) and chloramphenicol resistance *cat* ( $n = 1$ ,

0.5%) were detected in *S. pneumoniae* isolates. *S. pneumoniae* beta-lactam susceptibility genes were *pbp1a* ( $n = 195$ , 93.3%), *pbp2x* ( $n = 201$ , 96.2%) and *pbp2b* ( $n = 185$ , 88.5%). Genes associated with beta-lactam resistance *blaTEM* ( $n = 6$ , 8.1%) and aminoglycoside resistance *aph(3')-I* ( $n = 3$ , 4.1%) were identified in *H. influenzae* isolates.

#### *Vaccine Candidate Gene Detection*

Prevalence of vaccine candidate genes in the 34 *S. pneumoniae* isolates were *lytA* ( $n = 34$ , 100%), *piaA* ( $n = 34$ , 100%), *ply* ( $n = 34$ , 100%), *psaA* ( $n = 34$ , 100%), *pspA* ( $n = 9$ , 26.5%), *pspC/cbpA* ( $n = 28$ , 82.4%) and *nanA* ( $n = 32$ , 94.1%). Prevalence of vaccine candidate genes in the 32 *H. influenzae* isolates were *hpd* ( $n = 27$ , 84.4%), *ompP2* ( $n = 32$ , 100%), *ompP5* ( $n = 29$ , 90.6%), *tbpA/B* ( $n = 32$ , 100%), *hmw1* ( $n = 5$ , 15.6%), *hmw2* ( $n = 5$ , 15.6%) and *hia* ( $n = 11$ , 34.4%).

#### *MRSA Typing*

A total of 12 (2.9%) methicillin-resistant *S. aureus* isolates were identified within 409 nose isolates of *S. aureus* tested. This included 10 (3.8%) in summer nose swabs and 2 (1.4%) in winter nose swabs.

## **Species Organisation and Community Assembly Patterns**

#### *Nestedness*

Microbial species showed a more ordered incidence across samples, indicated by lower nestedness temperature,  $T$ , amongst individuals with no recent RTI ( $T = 8.47^\circ$  [Matrix fill = 0.20]) than amongst those with recent RTI ( $T = 12.54^\circ$  [Matrix fill = 0.22]). Figure 2 shows the more disordered turnover of species between samples for recently ill individuals (Fig. 2a) compared to healthy individuals (Fig. 2b). Incidence was also more ordered amongst winter samples ( $T = 7.04^\circ$  [Matrix fill = 0.18]) than summer samples ( $T = 12.39^\circ$  [Matrix fill = 0.24]), and amongst older individuals (18-64 year olds  $T = 8.92^\circ$  [Matrix fill = 0.23];  $\geq 65$  year olds  $T = 6.45^\circ$  [Matrix fill = 0.15]) than younger individuals (0-4 year olds  $T = 11.67^\circ$  [Matrix fill = 0.27]; 5-17 year olds  $T = 15.21^\circ$  [Matrix fill = 0.28]).

#### *Species Distribution*

Frequency distributions of microbial species richness deviated from random amongst summer samples ( $X^2 = 21.75$ ,  $df = 5$ ,  $p < 0.001$ ), 0-4 year olds ( $X^2 = 20.42$ ,  $df = 6$ ,  $p = 0.002$ ) and individuals without recent RTI ( $X^2 = 15.44$ ,  $df = 5$ ,  $p = 0.009$ ). These classifications had wider frequency distributions than Poisson expectation, with more occurrences of few or many species than would be expected from a random distribution. This wider distribution

indicates clumping of species. In particular *S. pneumoniae*, *M. catarrhalis* and *H. influenzae* were shown to co-occur (*S. pneumoniae* - *M. catarrhalis* OR 4.160 (2.914-7.293)  $p < 0.001$ ; *S. pneumoniae* - *H. influenzae* OR 11.641 (7.288-18.595)  $p < 0.001$ ; *H. influenzae* - *M. catarrhalis* OR 6.277 (3.441-11.268)  $p < 0.001$ ) (Table 5). Conversely, frequency distributions of microbial species richness did not deviate detectably from random amongst winter swabs samples ( $X^2 = 6.61$ ,  $df = 6$ ,  $p = 0.358$ ), 5-17 year olds ( $X^2 = 8.35$ ,  $df = 5$ ,  $p = 0.138$ ), 18-64 year olds ( $X^2 = 5.32$ ,  $df = 4$ ,  $p = 0.256$ ),  $\geq 65$  year olds ( $X^2 = 0.153$ ,  $df = 3$ ,  $p = 0.985$ ) and individuals with recent RTI ( $X^2 = 7.46$ ,  $df = 5$ ,  $p = 0.189$ ). For these classifications we therefore found no evidence of interactions between species, either facilitative or competitive.

### Community Assembly Patterns

Nose samples deviated from the neutral assembly model for samples from summer ( $X^2 = 12.36$ ,  $df = 2$ ,  $p = 0.015$ ), individuals aged 5-17 years ( $X^2 = 11.24$ ,  $df = 2$ ,  $p = 0.024$ ), 18-64 years ( $X^2 = 28.78$ ,  $df = 2$ ,  $p < 0.001$ ) and  $\geq 65$  years ( $X^2 = 29.69$ ,  $df = 2$ ,  $p < 0.001$ ) (Age models shown in Figure 3). Within these classifications, rarer species had lower abundances than neutral expectation, indicative of niche characteristics of a dominance hierarchy (cf. Fig. 3a for the 0-4 year age classification with no difference from neutral, and Fig. 3b-d for the other age classifications all deviating from neutral). No detectable deviations from neutral assembly were apparent for samples from winter ( $X^2 = 3.05$ ,  $df = 2$ ,  $p = 0.549$ ), individuals aged 0-4 years ( $X^2 = 1.68$ ,  $df = 2$ ,  $p = 0.794$ ), individuals with recent RTI ( $X^2 = 7.39$ ,  $df = 2$ ,  $p = 0.117$ ) and individuals without recent RTI ( $X^2 = 9.03$ ,  $df = 2$ ,  $p = 0.05$ ).

