UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Clinical and Experimental Sciences

Epidemiology and Ecology of Microbial Communities of the Upper Respiratory Tract

Volume 1 of 1

Abigail Lois Coughtrie

Thesis for the degree of Doctor of Philosophy

August 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Clinical and Experimental Sciences

Doctor of Philosophy

EPIDEMIOLOGY AND ECOLOGY OF MICROBIAL COMMUNITIES OF THE UPPER RESPIRATORY TRACT

by Abigail Lois Coughtrie

Respiratory tract infections (RTI) are responsible for over 4 million deaths per year worldwide. Microbial carriage in the upper respiratory tract is a precursor to respiratory infection and facilitates person-to-person transmission. A large community-based swabbing study was conducted, enabling the collection of a large number of swab samples that would provide key information concerning the epidemiology and ecology of respiratory tract communities. Traditional culture-based techniques, molecular methods and ecological and mathematical modelling methods were used.

Participation of members of the community within the swabbing study was shown to be greater within the self-swabbing group, in older individuals and in less deprived locations. Carriage of bacterial and viral species within the respiratory tract was shown to vary with participant age, recent RTI and the presence of other species. Self-taken swabs were largely non-inferior to healthcare professional (HCP)-taken swabs in assessing carriage of the targeted bacteria, offering a cheaper and more flexible alternative to HCP swabbing. Large numbers of capsular types (serotypes), sequence types and low levels of vaccine-targeted types demonstrate the genetic diversity of respiratory bacteria as well as their evolution in response to immunisation.

Microbial respiratory community structure was shown to be highly variable with less nested communities and facilitative relationships between species within young individuals and those with recent RTI potentially enhancing transmission and survival of carried species. Neutral and niche processes were both found to be important in respiratory tract community assembly. These insights into respiratory tract communities will allow predictions of microbial variation as a result of infection, varying age and season. Future work will involve 16S rDNA community analysis, further development of ecological methods and the conduction of larger multi-centre carriage studies.

Table of Contents

List of Tables

List of Figures

Declaration of Authorship

I, Abigail Coughtrie, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Epidemiology and Ecology of Microbial Communities of the Upper Respiratory Tract

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Either none of this work has been published before submission, or parts of this work have been published as:

Coughtrie, A.L., Whittaker, R.N., Begum, N., Anderson, R., Tuck, A., Faust, S.N., Jefferies, J.M.,Yuen, H.M., Roderick, P.J., Mullee, M.A., Moore, M.V., Clarke, S.C. (2012) Microbial analysis of the upper respiratory tract to support the development of effective vaccine policy: a population-based pilot study. *The Lancet*, 380, S31, p23. [PUBLISHED ABSTRACT FOR A CONFERENCE POSTER].

Coughtrie, A.L., Whittaker, R.N., Begum, N., Anderson, R., Tuck, A., Faust, S.N., Jefferies, J.M., Yuen, H.M., Roderick, P.J., Mullee, M.A., Moore, M.V., Clarke, S.C. 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.

Signed: ……

Date: ……..

Acknowledgments

I would like to thank my principal supervisor Stuart Clarke for giving me the opportunity to undertake this project and for all of his help and advice throughout. I also thank my co-supervisors Lex Kraaijeveld, Patrick Doncaster and Jeremy Webb for providing advice and support. In addition, I thank the investigators of the swabbing study for providing guidance and advice during the project and being instrumental in getting my first paper published.

Study design was undertaken by Stuart Clarke (principal investigator), Saul Faust, Mark Mullee, Michael Moore, Johanna Jefferies and Paul Roderick prior to commencement of my PhD. Study setup was undertaken with the assistance of Robert Whittaker, Research Facilitator, and extended to a second study time-point with the assistance of Denise Morris, Research Facilitator. I was involved in questionnaire and information sheet design, communication with GP practices and creation of sites files throughout the study. I was solely responsible for the input of data into SPSS and coding of questionnaire responses. Statistical guidance was obtained and randomisation of individuals for inclusion in the study was undertaken by Ho Ming Yuen, Medical Statistician at the University of Southampton. I then performed all statistical analyses in full with the use of this guidance. I was involved in the processing of swab samples and culture identification during the first time-point with the assistance of technical staff Andy Tuck, Robert Whittaker, Nelupha Begum, Rebecca Anderson, Shabana Hussein, Beverly Simms, Christine Tumman and Karen Cox. All swab processing and identification during the second time-point was undertaken by the technical staff.

I was fully involved in undertaking the entire sequencing protocol on a number of isolates. Due to the volume of isolates, sequencing was undertaken on the remaining isolates with the help of sequencing technicians Alice Still, Rebecca Anderson and Jenna Alnajar. I performed all bioinformatics analyses, molecular serotyping and ecological analyses in full. Guidance on the use of ecological methods was obtained from co-supervisors Lex Kraaijeveld and Patrick Doncaster.

Real-time PCR work was undertaken in collaboration with the Public Health England Southampton laboratory. Nusreen Ahmed and Peter Marsh aided me in the set up and completion of the work. All validation of real-time PCR assays was undertaken by the PHE laboratory prior to analysis of my samples, as these assays are used for routine

diagnostic work as well as research projects. I am thankful for the time they took out of their busy schedules to help me complete my work.

Additionally I would like to thank my colleagues and friends of the Infectious Disease Epidemiology Group, who have been a real pleasure to work with. My three years at the University have flown by and simply would not have been the same without their sense of humour and kindness. I feel truly lucky to have worked with such a lovely group of ladies.

I also thank my family, particularly my fiancé Nick who has been so supportive throughout my extensive studies but also for being patient with me during the stresses of my PhD and for cooking countless meals during my write up. I also thank my mum and dad who have always supported me in my ambitions and goals and for helping me get to where I am today.

Finally I would like to thank those who provided financial support. I thank the Rosetree's Trust and Stuart Clarke for funding my PhD and the Bupa Foundation for funding the swabbing study. I would also like to thank the Society for General Microbiology and the University of Southampton's Conference Attendance Fund for enabling me to travel to Asia for the very first time to present my work at the 15th International Symposium on Microbial Ecology in Seoul, South Korea.

Abbreviations

CFU	Colony-forming units
CI	Confidence Interval
EV	Enterovirus
FET	Fisher's Exact Test
GP	General practitioner
HCP	Healthcare professional
Hib	*Haemophilus influenzae* type b
IMD	Indices of multiple deprivation
IVA	Influenza A virus
IVB	Influenza B virus
LRTI	Lower respiratory tract infection
MLST	Multilocus sequence typing
MPV	Metapneumovirus
MRSA	Methicillin-resistant *Staphylococcus aureus*
NP	Nasopharyngeal
NTHi	Non-typeable *H. influenzae*
OR	Odds ratio
PCR	Polymerase chain reaction
PCRN	Primary care research network
PCV	Pneumococcal conjugate vaccine
PIV	Parainfluenza virus
RSV	Respiratory syncytial virus
RTI	Respiratory tract infection
RV	Rhinovirus
SS	Self-swabbing
ST	Sequence type
STGG	Skim milk, tryptone, glucose and glycerine
URTI	Upper respiratory tract infection
WMS	Whole mouth swab

Chapter 1. Introduction

1.1 Respiratory Tract Infections

Respiratory tract infections (RTI) are responsible for over 4 million deaths per year worldwide, and are the most common cause of death in children under the age of one year (Lozano et al., 2012). In the UK alone, an estimated 6 million people seek medical advice annually regarding an RTI, posing a significant economic burden to the National Health Service (National Institute for Health and Care Excellence, 2008). RTI can be categorised as upper (URTI) and lower (LRTI), depending on the site of infection.

1.1.1 Upper Respiratory Tract Infections

URTI encompass otitis media, pharyngitis, sinusitis and nasopharyngitis amongst others (Cimolai, 1988). The most common bacterial organisms responsible for URTI are *Streptococcus pneumoniae, Haemophilus influenzae* and *Moraxella catarrhalis* (Murphy et al., 2009). Viral pathogens also commonly cause URTI, with adenovirus, rhinovirus and coronavirus being particularly prevalent (Bicer et al., 2013). Viral URTI are extremely common in children and are often complicated by infections of the middle ear (Chonmaitree et al., 2008).

Otitis media, commonly caused by *S. pneumoniae* and *H. influenzae*, is an infection of the middle ear and can involve inflammation of the eardrum and fever (Cimolai, 1988, Vergison, 2008). This type of infection is common in children, with over 80% of young children from the USA experiencing at least one occurrence before 3 years of age and nearly half experiencing three occurrences or more before this age (Vergison, 2008). Sinusitis, an infection of the sinuses, is characterised by facial pain, nasal discharge, congestion and fever (Brook, 2001). Bacterial infection of the sinuses is usually preceded by a viral infection of the upper respiratory tract, such as colds or flu (Piccirillo, 2004). Pharyngitis, a common ailment, is frequently associated with Group A Streptococci (Bisno, 2001).

Whooping cough, also known as pertussis, is a contagious respiratory disease caused by *Bordetella pertussis.* The disease commonly causes severe symptoms and high

mortality rates in children, but can also affect adults, acting as a source for transmission to children (von König et al., 2002). Incidence in older children and adults has become increasingly common as a result of waning immunity following vaccination (Paisley et al., 2012).

1.1.2 Lower Respiratory Tract Infections

LRTI is the second highest cause of years of life lost (YLL) after ischaemic heart disease and caused approximately 2.8 million deaths per year worldwide in 2010 (Lozano et al., 2012). These infections can result from invasion by upper respiratory species as well as other pathogens. Pneumonia is the biggest cause of childhood mortality and is responsible for 19% of deaths in children aged 4 years or younger (WHO and UNICEF, 2006). Bacterial or viral infections of the lower respiratory tract occur at higher levels in disease states such as chronic obstructive pulmonary disease (COPD) and asthma where lung function is impaired (Sethi, 2010, Sykes and Johnston, 2008).

There are three categories of organisms responsible for LRTI: typical bacteria, atypical bacteria and respiratory viruses (Wolf and Daley, 2007b). Typical bacteria include *S. pneumoniae, M. catarrhalis, H. influenzae* and *Staphylococcus aureus*. Atypical bacteria, which show different clinical symptoms to typical bacteria, incorporate *Legionella species, Mycoplasma pneumoniae, Chlamydophila pneumoniae* and *Chlamydophila psittaci* (Wolf and Daley, 2007a). Respiratory viruses causing LRTI include respiratory syncytial virus (RSV), parainfluenza viruses and influenza viruses (Wolf and Daley, 2007b).

Mycobacterium tuberculosis, the causative agent of tuberculosis, remains a global problem with 8.8 million occurrences of disease reported in 2010. Prevalence of tuberculosis is highest in India, China, Pakistan and Bangladesh (WHO, 2011). Drug resistance is problematic in the treatment and control of disease as multiple drug resistant (MDR) and extensively drug resistant (XDR) strains are becoming increasingly common. MDR-TB is thought to occur in approximately 3.6% of all TB cases worldwide (WHO, 2010).

Influenza is a widespread cause of LRTI, with approximately 90 million identified cases of the virus in 2008 and a significant number of children under the age of 5 dying as a

result of Influenza infection (Nair et al., 2011). The young and elderly are particularly at risk of infection with elderly individuals demonstrating high rates of Influenza-associated mortality when compared with other age groups (Neuzil et al., 2002, Falsey et al., 1995). Influenza viruses encompass three distinct types (A, B and C), which are further subdivided into genetically varied subtypes. Influenza A is classified according to haemagglutinin and neuraminidase proteins of which there are 18 and 11 types, respectively (CDC, 2014b).

1.2 Invasive Disease

Septicaemia and meningitis occur when a local infection within the respiratory tract spreads into the blood and central nervous system (Greenough, 1996). Meningitis involves bacterial invasion of the cerebral spinal fluid (CSF) causing seizures, fatigue and temperature extremes (Baas Rubarth, 2010). This invasion results in host immune system stimulation and inflammation of the meninges, causing severe damage to the nervous system (Saez-Llorens and McCracken, 2003). Neonates are the most at risk of meningitis, with 8% of premature babies contracting the disease (Greenough, 1996).

Septicaemia involves microbial invasion of the bloodstream leading to a condition known as sepsis (Baas Rubarth, 2010). Sepsis involves a systemic inflammatory response syndrome (SIRS) caused by hyper-stimulation of the host immune responses as a result of bacterial invasion (Bone et al., 1992). Levels of cytokines, in particular, are thought to become elevated causing the main symptoms of SIRS (Rittirsch et al., 2008). This can lead to single or multiple organ failure, septic shock and death (Baas Rubarth, 2010). Septicaemia is referred to as bacteraemia when bacteria are present in the bloodstream, viraemia when viruses are present in the bloodstream, fungaemia when fungi are present in the bloodstream and, finally, parasitaemia when parasites are present in the bloodstream (Bone et al., 1992).

1.3 Epidemics and Pandemics of Respiratory Disease

1.3.1 Epidemics and Pandemics of Bacterial Infection

N. meningitidis, or meningococci, are responsible for widespread epidemics of meningitis. Prior to meningitis C (menC) vaccine introduction, this serogroup was

endemic in Europe (Miller et al., 2001). Serogroup A is particularly problematic in the Meningitis Belt of Africa during the dry season, with epidemics and pandemics occurring up to every two and five years respectively (Stephens et al., 2007). Serogroup W135 is highly associated with epidemics in the Hajj pilgrims in Mecca (Aguilera et al., 2002). Serogroup B epidemics remain highly problematic in New Zealand (Baker et al., 2001). In addition, serogroup B also remains endemic in Europe and the Americas (Perrett and Pollard, 2005).

Outbreaks of pneumococcal disease have been reported historically in closed setting environments such as nursing homes and jails (Hoge, 1994, Nuorti et al., 1998). More recently, an outbreak of pneumococcal meningitis in Ghana caused by the sequence type (ST) 217 showed similarities to meningococcal meningitis epidemics in this region of Africa (Leimkugel et al., 2005). Small outbreaks of *H. influenzae* serotype a have also been reported, including a group of five related cases in Alaska, USA, in 2003 (Hammitt et al., 2005). Furthermore outbreaks of disease caused by non-typeable *H. influenzae* (NTHi) have also been reported in closed-setting environments. An outbreak in the USA reported 13 cases within a nursing home in 2005 (Van Dort et al., 2007). In the Netherlands, a similar number were affected within a rehabilitation centre in 1988 (Sturm et al., 1990). Finally in the UK, an outbreak of NTHi disease was reported in a geriatric ward (Mehtar and Law, 1980).

Historically, pertussis epidemics were frequent (Cherry et al., 1988), but in recent years epidemics are thought to be the result of waning protection of the acellular pertussis vaccine (Klein et al., 2012). Epidemics of pertussis have occurred in the USA (Winter et al., 2012), Australia (Campbell et al., 2012), the Netherlands (de Greeff et al., 2010) and the UK (Public Health England, 2012). Methicillin-resistant *S. aureus* outbreaks and epidemics are common in both hospital and community settings (Koser et al., 2012, Tenover et al., 2006). Epidemic strains EMRSA-15 and EMRSA-16 have been particularly problematic in hospitals across the UK due to increased association with bacteraemia (Johnson et al., 2001).

1.3.2 Epidemics and Pandemics of Viral Infection

Epidemics and pandemics of viral infection are relatively common, occurring seasonally in many cases (Kim et al., 1973, Dushoff et al., 2004). Influenza pandemics are a serious public health threat and can be highly unpredictable. Epidemic spread is assisted by outbreaks in close contact settings such as nursing homes, schools and

nurseries even if vaccination is routine (Camilloni et al., 2010). Historical examples of influenza pandemics include the 1918, 1957/1968 and 2009 pandemics, which led to millions of deaths worldwide (Johnson and Mueller, 2002, Kawaoka et al., 1989, Miller et al., 2010). More recent Influenza epidemics and pandemics have also been important public health concerns. The H1N1 Influenza A pandemic of 2009, which originated from a genetically re-assorted swine/human Influenza virus, spread from Mexico to as many as 30 countries across the globe (Smith et al., 2009). During the epidemic children under the age of 5 years were highly affected by the virus (Miller et al., 2010). Furthermore, avian Influenza A strain H5N1 has caused epidemics in humans, affecting nearly 600 people and causing over 300 deaths since 2003 (Gambotto et al., 2008, WHO, 2012). Finally a H7N7 avian influenza strain outbreak occurred in 2003, affecting over 450 people in the Netherlands (Koopmans, 2004).

RSV and metapneumovirus can also cause epidemics of acute respiratory infections annually in most countries (Duppenthaler et al., 2003, Heininger et al., 2009). Coronaviruses are also an important cause of widespread epidemics and pandemics (Gallagher and Perlman, 2013). For example, the 2002-2003 severe acute respiratory syndrome coronavirus (SARS-CoV) epidemic, originating in China, affected more than 8,000 individuals in 25 countries and resulted in over 750 deaths in 2002 (Peiris et al., 2004, Peiris et al., 2003). High transmissibility, especially within hospitals, was thought to be responsible for the spread of the virus (Riley et al., 2003). The Middle Eastern respiratory syndrome coronavirus (MERS-CoV) has caused recent concern due to the severity of symptoms occurring in those affected, including pneumonia and renal failure (Zaki et al., 2012). Since the summer of 2012, 536 laboratory-confirmed cases have been reported in humans, mainly within Saudi Arabia (CDC, 2014a). Low transmissibility of the virus may, however, have prevented widespread transmission of the virus to other countries (Gautret et al., 2013).

1.3.3 Secondary Bacterial Infections

Secondary bacterial infections associated with viral infections are common (Brundage, 2006, Crowe et al., 2011). The seasonality of bacterial respiratory tract infections is thought mainly to be due to viral co-infections (Bogaert et al., 2011). During the 1918 influenza pandemic, bacterial pneumonia rates were high with *S. pneumoniae* being cultured from 29% of patients dying during the pandemic (Brundage, 2006). *S. pneumoniae, S. aureus, Streptococcus haemolyticus* and *H. influenzae* were also isolated from post-mortem blood cultures of fatal influenza (Spooner, 1919). Similar

trends were observed in the Asian influenza pandemic (1957-1958) and the 2009 H1N1 influenza outbreak (DeLeo and Musser, 2010). In addition, bacterial meningitis and staphylococcal toxic shock syndrome have both shown increased incidence during influenza outbreaks (Brundage, 2006, Cartwright et al., 1991, Sperber and Francis, 1987). It is thought that these secondary bacterial infections can result in greater complications and higher death rates than would be found with influenza alone (McCullers, 2006, Crowe et al., 2011). Bacterial secondary infections are thought to account for the largest proportion of deaths during the influenza season as opposed to the virus itself (Lee et al., 2010).

1.4 Microbial Carriage in the Respiratory Tract

1.4.1 The Biology of Respiratory Carriage

The human respiratory tract is home to a diverse set of microorganisms. The nasopharynx alone has been shown to comprise as many as 250 species across 13 phyla (Bogaert et al., 2011). This microbial community is thought to develop throughout the first year of life, eventually fulfilling a protective function (Vives et al., 1997). The respiratory tract is host to a multitude of organisms including *S. pneumoniae, H. influenzae, M. catarrhalis, S. aureus, N. meningitidis, Pseudomonas aeruginosa*, influenza A and B viruses, parainfluenza 1-3 viruses, RSV, rhinovirus, adenovirus, *Corynebacterium* species, *Fusobacterium* species and other *Streptococci* and *Staphylococci* (Bogaert et al., 2011, Bogaert et al., 2004a, Ruuskanen et al., 1989, Harrison et al., 1999). In addition to microbial species which are easily cultured by standard microbiology techniques, there are a number of species which might be important members of the respiratory community which cannot be assessed using standard microbiology techniques. Therefore, non-culture techniques including metagenomics are becoming increasingly vital in determining the true diversity of the respiratory tract (Bogaert et al., 2011).

Carriage involves dynamic fluctuations of microorganisms and strains within the respiratory tract niche (Weiss-Salz and Yagupsky, 2010). Single bacterial species or specific serotypes within a particular bacterial species can be carried transiently or for prolonged or sequential periods (Garcia-Rodriguez and Fresnadillo Martínez, 2002). Variation in carriage of respiratory tract organisms across the population has been shown to be highly significant.

Frequency of carriage is often found to be variable with age. *S. pneumoniae, H. influenzae* and *M. catarrhalis* carriage has been shown to increase within the first year of life, with prevalence of each bacterial species reaching over 80% within the first few months (Kwambana et al., 2011). Carriage of these three bacterial species subsequently decreases with age, with *S. pneumoniae* being isolated from19%, 6% and 0.8% of those aged <7 years, 7-15 years and ≥16 years; *H. influenzae* being isolated from 13%, 6% and 3% of individuals from the same age groups and *M. catarrhalis* being isolated from 27%, 4% and 2% of individuals from these age groups (Gunnarsson et al., 1998). *S.aureus* carriage, in contrast, demonstrates a very different pattern of variation with age. Within neonates, *S. aureus* carriage drops from 50% within the first week after birth to 20% at 9 weeks (Kwambana et al., 2011). After approximately 2 years of age, *S.aureus* carriage was found to increase (Regev-Yochay et al., 2004a). Furthermore, carriage is then thought to rise to a peak frequency at approximately 11 years of age and then decline throughout adult life (Bogaert et al., 2004c). Carriage of *N. menigitidis* is low in those aged 0-4 years (2.1%) and increases until the age of 15-19 years (24.5%) and then declines again with age (Cartwright et al., 1987). Furthermore, one study have shown a peak in frequency of *N. meningitidis* carriage in the first year of life as well as in those aged 15-19 years (Bogaert et al., 2005).

Furthermore, differences in the frequency of carriage according to swabbing site have also been observed. *S. pneumoniae* carriage is predominantly in the nasopharynx (NP) of children but similarly in the NP and oropharynx (OP) of their mothers (Greenberg et al., 2004). *H. influenzae* carriage, on the other hand, was similar in NP and OP swabs collected from children but much higher in the OP versus the NP in their mothers (Greenberg et al., 2004). In a further study in young children <15 months old, *S. pneumoniae* carriage was higher in the nose than the OP, whereas *H . influenzae* was carried equally in these two sites (Capeding et al., 1995). *N. meningitidis* carriage has been shown to be higher in the OP than the NP, within family members of patients with meningococcal and non-meningococcal meningitis from Sweden (Olcen et al., 1979). Furthermore, OP swabs showed greater sensitivity and specificity for Group A Streptococci than oral cavity swabs in children aged 3-18 years presenting with Streptococcal pharyngitis (Fox et al., 2006).

The presence of host symptoms during bacterial colonisation is thought to be important for increased transmission from host to host. *H. influenzae* carriage was found to be associated with symptoms of rhinitis within children <6 years in Portugal (Rodrigues et

al., 2013). Duration of carriage is also variable between different microbial species colonising the respiratory tract. *S. pneumoniae* and *H. influenzae* were found to colonise the respiratory tract for a mean of 96 and 74 days respectively in newborns from Papua New Guinea (Gratten et al., 1986). *M. catarrhalis*, on the other hand, was found to colonise asymptomatically for a median of 40.4 days in patients with COPD compared with a median of 31.0 days during exacerbations (Murphy et al., 2005). In *S. pneumoniae*, duration of carriage is negatively associated with the proportion of cases of invasive disease in relation to carriage (known as attack rate), with longer periods of carriage observed in serotypes with fewer invasive pneumococcal disease (IPD) cases relative to carriage (Sleeman et al., 2006).

1.4.2 Carriage Study Methodology

Assessing carriage of microorganisms within the respiratory tract can be undertaken using a variety of sample sites, including the nasopharynx, oropharynx, nose and throat (Trotter and Gay, 2003, Watt et al., 2004, Gamblin et al., 2013). Methods for collecting respiratory samples may also include the use of swabs, nose blowing, saliva, nasal washings or nasopharyngeal aspiration (Akmatov et al., 2011, Leach et al., 2008, Lieberman et al., 2006, Rapola et al., 1997, Greenberg et al., 2004, Allen et al., 2013, Wyllie et al., 2014). The WHO Pneumococcal Vaccine Trials Carriage Working Group recommend the use of a nasopharyngeal swab for detecting the presence of *S. pneumoniae* within the respiratory tract (O'Brien and Nohynek, 2003, Satzke et al., 2013). A recent systematic review showed that few studies fully adhere to these guidelines, with differences observed in the swab type, transport media and swab storage techniques (Gladstone et al., 2012).

Culture and non-culture methods can also be used for detecting carried microorganisms within respiratory samples. Culture is regarded as the gold standard method for detecting *S. pneumoniae* within respiratory swab samples (O'Brien and Nohynek, 2003, Satzke et al., 2013). No standard method has been outlined for the detection of other respiratory pathogens but the use of culture in clinical diagnostic laboratories is routine. Molecular methods such as the use of polymerase chain reaction (PCR) have increased sensitivity and specificity for detection of such microorganisms, allow decreased time for identification and enable the identification of non-culturable or slow-growing species (Caliendo, 2011). Antigen immunoassays are also used for the detection of species such as Group A Streptococci (Clegg et al.,

2003). Within a clinical setting, such rapid detection systems have demonstrated clear patient benefits in terms of fewer antibiotics used and less time spent in hospital (Woo et al., 1997).

1.5 Carriage Studies for the Evaluation of Vaccination Strategies

1.5.1 Methods for Assessing Vaccine Effectiveness

Methods for assessing effectiveness of vaccination strategies include case-control studies, disease and outbreak investigations as well as screening methods (Orenstein et al., 1985). Disease surveillance is a commonly used method of epidemiological post-licensing assessment of vaccination strategies. After the introduction of PCV-7, levels of IPD cases were monitored to ensure adequate control of invasive disease caused by this organism (Pilishvili et al., 2010). Similarly, following the implementation of the Hib vaccine, levels of *H. influenzae* type b disease were also scrutinised (Hargreaves et al., 1996). Finally, levels of *N. meningitidis* serogroup C disease were evaluated following the introduction of the MenC vaccine (Miller et al., 2001).

1.5.2 Relationship between Carriage and Disease

Carriage is the first essential step in the development of a RTI and enables transmission of an organism to a susceptible host (Weinberger et al., 2009). Duration of carriage is thought to affect the progression towards disease (Gray et al., 1980). Carriage and disease are intrinsically linked, with a number of studies showing reductions in both carriage and disease rates as a result of vaccination. The *H. influenzae* type b (Hib) conjugate vaccine reduced cases of invasive Hib disease by 95% (Mulholland et al., 1997) and Hib carriage by 60% (Adegbola et al., 1998) in a Gambian study involving children aged between one and two years old. Pneumococcal conjugate vaccination (PCV-7) reduced invasive disease caused by vaccine types by 99% and carriage of vaccine types by 95% in 1,301 children less than 5 years old in Alaska, USA (Keck et al., 2014). *Neisseria meningitidis* serogroup C vaccines reduced meningococcal disease cases by 97% in 15-17 year olds (Ramsay et al., 2001) and carriage of *N. meningitidis* serogroup C by 66% (Maiden and Stuart, 2002) in the UK. The reduction in disease incidence as a result of vaccination is thought to be largely

due to the reduction in carriage, as mucosal immune responses target serotype-specific capsular antigens (Zhang et al., 2004, Zhang et al., 2002).

Although duration of carriage has been found to vary between individuals, one study estimated that 74% of pneumococcal infections resulted from carriage of a specific serotype for a short period of time (less than one month), indicating that invasion is most likely to occur within one month of acquiring a new strain (Gray et al., 1980). Furthermore, an outbreak strain of *N. meningitidis* serogroup C was found to be carried in only 0.4% of 9,796 individuals with no isolates in those aged 2-4 years, the age group that experienced the highest incidence of disease during the outbreak (Fernandez et al., 1999). Carriage of *N. meningitidis*, however, has been shown to be generally high in university students who also experience high level s of disease incidence, with periods of prolonged carriage being relatively common (Ala'Aldeen et al., 2000).

Genetic and phenotypic differences between specific species types causing disease and those carried within the general population have been observed. For example, in Europe *N. meningitidis* sequence type ST-11 was mostly present in disease cases during the 1990s whereas ST-23 was more frequently carried without disease occurrence (Caugant et al., 2007). These sequence types, which have been shown to have disease/carriage ratios of 6.6 and 0.8 respectively, are not exclusively found in either disease or carriage and it is thought that multiple factors influence the propensity of different serogroups or types to cause disease (Caugant and Maiden, 2009).

S. pneumoniae serotypes associated more often with IPD in 2006 were 1, 4, 5, 7F, 14 whereas those associated more often with colonisation were 6A, 6B, 8, 9V, 15, 18C, 19F, 23F, 33, 38 (Brueggemann et al., 2003). One study demonstrated that serotypes 4, 14, 7F, 9V and 18C had high IPD incidence relative to numbers of acquisitions but short duration of carriage whereas serotypes 23F, 6A, 19F, 16F, 6B and 15B/C had low IPD incidence relative to numbers of acquisitions and longer duration of carriage (Sleeman et al., 2006). Despite the differences in rates of disease relative to carriage, no serotype is exclusively found in disease or carriage. Both host and microbial factors are thought to determine whether a strain in carried asymptomatically or whether it invades the host causing disease. Pneumococcal factors include the capsule, adhesins, phase variation, biofilm formation and interactions with other species (Simell et al., 2012).

1.5.3 Elucidating Respiratory Epidemiology

Carriage studies generate an understanding of specific microorganisms carried within the general population at any given time. Measurement of the genetic and phenotypic diversity of carried species and their distribution within the population is important for understanding the dynamics of the microbial community (Caugant et al., 2007). For example, determining the distribution of meningococcal or pneumococcal serogroups or serotypes across the UK enables us to understand the relevance of each capsular group or type and the effect of interventions on them (Maiden and Stuart, 2002, Tocheva et al., 2011, Choi et al., 2011, Lakshman et al., 2003).

Environmental factors can also be assessed for their impact on carriage, allowing for a greater understanding of what factors can put a patient at risk (Garcia-Rodriguez and Fresnadillo Martínez, 2002). Cigarette smoking or even passive smoking have been linked to greater colonisation by *S. pneumoniae*, *M. catarrhalis*, *H. influenzae* and *N. meningitidis* (Bakhshaee et al., 2012, Stanwell-Smith et al., 1994). Diversity and bacterial community disturbances have also shown to be increased in smokers, causing pathogens to thrive and protective organisms to wane (Charlson et al., 2010, MacLennan et al., 2006). Diet, season, age, social status, gender and presence of children within a family were among the factors also found to effect colonisation rates of key bacteria in the respiratory tract (Garcia-Rodriguez and Fresnadillo Martínez, 2002).

Knowledge of geographical and temporal distributions of respiratory microbes is important for understanding the dissemination and transmission of circulating strains. Annual carriage studies can enable the temporal distribution of circulating bacterial species to be tracked (Trotter and Gay, 2003). Geographical distribution is also essential in determining spatial epidemiology of circulating organisms and how this affects disease. Geographical distribution of meningococcal serogroups is highly divergent with serogroup A predominating in central Africa and serogroup Y in the USA (Harrison et al., 2014, Trotter and Greenwood, 2007). In addition, frequency of IPD serotypes after implementation of PCV-7 into the childhood immunisation schedule (post-PCV-7) were found to be divergent in eight areas of the USA despite similar serotype replacement, possibly due to differing levels of chronic disease and antibiotic resistance (Rosen et al., 2011). Understanding distribution across space and time can allow treatment and prevention strategies to be targeted correctly, improving their effectiveness. Determining patterns of meningococcal or pneumococcal serogroups or serotypes across the UK has enabled an understanding of the relevance of each group

and the effect of interventions on them (Maiden and Stuart, 2002, Choi et al., 2011). A recent study assessing the effect of 4CMenB and MenACWY-CRM *N. menigitidis* vaccines in England demonstrated clear reductions in the serogroups targeted by the vaccines within three months of vaccination (Read et al., 2014).

1.5.4 Evaluating Vaccine Effectiveness in Disease and Carriage

The direct relationship between carriage and disease enables carriage to be used as a proxy for protection and can enable the prediction of risk factors for respiratory disease (Simell et al., 2012). In terms of carriage, vaccine efficacy (VE_{acq}, acq = acquisition) is the extent of decline in colonisation rates of a particular species within a vaccinated population relative to unvaccinated controls (Rinta-Kokko et al., 2009). Comparing carriage in vaccinated individuals versus non-vaccinated controls may be problematic due to differences in the serotypes or serogroups carried within these populations. However, once herd immunity is present, carriage will be similar in vaccinated and unvaccinated individuals (Rinta-Kokko et al., 2009, Weinberg and Szilagyi, 2010). The use of carriage studies in evaluating vaccine effectiveness was highlighted by the Joint Committee on Vaccination and Immunisation (JCVI), an expert advisory board for the UK Department of Health, which emphasized a lack of carriage evidence for the effectiveness of serogroup B vaccine (4CMenB) in preventing serogroup B disease in the UK (JCVI, 2013). Prior to vaccine introduction, levels of carriage are assessed and act as a baseline measure from which any changes resulting from vaccination can be gauged (Mulholland and Satzke, 2012). Direct effects on targeted serotypes or serogroup, indirect effects on non-targeted serotypes or serogroups as well as herd immunity can be evaluated.

Countries across the globe are involved in tracking these changes in order to improve vaccines and keep up with epidemiological changes of *S. pneumoniae*. Studies in the USA, France, UK and Portugal have highlighted such epidemiological changes (Hicks et al., 2007, Couloigner et al., 2012, Flasche et al., 2011, Frazao et al., 2010). Furthermore, the evaluation of the 4CMenB and MenACWY-CRM *N. meningitidis* vaccines demonstrated direct effects of diminished carriage of the serogroups targeted by the vaccines in young adults as well as evidence for the potential for indirect effects of herd immunity which will in turn contribute to a reduction in disease (Read et al., 2014).

1.6 The Respiratory Tract Ecosystem

1.6.1 The Role of Diversity in Health and Disease

The respiratory tract is a highly diverse environment. The presence of individual species within the respiratory tract is influenced by an array of host and environmental factors, including season, living conditions, exposure to cigarette smoke, diet, sleeping position, age and vaccination (Harrison et al., 1999, Garcia-Rodriguez and Fresnadillo Martínez, 2002, Charlson et al., 2010, MacLennan et al., 2006). In addition, siblings and attendance at day care facilities can also increase the incidence of microbial carriage and disease of the respiratory tract (Vives et al., 1997). Previous colonisation by a species as well as the presence of competing and facilitative species also affects the species present within the respiratory tract (Spinola et al., 1986, Bosch et al., 2013). Genetics and chronic disease, including COPD and asthma, also play an important role in the respiratory microbiome as these affect colonising bacterial communities and interactions between these species (Erb-Downward et al., 2011). As such, seasonal, person-to-person and geographical variation of respiratory tract microbes is common. As a consequence of this it has been a challenge to identify a consistent principal microbial population within this environment (Bogaert et al., 2011).

Diversity is an essential feature of polymicrobial communities and is estimated by richness (the number of different species) and relative abundance of each species (Hubbell, 2001). Within an ecological community, species can be defined as either core or satellite species depending on their distribution and abundance (Hanski, 1982). This can be applied to microbial communities within the respiratory tract (van der Gast et al., 2011). It has often been observed that where there is greater diversity within an ecosystem, there is also greater stability (McCann, 2000). It has been suggested that disruption to such stable and diverse communities found within the respiratory tract environment allows pathogens to invade and cause disease (Frazao et al., 2010). Lower diversity has been observed in respiratory tract disease states, including Chronic Obstructive Pulmonary Disease (COPD) where *Pseudomonas* species were found to be most prevalent. Severity of COPD is also linked to the reductions in diversity (Rabe and Wedzicha, 2011, Erb-Downward et al., 2011). Furthermore, asthma is associated with disruptions to the microbial lung community with Proteobacteria, in particular, being dominant in children suffering from this condition (Hilty et al., 2010). It has also been observed that *S. pneumoniae*-positive respiratory tract populations demonstrate less diversity and heterogeneity than *S. pneumoniae*-negative populations, indicating

the possible disruptive effects of this organism on the respiratory tract community (Laufer et al., 2011).

Diversity is thought to enable protection from invasion and subsequent disease in the respiratory tract caused by opportunistic or obligate pathogens that may populate or be introduced into this niche (Tuttle et al., 2011, Gonzalez et al., 2011). Immune regulation and protection against epithelial impairment are also thought to be important roles of bacterial diversity (Erb-Downward et al., 2011). In patients with cystic fibrosis, microbial diversity is highly associated with lung function, with greater lung function correlating with a healthier and more stable microbiome (van der Gast et al., 2011).

Disruptions to microbial community diversity and richness can result from antibiotic use (Dethlefsen et al., 2008). Antibiotic treatment has been shown to cause a reduction in diversity of endotracheal aspirate samples in intubated individuals, with increased presence of *P. aeruginosa* (Flanagan et al., 2007). Carriage of *S. pneumoniae*, *H. influenzae* and *M. catarrhalis* have all been shown to decrease with antibiotic treatment within children under the age of three years in France, with Cefaclor and Erythromycin/sulfizoxazole reducing carriage of *H. influenzae*; Amoxicillin, Amoxicillin/clavulanate, Cefpodoxime and Erythromycin/sulfizoxazole reducing carriage of *S. pneumoniae* and a number of antibiotics reducing carriage of *M. catarrhalis* (Varon et al., 2000). The reduction in carriage of respiratory species as a result of antibiotics is often short-lived, with one study demonstrating *S. pneumoniae* carriage to drop from 68% to 29% within two weeks following antibiotic treatment but then rise to 78% and 87% two months and six months following treatment respectively (Leach et al., 1997). The microbial community is thought to be disrupted by antibiotic-induced increases in the frequency of resistant strains such as penicillin-resistant *S. pneumoniae* (Melander et al., 1998) and reductions in antibiotic-susceptible strains such as penicillin-susceptible *S. pneumoniae* (Varon et al., 2000). Studies have demonstrated clear effects of antibiotics on the microbiome of young children presenting with AOM, with increases in levels of Protoobacteria, *Pasteurellaceae*, Rothia and Actinomyces and decreases in levels of actinobacteria and Dolosigranulum (Lazarevic et al., 2013, Pettigrew et al., 2012, Hilty et al., 2012). However, no studies have demonstrated the effect of antibiotics on the microbiome in healthy individuals.

1.6.2 Assessing Diversity of an Ecological Niche

Inter-species diversity is not easily assessed using conventional culture methods as this selects for organisms that can easily grow under these limiting conditions (Tuttle et al., 2011). In addition, those that do grow are not necessarily the most important or even the most common organisms within this ecosystem (Hugenholtz, 2002). Molecular methods are now being used to better assess for diversity within specific niches, as they allow the identification of known, unculturable, and novel organisms (Woo et al., 2008). These include 16S rDNA variable region sequencing and PCR-mass spectrometry approaches that simultaneously identify groups of microorganisms within the same ecological niche (Dollhopf et al., 2001, Ecker et al., 2008).

Intra-species diversity involves the discrimination of bacterial types within samples using epidemiological typing methods, which is a useful tool for determining the diversity of populations of bacteria circulating within the respiratory tract. Serotyping and multilocus sequence typing (MLST) methods allow the discrimination of bacterial types according to capsular and housekeeping gene variability respectively (Maiden et al., 1998). Serotyping of *S. pneumoniae, H. influenzae* and *N. meningitidis* can be undertaken via PCR or more traditional serum agglutination assays to enable the identification of specific serotypes or serogroups of these bacterial species (Pai et al., 2006, LaClaire et al., 2003, Mothershed et al., 2004). However, serogrouping of *N. meningitidis* carriage isolates can be problematic due to low capsule expression within these isolates (Claus et al., 2002). MLST can be undertaken via PCR amplification, sequencing and analysis of the target housekeeping genes (Enright and Spratt, 1998) or more recently by bioinformatics analysis of whole genome sequences (Inouye et al., 2012).

1.7 Biofilms and Respiratory Tract Infections

Biofilms are highly organised structures allowing bacteria to survive within the host or the environment (Davey and O'Toole G, 2000). RTIs are thought to involve the formation of dense bacterial biofilms. Otitis media is a disease that has often been associated with biofilm formation by *S. pneumoniae, H. influenzae* and *M. catarrhalis* (Hall-Stoodley et al., 2006). In addition, significant variability in biofilm formation among serotypes and isolates of *S. pneumoniae, H. influenzae* and *M. catarrhalis* has also been observed. These structures can differ in their biomass, thickness, protein

expression, pili and lipooligosaccharide (LOS) expression (Allegrucci et al., 2006, Murphy and Kirkham, 2002, Pearson et al., 2006, Stewart and Franklin, 2008).

The treatment of infections involving biofilms is problematic due to their mixed nature as well as their structure and properties that render then insensitive to many antibiotics (Murphy et al., 2009, Bakaletz, 2007, Stewart and Costerton, 2001). The recurrent nature of otitis media infections is thought to be a result of the presence of bacterial biofilms within the middle ear and the difficulty in eliminating organisms present in this depressed metabolic state (Murphy et al., 2009). Furthermore, the inability to culture many causative agents of respiratory tract infections may also be due to the presence of bacteria within biofilm structures as reduced growth rates within this state are often observed (Morris, 2007). This reduced growth rate also explains the slower penetration of antibiotics into bacteria found within biofilms (Donlan and Costerton, 2002).

Genetic diversification of bacteria growing in biofilms has been shown to occur at increased rates compared with planktonic bacteria (Boles et al., 2004, Webb et al., 2004). These structures facilitate horizontal gene transfer events, including conjugative transfer of plasmids and release and uptake of naked DNA (Molin and Tolker-Nielsen, 2003, Davey and O'Toole G, 2000). *H. influenzae* has been found to increase biofilm formation in *S. pneumoniae* during co-colonisation (Weimer et al., 2010). Both the diverse genotypes of bacterial species and the polymicrobial nature of biofilms have important implications in the control of respiratory tract infections. It has been suggested that future vaccine strategies should focus on targeting bacterial components expressed within the biofilm state (Jefferies et al., 2011). Targeting of microbial species within biofilms remains a challenge due to the highly divergent patterns of gene expression compared to planktonic bacteria (Harro et al., 2010). Despite these challenges, the importance of biofilms in the persistence of a number of infections means that the development of vaccines targeting these structures is vital. Such vaccine candidates have been investigated for *S. aureus* (Joyce et al., 2003).

1.8 Dynamics of the Respiratory Tract Ecosystem

1.8.1 Strain Mutation and Recombination

Mechanisms of bacterial genome dynamics include gene duplication, horizontal gene transfer, gene deletion, recombination and mutation and can occur within a single species or between species co-habiting the same ecological niche (Mira et al., 2002). Mutation of a bacterial strain occurs when the sequence of its DNA is modified, which may prevent correct gene function (Ochman and Davalos, 2006). Mutation can allow adaptation of a species to its environment as well as contributing to diversification. Mutation of IS-*bexA* in *H. influenzae* type a, acquired from *H. influenzae* type b, benefits this bacterium by allowing greater survival in the blood as a result of increased capsular polyribosylribitol phosphate production (Kroll et al., 1994). Furthermore, mutations in the capsular genes (*cps*) of pneumococci during biofilm formation have resulted in the formation of non-phase variable colonies of bacteria that are better adapted to the biofilm environment and confer greater adhesive capabilities to the organism (Allegrucci and Sauer, 2008). However, such mutations may carry a metabolic cost, especially in terms of transmission capabilities, growth or virulence (Andersson, 2006).

Recombination allows the exchange of DNA sequences within a species or between closely-related species within a single niche, due to the high level of similarity in their DNA sequences (Ochman et al., 2000). *S. pneumoniae* serotype 4 has undergone recombination in a capsular switch event in order to express a 19A capsule. This event, in response to vaccination, allows the formation of a 'vaccine escape recombinant' as well as to gain penicillin resistance genes (Brueggemann et al., 2007). Five distinct vaccine escape recombinants, P1 to P5, have been identified, all of which have switched from serotype 4 to serotype 19A since the introduction of PCV-7 (Golubchik et al., 2012). Conjugation allows the transfer of genetic material via plasmids (Willetts and Wilkins, 1984). Plasmids often mediate the transfer of antibiotic-resistance genes between species. *H. influenzae* ampicillin-resistant strains are thought to have acquired resistance via the transfer of a plasmid from enteric species (De Graaff et al., 1976).

1.8.2 Bacterial-Viral Interactions

Viral infections facilitate the development of secondary bacterial infections by causing respiratory epithelial damage, exposure of bacterial receptors and bacterial aerosolisation (McCullers and Rehg, 2002, Harford et al., 1949, Brundage, 2006, McCullers, 2006). Immune responses against viral infection can also deplete vital immune cells, such as granulocytes, that protect against bacterial infection (Navarini et al., 2006). Glucocorticoid production prompted by influenza infection is thought to stifle the immune response, allowing bacterial invasion (Jamieson et al., 2010). Furthermore, the influenza neuraminidase enzyme enables adherence of bacterial pathogens to the respiratory tract mucosa, facilitating development of secondary bacterial pneumonia (Peltola et al., 2005). Severe bloodstream infections and high mortality rates have been associated with influenza and bacterial co-infection with *S. pneumoniae* and *S. aureus* (Tasher et al., 2011). Influenza A strain H3N2 and RSV have furthermore been associated with increased rates of invasive pneumococcal pneumonia (IPP) in the USA (Zhou et al., 2012).

The human coronavirus NL63 has also shown to increase attachment of *S. pneumoniae* to respiratory tract epithelia (Golda et al., 2011). Bacteriophages also affect the structure of bacterial ecosystems by determining the quantity and diversity of different groups of species that they infect (Gonzalez et al., 2011). Furthermore rhinovirus and metapneumovirus have also been found to co-infect with various respiratory bacteria leading to severe respiratory diseases such as pneumonia and exacerbations of COPD (Madhi et al., 2006, Wilkinson et al., 2006).

In addition to interactions which may lead to disease, bacteria and viruses are thought to interact within their normal life cycles. Colonisation by bacterial species is thought to be associated with the lengthening of or increased symptoms of rhinitis during viral infection, which enable enhanced transmission of these species from host to host (Rodrigues et al., 2013). Influenza A infection has been found to increase the transmission and dissemination of *S. pneumoniae* with mice suggesting a role in carriage as well as disease (Diavatopoulos et al., 2010). Previous studies on retroviruses have also shown that interactions of these viruses with the normal microbial flora stimulate immune evasion which is necessary for transmission of the virus (Kane et al., 2011). Therefore it is thought that bacterial-viral interactions bring about conditions that shape adaptation and reproductive success of both bacterial and viral species.

1.8.3 Bacterial-Bacterial Interactions

Bacterial antagonism, also known as bacterial interference, is a well-documented feature of respiratory tract colonisation. A negative relationship between *S. pneumoniae* and *S. aureus* carriage has been reported in a number of studies, specifically vaccine type *S. pneumoniae* (Bogaert et al., 2004c, Regev-Yochay et al., 2004a). *S. pneumoniae* can inhibit colonisation of *S. aureus* via the production of hydrogen peroxide (Regev-Yochay et al., 2006). Furthermore, clinical trials have demonstrated increased colonisation by *S. aureus* following PCV-7 implementation (van Gils et al., 2011a). However, another study has demonstrated no effect of PCV-7 introduction on carriage frequency of *S. aureus* (Cohen et al., 2007). The findings of no association between *S. pneumoniae* and *S. aureus* have been criticised due to the non-inclusion of age and other important variables within the multivariate regression model (Regev-Yochay et al., 2008, Cohen et al., 2008). Finally, a paediatric carriage study of young children in Portugal did not find a significant negative relationship between *S. pneumoniae* and *S. aureus* carriage, which was hypothesised to be a result of the multivariate regression model accounting for differences in age as well as the use of quantitative PCR which provided greater detail concerning bacterial density and interactions (Rodrigues et al., 2013).

Relationships between *S. pneumoniae* and *H. influenzae* have also been described as competitive, with colonisation by *H. influenzae* being negatively associated with *S. pneumoniae* colonisation (Pettigrew et al., 2008). A further study also demonstrated this negative relationship but only when *S. pneumoniae* was the first to colonise the mucosa; when *H. influenzae* was the initial coloniser a positive association was observed (Margolis et al., 2010). The relationship between *S. pneumoniae* and *H. influenzae* is also dependent on the presence of other species, with *M. catarrhalis* colonisation resulting in a positive association between the two species (Pettigrew et al., 2008). However, studies have also shown positive relationships between *S. pneumoniae* and *H. influenzae* when controlling for age as well as using quantitative PCR for the detection of these species (Chien et al., 2013).

M. catarrhalis has been shown to frequently co-colonise with both *S. pneumoniae* and *H. influenzae* (Bosch et al., 2013). One study in the Netherlands demonstrated that *M. catarrhaiis* and *H. influenzae* were more often found to co-colonise than to be found alone within children under the age of two years (Verhaegh et al., 2011). This positive relationship between *M. catarrhalis* and *H. influenzae* was also found within the biofilm

state, with quorum sensing between the two species resulting in greater persistence, antibiotic resistance and immune evasion (Armbruster et al., 2010). Alpha-haemolytic streptococci also have the ability to reduce colonisation of *S. pneumoniae*, NTHi and *M. catarrhalis* by 92%, 74% and 89% respectively in infants with recurrent and secretory otitis media (Tano et al., 2000).

Competitive processes are thought to accelerate evolution of antagonistic bacteria within the same ecological niche (Paterson et al., 2010). Antagonism can also exist within a single species of bacteria (Weiss-Salz and Yagupsky, 2010). This is thought to explain why certain pneumococcal serotypes tend to prevail in the nasopharyngeal niche. Bacteriocins, toxic proteins produced by the bacterium, are responsible for this intraspecies antagonism, targeting differing pneumococci strains (Dawid et al., 2007). Synergism can also promote bacterial survival, transmission and growth.

1.8.4 Host-Pathogen Interactions

Bacteria and their hosts are thought to be constantly evolving in response to one other in a process known as Red Queen dynamics (Jefferies et al., 2011). As host immune responses are modified by vaccination, new genotypes of bacteria appear through mutation and recombination, creating a situation of on-going co-evolution (Jefferies et al., 2011). This co-evolution is thought to lead to accelerated intra- and inter-species evolution, leading to evolutionary transformation and even speciation (Paterson et al., 2010). These 'mutator' bacteria are able to survive better within the host and adapt to the stressful conditions imposed by the vaccine-induced changes in the host immune system (Jayaraman, 2011). Phase variation is also an important feature of Red Queen Dynamics, as specific antigens undergo modification to allow for adaptation, immune evasion as well as enhanced virulence and transmission (Jayaraman, 2011). *N. meningitidis* and *H. influenzae* undergo phase variation in their *nadA* and *mod* genes respectively, allowing for changes in gene expression and virulence properties (Jayaraman, 2011). Furthermore, *S. pneumoniae* undergoes phase variation in colony opacity antigens in order to allow for more successful colonisation of the nasopharynx and adherence to epithelial receptors (Weiser et al., 1994, Cundell et al., 1995).

The immune system is also thought to maintain microbial diversity and promote co-existence of species and strains. Mathematical modelling of respiratory tract carriage has demonstrated immune responses specific to pneumococcal serotypes as well as

non-specific acquired immunity to the bacterium support serotype diversity and enable multiple serotypes of *S. pneumoniae* to co-colonise (Cobey and Lipsitch, 2012).

1.9 Vaccination Strategies

1.9.1 Non-Conjugate Vaccines against Respiratory Pathogens

The first vaccines introduced to control disease caused by respiratory pathogens were the inactivated whole cell *B. pertussis* vaccine introduced in 1926 and the inactivated toxoid vaccine against *Corynebacterium diphtheriae* introduced in 1923 (Ramon, 1923, Plotkin and Plotkin, 2011). More recently, capsular polysaccharide vaccines have been developed, which target the complex carbohydrate molecules of the bacterial capsule (Plotkin and Plotkin, 2011). The Pneumococcal Polysaccharide Vaccine (PPV) is used to target as many as 23 pneumococcal serotypes, and is commonly administered to the elderly and at-risk groups (Smit et al., 1977, Shapiro et al., 1991). The *H. influenzae* type b polysaccharide vaccine, containing polyribophosphate (PRP) polysaccharides, was also used for a short period of time to protect against this pathogen (Smith et al., 1973). Furthermore, *N.meningitidis* polysaccharide vaccines have included monovalent vaccines targeting serogroups A (Wahdan et al., 1973) and C (Artenstein et al., 1970) as well as a quadrivalent vaccine targeting A, C, Y and W135 (Lepow et al., 1986).

1.9.2 *H. influenzae* Type b Conjugate Vaccine

The first available conjugate vaccine was that against type b *H. influenzae* (Hib, multiple manufacturers). The vaccine is a capsular polysaccharide vaccine conjugated to either a tetanus toxoid, diphtheria protein or *N. meningitidis* serotype b outer membrane protein (Morris et al., 2008). The vaccine has been successful in reducing Hib disease in various countries, but economic barriers have prevented their use in a number of countries due to the expense of the vaccine (Girard et al., 2005). NTHi is now responsible for the majority of *H.* influenzae-related upper respiratory tract disease, highlighting the need for an NTHi vaccine (Wiertsema et al., 2011b). Numerous outer membrane components are being assessed for their potential use in vaccine design, including the minor outer membrane protein (OMP) P6, major OMPs and adhesins (Poolman et al., 2001, Michel et al., 2011). Furthermore, the PHiD-CV

(Synflorix®, GSK) mentioned in the following section also includes a non-typeable *H. influenzae* (NTHi) protein D (Croxtall and Keating, 2009).

1.9.3 Pneumococcal Conjugate Vaccines

Pneumococcal conjugate vaccines (PCVs) have been included in the National Immunisation Programmes of 44% of all World Health Organisation member states (CDC, 2013). They include PCV-7 and PCV-13 (Prevenar®, Wyeth) which were introduced in the UK in 2006 and 2010 respectively (Department of Health, 2002, Department of Health, 2010). Both vaccines include serotypes 4, 6B, 9V, 14, 18C, 19F and 23F (Clarke, 2006) but PCV-13 includes further serotypes 1, 3, 5, 6A, 7F and 19A (Department of Health, 2010). Targeted serotypes are those that are more frequently found to cause invasive disease, posing the greatest public health risk (Hicks et al., 2007). PHiD-CV (Synflorix®, GSK) is a further vaccine active against 10 serotypes in which eight are conjugated to a non-typeable *H. influenzae* (NTHi) cell surface lipoprotein known as protein D (Wysocki et al., 2009, Croxtall and Keating, 2009). The PHiD-CV vaccine has also been shown to confer protection against acute otitis media (AOM) infections, especially those caused by *S. pneumoniae* vaccine types. However, evidence for the protection against NTHi otitis media infections is unclear (Prymula et al., 2006, Talbird et al., 2010). The mode of action of this effect on NTHi otitis media infections is controversial (Dagan, 2008).

PCV15 (Merck), which is currently undergoing trials, contains serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, 22F, 23F and 33F and an aluminium phosphate adjuvant (Skinner et al., 2011). Furthermore, pneumococcal protein vaccines are being developed with the aim of targeting all *S. pneumoniae* serotypes. These include a trivalent vaccine containing pneumococcal choline-binding protein A (PcpA), pneumococcal histidine triad protein D (PhtD) and pneumolysin D1 which has been detoxified (PlyD1) (Ljutic et al., 2012). Furthermore, vaccines containing the pneumolysin toxoid alone or in combination with PhtD have recently been evaluated in phase I and II clinical trials (Leroux-Roels et al., 2014).

1.9.4 Meningococcal Vaccines

The meningococcal serogroup C conjugate vaccine (MenC) was included in the UK National Immunisation Programme in 1999 (Department of Health, 1999). There are thee separate vaccines which have been licensed for protection against meningococcal serogroup C disease: Meningitec® (Pfizer), NeisVac-C (Baxter) and Menjugate® (Novartis). The inclusion of MenC vaccines into the childhood immunisation programme has led to a significant reduction in the number of meningitis C cases as well as number of deaths within 2 years (Balmer et al., 2002). Following the introduction of the vaccine in the UK, catch up vaccination was also carried out in individuals aged less than 18 and was critical in the effectiveness of the vaccine (Trotter et al., 2004). In addition, reductions in carriage and evidence of herd immunity have been observed (Maiden et al., 2008).

The biggest challenge in controlling meningococcal disease has been in the development of a meningococcal serogroup B vaccine. This serogroup is responsible for the greatest number of cases of meningococcal disease in the UK, but its non-immunogenic and auto-antigenic capsular polysaccharides have prevented their use in vaccine design (Donnelly et al., 2010). The meningitis B vaccine Bexsero® (4CMenB, Novartis) has recently been recommended for routine use in the UK by the Joint Committee on Vaccination and Immunisation (JCVI). This vaccine contains factor H binding protein (fHBP), Neisserial adhesin A (NadA), Neisseria heparin-binding antigen (NHBA) and Outer Membrane Vesicle (OMV) (Serruto et al., 2012). A further meningitis B vaccine TRUMENBA® (Pfizer), containing two variants of fHBP, has also gained approval from the Food and Drug Administration (FDA) in the USA.

Furthermore, tetravalent conjugate vaccines against serogroups A, C, Y and W135 are also available in Europe, which include Menveo® (Novartis) and ACWY Vax® (GSK). Another tetravalent conjugate vaccine against serogroups A, C, Y and W135 (MCV4, Menactra®, Sanofi Pasteur) is available in the USA (Smith, 2008). This vaccine has also been licensed in Saudi Arabia since 2011 (Khalil et al., 2012). Finally, the MenAfriVac® (SynCo Bio) vaccine targeting meningococcal serogroup A, was introduced in 2010 in order to combat meningitis epidemics within the African Meningitis Belt (Frasch et al., 2012).

1.9.5 Vaccine Development Difficulties

Previously, problems associated with vaccines against respiratory organisms were due short-lived responses to polysaccharide vaccines as a result of the production of low-affinity antibodies and the lack of B cell memory (Siegrist, 2008). Furthermore, poor immunological responses to polysaccharide vaccines in young children is problematic (Siegrist, 2008). The inclusion of a conjugated protein within these vaccines has resolved this issue as well as preventing carriage of the targeted organism (Lipsitch, 1997). However, conjugate vaccines against *S. pneumoniae, N.meningitidis* serogroup C and *H. influenzae* type b target specific capsular polysaccharides, meaning that the remaining serogroups or serotypes are not targeted by the vaccines (Magee and Yother, 2001, Barbour et al., 1995). Furthermore, intra-species variation in *S. pneumoniae* polysaccharides limits the efficacy of vaccines targeted the organism. As a consequence, vaccination results in modifications in the diversity and structure of pneumococcal populations (Lipsitch et al., 2009, Pettigrew et al., 2008). Such diversity imposes selective pressures on the organism, inducing antigenic variation as a means of survival (Weiss-Salz and Yagupsky, 2010). Constant antigenic variation of colonising *S. pneumoniae* in the respiratory tract means that there is an on-going need to update and modify vaccines in order to 'keep up' with this genetic evolution, a process known as Red Queen dynamics (Jefferies et al., 2011).

1.10 Selection Pressures of Immunisation Programmes

Vaccines impose evolutionary pressures on the organisms that they target. Following the introduction of PCV-7, serotype replacement of vaccine types (VT) with non-vaccine types (NVT) has become apparent in cases of invasive disease in the USA and the UK (Hicks et al., 2007, Kaye et al., 2009). Furthermore, in the UK, increases in the number of cases of invasive disease caused by non PCV-13 serotypes have also be observed since the introduction of PCV-13 (Public Health England, 2015b). Serotypes colonising the respiratory tract have also undergone similar replacement, with 19A, 6A, 15B/C and 11A becoming more prevalent in the USA, UK and Canada (Kellner et al., 2008, Huang et al., 2009, Flasche et al., 2011, Tocheva et al., 2011). Increases in serotype 6C were also observed after introduction of PCV-7 (Tocheva et al., 2010). It has been noted, however, that in most countries serotype replacement has occurred to a greater extent in carriage than in invasive disease, perhaps due to lower invasive properties of the serotypes that have replaced the vaccines serotypes (Weinberger et

al., 2011). On the other hand, some replacement serotypes, such as 19A, have demonstrated higher invasiveness (Pilishvili et al., 2010).

The ability of respiratory tract organisms to evolve in response to vaccination is a worrying problem. The decline in *S. pneumoniae* vaccine types caused by vaccination has the potential to cause an increase in prevalence of potentially more virulent or antibiotic resistant non-vaccine types which were previously out-competed (Gladstone et al., 2011). Rates of invasive disease post PCV-7 introduction have shifted from vaccine types to non-vaccine types, in particular serotypes 3, 15, 19A, 22F and 33F (Hicks et al., 2007). The invasive potential of less prevalent serotypes could be problematic in the future, as more common serotypes are eliminated from this ecological niche. In particular, clonal replacement of ST199 with ST320 and ST695 within the 19A serotype was observed in the USA (Hanage et al., 2011). Such observations have led to the development of further vaccines covering these non-vaccine types, including PCV-13 which contains 19A, and PCV-15 which is undergoing pre-clinical assessment and covers serotypes 3, 19A, 22F and 33F in addition to those serotypes covered by PCV-13 (Skinner et al., 2011).

Diversification of particular strains has also occurred as a result of vaccination, with the discovery of novel serotypes, including pneumococcal serotypes 6C and 6D (McEllistrem and Nahm, 2012). Novel sequence types (ST) have also been reported, including ST4533, which caused a single case of life-threatening invasive illness (White et al., 2011).

As yet, no large-scale serotype shifts have been observed since the widespread use of the Hib vaccine, however a small number of shifts from *H. influenzae* type b to non-b and non-typeable *H. influenzae* (NTHi) have been reported in a number of countries (Adam et al., 2010, Bender et al., 2010). In the UK, *H. influenzae* type e and f invasive disease incidence has risen by 7.4% each year (0.03 cases per 100,000 in 2009/2010) and 11% over 10 years (0.09 cases per 100,000 in 2009/2010) respectively between 2001 and 2010 (Ladhani et al., 2012a). NTHi is now responsible for the majority of *H. influenzae* upper RTI with increases in NTHi-associated acute otitis media (AOM) (Wiertsema et al., 2011b). In the USA, NTHi-positive AOM samples have increased by 19% from 1995 to 2003 (Casey and Pichichero, 2004).

N. meningitidis capsular replacement from serogroup C to B following the meningitis C vaccine introduction has been observed in ST11 isolates collected from two outbreaks in Spain between 1997 and 2002 (Perez-Trallero et al., 2002). Similar replacement has

been observed in ST11 in the Czech Republic (Kriz et al., 1999) and Canada (Kertesz et al., 1998). Increases in serogroup Y and W135 have also been reported in the UK following the introduction of the meningitis C vaccine (Ladhani et al., 2012b, Maiden and Stuart, 2002). However, other studies have not identified shifts in these serogroups as a result of serogroup C vaccination, with a Canadian report demonstrating no change in levels of serogroup Y and W135 between 2002 and 2006 (Bettinger et al., 2009). Furthermore, clonal expansion of *N. meningitidis*, with the occurrence of novel STs, has been observed in the UK (Diggle and Clarke, 2005).

It is also thought that populations of co-colonising bacteria will be altered as a result of vaccination against specific organisms. For example, a relative increase in NTHi causing acute otitis media post PCV-7 implementation has been observed in the USA and Australia (Wiertsema et al., 2011b, Casey and Pichichero, 2004). However other reports are more inconclusive about changes in *H. influenzae* and *M. catarrhalis* following PCV-7 implementation, with studies showing no increase in these species (van Gils et al., 2011b) and others demonstrating increased levels of *H. influenzae* and *M. catarrhalis* in AOM following PCV-7 implementation (Leibovitz et al., 2004).

The alterations in populations of species and serotypes carried in the nasopharynx as a result of vaccination is thought to bring about an imbalance within this ecological niche (Weiss-Salz and Yagupsky, 2010). Although current replacement serotypes are generally found to be less invasive, as vaccine composition changes, there is the possibility that particular replacement serotypes may become more invasive. This has the potential to bring about an increase in the incidence of respiratory disease and the potential for epidemics and pandemics. Invasive non-b type *H. influenzae* types e and f have been described as well as invasive non-vaccine type pneumococci (Ladhani et al., 2012a, Scott et al., 2011). However, there is no evidence for this potential rise in invasiveness of less invasive serotypes. It is thought that *S. pneumoniae* serotypes achieve a new equilibrium state following PCV implementation which limits any potential further large changes in invasive serotypes (Hanage et al., 2010). The difficulty in predicting replacement and disease potential of particular serotypes means that surveillance studies, to include both surveillance of carriage and disease, are essential for monitoring any changes within the population in relation to the prevalence of disease (Weinberger et al., 2011).

1.11 Advances in Microbial Epidemiology Research Methods

1.11.1 Quantitative and Real-time PCR

Traditional microbiological identification methods rely on culture of microbes on specific growth medium, microscopy and serological assays (Mackay, 2004). These methods are limited with respect to slow-growing or non-culturable species, low sensitivity, biohazard risk, high background flora and poor interpretability (Whelen and Persing, 1996). Polymerase Chain Reaction (PCR) allows the amplification of small amounts of nucleic acid within a sample. Real-time PCR allows the quantification of nucleic acid within the sample, which can provide information on bacterial or viral load and gene expression levels (Klein, 2002). Assays have been developed for the detection of a number of bacterial and viral respiratory species, including *S. pneumoniae*, *N. meningitis, H. influenzae, M. catarrhalis, S. aureus,* parainfluenza viruses, adenovirus, influenza viruses, RSV, coronavirus, rhinovirus (Wu et al., 2013a, van de Pol et al., 2007, Greiner et al., 2003, Jonas et al., 2002, Luna et al., 2007, Do et al., 2010). These assays have allowed increased sensitivity for detection of these species within respiratory samples.

Real-time PCR is essential for treatment of tuberculosis as detection of the species can be undertaken in days rather than weeks which allows for optimal management of disease as well as prompt treatment (Brisson-Noel et al., 1991). Furthermore, in cases of atypical pneumonia, real-time PCR has proved essential in the identification of a number of species which are extremely difficult to culture, including *Legionella* species, *Mycoplasma pneumoniae* and *Chlamyida pneumoniae* (Campbell et al., 1992, Williamson et al., 1992, Matsiota-Bernard et al., 1994). Sensitivity for detection of *B. pertussis*, a fastidious and slow-growing species, has also been greatly enhanced by the use of real-time PCR (Muller et al., 1997).

The investigation of antibiotic susceptibility is also achieved using real-time PCR. Methicillin-resistant *S. aureus* strains can be successfully identified using *mecA* gene detection, which subsequently allows for appropriate infection management and treatment (Jonas et al., 2002). Antibiotic resistance in *M. tuberculosis* is highly problematic, with multiple drug resistant strains hampering efforts to control the disease (Morris et al., 1995). Real-time PCR assays to detect resistance to rifampicin, one of the first-line treatments for tuberculosis, have been developed in order to rapidly detect resistance and amend treatment regimens accordingly (Whelen et al., 1995).

Real-time PCR has also had a number of implications for understanding the epidemiology of respiratory tract microorganisms. Compared with culture detection, real-time PCR enables quantification of species with respiratory samples, which permits enhanced investigation of the associations between species abundance and host characteristics such as age and species interactions (Rodrigues et al., 2013). Further epidemiological methods which have been facilitated by the development of real-time PCR include microbial typing and genetics, which will be discussed in the following sections.

1.11.2 Microbial Typing

Traditionally, capsular typing of bacterial species was undertaken via serological techniques. The Quellung reaction for *S. pneumoniae* remains the gold standard for capsular typing of this species (Austrian, 1976). Slide agglutination of *H. influenzae* can also enable the discrimination of the different serotypes of this species, however interpretation of results is often challenging (Shively et al., 1981). Furthermore, slide agglutination of *N. meningitidis* also enables identification of sergroups of the species, but cross-reactivity and auto-agglutination are problematic (Craven et al., 1978).

More recently a number of different methods have been used which offer certain advantages over the more traditional serological assays in serotyping respiratory species, including batch processing of samples as well as direct detection of serotypes from clinical samples (Jauneikaite et al., 2015). Enzyme immunoassays (EIA) have been developed to detected specific serotypes or serogroups of a bacterial species using antigens. Discrimination of *S. pneumoniae* serotypes in nasopharyngeal samples by EIA was 100% sensitive (Lankinen et al., 2004). An enzyme-linked immunosorbent assay (ELISA) developed for typing *N. meningitidis* also enabled the detection of types and subtypes of the species (Abdillahi and Poolman, 1988).

Molecular methods offer additional advantages for typing of microbial species, including rapidity and reduced costs. PCR-based methods are widely used for serotyping or serogrouping species via amplification of a species-specific gene. PCR methods are capable of detecting up to 50 specific serotypes of *S. pneumoniae*, and include multiplex PCR (Pai et al., 2006) and nanofluidic real-time PCR (Dhoubhadel et al., 2014) methods. Serotyping of *H. influenzae* types a-f and identification of non-typeable *H. influenzae* is also undertaken using single PCR reactions (Falla et al.,

1994) and multiplex PCR (Cote et al., 1994). Finally serogroup specific PCR of *N. meningitidis* was found to be highly specific and sensitive for detecting serogroups of the species, and reported fewer non-typeable strains when compared with slide agglutination (Mothershed et al., 2004).

Microarrays, which detect nucleic acid via probes bound to a microarray chip surface, can enable the detection of multiple serotypes of *S. pneumoniae* within single samples (Tomita et al., 2011, Turner et al., 2011a). Finally, whole genome sequencing and *in silico* analysis, which will be discussed in the following section, allows increased sensitivity and resolution for the detection of *S. pneumoniae* serotypes (Gladstone, 2013).

In addition to capsular typing to determine serotypes or serogroups, typing of multiple genetic loci can enable species to be further characterised. Pulse-field gel electrophoresis (PFGE) is a commonly used method during outbreak situations due to its discriminatory power and reproducibility (Sabat et al., 2013). Ribotyping and amplified fragment length polymorphism (AFLP) are addition al methods which involve, like PFGE, the discrimination of types by using restriction enzymes (Vos et al., 1995, Syrogiannopoulos et al., 2001). Although useful in outbreak situations, these methods map rapid evolution and can be deceptive when studying circulating strains in long-term epidemiology (Maiden et al., 1998). The first typing method to overcome this limitation was multilocus enzyme electrophoresis (MLEE) which enables the long-term epidemiology of species to be assessed (Selander et al., 1986). However, as with all restriction enzyme methods, inter-laboratory comparison of MLEE results is a challenge (Maiden et al., 1998). Multilocus sequence typing (MLST), which involves sequencing seven or eight housekeeping genes of a specific isolate and assigning a sequence type (ST) to the isolate depending on the allelic profile (Maiden et al., 1998), has been used to understand the epidemiology of *S. pneumoniae* (Enright and Spratt, 1998), *H. influenzae* (Meats et al., 2003) and *N. meningitidis* (Maiden et al., 1998). Compared to the previous methods discussed, MLST is highly portable which has enabled inter-laboratory comparisons and exchange of molecular data (Maiden et al., 1998).

