University of Southampton Research Repository

UNIVERSITY OF SOUTHAMPTON
FACULTY OF MEDICINE
CLINICAL AND EXPERIMENTAL SCIENCES
VOLUME 1 OF 1

Characterisation of *Haemophilus influenzae* and *Haemophilus haemolyticus* in Chronic Obstructive Pulmonary Disease (COPD)

by

Karen Osman

Thesis for the degree of Doctor of Philosophy

March 2017

University of Southampton

ABSTRACT

Faculty of Medicine, Clinical and Experimental Sciences

Doctor of Philosophy

Characterisation of *Haemophilus influenzae* and *Haemophilus haemolyticus* inChronic Obstructive Pulmonary Disease (COPD) by Karen Cox

Chronic obstructive pulmonary disease (COPD) accounted for 5.71% of all global mortality in 2015. Non-typeable *Haemophilus influenzae* (NTHi) is a genetically diverse opportunistic pathogen that colonises the respiratory tract and is linked to acute exacerbations in COPD which eventually lead to a fatal irreversible degradation of lung function.

H. haemolyticus is the closest phylogenetic relative to NTHi and is also found to colonise the respiratory tract of subjects with COPD. However, *H. haemolyticus* is not thought to cause either invasive disease or exacerbations in COPD. The motivation for the work in this thesis is that there are currently no molecular or culture assays that can unequivocally differentiate these two species. The genetic diversity of NTHi has been reported. The objective of this study was to determine whether the genetic variation in atypical strains of the two species was sufficiently large to result in a genetic continuum that makes their delineation impossible and therefore questioning the current taxonomic classification.

Whole genome sequencing data generated from 1460 isolates phenotypically identified as *Haemophilus* spp. were used to determine the genomic heterogeneity of NTHi and *H. haemolyticus*. This included assessing the utility of gene markers for example, *fucK*, *smpB*, *iga*, *sodC*, *lgtC*, *hpd*, *omp2* and *omp6*, which have previously been shown to enable differentiation, as well as more extensive comparative genomic analyses. These data revealed the presence of previously poorly recognised atypical strains of NTHi and *H. haemolyticus.* Atypical isolates were initially characterised as those with an unexpected genotype for NTHi or *H. haemolyticus* as defined by previously described markers. The use of further adhesin genes and comparative genomics supported these findings. Comparison of NTHi and *H. haemolyticus* demonstrated distinctive clustering based on ANI with <92% similarity; a value below current taxonomic thresholds for species definition. In conclusion, despite extensive genomic variation in strains of NTHi they may be routinely separated from *H. haemolyticus* using whole genome analysis and multiple gene markers.

Table of Contents

Table of figures

Table of tables

Academic Thesis: Declaration of Authorship

I, Karen Cox, 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.

Characterisation of *Haemophilus influenzae* and *Haemophilus haemolyticus* in Chronic Obstructive Pulmonary Disease (COPD)

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. None of this work has been published before submission.

Signed: ……………………………………………………………………………

Date: ……………………………………………………………………………

Acknowledgements

I would first like to say how grateful I am for being able to undertake this opportunity. I would like to express my gratitude to my supervisors Dr Stuart Clarke, Dr David Cleary, Dr Johanna Jefferies and Dr Christopher Woelk for their constant support and encouragement most notably whilst I was in Australia. I would also like to thank the rest of the IDEPI research group especially Rebecca Anderson, Denise Morris and Jessica Jones for going above and beyond and being there for coffee drinking, wine quaffing and general spirit lifting.

Thank you to all the AERIS project team at GSK in Belgium for their funding and support.

A big thanks also to Dr Elena Vataga and the team at Iridis who without their help, patience and guidance none of the bioinformatics analysis would have been possible. Thank you to Alice Still and Rebecca Anderson for their work carrying out extractions, sequencing and other laboratory based assistance.

Thank you to Dr Janessa Pickering and Dr Lea-Ann Kirkham for being my antipodean sounding boards with regards to the world of *H. haemolyticus* and a big thank you also to Dr Aaron Darling for inviting me along to the Tea and Buns sessions at the University of Technology in Sydney.

A huge thank you to my family and friends. My mother as always has been my voice of reason and my rock. To Michelle, Pete, Andy and Fi, always there to support their sibling despite thinking I was quite mad. To Jet, Ali, Munki, Kathryn, Judith, Morgan and Phil, thank you for keeping me sane and focused, without you life would be very empty. To Mark, a big thank you for your patience, support and companionship throughout the last seven years of full time study.

This is dedicated to Seb and Rosie, my two beautiful children who have suffered four years of slightly rushed bedtime reading and cuddles so mummy can go and re-join her laptop. I hope at the very least I have made you realise that anything is possible!

Abbreviations

adk	adenosine kinase, used in MLST
AE/AECOPD	Acute exacerbation in COPD
AMRG	Acquired macrolide resistance genes
ANI	Average nucleotide identity
AOM	Acute otitis media
atpG	ATP synthase gamma subunit, used in MLST
Azi	azithromycin
bexA	gene responsible for capsule
bp	Base pair
CEACAM	Carcinoembryonic antigen-related cell adhesion molecule
Clari	clarithromycin
COPD	Chronic obstructive pulmonary disease
csv	comma separated value file
dNTP	deoxynucleotide
EBI	European Bioinformatics Institute
ECM	Extra cellular matrix
Ery	erythromycin
fuc	gene from the fucose operon
Gavi	The Vaccine Alliance
GWAS	Genome wide association study
GTRCAT	General time reversible and category models of nucleotide substitution
H_2S	Hydrogen sulphide
Hap	*Haemophilus* adhesin protein
hel	gene responsible for omp4
Hh	*H. haemolyticus*
Hib	*H. influenzae* serotype B
Hif	*H. influenzae* fimbrae
Hpd	*H. influenzae* protein D
Hmw	High molecular weight protein
ICAM	Intercellular Adhesion Molecule 1
iga	immunoglobulin a protease
LOS	Lipooligosaccharide
Mac	macrolide
MIC	minimum inhibitory concentration
MLST	multi locus sequence typing
MRSA	methicillin-resistant staphylococcus aureus
NAD	nicotinamide adenine dinucleotide (V factor)
NCBI	National Center for Biotechnology Information
NCTC	National Collection of Type Cultures
NGS	Next-generation sequencing
NTHi	Non-typeable *H. influenzae*
OM	Otitis media
Omp	outer membrane protein
PHE	Public Health England
PhiD-CV10	10 serotype pneumococcal vaccine
PCV	pneumococcal conjugate vaccine
pilA	pilin protein
SLPI	Secretory leucoprotease inhibitor
SNP	single nucleotide polymorphism

smpB	ssra binding protein
sodC	copper dismutase
spp.	*Species pluralis* (multiple species)
SRST2	Short Read Sequence Typing
ST	Sequence type
STGG	Medium containing skim milk, tryptone, glucose, and glycerin
UPGMB	MEGA 6.0 clustering algorithm based on Unweighted Pair Group Method with Arithmetic Mean
TB	Tuberculosis
Tpf	Type IV pilus
VRE	vancomycin resistant *Enterococcus*
WGA	Whole genome analysis
WHO	World Health Organisation

Genes are denoted in italics, proteins are capitalised and in standard font.

1 Introduction

1.1 Chronic Obstructive Pulmonary Disease (COPD)

Chronic obstructive pulmonary disease (COPD) is a multi-faceted condition that progressively and irreversibly damages the pulmonary airways and impedes airflow in the lungs. This occurs when alveoli are reduced in diameter and obstructed by excessive bronchial secretion (Chung and Adcock, 2008). Respiratory function is also disrupted by the failure of elastic recoil caused by inflammatory processes. This results in bronchioles being broken down into larger air spaces reducing the surface area for gaseous exchange (Chung and Adcock, 2008).

COPD is a disease that is increasingly gaining recognition due to reported global morbidity rates. Over 3.1 million deaths representing 5.7% of total deaths in 2015 were due to COPD making it the third largest cause of death globally, above lung cancer (3.1%) and HIV/AIDS (2.1%) and below ischaemic heart disease (16%) and cerebrovascular disease (11.3%) (Collaborators, 2016). In the UK alone COPD was responsible for over 40,000 deaths in 2015, this represented 6.71% of all deaths (Figure 1.1) (Collaborators, 2016). This accrues an annual financial burden on UK health services in the region of £810 million (NHS, 2012). COPD patients have been reported to suffer a loss of earnings of around £200,000 per patient due to forced early retirement because of ill health (Fletcher et al., 2011).

1.1.1 Epidemiology

COPD is a global phenomenon. India and China had the highest percentage of total deaths attributed to COPD in 2015 with 10.13% and 9.67% respectively as can be seen in Figure 1.1. However, whereas the trend in China shows a reduction in deaths since 1990, COPD deaths in India are slowly increasing (Figure 1.1). This increasing trend is also visible in the United States and the United Kingdom with COPD responsible for 3.9% of deaths in 1990 in the US rising to 5.7% in 2015, and with 5.1% of deaths being attributed to COPD in 1990 rising to 6.7% in 2015 in the UK. Globally, mortality fluctuates between 5.5% and 6.2% (Figure 1.1). These figures may reflect real increase or an increase in the deaths recognised as COPD related.

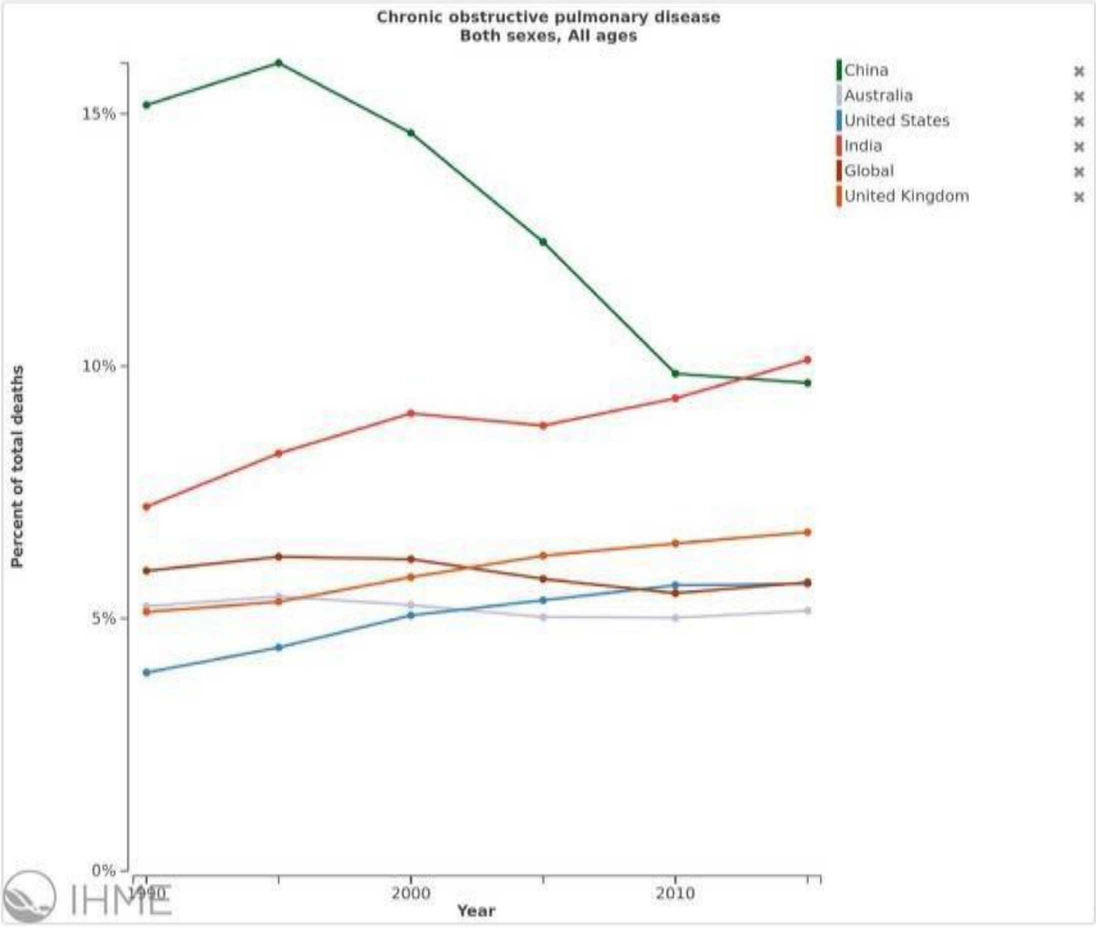


Figure 1.1 Percentage of deaths reported for COPD in selected countries compared to global rate from 1990 – 2015.

The global trend is a steady fluctuation between 5.5% and 6.2% however, a decreasing trend in the total deaths attributed to COPD can be seen in China and an increasing trend observed in India, the UK and US. (Figure obtained from the Institute for Health Metrics and Evaluation, http://www.healthdata.org)

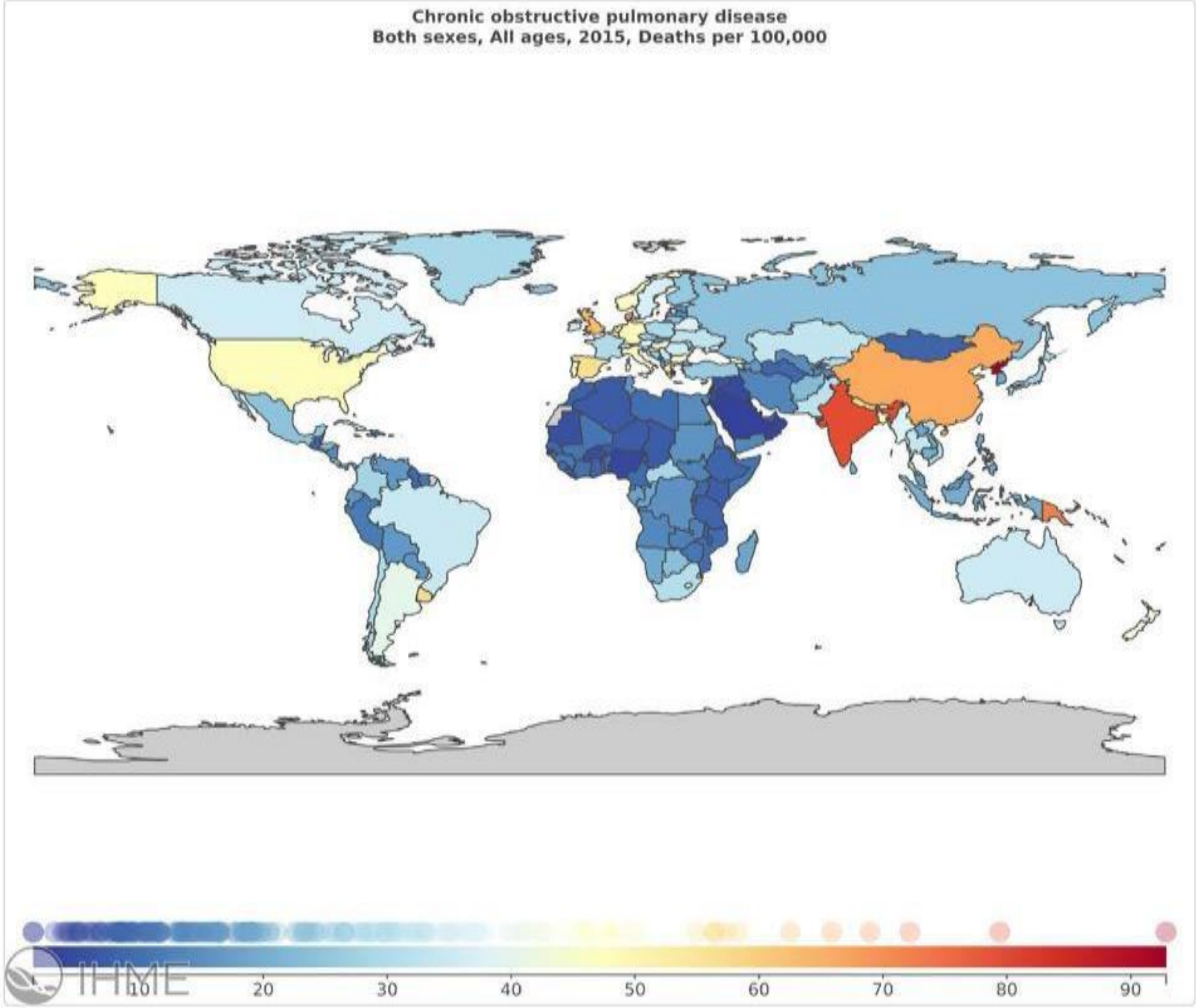


Figure 1.2 Global map indicating deaths per 100,000 of COPD.

(Figure obtained from the Institute for Health Metrics and Evaluation, http://www.healthdata.org)

1.1.2 Risk Factors

There is not one causal factor identified for the onset or progression of COPD but the largest risk factor in the developed world is thought to be smoking (Figure 1.3)(Lundback et al., 2003, Forey et al., 2011, Woodruff et al., 2016, Collaborators, 2016). Of the deaths caused by COPD reported by the global burden of disease study in 2015, 76% could be attributed to specific risk factors, the three largest being smoking (44.8%), ambient particulate matter (27.1%) and household air pollution (20.6%) (Figure 1.1). Alarge proportion of COPD deaths occurring in both upper – middle income and high income countries have been attributed to smoking however due to an increase in adults smoking in developing countries, more specifically male populations, the amount of COPD deaths attributed to smoking in lower income countries may also soon increase (Thun et al., 2012). Despite smoking being a significant risk factor in the developed world, the majority of COPD sufferers (90%) reside in low-middle income countries where smoking is a larger risk factor for the male population than it is female (Figure 1.3). It is thought that populations within these countries are also subject to socioeconomic factors such as lower income, overcrowding, a diet low in anti-oxidants and more frequent exposure to occupational dust/pollutants, in addition to exposure to biomass cooking methods resulting in household air pollution (Yin et al., 2011, Lee et al., 2012).

Cooking with biomass, or exposure to wood and charcoal smoke, has been associated with an increased risk of developing chronic respiratory disease in the female population. As such, it has been established that ambient particulate matter and household air pollution are the biggest risk factors for COPD mortality in the female population (Figure 1.3) (Orozco-Levi et al., 2006, Ramirez-Venegas et al., 2006, Perez-Padilla et al., 1996, Bhandari and Sharma, 2012). An estimated 2.8 billion people are thought to use biomass as a major source of fuel for cooking globally including 60% of households in Africa and South East Asia (Bonjour, 2013). Exposure to biomass smoke has been shown to contribute to COPD in China and India, where percentages of death from COPD are higher than anywhere else globally (Fernandes, 2014, Zhou and Chen, 2013, Zhong et al., 2007).

A significantly higher exposure to household air pollution from biomass use in rural Yunyan (88.1%) compared to urban Liwang (0.7%) correlated with a higher incidence of COPD in both the whole population and a female sub population of non–smokers (Liu et al., 2007).

Similarly, in India whereas only 3.6% of the female population were established as smokers compared to 26.3% of men, 42.5% of diagnosed cases of COPD were in women (Fernandes, 2014). The majority of these women (97.6%) had regular exposure to biomass smoke and subsequent household air pollution (Fernandes, 2014).

China accounts for a third of the world’s smokers however there is a disparity in gender with only 2.4%-8.5% of Chinese women smoking compared to 52.9%-56.7% of men (Zhou and Chen, 2013, Zhong et al., 2007). It has been reported that 5.1% of the female study population were diagnosed with COPD compared to 12.4% of the male study population therefore as stated above, cooking related exposure may account for some of this disparity (Zhou and Chen, 2013, Zhong et al., 2007).

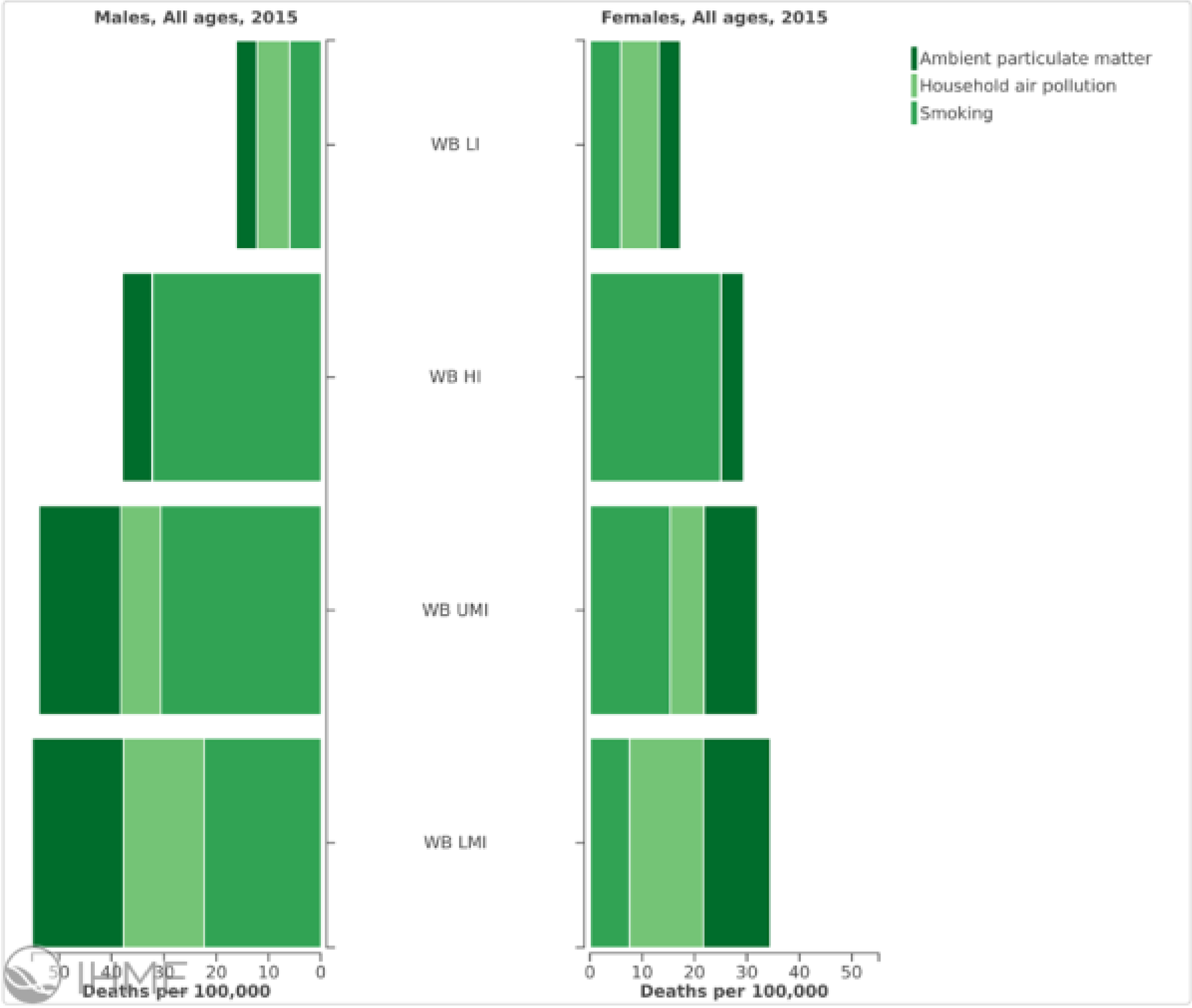


Figure 1.3 Deaths of COPD attributed to top three risk factors stratified by economic state of country as according to World Bank.

The highest reported deaths occur in countries with low to middle income with ambient particulate matter and household air pollution being the largest risk factors contributing to death by COPD in female deaths, larger than the risk factor of smoking. Smoking is the largest risk factor in the countries with high incomes. (Figure obtained from the Institute for Health Metrics and Evaluation, http://www.healthdata.org)

1.1.3 Infection and COPD

Acute exacerbations in COPD sufferers are periods of worsening symptoms that result in the deterioration of lung function and are associated with microbial and viral infections. *H. influenzae* has most notably been associated with the onset of acute exacerbations. Although it has been reported that the colonisation dynamics of *H. influenzae* in COPD patients does not to have an effect on exacerbation occurrence (Murphy et al., 2007, Bandi et al., 2003, Murphy et al., 2004, Albert et al., 2011), the presence of viral and bacterial pathogens has nonetheless been established in severe exacerbations requiring hospitalisation (Papi et al., 2006). Rhinovirus, influenza virus, respiratory syncytial virus, coronavirus, parainfluenzae and human metapneumovirus have all been isolated in exacerbation samples (Papi et al., 2006). Exacerbations have been associated with co-isolations, specifically with *H. influenzae* and rhinovirus, and are associated with greater degradation of lung function resulting compared to infection with lone bacterial or viral organisms (Wilkinson et al., 2006).

In some cases, it has been reported that COPD also and tuberculosis (TB) appear to display a reciprocal relationship with prior infection with *Mycobacterium tuberculosis* found to increase the risk of developing COPD (Lee et al., 2012, Inghammar et al., 2010). A delay in treatment in TB has been associated with not only a higher risk of development but also significantly earlier onset of COPD (Lee et al., 2012). Inversely a three-fold increase in developing active TB was discovered in COPD patients with patients developing both more likely to die within a year of COPD diagnosis (Inghammar et al., 2010).