## DISCUSSION

Understanding the prevalence and community distribution of bacterial and viral pathogens, particularly where carriage is a prerequisite for progression to disease, is a key undertaking in the on-going development of interventional responses to changing epidemiology. Here we set out to determine the distribution of pathobiont bacterial and viral species commonly associated with respiratory tract infections using culture-independent approaches and ecological models of community assemblage as well as co-occurrence. Using multivariate analysis of microbial prevalence, we demonstrated that age affects carriage for a number of bacteria and viruses, with young participants experiencing higher carriage of *S. pneumoniae*, *H. influenzae*, *M. catarrhalis* and rhinovirus/enterovirus. Recent RTI was also associated with increased prevalence of *S. pneumoniae*, rhinovirus/enterovirus and coronavirus. Epidemiological typing revealed few vaccine type *S. pneumoniae* and *H. influenzae* within

this set of isolates, which is indicative of the effectiveness of current PCV-13 and Hib vaccines in targeting specific serotypes as well as the effect of herd immunity in protecting individuals within age groups unlikely to have received these vaccines (41, 42). The high proportion of non-vaccine type (NVT) *S. pneumoniae* reflects the process of serotype replacement as a result of vaccination, although it is not thought to lead to future large increases in the incidence of invasive disease caused by NVTs (43). The identification of increasingly common NVT serotypes 6C, 11A/D and 23B and non-typeable *H. influenzae* provides important information for the epidemiological assessment of vaccines and for the development of future vaccines.

The lack of ubiquity of potential vaccine targets within the isolates collected in this study may highlight certain difficulties in the development of protein-based vaccines against these species (32). This knowledge will help to inform the targeting of future treatment and prevention strategies, including the development of new vaccines against *S. aureus* and *M. catarrhalis* (44, 45) and vaccines targeting greater numbers of *S. pneumoniae* serotypes (46). High intra-species diversity within these respiratory isolates is potentially a result of the pressures imposed by current vaccination strategies (47) as well as the high levels of recombination in *S. pneumoniae* (48). Few antibiotic resistance genes within this group of isolates reflects the low and stable levels of antibiotic-resistant respiratory species in the UK (49). Continued monitoring of levels of antibiotic resistance within respiratory isolates within the UK is essential for clinical treatment of infection (49).

The ecological analyses showed a relatively high degree of ordering (nesting) in the incidence of species amongst the respiratory tract communities of older participants, those without recent RTI, and amongst winter samples, in contrast to a more disordered turnover amongst young children, those with recent RTI, and summer samples. Lower levels of nestedness in these individuals may be related to the reduced isolation of species, allowing greater transmission, and the role of the immune system in maintaining species diversity. Furthermore, communities within the respiratory tract were found to have a clumped frequency distribution of species (either few or many species present) amongst young individuals, consistent with facilitative relationships occurring between the species within the upper respiratory tract. Facilitative relationships may enhance transmission opportunities and survival of species carried in the respiratory tract. Respiratory species abundances were more consistent with a dominance hierarchy than a neutral assembly model for older participants, those without recent RTI and summer samples, but not for other classifications. There are likely to be both niche and neutral processes involved in microbial community formation in the upper respiratory tract, with speciation and the environment playing key

roles (50, 51). The evidence that niche processes tend towards dominance hierarchies rather than resource segregation suggests that the microbial species compete for similar resources; moreover, communities may have fragile coexistence equilibria, with some dominant competitors functioning as keystone species capable of triggering cascades of extinctions (15). Further development of ecological methodologies will allow predictions of microbial variation as a result of infection, season and increasing age. Monitoring communities in states of health and disease is important for generating knowledge that will be useful in clinical practice (52).

Inferences from this study should be understood in the context of a number of methodological constraints. The characteristics of the 21% ( $n = 2103$ ) of responders may differ from those of non-responders. The study focused on a limited number of species, and did not account for species such as Group A *Streptococci* which may play an important role within this sample site as well as other non-culturable species. Furthermore, the swab samples used were not optimised for the detection of viral species. The questionnaire used for the study was designed for simplicity and ease of completion. This meant that information was not collected on a number of important factors, including smoking status, siblings, and attendance at nursery or day-care facilities. Samples assessed by real-time PCR were limited by the number of species that were targeted as well as the non-random nature of their selection. The ecological models ignore acquired or specific host immunity, which can influence the ecology of the respiratory tract (53). The combined analysis of bacterial and viral species, and the combined analysis of all RTI, may also be simplistic. Neutral theory assumes no variation in the total number of species within a community whilst niche theory ignores dispersal (54, 55). The quantification of nestedness provides only a single measure of matrix temperature per classification of samples, ruling out significance testing.

## Conclusion

The study has provided key epidemiological information on spatio-temporal patterns and trends amongst circulating bacterial species and types. Specifically we have demonstrated for the first time the impact of age and season on the distribution of both bacterial and viral pathogens, and shown clear inter-genus associations of co-carriage between *S. pneumoniae*, *M. catarrhalis* and *H. influenzae*. These epidemiological findings on circulating respiratory tract pathogens, that includes the distribution of strains and serotypes amongst different age demographics, underpins our ability to assess risk in these populations and therefore ultimately informs the development of new treatment and prevention strategies as well as targeted and effective antibiotics and vaccination policies.

## **Acknowledgments**

The authors thank the Bupa Foundation for providing the funding to SCC in order to undertake the study and the Rosetrees Trust for their funding contribution for ALC's PhD studentship. The authors also thank Shabana Hussain, Beverly Simms, Christine Tumman and Karen Cox for technical assistance throughout the study. The authors thank the Southampton NIHR Wellcome Trust Clinical Research Facility for part-funding SNF. The authors also thank the NIHR Comprehensive Local Research Network (NIHR CLRN), Solent NHS Trust, the South West (East Hub) Primary Care Research Network (PCRN), and Danvers International for their support.

## **Competing Financial Interests**

SNF receives support from the National Institute for Health Research funding via the Southampton NIHR Wellcome Trust Clinical Research Facility and the Southampton NIHR Respiratory Biomedical Research Unit. SNF, SCC and JMJ act as principal investigator for clinical trials and other studies conducted on behalf of University Hospital Southampton NHS Foundation Trust/University of Southampton that are sponsored by vaccine manufacturers but receives no personal payments from them. SNF, JMJ and SCC have participated in advisory boards for vaccine manufacturers but receive no personal payments for this work. SNF, SCC and JMJ have received financial assistance from vaccine manufacturers to attend conferences. All grants and honoraria are paid into accounts within the respective NHS Trusts or Universities, or to independent charities. DWC employed for 18 months on a GSK funded research project in 2014/15. All other authors have no conflicts of interest.

## **Author Contributions**

ALC contributed to study set-up, data collection, data analysis and writing of the manuscript. CPD and ARK contributed to study-design, data analysis and manuscript preparation. RNW contributed to study set-up, data collection. NB and RA contributed to data collection. DM contributed to study design, set-up and data collection. RA and AT contributed to study design and data collection. SNF, JMJ, HMY, PJR, MAM and MVM contributed to study design, data analysis, proofreading of manuscript. DWC contributed to manuscript writing and proofreading. SCC contributed to study design, data collection, data analysis, and writing and proofreading of the manuscript.