1.11.3 Genome Sequencing

One of the largest technological advances within the field of infectious disease epidemiology is the development of whole genome sequencing technologies. Sanger

sequencing, which involves the incorporation of chain-terminating inhibitors and sequence determination via separation of DNA fragments using gel electrophoresis, was the first of the sequencing methods to be developed (Sanger et al., 1977). This method was lengthy, expensive and labour-intensive (Metzker, 2005). As a result a number of new technologies have been developed, which incorporate a range of template preparation and nucleic acid sequencing methods (Metzker, 2010). Template preparation involves the generation of good quality non-biased nucleic acid and can be undertaken via clonal amplification of templates on a bead surface (emulsion PCR) (Williams et al., 2006) or on a solid surface (solid-phase amplification) (Fedurco et al., 2006) or using single molecule templates (Harris et al., 2008). Methods for sequencing these nucleic acid templates involve cyclic reversible termination (Bentley et al., 2008), pyrosequencing (Ronaghi et al., 1998) and real-time sequencing (Eid et al., 2009). The development of these next-generation sequencing technologies has a number of applications in the field of microbial epidemiology. Metagenomics, the analysis of sequences of microbial DNA which have been extracted directly from a sample, enables the study of microbial community composition and greater understanding of the contribution of non-culturable species (Singh et al., 2009). Microbial communities such as those of the nasopharynx (Bogaert et al., 2011), the gut (Dethlefsen et al., 2008) and the lung (Rogers et al., 2004) have been analysed in order to understand species interactions and the effect of interventions such as antibiotics on community structure.

With the decreasing cost and increasing use of sequencing technologies, there is the need for more advanced bioinformatics analysis tools to handle the data being generated. The main challenge involved the volume of data and the varied types of data being produced by these technologies (Pop and Salzberg, 2008). Assembly of sequences, which involves the assembly of short reads into longer sequences (scaffolds and contigs) using complex algorithms, is particularly challenging due to variations in read length, paired-end vs. single-end and genome size (Zhang et al., 2011). A wide variety of assemblers have been developed, including Velvet (Zerbino and Birney, 2008), AbySS (Simpson et al., 2009) and SSAKE (Warren et al., 2007), to name a few, which vary in their ability to produce good quality sequences from different types of sequencing data. Annotation of sequences involves assigning biological meaning and function to the sequences of DNA within a species. Annotation can be done manually, however, the volume of data generated using next-generation sequencing requires a more automated approach (Rust et al., 2002). Automated annotation methods include gene prediction using software such as GLIMMER (Delcher et al., 1999) followed by sequence alignment using BLAST (Altschul et al.,

1990) and FASTA (Pearson, 1990).Further annotation methods include Rapid Annotation using Subsystems Technology (RAST) which predicts sets of functional roles represented in the genome which have been manually curated (Aziz et al., 2008). Despite attempts to fully automate the annotation process, the methods are not without their flaws and some manual corrections are often required to produce good quality annotations (Richardson and Watson, 2013).

1.11.4 Statistical Methods in Cross-sectional Studies

Statistical methods are essential for the design and sample size calculations within research studies as well as for evaluating the outcomes of the study (Sprent, 2003). Within research studies, a representative sample of a population is used in order to infer associations and patterns within the general population. This sample is ideally randomly chosen and unbiased (Kadam and Bhalerao, 2010). Stratified random sampling can also be performed when subpopulations differ within the population being sampled (Cochran, 1946). Sample size is based on the underlying prevalence within the population, level of significance and power (Kirby et al., 2002). However, researchers also consider response rates, attrition and study objectives when considering an appropriate sample size (Larsen et al., 1985). The sample size of a study is important for obtaining high quality results which can easily be interpreted and reported transparently (von Elm et al., 2007).

Statistical analysis of cross-sectional study outcomes is most commonly performed using logistic regression (Barros and Hirakata, 2003). This method predicts binary outcomes based on multiple variables and allows the estimation of an odds ratio (OR) (Cox, 1958). However, the OR can overestimate the prevalence ratio (PR) when the outcome being observed in common (Thompson et al., 1998). Cox regression, Poisson regression and log-binomial have been suggested as alternatives to binary logistic regression (Skov et al., 1998, Traissac et al., 1999). These methods are thought to produce easily interpretable prevalence ratios (Barros and Hirakata, 2003). However, others argue that these methods cannot adequately estimate the PR whilst adjusting for covariates (Stromberg, 1995).

1.12 Aims and Objectives

Microorganisms carried in the URT are capable of causing respiratory infection, meningitis or sepsis and pose a significant burden in terms of morbidity and mortality in the UK and elsewhere. Vaccines targeting these microbes remain the most effective intervention against respiratory infection, meningitis and sepsis. It is therefore crucial to understand the carriage of microbes within the population in order to inform vaccine and antibiotic development and policy.

The respiratory tract is a complex ecosystem with multiple species and strains co-inhabiting the same niche. This diversity has led to concern that the introduction of bacterial conjugate vaccines targeting specific strains will lead to other bacterial strains or species colonising the niche which has been vacated by vaccine-targeted strains. This may increase the prevalence of pathogens not targeted by vaccination and may potentially reduce the effectiveness of such vaccines. With viral epidemics occurring frequently, greater knowledge is required about the effects of viral pathogens on the niche dynamics of the upper respiratory tract. It is well known that Influenza viruses have caused significant bacterial secondary infections in the past which has led to enhanced morbidity and mortality. Understanding such interactions is key to determining the dynamics of the respiratory tract and understanding the true diversity within this niche.

The primary aim of the project was to gain insight into the epidemiology and ecology of respiratory tract communities. Secondary study aims were as follows:

1. Understand participation of members of the general population to a large community-based swabbing study.
2. Determine carriage prevalence of target bacterial and viral species and build a mathematical model for microbial respiratory tract carriage.
3. Gain insight into the inter-species and intra-species microbial diversity of the upper respiratory tract community.
4. Determine species distribution and community assembly patterns within the respiratory tract community.

In order to achieve these aims, traditional culture-based techniques, molecular methods and ecological and mathematical modelling methods were used. Data collected from a large community-based respiratory carriage study would be used, which would provide key epidemiological information for the evaluation of vaccine and antibiotic design and policy. Furthermore, data would provide crucial information on

participation and carriage rates, which would inform future carriage study design, including larger multi-centre carriage studies.

Chapter 2. Materials and Methods

2.1 Participant Recruitment and Sample Collection

2.1.1 Study Rationale

A large community-based swabbing study was undertaken in order to understand the epidemiology of bacterial carriage within the upper respiratory tract across all age groups. Previous studies on bacterial carriage have focused on fewer species within specific age groups. This study, however, aimed to gain a better understanding of bacterial carriage across all age groups, with the hope of producing results which would inform future improvements in vaccine and antibiotic development and policy. The main objectives of the swabbing study were to compare different swabbing techniques for optimum detection of bacterial species, by comparing swab type (nose, whole mouth swab and nasopharyngeal) and swabbing method (self-swabbing and healthcare professional swabbing). Furthermore the study aimed to determine levels of participation by members of the general population and prevalence of bacterial carriage to help inform sample size calculations for future larger multi-centre carriage studies. The swabbing study would also assess all features of a future larger study, including participant recruitment and study documentation.

The swabbing study was funded by The Bupa Foundation and was awarded to the Principal Investigator, Dr Stuart Clarke and study co-investigators. Ethical Approval was obtained from the National Research Ethics Service (NRES) Committee South Central Hampshire A (REC, 11/SC/0518).

2.1.2 Duration and Timing of Swabbing Time-points

Swab samples were collected during two study time-points, May to August 2012 and February to April 2013. Swab samples were processed between 18th May 2012 to 31st August 2012 and 15th February 2013 to 15th April 2013.

2.1.3 Participating General Practitioner Practices

Twenty GP practices across Hampshire were invited to participate in the swabbing study. The practices were chosen to reflect a mix of urban and rural locations as well as deprivation levels. The practices formed part of the South West (East Hub) Primary Care Research Network (PCRN). Practices were located within NHS Southampton, Solent NHS Trust, NHS Hampshire and NHS Portsmouth. The locations of these practices are shown in Figure 1.

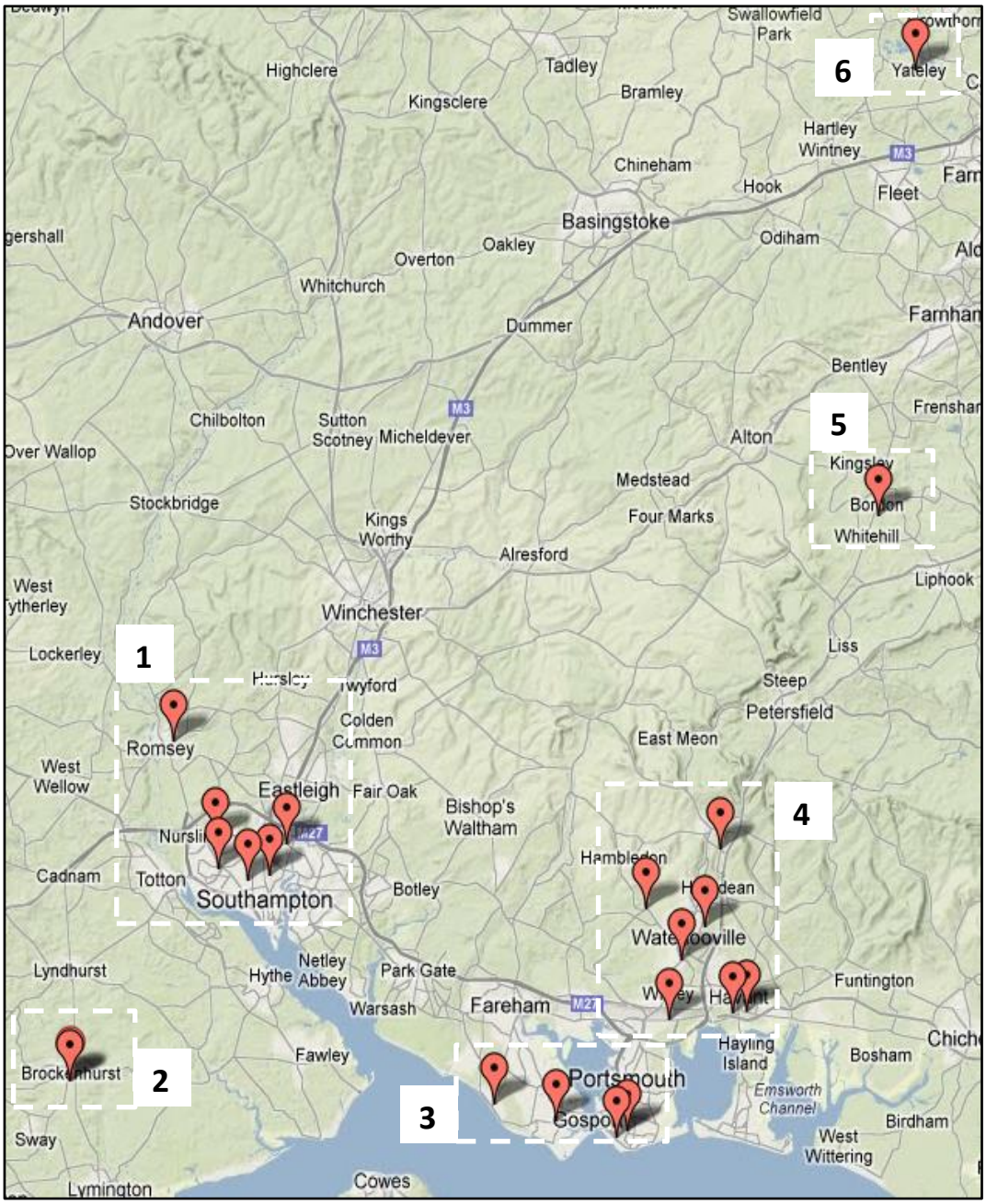


Figure 1. Location of the 20 Participating GP Practices

Red markers show the location of each of the 20 GP practices. The map represents the South Central area of the United Kingdom. GP practices have been grouped into geographical areas for analysis purposes: 1 = Southampton/Romsey, 2 = New Forest, 3 = Portsmouth, 4 = Outer Portsmouth, 5 = Borden and 6 = Yateley. The map was generated using Google Maps.

2.1.4 Sample Size

Sample size was based on the precision with which we can estimate true carriage rates. A 25% response rate was assumed based on results from a previous staphylococcal carriage study (Gamblin et al., 2013). A total of 2,020 patients aged 0-4 years and 3,200 patients aged 5 years and older were invited to take part within two separate swabbing groups, anticipating 505 patients aged 0-4 years and 800 patients aged 5 years and older to respond within each group, accounting for lower carriage rates observed within the older age group. A predicted carriage rate of *S. pneumoniae* of 30% in 505 participants aged 0-4 years would enable true carriage to be determined to within ±4.0% (95% confidence) (Tocheva et al., 2011). A predicted carriage rate of *S. pneumoniae* of 20% in 800 participants aged 5 years and older would enable true carriage to be determined to within ±2.8% (95% confidence) (Watt et al., 2004).

2.1.5 Participant Recruitment

Each GP practice generated a list of their entire patient cohort. Patients deemed unfit for participation by their GP were removed from the list. The list was then divided into two: one containing all patients aged 0-4 years and one containing all patients aged 5 years and older. Each practice list was then randomised using the *ralloc* command in Stata (StataCorp, version 12), allocating a random number between 1 and 5 to each patient (this was undertaken by Ho Ming Yuen). Approximately 580 eligible participants per practice, 220 patients aged 0-4 years and 360 aged 5 years or older, were randomly selected from each GP list using the random number allocated to each participant. Each of the two age groups was then split equally between two study groups: Healthcare professional (HCP) swabbing and self-swabbing. For the second time-point, only the self-swabbing group was recruited (from a separate set of individuals to the first time-point), with the random selection of 290 eligible participants per practice during this time-point. Participants from the second time-point were selected using a different random number to the first time-point to ensure all participants were different to the first time-point. The overall structure of the study is shown in Figure 2.

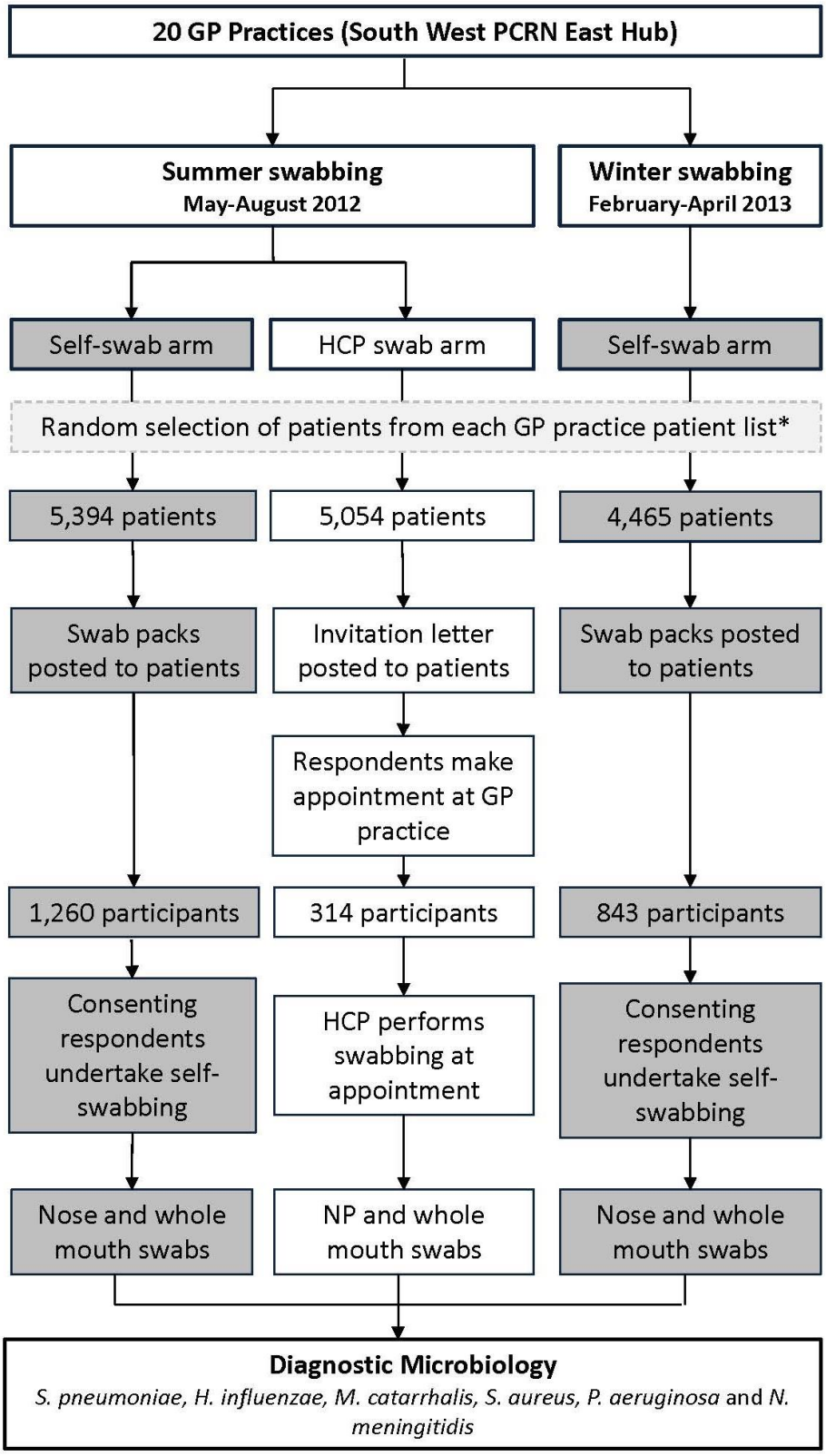


Figure 2. Structure of the Swabbing Study

GP = general practitioner, PCRN = primary care research network, HCP = healthcare professional, NP = nasopharyngeal. *Random selection of patients using random number generation, allowing mutually exclusive groups to be selected for summer self-swabbing, summer HCP swabbing and winter self-swabbing.

2.1.6 Study Groups

Participants were split equally between the two study groups, as mentioned previously. Participants in the HCP swabbing study group were invited via letter to visit their GP practice to have their swabs taken by a registered HCP. At the swabbing appointment, patients were asked to read the study information sheet and sign the consent form if they agreed to participate in the study. Consenting patients then had a nasopharyngeal (NP) swab and a whole mouth swab (WMS) taken by the HCP. Each participant was also asked to complete a patient questionnaire at their appointment. Swabs were returned to the Infectious Disease Epidemiology Group laboratory in the Faculty of Medicine at the University of Southampton using pre-existing NHS transport arrangement, if available, or via taxi on the day of swabbing.

All participants in the self-swabbing study group were sent self-swabbing packs by Danvers International (London, UK), a medical courier company. These packs contained an age-appropriate participant information sheet, consent form, instruction sheet, nose swab and WMS and patient questionnaire. Participants were invited to take their own nose swab and WMS if they consented to participate in the study and return them, along with consent forms and questionnaires, via first class freepost return to the Infectious Disease Epidemiology Group laboratory in the Faculty of Medicine at the University of Southampton. All swab packs were assembled and mailed to participants by Danvers International.

Questionnaires were designed to be short (one page A4) and easy to complete in order to maximise the numbers of responses. The questionnaires, shown in Appendix 1, were identical in both study groups and gathered the following information pertinent to bacterial carriage:

- Participant age (years)
- Location (GP practice attended)
- Swabbing group (HCP swabbing or self-swabbing)
- Use of antibiotics (yes, no or do not know) and type used (amoxil, co-amoxiclav, erythromycin, azithromycin or other) within the past calendar month (recent).
- Respiratory tract infection (yes, no) and type of infection (cold, flu, ear infection or other chest infection) within the past calendar month (recent).
- Vaccination status (up-to-date, not up-to-date or do not know).

Other factors, including smoking, family size and structure, although considered important were not included so as to ensure the questionnaire was not too long and easy to complete. Furthermore, the study was a pilot study which aimed primarily to compare self-swabbing and HCP swabbing.

2.1.7 Swabbing

Swab heads were composed of viscose (rayon) and swab shafts were composed of polystyrene (nose and whole mouth swabs) or aluminium (NP swabs). Self-swabbing participants were instructed on how to self-perform nose and whole mouth swabbing via a set of instructions within their swabbing packs, as follows. After hand washing participants were asked to carefully remove a swab from its packaging. The swab was to be inserted approximately 1cm into the first nostril, and rubbed firmly against the inside of the nostril, withdrawn and the process repeated for the second nostril. A second sterile swab was then inserted into the mouth and firmly rubbed against the inside of both cheeks. Healthcare professionals were asked to take NP and whole mouth swabs according to standard sterile technique. Once taken, swabs were placed in polypropylene tubes, containing amies transport medium with charcoal and labelled according to swab type using sticky labels provided.

2.2 Sample Processing

HCP-taken swabs were returned for analysis on the day of swabbing by taxi or within 1-2 days by pre-existing NHS delivery service. Self-taken swabs were returned by first-class freepost return directly to the University laboratory (1-2 days). Once received in the laboratory swabs, questionnaires, consent forms and agar plates were labelled with a unique participant identification number for processing.

Swabs were removed from the polypropylene tubes containing transport medium and inoculated into 1ml of STGG (skim milk, tryptone, glucose, and glycerine, HPA) by rubbing the swab vigorously against the side of the tube and vortexing to ensure transfer of bacteria into the tube. Ten microliters of the inoculated STGG was then streaked out onto the following agar plates: Columbia Blood Agar with Horse Blood (CBA, Oxoid, PB0124), Columbia Blood Agar with Colisitin and Naladixic Acid (CNA, Oxoid, PB0308), Columbia Agar with Chocolated Horse Blood (CHOC, Oxoid, PB0124), Columbia Agar with Chocolated Horse Blood and Bacitracin (BAC, Oxoid, PB0220), *Pseudomonas*-selective

Agar (CFC, Oxoid, PB0291) and Lysed GC Selective Agar (GC, Oxoid, PB0962). Discs containing 5µg optochin (Oxoid, DD0001) were then placed on the CNA plates to determine sensitivity to optochin. CBA, CNA, CHOC, BAC and GC plates were incubated at 37°C with 5% CO_2 for 24 hours. CFC plates were incubated at 37°C for 24 hours. The STGG vials containing the swab contents were then frozen at -80°C for further analysis.

2.3 Identification of Respiratory Bacteria using Culture

2.3.1 Microbial Identification

Swabs were analysed for the presence of the six following microorganisms: *S. pneumoniae, H. influenzae, M. catarrhalis, S. aureus, N. meningitidis* and *P. aeruginosa*. After 24 hours, plates were read for the presence of the six target organisms and standard microbiological identification techniques were used, as shown in Table 1. A sweep of each CHOC plate was also inoculated into STGG and frozen at -80°C.

Table 1. Standard Microbiological Identification Techniques for *S. pneumoniae, H. influenzae, M. catarrhalis, P. aeruginosa, S. aureus* and *N. meningitidis*

Bacterial Species	Identification Method
S. pneumoniae	Characteristic optochin-sensitive colonies on CNA
H. influenzae	Characteristic colonies requiring X+V on BAC
M. catarrhalis	Gram-negative oxidase-positive, tributyrin-positive and DNase-positive diplococci on CHOC, CBA or CNA
P. aeruginosa	Characteristic green oxidase-positive colonies on CFC
S. aureus	Characteristic coagulase-positive colonies on CNA
N. meningitidis	Characteristic oxidase-positive Gram-negative diplococci on GC with the correct API NH profile

CBA = Columbia Blood Agar, CNA = CBA with Colisitin and Naladixic Acid, BAC = CBA with Chocolated Horse Blood and Bacitracin, CHOC = CBA with Chocolated Horse Blood, CFC = *Pseudomonas*-selective agar, GC = Lysed GC Selective Agar

X and V testing of *H. influenzae* was undertaken by inoculating the surface of a Blood Agar Base (Oxoid, PO1046) plate with a suspension of suspected *H. influenzae* culture in sterile distilled water. X (Oxoid, DD0003), V (Oxoid, DD0004) and X+V (Oxoid, DD0005) discs were placed on the plate at equal distances from each other. The plate was then incubated at 37°C with 5% CO_2 for 24 hours. Colonies growing around the X+V disc alone were identified as *H. influenzae*.

Oxidase testing of *M. catarrhalis*, *P. aeruginosa* and *N. meningitis* was undertaken by transferring a small amount of suspect bacterial material onto an oxidase strip (Mast Diagnostics, ET04). Oxidase-positive colonies turned the oxidase strip a dark purple colour.

Coagulase activity testing of *S. aureus* was undertaken using a Pastorex Staph Plus Kit (Bio-Rad, 56353). A small amount of bacterial material was mixed with the test reagent (containing monoclonal antibodies) and mixed. The presence of an agglutination reaction was indicative of a coagulase-positive colony.

Tributyrin testing of *M. catarrhalis* was undertaken by placing a tributyrin tablet (Rosco Diagnostica, 48821) into a 5ml turbid suspension of each suspect colony in saline. Tributyrin–positive colonies turned yellow after incubation at 37°C with 5% CO_2 for 24 hours.

DNase testing of *M. catarrhalis* was undertaken by streaking a small amount of bacterial material onto a DNase methyl green agar plate (Oxoid, PO1000A). Plates were incubated at 37°C for 24 hours and then flooded with 1M Hydrochloric Acid (HCl). DNase-positive colonies were identified by the presence of clear zones immediately around the colonies.

API NH testing of *N. meningitidis* was undertaken by inoculating an API NH strip (Biomerieux, 10400) with a 5ml turbid suspension of each suspect colony in saline. After incubation at 37°C for 24 hours, results were recorded and submitted to the apiweb (https://apiweb.biomerieux.com/) to determine the identity of the colony in question.

After identification, isolates were re-streaked onto a purity plate and re-incubated for 24h. Each of the pure isolates were then inoculated into 1ml STGG and frozen at -80°C for further analysis. Results were recorded on the laboratory sample sheet, shown in Appendix 2, as presence (1) or absence (0). No colony counting was performed directly from culture samples.

2.3.2 Statistical Analysis

Microbiology culture identification results as well as information gathered from the participant questionnaire were entered into SPSS (IBM, version 20). Participation of individuals invited as well as carriage and co-carriage of bacterial species were then assessed.

2.3.2.1 Study Participation

Participation was calculated as a percentage of total individuals invited for each study group (self-swabbing and HCP swabbing), swabbing time-point, GP practice and age group (0-4 years and ≥5 years) in order to gauge the participation of different types of individuals. This was calculated as follows:

$$Participation\ (\%) = \frac{N°\ of\ participants}{Total\ n°\ of\ individuals\ invited} \times 100$$

Participation was also assessed according to population density and deprivation. Deprivation was assessed via the UK Indices of Multiple Deprivation (IMD) 2010 score of each participant's GP practice. Deprivation is a measure of poverty and is calculated based on income, crime levels, living conditions, education and training, employment levels, health and barriers to house and services. An IMD Score and a rank IMD score were obtained for each practice. Population density (individuals per km^2) was assessed via Office of National Statistics (ONS) Local Area classification codes of each participants practice.

2.3.2.2 Prevalence of Bacterial Carriage and Co-carriage

Overall prevalence of bacterial species carriage and co-carriage within different swab types was calculated and compared using the following statistical tests. Chi-squared and fisher's exact tests were used to compare samples from different individuals (NP versus nose, NP versus self-taken WMS, self-taken WMS versus HCP-taken WMS, nose versus HCP taken WMS). McNemar's Chi-squared test was used to compare samples from the

same individuals (nose versus self-taken WMS and NP versus HCP-taken WMS). The formulas for these tests are shown below:

Fisher's Exact Test:

$$P_{FET} = \frac{\binom{A+C}{A}\binom{B+D}{B}}{\binom{N}{A+B}} = \frac{(A+B)!\,(C+D)!\,(A+C)!\,(B+D)!}{A!\;B!\;C!\;D!\;N!}$$

A = number of carriage-positive variable-positive individuals
B = number of carriage-negative variable-positive individuals
C = number of carriage-positive variable-negative individuals
D = number of carriage-negative variable-negative individuals
N = A+B+C+D

Chi-squared Test:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)}{E_i}$$

O_i = number of observations
E_i = expected frequency
n = number of cells in the table

McNemar's Chi-squared Test:

$$X^2 = \frac{(b-c)^2}{b+c}$$

b = test 1 positive, test 2 negative
c = test 1 negative, test 2 positive

Binary logistic regression analysis was undertaken using SPSS (version 21, IBM) in order to determine the effect of demographic variables on carriage and co-carriage of these bacterial species. Models were constructed for each bacterial species as well as co-carriage of two or more bacterial species. Univariate regression analysis was used to pre-select variables to be included in the final model (alpha = 0.10). Multivariable binomial logistic regression was then performed using these pre-selected variables followed by backward elimination of variables that did not meet the required level of significance within the final model (alpha = 0.05). Categorical variables were set as 'simple contrast' with the first category or 'no' category as the reference. 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 demonstrate that the model predicts carriage significantly better than chance.

Variables assessed for inclusion via univariate analyses included age (in years), recent RTI (within a month prior to swabbing), geographical location (six geographical areas detailed in Figure 1), recent use of antibiotics (within a month prior to swabbing), vaccinations up-to-date (with the national immunisation schedule), season (winter or summer swabbing) and presence of other microbial species within the same sample.

2.4 Identification of Respiratory Bacteria and Viruses using Real-time PCR

Real-time PCR work was undertaken in collaboration with the Public Health England (PHE) Southampton laboratory. All validation of real-time PCR assays was undertaken by the PHE laboratory, prior to analysis of these samples.

2.4.1 Selection of Samples

A sample size of 380 samples was chosen to be analysed by real-time PCR due to the predicted overall prevalence of bacterial carriage in nose and NP swabs of approximately 30% (as determined by culture). This would enable carriage to be determined to within ±4.6% with 95% confidence. Equal numbers of nose and nasopharyngeal swabs (190 of each) were analysed in order to be able to undertake comprehensive analysis of both datasets. The aim was for each swab set to be composed of equal numbers of the four age groups: 0-4, 5-17, 18-64 and 65+ years. If this was not possible, for example there were only 22 NP swabs available in the 5-17 years age group, numbers would be made up from the other age categories. It was also aimed to obtain equal sample numbers from each of the 20 GP practices. If this was not possible, for example due to practices not participating in one of the swabbing time-point or study groups, numbers would be made up from the other GP practices. The selection of samples was non-random; this is not the optimal method for sample selection. However, as the full set of samples was shaped by study participation (response rates), it was thought that random selection from this sample may not the best way of obtaining a diverse sample set.

2.4.2 Nucleic Acid Extraction

Nucleic acid extraction of STGG samples was undertaken using the QIAxtractor automated nucleic acid purification instrument (Qiagen) and QIAamp Viral RNA Mini kits (Qiagen, 52906). The vaccum protocol was used on the QIAxtractor, as follows. Firstly, 560µl of Buffer AVL containing carrier RNA was added to 200µl of each STGG sample before being mixed six times and incubated at room temperature for 10 minutes. Then 560µl of 100% ethanol was added to each sample and mixed six times before incubation for 30 seconds. 1,300µl of each sample was then loaded onto the capture plate and vacuumed for 5 minutes at 25kPa. 750µl of Buffer AW1 was then added to the samples and vacuumed for 5 minutes at 25kPa. Then, 750µl of Buffer AW2 was then added to the samples and vacuumed for 3 minutes at 40kPa then 5 minutes at 25kPa. Finally, 120µl of Buffer AVE was added to the tubes, which were then incubated for 5 minutes at room temperature and vacuumed for 1 minute at 35kPa.

2.4.3 Real-time PCR

2.4.3.1 Primers and Probes

Primers and probes used in the real-time PCR reaction are listed in Table 2 (bacterial) and Table 3 (viral). These were made up to stock concentrations of 100µM in Ultra High Quality (UHQ) water or as recommended by the manufacturer.

Table 2. Primers and Probes used in the Bacterial Real-time PCR Reactions

Species	Target gene	Sequence (5' to 3')	Performance/Limit of Detection	Reference
S. aureus	*femB*-F	GAC ATT TGA TAG TCA ACG TAA ACG TAA TAT T	100% specificity 100% sensitivity	(Saeed et al., 2010) (Jonas et al., 2002)
	femB-R	GCT CTT CAG TTT CAC GAT ATA AAT CTA AGA		
	femB-Pro	FAM-TCA TCA CGT TCA AGG AAT CTG ACT TTA ACA CCA TAG T-TAMRA		
M. catarrhalis	*copB*-F	GTG AGT GCC GCT TTT ACA ACC	91% specificity 100% sensitivity	(Greiner et al., 2003)
	copB-R	TGT ATC GCC TGC CAA GCA A		
	copB-Pro	FAM-TGC TTT TGC AGC TGT TAG CCA GCC TAA-TAMRA		
H. influenzae	*hel*-F	CCG GGT GCG GTA GAA TTT AAT AA	6 CFU per 100ml*	(Rogers et al., 2014)
	hel-R	CTG ATT TTT CAG TGC TGT CTT TGC		
	hel-Pro	FAM-ACA GCC ACA ACG GTA AAG TGT TCT ACG-TAMRA		
S. pneumoniae	*lytA*-F	ACG CAA TCT AGC AGA TGA AGC A	97.1% specificity 97.5% sensitivity	(Wu et al., 2013a)
	l*ytA*-R	TCG TGC GTT TTA ATT CCA GCT		
	l*ytA*-Pro	FAM-TGC CGA AAA CGC TTG ATA CAG GGA G-TAMRA		
P. aeruginosa	*regA*-F	TGC TGG TGG CAC AGG ACA T	15.5 gene copies per 100ml	(Shannon et al., 2007)
	regA-R	TTG TTG GTG CAG TTC CTC ATT G		
	regA-Pro	FAM-CAG ATG CTT TGC CTC AA-TAMRA		
N. meningitidis	*ctrA*-F	GCT GCG GTA GGT GGT TCA A	100% specificity 88.4% sensitivity	(Corless et al., 2001)
	ctrA-R	TTG TCG CGG ATT TGC AACT A		
	ctrA-Pro	FAM-CAT TGC CAC GTG TCA GCT GCA CAT-TAMRA		

Forward primer (-F), reverse primer (-R) and probe (-Pro) sequences. *fem*B = aminoacyltransferase gene; *cop*B = outer membrane protein gene; *hel* = lipoprotein e (P4) gene; *lyt*A = autolysin gene; *reg*A = toxin A synthesis regulating gene; *ctr*A = capsular transport gene. FAM = 6-carboxyfluorescein (fluorophore at 5' end of probe); TAMRA = tetramethylrhodamine (quencher at 3' end of probe). CFU = colony-forming units.
*Unpublished information provided by PHE Southampton Laboratory.

Table 3. Primers and Probes used in the Viral Real-time PCR Reactions

Species (Target gene)	Name	Sequence (5' to 3')	Limit of Detection	Reference
Influenza A virus (M)	AM-F AM-R AM2-Pro	GAG TCT TCT AAC MGA GGT CGA AAC GTA GGG CAC GGT GAG CGT RAA FAM-CCC CTC AAA GCC GA-MGBNFQ	Not provided	(Wu et al., 2013b) (Ellis and Curran, 2011)
Influenza B virus (HA)	InfB-F InfB-R InfB-Pro	AAA TAC GGT GGA TTA AAT AAA AGC AA CCA GCA ATA GCT CCG AAG AAA VIC-CAC CCA TAT TGG GCA ATT TCC TAT GGC-TAMRA	11 copies RNA 0.006 TCID50	(van Elden et al., 2001)
RSV A (N)	RSVA-F RSVA-R RSVA-Pro	GCT CTT AGC AAA GTC AAG TTG AAT GA TGC TCC GTT GGA TGG TGT ATT 6FAM-ACA CTC AAC AAA GAT CAA CTT CTG TCA TCC AGC-NFQ TaqMan®	2 copies RNA 0.023 PFU	(Hu et al., 2003)
RSV B (N)	RSVB-F RSVB-R RSVB-Pro	GAT GGC TCT TAG CAA AGT CAA GTT AA TGT CAA TAT TAT CTC CTG TAC TAC GTT GAA 6FAM-TGA TAC ATT AAA TAA GGA TCA GCT GCT GTC ATC CA-NFQ TaqMan®	9 copies RNA 0.018 PFU	(Hu et al., 2003)
MPV (L)	MPV-F MPV-R MPV-Pro	CAT ATA AGC ATG CTA TAT TAA AAG AGT CTC CCT ATT TCT GCA GCA TAT TTG TAA TCA G VIC-TGY AAT GAT GAG GGT GTCACT GCG GTT G-NFQ TaqMan®	5-10 copies RNA 0.006-0.01 TCID50	(Maertzdorf et al., 2004)
Adenovirus (Hexon)	AD-F AD-R AD-Pro	GCC CCA GTG GTC TTA CAT GCA CAT C GCC ACG GTG GGG TTT CTA AAC TT FAM-TGC ACC AGA CCC GGG CTC AGG TAC TCC GA-TAMRA	15 copies DNA	(Heim et al., 2003)
BMV (RNA2)	BMV-F BMV-R BMV-Pro	GTT CAC CGA TAG ACC GCT G AAG AGC CCG GAA TGT CAA GA CY5-CCT CAA GCT GAA ATG GCA CGG ATG-BHQ2	Not provided	(Ferns and Garson, 2006)

RSV = Respiratory syncytial virus; MPV = metapneumovirus; BMV = Brome Mosaic Virus; Forward primer (-F), reverse primer (-R) and probe (-Pro) sequences. FAM = 6-carboxyfluorescein; MGB = minor groove binding; NFQ = non-fluorescent quencher; VIC = fluorophore; TAMRA = tetramethylrhodamine (quencher); CY5 = cyanine 5 fluorescent dye; BHQ = black hole quencher; PFU = plaque-forming unit; TCID50 = 50% tissue culture infective dose; RNA = ribonucleic acid; DNA = deoxyribonucleic acid; M = matrix protein; HA = haemagglutinin; N = nucleocapsid; L = large polymerase.

2.4.3.2 Commercial PCR Kits

Commercial PCR kits were used for the detection of Coronavirus and Parainfluenza Virus (71-045, Argene) as well as Enterovirus/Rhinovirus (i.e. targeting the Enterovirus genus) (71-042, Argene). The limits of detection, as described in the product information sheets of these kits, are shown in Table 4 below.

Table 4. Limits of Detection of the Commercial PCR Kits

Kit Number	Species	Limit of Detection (TCID/50ml)
71-045	Coronavirus 229E	0.20
	Coronavirus OC43	86.85
	Coronavirus NL63	0.002
	Coronavirus HKU1	18.07*
	Parainfluenza virus 1	379.79
	Parainfluenza virus 2	529.93
	Parainfluenza virus 3	245.28
	Parainfluenza virus 4	0.21
71-042	Human Rhinovirus B 14	0.43
	Human Rhinovirus A 1B	3.30
	Echovirus 25	37.96

*copies/ml, TCID = tissue culture infective dose

2.4.3.3 Working Solutions and Master Mix Preparation

Working stock solutions of the primers and probes were made up prior to PCR. For the bacterial monoplex reactions these consisted of 21µl of 100µM forward primer, 21µl of 100µM reverse primer, 12µl of 100µM probe and 1,694µl of UHQ water. For the Influenza A-B and BMV reaction the working stock consisted of 20µl of 100µM AM-F, 40µl of 100µM AM-R, 8µl of 100µM AM2-Pro, 12µl of 100µM InfB-F, 12µl of 100µM InfB-R, 4µl of 100µM InfB-Pro, 10µl of 100µM BMV, 10µl of 100µM BMV-F, 10µl of 100µM BMV-R, 2µl of 100µM BMV-Pro, 10µl ROX and 1,212µl UHQ water. For the RSVA-B and MPV reaction the working stock consisted of 15µl of 100µM RSVA-F, 15µl of 100µM RSVA-R, 2.5µl of

100µM RSVA-Pro, 15µl of 100µM RSVB-F, 15µl of 100µM RSVB-R, 2.5µl of 100µM RSVB-Pro, 15µl of 100µM MPV-F, 15µl of 100µM MPV-R, 5µl of 100µM MPV-Pro, 10µl ROX and 1,290µl UHQ water. For the Adenovirus reaction the working stock consisted of 25µl of 100µM AD-F, 25µl of 100µM AD-R, 20µl of 100µM AD-Pro and 1,430µl UHQ water.

Bacterial real-time PCR reactions were done in monoplex 20µl reactions, consisting of 5µl of the primer/probe working stock solution, 10µl Lightcycler 480 Probes Master (4887301001, Roche) and 5µl extracted DNA. Viral real-time PCR reactions were done in multiplex 25µl reactions. The Influenza A-B and BMV reaction consisted of 12.5µl Superscript III One-Step RT-PCR System 2x Reaction Mix, 0.8µl SuperScript III RT/Platinum Taq Enzyme Mix (both 11732-088, Invitrogen), 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 TaqMan 2x Universal PCR Mastermix (4364338, Applied Biosystems), 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.

2.4.3.4 Real-time PCR

Each monoplex bacterial real-time PCR was run on the Rotor-Gene Q (Qiagen) 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) 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.

2.4.4 Standard Curves for Quantification of Bacterial Species

To enable quantification of bacterial species, a standard curve was set up for each of the following bacterial references: *S. pneumoniae* ATCC 49619*, S. aureus* (clinical isolate)*, H. influenzae* NCTC 11931*, M. catarrhalis* (clinical isolate)*, P. aeruginosa* NCTC 10662 and *N. meningitidis* MC58. Firstly, a 1-2 McFarland solution of each bacterial species was made up in 1ml UHQ water (Ambion, AM9937). A 10-fold series dilution was then undertaken to dilute the bacterial solutions to 10^{-9}. Each of the dilutions was then plated out in triplicate 10µl spots onto CBA, except for *H. influenzae* which was plated out onto CHOC. The plates were then incubated at 37°C for 24 hours, with 5% CO_2 for *S. pneumoniae, H. influenzae, M. catarrhalis* and *N. meningitidis*. After incubation, colonies forming units (CFU) were counted on each of the 'readable' (between 0-300 colonies) plates and CFU/ml values were calculated for each dilution.

Each dilution was also extracted using the QIAxtractor nucleic extraction method described previously. Once extracted each of the dilutions were run in duplicate on the Rotor-Gene real-time PCR instrument (Qiagen). For each extraction the CFU/ml values were entered into the Rotor-Gene Q software, which were then correlated to a given C_T value, creating a standard curve. This curve could then be loaded during real-time PCR detection of swab samples in order to quantify each bacterial species according to its C_T value. Standard curves for each of the bacterial species are given in Appendix 3.

2.4.5 Statistical Analysis

Real-time PCR identification results as well as information gathered from the participant questionnaire were entered into SPSS (IBM, version 20). Carriage and co-carriage of bacterial and viral species were then assessed.

Overall prevalence of bacterial and viral species carriage and co-carriage within different swab types was calculated. For bacterial species, where the numbers of positive samples were large enough, Chi-squared and fisher's exact tests were used to compare samples from different individuals (NP versus nose, NP versus self-taken WMS, self-taken WMS versus HCP-taken WMS, nose versus HCP taken WMS) and McNemar's Chi-squared test was used to compare samples from the same individuals (nose versus self-taken WMS and NP versus HCP-taken WMS). Furthermore, McNemar's test was used to determine

differences between overall bacterial prevalence between culture and real-time PCR detection methods. The formulas for these tests are shown in section 2.3.2.2.

Due to the small numbers samples positive for respiratory viral species, only descriptive statistics and univariate analyses could be performed for comparing prevalence between swab types and for understanding the effect of demographic variables on the carriage and co-carriage of these species. The Mann-Whitney U-test was performed to assess the crude association between carriage and age. Chi-squared (X^2) or Fisher's Exact (FET) tests were used to determine crude associations between carriage and binary or categorical variables. Variables assessed using univariate analyses included age (in years), recent RTI (within a month prior to swabbing), geographical location (six geographical areas detailed in Figure 1), recent use of antibiotics (within a month prior to swabbing), vaccinations up-to-date (with the national immunisation schedule), season (winter or summer swabbing) and presence of other microbial species within the same sample. Formulas for X^2 and FET are shown in section 2.3.2.2.

Co-carriage of bacterial and viral species within the same sample using real-time PCR detection was performed using multivariable binary logistic regression analysis as described in section 2.3.2.2.

In addition, a one-way analysis of variance (ANOVA) was used to determine within sample and between sample variance in bacterial species concentrations from real-time PCR data. ANOVA *F* test values, degrees of freedom (df) and *p*-values were recorded.

2.5 Typing of Bacterial Isolates

2.5.1 Selection of Isolates

All nose and NP isolates of *S. pneumoniae* and *H. influenzae* collected during both swabbing seasons were used to undertake molecular typing. Furthermore, all *S. aureus* isolates from all swabs types collected during both swabbing seasons were used to undertake MRSA typing and to determine the prevalence of methicillin-resistant *S. aureus* (MRSA).

2.5.2 PCR Serotyping of *S. pneumoniae* and *H. influenzae* Isolates

2.5.2.1 S. pneumoniae

PCR serotyping of *S. pneumoniae* was undertaken using eight separate PCR reaction pools containing four to five serotypes per pool and covering 40 of the most common pneumococcal serotypes using the Centre for Disease Control and Prevention (CDC) protocol (Pai et al., 2006). Each pool also contained *cpsA* primers specific for *S. pneumoniae* as well as UHQ water, 50mM $MgCl_2$ and 2x Red PCR Mix.

Firstly master mixes of each PCR reaction were made up for each of the reaction pools, which are described in full in the Appendix 4. Then, 11.5µl of the appropriate master mix was aliquoted into PCR tubes and mixed with 1µl of isolate DNA or positive/negative control. PCR tubes were then loaded onto the thermocycler and run as follows: 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.

Following PCR, 12.5µl PCR products were loaded in the wells of a 2% high-resolution agarose gel (AGTC Bioproducts, AGD1) containing GelRed Nucleic Acid Stain (Biotium, 41003) and run for 1h45 at 125V. Hyperladder IV (Bioline, BIO-33056) was also loaded onto the gel for the determination of product sizes. The gel was then imaged using a UV gel imager and product sizes were determined. Reactions for the 134 pneumococcal isolates were undertaken sequentially through one to eight until a positive PCR reaction was found.

2.5.2.2 H. influenzae

Serotyping of *H. influenzae* was undertaken using previously published methods (Falla et al., 1994, Davis et al., 2011). Firstly, speciation PCR was undertaken on *H. influenzae* isolates using outer membrane protein P2 (*omp*P2) primers. Primers were diluted to a concentration of 10mM using distilled water. A mastermix was then made up using 0.625µl of each primer (forward and reverse), 6.25µl 2x Red PCR Mix and 4µl of distilled water per reaction. 11.5µl of mastermix was then aliquoted into each PCR tube along with

1µl of isolate DNA or positive/negative control. The PCR was then run on the thermocycler using the following parameters: 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. PCR products were then run on a 2% agarose gel for 70 minutes at 120V. The gel was then imaged using a UV gel imager and product sizes were determined. Banding patterns were then assessed to determine *H. influenzae* positive isolates.

Subsequently, an encapsulation PCR was undertaken on the *omp*P2-positive isolates by targeting the *bex*B gene. Primers were again diluted to a concentration of 10mM using distilled water. A mastermix was then made up using 1µl of each primer (forward and reverse), 10µl 2x Red PCR Mix and 6.4µl of distilled water per reaction. 18.4µl of mastermix was then aliquoted into each PCR tube along with 1.6µl of isolate DNA or positive/negative control. The PCR was then run on the thermocycler using the following parameters: 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. PCR products were then run on a 2% agarose gel for 70 minutes at 120V. The gel was then imaged using a UV gel imager and product sizes were determined. Banding patterns were then assessed to determine encapsulated *H. influenzae* isolates.

Finally a serotyping PCR was undertaken on each of the encapsulated isolates using primers specific to serotypes a-f. Primers were once more diluted to a concentration of 10mM using distilled water. A mastermix was then made up for each serotype using 0.625µl of each primer (forward and reverse), 6.25µl 2x Red PCR Mix and 4µl of distilled water per reaction. 11.5µl of mastermix was then aliquoted into each PCR tube along with 1µl of isolate DNA or positive/negative control. The PCR was then run on the thermocycler using the following parameters: 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. PCR products were then run on a 2% agarose gel for 70 minutes at 120V. The gel was then imaged using a UV gel imager and product sizes were determined. Banding patterns were then assessed to determine the serotype of each *H. influenzae* isolates.

2.5.3 MRSA Typing of *S. aureus*

For MRSA typing, all of the *S. aureus* isolates from all four swab types and both time-points were used. *S. aureus* isolates were tested for methicillin resistance by plating out 10µl 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.

2.6 Whole Genome Sequencing

Whole genome sequencing was undertaken in-house using the Illumina MiSeq sequencer. Sample preparation for sequencing was undertaken using Nextera XT Sample Preparation (FC-131-1024, Illumina) and Index (FC-131-1001) kits.

2.6.1 Selection of Isolates

All nose and NP isolates of *S. pneumoniae* and *H. influenzae* from both swabbing time-points were used for *in silico* typing methods. For further in-depth analysis of vaccine candidate and antigen genes, a selection of *S. pneumoniae* and *H. influenzae*, were selected as follows: all co-colonisers of these species were selected from the first swabbing time-point along with a small set (3-5 isolates) of non co-colonisers from the first swabbing time-point.

2.6.2 DNA Extraction

Isolates were plated out onto CBA (CHOC for *H. influenzae*) and incubated at 37°C with 5% CO_2 for 24 hours. Plates were checked for purity and mixed isolates were re-streaked from a single colony to obtain pure cultures. DNA was extracted from a sweep of approximately 10 colonies of pure culture using the Qiagen QIAamp DNA Mini Kit (51306, Qiagen) Blood Protocol, as follows.

The sweep of pure culture colonies was transferred to 200µl of lysis buffer and vortexed briefly before being incubated at 37°C for one hour. After incubation, 10µl of proteinase K was added and tubes were then incubated in a water bath at 56°C for 1-2 hours. 200µl of Buffer AL was then added to the tubes followed by incubation on a heat block at 90°C for 10mins. 200µl of 100% ethanol was then added to the samples, followed by centrifugation and transfer of the liquid to a spin column. Tubes were then centrifuged at 6000xg for 1

minute and the filtrate discarded. Subsequently, 500µl of AW1 buffer was added to the tubes and centrifuged at 6000xg for 1 minute before discarding the filtrate. Then, 500µl of AW2 buffer was added to the tubes and centrifuged at maximum speed for 3 minutes before discarding the filtrate and repeating the step. The spin column was then eluted by adding 200µl of UHQ water, incubating at room temperature for 1 minute and centrifuging at 6000xg for 1 minute. This final step was repeated to increase the elution volume. The extracted DNA was then stored at -20°C.

2.6.3 DNA Quantification

DNA extracted from the nose isolates was quantified using the Quant-iT dsDNA HS Assay (Invitrgoen, Q-33120), as follows. Quant-iT reagent and Quant-iT buffer were mixed in a 1:200 ratio to create a working solution. 195µl of this working solution was then added to 5µl of extracted DNA, vortexed and incubated at room temperature for 2 minutes before being read on the Qubit fluorometer. Controls were prepared to calibrate the machine before the samples were read by adding 190µl of working solution to 10µl of Control 1 and Control 2. Sample concentrations for each isolate extraction were given in µg/ml. DNA was then diluted to 0.2ng/µl for DNA sequencing.

2.6.4 Genomic DNA Tagmentation

Five microliters of 0.2ng/µl DNA was added to PCR tubes along with 10µl Tagment DNA Buffer (TD) and 5µl Amplicon Tagment Mix (ATM). The tubes were then placed on a thermocycler which heated the tubes to 55°C for five minutes and held them at 10°C. Subsequently 5µl of Neutralize Tagment Buffer (NT) was added to each tube and incubated at room temperature for 5 minutes.

2.6.5 PCR Amplification of Tagmented DNA

Fifteen microliters of Nextera PCR Master Mix (NPM) was added to each well of tagmented DNA along with 5µl of appropriate index primers 1 (i7) and 2 (i5). The centrifuge tubes were then placed on the thermocycler and run as follows: 72°C for 3

minutes, 95°C for 30 seconds followed by 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds, and then finally 72°C for 5 minutes and hold at 10°C.

2.6.6 PCR Clean-up

Amplified tagmented DNA was cleaned up using the AMPure XP kit (A63880, Agencourt), as follows. 30µl of beads was added to each well containing the tagmented DNA and incubated for 5 minutes at room temperature. The tubes were then placed on a magnetic stand for 2 minutes (or until the liquid has become clear) and the supernatant removed. The beads were then resuspended using 200µl of 80% ethanol and incubated for 30 seconds. This step was repeated and the tubes air-dried for 30 minutes. 52.5µl of Resuspension Buffer (RSB) was then added to each well before being incubated for 2 minutes at room temperature. Tubes were again placed on the magnetic stand and 50µl of the supernatant transferred to new PCR tubes.

2.6.7 Library Normalisation

45µl of Library Normalisation Additives 1/ Library Normalisation Beads 1 (LNA1/LNB1) mix (in a 5.5:1 ratio) was added to 20µl of the clean amplified tagmented DNA in 1.5ml centrifuge tubes and placed on a shaker at 250rpm for 30 minutes. After shaking, tubes were placed on a magnetic stand for 2 minutes (or until the liquid has become clear) and then the supernatant removed. 45µl of Library Normalisation Wash 1 (LNW1) when then added to the tubes before being place on the magnetic stand and supernatant removed once the liquid has cleared. This step was then repeated with 30µl of 0.1M NaOH before transferring 30µl of the supernatant to 30µl of Library Normalisation Storage buffer 1 (LNS1).

2.6.8 Library Pooling and Loading the MiSeq Cartridge

Five microliters of each normalised DNA library was pooled into a single microcentrifuge tube. 12µl of this Pooled Amplicon Library (PAL) was then added to 588µl of Hybridisation Buffer 1 (HT1), vortexed and incubated at 96°C for 2 minutes. Then this Diluted Amplicon

Library (DAL) tube was placed in an ice bath for 1-2 minutes and the whole volume loaded into the MiSeq cartridge sample well.

2.6.9 De novo Sequencing on the MiSeq

The flow cell and reagent cartridge containing the DAL were placed into the MiSeq Sequencer. The *de novo* assembly workflow was selected and the run commenced after pre-run checks.

Illumina technology uses the sequencing by synthesis method whereby fluorescently labelled nucleotides are incorporated one by one and imaged in order to determine the nucleic acid sequence. Solid-phase bridge amplification is used to immobilise and amplify the nucleic acid fragments (which are tagged according to sample) on the flow cell surface (Metzker, 2010, Illumina, 2010).

2.7 Whole Genome Sequence Analysis

2.7.1 Trimming of Sequencing Adaptors

Illumina Nextera sequencing adapters were removed from the raw sequences using Trimmomatic (v0.32, available at http://www.usadellab.org/cms/?page=trimmomatic) (Bolger et al., 2014). This tool produces paired reads from the raw sequence files. This was undertaken as follows:

```
$ module load trimmomatic

$ module load jdk

$ java –jar /local/software/trimmomatic/0.32/trimmomatic-0.32.jar
PE –threads 4 –phred33 –trim.log file_1.fastq.gz file_2.fastq.gz
file_1_paired.fastq.gz file_1_unpaired.fastq.gz
file_2_paired.fastq.gz file_2_unpaired.fastq.gz
SLIDINGWINDOW:10:20 MINLEN:50 ILLUMINACLIP:nextera.fa:1:40:15
```

2.7.2 Determining the Quality of Sequences

Trimmed sequence files were then checked for sequence quality using the FastQC software. Detailed information about each sequence is given, including number of sequences generated, per base sequence quality, Kmer content, per base N content, per base GC content and overrepresented sequences.

2.7.3 Genome Assembly using Velvet

Trimmed sequence files were assembled using the Velvet Optimiser script (Victorian Bioinformatics Consortium, version 2.2.5, available from https://github.com/Victorian-Bioinformatics-Consortium/VelvetOptimiser.git), as follows:

```
$ module load sanger

$ velvetopti.sh –f paired_sequence_1.fastq –p –r
paired_sequence_2.fastq –s sequence_shuffled.fastq
```

This script produced a contiguous sequence file (contigs.fa) for the paired trimmed sequence files for each isolate.

2.7.4 Assembly Improvement

Following assembly, contig sequences were improved using the Assembly Improver perl module (available at http://search.cpan.org/~ajpage/Bio_AssemblyImprovement-1.140300/lib/Bio/AssemblyImprovement.pm), which uses Abacas/SSpace and GapFiller. This removes gaps from the sequences and reorders contigs.

```
$ improve_assembly -a contigs.fa -f sequence_1.fastq -r
sequence_2.fastq
```

2.7.5 Determining the Quality of Assemblies

After assembly, contigs were quality controlled using the assemblathon_stats.pl script (University of California Davis, available at https://github.com/ucdavis-bioinformatics/assemblathon2-analysis) (Earl et al., 2011, Bradnam et al., 2013). This script calculates the basic metrics for genome assembly including number of contigs and scaffolds, number of contigs within the scaffolds, N50 contig size and mean contig size. The script was run as follows:

```
$ perl assemblathon_stats.pl contigs.fa > contigs.txt
```

The information provided via this script determines the quality of each assembly and whether further analyses would be reliable.

2.7.6 Serotyping

Contiguous sequence files (contig.fa) were then serotyped using the *in silico* PCR experiment simulation system (ipcress, European Bioinformatics Institute, available at http://manpages.ubuntu.com/manpages/precise/man1/ipcress.1.html). The system is part of the exonerate tool. Firstly, a tab-delimited file containing primer sequences was generated. This file included each primer name, forward and reverse primer sequences and maximum and minimum product lengths. Primer sequences for *in silico* serotyping of *S. pneumoniae* and *H. influenzae* are shown in Appendices 5 and 6, respectively. *In silico* PCR was then undertaken as follows:

```
$ module load exonerate
```

```
$ ipcress primers.txt contigs.fa –p F –P T > result_file.txt
```

Diversity of the serotypes detected within this set of isolates was assessed using Simpson's Index of Diversity, as follows:

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

n = Total number of organisms of a particular species
N = total number of organisms of all species

2.7.7 Sequence Typing

2.7.7.1 Short Read Sequence Typing using SRST2

Raw sequence files were then used to determine the sequence type of each isolate of *S. pneumoniae* and *H. influenzae* using short read sequence typing (SRST, available at https://github.com/katholt/srst2) (Inouye et al., 2012, Inouye et al., 2014a). SRST allows the identification of MLST alleles of each of the housekeeping genes based on DNA sequence. The housekeeping genes used in MLST analysis are shown below in Table 5. SRST was undertaken as follows:

Firstly srst2 was loaded in order to perform the analysis.

```
$ module load srst2
```

Then, the MLST dataset of the bacterial species to be analysed, was downloaded from the MLST database using the *getmlst.py* script. The script gathers the sequences of each MLST gene allele as well as MLST profiles from the MLST website found at http://pubmlst.org/data/.

```
$ getmlst.py --species "species name"
```

Finally, the srst2 script was used to determine the ST of each of the bacterial isolates using paired raw sequence files and the MLST dataset.

```
$ srst2 --input_pe sequence_1.fastq.gz sequence_2.fastq.gz --
output_pe isolate_reference --log --mlst_db species.fasta --
mlst_definitions species.txt
```

The script searches the paired files for each of the MLST genes and according to the sequence of each gene, determines a sequence type. Any mismatches or uncertainties are reported, along with a measure of depth. This enables the certainty in the designated ST to be determined.

Diversity of the STs detected within this set of isolates was assessed using Simpson's Index of Diversity, as shown in section 2.7.6.

Table 5. MLST genes for *S. pneumoniae* and *H. influenzae*

Species	MLST Gene	Gene Role
S. pneumoniae	*aroE*	Shikimate dehydrogenase
	gdh	Glucose-6-phophate dehydrogenase
	gki	Glucose kinase
	recP	Transketolase
	spi	Signal peptidase I
	xpt	Xanthine phosphoribosyltransferase
	ddl	D-alanine-D-alanine ligase
H. influenzae	*adk*	Adenylate kinase
	atpG	ATP synthase gamma chain
	frdB	Fumarate reductase
	fucK	Fuculose kinase
	mdh,	Malate dehydrogenase
	pgi	Phosphoglucose isomerase
	recA	Recombinase A

MLST = multilocus sequence typing

2.7.7.2 eBURST Analysis

MLST data was visualised and compared against the MLST databases using eBURST (version 3, available at http://eburst.mlst.net/) (Feil et al., 2004). This allowed the relationships between the nose isolates to be determined and for these to be compared with historical and geographically-distinct isolates found within the database. By comparing against the MLST database any changes that may have occurred over time were determined.

2.7.8 Antibiotic Resistance Analysis

SRST2 was used to screen the whole genome sequences of the selected *S. pneumoniae,* and *H. influenzae* isolates for a defined set of antibiotic resistance genes from ResFinder (http://cge.cbs.dtu.dk/services/ResFinder/) and Comprehensive Antibiotic Resistance Database (CARD, http://arpcard.mcmaster.ca) and available from the SRST2 website.

The full list of antibiotic resistance genes are listed in the original articles (Zankari et al., 2012, McArthur et al., 2013). This was undertaken as follows:

```
$ module load srst2

$ srst2 --input_pe sequence_1.fastq.gz sequence_2.fastq.gz --
output isolate --gene_db resistance.fasta
```

Antibiotic resistance testing using SRST2 only targets acquired resistance genes, therefore in addition to these genes, *pbp*1a, *pbp*2x and *pbp*2b genes were also identified within *S. pneumoniae* isolates using ipcress (script detailed in section 2.7.6) and primers described previously (Nagai et al., 2001).

2.7.9 Vaccine Candidates and Antigens

SRST2 was used to screen the whole genome sequences of the selected *S. pneumoniae,* and *H. influenzae* isolates for a defined set of genes from the Virulence Factors of Pathogenic Bacteria Database (VFBD) (Chen et al., 2005, Chen et al., 2012). The full list of genes, DNA sequences and protein sequences is available on the VFDB website (http://www.mgc.ac.cn/VFs/main.htm). Antigens which are or have been considered as vaccine candidates are of particular interest. *S. pneumoniae* vaccine candidates include *piaA, pspA, pspC/cbpA, psaA, nanA, lytA* and *ply* (Brown et al., 2001, Tai, 2006, Ogunniyi et al., 2007, Bogaert et al., 2004b). *H. influenzae* vaccine candidates include hpd, ompP2, ompP5, tbpA/B, hmw1/2 and Hia (Poolman et al., 2001, Forsgren et al., 2008). This was undertaken as follows:

```
$ module load srst2

$ srst2 --input_pe sequence_1.fastq.gz sequence_2.fastq.gz --
output isolate --gene_db virulence_factors.fasta
```

In addition to these genes, the *H. influenzae* protein D (*hdp*) gene was also identified using ipcress (script detailed in section 2.7.6) and primers described previously (Binks et al., 2012).

2.8 Ecological Analysis

2.8.1 Nestedness

Nestedness is a measure of organisation within an ecological system and allows the level of order in species extinction to be determined. A nestedness tool, PlotTemp.R, was used within the R graphical user interface (CRAN, version 3.0.1) to determine the level of order within real-time PCR datasets. The PlotTemp.R script, which was created by Patrick Doncaster and modified for use with the swab sample data, is shown in full in Appendix 7. This script required Vegan and Permute packages as well as the ResetPar.R. This was undertaken as follows:

```
$ library (vegan)

$ source ("PlotTemp.R")

$ plot_temp("dataset.txt")
```

Real-time PCR datasets were converted to matrices of species presences (1) or absences (0) across samples. The plot_temp script then takes the matrix and 'packs' it so that samples and species are shuffled according to incidence and richness. Presences are represented by red squares and absences by white squares.

A line of perfect fill is also constructed and is based on all absences being organised to the right and presences to the left. Surprise presences are presences found to the right of the line of perfect fill and surprise absences are absences found to the left of the line of perfect fill.

The script also calculates a matrix temperature, ranging from 0-100°, determined by the sum of the distances of surprise absences or presences from the line of perfect fill. The temperature is interpreted to determine the level of nestedness within the system. A low temperature shows a highly nested matrix and a high temperature shows a highly random matrix.

Data was analysed according to age, recent RTI and season.

2.8.2 Species Distribution

Species distribution of the different bacterial and viral species across swab samples was analysed by plotting the observed frequency of species types (i.e. number of observations of 1, 2, 3, etc. species within swab samples). This distribution was compared to the Poisson (random) distribution, shown below:

$$P(x) = e^{-\bar{x}} \frac{\bar{x}^x}{x!}$$

e = Euler's number (2.71828)
$\bar{x}$ = mean number of species observed within samples
x = actual number of species within samples

Deviations of the observed distribution from the Poisson distribution were then calculated using the X^2 Goodness of Fit test, shown below:

$$\chi^2 = \sum_{i=1}^{a} \frac{(O_i - E_i)^2}{E_i}$$

X^2 = Pearson's cumulative test statistic
O_i = Observed frequency
E_i = Expected frequency
a = number of cells in table

The X^2 test would then enable us to determine if the observed distribution differed from the Poisson distribution. A significant deviation ($p < 0.05$) was then interpreted. If the observed distribution was narrower than the Poisson distribution (regular distribution with many swab samples containing a similar number of species) then there was indirect evidence of mutual repulsion in competition. If the observed distribution was broader then the Poisson distribution (clumped distribution with many swab samples containing very few or many species) then there was indirect evidence of mutual attraction in facilitation.

Data was analysed according to age, recent RT and season.

2.8.3 Application of Ecological Theory

Species abundance distribution curves of bacterial and viral abundances (n) as well as bacterial concentrations (CFU/ml) were plotted and compared to McArthur's Broken Stick model, a neutral model of community assembly (Macarthur, 1957), as shown below:

$$Rank = S \times e^{-S \times \frac{n}{N}}$$

S = total number of species
n = number of individuals
N = total number of individuals

The species abundance distribution curves were then compared to this theoretical species abundance distribution curve using X^2 Goodness of Fit test in order to determine any deviations from the theoretical distributions.

Chapter 3. Participation in a Community Swabbing Study

3.1 Introduction

Participation, the proportion of individuals taking part in the study relative to the total number of individuals invited, was assessed in order to determine the number of samples that can be collected by undertaking a large community-based swabbing study. By gauging the level of participation from members of the community, this will facilitate the set up of larger multi-site national and international studies. The level of participation within a study is important as low participation rates can result in sampling bias, whereby a certain group of individuals are over-sampled and another group are under-sampled. Therefore greater participation rates tend to produce more accurate sampling results and will be more representative. Furthermore increased participation also increases sample size and hence power of a study in addition to reducing the per capita costs of conducting the study.

Participation was calculated for each study arm (self-swabbing and HCP swabbing), GP practice and age group in order to determine whether swab collection method, location and age affected the participation of individuals. The level of deprivation and urban/rural location were also analysed in relation to participation rates. It was hypothesised that participation rates would be greater within the self-swabbing group due to the convenience of performing self-swabbing when compared with HCP-performed swabbing. It was also hypothesised that older patients and those in less deprived locations would be more willing to participate when compared with younger participants and those from more deprived areas potentially due to more time available for participation in the study. Finally, it was hypothesised that participants in more urban locations would be more likely to respond, especially in the HCP swabbing group, due to greater proximity to healthcare services.

3.2 Analysis of Community Participation

3.2.1 Characteristics of Participants

A total of 2,417 patients from 20 GP practices across the South West PCRN East Hub participated in the swabbing study across the two swabbing time-points. Different individuals were swabbed in each study arm and time-point. The characteristics of the study participants are shown in Table 6.

Table 6. Characteristics of the Study Participants

	Summer 2012 *n* (%)		**Winter 2013** *n* (%)	**Total** *n* (%) (N = 2,417)
	Self-swabbing (N = 1,260)	HCP swabbing (N = 314)	Self-swabbing (N = 843)	
Age (years)				
Mean	37.4	50.1	40.7	40.2
Minimum	0	0	0	0
Maximum	94	88	95	95
0-4	329 (26.1)	56 (17.8)	168 (19.9)	553 (22.9)
5-17	137 (10.9)	24 (7.6)	127 (15.1)	288 (11.9)
18-64	465 (36.9)	89 (28.3)	286 (33.9)	840 (34.8)
65+	311 (24.7)	145 (46.2)	244 (28.9)	700 (29.0)
Unknown/missing	18 (1.4)	0 (0.0)	18 (2.1)	36 (1.5)
Recent Antibiotic Treatment				
Yes	101 (8.0)	26 (8.3)	79 (9.4)	206 (8.5)
No	1124 (89.2)	286 (91.1)	743 (88.1)	2153 (89.1)
Unknown/missing	35 (2.8)	2 (0.6)	21 (2.5)	58 (2.4)
Recent Respiratory Tract Infection (RTI)				
Yes	365 (29.0)	61 (19.4)	367 (43.5)	793 (32.8)
No	860 (68.3)	250 (79.6)	453 (53.7)	1563 (64.7)
Unknown/missing	35 (2.8)	3 (1.0)	23 (2.7)	61 (2.5)
Vaccination Status				
Up-to-date	1022 (81.1)	270 (86.0)	680 (80.7)	1972 (81.6)
Not up-to-date	40 (3.2)	10 (3.2)	33 (3.9)	83 (3.4)
Unknown/missing	198 (15.7)	34 (10.8)	130 (15.4)	362 (15.0)

Recent antibiotic treatment or RTI refer to those having occurred with one month of swab sample collection. HCP = healthcare professional.

3.2.2 Participation of the Study Groups

Overall participation, as shown in Figure 3, was greater in the self-swabbing group in both time-points with 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 compared with the HCP swabbing group at 6.2% (n = 314, 95% CI 5.54% to 6.86%). Individual GP practice participation, as shown in Figure 4, varied from 9.3% (n = 27) to 33.1% (n = 96) in the summer self-swabbing group, 1.0% (n = 3) to 12.3% (n = 34) in the summer HCP swabbing group and 9.1% (n = 21) to 30.4% (n = 80) in the winter self-swabbing group. The anticipated level of participation, of 25% or higher, was not achieved within the HCP swabbing group by any of the GP practices. Only three practices achieved this anticipated level in both self-swabbing time-points and eight achieved this in one self-swabbing time-point. Nine practices failed to achieve 25% in either time-point.