H. influenzae is reported to be the most prevalent bacterial species isolated from the COPD lung although *Streptococcus pneumoniae*, *Moraxella catarrhalis*, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Enterobacter* spp. have also been observed (Papi et al., 2006, Wilkinson et al., 2006, Sethi et al., 2006). It has been reported that the presence of disease and inflammation results in reduced diversity in the COPD lung (Erb-Downward et al., 2011) however, it should be noted that *Haemophilus* and *Pseudomonas* spp. have been found to be present in ‘healthy’ smokers lungs (Erb-Downward et al., 2011). The prevalence of *Pseudomonas*, *Stenotrophomonas* or *Haemophilus* spp. has been shown to drive the microbial variation in the COPD lung, however it is thought that presence of *Haemophilus* spp. is the most important for this variation (Erb-Downward et al., 2011, Wang et al., 2016).

NTHi has repeatedly been isolated in the COPD lung and an acquisition of a new strain has been associated with the onset of acute exacerbation (Garmendia et al., 2014, Bandi et al., 2003, Murphy et al., 2004, Fernaays et al., 2006b, Leanord and Williams, 2002, Murphyet al., 2007). NTHi is capable of persistent colonisation in the lower respiratory tract of patients with COPD by employing such survival strategies as adaptation of outer membrane proteins between planktonic and biofilm environments and intercellular invasion, therefore forming a reservoir for infection (Murphy et al., 2004, Bandi et al., 2001, Murphy and Kirkham, 2002, Salcedo and Cid, 2011). Phagocytosis of NTHi by alveolar but not blood macrophages is impaired in COPD, possibly also explaining its persistence capabilities (Berenson et al., 2006a, Berenson et al., 2006b).

1.2 H. influenzae

H. influenzae is a Gram negative opportunistic pathogen that colonises the respiratory tract. It is a known causative organism in invasive diseases such as meningitis, pneumonia, bacteremia and epiglottitis as well as a major cause of otitis media (OM) (McVernon et al., 2006, Cardines et al., 2007, Cripps, 2010, Monasta et al., 2012). *H. influenzae* occurs as both encapsulated and non-encapsulated strains. Non-encapsulated strains are referred to as non-typeable *H. influenzae* (NTHi). There are seven encapsulated serotypes designated a-f which represent the variation in composition of the polysaccharide capsule (Pittman, 1931, Kilian, 1976b).

In the UK, prior to 1992, 90% of invasive *Haemophilus* disease was caused by *H. influenzae* serotype b (Hib) (Anderson et al., 1995a). Hib was the largest cause of childhood invasive disease presenting most often as meningitis (50-65%) and resulted in 25.1-35.5 cases per 100,000. This was drastically reduced to 0.06 per 100,000 in 2012 as a direct result of the introduction of the vaccination and indirectly through associated herd immunity. (Anderson et al., 1995a, Hargreaves et al., 1996, Booy et al., 1993). Within a year of the introduction of the vaccination in the UK, cases of Hib invasive disease had fallen to 2 per 100,000 (Hargreaves et al., 1996). In the US cases fell from 37 per 100,000 in 1989 prior to the introduction of the vaccine to 11 cases per 100,000 in 1991 (Adams et al., 1993).

In 2000 Gavi, the Vaccine Alliance was set up to fund immunisation programmes in lower income countries. Mostly participating governments funded this with 20% being funded by private contributions. From the onset of the Gavi programme the Hib vaccination was offered in the form of the pentavalent vaccination. This provided protection against diphtheria, tetanus, pertussis, hepatitis B and Hib. Initially, uptake was slow leading to a US$37 million Hib vaccination initiative in 2005 (alliance, 2016). By 2014 all 73 Gavi eligible countries had rolled out the vaccination programme with South Sudan being the last country to adopt the vaccine in July 2014 (alliance, 2016). The uptake of the pentavalent vaccine coverage across Gavi supported countries was seen to increase from 1% in 2000 to 68% in 2015 (GAVI, 2016). The overall global uptake of the vaccination has increased from 42% in 2010 to 64% in 2015 (Casey et al., 2016). Invasive Hib disease has since reduced significantly, although invasive *Haemophilus* disease still occurs. Although all serotypes have the potential to cause disease, NTHi has taken over as the major cause of

Haemophilus invasive disease accounting for 84% of UK cases in 2013 (Ladhani et al., 2010, PHE, 2014).

In addition to invasive disease NTHi is also a leading cause of an infection of the middle ear, otitis media (OM). Approximately 709 million cases of acute OM are reported annually around the world of which NTHi being the predominant causes (Monasta et al., 2012, Barkai et al., 2009, Kong and Coates, 2009). At around 73%, OM is most extensively suffered by children of indigenous populations of Australia (Leach et al., 2015). Over half of the global cases occur in children under the age of 5 with approximately 31 million cases going onto become chronic OM with associated hearing loss, 21 thousand people dying from complications arising from OM each year and required treatment resulting in annual costs of between AUS$100-400 million in Australia and US$2.8 billion in the US (Kong and Coates, 2009, Taylor et al., 2009, Monasta et al., 2012, Ahmed et al., 2014).

Elsewhere nasopharyngeal carriage of NTHi has been has been reported in 26%-52.4%of infants under the age of 5 years (Cohen et al., 2012, Angoulvant et al., 2015, Brandileone et al., 2016). Vaccination for the Gram positive opportunistic pathogen of the respiratory tract, *S. pneumoniae*, has been reported to have an effect on the carriage of *H. influenzae* (Cohen et al., 2012, Angoulvant et al., 2015, Brandileone et al., 2016). *S. pneumoniae* is also a causative organism in disease such as meningitis, pneumonia and OM (middle ear infection) and is found to co–colonise with NTHi in the nasopharynx (Barkai et al., 2009, Leach et al., 2016, Kastenbauer and Pfister, 2003, Vu et al., 2011).

S. pneumoniae, is identified by serotype due to variation in polysaccharide capsule. There are currently at least 94 (Kapatai et al., 2016) *S. pneumoniae* serotypes and three vaccines which encompass 10, 13 and 23 of the most virulent serotypes, referred to as PhiD-CV10 Synflorix ™, Prevenar 13™ and Pneumovax 23™ respectively (Habib et al., 2014, Brandileone et al., 2016, Cohen et al., 2012, Leach et al., 2016). A fourth vaccine Prevenar 7™ was replaced by Prevenar 13™ in April 2010. The PhiD-CV10 Synflorix™ pneumococcal vaccine targets ten of the most invasive serotypes of *S. pneumoniae* of which eight are conjugated to NTHi protein D (Croxtall and Keating, 2009). However, the impact of an immunogenic response from the protein D conjugate is currently unclear with studies into the effects on NTHi carriage reporting contradictory findings. In Brazilian infants, the introduction of the 10-valent PHiD-CV10 vaccine resulted in a significant (*p*=<0.0001) increase in nasopharyngeal carriage of NTHi from 26%-43.6% over a three year period

(Brandileone et al., 2016). Conversely, no effect on NTHi carriage was observed in a Finnish study of 4117 infants after introduction of the PHiD-CV10 vaccine (Vesikari et al., 2016, Cohen et al., 2012). Similarly, no variation in NTHi colonisation was observed between 780 children (<2yrs) vaccinated with either PHiD-CV10 or Prevenar 7™ (van den Bergh et al., 2013). Carriage of both *S. pneumoniae* and NTHi was seen to significantly decrease in indigenous Australian children with OM after vaccination with Prevenar 13™ compared to a group vaccinated with the 10-valent PHiD-CV10 vaccine (Leach et al., 2016). NTHi carriage in Kenyan infants immunised with PHiD-CV10 was seen to reduce from 54% to 40% (Hammitt et al., 2014). Conversely in France, a significant increase in NTHi carriage (from 48.2%-52.4%) was observed in a paediatric study group after immunisation with PCV-13 but this was not the case in a study group vaccinated with PCV-7 (Angoulvant et al., 2015, Cohen et al., 2012). In summary, the interplay between pneumococcal vaccination and carriage of NTHi is not fully understood.

1.2.1 NTHi and *H. haemolyticus*

Phylogenetically, *H. haemolyticus* is the closest species to NTHi. The two species share morphology and the growth factor requirements of nicotinamide adenine dinucleotide (NAD) and haemin (Forbes KJ, 2002, Kilian, 1976b). Historically, *H. haemolyticus* has been considered to be a harmless commensal of the nasopharynx previously identified in 11.7% of children although it has been noted significantly absent in paediatric patients suffering from otitis media (Fox et al., 2014, Kirkham et al., 2010). Despite its non-pathogenic status there have been eight cases of *H. haemolyticus* isolated from invasive disease (Jordan et al., 2011, Anderson et al., 2012, Morton et al., 2012). In addition to carriage in the upper respiratory tract *H. haemolyticus* has also been isolated from the lower respiratory tract of patients with COPD but no link with the onset of acute exacerbations has been identified (Murphy et al., 2007). Interestingly *H. haemolyticus* has been shown to decrease the attachment of NTHi to epithelial cells suggesting competition between the two species for colonisation (Pickering et al., 2016). Additionally, NTHI display a varying ability to invade epithelial cells on binding invariably *H. haemolyticus* results in cytotoxicity (Pickering et al., 2016, Singh et al., 2016a).

H. haemolyticus can be distinguished from *H. influenzae* by the presence of haemolysis surrounding colonies grown on blood agar (Kilian, 1976a). However, the loss of the haemolysis phenotype has been shown and results in strains of non haemolytic *H.*

haemolyticus (Kilian, 1976a). This finding rendered culture methods employed to distinguish between the two species unreliable, highlighting the potential for misidentification of *H. haemolyticus* as NTHi. This resulted in the need for more robust molecular methods for differentiation between the two species. Misidentification of *H. haemolyticus* as NTHi has been established in several studies that retrospectively investigated *Haemophilus* spp. from cases of disease using molecular methods. *H. haemolyticus* misidentified as NTHi in clinical isolates were reported at rates of 5.9% and 1.5% in Germany and Australia, respectively (Frickmann et al., 2014, Zhang et al., 2014). An Australian study identified 11.7% of nasopharyngeal carriage isolates as *H. haemolyticus* and a study of isolates from cystic fibrosis patients in Denmark re-classified 0.5% of isolates (Kirkham et al., 2010, Fenger et al., 2012). An earlier study investigating isolates from the sputa of COPD lungs and paediatric nasopharyngeal carriage isolates reported a much higher level of misidentification with 39.5% and 27.3% of isolates respectively re-classified as *H. haemolyticus* after molecular investigation (Murphy et al., 2007). *H. haemolyticus* were identified in 12 out of 63 *Haemophilus* spp. isolates from paediatric oropharyngeal samples but no *H. haemolyticus* were identified in nasopharyngeal or middle ear fluid samples from the same study (Chang et al., 2010).

Other biochemical methods, such as hydrogen sulphide production, have been suggested useful to differentiate between the two species. However, ambiguity may still arise given this phenotype is variably present found in 70% of *H. haemolyticus* and 12.5% of NTHi (McCrea et al., 2008).

NTHi cannot be serotyped due to the absence of a capsule although classification can be achieved using multi locus sequence typing (MLST). MLST uses partial sequences from seven housekeeping genes thought to be conserved within *H. influenzae*; adenylate kinase (*adk*), ATP synthase F1 subunit gamma (*atp*G), fumarate reductase iron-sulfur protein (*frd*B), fuculokinase (*fuc*K), malate dehydrogenase (*mdh*), glucose-6-phosphate isomerase (*pgi*) and RecA protein (*recA*) (Meats et al., 2003). The concatenation of partial sequences found at these loci allow for a sequence type (ST) to be defined, with new STs being assigned on the discovery of previously unseen sequences. However, strains in which the and indeed the entire *fuc* operon have been reported to be absent (Ridderberg et al.,2010, Latham et al., 2015). This not only brings into question the accuracy of MLST for NTHi but also whether MLST accurately reflects the heterogeneity of the population (Ridderberg et al., 2010, Binks et al., 2012). MLST does, however, provide a tool to cluster typical strains of NTHi offering an initial categorisation at a molecular level for related isolates.

1.2.2 Mechanisms for Diversity in NTHi

In comparison to its encapsulated clonal counterparts, NTHi is genomically heterogeneic in nature and displays variation throughout strains (Power et al., 2012, Meats et al., 2003, Erwin et al., 2008, De Chiara et al., 2014, McCrea et al., 2008, Musser et al., 1986a, Musser et al., 1990). It is this heterogeneity of NTHi which makes full understanding of the virulence mechanisms problematic, in turn hindering the identification of suitable targets for immunisation (LaCross et al., 2013). It has been hypothesised that NTHi is adaptable to the surrounding environment drawing upon a subset of genes from the larger pan-genome within the species (Shen et al., 2005, Fernaays et al., 2006b, Garmendia et al., 2014). This genetic variability has been attributed to horizontal gene transfer (HGT) between NTHi most feasibly through transformation (Power et al., 2012, Connor et al., 2012). Natural transformation is a result of exogenous DNA being taken from the external environment and incorporated into the bacterial genome (Figure 1.4) (Carruthers et al., 2012, Barany and Kahn, 1985).

Transformation can result in recombination or insertion mediated integration, depending on the homology of the exogenous DNA with the existing genome. Recombination events occur between homologous regions of the transformed DNA and the host chromosome, whereas insertion can occur in the absence of homology. Both mechanisms have been reported in NTHi (Mhlanga-Mutangadura et al., 1998, Satola et al., 2008, St Geme et al., 1998, Connor et al., 2012, de Gier et al., 2015, Kroll et al., 1988, McCrea et al., 2010, McCrea et al., 2008, Maughan and Redfield, 2009a, Skaare et al., 2010, Witherden et al., 2014). The ability of NTHi to uptake DNA by recombination has been reported to vary by 1,000 fold between strains offering a potential mechanism and explanation for the extent of genetic variation (Maughan and Redfield, 2009a). This could be linked to absence of competence capability within some strains *H. influenzae* which is required for transformation (Maughan and Redfield, 2009b). Competence is activated by certain conditions such as limited availability of nutrients and in *H. influenzae* has been shown to occur at the end of exponential growth phase proceeding the stationery growth phase

(Goodgal and Herriott, 1961). Mutations within the genes responsible for competence disrupt the ability of the bacteria to uptake DNA and complete transformation (Maughan and Redfield, 2009b, Maughan and Redfield, 2009a).

Although capsulated strains have been shown to be less genetically diverse than non-encapsulated versions, the lack of the capsule has been shown not to be the sole cause for the extent of heterogeneity within NTHi (Kroll et al., 1988, Maughan and Redfield, 2009a). An investigation into the evolutionary relationship of *H. influenzae* reported certain strains of NTHi have clustering with Hib indicating a close phylogenetic relationship (Connor et al., 2012). This supports the proposed evolution of NTHi from encapsulated strains, specifically Hib (Musser et al., 1986b, Connor et al., 2012). Recombination disrupting the *bexA* gene responsible for transport of capsular polysaccharide has been attributed to the loss of capsule from Hib strains (Kroll et al., 1988). Insertion elements associated with Hib have also been located within certain NTHi strains (Satola et al., 2008).

The relationship between non-encapsulated strains and their evolution from Hib is not fully understood but non-encapsulated strains show larger genetic diversity than encapsulated strains

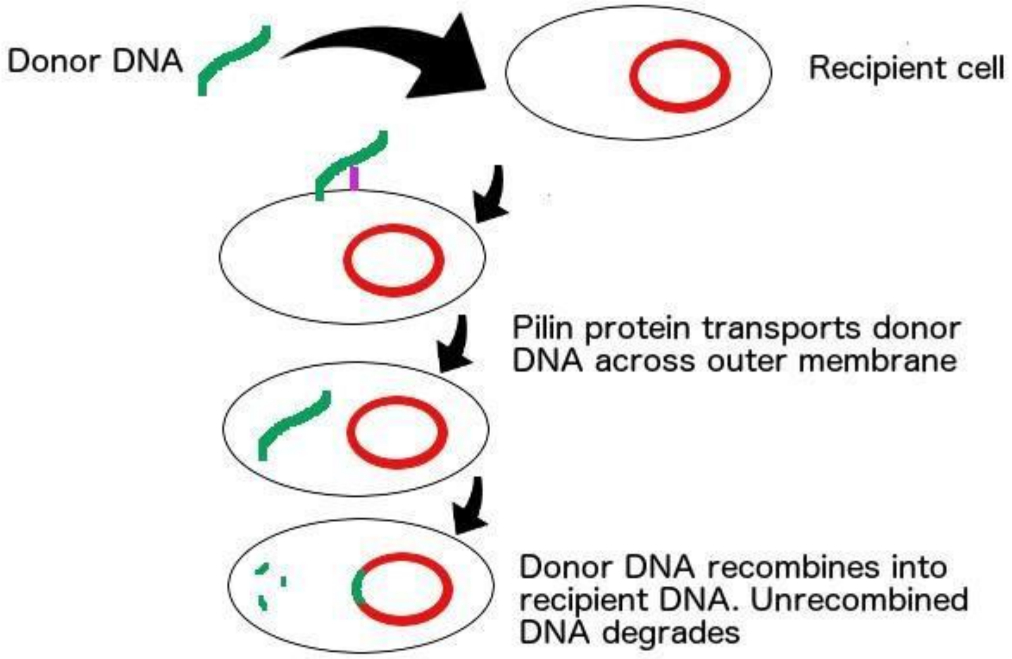


Figure 1.4 Horizontal gene transfer by transformation.

Donor DNA is taken up by pilin proteins and transported across the outer membrane into the recipient cell. DNA is recombined into the recipient chromosomal DNA and the remainder of the donor DNA degraded.

1.2.3 The Genetic Relationship Between *H. haemolyticus* and NTHi

NTHi displays large genetic variability. Inter species recombination between NTHi and *H. haemolyticus* has been described, potentially enabling the blurring of the two species (McCrea et al., 2008, de Gier et al., 2015, Witherden et al., 2014, Witherden and Tristram, 2013). The need for molecular methods for accurate identification of NTHi and *H. haemolyticus* has stemmed from the inability of culture methods to do so. This is observed in the failure to develop a molecular assay that can unequivocally differentiate between the two species.

1.2.3.1 Gene Markers

A number of candidate genes for speciation have been suggested and are summarised in Figure 1.5. Certain genes, such as *fuc*K, *lgt*C, *fucP* and *iga* are present in NTHi and absent within *H. haemolyticus, sodC* is typically present in *H. haemolyticus* but not in NTHi(Figure 1.5) (Fung et al., 2006, McCrea et al., 2010, Latham et al., 2015, Binks et al., 2012, Hinz et al., 2015, McCrea et al., 2008, Mukundan et al., 2007) An alternative method is to find conserved sequence variations between the two species in mutually prevalent genes such as *omp6*, *smpB*, and *hpd* (Figure 1.5) (Reddington et al., 2015, Binks et al., 2012, Mukundan et al., 2007, Sandstedt et al., 2008, Hare et al., 2012, Wang et al., 2011, Theodore etal., 2012, Pickering et al., 2014a).

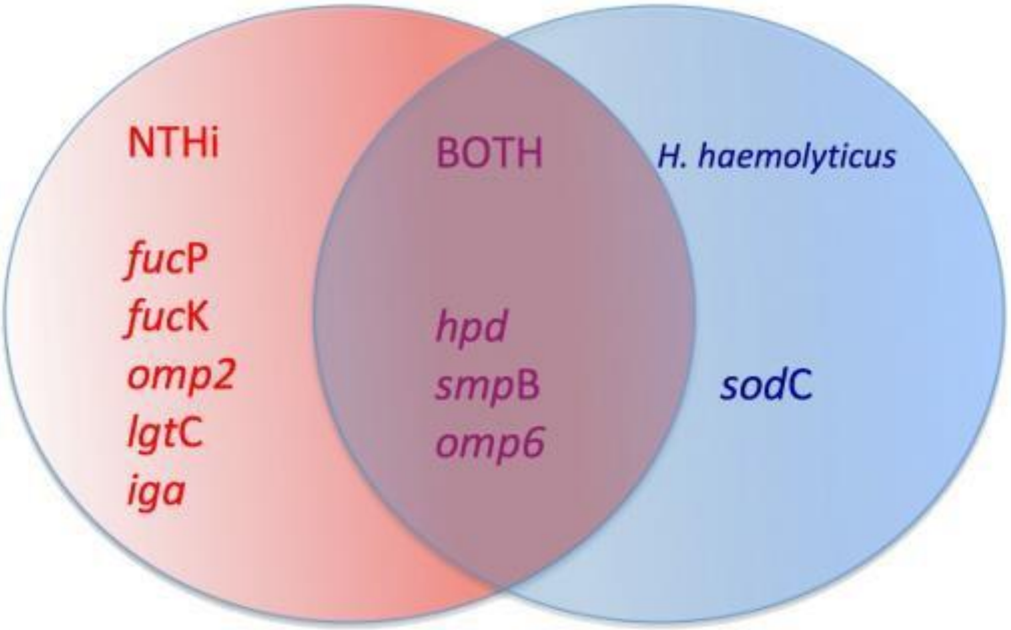


Figure 1.5 Expected genotypes of published gene markers for NTHi and *H. haemolyticus.* PCR assays used to differentiate between NTHi and *H. haemolyticus*. Genes sat under the heading of one species are expected to be found present in that species and absent in the other. The genes sat in between species are those present in both species are assays that exploit sequence variation between the species (Fung et al., 2006, McCrea et al., 2010, McCrea et al., 2008, Latham et al., 2015, Binks et al., 2012, Hinz et al., 2015, Mukundan et al., 2007, Reddington et al., 2015, Price et al., 2015, Sandstedt et al., 2008, Hare et al., 2012, Wang et al., 2011, Theodore et al., 2012, Pickering et al., 2014a).

1.2.3.2 Protease (*iga* and *iga*B)

NTHi produces a protease that cleaves secretory human immunoglobulin A (IgA) allowing the bacteria to evade detection by IgA antibodies. Two genes are thought to be responsible for the expression of IgA protease, *iga* and *igaB* (Binks et al., 2012, Fernaays et al., 2006a, Fernaays et al., 2006b, McCrea et al., 2008, Murphy et al., 2011, Vitovski et al., 1999).The *igaB* gene in NTHi is homologous to that found in *Neisseria meningitidis* suggesting recombination between the two species (Fernaays et al., 2006a).

Strains containing *igaB* in addition to *iga* are thought to exhibit an increase in IgA protease production and are reportedly more associated with strains isolated from the lungs of patients with COPD (Vitovski et al., 2002). Indeed, invasive *N. meningitidis* strains have been shown to produce higher levels of IgA protease compared to strains isolated from carriage (Vitovski et al., 1999). The *iga* and *igaB* genes may also have secondary roles. IgA protease enables the bacteria to invade host cells whereas the IgaB protease enables intracellular persistence once host cells have been invaded; both genes together give an optimal survival factor facilitating internalisation and persistence (Clementi et al., 2014). Although more often found in NTHi, the *iga* gene has been identified in small proportions (7.7%) of *H. haemolyticus* (Binks et al., 2012). The limited ability for invasion of H. haemolyticus is potentially explained by the absence of *iga* (Pickering et al., 2016, Singhet al., 2016a). However, the absence of *iga* in some NTHi strains (9%) has been reported making it an inconclusive target for species differentiation (Binks et al., 2012).

1.2.3.3 Lipoligosacharide and Galactosyltransferase (*lgtC*)

The lipoligosaccharide (LOS) is an endotoxin found within the membrane of *H. influenzae* and associated with virulence (Preston et al., 1996, Kimura and Hansen, 1986). Galactosyltransferase forms part of the LOS and is expressed by the *lgt*C gene which has been found to be fifty-four times more prevalent in NTHi than in *H. haemolyticus* but was identified in 2% of *H. haemolyticus* strains (McCrea et al., 2008). Although *lgtC* does not unambiguously identify NTHi it is still very useful in assisting in identification (McCrea et al., 2008, Fox et al., 2014).

1.2.3.4 [Cu,Zn]-Superoxide Dismutase Gene (*sod*C)

The [Cu,Zn]-superoxide dismutase (Sod) protects bacteria from the oxygen free radical, superoxide produced by the immune system of the host (Vanaporn et al., 2011). The *sodC* gene has also been described in encapsulated strains of *H. influenzae* and a cryptic *Haemophilus* genospecies, separate from NTHi and *H. haemolyticus,* which has been isolated from urogenital and neonatal infections (Langford et al., 2002). The encapsulated strains however do not produce the Sod enzyme due to a substitution of a histidine residue with a tyrosine resulting in a non-functional protein (Kroll et al., 1991). The cryptic genospecies does express the enzyme but has not been isolated from the respiratory tract (Langford et al., 2002). Some NTHi strains have been identified as positive for a homologue of *sod*C with alleles homologous to those more associated with wither H. haemolyticus or encapsulated strains, suggestive of recombination at this locus (McCrea et al., 2010).