## TABLES

**Table 1.** Characteristics of study participants

	<b>Summer</b> ( <i>N</i> = 1,260)	<b>Winter</b> ( <i>N</i> = 843)	<b>Total</b> ( <i>N</i> = 2,103)
<b>Age (years), <i>n</i> (%)</b>			
Mean (SD)	37.4 (29.1)	40.7 (29.8)	38.7 (29.4)
0-4	329 (26.1)	168 (19.9)	497 (23.6)
5-17	137 (10.9)	127 (15.1)	264 (12.6)
18-64	465 (36.9)	286 (33.9)	751 (35.7)
65+	311 (24.7)	244 (28.9)	555 (26.4)
Missing	18 (1.4)	18 (2.1)	36 (1.7)
<b>Recent antibiotics use, <i>n</i> (%)</b>			
Yes	101 (8.0)	79 (9.4)	180 (8.6)
No	1124 (89.2)	743 (88.1)	1867 (88.8)
Missing	35 (2.8)	21 (2.5)	56 (2.7)
<b>Recent RTI, <i>n</i> (%)</b>			
Yes	365 (29.0)	367 (43.5)	732 (34.8)
No	860 (68.3)	453 (53.7)	1313 (62.4)
Missing	35 (2.8)	23 (2.7)	58 (2.8)
<b>Vaccination Status, <i>n</i> (%)</b>			
Up-to-date	1022 (81.1)	680 (80.7)	1702 (80.9)
Not up-to-date	40 (3.2)	33 (3.9)	73 (3.5)
Missing	198 (15.7)	130 (15.4)	328 (15.6)

SD = standard deviation, RTI = respiratory tract infection, recent is having occurred with one month of swab sample collection (indicated on the questionnaire).

**Table 2.** Carriage of bacterial species in nose swabs detected by culture, stratified by age group

<b>Carriage, n (%)</b>	<b>Age Group (years)</b>	<b>Summer (N = 1,260)</b>	<b>Winter (N = 843)</b>	<b>Total (N = 2,103)</b>
<i>S. pneumoniae</i>	All	138 (11.0)	74 (8.8)	212 (10.1)
	0-4	108 (32.8)	51 (30.4)	159 (32.0)
	5-17	18 (13.1)	18 (14.4)	36 (13.7)
	18-64	5 (1.1)	2 (0.7)	7 (0.9)
	≥65	6 (2.0)	2 (0.8)	8 (1.5)
<i>M. catarrhalis</i>	All	31 (2.5)	64 (7.6)	95 (4.5)
	0-4	19 (5.8)	31 (18.5)	50 (10.1)
	5-17	1 (0.7)	15 (12.0)	16 (6.1)
	18-64	7 (1.5)	6 (2.1)	13 (1.7)
	≥65	4 (1.3)	9 (3.7)	13 (2.4)
<i>H. influenzae</i>	All	34 (2.7)	45 (5.4)	79 (3.8)
	0-4	24 (7.3)	31 (18.5)	55 (11.1)
	5-17	7 (5.1)	10 (8.0)	17 (6.5)
	18-64	1 (0.2)	2 (0.7)	3 (0.4)
	≥65	2 (0.7)	0 (0.0)	2 (0.4)
<i>S. aureus</i>	All	269 (21.5)	149 (17.7)	418 (19.9)
	0-4	32 (9.7)	16 (9.5)	48 (9.7)
	5-17	48 (35.0)	21 (16.8)	69 (26.3)
	18-64	115 (24.8)	68 (23.8)	183 (24.4)
	≥65	71 (23.2)	41 (16.8)	112 (20.4)
<i>P. aeruginosa</i>	All	20 (1.6)	7 (0.8)	27 (1.3)
	0-4	9 (2.7)	2 (1.2)	11 (2.2)
	5-17	1 (0.7)	1 (0.8)	2 (0.8)
	18-64	6 (1.3)	1 (0.3)	7 (0.9)
	≥65	3 (1.0)	2 (0.8)	5 (0.9)
<i>N. meningitidis</i>	All	0 (0.0)	1 (0.1)	0 (0.0)
	0-4	0 (0.0)	0 (0.0)	0 (0.0)
	5-17	0 (0.0)	0 (0.0)	0 (0.0)
	18-64	0 (0.0)	1 (0.3)	1 (0.1)
	≥65	0 (0.0)	0 (0.0)	0 (0.0)

**Table 3.** Carriage of bacterial in nose swabs detected by real-time PCR, stratified by age group

<b>Carriage, n (%)</b>	<b>Age Group (years)</b>	<b>Summer (N = 115)</b>	<b>Winter (N = 75)</b>	<b>Total (N = 190)</b>
<i>S. pneumoniae</i>	All	40 (34.8)	24 (32.0)	64 (33.7)
	0-4	21 (52.5)	11 (55.0)	32 (53.3)
	5-17	11 (35.5)	8 (42.1)	19 (38.0)
	18-64	2 (9.1)	3 (16.7)	5 (12.5)
	≥65	6 (27.3)	2 (11.1)	8 (20.0)
<i>M. catarrhalis</i>	All	28 (24.3)	17 (22.7)	45 (23.7)
	0-4	18 (45.0)	8 (40.0)	26 (43.4)
	5-17	6 (19.4)	4 (21.1)	10 (20.0)
	18-64	1 (4.5)	3 (16.7)	4 (10.0)
	≥65	3 (13.6)	2 (11.1)	5 (12.5)
<i>H. influenzae</i>	All	34 (29.6)	33 (44.0)	67 (35.3)
	0-4	19 (47.5)	15 (75.0)	34 (56.7)
	5-17	9 (29.0)	9 (47.4)	18 (36.0)
	18-64	2 (9.1)	6 (33.3)	8 (20.0)
	≥65	4 (18.2)	3 (16.7)	7 (17.5)
<i>S. aureus</i>	All	50 (43.5)	33 (44.0)	83 (43.7)
	0-4	10 (25.0)	4 (20.0)	14 (23.3)
	5-17	17 (54.8)	8 (42.1)	25 (50.0)
	18-64	12 (54.5)	12 (66.7)	24 (60.0)
	≥65	11 (50.5)	9 (50.5)	20 (50.0)
<i>P. aeruginosa</i>	All	54 (47.0)	15 (20.0)	69 (36.3)
	0-4	17 (42.5)	1 (5.0)	18 (30.0)
	5-17	16 (51.6)	4 (21.1)	20 (40.0)
	18-64	12 (54.5)	7 (38.9)	19 (47.5)
	≥65	9 (40.9)	3 (16.7)	12 (30.0)
<i>N. meningitidis</i>	All	0 (0.0)	0 (0.0)	0 (0.0)
	0-4	0 (0.0)	0 (0.0)	0 (0.0)
	5-17	0 (0.0)	0 (0.0)	0 (0.0)
	18-64	0 (0.0)	0 (0.0)	0 (0.0)
	≥65	0 (0.0)	0 (0.0)	0 (0.0)