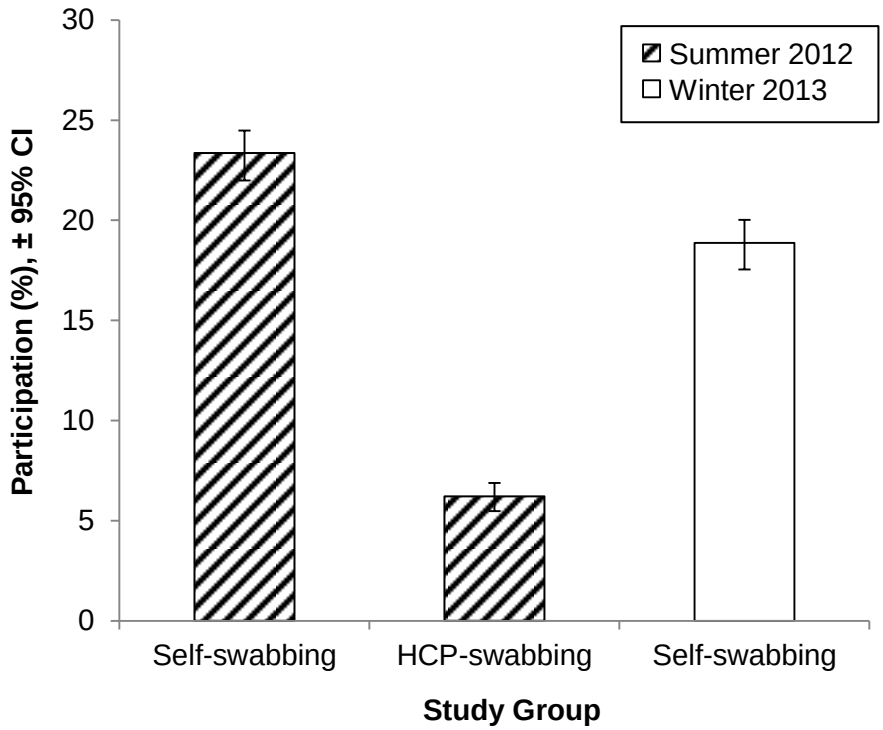


Figure 3. Overall Participation Rates by Study Arm

HCP = healthcare professional

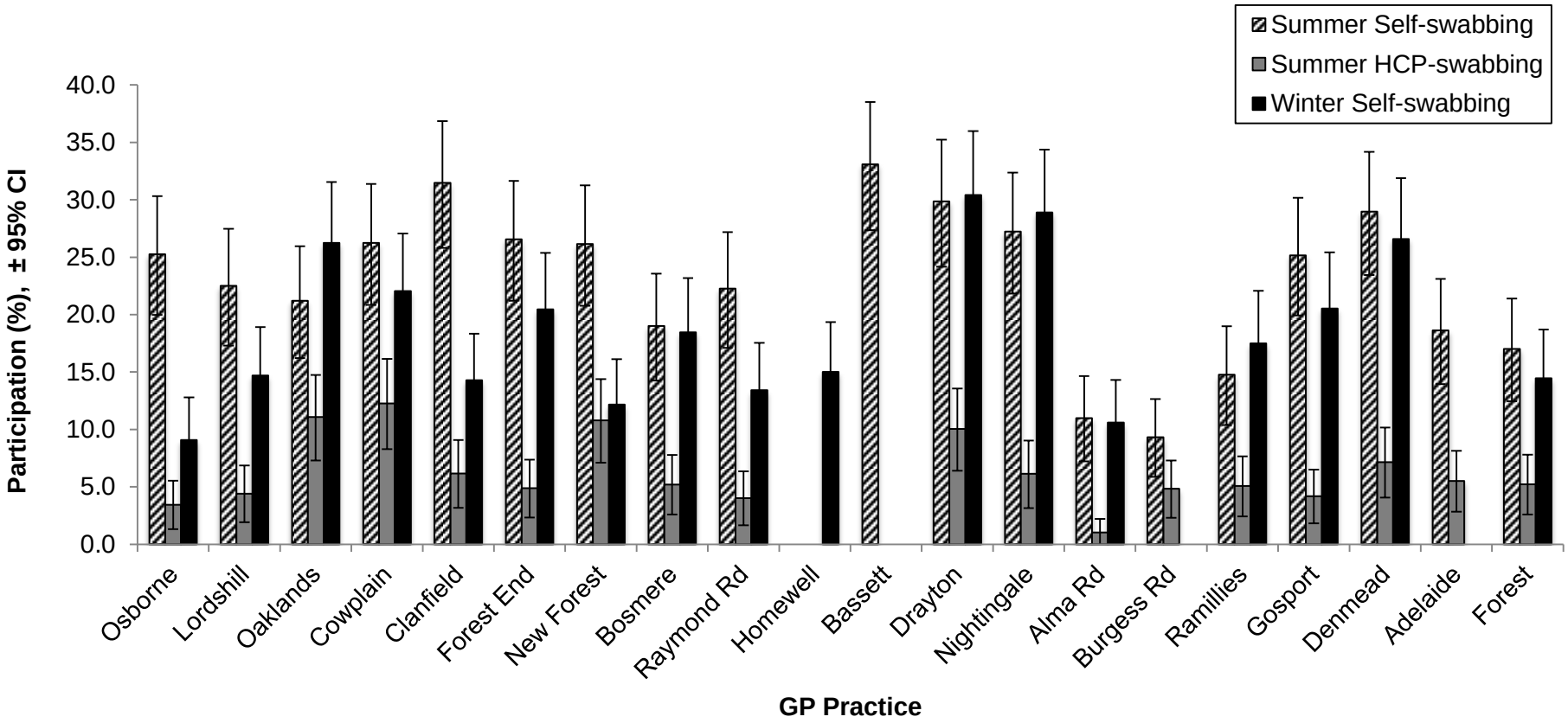


Figure 4. Participation of the 20 GP practices by Study Arm and Time-point

HCP = Healthcare professional

3.2.3 Age Distribution of Participants

Participation, as shown in Figure 5, was greater in those aged 5 years and older when compared with those aged 0-4 years, in both study groups and time-points. A larger discrepancy between age group participation was found within the self-swabbing group (summer 2012 Δ = -11.2, winter 2013 Δ = -14.3) versus the HCP swabbing group (summer 2012 Δ = -5.3). This is a largely due to the greater participation by patients aged 5 years and older within the self-swabbing group versus the HCP swabbing group, whereas participation of patients aged 0-4 years varied less between study groups.

The distribution of participants according to age, as shown in Figure 6, shows two peaks in the number of participants. This first peak, representing patients aged 0-10 years, is a result of the sampling method which involved inviting a large number of children aged 0-4 years to take part in the study. The second peak, representing individuals aged 61-70 years, demonstrates that participation increases with age in adults but decreases again after the age of 70 years.

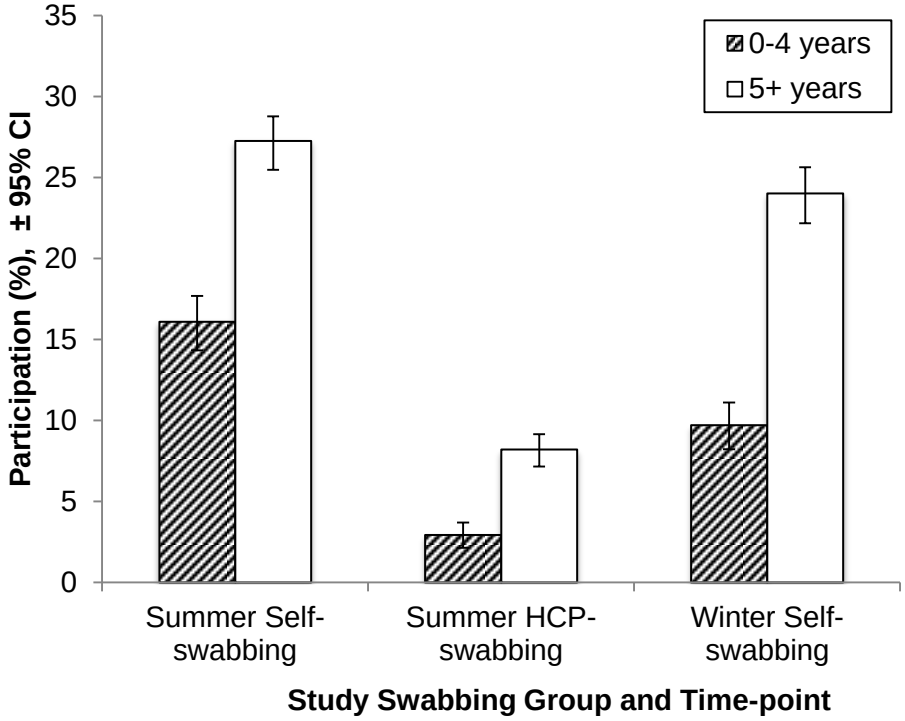


Figure 5. Participation by Age Group

HCP = Healthcare professional

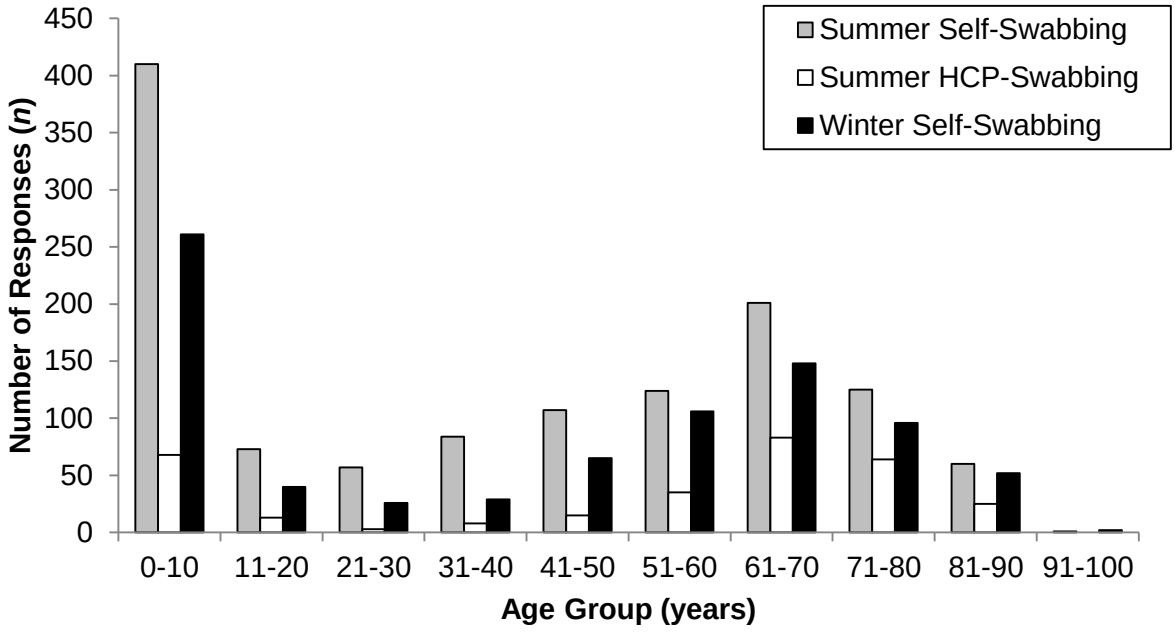


Figure 6. Distribution of Participants by Age Group

HCP = Healthcare professional

3.3 The Impact of Deprivation on Participation

The UK IMD 2010 score is a measure of deprivation within small geographical areas referred to as lower layer super output areas (LSOA) within the United Kingdom. These are based on levels of crime, education, income, health and disability, living environment, employment and barriers to housing and services. IMD scores of the GP practices' LSOA ranged from 2.70 to 55.60, with a mean of 18.04. Participation, as shown in Figure 7, is negatively correlated with IMD score within the self-swabbing group in summer 2012. This negative relationship indicates that more deprived areas participated less in the study than less deprived areas. However, a wide range of participation levels are observed at similar deprivation scores indicating other factors unrelated to deprivation may also be influencing participation.

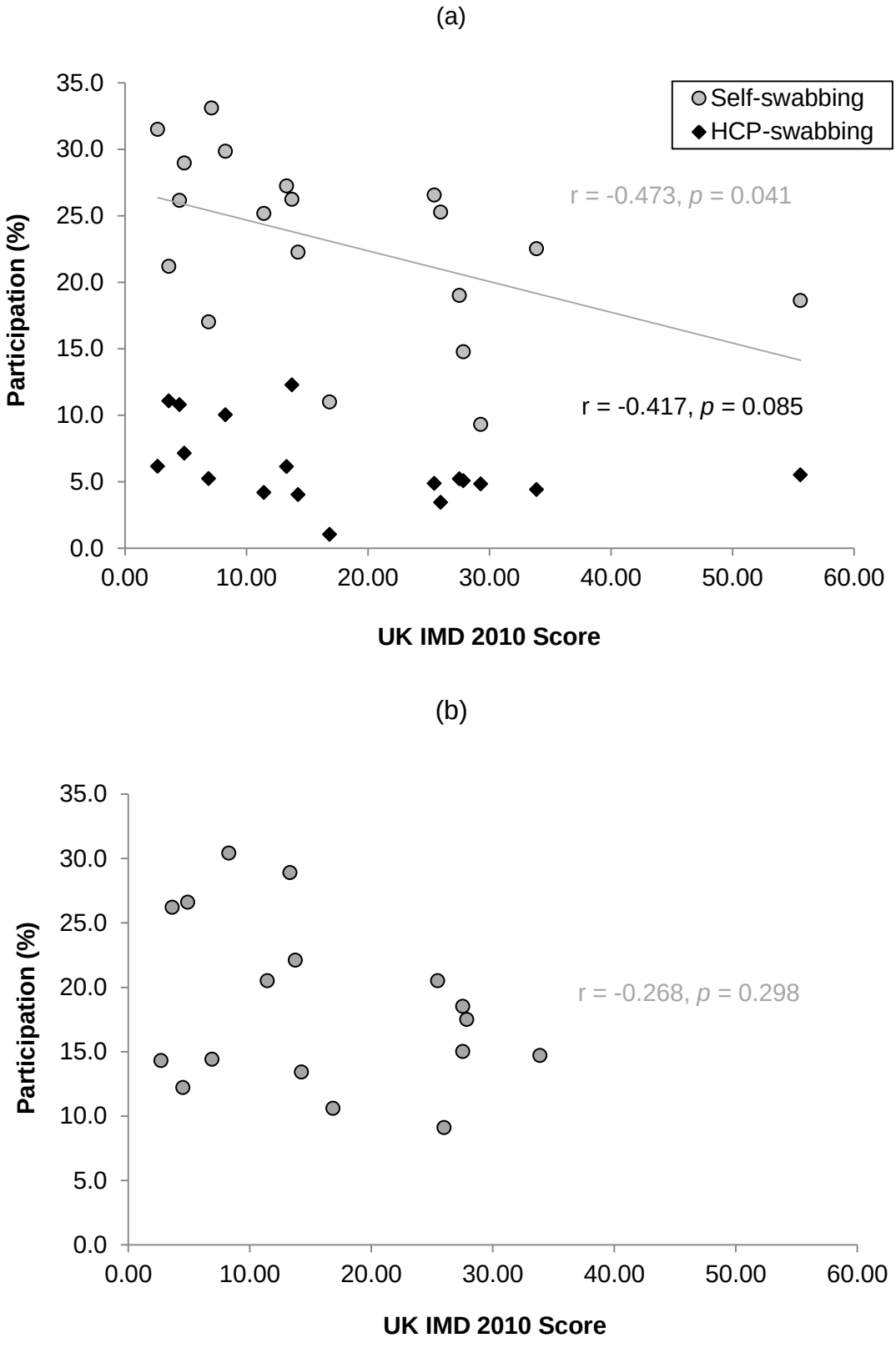


Figure 7. Pearson's Correlation Analysis of Practice Participation as a function of UK IMD 2010 Score for (a) Summer 2012 and (b) Winter 2013 Swabbing Time-points

HCP = Healthcare professional, IMD = Indices of Multiple Deprivation

3.4 The Impact of Population Density on Participation

Population density, the number of individuals per square kilometre, is a measure of urban/rural location with rural locations having lower numbers of individuals per square kilometre and urban locations having greater numbers of individuals per square kilometre. Population density of the GP practices' geographic locations ranged from 173 individuals per km^2 to 5146 individuals per km^2, with a mean of 2785 individuals per km^2. Participation, as shown in Figure 8, does not correlate with the population density. Again however, a wide range of participation levels are observed at similar population densities indicating other factors unrelated to population density may also be influencing participation.

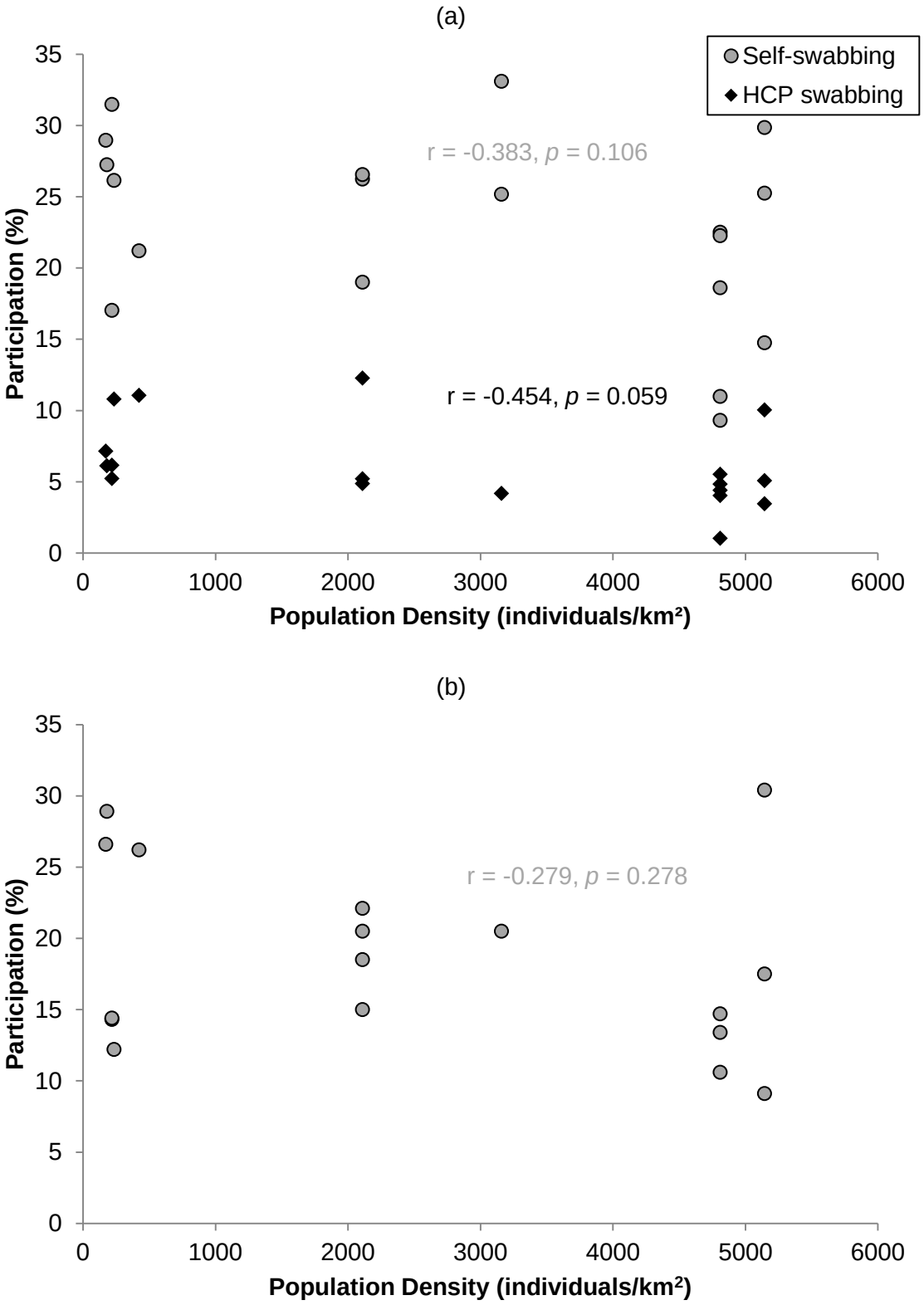


Figure 8. Pearson's Correlation Analysis of Practice Participation as a function of Population Density for (a) Summer 2012 and (b) Winter 2013 Swabbing Time-points

HCP = Healthcare professional

3.5 Discussion

Characteristics of the participants were largely similar between both of the study groups and both time-points, reflecting appropriate stratified random sampling of the population. Numbers of participants reporting recent antibiotic use was between 8.0% and 9.4%, dependent on study group. The national average of UK community and hospital prescriptions in 2013 was 27.4 defined daily doses (DDD) per 1,000 inhabitant per day (Public Health England, 2014a). The lack of data on antibiotic doses within this study limits the comparison of our data to national prescription data.

Numbers of participants reporting an up-to-date vaccination status was between 80.7% and 86.0%, dependent on study group. This was lower than the published vaccination coverage for the UK and Hampshire between October and December 2014, which were approximately 92.7% and 94.3% respectively (Public Health England, 2015c). This difference in vaccination coverage may reflect the limitations of the question posed within the study questionnaire as a number of individuals may not be aware of which vaccinations they are meant to have had. This information should be collected more precisely, with details of individual vaccinations, and from healthcare professionals to enable accurate data capture. Furthermore self-reported vaccination status, although a useful tool, is not as accurate as HCP reported vaccination status, as shown in a study comparing self-reported influenza vaccination status and data from an immunisation registry, where the positive predictive value and negative predictive value of self-reporting were 89% and 96% respectively (Irving et al., 2009).

Numbers of participants reporting a recent RTI was between 19.4% and 43.5%, dependent on study group. Overall, this was similar to the consultation rate for all respiratory infections in 2000 across 108 GP practices in England, which was 27.3% (Ashworth et al., 2004). Recent RTI was greater in the winter swabbing group versus either of the summer swabbing groups, which is expected as upper RTI, especially colds and flu, are often more frequent in the winter months (Dushoff et al., 2004, Eccles, 2002).

The mean age of participants within the study ranged from 37.4 to 50.1 years, dependent on study group. The distribution of ages differed from the age distribution of Hampshire (mean age 42 years, 5.7% 0-4 years, 76.3% 5-64 years and 18% ≥65 years) and the UK as a whole (mean age 39 years, 6% 0-4 years, 77.7% 5-64 years and 16.3% ≥65 years) with more individuals aged 0-4 years and ≥65 years and fewer individuals aged 5-64 years represented within the study (Hampshire County Council, 2011). This difference in

distribution reflects the design of the study, with stratified random sampling in order to obtain greater numbers of participants aged 0-4 years. Age distribution of responders showed similar patterns in both study groups and both time-points, with elderly participants accounting for the greater levels of participation in those aged 5 years and older. The greater numbers of participants in this age group are most likely due to the amount of time available to elderly participants as they are generally free from work or childcare commitments. Many elderly people are more likely to have experienced severe RTI, which might mean that they are able to relate to the study aims (Nicholson et al., 1997). Previous studies have shown increased response rates with increasing age in the adult population, with those aged 40 years and older more likely to respond to questionnaires sent via mail than those aged less than 40 years old (Eaker et al., 1998). However, elderly patients 65 years and older showed decreased response rates compared to those aged 30-64 years old (Murthy et al., 2004).

Participation in the self-swabbing group was much greater than the HCP swabbing group, when compared in the summer time-point. This is most likely due to the limited availability of swabbing appointments, which were within GP practice opening hours, and the amount of time required to attend a swabbing appointment, including time off work. The time taken to contact the GP practice and attend a swabbing clinic is most likely a disincentive for participation.This may have led to a large proportion of responders with the HCP group being those who were most available for such appointments, including the elderly, unemployed, stay at home parents and young children. As such individuals of school or working age are underrepresented, as shown in Figure 6 with low levels of participation for ages 11-60 years. Furthermore the NP swab is known to cause slight discomfort, which might discourage some individuals from taking part. The ease of self-swabbing, on the other hand, which can be done easily at each participant's convenience, is much more appealing. For many participants, especially children, self-swabbing is a much easier option which could explain the higher participation in this group.

Self-swabbing has previously been used for the detection of respiratory viruses (Smieja et al., 2010, Akmatov et al., 2012), *S. aureus* (Gamblin et al., 2013) and Group A Streptococci (Murray et al., 2015). Furthermore, self-swabbing is also thought to be essential in the collection of samples during the symptomatic stages of viral respiratory infections (Akmatov et al., 2012). Studies have also demonstrated that thorough swabbing technique for sampling viruses, with high levels of human epithelial cell with samples, was achieved via self-swabbing (Smieja et al., 2010, Akmatov et al., 2012). Finally the cost per participant for self-swabbing versus HCP-swabbing is lower, making the method much more cost-effective (Coughtrie et al., 2014). Despite the advantages of self-swabbing, the

mishandling of swabs and poor swabbing technique are entirely possible with self-swabbing, which may produce poor microbial recovery (Akmatov and Pessler, 2011). Furthermore, delays in returning the swabs once taken may also be problematic in terms of microbial recovery, as transport media will only maintain species survival for a limited period of time (Ellner and Ellner, 1966).

The IMD score varied between GP practices' LSOA from 2.7 to 55.6, which was similar to the IMD scores from all LSOA across the country (0.53 to 87.80, mean 21.67). IMD was negatively correlated with participation. This has previously been observed in community swabbing studies (Gamblin et al., 2013). This is potentially due lack of time available for participation, with individuals living in more deprived areas undertaking multiple employments or lacking access to childcare. Previous reports have shown socioeconomic status to be an important determinant in participation in clinical trials, which may be a result of poor health status of individuals from more deprived areas, which can result in ineligibility for trial participation (Unson et al., 2004). However, this study did not include stringent exclusion criteria as seen in clinical trials. Furthermore, lack of finances and less access to healthcare may also reduce trial participation in individuals with lower socioeconomic status (Gross et al., 2005). However, within this study, the large range of participation levels across similar deprivation scores would seem to suggest that other factors may have influenced study participation.

Further factors which may influence study participation include ethnicity and female gender, which have also been shown to influence participation in cancer trials (Murthy et al., 2004). However, using multivariate analysis, one study demonstrated that the difference in participation in different ethnic groups may be explained by socioeconomic status (Gross et al., 2005). Further studies have in fact shown no differences in participation rates by ethnic group (Corbie-Smith et al., 2003). Furthermore, differences in participation by gender were thought to be limited to elderly patients (Murthy et al., 2004). Study design has also been shown to affect participation rates, with the use of telephone surveys, interviews, preliminary warnings and financial incentives all increasing the level of response (Yu and Cooper, 1983).

Chapter 4. Microbial Carriage in the Upper Respiratory Tract

4.1 Introduction

Carriage studies permit the estimation of the prevalence of microbial species within the upper respiratory tract at a given time. Such carriage studies have been undertaken in the United Kingdom in order to determine levels of carriage of *S. pneumoniae* (Tocheva et al., 2013, Flasche et al., 2011, Lakshman et al., 2003), *H. influenzae* (Barbour et al., 1995, McVernon et al., 2004) and *N. meningitidis* (Maiden and Stuart, 2002, Cartwright et al., 1987, Read et al., 2014) before and after the introduction of conjugate vaccines. However, a standard method for assessing carriage is only currently available for *S. pneumoniae* (Satzke et al., 2013, Gladstone et al., 2012, O'Brien and Nohynek, 2003). The aim of the current study was to compare swabbing methodologies for detecting six key upper respiratory tract bacterial species, *S. pneumoniae, H. influenzae, M. catarrhalis, S. aureus, P. aeruginosa* and *N. meningitidis*. In order to assess the best swabbing method for detecting carriage within the upper respiratory tract, two study groups were compared: self-swabbing, via nose and whole mouth swabs (WMS), and HCP swabbing, via NP and WMS. It was hypothesised that self-swabbing would be a non-inferior method for sampling the upper respiratory tract bacterial flora as HCP swabbing. In this study, non-inferior signifies no significant difference in the detection of the target bacteria between self-swabbing and HCP swabbing. Evidence for the usefulness of self-swabbing in carriage studies has been observed previously for *S. aureus* (Gamblin et al., 2013).

In addition. real-time PCR, a method which enables the detection of respiratory viruses (van de Pol et al., 2007, van Elden et al., 2001), was performed. Using carriage data and demographic data collected throughout the study, the effects of demographic variables such as age, recent RTI, vaccinations, antibiotics and geographical location (areas shown in Figure 1) on carriage would be determined and a mathematical model of upper respiratory tract carriage would be constructed. It was hypothesised that age and recent RTI were likely to have a large effect on carriage as the microbial flora changes throughout life and in the event of infection (Vives et al., 1997, Peltola et al., 2005). However, it was hypothesised that antibiotics and vaccination were unlikely to have a large effect due to low and stable levels of antibiotic resistance with respiratory species (Reynolds, 2009) and high levels of vaccine uptake within the United Kingdom (95% vaccination coverage at five years of age) (Public Health England, 2014b). Furthermore, vaccination has been shown to change the distribution of *S. pneumoniae* serotypes but

not the overall prevalence of carriage of this species (Gladstone et al., 2015). Due to the effects of herd immunity, similar patterns of carriage are likely to be observed within the unvaccinated population compared with the vaccinated population (Miller et al., 2011).

4.2 Sample Information

Details of swab samples analysed by culture and real-time PCR are shown in Table 7 below.

Table 7. Real-time PCR and Culture Sample Information

	Real-time PCR *n* (%) (N = 380)	**Culture** *n* (%) (N = 2,417)
Age (years)		
Mean (SD)	34.0 (30.8)	40.2 (29.6)
Minimum	0	0
Maximum	92	95
0-4	115 (30.3)	553 (22.9)
5-17	72 (18.9)	288 (11.9)
18-64	93 (24.5)	840 (34.8)
65+	100 (26.3)	700 (29.0)
Unknown	0 (0.0)	36 (1.5)
Recent Antibiotic Treatment		
Yes	34 (8.9)	206 (8.5)
No	345 (90.8)	2153 (89.1)
Unknown	1 (0.3)	58 (2.4)
Recent RTI		
Yes	118 (31.1)	793 (32.8)
No	262 (68.9)	1563 (64.7)
Unknown	0 (0.0)	61 (2.5)
Vaccination Status		
Up-to-date	332 (87.4)	1972 (81.6)
Not up-to-date	9 (2.6)	83 (3.4)
Unknown	39 (10.3)	362 (15.0)

SD = standard deviation, RTI = respiratory tract infection, PCR = polymerase chain reaction.

The sample size for real-time PCR detection (n = 190 for each swab type) was not designed to be able to provide a highly precise measure of carriage prevalence but rather to enable the preliminary application of novel ecological methods to respiratory samples (see Chapter 6). Therefore, although carriage prevalence is compared between culture

and real-time PCR, mathematical regression modelling would utilise results from culture detection (where available) as a higher sample size for this method would enable greater statistical power.

4.3 Analysis Methodologies

4.3.1 Bacterial Carriage

Firstly, the prevalence of bacterial carriage was calculated for the six target bacteria. Overall carriage of each of the bacterial species was assessed in relation to swab type (nose, self-taken WMS, NP and HCP-taken WMS) and detection method (culture and real-time PCR). Chi-squared and fisher's exact tests were used to compare samples from different individuals (NP versus nose, NP versus self-taken WMS, self-taken WMS versus HCP-taken WMS, nose versus HCP taken WMS). McNemar's Chi-squared test was used to compare samples from the same individuals (nose versus self-taken WMS and NP versus HCP-taken WMS). This analysis was designed to compare overall crude prevalence between methods in all individuals and hence does not incorporate stratification according to potential confounders such as age. This analysis would enable the identification of the most effective swabbing method for each species within the respiratory tract, which would then be used to perform multivariate analysis. Confounders would be assessed within these subsequent multivariate analyses.

Subsequently, multivariate binary logistic regression analysis of carriage was performed for the bacterial species using the data collected from the swabbing or detection method deemed most effective. Demographic information collected during the study (see questionnaire in Appendix 1; variables include participant age, recent RTI, vaccination status and recent antibiotic treatment) was used to perform the regression analysis. RTI status was classified into 'yes', 'no' and 'do not know/missing'. Vaccination status was classified as 'up-to-date', 'not up-to-date' and 'do not know/missing'. Finally recent antibiotic treatment was classified into 'yes', 'no' and 'do not know/missing'. Recent (for RTI and antibiotic treatment) refers to within the month prior to swabbing, which was self-reported by participants. *N. meningitidis* carriage could not be analysed as a single isolate was collected from the winter 2013 swabbing time-point.

Binary logistic regression modelling is a method for modelling outcomes using binary or categorical variables (Harrell, 2001). The method measures the relationship between a dependent variable (carriage of a specific species) and a set of independent variables (e.g. age, recent RTI, antibiotics use). The method calculates an odds ratio (OR) for each independent variable, accounting for all other independent variables in the model. This method was used in order to understand the effects of different variables on carriage of a specific species. The method also allows any confounding variables to be controlled for by construction of a multi-variable model (McNamee, 2005).

Where numbers were sufficient, >10 cases per independent variable (Hosmer et al., 2013), multivariable logistic regression analysis was performed. Univariate regression analysis was firstly undertaken for each of the bacterial species and significant covariates (10% significance level) from the univariate analysis were then included in the multivariable model. Backward elimination of covariates which did not meet the 5% significance level within the multivariable model was then undertaken. An odds ratio (OR) and 95% confidence intervals were calculated for each variable, controlling for all other variables in the model. An OR of 1 indicates that the covariate does not affect carriage; an OR of less than 1 indicates that the covariate has a negative effect (decrease) on carriage and an OR of greater than 1 indicates that the covariate has a positive effect (increase) on carriage. *P*-values were calculated to determine the strength of positive or negative associations with carriage of that particular species.

For all covariates, a reference category was assigned for comparison. This reference category was the Southampton/Romsey area for geographical location, the summer for season and the ‘no’ group for recent RTI, recent antibiotics use, vaccinations up-to-date and the presence of other species.

4.3.2 Viral Prevalence

Prevalence of respiratory viruses within upper respiratory tract samples was calculated for the following species: rhinovirus/enterovirus, respiratory syncytial virus (RSV), coronavirus, adenovirus, influenza viruses A and B, parainfluenza viruses 1-3 and metapneumovirus. Identification of these viral species was undertaken using real-time PCR on a subset of swab samples collected during the swabbing study (N = 380). The selection of this subset of samples was described previously (see section 2.4.1). Viral

prevalence was assessed within each swabbing method in order to ascertain the variation in prevalence between these methods.

Prevalence of each viral species was described in relation to participant demographics in order to determine patterns of prevalence of the different viral species and the effects of such demographic variables. Given that numbers of positive samples were low, only univariate analyses could be performed and any differences that may be observed are interpreted with caution.

4.3.3 Co-carriage

Firstly, chi-squared and fisher's exact tests were used to compare the prevalence of co-carriage between samples from different individuals (NP versus nose, NP versus self-taken WMS, self-taken WMS versus HCP-taken WMS, nose versus HCP taken WMS). McNemar's Chi-squared test was used to compare samples from the same individuals (nose versus self-taken WMS and NP versus HCP-taken WMS). This analysis was designed to compare overall crude prevalence of co-carriage between methods in all individuals and hence does not incorporate stratification according to potential confounders such as age. This analysis would enable the identification of the most effective sampling method for co-carriage of species within the respiratory tract which would then be used to perform multivariate analysis. Confounders would be assessed within these subsequent multivariate analyses.

Subsequently, multivariate binary logistic regression analysis of carriage was performed for the bacterial species co-carriage as well as bacterial-viral co-carriage using the data collected from the swabbing or detection method deemed most effective. Multivariable analysis was informed by univariate regression analysis, as described previously for bacterial carriage (section 4.3.1). For viral species co-carriage, multivariable analysis could not be performed due to too few virus-positive samples, and therefore univariate analysis using Chi-squared and Fisher's Exact tests (for categorical variables) and the Mann-Whitney U-test (for continuous variables) were performed.

4.4 Carriage of Bacterial Species in the Respiratory Tract

4.4.1 Analysis of Swabbing Methodologies for Detecting the Prevalence of Bacterial Carriage

Bacterial carriage was assessed within each swabbing method in order to ascertain the variation in carriage between these methods and well as to determine the most effective swabbing method for detecting the presence of each of the bacterial species.

4.4.1.1 Swab Type

Carriage of each of the bacterial species was determined for all four swab types and compared using Chi-squared and Fisher's Exact tests (for samples from different individuals) or McNemar's Chi-squared test (for samples from the same individuals). Carriage prevalence of each bacterial species within the swab types is shown in Figure 9 to Figure 14, with *p*-values for each pairwise comparison.

Culture results demonstrated *S. pneumoniae* carriage to be highest in the nose and NP swabs, which were significantly higher than both mouth swabs. *M. catarrhalis* carriage was significantly higher in the HCP-taken WMS than all other swab types. There was no significant difference in *P. aeruginosa* carriage between any of the swab types. *S. aureus* carriage was significantly higher in the nose swab than any other swab type. Finally *H. influenzae* carriage was highest in the nose and NP, which were both significantly higher than the self-taken WMS but not the HCP taken WMS.

Real-time PCR results demonstrated *S. pneumoniae* and *H. influenzae* carriage to be higher in the nose compared to the NP swabs. However, carriage of *M. catarrhalis* and *S. aureus* did not differ significantly between nose and NP swabs. Furthermore, carriage of *P. aeruginosa* was significantly higher in NP swabs compared with nose swabs.

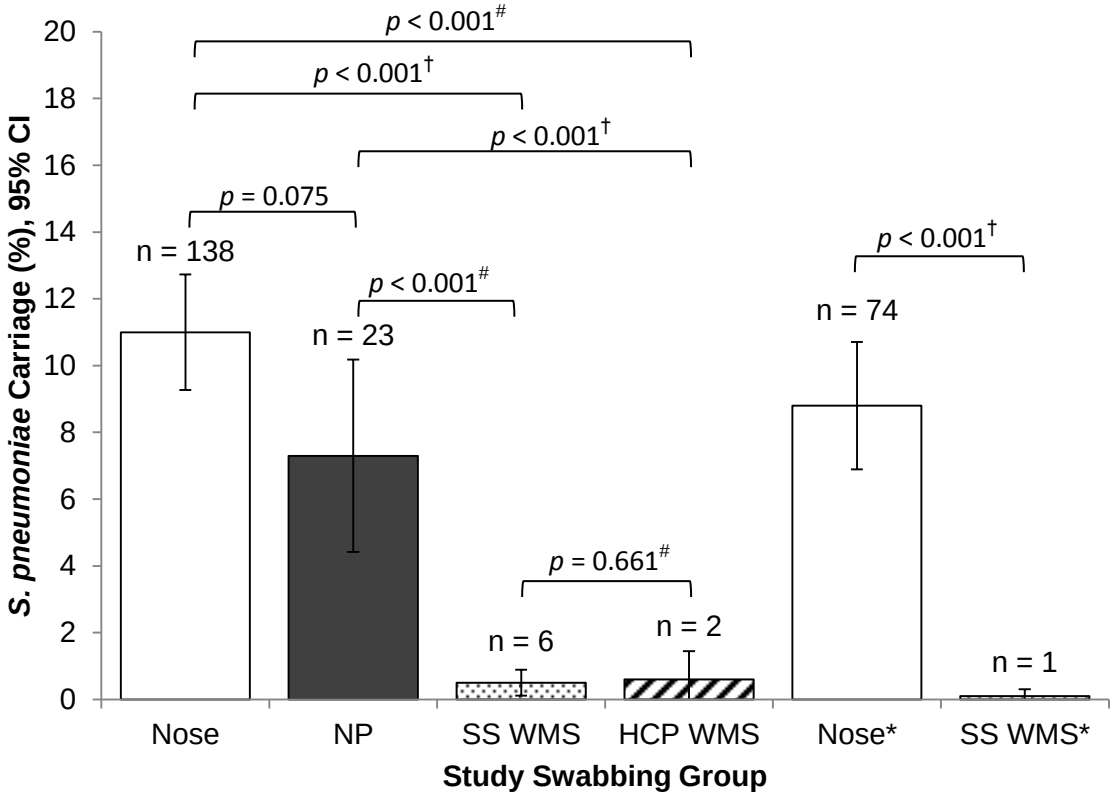


Figure 9. *S. pneumoniae* carriage prevalence by swab type detected by culture

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal, WMS = whole mouth swab, SS = self-taken swab, HCP = HCP-taken swab. *P*-values are two-tailed χ^2 p-values unless otherwise indicated: † McNemar's χ^2 *p*-value, # Fisher's exact test *p*-value.

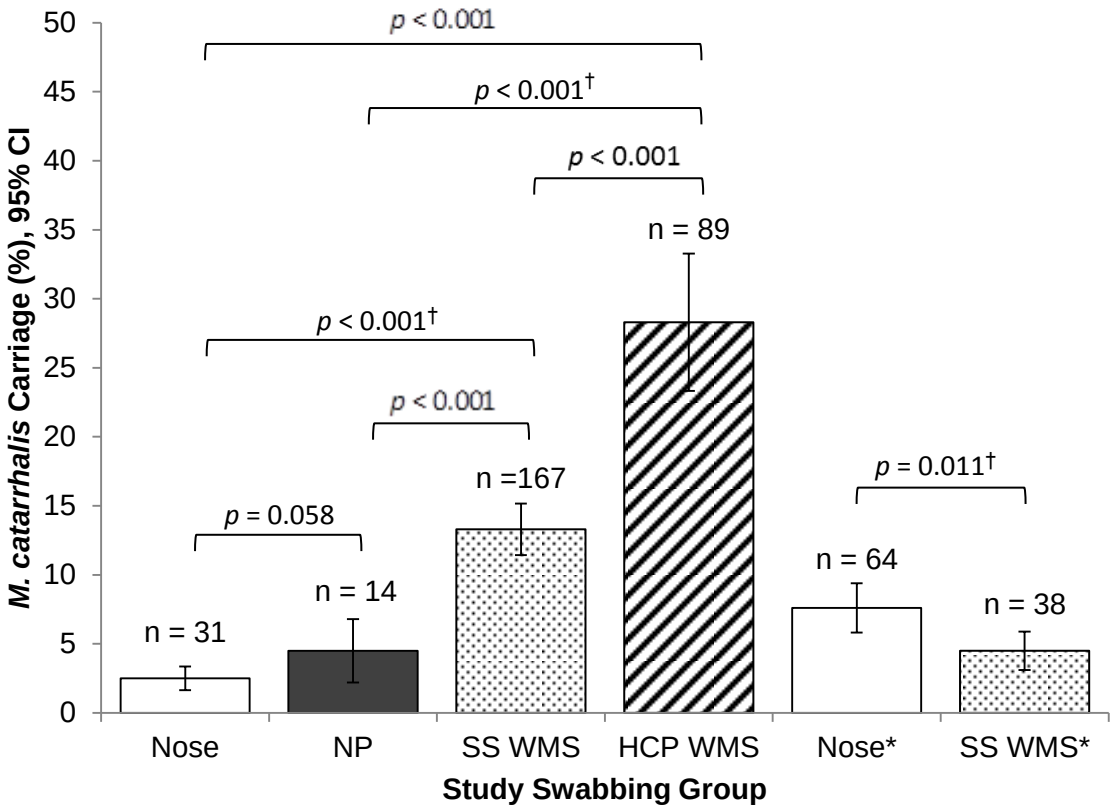


Figure 10. *M. catarrhalis* carriage prevalence by swab type detected by culture

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal, WMS = whole mouth swab, SS = self-taken swab, HCP = HCP-taken swab. *P*-values are two-tailed χ^2 p-values unless otherwise indicated: † McNemar's χ^2 *p*-value, # Fisher's exact test *p*-value.

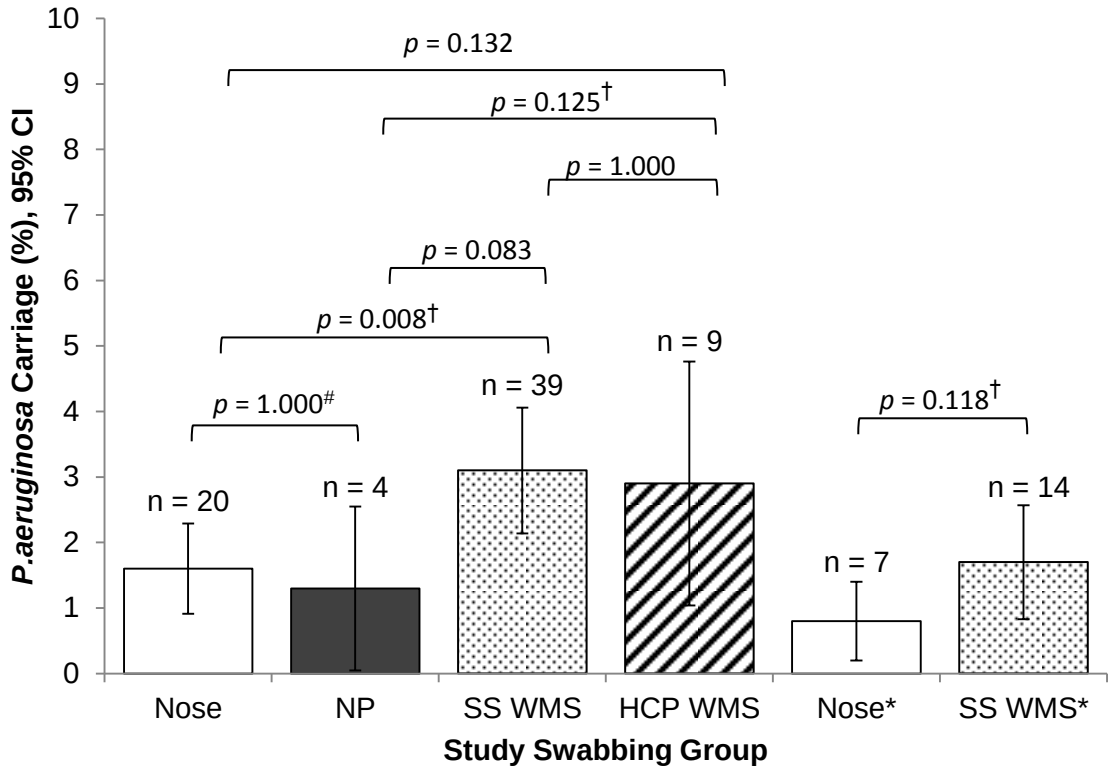


Figure 11. *P. aeruginosa* carriage prevalence by swab type detected by culture

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal, WMS = whole mouth swab, SS = self-taken swab, HCP = HCP-taken swab. *P*-values are two-tailed χ^2 p-values unless otherwise indicated: † McNemar's χ^2 *p*-value, # Fisher's exact test *p*-value.

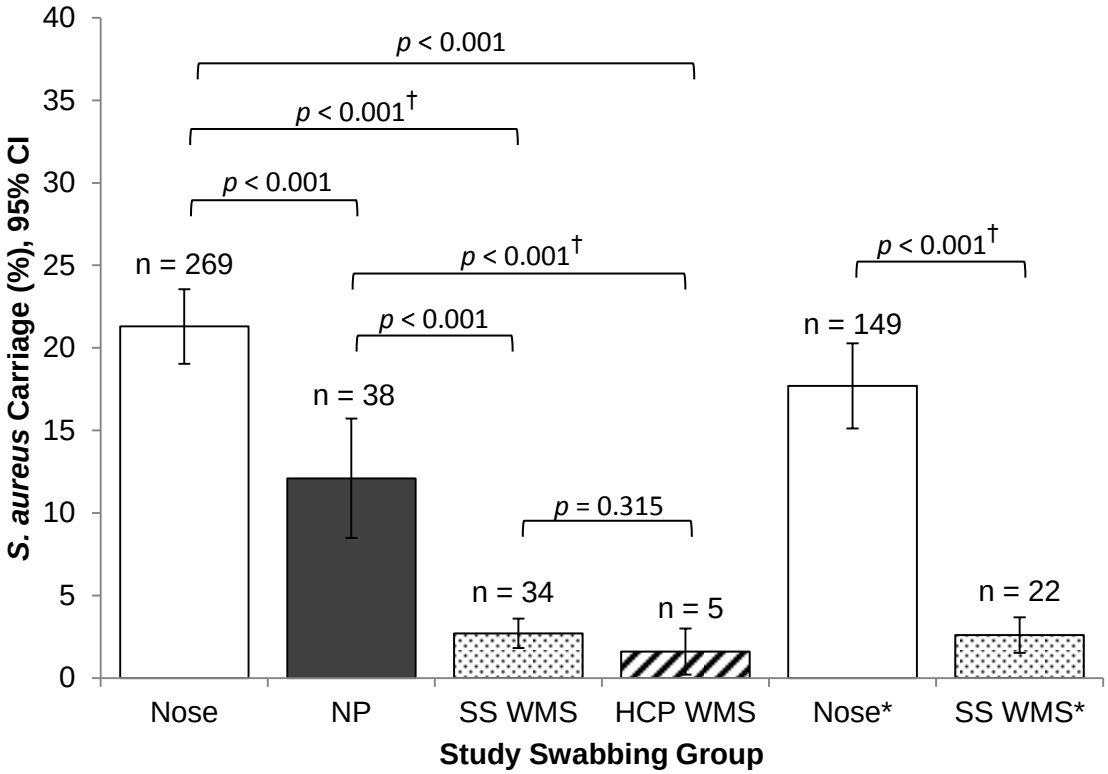


Figure 12. *S. aureus* carriage prevalence by swab type detected by culture

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal, WMS = whole mouth swab, SS = self-taken swab, HCP = HCP-taken swab. *P*-values are two-tailed χ^2 p-values unless otherwise indicated: † McNemar's χ^2 *p*-value, # Fisher's exact test *p*-value.

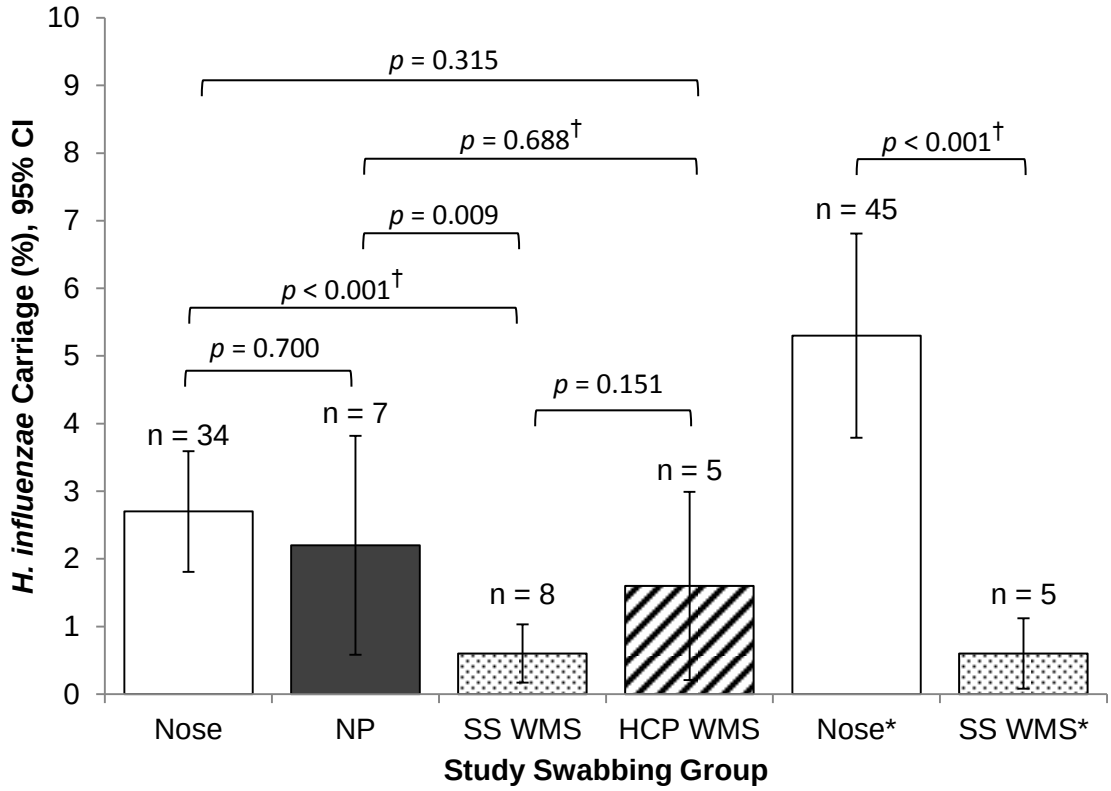


Figure 13. *H. influenzae* carriage prevalence by swab type detected by culture

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal, WMS = whole mouth swab, SS = self-taken swab, HCP = HCP-taken swab. *P*-values are two-tailed χ^2 p-values unless otherwise indicated: † McNemar's χ^2 *p*-value, # Fisher's exact test *p*-value.

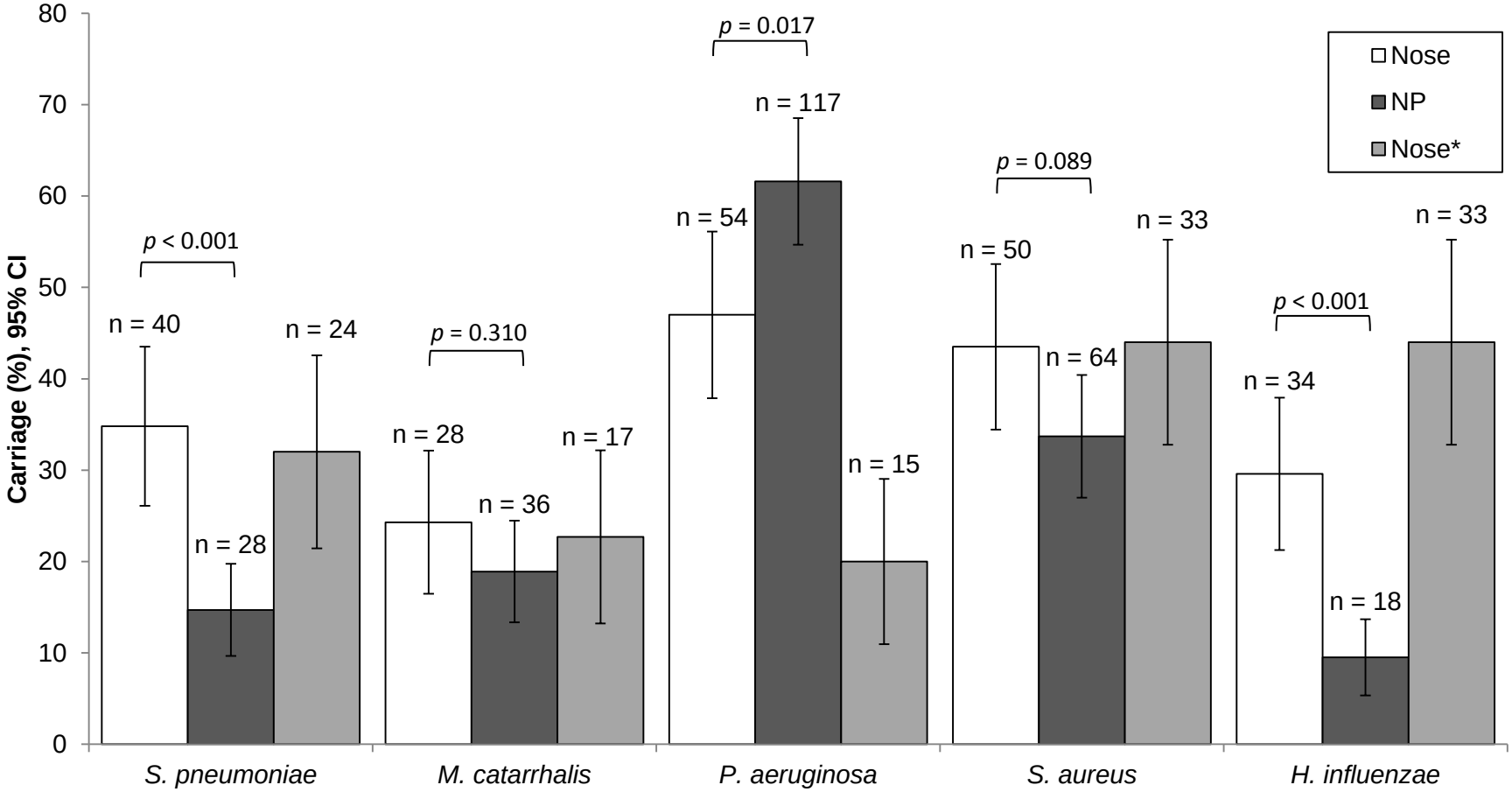


Figure 14. Carriage of bacterial species in nose and NP swabs detected by real-time PCR

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal. *P*-values are two-tailed χ^2 p-values.

4.4.1.2 Detection Method

Carriage of each of the bacterial species was compared between detection methods (culture and real-time PCR) for the subset of samples analysed by both culture and real-time PCR. Only nose and NP samples were analysed using real-time PCR, therefore no data is available for either WMS. This would enable the determination of any differences in the level of detection between these two methods. McNemar's Chi-squared test was used to compare results for culture and real-time PCR, which is shown in Table 8.

Table 8. Comparison of Culture and Real-time PCR Methods for the Detection of Bacterial Species using McNemar's Chi-squared Test

n (%)			Culture			
			Nose (N = 190)		NP (N = 190)	
			No	Yes	No	Yes
Real-time PCR	***S. pneumoniae***	No	125 (65.8)	1 (0.5)	161 (84.7)	1 (0.5)
		Yes	35 (18.4)	29 (15.3)	8 (4.2)	20 (10.5)
		p-value	<0.001		0.039	
	M. catarrhalis	No	140 (73.7)	5 (2.6)	151 (79.5)	3 (1.6)
		Yes	37 (19.5)	8 (4.2)	27 (14.2)	9 (4.7)
		p-value	<0.001		<0.001	
	P. aeruginosa	No	118 (62.1)	3 (1.6)	73 (38.4)	0 (0.0)
		Yes	67 (35.3)	2 (1.1)	114 (60.0)	3 (1.6)
		p-value	<0.001		<0.001	
	S. aureus	No	98 (51.6)	9 (4.7)	125 (65.8)	1 (0.5)
		Yes	41 (21.6)	42 (22.1)	45 (23.7)	19 (10.0)
		p-value	<0.001		<0.001	
	H. influenzae	No	121 (63.7)	2 (1.1)	170 (89.5)	2 (1.1)
		Yes	57 (30.0)	10 (5.3)	13 (6.8)	5 (2.6)
		p-value	<0.001		0.007	

NP = Nasopharyngeal. *P*-values are McNemar's Chi-squared *P*-values.

The results from McNemar's Chi-squared test show real-time PCR to detect more bacterial isolates than culture in all species tested. As such, high levels of culture-negative

PCR-positive samples were observed in all of these species. However, some culture-positive PCR-negative samples are also present.

4.4.2 Patterns of Carriage according to Participant Demographics

Culture results would be used for regression modelling due to the greater numbers of overall and positive samples available, which increases the statistical power of a mathematical model. Furthermore, the lack of detection of dead cells using culture detection, which may be problematic in PCR, prevents the assessment of potential false positive results. Furthermore, self-taken swabs would be used for regression modelling, again due to greater numbers of samples available compared with HCP-taken swabs and due to the non-inferiority of self-taken swabs in detecting carriage of these bacterial species compared to HCP-taken swabs (with the exception of *M. catarrhalis*). In addition, due to these advantages, self-swabbing was continued during the second time-point of the study and therefore differences in carriage between these time-points could be assessed accounting for swabbing season. Nose swabs were therefore chosen for *S. pneumoniae, H. influenzae* and *S. aureus* due to significantly higher prevalence of carriage in these swabs compared with self-taken WMS. *M. catarrhalis* and *P. aeruginosa* carriage, however, were assessed using self-taken WMS due to significantly higher prevalence of carriage in these swabs compared with nose swabs.

4.4.2.1 S. pneumoniae

Summary statistics and multivariable binary logistic regression modelling results for *S. pneumoniae* using culture-detected nose swabs are presented in Table 9. Covariates that were found to significantly increase the odds of *S. pneumoniae* carriage were recent RTI, *M. catarrhalis* carriage, *H. influenzae* carriage and geographical location (relative to Southampton/Romsey).Covariates that were found to significantly decrease the odds of *S. pneumoniae* carriage were age, winter season (relative to summer) and *S. aureus* carriage. The final model gave a pseudo-R^2 (Nalgerkerke) of 0.396. The area under the curve (AUC) for this final model was 0.885 (95% CI 0.862, 0.908) which was significantly different from 0.5 ($p < 0.001$), demonstrating that the model predicts *S. pneumoniae* carriage significantly better than chance.

Table 9. Logistic Regression Model of *S. pneumoniae* Carriage in Self-taken Nose Swabs

Covariate	N	*S. pneumoniae* Carriage		Crude		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)	-	7.3 (16.7)	42.2 (28.2)	0.934	<0.001	0.943 (0.932, 0.955)	<0.001
Recent RTI							
Yes	729	140 (19.2)	589 (80.8)	4.334 (3.192, 5.885)	<0.001	2.071 (1.448, 2.963)	<0.001
No	1308	68 (5.2)	1240 (94.8)	Ref.	-	Ref.	-
Recent Antibiotics							
Yes	179	15 (8.4)	164 (91.6)	0.790 (0.456, 1.368)	0.4000	-	-
No	1860	193 (10.4)	1667 (89.6)	Ref.	-	-	-
Vaccinations up-to-date							
Yes	1695	199 (11.7)	1496 (88.3)	3.104 (0.968, 9.951)	0.057	-	-
No	73	3 (4.1)	70 (95.9)	Ref.	-	-	-
Geographical Location							
1	487	31 (6.4)	456 (93.6)	Ref.	<0.001	Ref.	0.015
2	106	20 (18.9)	86 (81.1)	3.421 (1.863, 6.280)	<0.001	3.055 (1.480, 6.303)	0.003
3	400	58 (14.5)	342 (85.5)	2.495 (1.578, 3.944)	<0.001	2.224 (1.322, 3.741)	0.003
4	851	83 (9.8)	768 (90.2)	1.590 (1.036, 2.440)	0.034	1.621 (0.998, 2.633)	0.051
5	89	6 (6.7)	83 (93.3)	1.063 (0.430, 2.628)	0.894	1.052 (0.393, 2.817)	0.920
6	126	11 (8.7)	115 (91.3)	1.407 (0.687, 2.884)	0.351	1.589 (0.709, 3.560)	0.261
Season							
Summer	1254	138 (11.0)	1116 (89.0)	Ref.	-	Ref.	-
Winter	841	74 (8.8)	767 (91.2)	0.780 (0.580, 1.050)	0.101	0.550 (0.376, 0.804)	0.002
H. influenzae							
Yes	79	41 (51.9)	38 (48.1)	11.641 (7.288, 18.595)	<0.001	4.195 (2.460, 7.154)	<0.001
No	2016	171 (8.5)	1845 (91.5)	Ref.	-	Ref.	-
M. catarrhalis							
Yes	95	30 (31.6)	65 (68.4)	4.610 (2.914, 7.293)	<0.001	1.974 (1.130, 3.449)	0.017
No	2000	182 (9.1)	1818 (90.0)	Ref.	-	Ref.	-
S. aureus							
Yes	418	14 (3.3)	404 (96.7)	0.259 (0.149, 0.450)	<0.001	0.354 (0.193, 0.650)	0.001
No	1677	198 (11.8)	1479 (88.2)	Ref.	-	Ref.	-
P. aeruginosa							
Yes	27	5 (18.5)	22 (81.5)	0.808 (0.108, 6.016)	0.835	-	-
No	2068	207 (10.0)	1862 (90.0)	Ref.	-	-	-

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

4.4.2.2 M. catarrhalis

Summary statistics and multivariable binary logistic regression modelling results for *M. catarrhalis* using culture-detected self-taken WMS swabs are shown in Table 10. Covariates that were found to significantly increase the odds of *M. catarrhalis* carriage were *H. influenzae* carriage and geographical location (relative to Southampton/Romsey). Covariates that were found to significantly decrease the odds of *M. catarrhalis* carriage were recent RTI and winter season (relative to summer). The final model gave a pseudo-R^2 (Nalgerkerke) of 0.670. The area under the curve (AUC) for this final model was 0.670 (95% CI 0.630, 0.709) which was significantly different from 0.5 ($p < 0.001$), demonstrating that the model predicts *M. catarrhalis* carriage significantly better than chance.

Table 10. Logistic Regression Model of *M. catarrhalis* Carriage in Self-taken Whole Mouth Swabs

Covariate	N	*M. catarrhalis* Carriage		Crude		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)	-	41.5 (29.2)	38.3 (29.4)	1.004 (0.999, 1.009)	0.149	-	-
Recent RTI							
Yes	729	53 (7.3)	676 (92.7)	0.614 (0.442, 0.852)	0.004	0.688 (0.492, 0.964)	0.030
No	1307	148 (11.2)	1159 (88.7)	Ref.	-	Ref.	-
Recent Antibiotics							
Yes	178	19 (10.7)	159 (89.3)	1.095 (0.664, 1.805)	0.722	-	-
No	1860	183 (9.8)	1677 (90.2)	Ref.	-	-	-
Vaccinations up-to-date							
Yes	1694	166 (9.8)	1528 (90.2)	2.535 (0.789, 8.140)	0.118	-	-
No	73	3 (4.1)	70 (95.9)	Ref.	-	-	-
Geographical Location							
1	487	43 (8.8)	444 (91.2)	Ref.	0.022	Ref.	0.014
2	105	19 (18.1)	86 (81.9)	2.281 (1.268, 4.104)	0.006	2.262 (1.237, 4.138)	0.008
3	400	38 (9.5)	362 (90.5)	1.084 (0.686, 1.713)	0.730	1.007 (0.628, 1.615)	0.977
4	851	75 (8.8)	776 (91.2)	0.998 (0.674, 1.478)	0.992	1.129 (0.754, 1.692)	0.555
5	89	14 (15.7)	75 (84.3)	1.927 (1.005, 3.695)	0.048	2.188 (1.117, 4.287)	0.022
6	126	15 (11.9)	111 (88.1)	1.395 (0.748, 2.603)	0.295	1.797 (0.948, 3.409)	0.073
Season							
Summer	1254	167 (13.3)	1087 (86.7)				
Winter	840	38 (4.5)	802 (95.5)	0.308 (0.214, 0.444)	<0.001	0.311 (0.213, 0.453)	<0.001
S. pneumoniae							
Yes	7	0 (0.0)	7 (100.0)	-	-	-	-
No	2087	205 (9.8)	1882 (90.2)	-	-	-	-
H. influenzae							
Yes	13	5 (38.5)	8 (61.5)	5.878 (1.905, 18.139)	0.002	6.640 (1.972, 22.359)	0.002
No	2081	200 (9.6)	1881 (90.4)	Ref.	-	Ref.	-
S. aureus							
Yes	56	5 (8.9)	51 (91.1)	0.901 (0.355, 2.284)	0.826	-	-
No	2038	200 (9.8)	1838 (90.2)	Ref.	-	-	-
P. aeruginosa							
Yes	53	3 (5.7)	50 (94.3)	0.546 (0.169, 1.767)	0.313	-	-
No	2041	202 (9.9)	1839 (90.1)	Ref.	-	-	-

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

4.4.2.3 H. influenzae

Summary statistics and multivariable binary logistic regression modelling results for *H. influenzae* using culture-detected nose swabs are shown in Table 11. Covariates that were found to significantly increase the odds of *H. influenzae* carriage were *S. pneumoniae* carriage and winter season (relative to summer).The only covariate that was found to significantly decrease the odds of *H. influenzae* carriage was age. The final model gave a pseudo-R^2 (Nalgerkerke) of 0.295. The area under the curve (AUC) for this final model was 0.874 (95% CI 0.838, 0.910) which was significantly different from 0.5 ($p < 0.001$), demonstrating that the model predicts *H. influenzae* carriage significantly better than chance.

Table 11. Logistic Regression Model of *H. influenzae* Carriage in Self-taken Nose Swabs

Covariate	N	*H. influenzae* Carriage		*Crude*		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)	-	6.8 (14.1)	39.9 (29.1)	0.934 (0.916, 0.953)	<0.001	0.948 (0.930, 0.966)	<0.001
Recent RTI							
Yes	729	54 (7.4)	675 (92.6)	4.676 (2.824, 7.744)	<0.001	-	-
No	1308	22 (1.7)	1286 (98.3)	Ref.	-	-	-
Recent Antibiotics							
Yes	179	6 (3.4)	173 (96.6)	0.887 (0.380, 2.071)	0.781	-	-
No	1860	70 (3.8)	1790 (96.2)	Ref.	-	-	-
Vaccinations up-to-date							
Yes	1695	74 (4.4)	1621 (95.6)	3.287 (0.451, 23.979)	0.241	-	-
No	73	1 (1.4)	72 (98.6)	Ref.	-	-	-
Geographical Location							
1	487	13 (2.7)	474 (97.3)	Ref.	0.685	-	-
2	106	2 (2.8)	103 (97.2)	1.062 (0.297, 3.794)	0.926	-	-
3	400	18 (4.5)	382 (95.5)	1.718 (0.831, 3.551)	0.144	-	-
4	851	36 (4.2)	815 (95.8)	1.611 (0.846, 3.067)	0.147	-	-
5	89	3 (3.4)	86 (96.6)	1.272 (0.355, 4.557)	0.712	-	-
6	126	4 (3.2)	122 (96.8)	1.195 (0.383, 3.731)	0.759	-	-
Season							
Summer	1254	34 (2.7)	1220 (97.3)	Ref.	-	Ref.	-
Winter	841	45 (5.4)	796 (94.6)	2.029 (1.288, 3.195)	0.002	2.495 (1.529, 4.071)	<0.001
S. pneumoniae							
Yes	212	41 (19.3)	171 (80.7)	11.641 (7.288, 18.595)	<0.001	4.602 (2.755, 7.688)	<0.001
No	1883	38 (2.0)	1845 (98.0)	Ref.	-	Ref.	-
M. catarrhalis							
Yes	95	16 (16.8)	79 (83.2)	6.227 (3.441, 11.268)	<0.001	-	-
No	2000	63 (3.2)	1937 (96.9)	Ref.	-	-	-
S. aureus							
Yes	418	6 (1.4)	412 (98.6)	0.320 (0.138, 0.741)	0.008	-	-
No	1677	73 (4.4)	1604 (95.6)	Ref.	-	-	-
P. aeruginosa							
Yes	27	0 (0.0)	27 (100.0)	-	-	-	-
No	2068	79 (3.8)	1989 (96.2)	-	-	-	-

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

4.4.2.4 P. aeruginosa

Summary statistics and multivariable binary logistic regression modelling results for *P. aeruginosa* using culture-detected self-taken WMS are shown in Table 12. Covariates that were found to significantly decrease the odds of *P. aeruginosa* carriage were age and winter season (relative to summer). The final model gave a pseudo-R^2 (Nalgerkerke) of 0.024. The area under the curve (AUC) for this final model was 0.624 (95% CI 0.546, 0.701) which was significantly different from 0.5 (p = 0.003), demonstrating that the model predicts *P. aeruginosa* carriage significantly better than chance.