The *sodC* gene encoding the Sod enzyme has previously been found to be present in 95% - 100% of *H. haemolyticus* strains (Fung et al., 2006, McCrea et al., 2010, Latham et al., 2015). In NTHi, *sodC* has been reported in 9.2% of isolates (McCrea et al., 2010, Fung et al., 2006). Despite the presence of *sodC* in a small proportion of NTHi there is clear merit in exploring this as a potentially useful marker to identify *H. haemolyticus.*

1.2.3.5 Outer Membrane Proteins 2 and 6 (*omp2*, *omp6*)

Outer membrane protein 6 (*omp6*) is thought to be important connecting the outer membrane to the cell wall and outer membrane protein 2 (*omp2*) is shown to make up 50% of all the outer membrane protein (Murphy et al., 2006). Four separate single nucleotide polymorphisms (SNPs) within the *omp6* are able to speciate between NTHi and *H. haemolyticus* in a previously developed PCR assay. However additional sequence variation within *omp6* has been reported in 10% of NTHi (Chang et al., 2010, Binks et al., 2012, Chang et al., 2011, Mukundan et al., 2007, Murphy et al., 2007). Speciation between NTHi and *H. haemolyticus* has also been achieved by determining the presence of *omp2* in NTHi and absence in *H. haemolyticus*. Strains of NTHi absent for *omp2* have been reported, despite this, *omp2* has been found to be a better source of differentiation than *omp6* (Binks et al., 2012, Chang et al., 2010).

1.2.3.6 The Fucose Members (*fucK, fucP*)

Presence of the *fuc*K gene, one of the seven housekeeping genes used within MLST to type *H. influenzae*, has also been used as a tool to differentiate between *H. influenzae* and *H. haemolyticus* (Norskov-Lauritsen, 2009, Binks et al., 2012, Norskov-Lauritsen et al., 2009). An assay for another member of the *fuc* operon, *fucP*, has been developed which also exploits its presence in NTHi and absence in *H. haemolyticus* (Price et al., 2015). However, a number of strains that would otherwise be identified as NTHi were it not for the absence of the *fuc* operon, and therefore both absent for *fuc*K and *fuc*P, have been reported (Ridderberg et al., 2010). The absence of the *fuc*K gene renders strains non-typeable by MLST and can potentially lead to misidentification as *H. haemolyticus*.

1.2.3.7 Protein D (*hpd*)

The *hpd* gene encodes for protein D. Although present in both NTHi and *H. haemolyticus* it has been found to be absent in some clinical NTHi strains (Smith-Vaughan et al., 2014). Despite this it displays inter species sequence variation, which has been used to successfully differentiate NTHi and *H. haemolyticus* to a high degree of sensitivity (Wang et al., 2011, Binks et al., 2012, Pickering et al., 2014a).

1.2.3.8 RNA Binding Protein Small Protein B –*smpB*

A new assay, based on sequence variation between NTHi and *H. haemolyticus* in the conserved *smpB* gene, has been recently described (Reddington et al., 2015). Its' broader use has not yet been reported.

In summary, single molecular markers are unable to unequivocally speciate between NTHi and *H. haemolyticus* due to atypical strains and the heterogeneity and recombination displayed by NTHi (Cody et al., 2003, Connor et al., 2012, LaCross et al., 2013, Lacross et al., 2008).

1.3 Antimicrobial Resistance

Antibiotic resistance is a global cause for concern and results from horizontal gene transfer in bacteria or mutation due to selective pressure. The ability of *H. haemolyticus* and NTHi to share genes with each other may offer a potential pool of resistance genes in *H. haemolyticus* accessible by the more pathogenic NTHi or vice versa. Recombination has already been accredited to the transfer of the *ftsl* gene responsible for penicillin resistance between NTHi and *H. haemolyticus* (Witherden et al., 2014, Marti et al., 2016). *H. haemolyticus* has recently been found to develop resistance to ciprofloxacin (Witherden et al., 2014, Marti et al., 2016).

1.3.1 Azithromycin and the Macrolides

National guidelines for the management of exacerbation in COPD inform the local strategy for treatment of NTHi infections in COPD exacerbations with doxycycline or azithromycin at University Hospital Southampton NHS Trust (Solutions, 2014, NICE, 2010). Further use of azithromycin as a prophylactic has been proposed to reduce inflammation and in turn reduce the number of exacerbation events (Albert et al., 2011). This preventative method has been contested due to the potential for increasing macrolide resistance in respiratory bacterial species exposed including NTHi, and increased risk of death in people with cardiovascular disease (Crosbie and Woodhead, 2013, Mammen and Sethi, 2012, Serisier, 2013). Exposure to azithromycin has been shown to be the driving factor for resistance and long-term usage has shown to increase the number of resistant strains of *S. pneumoniae*, *S. aureus* and *H. influenzae* (Albert et al., 2011, Malhotra-Kumar et al., 2007, Bergman et al., 2006, Leach et al., 1997, Hare et al., 2013, Valery et al., 2013, Phaff et al., 2006, Pettigrew et al., 2016).

Erythromycin A is a 14 ring structured macrolide derived from *Streptomyces erythreus* and has been important in fighting infection since 1952 especially in patients that displayed intolerance to penicillin (Welch et al., 1952, Haight and Finland, 1952, Vannucchi, 1952). However, in the 1960's several reports were made of bacterial species such as *S. pneumoniae*, *Streptococcus pyogenes*, *S. aureus*, *Escherichia coli* and *Bacillus subtilis* developing resistance to erythromycin (Taubman et al., 1963, Taubman et al., 1966, McGehee et al., 1968, Ravin and Iyer, 1961, Ravin et al., 1969, Hall, 1960, Dobrzanski and

Pulawska, 1960, Wahl and Fouace, 1961b, Wahl and Fouace, 1961a, Weaver and Pattee, 1964, Rychlik and Cerna, 1968). Resistance coupled with the requirement to improve acid stability to prevent inactivation of erythromycin A in the stomach environment led to further research and the development of the erythromycin A oxime. This led to a succession of semi-synthetic derivatives such as clarithromycin, dirithromycin and roxithromycin that in turn led to the development of the 15 ring structure semi-synthetic macrolide named azithromycin (Hardy et al., 1988, Fernandes and Hardy, 1988). The 15 ring structure compared to the 14 ring structure of previous macrolides was achieved by the use of formaldehyde and formic acid to methylate erythromycin A (Lazarevski et al., 1998, Djokic et al., 1987). Azithromycin proved to be more effective than erythromycin against bacterial species such as *H. influenzae*, *N. gonorrhoea, M. catarrhalis* and many of the *Enterobacteriaceae, Campylobacter* and *Legionella sp.,* and with a lower MIC (Retsema et al., 1987, Hardy et al., 1988, Fernandes and Hardy, 1988).

Azithromycin is classed as bacteriostatic and works by inhibiting nascent proteins from exiting the protein exit tunnel within the 50s ribosomal subunit. Azithromycin is shown to bind to a site close to the constriction of the peptide exit tunnel formed by the L4 and L22 proteins (Lovmar et al., 2004, Garza-Ramos et al., 2001). In addition, azithromycin alters the assembly of the 50s subunit resulting in ribosome degradation (Petropoulos et al., 2009, Chittum and Champney, 1995). Azithromycin has displayed species specific binding patterns with two molecules binding to each ribosome in *Deinococcus radiodurans,* but only one per ribosome in the case of *E. coli* (Petropoulos et al., 2009).

Supplementary to its antimicrobial actions, azithromycin has shown anti-inflammatory therapeutic effects in conditions such as the emphysetamous lung, meibomian gland disease as well as *in vitro* studies on COPD sputum cells and bronchial epithelial cells (Shinkai et al., 2006, Marjanovic et al., 2011, Zhang et al., 2015, Segal et al., 2016). Prophylactic use of azithromycin in patients with chronic inflammatory conditions, such as asthma and bronchiectasis in both cystic fibrosis and non-cystic fibrosis cases has been shown to minimise the risk of exacerbation occurrence (Menzel et al., 2016, Albert et al., 2011, Hare et al., 2013, Altenburg et al., 2013, Hansen et al., 2009). Additionally, azithromycin has been shown to decrease the viral load of rhinovirus, linked to the onset of acute exacerbations in COPD patients (Menzel et al., 2016). The antiviral effect from azithromycin administration was induced by interferon expression and was observed in asthma and COPD bronchial cells but was not replicated in healthy bronchial cells (Menzel et al., 2016, Gielen et al., 2010).

More pertinent for this study is the prophylactic use of azithromycin in COPD. Previous cohort studies had varied durations of administration including daily, every 3 days, every 21 days to 12 months of daily oral therapy (Table 1.1). A reduction in acute exacerbation frequency in those COPD patients receiving macrolide therapy (erythromycin, clarithromycin or azithromycin) has been observed in the majority of studies (Table 1.1) (Albert et al., 2011, Banerjee et al., 2005, Suzuki et al., 2001, Yamaya et al., 2008, He et al., 2010, Blasi et al., 2010, Gomez et al., 2000). The number of patient numbers in each study varied from 11-495 in the macrolide treatment groups and 11-502 in the control groups (Table 1.1). The variation of the studies and the uniformity of the outcome revealed that macrolide therapy may be a good course of prophylactic treatment for reducing the risk of the onset of exacerbations. Potentially however this could lead to the increase of resistance to azithromycin.

Table 1.1. Effects of macrolide treatment on COPD patients.

Macrolide therapy	Treatment Period (mths)	Pts (mac)	Pts in (placebo)	Significant reduction of	Significant difference in hospital admissions?	Other findings?	Ref
Clari 500mg daily	3	26	34	No	N/A	No improvement of health status No reduction of bacterial content in sputum sample	(Banerjee et al., 2005)
Ery 200mg-400mg daily	12	55	54	Yes. $p=0.0007$	Yes $p=0.0007$	Significant reduction in common cold in treatment group $p=0.002$	(Suzuki et al., 2001)
Clari 200-1200mg daily	42.9 (mean)	45	78	Yes. $p=0.01$	Yes. $p=0.04$	Sputum volume associated withexacerbations $p=0.007$	(Yamaya et al., 2008)
Ery 125mg 3 x daily	6	16	15	Yes. $p=0.042$	N/A	Reduction in neutrophils in treatment group. $p=0.0005$. Delay in time to first AE in treatment group $p=0.032$	(He et al., 2010)
Azi 500mg 3/21 days	3	54	40	Yes.	Yes.		(Gomez et al., 2000)
Azi 500mg 3/7 days	6	11	11	Yes. $p=0.001$	Yes. $p=0.02$	Delay in time to first AE in treatment group $p=<0.001$	(Blasi et al., 2010)
Azi 250mg daily	12	495	502	Yes. $p=<0.001$	N/A	Treatment group less likely to be colonised with pathogenic bacteria but more likely to be colonised with macrolide resistant strains. Hearing decrement in treatment group significantly larger than placebo group.$p=0.04$	(Albert et al., 2011)
Azi 500mg 3/7 days	12	20	N/A	Yes. $p=<0.001$	Yes.	Reduction of *H. influenzae* but not *S. pneumoniae* with long term use of azithromycin.	(Pomares et al., 2011)

mac – macrolide treatment group, placebo = placebo treatment group

1.3.2 Genetic Mechanisms for Resistance to Macrolides

Resistance to azithromycin and macrolides has generally centred around two mechanisms, disruption to the binding sites within the bacterial 50s ribosomal protein and efflux pumps that remove the antibiotic before it has a detrimental effect. The L4 and L22 proteins make up the protein exit tunnel in the 50s ribosome which is blocked by azithromycin in successful therapeutic use. Mutations within these proteins have been observed as a resistance mechanism within NTHi (Pettigrew et al., 2016, Clark et al., 2002). Decrease in susceptibility to azithromycin has been associated with L4 substitutions A2058G and A2059C in *E. coli*, *T. palladium*, *S. pneumoniae* and *H. influenzae* (Clark et al., 2002, Pericet al., 2003, Tait-Kamradt et al., 2000b, Diner and Hayes, 2009, Katz and Klausner, 2008, Van Damme et al., 2009, Lukehart et al., 2004). Macrolide resistance has also been observed in *E. coli* involving multiple different L4 mutations however substitution K90W in *E. coli* has been identified as explicit to increased resistance to erythromycin (Diner and Hayes,2009). Further to those identified in *E.coli*, L4 mutations in NTHi have been reported as substitutions K61Q, T64K, A69S, T82I, G65D and an insertion of GT at position 63. In *S. pneumoniae*, again in the same region as NTHi, substitution of GTG for TPS at position 69 has been reported along with a 6 amino acid insertion (REKGTG) at position 69 (Tait-Kamradt et al., 2000b).

L22 mutations associated with resistance to azithromycin in NTHi include deletions at position 82 of M and MKR, substitutions R88P, A94E and an insertion of 7 amino acids at position 72 (Clark et al., 2002). Furthermore, mutations within the 23S rRNA have been associated with increased macrolide resistance in NTHi (Clark et al., 2002).

Acquired macrolide resistance genes (AMRG) have been frequently reported in *S. pneumoniae* and include the *erm* and mef(A/E genes). The erm genes encode for a methylase that inhibits the 23S rRNA of the 50s ribosome and the *mef*(A/E) genes encode for a drug efflux pump (Pankuch et al., 1998, Tait-Kamradt et al., 1997, Del Grosso et al., 2002). The *mef* gene has been shown to be transferable between many bacterial species including *S. pneumoniae* and *H. influenzae* which are both found to co-colonise in the respiratory tract (Luna et al., 2000, Brandileone et al., 2016, Leach et al., 2016). There are conflicting reports on the presence or absence of the AMRGs in NTHi. One study claimed to have identified at least one AMRG using PCR assays in 106 NTHI isolates investigated from paediatric cystic fibrosis patients all showing intermediate to full resistance. However, an Australian study was unable to ascertain the presence of any AMRG in 172 clinical isolates, 59 of which were from cystic fibrosis and 98.8% displaying intermediate or full resistance to azithromycin (Roberts et al., 2011, Clark et al., 2002, Atkinson et al., 2015, Peric et al., 2003).

The genotype responsible for azithromycin resistance in NTHi has not always been successfully identified therefore resistance mechanisms in NTHi require further investigation (Atkinson et al., 2015, Clark et al., 2002). There is no current data on resistance to azithromycin in *H. haemolyticus*. (Witherden et al., 2014).

1.3.3 Adhesins of NTHi

For microbial populations to colonise the host they must overcome mucins, penetrate the epithelial layer, infiltrate defensive protein structures in the extra cellular membrane and evade the immune system (Goldsby, 2000). NTHi, like many other bacteria, have evolved to overcome these defences, and adhesin proteins have been shown to play a pivotal role.

Although *H. haemolyticus* and NTHi are genetically very similar and have shown the ability to transfer genetic content, there is still a variation in the behaviour of both species (Pickering et al., 2016). *H. haemolyticus* are largely non-pathogenic commensals and has shown cytotoxicity on binding to epithelial cells whereas NTHi are opportunistic pathogens and have shown adhesion and invasion capabilities (Pickering et al., 2016, Singh et al., 2016a). Defining the adhesin genotype therefore may deliver a reason for why these bacterial species act differently and potentially provide further gene markers for differentiation.

1.3.4 The Trimeric Autotransporters – Hmw, Hia, Hap and Hsf

High molecular weight (Hmw) and *H. influenzae* adhesion (Hia) along with *Haemophilus* adhesion protein (Hap) are trimeric autotransporters observed throughout Gram negative bacteria that secrete proteins across the periplasm and outer membranes of the cell wall to be expressed on the cell surface for interaction with host cells (Figure 1.6) (Meng et al., 2008, Surana et al., 2004, St Geme and Cutter, 2000, Spahich et al., 2014, Spahich et al., 2012, Grass and St Geme, 2000, Laarmann et al., 2002, Cutter et al., 2002). The autotransporters consist of a C-terminal anchor domain which embeds itself into the outer membrane creating a pore out to the extracellular area enabling the passenger domain to travel through the membrane emerging at the bacterial cell surface facilitating interaction with host cells (Figure 1.6)(Linke et al., 2006, Surana et al., 2004).

Hmw proteins 1 and 2 are NTHi adhesins that are capable of binding to a variety of cells with a high level of adherence and have been shown to outcompete *hmw* deficient strains in colonising rhesus macaques (St Geme et al., 1998, Yang et al., 1998, Vuong et al., 2013, St Geme et al., 1993, Rempe et al., 2016). Conversely though *hmw* positive strains have been reported as significantly (p=0.0001 – p=0.004) more predominant in the middle ear of otitis media than in nasopharyngeal or throat swabs in healthy patients indicating a potential survival advantage of *hmw* positive strains in the middle ear or selective pressure due to environment (Vuong et al., 2013, St Geme et al., 1998, Ecevit et al., 2004, Davis et al., 2014b, Xie et al., 2006). This has been disputed by the finding in a study of 97 strains varying in geographical and disease origins which observed no association between presence of *hmw* and site of isolation (De Chiara et al., 2014).

Strains of NTHi positive for *hia* are generally mutually exclusive to *hmw* strains although a small number (3.1%-8.3%) containing both *hmw* and *hia* genotypes have been observed (Barenkamp and St Geme, 1996, St Geme et al., 1998, Satola et al., 2008, St Geme and Cutter, 2000, Ecevit et al., 2004, Cardines et al., 2012). Hia positive strains are not as numerous as their *hmw* counterparts with 8.3% -33% of strains found to harbour *hia* compared to 45% - 89% positive for *hmw* genes (St Geme et al., 1998, Ecevit et al.,2004, Satola et al., 2008, Cardines et al., 2012). Unlike *hmw* positive strains *hia* has not been observed to be more prevalent at any site of collection within the respiratory tract or within any disease state (Ecevit et al., 2004).

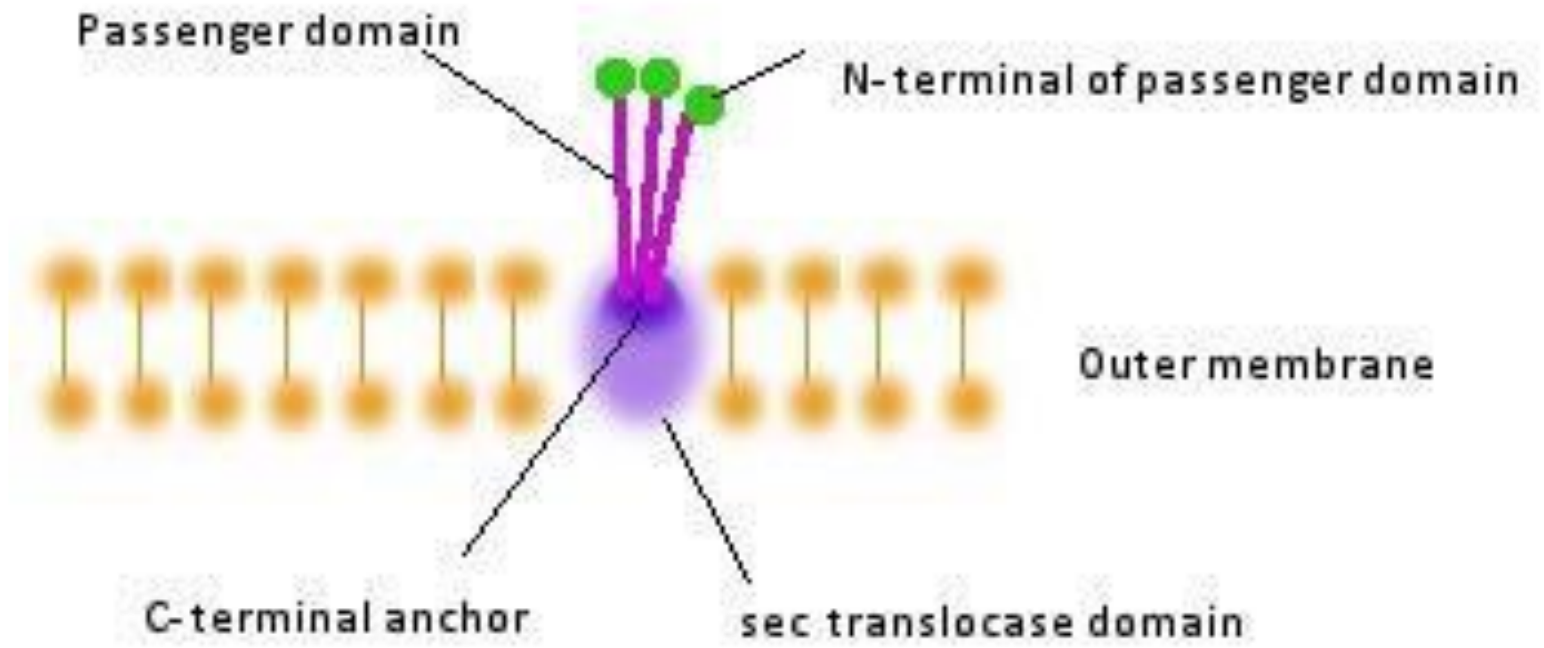


Figure 1.6 Trimeric auto transporter structure.

The passenger domain travels up through the C-terminal membrane anchor to the bacterial cell surface.

Hmw, Hia and Hap are associated with highly adherent and invasive strains of NTHi suggesting that they play an important role in virulence (Vuong et al., 2013, Rodriguez et al., 2003, St Geme and Cutter, 2000). NTHi positive for *hia* and *hmw* have been identified as causative strains in paediatric meningitis (Cardines et al., 2007). Hib, although a frequent cause of meningitis and invasive disease, when present in carriage is not known to contain Hmw, Hia and Hap proteins.

Insertion elements are short DNA sequences that can move from one location on the genome to another. Insertion elements can cause mutations by relocating inside a functional gene. Insertion element IS1016 has been identified in 9.6% of NTHi strains, all of which are *hmw* negative and 66-95% are *hia* positive. (Winter and Barenkamp, 2014, St Geme et al., 1998, Barenkamp and St Geme, 1996, Satola et al., 2008, Singh et al., 2014, Singh et al., 2015, Ecevit et al., 2004). IS1016 also contains a portion of the cap locus but does not translate into a capsule, suggesting a potential origin of the insertion element from capsulated strains (Satola et al., 2008).

Haemophilus surface fibril (Hsf) protein is a trimeric autotransporter and homologue of Hia found within encapsulated strains (Singh et al., 2014, Ecevit et al., 2004, Satola et al., 2008, Singh et al., 2015, Laarmann et al., 2002). The Hsf protein binds to a glycoprotein called vitronectin within the extra cellular matrix (ECM)(Singh et al., 2014). This binding interaction has been shown to increase levels of adherence and internalisation of Hib strains as well as masking them from the membrane attack complex and therefore evading phagocytosis (Singh et al., 2014).

Hia has two binding domains homologous to two of the three binding domains present in Hsf (Cotter et al., 2005, Radin et al., 2009, Laarmann et al., 2002). Although the binding properties of Hia within the ECM have yet to be determined there is the possibility that Hia may bind to vitronectin in a similar manner to Hsf (Singh et al., 2014). A single amino acid change within Hsf has been shown to affect the binding capacity of the acidic binding pockets in conserved encapsulated strains, *hia* is however more associated with the genetically heterogenic population of NTHi and has shown a marked variance across strains (Radin et al., 2009). Similarly, *hmw* and *hap* also display considerable sequence variation within NTHi (Giufre et al., 2006, Winter and Barenkamp, 2014, Davis et al., 2014b, Cutter et al., 2002).

hap has been observed in 96-100% of NTHi and interacts with members of the ECM namely laminin, collagen IV and is the primary ligand of fibronectin (De Chiara et al., 2014, Vuong et al., 2013). Additionally, Hap is capable of autoproteolysis and paradoxically, of creating microcolonies (Meng et al., 2011, Fink et al., 2003, Fink et al., 2001, Hendrixson and St Geme, 1998, Klemm et al., 2006). Microcolonies form due to bacterial aggregation, in the case of Hap this is mediated by the secretory leucocyte protease inhibitor (SLPI), an enzyme found in respiratory secretions that inhibits the autoproteolysis mechanisms resulting in autogenic binding of Hap serine protease (Hap$_S$) (Hendrixson and St Geme, 1998, Klemm et al., 2006). Once Hap$_S$ accrues to an optimal concentration, autoproteolysis occurs overcoming the actions of SLPI, Hap$_S$ is cleaved from the bacterial cell and released as an extracellular protein (Hendrixson and St Geme, 1998). The autoproteolytic ability displayed by Hap is counterintuitive for microcolony formation or adherence but is thought to both enable disruption of the epithelial environment to mediate colonisation or disperse individual cells from microcolonies for migration purposes (Hendrixson and St Geme, 1998, Fink and St Geme, 2003, Kenjale et al., 2009).

NTHi have been observed in biofilms which are thought to play an important role in pathogenesis and microbial survival within OM and mucosal surfaces (Hall-Stoodley et al., 2006, Post, 2001, Ehrlich et al., 2002). The prevalence of *hmwA* positive strains within OM suggests that Hmw proteins may play some role within these associated biofilms (Hall-Stoodley et al., 2006, Post, 2001, Ehrlich et al., 2002, Davis et al., 2014b). Microcolony formation by Hap, is thought to be an important step in biofilm progression however one *in vitro* study has reported that this is not the case and there is no observed requirement for Hap in biofilm formation (Euba et al., 2015, Meng et al., 2011, Hendrixson and St Geme, 1998). Proteins Hmw and Hia have also been associated with biofilm formation in cystic fibrosis patients with Hia positive strains resulting in denser biofilms than Hmw positive strains (Cardines et al., 2012).