**Table 4.** Presence of viral species in nose swabs detected by real-time PCR

<b>Presence, <i>n</i> (%)</b>	<b>Summer (<i>N</i> = 115)</b>	<b>Winter (<i>N</i> = 75)</b>	<b>Total (<i>N</i> = 190)</b>
Influenza virus A	0 (0.0)	2 (2.7)	2 (1.1)
Influenza virus B	0 (0.0)	0 (0.0)	0 (0.0)
Respiratory syncytial virus	7 (6.1)	6 (8.0)	13 (6.8)
Adenovirus	5 (4.3)	1 (1.3)	6 (3.2)
Rhinovirus/enterovirus	19 (16.5)	6 (8.0)	25 (13.2)
Coronavirus	12 (10.4)	7 (9.3)	19 (10.0)
Parainfluenza virus	0 (0.0)	1 (1.3)	1 (0.5)
Metapneumovirus	0 (0.0)	0 (0.0)	0 (0.0)

**Table 5.** Unadjusted and multivariate-adjusted logistic regression models for nasal carriage of bacterial species

Covariate OR (95% CI) P-value	<i>S. pneumoniae</i>		<i>H. influenzae</i>		<i>M. catarrhalis</i>		<i>P. aeruginosa</i>		<i>S. aureus</i>	
	Unadjusted	Multivariate-adjusted	Unadjusted	Multivariate-adjusted	Unadjusted	Multivariate-adjusted	Unadjusted	Multivariate-adjusted	Unadjusted	Multivariate-adjusted
Age (years)	0.934 (0.932, 0.955) <0.001	0.943 (0.932, 0.955) <0.001	0.934 (0.916, 0.953) <0.001	0.948 (0.930, 0.966) <0.001	0.975 (0.967, 0.983) <0.001	0.981 (0.972, 0.990) <0.001	0.994 (0.981, 1.008) 0.384	-	1.005 (1.002, 1.009) 0.006	-
Recent RTI	4.334 (3.192, 5.885) <0.001	2.071 (1.448, 2.963) <0.001	4.676 (2.824, 7.744) <0.001	-	2.551 (1.670, 3.898) <0.001	-	1.199 (0.536, 2.682) 0.659	-	0.838 (0.665, 1.056) 0.134	-
Recent antibiotics use	0.790 (0.456, 1.368) 0.4000	-	0.887 (0.380, 2.071) 0.781	-	0.586 (0.235, 1.462) 0.252	-	0.449 (0.060, 3.342) 0.434	-	0.947 (0.641, 1.400) 0.786	-
Vaccinations up-to-date	3.104 (0.968, 9.951) 0.057	-	3.287 (0.451, 23.979) 0.241	-	1.217 (0.375, 3.944) 0.744	-	0.947 (0.126, 7.122) 0.958	-	0.835 (0.474, 1.474) 0.535	-
Geo location 2	3.421 (1.863, 6.280) <0.001	3.055 (1.480, 6.303) 0.003	1.062 (0.297, 3.794) 0.926	-	2.758 (1.126, 6.753) 0.026	-	0.832 (0.182, 3.811) 0.813	-	0.671 (0.366, 1.232) 0.198	-
Geo location 3	2.495 (1.578, 3.944) <0.001	2.224 (1.322, 3.741) 0.003	1.718 (0.831, 3.551) 0.144	-	1.872 (0.939, 3.731) 0.075	-	0.771 (0.296, 2.007) 0.594	-	1.318 (0.951, 1.826) 0.098	-
Geo location 4	1.590 (1.036, 2.440) 0.034	1.621 (0.998, 2.633) 0.051	1.611 (0.846, 3.067) 0.147	-	1.754 (0.948, 3.246) 0.074	-	0.204 (0.065, 0.645) 0.007	-	1.183 (0.893, 1.569) 0.242	-
Geo location 5	1.063 (0.430, 2.628) 0.894	1.052 (0.393, 2.817) 0.920	1.272 (0.355, 4.557) 0.712	-	1.179 (0.332, 4.188) 0.800	-	0.492 (0.063, 3.857) 0.499	-	0.823 (0.445, 1.523) 0.536	-
Geo location 6	1.407 (0.687, 2.884) 0.351	1.589 (0.709, 3.560) 0.261	1.195 (0.383, 3.731) 0.759	-	1.108 (0.358, 3.425) 0.859	-	-	-	0.985 (0.593, 1.635) 0.953	-
Winter	0.780 (0.580, 1.050) 0.101	0.550 (0.376, 0.804) 0.002	2.029 (1.288, 3.195) 0.002	2.495 (1.529, 4.071) <0.001	3.250 (2.097, 5.037) <0.001	3.503 (2.223, 5.519) <0.001	0.518 (0.218, 1.230) 0.136	-	0.788 (0.631, 0.985) 0.036	-
<i>S. pneumoniae</i>	-	-	11.641 (7.288, 18.595) <0.001	4.602 (2.755, 7.688) <0.001	6.227 (2.914, 7.293) <0.001	2.674 (1.580, 4.525) <0.001	2.043 (0.766, 5.453) 0.154	-	0.259 (0.149, 0.450) <0.001	0.282 (0.162, 0.492) <0.001
<i>H. influenzae</i>	11.641 (7.288, 18.595) <0.001	4.195 (2.460, 7.154) <0.001	-	-	6.227 (3.441, 11.268) <0.001	-	-	-	0.320 (0.138, 0.741) 0.008	-
<i>M. catarrhalis</i>	4.610 (2.914, 7.293)	1.974 (1.130, 3.449)	6.227 (3.441, 11.268)	-	-	-	0.808 (0.108, 6.016)	-	0.213 (0.086, 0.529)	0.258 (0.103, 0.642)

	<0.001	0.017	<0.001				0.835		0.001	0.004
	0.259	0.354	0.320		0.213	0.238	1.411			
<i>S. aureus</i>	(0.149, 0.450)	(0.193, 0.650)	(0.138, 0.741)	-	(0.086, 0.529)	(0.086, 0.660)	(0.593, 3.360)	-	-	-
	<0.001	0.001	0.008		0.001	0.006	0.437			
	0.546				0.808				1.411	
<i>P. aeruginosa</i>	(0.169, 1.767)	-	-	-	(0.108, 6.016)	-	-	-	(0.593, 3.360)	-
	0.313				0.835				0.437	

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geo location = distinct graphical areas. Reference categories are 'No' for recent RTI, recent antibiotics use, vaccinations up-to-date and presence of other bacterial species, Geo location 1 for geographical location and summer for season. Final model fit: *S. pneumoniae* pseudo-R<sup>2</sup> = 0.396 and area under curve (AUC) = 0.885 (95% CI 0.862, 0.908); *H. influenzae* pseudo-R<sup>2</sup> = 0.295 and AUC = 0.874 (0.838, 0.910); *S. aureus* pseudo-R<sup>2</sup> = 0.064 and AUC = 0.557 (0.528, 0.586); *M. catarrhalis* pseudo-R<sup>2</sup> = 0.155 and area under curve (AUC) = 0.772 (95% CI 0.721, 0.824); *P. aeruginosa* no final model.