Table 12. Summary Statistics of *P. aeruginosa* Carriage in Self-taken Whole Mouth Swabs

Covariate	N	*P. aeruginosa* Carriage		Crude		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)	-	28.9 (30.6)	38.9 (29.3)	0.998 (0.979, 0.998)	0.018	0.989 (0.979, 0.998)	0.024
Recent RTI							
Yes	729	23 (3.2)	706 (96.8)	1.488 (0.851, 2.603)	0.164	-	-
No	1307	28 (2.1)	1279 (97.9)	Ref.	-	-	-
Recent Antibiotics							
Yes	178	5 (2.8)	173 (97.2)	1.140 (0.447, 2.906)	0.784	-	-
No	1860	46 (2.5)	1814 (97.5)	Ref.	-	-	-
Vaccinations up-to-date							
Yes	1694	44 (2.6)	1650 (97.4)	1.920 (0.261, 14.132)	0.522	-	-
No	73	1 (1.4)	72 (98.6)	Ref.	-	-	-
Geographical Location							
1	487	14 (2.9)	473 (97.1)	Ref.	0.506	-	-
2	105	3 (2.9)	102 (97.1)	0.994 (0.280, 3.521)	0.992	-	-
3	400	12 (3.0)	388 (97.0)	1.045 (0.478, 2.285)	0.912	-	-
4	851	17 (2.0)	834 (98.0)	0.689 (0.336, 1.410)	0.307	-	-
5	89	4 (4.5)	85 (95.5)	1.590 (0.511, 4.946)	0.423	-	-
6	126	1 (0.8)	125 (99.2)	0.270 (0.035, 2.075)	0.208	-	-
Season							
Summer	1254	39 (3.1)	1215 (96.9)	Ref.	-	Ref.	-
Winter	840	14 (1.7)	826 (98.3)	0.528 (0.285, 0.979)	0.043	0.525 (0.277, 0.992)	0.047
S. pneumoniae							
Yes	7	0 (0.0)	7 (100.0)	-	-	-	-
No	2087	53 (2.5)	2034 (97.5)	-	-	-	-
H. influenzae							
Yes	13	0 (0.0)	13 (100.0)	0.546 (0.169, 1.767)	0.313	-	-
No	2081	53 (2.5)	2028 (97.5)	Ref.	-	-	-
M. catarrhalis							
Yes	205	3 (1.5)	202 (98.5)	-	-	-	-
No	1889	50 (2.6)	1839 (97.4)	-	-	-	-
S. aureus							
Yes	56	3 (5.4)	53 (94.6)	2.251 (0.680, 7.446)	0.184	-	-
No	2038	50 (2.5)	1988 (97.5)	Ref.	-	-	-

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

4.4.2.5 S. aureus

Summary statistics and multivariable binary logistic regression modelling results for *S. aureus* using culture-detected nose swabs are shown in Table 13. Covariates that were found to significantly decrease the odds of *S. aureus* carriage were *S. pneumoniae* carriage and *M. catarrhalis* carriage. No covariates were found to significantly increase the odds of *S. aureus* carriage. The final model gave a pseudo-R^2 (Nalgerkerke) of 0.064. The area under the curve (AUC) for this final model was 0.557 (95% CI 0.528, 0.586) which was significantly different from 0.5 ($p < 0.001$), demonstrating that the model predicts *S. aureus* carriage significantly better than chance.

Table 13. Logistic Regression Model of *S. aureus* Carriage in Self-taken Nose Swabs

Covariate	N	*S. aureus* Carriage		Crude		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)	-	42.3 (27.0)	37.8 (29.9)	1.005 (1.002, 1.009)	0.006	-	-
Recent RTI							
Yes	729	132 (18.1)	597 (81.9)	0.838 (0.665, 1.056)	0.134	-	-
No	1308	273 (20.9)	1035 (79.1)	Ref.	-	-	-
Recent Antibiotics							
Yes	179	34 (19.0)	145 (81.0)	0.947 (0.641, 1.400)	0.786	-	-
No	1860	369 (19.8)	1491 (80.2)	Ref.	-	-	-
Vaccinations up-to-date							
Yes	1695	322 (19.0)	1373 (81.0)	0.835 (0.474, 1.474)	0.535	-	-
No	73	16 (21.9)	57 (78.1)	Ref.	-	-	-
Geographical Location							
1	487	90 (18.5)	397 (81.5)	Ref.	0.165	-	-
2	106	14 (13.2)	92 (86.6)	0.671 (0.366, 1.232)	0.198	-	-
3	400	92 (23.0)	308 (77.0)	1.318 (0.951, 1.826)	0.098	-	-
4	851	180 (21.2)	671 (78.8)	1.183 (0.893, 1.569)	0.242	-	-
5	89	14 (15.7)	75 (84.3)	0.823 (0.445, 1.523)	0.536	-	-
6	126	23 (18.3)	103 (81.7)	0.985 (0.593, 1.635)	0.953	-	-
Season							
Summer	1254	269 (21.5)	985 (78.5)	Ref.	-	-	-
Winter	841	149 (17.7)	692 (82.3)	0.788 (0.631, 0.985)	0.036	-	-
S. pneumoniae							
Yes	212	14 (6.6)	198 (93.4)	0.259 (0.149, 0.450)	<0.001	0.282 (0.162, 0.492)	<0.001
No	1883	404 (21.5)	1479 (78.5)	Ref.	-	Ref.	-
H. influenzae							
Yes	79	6 (7.6)	73 (92.4)	0.320 (0.138, 0.741)	0.008	-	-
No	2016	412 (20.4)	1604 (79.6)	Ref.	-	-	-
M. catarrhalis							
Yes	95	5 (5.3)	90 (94.7)	0.213 (0.086, 0.529)	0.001	0.258 (0.103, 0.642)	0.004
No	2000	413 (20.7)	1587 (79.4)	Ref.	-	Ref.	-
P. aeruginosa							
Yes	27	7 (25.9)	20 (74.1)	1.411 (0.593, 3.360)	0.437	-	-
No	2068	411 (19.9)	1657 (80.1)	Ref.	-	-	-

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

4.5 Prevalence of Viral Species in the Respiratory Tract

Overall, viruses were detected in 28.9% (n = 110) of samples using real-time PCR. Viral carriage was highest for rhinovirus/enterovirus (12.9%, n = 49), then RSV (9.5%, n = 36), coronavirus (6.8%, n = 26), adenovirus (1.8%, n = 7), influenza A virus (0.5%, n = 2) and finally parainfluenza virus (0.3%, n = 1). Metapneumovirus and influenza B virus were not detected in any of the samples. Due to small numbers of positive samples, multivariable analyses were not performed for these viral species. Therefore, only summary statistics and univariate analysis of species prevalence was performed.

4.5.1 Analysis of Swabbing Methodologies for Detecting the Prevalence of Viral Species

The differences in prevalence between swabbing methods are summarised in Table 14. Viral prevalence, as detected by real-time PCR did not differ between nose and NP swabs in any of the species tested. Low numbers of positive samples mean that confidence intervals for viral prevalence are wide and cannot allow the accurate determination of the prevalence of these species between swab types and swabbing season. Overlapping confidence intervals mean that comparisons between groups are difficult and any potential differences may not be discernible.

Table 14. Viral Carriage by Swabbing Method

Viral species % (n) (95% CI)	Summer 2012		Winter 2013
	Nose (N = 115)	**NP (N = 190)**	**Nose (N = 75)**
IVA	0 (0) NA	0 (0) NA	2.7 (2) (-1.0, 6.4)
RSV	6.1 (7) (1.7, 10.5)	12.1 (23) (7.5, 16.7)	8.0 (6) (1.9, 14.1)
ADV	4.3 (5) (0.6, 8.0)	0.5 (1) (-0.5, 1.5)	1.3 (1) (-1.3, 3.9)
RV/EV	16.5 (19) (9.7, 23.3)	12.6 (24) (7.9, 17.3)	8.0 (6) (1.9, 14.1)
COV	10.4 (12) (4.8, 16.0)	3.7 (7) (1.0, 6.4)	9.3 (7) (2.7, 15.9)
PIV	0 (0) NA	0 (0) NA	1.3 (1) (-1.3, 3.9)

NA = not applicable, NP = Nasopharynx, IVA = influenza virus A, RSV = respiratory syncytial virus, ADV = adenovirus, RV/EV = rhinovirus/enterovirus, COV = coronavirus, PIV = parainfluenza virus

4.5.2 Patterns of Viral Species Prevalence according to Participant Demographics

Nose swab results would be used for these analyses as greater numbers of positive samples were obtained in this swab compared to NP swabs. Furthermore, the nose swab was collected during both study time-points and hence seasonal variation could be analysed via this swab type.

4.5.2.1 Respiratory Syncytial Virus

Summary statistics and univariate analyses were undertaken for this species, as shown in Table 15. Without adjusting for other covariates, prevalence of RSV was found to differ significantly by geographical location. However, further samples would need to be collected in order to confirm these findings via multivariate analysis.

Table 15. Summary Statistics of RSV Prevalence in Nose Swabs

		RSV		
	N	**Yes n (%)**	**No n (%)**	**P-value**
Age (years)				
Mean (SD)	-	26.2 (28.1)	27.9 (29.0)	0.836[#]
Recent RTI				
Yes	77	4 (5.2)	73 (94.8)	0.566
No	113	9 (8.0)	104 (92.0)	
Recent Antibiotics				
Yes	19	1 (5.3)	18 (94.7)	1.000*
No	170	12 (7.1)	158 (92.9)	
Vaccinations up-to-date				
Yes	163	10 (6.1)	153 (93.9)	0.241*
No	4	1 (25.0)	3 (75.0)	
Geographical Location				
1	56	8 (14.3)	48 (85.7)	0.012*
2	10	1 (10.0)	9 (90.0)	
3	37	3 (8.1)	34 (91.9)	
4	68	0 (0.0)	68 (100.0)	
5	9	1 (11.1)	8 (88.9)	
6	10	0 (0.0)	10 (100.0)	
Season				
Summer	115	7 (6.1)	108 (93.9)	0.770
Winter	75	6 (8.0)	69 (92.0)	
S. pneumoniae				
Yes	64	5 (7.8)	59 (92.2)	0.764*
No	126	8 (6.3)	118 (93.7)	
H. influenzae				
Yes	67	7 (10.4)	60 (89.6)	0.227*
No	123	6 (4.9)	117 (95.1)	
M. catarrhalis				
Yes	45	5 (11.1)	40 (88.9)	0.193*
No	8	8 (5.5)	137 (94.5)	
S. aureus				
Yes	83	6 (7.2)	77 (92.8)	1.000
No	107	7 (6.5)	100 (93.5)	
P. aeruginosa				
Yes	69	2 (2.9)	67 (97.1)	0.139*
No	121	11 (9.1)	110 (90.9)	
ADV				
Yes	6	0 (0.0)	6 (100.0)	1.000*
No	184	13 (7.1)	171 (92.9)	
RV/EV				
Yes	25	2 (8.0)	23 (92.0)	0.682*
No	165	11 (6.7)	154 (93.3)	
COV				
Yes	19	0 (0.0)	19 (100.0)	0.369*
No	171	13 (7.6)	158 2.4)	

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, [#]Mann Whitney Test.

4.5.2.2 Adenovirus

Summary statistics and univariate analyses were undertaken for this species, as shown in Table 16. Without adjusting for other covariates, prevalence of adenovirus was found to differ significantly in the presence of *S. pneumoniae, H. influenzae* and *M. catarrhalis*. However, further samples would need to be collected in order to confirm these findings via multivariate analysis.

Table 16. Summary Statistics of Adenovirus Prevalence in Nose Swabs

		ADV		
	N	**Yes n (%)**	**No n (%)**	**P-value**
Age (years)				
Mean (SD)	-	27.2 (39.2)	27.8 (28.6)	0.450#
Recent RTI				
Yes	77	3 (3.9)	74 (96.1)	0.688*
No	113	3 (2.7)	110 (97.3)	
Recent Antibiotics				
Yes	19	0 (0.0)	19 (100.0)	1.000*
No	170	6 (3.5)	164 (96.5)	
Vaccinations up-to-date				
Yes	163	6 (3.7)	157 (96.3)	1.000*
No	4	0 (0.0)	4 (100.0)	
Geographical Location				
1	56	1 (1.8)	55 (98.2)	
2	10	1 (10.0)	9 (90.0)	
3	37	0 (0.0)	37 (100.0)	
4	68	4 (5.9)	64 (94.1)	0.378*
5	9	0 (0.0)	9 (100.0)	
6	10	0 (0.0)	10 (100.0)	
Season				
Summer	115	5 (4.3)	110 (95.7)	0.406*
Winter	75	1 (1.3)	74 (98.7)	
S. pneumoniae				
Yes	64	5 (7.8)	59 (92.2)	0.017*
No	126	1 (0.8)	125 (99.2)	
H. influenzae				
Yes	67	5 (7.5)	62 (92.5)	0.021*
No	123	1 (0.8)	122 (99.2)	
M. catarrhalis				
Yes	45	4 (8.9)	41 (91.1)	0.029*
No	8	2 (1.4)	143 (98.6)	
S. aureus				
Yes	83	1 (1.2)	82 (98.8)	0.234*
No	107	5 (4.7)	102 (95.3)	
P. aeruginosa				
Yes	69	1 (1.4)	68 (98.6)	0.420*
No	121	5 (4.1)	116 (95.9)	
RSV				
Yes	6	0 (0.0)	13 (100.0)	1.000*
No	184	6 (3.4)	171 (96.6)	
RV/EV				
Yes	25	2 (8.0)	23 (92.0)	0.179*
No	165	4 (2.4)	161 (97.6)	
COV				
Yes	19	0 (0.0)	19 (100.0)	1.000*
No	171	6 (3.5)	165 (96.5)	

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, #Mann Whitney Test.

4.5.2.3 Rhinovirus/enterovirus

Summary statistics and univariate analyses were undertaken for this species, as shown in Table 17. Without adjusting for other covariates, prevalence of Rhinovirus/Enterovirus was found to differ significantly with age, recent RTI as well as in the presence of *S. pneumoniae* and *H. influenzae*. However, further samples would need to be collected in order to confirm these findings via multivariate analysis.

Table 17. Summary Statistics of Rhinovirus/Enterovirus Carriage in Nose Swabs

	N	RV/EV Yes n (%)	RV/EV No n (%)	P-value
Age (years)				
Mean (SD)	-	10.9 (20.9)	30.4 (29.1)	<0.001#
Recent RTI				
Yes	77	15 (19.5)	62 (80.5)	0.048
No	113	10 (8.8)	103 (91.2)	
Recent Antibiotics				
Yes	19	0 (0.0)	19 (100.0)	0.082*
No	170	25 (14.7)	145 (985.3)	
Vaccinations up-to-date				
Yes	163	22 (13.5)	141 (86.5)	0.450*
No	4	1 (25.0)	3 (75.0)	
Geographical Location				
1	56	9 (16.1)	47 (83.9)	0.431*
2	10	0 (0.0)	10 (100.0)	
3	37	7 (18.9)	30 (81.1)	
4	68	7 (10.3)	61 (89.7)	
5	9	0 (0.0)	9 (100.0)	
6	10	2 (20.0)	8 (80.0)	
Season				
Summer	115	19 (16.5)	96 (83.5)	0.124
Winter	75	6 (8.0)	69 (92.0)	
S. pneumoniae				
Yes	64	16 (25.0)	48 (75.0)	0.001
No	126	9 (7.1)	117 (92.9)	
H. influenzae				
Yes	67	13 (19.4)	54 (80.6)	0.073
No	123	12 (9.8)	111 (90.2)	
M. catarrhalis				
Yes	45	13 (28.9)	32 (71.1)	0.001
No	8	12 (8.3)	133 (91.7)	
S. aureus				
Yes	83	8 (9.6)	75 (90.4)	0.280
No	107	17 (15.9)	90 (84.1)	
P. aeruginosa				
Yes	69	13 (18.8)	56 (81.2)	0.117
No	121	12 (9.9)	109 (90.1)	
RSV				
Yes	6	2 (15.4)	11 (84.6)	0.179*
No	184	23 (13.0)	154 (87.0)	
ADV				
Yes	25	2 (33.3)	4 (66.7)	0.682*
No	165	23 (12.5)	161 (87.5)	
COV				
Yes	19	5 (26.3)	14 (73.7)	0.142*
No	171	20 (11.7)	151 (88.3)	

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, #Mann Whitney Test.

4.5.2.4 Coronavirus

Summary statistics and univariate analyses were undertaken for this species, as shown in Table 18. Without adjusting for other covariates, prevalence of coronavirus was found to differ significantly with recent RTI, by geographical location and in the presence of *S. pneumoniae* and *S. aureus*. However, further samples would need to be collected in order to confirm these findings via multivariate analysis.

Table 18. Summary Statistics of Coronavirus Carriage in Nose Swabs

	N	COV		P-value
		Yes n (%)	No n (%)	
Age (years)				
Mean (SD)	-	16.6 (25.0)	29.1 (29.1)	0.100#
Recent RTI				
Yes	77	13 (16.9)	64 (83.1)	0.013
No	113	6 (5.3)	107 (94.7)	
Recent Antibiotics				
Yes	19	1 (5.3)	18 (94.7)	0.699*
No	170	18 (10.6)	152 (89.4)	
Vaccinations up-to-date				
Yes	163	15 (9.2)	148 (90.8)	1.000*
No	4	0 (0.0)	4 (100.0)	
Geographical Location				
1	56	7 (12.5)	49 (87.5)	0.038*
2	10	0 (0.0)	10 (100.0)	
3	37	8 (21.6)	29 (78.4)	
4	68	2 (2.9)	66 (97.1)	
5	9	1 (11.1)	8 (88.9)	
6	10	1 (10.0)	9 (90.0)	
Season				
Summer	115	12 (10.4)	103 (89.6)	1.000
Winter	75	7 (9.3)	68 (90.7)	
S. pneumoniae				
Yes	64	12 (18.8)	52 (81.3)	0.006
No	126	7 (5.6)	119 (94.4)	
H. influenzae				
Yes	67	9 (13.4)	58 (86.6)	0.312
No	123	10 (8.1)	113 (91.9)	
M. catarrhalis				
Yes	45	8 (17.8)	37 (82.2)	0.083*
No	8	11 (7.6)	134 (92.4)	
S. aureus				
Yes	83	16 (19.3)	67 (80.7)	<0.001
No	107	3 (2.8)	104 (97.2)	
P. aeruginosa				
Yes	69	7 (10.1)	62 (89.9)	1.000
No	121	12 (9.9)	109 (90.1)	
RSV				
Yes	6	0 (0.0)	13 (100.0)	0.369*
No	184	19 (10.7)	158 (89.3)	
ADV				
Yes	25	0 (0.0)	6 (100.0)	1.000*
No	165	19 (10.3)	165 (89.7)	
RV/EV				
Yes	19	5 (20.0)	20 (80.0)	0.142*
No	171	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, #Mann Whitney Test.

4.6 Co-carriage of Microbial Species in the Respiratory Tract

Co-carriage, also known as co-colonisation, is the presence of two or more species within the same sample site. Samples containing co-carried species were identified from culture and real-time PCR results. In samples assessed by culture, 5.0% (n = 121) of all samples demonstrated co-carriage of bacterial species. In samples assessed by real-time PCR, 51.1% (n = 194) of all samples demonstrated co-carriage of bacterial and/or viral species. The co-carriage of multiple viruses (2.9%, n = 11) was less common than the co-carriage of multiple bacteria (40.8%, n = 155). The co-carriage of at least one bacterial species and one viral species was also common (25.8%, n = 98).

4.6.1 Analysis of Swabbing Methodologies for Detecting Co-carriage

Co-carriage was assessed within each swabbing method in order to ascertain the variation in co-carriage between these methods and well as to determine the most effective method in detecting the presence of multiple microbial species within the upper respiratory tract.

Culture results demonstrated co- carriage of bacterial species to be highest in the nose swabs, which was significantly higher than both mouth swabs but not significantly higher than NP swabs. Real-time PCR results demonstrated co-carriage of viral species was higher in nose swabs, however, due to small numbers of isolates the significance of the differences between these swab types cannot be determined. Furthermore, bacterial and viral co-colonisation (at least one bacteria and one virus), as detected by real-time PCR was not significantly different between swab types. The differences in co-carriage between swabbing methods are summarised in Figure 15.

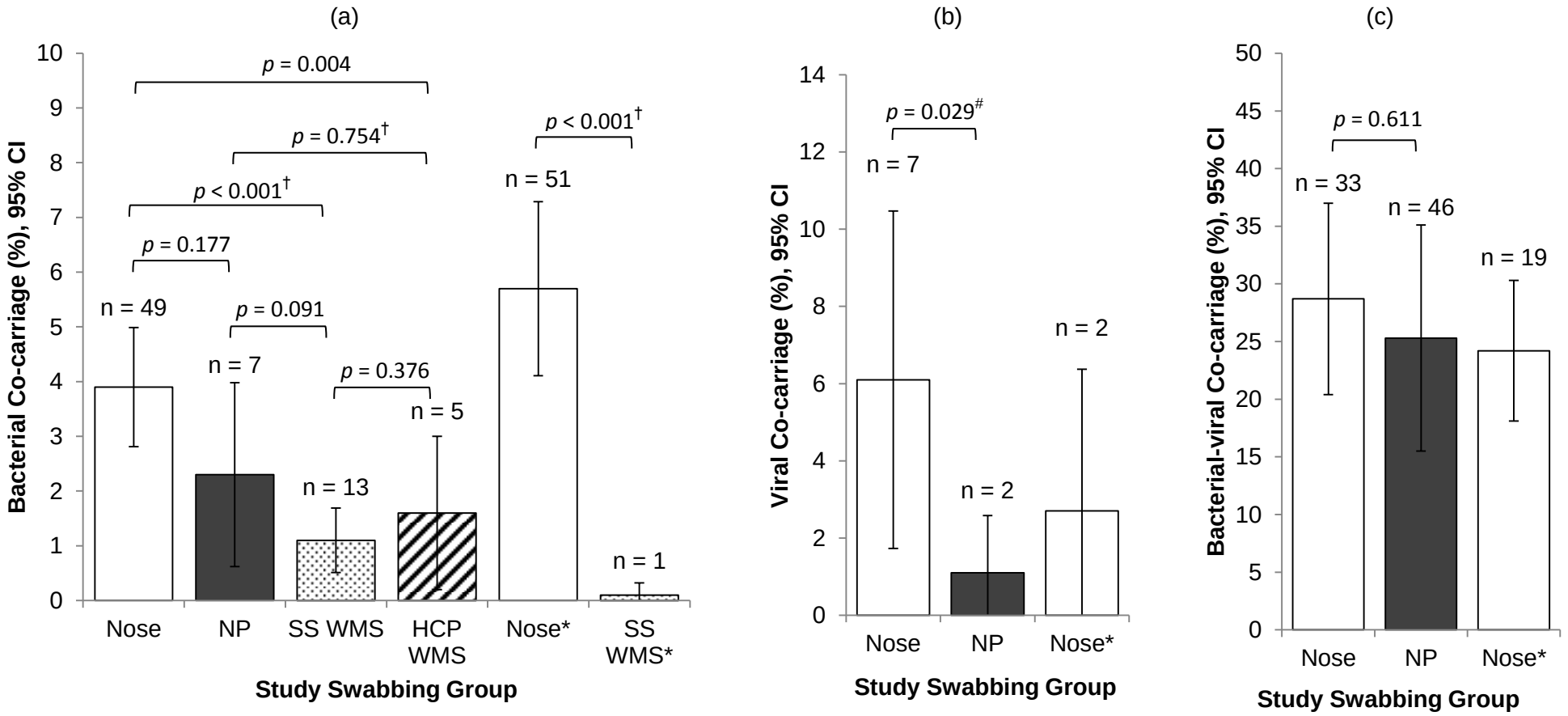


Figure 15. Co-carriage by Swab Type for (a) Bacteria detected by culture, (b) Viruses detected by real-time PCR and (c) Bacteria & viruses detected by real-time PCR

Brackets demonstrate pairwise comparison of swab types. Swabs samples were collected in summer and winter (*) seasons. NP = nasopharyngeal, WMS = whole mouth swab, SS = self-taken, HCP = HCP-taken. *P*-values are two-tailed χ^2 p-values unless otherwise indicated: † McNemar's χ^2 *p*-value, # Fisher's exact *p*-value.

4.6.2 Patterns of Co-carriage according to Participant Demographics

Co-carriage was assessed in relation to participant demographics in order to ascertain the variation in co-carriage between individuals and well as to determine the effects of such demographic variables. For bacterial co-carriage, nose swabs detected by culture were analysed due to the greater numbers of samples available for this swab type and detection method as well as greater prevalence of co-carriage within nose swabs when compared to other culture-detected swab types. For viral co-carriage, PCR-detected nose swabs were analysed due to the greater number of virus-positive samples obtained from this swab type compared to NP swabs. Finally for bacterial and viral co-carriage, nose swabs detected by real-time PCR were used as greater numbers of virus-positive samples were obtained from this swab type compared to NP swabs. Co-carriage of viral species was less common and therefore multivariable analysis could not be performed.

A multivariable binary logistic regression model was constructed for co-carriage of bacterial species in nose swabs detected by culture. Multivariable binary logistic regression results are shown in Table 19. The only covariate to significantly increase the odds of bacterial co-carriage was recent RTI. Furthermore, the only covariate to significantly decrease the odds of bacterial co-carriage was age. The final model gave a pseudo-R^2 (Nalgerkerke) of 0.186. The area under the curve (AUC) for this final model was 0.806 (95% CI 0.763, 0.849) which was significantly different from 0.5 ($p < 0.001$), demonstrating that the model predicts bacterial co-carriage significantly better than chance.

Descriptive statistics and univariate analyses of viral species co-carriage were performed using PCR-detected nose swabs, as shown in Table 20. Without adjusting for other covariates, prevalence of multiple viruses was found to differ significantly by participant age. However, further samples would need to be collected in order to confirm these findings via multivariate analysis.

A multivariable binary logistic regression model was constructed for co-carriage of bacterial and viral species in nose swabs detected by real-time PCR. Multivariable binary logistic regression results are shown in Table 21. The only covariate to significantly increase the odds of co-carriage was recent RTI. Furthermore, the only covariate to significantly decrease the odds of co-carriage was age. The final model gave a pseudo-R^2 (Nalgerkerke) of 0.119. The area under the curve (AUC) for this final model was 0.702 (95% CI 0.619, 0.785) which was significantly different from 0.5 ($p < 0.001$),

demonstrating that the model predicts co-carriage of viral and bacterial species significantly better than chance.

Table 19. Logistic Regression Model for Bacterial Co-carriage in Culture-detected Nose Swabs

Covariate	N	Bacterial Co- Carriage		Crude		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)	-	10.9 (20.3)	40.0 (29.1)	0.954 (0.943, 0.966)	<0.001	0.959 (0.948, 0.970)	<0.001
Recent RTI							
Yes	729	67 (9.2)	662 (90.8)	4.627 (2.947, 7.263)	<0.001	2.519 (1.612, 3.938)	<0.001
No	1308	28 (2.1)	1280 (97.9)	Ref.	-	Ref.	-
Recent Antibiotics							
Yes	179	11 (6.1)	168 (93.9)	1.384 (0.724, 2.647)	0.325	-	-
No	1860	84 (4.5)	1776 (95.5)	Ref.	-	-	-
Vaccinations up-to-date							
Yes	1695	88 (5.2)	1607 (94.8)	3.943 (0.542, 28.704)	0.176	-	-
No	73	1 (1.4)	72 (98.6)	Ref.	-	-	-
Geographical Location							
1	487	16 (3.3)	471 (96.7)	Ref.	0.093	-	-
2	106	8 (7.5)	98 (92.5)	2.403 (1.001, 5.771)	0.050	-	-
3	400	27 (6.8)	373 (93.3)	2.131 (1.131, 4.013)	0.019	-	-
4	851	39 (4.6)	812 (95.4)	1.414 (0.782, 2.558)	0.252	-	-
5	89	3 (3.4)	86 (96.6)	1.027 (0.293, 3.600)	0.967	-	-
6	126	3 (2.4)	123 (97.6)	0.718 (0.206, 2.503)	0.603	-	-
Season							
Summer	1254	49 (3.9)	1205 (96.1)	Ref.	-	-	-
Winter	841	51 (6.1)	790 (93.9)	1.588 (1.062, 2.374)	0.024	-	-

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

Table 20. Summary Statistics of Viral Co-carriage Prevalence in Real-time PCR-detected Nose Swabs

Covariate	N	Viral Co- Carriage		P-value
		Yes n (%)	No n (%)	
Age (years)				
Mean (SD)	-	2.4 (2.3)	29.1 (30.0)	<0.001#
Recent RTI				
Yes	77	4 (5.2)	73 (94.8)	1.000*
No	113	5 (4.4)	108 (95.6)	
Recent Antibiotics				
Yes	19	0 (0.0)	19 (100.0)	0.602*
No	170	9 (5.3)	161 (94.7)	
Vaccinations up-to-date				
Yes	163	8 (4.9)	155 (95.1)	1.000*
No	4	0 (0.0)	4 (100.0)	
Geographical Location				
1	56	3 (5.4)	53 (94.6)	0.562*
2	10	0 (0.0)	10 (100.0)	
3	37	4 (10.8)	33 (89.2)	
4	68	2 (2.9)	66 (97.1)	
5	9	0 (0.0)	9 (100.0)	
6	10	0 (0.)	10 (100.0)	
Season				
Summer	115	7 (6.1)	108 (93.9)	0.487*
Winter	75	2 (2.7)	73 (97.3)	

SD = standard deviation; RTI = Respiratory tract infection. Geographical areas are listed in methods section. P-values are 2-tailed Chi-squared P-values unless otherwise indicated, *Fisher's Exact Test, #Mann Whitney Test.

Table 21. Logistic Regression Model for Bacterial and Viral Co-carriage in Real-time PCR-detected Nose Swabs

Covariate	N	Bacterial Co- Carriage		Crude		Multivariate-adjusted	
		Yes n (%)	No n (%)	OR (95% CI)	P-value	OR (95% CI)	P-value
Age (years)							
Mean (SD)		17.5 (25.8)	31.7 (29.1)	0.980 (0.968, 0.993)	0.003	0.985 (0.971, 0.998)	0.024
Recent RTI							
Yes	77	31 (40.3)	46 (59.7)	2.952 (1.530, 5.697)	0.001	2.381 (1.201, 4.721)	0.013
No	113	21 (18.6)	92 (81.4)	Ref.	-	Ref.	-
Recent Antibiotics							
Yes	19	3 (15.8)	16 (84.2)	0.463 (0.129, 1.660)	0.237		
No	170	49 (28.8)	121 (71.2)	Ref.	-		
Vaccinations up-to-date							
Yes	163	44 (27.0)	119 (73.0)	0.370 (0.051, 2.706)	0.327		
No	4	2 (50.0)	2 (50.0)	Ref.	-		
Geographical Location							
1	56	19 (33.9)	37 (66.1)	Ref.	0.244		
2	10	2 (20.0)	8 (80.0)	0.487 (0.094, 2.523)	0.391		
3	37	14 (37.8)	23 (62.2)	1.185 (0.499, 2.814)	0.700		
4	68	12 (17.6)	56 (82.4)	0.417 (0.181, 0.960)	0.040		
5	9	2 (22.2)	7 (77.8)	0.556 (0.105, 2.944)	0.490		
6	10	3 (30.0)	7 (70.0)	0.835 (0.194, 3.598)	0.808		
Season							
Summer	115	33 (28.7)	82 (71.3)	Ref.	-		
Winter	75	19 (25.3)	56 (74.7)	0.843 (0.436, 1.629)	0.612		

OR = odds ratio; 95% CI = 95% confidence intervals; SD = standard deviation; RTI = respiratory tract infection. Geographical areas are listed in methods section. Ref. = reference category.

4.7 Discussion

The results from this swabbing study demonstrate clear differences between swabbing methodologies as well as according to different participant demographics, highlighting the variability in carriage between swab sites and individuals. The study aimed to determine the swab type which would allow the greatest detection of the target species and hence determine patterns of carriage amongst individuals with the greatest statistical power.

The current gold standard for detection of *S. pneumoniae* within the respiratory tract is an NP swab which is cultured onto selective nutrient agar (Satzke et al., 2013). This study demonstrated self-swabbing to be non-inferior to HCP swabbing in detecting many of the target organisms. Nose swabs were as effective as NP swabs in detecting *S. pneumoniae, M. catarrhalis, S. aureus, P. aeruginosa, H. influenzae*, adenovirus and coronavirus. This demonstrates that nose swabs might be an acceptable alternative to HCP-taken NP swabs. A previous study in a paediatric population demonstrated levels of *S. pneumoniae* and *H. influenzae* to be similar in nose and NP swabs (Rapola et al., 1997, Ridgway et al., 1995). There is also evidence supporting the use of self-taken nasal swabs for detection of respiratory viruses (Smieja et al., 2010, Akmatov et al., 2012). These findings would mean a decrease in study costs as well as increased levels of participation within large swabbing studies (Gamblin et al., 2013).

On the other hand, whole mouth swabs detected greater carriage of *M. catarrhalis* and *P. aeruginosa* than nose or NP swabs. This swab type would therefore be recommended for studying carriage of these two species. Mouth sampling methods have previously been shown to be useful in detecting species such as *N. meningitidis* as throat swabs are more difficult to perform (Jordens et al., 2002). The low number of *N. meningitidis* isolates within this study is potentially a result of swab type, as the oropharynx is the optimal sampling site for this organism (Olcen et al., 1979). Furthermore the distribution of responders may have limited the number of *N. meningitidis* isolates as teenagers, the most frequently colonised age group, responded in fewer numbers (n = 92 aged 13-19 years) than other age groups (Cartwright et al., 1987).

PCR results demonstrated higher levels of carriage within nose and NP swabs than detected by culture. This may be due to a number of factors including the lower limit of detection, the detection of non-viable cells' nucleic acid as well as potential detection of closely related species such as *S. pseudopneumoniae* within PCR assays (Satzke et al., 2013). Both false-positives and false-negatives may occur in real-time PCR detection with

problems such as environmental contamination as well as PCR inhibiting agents (Yang and Rothman, 2004). Furthermore, the volume of STGG may also have contributed to higher carriage prevalence using real-time PCR-detection as 10µl of STGG was used for culture compared with 200µl of STGG for real-time PCR. However, the volume of STGG used in culture is acceptable for detecting *S. pneumoniae* carriage according to the World Health Organization Pneumococcal Carriage Working Group recommendations and offers greater swab sample preservation over direct plating of swab samples (Satzke et al., 2013). Nevertheless, increased sample volume would allow greater sensitivity for detection by culture (Satzke et al., 2013). The use of different volumes within culture and real-time PCR methods and the detection of live cells in culture methods versus nucleic acid in real-time PCR means that the comparison of these methods is flawed and may have resulted in culture results with artificially low prevalence of bacterial species.

The study of clinical infection, such as meningitis and septicaemia, where microorganisms can be identified in usually sterile sites, benefits from the use of non-culture methods such as real-time PCR. In such cases, antibiotics are rapidly administered to patients leading to culture-negative results. However, real-time PCR is able to detect the presence of *S. pneumoniae, H. influenzae* and *N. meningitidis* in antibiotic-treated individuals enabling the aetiological agent of disease to be identified (Wu et al., 2013a). Furthermore, viral infection diagnosis, which was traditionally carried out via cell culture, also benefits from greater sensitivity and rapidity of real-time PCR (Heim et al., 2003). However, false-positives in non-sterile sites such as the upper and lower respiratory tract are problematic when studying infection as their clinical significance may be questionable (Guclu et al., 2005).

Within this study, however, the small sample size (n=380) and the non-random method for obtaining the sample for use in PCR are limitations. Fewer samples result in reduced statistical power for detecting differences between groups and for the mathematical modelling of carriage. The sample size for real-time PCR was chosen in order to perform preliminary analysis of samples using ecological methods (Chapter 6) and hence does not have the same power for the precise analysis of carriage prevalence that is offered by culture-detected samples. The non-random sample used for real-time PCR may result in biased results and hence results should be interpreted with caution. For this reason, where possible, culture results were used in favour of PCR for describing and analysing the epidemiology of carriage.

Viral species were found to be present in fewer respiratory tract samples than bacterial species. This might be a result of viruses being associated with infection rather than being

carried asymptomatically for long periods of time. Rhinovirus/enterovirus, coronavirus and RSV were the most commonly detected viral species within samples. High levels of rhinovirus, coronavirus and RSV in young children have previously been observed (van den Bergh et al., 2012, Tanner et al., 2012, van Benten et al., 2003). Lower levels of viral detection within this study may be due to the larger age range of participants as well as the swab types and transport media used, which were not optimised for virus survival but rather bacterial survival. Flocked swabs have been shown to have high efficacy in viral detection as they enable greater numbers of cells from the nasal epithelium, which harbour the viruses, to be gathered (Daley et al., 2006).

Age was an important factor in both bacterial and viral colonisation. *S. pneumoniae, H. influenzae, P. aeruginosa* and rhinovirus/enterovirus were all found to significantly decrease with increasing age. A study of healthy participants aged <7 years, 7-15 years and >16 years showed decreased carriage of *S. pneumoniae, H. influenzae* and *M. catarrhalis* with age (Gunnarsson et al., 1998). Furthermore rhinovirus infections are common in young children (Korppi et al., 2004). The high prevalence of these species within the young is thought to be a result of the increased chances of transmission of the organisms between hosts within this age group (Bogaert et al., 2004a). Transmissibility and carriage have been demonstrated to occur at higher rates in environments such as nurseries where children are able to interact and facilitate the spread of microbial species (Sa-Leado et al., 2008).

Recent respiratory tract infection was associated with increased carriage of *S. pneumoniae,* rhinovirus/enterovirus and coronavirus as well as decreased carriage of *M. catarrhalis*. One proposed explanation is that viral RTI can damage the protective barriers of the upper respiratory tract as well as diverting immune cells hence allowing increased numbers of bacteria and viruses to colonise (Brundage, 2006). This positive association between viral infection and bacterial colonisation is thought to enable the increased transmission of carried species such as *S. pneumoniae* via symptoms of rhinitis (Klugman et al., 2009, Rodrigues et al., 2013).

The use of antibiotics did not significantly affect carriage of any of the target microbial species in the upper respiratory tract. This could be a result of there being no effect of antibiotics on the overall carriage rates of the targeted species. Antibiotics are known to have short-term effects on some microbial communities such as those of the human gut (Panda et al., 2014). Azithromycin use has been associated with decreased carriage of *S. pneumoniae, H. influenzae* and *M. catarrhalis* in young children with bronchiectasis within two weeks of exposure to the antibiotic (Hare et al., 2013). Furthermore intravenous or

oral antibiotic use two months prior to swab sampling was found to significantly reduce carriage of *S. pneumoniae* in children less than two years of age using a multivariable model which adjusted for age (van den Bergh et al., 2012). The recent nature of antibiotic use within the study population cannot be confirmed as exposure dates were not requested from participants. Therefore antibiotic use may not have been recent enough to have a detectable effect on carriage.

Furthermore, vaccination status did not significantly affect microbial carriage in the respiratory tract. This may reflect the lack of an effect of vaccination on overall carriage of these species. Vaccination has been shown to change the distribution of *S. pneumoniae* serotypes but not the overall prevalence of carriage of this species (Gladstone et al., 2015). Furthermore due to the effects of herd immunity, similar patterns of carriage are likely to be observed within the unvaccinated population compared with the vaccinated population (Miller et al., 2011). However, the question posed within the questionnaire is flawed as it does not provide information on individual vaccines such as PCV-13 and Hib which are relevant to the species being studied.

Seasonal differences in carriage were observed in *S. pneumoniae*, *M. catarrhalis*, *P.aeruginosa* and *H. influenzae* with the first three demonstrating significantly higher carriage prevalence in the summer versus the winter and *H. influenzae* demonstrating higher carriage in the winter. Nasopharyngeal carriage of *S. pneumoniae* and *H. influenzae* has previously been shown to be more frequent in winter in young children (Harrison et al., 1999, Gray et al., 1982). In contrast, in a further study, *S. pneumoniae* carriage was found to be higher in the summer compared with the winter in older children aged 7-15 years (Gunnarsson et al., 1998). Furthermore, other studies have demonstrated no effect of season on nasopharyngeal carriage of *S. pneumoniae* in children less than two years of age (Syrjanen et al., 2001). In addition, *M. catarrhalis* prevalence has been show n to be highest in the autumn and winter seasons in a study of children up to two years of age (Verhaegh et al., 2011). However, another study showed no difference in *M. catarrhalis* carriage between seasons (Ejlertsen et al., 1994). These conflicting findings concerning the seasonality of bacterial respiratory species carriage indicate that other factors may be contributing to these observed patterns, such as the prevalence of viral species.

Viral prevalence was not significantly different between seasons, however influenza A and parainfluenza viruses were only present in samples collected in the winter. Influenza A is known to peak in the winter seasons and is thought to be a result of a drop in absolute

humidity which encourages transmission and survival of the virus (Shaman and Kohn, 2009).

Geographical differences in carriage were also observed for *S. pneumoniae, M. catarrhalis*, RSV and coronavirus. Geographical differences in rates of IPD have been observed between countries (Hausdorff et al., 2001). However, it is thought that methodologies may explain any differences in overall carriage with individuals from certain geographical areas performing superior swabbing technique than others. Unfortunately this is one of the pitfalls of the self-swabbing method which is more difficult to control than healthcare professional led swabbing where training of staff is a prerequisite to the role.

In addition to the demographic variables studied, there are further variables which are important in understanding the epidemiology of bacterial carriage. These include smoking status where both smoking as well as passive smoking have been shown to increase carriage of *S. pneumoniae* (Greenberg et al., 2006, Bakhshaee et al., 2012). Furthermore, family structure can also affect carriage of *S. pneumoniae* with higher carriage in individuals with young siblings (Regev-Yochay et al., 2004b).

Co-carriage was common within the upper respiratory tract samples collected in this study. Young age and recent RTI significantly increased the prevalence of co-carriage. Multiple bacteria as well as multiple viruses within swabs samples collected from children have previously shown to be common, especially in children of one year of age (van den Bergh et al., 2012). A number of positive relationships were observed between *S. pneumoniae* and *H. influenzae, S. pneumoniae* and *M. catarrhalis, H. influenzae* and *M. catarrhalis, S. pneumoniae* and adenovirus*, S. pneumoniae* and rhinovirus/enterovirus*, M. catarrhalis* and rhinovirus/enterovirus*, H. influenzae* and adenovirus, *M. catarrhalis* and adenovirus, *S. pneumoniae* and coronavirus as well as *S. aureus* and coronavirus. These positive associations were reflected in a study using nasopharyngeal samples collected from children under the age of 2 years, in addition to positive associations between influenza viruses and *S. aureus* (van den Bergh et al., 2012, Bosch et al., 2013).

Adenovirus has previously been associated with *S. pneumoniae* and *M. catarrhalis* colonisation in children with recurrent otitis media (Wiertsema et al., 2011a) and *H. influenzae* carriage in healthy children (Moore et al., 2010). Furthermore, rhinovirus has been positively associated with *M. catarrhalis* carriage in healthy children (Moore et al., 2010) and *S. pneumoniae* carriage in children with recurrent otitis media (Pitkaranta et al., 2006). Positive associations between bacterial colonisers *S. pneumoniae, H. influenzae* and *M. catarrhalis* have also been observed although order of species colonisation and presence of other species are thought to affect the nature of these associations (Margolis

et al., 2010, Pettigrew et al., 2008, Verhaegh et al., 2011). The positive associations between these species, especially those observed between viruses and bacterial colonisers, are thought to enable the enhanced survival and transmission of colonising species between individuals via the symptoms of viral infection e.g. rhinitis (Bogaert et al., 2004a).

On the other hand, *S. aureus* was negatively associated with *S. pneumoniae* and *M. catarrhalis* carriage, demonstrating the competitive nature of these relationships. This negative relationship between *S. pneumoniae* and *S. aureus* has previously been reported in young children and is thought to be due to microbial interference (Regev-Yochay et al., 2006). However, it is thought that age is the driving force behind the negative associations observed between carriage of *S. pneumoniae* and *S. aureus* and hence should be accounted for when observing these relationships (Rodrigues et al., 2013). Within this study, after accounting for age within the multivariable regression models, a number of relationships remained significant. Due to the small numbers of viral species isolated and hence the lack of power to perform multivariable modelling for these species, the significance of viral relationships with other species cannot be confirmed.

Although this swabbing study allowed the collection of a large amount of data, response bias can be problematic if the characteristics of those responding are different to the characteristics of non-responders (Etter and Perneger, 1997). Unfortunately this type of bias is inherent in clinical studies involving mail-outs. Additionally the HCP swabbing group may also be biased towards individuals who attend GP clinics more often. Self-swabbing, on the other hand, may overcome this bias as attendance at the GP clinic is not necessary for participation (Gamblin et al., 2013). The non-inclusion of Group A *Streptococci* within this study is also a potential limitation as these bacteria can also cause invasive respiratory disease (Lamagni et al., 2008). However, this bacterium was not considered a priority within this study at study design phase and the laborious bacterial culture methods within such carriage studies mean that all species cannot be detected and hence decisions regarding which species to include have to be made.

Chapter 5. Phenotypic and Molecular Diversity of Respiratory Bacteria

5.1 Introduction

Molecular and phenotypic typing of bacterial isolates is essential for determining the types and characteristics of circulating strains. This has been shown to be important for monitoring the changes in *S. pneumoniae* types associated with IPD (Singleton et al., 2007, Pichon et al., 2013) as well as carriage (Huang et al., 2009, Tocheva et al., 2011) after the introduction of PCVs. Such monitoring enables the evaluation of current vaccination strategies as well as informing future vaccine design and policy. Historically, antibiotic resistance in isolates of *S. pneumoniae* and *H. influenzae* causing respiratory disease was common in the USA (Hofmann et al., 1995, Doern et al., 1997). However, vaccines have reduced rates of disease caused by highly antibiotic-resistant strains of *S. pneumoniae* in the USA (Kyaw et al., 2006). Furthermore, monitoring antibiotic resistance patterns is vital for the prevention of widespread antibiotic resistance as well as informing antibiotic prescription policies.

It is hypothesised that levels of antibiotic resistance will be low within these isolates as there has been no observed alteration in antibiotic resistance as a result of PCV introduction in the UK (Miller et al., 2011). It is also hypothesised that bacterial types will be largely those not covered by current vaccines, including non-typeable *H. influenzae* and non-PCV-13 vaccine type *S. pneumoniae*. It is expected that these carriage isolates will be diverse in nature due to the selection pressures of vaccination which transiently affect the distribution of serotypes with the respiratory tract (Hanage et al., 2010). Therefore, antibiotic resistance levels were determined using selective plates and *in silico* whole genome analysis, serotyping was undertaken using a combination of traditional PCR and *in silico* analysis of capsular genes and finally multilocus sequence typing and vaccine candidate and antigen analysis was undertaken using *in silico* analysis. This would enable the characterisation of *S. pneumoniae, H. influenzae* and *S. aureus* isolates found within the respiratory tract of members of the general population.

5.2 Isolates Analysed

A total of 237 *S. pneumoniae* isolates from nose and NP swabs and 87 *H. influenzae* isolates from nose and NP swabs were collected during the two time-points of the swabbing study. The aim was to analyse all *S. pneumoniae* and *H. influenzae* isolates in terms of serotype, multilocus sequence type and antibiotic resistance profile. A selection of these isolates was then chosen for further in-depth analysis of vaccine candidate genes and antigens. This included 34 *S. pneumoniae* and 32 *H. influenzae*. These isolates were those which co-colonised with other targeted species, as a further aim was to investigate co-colonisation. All *S. aureus* isolates were tested for resistance to methicillin.

5.3 Sequencing Quality Control

The first step in whole genome sequencing quality control is to determine if each of the sequencing runs was of optimal quality. This was determined using data collected on the MiSeq during the sequencing run and optimal values specified by Illumina. For a 2x150bp run, 4.1-5.1GB of output, 24-30 millions clusters passing filter and >80% of bases higher than Q30 was indicative of good quality sequencing. For a 2x250bp run, 7.5-8.5GB of output, 24-30 millions clusters passing filter and >75% of bases higher than Q30 was indicative of good quality sequencing according to Illumina quality specifications (Illumina, 2015).

The second step in whole genome sequencing quality control is to determine the quality of the sequences themselves as well as the assemblies. This involves determining sequence and assembly metrics for all of the genomes sequenced. The quality control metrics for all of the genomes sequenced are listed in Table 22.

Quality metrics were improved post-upgrade (from the 2 x 150bp pair-end read protocol to the 2 x 250bp pair-end read protocol) with greater coverage, fewer and larger contigs as well as higher N50 values. Both species showed similar sequence and assembly metrics with slightly better quality in *H. influenzae* when compared to *S. pneumoniae*. This may be a result of the thick capsule of *S. pneumoniae* which makes DNA extraction more difficult.

Table 22. Quality Control Metrics for Whole Genome Sequences of *S. pneumoniae* and *H. influenzae*

	Quality Control Metric Mean (SD)		
	S. pneumoniae pre-upgrade* (n = 101)	*S. pneumoniae* post-upgrade (n = 131)	*H. influenzae* (n = 80)
Number of reads	304,177 (253,725)	538,941 (274,326)	506,376 (165,269)
Fold-coverage	29.5 (27.8)	66.4 (33.8)	69.2 (23.0)
Number of scaffolds	455 (415)	154 (130)	177 (381)
Number of contigs	462 (419)	155 (129)	181 (398)
Mean contig Size (bp)	7,364 (6,454)	18,683 (8,598)	31,078 (16,953)
Total size of contigs (bp)	1,933,339 (291,341)	2,133,271 (114,139)	1,996,983 (448,585)
Longest contig (bp)	53,861 (34,403)	145,243 (75,069)	266,941 (180,860)
Shortest contig (bp)	209 (125)	306 (58)	293 (80)
N50 contig length (bp)	16,842 (14,238)	49,272 (27,151)	104,049 (74,332)

*Pre-upgrade runs were those performed using a 2 x 150bp protocol or those performed before optimisation of the 2 x 250bp protocol. SD = standard deviation, contigs = continguous sequences, bp = base pairs. N50 is a weighted median measure of the average length of the sequences.

5.4 S. pneumoniae

5.4.1 Distribution of Serotypes

S. pneumoniae isolates differ in their capsular type (serotype). Determining serotype distribution and prevalence is important for determining the genetic diversity of isolates and to determine the level of serotypes covered or not covered by PCVs. All nose and NP *S. pneumoniae* isolates (n = 232) were serotyped using *in silico* whole genome analysis and all nose isolates from the first (summer) swabbing time-point (n = 135) were serotyped by PCR amplification of capsular genes.

Comparison of PCR and *in silico* results showed the two methods to be highly concordant. A total of 109 isolates gave the same result using both methods. Four isolates were classed as unknown using PCR but were successfully serotyped using *in silico* analysis. Twenty-one isolates were classed as unknown using *in silico* analysis but were successfully serotyped using PCR. A single isolate gave differing results using the two methods, 15A/F by PCR and 17F by *in silico* analysis. Where a serotype was identified by both methods, results were highly consistent and thus nose isolates from the second swabbing time-point and NP isolates from the first swabbing time-point were serotyped by *in silico* analysis alone.

Overall, serotyping showed a diverse group of 27 *S. pneumoniae* serotypes. The most common serotypes were 11A/D (n = 28, 12.0%), 15B/C (n = 20, 8.6%), 6C (n = 20, 8.6%), 23B (n = 17, 7.3%), 35F/47F (n = 15, 6.4%) and 24A/B/F (n = 14, 6.0%). The two samples containing two phenotypically distinct *S. pneumoniae* isolates both showed two different serotypes for the two phenotypes. The first sample contained serotypes 3 and 35B and the second sample contained 11A/D and 15B/C. Twelve isolates were of unknown serotype; which included five unencapsulated *S. pneumoniae* and seven serotypes not covered by the PCR assays. The frequency of serotypes is summarised in Figure 16.

Monitoring vaccine types (VT) and non-vaccine types (NVT) is essential for observing the effects of vaccination on the population. Within these isolates, 10.0% (n = 23) were vaccine types. This included serotypes 3 (n = 6), 7A/F (n = 1), 6A/B (n = 1), 18 (n = 1), 19A (n = 13) and 19F (n = 1). Therefore 90.0% (n = 210) were non-vaccine types (NVT). The small proportion of VT within this set of isolates is consistent with the reductions in PCV serotypes observed since the introduction of these vaccines (Kaye et al., 2009). The serotypes showed a Simpson's Index of diversity (1-D) of 0.95.

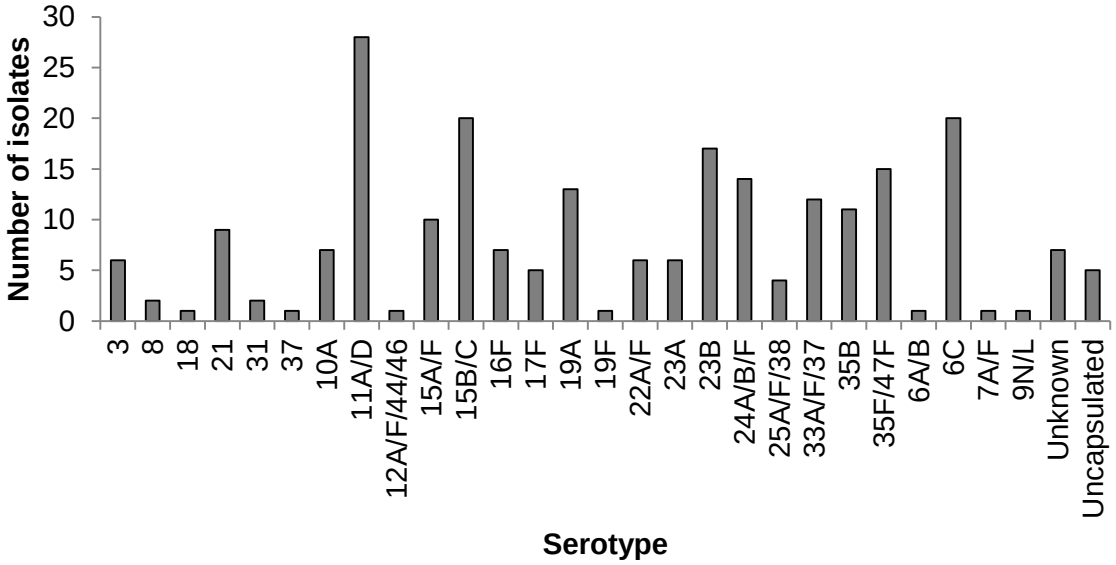


Figure 16. Distribution of Pneumococcal Serotypes detected by PCR and *in silico* PCR

For the single isolate where results were discordant by these methods, the *in silico* PCR result has been plotted.

Of the 232 *S. pneumoniae* isolates, 179 (77.2%) of these were from individuals aged 0-4 years, 52 (22.4%) were from individuals aged ≥5 years and a single isolate was from an individual of unknown age. Within those aged 0-4 years, 8.4% (n=15) of isolates were vaccine types and 88.8% (n=159) of isolates were non-vaccine types. Within those aged ≥5 years, 15.4% (n=8) were vaccine types and 71.2% (n=37) were non-vaccine types.

5.4.2 Sequence Type Diversity

Sixty-three distinct sequence types (ST) of *S. pneumoniae* were identified within the entire set of isolates (n=232). There were also 14 isolates of unknown ST, of which six were new combinations of alleles. The remaining eight isolates could not be sequence typed due to poor quality sequences (many mismatches and uncertainties within the sequences) and thus were excluded from further analysis. The most common STs were 62 (11.2%, n = 26), 199 (9.5%, n = 22), 162 (5.2%, n = 12), 439 (6.0%, n = 14) and 1635 (5.6%, n = 13). The set of *S. pneumoniae* STs gave a Simpson's Index of Diversity (1-D) of 0.96.

A number of sequence types were observed to be frequently associated with specific serotypes; these are shown in Table 23.

Table 23. Frequent Associations of Pneumococcal Sequence Types with Serotypes

ST	Frequently associated serotypes (n/N, % within serotype)
ST199	19A (9/13, 69.2%); 15B/C (12/20, 60.0%)
ST414	16F (3/7, 42.9%)
ST433	22A/F (4/6, 66.7%)
ST1635	35F/47F (12/15, 80.0%)
ST162	24A/B/F (10/14, 71.4%)
ST100	33A/F/37 (6/12, 50.0%)
ST393	25A/F/37 (3/4, 75.0%)
ST1262	15B/C (8/20, 40.0%)
ST438	23A (5/5, 100%)
ST439	23B (14/17, 82.4%)
ST1877	21 (5/8, 62.5%)
ST2068	10A (3/7, 42.9%)

ST = Sequence type

Analysis using eBURST revealed *S. pneumoniae* isolates as 9 clonal groups (groups of STs based on similarity to a single central allelic profile and share at least 5 of the same alleles, with single locus variants [SLV] differing by one allele and double locus variants [DLV] differing by two alleles) and 48 singleton STs (STs that do not share at least 5 loci with any other ST). This demonstrates that the isolates are not highly related; they are in fact a highly genetically diverse group of isolates. Figure 17 demonstrates the genetic relatedness of these isolates.

The set of *S. pneumoniae* isolates was compared to the *S. pneumoniae* MLST dataset, which showed this set of isolates to spread across 35 clonal groups within the entire MLST dataset, demonstrating that a number of singletons from the set of isolates were actually part of clonal groups within the entire MLST dataset rather than being true singletons. Two of the isolates were singletons within the *S. pneumoniae* database and these were unknown sequence types. Twenty-six of the sequence types were also found to be founders of clonal groups (representing the single central allelic profile of a clonal group and the ST which differs most from other STs within the clonal group) within the *S.*

pneumoniae database. A number of the *S. pneumoniae* isolates are clustered into clonal complexes ST 460, ST 1379 and ST 30.The distribution of these isolates across the *S. pneumoniae* MLST database is shown in Figure 18. This eBURST figure is limited by the screen size, therefore in order to capture the entire dataset the resolution to visualise individual STs cannot be improved (seen as black lines within the figure).

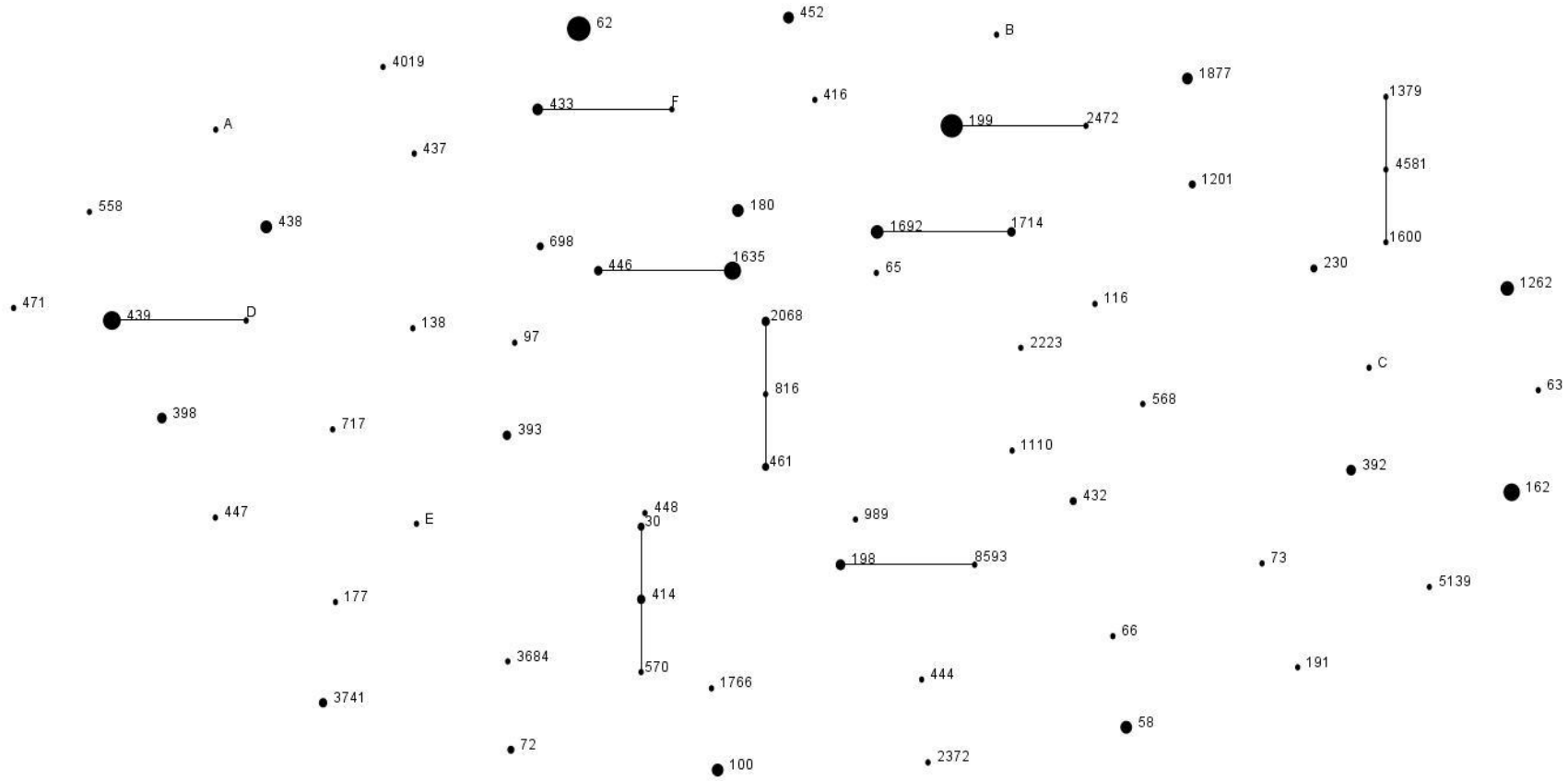


Figure 17. eBURST Diagram of the *S. pneumoniae* Sequence Types (STs) isolated from study swab samples

Black circles represent STs and the size of each circle is relative to abundance of each ST within the dataset. Lines indicate that isolates share 6 out of the 7 MLST genes (SLV) and belong to the same clonal complex. A-F = unknown STs. Positioning of STs does not reflect genetic relatedness if STs are not part of the same clonal complex.

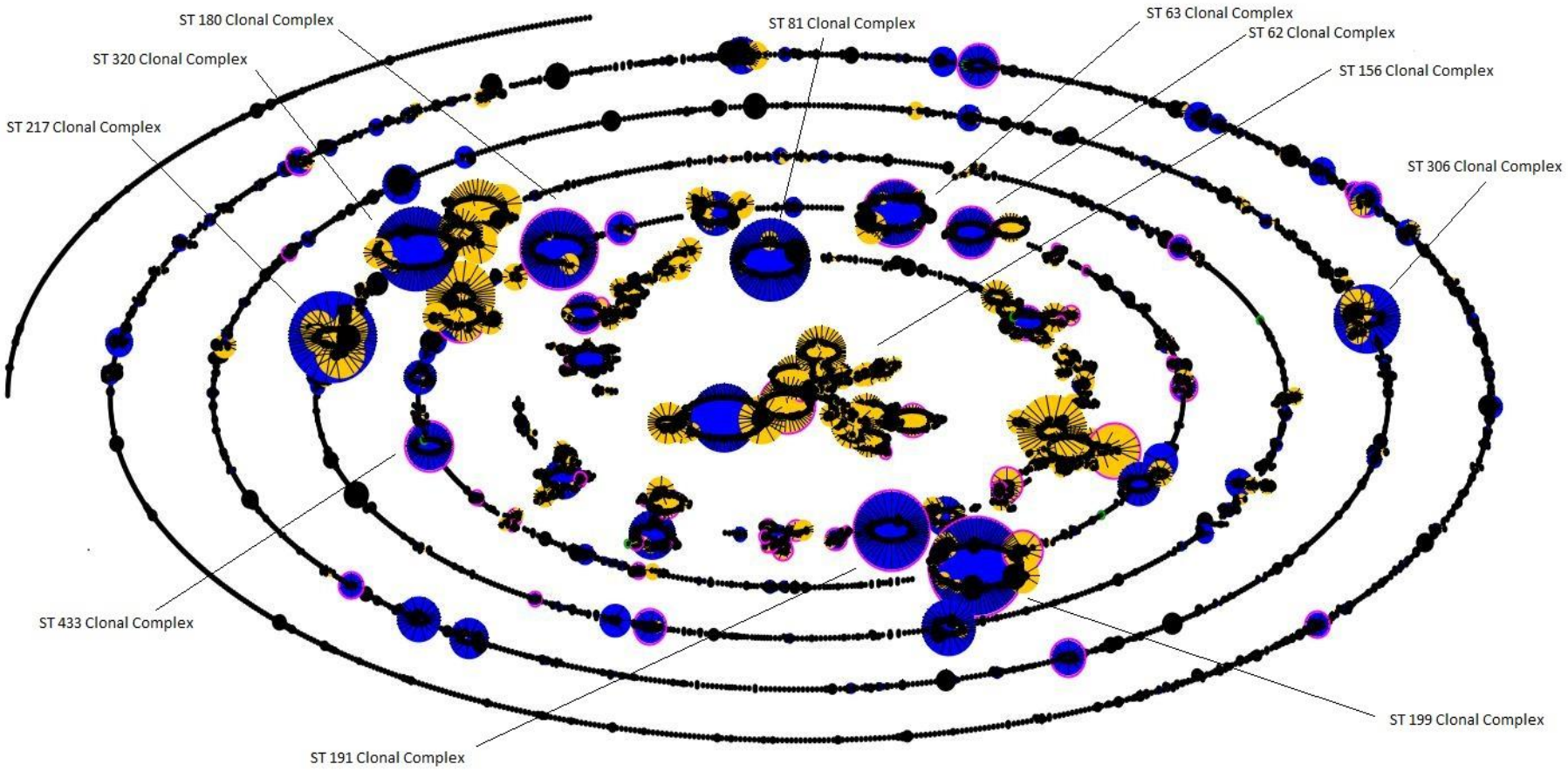


Figure 18. eBURST Diagram Comparing *S. pneumoniae* Isolates to the *S. pneumoniae* MLST Database

MLST Dataset: Blue circles = clonal group founders, yellow circles = subgroup founder, black circles = all other STs. Study Dataset: Pink outline = STs present in the MLST database, green outline = STs not present in the MLST database. Size of each circle is relative to abundance of each ST within the datasets.

5.4.3 Antibiotic Resistance

S. pneumoniae isolates were tested for resistance to a range of antibiotics using *in silico* detection of resistance genes from whole genome sequences. Resistance genes were identified in 5.2% (n = 12) of isolates, with the number of genes present within each individual isolate ranging from one to six. Table 24 below summarises the resistance genes present within this set of *S. pneumoniae* isolates. For the full list of 1,913 alleles tested please refer to the original articles (Zankari et al., 2012, McArthur et al., 2013). Since only acquired genes are targeted in SRST2, penicillin-binding protein genes (*pbp*) were targeted separately.

Table 24. Antibiotic Resistance and Susceptibility Genes present within *S. pneumoniae* isolates (N=232)

Gene	Antibiotic	Allele(s)	Number of Isolates containing gene (n)	Prevalence of specific gene % (95% CI)
cat	Chloramphenicol	pC194	1	0.4 (-0.4, 1.2)
ermB	Erythromycin	7, 16	6	2.6 (0.6, 4.7)
mefA	Macrolide	3,10	4	1.7 (0.0, 3.4)
msrD	Macrolide	2, 3	4	1.7 (0.0, 3.4)
tet32	Tetracycline	2	1	0.4 (-0.4, 1.2)
tetM	Tetracycline	2, 4, 8, 10, 12	10	4.3 (1.7, 6.9)
tetS	Tetracycline	1, 3	2	0.9 (-0.3, 2.1)
pbp1a	Beta-lactam	NA	214	92.2 (88.8, 95.7)
pbp2x	Beta-lactam	NA	223	96.1 (93.6, 98.6)
pbp2b	Beta-lactam	NA	204	87.9 (83.7, 92.1)

Cat = chloramphenicol acetyltransferase, erm = erythromycin resistance methylase, mef = macrolide efflux, tet = tetracycline restsance, msr = macrolide and streptogramin B resistant, pbp = penicillin-binding protein, NA = not applicable.

Acquired resistance genes were identified against three classes of antibiotics: phenicols, macrolides and tetracycline. Tetracycline resistance was the most common acquired resistance, with *tetM* being present in ten out of the twelve isolates that contained resistance genes. Multiple alleles of *tetM* were commonly found within the same isolate, with seven isolates containing three alleles. No beta-lactam acquired antibiotic resistance genes, such as beta-lactamase genes, were identified within this set of isolates.

Identification of *pbp* genes indicates that the isolates are susceptible strains with non-altered *pbp* genes whereas non-detection is indicative of modified genes which result in reduced antibiotic susceptibility (Nagai et al., 2001). For 13.8% (32/232) of isolates tested, at least one *pbp* gene was not detected: eight isolates lacked all three genes, six isolates lacked *pbp1a* and *pbp2b*, one isolate lacked *pbp2x* and *pbp2b*, 13 isolates lacked *pbp2b* only and 4 isolates lacked *pbp1a* only. These 32 isolates are therefore thought to have varying levels of penicillin non-susceptibility.

5.4.4 Analysis of Vaccine Candidates and Antigens

Out of all of the genes tested, 95 were identified within the entire set of *S. pneumoniae* genomes. Individual isolates possessed between 26 and 55 of the 95 genes. Genes being considered as vaccine candidates found within this set of isolates are shown in Table 25. All other detected antigens are shown in full in Appendix 8. For the full list of the 1,077 alleles tested please refer to the VFDB website (http://www.mgc.ac.cn/VFs/main.htm) and the SRST2 website (https://github.com/katholt/srst2).

Table 25. Vaccine candidates found within the *S. pneumoniae* isolates (N=34)

Gene	Role	n (%)
lytA	Autolysin	34 (100)
piaA	ABC transporter	34 (100)
ply	Pneumolysin	34 (100)
psaA	Pneumococcal surface antigen A	34 (100)
pspA	Pneumococcal surface protein A	9 (26.5)
pspC/cbpA	Choline binding protein	28 (82.4)
nanA	Neuraminidase A	32 (94.1)

The presence of vaccine candidates *lytA, piaA, ply* and *psaA* in all isolates tested demonstrates the conserved nature of these antigens. Further vaccine candidates *pspA, pspC/cbpA* and *nanA* were not present in all isolates tested, with *pspA* only being present in nine isolates. Further common antigens present in all isolates were also identified. These may be considered as vaccine candidates in the future.

5.5 H. influenzae

5.5.1 Distribution of Serotypes

H. influenzae isolates also varied in their capsular serotype, and like *S. pneumoniae*, it is important to understand the diversity of isolates in terms of their capsular type. All *H. influenzae* isolates (n = 80) were serotyped using *in silico* analysis and the nose isolates from the first (summer) swabbing time-point were also serotyped using PCR detection of capsular genes. The two methods were 100% concordant, classifying all nose isolates from the first swabbing time-point as non-typeable (NTHi). Overall, a single isolate was *bexB* positive, therefore a typeable but non a-f serotype. All other isolates were NTHi which means that they are unencapsulated. Serotype distribution could not be assessed in relation to participant characteristics, swabbing methodology and geographical location due to all but one isolate being NTHi.

5.5.2 Sequence Type Diversity

Forty-six distinct sequence types of *H. influenzae* were identified within the entire set of isolates. The *H. influenzae* isolates were highly diverse with only one or two isolates commonly being identified within most STs. The more common STs were 474 (7.5%, n = 6), 57 (3.8%, n = 3), 348 (3.8%, n = 3), 569 (3.8%, n = 3) and 1215 (3.8%, n = 3). Fourteen isolates were of unknown sequence type, which were due to new combinations of alleles. The set of *H. influenzae* isolates gave a Simpson's Index of Diversity (1-D) of 0.99.

eBURST analysis showed the isolates to be a collection of 5 clonal groups (groups of STs based on similarity to a single central allelic profile and share at least 5 of the same alleles) and 49 singletons (STs that do not share at least 5 loci with any other ST). This demonstrates that the *H. influenzae* isolates are not highly related; they are in fact a highly genetically diverse group of isolates. Figure 19 demonstrates the genetic relatedness of these isolates.

The set of *H. influenzae* isolates was compared to the *H. influenzae* MLST dataset, which showed this set of isolates to be spread across 37 clonal groups within the entire database, demonstrating that a number of singletons from the set of isolates were actually

part of clonal groups within the entire MLST dataset rather than being true singletons.. The other 23 STs were singletons within the MLST database. Seventeen of the sequence types were also found to be founders of clonal groups (representing the single central allelic profile of a clonal group and the ST which differs most from other STs within the clonal group) within the *H. influenzae* MLST dataset. The distribution of the study isolates across the *H. influenzae* MLST database is shown below in Figure 20.