Both Hmw and Hia proteins are immunogenic, however phase variation has been observed in both genes resulting in mediated protein expression (Winter and Barenkamp, 2014, Atack et al., 2015, Cholon et al., 2008). Reduction in Hmw protein expression has been observed in persistent strains isolated from COPD over time and correlated to an increase in 7bp repeats (Davis et al., 2014a, Cholon et al., 2008, Giufre et al., 2008). Reduced Hia expression has been associated with the extension of a polythymidine tract within the promoter region of *hia* (Atack et al., 2015).

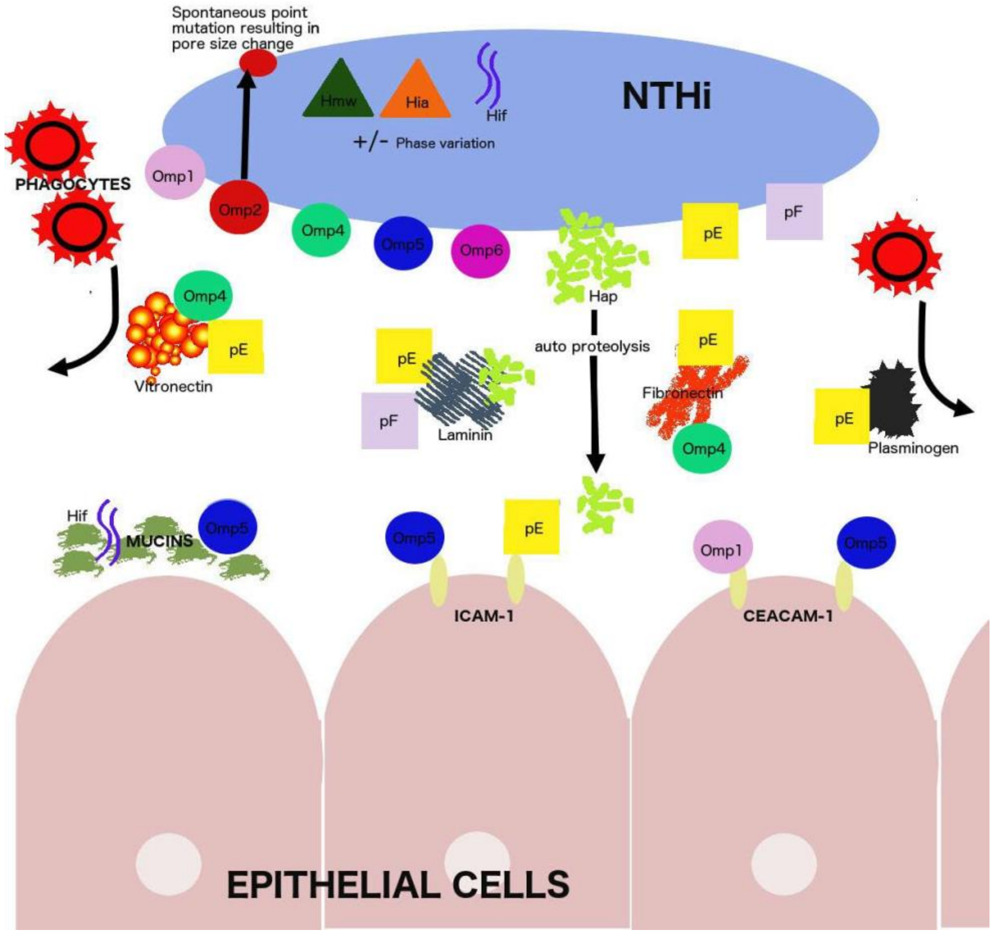


Figure 1.7 Known secondary roles and receptor interactions of the NTHi adhesins.

Hif (Hf), Omp1(P1), Omp2(P2), Omp4(4), Omp5(P5), protein E (PE), protein F (PF), Hia, Hmw and Hap. Hif binds to mucins and displays phase variation. Omp5 also binds to mucins and has ICAM-1 and CEACAM- 1 cell receptors. Omp1 binds to CEACAM-1 only and Protein E binds to ICAM-1, laminin, plasminogen, vitronectin and fibronectin. Hia shows phase variation and may also have an interaction with vitronectin. Omp4 binds to vitronectin and fibronectin. Hap also binds to laminin and is able to form microcolonies. Omp2 creates spontaneous point mutations. And differing pore sizes. Hmw displays phase variation.

1.3.5 Outer Membrane Proteins 1, 2, 4, 5, 6 and Proteins E and F

The outer membrane proteins vary in their roles for adherence and virulence. Nasopharyngeal mucins clear bacteria from the respiratory epithelium and inhibit colonisation however porin protein Omp2 (*omp2*) and fimbriae protein Omp5 (*ompA*)are able attach to nasopharyngeal mucins assisting in colonisation of NTHi (Reddy et al., 1996, Jiang et al., 1999, Gallaher et al., 2006).

Omp5 encoded for by *ompA* has been associated with both the epithelial cell receptors intercellular adhesin molecule 1 (ICAM-1) and carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM-1) (Hill et al., 2001, Tchoupa et al., 2015). Omp1 is reported to be the primary ligand for CEACAM-1 with no residual binding to Omp1 mutant strains observed unlike Omp5 which did display residual binding to CEACAM-1 in Omp5 mutant strains (Tchoupa et al., 2015, Garmendia et al., 2014, Hill et al., 2001, Euba et al., 2015). Cell internalisation has also been associated with Omp1 (Tchoupa et al., 2015). Omp5 binding to, and up-regulating ICAM-1 has been observed but differing reports on the actual adherence capacity of Omp5 strains leave its overall role not fully understood. No association has been observed between Omp5 strains and high adherence and no change in adherence capacity resulted from sequence variation within *ompA*, the gene that encodes Omp5 (Avadhanula et al., 2006, Garmendia et al., 2014, Vuong et al., 2013). Omp5 has however been reported to be critical in a significant (p=0.008) increase in adherence capacity in NTHi when adhering to epithelial cells pre-infected with respiratory syncytial virus (Jiang et al., 1999).

Contradictory results have again been reported for the role of Omp5 in biofilms with Omp5 being both identified as an important biofilm protein and required for growth but also conversely to be nonessential for biofilm formation (Euba et al., 2015, Puig et al., 2014, Murphy and Kirkham, 2002, Wu et al., 2014).

Similar to Omp5, Protein E is also able to up regulate ICAM-1 along with the chemokine IL-8 to self-propagate binding sites with a 50% reduction in adherence observed in a Protein E mutant strain (Frick et al., 2000, Ronander et al., 2009). Protein E can simultaneously bind to plasminogen, vitronectin and laminin in the ECM and is thought to be present in 96.9% of NTHi (Singh et al., 2013, Hallstrom et al., 2011, Singh et al., 2010). Plasminogenis activated into the serine protease plasmin once bound to a ligand, which impedes the complement pathway in multiple ways providing an additional immune evasion opportunity to that of vitronectin binding (Barthel et al., 2012b, Barthel et al., 2012a, Singh et al., 2011). Plasmin also degrades the ECM enabling further invasion and bacterial cell migration (Godier and Hunt, 2013).

Porin protein Omp2 constitutes approximately 50% of the outer membrane and like Omp5 also adheres to nasopharyngeal mucins (Reddy et al., 1996, Burns and Smith, 1987). Omp2 displays sequence variation which results in a change in pore size and a decrease in antimicrobial susceptibility to broad-spectrum treatments (Sikkema and Murphy, 1992, Regelink et al., 1999, Hiltke et al., 2002). Spontaneous point mutations in loop 6 of *omp2* have been observed in separately isolated strains suggesting that this could be an immune evasion mechanism (Duim et al., 1994). Omp2 and Omp1 are found in both NTHi and Hib and have also both been shown to be important for biofilm production (Munson et al., 1992, Wu et al., 2014).

Omp4 encoded by *hel* binds to vitronectin, fibronectin and laminin. Complement immune evasion through vitronectin binding has been observed alongside increased adherence in Omp4 (Su et al., 2016).

Although also an outer membrane protein Omp6 is not fully expressed on the surface, with the majority of Omp6 internally positioned into the periplasmic space of the cell and a smaller portion exposed (Michel et al., 2013). It has an association with peptidoglycan and is thought to play an important role in cell integrity by joining the cell wall and outer membrane (Murphy et al., 2006). Omp6 has also been associated with biofilm growth and has also has the ability to self-regulate its own expression by binding to its own gene (Murphy and Kirkham, 2002, Gallaher et al., 2006, Webster et al., 2006, Sikkema et al., 1992).

1.3.6 Type IV pili and *H. influenzae* fimbria (Hif)

NTHi produces pili in the form of a type IV secretion system encoded by genes *pil*A-D and *com*A-F with *pil*A being responsible for the major pili protein. The type IV pili have been associated with high levels of adherence to epithelial cells with ICAM-1 being the main receptor for PilA (Novotny and Bakaletz, 2016). Mutants of all *pil* and *com* genes except *com*C were observed to significantly (p=<0.5) reduce the ability of NTHi to adhere to human bronchial epithelial cells (Carruthers et al., 2012). The type IV pili has furthermore been identified as important for biofilm formation (Jurcisek and Bakaletz, 2007, Gallaher et al., 2006, Carruthers et al., 2012, Murphy and Kirkham, 2002).

H. influenzae fimbria (Hif) are surface protrusions encoded by *hif*A-E that are associated with adherence and prevent the physical removal of the strains via respiratory mucins (Weber et al., 1991, Kubiet et al., 2000). A 95% reduction in adhesion to mucins has been observed in non-fimbriated compared to fimbriated strains (Kubiet et al., 2000). Hif has been identified in Hib and NTHi but is more associated with non-invasive strains of the latter (Mhlanga-Mutangadura et al., 1998). Mediation of Hif levels are reported to be facilitated by phase variation of the *hifA* and *hifB* genes possibly again to avoid immune clearance or to enable the chronic colonisation of patients where bronchial secretions are not cleared such as in the case of cystic fibrosis. (Pichichero et al., 1982, Weber et al.,1991, van Ham et al., 1993, Kubiet et al., 2000). Adhesion to mucins differs between sample type with NTHi isolated from sputum more adhesive to mucins than those isolated from blood samples indicating an environmental response (Kubiet and Ramphal, 1995). Strains containing *hifB and hifC* have been more prevalently isolated from the throat of healthy subjects than those isolated from middle ear effusions samples from OM patients (Ecevit et al., 2004).

Protein F is thought to be a ubiquitous protein throughout NTHi and interacts with laminin in the ECM (Jalalvand et al., 2014, Jalalvand et al., 2013). A reduction of 64% in adherence to bronchial epithelial cells was observed in mutant Protein F strains (Jalalvand et al., 2013).

1.3.7 Immunogenicity and Potential Vaccine Candidates

As well as their pathogenic properties, interest in adhesins stems from the ability of many to evoke an immune response and as such, a number of these proteins have been investigated further for vaccine development purposes. A vaccine composing of epitopes from both PilA and Omp5 significantly reduced (p=<0.05) biofilm within the middle ear of chinchilla model and has since been granted a US patent (Novotny et al., 2009, Bakaletz, 2014, Novotny et al., 2011).

Hia and Hmw are highly immunogenic and are targets for opsonophagocytic activity (Winter and Barenkamp, 2014, Yang et al., 1998). *hia* however, is only observed in 8.3%-33% of strains and therefore may not be effective as a single target for immunisation purposes (Ecevit et al., 2004, Satola et al., 2008, St Geme et al., 1998, Cardines et al., 2012). Phase variation within *hia* and *hmwA* moreover mediates immune evasion in both proteins further reducing the likelihood of Hia being sufficiently prevalent to be useful as a sole vaccine candidate (Davis et al., 2014a, Atack et al., 2015, Cholon et al., 2008). HmwA and Hia are mutually exclusive within strains and anti-sera to Hmw was unable to eradicate Hia positive strains and vice versa (Winter and Barenkamp, 2014).

The outer membrane proteins are prime vaccine candidates due to their exposure on the cell surface. Immunoprecipitation studies of intranasal immunisation with outer member vesicles (OMV) from NTHi identified Omp1, Omp2, Omp5 and Omp6 as important antigenic proteins (Roier et al., 2012, Roier et al., 2015). Omp2 holds particular interest as it is the predominant surface protein and has been shown to invoke an immune response in multiple studies, though a drawback of using *omp2* as a vaccine component is the reported sequence variation, which has resulted in strain specific immunity (Murphy and Bartos, 1988, Troelstra et al., 1994, Yi and Murphy, 1997, Munson et al., 1992). Using recombinant Omp2 has resulted in a more cross-reactive response and a further investigation resulted in epitopes from the external loop structures being identified as potential vaccine candidates (Ostberg et al., 2009, Neary and Murphy, 2006, Neary et al., 2001). External loop structures five and six have displayed antigenic capabilities and although point mutations and resulting amino acid alterations have been reported, a conserved epitope from loop six found in a third of strains culminated in a multi strain response (Yi and Murphy, 1997,Duim et al., 1994, Neary and Murphy, 2006, Neary et al., 2001, Regelink et al., 1999). Omp5 and Omp1 are also heterologous in nature within NTHi therefore suitability for vaccine candidature has been questioned (Munson et al., 1992, Novotny and Bakaletz, 2003, Bolduc et al., 2000, Roier et al., 2012, Roier et al., 2014).

Proteins F, Omp4 and Hap are both thought to be ubiquitous within strains of NTHi, with Protein E present in 98.6% and all have been shown to solicit an immune response (Green et al., 1991, Hotomi et al., 2005, Roier et al., 2012, Roier et al., 2014, Green et al., 2005, Jalalvand et al., 2013, Su et al., 2013, Singh et al., 2011). Omp4 is an enzyme and therefore mutations to remove the enzymatic ability to ensure suitability for vaccine have been performed and resulted in successful recombinant candidates (Green et al., 2005). Recombinant Omp4 when bound to a cholera toxin resulted in clearance of intranasal carriage of NTHi in mice (Su et al., 2016, Hotomi et al., 2005). Immunisation with Protein F and Protein E resulted in pulmonary clearance of NTHi in mice (Ronander et al., 2009, Jalalvand et al., 2014). Hap$_s$ when attached to a cholera toxin adjuvant (CT-E29H) was shown to stimulate an enhanced immune response in intranasally administered mice reducing nasopharyngeal carriage (Cutter et al., 2002).

Due to the ability of Omp6 to elicit bactericidal antibodies there has been a lot of interest in the potential for Omp6 as a vaccine candidate despite the majority of the protein being internally positioned and not immunogenic (Michel et al., 2013, Wu et al., 2005, Sabirov et al., 2004, Sabirov et al., 2009, Pichichero et al., 2010, Hotomi et al., 1996, Hotomi et al., 2002, Hotomi et al., 1999). The surface exposed section of Omp6 is immunogenic with up regulation in IL-10, TNF-α and IL-8 documented and only a small percentage of surface exposed Omp6 found necessary to invoke an efficient immune response (Berenson et al., 2005). Investigations into intranasal immunisation in murine models with recombinant Omp6 and cholera toxin or adamantylamide dipeptide as adjuvants revealed clearance of NTHi from the sinuses, nasopharynx and middle ear intimating a method of protection against OM, sinusitis and lung infection (Sabirov et al., 2004, Bertot et al., 2004, Hotomiet al., 1996, Hotomi et al., 2002). Natural maternal antibodies to Omp6 and NTHi have been shown to reduce the risk of OM in breastfed babies however an increase in protection to NTHi infection passed onto the offspring in a mouse model has also been reported after maternal mice were intranasally immunised with Omp6 (Sabirov et al., 2009, Yamauchi et al., 2006). Evidence has shown that Omp6 also shows sequence variation and is not conserved in all strains of NTHi with 5.6% (eight from 151 strains) and 4.9% (eight from163 strains) displaying structural changes (Chang et al., 2011, Chang et al., 2010, Karalus and Murphy, 1999).

In summary, adhesins enable bacteria to adhere to cells within the host, the first crucial step of colonisation (St Geme et al., 1998, Yang et al., 1998, Vuong et al., 2013, Carruthers et al., 2012, Barenkamp and St Geme, 1996, Satola et al., 2008, Ronander et al., 2008, Jalalvand et al., 2013). Despite their ability to invoke immune responses, adhesins are important tools for colonisation and pathogenesis of NTHi utilising phase variation and binding to different glycoproteins within the ECM to avoid immune clearance (Cutter et al., 2002, Ecevit et al., 2004, Hallstrom et al., 2009, Barthel et al., 2012b, Berenson et al., 2005, Winter and Barenkamp, 2014, Atack et al., 2015, Cholon et al., 2008, Novotny et al., 2009, Fink et al., 2002).

The result of these interactions also enables further degradation of the ECM to enable cell migration and deeper infiltration of the basement membrane (Godier and Hunt, 2013). Contradictory studies of genotypes responsible for biofilm formation reveal that the adhesins role in NTHi biofilms is yet to be fully determined however biofilms have been observed within the middle ear of otitis media and respiratory tract of COPD patients (Cardines et al., 2012, Carruthers et al., 2012, Gallaher et al., 2006, Murphy and Kirkham, 2002, Euba et al., 2015, Jurcisek and Bakaletz, 2007).

Currently there is limited research into *H. haemolyticus* and the capabilities and mechanisms for adherence. It has, however, been reported that *H. haemolyticus* is unable to invade epithelial cells and shows different behaviour to NTHi on binding with epithelial cells (Singh et al., 2016a, Pickering et al., 2016). After 24 h *H. haemolyticus* displays a cytotoxic response unlike the invasive action of NTHi (Pickering et al., 2016, Clementi et al., 2014, Singh et al., 2016b). This could indicate that adhesin genotype may be one area that may differ substantially between the two species.

1.3.8 Advances in Next Generation Sequencing and Comparative Genomics

Since the discovery of structural DNA in 1953, methods to investigate genes and genomes have continually developed and improved. First generation sequencing started with the development of the Sanger chain termination method in 1977, the first automated systems were introduced by Applied Biosystems in 1987 and by 2005 the development of the Genome Analyser™ by Solexa, and the birth of ‘Next Generation Sequencing’(NGS) enabled massive scale parallel sequencing and increased the resulting read length significantly. NGS technologies now enable a reasonably priced method for in depth genomic analyses with sequencing by synthesis currently the most widely adopted method as illustrated in Figure 1.8.

Reductions in sequencing costs have seen the technology out compete Moore's Law prediction that has seen the cost of a human genome fall from US$340000 in 2008 to US$4200 in 2015 (Muir et al., 2016). The advancements and accessibility of sequencing have yielded much greater insights into bacterial species, enabling outbreak analysis in both nosocomial infections such as *P. aeruginosa*, VRE or MRSA or food/water borne occurrences from bacteria such as *C. jejuni* (Lahti et al., 2016, Harvala et al., 2016, Revez et al., 2014a, Revez et al., 2014b). Whole genome sequencing has enabled the monitoring of evolution and spread of pathogenic strains and antimicrobial resistance (Almofti et al., 2011, Caro-Quintero et al., 2009, Feil et al., 2004, Morley et al., 2015). Behaviour of bacterial species in infection and their adaptation can also be traced using whole genome sequencing (McNally et al., 2013, Darch et al., 2015, Kozitskaya et al., 2004, Morley et al., 2015). In addition, *in silico* taxonomy methods enable comparison of whole genomes to identify species similarity (Richter and Rossello-Mora, 2009, Konstantinidis and Tiedje, 2005b).

The distributed genome hypothesis describes how bacteria employs a strategy of interchanging genes from a supragenome pool to ensure persistence in chronic diseases (Ehrlich et al., 2010). The supragenome, alternatively named the pan genome, is not therefore contained within one particular strain but spread across naturally transformable strains which are able to utilise a full complement of genes to their advantage resulting ina genetically divergent species (Ehrlich et al., 2010). Whole genome sequencing enables the compilation of the pan genome, however it also enables identification of the core genes that are found in 100% of strains in addition to the accessory genes i.e those that are interchangeable. For the first time the full extent of the pan genome along with defining the core and accessory components has been ascertained by accumulation of sequencing data for a species (Hogg et al., 2007, van Tonder et al., 2014). This will facilitate better understanding of gene distribution, virulence and disease potential for clonal strains (Walkowiak et al., 2016, van Vliet, 2016, Liu et al., 2016).

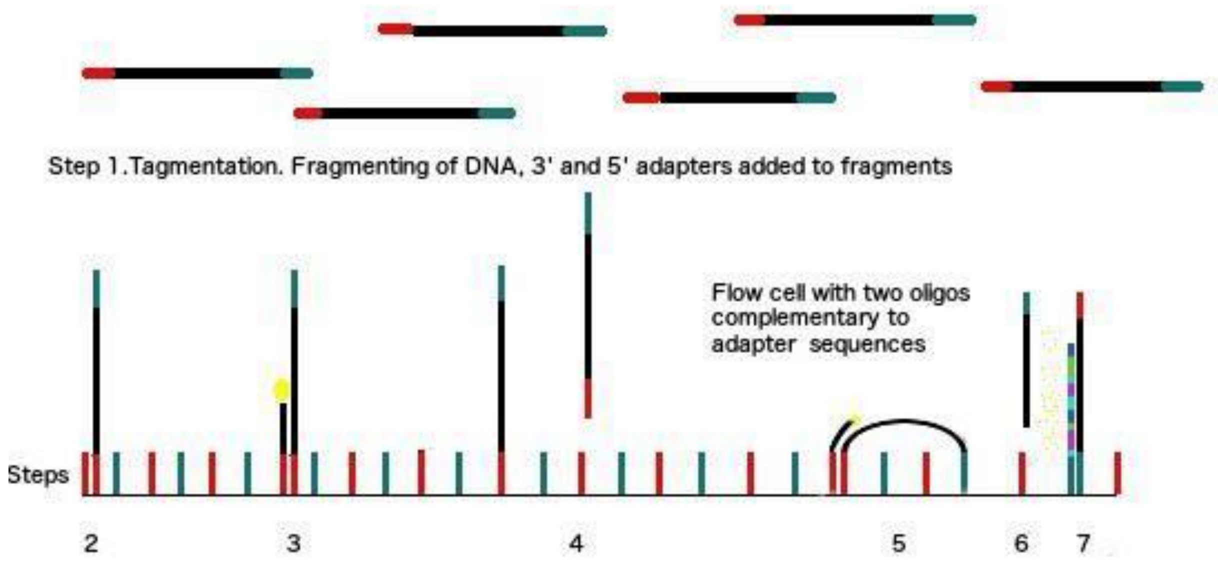


Figure 1.8 Sequence by synthesis method.

Initially tagmentation involves fragmenting DNA and adding adapters to the ends of the fragments. The adapters contain indices and regions complimentary to two types of oligo that are attached in a lawn to a flow cell. The adapters on the fragments attach to the first oligo (step 2). A polymerase extends a complimentary strand to the original fragment (step 3) which on completion of the complimentary strand, the original is washed away (step 4). The resulting strands bend and attach to the remaining oligo and polymerases extend the strand creating a double stranded bridge (step 5), the bridge is denatured resulting in two strands attached to the flow cell that undergo bridge amplification. This step is carried out repeatedly resulting in clusters of identical strands through clonal amplification (step 5). When clonal amplification is complete the reverse strands are cleaved and washed off. Four of each dNTP tagged each with unique fluorescence are released and competitively attempt to bind to the strand at the next available position with only the complimentary base to the strand being successful. A light source excites the unique fluorescent signal on the successfully attached base informing the sequencer of the nucleotide at that position (step 7). This continues until the read length is reached (step 7). Once the strands have been read the process is repeated only this time using the second oligo and reverse strand. The stronger the signal intensity the higher the base call which is the quality measurement for the certainty of the base pair noted at that position within the read.

1.4 Aims, Objectives and Hypotheses

1.4.1 Summary

NTHi has been identified as having a role in exacerbations in COPD and is the main pathogen isolated from the COPD lung (Bandi et al., 2003, Berenson et al., 2006a, Erb-Downward et al., 2011, Moghaddam et al., 2011, Murphy et al., 2004, Murphy et al., 2007, Wang et al., 2016). *H. haemolyticus* has been isolated from COPD patients as well as healthy children and those prone to otitis media (Murphy et al., 2007, Kirkham et al., 2010). Standard culture based microbiology methods are limited in their ability to both identify and monitor evolution and genetic variation within *H. influenzae* and *H. haemolyticus*. Molecular methods offer a more detailed insight and can generate better resolution. The two species appear to be too similar to discriminate using any one method; this infers a high rate of horizontal gene transfer between the two species potentially allowing for the spread of resistance and virulence factors (Binks et al., 2012, de Gier et al., 2015, Ridderberg et al., 2010).