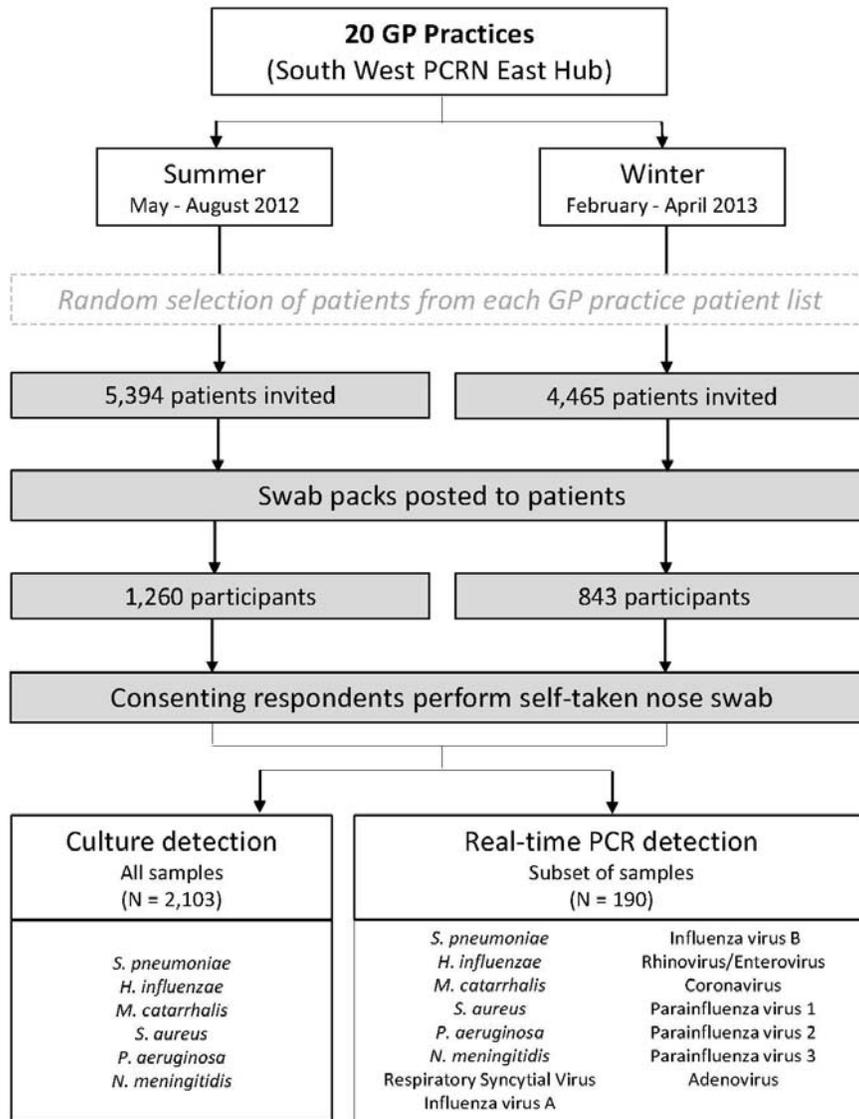
**Table 6.** Summary Statistics of Viral Prevalence in Nose Swabs

	N	RSV			ADV			RV/EV			COV		
		Yes n (%)	No n (%)	P	Yes n (%)	No n (%)	P	Yes n (%)	No n (%)	P	Yes n (%)	No n (%)	P
<b>Age (years)</b>													
Mean (SD)	-	26.2 (28.1)	27.9 (29.0)	0.836#	27.2 (39.2)	27.8 (28.6)	0.450#	10.9 (20.9)	30.4 (29.1)	<0.001#	16.6 (25.0)	29.1 (29.1)	0.100#
<b>Recent RTI</b>													
Yes	77	4 (5.2)	73 (94.8)	0.566	3 (3.9)	74 (96.1)	0.688*	15 (19.5)	62 (80.5)	0.048	13 (16.9)	64 (83.1)	0.013
No	113	9 (8.0)	104 (92.0)		3 (2.7)	110 (97.3)		10 (8.8)	103 (91.2)		6 (5.3)	107 (94.7)	
<b>Recent Antibiotics</b>													
Yes	19	1 (5.3)	18 (94.7)	1.000*	0 (0.0)	19 (100.0)	1.000*	0 (0.0)	19 (100.0)	0.082*	1 (5.3)	18 (94.7)	0.699*
No	170	12 (7.1)	158 (92.9)		6 (3.5)	164 (96.5)		25 (14.7)	145 (985.3)		18 (10.6)	152 (89.4)	
<b>Vaccinations up-to-date</b>													
Yes	163	10 (6.1)	153 (93.9)	0.241*	6 (3.7)	157 (96.3)	1.000*	22 (13.5)	141 (86.5)	0.450*	15 (9.2)	148 (90.8)	1.000*
No	4	1 (25.0)	3 (75.0)		0 (0.0)	4 (100.0)		1 (25.0)	3 (75.0)		0 (0.0)	4 (100.0)	
<b>Geographical Location</b>													
1	56	8 (14.3)	48 (85.7)	0.012*	1 (1.8)	55 (98.2)	0.378*	9 (16.1)	47 (83.9)	0.431*	7 (12.5)	49 (87.5)	0.038*
2	10	1 (10.0)	9 (90.0)		1 (10.0)	9 (90.0)		0 (0.0)	10 (100.0)		0 (0.0)	10 (100.0)	
3	37	3 (8.1)	34 (91.9)		0 (0.0)	37 (100.0)		7 (18.9)	30 (81.1)		8 (21.6)	29 (78.4)	
4	68	0 (0.0)	68 (100.0)		4 (5.9)	64 (94.1)		7 (10.3)	61 (89.7)		2 (2.9)	66 (97.1)	
5	9	1 (11.1)	8 (88.9)		0 (0.0)	9 (100.0)		0 (0.0)	9 (100.0)		1 (11.1)	8 (88.9)	
6	10	0 (0.0)	10 (100.0)		0 (0.0)	10 (100.0)		2 (20.0)	8 (80.0)		1 (10.0)	9 (90.0)	
<b>Season</b>													
Summer	115	7 (6.1)	108 (93.9)	0.770	5 (4.3)	110 (95.7)	0.406*	19 (16.5)	96 (83.5)	0.124	12 (10.4)	103 (89.6)	1.000
Winter	75	6 (8.0)	69 (92.0)		1 (1.3)	74 (98.7)		6 (8.0)	69 (92.0)		7 (9.3)	68 (90.7)	
<b>S. pneumoniae</b>													
Yes	64	5 (7.8)	59 (92.2)	0.764*	5 (7.8)	59 (92.2)	0.017*	16 (25.0)	48 (75.0)	0.001	12 (18.8)	52 (81.3)	0.006
No	126	8 (6.3)	118 (93.7)		1 (0.8)	125 (99.2)		9 (7.1)	117 (92.9)		7 (5.6)	119 (94.4)	
<b>H. influenzae</b>													
Yes	67	7 (10.4)	60 (89.6)	0.227*	5 (7.5)	62 (92.5)	0.021*	13 (19.4)	54 (80.6)	0.073	9 (13.4)	58 (86.6)	0.312
No	123	6 (4.9)	117 (95.1)		1 (0.8)	122 (99.2)		12 (9.8)	111 (90.2)		10 (8.1)	113 (91.9)	
<b>M. catarrhalis</b>													
Yes	45	5 (11.1)	40 (88.9)	0.193*	4 (8.9)	41 (91.1)	0.029*	13 (28.9)	32 (71.1)	0.001	8 (17.8)	37 (82.2)	0.083*
No	8	8 (5.5)	137 (94.5)		2 (1.4)	143 (98.6)		12 (8.3)	133 (91.7)		11 (7.6)	134 (92.4)	
<b>S. aureus</b>													
Yes	83	6 (7.2)	77 (92.8)	1.000	1 (1.2)	82 (98.8)	0.234*	8 (9.6)	75 (90.4)	0.280	16 (19.3)	67 (80.7)	<0.001
No	107	7 (6.5)	100 (93.5)		5 (4.7)	102 (95.3)		17 (15.9)	90 (84.1)		3 (2.8)	104 (97.2)	
<b>P. aeruginosa</b>													
Yes	69	2 (2.9)	67 (97.1)	0.139*	1 (1.4)	68 (98.6)	0.420*	13 (18.8)	56 (81.2)	0.117	7 (10.1)	62 (89.9)	1.000
No	121	11 (9.1)	110 (90.9)		5 (4.1)	116 (95.9)		12 (9.9)	109 (90.1)		12 (9.9)	109 (90.1)	
<b>ADV</b>													