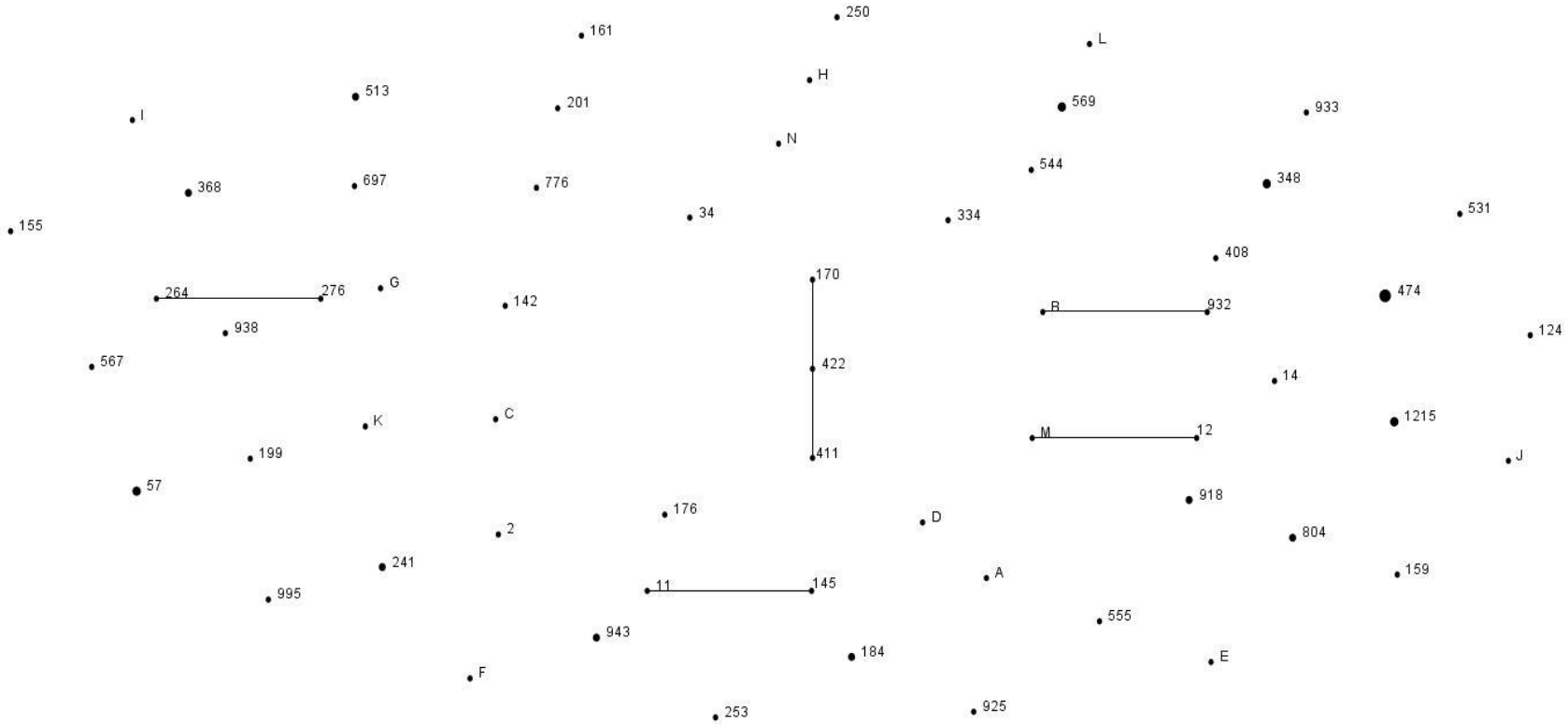


Figure 19. eBURST Diagram of *H. influenzae* Sequence Types (STs)

Black circles represent STs and the size of each circle is relative to abundance of each ST within the dataset. Lines indicate that isolates share 6 out of the 7 MLST genes and belong to the same clonal complex. A-N = unknown STs. Positioning of STs does not reflect genetic relatedness if STs are not part of the same clonal complex.

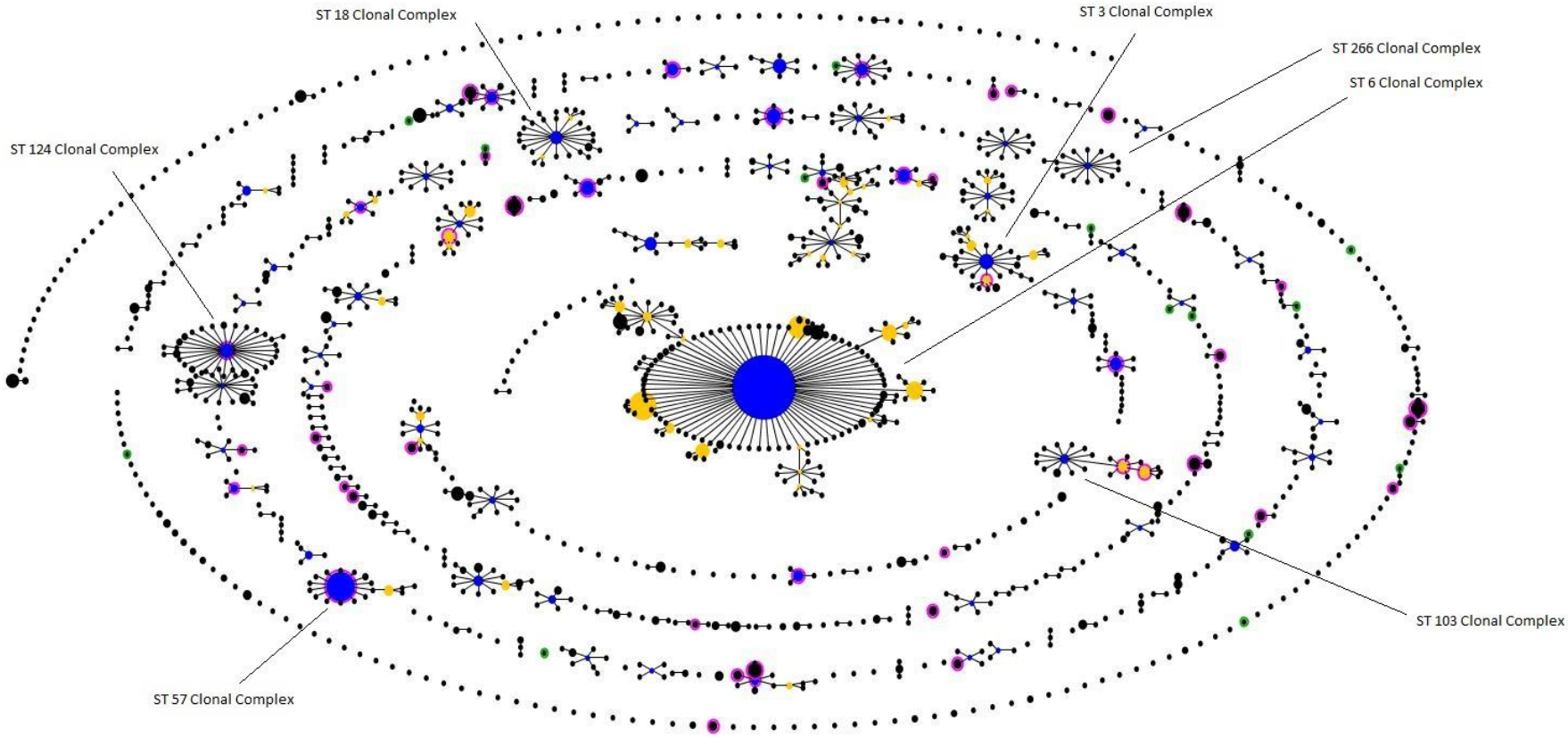


Figure 20. eBURST Diagram Comparing the *H. influenzae* Isolates to the *H. influenzae* MLST Database

MLST Dataset: Blue circles = clonal group founders, yellow circles = subgroup founder, black circles = all other STs; Study Dataset: Pink circles = STs present in the MLST database, green circles = STs not present in the MLST database. Size of each circle is relative to abundance of each ST within the datasets.

5.5.3 Antibiotic Resistance

H. influenzae isolates were tested for resistance to a range of antibiotics using *in silico* detection of resistance genes from whole genome sequences. Resistance genes were identified in 7.6% (n = 6) of isolates, with one or two resistance genes being present within each isolate. Table 26 below summarises the resistance genes present within this set of *H. influenzae* isolates. For the full list of 1,913 alleles tested please refer to the original articles (Zankari et al., 2012, McArthur et al., 2013). Only acquired genes are targeted in SRST2.

Table 26. Antibiotic Resistance Genes present within *H. influenzae* Isolates (N=80)

Gene	Antibiotic	Alleles	Isolates (n)	Prevalence of specific gene (%, 95% CI)
blaTEM	Beta-lactam	1	6	7.6 (1.8, 13.4)
aph(3')-I	Aminoglycoside	A	3	3.8 (-0.4, 8.0)

BlaTEM = extended spectrum beta-lactamase, aph = aminoglycoside phosphotransferase.

Acquired resistance genes were only identified against two classes of antibiotics: beta-lactams and aminoglycosides. Beta-lactam resistance was the most common acquired resistance, with *blaTEM* being present in all six of the isolates that contained resistance genes.

5.5.4 Analysis of Vaccine Candidates and Antigens

Out of all of the genes tested, 130 were identified within the entire set of *H. influenzae* genomes. Individual isolates possessed between 87 and 105 of the 130 genes. Genes being considered as vaccine candidates found within this set of isolates are shown in Table 27. All other detected antigens are shown in full in Appendix 9. For the full list of the 556 alleles tested please refer to the VFDB website (http://www.mgc.ac.cn/VFs/main.htm) and the SRST2 website (https://github.com/katholt/srst2).

Table 27. Vaccine candidates found within the *H. influenzae* isolates (N=32)

Gene	Role	n (%)
hpd	Surface lipoprotein	27 (84.4)
ompP2	Outer membrane protein	32 (100)
ompP5	Outer membrane protein	29 (90.6)
tbpA/B	Periplasmic binding	32 (100)
hmw1	Adhesin	5 (15.6)
hmw2	Adhesin	5 (15.6)
hia	Adhesin	11 (34.4)

Hpd = Haemophilus protein D, omp = outer membrane protein, tbp = transferrin-binding protein, hmw = high molecular weight, hia = *H. influenzae* adhesin.

The presence of vaccine candidate genes *ompP2* and *tbpA/B* in all isolates tested demonstrates the conserved nature of these antigens. Further vaccine candidates *hpd, ompP5, hmw1, hmw2* and *hia* were not present in all isolates tested, with *hmw1/2* only being present in five isolates. Further common antigens present in all isolates were also identified. These may be considered as vaccine candidates in the future.

5.6 Methicillin Resistant *S. aureus*

All *S. aureus* isolates were tested for methicillin resistance using MRSA select plates. Results are shown below in Table 28.

Table 28. Number (n) of Methicillin-resistant *S. aureus* isolates identified within each swab type and swabbing time-point

Time-point	Swab Type	Total (*N*)	*n*	% (95% CI)
Summer 2012	Nose	264	10	3.8 (1.5, 6.1)
	SS WMS	33	0	0.0 (NA)
	NP	35	3	8.6 (-0.7, 17.9)
	HCP WMS	2	0	0.0 (NA)
Winter 2013	Nose	145	2	1.4 (-0.5, 3.3)
	SS WMS	20	0	0.0 (NA)

SS = self-swabbing, WMS = whole mouth swab, HCP = healthcare professional swabbing, NA = not applicable, 95% CI = 95% confidence interval of proportion.

Overall, levels of MRSA were low within the set of samples tested and did not vary significantly between the different swab types.

5.7 Discussion

Molecular and phenotypic typing methods were used to assess genotypes and phenotypes of bacteria found within the upper respiratory tract of members of the general population. Antibiotic resistance is an important global issue and tracking the resistance of bacterial isolates is important for controlling the issue and reducing costs associated with resistance (Reynolds et al., 2014, Carlet et al., 2014). Antibiotic resistance genes were detected within these carriage isolates of *S. pneumoniae, H. influenzae* and *S. aureus*. Macrolide resistance genes identified within these *S. pneumoniae* isolates were *ermB* (2.6%), *mefA* (1.7%) and *msrD* (1.7%). The *ermB* gene encodes a ribosomal methylase where as the *mefA/msrD* genes encode antibiotic efflux pumps (Hoban et al., 2001, Chancey et al., 2015). Approximately 17% of *S. pneumoniae* bacteraemia isolates from the United Kingdom and Ireland have been shown to express erythromycin resistance (Reynolds et al., 2004). The expression of the *ermB* gene is highly associated with erythromycin resistance in a number of European countries, with 98.5% of 202 erythromycin-resistant isolates from France possessing the gene (Reinert et al., 2005). However, in the USA, *mefA* is a more common mechanism of erythromycin resistance, with 70.9% of 3,044 erythromycin-resistant isolates possessing the gene (Brown et al., 2004).

Tetracycline resistance genes identified within these *S. pneumoniae* isolates were *tetM* (4.3%), *tetS* (0.9%) and *tet32* (0.4%). These genes are involved in protecting the bacterial ribosome from the effects of the antibiotic (Roberts, 2005). Approximately 4.3% of *S. pneumoniae* bacteraemia isolates from the United Kingdom and Ireland have been shown to express tetracycline resistance (Reynolds et al., 2004). Furthermore, the chloramphenicol resistance gene *cat,* which encodes an acetyltransferase that modifies the structure of the antibiotic, has previously been identified in 0.4% of *S. pneumoniae* isolates (Dang-Van et al., 1978). Approximately 5.4% of *S. pneumoniae* isolates in the UK between 1992 and 1995 were chloramphenicol resistant (Goldsmith et al., 1997).

Penicillin-binding protein (PBP) genes were also identified within a large proportion of the study isolates. These conserved proteins bind beta-lactam antibiotics and, in their non-mutated forms, confer susceptibility to these antibiotics. The lack of detection of *pbp* genes in their non-mutated form may be indicative of the presence of mutated forms and hence beta-lactam resistance (Nagai et al., 2001). Prevalence of penicillin non-susceptible *S. pneumoniae* isolates in the UK between 2007-2008 was approximately 4.7% (Reynolds, 2009).

Beta lactamase genes identified in *H. influenzae* isolates were *blaTEM* (7.6%). TEM beta-lactamases are extended spectrum beta-lactamases which confer resistance to penicillin and first-generation cephalosporin antibiotics (Medeiros, 1984). Beta-lactamase production, particularly TEM-1, is common in *H. influenzae* with 15.7% prevalence of TEM-1 in the UK between 1999 and 2003 (Farrell et al., 2005). Furthermore, aminoglycoside resistance genes identified in *H. influenzae* isolates were *aph(3')-I* (3.8%). This gene encodes an aminoglycoside phosphotransferase which modifies the structure of aminoglycoside antibiotics gentamicin B, neomycin and kanamycin amongst others (Shaw et al., 1993). Aminoglycoside resistance via *aph(3')-I* has been reported in other European countries such as Spain (Gomez-Lus and Vergara, 1995) but, to our knowledge, is not common in the UK.

The identification of few antibiotic resistance genes within these isolates is thought to reflect the low and stable levels of antibiotic resistant respiratory species within the UK (Reynolds, 2009). Antibiotic prescribing within the community in the UK is similar to a number of European countries, with 21.5 defined daily doses (DDD) per 1,000 inhabitants per day compared with the EU average of 22.4 DDD per 1,000 inhabitants per day (ECDC, 2014). However, antibiotic prescribing is increasing with a 6% rise in the overall number of antibiotic prescriptions between 2010 and 2013 (Public Health England, 2014a).

A limitation to examining antibiotic gene presence is that phenotypic expression of resistance does not always correlate with gene presence. Antibiotic resistance genes can be attenuated within a microbial species in order to reduce the biological cost associated with gene expression (Depardieu et al., 2007). A study of *Enterococcus faecium* in livestock demonstrated that *ermB* can be present in erythromycin susceptible isolates of the species (Enne et al., 2006). Furthermore, the *tetM* gene is not always expressed in *S. pneumoniae* and *Streptococcus pyogenes* (Giovanetti et al., 2007). This indicates that phenotypic assessment of antimicrobial susceptibility is necessary to determine true levels

of resistance. A further limitation to the antibiotic susceptibility testing undertaken using SRST2 is that only acquired resistance genes are detected. Therefore other non-acquired resistance genes could not be assessed. This might include mutated forms of the *H. influenzae pbp3* gene, which has been shown to be a common resistance mechanism in this organism in a number of countries (Ubukata et al., 2001, Skaare et al., 2014).

Vaccine candidate genes present in all *S. pneumoniae* study isolates were *lytA, piaA, ply* and *psaA*. However, vaccine candidate genes *pspA* (26.5%), *pspC/cbpA* (82.4%) and *nanA* (94.1%) were not present in all isolates tested. *LytA*, *piaA* and *psaA* have been shown to be protective against infection by *S. pneumoniae* in mice and are also highly conserved (Yuan et al., 2011, Whatmore and Dowson, 1999, Brown et al., 2001, Talkington et al., 1996). *Ply* has also been shown to induce immune responses in clinical trials in humans (Kamtchoua et al., 2013). Low detection levels of *pspA, pspC/cbpA* and *nanA* has also been reported via quantitative PCR with 94.45%, 57.36% and 97.43% detection respectively in 50 nasopharyngeal samples (Sakai et al., 2013).

Vaccine candidate genes present in all *H. influenzae* study isolates were *tbpA/B* and *ompP2.* However, vaccine candidate genes *hpd* (84.4%), *ompP5* (90.6%), *hmw1/2* (15.6%) and *hia* (34.4%) were not present in all isolates tested. Despite being highly expressed, ompP2 proteins are thought to be highly variable between strains of *H. influenzae* (Duim et al., 1996). However, conserved regions within these proteins may enable them to be considered as potential vaccine candidates (Hiltke et al., 2002). Immunisation of rats with TbpB was found to reduce bacterial presence within the lung and, similarly to ompP2, variation within this protein may be problematic for its use within a vaccine (Webb and Cripps, 1999). Vaccination of non-typeable *H. influenzae* is thought to be most likely achieved via the use of multiple antigens (Barenkamp, 1996).

The presence of these antigens in carriage isolates is not necessarily reflective of pathogenicity, as similar sets of genes can be found during both colonisation and disease (Donkor et al., 2012). The use of in-depth whole genome sequencing and single nucleotide polymorphism (SNP) analysis would enable greater insight into the mechanisms of carriage and pathogenicity, which has been investigated in *S. aureus* (Young et al., 2012).

Serotyping revealed a diverse set of *S. pneumoniae* and *H. influenzae* isolates with non-vaccine types being predominant. Increases in *S. pneumoniae* serotypes 11A/D, 15B/C, 6C, 23B, 35F/47F and 24A/B/F have previously been reported within invasive disease and

paediatric carriage studies (Tocheva et al., 2013, Miller et al., 2011). The majority of serotypes within this study were non-vaccine types, reflecting the effectiveness of vaccination in reducing carriage of the targeted serotypes. Few vaccine types were collected during the study (10.0%). This is similar to further carriage studies within the UK paediatric population, where 11% of serotypes were found to be vaccine types (Gladstone et al., 2015). This is consistent with changes incurred as a result of PCV introduction (Gladstone et al., 2011). Distribution of serotypes is thought to vary with age with serogroups 6, 9 and 23 being carried for longer periods of time in young children than other serogroups (Cobey and Lipsitch, 2012, Hogberg et al., 2007). Since a large proportion of serotypes isolated within this study were from individuals less than five years of age, the longer duration of carriage within this age group would therefore explain the high prevalence of serotypes 6C (n=20), 23B (n=17) and 19A (n=13).

S. pneumoniae ST199 was found to be a common sequence type within this collection of isolates. This ST has been shown to increase post-PCV-7 introduction but decrease post-PCV-13 introduction due to its association with serotype 19A, which is included in the latter vaccine (Hanage et al., 2011, Pichon et al., 2013). The high levels of ST199 in this study might reflect the numbers of isolates of this sequence type expressing the serotype 15B/C capsule in addition to those expressing the serotype 19A capsule.

Non-typeable *H. influenzae* were found to predominate within this set of isolates. The frequency of non-typeable *H. influenzae* in disease and carriage studies has previously been recorded in other studies (Giufre et al., 2011, Puig et al., 2014). This is consistent with the fall in serotype b levels and rise in NTHi since the introduction of the *H. influenzae* type b (Hib) vaccine (Agrawal and Murphy, 2011). Prevalence of *H. influenzae* type b carriage in the UK has decreased from 4.0% in 1992 to 0.0% in 2002 (McVernon et al., 2004). The proportion of non-typeable *H. influenzae* infections relative to all infections caused by the species has increased in the UK from 3.2% in 1990 to 69.6% in 2013 (Public Health England, 2015a).

Where a serotype was identified using both PCR and *in silico* PCR, a single isolate showed discordant results between the two methods. No cross-reactivity between serotypes 15 and 17 has been documented (Pai et al., 2006). The potential for multiple serotypes of *S. pneumoniae* within samples might, however, explain this result (Turner et al., 2011a). The methodology used here, which involved the isolation of single phenotypic colonies of bacteria from STGG samples for use in whole genome sequencing and serotyping, is a potential limitation. This method, although it is the recommended method

and enables a detailed analysis of a particular colony, does not adequately allow for the detection of multiple serotypes or sequence types within samples (Satzke et al., 2013).

Analysis of the MLST data shows a diverse set of *S. pneumoniae* and *H. influenzae* isolates with more common STs, including clonal group founders, being found as well as a number of unknown STs. A larger number of unknown *H. influenzae* STs may be due to fewer STs being deposited within these MLST databases as these species were previously undescribed and greater efforts were previously being concentrated on understanding the epidemiology of *S. pneumoniae,* a well established pathogen. More recently, greater understanding of *H. influenzae* as pathogens has led to greater efforts in understanding the epidemiology of these species (Catlin, 1990, Turk, 1984). The presence of unknown STs is potentially a result of a lack of previous reports of these STs or the evolution of these species into varied and distinct types or both. This evolution may be important for vaccine and antibiotic development and should be taken into account when determining novel targets and biomarkers.

Chapter 6. Ecology of Microbial Respiratory Tract Communities

6.1 Introduction

A number of microbial communities have been extensively studied, including those of the lung (Rogers et al., 2004), the stomach (Dethlefsen et al., 2008) and the vagina (Zhou et al., 2004), in order to understand the microbial species present, the interactions occurring between them, and the effect of the host environment. Studying microbial communities has important implications for understanding disease states (Ley et al., 2006) as well as the effects of human interventions, such as antibiotics (Dethlefsen et al., 2008). The upper respiratory tract community includes species which are common to both health and disease states. The application of ecological analysis tools to upper respiratory tract swab samples will enable the interactions between species and the effect of host characteristics to be deciphered.

Nestedness is a measurement of organisation within an ecological system (Darlington, 1957, Patterson and Atmar, 1986). Organisation refers to the level of order in species extinction and colonisation within the ecological system (Atmar and Patterson, 1993). A more organised system will have a specific order of extinction whereas a less organised system will have a random order of extinction (Ulrich et al., 2009). A random order of extinction implies that species have equal chances of becoming extinct and extinction of one does not affect the presence of another species. On the other hand, a nested system is one where species found within ecologically-impoverished areas are subsets of species within ecologically-enriched areas. This implies that extinction occurs within a specific order and extinction of one species directly affects extinction of another (Atmar and Patterson, 1993). Nestedness is hence associated with β-diversity, the variation of species across habitats (Whittaker, 1960, Ulrich and Almeida-Neto, 2012). This method has been applied to distribution of bird species (Fernandez-Juricic, 2002) and fish parasites (Rohde et al., 1998) as well as interaction networks of plants and animals (Bascompte et al., 2003) in order to decipher community assembly patterns.

Species distribution patterns, either using presence-absence data or abundance data, allow the analysis of biodiversity (Elith et al., 2006). Such patterns have been useful in aiding conservation efforts, as geographical patterns of species distribution can be monitored over time (Ferrier, 2002). The environmental and evolutionary factors influencing the distribution of species can also be determined (Ricklefs, 2004). Species

distribution patterns can be applied to microbial communities in order to understand the biodiversity of microbial species within different environments (Bohannan and Hughes, 2003). Species distribution within microbial communities of the lung has been studied with the aim of understanding diseases such as cystic fibrosis and COPD (van der Gast et al., 2011, Erb-Downward et al., 2011).

There are two main ecological theories explaining species abundances and levels of diversity within communities. Niche models involve ‘niches’, or separated habitats, which have distinct sets of nutrients and environmental influences (Elton, 1927). The characteristics of a niche determine levels of species abundances (Grinnell, 1928) as well as subdivision of resources (Schoener, 1974). Neutral models, first developed by Stephen Hubbell (Hubbell, 2001), explain species abundances and diversity in terms of stochastic influences (including speciation, dispersal, immigration, birth and death). All species are equal and differences between niches do not determine the distribution of these species, which is left up to chance (McGill, 2003).

MacArthur’s Broken Stick model was developed by Robert MacArthur to describe the abundance of bird species (Macarthur, 1957). This model involves neutral concepts as it states that an environment is analogous to a stick of a particular size (or length) which is broken sequentially at random. The lengths of the fragments of stick produced by random breakage are relative to the abundance of each species within this environment (King, 1964).

The aim was to apply the ecological methods mentioned above to human upper respiratory tract samples in order to understand community assembly, extinction and colonisation patterns. This would involve the use of nestedness in order to understand the organisation within the respiratory tract and order in species extinction. Furthermore, species distribution curves would be used to determine whether species are randomly distributed between individuals or whether facilitative or competitive interactions may be important. In addition, analysis of abundance levels and application of MacArthur’s Broken Stick model will enable the determination of community assembly patterns in order to understand whether communities have niche or neutral characteristics. This was a preliminary application of these ecological methods in a novel context and hence is not be without limitations.

It is hypothesised that communities of the upper respiratory tract will be highly nested in adults or those without recent infection. On the other hand, less nested communities were expected in those with recent infection and in young children. This is thought to be the

result of increased colonisation and transmission of different species within the latter, resulting in greater flux of species (Ulrich et al., 2009, Sa-Leado et al., 2008).

It is also hypothesised that species distribution will be more regular (similar numbers of species colonising different individuals) in adults or those without recent infection, which is indirectly indicative of competitive relationships. On the other hand, it is hypothesised that species distribution will be clumped (instances of few or many species colonising different individuals) in those with recent infection and in young children, which is indirectly indicative of facilitative relationships (Haase, 1995, He et al., 1997). This is thought to be a result of facilitation enabling the transmission of species, which is believed to occur frequently in young children and in the event of viral upper respiratory infections (Rodrigues et al., 2013).

Finally it is hypothesised that communities within the upper respiratory tract will demonstrate niche characteristics as a result of environmental factors and host immunity playing a significant role in shaping the community (Garcia-Rodriguez and Fresnadillo Martínez, 2002). Previous ecological theory application to microbial communities in the soil have revealed the importance of pH, reflecting the role of niche dynamics (Dumbrell et al., 2010). Furthermore, animal gastrointestinal microbial communities have reflected niche patterns of community assembly, with diet and species interactions playing important roles (Jeraldo et al., 2012). A niche, in ecological terms, represents an entity within which species interactions and the environment determine the distribution of species abundances (Grinnell, 1928). Within the respiratory tract, different sites demonstrate varying species abundances and are thought to represent distinct niches (Charlson et al., 2011).

6.2 Samples and Covariates used in Ecological Analyses

Nose and NP swab samples detected by real-time PCR were used throughout this chapter rather than culture-detected samples due to the greater numbers of microbial species that could be detected by this method, including viral species. Thus analysis included the following species: *S. pneumoniae, H. influenzae, M. catarrhalis, P. aeruginosa, S. aureus*, RSV, influenza virus, parainfluenza virus, rhinovirus/enterovirus, coronavirus and adenovirus. Covariates used throughout this chapter were those that were found to be significant in individual species multivariable analyses in Chapter 4, in

order to reduce the effect of potential confounders. Thus analysis included the following covariates: age, recent RTI and season.

6.3 Nestedness of Respiratory Communities

Nestedness analysis transforms a matrix of presence-absence data into a packed matrix within which species and samples are shuffled according to incidence and richness respectively (Ulrich and Almeida-Neto, 2012). Samples, presented on the y-axis, are shuffled so that those with the highest levels of richness (greater numbers of different species within individuals) move towards the top of the matrix and species, presented on the x-axis, are shuffled so that those showing the highest incidence (greatest numbers of a given species across individuals) move towards the left of the matrix. This process aims to maximise the number of presences to the top left of the matrix and absences in the bottom right. This concept is illustrated using simplified examples in Figure 21 for both highly nested and highly random samples.

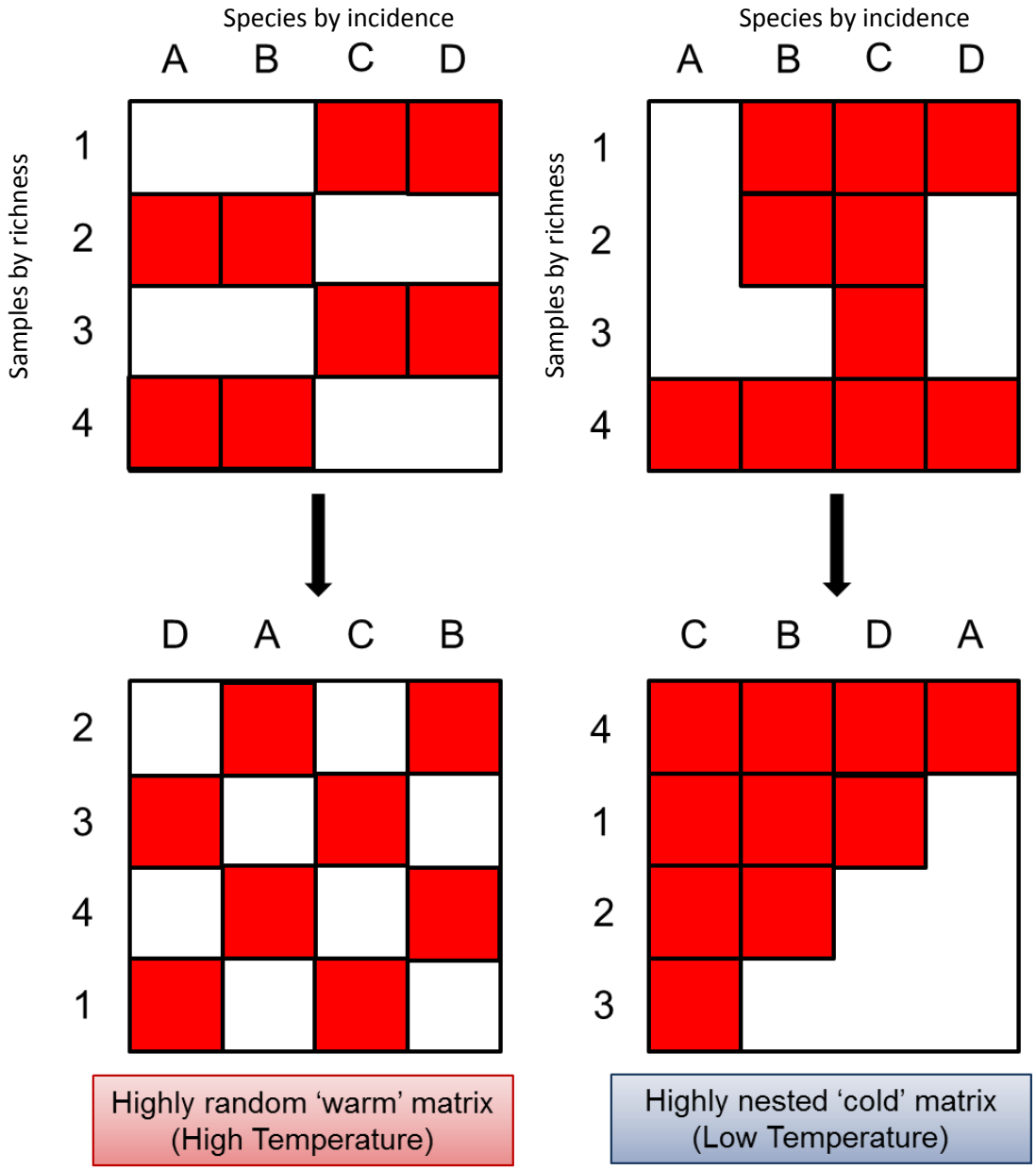


Figure 21. The Process of Shuffling during Nestedness Analysis in Highly Random and Highly Nested Communities

Top matrices represent the original matrices and bottom matrices represent the shuffled matrices. A-D = species (ordered by incidence), 1-4 = samples (ordered by richness).

The matrix shows species presences as red squares and absences as white squares. A line of perfect fill is constructed based on all absences organised to the right and all presences to the left (Ulrich et al., 2009). Any absences found to the left or presences found to the right are known as surprise absences or surprise presences, respectively. A temperature (T) is also calculated, which is determined by the sum of the distances of surprise absences or surprise presences from the line of perfect fill (Patterson and Atmar, 1986, Rodriguez-Girones and Santamaria, 2006). This matrix temperature ranges from 0 to 100, indicating if the matrix is cold (highly nested) or hot (highly random) respectively. A matrix fill is also given, which indicates what proportion of the matrix is represented by presences. This tool was applied to upper respiratory tract microbial colonisation data collected from nose and NP swabs that were analysed by real-time PCR. The aim was to understand the factors affecting organisation within the respiratory tract and to determine whether the level of organisation was the same in different swab types as well as in different individuals.

6.3.1 Patterns of Nestedness within Communities

Levels of nestedness were calculated for nose and NP swabs detected using real-time PCR. Samples were assessed according to age, recent RTI and season. This would enable the determination of the effects of such factors on the organisation of the respiratory tract community. Nestedness results are summarised below in Table 29.

Table 29. Nestedness Results for Nose and NP Samples Detected by Real-time PCR

Covariate		Nose		NP	
		N	T (matrix fill)	N	T (matrix fill)
Season	**Summer**	115	12.39 (0.24)	190	NA
	Winter	75	7.04 (0.18)	0	NA
Age (years)	**0-4**	60	11.67 (0.27)	56	22.37 (0.29)
	5-17	50	15.21 (0.28)	22	12.41 (0.19)
	18-64	40	8.92 (0.23)	88	5.69 (0.19)
	≥65	40	6.45 (0.15)	143	7.97 (0.15)
RTI	**Yes**	77	12.54 (0.22)	59	19.33 (0.30)
	No	113	8.47 (0.20)	247	5.23 (0.16)

Temperature (T) = measure of nestedness. Low temperatures represent more nested (organised) communities and high temperatures represent less nested (random) communities. Matrix fill = proportion of species presences within the matrix. NP = nasopharyngeal swab. RTI = recent respiratory tract infection (within the past month). NA = not applicable.

6.3.2 Swabbing Season

Nestedness temperatures were determined for samples collected in both summer and winter swabbing seasons in order to assess the effect of swabbing season on the organisation of respiratory tract communities. Lower temperatures were observed within the winter season when compared with summer season, as shown in Figure 22. This trend demonstrates greater species organisation within the winter. Furthermore, greater numbers of species were observed in winter samples compared with summer samples. This demonstrates increased numbers of bacterial and viral species circulating amongst individuals within the winter months. Lower matrix fill was also observed within winter samples, which indicates fewer species presences despite greater number of different species.

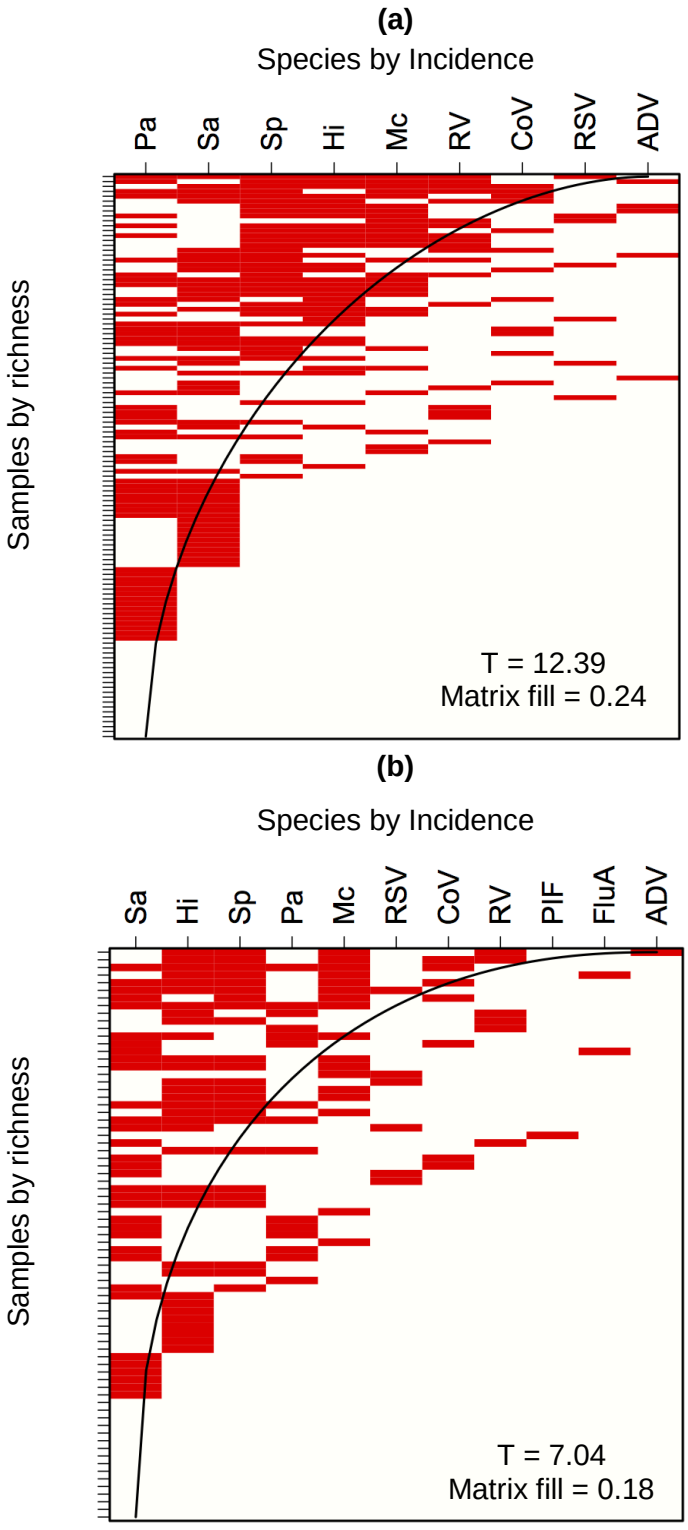


Figure 22. Incidence Matrices of Nose Swab Samples detected by Real-time PCR in (a) summer and (b) winter swabbing time-points

■ = presences, □= absences, − = line of perfect fill, T = 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.

6.3.3 Participant Age

Nestedness temperatures were determined for participants within four age categories (0-4, 5-17, 18-64 and ≥65 years) in order to assess the effect of age on the organisation of respiratory tract communities. A general trend of decreasing temperature with age was observed, as shown in Figure 23, reflecting greater nestedness within older participants. Matrix fill was generally higher in those aged 0-4 years and ≥65 years and lower in those aged 5-17 years and 18-64 years. Order of species incidence varied with age, with the different age groups experiencing different species orders. *S. pneumoniae* and *H. influenzae* have the highest incidence in the young participants (0-4 years) whereas *S. aureus* and *P. aeruginosa* has the highest incidences in older children, adults and elderly participants.

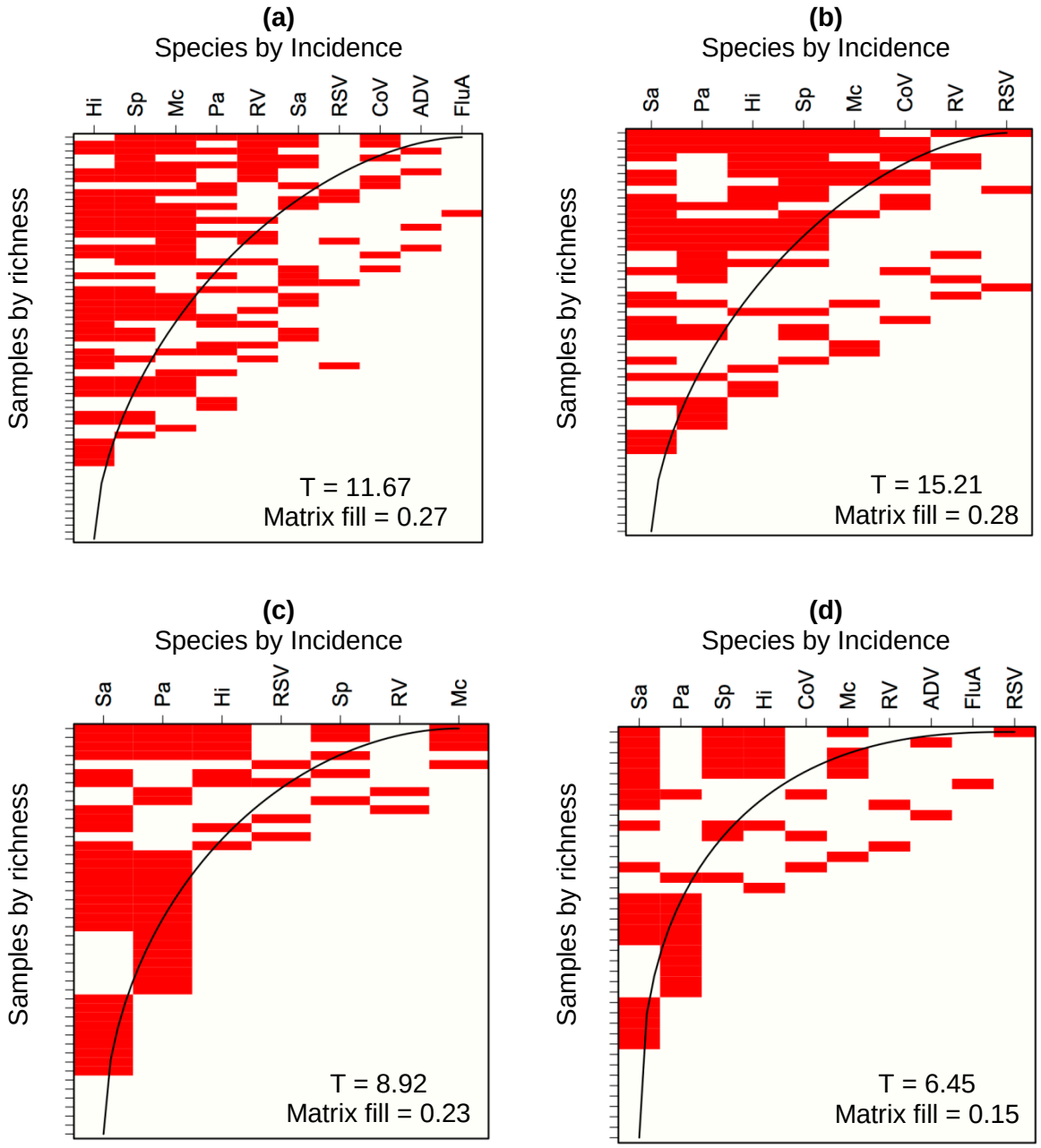


Figure 23. Incidence Matrices in Nose Swab Samples detected by Real-time PCR in Participants Aged (a) 0-4 years (b) 5-17 years (c) 18-64 years and (d) 65+ years

■ = presences, □= absences, − = line of perfect fill, T = 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.

6.3.4 Recent RTI

Nestedness temperatures were determined for both participants who had experienced a recent RTI and participants who had not in order to assess the effect on RTI on the organisation of respiratory tract communities. Individuals with recent RTI demonstrated higher matrix temperatures and thus lower levels of nestedness when compared with samples collected form participants with no recent RTI (Figure 24). Matrix fill was also higher in participants with recent RTI, which demonstrates greater numbers of species present in the upper respiratory tract. Order of species incidence was modified by recent RTI, with *H. influenzae* being more dominant in participants with recent RTI and *S. aureus* being more dominant in cases without recent RTI. Certain viral species, notably parainfluenza viruses and influenza A, are also only present in samples from participants with recent RTI.

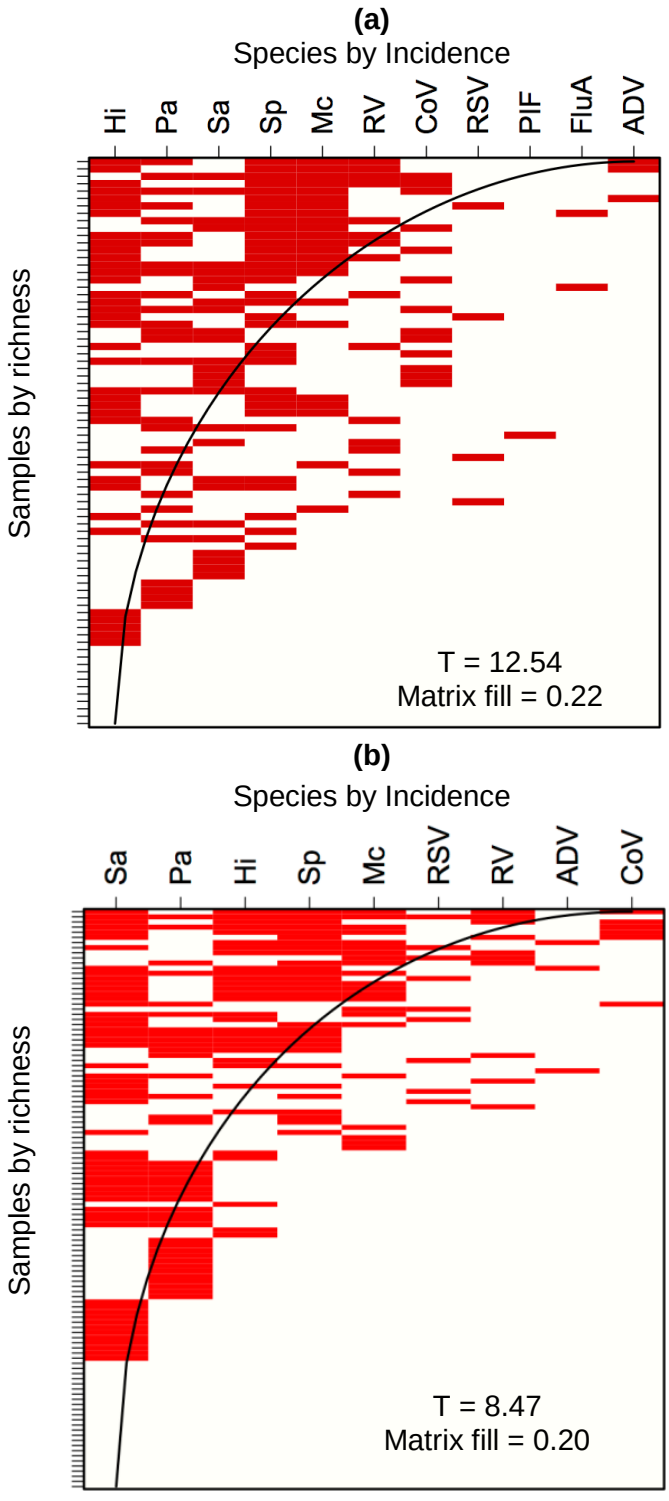


Figure 24. Incidence Matrices of Nose Swab Samples detected by Real-time PCR in Participants (a) with recent RTI (b) with no recent RTI

■ = presences, □= absences, − = line of perfect fill, T = 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.

6.4 Species Distribution within Respiratory Communities

Multiple bacterial and viral species inhabit the ecological niche of the human upper respiratory tract (Bogaert et al., 2011). Ecological species distribution patterns have been observed in insect-plant interactions in order to understand how the number of insects associated with a single tree reflect the overall abundance of that type of tree (Southwood, 1961). These methods were applied to patterns of species distribution within the respiratory tract of different individuals in order to provide an insight into the types of relationships occurring between the different species and to understand how each individual's characteristics might influence the number of species associated with that individual.

This was done by plotting the frequency of different numbers of species types within each sample and comparing it to the Poisson (random) distribution. Deviations of observed data from the Poisson distribution, using the Chi-squared goodness-of-fit test, would then be interpreted. A narrower distribution than the Poisson distribution, with many individuals being colonised by a similar number of species, demonstrates a regular pattern of species distribution and provides indirect evidence for mutual repulsion and competition between the species. On the other hand, a wider distribution than the Poisson distribution, with many individuals having either few or many species, demonstrates a clumped pattern of species distribution and provides indirect evidence for mutual attraction and facilitation between the species. Such information is important for determining interactions occurring between microbial species within the respiratory tract in order to be able to predict the effects of an individual's characteristics on the respiratory community.

6.4.1 Patterns of Species Distribution within Individuals

Species distribution patterns were determined within nose and NP swab types. Samples were assessed according to season, participant age and recent RTI. This would enable the determination of the effects of such factors on the distribution of species within the respiratory tract community and hence to understand the relationships occurring between different species. Overall species distribution results are shown in Figure 25 and detailed in full in Appendix 10.

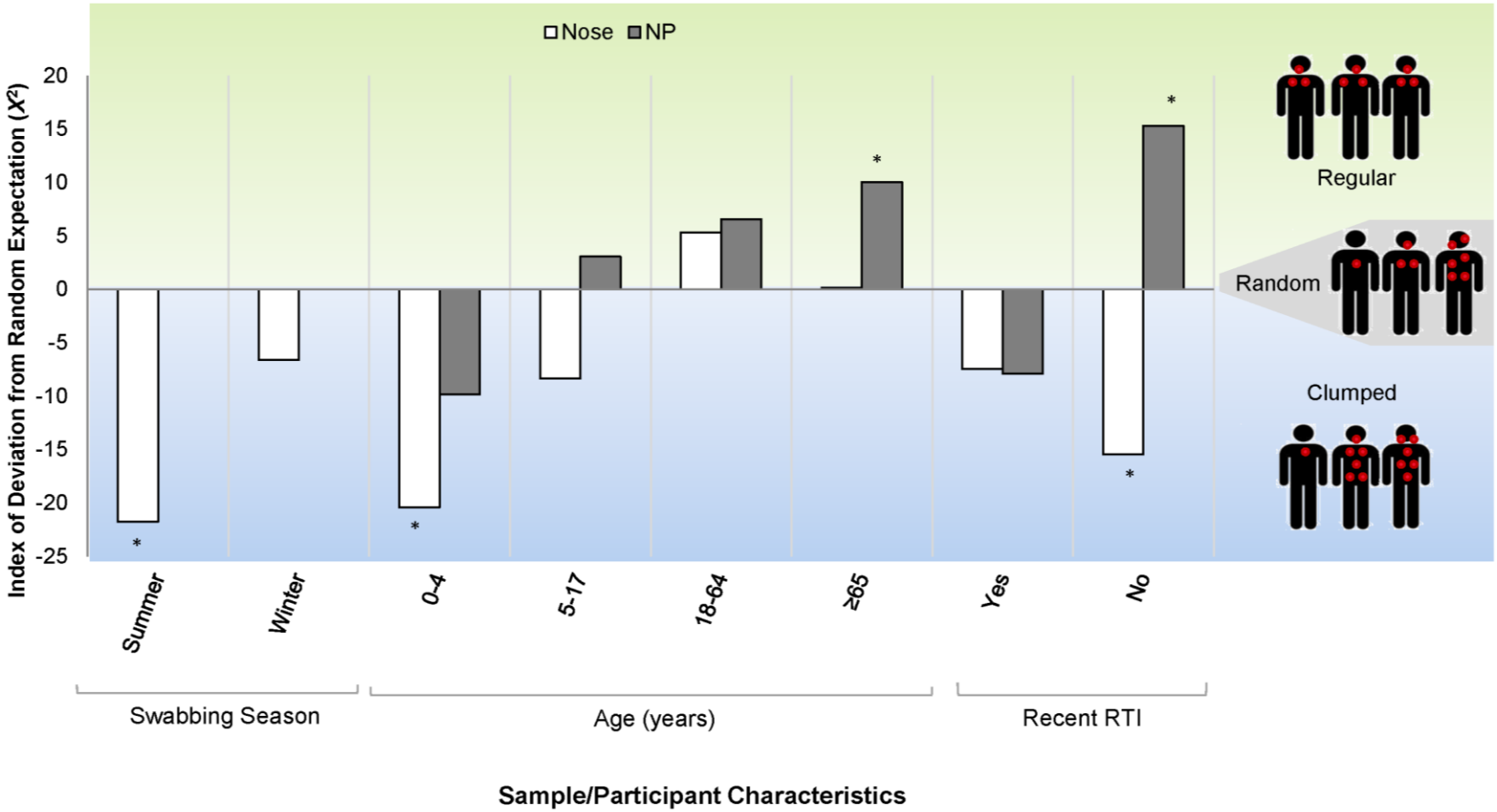


Figure 25. Deviation of Observed Species Distributions from Random Expectation

* Significant deviations from the random distribution (X^2 *P*-value < 0.05). RTI = respiratory tract infection. NP = nasopharyngeal swab. Red dots demonstrate microbial species distributed across individuals. Seasonal data not available for NP swabs.

6.4.2 Swabbing Season

Species distribution patterns were determined for samples collected in both summer and winter swabbing seasons in order to assess the effect of season on the relationships occurring between different microbial species within the respiratory tract community. Distribution of species within nose samples from the summer was wider than the Poisson distribution (X^2 = 21.75, df = 5, p < 0.001) whereas distribution of species within samples from the winter did not deviate from the Poisson distribution. This demonstrates that species in summer samples have a clumped pattern of species distribution within individuals which provides indirect evidence of facilitative relationships between species. On the other hand, distribution of species within winter samples is more random. NP swabs were only collected in the summer and therefore could not be compared according to season.

6.4.3 Participant Age

Species distribution patterns were determined for participants within four age categories (0-4, 5-17, 18-64 and ≥65 years) in order to assess the effect of age on the relationships occurring between different members of the respiratory tract community. Species distribution deviated from random in participants aged 0-4 years in nose samples (X^2 = 20.42, df = 6, p = 0.002). These individuals showed a wider distribution of microbial species than the random distribution. This demonstrates that the species have a clumped pattern of species distribution within individuals which is indirect evidence of facilitative relationships. NP samples also showed a clumped distribution but did not deviate significantly from the Poisson distribution.

Samples collected from older participants, those aged 18-64 years and ≥65 years, showed narrower and hence more regular species distributions than the random distribution. However, only NP samples from individuals aged ≥65 years deviated significantly from random (X^2 = 10.03, df = 3, p = 0.018). This demonstrates that species have a more regular pattern of species distribution within these older individuals which provides indirect evidence of competitive relationships.

6.4.4 Recent RTI

Species distribution patterns were determined for both participants who had experienced a recent RTI and participants who had not in order to assess the effect of RTI on the relationships occurring between different members of the respiratory tract community. Nose samples showed a wider distribution of species in individuals without recent RTI compared to the random distribution (X^2 = 15.44, df = 5, p = 0.009) and hence a clumped distribution of species. However, NP samples showed a narrower distribution in individuals without recent RTI compared with the random distribution (X^2 = 15.30, df = 4, p = 0.004) and hence a regular distribution of species. Participants with recent RTI, on the other hand, showed a more clumped distribution of species within both swab types but this distribution did not deviate significantly from random.

6.5 Bacterial Abundances within Respiratory Communities

Abundance in microbial populations can be measured across samples using patterns of species presence and absence or within samples using relative or absolute quantification. Assessing microbial abundance patterns is important for determining the key species within a biological sample. Furthermore applications include quantification for monitoring the efficacy of antimicrobial treatment (Honeyborne et al., 2011) as well as determining the effects of species abundance in disease (Esposito et al., 2013). Quantitative real-time PCR data from nose and NP swab samples were used in order to determine bacterial species abundances within and between samples. Ecological theory would be applied in order to decipher the interactions occurring between species within the upper respiratory tract.

6.5.1 Bacterial Species Abundances

Quantitative real-time PCR was undertaken on a set of nose and NP samples in order to determine the bacterial colony-forming units per millilitre (CFU/ml) within the swab samples stored in STGG. Mean, minimum and maximum CFU/ml for each species tested are shown below in Figure 26. Differences between concentrations of the different bacterial species within swab samples were assessed using one-way ANOVA. Bacterial

concentrations within the nose were shown to vary ($F(4, 945) = 4.793$, $p = 0.001$), with higher concentrations of *S. aureus* and lower concentrations of *P. aeruginosa* compared to the other bacterial types. Bacterial concentrations within the NP, however, were not different between species ($F(4, 940) = 0.544$, $p = 0.704$).

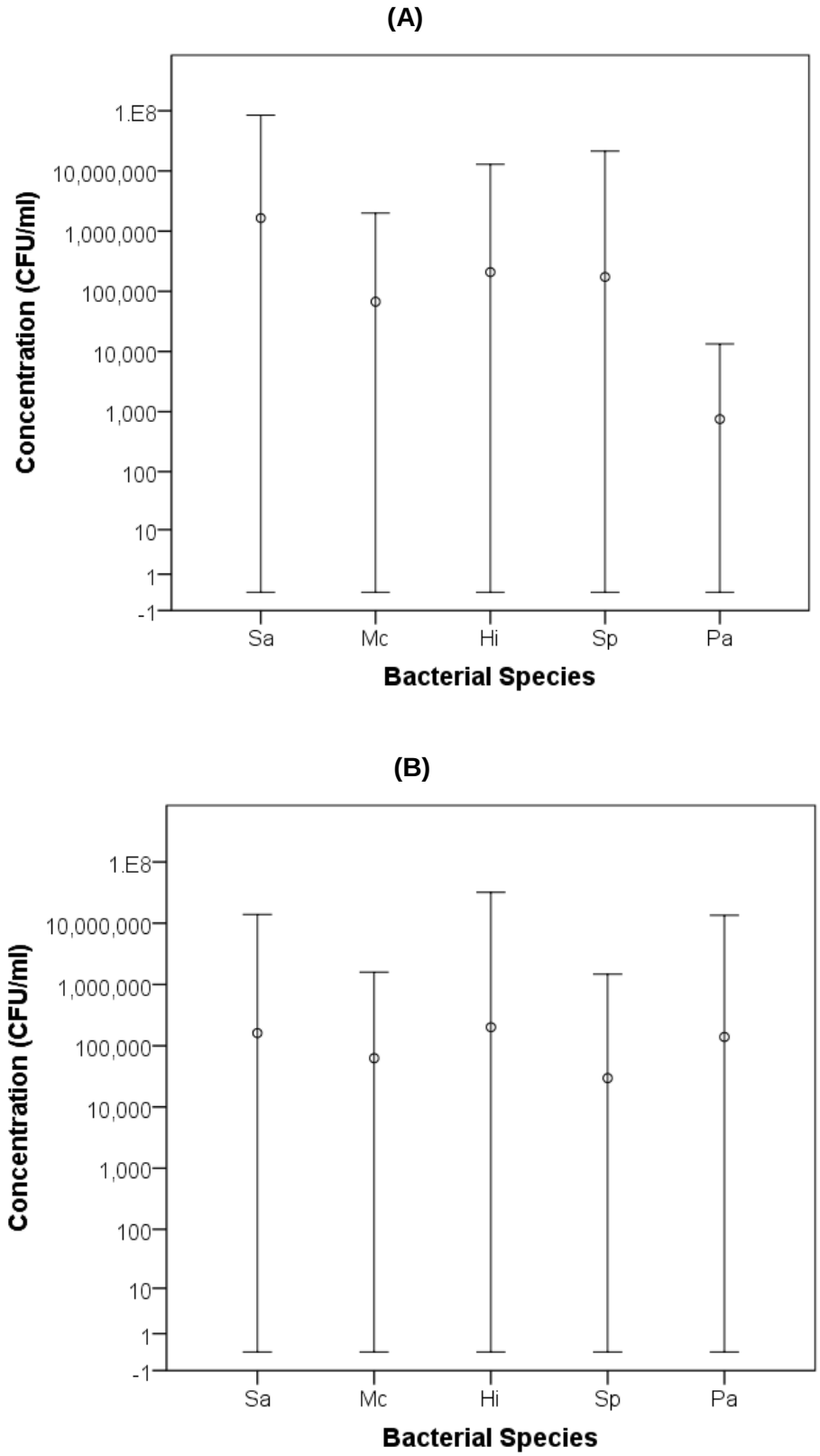


Figure 26. Bacterial Species Concentrations (CFU/ml) within (A) Nose and (B) NP swabs

Circles present mean concentrations across the 190 samples of each swab type and horizontal lines represent minimum and maximum values. Sa = *S. aureus*, Mc = *M. catarrhalis*, Hi = *H. influenzae*, Sp = *S. pneumoniae*, Pa = *P. aeruginosa*. CFU = colony-forming unit.

6.5.2 Ecological Theory an Patterns of Community Assembly

MacArthur's Broken Stick model, as described previously, was applied to the bacterial and viral presence-absence and abundance (CFU/ml) data collected using quantitative real-time PCR. The model consists of ranking species in order of abundance across samples and comparing them to the MacArthur's theoretical rank abundance.

Both observed ranks and theoretical ranks were then plotted against bacterial abundances and a Chi-squared goodness-of-fit test was used to determine deviations from MacArthur's Broken Stick model. MacArthur's Broken Stick models neutral characteristics, and will hereafter be referred to as the 'neutral model'. Deviations from the neutral model indicate niche characteristics of a dominance hierarchy if rare species have abundances below the line or resource segregation if rare species have abundances above the line. Observations that did not deviate from this model were thought to demonstrate neutral characteristics (Doncaster, 2009).

6.5.2.1 Species Abundance Deviations from the Neutral Theory

Species abundance patterns and deviations from the neutral model were determined within nose and NP swab types. Samples were assessed according to swabbing season, participant age and recent RTI. This would enable the determination of the effects of such factors on the abundance patterns of species within the respiratory tract community and hence to understand the contribution of niche or neutral factors in shaping these respiratory tract communities. Overall results of deviations from the neutral model using the Chi-squared test are shown in Table 30 below.

Table 30. Chi-squared Results showing Species Abundance Deviations from MacArthur's Broken Stick Model of Neutral Community Assembly

		Nose			NP		
		X^2	*p*	df	X^2	*p*	df
Season	**Summer**	12.356	0.015	2	NA	NA	NA
	Winter	3.052	0.549	2	NA	NA	NA
Age (years)	**0-4**	1.683	0.794	2	2.115	0.715	2
	5-17	11.243	0.024	2	1.767	0.779	2
	18-64	28.784	<0.001	2	21.659	<0.001	2
	≥65	29.689	<0.001	2	2.061	0.357	0
RTI	**Yes**	7.385	0.117	2	2.090	0.719	2
	No	9.025	0.050	2	10.109	0.039	2

Results show deviations of bacterial abundances (colony-forming units per millilitre) from MacArthur's Broken Stick model, a neutral model of community assembly, using Chi-squared (X^2) goodness-of-fit test. df = degrees of freedom. NA = not applicable. RTI = respiratory tract infection, NP = nasopharyngeal swab.

6.5.2.2 Swabbing Season

Species abundance patterns and deviations from the neutral model were determined for samples collected in both summer and winter swabbing seasons in order to assess the effect of swabbing season on species abundance patterns. Species abundances within nose swabs collected from the summer deviated from the neutral model with fewer individuals then the neutral model, which is indicative of niche characteristics of a dominance hierarchy. On the other hand, samples collected from the winter did not deviate from the neutral model and therefore fit more closely to neutral theory than niche theory.

6.5.2.3 Participant Age

Species abundance patterns and deviations from the neutral model were determined for samples collected from participants within four age groups (0-4, 5-17, 18-64 and ≥65 years) in order to assess the effect of age on species abundance patterns. Bacterial abundances in nose swabs from participants aged 5-17 years, 18-64 years and ≥65 years deviated from the neutral model (Figure 27). All of these age groups demonstrated fewer individuals than the neutral model. Bacterial abundances in NP swabs in participants aged 18-64 years also deviated from the neutral model. Again, this age group demonstrated fewer individuals than this model reflecting niche characteristics of a dominance hierarchy.

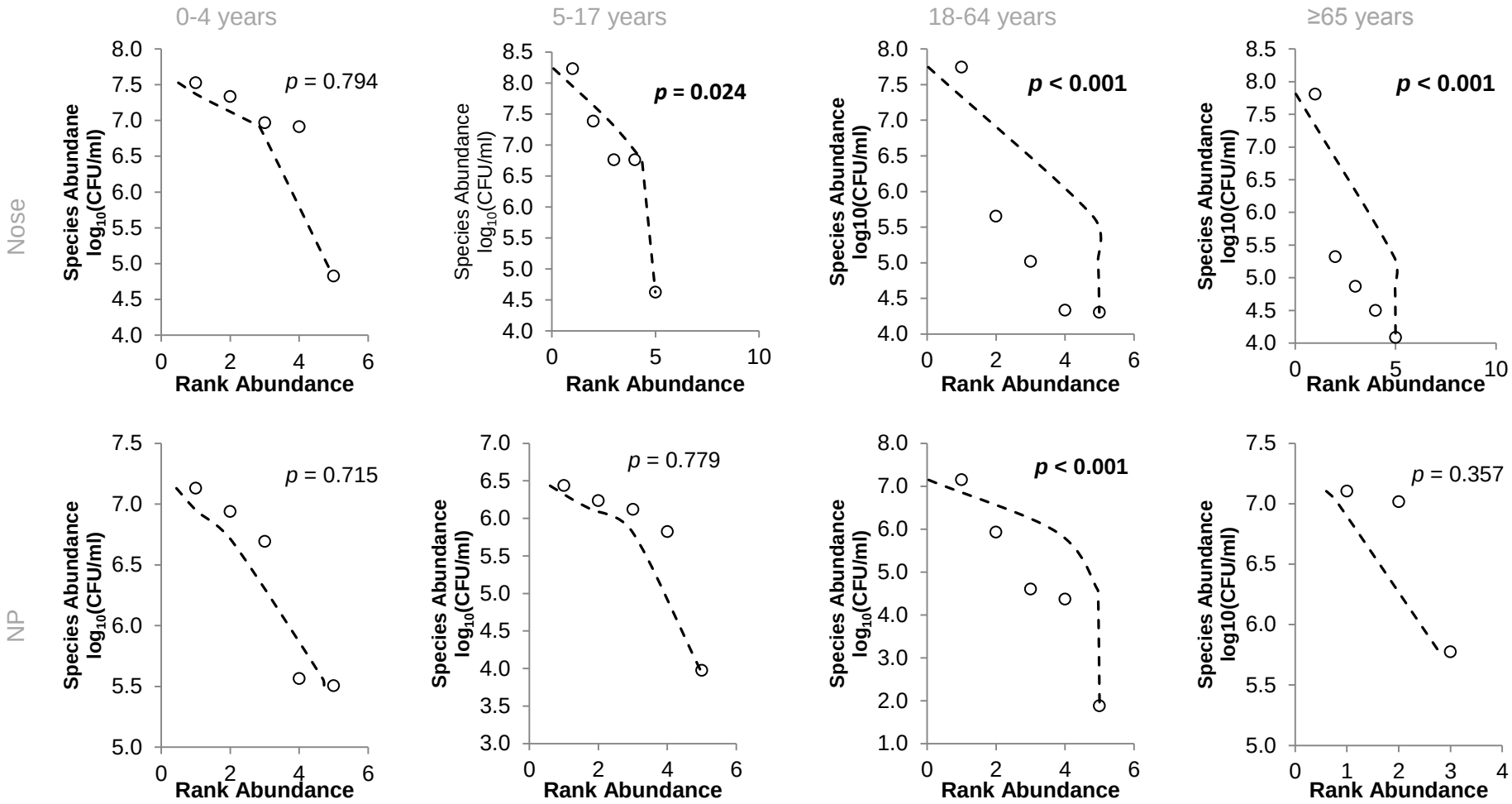


Figure 27. Species Abundance Curves (CFU/ml) for Nose and NP Swabs according to Age Group

o = species abundance, -- = MacArthur's Broken Stick model of neutral community assembly, deviations with X^2 *P*-value < 0.05 in bold. CFU = colony-forming unit.

6.5.2.4 Recent RTI

Species abundance patterns and deviations from the neutral model were determined for samples collected from participants who had recently experienced a RTI and participants who had not in order to assess the effect of RTI on species abundance patterns. Neither participants with recent RTI nor participants without recent RTI appeared to demonstrate species abundances that deviated from the neutral model within nose samples. NP samples, on the other hand, demonstrated species abundances that deviated from the neutral model in participants without recent RTI. These participants demonstrated samples with fewer individuals than this model, which is again indicative of niche characteristics of a dominance hierarchy.

6.6 Discussion

Ecological analysis of the upper respiratory tract community has allowed key patterns to be deciphered. Patterns of nestedness within these samples demonstrated bacterial communities to be more nested within samples collected within the winter season, from older individuals as well as individuals without recent RTI. There are a number of explanations which have been described for understanding patterns of nestedness, these include the abundance of species whereby more abundant species are more likely to colonise greater numbers of individuals (known as passive sampling or neutrality) (Andren, 1994, Ulrich and Zalewski, 2007). Furthermore, isolation is another possible explanation whereby greater isolation of sites (here, individuals) affects the dispersal ability of the species (Darlington, 1957). Additionally, selective pressures of the environment can also determine species distribution (Ulrich et al., 2009).

Isolation is a potential explanation for low nestedness within young individuals as this age group is thought to enable greater transmission of species between individuals and hence has lower isolation of species within individuals of this age group. Attendance at day care is one environment in which greater species dispersal occurs as a result of the proximity of children (Sa-Leado et al., 2008). On the other hand, adults do not offer the same frequency of transmission opportunities and hence isolation of species may be more common. Similarly individuals with recent RTI may also enable greater transmission of species between individuals (due to symptoms of viral infection such as coryza) and have

lower isolation of species than those without recent RTI. Finally the selective pressures of the immune system may also determine nestedness of species. Previous modelling has shown that vaccine-induced selective pressures of PCV initially resulted in greater diversity of *S. pneumoniae* but over time diversity returned to pre-PCV levels (Hanage et al., 2010). Furthermore, again in *S. pneumoniae*, both PCV-induced serotype-specific immunity and acquired immunity have been shown to be involved in maintaining diversity of the species, sustaining competition and eliminating any fitness advantages between serotypes (Cobey and Lipsitch, 2012). Although we cannot determine the effects of the immune system within this ecological analysis, it is likely to play an important role in shaping the microbial community structure.

Bacterial species *S. pneumoniae, H. influenzae* and *M. catarrhalis* as well as viral species coronavirus and rhinovirus were commonly found higher up the order of species incidence in samples with lower levels of nestedness. This signifies that these species are more abundant than the other targeted species in younger individuals, those with RTI or from samples collected during the summer season. On the other hand, *S. aureus* and RSV were commonly found higher up the order of species incidence in samples with higher levels of nestedness. This signifies that these species are more abundant than the other targeted species in older individuals, those without RTI or from samples collected during the winter season. These observations have already been discussed in Chapter 4. Influenza A and parainfluenza viruses were only found in individuals with recent RTI which highlights the role of these species in causing infection. In most samples, viral species were commonly found lower down in the order of species incidence than bacteria which indicates that bacteria colonise and persist for detectably longer than viruses.