The misidentification of *H. haemolyticus* as NTHi has implications for treatment as inappropriate administration of antimicrobials could result in increased resistance in *H. haemolyticus* (Marti et al., 2016). This could in turn allow the transfer of resistance mechanisms through horizontal gene transfer from *H. haemolyticus* to the pathobiont *H. influenzae* (Witherden et al., 2014). In addition, the misidentification of *H. haemolyticus* as NTHi may lead to the underestimation of the pathogenic role of *H. haemolyticus* in disease (Jordan et al., 2011, Anderson et al., 2012, Morton et al., 2012).

Molecular methods give a more in depth view with MLST allowing for clustering of related strains (Meats et al., 2003, Maiden, 2006). However, this has been found to be inadequate in identifying atypical strains of NTHi as strains lacking the *fuc*K gene cannot be typed (de Gier et al., 2015, Ridderberg et al., 2010). Traditional PCR allows for the presence of target genes to be detected however *in silico PCR* allows allele concatenation to inform the relationship between strains with more resolution.

The relatively recent advances in sequencing technology, with the associated reduction in costs have enabled the whole genome analysis (WGA) of large sample sets. WGA enables a complete view of the variation of NTHi within collections of samples. Comparing whole genomes of strains will potentially identify more specific regions of interest and further characterise the relationship between *H. haemolyticus* and NTHi.

1.4.2 Hypothesis

Atypical strains of non-typeable *H. influenzae* (NTHi) isolated from the COPD lungdisplay genetic diversity resulting in a species continuum between NTHi and *H. haemolyticus.*

1.4.3 Aims

The overall aim of this study is to investigate the genetic diversity of *H. influenzae* and *H. haemolyticus* isolated from the COPD lung, identifying regions of interest that may assist in the understanding of virulence, pathogenesis and resistance within these strains and also potentially inform vaccine development.

1.4.4 Objectives

Specific objectives are as follows:

1. To ascertain whether there is a continual spectrum of genotype between *H. influenzae* and *H. haemolyticus* present in COPD.
2. To identify the role of adhesins from NTHi and *H. haemolyticus* in COPD.
3. To investigate the potential genotypes describing antibiotic resistance to azithromycin within NTHi isolated from COPD.
4. To identify the pan-genome of NTHi strains isolated from COPD and the core genome associated with NTHI and *H. haemolyticus*

2 Methods

2.1 Study Details and Sample Section

2.1.1 Isolates - AERIS Study

The isolates to be analysed in this study were generated from the AERIS COPD study (REC ref: 11/H0502/9) as previously detailed (Bourne et al., 2014). In summary, sputa were collected monthly and on event of an exacerbation, where possible, from a planned cohort of 120 patients with moderate to severe COPD. Sputa collected from patients were processed within a 6 hour window inoculated onto culture media and identified using standard microbiology methods in the Southampton Public Health England (PHE) laboratory. A single colony of *H. influenzae* from the culture media was suspended into skimmed milk, tryptone, glucose, and glycerin (STGG) (Bio-trading, Mijdrecht, The Netherlands) and stored at -70^{o}C. Where possible, ten isolated colonies were separately collected and stored for each sputum sample. These samples were further identified as *H. influenzae* using a PCR assay targeting *lgtC* by GSK before being delivered back to the University of Southampton for further investigation.

2.1.2 Reference Isolates

Reference isolates were purchased from the PHE National Culture of Type Cultures (NCTC) and used throughout assays (Table 2.1).

Table 2.1 Details of NCTC reference strains used throughout study

NCTC Number	Species	Details of isolation
4842	*H. influenzae*	Isolated from respiratory tract, 1935. Biotype III
7279	*H. influenzae*	Meningitis, 1947. Biotype I.
8467	*H. influenzae*	Cerebrospinal fluid, meningitis, 1944. Biotype I.
10659	*H. haemolyticus*	Sputum, chest infection, 1969.
10839	*H. haemolyticus*	Infant nasopharynx. 1972.

2.2 Phenotypic and Culture Based Methods for Differentiating NTHi from *H. haemolyticus*

2.2.1 Initial Identification of NTHi

Identification of NTHi was achieved by colony morphology on chocolate agar (Oxoid, Basingstoke, UK) and requirement of X+V factors (Oxoid, Basingstoke, UK) on blood base agar (Oxoid, Basingstoke, UK). Samples were inoculated onto chocolate agar and incubated at 37°C in 5% CO_2. Colonies that were morphologically identified as potential NTHi were further inoculated onto blood base agar for identification of growth factor requirements as detailed below. Positive isolates for X+V factors only were suspended into STGG (BioTrading, Mijdrecht, The Netherlands) and stored at -80°C. For all AERIS isolates culture identification was initially carried out in the Public Health England laboratories at University Hospital Southampton NHS Foundation Trust.

2.2.2 X+V Growth Factors

All isolates in STGG were defrosted at room temperature and then streaked onto a chocolate agar plate (5% haemolysed blood agar, Oxoid, Basingstoke, UK) using a 10µl sterile loop. After overnight incubation at 37°C at 5% CO_2 duplicate colonies were picked from the chocolate agar plate using a 1µl sterile loop and inoculated onto the middle of blood base agar plate (Oxoid, Basingstoke, UK). A sterile cotton wool swab was dipped into distilled water and used to spread the bacteria over the plate. Three discs containing X+V, X and V growth factors (Oxoid, Basingstoke, UK) were positioned in a triangular formation with sufficient room between them to ensure no merging of the growth factors occurred. The plates were incubated at 37°C at 5% CO_2 overnight. An isolate was determined as positive for X+V growth factor requirement if the only growth on the plate occurred around the disc containing both X+V factors. If growth was observed around the other discs the isolate was discarded from the study.

2.2.3 Haemolysis

Haemolysis was determined by growth on blood agar. Isolates were first defrosted at room temperature. A 10µl sterile loop was then used to inoculate Columbia blood agar (5% horse blood, non-haemolysed, Oxoid, Basingstoke, UK). The plates were incubated at 37°C in5%

CO_2 overnight and checked for distinguishable haemolysis around colonies. The plates were re-incubated and checked again at 48 h and 72 h.

2.2.4 Hydrogen Sulphide Production

The presence of hydrogen sulphide was ascertained by the inclusion of a lead acetate strip (Sigma-Aldrich, Shaftesbury, UK) in the lid of a chocolate agar plate. A 1µl loop was used to inoculate the plate from the STGG containing the isolate. This, along with the lead acetate strip was incubated overnight at 37°C at 5% CO_2. After incubation, the lead acetate strip was checked for blackening indicating the isolate positive for production of hydrogen sulphide. All plates were re-incubated and re-investigated after 48 h.

2.2.5 Antibiotic Susceptibility

Isolates were removed from frozen storage and allowed to defrost at room temperature. Samples were inoculated onto chocolate agar (Oxoid, Basingstoke, UK) and incubated at 37°C at 5% CO_2 overnight. Resulting growth was resuspended into Mueller-Hinton broth (Oxoid, Basingstoke, UK) at a density of MacFarland standard 0.5. This was used to inoculate *Haemophilus* test media (HTM) plates (BD, Oxford, UK) in a thorough lawn pattern. Plates were left to dry for no longer than 10 minutes. Etests® (Biomerieux, Basingstoke, UK) for azithromycin were placed on the centre of the plate. The HTM plates were incubated at 37°C at 5% CO_2 overnight. The minimum inhibitory concentration (MIC) of each isolate was read as the point on the gradient where the bottom of the ellipse of growth made contact with the Etest®.

2.3 Next Generation Sequencing of *Haemophilus* spp. isolates.

2.3.1 Extraction of Genomic DNA

H. influenzae isolates stored in STGG were removed from frozen storage at -70°C and allowed to defrost at room temperature. Using a 10µl sterile loop in a class II hood the samples were inoculated onto chocolate agar (Oxoid, Basingstoke, UK) in a streaking formation. The agar was incubated at 37°C in 5% CO_2 overnight. Ten colonies from the resulting growth were suspended in 180µl of ATL buffer supplied within the QiaAmp mini DNA extraction kit (QiaGen, Manchester, UK). To each suspension, 20µl of proteinase K (QiaGen, Manchester, UK) was added and these were then placed into a water bath at 56°C for 2 h. DNA extraction was then performed using the QiaAmp mini kit (QiaGen, Manchester, UK) as per the manufacturer's instructions.

2.3.2 DNA Quantification

DNA quantification was performed using a Qubit® fluorimeter (Life Technologies,Paisley, UK) according to the manufacturer's instructions.

2.3.3 Short Read Sequencing - MiSeq™

Extractions were diluted to 0.2ng ul^{-1} using the following equation to calculate the required amount of DNAse/RNAse free water (Gibco, Paisley, UK):

$$\left(\frac{sample\ concentration}{0.2}\right) - 1$$

De novo sequencing was performed in batches of 24 extractions using a MiSeq™ benchtop sequencer (Illumina®, Saffron Walden, UK). DNA library preparation was completed using the Nextera® XT DNA Prep Kit (Illumina, Saffron Walden, UK) following the manufacturer's instructions. Indices from the Nextera® XT DNA Prep Kit were added to tag each sample (

Table 2.2, Table 2.3). Amplification steps were carried out on the Veriti™ Thermocycler (Applied Biosystems®, Paisley, UK) starting with a hold temperature of 72°C for 3 min followed by 12 cycles of 95 °C for 40s and 55°C for 30 s, then 5 min at 72°C and a hold temperature of 10°C. PCR purification was carried out using AMPure® XP beads (Beckman Coulter, High Wycombe, UK) according to manufacturer's instructions. Following washing in ethanol, beads were resuspended in 50µl of resuspension buffer to elute the DNA. Eluted DNA was then transferred to PCR strip tubes. Library normalisation and preparation was carried out using the Nextera® XT DNA prep kit (Illumina,Saffron Walden, UK). The samples were loaded onto the MiSeq™ cartridge, which was placed on the MiSeq™ for sequencing. This resulted in fastq files containing the resulting paired end short read (2x250bp) sequences. Successful run thresholds were cluster densities of ~1000K/mm^2 and quality scores of >80% reads at Q30.

Table 2.2. Indices used to identify sequences from sample on Illumina MiSeq.

Index 1 Primer set	Sequence	Index 2 primer set	Sequence
N701	TAAGGGCGA	N502	CTCTCTCTAT
N702	CGTACTAG	N503	TATCCTCT
N703	AGGCAGAA	N504	AGAGTAGA
N704	TCCTGAGC	N505	GTAAGGAG
N705	GGACTCCT	N506	ACTGCATA
N706	TAGGCATG	N507	AAGGAGTA
N707	CTCTCTAC	N508	CTAAGCCT
N708	CAGAGAGG		
N709	GCTACGCT		
N710	CGAGGCTG		
N711	AAGAGGCA		
N712	GTAGAGGA		

Table 2.3. Example of sample sheet indicating index 1 and 2 primers in a mix specific to each sample.

Indices	N502	N503	N504	N505
N701	Sample 1	Sample 2	Sample 3	Sample 4
N702	Sample 5	Sample 6	Sample 7	Sample 8
N703	Sample 9	Sample 10	Sample 11	Sample 12
N704	Sample 13	Sample 14	Sample 15	Sample 16
N705	Sample 17	Sample 18	Sample 19	Sample 20
N706	Sample 21	Sample 22	Sample 23	Sample 24

2.3.4 Sequencing Statistics

Quality analysis was performed using fastQC (Andrews, 2012). Depth or coverage was ascertained by using the following equation:

$$C=LN/G$$

Where L is the average read length, G is the expected genome size, in this case 1,830,138bp defined by the first whole *H. influenzae* genome sequenced (KW20 RD) and N equals the number of reads (Fleischmann et al., 1995).

2.3.5 Assigning Sequence Type to Isolates Through Multi Locus Sequence Typing

Paired end read analysis was performed using the *Haemophilus* MLST database using SRST2 on short read sequence data (Inouye et al., 2012). Sequence type was determined and recorded using the seven housekeeping genes used for sequence typing *H. influenzae* (*adk, atpG, frdB*, *fucK*, *mdh*, *pgi* and *recA)* (Meats et al., 2003). MLST data was visualised and compared to public database of MLST data using eBurst (Feil et al., 2004). Phyloviz was used to generate a visualisation of neighbour joining tree of MLST data (Francisco et al., 2012).

2.3.6 Assembling Genomes and Assessing Quality

Trimmomatic was used to trim adapter sequences from Fastq raw reads (Bolger et al., 2014). Following trimming, *de novo* assembly was done using MaSuRCA (Zimin et al., 2013). A script was written to enable batch assemblies with MaSuRCA. The script enabled the preparation of folder structures for assembly, addition of required file and extraction and renaming of final assembled genome files (Appendix 1). Assemblies were then assessed for quality.

All assemblies were passed through the assemblathon perl script to calculate statistics related to the success of the assembly (Earl et al., 2011). The metrics that were used are as follows:

- Number of contigs – a contig is a continuous sequence that has been assembled from the shorter reads in the sequencing data. Ideally a whole genome would sit in one contig however this is currently unachievable from short read sequence data. Therefore, the larger the number of contigs the more fragmented the assembly. Ideally, when taken into account with the other two statistics, fewer contigs means a less fragmented and more successful the assembly. Assemblies were re-investigated where an assembly returned more than 300 contigs, size of assembly, N50 and fastQC results were consulted and a decision made for a need for re-sequencing the isolate.
- N50 – This figure is the length at which all contigs of this length and larger contain at least half of the size of all lengths of the assembly. The N50 offers a rough median average compared to the resulting assembly size. The larger the N50 length the less fragmented the assembly. This is used in context with the other two statistics however a minimum of 30,000 is preferable (Earl et al., 2011).
- NG50 – This figure is similar to the N50 but offers greater accuracy as it is compared to the expected genome size rather than the assembly size (Earl et al., 2011).
- Size of the assembly – This should be around the known size for the *H. influenzae* genome. Severe deviations from the 1,830,138 base pairs of the initial *H. influenzae* KW20 Rd genome assembled should be re-investigated (Fleischmann et al., 1995).

2.3.7 Reporting Genotypes

PROKKA was used to annotate assembled genomes using default settings resulting in annotated files (Seemann, 2014).

Mapping of paired-end reads using specifically chosen GenBank references (Table 2.4)was done using SRST2 (Inouye et al., 2014, Benson et al., 2016). CD-HITS was used to prepare the GenBank reference for use with SRST2 (Appendix 1) (Inouye et al., 2014, Benson et al., 2016, Fu et al., 2012). Minimum coverage was set to 60. Maximum divergence set to 40. Resulting fasta files contained mapped consensus allele for each isolate for each target gene if present.

The *in silico PCR* experiment simulation system (IPCRESS) tool allows multiple primers to be applied to sequencing data simultaneously. IPCRESS mimics traditional PCR with sequences complementary to the primers being identified within the assembled genome data and resulting in sequences between the primers (Table 2.5). It is part of the Exonerate package developed and supplied by the European Bioinformatics Institute (EBI), which results in a putative sequence (Slater and Birney, 2005). IPCRESS was used to ascertain gene presence or absence where stated and was used in the command line (Appendix 1). Mismatching of primers was set to three base pairs.

For antibiotic resistance genotyping published primers (Table 2.5) for resistance genes *erm*(A/B/C/F) and *mef*(A/E) along with L4 and L22 were applied using IPCRESS with mismatch set to a maximum of 3 base pairs. Primers were tested against positive control sequences shown in Table 2.6.

Confirmation of gene presence was also obtained by searching for gene names within annotations using exonerate to default settings and the grep function (Slater and Birney, 2005).

Confirmation of allele identification was carried out when required by comparing the mapping result to the BLAST database online using the megablast algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi)(Figure 2.1)(Altschul et al., 1990).

Multiple small sequences were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and pairwise protein sequence alignments were carried out using EMBOSS NEEDLE (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) to default settings (Sievers et al., 2011, Rice et al., 2000).

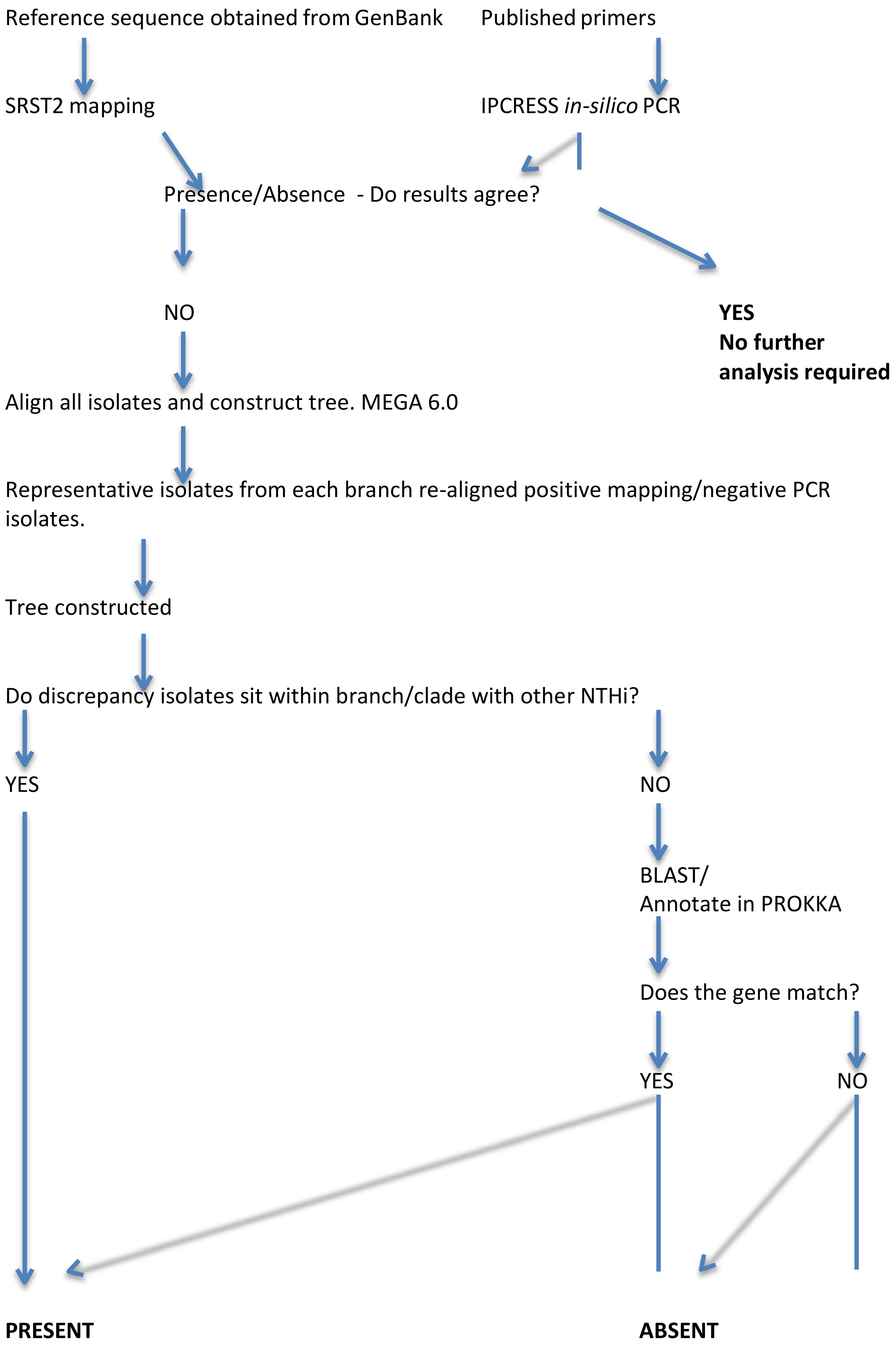


Figure 2.1 Identifying presence and absence of gene markers in sequence data.

Table 2.4 GenBank reference sequences used for mapping

Gene	**GenBank Accession**	**Ref**
*fuc*K	CP009610.1 (287354..28876)	Mell *et al.*, 2013, (unpublished)
*fuc*P	CP000057.2 (823471…824757)	(Harrison et al., 2005)
hap	U11024.1	(St Geme et al., 1994)
hia	U38617.2	(Barenkamp and St Geme, 1996)
*hmw*A	AY601283.1	(Ecevit et al., 2005)
*hmw*2A	U08875 (352…4785)	(Barenkamp and St Geme, 1994)
hif	U19730.1	Green,B.A. and Olmsted,S.B. 1995 (unpublished)
hpd	CP002277.1 (1826771…1827865)	VanWagoner *et al.*, Strain R2866 , 2004 (unpublished)
hsf	U41852.2	(St Geme et al., 1996)
iga (conserved beta core)	M87492 (4124…4978)	(Sandstedt et al., 2008)
*lgt*C	CP002277.1 (315813…316835)	VanWagoner *et al.*, Strain R2846 , 2004 (unpublished)
*omp*1	CP000057.2 (491079…492455)	(Harrison et al., 2005)
omp2 (P2)	CP000057.2 (216249…217846)	(Harrison et al., 2005)
*omp*4 (hel)	M68502.1	(Green et al., 1991)
*omp*5 (P5)	L20309.1	(Munson et al., 1993)
omp6 (P6)	M1939.1	(Nelson et al., 1988)
*pil*A	CP000057.2 (396415…396864)	(Harrison et al., 2005)
protein E	CP002276.1 (463373…463855)	VanWagoner *et al.*, Strain R2866 , 2004 (unpublished)
*smp*B	NC_000907.1 (1038488…1038973)	(Fleischmann et al., 1995)
*sod*C	AFQQ01000001.1 (155446…156018)	(Jordan et al., 2011)

Table 2.5 *In silico* PCR primers used for ascertaining presence/absence of gene markers and antibiotic resistance alleles within *Haemophilus* spp.

Gene	**Forward**	**Reverse**	**Ref**
iga (β core)	TGAATAACGAGGGGCAATATAAC	TCACCGCACTTAATCACTGAAT	(Sandstedt et al., 2008)
hpd	CAAAGTGTTGAAAAAATATGGCTATGA	GTTGCACCTGATTTATTCAATAATGC	(Pickering et al., 2014a)
*fuc*K	ACCACTTTCGGCGTGGATGG	AAGATTTCCCAGGTGCCAGA	(Meats et al., 2003)
*fuc*P	GCCGCTTCTGAGGCTGG	AACGACATTACCAATCCGATGG	(Price et al., 2015)
*lgt*C	CGGACTGTCAGTCAGACAATG	CTCAAAATGATCATACCAAGATG	(McCrea et al., 2008)
omp2	GGTGAAGTAAAACTTGGTC	GCGCCTAATACTAAACCATC	(Hiltke et al., 2002)
omp6	ATGAACAAATTTGTTAAATCA	TTAGTACGCTAACACTGC	(Murphy et al., 2007)
*smp*B	ATTAAATGTTGCATCAACG	GACTTTTGCCCACGCAC	(Reddington et al., 2015)
*sod*C	GTGCGGTATGTTCAGTTG	AGGCCATAGTTAGATTCAGTAAT	(Latham et al., 2015)
ermA	ACAAGACAACGTAATAGAAATCG	CTTTTATATTCTCAGAGGGGTTTAC	(Atkinson et al., 2015)
ermB	TCATCCTAAACCAAAAGTAAACAG	TTAGTAAACAGTTGACGATATTCTC	(Atkinson et al., 2015)
ermC	AGTACAGATATAATACGCAAAATTG	ATGCCAATGAGCGTTTTG	(Atkinson et al., 2015)
ermF	TTGAAACTTGTCTATGAGGTAGG	CTCTAACAGACAGGAAATAAATGC	(Atkinson et al., 2015)
mefA	GGATCGTCATGATAGGAAGAAG	CAGGTAGCTCCATATAGAATGC	(Atkinson et al., 2015)
mefE	CCTAAGCTGGGTAATCAAGTG	CTAAGAGTAATAAGGCAAACAATCC	(Atkinson et al., 2015)
L4	TTAAGCCGGCAGTTAAAGC	CACTTAGCAAACGTTCTTG	(Peric et al., 2003)
L22	CGGCAGATAAGAAAGCTAAG	TGGATGTACTTTTTGACCC	(Peric et al., 2003)
23S	LN831035.1(270311-273445) GenBank reference used in SRST2 for gene mapping.	Unpublished. Wellcome Trust Sanger Institute, March 2015	GenBank

Table 2.6 GenBank reference sequences used as positive controls for primers forantibiotic resistance.

Primers for gene	GenBank reference	Ref
ermA	X03216.1	(Murphy et al., 1985)
ermB	M11180.2	(Shaw and Clewell, 1985)
ermC	NC_001395.1	(Catchpole et al., 1988)
ermF	AF21923.1	(Fletcher et al., 1995)
mefA	U70055.1	(Clancy et al., 1996)
mefE	U83667.1	(Tait-Kamradt et al., 1997)

2.3.8 Phylogenetic Analysis of Consensus Alleles from Gene Markers

Consensus alleles were collated into one fasta file and aligned in MEGA 6.0 using MUSCLE UPGMB clustering method on default settings (Tamura et al., 2013, Edgar, 2004). Neighbour joining or maximum likelihood trees were created using MUSCLE with bootstrap reiterations set to 500 resulting in Newick files (Edgar, 2004). Newick files were used to create radial trees in FigTree 1.4.2 or used alongside metadata files in microreact (http://www.microreact.org) (Rambaut).