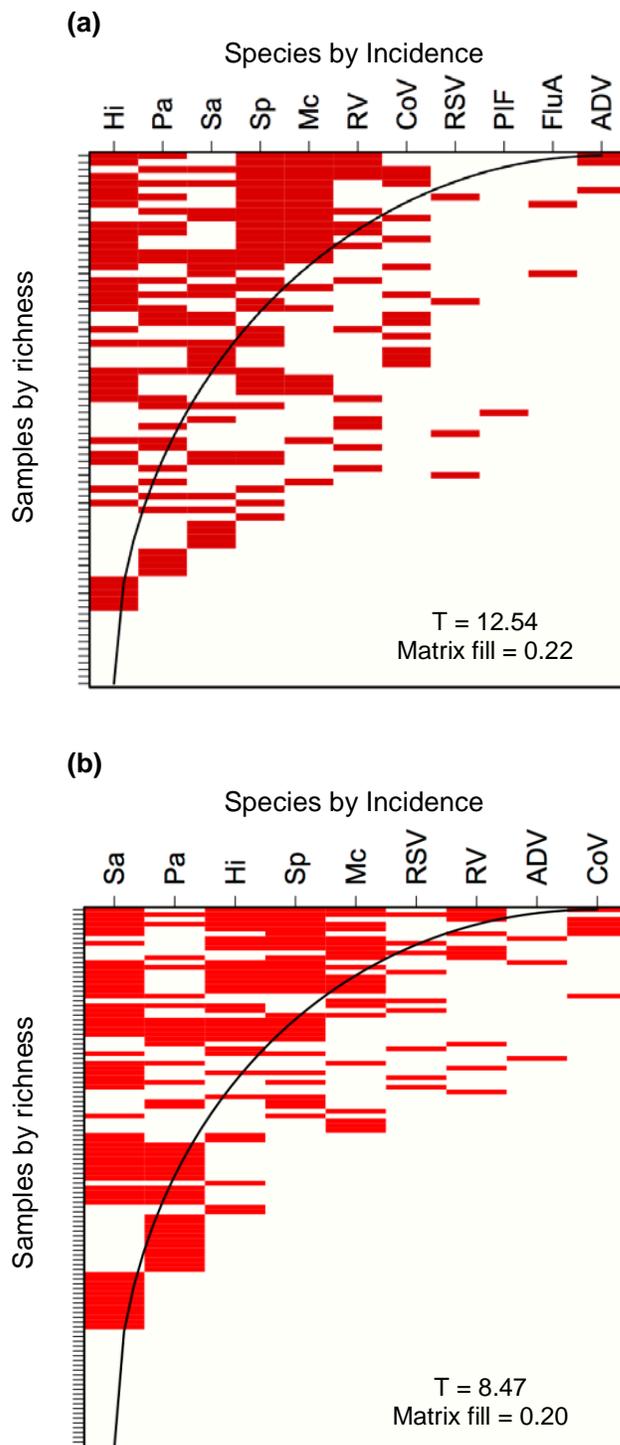
Yes	6	0 (0.0)	6 (100.0)	1.000*	0 (0.0)	13 (100.0)	1.000*	2 (15.4)	11 (84.6)	0.179*	0 (0.0)	13 (100.0)	0.369*
No	184	13 (7.1)	171 (92.9)		6 (3.4)	171 (96.6)		23 (13.0)	154 (87.0)		19 (10.7)	158 (89.3)	
<b>RV/EV</b>													
Yes	25	2 (8.0)	23 (92.0)	0.682*	2 (8.0)	23 (92.0)	0.179*	2 (33.3)	4 (66.7)	0.682*	0 (0.0)	6 (100.0)	1.000*
No	165	11 (6.7)	154 (93.3)		4 (2.4)	161 (97.6)		23 (12.5)	161 (87.5)		19 (10.3)	165 (89.7)	
<b>COV</b>													
Yes	19	0 (0.0)	19 (100.0)	0.369*	0 (0.0)	19 (100.0)	1.000*	5 (26.3)	14 (73.7)	0.142*	5 (20.0)	20 (80.0)	0.142*
No	171	13 (7.6)	158 (92.4)		6 (3.5)	165 (96.5)		20 (11.7)	151 (88.3)		14 (8.5)	151 (91.5)	

SD = standard deviation; RTI = Respiratory tract infection; RSV = Respiratory Syncytial Virus; ADV = adenovirus; RV/EV = rhinovirus/enterovirus; COV = coronavirus. Geographical areas are listed in methods section. *P*-values are 2-tailed Chi-squared *P*-values unless otherwise indicated, \*Fisher's Exact Test *P*-value, #Mann Whitney Test *P*-value.