Species distribution patterns were found to be significantly different from the random distribution within young individuals (0-4 years) and within samples collected from the summer season, with a wider distribution and clumping of species. Older age groups tended to have more regular species distribution patterns, although a significant deviation from random was only observed within the ≥65 year age group. Furthermore, winter samples also showed a wider distribution of species but not significantly different from random. Individuals without recent RTI demonstrated different results for the two swab types, with a wider distribution and clumping of species in the nose and a more regular distribution within the NP. For individuals with recent RTI, a more clumped distribution of species was observed, however this did not deviate from random in either swab type.

Clumping of species demonstrates that individuals have few or many colonising species and indirectly indicates the presence of facilitative relationships occurring between

species (Tirado and Pugnaire, 2005). Facilitative relationships occurring between species in young individuals is thought to be the result of increased transmission opportunities within these individuals. Facilitative relationships may have evolved in order to enable survival and spread of species carried within the respiratory tract (Bogaert et al., 2004a, Sa-Leado et al., 2008). Although not significant in this study, the clumping of species within individuals with recent RTI is potentially the result of viral infections facilitating bacterial colonisation and even disease. Influenza virus and coronavirus have previously been shown to facilitate bacterial invasion and have been associated with secondary bacterial infections (McCullers and Rehg, 2002, Golda et al., 2011). In terms of bacterial colonisation, viral infection is thought to promote the spread of colonising species between hosts (Moore et al., 2010, Jacoby et al., 2007).

Regular distribution of species demonstrates individuals have similar numbers of colonising species and indirectly indicates the presence of competitive relationships between species (Getzin et al., 2006). Competitive relationships occurring between species in older individuals is potentially the result of greater carriage of species such as *S. aureus* which are though to compete with other respiratory species (Regev-Yochay et al., 2004a). Furthermore, maturation of the immune system as well as the lower prevalence of viruses such as rhinovirus might also influence the competitive relationships observed between species (Siegrist, 2008).

Contradicting patterns of species distribution in individuals without recent RTI might be the result of sampling site. The nose demonstrates more clumping of species, where as the NP demonstrates a more regular distribution in these individuals. The nose might display such facilitative relationships due to greater potential for transmission of species within this more exposed anatomical site via hand-to-nose contact and direct transmission of infectious agents to the nasal mucosal layer which is thought to be important in the transmission of respiratory viruses and bacteria (Goldmann, 2000, Monto, 2002). However, other modes of transmission are also important including aerosol transmission (Musher, 2003).

Species abundances demonstrated no deviation from the neutral model in young individuals (0-4 years), individuals with recent RTI and within samples collected from the winter season. However, in older individuals (≥5 years), in those without recent RTI and in samples collected from the summer season species abundances deviated from the neutral model with fewer individuals than would be expected by this model, demonstrating niche characteristics of a dominance hierarchy. Samples demonstrating more neutral community assembly patterns indicate the importance of stochastic factors such as

speciation, dispersal, immigration, birth and death in controlling levels of species abundance and diversity within the respiratory tract microbial community (Hubbell, 2001). However, samples demonstrating more niche community assembly patterns indicate the importance of environmental factors in determining the abundances and distribution of species (Mikkelson, 2005). In addition, other analyses performed within this chapter and other chapters would seem to indicate that environmental factors also play an important role within the microbial community. It is likely that there are both neutral and niche characteristics to the upper respiratory tract community with stochastic factors and environmental factors being involved. Other studies have demonstrated the role of both niche and neutral processes with gastrointestinal and microbial communities (Jeraldo et al., 2012, Dumbrell et al., 2010).

The application of ecological theory to multi-species microbial carriage data is novel and hence the models have their limitations. The models are simplistic in nature as they do not account for acquired or specific host immunity which is important in shaping the ecology of the respiratory tract (Cobey and Lipsitch, 2012). Furthermore, as discussed previously, the real-time PCR samples were non-randomly sampled from the collection of swab samples from the swabbing study and hence may have led to some level of bias. Furthermore, different types of RTI have been analysed together within these models, however in reality these infections may have very different effects on the respiratory tract community and should preferably be analysed separately. The inclusion of both bacteria and viruses together within these models is also a potential limitation due to the differing nature of these microorganisms. However, due to the fact that they colonise the same habitats and interact with each other, it seems feasible to include them as a single ecology. Further analysis should include sensitivity analysis in order to determine whether inclusion of bacteria and viruses together or separately is optimal for understanding the respiratory community.

Nestedness analysis is limited by the fact that no statistical measure of difference between temperatures can be used. This means that the results cannot be statistically compared and can therefore only be descriptive. Furthermore, additional information is thought to be required in order to determine the underlying mechanisms of nestedness such as level of immunity (Ulrich et al., 2009, Cobey and Lipsitch, 2012). Furthermore, species distribution analysis is unable to demonstrate the true complexity of microbial, which may be different between any pair of species within the respiratory tract and affected by other factors which have not been considered here. Neutral theory analysis is also limited by the fact that quantification of viruses was not possible due to difficulties in culturing and accurately quantifying viral species abundance.

Furthermore, these ecological analyses focus on a limited number of targeted species and hence a proportion of the species within the respiratory microbiome are missing within these models. This may mean that these models could be different in the presence of these other species. However, the inclusion of a larger number of species in the analysis of real-time PCR results would be unfeasible and hence other methods such as 16S rDNA sequencing should be considered. The respiratory microbiome is highly variable and has shown to be affected by environmental factors such as smoking (Morris et al., 2013), season (Bogaert et al., 2011) and viral infection (Allen et al., 2014).

Chapter 7. Discussion

The primary aim of the project was to gain insight into the epidemiology and ecology of upper respiratory tract communities. This included gaining an understanding of the participation of members of the general population to a large community-based swabbing study. Furthermore, to achieve this aim, the determination of carriage prevalence of target bacterial and viral species and the construction of a mathematical model for microbial respiratory tract carriage were undertaken. Additionally, this involved gaining insight into the inter-species and intra-species microbial diversity of the upper respiratory tract community. Finally, species distribution and community assembly patterns within the respiratory tract community were determined.

In order to achieve these aims, a large community-based swabbing study was undertaken. This involved the collection of swab samples from members of the general population, including self-taken nose and whole mouth swabs as well as HCP-taken nasopharyngeal and whole mouth swabs. A total of 2,417 participating individuals from 20 GP practices across Hampshire in the United Kingdom enabled the collection of a large set of swab samples. These samples were then analysed using established culture-based methods, molecular techniques, whole genome sequencing as well as ecological and mathematical modelling approaches.

7.1 Participation in a Community Swabbing Study

Determination of levels of participation of members of the general population within this swabbing was essential for gaining an accurate measure of microbial carriage within the upper respiratory tract. Level of participation was assessed using information gathered on numbers of invitation letters or swab packs sent out to self-swabbing and HCP swabbing groups respectively. Results from the swabbing study showed participation to be higher for participants undertaking self-swabbing than those undertaking HCP swabbing (Coughtrie et al., 2014). Additionally higher responses were achieved in older children and adults as well as in less deprived geographical areas, demonstrating the associations between age and deprivation with participation in a community-based study (Goldberg et al., 2001). The usefulness of self-swabbing has been reported in studies of respiratory viruses as well as *S. aureus* (Akmatov et al., 2012, Gamblin et al., 2013). Self-swabbing has also been shown to dramatically reduce the costs of a study compared to HCP

swabbing (Akmatov et al., 2011). These results have implications for the set up of future large carriage studies as self-swabbing is shown to be a potential alternative to more traditional and costly HCP swabbing methods.

Although numbers of responses were fewer than anticipated, a large number of samples were obtained which has provided important information regarding the epidemiology and ecology of microbial species carried within the upper respiratory tract. Furthermore, the targeting of a large variety of individuals, in terms of age range, geographical area deprivation as well as demographic characteristics, has enabled the collection of a diverse set of samples. A number of carriage studies have previously compared swabbing methodologies for identification of single bacterial species (Watt et al., 2004, Uemura et al., 2004, Vaneechoutte et al., 1990). Other studies have compared carriage of multiple bacterial species within a single swab site (Kwambana et al., 2011, Mackenzie et al., 2010, Harrison et al., 1999). Within this study, a larger set of species were targeted using self- and HCP-swabbing methods and within four different swab types. This has enabled a more realistic insight into carriage within the respiratory tract to be generated, which is closer to the true microbial community of these sample sites.

There are, however, several limitations to this study which should be taken into account in the design of future large carriage studies. Numbers of responses might be further improved via the use of reminder letters, which would help to capture the proportion of individuals who may have forgotten or not had time to take part when the first letter or swab pack was sent out. Reminders were found to increase response rates in mail-based studies (Asch et al., 1997). The use of incentives for participation is a further method to enable greater levels of participation without necessarily incurring excessive additional costs. For example, the use of prize draws to incentivise participation might raise the level of response and the number of samples obtained. Financial incentives have also previously been shown to increase the participation of individuals within mail-based studies, with greater effort in survey responses (James and Bolstein, 1990). Additionally, the collection of information about non-responders would enable increased understanding of the reasons for non-response, which might in turn enable better study design. Examples of this might include the use of additional small questionnaires sent specifically to non-responders, to capture additional information about these individuals as well as to check for any bias which might result from non-response. A study analysing non-response in mail-based studies showed questionnaires and telephone calls to be successful in obtaining non-responder information (Kotaniemi et al., 2001).

7.2 Microbial Carriage in the Upper Respiratory Tract

Culture-based identification techniques were used to identify six common bacterial respiratory tract colonisers: *S. pneumoniae, H. influenzae, M. catarrhalis, S. aureus, P. aeruginosa* and *N. meningitidis*. Real-time PCR was also used on a subset of samples in order to identify the same six bacteria as well as eight viral species: influenza A virus, influenza B virus, adenovirus, metapneumovirus, RSV, parainfluenza virus, coronavirus and rhinovirus/enterovirus. A vast amount of microbial carriage data was collected which has provided important information on microbial prevalence within the human upper respiratory tract and has helped to determine specific patterns of carriage within this set of individuals.

Carriage results demonstrated the non-inferiority of a self-taken nose swab for the detection of *S. pneumoniae, M. catarrhalis, S. aureus, P. aeruginosa, H. influenzae*, adenovirus and coronavirus when compared with a HCP-taken NP swab. Other studies have also shown nose swabs to be as effective as NP swabs in detecting carriage of *S. pneumoniae* and *H. influenzae* (Rapola et al., 1997). However, WMS were shown to enable greater detection of *M. catarrhalis* and *P. aeruginosa* than nose or NP swabs. The detection of a single *N. meningitidis* isolate in swab samples may be due to the type of swabs used, as an oropharyngeal swab is preferable (Olcen et al., 1979). Furthermore, delays in swab transport may also have resulted in fewer isolates than expected as *N. meningitidis* survive for shorter periods of time (24h) than other respiratory species such as *S. pneumoniae* (48h), although the use of charcoal can promote survival (Farhat et al., 2001, Caugant et al., 2007). Furthermore direct plating of swabs may also be a better approach for this species (Cunningham et al., 2001). Viral prevalence was found to be lower than that of bacterial species, which is thought to be a result of viruses being less associated with carriage but rather symptomatic infection (Jartti et al., 2008).

Multivariate analysis demonstrated age to significantly affect carriage of a number of bacteria and viruses, with young participants experiencing higher carriage of *S. pneumoniae, H. influenzae, P. aeruginosa* and rhinovirus/enterovirus. Recent RTI was also associated with increased prevalence of *S. pneumoniae*, rhinovirus/enterovirus and coronavirus but decreased prevalence of *M. catarrhalis*. Co-carriage was also common between a number of bacteria and viruses and was significantly associated with young children and the occurrence of a recent RTI.

This knowledge will help to inform the targeting of future treatment and prevention strategies. Such knowledge is key to the development of new vaccines targeting *S. aureus* and *M. catarrhalis* (Harro et al., 2012, Otsuka et al., 2014) as well as new vaccines targeting greater numbers of *S. pneumoniae* serotypes (Skinner et al., 2011). This study was a cross-sectional study, which enabled the prevalence of microbial carriage to be assessed. This type of study design has a number of advantages: the association between outcomes and exposure can be explored, several outcomes can be studied within the same study and they are often cheap and easy to undertake (Mann, 2003).

Despite the advantages of cross-sectional studies, there may also be advantages of performing other types of studies to understand microbial carriage in the upper respiratory tract. The extension of this study into a longitudinal study across a larger number of geographical areas and countries would enable the capture of longitudinal data as well as to determine the generalisability of patterns and trends across geographic regions both nationally and internationally. Longitudinal studies have been shown to provide important information regarding the transmission patterns of respiratory pathogens amongst individuals (Hill et al., 2010). The collection of carriage data from individuals at regular time-points, in order to assess incidence, could provide key information on the rate, and hence risk, of carriage over time. A recent incidence study has shown the risk of non-vaccine type *S. pneumoniae* in causing IPD in eight centres across the USA (Hicks et al., 2007).

Swab types, transport media and storage media used in the study, although enable viral survival, were not thought to be optimal for the detection of viral species. Swab types are not optimised for the collection of respiratory epithelial cells which the viruses infect. Flocked swabs, on the other hand, have been shown to provide enhanced collection of respiratory epithelial cells and hence the collection of respiratory viruses compared to rayon swabs (Daley et al., 2006). Transport media is also crucial for preventing damage of viral particles as well as virus survival which may be important where viruses will subsequently be grown in cell culture (Johnson, 1990). Viral transport medium (VTM) is commonly used for preserving viruses during sample transport (Madeley, 1977). However, in contrast to STGG, which can support both viral and bacterial species, VTM does not support bacterial growth due to the inclusion of antibiotic agents (Turner et al., 2011b). The use of additional specific viral transport and storage media could provide more accurate data on viral prevalence, but this would need to be achieved via additional swabs.

The use of real-time PCR is thought to more sensitive in detecting a number of bacterial species than more traditional culture-based methods (McAvin et al., 2001, Ueyama et al., 1995). However, the detection of nucleic acids to identify species may also result in the detection of dead bacterial cells which may artificially raise the frequency of carriage of these bacterial respiratory species (Rogers et al., 2008). However, the interpretation of such culture-negative PCR-positive samples is difficult. False-positives in non-sterile sites such as the upper and lower respiratory tract are problematic when studying infection as their clinical significance may be questionable (Guclu et al., 2005).

7.3 Phenotypic and Molecular Diversity of Respiratory Bacteria

Molecular and phenotypic typing of bacterial species was performed using PCR serotyping, antibiotic sensitivity testing as well as *in silico* serotyping, multi-locus sequence typing, vaccine candidate and antibiotic resistance gene analysis. These analyses revealed high levels of intra-species diversity, low levels of antibiotic resistance genes and low levels of vaccine type *S. pneumoniae* and *H. influenzae*. High intra-species diversity within these respiratory isolates is potentially a result of the pressures imposed by current vaccination strategies (Croucher et al., 2011) as well as the high levels of recombination in *S. pneumoniae* (Golubchik et al., 2012). Few antibiotic resistance genes within this group of isolates reflects the low and stable levels of antibiotic-resistant respiratory species in the UK (Reynolds, 2009). Continued monitoring of levels of antibiotic resistance within respiratory isolates within the UK is essential for clinical treatment of infection (Reynolds, 2009). The lack of widespread vaccine candidate gene presence within the isolates collected in this study demonstrates one of a number of difficulties in the development of protein-based vaccines against these species (Tai, 2006).

Few vaccine type *S. pneumoniae* and *H. influenzae* within this set of isolates is indicative of the effectiveness of current PCV-13 and Hib vaccines in targeting specific serotypes (Tocheva et al., 2011, Miller et al., 2011). The high proportion of non vaccine-type *S. pneumoniae* is reflective of the process of serotype replacement which has seen as shift in serotypes as a result of vaccination. Serotype replacement in carriage is not thought to lead to future large increases in the incidence of invasive disease caused by non vaccine type *S. pneumoniae* (Hanage et al., 2010). The identification of increasingly common non

vaccine-type serotypes such as *S. pneumoniae* serotypes 6C, 11A/D and 23B and non-typeable *H. influenzae* have provided important information instrumental for the epidemiological assessment of vaccines and for the development of future vaccines.

In silico analysis has been shown to have increased sensitivity and resolution for detection of *S. pneumoniae* serotypes (Gladstone, 2013). The application of these rapid typing methods is vital in clinical microbiology and in the understanding patterns of antimicrobial resistance and virulence (Inouye et al., 2014b). Furthermore whole genome sequencing technology has been shown to be useful in monitoring outbreak situations (Grad et al., 2012). However, currently whole genome sequencing remains expensive and substantial changes in the infrastructure, training of staff and computing resources would be required before the method could be used routinely in clinical laboratories (Didelot et al., 2012). Furthermore, interpretation of sequencing results may also be difficult, especially with sequencing technologies such as 16S rDNA analysis (Woo et al., 2008).

In silico methods for assessing presence of capsular, antibiotic resistance and vaccine candidate genes, although a powerful tool, do not provide gene expression information. Further serological typing, minimum inhibitory concentration (MIC) testing and vaccine candidate gene expression analysis (such as microarray) would provide a more accurate indication of the expression of these genes (Austrian, 1976, Jorgensen et al., 1990, Peterson et al., 2000). A further limitation of this analysis is that a limited set of candidate vaccine antigens was analysed within a small set of bacterial isolates. Further bioinformatics analyses including the use of single nucleotide polymorphism (SNP) analysis on a larger set of isolates would provide additional more in-depth information on the incidence of these and other vaccine candidate antigens (Landegren et al., 1998).

7.4 Ecology of Microbial Respiratory Tract Communities

Ecological analysis of respiratory tract samples involved the use of established methods for understanding the assembly and interactions of microbial species within a novel context, that of the respiratory community. Nestedness, species distribution and species abundance analyses were undertaken using real-time PCR data from nose and NP swabs.

These analyses revealed respiratory tract communities to be highly nested in older participants, those without recent RTI and within winter samples but less nested in young

children, those with recent RTI and summer samples. Lower levels of nestedness in these individuals is thought to 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 distribution of species within young individuals, demonstrating facilitative relationships occurring between the species within the upper respiratory tract. Facilitative relationships are thought to enhance transmission opportunities and survival of species carried in the respiratory tract. Respiratory species abundances were found to fit more closely to the neutral model of community assembly within young individuals, those with RTI and winter samples but deviated from this model in older participants, those without recent RTI and summer samples with species abundances fewer than could be explained by the neutral model, demonstrating niche characteristics of a dominance hierarchy. There is likely to be both neutral and niche processes involved the microbial community formation in the upper respiratory tract with speciation and the environment playing key roles (Jeraldo et al., 2012, Dumbrell et al., 2010).

The application of ecology theory to sets of respiratory tract samples has allowed the understanding of species patterns and interactions within communities. The development of such 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 (Costello et al., 2009). Ecological theory may also go hand in hand with phylogenetic methods in order to understand the mechanisms underlying community structure within non-culturable communities (Jeraldo et al., 2012).

In order to improve ecological analyses, viral abundance data (virus-like particles per mL) could be collected. This has previously been achieved in other studies via culture-based, microscopy or flow cytometry techniques (Suttle, 2007). However, culture-based methods do not allow the accurate quantification of total viral abundance within a sample and transmission electron microscopy is a highly inaccurate and variable method (Suttle, 2007). Flow cytometry is a useful method, however, methodological inconsistencies have been problematic in the use of this method (Brussaard, 2004). The use of 16S community analysis would also offer greater information on the whole bacterial community of the respiratory tract, as even increased numbers of targets within culture and real-time PCR will inevitably miss more obscure, novel or non-culturable species (Woo et al., 2008). However, 16S analysis does not target viral species. The use of ecological theory to understand community assembly patterns also has its limitations. Neutral theory assumes

there is no variation in the number of total species within a community whilst niche theory does not explain the process of dispersal (Mikkelson, 2005, Chase and Leibold, 2003). It had been proposed that niche and neutral theories should be used in conjunction, a process known as synthesis (Mikkelson, 2005, Hubbell, 2001).

7.5 Future Work

Future work will be conducted to address the limitations and expand the knowledge gained from this study. A larger community-based swabbing study is likely to be conducted across multiple geographical areas within the United Kingdom. This will reduce the inevitable bias associated with conducting a study within a very small geographical area and enable the generalisability of the findings from this study to be established (Bellomo et al., 2009). Further work will also entail the development of ecological aspects of this project using 16S rDNA sequencing. This method will enable the collection of abundance data from the entire bacterial community and will provide results which are not biased by detection of specific species (Woo et al., 2008). Finally, further work will be undertaken in order to further develop the ecological modelling of multiple respiratory tract species in order to account for acquired or specific host immunity as has been undertaken on *S. pneumoniae* (Cobey and Lipsitch, 2012).

7.6 Final Conclusion

This project has produced important epidemiological and ecological information that will be instrumental in the fight against respiratory disease, meningitis and septicaemia. Firstly, it has provided information that will help to inform future carriage study design, which will not only reduce the costs of such studies but also improve outcomes. Such information is also applicable to larger multi-centre carriage studies, which will help to tackle respiratory disease, septicaemia and meningitis on a global scale. Secondly, the project has provided key epidemiological information regarding patterns and trends in circulating bacterial species and types. This information will help to inform the development of new treatment and prevention strategies as well as targeted and effective antibiotics and vaccination policies. Finally, the project has produced novel ecological information regarding the interactions and community structure of microbial respiratory tract communities. This information paves the way for ecological modelling of community

composition that may facilitate predictions of microbial variation as a result of interventions such as vaccination. Future work will focus on 16S rDNA community analysis and ecological methodologies in order to gain insight into the true diversity of the bacterial upper respiratory tract community and to detect novel and non-culturable species as well as the set up of future large carriage studies.

Appendices

Appendix 1. Study Questionnaire

Medicine

UNIVERSITY OF Southampton

QUESTIONNAIRE - ADULT

Title of Project: Analysis of the microbial community of the upper respiratory tract to support the development of effective vaccine policy

Study number .....

Please tick the relevant boxes and provide extra information where necessary.

Have you received all the vaccines you are due to have had so far in line with the national schedule?

Yes ☐ No ☐ Don't know ☐

Have you had a cold, flu, ear infection, or other respiratory infection during the past month?

Cold ☐ Flu ☐ Ear infection ☐

Other chest infection ☐ None ☐

Have you received antibiotics over the past month?

Yes ☐ No ☐

If yes, do you remember the name: ……

Amoxil ☐ Augmentin (=coamoxiclav) ☐ Erythromycin ☐

Azithromycin ☐ Other (know name, please state below) ☐ Other (don't know name) ☐

'Other' antibiotic name: ……

What is your age? ..... years

Thank you for completing the questionnaire.

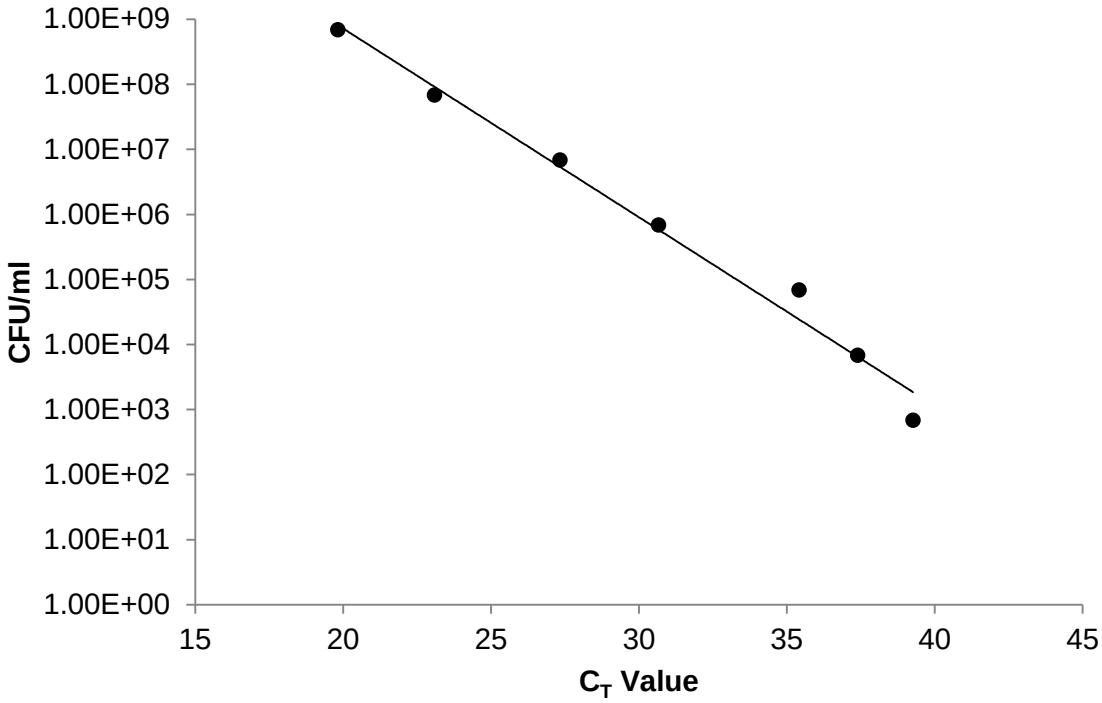
REC Number: 11/SC/0518
Version 1, 15/11/11

Appendix 2. Study Laboratory Sample Sheet

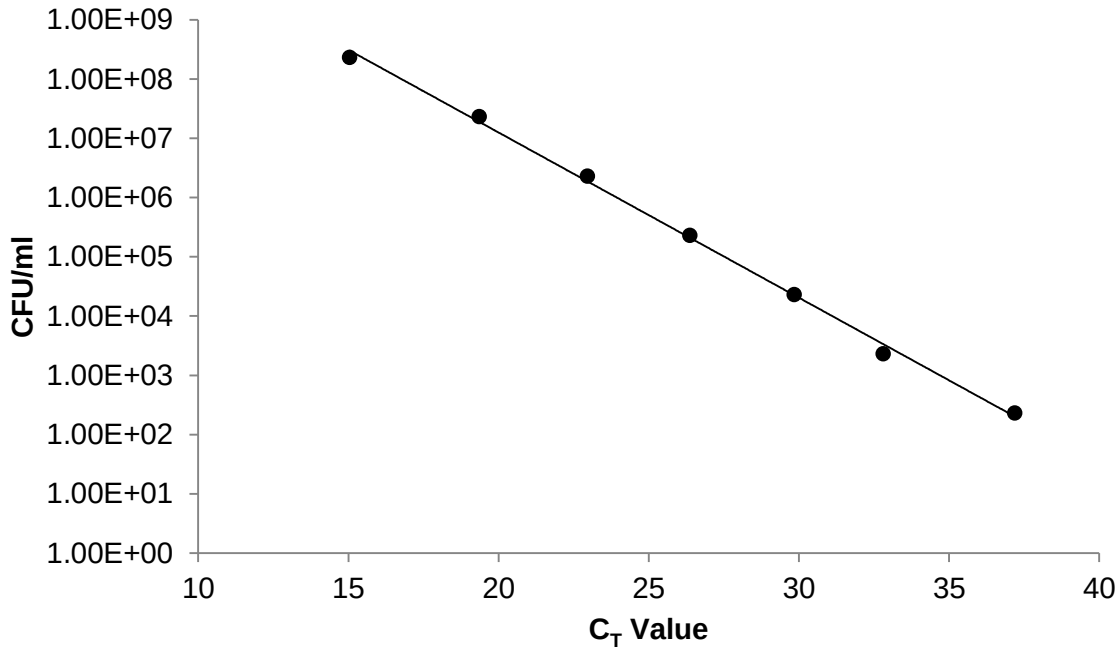
ID:	CBA	CHOC	CNA (Selective for *Strep* and *Staph* spp)	BACH (Selective for *Haemophilus* spp)	Pseudo CFC (*P.aeruginosa* agar)
1° Culture					VCAT-GC Agar (*N.meningiditis agar*)

Appendix 3. Standard Curves for Quantification of Bacterial Species

(A) *S. aureus*

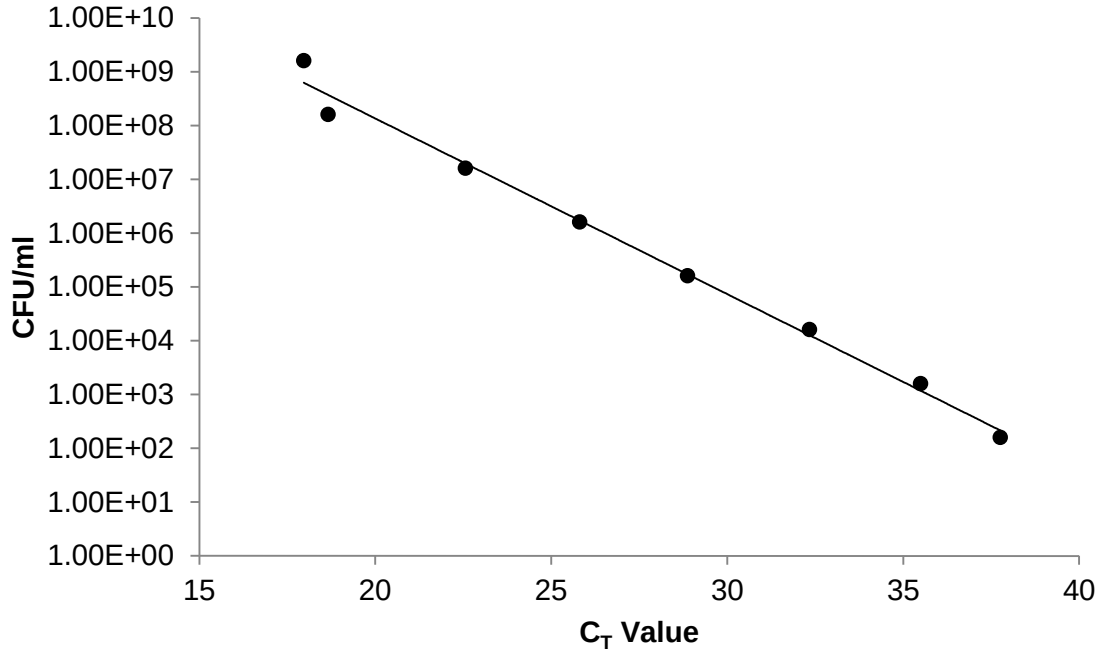


(B) *S. pneumoniae*

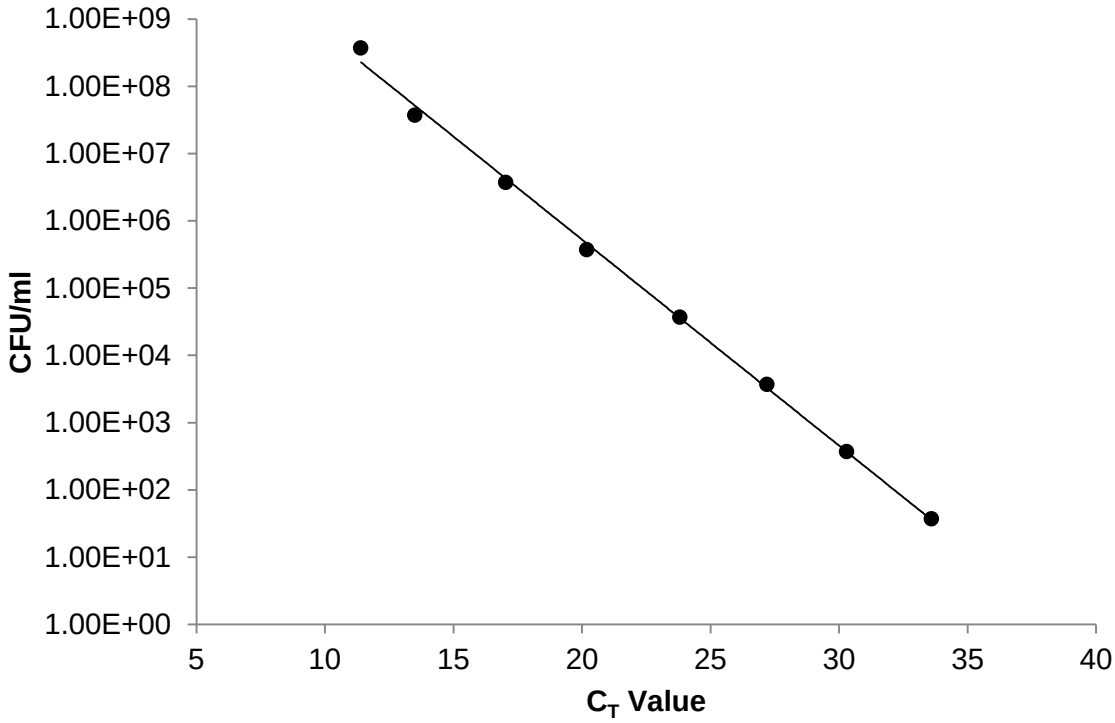


Appendix 3 continued. Standard Curves for Quantification of Bacterial Species

(C) *H. influenzae*

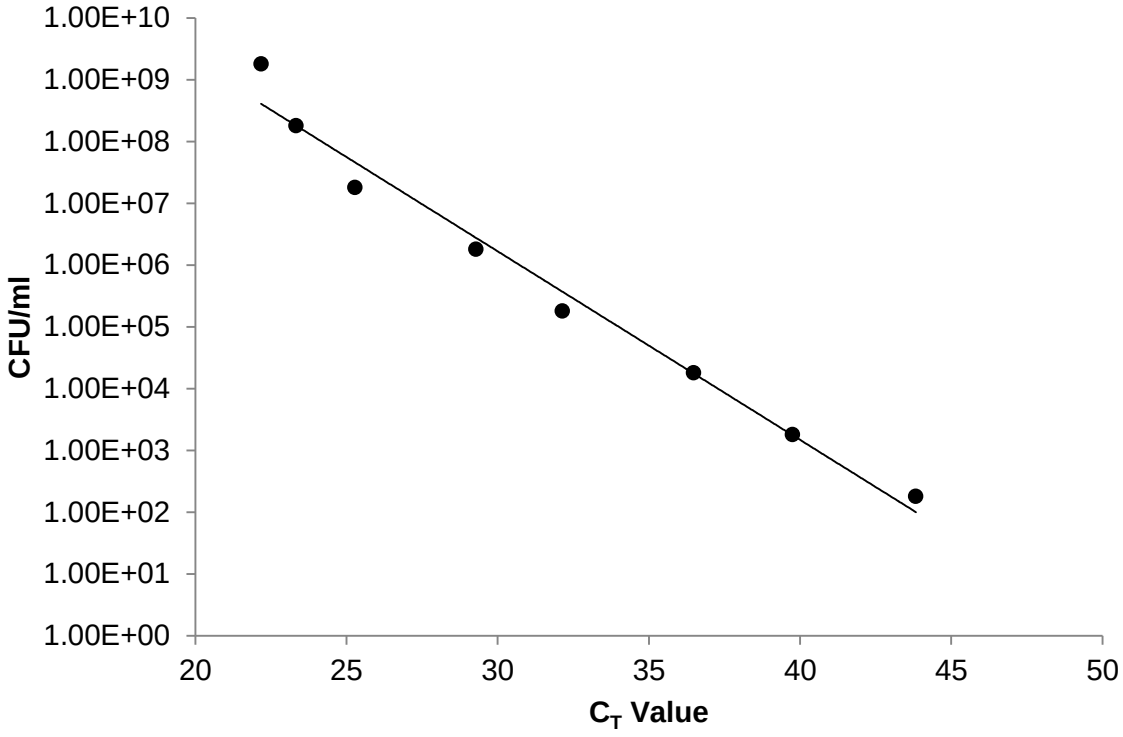


(D) *N. meningitidis*

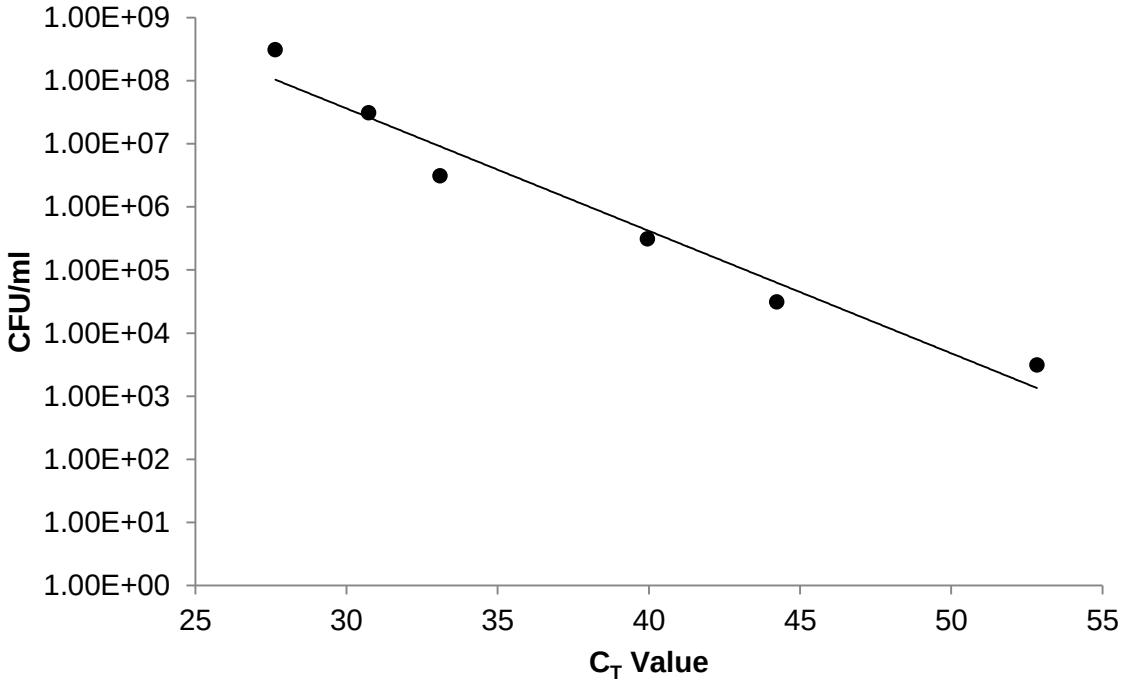


Appendix 3 continued. Standard Curves for Quantification of Bacterial Species

(E) *P. aeruginosa*



(F) *M. catarrhalis*



Appendix 4. Mastermix Components for PCR Serotyping of *S. pneumoniae* (50 reactions)

	Pool 1 Component	Volume(µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	6(6A/6B/6C)-F 6(6A/6B/6C)-R	1.875
	3-F 3-R	
	19A-F 19A-R	
	22F/22A-F 22F/22A-R	3.125
	16F-F 16F-R	2.500
Other	PCR water	213.8
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

	Pool 2 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	8-F 8-R	1.250
	33F/33A/37-F 33F/33A/37-R	1.875
	15A/15F-F 15A/15F-R	
	7F/7A-F 7F/7A-R	2.500
	23A-F 23A-R	3.125
Other	PCR water	215.0
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

	Pool 3 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	19F-F 19F-R	3.125
	12F/12A/44/46-F 12F/12A/44/46-R	
	11A/11D-F 11A/11D-R	1.875
	38/25F-F 38/25F-R	
	35B-F 35B-R	3.125
Other	PCR water	210.0
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

Appendix 4 continued. Mastermix Components for PCR serotyping of *S. pneumoniae* (50 reactions)

	Pool 4 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	24A/24B/24F-F 24A/24B/24F-R	
	7C/7B/40-F 7C/7B/40-R	1.875
	4-F 4-R	
	18A/18B/18C/18F-F 18A/18B/18C/18F-R	
	9V/9A-F 9V/9A-R	3.125
Other	PCR water	217.5
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

	Pool 5 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	14-F 14-R	1.875
	1-F 1-R	
	23F-F 23F-R	3.125
	15B/15C-F 15B/15C-R	1.875
	10A-F 10A-R	3.125
Other	PCR water	212.5
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

	Pool 6 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	39-F 39-R	1.250
	10F/10C/33C-F 10F/10C/33C-R	1.875
	5-F 5-R	
	35F/47F-F 35F/47F-R	
	17F-F 17F-R	3.125
Other	PCR water	216.3
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

Appendix 4 continued. Mastermix Components for PCR serotyping of *S. pneumoniae* (50 reactions)

	Pool 7 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	23B-F 23B-R	1.250
	35A/35C/42-F 35A/35C/42-R	1.875
	34-F 34-R	
	9N/9L-F 9N/9L-R	3.125
	31-F 31-R	
Other	PCR water	213.8
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

	Pool 8 Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	21-F 21-R	1.250
	2-F 2-R	1.875
	20-F 20-R	
	13-F 13-R	2.500
Other	PCR water	217.5
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

	Pool 6C Component	Volume (µl)
Primers (100mM)	CPSA-F CPSA-R	0.625
	6A/6B/6C-F 6A/6B/6C-R	1.875
	6C-F 6C-R	3.125
Other	PCR water	226.3
	50mM $MgCl_2$	25.0
	2x Red PCR Mix	312.5

Appendix 5. Primers for *in silico* Serotyping of *S. pneumoniae*

```
srt1 CTCTATAGAATGGAGTATATAAACTATGGTTA CCAAAGAAAATACTAACATTATCACAATATTGGC 400 200
srt2 TATCCCAGTTCAATATTTCTCCACTACACC ACACAAAATATAGGCAGAGAGAGACTACT 400 200
srt3 ATGGTGTGATTTCTCCTAGATTGGAAAGTAG CTTCTCCAATTGCTTACCAAGTGCAATAACG 500 300
srt4 CTGTTACTTGTTCTGGACTCTCGATAATTGG GCCCACTCCTGTTAAAATCCTACCCGCATTG 600 300
srt5 ATACCTACACAACTTCTGATTATGCCTTTGTG GCTCGATAAACATAATCAATATTTGAAAAAGTATG 500 300
srg6A_6B_6C_6D AATTTGTATTTTATTCATGCCTATATCTGG TTAGCGGAGATAATTTAAAATGATGACTA 350 200
srg6C_6D CATTTTAGTGAAGTTGGCGGTGGAGTT  AGCTTCGAAGCCCATACTCTTCAATTA 900 600
srts7C_7B_40 CTATCTCAGTCATCTATTGTTAAAGTTTACGACGGGA GAACATAGATGTTGAGACATCTTTTGTAATTTC 400 200
srg7F_7A TCCAAACTATTACAGTGGGAATTACGG ATAGGAATTGAGATTGCCAAAGCGAC 700 500
srt8 GAAGAAACGAAACTGTCAGAGCATTTACAT  CTATAGATACTAGTAGAGCTGTTCTAGTCT 300 100
srg9N_9L GAACTGAATAAGTCAGATTTAATCAGC  ACCAAGATCTGACGGGCTAATCAAT 600 400
srg9V_9A GGGTTCAAAGTCAGACAGTGAATCTTAA CCATGAATGAAATCAACATTGTCAGTAGC 900 700
srt10A GGTGTAGATTTACCATTAGTGTCGGCAGAC GAATTTCTTCTTTAAGATTCGGATATTTCTC  700 500
srts10F_10C_33C GGAGTTTATCGGTAGTGCTCATTTTAGCA CTAACAAATTCGCAACACGAGGCAACA 350 200
srg11A_11D GGACATGTTCAGGTGATTTCCCAATATAGTG GATTATGAGTGTAATTTATTCCAACTTCTCCC  600 400
srts12F_12A_44_46 GCAACAAACGGCGTGAAAGTAGTTG CAAGATGAATATCACTACCAATAACAAAAC 500 300
srt13 TACTAAGGTAATCTCTGGAAATCGAAAGG  CTCATGCATTTTATTAACCGCTTTTTGTTC 800 600
srt14 GAAATGTTACTTGGCGCAGGTGTCAGAATT GCCAATACTTCTTAGTCTCTCAGATGAAT  300 100
srg15A_15F ATTAGTACAGCTGCTGGAATATCTCTTC GATCTAGTGAACGTACTATTCCAAAC  600 400
srg15B_15C TTGGAATTTTTTAATTAGTGGCTTACCTA CATCCGCTTATTAATTGAAGTAATCTGAACC 700 400
srt16F GAATTTTTCAGGCGTGGGTGTTAAAAG CAGCATATAGCACCGCTAAGCAAATA 800 600
srt17F TTCGTGATGATAATTCCAATGATCAAACAAGAG GATGTAACAAATTTGTAGCGACTAAGGTCTGC 800 600
srg18A_18B_18C_18F CTTAATAGCTCTCATTATTCTTTTTTTAAGCC TTATCTGTAAACCATATCAGCATCTGAAAC 700 500
srt19A GAGAGATTCATAATCTTGCACTTAGCCA CATAATAGCTACAAATGACTCATCGCC 700 500
srt19F GTTAAGATTGCTGATCGATTAATTGATATCC GTAATATGTCTTTAGGGCGTTTATGGCGATAG 400 250
srt20 GAGCAAGAGTTTTTCACCTGACAGCGAGAAG CTAAATTCCTGTAATTTAGCTAAAACTCTTATC 600 450
srt21 CTATGGTTATTTCAACTCAATCGTCACC  GGCAAACTCAGACATAGTATAGCATAG 300 100
srg22F_22A GAGTATAGCCAGATTATGGCAGTTTTATTGTC CTCCAGCACTTGCGCTGGAAACAACAGACAAC 800 600
srt23A TATTCTAGCAAGTGACGAAGATGCG CCAACATGCTTAAAAACGCTGCTTTAC 850 650
srt23B CCACAATTAGCGCTATATTCATTCAATCG GTCCACGCTGAATAAAATGAAGCTCCG 300 100
srt23F GTAACAGTTGCTGTAGAGGGAATTGGCTTTTC CACAACACCTAACACTCGATGGCTATATGATTC 500 300
srg24A_24B_24F GCTCCCTGCTATTGTAATCTTTAAAGAG GTGTCTTTTATTGACTTTATCATAGGTCGG 200 50
srt31 GGAAGTTTTCAAGGATATGATAGTGGTGGTGC CCGAATAATATATTCAATATATTCCTACTC 800 600
srts33F_33A_37 GAAGGCAATCAATGTGATTGTGTCGCG CTTCAAAATGAAGATTATAGTACCCTTCTAC 500 300
srt34 GCTTTTGTAAGAGGAGATTATTTTCACCCAAC CAATCCGACTAAGTCTTCAGTAAAAAACTTTAC 500 300
srts35A_35C_42 ATTACGACTCCTTATGTGACGCGCATA CCAATCCCAAGATATATGCAACTAGGTT 400 200
srt35B GATAAGTCTGTTGTGGAGACTTAAAAAGAATG CTTTCCAGATAATTACAGGTATTCCTGAAGCAAG 800 600
srt35F_47F GAACATAGTCGCTATTGTATTTTATTTAAAGCAA GACTAGGAGCATTATTCCTAGAGCGAGTAAACC 700 500
srts38_25F_25A CGTTCTTTTATCTCACTGTATAGTATCTTTATG ATGTTTGAATTAAAGCTAACGTAACAATCC 700 500
srt39 TCATTGTATTAACCCTATGCTTTATTGGTG GAGTATCTCCATTGTATTGAAATCTACCAA 200 50
cpsA GCAGTACAGCAGTTTGTTGGACTGACC GAATATTTTCATTATCAGTCCCAGTC 300 100
cps15bM ATGGTTACTAAAGATAAAGGATTAACA TTAACTAAACCTTTTTAATAGTCTGTTT 1200 800
wzy33f  ATGCATGTTAGATTAGATGGTTTGCTGGACTA TTAATTAGTAATGTACTGTTTCTCATGAC 1400 1100
wzy33d  ATGTCAGTTAGAAAACTATCTTTAATAGATATA TTACGTTCCAGTTTTATCTAATGTTACAA 1300 1100
wcwV22  ATGAAAAAAGTATTAATTGTCCGAA TTACCTTTCATCAAATGTCTCAATTATTTT 1300 1000
wzy35F_47F GTGATGAAAGTAGTAATAAAAAGAAAGTCTTT CTATATAATTTCTTCAAAAGAATTTTCCC 1400 1000
wcjE11  ATGAAAATATGTTTAGTTGGGTCTAGTGGA        TTAGAAAATCCCTCGAGATTAATTGCTTTT 500 300
wcwR11  ATGAAAAATGATAGGAAAGCTCAACTAATA        TTACTTCTCTAATGCTATATTCCTCGCTGG 800 100
wcwC11  ATGAAAGGGAAGAGTTTTCTTTTGAAGATG        TCATGTTTGTGTCTCTCCATTCAGCCAATT 800 600
gct11 ATGAAAAGAGTAATTACTTATGGAACCTTT  CTAATCTGCTAAATCTTCTTTAATTTTAGT        500 200
WciP6 ATGGGAAAGTCAGTTGCAATTTTAATG TTATAAGAGTTCTTTCATAATATATTTTTT 1200 900
WciNbeta TTTTATAATGTTCTAAAGGGAGATATGAGT GCTTTTTTAGCAGGCGACATAGTTTTTCTT 2200 1200
38wcyD GTGAAAAAAATAGTATTTTTGTGTCCATACTTT TCATTCCGACCTCCAAATTTTATTCCATAATCG 1100 800
22FwcwC ATGAAAGGATTACAAATGATAATCAAA TTTAATACTACTTTCTTTCCATTAAAAT 800 500
```

Appendix 6. Primers for *in silico* Serotyping of *H. influenzae*

```
srtA CTACTCATTGCAGCATTTGC GAATATGACCTGATCTTCTG 0 5000
srtB GCGAAAGTGAACTCTTATCTCTC GCTTACGCTTCTATCTCGGTGAA 0 5000
srtC TCTGTGTAGATGATGGTTCA CAGAGGCAAGCTATTAGTGA 0 5000
srtD TGATGACCGATACAACCTGT TCCACTCTTCAAACCATTCT 0 5000
srtE TTTGGTAACGAATGTAGTGGTAG ATAGCTTTACTGTATAAGTCTAG  0 5000
srtF GCTACTATCAAGTCCAAATC CGCAATTATGGAAGAAAGCT 0 5000
bexB GGTGATTAACGCGTTGCTTATGCG TTGTGCCTGTGCTGGAAGGTTATG 0 5000
ompP ATAACAACGAAGGGACTAACG ACCTACACCCACTGATTTTTC 0 5000
```

Appendix 7. PlotTemp.R Script used for Nestedness Analysis

```
# Calculate and plot incidence temperature, using packages: vegan and permute
 # Function plot_temp("filename") plots incidence temperature from a filename in
    quotes
 # NOTE: Top-left cell must be empty! I.e. No header for samples.
 # NOTE: Need to load package 'vegan'
 # Call and run the function with the commands:
 # source("PlotTemp.R")
 # plot_temp("filename")
   plot_temp <- function(fichero)
   {
   if (class(fichero) != "character") stop("request filename in quotes")
   diatoms <- read.table(fichero, header=T)
   library(vegan)
   library(permute)
   tempspecies <- nestedtemp(species)
   # plot of the nestedness temperature matrix with no labels or title,
   # square plot (use "asp = 1" for squares inside)
   par(col = "black", lwd = 2, pty = "s", mar=(c(2,4,8,2)+0.1))
   # names=TRUE writes the names of the species and the sample numbers of the
   plot(tempspecies, kind="incid", ann = FALSE, names=T)
   # Bring legends close to axes
   mtext(side = 3, text = "Species by incidence", line = 6, cex = 1.25)
   mtext(side = 2, text = "Samples by richness", line = 3, cex = 1.25)
   # gives temperature and matrix fill of the analysed matrix
   source("ResetPar.R")
   tempspecies
   }
```

Appendix 8. Antigen alleles present in *S. pneumoniae* isolates

Allele	Count (n)	% of isolates
cbpD_152	34	100
cbpG_335	28	82.35294
cppA_400	34	100
cps2A_127	1	2.941176
cps2B_408	6	17.64706
cps2C_429	11	32.35294
cps2D_433	11	32.35294
cps2E_147	11	32.35294
cps2L_351	20	58.82353
cps2P_443	8	23.52941
cps2T_202	2	5.882353
eno_157	34	100
eno_159	34	100
htrA/degP_201	34	100
hysA_40	33	97.05882
iga_10	6	17.64706
iga_12	6	17.64706
iga_5	13	38.23529
iga_8	11	32.35294
lmb_316	34	100
lytA_302	34	100
lytA_303	34	100
lytB_69	34	100
lytB_77	34	100
lytC_119	33	97.05882
nanA_42	26	76.47059
nanA_51	11	32.35294
pavA_98	34	100
pce/cbpE_86	34	100
piaA_265	33	97.05882
piaA_298	34	100
plr/gapA_247	34	100
plr/gapA_263	34	100
plr/gapA_272	34	100
plr/gapA_274	34	100
ply_136	34	100
psaA_312	34	100
pspA_64	1	2.941176
pspA_72	3	8.823529
pspA_88	1	2.941176
pspA_90	4	11.76471
pspC/cbpA_70	27	79.41176

Appendix 8 continued. Antigen alleles present in *S. pneumoniae* isolates

Allele	Count (n)	% of isolates
pspC/cbpA_73	27	79.41176
pspC/cbpA_74	20	58.82353
pspC/cbpA_76	3	8.823529
rfbA_348	15	44.11765
rfbB_257	11	32.35294
rfbD_359	8	23.52941
rmlA_349	19	55.88235
rmlA_352	5	14.70588
rmlA_353	5	14.70588
rmlB_257	9	26.47059
rmlB_261	4	11.76471
rmlC_445	8	23.52941
rmlD_357	14	41.17647
rrgA_54	2	5.882353
rrgC_205	2	5.882353
slrA_375	34	100
SMU.322c_331	15	44.11765
SP_0346_127	16	47.05882
SP_0347_408	18	52.94118
SP_0348_430	6	17.64706
SP_0348_431	6	17.64706
SP_0349_435	2	5.882353
SPH_0455_127	11	32.35294
SPH_0456_408	9	26.47059
SPH_0457_429	9	26.47059
SPH_0458_433	11	32.35294
SPH_0459_147	9	26.47059
SPH_0460_398	4	11.76471
SPH_0461_368	4	11.76471
SPH_0462_346	4	11.76471
SPH_0463_154	4	11.76471
SPH_0464_134	4	11.76471
SPH_0465_243	4	11.76471
srtA_397	33	97.05882
srtB_338	2	5.882353
srtC_339	2	5.882353
srtD_360	2	5.882353
STER_1224_353	2	5.882353
STER_1444_347	8	23.52941
tig/ropA_163	34	100
tig/ropA_165	34	100
tig/ropA_166	33	97.05882
tig/ropA_167	3	8.823529
wchA_146	9	26.47059

Appendix 8 continued. Antigen alleles present in *S. pneumoniae* isolates

Allele	Count (n)	% of isolates
wchJ_465	3	8.823529
wzd_428	11	32.35294
wze_435	9	26.47059
wzg_127	5	14.70588
wzh_407	19	55.88235
zmpB_11	3	8.823529
zmpB_14	1	2.941176
zmpB_6	7	20.58824
zmpC_16	6	17.64706

Appendix 9. Antigen alleles present in *H. influenzae* isolates

Allele	Count (n)	% of isolates
CGSHiEE_06765_54	3	9.375
CGSHiEE_07700_95	3	9.375
CGSHiEE_08890_19	1	3.125
CGSHiGG_06430_10	4	12.5
CGSHiGG_08950_54	16	50
comE/pilQ_75	32	100
galE_141	32	100
galU_185	32	100
gmhA/lpcA_251	18	56.25
gmhA/lpcA_252	18	56.25
gmhA/lpcA_253	32	100
hemH_153	32	100
hemM_243	32	100
hemN_110	32	100
hemR_27	32	100
hemX_96	32	100
hemY_82	32	100
hgbA_20	1	3.125
hgpB_16	32	100
hgpB_17	32	100
hgpC_11	28	87.5
hgpC_12	32	100
hgpC_9	31	96.875
hgpD_13	4	12.5
HI0867_95	20	62.5
HI1064_54	6	18.75
hia/hsf_246	11	34.375
hifA_247	3	9.375
hifB_241	10	31.25
hifC_26	10	31.25
hifD_240	7	21.875
hifE_79	2	6.25
hitA_148	32	100
hitB_59	32	100
hitC_120	32	100
hmw1A_4	1	3.125
hmw1B_52	4	12.5
hmw1C_33	2	6.25
hmw2A_3	1	3.125
hmw2B_52	2	6.25
hmw2C_33	4	12.5
htrB_167	32	100
hxuA_22	32	100
hxuA_24	32	100

Appendix 9 continued. Antigen alleles present in *H. influenzae* isolates

Allele	Count (n)	% of isolates
hxuA_25	32	100
hxuB_41	32	100
hxuC_28	32	100
hxuC_29	32	100
iga1_1	6	18.75
iga1_2	6	18.75
kdkA_26	32	100
kdkA_226	32	100
kdsA_193	32	100
kdsA_194	12	37.5
kdsB_216	32	100
kdtA_85	32	100
kfiC_222	23	71.875
kpsF_144	32	100
lex2A_265	27	84.375
lex2B_223	30	93.75
lgtA_154	31	96.875
lgtC_151	30	93.75
lgtF_217	32	100
lic2A_164	32	100
lic2A_178	32	100
lic2A_199	18	56.25
lic3A_146	32	100
licA_115	32	100
licB_189	32	100
licC_231	32	100
licD_207	24	75
lpsA_196	16	50
lpt6_47	32	100
lpxA_210	32	100
lpxB_107	32	100
lpxC_174	32	100
lpxC_175	31	96.875
lpxD_136	32	100
lpxH_227	32	100
lpxK_147	32	100
lsgA_100	32	100
lsgB_176	32	100
lsgC_125	32	100
lsgD_212	32	100
lsgE_186	32	100
lsgF_203	32	100
manA_268	24	75
manB_48	16	50

Appendix 9 continued. Antigen alleles present in *H. influenzae* isolates

Allele	Count (n)	% of isolates
mrsA/glmM_76	32	100
msbA_38	32	100
msbA_39	32	100
msbB_159	32	100
neuA_236	32	100
NTHI1224_54	7	21.875
oapA_81	30	93.75
ompP2_111	28	87.5
ompP2_116	32	100
ompP2_118	18	56.25
ompP2_126	32	100
ompP5_117	29	90.625
opsX/rfaC_129	32	100
orfE_187	24	75
orfM_250	32	100
orfO_270	2	6.25
pgi_42	32	100
pilA_262	30	93.75
pilA_263	28	87.5
pilB_67	32	100
pilC_93	32	100
pilD_234	32	100
rfaD_170	2	6.25
rfaD_171	32	100
rfaE_62	32	100
rfaE_63	17	53.125
rfaF_132	32	100
rffG_130	9	28.125
rffG_145	24	75
siaA_173	2	6.25
siaA_190	16	50
tbpA_23	32	100
tbp_34	11	34.375
tbp_35	10	31.25
tbp_36	14	43.75
tbp_60	5	15.625
waaQ_119	32	100
wbaP/rfbP_64	24	75
wbaP/rfbP_65	24	75
wecA_121	32	100
yhbX_58	32	100
yhxB/manB_48	16	50
hpd	27	84.4

Appendix 10. Chi-squared Test Results showing Deviations from the Random (Poisson) Distribution in real-time PCR samples

		Nose			NP		
		X^2	*p*	df	X^2	*p*	df
Season	**Summer**	21.753	<0.001	5	NA	NA	NA
	Winter	6.614	0.358	6	NA	NA	NA
Age (years)	**0-4**	20.417	**0.002**	6	9.856	0.079	5
	5-17	8.354	0.138	5	3.055	0.383	3
	18-64	5.323	0.256	4	6.563	0.087	3
	≥65	0.153	0.985	3	10.030	**0.018**	3
RTI	**Yes**	7.461	0.189	5	7.927	0.160	5
	No	15.437	**0.009**	6	15.302	**0.004**	4

X^2 = Chi-squared, df = degrees of freedom, Deviations with X^2 *P*-value < 0.05 are highlighted in bold.

RTI = respiratory tract infection. NP = nasopharyngeal swab. NA = not applicable.

References

ABDILLAHI, H. & POOLMAN, J. T. 1988. Typing of group-B Neisseria meningitidis with monoclonal antibodies in the whole-cell ELISA. *J Med Microbiol,* 26**,** 177-80.

ADAM, H. J., RICHARDSON, S. E., JAMIESON, F. B., RAWTE, P., LOW, D. E. & FISMAN, D. N. 2010. Changing epidemiology of invasive Haemophilus influenzae in Ontario, Canada: Evidence for herd effects and strain replacement due to Hib vaccination. *Vaccine,* 28**,** 4073-8.

ADEGBOLA, R. A., MULHOLLAND, E. K., SECKA, O., JAFFAR, S. & GREENWOOD, B. M. 1998. Vaccination with a Haemophilus influenzae type b conjugate vaccine reduces oropharyngeal carriage of H-influenzae type b among Gambian children. *J Infect Dis,* 177**,** 1758-61.

AGRAWAL, A. & MURPHY, T. F. 2011. Haemophilus influenzae Infections in the H. influenzae Type b Conjugate Vaccine Era. *J Clin Microbiol,* 49**,** 3728-32.

AGUILERA, J. F., PERROCHEAU, A., MEFFRE, C. & HAHNE, S. 2002. Outbreak of serogroup W135 meningococcal disease after the Hajj pilgrimage, Europe, 2000. *Emerg Infect Dis,* 8**,** 761-7.

AKMATOV, M. K., GATZEMEIER, A., SCHUGHART, K. & PESSLER, F. 2012. Equivalence of self- and staff-collected nasal swabs for the detection of viral respiratory pathogens. *PLoS One,* 7**,** e48508.

AKMATOV, M. K., KREBS, S., PREUSSE, M., GATZEMEIER, A., FRISCHMANN, U., SCHUGHART, K. & PESSLER, F. 2011. E-mail-based symptomatic surveillance combined with self-collection of nasal swabs: a new tool for acute respiratory infection epidemiology. *Int J Infect Dis,* 15**,** e799-803.

AKMATOV, M. K. & PESSLER, F. 2011. Self-collected nasal swabs to detect infection and colonization: a useful tool for population-based epidemiological studies? *Int J Infect Dis,* 15**,** e589-93.

ALA'ALDEEN, D. A. A., NEAL, K. R., AIT-TAHAR, K., NGUYEN-VAN-TAM, J. S., ENGLISH, A., FALLA, T. J., HAWKEY, P. M. & SLACK, R. C. B. 2000. Dynamics of meningococcal long-term carriage among university students and their implications for mass vaccination. *J Clin Microbiol,* 38**,** 2311-6.

ALLEGRUCCI, M., HU, F. Z., SHEN, K., HAYES, J., EHRLICH, G. D., POST, J. C. & SAUER, K. 2006. Phenotypic characterization of Streptococcus pneumoniae biofilm development. *J Bacteriol,* 188**,** 2325-35.

ALLEGRUCCI, M. & SAUER, K. 2008. Formation of Streptococcus pneumoniae non-phase-variable colony variants is due to increased mutation frequency present under biofilm growth conditions. *J Bacteriol,* 190**,** 6330-9.

ALLEN, E. K., KOEPPEL, A. F., HENDLEY, J. O., TURNER, S. D., WINTHER, B. & SALE, M. M. 2014. Characterization of the nasopharyngeal microbiota in health and during rhinovirus challenge. *Microbiome,* 2**,** 22.

ALLEN, E. K., PITKARANTA, A., MAKI, M., HENDLEY, J. O., LAAKSO, S., SALEI, M. M. & WINTHER, B. 2013. Bacteria in the nose of young adults during wellness and rhinovirus

colds: detection by culture and microarray methods in 100 nasal lavage specimens. *Int Forum Allergy Rhinol,* 3**,** 731-9.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ANDERSSON, D. I. 2006. The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol,* 9**,** 461-5.

ANDREN, H. 1994. Can One Use Nested Subset Pattern to Reject the Random Sample Hypothesis - Examples from Boreal Bird Communities. *Oikos,* 70**,** 489-91.

ARMBRUSTER, C. E., HONG, W., PANG, B., WEIMER, K. E., JUNEAU, R. A., TURNER, J. & SWORDS, W. E. 2010. Indirect pathogenicity of Haemophilus influenzae and Moraxella catarrhalis in polymicrobial otitis media occurs via interspecies quorum signaling. *MBio,* 1**,** e00102-10.

ARTENSTEIN, M. S., GOLD, R., ZIMMERLY, J. G., WYLE, F. A., SCHNEIDER, H. & HARKINS, C. 1970. Prevention of meningococcal disease by group C polysaccharide vaccine. *N Engl J Med,* 282**,** 417-20.

ASCH, D. A., JEDRZIEWSKI, M. K. & CHRISTAKIS, N. A. 1997. Response rates to mail surveys published in medical journals. *J Clin Epidemiol,* 50**,** 1129-36.

ASHWORTH, M., LATINOVIC, R., CHARLTON, J., COX, K., ROWLANDS, G. & GULLIFORD, M. 2004. Why has antibiotic prescribing for respiratory illness declined in primary care? A longitudinal study using the General Practice Research Database. *J Public Health (Oxf.),* 26**,** 268-74.

ATMAR, W. & PATTERSON, B. D. 1993. The Measure of Order and Disorder in the Distribution of Species in Fragmented Habitat. *Oecologia,* 96**,** 373-82.

AUSTRIAN, R. 1976. The quellung reaction, a neglected microbiologic technique. *Mt Sinai J Med,* 43**,** 699-709.

AZIZ, R. K., BARTELS, D., BEST, A. A., DEJONGH, M., DISZ, T., EDWARDS, R. A., FORMSMA, K., GERDES, S., GLASS, E. M., KUBAL, M., MEYER, F., OLSEN, G. J., OLSON, R., OSTERMAN, A. L., OVERBEEK, R. A., MCNEIL, L. K., PAARMANN, D., PACZIAN, T., PARRELLO, B., PUSCH, G. D., REICH, C., STEVENS, R., VASSIEVA, O., VONSTEIN, V., WILKE, A. & ZAGNITKO, O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75.

BAAS RUBARTH, L. 2010. Sepsis, Pneumonia, and Meningitis: What Is the Difference? *Newborn Infant Nurs Rev,* 10**,** 177-81.

BAKALETZ, L. O. 2007. Bacterial Biofilms in Otitis Media. *Pediatr Infect Dis J,* 26**,** S17-9.

BAKER, M. G., MARTIN, D. R., KIEFT, C. E. & LENNON, D. 2001. A 10-year serogroup B meningococcal disease epidemic in New Zealand: descriptive epidemiology, 1991-2000. *J Paediatr Child Health,* 37**,** S13-9.

BAKHSHAEE, M., NADERI, H. R., GHAZVINI, K., SOTOUDEH, K., AMALI, A. & ASHTIANI, S. J. 2012. Passive smoking and nasopharyngeal colonization by Streptococcus pneumoniae, Haemophilus influenzae, and Moraxella catarrhalis in daycare children. *Eur Arch Otorhinolaryngol,* 269**,** 1127-32.

BALMER, P., BORROW, R. & MILLER, E. 2002. Impact of the meningococcal C conjugate vaccine in the UK. *J Med Microbiol,* 51**,** 717-22.

BARBOUR, M. L., MAYON-WHITE, R. T., COLES, C., CROOK, D. W. M. & MOXON, E. R. 1995. The Impact of Conjugate Vaccine on Carriage of Haemophilus influenzae Type b. *J Infect Dis,* 171**,** 93-6.

BARENKAMP, S. J. 1996. Immunization with high-molecular-weight adhesion proteins of nontypeable Haemophilus influenzae modifies experimental otitis media in chinchillas. *Infect Immun,* 64**,** 1246-51.

BARROS, A. J. & HIRAKATA, V. N. 2003. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol,* 3**,** 21.

BASCOMPTE, J., JORDANO, P., MELIAN, C. J. & OLESEN, J. M. 2003. The nested assembly of plant-animal mutualistic networks. *Proc Natl Acad Sci U S A,* 100**,** 9383-7.

BELLOMO, R., WARRILLOW, S. J. & READE, M. C. 2009. Why we should be wary of single-center trials. *Crit Care Med,* 37**,** 3114-9.

BENDER, J. M., COX, C. M., MOTTICE, S., SHE, R. C., KORGENSKI, K., DALY, J. A. & PAVIA, A. T. 2010. Invasive Haemophilus influenzae disease in Utah children: an 11-year population-based study in the era of conjugate vaccine. *Clin Infect Dis,* 50**,** e41-6.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., CHEETHAM, R. K., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X. H., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M. J., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X. L., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M. D., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CATENAZZI, M. C. E., CHANG, S., COOLEY, R. N., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FAJARDO, K. V. F., FUREY, W. S., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

BETTINGER, J. A., SCHEIFELE, D. W., LE SAUX, N., HALPERIN, S. A., VAUDRY, W., TSANG, R. & CANADIAN IMMUNIZATION MONITORING PROGRAM, A. 2009. The impact of childhood meningococcal serogroup C conjugate vaccine programs in Canada. *Pediatr Infect Dis J,* 28**,** 220-4.

BICER, S., GIRAY, T., COL, D., ERDAG, G. C., VITRINEL, A., GUROL, Y., CELIK, G., KASPAR, C. & KUCUK, O. 2013. Virological and clinical characterizations of respiratory infections in hospitalized children. *Ital J Paediatr,* 39**,** 22.

BINKS, M. J., TEMPLE, B., KIRKHAM, L. A., WIERTSEMA, S. P., DUNNE, E. M., RICHMOND, P. C., MARSH, R. L., LEACH, A. J. & SMITH-VAUGHAN, H. C. 2012. Molecular surveillance of true nontypeable Haemophilus influenzae: an evaluation of PCR screening assays. *PLoS One,* 7**,** e34083.

BISNO, A. L. 2001. Acute Pharyngitis. *New Engl J Med,* 344**,** 205-11.

BOGAERT, D., DE GROOT, R. & HERMANS, P. W. M. 2004a. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis,* 4**,** 144-54.

BOGAERT, D., HERMANS, P. W., ADRIAN, P. V., RUMKE, H. C. & DE GROOT, R. 2004b. Pneumococcal vaccines: an update on current strategies. *Vaccine,* 22**,** 2209-20.

BOGAERT, D., HERMANS, P. W., BOELENS, H., SLUIJTER, M., LUIJENDIJK, A., RUMKE, H. C., KOPPEN, S., VAN BELKUM, A., DE GROOT, R. & VERBRUGH, H. A. 2005. Epidemiology of nasopharyngeal carriage of Neisseria meningitidis in healthy Dutch children. *Clin Infect Dis,* 40**,** 899-902.

BOGAERT, D., KEIJSER, B., HUSE, S., ROSSEN, J., VEENHOVEN, R., VAN GILS, E., BRUIN, J., MONTIJN, R., BONTEN, M. & SANDERS, E. 2011. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One,* 6**,** e17035.

BOGAERT, D., VAN BELKUM, A., SLUIJTER, M., LUIJENDIJK, A., DE GROOT, R., RUMKE, H. C., VERBRUGH, H. A. & HERMANS, P. W. 2004c. Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children. *Lancet,* 363**,** 1871-2.

BOHANNAN, B. J. & HUGHES, J. 2003. New approaches to analyzing microbial biodiversity data. *Curr Opin Microbiol,* 6**,** 282-7.