Core alignments from ROARY were used to create maximum likelihood trees constructed in RAxML using the GTRCAT model for heterogeneity on CIPRES (https://www.phylo.org/) on default settings (Stamatakis, 2014, Miller et al., 2015).

2.3.9 Pan Genome and Gene Association

ROARY was used to calculate the pan genome from annotated genomes using default settings (Page et al., 2015).

Scoary is a python script for developed to calculate association of genes present with a phenotypic trait from a metadata file and ROARY output gene_presence_absence.csv. Metadata files were created in text files containing information on chosen isolates and the relevant trait alongside the gene_presence_absence.csv ROARY output containing a presence/absence matrix of annotated genes in all the isolates (Page et al., 2015, Brynildsrud et al., 2016). Scoary was used with all default parameters except the standard number set to -12.

2.3.10 Visual Representation

R studio was used for graphical representation for graphs, barcharts and dotplots using ggplots2 (http://ggplot2.org/) (Team, 2015) .
Pan genome visualisations of heatmaps for distribution and plots were created using create_pan_genome_plots.R (https://github.com/sanger-pathogens/Roary/blob/master/bin/create_pan_genome_plots.R) and python script roary_plots.py in the command line (https://github.com/sanger-

pathogens/Roary/blob/master/contrib/roary_plots/roary_plots.py) in R studio (https://www.rstudio.com/) (Team, 2015)

2.4 Whole genome taxonomy

2.4.1 Speciation using MetaPhlAn

MetaPhlAn enables species identification from metagenomic sequencing data by comparison of the sample data to identified marker genes held within the MetaPhlAn database (Segata et al., 2012). MetaPhlAn returns the percentage similarity of the samples to the genus and species as far as possible. MetaPhlAn was used to ascertain whether there was any part of the unassembled sequences that belonged to a different genus other than *Haemophilus* or a different species than *H. influenzae* (Segata et al., 2012). In addition, MetaPhlAn establishes the genetic similarity of an isolate to the MetaPhlAn held reference *H. influenzae.*

2.4.2 Average Nucleotide Identity

Average nucleotide identity (ANI) is an *in silico* method that randomly fragments genomes into small sequences that are compared against each other. ANI was calculated using the MUMmer alignment algorithm using the python module pyani (https://github.com/widdowquinn/pyani) (Altschul et al., 1990, Kurtz et al., 2004). The recommended threshold to identify strains from the same species is >95%. (Goris et al., 2007, Konstantinidis and Tiedje, 2007, Konstantinidis and Tiedje, 2005a).

2.5 Statistical Analysis

Chi squared analysis or Fisher's exact test where appropriate were carried out in SPSS v 23 (IBM, New York, USA). P values <0.05 were considered significant. Duplicates from isolates were removed for use in SPSS using parameters patient ID, date, ST and presence or absence of adhesin genes investigated to ascertain relationship between adhesin genes and onset of exacerbation events.

3 Differentiating between *H. haemolyticus* and NTHi genotypes

3.1 Introduction

At a phenotypic level, NTHi and *H. haemolyticus* can only be differentiated by the haemolytic capability of the latter on blood agar. However, this phenotype can be lost thereby rendering culture an imperfect technique for differentiating between these two morphologically similar species. This has led to the misidentification of *H. haemolyticus* as NTHi with an error rate of 0.5%- 39.5% as discussed in the Introduction (section1.2.1).

The importance of differentiating between both species arises from the potential for ineffective or inappropriate administration of antibiotics, affecting antimicrobial stewardship, which in turn could impact on increasing resistance levels. There is also arisk of potentially missing invasive cases of *H. haemolyticus* (Anderson et al., 2012, Mortonet al., 2012).

To overcome the issue of phenotypic misidentification, PCR assays to differentiate between the two species have been developed. However, due to the genetic heterogeneity of NTHi, no single assay has been able to unambiguously define the species and thereby unequivocally differentiate from *H. haemolyticus* (Binks et al., 2012, Pickering et al., 2014a, McCrea et al., 2008, Sandstedt et al., 2008). Multi-locus sequence typing (MLST) has been used to characterise the *H. influenzae* species, including NTHi. Seven housekeeping gene fragments are used, yet one of the schema, *fucK*, is reportedly absent in some strains of NTHi. Since NTHi is known to be heterogeneous, MLST has been questioned for its usefulness in characterising NTHi that do not return a typical genotype. Despite these reservations, it is hypothesised here that the use of multiple genes and MLST should enable differentiation between NTHi and *H. haemolyticus* and therefore ascertain the amount of *H. haemolyticus* within the *Haemophilus* spp. isolated within this study.

3.2 Hypothesis

Multiple genetic markers will be sufficient to differentiate NTHi from *H. haemolyticus*.

3.3 Aims

In this chapter, the rate of misidentification of *H. haemolyticus* as NTHi was retrospectively determined in a longitudinal study of NTHi isolated from sputa collected from COPD patients. The prevalence and variance in atypical NTHi was investigated by characterising isolates that display an atypical genotype in speciating PCR targets. In this chapter, the aims are:

- to identify atypical NTHi and potential *H. haemolyticus* using culture methods measuring haemolysis and hydrogen sulphide production.
- to establish sequence types for all isolates using the *H. influenzae* MLST schema, and
- to determine the genetic conservation of previously reported marker genes for the species.

3.4 Results

3.4.1 Multi Locus Sequence Typing

In this study, MLST was carried out on whole genome sequence data from 1460 isolates (Materials and Methods, section 2.3.5). These represented all isolates from sputa from 24 COPD patients over two years either at routine monthly appointments or during an acute exacerbation event (Materials and Methods, section 2.1). The subgroup of 24 patients were selected by a group of experts made up of GSK, University of Southampton and NHS professionals. This was based on the presence of bacteria isolated in the first six months of the study. Approximately ten colonies of *Haemophilus* spp. were picked from each culture plate.

Previous studies have used MLST to characterise the population structure of *H.influenzae* and reported the diversity of non-encapsulated strains compared to the more conserved lineages of capsulated *H. influenzae* (De Chiara et al., 2014, Kaur et al., 2011, Erwin et al., 2008). One of seven loci used in the MLST schema, *fucK*, has been shown to be absent in some strains of NTHi, an example of the heterogeneity within non-encapsulated strains of *H. influenzae* and rendering MLST useless to categorise these strains (de Gier et al., 2015, Ridderberg et al., 2010).

From the MLST data, 28 sequence types (ST) were identified that had been previously documented in the MLST database, and a further nine STs were novel to this study. Eight of the nine were curated to the MLST database as STs 154, 353, 356, 357, 1314, 1441, 1442 and 1664 (http://www.pubmlst.org). The ninth novel group that did not return a known ST were *fucK* negative isolates. The MLST schema includes *fucK* within the seven housekeeping genes and all seven are required to enable the allocation of an ST. A large number of isolates (n=146) were negative for the *fucK* gene and these isolates fell into two further groups. One group of 54 isolates returned identical known ST sequences in the remaining six loci of the MLST schema that were isolated from the same patient. The other 92isolates returned no previously documented sequence for any of the remaining six loci. The group of 54 isolates was characterised as *fucK* negative atypical NTHi and the second group of 92 as potential *H. haemolyticus*. A breakdown of all ST data can be found in Appendix2.

The relationship of the STs from this study to all STs held within the MLST database was carried out and displayed in an eBurst diagram (http://www.mlst.net) in Figure 3.1 (Materials and Methods, section 2.3.5). The spread of the STs throughout the full database shows the diversity of the isolates. A clonal group is classed as those STs sharing documented sequences at six out of the seven loci from the MLST schema (Feil et al., 2004).

Figure 3.2 shows the limited clonal group relationship between the STs found in this study. Only eleven STs from this study were related to each other by clonal group, four groups of two STs from a clonal group and one group of three STs. For five STs from the study there is no ST within the MLST database with six out of seven shared loci and these are referred to as singletons. This limited sharing of alleles between isolates may imply that the majority of the isolates and their associated STs are unrelated.

NTHi displayed diversity throughout patients with only six STs being isolated in more than one patient and only one ST in more than two patients (Kaur et al., 2011)(Figure 3.3). ST 57 was isolated in five patients. A previous study identified ST 57 as the predominant ST in 170 commensal and otitis media derived isolates from three different geographical locations (LaCross et al., 2013). Potential *H. haemolyticus* as defined by *fucK* negative status and with no previously identified MLST alleles as described above was isolated from eleven patients (Figure 3.3).

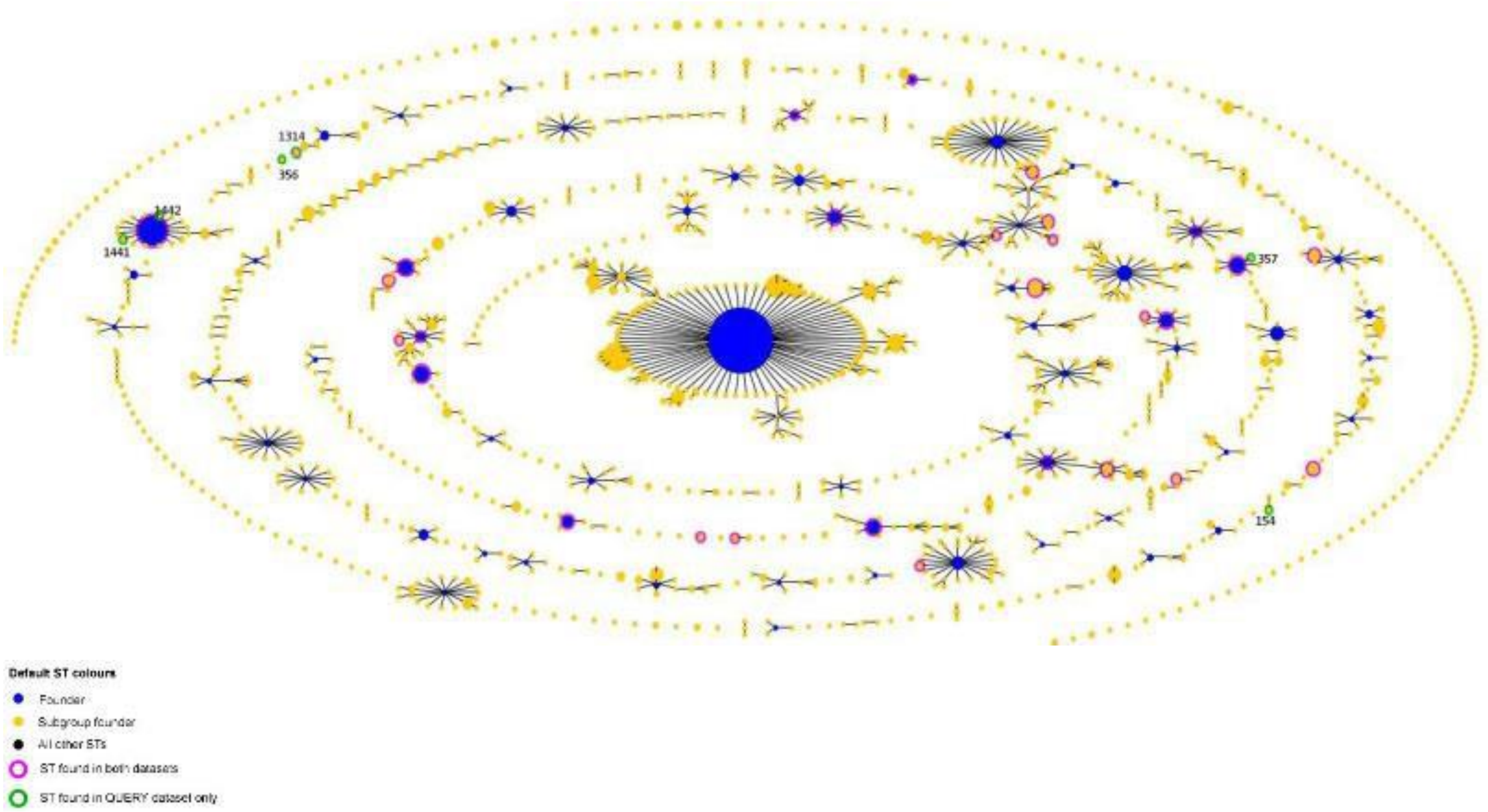


Figure 3.1 eBurst diagram displaying the distribution of STs from this study throughout the full MLST database for *H. influenzae*.

Each ST from this study is marked with either a pink or green circle. STs novel to the study are marked with a green circle and annotated with the ST number. Pink circles represent those STs found in the MLST database also found in the study. The diversity of the isolates can be seen by the wide distribution of the STs throughout the diagram. The clonal founders are marked in blue, to give perspective the central clonal group consists predominantly of encapsulated Hib STs demonstrating the conservative nature of the capsulated strains.

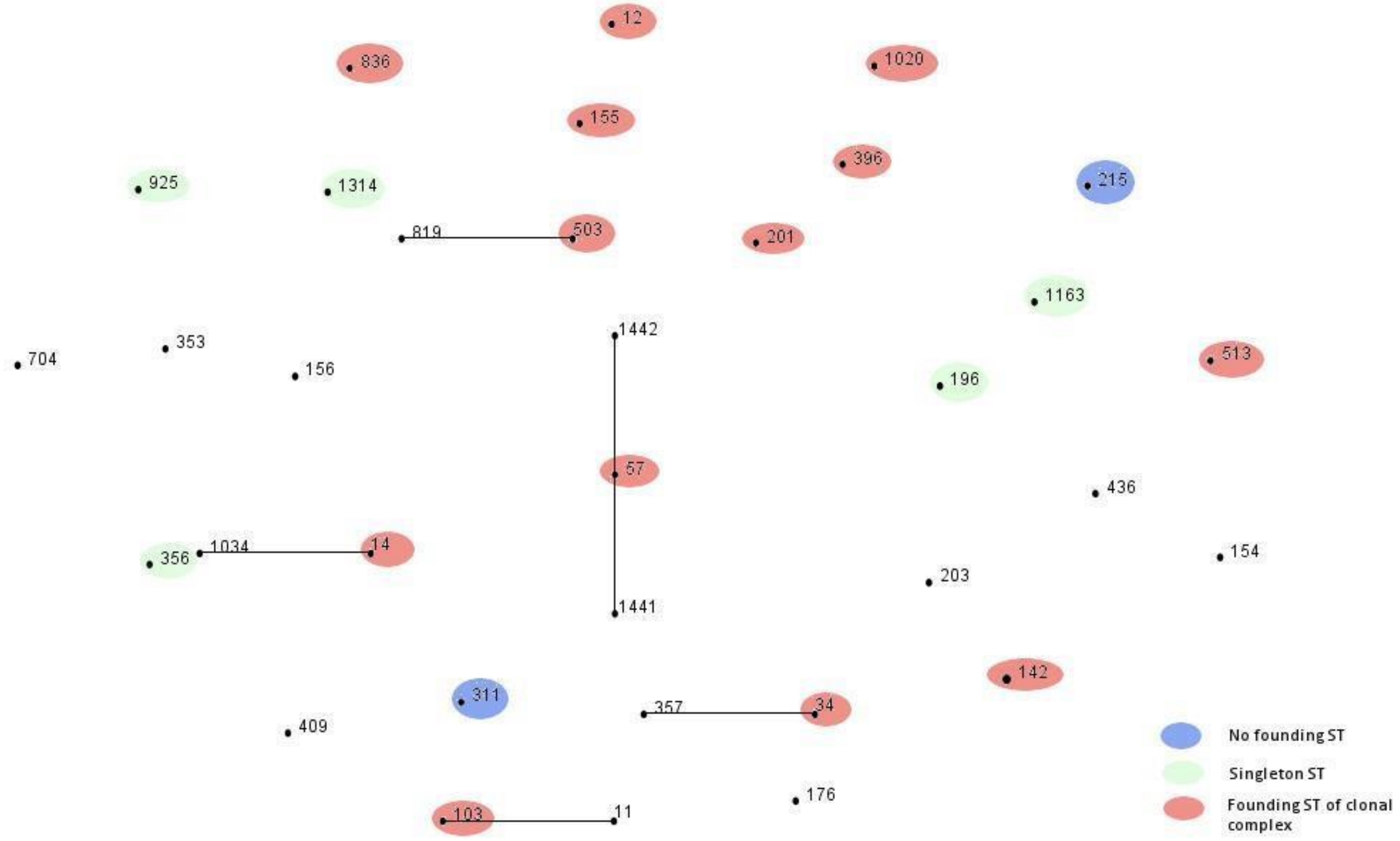


Figure 3.2 STs from MLST analysis showing distribution in clonal groups (STs sharing six out of seven MLST loci).

Only 11 STs were found to be related by clonal group. The maximum number of STs from one clonal group was three with ST 57 as clonal founder and STs 1441 and 1442 wthat were both novel to this study. The remaining STs were not related by clonal group. Thirteen STs were clonal founders of groups indicating a predominant ST in the Hi population. Five were singletons and not related to any other documented ST by clonal group and two STs were from groups without an identified clonal founder. The lack of clonal group relationships between STs from this study demonstrates the diversity of the NTHi isolated within this study.

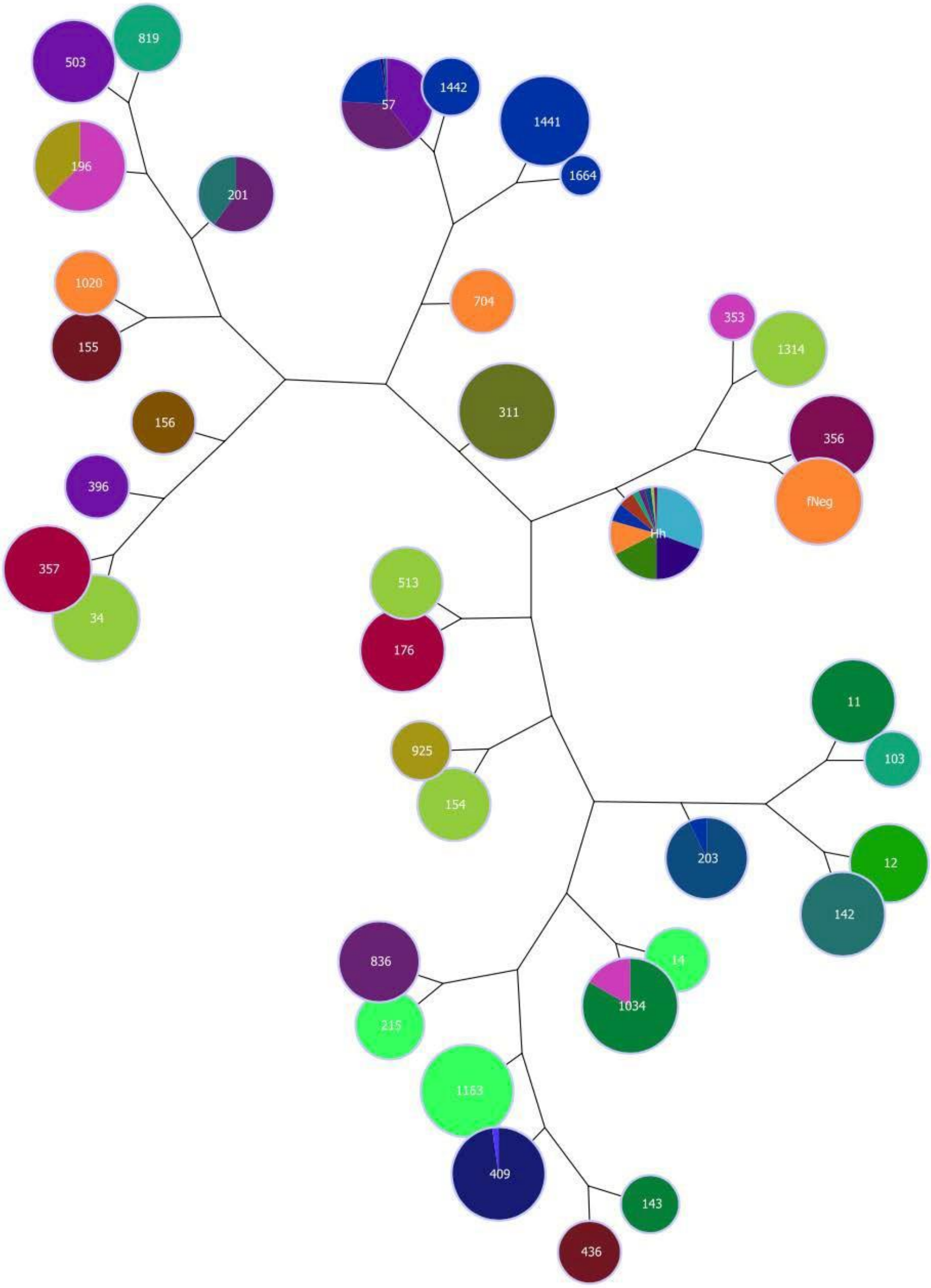


Figure 3.3 ST distribution by patient. Neighbour joining tree created in Phyloviz from MLST allele data downloaded from http://www.pubmlst.org (Materials and Methods, section 2.3.5).

ST is indicated by the number in each circle, each patient is represented by a different colour. STs 109, 201, 409, 203, 1034 and 57 were isolated in more than one patient. The novel ST's to this study are indicated with black lettering. Potential *H. haemolyticus* was observed in 11 patients, the Hh circle indicates the number of patients only and does not imply the diversity within the potential *H. haemolyticus* isolates.

3.4.2 Culture Based Testing for Haemolysis and Measures of Hydrogen Sulphide Production to Identify *H. haemolyticus* Isolates

The ability of culture based identification to differentiate between NTHi and *H. haemolyticus* has been questioned due to the morphological similarity, the dependence for both species on the growth factors nicotinamide adenine dinucleotide and haemin, and more importantly, the loss of haemolysis from strains of *H. haemolyticus* after passaging (Murphy et al., 2007, Fenger et al., 2012, Frickmann et al., 2014, Zhang et al., 2014). Previously however even haemolytic isolates of *H. haemolyticus* have been misidentified as NTHi (Murphy et al., 2007).

In order to identify any haemolytic samples in the present study, a culture based haemolysis assay was carried out on all 1460 isolates (Materials and Methods, section 2.2.3). The haemolysis assay involved inoculating each isolate onto a blood agar plate and incubating for up to 72 h. Plates were checked every 24 h for signs of haemolysis classified as clear zones in the agar around colonies. Twelve isolates were identified as having haemolytic capability and therefore discounted as NTHi. Only nine of the twelve haemolytic isolates displayed haemolysis after 24 h. Three isolates did not exhibit haemolysis until 48 h. No further isolates developed haemolysis between the 48 and 72 h indicating a potential optimum 48 h incubation time for three *H. haemolyticus* that display delayed haemolysis. All twelve isolates were part of the potential *H. haemolyticus* group identified by MLST.

Another phenotype that has been previously reported to differentiate between the two species is the ability for most *H. haemolyticus* to produce hydrogen sulphide (H_2S) although this phenotype has also been reported in a small proportion of NTHi (McCrea et al., 2008, Kilian, 1976b, Pittman, 1953). The production of H_2S can be confirmed by the result of a darkened lead acetate strip on the lid of a chocolate agar plate inoculated with the isolate after a 24 h incubation (Materials and Methods, section 2.2.4). This assay was carried out on a smaller subset of 111 isolates that represented each ST type from MLST analysis and all potential H. haemolyticus as identified from MLST analysis above. These isolates were taken from a sub cohort of 16 patients from the first six months of the study (Appendix 2). This resulted in a subset of 111 isolates of which 43 were potential *H. haemolyticus.* From the 43 potential *H. haemolyticus* 39 (90.7%) were able to produce hydrogen sulphide and from the 68 NTHi, 33 (48.5%) were positive for H_2S production.

Table 3.1 Results of haemolysis, hydrogen sulphide production and MLST analysis from 111 *Haemophilus* spp. isolates representing all STs or strains without an ST isolated from 16 different patients over 6 months.

n=111	Haemolysis status	H_2S status	ST (MLST) status	Species
12	+	+	-	*H. haemolyticus*
35	-	-	+	NTHi
33	-	+	+	Atypical
27	-	+	-	Atypical
4	-	-	-	Atypical

As discussed in the introduction and above, typical NTHi are defined as having a recognised ST by MLST analysis, no visible haemolysis on blood agar and no H_2S production capability. For *H. haemolyticus* it is the opposite. From the 43 potential *H. haemolyticus*, 12 isolates were typical for the expected results. Four of the remaining 31 were negative for MLST, haemolysis and H_2S and the remaining 27 were negative for haemolysis and MLST but positive for H_2S production. Thirty-five of the 68 NTHi were typical for the expected NTHi result. The remaining 33 were negative for haemolysis but positive for MLST and H_2S production. MLST identification and haemolysis were found to be significantly inversely associated (p=0.001, Chi^2 test). MLST identification and production of H_2S were also found to be significantly inversely associated (*p*=<0.001) (Appendix 3).