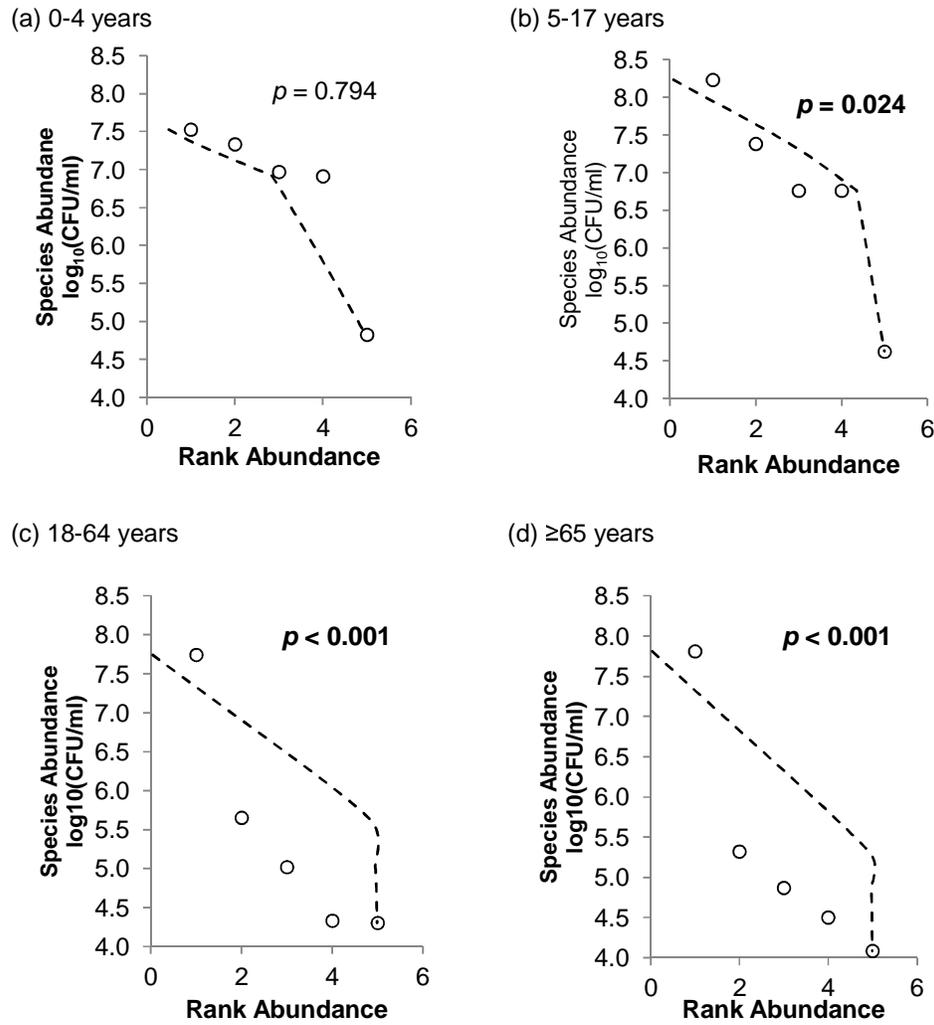
## FIGURES



**Figure 1.** Study design. GP = general practitioner



**Figure 2.** Incidence matrices of nose swab samples detected by real-time PCR in participants (a) with recent RTI (b) with no recent RTI. Red cells show presences, white show absences, curve delineates perfect fill, T = nestedness temperature. Pa = *P.aeruginosa*, Sa = *S. aureus*, Sp = *S. pneumoniae*, Hi = *H. influenzae*, Mc = *M. catarrhalis*, RV = rhinovirus/enterovirus, CoV = coronavirus, RSV = respiratory syncytial virus, ADV = adenovirus, PIF = parainfluenza virus, FluA = influenza virus A.



**Figure 3.** Species abundance curves (CFU/ml) for nose swabs by age group. Dots show species abundance, dashed line is MacArthur's Broken Stick model of neutral community assembly; deviations with  $\chi^2$   $P$ -value  $< 0.05$  in bold. CFU = colony-forming unit.

## REFERENCES

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2095-128.
2. NICE. Respiratory tract infections - antibiotic prescribing costing report. London, UK: NHS; 2008.
3. Weiss-Salz I, Yagupsky P. Asymptomatic Carriage of Respiratory Pathogens: "The Wolf shall Dwell with the Lamb...and a Little Child shall Lead them" (Isaiah 11: 6). *Open Infect Dis J*. 2010;4(1):11-5.
4. Bosch AA, Biesbroek G, Trzcinski K, Sanders EA, Bogaert D. Viral and bacterial interactions in the upper respiratory tract. *PLoS pathogens*. 2013;9(1):e1003057.
5. Hubbell SP. *The Unified Theory of Biodiversity and Biogeography*. Princeton, USA: Princeton University Press; 2001.
6. Tuttle MS, Mostow E, Mukherjee P, Hu FZ, Melton-Kreft R, Ehrlich GD, et al. Characterization of bacterial communities in venous insufficiency wounds by use of conventional culture and molecular diagnostic methods. *Journal of clinical microbiology*. 2011;49(11):3812-9.
7. Dollhopf SL, Hashsham SA, Tiedje JM. Interpreting 16S rDNA T-RFLP Data: Application of Self-Organizing Maps and Principal Component Analysis to Describe Community Dynamics and Convergence. *Microbial ecology*. 2001;42(4):495-505.
8. Mackay IM. Real-time PCR in the microbiology laboratory. *Clin Microbiol Infect*. 2004;10(3):190-212.
9. Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *Journal of clinical microbiology*. 2006;44(1):124-31.
10. LaClaire LL, Tondella MLC, Beall DS, Noble CA, Raghunathan PL, Rosenstein NE, et al. Identification of *Haemophilus influenzae* serotypes by standard slide agglutination serotyping and PCR-based capsule typing. *Journal of clinical microbiology*. 2003;41(1):393-6.
11. Mothershed EA, Sacchi CT, Whitney AM, Barnett GA, Ajello GW, Schmink S, et al. Use of real-time PCR to resolve slide agglutination discrepancies in serogroup identification of *Neisseria meningitidis*. *Journal of clinical microbiology*. 2004;42(1):320-8.
12. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(6):3140-5.
13. Sheppard SK, Jolley KA, Maiden MC. A Gene-By-Gene Approach to Bacterial Population Genomics: Whole Genome MLST of *Campylobacter*. *Genes (Basel)*. 2012;3(2):261-77.
14. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*. 2012;158(Pt 4):1005-15.
15. Doncaster CP. Ecological equivalence: a realistic assumption for niche theory as a testable alternative to neutral theory. *PloS one*. 2009;4(10):e7460.
16. Elton CS. *Animal Ecology*. Huxley JS, editor. London, UK: Sidgwick & Jackson Limited; 1927. 207 p.
17. Schoener TW. Resource partitioning in ecological communities. *Science*. 1974;185(4145):27-39.
18. McGill BJ. A test of the unified neutral theory of biodiversity. *Nature*. 2003;422(6934):881-5.