BOLES, B. R., THOENDEL, M. & SINGH, P. K. 2004. Self-generated diversity produces "insurance effects" in biofilm communities. *Proc Natl Acad Sci U S A,* 101**,** 16630-5.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BONE, R. C., BALK, R. A., CERRA, F. B., DELINGER, R. P., FEIN, A. M., KNAUS, W. A., SCHEIN, R. M. H. & SIBBALD, W. J. 1992. Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. *Chest,* 101**,** 1644-55.

BOSCH, A. A., BIESBROEK, G., TRZCINSKI, K., SANDERS, E. A. & BOGAERT, D. 2013. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog,* 9**,** e1003057.

BRADNAM, K. R., FASS, J. N., ALEXANDROV, A., BARANAY, P., BECHNER, M., BIROL, I., BOISVERT, S., CHAPMAN, J. A., CHAPUIS, G., CHIKHI, R., CHITSAZ, H., CHOU, W. C., CORBEIL, J., DEL FABBRO, C., DOCKING, T. R., DURBIN, R., EARL, D., EMRICH, S., FEDOTOV, P., FONSECA, N. A., GANAPATHY, G., GIBBS, R. A., GNERRE, S., GODZARIDIS, E., GOLDSTEIN, S., HAIMEL, M., HALL, G., HAUSSLER, D., HIATT, J. B., HO, I. Y., HOWARD, J., HUNT, M., JACKMAN, S. D., JAFFE, D. B., JARVIS, E. D., JIANG, H., KAZAKOV, S., KERSEY, P. J., KITZMAN, J. O., KNIGHT, J. R., KOREN, S., LAM, T. W., LAVENIER, D., LAVIOLETTE, F., LI, Y., LI, Z., LIU, B., LIU, Y., LUO, R., MACCALLUM, I., MACMANES, M. D., MAILLET, N., MELNIKOV, S., NAQUIN, D., NING, Z., OTTO, T. D., PATEN, B., PAULO, O. S., PHILLIPPY, A. M., PINA-MARTINS, F., PLACE, M., PRZYBYLSKI, D., QIN, X., QU, C., RIBEIRO, F. J.,

RICHARDS, S., ROKHSAR, D. S., RUBY, J. G., SCALABRIN, S., SCHATZ, M. C., SCHWARTZ, D. C., SERGUSHICHEV, A., SHARPE, T., SHAW, T. I., SHENDURE, J., SHI, Y., SIMPSON, J. T., SONG, H., TSAREV, F., VEZZI, F., VICEDOMINI, R., VIEIRA, B. M., WANG, J., WORLEY, K. C., YIN, S., YIU, S. M., YUAN, J., ZHANG, G., ZHANG, H., ZHOU, S. & KORF, I. F. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience,* 2**,** 10.

BRISSON-NOEL, A., AZNAR, C., CHUREAU, C., NGUYEN, S., PIERRE, C., BARTOLI, M., BONETE, R., PIALOUX, G., GICQUEL, B. & GARRIGUE, G. 1991. Diagnosis of tuberculosis by DNA amplification in clinical practice evaluation. *Lancet,* 338**,** 364-6.

BROOK, I. 2001. Sinusitis - overcoming bacterial resistance. *Int J Pediatr Otorhinolaryngol,* 58**,** 27-36.

BROWN, J. S., OGUNNIYI, A. D., WOODROW, M. C., HOLDEN, D. W. & PATON, J. C. 2001. Immunization with components of two iron uptake ABC transporters protects mice against systemic Streptococcus pneumoniae infection. *Infect Immun,* 69**,** 6702-6.

BROWN, S. D., FARRELL, D. J. & MORRISSEY, I. 2004. Prevalence and molecular analysis of macrolide and fluoroquinolone resistance among isolates of Streptococcus pneumoniae collected during the 2000-2001 PROTEKT US study. *J Clin Microbiol,* 42**,** 4980-7.

BRUEGGEMANN, A. B., GRIFFITHS, D. T., MEATS, E., PETO, T., CROOK, D. W. & SPRATT, B. G. 2003. Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis,* 187**,** 1424-32.

BRUEGGEMANN, A. B., PAI, R., CROOK, D. W. & BEALL, B. 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog,* 3**,** e168.

BRUNDAGE, J. F. 2006. Interactions between influenza and bacterial respiratory pathogens: implications for pandemic preparedness. *Lancet Infect Dis,* 6**,** 303-12.

BRUSSAARD, C. P. 2004. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol,* 70**,** 1506-13.

CALIENDO, A. M. 2011. Multiplex PCR and emerging technologies for the detection of respiratory pathogens. *Clin Infect Dis,* 52**,** S326-30.

CAMILLONI, B., NERI, M., LEPRI, E., BASILEO, M., SIGISMONDI, N., PUZELLI, S., DONATELLI, I. & IORIO, A. M. 2010. An influenza B outbreak during the 2007/2008 winter among appropriately immunized elderly people living in a nursing home. *Vaccine,* 28**,** 7536-41.

CAMPBELL, L. A., PEREZ MELGOSA, M., HAMILTON, D. J., KUO, C. C. & GRAYSTON, J. T. 1992. Detection of Chlamydia pneumoniae by polymerase chain reaction. *J Clin Microbiol,* 30**,** 434-9.

CAMPBELL, P., MCINTYRE, P., QUINN, H., HUESTON, L., GILBERT, G. L. & MCVERNON, J. 2012. Increased population prevalence of low pertussis toxin antibody levels in young children preceding a record pertussis epidemic in Australia. *PLoS One,* 7**,** e35874.

CAPEDING, M. R., NOHYNEK, H., SOMBRERO, L. T., PASCUAL, L. G., SUNICO, E. S., ESPARAR, G. A., ESKO, E., LEINONEN, M. & RUUTU, P. 1995. Evaluation of sampling sites for detection of upper respiratory tract carriage of Streptococcus pneumoniae and Haemophilus influenzae among healthy Filipino infants. *J Clin Microbiol,* 33**,** 3077-9.

CARLET, J., PULCINI, C. & PIDDOCK, L. J. 2014. Antibiotic resistance: A geo-political issue. *Clin Microbiol Infec,* 20**,** 949-53.

CARTWRIGHT, K. A., JONES, D. M., SMITH, A. J., STUART, J. M., KACZMARSKI, E. B. & PALMER, S. R. 1991. Influenza A and meningococcal disease. *Lancet,* 338**,** 554-7.

CARTWRIGHT, K. A., STUART, J. M., JONES, D. M. & NOAH, N. D. 1987. The Stonehouse survey: nasopharyngeal carriage of meningococci and Neisseria lactamica. *Epidemiol Infect,* 99**,** 591-601.

CASEY, J. R. & PICHICHERO, M. E. 2004. Changes in frequency and pathogens causing acute otitis media in 1995-2003. *Pediatr Infect Dis J,* 23**,** 824-8.

CATLIN, B. W. 1990. Branhamella-Catarrhalis - an Organism Gaining Respect as a Pathogen. *Clin Microbiol Rev,* 3**,** 293-320.

CAUGANT, D. A. & MAIDEN, M. C. 2009. Meningococcal carriage and disease--population biology and evolution. *Vaccine,* 27 Suppl 2**,** B64-70.

CAUGANT, D. A., TZANAKAKI, G. & KRIZ, P. 2007. Lessons from meningococcal carriage studies. *FEMS Microbiol Rev,* 31**,** 52-63.

CDC 2013. Progress in introduction of pneumococcal conjugate vaccine - worldwide, 2000-2012. *Morb Mortal Wkly Rep,* 62**,** 308-11.

CDC 2014a. First Confirmed Cases of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Infection in the United States, Updated Information on the Epidemiology of MERS-CoV Infection, and Guidance for the Public, Clinicians, and Public Health Authorities - May 2014. Atlanta, USA.

CDC. 2014b. *Types of Influenza Viruses* [Online]. Atlanta, USA. [Accessed 15 September 2014.

CHANCEY, S. T., BAI, X., KUMAR, N., DRABEK, E. F., DAUGHERTY, S. C., COLON, T., OTT, S., SENGAMALAY, N., SADZEWICZ, L., TALLON, L. J., FRASER, C. M., TETTELIN, H. & STEPHENS, D. S. 2015. Transcriptional attenuation controls macrolide inducible efflux and resistance in Streptococcus pneumoniae and in other Gram-positive bacteria containing mef/mel(msr(D)) elements. *PLoS One,* 10**,** e0116254.

CHARLSON, E. S., BITTINGER, K., HAAS, A. R., FITZGERALD, A. S., FRANK, I., YADAV, A., BUSHMAN, F. D. & COLLMAN, R. G. 2011. Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory Tract. *Am J Respir Crit Care Med,* 184**,** 957-63.

CHARLSON, E. S., CHEN, J., CUSTERS-ALLEN, R., BITTINGER, K., LI, H., SINHA, R., HWANG, J., BUSHMAN, F. D. & COLLMAN, R. G. 2010. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One,* 5**,** e15216.

CHASE, J. M. & LEIBOLD, M. A. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches,* Chicago, USA, University of Chicago Press.

CHEN, L., XIONG, Z., SUN, L., YANG, J. & JIN, Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res,* 40**,** D641-5.

CHEN, L., YANG, J., YU, J., YAO, Z., SUN, L., SHEN, Y. & JIN, Q. 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res,* 33**,** D325-8.

CHERRY, J. D., BRUNELL, P. A., GOLDEN, G. S. & KARZON, D. T. 1988. Report of the task force on pertussis and pertussis immunization - 1988. *Paediatrics,* 81**,** 939-84.

CHIEN, Y. W., VIDAL, J. E., GRIJALVA, C. G., BOZIO, C., EDWARDS, K. M., WILLIAMS, J. V., GRIFFIN, M. R., VERASTEGUI, H., HARTINGER, S. M., GIL, A. I., LANATA, C. F. & KLUGMAN, K. P. 2013. Density interactions among Streptococcus pneumoniae, Haemophilus influenzae and Staphylococcus aureus in the nasopharynx of young Peruvian children. *Pediatr Infect Dis J,* 32**,** 72-7.

CHOI, Y. H., JIT, M., GAY, N., ANDREWS, N., WAIGHT, P., MELEGARO, A., GEORGE, R. & MILLER, E. 2011. 7-Valent Pneumococcal Conjugate Vaccination in England and Wales: Is It Still Beneficial Despite High Levels of Serotype Replacement? *PLoS One,* 6**,** e26190.

CHONMAITREE, T., REVAI, K., GRADY, J. J., CLOS, A., PATEL, J. A., NAIR, S., FAN, J. & HENRICKSON, K. J. 2008. Viral upper respiratory tract infection and otitis media complication in young children. *Clin Infect Dis,* 46**,** 815-23.

CIMOLAI, N. 1988. Bacteriology of the Upper Respiratory Tract: What Is Important? *Can Fam Physician,* 34**,** 2155-9.

CLARKE, S. C. 2006. Control of pneumococcal disease in the United Kingdom--the start of a new era. *J Med Microbiol,* 55**,** 975-80.

CLAUS, H., MAIDEN, M. C., MAAG, R., FROSCH, M. & VOGEL, U. 2002. Many carried meningococci lack the genes required for capsule synthesis and transport. *Microbiology,* 148**,** 1813-9.

CLEGG, H. W., DALLAS, S. D., RODDEY, O. F., MARTIN, E. S., SWETENBURG, R. L., KOONCE, E. W., FELKNER, M. B., RYAN, A. G. & PRESBYTERIAN PEDIATRIC RESEARCH, G. 2003. Extrapharyngeal group A Streptococcus infection: diagnostic accuracy and utility of rapid antigen testing. *Pediatr Infect Dis J,* 22**,** 726-31.

COBEY, S. & LIPSITCH, M. 2012. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science,* 335**,** 1376-80.

COCHRAN, W. G. 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann Math Stat,* 17**,** 164-77.

COHEN, R., LEVY, C., THOLLOT, F., DE LA ROCQUE, F., KOSKAS, M., BONNET, E., FRITZELL, B. & VARON, E. 2007. Pneumococcal Conjugate Vaccine Does Not Influence Staphylococcus aureus Carriage in Young Children with Acute Otitis Media. *Clin Infect Dis,* 45**,** 1583-7.

COHEN, R., LEVY, C. & VARON, E. 2008. Does pneumococcal conjugate vaccine influence Staphylococcus aureus carriage in children? Reply to Regev-Yochay et al. *Clin Infect Dis,* 47**,** 291-2.

CORBIE-SMITH, G., VISCOLI, C. M., KERNAN, W. N., BRASS, L. M., SARREL, P. & HORWITZ, R. I. 2003. Influence of race, clinical, and other socio-demographic features on trial participation. *J Clin Epidemiol,* 56**,** 304-9.

CORLESS, C. E., GUIVER, M., BORROW, R., EDWARDS-JONES, V., FOX, A. J. & KACZMARSKI, E. B. 2001. Simultaneous detection of Neisseria meningitidis, Haemophilus influenzae, and Streptococcus pneumoniae in suspected cases of meningitis and septicemia using real-time PCR. *J Clin Microbiol,* 39**,** 1553-8.

COSTELLO, E. K., LAUBER, C. L., HAMADY, M., FIERER, N., GORDON, J. I. & KNIGHT, R. 2009. Bacterial community variation in human body habitats across space and time. *Science,* 326**,** 1694-7.

COTE, S., SANSCHAGRIN, F., DARGIS, M., SIMARD, J. L., ROY, P. H., MACDONALD, N. E., RABIN, H. R., BERGERON, M. G. & MALOUIN, F. 1994. Molecular typing of Haemophilus influenzae using a DNA probe and multiplex PCR. *Mol Cell Probes,* 8**,** 23-37.

COUGHTRIE, A. L., WHITTAKER, R. N., BEGUM, N., ANDERSON, R., TUCK, A., FAUST, S. N., JEFFERIES, J. M., YUEN, H. M., RODERICK, P. J., MULLEE, M. A., MOORE, M. V. & CLARKE, S. C. 2014. Evaluation of swabbing methods for estimating the prevalence of bacterial carriage in the upper respiratory tract: a cross sectional study. *BMJ Open,* 4**,** e005341.

COULOIGNER, V., LEVY, C., FRANCOIS, M., BIDET, P., HAUSDORFF, W. P., PASCAL, T., BOUCHERAT, M., BINGEN, E., MARIANI, P., PIERROT, S., BILLE, E., CARBONNELLE, E., VARON, E. & COHEN, R. 2012. Pathogens implicated in acute otitis media failures after 7-valent pneumococcal conjugate vaccine implementation in France: distribution, serotypes, and resistance levels. *Pediatr Infect Dis J,* 31**,** 154-8.

COX, D. R. 1958. The regression analysis of binary sequences. *J Roy Stat Soc B,* 20**,** 215–42.

CRAVEN, D. E., FRASCH, C. E., ROBBINS, J. B. & FELDMAN, H. A. 1978. Serogroup identification of Neisseria meningitidis: comparison of an antiserum agar method with bacterial slide agglutination. *J Clin Microbiol,* 7**,** 410-4.

CROUCHER, N. J., HARRIS, S. R., FRASER, C., QUAIL, M. A., BURTON, J., VAN DER LINDEN, M., MCGEE, L., VON GOTTBERG, A., SONG, J. H., KO, K. S., PICHON, B., BAKER, S., PARRY, C. M., LAMBERTSEN, L. M., SHAHINAS, D., PILLAI, D. R., MITCHELL, T. J., DOUGAN, G., TOMASZ, A., KLUGMAN, K. P., PARKHILL, J., HANAGE, W. P. & BENTLEY, S. D. 2011. Rapid pneumococcal evolution in response to clinical interventions. *Science,* 331**,** 430-4.

CROWE, S., UTLEY, M., WALKER, G., GROVE, P. & PAGEL, C. 2011. A model to evaluate mass vaccination against pneumococcus as a countermeasure against pandemic influenza. *Vaccine,* 29**,** 5065-77.

CROXTALL, J. D. & KEATING, G. M. 2009. Pneumococcal polysaccharide protein D-conjugate vaccine (Synflorix; PHiD-CV). *Paediatr Drugs,* 11**,** 349-57.

CUNDELL, D. R., WEISER, J. N., SHEN, J., YOUNG, A. & TUOMANEN, E. I. 1995. Relationship between Colonial Morphology and Adherence of Streptococcus-Pneumoniae. *Infect Immun,* 63**,** 757-61.

CUNNINGHAM, R., MATTHEWS, R., LEWENDON, G., HARRISON, S. & STUART, J. M. 2001. Improved rate of isolation of Neisseria meningitidis by direct plating of pharyngeal swabs. *J Clin Microbiol,* 39**,** 4575-6.

DAGAN, R. 2008. The impact of pneumococcal conjugate vaccines on otitis media-can we extrapolate to pneumonia? *Vaccine,* 26 Suppl 2**,** B16-8.

DALEY, P., CASTRICIANO, S., CHERNESKY, M. & SMIEJA, M. 2006. Comparison of flocked and rayon swabs for collection of respiratory epithelial cells from uninfected volunteers and symptomatic patients. *J Clin Microbiol,* 44**,** 2265-7.

DANG-VAN, A., TIRABY, G., ACAR, J. F., SHAW, W. V. & BOUANCHAUD, D. H. 1978. Chloramphenicol resistance in Streptococcus pneumoniae: enzymatic acetylation and possible plasmid linkage. *Antimicrob Agents Chemother,* 13**,** 577-83.

DARLINGTON, P. J. 1957. *Zoogeography: the geographical distribution of animals,* New York, Wiley.

DAVEY, M. E. & O'TOOLE G, A. 2000. Microbial biofilms: from ecology to molecular genetics. *Microbiol Mol Biol Rev,* 64**,** 847-67.

DAVIS, G. S., SANDSTEDT, S. A., PATEL, M., MARRS, C. F. & GILSDORF, J. R. 2011. Use of bexB to detect the capsule locus in Haemophilus influenzae. *J Clin Microbiol,* 49**,** 2594-601.

DAWID, S., ROCHE, A. M. & WEISER, J. N. 2007. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. *Infect Immun,* 75**,** 443-51.

DE GRAAFF, J., ELWELL, L. P. & FALKOW, S. 1976. Molecular nature of two beta-lactamase-specifying plasmids isolated from Haemophilus influenzae type b. *J Bacteriol,* 126**,** 439-46.

DE GREEFF, S. C., DE MELKER, H. E., VAN GAGELDONK, P. G., SCHELLEKENS, J. F., VAN DER KLIS, F. R., MOLLEMA, L., MOOI, F. R. & BERBERS, G. A. 2010. Seroprevalence of pertussis in The Netherlands: evidence for increased circulation of Bordetella pertussis. *PLoS One,* 5**,** e14183.

DELCHER, A. L., HARMON, D., KASIF, S., WHITE, O. & SALZBERG, S. L. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res,* 27**,** 4636-41.

DELEO, F. R. & MUSSER, J. M. 2010. Axis of coinfection evil. *J Infect Dis,* 201**,** 488-90.

DEPARDIEU, F., PODGLAJEN, I., LECLERCQ, R., COLLATZ, E. & COURVALIN, P. 2007. Modes and modulations of antibiotic resistance gene expression. *Clin Microbiol Rev,* 20**,** 79-114.

DEPARTMENT OF HEALTH 1999. Introduction of Immunisation against Group C Meningococcal Infection. London, UK.

DEPARTMENT OF HEALTH 2002. Extending Meningitis C Vaccine to 20-24 Year Olds; Pneumococcal Vaccine for at-risk under 2 Year Olds. London, UK.

DEPARTMENT OF HEALTH 2010. Introduction of Prevenar 13 into the Childhood Immunisation Programme. London, UK.

DETHLEFSEN, L., HUSE, S., SOGIN, M. L. & RELMAN, D. A. 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol,* 6**,** e280.

DHOUBHADEL, B. G., YASUNAMI, M., YOSHIDA, L. M., THI, H. A. N., THI, T. H. V., THI, T. A. N., WATANABE, K., SUZUKI, M., MORIMOTO, K., DANG, D. A. & ARIYOSHI, K. 2014. A novel high-throughput method for molecular serotyping and serotype-specific quantification of Streptococcus pneumoniae using a nanofluidic real-time PCR system. *J Med Microbiol,* 63**,** 528-39.

DIAVATOPOULOS, D. A., SHORT, K. R., PRICE, J. T., WILKSCH, J. J., BROWN, L. E., BRILES, D. E., STRUGNELL, R. A. & WIJBURG, O. L. 2010. Influenza A virus facilitates Streptococcus pneumoniae transmission and disease. *FASEB J,* 24**,** 1789-98.

DIDELOT, X., BOWDEN, R., WILSON, D. J., PETO, T. E. & CROOK, D. W. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet,* 13**,** 601-12.

DIGGLE, M. A. & CLARKE, S. C. 2005. Increased genetic diversity of Neisseria meningitidis isolates after the introduction of meningococcal serogroup C polysaccharide conjugate vaccines. *J Clin Microbiol,* 43**,** 4649-53.

DO, D. H., LAUS, S., LEBER, A., MARCON, M. J., JORDAN, J. A., MARTIN, J. M. & WADOWSKY, R. M. 2010. A one-step, real-time PCR assay for rapid detection of rhinovirus. *J Mol Diagn,* 12**,** 102-8.

DOERN, G. V., BRUEGGEMANN, A. B., PIERCE, G., HOLLEY, H. P. & RAUCH, A. 1997. Antibiotic resistance among clinical isolates of Haemophilus influenzae in the United States in 1994 and 1995 and detection of beta-lactamase-positive strains resistant to amoxicillin-clavulanate: Results of a national multicenter surveillance study. *Antimicrob Agents Chemother,* 41**,** 292-7.

DOLLHOPF, S. L., HASHSHAM, S. A. & TIEDJE, J. M. 2001. Interpreting 16S rDNA T-RFLP Data: Application of Self-Organizing Maps and Principal Component Analysis to Describe Community Dynamics and Convergence. *Microb Ecol,* 42**,** 495-505.

DONCASTER, C. P. 2009. Ecological equivalence: a realistic assumption for niche theory as a testable alternative to neutral theory. *PLoS One,* 4**,** e7460.

DONKOR, E. S., STABLER, R. A., HINDS, J., ADEGBOLA, R. A., ANTONIO, M. & WREN, B. W. 2012. Comparative phylogenomics of Streptococcus pneumoniae isolated from invasive disease and nasopharyngeal carriage from West Africans. *BMC Genomics,* 13**,** 569.

DONLAN, R. M. & COSTERTON, J. W. 2002. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev,* 15**,** 167-93.

DONNELLY, J., MEDINI, D., BOCCADIFUOCO, G., BIOLCHI, A., WARD, J., FRASCH, C., MOXON, E. R., STELLA, M., COMANDUCCI, M., BAMBINI, S., MUZZI, A., ANDREWS, W., CHEN, J., SANTOS, G., SANTINI, L., BOUCHER, P., SERRUTO, D., PIZZA, M., RAPPUOLI, R. & GIULIANI, M. M. 2010. Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines. *Proc Natl Acad Sci U S A,* 107**,** 19490-5.

DUIM, B., VOGEL, L., PUIJK, W., JANSEN, H. M., MELOEN, R. H., DANKERT, J. & VANALPHEN, L. 1996. Fine mapping of outer membrane protein P2 antigenic sites which vary during persistent infection by Haemophilus influenzae. *Infect Immun,* 64**,** 4673-9.

DUMBRELL, A. J., NELSON, M., HELGASON, T., DYTHAM, C. & FITTER, A. H. 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J,* 4**,** 337-45.

DUPPENTHALER, A., GORGIEVSKI-HRISOHO, M., FREY, U. & AEBI, C. 2003. Two-year periodicity of respiratory syncytial virus epidemics in Switzerland. *Infection,* 31**,** 75-80.

DUSHOFF, J., PLOTKIN, J. B., LEVIN, S. A. & EARN, D. J. D. 2004. Dynamical resonance can account for seasonality of influenza epidemics. *Proc Natl Acad Sci U S A,* 101**,** 16915-6.

EAKER, S., BERGSTROM, R., BERGSTROM, A., ADAMI, H. O. & NYREN, O. 1998. Response rate to mailed epidemiologic questionnaires: a population-based randomized trial of variations in design and mailing routines. *Am J Epidemiol,* 147**,** 74-82.

EARL, D., BRADNAM, K., ST JOHN, J., DARLING, A., LIN, D., FASS, J., YU, H. O., BUFFALO, V., ZERBINO, D. R., DIEKHANS, M., NGUYEN, N., ARIYARATNE, P. N., SUNG, W. K., NING, Z., HAIMEL, M., SIMPSON, J. T., FONSECA, N. A., BIROL, I., DOCKING, T. R., HO, I. Y., ROKHSAR, D. S., CHIKHI, R., LAVENIER, D., CHAPUIS, G., NAQUIN, D., MAILLET, N., SCHATZ, M. C., KELLEY, D. R., PHILLIPPY, A. M., KOREN, S., YANG, S. P., WU, W., CHOU, W. C., SRIVASTAVA, A., SHAW, T. I., RUBY, J. G., SKEWES-COX, P., BETEGON, M., DIMON, M. T., SOLOVYEV, V., SELEDTSOV, I., KOSAREV, P., VOROBYEV, D., RAMIREZ-GONZALEZ, R., LEGGETT, R., MACLEAN, D., XIA, F., LUO, R., LI, Z., XIE, Y., LIU, B., GNERRE, S., MACCALLUM, I., PRZYBYLSKI, D., RIBEIRO, F. J., YIN, S., SHARPE, T., HALL, G., KERSEY, P. J., DURBIN, R., JACKMAN, S. D., CHAPMAN, J. A., HUANG, X., DERISI, J. L., CACCAMO, M., LI, Y., JAFFE, D. B., GREEN, R. E., HAUSSLER, D., KORF, I. & PATEN, B. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res,* 21**,** 2224-41.

ECCLES, R. 2002. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Otolaryngol,* 122**,** 183-91.

ECDC 2014. Summary of the latest data on antibiotic consumption in the European Union. Stockholm, Sweden: ECDC.

ECKER, D. J., SAMPATH, R., MASSIRE, C., BLYN, L. B., HALL, T. A., ESHOO, M. W. & HOFSTADLER, S. A. 2008. Ibis T5000: a universal biosensor approach for microbiology. *Nat Rev Microbiol,* 6**,** 553-8.

EID, J., FEHR, A., GRAY, J., LUONG, K., LYLE, J., OTTO, G., PELUSO, P., RANK, D., BAYBAYAN, P., BETTMAN, B., BIBILLO, A., BJORNSON, K., CHAUDHURI, B., CHRISTIANS, F., CICERO, R., CLARK, S., DALAL, R., DEWINTER, A., DIXON, J., FOQUET, M., GAERTNER, A., HARDENBOL, P., HEINER, C., HESTER, K., HOLDEN, D., KEARNS, G., KONG, X., KUSE, R., LACROIX, Y., LIN, S., LUNDQUIST, P., MA, C., MARKS, P., MAXHAM, M., MURPHY, D., PARK, I., PHAM, T., PHILLIPS, M., ROY, J., SEBRA, R., SHEN, G., SORENSON, J., TOMANEY, A., TRAVERS, K., TRULSON, M., VIECELI, J., WEGENER, J., WU, D., YANG, A., ZACCARIN, D., ZHAO, P., ZHONG, F., KORLACH, J. & TURNER, S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science,* 323**,** 133-8.

EJLERTSEN, T., THISTED, E., EBBESEN, F., OLESEN, B. & RENNEBERG, J. 1994. Branhamella catarrhalis in children and adults. A study of prevalence, time of colonisation, and association with upper and lower respiratory tract infections. *J Infect,* 29**,** 23-31.

ELITH, J., GRAHAM, C. H., ANDERSON, R. P., DUDÍK, M., FERRIER, S., GUISAN, A., HIJMANS, R. J., HUETTMANN, F., LEATHWICK, J. R., LEHMANN, A., LI, J., LOHMANN, L. G., LOISELLE, B. A., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J. M., PETERSON, A. T., PHILLIPS, S. J., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPIRE, R. E., SOBERÓN, J., WILLIAMS, S., WISZ, M. S. & ZIMMERMANN, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography,* 29**,** 129-51.

ELLIS, J. S. & CURRAN, M. D. 2011. Simultaneous molecular detection and confirmation of influenza AH5, with internal control. *Methods Mol Biol,* 665**,** 161-81.

ELLNER, P. D. & ELLNER, C. J. 1966. Survival of bacteria on swabs. *J Bacteriol,* 91**,** 905-6.

ELTON, C. S. 1927. *Animal Ecology,* London, UK, Sidgwick & Jackson Limited.

ENNE, V. I., PLEYDELL. E. & BENNETT, P. M. 2006. Prevalence of antibiotic resistance and antibiotic resistance genes among Enterococcus faecium from pigs and poultry in the United Kingdom. *European Congress of Clinical Microbiology and Infectious Diseases.* Nice, France.

ENRIGHT, M. C. & SPRATT, B. G. 1998. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology,* 144**,** 3049-60.

ERB-DOWNWARD, J. R., THOMPSON, D. L., HAN, M. K., FREEMAN, C. M., MCCLOSKEY, L., SCHMIDT, L. A., YOUNG, V. B., TOEWS, G. B., CURTIS, J. L., SUNDARAM, B., MARTINEZ, F. J. & HUFFNAGLE, G. B. 2011. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One,* 6**,** e16384.

ESPOSITO, S., MARCHESE, A., TOZZI, A. E., ROSSI, G. A., DA DALT, L., BONA, G., PELUCCHI, C., SCHITO, G. C., PRINCIPI, N. & ITALIAN PNEUMOCOCCAL CAP GROUP 2013. DNA bacterial load in children with bacteremic pneumococcal community-acquired pneumonia. *Eur J Clin Microbiol Infect Dis,* 32**,** 877-81.

ETTER, J. F. & PERNEGER, T. V. 1997. Analysis of non-response bias in a mailed health survey. *J Clin Epidemiol,* 50**,** 1123-8.

FALLA, T. J., CROOK, D. W., BROPHY, L. N., MASKELL, D., KROLL, J. S. & MOXON, E. R. 1994. PCR for capsular typing of Haemophilus influenzae. *J Clin Microbiol,* 32**,** 2382-6.

FALSEY, A. R., CUNNINGHAM, C. K., BARKER, W. H., KOUIDES, R. W., YUEN, J. B., MENEGUS, M., WEINER, L. B., BONVILLE, C. A. & BETTS, R. F. 1995. Respiratory Syncytial Virus and Influenza-a Infections in the Hospitalized Elderly. *J Infect Dis,* 172**,** 389-94.

FARHAT, S. E., THIBAULT, M. & DEVLIN, R. 2001. Efficacy of a swab transport system in maintaining viability of Neisseria gonorrhoeae and Streptococcus pneumoniae. *J Clin Microbiol,* 39**,** 2958-60.

FARRELL, D. J., MORRISSEY, I., BAKKER, S., BUCKRIDGE, S. & FELMINGHAM, D. 2005. Global distribution of TEM-1 and ROB-1 beta-lactamases in Haemophilus influenzae. *J Antimicrob Chemother,* 56**,** 773-6.

FEDURCO, M., ROMIEU, A., WILLIAMS, S., LAWRENCE, I. & TURCATTI, G. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res,* 34**,** e22.

FEIL, E. J., LI, B. C., AANENSEN, D. M., HANAGE, W. P. & SPRATT, B. G. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol,* 186**,** 1518-30.

FERNANDEZ-JURICIC, E. 2002. Can human disturbance promote nestedness? A case study with breeding birds in urban habitat fragments. *Oecologia,* 131**,** 269-78.

FERNANDEZ, S., ARREAZA, L., SANTIAGO, I., MALVAR, A., BERRON, S., VAZQUEZ, J. A., HERVADA, X. & GESTAL, J. J. 1999. Carriage of a new epidemic strain of Neisseria meningitidis and its relationship with the incidence of meningococcal disease in Galicia, Spain. *Epidemiol Infect,* 123**,** 349-57.

FERNS, R. B. & GARSON, J. A. 2006. Development and evaluation of a real-time RT-PCR assay for quantification of cell-free human immunodeficiency virus type 2 using a Brome Mosaic Virus internal control. *J Virol Methods,* 135**,** 102-8.

FERRIER, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Syst Biol,* 51**,** 331-63.

FLANAGAN, J. L., BRODIE, E. L., WENG, L., LYNCH, S. V., GARCIA, O., BROWN, R., HUGENHOLTZ, P., DESANTIS, T. Z., ANDERSEN, G. L., WIENER-KRONISH, J. P. & BRISTOW, J. 2007. Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with Pseudomonas aeruginosa. *J Clin Microbiol,* 45**,** 1954-62.

FLASCHE, S., VAN HOEK, A. J., SHEASBY, E., WAIGHT, P., ANDREWS, N., SHEPPARD, C., GEORGE, R. & MILLER, E. 2011. Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in England: a cross-sectional study. *PLoS Med,* 8**,** e1001017.

FORSGREN, A., RIESBECK, K. & JANSON, H. 2008. Protein D of Haemophilus influenzae: a protective nontypeable H. influenzae antigen and a carrier for pneumococcal conjugate vaccines. *Clin Infect Dis,* 46**,** 726-31.

FOX, J. W., MARCON, M. J. & BONSU, B. K. 2006. Diagnosis of streptococcal pharyngitis by detection of Streptococcus pyogenes in posterior pharyngeal versus oral cavity specimens. *J Clin Microbiol,* 44**,** 2593-4.

FRASCH, C. E., PREZIOSI, M. P. & LAFORCE, F. M. 2012. Development of a group A meningococcal conjugate vaccine, MenAfriVac(TM). *Hum Vaccin Immunother,* 8**,** 715-24.

FRAZAO, N., SA-LEAO, R. & DE LENCASTRE, H. 2010. Impact of a single dose of the 7-valent pneumococcal conjugate vaccine on colonization. *Vaccine,* 28**,** 3445-52.

GALLAGHER, T. & PERLMAN, S. 2013. Public health: Broad reception for coronavirus. *Nature,* 495**,** 176-7.

GAMBLIN, J., JEFFERIES, J. M., HARRIS, S., AHMAD, N., MARSH, P., FAUST, S. N., FRASER, S., MOORE, M., RODERICK, P., BLAIR, I. & CLARKE, S. C. 2013. Nasal self-swabbing for estimating the prevalence of Staphylococcus aureus in the community. *J Med Microbiol,* 62**,** 437-40.

GAMBOTTO, A., BARRATT-BOYES, S. M., DE JONG, M. D., NEUMANN, G. & KAWAOKA, Y. 2008. Human infection with highly pathogenic H5N1 influenza virus. *Lancet,* 371**,** 1464-75.

GARCIA-RODRIGUEZ, J. A. & FRESNADILLO MARTÍNEZ, M. J. 2002. Dynamics of nasopharyngeal colonization by potential respiratory pathogens. *J Antimicrob Chemother,* 50**,** 59-74.

GAUTRET, P., CHARREL, R., BELHOUCHAT, K., DRALI, T., BENKOUITEN, S., NOUGAIREDE, A., ZANDOTTI, C., MEMISH, Z. A., AL MASRI, M., GAILLARD, C., BROUQUI, P. & PAROLA, P. 2013. Lack of nasal carriage of novel corona virus (HCoV-EMC) in French Hajj pilgrims returning from the Hajj 2012, despite a high rate of respiratory symptoms. *Clin Microbiol Infec,* 19**,** E315-7.

GETZIN, S., DEAN, C., HE, F. L., TROFYMOW, J. A., WIEGAND, K. & WIEGAND, T. 2006. Spatial patterns and competition of tree species in a Douglas-fir chronosequence on Vancouver Island. *Ecography,* 29**,** 671-682.

GIOVANETTI, E., BRENCIANI, A., BACCIAGLIA, A., MONTANARI, M. P. & VARALDO, P. E. 2007. Expressed and unexpressed tet(M) genes and the erm(B)-carrying Tn1116 transposon in Streptococcus pyogenes and Streptococcus pneumoniae. *Antimicrob Agents Chemother,* 51**,** 4535-6.

GIRARD, M. P., CHERIAN, T., PERVIKOV, Y. & KIENY, M. P. 2005. A review of vaccine research and development: human acute respiratory infections. *Vaccine,* 23**,** 5708-24.

GIUFRE, M., CARDINES, R., CAPORALI, M. G., ACCOGLI, M., D'ANCONA, F. & CERQUETTI, M. 2011. Ten years of Hib vaccination in Italy: prevalence of non-encapsulated Haemophilus influenzae among invasive isolates and the possible impact on antibiotic resistance. *Vaccine,* 29**,** 3857-62.

GLADSTONE, R. A. 2013. *Phenotypic and Genotypic Analysis of Streptococcus pneumoniae Diversity during the Introduction of Pneumococcal Conjugate Vaccines in the UK.* PhD, University of Southampton.

GLADSTONE, R. A., JEFFERIES, J. M., FAUST, S. N. & CLARKE, S. C. 2011. Continued control of pneumococcal disease in the UK - the impact of vaccination. *J Med Microbiol,* 60**,** 1-8.

GLADSTONE, R. A., JEFFERIES, J. M., FAUST, S. N. & CLARKE, S. C. 2012. Sampling methods for the study of pneumococcal carriage: a systematic review. *Vaccine,* 30**,** 6738-44.

GLADSTONE, R. A., JEFFERIES, J. M., TOCHEVA, A. S., BEARD, K. R., GARLEY, D., CHONG, W. W., BENTLEY, S. D., FAUST, S. N. & CLARKE, S. C. 2015. Five winters of pneumococcal serotype replacement in UK carriage following PCV introduction. *Vaccine,* 33**,** 2015-21.

GOLDA, A., MALEK, N., DUDEK, B., ZEGLEN, S., WOJARSKI, J., OCHMAN, M., KUCEWICZ, E., ZEMBALA, M., POTEMPA, J. & PYRC, K. 2011. Infection with human

coronavirus NL63 enhances streptococcal adherence to epithelial cells. *J Gen Virol,* 92**,** 1358-68.

GOLDBERG, M., CHASTANG, J. F., LECLERC, A., ZINS, M., BONENFANT, S., BUGEL, I., KANIEWSKI, N., SCHMAUS, A., NIEDHAMMER, I., PICIOTTI, M., CHEVALIER, A., GODARD, C. & IMBERNON, E. 2001. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol,* 154**,** 373-84.

GOLDMANN, D. A. 2000. Transmission of viral respiratory infections in the home. *Pediatr Infect Dis J,* 19**,** S97-102.

GOLDSMITH, C. E., MOORE, J. E. & MURPHY, P. G. 1997. Pneumococcal resistance in the UK. *J Antimicrob Chemother,* 40**,** 11-8.

GOLUBCHIK, T., BRUEGGEMANN, A. B., STREET, T., GERTZ, R. E., JR., SPENCER, C. C., HO, T., GIANNOULATOU, E., LINK-GELLES, R., HARDING, R. M., BEALL, B., PETO, T. E., MOORE, M. R., DONNELLY, P., CROOK, D. W. & BOWDEN, R. 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat Genet,* 44**,** 352-5.

GOMEZ-LUS, R. & VERGARA, Y. 1995. Aminoglycoside resistance in Haemophilus influenzae. *J Chemother,* 7**,** 97-9.

GONZALEZ, A., CLEMENTE, J. C., SHADE, A., METCALF, J. L., SONG, S., PRITHIVIRAJ, B., PALMER, B. E. & KNIGHT, R. 2011. Our microbial selves: what ecology can teach us. *EMBO Rep,* 12**,** 775-84.

GRAD, Y. H., LIPSITCH, M., FELDGARDEN, M., ARACHCHI, H. M., CERQUEIRA, G. C., FITZGERALD, M., GODFREY, P., HAAS, B. J., MURPHY, C. I., RUSS, C., SYKES, S., WALKER, B. J., WORTMAN, J. R., YOUNG, S., ZENG, Q., ABOUELLEIL, A., BOCHICCHIO, J., CHAUVIN, S., DESMET, T., GUJJA, S., MCCOWAN, C., MONTMAYEUR, A., STEELMAN, S., FRIMODT-MOLLER, J., PETERSEN, A. M., STRUVE, C., KROGFELT, K. A., BINGEN, E., WEILL, F. X., LANDER, E. S., NUSBAUM, C., BIRREN, B. W., HUNG, D. T. & HANAGE, W. P. 2012. Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. *Proc Natl Acad Sci U S A,* 109**,** 3065-70.

GRATTEN, M., GRATTEN, H., POLI, A., CARRAD, E., RAYMER, M. & KOKI, G. 1986. Colonisation of Haemophilus influenzae and Streptococcus pneumoniae in the upper respiratory tract of neonates in Papua New Guinea: primary acquisition, duration of carriage, and relationship to carriage in mothers. *Biol Neonate,* 50**,** 114-20.

GRAY, B. M., CONVERSE, G. M., 3RD & DILLON, H. C., JR. 1980. Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J Infect Dis,* 142**,** 923-33.

GRAY, B. M., TURNER, M. E. & DILLON, H. C., JR. 1982. Epidemiologic studies of Streptococcus pneumoniae in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. *Am J Epidemiol,* 116**,** 692-703.

GREENBERG, D., BROIDES, A., BLANCOVICH, I., PELED, N., GIVON-LAVI, N. & DAGAN, R. 2004. Relative importance of nasopharyngeal versus oropharyngeal sampling

for isolation of Streptococcus pneumoniae and Haemophilus influenzae from healthy and sick individuals varies with age. *J Clin Microbiol,* 42**,** 4604-9.

GREENBERG, D., GIVON-LAVI, N., BROIDES, A., BLANCOVICH, I., PELED, N. & DAGAN, R. 2006. The contribution of smoking and exposure to tobacco smoke to Streptococcus pneumoniae and Haemophilus influenzae carriage in children and their mothers. *Clin Infect Dis,* 42**,** 897-903.

GREENOUGH, A. 1996. Bacterial sepsis and meningitis. *Semin Neonatol,* 1**,** 147-59.

GREINER, O., DAY, P. J., ALTWEGG, M. & NADAL, D. 2003. Quantitative detection of Moraxella catarrhalis in nasopharyngeal secretions by real-time PCR. *J Clin Microbiol,* 41**,** 1386-90.

GRINNELL, J. 1928. *Presence and absence of animals,* California, USA, University of California Chronicle.

GROSS, C. P., FILARDO, G., MAYNE, S. T. & KRUMHOLZ, H. M. 2005. The impact of socioeconomic status and race on trial participation for older women with breast cancer. *Cancer,* 103**,** 483-91.

GUCLU, A. U., BAYSALLAR, M., GOZEN, A. G., KILIC, A., BALKAN, A. & DOGANCI, L. 2005. Polymerase chain reaction vs. conventional culture in detection of bacterial pneumonia agents. *Ann Microbiol,* 55**,** 313-6.

GUNNARSSON, R. K., HOLM, S. E. & SODERSTROM, M. 1998. The prevalence of potential pathogenic bacteria in nasopharyngeal samples from healthy children and adults. *Scand J Prim Health Care,* 16**,** 13-7.

HAASE, P. 1995. Spatial Pattern-Analysis in Ecology Based on Ripley K-Function - Introduction and Methods of Edge Correction. *J Veg Sci,* 6**,** 575-82.

HALL-STOODLEY, L., HU, F. Z., GIESEKE, A., NISTICO, L., NGUYEN, D., HAYES, J., FORBES, M., GREENBERG, D. P., DICE, B., BURROWS, A., WACKYM, P. A., STOODLEY, P., POST, J. C., EHRLICH, G. D. & KERSCHNER, J. E. 2006. Direct Detection of Bacterial Biofilms on the Middle-EarMucosa of Children With Chronic Otitis Media. *JAMA,* 296**,** 202-11.

HAMMITT, L. L., BLOCK, S., HENNESSY, T. W., DEBYLE, C., PETERS, H., PARKINSON, A., SINGLETON, R. J. & BUTLER, J. C. 2005. Outbreak of Invasive Haemophilus influenzae Serotype a Disease. *Pediatr Infect Dis J,* 24**,** 453-6.

HAMPSHIRE COUNTY COUNCIL. 2011. *Demography* [Online]. Hampshire, UK. Available: http://www3.hants.gov.uk/chapter_1_demography-2.pdf 14/06/2015].

HANAGE, W. P., BISHOP, C. J., LEE, G. M., LIPSITCH, M., STEVENSON, A., RIFAS-SHIMAN, S. L., PELTON, S. I., HUANG, S. S. & FINKELSTEIN, J. A. 2011. Clonal replacement among 19A Streptococcus pneumoniae in Massachusetts, prior to 13 valent conjugate vaccination. *Vaccine,* 29**,** 8877-81.

HANAGE, W. P., FINKELSTEIN, J. A., HUANG, S. S., PELTON, S. I., STEVENSON, A. E., KLEINMAN, K., HINRICHSEN, V. L. & FRASER, C. 2010. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. *Epidemics,* 2**,** 80-4.

HANSKI, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos,* 38**,** 210-21.

HARE, K. M., SINGLETON, R. J., GRIMWOOD, K., VALERY, P. C., CHENG, A. C., MORRIS, P. S., LEACH, A. J., SMITH-VAUGHAN, H. C., CHATFIELD, M., REDDING, G., REASONOVER, A. L., MCCALLUM, G. B., CHIKOYAK, L., MCDONALD, M. I., BROWN, N., TORZILLO, P. J. & CHANG, A. B. 2013. Longitudinal nasopharyngeal carriage and antibiotic resistance of respiratory bacteria in indigenous Australian and Alaska native children with bronchiectasis. *PLoS One,* 8**,** e70478.

HARFORD, C. G., LEIDLER, V. & HARA, M. 1949. Effect of the lesion due to Influenza Virus on the resistance of mice to inhaled pneumococci. *J Exp Med,* 89**,** 53-68.

HARGREAVES, R. M., SLACK, M. P., HOWARD, A. J., ANDERSON, E. & RAMSAY, M. E. 1996. Changing patterns of invasive Haemophilus influenzae disease in England and Wales after introduction of the Hib vaccination programme. *BMJ,* 312**,** 160-1.

HARRELL, F. E. 2001. *Regression Modeling Strategies,* New York, USA, Springer.

HARRIS, T. D., BUZBY, P. R., BABCOCK, H., BEER, E., BOWERS, J., BRASLAVSKY, I., CAUSEY, M., COLONELL, J., DIMEO, J., EFCAVITCH, J. W., GILADI, E., GILL, J., HEALY, J., JAROSZ, M., LAPEN, D., MOULTON, K., QUAKE, S. R., STEINMANN, K., THAYER, E., TYURINA, A., WARD, R., WEISS, H. & XIE, Z. 2008. Single-molecule DNA sequencing of a viral genome. *Science,* 320**,** 106-9.

HARRISON, L. H., SHUTT, K. A., ARNOLD, K. E., STERN, E. J., PONDO, T., KIEHLBAUCH, J. A., MYERS, R. A., HOLLICK, R. A., SCHMINK, S., VELLO, M., STEPHENS, D. S., MESSONNIER, N. E., MAYER, L. W. & CLARK, T. A. 2014. Meningococcal Carriage Among Georgia and Maryland High School Students. *J Infect Dis,* 211**,** 1761-8.

HARRISON, L. M., MORRIS, J. A., TELFORD, D. R., BROWN, S. M. & JONES, K. 1999. The nasopharyngeal bacterial flora in infancy: effects of age, gender, season, viral upper respiratory tract infection and sleeping position. *FEMS Immunol Med Mic,* 25**,** 19-28.

HARRO, C. D., BETTS, R. F., HARTZEL, J. S., ONORATO, M. T., LIPKA, J., SMUGAR, S. S. & KARTSONIS, N. A. 2012. The immunogenicity and safety of different formulations of a novel Staphylococcus aureus vaccine (V710): results of two Phase I studies. *Vaccine,* 30**,** 1729-36.

HARRO, J. M., PETERS, B. M., O'MAY, G. A., ARCHER, N., KERNS, P., PRABHAKARA, R. & SHIRTLIFF, M. E. 2010. Vaccine development in Staphylococcus aureus: taking the biofilm phenotype into consideration. *FEMS Immunol Med Mic,* 59**,** 306-23.

HAUSDORFF, W. P., SIBER, G. & PARADISO, P. R. 2001. Geographical differences in invasive pneumococcal disease rates and serotype frequency in young children. *Lancet,* 357**,** 950-2.

HE, F. L., LEGENDRE, P. & LAFRANKIE, J. V. 1997. Distribution patterns of tree species in a Malaysian tropical rain forest. *J Veg Sci,* 8**,** 105-14.

HEIM, A., EBNET, C., HARSTE, G. & PRING-AKERBLOM, P. 2003. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J Med Virol,* 70**,** 228-39.

HEININGER, U., KRUKER, A. T., BONHOEFFER, J. & SCHAAD, U. B. 2009. Human metapneumovirus infections--biannual epidemics and clinical findings in children in the region of Basel, Switzerland. *Eur J Pediatr,* 168**,** 1455-60.

HICKS, L. A., HARRISON, L. H., FLANNERY, B., HADLER, J. L., SCHAFFNER, W., CRAIG, A. S., JACKSON, D., THOMAS, A., BEALL, B., LYNFIELD, R., REINGOLD, A., FARLEY, M. M. & WHITNEY, C. G. 2007. Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. *J Infect Dis,* 196**,** 1346-54.

HILL, P. C., TOWNEND, J., ANTONIO, M., AKISANYA, B., EBRUKE, C., LAHAI, G., GREENWOOD, B. M. & ADEGBOLA, R. A. 2010. Transmission of Streptococcus pneumoniae in rural Gambian villages: a longitudinal study. *Clin Infect Dis,* 50**,** 1468-76.

HILTKE, T. J., SETHI, S. & MURPHY, T. F. 2002. Sequence stability of the gene encoding outer membrane protein P2 of nontypeable Haemophilus influenzae in the human respiratory tract. *J Infect Dis,* 185**,** 627-31.

HILTY, M., BURKE, C., PEDRO, H., CARDENAS, P., BUSH, A., BOSSLEY, C., DAVIES, J., ERVINE, A., POULTER, L., PACHTER, L., MOFFATT, M. F. & COOKSON, W. O. 2010. Disordered microbial communities in asthmatic airways. *PLoS One,* 5**,** e8578.

HILTY, M., QI, W., BRUGGER, S. D., FREI, L., AGYEMAN, P., FREY, P. M., AEBI, S. & MUHLEMANN, K. 2012. Nasopharyngeal microbiota in infants with acute otitis media. *J Infect Dis,* 205**,** 1048-55.

HOBAN, D. J., DOERN, G. V., FLUIT, A. C., ROUSSEL-DELVALLEZ, M. & JONES, R. N. 2001. Worldwide prevalence of antimicrobial resistance in Streptococcus pneumoniae, Haemophilus influenzae, and Moraxella catarrhalis in the SENTRY Antimicrobial Surveillance Program, 1997-1999. *Clin Infect Dis,* 32 Suppl 2**,** S81-93.

HOFMANN, J., CETRON, M. S., FARLEY, M. M., BAUGHMAN, W. S., FACKLAM, R. R., ELLIOTT, J. A., DEAVER, K. A. & BREIMAN, R. F. 1995. The prevalence of drug-resistant Streptococcus pneumoniae in Atlanta. *New Engl J Med,* 333**,** 481-6.

HOGBERG, L., GELI, P., RINGBERG, H., MELANDER, E., LIPSITCH, M. & EKDAHL, K. 2007. Age- and serogroup-related differences in observed durations of nasopharyngeal carriage of penicillin-resistant pneumococci. *J Clin Microbiol,* 45**,** 948-52.

HOGE, C. W. R., M.R.; DOMINGUEZ, E.A.; BREMER, J.C.; MASTRO, T.D.; HENDRICKS, K.A.; MUSHER, D.M.; ELLIOTT, J.A.; FACKLAM, R.R.; BREIMAN, R.F. 1994. An Epidemic of Pneumococcal Disease in an Overcrowded, Inadequately Ventilated Jail. *New Engl J Med,* 331**,** 643-8.

HONEYBORNE, I., MCHUGH, T. D., PHILLIPS, P. P., BANNOO, S., BATESON, A., CARROLL, N., PERRIN, F. M., RONACHER, K., WRIGHT, L., VAN HELDEN, P. D., WALZL, G. & GILLESPIE, S. H. 2011. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum Mycobacterium tuberculosis bacillary load during treatment. *J Clin Microbiol,* 49**,** 3905-11.

HOSMER, D. W., LEMESHOW, S. & STURDIVANT, R. X. 2013. *Applied Logistic Regression,* New Jersey, USA, Wiley.

HU, A., COLELLA, M., TAM, J. S., RAPPAPORT, R. & CHENG, S. M. 2003. Simultaneous detection, subgrouping, and quantitation of respiratory syncytial virus A and B by real-time PCR. *J Clin Microbiol,* 41**,** 149-54.

HUANG, S. S., HINRICHSEN, V. L., STEVENSON, A. E., RIFAS-SHIMAN, S. L., KLEINMAN, K., PELTON, S. I., LIPSITCH, M., HANAGE, W. P., LEE, G. M. & FINKELSTEIN, J. A. 2009. Continued impact of pneumococcal conjugate vaccine on carriage in young children. *Pediatrics,* 124**,** e1-11.

HUBBELL, S. P. 2001. *The Unified Theory of Biodiversity and Biogeography,* Priceton, USA, Princeton University Press.

HUGENHOLTZ, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol,* 3**,** 1-8.

ILLUMINA 2010. Illumina Sequencing Technology. *Technology Spotlight: Illumina® Sequencing.*

ILLUMINA 2015. MiSeq® System - System Specification Sheet: Sequencing.

INOUYE, M., CONWAY, T. C., ZOBEL, J. & HOLT, K. E. 2012. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC Genomics,* 13**,** 338.

INOUYE, M., DASHNOW, H., POPE, B. & HOLT, K. 2014a. *SRST2 - Short Read Sequence Typing for Bacterial Pathogens* [Online]. Melbourne, Australia. Available: http://katholt.github.io/srst2/.

INOUYE, M., DASHNOW, H., RAVEN, L., SCHULTZ, M. B., POPE, B. J., TOMITA, T., ZOBEL, J. & HOLT, K. E. 2014b. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med,* 6**,** 90.

IRVING, S. A., DONAHUE, J. G., SHAY, D. K., ELLIS-COYLE, T. L. & BELONGIA, E. A. 2009. Evaluation of self-reported and registry-based influenza vaccination status in a Wisconsin cohort. *Vaccine,* 27**,** 6546-9.

JACOBY, P., WATSON, K., BOWMAN, J., TAYLOR, A., RILEY, T. V., SMITH, D. W., LEHMANN, D. & KALGOORLIE OTITIS MEDIA RESEARCH PROJECT, T. 2007. Modelling the co-occurrence of Streptococcus pneumoniae with other bacterial and viral pathogens in the upper respiratory tract. *Vaccine,* 25**,** 2458-64.

JAMES, J. M. & BOLSTEIN, R. 1990. The Effect of Monetary Incentives and Follow-up Mailings on the Response Rate and Response Quality in Mail Surveys. *Public Opin Quart,* 54**,** 346-61.

JAMIESON, A. M., YU, S., ANNICELLI, C. H. & MEDZHITOV, R. 2010. Influenza virus-induced glucocorticoids compromise innate host defense against a secondary bacterial infection. *Cell Host Microbe,* 7**,** 103-14.

JARTTI, T., JARTTI, L., PELTOLA, V., WARIS, M. & RUUSKANEN, O. 2008. Identification of respiratory viruses in asymptomatic subjects: asymptomatic respiratory viral infections. *Pediatr Infect Dis J,* 27**,** 1103-7.

JAUNEIKAITE, E., TOCHEVA, A. S., JEFFERIES, J. M., GLADSTONE, R. A., FAUST, S. N., CHRISTODOULIDES, M., HIBBERD, M. L. & CLARKE, S. C. 2015. Current methods for capsular typing of Streptococcus pneumoniae. *J Microbiol Methods,* 113**,** 41-9.

JAYARAMAN, R. 2011. Phase variation and adaptation in bacteria: A 'Red Queen's Race'. *Current Science,* 100**,** 1163-71.

JCVI 2013. Minutes of the meeting on Wednesday 12 June 2013. London, UK.

JEFFERIES, J. M., CLARKE, S. C., WEBB, J. S. & KRAAIJEVELD, A. R. 2011. Risk of red queen dynamics in pneumococcal vaccine strategy. *Trends Microbiol,* 19**,** 377-81.

JERALDO, P., SIPOS, M., CHIA, N., BRULC, J. M., DHILLON, A. S., KONKEL, M. E., LARSON, C. L., NELSON, K. E., QU, A., SCHOOK, L. B., YANG, F., WHITE, B. A. & GOLDENFELD, N. 2012. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proc Natl Acad Sci U S A,* 109**,** 9692-8.

JOHNSON, A. P., AUCKEN, H. M., CAVENDISH, S., GANNER, M., WALE, M. C. J., WARNER, M., LIVERMORE, D. M., COOKSON, B. D. & PARTICPANTS, U. E. 2001. Dominance of EMRSA-15 and-16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS) - J Antimicrob Chemother 2001; 48 : 143-144. *J Antimicrob Chemother,* 48**,** 143-4.

JOHNSON, F. B. 1990. Transport of viral specimens. *Clin Microbiol Rev,* 3**,** 120-31.

JOHNSON, N. P. & MUELLER, J. 2002. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bull Hist Med,* 76**,** 105-15.

JONAS, D., SPECK, M., DASCHNER, F. D. & GRUNDMANN, H. 2002. Rapid PCR-based identification of methicillin-resistant Staphylococcus aureus from screening swabs. *J Clin Microbiol,* 40**,** 1821-3.

JORDENS, J. Z., WILLIAMS, J. N., JONES, G. R. & HECKELS, J. E. 2002. Detection of meningococcal carriage by culture and PCR of throat swabs and mouth gargles. *J Clin Microbiol,* 40**,** 75-9.

JORGENSEN, J. H., DOERN, G. V., MAHER, L. A., HOWELL, A. W. & REDDING, J. S. 1990. Antimicrobial Resistance among Respiratory Isolates of Haemophilus-Influenzae, Moraxella-Catarrhalis, and Streptococcus-Pneumoniae in the United-States. *Antimicrob Agents Chemother,* 34**,** 2075-80.

JOYCE, J. G., ABEYGUNAWARDANA, C., XU, Q., COOK, J. C., HEPLER, R., PRZYSIECKI, C. T., GRIMM, K. M., ROPER, K., IP, C. C., COPE, L., MONTGOMERY, D., CHANG, M., CAMPIE, S., BROWN, M., MCNEELY, T. B., ZORMAN, J., MAIRA-LITRAN, T., PIER, G. B., KELLER, P. M., JANSEN, K. U. & MARK, G. E. 2003. Isolation, structural characterization, and immunological evaluation of a high-molecular-weight exopolysaccharide from Staphylococcus aureus. *Carbohydr Res,* 338**,** 903-22.

KADAM, P. & BHALERAO, S. 2010. Sample size calculation. *Int J Ayurveda Res,* 1**,** 55-7.

KAMTCHOUA, T., BOLOGA, M., HOPFER, R., NEVEU, D., HU, B., SHENG, X., CORDE, N., POUZET, C., ZIMMERMANN, G. & GURUNATHAN, S. 2013. Safety and immunogenicity of the pneumococcal pneumolysin derivative PlyD1 in a single-antigen protein vaccine candidate in adults. *Vaccine,* 31**,** 327-33.

KANE, M., CASE, L. K., KOPASKIE, K., KOZLOVA, A., MACDEARMID, C., CHERVONSKY, A. V. & GOLOVKINA, T. V. 2011. Successful transmission of a retrovirus depends on the commensal microbiota. *Science,* 334**,** 245-9.

KAWAOKA, Y., KRAUSS, S. & WEBSTER, R. G. 1989. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *J Virol,* 63**,** 4603-8.

KAYE, P., MALKANI, R., MARTIN, S., SLACK, M., TROTTER, C., JIT, M., GEORGE, R. & MILLER, E. 2009. Invasive Pneumococcal Disease (IPD) in England & Wales after 7-valent conjugate vaccine (PCV7); potential impact of 10 and 13-valent vaccines. London, UK.

KECK, J. W., WENGER, J. D., BRUDEN, D. L., RUDOLPH, K. M., HURLBURT, D. A., HENNESSY, T. W. & BRUCE, M. G. 2014. PCV7-induced changes in pneumococcal carriage and invasive disease burden in Alaskan children. *Vaccine,* 32**,** 6478-84.

KELLNER, J. D., SCHEIFELE, D., VANDERKOOI, O. G., MACDONALD, J., CHURCH, D. L. & TYRRELL, G. J. 2008. Effects of routine infant vaccination with the 7-valent pneumococcal conjugate vaccine on nasopharyngeal colonization with streptococcus pneumoniae in children in Calgary, Canada. *Pediatr Infect Dis J,* 27**,** 526-32.

KERTESZ, D. A., COULTHART, M. B., RYAN, J. A., W.M., J. & ASHTON, F. E. 1998. Serogroup B, Electrophoretic Type 15 Neisseria meningitidis in Canada. *J Infect Dis,* 177**,** 1754-7.

KHALIL, M., AL-MAZROU, Y., FINDLOW, H., CHADHA, H., BOSCH CASTELLS, V., JOHNSON, D. R. & BORROW, R. 2012. Safety and immunogenicity of a meningococcal quadrivalent conjugate vaccine in five- to eight-year-old Saudi Arabian children previously vaccinated with two doses of a meningococcal quadrivalent polysaccharide vaccine. *Clin Vaccine Immunol,* 19**,** 1561-6.

KIM, H. W., ARROBIO, J. O., BRANDT, C. D., JEFFRIES, B. C., PYLES, G., REID, J. L., CHANOCK, R. M. & PARROTT, R. H. 1973. Epidemiology of Respiratory Syncytial Virus-Infection in Washington, Dc .1. Importance of Virus in Different Respiratory Tract Disease Syndromes and Temporal Distribution of Infection. *Am J Epidemiol,* 98**,** 216-25.

KING, C. E. 1964. Relative Abundance of Species and MacArthur's Model. *Ecology,* 45**,** 716-27.

KIRBY, A., GEBSKI, V. & KEECH, A. C. 2002. Determining the sample size in a clinical trial. *Med J Aust,* 177**,** 256-7.

KLEIN, D. 2002. Quantification using real-time PCR technology: applications and limitations. *Trends Mol Med,* 8**,** 257-60.

KLEIN, N. P., BARTLETT, J., ROWHANI-RAHBAR, A., FIREMAN, B. & BAXTER, R. 2012. Waning protection after fifth dose of acellular pertussis vaccine in children. *New Engl J Med,* 367**,** 1012-9.

KLUGMAN, K. P., CHIEN, Y. W. & MADHI, S. A. 2009. Pneumococcal pneumonia and influenza: a deadly combination. *Vaccine,* 27**,** C9-14.

KOOPMANS, M. 2004. Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. *Lancet,* 363**,** 587-93.

KORPPI, M., KOTANIEMI-SYRJANEN, A., WARIS, M., VAINIONPAA, R. & REIJONEN, T. M. 2004. Rhinovirus-associated wheezing in infancy: comparison with respiratory syncytial virus bronchiolitis. *Pediatr Infect Dis J,* 23**,** 995-9.

KOSER, C. U., HOLDEN, M. T., ELLINGTON, M. J., CARTWRIGHT, E. J., BROWN, N. M., OGILVY-STUART, A. L., HSU, L. Y., CHEWAPREECHA, C., CROUCHER, N. J., HARRIS, S. R., SANDERS, M., ENRIGHT, M. C., DOUGAN, G., BENTLEY, S. D., PARKHILL, J., FRASER, L. J., BETLEY, J. R., SCHULZ-TRIEGLAFF, O. B., SMITH, G. P. & PEACOCK, S. J. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *New Engl J Med,* 366**,** 2267-75.

KOTANIEMI, J. T., HASSI, J., KATAJA, M., JONSSON, E., LAITINEN, L. A., SOVIJARVI, A. R. A. & LUNDBACK, B. 2001. Does non-responder bias have a significant effect on the results in a postal questionnaire study? *Eur J Epidemiol,* 17**,** 809-17.

KRIZ, P., GIORGINI, D., MUSILEK, M., LARRIBE, M. & TAHA, M. K. 1999. Microevolution through DNA exchange among strains of Neisseria meningitidis isolated during an outbreak in the Czech Republic. *Res Microbiol,* 150**,** 273-80.

KROLL, J. S., MOXON, E. R. & LOYNDS, B. M. 1994. Natural genetic transfer of a putative virulence-enhancing mutation to Haemophilus influenzae type a. *J Infect Dis,* 169**,** 676-9.

KWAMBANA, B. A., BARER, M. R., BOTTOMLEY, C., ADEGBOLA, R. A. & ANTONIO, M. 2011. Early acquisition and high nasopharyngeal co-colonisation by Streptococcus pneumoniae and three respiratory pathogens amongst Gambian new-borns and infants. *BMC Infect Dis,* 11**,** 175.

KYAW, M. H., LYNFIELD, R., SCHAFFNER, W., CRAIG, A. S., HADLER, J., REINGOLD, A., THOMAS, A. R., HARRISON, L. H., BENNETT, N. M., FARLEY, M. M., FACKLAM, R. R., JORGENSEN, J. H., BESSER, J., ZELL, E. R., SCHUCHAT, A., WHITNEY, C. G. & SURVEILLANCE, A. B. C. 2006. Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. *New Engl J Med,* 354**,** 1455-63.

LACLAIRE, L. L., TONDELLA, M. L. C., BEALL, D. S., NOBLE, C. A., RAGHUNATHAN, P. L., ROSENSTEIN, N. E., POPOVIC, T. & SURVEILLANCE, A. B. C. 2003. Identification of Haemophilus influenzae serotypes by standard slide agglutination serotyping and PCR-based capsule typing. *J Clin Microbiol,* 41**,** 393-6.

LADHANI, S. N., COLLINS, S., VICKERS, A., LITT, D. J., CRAWFORD, C., RAMSAY, M. E. & SLACK, M. P. 2012a. Invasive Haemophilus influenzae serotype e and f disease, England and Wales. *Emerg Infect Dis,* 18**,** 725-32.

LADHANI, S. N., LUCIDARME, J., NEWBOLD, L. S., GRAY, S. J., CARR, A. D., FINDLOW, J., RAMSAY, M. E., KACZMARSKI, E. B. & BORROW, R. 2012b. Invasive meningococcal capsular group Y disease, England and Wales, 2007-2009. *Emerg Infect Dis,* 18**,** 63-70.

LAKSHMAN, R., MURDOCH, C., BURKINSHAW, R., SHAW, L. & FINN, A. 2003. Pneumococcal nasopharyngeal carriage in children following heptavalent pneumococcal conjugate vaccination in infancy. *Arch Dis Child,* 88**,** 211-4.

LAMAGNI, T. L., NEAL, S., KESHISHIAN, C., ALHADDAD, N., GEORGE, R., DUCKWORTH, G., VUOPIO-VARKILA, J. & EFSTRATIOU, A. 2008. Severe

Streptococcus pyogenes infections, United Kingdom, 2003-2004. *Emerg Infect Dis,* 14**,** 202-9.

LANDEGREN, U., NILSSON, M. & KWOK, P. Y. 1998. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res,* 8**,** 769-76.

LANKINEN, K. S., RINTAMAKI, S., SYRJANEN, R., KILPI, T., RUUTU, P. & LEINONEN, M. 2004. Type-specific enzyme immunoassay for detection of pneumococcal capsular polysaccharide antigens in nasopharyngeal specimens. *J Microbiol Methods,* 56**,** 193-9.

LARSEN, S., OSNES, M., EIDSAUNET, W. & SANDVIK, L. 1985. Factors influencing the sample size, exemplified by studies on gastroduodenal tolerability of drugs. *Scand J Gastroenterol,* 20**,** 395-400.

LAUFER, A. S., METLAY, J. P., GENT, J. F., FENNIE, K. P., KONG, Y. & PETTIGREW, M. M. 2011. Microbial communities of the upper respiratory tract and otitis media in children. *MBio,* 2**,** e00245-10.

LAZAREVIC, V., MANZANO, S., GAIA, N., GIRARD, M., WHITESON, K., HIBBS, J., FRANCOIS, P., GERVAIX, A. & SCHRENZEL, J. 2013. Effects of amoxicillin treatment on the salivary microbiota in children with acute otitis media. *Clin Microbiol Infect,* 19**,** E335-42.

LEACH, A. J., SHELBY-JAMES, T. M., MAYO, M., GRATTEN, M., LAMING, A. C., CURRIE, B. J. & MATHEWS, J. D. 1997. A prospective study of the impact of community-based azithromycin treatment of trachoma on carriage and resistance of Streptococcus pneumoniae. *Clin Infect Dis,* 24**,** 356-62.

LEACH, A. J., STUBBS, E., HARE, K., BEISSBARTH, J. & MORRIS, P. S. 2008. Comparison of nasal swabs with nose blowing for community-based pneumococcal surveillance of healthy children. *J Clin Microbiol,* 46**,** 2081-2.

LEE, M. H., ARRECUBIETA, C., MARTIN, F. J., PRINCE, A., BORCZUK, A. C. & LOWY, F. D. 2010. A postinfluenza model of Staphylococcus aureus pneumonia. *J Infect Dis,* 201**,** 508-15.

LEIBOVITZ, E., JACOBS, M. R. & DAGAN, R. 2004. Haemophilus influenzae: a significant pathogen in acute otitis media. *Pediatr Infect Dis J,* 23**,** 1142-52.

LEIMKUGEL, J., ADAMS FORGOR, A., GAGNEUX, S., PFLUGER, V., FLIERL, C., AWINE, E., NAEGELI, M., DANGY, J.-P., SMITH, T., HODGSON, A. & PLUSCHKE, G. 2005. An Outbreak of Serotype 1 Streptococcus pneumoniae Meningitis in Northern Ghana with Features That Are Characteristic of Neisseria meningitidis. *J Infect Dis,* 192**,** 192-9.

LEPOW, M. L., BEELER, J., RANDOLPH, M., SAMUELSON, J. S. & HANKINS, W. A. 1986. Reactogenicity and immunogenicity of a quadrivalent combined meningococcal polysaccharide vaccine in children. *J Infect Dis,* 154**,** 1033-6.

LEROUX-ROELS, G., MAES, C., DE BOEVER, F., TRASKINE, M., RUGGEBERG, J. U. & BORYS, D. 2014. Safety, reactogenicity and immunogenicity of a novel pneumococcal protein-based vaccine in adults: A phase I/II randomized clinical study. *Vaccine,* 32**,** 6838-46.

LEY, R. E., TURNBAUGH, P. J., KLEIN, S. & GORDON, J. I. 2006. Microbial ecology: human gut microbes associated with obesity. *Nature,* 444**,** 1022-3.

LIEBERMAN, D., SHLEYFER, E., CASTEL, H., TERRY, A., HARMAN-BOEHM, I., DELGADO, J. & PELED, N. 2006. Nasopharyngeal versus oropharyngeal sampling for isolation of potential respiratory pathogens in adults. *J Clin Microbiol,* 44**,** 525-8.

LIPSITCH, M. 1997. Vaccination against colonizing bacteria with multiple serotypes. *Proc Natl Acad Sci U S A,* 94**,** 6571-6.

LIPSITCH, M., COLIJN, C., COHEN, T., HANAGE, W. P. & FRASER, C. 2009. No coexistence for free: neutral null models for multistrain pathogens. *Epidemics,* 1**,** 2-13.