3.4.3 Molecular Markers for Speciation – Confirming Presence/Absence Status

Due to the inability of culture methods to identify between NTHi and *H. haemolyticus*, gene markers have been identified against which to develop PCR assays to differentiate between NTHi and *H. haemolyticus*. However as previously mentioned NTHi displays a large amount of genetic heterogeneity and therefore singular assays are not able to unequivocally discriminate between them from *H. haemolyticus* (Chang et al., 2010, Chang et al., 2011, Cardines et al., 2007, Connor et al., 2012). Molecular markers *omp2*, *omp6*, *lgtC*, *iga*, *sodC*, *fucK*, *fucP*, *hpd* and *smpB* previously selected for this purpose, were assessed for their presence and absence in the 1460 isolates from this study (Binks et al., 2012, McCrea et al., 2008, Reddington et al., 2015, Price et al., 2015, Meats et al., 2003, Latham et al., 2015, Zhang et al., 2014, Murphy et al., 2007, Pickering et al., 2014a, Wang et al., 2011, de Gier et al., 2016). This was done by using reference nucleotide sequences obtained from GenBank (https:// www.ncbi.nlm.nih.gov/genbank/) for each gene to map against sequencing data for each isolate (Materials and Methods, section 2.3.7). *In silico* PCR was also done using previously published primers on draft assembled genomes of all 1460 isolates as a comparable method for presence or absence of the genes within the isolates (Materials and Methods, section 2.3.7). From this data, genotypes for the isolates were defined depending on the presence and absence status of all the marker genes.

3.1.1.1 The Fucose Genes – Molecular Markers for NTHi

fucP and *fucK* are genes from the fucose operon; these are not found in *H. haemolyticus* but are found in typical strains of NTHi (Price et al., 2015, Binks et al., 2012). The MLST schema also uses *fucK* as one of the seven housekeeping genes to characterise *H. influenzae* (Meats et al., 2003).

Mapping was carried out on the isolates using reference sequence CP009610: 287354-288766 from GenBank for *fucK* (Materials and Methods, `Table 2.4`) and *in silico* PCRwas carried out using primers from the MLST schema first described by Meats *et al.* (2003) (Meats et al., 2003) (Materials and Methods, section 2.3.7). There were no discrepancies between the mapping and *in silico* PCR results. All 92 potential *H. haemolyticus* and 54/1368 (3.9%) NTHi were found to be negative for *fucK*. Mapping to *fucP* was carried out using GenBank reference CP00057.2: 823471-824757 (Materials and Methods, `Table 2.4`) and primers from Price *et al.* (2015) were used for *in silico* PCR (Materials and Methods, section 2.3.7). The group of 54 *fucK* negative isolates were also negative for *fucP.*

In addition, four isolates were negative for the presence of *fucP* using *in silico* PCR but positive for *fucP* when using mapping. To confirm the presence of the *fucP* gene, the mapped sequences from the *in silico* PCR negative isolates were aligned with mapped sequences taken from the positive *in silico* PCR isolates (Materials and Methods, section 2.3.7). The position of the four *in silico* PCR negative isolates within a resulting phylogenetic tree of the alignment shows them cluster with other NTHi positive for both mapping and *in silico* PCR (Figure 3.4). It is likely that the negative PCR result was returned due to contig breaks in the assembled genome occurring within the complimentary sequence for the PCR *in silico* primers. The 54 isolates negative for both *fucK* and *fucP* mapping and *in silico* PCR were also identified as *fucK* negative through MLST analysis.

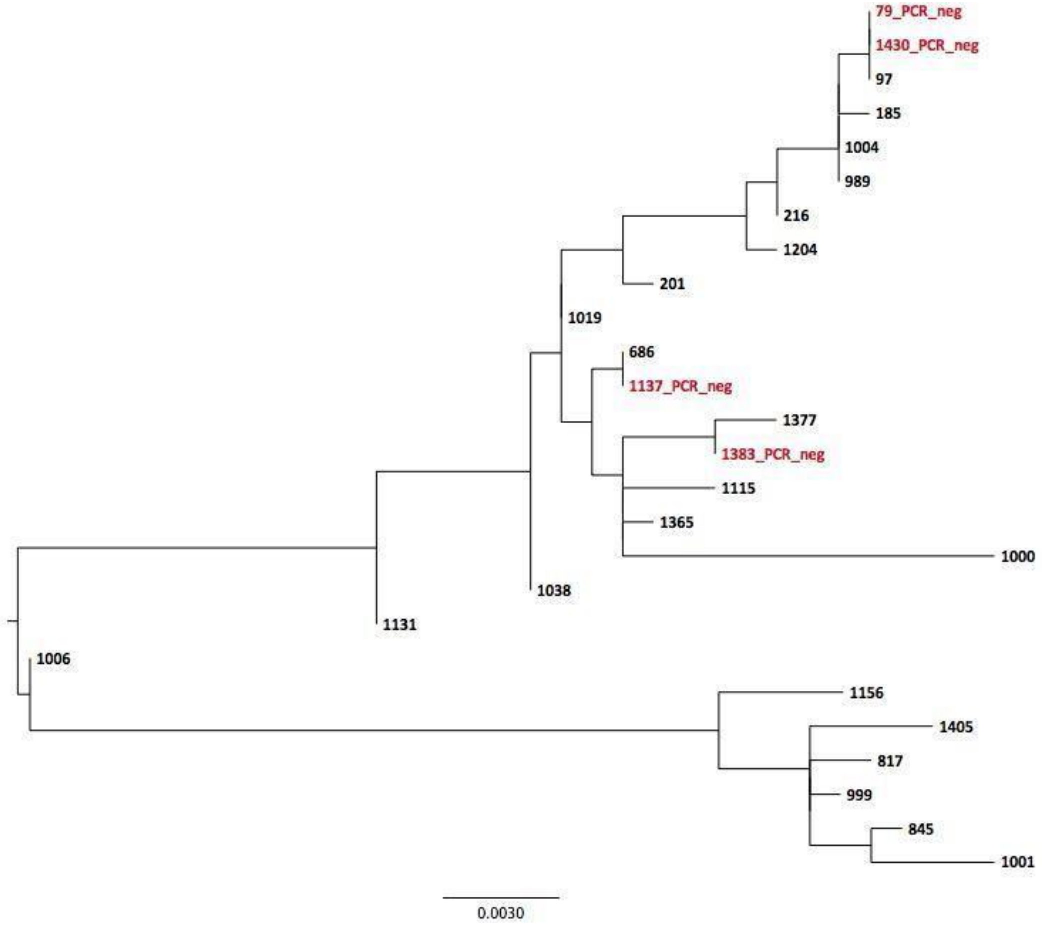


Figure 3.4 Alignment of *fucP* in silico PCR negative isolates.

fucP sequences resulting from mapping of four *in silico* PCR negative aligned against representative positive *in silico* PCR sequences in MUSCLE. A neighbour-joining tree was constructed in MEGA 6.0 and visualised in FigTree v1.4.2 from resulting alignment. All four isolates clustered within other NTHi sequences for *fucP*.

3.4.3.1 *lgtC*, *iga* beta core and *omp2* – Molecular Markers for NTHi

These three genes are not typically shown to be found in *H. haemolyticus* and have historically been used as a method of identifying NTHi in PCR assays (Binks et al., 2012, Sandstedt et al., 2008). Mapping for *lgtC* was carried out using GenBank reference sequence CP002277: 315813-316835 (Materials and Methods, Table 2.4) and primers were taken from McCrea *et al.* (2008)(McCrea et al., 2008) (Materials and methods, section 2.3.7).

All of the potential *H. haemolyticus* isolates were negative for *lgtC* in both mapping and *in silico* PCR. From the mapping 1301 (95.1%) NTHi returned a sequence for *lgtC*. Only 140/1368 (10.2%) NTHi isolates were positive for the *lgtC in silico* PCR. Alignments of all mapped sequences confirmed the correct identification in the 1301 isolates. On further investigation, the initial 231 base pairs of *lgtC* consists of tetrameric repeats (Figure 3.5). As the assembled genomes used in this study are not closed but draft genomes, and there may be issues with assembly of repetitive regions such as these then this could result in a negative *in silico* PCR result.

```
CLUSTAL O(1.2.2) multiple sequence alignment


NC-4842         ATGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA
100             ATGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA
1000            ATGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA
                ************************************************************

NC-4842         CAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA
100             CAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA
1000            CAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGA
                ************************************************************

NC-4842         CAGACAATGAATATTATATTTTCGTCTGACAACTATTATGCACCTTATTTAGCTGTAAGT
100             CAGACAATGAATATTATATTTTCGTCTGACAACTATTATGCACCTTATTTAGCTGTAAGT
1000            CAGACAATGAATATTATATTTTCATCTGACAACTATTATGCAACTTATTTAGCTGTAAGT
                ***********************.******************.*****************
```

Figure 3.5 Tetrameric repeats observed in *lgtC* sequence.

The first 121 base pairs ofthe *lgtC* gene consist mainly of tetrameric repeats. *In silico* PCR was applied to draftassembled genomes. Tetrameric repeats may have been troublesome to assemble in some cases, or may have span contig boundaries making the detection of the sequence for the primers inaccessible. NC-4842 is a reference isolate for NTHi. 100 and 1000 are representative isolates from PCR and mapping positive isolates.

The *iga* region is also known to be variable so a conserved *iga* beta core region was used both for mapping (GenBank reference M87492:4124-4978, Materials and Methods, Table 2.4) and a target for *in silico* PCR using previously published primers from Sandstedt *et al.* (2008) (Poulsen et al., 1992, Vitovski et al., 2002, Sandstedt et al., 2008) (Materials and Methods, section 2.3.7). All 1368 NTHi were positive for a sequence from mapping and six were negative for *in silico* PCR, these were deemed positive by alignment to the other positive mapping isolates as shown in Figure 3.6. A sequence for *iga* was observed in twelve *H. haemolyticus isolates*, these were also aligned with the positive NTHi and found to cluster as a separate clade as can be seen in Figure 3.6. Sequences returned from *H. haemolyticus* were much smaller on average than NTHi sequences with a mean average length of 581bp (524-720bp) compared to a mean average of 841bp (748-865bp) in NTHi. Twelve *H. haemolyticus* containing the truncated *iga* sequence were isolated from seven different patients with one patient displaying an isolate with a truncated *iga* sequence over two time points.

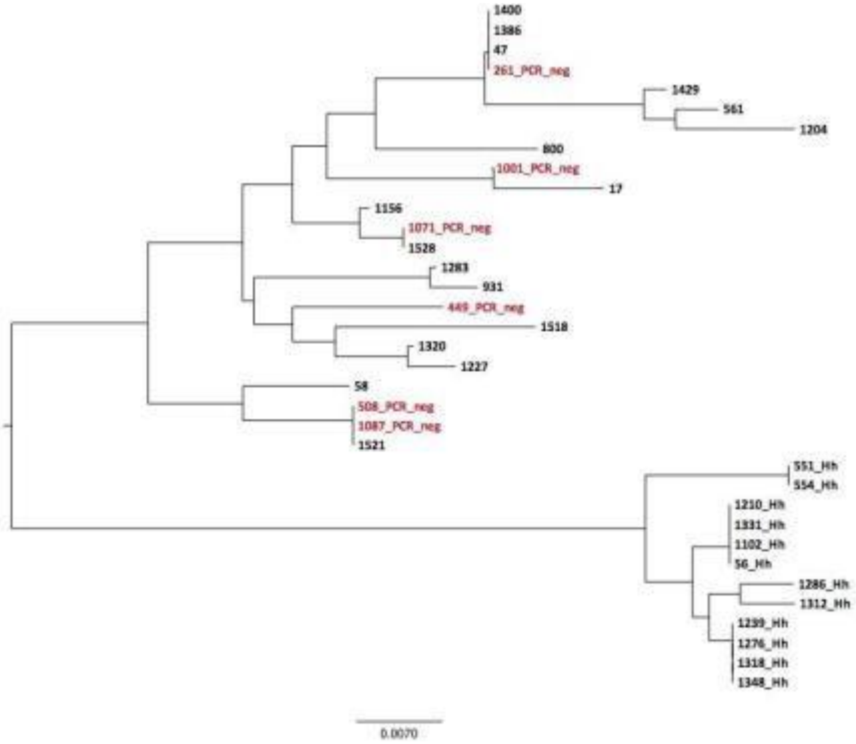


Figure 3.6 *iga* sequences from NTHi and *H. haemolyticus*.
Neighbour joining tree constructed in MEGA 6.06. Bootstrapping was set to 500.Red tips indicate mapping sequences from *iga in silico* PCR negative isolates which cluster with other *iga* mapping sequences from PCR positive isolates on branches indicating homology of *iga* sequence with other NTHi. The twelve *H. haemolyticus* marked "_Hh" clustered as a separate clade indicating a sequence variation from the NTHi. Scalebar at 0.007 substitutions per 100 base pairs.

Two GenBank reference sequences (GenBank reference CP00057: 216249-217346 and M93269.1, Materials and Methods, Table 2.4) were used to map the isolates to the *omp2* gene (Materials and Methods, section 2.3.7). *In silico* PCR was carried out using primers from Hiltke *et al.* (Hiltke et al., 2002) (Materials and Methods, section 2.3.7). No potential *H. haemolyticus* returned a sequence for either mapping attempt or *in silico* PCR. From the NTHi group 1250/1368 (91.4%) were positive from the initial mapping and a further fourteen isolates returned a sequence for the second mapping attempt which resulted in positive mapping sequences for 1264/1368 (92.4%), these were aligned to ensure homology between the sequences (Figure 3.7). Ninety-two of the 104 negative for mapping were also negative for *in silico* PCR. Fifteen isolates that were positive for *in silico* PCR but negative for mapping were deemed as negative. Interestingly among the negatives were all isolates of ST 353 (n=2), ST 356 (n=51) and ST 1314 (n=23), which were all novel STs identified in this study, as well as ST 819 (n=14) identified by MLST analysis.

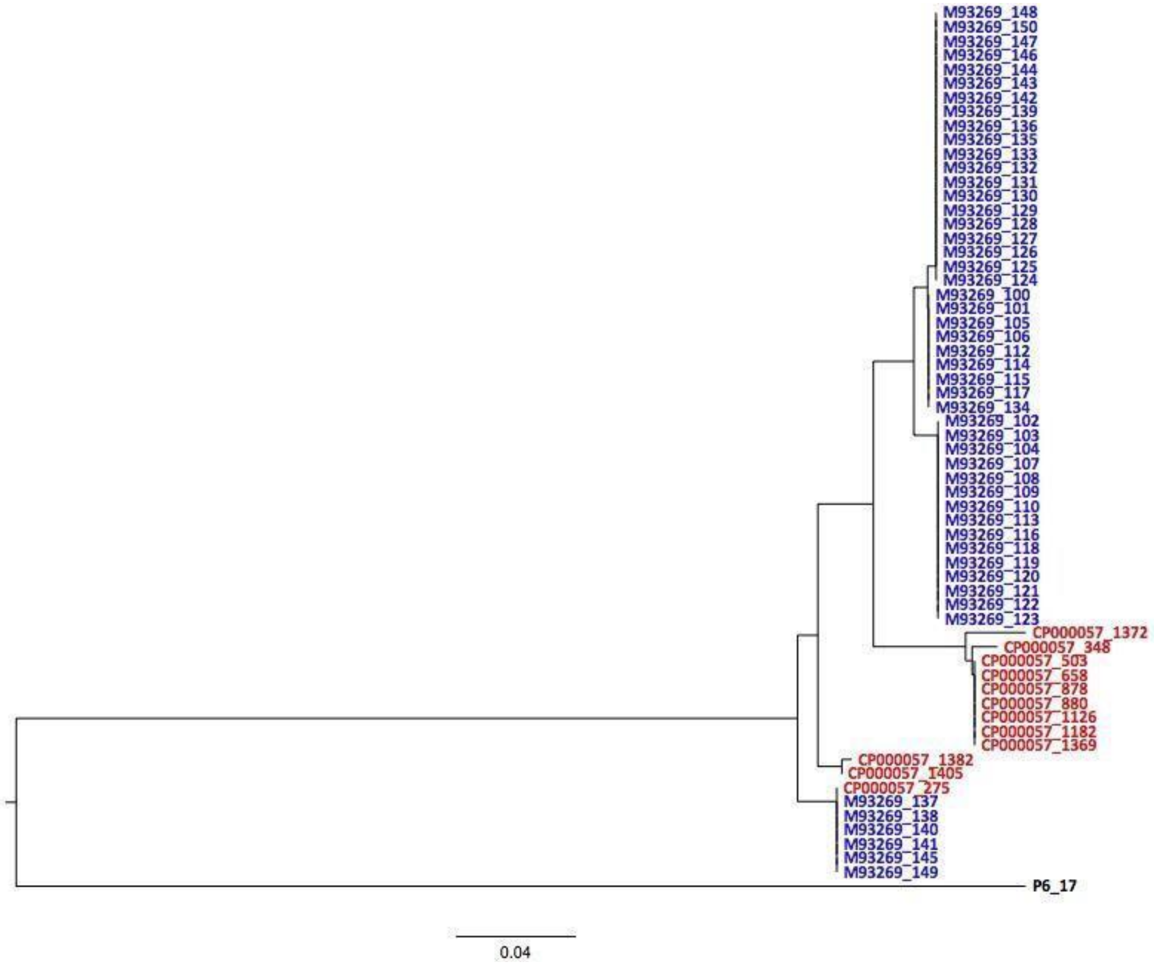


Figure 3.7 *omp2* sequences from two different GenBank reference sequences.

Neighbour joining tree was created in MEGA 6.06 bootstrapping set to 500. Red indicates sequences for *omp2* from mapping using CP000057.1 reference sequence only and negative for M93269 mapping. Isolates in blue depict sequences resulting from mapping using reference sequence M93269. The clustering of the sequences within the same clade despite resulting from different reference sequences displays the homology between the resulting sequences from the isolates. Neighbour joining tree of representative isolates from initial mapping of fastqs of 1460 isolates to GenBank reference M93269 alongside sequences from secondary mapping to CP00057.1. CP00057.1 isolates were negative for initial mapping. A sequence for *omp6* was used to root tree. Alignments were carried out using MUSCLE. Scale bar at 0.04 substitutions per 100 bp.

3.1.1.2 *sodC* – Molecular Marker for *H. haemolyticus*

The presence of *sodC* has been observed in 100% of *H. haemolyticus* however it has also been reported in NTHi but at 85% identity to *sodC* of *H. haemolyticus* (Fung et al., 2006, Latham et al., 2015, McCrea et al., 2010). GenBank reference AFQQ01000001.1: 155446-156018 was used for mapping and PCR primers from Latham *et al.* (2015) were used for *in silico* PCR (Latham et al., 2015)(Materials and Methods, section 2.3.7). All potential *H. haemolyticus* were positive for mapping and one was negative for *in silico* PCR but on alignment of returned mapping sequences was confirmed as positive (Figure 3.8). All 1368 NTHi were negative for *in silico* PCR but two NTHi isolates returned a sequence for mapping. These were aligned with all the positive *sodC* sequences from *H. haemolyticus* and were seen to cluster within the positive *H. haemolyticus* isolates. To ascertain % identity NEEDLE was used to align the sequences with a representative *H. haemolyticus* isolate (number 894) (Materials and Methods, section 2.3.7). Isolate numbered 1423 (NTHi PCR negative, mapping positive) shown on Figure 3.8 returned a 61.8% identity to the sequence for isolate 1422 (*H. haemolyticus* - mapping and PCR positive for *sodC* shown to cluster on branch with isolate 1423) and 60% for isolate 894 (*H. haemolyticus* mapping and PCR positive representing main cluster of *H. haemolyticus*). NTHi isolate numbered 1273 (NTHi positive and PCR negative for *sodC*) on Figure 3.8 returned an 86% identity to the *sodC* sequence for isolate number 894.

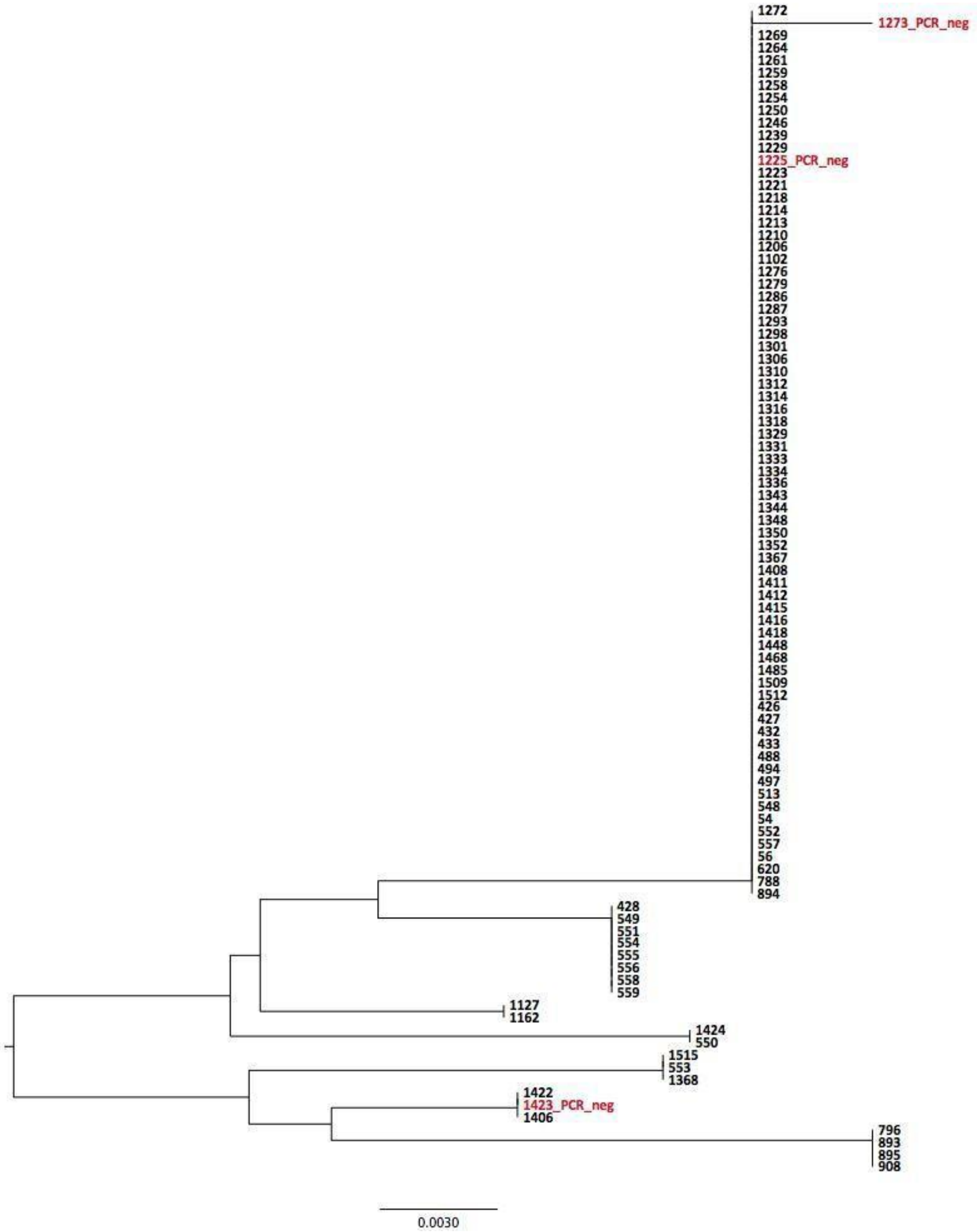


Figure 3.8 Sequences from mapping of 1460 isolates to *sodC.*

Three isolates were *in silico* PCR negative as highlighted in red. Isolate 1225 is *H. haemolyticus* and clusters within the other main *H. haemolyticus* clade. Isolates 1423 and 1273 are NTHi and returned STs 11 and 1442 respectively in MLST analysis. Isolate 1273 can be seen to be a variant of the main clade of *H. haemolyticus* and on further investigation is 86% identity to positive isolate 894. Isolate 1423 sits with two isolates outside the main branch but is 61% of the sequence to isolate 1422. Therefore, the two NTHi sequences are different truncated versions of *sodC*. Neighbour joining tree constructed in MEGA 6.06 from alignment carried out in MUSCLE. Boot strapping set at 500. Scale bar shows substitutions at 0.0030 per 100 base pairs.

3.4.4 *omp6*, *smpB* and *hpd* – the Allelic Variation Markers

The genes *omp6*, *hpd* and *smpB* are thought to be present in both species but have conserved sequences that enable differentiation between NTHi and *H. haemolyticus*. Reference sequences (GenBank CP002277.1: 1826771-1827865 – *hpd*, GenBank NC_000907.1:1038488-1038973 – *smpB*, GenBank M19391.1 – *omp6,* Materials and Methods, Table 2.4) were used to carry out mapping of the 1460 isolates and *in silico* PCR was performed using previously published primers from Pickering *et al.* (2014) for *hpd*, Reddington *et al.* (2015) for *smpB* and Murphy *et al.* (2007) for *omp6* (Pickering et al., 2014a, Reddington et al., 2015, Murphy et al., 2007) (Materials and Methods, section 2.3.7).

All 1368 NTHi and 92 potential *H. haemolyticus* isolates returned a mapped consensus for *omp6* and all 1368 NTHi isolates and all except one potential *H. haemolyticus* isolate were *in silico* PCR positive. The mapped consensus for the one isolate was compared to the online Blast database and returned 96% identity to *omp6* (Altschul et al., 1990)(Materials and Methods, section 2.3.7).