19. Southwood TRE. The Number of Species of Insect Associated with Various Trees. *J Anim Ecol.* 1961;30(1):1-8.
20. Verbruggen E, Van Der Heijden MG, Weedon JT, Kowalchuk GA, Røling WF. Community assembly, species richness and nestedness of arbuscular mycorrhizal fungi in agricultural soils. *Molecular ecology.* 2012;21(10):2341-53.
21. Rohde K, Worthen WB, Heap M, Huguency B, Guegan JF. Nestedness in assemblages of metazoan ecto- and endoparasites of marine fish. *Int J Parasitol.* 1998;28(4):543-9.
22. Coughtrie AL, Whittaker RN, Begum N, Anderson R, Tuck A, Faust SN, et al. Evaluation of swabbing methods for estimating the prevalence of bacterial carriage in the upper respiratory tract: a cross sectional study. *BMJ open.* 2014;4(10):e005341.
23. Falla TJ, Crook DW, Brophy LN, Maskell D, Kroll JS, Moxon ER. PCR for capsular typing of *Haemophilus influenzae*. *Journal of clinical microbiology.* 1994;32(10):2382-6.
24. Davis GS, Sandstedt SA, Patel M, Marrs CF, Gilsdorf JR. Use of *bexB* to detect the capsule locus in *Haemophilus influenzae*. *Journal of clinical microbiology.* 2011;49(7):2594-601.
25. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome research.* 2008;18(5):821-9.
26. Inouye M, Dashnow H, Pope B, Holt K. SRST2 - Short Read Sequence Typing for Bacterial Pathogens Melbourne, Australia2014 [Available from: <http://katholt.github.io/srst2/>].
27. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *The Journal of antimicrobial chemotherapy.* 2012;67(11):2640-4.
28. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy.* 2013;57(7):3348-57.
29. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic acids research.* 2005;33(Database issue):D325-8.
30. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic acids research.* 2012;40(Database issue):D641-5.
31. Brown JS, Ogunniyi AD, Woodrow MC, Holden DW, Paton JC. Immunization with components of two iron uptake ABC transporters protects mice against systemic *Streptococcus pneumoniae* infection. *Infection and immunity.* 2001;69(11):6702-6.
32. Tai SS. *Streptococcus pneumoniae* protein vaccine candidates: properties, activities and animal studies. *Critical reviews in microbiology.* 2006;32(3):139-53.
33. Ogunniyi AD, Grabowicz M, Briles DE, Cook J, Paton JC. Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. *Infection and immunity.* 2007;75(1):350-7.
34. Bogaert D, Hermans PW, Adrian PV, Rumke HC, de Groot R. Pneumococcal vaccines: an update on current strategies. *Vaccine.* 2004;22(17-18):2209-20.
35. Poolman JT, Bakaletz LO, Cripps A, Denoel PA, Forsgren A, Kyd J, et al. Developing a nontypeable *Haemophilus influenzae* (NTHi) vaccine. *Vaccine.* 2001;19(S1):S108-15.
36. Forsgren A, Riesbeck K, Janson H. Protein D of *Haemophilus influenzae*: a protective nontypeable H. influenzae antigen and a carrier for pneumococcal conjugate vaccines. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2008;46(5):726-31.
37. Nagai K, Shibasaki Y, Hasegawa K, Davies TA, Jacobs MR, Ubukata K, et al. Evaluation of PCR primers to screen for *Streptococcus pneumoniae* isolates and beta-lactam resistance, and to detect common macrolide resistance determinants. *The Journal of antimicrobial chemotherapy.* 2001;48(6):915-8.
38. Binks MJ, Temple B, Kirkham LA, Wiertsema SP, Dunne EM, Richmond PC, et al. Molecular surveillance of true nontypeable *Haemophilus influenzae*: an evaluation of PCR screening assays. *PloS one.* 2012;7(3):e34083.

39. MacArthur RH. On the Relative Abundance of Bird Species. *Proceedings of the National Academy of Sciences of the United States of America*. 1957;43(3):293-5.
40. King CE. Relative Abundance of Species and MacArthur's Model. *Ecology*. 1964;45(4):716-27.
41. Tocheva AS, Jefferies JM, Rubery H, Bennett J, Afimeke G, Garland J, et al. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of *Streptococcus pneumoniae* in young children. *Vaccine*. 2011;29(26):4400-4.
42. Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. *Lancet Infect Dis*. 2011;11(10):760-8.
43. Hanage WP, Finkelstein JA, Huang SS, Pelton SI, Stevenson AE, Kleinman K, et al. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. *Epidemics*. 2010;2(2):80-4.
44. Harro CD, Betts RF, Hartzel JS, Onorato MT, Lipka J, Smugar SS, et al. The immunogenicity and safety of different formulations of a novel *Staphylococcus aureus* vaccine (V710): results of two Phase I studies. *Vaccine*. 2012;30(9):1729-36.
45. Otsuka T, Kirkham C, Johnson A, Jones MM, Murphy TF. Substrate Binding Protein SBP2 of a Putative ABC Transporter as a Novel Vaccine Antigen of *Moraxella catarrhalis*. *Infection and immunity*. 2014;82(8):3503-12.
46. Skinner JM, Indrawati L, Cannon J, Blue J, Winters M, Macnair J, et al. Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model. *Vaccine*. 2011;29(48):8870-6.
47. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2011;331(6016):430-4.
48. Golubchik T, Brueggemann AB, Street T, Gertz RE, Jr., Spencer CC, Ho T, et al. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nature genetics*. 2012;44(3):352-5.
49. Reynolds R. Antimicrobial resistance in the UK and Ireland. *The Journal of antimicrobial chemotherapy*. 2009;64(S1):19-23.
50. Jeraldo P, Sipos M, Chia N, Brulc JM, Dhillon AS, Konkel ME, et al. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(25):9692-8.
51. Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J*. 2010;4(3):337-45.
52. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science*. 2009;326(5960):1694-7.
53. Cobey S, Lipsitch M. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science*. 2012;335(6074):1376-80.
54. Mikkelsen GM. Ecological niches: Linking classical and contemporary approaches. *Biol Philos*. 2005;20(2-3):557-66.
55. Chase JM, Leibold MA. *Ecological Niches: Linking Classical and Contemporary Approaches*. Chicago, USA: University of Chicago Press; 2003.