LJUTIC, B., OCHS, M., MESSHAM, B., MING, M., DOOKIE, A., HARPER, K. & AUSAR, S. F. 2012. Formulation, stability and immunogenicity of a trivalent pneumococcal protein vaccine formulated with aluminum salt adjuvants. *Vaccine,* 30**,** 2981-8.

LOZANO, R., NAGHAVI, M., FOREMAN, K., LIM, S., SHIBUYA, K., ABOYANS, V., ABRAHAM, J., ADAIR, T., AGGARWAL, R., AHN, S. Y., ALVARADO, M., ANDERSON, H. R., ANDERSON, L. M., ANDREWS, K. G., ATKINSON, C., BADDOUR, L. M., BARKER-COLLO, S., BARTELS, D. H., BELL, M. L., BENJAMIN, E. J., BENNETT, D., BHALLA, K., BIKBOV, B., BIN ABDULHAK, A., BIRBECK, G., BLYTH, F., BOLLIGER, I., BOUFOUS, S., BUCELLO, C., BURCH, M., BURNEY, P., CARAPETIS, J., CHEN, H., CHOU, D., CHUGH, S. S., COFFENG, L. E., COLAN, S. D., COLQUHOUN, S., COLSON, K. E., CONDON, J., CONNOR, M. D., COOPER, L. T., CORRIERE, M., CORTINOVIS, M., DE VACCARO, K. C., COUSER, W., COWIE, B. C., CRIQUI, M. H., CROSS, M., DABHADKAR, K. C., DAHODWALA, N., DE LEO, D., DEGENHARDT, L., DELOSSANTOS, A., DENENBERG, J., DES JARLAIS, D. C., DHARMARATNE, S. D., DORSEY, E. R., DRISCOLL, T., DUBER, H., EBEL, B., ERWIN, P. J., ESPINDOLA, P., EZZATI, M., FEIGIN, V., FLAXMAN, A. D., FOROUZANFAR, M. H., FOWKES, F. G., FRANKLIN, R., FRANSEN, M., FREEMAN, M. K., GABRIEL, S. E., GAKIDOU, E., GASPARI, F., GILLUM, R. F., GONZALEZ-MEDINA, D., HALASA, Y. A., HARING, D., HARRISON, J. E., HAVMOELLER, R., HAY, R. J., HOEN, B., HOTEZ, P. J., HOY, D., JACOBSEN, K. H., JAMES, S. L., JASRASARIA, R., JAYARAMAN, S., JOHNS, N., KARTHIKEYAN, G., KASSEBAUM, N., KEREN, A., KHOO, J. P., KNOWLTON, L. M., KOBUSINGYE, O., KORANTENG, A., KRISHNAMURTHI, R., LIPNICK, M., LIPSHULTZ, S. E., OHNO, S. L., et al. 2012. 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,* 380**,** 2095-128.

LUNA, L. K. D., HEISER, V., REGAMEY, N., PANNING, M., DREXLER, J. F., MULANGU, S., POON, L., BAUMGARTE, S., HAIJEMA, B. J., KAISER, L. & DROSTEN, C. 2007. Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *J Clin Microbiol,* 45**,** 1049-52.

MACARTHUR, R. H. 1957. On the Relative Abundance of Bird Species. *Proc Natl Acad Sci U S A,* 43**,** 293-5.

MACKAY, I. M. 2004. Real-time PCR in the microbiology laboratory. *Clin Microbiol Infec,* 10**,** 190-212.

MACKENZIE, G. A., LEACH, A. J., CARAPETIS, J. R., FISHER, J. & MORRIS, P. S. 2010. Epidemiology of nasopharyngeal carriage of respiratory bacterial pathogens in

children and adults: cross-sectional surveys in a population with high rates of pneumococcal disease. *BMC Infect Dis,* 10**,** 304.

MACLENNAN, J., KAFATOS, G., NEAL, K., ANDREWS, N., CAMERON, J. C., ROBERTS, R., EVANS, M. R., CANN, K., BAXTER, D. N., MAIDEN, M. C., STUART, J. M. & UNITED KINGDOM MENINGOCOCCAL CARRIAGE, G. 2006. Social behavior and meningococcal carriage in British teenagers. *Emerg Infect Dis,* 12**,** 950-7.

MADELEY, C. R. 1977. Guide to the collection and transport of virological specimens. Geneva, Switzerland: World Health Organization.

MADHI, S. A., LUDEWICK, H., KUWANDA, L., VAN NIEKERK, N., CUTLAND, C., LITTLE, T. & KLUGMAN, K. P. 2006. Pneumococcal Coinfection with Human Metapneumovirus. *J Infect Dis,* 193**,** 1236-43.

MAERTZDORF, J., WANG, C. K., BROWN, J. B., QUINTO, J. D., CHU, M., DE GRAAF, M., VAN DEN HOOGEN, B. G., SPAETE, R., OSTERHAUS, A. D. M. E. & FOUCHIER, R. A. M. 2004. Real-time reverse transcriptase PCR assay for detection of human metapneumoviruses from all known genetic lineages. *J Clin Microbiol,* 42**,** 981-6.

MAGEE, A. D. & YOTHER, J. 2001. Requirement for capsule in colonization by Streptococcus pneumoniae. *Infect Immun,* 69**,** 3755-61.

MAIDEN, M. C., IBARZ-PAVON, A. B., URWIN, R., GRAY, S. J., ANDREWS, N. J., CLARKE, S. C., WALKER, A. M., EVANS, M. R., KROLL, J. S., NEAL, K. R., ALA'ALDEEN, D. A., CROOK, D. W., CANN, K., HARRISON, S., CUNNINGHAM, R., BAXTER, D., KACZMARSKI, E., MACLENNAN, J., CAMERON, J. C. & STUART, J. M. 2008. Impact of meningococcal serogroup C conjugate vaccines on carriage and herd immunity. *J Infect Dis,* 197**,** 737-43.

MAIDEN, M. C. & STUART, J. M. 2002. Carriage of serogroup C meningococci 1 year after meningococcal C conjugate polysaccharide vaccination. *Lancet,* 359**,** 1829-31.

MAIDEN, M. C. J., BYGRAVES, J. A., FEIL, E., MORELLI, G., RUSSELL, J. E., URWIN, R., ZHANG, Q., ZHOU, J. J., ZURTH, K., CAUGANT, D. A., FEAVERS, I. M., ACHTMAN, M. & SPRATT, B. G. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A,* 95**,** 3140-5.

MANN, C. J. 2003. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J,* 20**,** 54-60.

MARGOLIS, E., YATES, A. & LEVIN, B. R. 2010. The ecology of nasal colonization of Streptococcus pneumoniae, Haemophilus influenzae and Staphylococcus aureus: the role of competition and interactions with host's immune response. *BMC Microbiol,* 10**,** 59.

MATSIOTA-BERNARD, P., PITSOUNI, E., LEGAKIS, N. & NAUCIEL, C. 1994. Evaluation of commercial amplification kit for detection of Legionella pneumophila in clinical specimens. *J Clin Microbiol,* 32**,** 1503-5.

MCARTHUR, A. G., WAGLECHNER, N., NIZAM, F., YAN, A., AZAD, M. A., BAYLAY, A. J., BHULLAR, K., CANOVA, M. J., DE PASCALE, G., EJIM, L., KALAN, L., KING, A. M., KOTEVA, K., MORAR, M., MULVEY, M. R., O'BRIEN, J. S., PAWLOWSKI, A. C., PIDDOCK, L. J., SPANOGIANNOPOULOS, P., SUTHERLAND, A. D., TANG, I., TAYLOR, P. L., THAKER, M., WANG, W., YAN, M., YU, T. & WRIGHT, G. D. 2013. The

comprehensive antibiotic resistance database. *Antimicrob Agents Chemother,* 57**,** 3348-57.

MCAVIN, J. C., REILLY, P. A., ROUDABUSH, R. M., BARNES, W. J., SALMEN, A., JACKSON, G. W., BENINGA, K. K., ASTORGA, A., MCCLESKEY, F. K., HUFF, W. B., NIEMEYER, D. & LOHMAN, K. L. 2001. Sensitive and specific method for rapid identification of Streptococcus pneumoniae using real-time fluorescence PCR. *J Clin Microbiol,* 39**,** 3446-51.

MCCANN, K. S. 2000. The diversity-stability debate. *Nature,* 405**,** 228-33.

MCCULLERS, J. A. 2006. Insights into the interaction between influenza virus and pneumococcus. *Clin Microbiol Rev,* 19**,** 571-82.

MCCULLERS, J. A. & REHG, J. E. 2002. Lethal Synergism between influenza virus and Streptococcus pneumoniae: Characterization of a mouse model and the role of platelet-activating factor receptor. *J Infect Dis,* 186**,** 341-50.

MCELLISTREM, M. C. & NAHM, M. H. 2012. Novel Pneumococcal Serotypes 6C and 6D: Anomaly or Harbinger. *Clin Infect Dis***,** 1379-86.

MCGILL, B. J. 2003. A test of the unified neutral theory of biodiversity. *Nature,* 422**,** 881-5.

MCNAMEE, R. 2005. Regression modelling and other methods to control confounding. *Occup Environ Med,* 62**,** 500-6.

MCVERNON, J., HOWARD, A. J., SLACK, M. P. & RAMSAY, M. E. 2004. Long-term impact of vaccination on Haemophilus influenzae type b (Hib) carriage in the United Kingdom. *Epidemiol Infect,* 132**,** 765-7.

MEATS, E., FEIL, E. J., STRINGER, S., CODY, A. J., GOLDSTEIN, R., KROLL, J. S., POPOVIC, T. & SPRATT, B. G. 2003. Characterization of encapsulated and noncapsulated Haemophilus influenzae and determination of phylogenetic relationships by multilocus sequence typing. *J Clin Microbiol,* 41**,** 1623-36.

MEDEIROS, A. A. 1984. Beta-lactamases. *Br Med Bull,* 40**,** 18-27.

MEHTAR, S. & LAW, C. A. 1980. An outbreak of beta-lactamase producing Haemophilus influenzae (biotype III) on a geriatric ward. *J Hosp Infect,* 1**,** 357-8.

MELANDER, E., MOLSTAD, S., PERSSON, K., HANSSON, H. B., SODERSTROM, M. & EKDAHL, K. 1998. Previous antibiotic consumption and other risk factors for carriage of penicillin-resistant Streptococcus pneumoniae in children. *Eur J Clin Microbiol Infect Dis,* 17**,** 834-8.

METZKER, M. L. 2005. Emerging technologies in DNA sequencing. *Genome Res,* 15**,** 1767-76.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

MICHEL, L. V., KALMETA, B., MCCREARY, M., SNYDER, J., CRAIG, P. & PICHICHERO, M. E. 2011. Vaccine candidate P6 of nontypable Haemophilus influenzae is not a transmembrane protein based on protein structural analysis. *Vaccine,* 29**,** 1624-7.

MIKKELSON, G. M. 2005. Ecological niches: Linking classical and contemporary approaches. *Biol Philos,* 20**,** 557-66.

MILLER, E., ANDREWS, N. J., WAIGHT, P. A., SLACK, M. P. & GEORGE, R. C. 2011. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. *Lancet Infect Dis,* 11**,** 760-8.

MILLER, E., HOSCHLER, K., HARDELID, P., STANFORD, E., ANDREWS, N. & ZAMBON, M. 2010. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet,* 375**,** 1100-8.

MILLER, E., SALISBURY, D. & RAMSAY, M. 2001. Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story. *Vaccine,* 20**,** S58-67.

MIRA, A., KLASSON, L. & ANDERSSON, S. G. E. 2002. Microbial genome evolution: sources of variability. *Curr Opin Microbiol,* 5**,** 508-12.

MOLIN, S. & TOLKER-NIELSEN, T. 2003. Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Curr Opin Biotech,* 14**,** 255-61.

MONTO, A. S. 2002. Epidemiology of viral respiratory infections. *Am J Med,* 112 4-12S.

MOORE, H. C., JACOBY, P., TAYLOR, A., HARNETT, G., BOWMAN, J., RILEY, T. V., REUTER, K., SMITH, D. W., LEHMANN, D. & PROJEC, K. O. M. R. 2010. The Interaction Between Respiratory Viruses and Pathogenic Bacteria in the Upper Respiratory Tract of Asymptomatic Aboriginal and Non-Aboriginal Children. *Pediatr Infect Dis J,* 29**,** 540-5.

MORRIS, A., BECK, J. M., SCHLOSS, P. D., CAMPBELL, T. B., CROTHERS, K., CURTIS, J. L., FLORES, S. C., FONTENOT, A. P., GHEDIN, E., HUANG, L., JABLONSKI, K., KLEERUP, E., LYNCH, S. V., SODERGREN, E., TWIGG, H., YOUNG, V. B., BASSIS, C. M., VENKATARAMAN, A., SCHMIDT, T. M., WEINSTOCK, G. M. & LUNG, H. I. V. M. P. 2013. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med,* 187**,** 1067-75.

MORRIS, D. P. 2007. Bacterial Biofilm in Upper Respiratory Tract Infections. *Curr Infect Dis Rep* 9**,** 186-92.

MORRIS, S., BAI, G. H., SUFFYS, P., PORTILLOGOMEZ, L., FAIRCHOK, M. & ROUSE, D. 1995. Molecular Mechanisms of Multiple-Drug Resistance in Clinical Isolates of Mycobacterium-Tuberculosis. *J Infect Dis,* 171**,** 954-60.

MORRIS, S. K., MOSS, W. J. & HALSEY, N. 2008. Haemophilus influenzae type b conjugate vaccine use and effectiveness. *Lancet Infect Dis,* 8**,** 435-43.

MOTHERSHED, E. A., SACCHI, C. T., WHITNEY, A. M., BARNETT, G. A., AJELLO, G. W., SCHMINK, S., MAYER, L. W., PHELAN, M., TAYLOR, T. H., BERNHARDT, S. A., ROSENSTEIN, N. E. & POPOVIC, T. 2004. Use of real-time PCR to resolve slide agglutination discrepancies in serogroup identification of Neisseria meningitidis. *J Clin Microbiol,* 42**,** 320-8.

MULHOLLAND, K., HILTON, S., ADEGBOLA, R., USEN, S., OPARAUGO, A., OMOSIGHO, C., WEBER, M., PALMER, A., SCHNEIDER, G., JOBE, K., LAHAI, G.,

JAFFAR, S., SECKA, O., LIN, K., ETHEVENAUX, C. & GREENWOOD, B. 1997. Randomised trial of Haemophilus influenzae type-b tetanus protein conjugate vaccine [corrected] for prevention of pneumonia and meningitis in Gambian infants. *Lancet,* 349**,** 1191-7.

MULHOLLAND, K. & SATZKE, C. 2012. Serotype replacement after pneumococcal vaccination. *Lancet,* 379**,** 1387.

MULLER, F. M., HOPPE, J. E. & WIRSING VON KONIG, C. H. 1997. Laboratory diagnosis of pertussis: state of the art in 1997. *J Clin Microbiol,* 35**,** 2435-43.

MURPHY, T. F., BAKALETZ, L. O. & SMEESTERS, P. R. 2009. Microbial interactions in the respiratory tract. *Pediatr Infect Dis J,* 28**,** S121-6.

MURPHY, T. F., BRAUER, A. L., GRANT, B. J. & SETHI, S. 2005. Moraxella catarrhalis in chronic obstructive pulmonary disease: burden of disease and immune response. *Am J Respir Crit Care Med,* 172**,** 195-9.

MURPHY, T. F. & KIRKHAM, C. 2002. Biofilm formation by nontypeable Haemophilus influenzae: strain variability, outer membrane antigen expression and role of pili. *BMC Microbiol,* 2**,** 1-8.

MURRAY, M. A., SCHULZ, L. A., FURST, J. W., HOMME, J. H., JENKINS, S. M., UHL, J. R., PATEL, R., COCKERILL, F. C., MYERS, J. F. & PRITT, B. S. 2015. Equal performance of self-collected and health care worker-collected pharyngeal swabs for group a streptococcus testing by PCR. *J Clin Microbiol,* 53**,** 573-8.

MURTHY, V. H., KRUMHOLZ, H. M. & GROSS, C. P. 2004. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA,* 291**,** 2720-6.

MUSHER, D. M. 2003. Medical progress: How contagious are common respiratory tract infections? *New Engl J Med,* 348**,** 1256-66.

NAGAI, K., SHIBASAKI, Y., HASEGAWA, K., DAVIES, T. A., JACOBS, M. R., UBUKATA, K. & APPELBAUM, P. C. 2001. Evaluation of PCR primers to screen for Streptococcus pneumoniae isolates and beta-lactam resistance, and to detect common macrolide resistance determinants. *J Antimicrob Chemother,* 48**,** 915-8.

NAIR, H., BROOKS, W. A., KATZ, M., ROCA, A., BERKLEY, J. A., MADHI, S. A., SIMMERMAN, J. M., GORDON, A., SATO, M., HOWIE, S., KRISHNAN, A., OPE, M., LINDBLADE, K. A., CAROSONE-LINK, P., LUCERO, M., OCHIENG, W., KAMIMOTO, L., DUEGER, E., BHAT, N., VONG, S., THEODORATOU, E., CHITTAGANPITCH, M., CHIMAH, O., BALMASEDA, A., BUCHY, P., HARRIS, E., EVANS, V., KATAYOSE, M., GAUR, B., O'CALLAGHAN-GORDO, C., GOSWAMI, D., ARVELO, W., VENTER, M., BRIESE, T., TOKARZ, R., WIDDOWSON, M. A., MOUNTS, A. W., BREIMAN, R. F., FEIKIN, D. R., KLUGMAN, K. P., OLSEN, S. J., GESSNER, B. D., WRIGHT, P. F., RUDAN, I., BROOR, S., SIMÕES, E. A. & CAMPBELL, H. 2011. Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. *Lancet,* 378**,** 1917-30.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2008. Respiratory tract infections - antibiotic prescribing costing report. London: NHS.

NAVARINI, A. A., RECHER, M., LANG, K. S., GEORGIEV, P., MEURY, S., BERGTHALER, A., FLATZ, L., BILLE, J., LANDMANN, R., ODERMATT, B.,

HENGARTNER, H. & ZINKERNAGEL, R. M. 2006. Increased susceptibility to bacterial superinfection as a consequence of innate antiviral responses. *Proc Natl Acad Sci U S A,* 103**,** 15535-9.

NEUZIL, K. M., ZHU, Y. W., GRIFFIN, M. R., EDWARDS, K. M., THOMPSON, J. M., TOLLEFSON, S. J. & WRIGHT, P. F. 2002. Burden of interpandemic influenza in children younger than 5 years: A 25-year prospective study. *J Infect Dis,* 185**,** 147-52.

NICHOLSON, K. G., KENT, J., HAMMERSLEY, V. & CANCIO, E. 1997. Acute viral infections of upper respiratory tract in elderly people living in the community: comparative, prospective, population based study of disease burden. *BMJ,* 315**,** 1060-4.

NUORTI, J. P., BUTLER, J. C., CRUTCHER, J. M., GUEVARA, R., WELCH, D., HOLDER, P. & ELLIOTT, J. A. 1998. An Outbreak of Multidrug-Resistant Pneumococcal Pneumonia and Bacteremia Among Unvaccinated Nursing Home Residents. *New Engl J Med,* 338**,** 1861-8.

O'BRIEN, K. L. & NOHYNEK, H. 2003. Report from a WHO Working Group: standard method for detecting upper respiratory carriage of Streptococcus pneumoniae. *Pediatr Infect Dis J,* 22**,** e1-11.

OCHMAN, H. & DAVALOS, L. M. 2006. The nature and dynamics of bacterial genomes. *Science,* 311**,** 1730-3.

OCHMAN, H., LAWRENCE, J. G. & GROISMAN, E. A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature,* 405**,** 299-304.

OGUNNIYI, A. D., GRABOWICZ, M., BRILES, D. E., COOK, J. & PATON, J. C. 2007. Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of Streptococcus pneumoniae. *Infect Immun,* 75**,** 350-7.

OLCEN, P., KJELLANDER, J., DANIELSSON, D. & LINDQUIST, B. L. 1979. Culture diagnosis of meningococcal carriers: yield from different sites and influence of storage in transport medium. *J Clin Pathol,* 32**,** 1222-5.

ORENSTEIN, W. A., BERNIER, R. H., DONDERO, T. J., HINMAN, A. R., MARKS, J. S., BART, K. J. & SIROTKIN, B. 1985. Field evaluation of vaccine efficacy. *Bull World Health Organ,* 63**,** 1055-68.

OTSUKA, T., KIRKHAM, C., JOHNSON, A., JONES, M. M. & MURPHY, T. F. 2014. Substrate Binding Protein SBP2 of a Putative ABC Transporter as a Novel Vaccine Antigen of Moraxella catarrhalis. *Infect Immun,* 82**,** 3503-12.

PAI, R., GERTZ, R. E. & BEALL, B. 2006. Sequential multiplex PCR approach for determining capsular serotypes of Streptococcus pneumoniae isolates. *J Clin Microbiol,* 44**,** 124-31.

PAISLEY, R. D., BLAYLOCK, J. & HARTZELL, J. D. 2012. Whooping cough in adults: an update on a reemerging infection. *Am J Med,* 125**,** 141-3.

PANDA, S., EL KHADER, I., CASELLAS, F., LOPEZ VIVANCOS, J., GARCIA CORS, M., SANTIAGO, A., CUENCA, S., GUARNER, F. & MANICHANH, C. 2014. Short-term effect of antibiotics on human gut microbiota. *PLoS One,* 9**,** e95476.

PATERSON, S., VOGWILL, T., BUCKLING, A., BENMAYOR, R., SPIERS, A. J., THOMSON, N. R., QUAIL, M., SMITH, F., WALKER, D., LIBBERTON, B., FENTON, A., HALL, N. & BROCKHURST, M. A. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature,* 464**,** 275-8.

PATTERSON, B. D. & ATMAR, W. 1986. Nested Subsets and the Structure of Insular Mammalian Faunas and Archipelagoes. *Biol J Linn Soc,* 28**,** 65-82.

PEARSON, M. M., LAURENCE, C. A., GUINN, S. E. & HANSEN, E. J. 2006. Biofilm formation by Moraxella catarrhalis in vitro: roles of the UspA1 adhesin and the Hag hemagglutinin. *Infect Immun,* 74**,** 1588-96.

PEARSON, W. R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol,* 183**,** 63-98.

PEIRIS, J. S., GUAN, Y. & YUEN, K. Y. 2004. Severe acute respiratory syndrome. *Nat Med,* 10**,** S88-97.

PEIRIS, J. S. M., LAI, S. T., POON, L. L. M., GUAN, Y., YAM, L. Y. C., LIM, W., NICHOLLS, J., YEE, W. K. S., YAN, W. W., CHEUNG, M. T., CHENG, V. C. C., CHAN, K. H., TSANG, D. N. C., YUNG, R. W. H., NG, T. K. & YUEN, K. Y. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet,* 361**,** 1319-25.

PELTOLA, V. T., MURTI, G. & MCCULLERS, J. A. 2005. Influenza Virus Neuraminidase Contributes to Secondary Bacterial Pneumonia. *J Infect Dis,* 192**,** 249-57.

PEREZ-TRALLERO, E., VICENTE, D., MONTES, M. & CISTERNA, R. 2002. Positive effect of meningococcal C vaccination on serogroup replacement in Neisseria meningitidis. *Lancet,* 360**,** 953.

PERRETT, K. P. & POLLARD, A. J. 2005. Towards an improved serogroup B Neisseria meningitidis vaccine. *Expert Opin Biol Ther,* 5**,** 1611-25.

PETERSON, S., CLINE, R. T., TETTELIN, H., SHAROV, V. & MORRISON, D. A. 2000. Gene expression analysis of the Streptococcus pneumoniae competence regulons by use of DNA microarrays. *J Bacteriol,* 182**,** 6192-202.

PETTIGREW, M. M., GENT, J. F., REVAI, K., PATEL, J. A. & CHONMAITREE, T. 2008. Microbial Interactions during Upper Respiratory Tract Infections. *Emerg Infect Dis,* 14**,** 1584-91.

PETTIGREW, M. M., LAUFER, A. S., GENT, J. F., KONG, Y., FENNIE, K. P. & METLAY, J. P. 2012. Upper Respiratory Tract Microbial Communities, Acute Otitis Media Pathogens, and Antibiotic Use in Healthy and Sick Children. *Appl Environ Microbiol,* 78**,** 6262-70.

PICCIRILLO, J. F. 2004. Clinical practice. Acute bacterial sinusitis. *New Engl J Med,* 351**,** 902-10.

PICHON, B., LADHANI, S. N., SLACK, M. P., SEGONDS-PICHON, A., ANDREWS, N. J., WAIGHT, P. A., MILLER, E. & GEORGE, R. 2013. Changes in Molecular Epidemiology of Streptococcus pneumoniae Causing Meningitis following Introduction of Pneumococcal Conjugate Vaccination in England and Wales. *J Clin Microbiol,* 51**,** 820-7.

PILISHVILI, T., LEXAU, C., FARLEY, M. M., HADLER, J., HARRISON, L. H., BENNETT, N. M., REINGOLD, A., THOMAS, A., SCHAFFNER, W., CRAIG, A. S., SMITH, P. J., BEALL, B. W., WHITNEY, C. G. & MOORE, M. R. 2010. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J Infect Dis,* 201**,** 32-41.

PITKARANTA, A., ROIVAINEN, M., BLOMGREN, K., PELTOLA, J., KAIJALAINEN, T., RATY, R., ZIEGLER, T., RONKKO, E., HATAKKA, K., KORPELA, R., POUSSA, T., LEINONEN, M. & HOVI, T. 2006. Presence of viral and bacterial pathogens in the nasopharynx of otitis-prone children. A prospective study. *Int J Pediatr Otorhinolaryngol,* 70**,** 647-54.

PLOTKIN, S. A. & PLOTKIN, S. L. 2011. The development of vaccines: how the past led to the future. *Nat Rev Microbiol,* 9**,** 889-93.

POOLMAN, J. T., BAKALETZ, L. O., CRIPPS, A., DENOEL, P. A., FORSGREN, A., KYD, J. & LOBET, Y. 2001. Developing a nontypeable Haemophilus influenzae (NTHi) vaccine. *Vaccine,* 19**,** S108-15.

POP, M. & SALZBERG, S. L. 2008. Bioinformatics challenges of new sequencing technology. *Trends Genet,* 24**,** 142-9.

PRYMULA, R., PEETERS, P., CHROBOK, V., KRIZ, P., NOVAKOVA, E., KALISKOVA, E., KOHL, I., LOMMEL, P., POOLMAN, J., PRIEELS, J.-P. & SCHUERMAN, L. 2006. Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both Streptococcus pneumoniae and non-typable Haemophilus influenzae: a randomised double-blind efficacy study. *Lancet,* 367**,** 740-8.

PUBLIC HEALTH ENGLAND 2012. Laboratory confirmed cases of pertussis reported to the enhanced pertussis surveillance programme (England and Wales): annual report 2011.

PUBLIC HEALTH ENGLAND 2014a. English surveillance programme for antimicrobial utilisation and resistance (ESPAUR).

PUBLIC HEALTH ENGLAND 2014b. Quarterly vaccination coverage statistics for children aged up to five years in the UK (COVER programme): April to June 2014.

PUBLIC HEALTH ENGLAND 2015a. Laboratory reports of Haemophilus influenzae infection by serotype and year England 1990 to 2014.

PUBLIC HEALTH ENGLAND 2015b. Pneumococcal disease infections caused by serotypes not in Prevenar 13 vaccine.

PUBLIC HEALTH ENGLAND 2015c. Vaccination coverage report, UK (COVER programme): October to December 2014.

PUIG, C., MARTI, S., FLEITES, A., TRABAZO, R., CALATAYUD, L., LINARES, J. & ARDANUY, C. 2014. Oropharyngeal Colonization by Nontypeable Haemophilus influenzae Among Healthy Children Attending Day Care Centers. *Microb Drug Resist,* 20**,** 450-5.

RABE, K. F. & WEDZICHA, J. A. 2011. Controversies in treatment of chronic obstructive pulmonary disease. *Lancet,* 378**,** 1038-47.

RAMON, G. 1923. Sur le pouvoir floculant et sur les propriétés immunisantes d'une toxine diphtérique rendue anatoxique (anatoxine). *C R Acad Sci Paris,* 177**,** 1338-40.

RAMSAY, M. E., ANDREWS, N., KACZMARSKI, E. B. & MILLER, E. 2001. Efficacy of meningococcal serogroup C conjugate vaccine in teenagers and toddlers in England. *Lancet,* 357**,** 195-6.

RAPOLA, S., SALO, E., KIISKI, P., LEINONEN, M. & TAKALA, A. K. 1997. Comparison of Four Different Sampling Methods for Detecting Pharyngeal Carriage of Streptococcus pneumoniae and Haemophilus influenzae in Children. *J Clin Microbiol,* 35**,** 1077-9.

READ, R. C., BAXTER, D., CHADWICK, D. R., FAUST, S. N., FINN, A., GORDON, S. B., HEATH, P. T., LEWIS, D. J., POLLARD, A. J., TURNER, D. P., BAZAZ, R., GANGULI, A., HAVELOCK, T., NEAL, K. R., OKIKE, I. O., MORALES-AZA, B., PATEL, K., SNAPE, M. D., WILLIAMS, J., GILCHRIST, S., GRAY, S. J., MAIDEN, M. C., TONEATTO, D., WANG, H., MCCARTHY, M., DULL, P. M. & BORROW, R. 2014. Effect of a quadrivalent meningococcal ACWY glycoconjugate or a serogroup B meningococcal vaccine on meningococcal carriage: an observer-blind, phase 3 randomised clinical trial. *Lancet,* Published online ahead of print.

REGEV-YOCHAY, G., BOGAERT, D., MALLEY, R., HERMANS, P. W. M., VEENHOVEN, R. H., SANDERS, E. A. M., LIPSITCH, M. & RUBINSTEIN, E. 2008. Does pneumococcal conjugate vaccine influence Staphylococcus aureus carriage in children? *Clin Infect Dis,* 47**,** 289-91.

REGEV-YOCHAY, G., DAGAN, R., RAZ, M., CARMELI, Y., SHAINBERG, B., DERAZNE, E., RAHAVI, G. & RUBINSTEIN, E. 2004a. Association between carriage of Streptococcus pneumoniae and Staphylococcus aureus in children. *JAMA,* 292**,** 716-20.

REGEV-YOCHAY, G., RAZ, M., DAGAN, R., PORAT, N., SHAINBERG, B., PINCO, E., KELLER, N. & RUBINSTEIN, E. 2004b. Nasopharyngeal carriage of Streptococcus pneumoniae by adults and children in community and family settings. *Clin Infect Dis,* 38**,** 632-9.

REGEV-YOCHAY, G., TRZCINSKI, K., THOMPSON, C. M., MALLEY, R. & LIPSITCH, M. 2006. Interference between Streptococcus pneumoniae and Staphylococcus aureus: In vitro hydrogen peroxide-mediated killing by Streptococcus pneumoniae. *J Bacteriol,* 188**,** 4996-5001.

REINERT, R. R., REINERT, S., VAN DER LINDEN, M., CIL, M. Y., AL-LAHHAM, A. & APPELBAUM, P. 2005. Antimicrobial susceptibility of Streptococcus pneumoniae in eight European countries from 2001 to 2003. *Antimicrob Agents Chemother,* 49**,** 2903-13.

REYNOLDS, C. A., FINKELSTEIN, J. A., RAY, G. T., MOORE, M. R. & HUANG, S. S. 2014. Attributable healthcare utilization and cost of pneumonia due to drug-resistant streptococcus pneumonia: a cost analysis. *Antimicrob Resist Infect Control,* 3**,** 1-8.

REYNOLDS, R. 2009. Antimicrobial resistance in the UK and Ireland. *J Antimicrob Chemother,* 64**,** 19-23.

REYNOLDS, R., POTZ, N., COLMAN, M., WILLIAMS, A., LIVERMORE, D., MACGOWAN, A. & SURVEILLANCE, B. E. W. P. O. B. R. 2004. Antimicrobial susceptibility of the pathogens of bacteraemia in the UK and Ireland 2001-2002: the BSAC Bacteraemia Resistance Surveillance Programme. *J Antimicrob Chemother,* 53**,** 1018-32.

RICHARDSON, E. J. & WATSON, M. 2013. The automatic annotation of bacterial genomes. *Brief Bioinform,* 14**,** 1-12.

RICKLEFS, R. E. 2004. A comprehensive framework for global patterns in biodiversity. *Ecol Lett,* 7**,** 1-15.

RIDGWAY, E. J., TREMLETT, C. H. & ALLEN, K. D. 1995. Capsular serotypes and antibiotic sensitivity of Streptococcus pneumoniae isolated from primary-school children. *J Infect,* 30**,** 245-51.

RILEY, S., FRASER, C., DONNELLY, C. A., GHANI, A. C., ABU-RADDAD, L. J., HEDLEY, A. J., LEUNG, G. M., HO, L. M., LAM, T. H., THACH, T. Q., CHAU, P., CHAN, K. P., LO, S. V., LEUNG, P. Y., TSANG, T., HO, W., LEE, K. H., LAU, E. M., FERGUSON, N. M. & ANDERSON, R. M. 2003. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science,* 300**,** 1961-6.

RINTA-KOKKO, H., DAGAN, R., GIVON-LAVI, N. & AURANEN, K. 2009. Estimation of vaccine efficacy against acquisition of pneumococcal carriage. *Vaccine,* 27**,** 3831-7.

RITTIRSCH, D., FLIERL, M. A. & WARD, P. A. 2008. Harmful molecular mechanisms in sepsis. *Nat Rev Immunol,* 8**,** 776-87.

ROBERTS, M. C. 2005. Update on acquired tetracycline resistance genes. *FEMS Microbiol Lett,* 245**,** 195-203.

RODRIGUES, F., FOSTER, D., NICOLI, E., TROTTER, C., VIPOND, B., MUIR, P., GONCALVES, G., JANUARIO, L. & FINN, A. 2013. Relationships Between Rhinitis Symptoms, Respiratory Viral Infections and Nasopharyngeal Colonization With Streptococcus pneumoniae, Haemophilus influenzae and Staphylococcus aureus in Children Attending Daycare. *Pediatr Infect Dis J,* 32**,** 227-232.

RODRIGUEZ-GIRONES, M. A. & SANTAMARIA, L. 2006. A new algorithm to calculate the nestedness temperature of presence-absence matrices. *J Biogeogr,* 33**,** 924-35.

ROGERS, G. B., CARROLL, M. P., SERISIER, D. J., HOCKEY, P. M., JONES, G. & BRUCE, K. D. 2004. Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16s ribosomal DNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol,* 42**,** 5176-83.

ROGERS, G. B., STRESSMANN, F. A., KOLLER, G., DANIELS, T., CARROLL, M. P. & BRUCE, K. D. 2008. Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. *Diagn Microbiol Infect Dis,* 62**,** 133-41.

ROGERS, G. B., ZAIN, N. M., BRUCE, K. D., BURR, L. D., CHEN, A. C., RIVETT, D. W., MCGUCKIN, M. A. & SERISIER, D. J. 2014. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc,* 11**,** 496-503.

ROHDE, K., WORTHEN, W. B., HEAP, M., HUGUENY, B. & GUEGAN, J. F. 1998. Nestedness in assemblages of metazoan ecto- and endoparasites of marine fish. *Int J Parasitol,* 28**,** 543-9.

RONAGHI, M., UHLEN, M. & NYREN, P. 1998. A sequencing method based on real-time pyrophosphate. *Science,* 281**,** 363-5.

ROSEN, J. B., THOMAS, A. R., LEXAU, C. A., REINGOLD, A., HADLER, J. L., HARRISON, L. H., BENNETT, N. M., SCHAFFNER, W., FARLEY, M. M., BEALL, B. W. & MOORE, M. R. 2011. Geographic variation in invasive pneumococcal disease following pneumococcal conjugate vaccine introduction in the United States. *Clin Infect Dis,* 53**,** 137-43.

RUST, A. G., MONGIN, E. & BIRNEY, E. 2002. Genome annotation techniques: new approaches and challenges. *Drug Discov Today,* 7**,** S70-6.

RUUSKANEN, O., AROLA, M., PUTTO-LAURILA, A., MERTSOLA, J., MEURMAN, O., VILJANEN, M. K. & HALONEN, P. 1989. Acute otitis media and respiratory virus infections. *Pediatr Infect Dis J,* 8**,** 94-9.

SA-LEADO, R., NUNES, S., BRITO-AVO, A., ALVES, C. R., CARRICO, J. A., SALDANHA, J., ALMEIDA, J. S., SANTOS-SANCHES, I. & DE LENCASTRE, H. 2008. High rates of transmission of and colonization by Streptococcus pneumoniae and Haemophilus influenzae within a day care center revealed in a longitudinal study. *J Clin Microbiol,* 46**,** 225-34.

SABAT, A. J., BUDIMIR, A., NASHEV, D., SA-LEAO, R., VAN DIJL, J., LAURENT, F., GRUNDMANN, H., FRIEDRICH, A. W. & MARKERS, E. S. G. O. E. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill,* 18**,** 20380.

SAEED, K., AHMAD, N., PALLETT, A., GUIVER, M. & MARSH, P. 2010. Specific staphylococcal polymerase chain reaction can be a complementary tool for identifying causative organisms and guiding antibiotic management in orthopaedic infections. *Curr Orthop Pract,* 21**,** 628-31.

SAEZ-LLORENS, X. & MCCRACKEN, G. H., JR. 2003. Bacterial meningitis in children. *Lancet,* 361**,** 2139-48.

SAKAI, F., TALEKAR, S. J., KLUGMAN, K. P., VIDAL, J. E. & GROUP, I. 2013. Expression of Streptococcus pneumoniae Virulence-Related Genes in the Nasopharynx of Healthy Children. *PLoS One,* 8**,** e67147.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SATZKE, C., TURNER, P., VIROLAINEN-JULKUNEN, A., ADRIAN, P. V., ANTONIO, M., HARE, K. M., HENAO-RESTREPO, A. M., LEACH, A. J., KLUGMAN, K. P., PORTER, B. D., SÁ-LEÃO, R., SCOTT, J. A., NOHYNEK, H. & O'BRIEN, K. L. 2013. Standard method for detecting upper respiratory carriage of Streptococcus pneumoniae: Updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. *Vaccine,* 32**,** 165-79.

SCHOENER, T. W. 1974. Resource partitioning in ecological communities. *Science,* 185**,** 27-39.

SCOTT, J. R., HANAGE, W. P., LIPSITCH, M., MILLAR, E. V., MOULTON, L. H., HINDS, J., REID, R., SANTOSHAM, M. & O'BRIEN, K. L. 2011. Pneumococcal sequence type replacement among American Indian children: A comparison of pre- and routine-PCV7 eras. *Vaccine,* 30**,** 2376-81.

SELANDER, R. K., CAUGANT, D. A., OCHMAN, H., MUSSER, J. M., GILMOUR, M. N. & WHITTAM, T. S. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol,* 51**,** 873-84.

SERRUTO, D., BOTTOMLEY, M. J., RAM, S., GIULIANI, M. M. & RAPPUOLI, R. 2012. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. *Vaccine,* 30 Suppl 2**,** B87-97.

SETHI, S. 2010. Infection as a comorbidity of COPD. *Eur Respir J,* 35**,** 1209-15.

SHAMAN, J. & KOHN, M. 2009. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc Natl Acad Sci U S A,* 106**,** 3243-8.

SHANNON, K. E., LEE, D. Y., TREVORS, J. T. & BEAUDETTE, L. A. 2007. Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Sci Total Environ,* 382**,** 121-9.

SHAPIRO, E. D., BERG, A. T., AUSTRIAN, R., SCHROEDER, D., PARCELLS, V., MARGOLIS, A., ADAIR, R. K. & CLEMENS, J. D. 1991. The protective efficacy of polyvalent pneumococcal polysaccharide vaccine. *New Engl J Med,* 325**,** 1453-60.

SHAW, K. J., RATHER, P. N., HARE, R. S. & MILLER, G. H. 1993. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev,* 57**,** 138-63.

SHIVELY, R. G., SHIGEI, J. T., PETERSON, E. M. & DE LA MAZA, L. M. 1981. Typing of Haemophilus influenzae by coagglutination and conventional slide agglutination. *J Clin Microbiol,* 14**,** 706-8.

SIEGRIST, C.-A. 2008. Vaccine Immunology. *In:* PLOTKIN, S., ORENSTEIN, W. & OFFIT, P. (eds.) *Vaccines.* 5 ed. Philadelphia, PA: Saunders Elsevier.

SIMELL, B., AURANEN, K., KAYHTY, H., GOLDBLATT, D., DAGAN, R., O'BRIEN, K. L. & GROUP, P. C. 2012. The fundamental link between pneumococcal carriage and disease. *Expert Rev Vaccines,* 11**,** 841-55.

SIMPSON, J. T., WONG, K., JACKMAN, S. D., SCHEIN, J. E., JONES, S. J. M. & BIROL, I. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res,* 19**,** 1117-23.

SINGH, J., BEHAL, A., SINGLA, N., JOSHI, A., BIRBIAN, N., SINGH, S., BALI, V. & BATRA, N. 2009. Metagenomics: Concept, methodology, ecological inference and recent advances. *Biotechnol J,* 4**,** 480-94.

SINGLETON, R. J., HENNESSY, T. W., BULKOW, L. R., HAMMITT, L. L., ZULZ, T., HURLBURT, D. A., BUTLER, J. C., RUDOLPH, K. & PARKINSON, A. 2007. Invasive pneumococcal disease caused by nonvaccine serotypes among Alaska native children with high levels of 7-valent pneumococcal conjugate vaccine coverage. *JAMA,* 297**,** 1784-92.

SKAARE, D., ANTHONISEN, I. L., CAUGANT, D. A., JENKINS, A., STEINBAKK, M., STRAND, L., SUNDSFJORD, A., TVETEN, Y. & KRISTIANSEN, B. E. 2014. Multilocus sequence typing and ftsI sequencing: a powerful tool for surveillance of penicillin-binding

protein 3-mediated beta-lactam resistance in nontypeable Haemophilus influenzae. *BMC Microbiol,* 14**,** 131.

SKINNER, J. M., INDRAWATI, L., CANNON, J., BLUE, J., WINTERS, M., MACNAIR, J., PUJAR, N., MANGER, W., ZHANG, Y., ANTONELLO, J., SHIVER, J., CAULFIELD, M. & HEINRICHS, J. H. 2011. Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model. *Vaccine,* 29**,** 8870-6.

SKOV, T., DEDDENS, J., PETERSEN, M. R. & ENDAHL, L. 1998. Prevalence proportion ratios: estimation and hypothesis testing. *Int J Epidemiol,* 27**,** 91-5.

SLEEMAN, K. L., GRIFFITHS, D., SHACKLEY, F., DIGGLE, L., GUPTA, S., MAIDEN, M. C., MOXON, E. R., CROOK, D. W. & PETO, T. E. 2006. Capsular serotype-specific attack rates and duration of carriage of Streptococcus pneumoniae in a population of children. *J Infect Dis,* 194**,** 682-8.

SMIEJA, M., CASTRICIANO, S., CARRUTHERS, S., SO, G., CHONG, S., LUINSTRA, K., MAHONY, J. B., PETRICH, A., CHERNESKY, M., SAVARESE, M. & TRIVA, D. 2010. Development and evaluation of a flocked nasal midturbinate swab for self-collection in respiratory virus infection diagnostic testing. *J Clin Microbiol,* 48**,** 3340-2.

SMIT, P., OBERHOLZER, D., HAYDEN-SMITH, S., KOORNHOF, H. J. & HILLEMAN, M. R. 1977. Protective efficacy of pneumococcal polysaccharide vaccines. *JAMA,* 238**,** 2613-6.

SMITH, D. H., PETER, G., INGRAM, D. L., HARDING, A. L. & ANDERSON, P. 1973. Responses of children immunized with the capsular polysaccharide of Hemophilus influenzae, type b. *Pediatrics,* 52**,** 637-44.

SMITH, G. J., VIJAYKRISHNA, D., BAHL, J., LYCETT, S. J., WOROBEY, M., PYBUS, O. G., MA, S. K., CHEUNG, C. L., RAGHWANI, J., BHATT, S., PEIRIS, J. S., GUAN, Y. & RAMBAUT, A. 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature,* 459**,** 1122-5.

SMITH, M. J. 2008. Meningococcal tetravalent conjugate vaccine. *Expert Opin Biol Ther,* 8**,** 1941-6.

SOUTHWOOD, T. R. E. 1961. The Number of Species of Insect Associated with Various Trees. *J Anim Ecol,* 30**,** 1-8.

SPERBER, S. J. & FRANCIS, J. B. 1987. Toxic shock syndrome during an influenza outbreak. *JAMA,* 257**,** 1086-7.

SPINOLA, S. M., PEACOCK, J., DENNY, F. W., SMITH, D. L. & CANNON, J. G. 1986. Epidemiology of colonization by nontypable Haemophilus influenzae in children: a longitudinal study. *J Infect Dis,* 154**,** 100-9.

SPOONER, L. H. S., L.H.; HEATH, E.H. 1919. A bacteriologic study of the influenza epidemic at Camp Devens, Mass. *JAMA,* 72**,** 155-59.

SPRENT, P. 2003. Statistics in medical research. *Swiss Med Wkly,* 133**,** 522-9.

STANWELL-SMITH, R. E., STUART, J. M., HUGHES, A. O., ROBINSON, P., GRIFFIN, M. B. & CARTWRIGHT, K. 1994. Smoking, the environment and meningococcal disease: a case control study. *Epidemiol Infect,* 112**,** 315-28.

STEPHENS, D. S., GREENWOOD, B. & BRANDTZAEG, P. 2007. Epidemic meningitis, meningococcaemia, and Neisseria meningitidis. *Lancet,* 369**,** 2196-210.

STEWART, P. S. & COSTERTON, J. W. 2001. Antibiotic resistance of bacteria in biofilms. *Lancet,* 358**,** 135-8.

STEWART, P. S. & FRANKLIN, M. J. 2008. Physiological heterogeneity in biofilms. *Nat Rev Microbiol,* 6**,** 199-210.

STROMBERG, U. 1995. Prevalence odds ratio v prevalence ratio--some further comments. *Occup Environ Med,* 52**,** 143.

STURM, A. W., MOSTERT, R., ROUING, P. J., VAN KLINGEREN, B. & VAN ALPHEN, L. 1990. Outbreak of multiresistant non-encapsulated Haemophilus influenzae infections in a pulmonary rehabilitation centre. *Lancet,* 335**,** 214-6.

SUTTLE, C. A. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol,* 5**,** 801-12.

SYKES, A. & JOHNSTON, S. L. 2008. Etiology of asthma exacerbations. *J Allergy Clin Immunol,* 122**,** 685-8.

SYRJANEN, R. K., KILPI, T. M., KAIJALAINEN, T. H., HERVA, E. E. & TAKALA, A. K. 2001. Nasopharyngeal carriage of Streptococcus pneumoniae in Finnish children younger than 2 years old. *J Infect Dis,* 184**,** 451-9.

SYROGIANNOPOULOS, G. A., DOIT, C., GRIVEA, I. N., GESLIN, P. & BINGEN, E. 2001. Clonal relationships among penicillin-susceptible, multiresistant serotype 6B Streptococcus pneumoniae isolates recovered in Greece and France. *Eur J Clin Microbiol Infect Dis,* 20**,** 61-4.

TAI, S. S. 2006. Streptococcus pneumoniae protein vaccine candidates: properties, activities and animal studies. *Crit Rev Microbiol,* 32**,** 139-53.

TALBIRD, S. E., ISMAILA, A. S. & TAYLOR, T. N. 2010. A steady-state, population-based model to estimate the direct and indirect effects of pneumococcal vaccines. *Vaccine,* 28**,** G3-13.

TALKINGTON, D. F., BROWN, B. G., THARPE, J. A., KOENIG, A. & RUSSELL, H. 1996. Protection of mice against fatal pneumococcal challenge by immunization with pneumococcal surface adhesin A (PsaA). *Microb Pathog,* 21**,** 17-22.

TANNER, H., BOXALL, E. & OSMAN, H. 2012. Respiratory viral infections during the 2009-2010 winter season in Central England, UK: incidence and patterns of multiple virus co-infections. *Eur J Clin Microbiol Infect Dis,* 31**,** 3001-6.

TANO, K., GRAHN-HAKANSSON, E., HOLM, S. E. & HELLSTROM, S. 2000. Inhibition of OM pathogens by alpha-hemolytic streptococci from healthy children, children with SOM and children with rAOM. *Int J Pediatr Otorhinolaryngol,* 56**,** 185-90.

TASHER, D., STEIN, M., SIMOES, E. A., SHOHAT, T., BROMBERG, M. & SOMEKH, E. 2011. Invasive Bacterial Infections in Relation to Influenza Outbreaks, 2006-2010. *Clin Infect Dis*.

TENOVER, F. C., MCDOUGAL, L. K., GOERING, R. V., KILLGORE, G., PROJAN, S. J., PATEL, J. B. & DUNMAN, P. M. 2006. Characterization of a strain of community-associated methicillin-resistant Staphylococcus aureus widely disseminated in the United States. *J Clin Microbiol,* 44**,** 108-18.

THOMPSON, M. L., MYERS, J. E. & KRIEBEL, D. 1998. Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done! *Occup Environ Med,* 55**,** 272-7.

TIRADO, R. & PUGNAIRE, F. I. 2005. Community structure and positive interactions in constraining environments. *Oikos,* 111**,** 437-444.

TOCHEVA, A. S., JEFFERIES, J. M., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2010. Increase in serotype 6C pneumococcal carriage, United Kingdom. *Emerg Infect Dis,* 16**,** 154-5.

TOCHEVA, A. S., JEFFERIES, J. M., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2013. Distribution of carried pneumococcal clones in UK children following the introduction of the 7-valent pneumococcal conjugate vaccine: A 3-year cross-sectional population based analysis. *Vaccine,* 31**,** 3187-90.

TOCHEVA, A. S., JEFFERIES, J. M., RUBERY, H., BENNETT, J., AFIMEKE, G., GARLAND, J., CHRISTODOULIDES, M., FAUST, S. N. & CLARKE, S. C. 2011. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of Streptococcus pneumoniae in young children. *Vaccine,* 29**,** 4400-4.

TOMITA, Y., OKAMOTO, A., YAMADA, K., YAGI, T., HASEGAWA, Y. & OHTA, M. 2011. A new microarray system to detect Streptococcus pneumoniae serotypes. *J Biomed Biotechnol,* 2011**,** 352736.

TRAISSAC, P., MARTIN-PREVEL, Y., DELPEUCH, F. & MAIRE, B. 1999. Régression logistique vs autres modèles linéaires généralisés pour l'estimation de rapports de prévalences. *Rev Epidemiol Sante Publique,* 47**,** 593-604.

TROTTER, C. L., ANDREWS, N. J., KACZMARSKI, E. B., MILLER, E. & RAMSAY, M. E. 2004. Effectiveness of meningococcal serogroup C conjugate vaccine 4 years after introduction. *Lancet,* 364**,** 365-7.

TROTTER, C. L. & GAY, N. J. 2003. Analysis of longitudinal bacterial carriage studies accounting for sensitivity of swabbing: an application to Neisseria meningitidis. *Epidemiol Infect,* 130**,** 201-5.

TROTTER, C. L. & GREENWOOD, B. M. 2007. Meningococcal carriage in the African meningitis belt. *Lancet Infect Dis,* 7**,** 797-803.

TURK, D. C. 1984. The pathogenicity of Haemophilus influenzae. *J Med Microbiol,* 18**,** 1-16.

TURNER, P., HINDS, J., TURNER, C., JANKHOT, A., GOULD, K., BENTLEY, S. D., NOSTEN, F. & GOLDBLATT, D. 2011a. Improved detection of nasopharyngeal

cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. *J Clin Microbiol,* 49**,** 1784-9.

TURNER, P., PO, L., TURNER, C., GOLDBLATT, D. & NOSTEN, F. 2011b. Detection of respiratory viruses by PCR assay of nasopharyngeal swabs stored in skim milk-tryptone-glucose-glycerol transport medium. *J Clin Microbiol,* 49**,** 2311-3.

TUTTLE, M. S., MOSTOW, E., MUKHERJEE, P., HU, F. Z., MELTON-KREFT, R., EHRLICH, G. D., DOWD, S. E. & GHANNOUM, M. A. 2011. Characterization of bacterial communities in venous insufficiency wounds by use of conventional culture and molecular diagnostic methods. *J Clin Microbiol,* 49**,** 3812-9.

UBUKATA, K., SHIBASAKI, Y., YAMAMOTO, K., CHIBA, N., HASEGAWA, K., TAKEUCHI, Y., SUNAKAWA, K., INOUE, M. & KONNO, M. 2001. Association of amino acid substitutions in penicillin-binding protein 3 with beta-lactam resistance in beta-lactamase-negative ampicillin-resistant Haemophilus influenzae. *Antimicrob Agents Chemother,* 45**,** 1693-9.

UEMURA, E., KAKINOHANA, S., HIGA, N., TOMA, C. & NAKASONE, N. 2004. Comparative Characterization of Staphylococcus aureus Isolates from Throats and Noses of Healthy Volunteers. *Jpn J Infect Dis,* 57**,** 21-4.

UEYAMA, T., KURONO, Y., SHIRABE, K., TAKESHITA, M. & MOGI, G. 1995. High-Incidence of Haemophilus-Influenzae in Nasopharyngeal Secretions and Middle-Ear Effusions as Detected by Pcr. *J Clin Microbiol,* 33**,** 1835-8.

ULRICH, W. & ALMEIDA-NETO, M. 2012. On the meanings of nestedness: back to the basics. *Ecography,* 35**,** 865-71.

ULRICH, W., ALMEIDA, M. & GOTELLI, N. J. 2009. A consumer's guide to nestedness analysis. *Oikos,* 118**,** 3-17.

ULRICH, W. & ZALEWSKI, M. 2007. Are ground beetles neutral? *Basic Appl Ecol,* 8**,** 411-20.

UNSON, C. G., OHANNESSIAN, C., KENYON, L., CASE, A., REISINE, S. & PRESTWOOD, K. 2004. Barriers to eligibility and enrollment among older women in a clinical trial on osteoporosis: Effects of ethnicity and SES. *J Aging Health,* 16**,** 426-43.

VAN BENTEN, I., KOOPMAN, L., NIESTERS, B., HOP, W., VAN MIDDELKOOP, B., DE WAAL, L., VAN DRUNEN, K., OSTERHAUS, A., NEIJENS, H. & FOKKENS, W. 2003. Predominance of rhinovirus in the nose of symptomatic and asymptomatic infants. *Pediatr Allergy Immunol,* 14**,** 363-70.

VAN DE POL, A. C., VAN LOON, A. M., WOLFS, T. F., JANSEN, N. J., NIJHUIS, M., BRETELER, E. K., SCHUURMAN, R. & ROSSEN, J. W. 2007. Increased detection of respiratory syncytial virus, influenza viruses, parainfluenza viruses, and adenoviruses with real-time PCR in samples from patients with respiratory symptoms. *J Clin Microbiol,* 45**,** 2260-2.

VAN DEN BERGH, M. R., BIESBROEK, G., ROSSEN, J. W. A., PITERS, W. A. A. D., BOSCH, A. A. T. M., VAN GILS, E. J. M., WANG, X. H., BOONACKER, C. W. B., VEENHOVEN, R. H., BRUIN, J. P., BOGAERT, D. & SANDERS, E. A. M. 2012. Associations between Pathogens in the Upper Respiratory Tract of Young Children: Interplay between Viruses and Bacteria. *PLoS One,* 7**,** e47711.

VAN DER GAST, C. J., WALKER, A. W., STRESSMANN, F. A., ROGERS, G. B., SCOTT, P., DANIELS, T. W., CARROLL, M. P., PARKHILL, J. & BRUCE, K. D. 2011. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J,* 5**,** 780-91.

VAN DORT, M., WALDEN, C., WALKER, E. S., REYNOLDS, S. A., LEVY, F. & SARUBBI, F. A. 2007. An outbreak of infections caused by non-typeable Haemophilus influenzae in an extended care facility. *J Hosp Infect,* 66**,** 59-64.

VAN ELDEN, L. J. R., NIJHUIS, M., SCHIPPER, P., SCHUURMAN, R. & VAN LOON, A. M. 2001. Simultaneous detection of influenza viruses A and B using real-time quantitative PCR. *J Clin Microbiol,* 39**,** 196-200.

VAN GILS, E. J., HAK, E., VEENHOVEN, R. H., RODENBURG, G. D., BOGAERT, D., BRUIN, J. P., VAN ALPHEN, L. & SANDERS, E. A. 2011a. Effect of seven-valent pneumococcal conjugate vaccine on Staphylococcus aureus colonisation in a randomised controlled trial. *PLoS One,* 6**,** e20229.

VAN GILS, E. J., VEENHOVEN, R. H., RODENBURG, G. D., HAK, E. & SANDERS, E. A. 2011b. Effect of 7-valent pneumococcal conjugate vaccine on nasopharyngeal carriage with Haemophilus influenzae and Moraxella catarrhalis in a randomized controlled trial. *Vaccine,* 29**,** 7595-8.

VANEECHOUTTE, M., VERSCHRAEGEN, G., CLAEYS, G., WEISE, B. & VAN DEN ABEELE, A. M. 1990. Respiratory tract carrier rates of Moraxella (Branhamella) catarrhalis in adults and children and interpretation of the isolation of M. catarrhalis from sputum. *J Clin Microbiol,* 28**,** 2674-80.

VARON, E., LEVY, C., DE LA ROCQUE, F., BOUCHERAT, M., DEFORCHE, D., PODGLAJEN, I., NAVEL, M. & COHEN, R. 2000. Impact of antimicrobial therapy on nasopharyngeal carriage of Streptococcus pneumoniae, Haemophilus influenzae, and Branhamella catarrhalis in children with respiratory tract infections. *Clin Infect Dis,* 31**,** 477-81.

VERGISON, A. 2008. Microbiology of otitis media: a moving target. *Vaccine,* 26**,** G5-10.

VERHAEGH, S. J., SNIPPE, M. L., LEVY, F., VERBRUGH, H. A., JADDOE, V. W., HOFMAN, A., MOLL, H. A., VAN BELKUM, A. & HAYS, J. P. 2011. Colonization of healthy children by Moraxella catarrhalis is characterized by genotype heterogeneity, virulence gene diversity and co-colonization with Haemophilus influenzae. *Microbiology,* 157**,** 169-78.

VIVES, M., GARCIA, M. E., SAENZ, P., MORA, M. A., MATA, L., SABHARWAL, H. & SVANBORG, C. 1997. Nasopharyngeal colonization in Costa Rican children during the first year of life. *Pediatr Infect Dis J,* 16**,** 852-8.

VON ELM, E., ALTMAN, D. G., EGGER, M., POCOCK, S. J., GOTZSCHE, P. C., VANDENBROUCKE, J. P. & INITIATIVE, S. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology,* 18**,** 800-4.

VON KÖNIG, C. H., HALPERIN, S., RIFFELMANN, M. & GUISO, N. 2002. Pertussis of adults and infants. *Lancet Infect Dis,* 2**,** 744-50.

VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J., KUIPER, M. & ET AL. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res,* 23**,** 4407-14.

WAHDAN, M. H., RIZK, F., EL-AKKAD, A. M., EL-GHOROURY, A. A., HABLAS, R., GIRGIS, N. I., AMER, A., BOCTAR, W., SIPPEL, J. E., GOTSCHLICH, E. C., TRIAU, R., SANBORN, W. R. & CVJETANOVIC, B. 1973. A controlled field trial of a serogroup A meningococcal polysaccharide vaccine. *Bull World Health Organ,* 48**,** 667-73.

WARREN, R. L., SUTTON, G. G., JONES, S. J. M. & HOLT, R. A. 2007. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics,* 23**,** 500-1.

WATT, J. P., O'BRIEN, K. L., KATZ, S., BRONSDON, M. A., ELLIOTT, J., DALLAS, J., PERILLA, M. J., REID, R., MURROW, L., FACKLAM, R., SANTOSHAM, M. & WHITNEY, C. G. 2004. Nasopharyngeal versus oropharyngeal sampling for detection of pneumococcal carriage in adults. *J Clin Microbiol,* 42**,** 4974-6.

WEBB, D. C. & CRIPPS, A. W. 1999. Immunization with recombinant transferrin binding protein B enhances clearance of nontypeable Haemophilus influenzae from the rat lung. *Infect Immun,* 67**,** 2138-44.

WEBB, J. S., LAU, M. & KJELLEBERG, S. 2004. Bacteriophage and phenotypic variation in Pseudomonas aeruginosa biofilm development. *J Bacteriol,* 186**,** 8066-73.

WEIMER, K. E., ARMBRUSTER, C. E., JUNEAU, R. A., HONG, W., PANG, B. & SWORDS, W. E. 2010. Coinfection with Haemophilus influenzae promotes pneumococcal biofilm formation during experimental otitis media and impedes the progression of pneumococcal disease. *J Infect Dis,* 202**,** 1068-75.

WEINBERG, G. A. & SZILAGYI, P. G. 2010. Vaccine epidemiology: efficacy, effectiveness, and the translational research roadmap. *J Infect Dis,* 201**,** 1607-10.

WEINBERGER, D. M., MALLEY, R. & LIPSITCH, M. 2011. Serotype replacement in disease after pneumococcal vaccination. *Lancet,* 378**,** 1962-73.

WEINBERGER, D. M., TRZCINSKI, K., LU, Y. J., BOGAERT, D., BRANDES, A., GALAGAN, J., ANDERSON, P. W., MALLEY, R. & LIPSITCH, M. 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog,* 5**,** e1000476.

WEISER, J. N., AUSTRIAN, R., SREENIVASAN, P. K. & MASURE, H. R. 1994. Phase Variation in Pneumococcal Opacity - Relationship between Colonial Morphology and Nasopharyngeal Colonization. *Infect Immun,* 62**,** 2582-9.

WEISS-SALZ, I. & YAGUPSKY, P. 2010. 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,* 4**,** 11-5.

WHATMORE, A. M. & DOWSON, C. G. 1999. The autolysin-encoding gene (lytA) of Streptococcus pneumoniae displays restricted allelic variation despite localized recombination events with genes of pneumococcal bacteriophage encoding cell wall lytic enzymes. *Infect Immun,* 67**,** 4551-6.

WHELEN, A. C., FELMLEE, T. A., HUNT, J. M., WILLIAMS, D. L., ROBERTS, G. D., STOCKMAN, L. & PERSING, D. H. 1995. Direct Genotypic Detection of Mycobacterium

Tuberculosis Rifampin Resistance in Clinical Specimens by Using Single-Tube Heminested PCR. *J Clin Microbiol,* 33**,** 556-61.

WHELEN, A. C. & PERSING, D. H. 1996. The role of nucleic acid amplification and detection in the clinical microbiology laboratory. *Annu Rev Microbiol,* 50**,** 349-73.

WHITE, B., DIGGLE, M., TODD, A., DUNDAS, S. & INVERARITY, D. 2011. A novel pneumococcus with a new association. *Travel Med Infect Dis,* 9**,** 84-7.

WHITTAKER, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol Monogr,* 30**,** 279-338.

WHO 2010. Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 Global Report on Surveillance and Response. Geneva, Switzerland.

WHO 2011. Global Tuberculosis Control 2011. Geneva, Switzerland.

WHO 2012. Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2012. Geneva, Switzerland.

WHO & UNICEF 2006. Pneumonia The Forgotten Killer of Children. Geneva, Switzerland.

WIERTSEMA, S. P., CHIDLOW, G. R., KIRKHAM, L. A., CORSCADDEN, K. J., MOWE, E. N., VIJAYASEKARAN, S., COATES, H. L., HARNETT, G. B. & RICHMOND, P. C. 2011a. High detection rates of nucleic acids of a wide range of respiratory viruses in the nasopharynx and the middle ear of children with a history of recurrent acute otitis media. *J Med Virol,* 83**,** 2008-17.

WIERTSEMA, S. P., KIRKHAM, L. A., CORSCADDEN, K. J., MOWE, E. N., BOWMAN, J. M., JACOBY, P., FRANCIS, R., VIJAYASEKARAN, S., COATES, H. L., RILEY, T. V. & RICHMOND, P. 2011b. Predominance of nontypeable Haemophilus influenzae in children with otitis media following introduction of a 3+0 pneumococcal conjugate vaccine schedule. *Vaccine,* 29**,** 5163-70.

WILKINSON, T. M. A., HURST, J. R., PERERA, W. R., WILKS, M., DONALDSON, G. C. & WEDZICHA, J. A. 2006. Effect of Interactions Between Lower Airway Bacterial and Rhinoviral Infection in Exacerbations of COPD. *Chest,* 129**,** 317-24.

WILLETTS, N. & WILKINS, B. 1984. Processing of plasmid DNA during bacterial conjugation. *Microbiol Rev,* 48**,** 24-41.

WILLIAMS, R., PEISAJOVICH, S. G., MILLER, O. J., MAGDASSI, S., TAWFIK, D. S. & GRIFFITHS, A. D. 2006. Amplification of complex gene libraries by emulsion PCR. *Nat Methods,* 3**,** 545-50.

WILLIAMSON, J., MARMION, B. P., WORSWICK, D. A., KOK, T. W., TANNOCK, G., HERD, R. & HARRIS, R. J. 1992. Laboratory diagnosis of Mycoplasma pneumoniae infection. 4. Antigen capture and PCR-gene amplification for detection of the Mycoplasma: problems of clinical correlation. *Epidemiol Infect,* 109**,** 519-37.

WINTER, K., HARRIMAN, K., ZIPPRICH, J., SCHECHTER, R., TALARICO, J., WATT, J. & CHAVEZ, G. 2012. California pertussis epidemic, 2010. *J Pediatr,* 161**,** 1091-6.

WOLF, J. & DALEY, A. J. 2007a. Microbiological aspects of bacterial lower respiratory tract illness in children: atypical pathogens. *Paediatr Respir Rev,* 8**,** 212-9.

WOLF, J. & DALEY, A. J. 2007b. Microbiological aspects of bacterial lower respiratory tract illness in children: typical pathogens. *Paediatr Respir Rev,* 8**,** 204-10.

WOO, P. C., CHIU, S. S., SETO, W. H. & PEIRIS, M. 1997. Cost-effectiveness of rapid diagnosis of viral respiratory tract infections in pediatric patients. *J Clin Microbiol,* 35**,** 1579-81.

WOO, P. C. Y., LAU, S. K. P., TENG, J. L. L., TSE, H. & YUEN, K.-Y. 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infec,* 14**,** 908-34.

WU, H. M., CORDEIRO, S. M., HARCOURT, B. H., CARVALHO, M., AZEVEDO, J., OLIVEIRA, T. Q., LEITE, M. C., SALGADO, K., REIS, M. G., PLIKAYTIS, B. D., CLARK, T. A., MAYER, L. W., KO, A. I., MARTIN, S. W. & REIS, J. N. 2013a. Accuracy of real-time PCR, Gram stain and culture for Streptococcus pneumoniae, Neisseria meningitidis and Haemophilus influenzae meningitis diagnosis. *BMC Infect Dis,* 13**,** 1-10.

WU, L. T., THOMAS, I., CURRAN, M. D., ELLIS, J. S., PARMAR, S., GOEL, N., SHARMA, P. I., ALLAIN, J. P. & LEE, H. H. 2013b. Duplex molecular assay intended for point-of-care diagnosis of influenza A/B virus infection. *J Clin Microbiol,* 51**,** 3031-8.

WYLLIE, A. L., CHU, M. L., SCHELLENS, M. H., VAN ENGELSDORP GASTELAARS, J., JANSEN, M. D., VAN DER ENDE, A., BOGAERT, D., SANDERS, E. A. & TRZCINSKI, K. 2014. Streptococcus pneumoniae in saliva of Dutch primary school children. *PLoS One,* 9**,** e102045.

WYSOCKI, J., TEJEDOR, J. C., GRUNERT, D., KONIOR, R., GARCIA-SICILIA, J., KNUF, M., BERNARD, L., DIEUSSAERT, I. & SCHUERMAN, L. 2009. Immunogenicity of the 10-valent pneumococcal non-typeable Haemophilus influenzae protein D conjugate vaccine (PHiD-CV) when coadministered with different neisseria meningitidis serogroup C conjugate vaccines. *Pediatr Infect Dis J,* 28**,** S77-88.

YANG, S. & ROTHMAN, R. E. 2004. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect Dis,* 4**,** 337-48.

YOUNG, B. C., GOLUBCHIK, T., BATTY, E. M., FUNG, R., LARNER-SVENSSON, H., VOTINTSEVA, A. A., MILLER, R. R., GODWIN, H., KNOX, K., EVERITT, R. G., IQBAL, Z., RIMMER, A. J., CULE, M., IP, C. L., DIDELOT, X., HARDING, R. M., DONNELLY, P., PETO, T. E., CROOK, D. W., BOWDEN, R. & WILSON, D. J. 2012. Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease. *Proc Natl Acad Sci U S A,* 109**,** 4550-5.

YU, J. & COOPER, H. 1983. A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *J Marketing Res,* 20**,** 36-44.

YUAN, Z. Q., LV, Z. Y., GAN, H. Q., XIAN, M., ZHANG, K. X., MAI, J. Y., YU, X. B. & WU, Z. D. 2011. Intranasal immunization with autolysin (LytA) in mice model induced protection against five prevalent Streptococcus pneumoniae serotypes in China. *Immunol Res,* 51**,** 108-15.

ZAKI, A. M., VAN BOHEEMEN, S., BESTEBROER, T. M., OSTERHAUS, A. D. & FOUCHIER, R. A. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New Engl J Med,* 367**,** 1814-20.

ZANKARI, E., HASMAN, H., COSENTINO, S., VESTERGAARD, M., RASMUSSEN, S., LUND, O., AARESTRUP, F. M. & LARSEN, M. V. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother,* 67**,** 2640-4.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res,* 18**,** 821-9.

ZHANG, Q., ARNAOUTAKIS, K., MURDOCH, C., LAKSHMAN, R., RACE, G., BURKINSHAW, R. & FINN, A. 2004. Mucosal immune responses to capsular pneumococcal polysaccharides in immunized preschool children and controls with similar nasal pneumococcal colonization rates. *Pediatr Infect Dis J,* 23**,** 307-13.

ZHANG, Q., PETTITT, E., BURKINSHAW, R., RACE, G., SHAW, L. & FINN, A. 2002. Mucosal immune responses to meningococcal conjugate polysaccharide vaccines in infants. *Pediatr Infect Dis J,* 21**,** 209-13.

ZHANG, W., CHEN, J., YANG, Y., TANG, Y., SHANG, J. & SHEN, B. 2011. A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. *PLoS One,* 6**,** e17915.

ZHOU, H., HABER, M., RAY, S., FARLEY, M. M., PANOZZO, C. A. & KLUGMAN, K. P. 2012. Invasive Pneumococcal Pneumonia and Respiratory Virus Co-infections. *Emerg Infect Dis,* 18**,** 294-7.

ZHOU, X., BENT, S. J., SCHNEIDER, M. G., DAVIS, C. C., ISLAM, M. R. & FORNEY, L. J. 2004. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology,* 150**,** 2565-73.