All 1368 NTHi isolates returned a sequence for *smpB* and 90/92 potential *H. haemolyticus* returned a sequence for mapping. The two isolates negative for mapping were investigated further by annotating the assembled genomes and searching for the 'Ssra binding protein', the protein expressed by *smpB*, within the annotation. This was found in both isolates and the associated sequence compared using BlastN with the megablast algorithm, both returned 86% identity to *smpB* (Altschul et al., 1990). Seven isolates were negative for *in silico* PCR for *smpB* but when all mapping sequences were aligned these were seen to cluster with other positive isolates within the clades containing the same species (Figure 3.9).

The *hpd* analysis demonstrated the presence of the gene in 1348/1368 (98.5%) of NTHi and all 92 potential *H. haemolyticus* returning a sequence. *In silico* PCR resulted in 1335/1368 positive results for *hpd*. From these 33 negatives, 20 were also negative for mapping and constituted all ST 925 (n=7) and ST 819 (n=14) isolates. The remaining thirteen isolates positive for mapping were aligned with all positive mapping sequences also positive for *in silico* PCR. All were seen to cluster with the other *hpd* positive NTHi sequences confirming the presence of the gene within the *in silico* negative isolates (Figure 3.10).

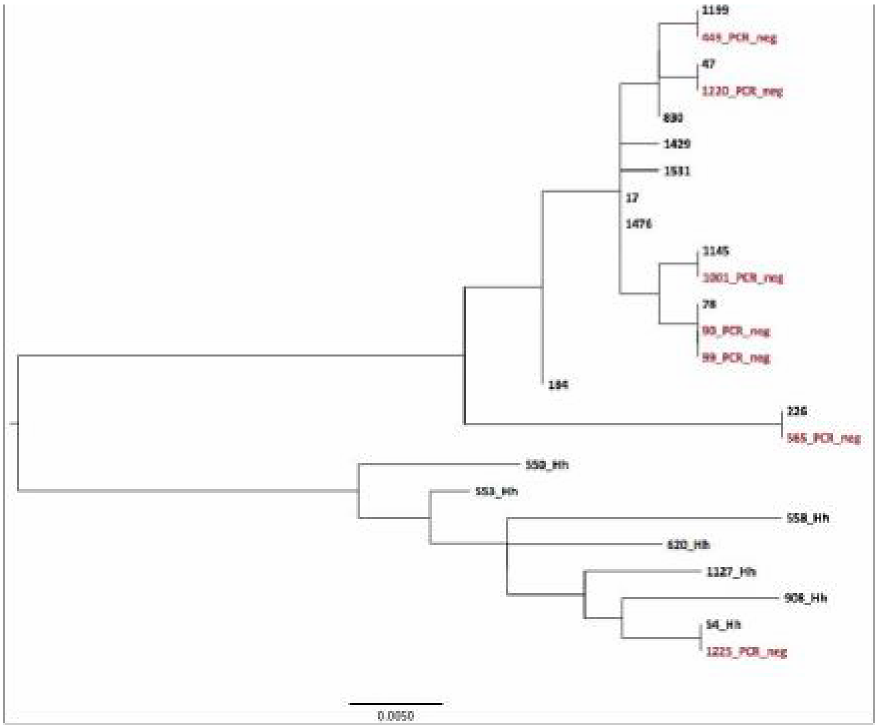


Figure 3.9 *smpB* mapping sequences for *in silico* PCR negative isolates.

Maximum likelihood tree constructed in MEGA 6.06 from MUSCLE alignment.
In silico PCR negative isolates for *smpB* and sequences from representative both positive for mapping and *in silico* PCR were aligned. Homology of *smpB* sequence with other isolates displayed by *in silico* PCR negative isolates clustering on the same branch as other positive isolates. All PCR negative isolates cluster with sequences for NTHi and *H. haemolyticus*. Hh= *H. haemolyticus.* Scale bar shows substitutions at 0.005 bp per 100.

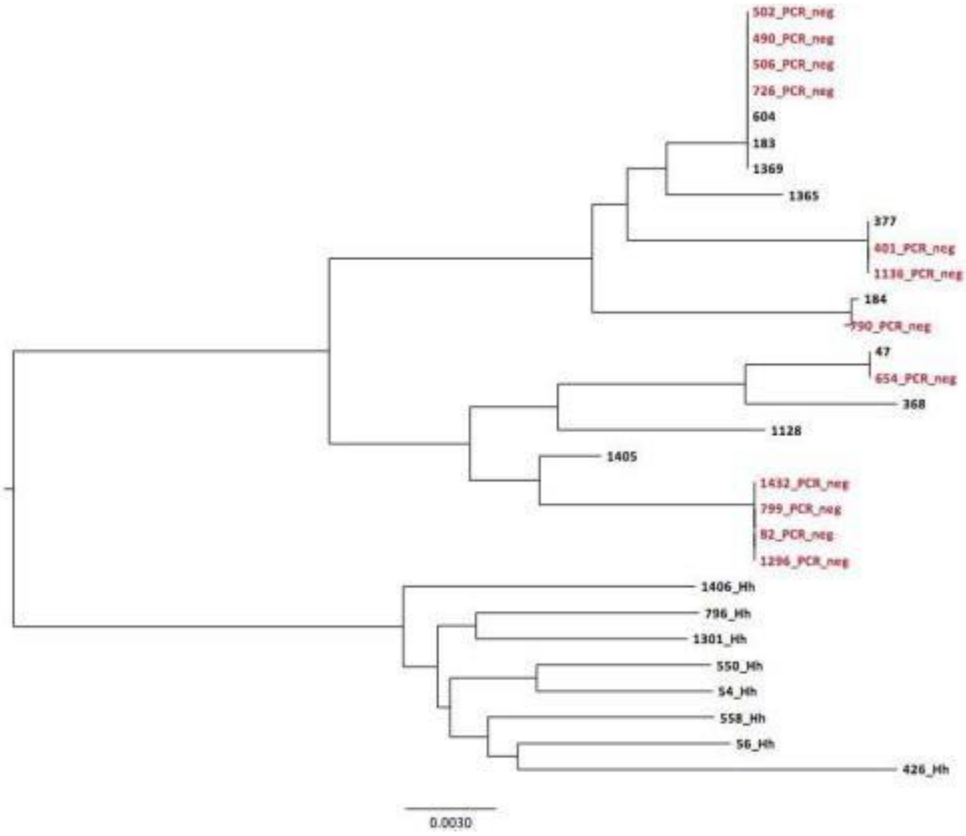


Figure 3.10 *hpd* mapping sequences for *in silico* PCR negative isolates.

Neighbour joining tree constructed in MEGA 6.06 from MUSCLE alignment using *hpd* sequences derived from *in silico* PCR negative isolates and isolates that were both positive for mapping and *in silico* PCR. Homology of *hpd* sequence with other isolates displayed by *in silico* PCR negative isolates clustering on the same branch as other positive isolates. Hh= *H. haemolyticus.*Scale bar shows substitutions at 0.003 bp per 100.

3.4.5 Summary of Molecular Markers Reveals Different Genotypes within NTHi and *H. haemolyticus*

To summarise the atypical traits observed, *omp2* negative isolates consisted of all isolates from the study that were of STs 353, 1314, 356 and 819. Fourteen other *omp2* negative isolates from other STs were observed but these were isolates from time points where *omp2* positive isolates were observed also. There were NTHi 20 isolates negative for *hpd*, these were all either from ST 925 or ST 819. All *fucK* and *fucP* negative isolates were from the same patient at all five sampling time points. Two potential *H. haemolyticus* isolates displayed a truncated version of *smpB* and two NTHi isolates were positive for truncated versions of *sodC*. There were 67 isolates negative for *lgtC* and twelve potential *H. haemolyticus* which were found to harbour a truncated *iga* sequence. There were no atypical results for *omp6.* Interestingly, the atypical isolates also predominantly group into ST types.

The collation of the presence and absence of the molecular markers identifies ten genotypes incorporating atypical isolates of NTHi and two of *H. haemolyticus*, see Table 3.2. The ten resulting atypical genotypes for NTHi involve six with more than one atypical trait. The *omp2* negative isolates can be seen to span five genotypes. Only three genotypes were returned for *H. haemolyticus* however both species were similar in proportions of atypical isolates with NTHi displaying atypical traits for the molecular markers in 226/1368 (16.5%) of isolates and *H. haemolyticus* displaying atypical traits in 14/92 (15.2%). However due to the discrepancy in numbers between the two groups this is an observation that would require further investigation in a larger set of *H. haemolyticus*. Hypothetically though this could indicate a similar level of heterogeneity in *H. haemolyticus*.

Table 3.2 Presence/absence genotypes returned from mapping of reference sequences for molecular markers *hpd*, *omp2*, *omp6*, *lgtC*, *iga*, *smpB*, *sodC*, *fucP* and *fucK* to 1460 culture identified *Haemophilus* spp.

G=genotype. Blue highlights those genes displaying atypical results from the expected or typical genotype. G1 is typical genotype for NTHi. Genotype Hh_G1 is typical genotype for *H. haemolyticus*.

	smpB	*lgtC*	*iga*	*fucP*	*P2*	*P6*	*sodC*	*hpd*	*fucK*	Total (%)
NTHi G1	+	+	+	+	+	+	-	+	+	**1142 (83)**
NTHI G2	+	+	+	+	-	+	-	+	+	**84 (6.1)**
NTHi G3	+	-	+	+	+	+	-	+	+	**60 (4.4)**
NTHi G4	+	-	+	+	-	+	-	+	+	**5 (0.4)**
NTHi G5	+	+	+	+	+	+	-	-	+	**8 (0.6)**
NTHi G6	+	+	+	+	-	+	-	-	+	**12 (0.9)**
NTHi G7	+	-	+	+	-	+	-	-	+	**1 (0.1)**
NTHi G8	+	+	+	-	+	+	-	+	-	**53 (3.9)**
NTHi G9	+	-	+	-	+	+	-	+	-	**1 (0.1)**
NTHi G10	+	+	+	+	-	+	**Truncated**	+	+	**1 (0.1)**
NTHi G11	+	+	+	+	+	+	**Truncated**	+	+	**1 (0.1)**
Totals/1368	**1368**	**1301**	**1368**	**1314**	**1264**	**1368**	**2**	**1347**	**1314**	
Hh_G1	+	-	-	-	-	+	+	+	-	**78 (84.8)**
Hh_G2	**86% ID**	-	-	-	-	+	+	+	-	**2 (2.2)**
Hh_G3	+	-	**Truncated**	-	+	+	+	+	-	**12 (13.0)**
Totals /92	**90**	**0**	**1**	**0**	**1**	**91**	**92**	**92**	**0**	

3.4.6 Allelic variations in ***hpd***, ***smpB*** and ***omp6*** differentiate between the two species

The *hpd*, *smpB* and *omp6* genes are expected in both NTHi and *H. haemolyticus* but varied enough as to enable species differentiation (Binks et al., 2012, Reddington et al., 2015, Pickering et al., 2014a). PCR assays and high resolution melt curve analysis have been designed to exploit this phenomenon however *hpd* has been reported as absent in some strains and the utility of *omp6* has been questioned due to the level of variation displayed (Janson et al., 1993, Pickering et al., 2014a, Smith-Vaughan et al., 2014, Wang et al., 2011, Chang et al., 2010, Chang et al., 2011, Nelson et al., 1991, Binks et al., 2012).

Mapped consensus sequences for each of the three genes were aligned to visualise the allelic variation between the two species (Materials and Methods, section 2.3.8). The alignments were used to construct phylogenetic trees with *H. haemolyticus* being marked. As can be seen in Figure 3.11 - Figure 3.13 the potential *H. haemolyticus* group was seen to cluster separately from NTHi in all three cases. The NTHi showed heterogeneity in all 3 cases with the sequences from all NTHi isolates separating out into different clades whereas the sequences for each gene from the potential *H. haemolyticus* were observed to cluster together in one clade, this observation suggests a much higher level of sequence heterogeneity for *hpd*, *smpB* and *omp6* in NTHi than in *H. haemolyticus*. This could also however be a result of the smaller number of *H. haemolyticus* available for analysis.

In both *smpB* and *hpd* a subset of isolates clustered separately from the remaining NTHi (Figure 3.11,Figure 3.12). This group were all *omp2* negative and were allocated either ST 1314, 353 or 356, and these STs were all novel to this study. There were no other occurrences of these STs from this group.

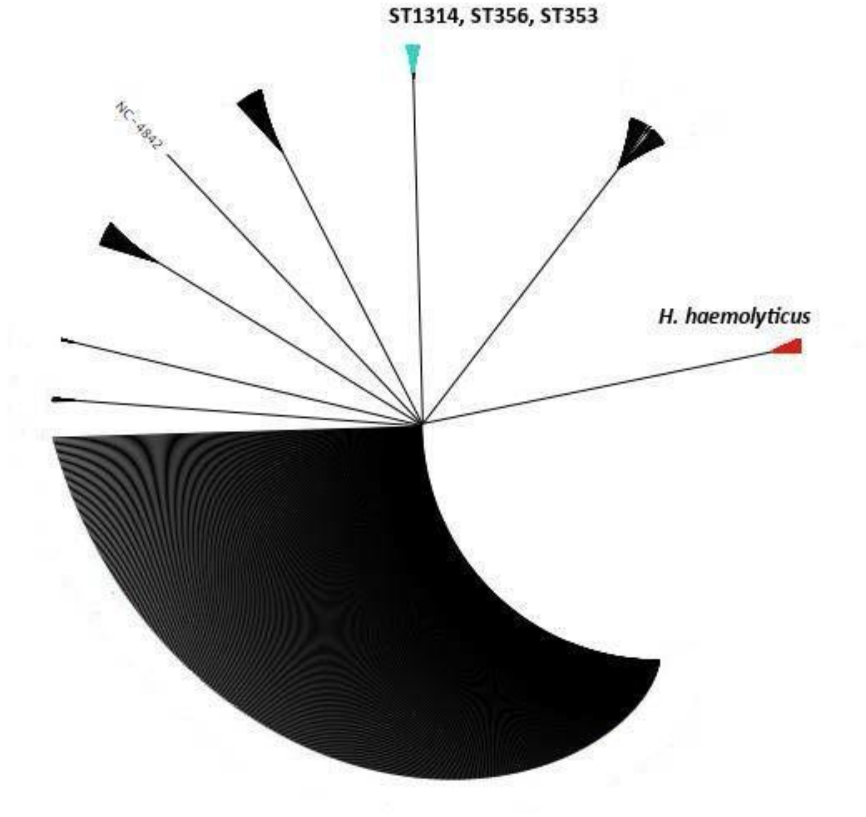


Figure 3.11 ***hpd*** **- showing the diversity in NTHi from** ***H. haemolyticus*** **in sequence.** Neighbour joining tree created in MEGA 6.06 bootstrapping set to 500The *H. haemolyticus* are represented in red, the NTHi in black. Both species cluster into separate clades showing variation between species. ST 1314, ST 353 and ST 356 cluster together, these are novel to this study and have shown to be *omp2* negative, a trait which is more associated with *H. haemolyticus.* NC-4842 is a reference strain for NTHi from Public Health England. Both NC-10659 and NC-10839 which are reference strains for *H. haemolyticus* clustered within the *H. haemolyticus* clade*.* Alignment carried out using MUSCLE.

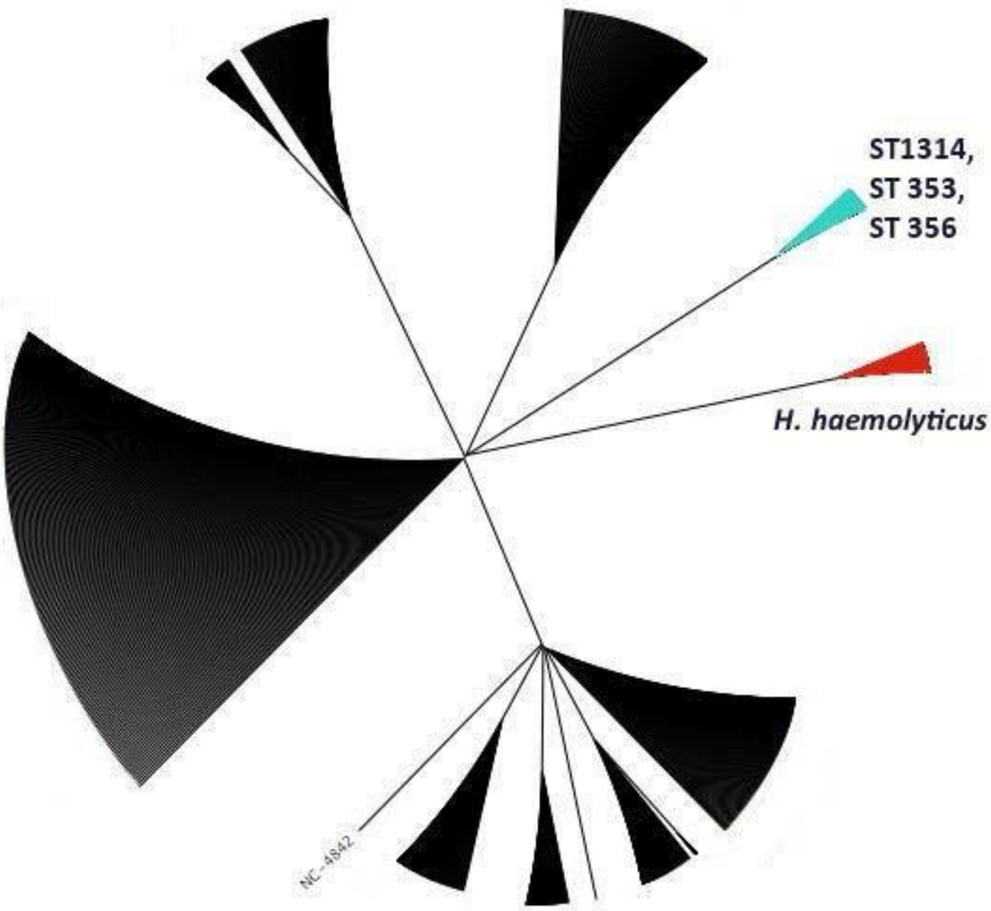


Figure 3.12 *smpB* sequences displaying the independent clustering of NTHi (black) from *H.haemolyticus* (red).

Neighbour joining tree created in MEGA 6.06 bootstrapping set to 500. A visibly divergent branch from the NTHI clade consists of STs 356, 353 and 1314 all of which were negative for *omp2*, a trait more associated with *H. haemolyticus*, and novel to this study (green). Other isolates displaying atypical NTHi genotype traits cluster with the expected NTHi genotype. Isolates displaying both *hpd* negative traits and *hpd* and *omp2* negative traits are not found to cluster separately but sit within the larger section of the tree amongst other typical isolates. The isolates that were *fucP* and *fucK* negative also sit amongst other typical NTHi isolates. Position of reference sequence is shown for NC-4842 reference strain (NTHi). Both *H. haemolyticus* reference strains NC-10659 and NC-10839 were present within the *H. haemolyticus* cluster. Alignments carried out in MUSCLE.

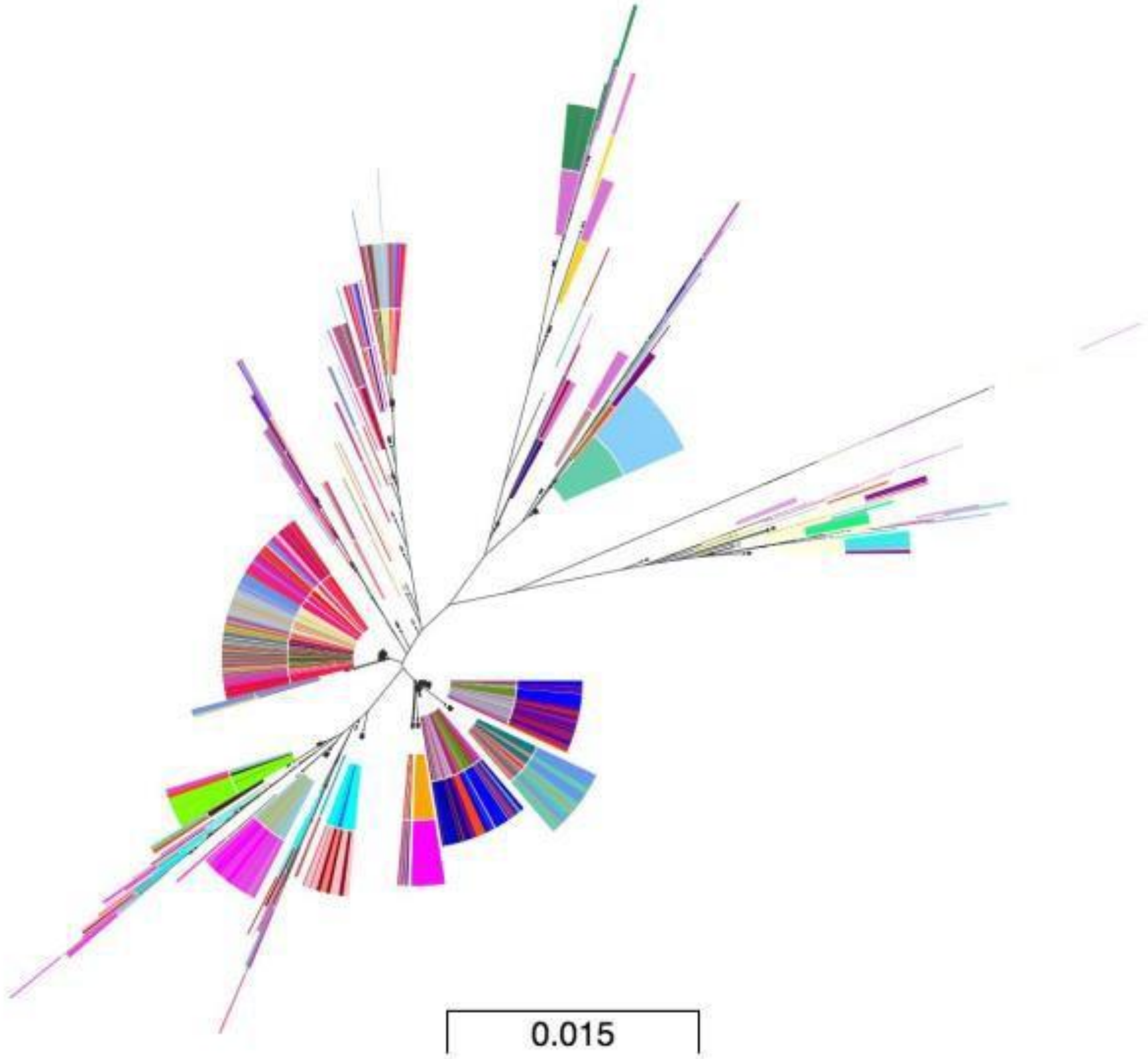


Figure 3.13 The sequences from mapping for *omp6* shown with the corresponding ST. A neighbour joining tree was constructed in MEGA 6.06 from a MUSCLE alignment and bootstrapping was set at 500. The Newick file and metadata were investigated using microreact. This phylogeny can be further investigated dynamically using microreact using the link (https://microreact.org/project/ryTBYIEHx) Potential *H. haemolyticus* are labelled in light yellow in the right hand clade the remaining clades are NTHi. The coloured inner circle represents different STs and the outer circle of colours represents patients. Clades containing NTHi display different STs within them showing the heterogeneity of the *omp6* sequences found within NTHi. *H. haemolyticus* clusters into a clade on its own showing a more conserved sequence within the species.

3.5 Discussion

Here the analysis of 1460 *Haemophilus* spp. isolates from the AERIS study is presented, a longitudinal cohort of 24 COPD patients (Bourne et al., 2014). Initially the isolates were examined and classified based on haemolysis. A 48 h haemolytic assay on blood agar identified 12 (0.8%) haemolytic isolates. Three of which were identified after the currently recommended 24 h incubation period for haemolysis assays on blood agar for identification of haemolytic activity (PHE, 2015). No isolate showed further haemolysis at 72 h therefore we have shown that delayed haemolysis may occur in some *H. haemolyticus* and an optimum 48 h period may ensure all haemolytic isolates are identified (Murphy et al., 2007).

The visible beta haemolysis of *H. haemolyticus* on blood agar, which has historically been the phenotype used to differentiate between the two species, can be lost (Kilian, 1976a). Routine diagnostics to determine the correct causative organism in cases of infection, and in turn the appropriate prescription of antimicrobial treatment, are therefore affected by the inability to differentiate successfully between NTHi and *H. haemolyticus*. Therefore, although the haemolysis assay data shows that some strains may show delay in haemolysis, haemolysis itself is not a reliable identifier of *H. haemolyticus*. However, this basic culture assay could help discard haemolytic samples from further investigation. Obviously though this does not rectify the misidentification of non–haemolytic *H. haemolyticus* which has led to the development of molecular assays due to the need for a better differentiation method. Due to the heterogeneity of NTHi, the development of an unambiguous PCR assay has been found to be challenging (Latham et al., 2015, Binks et al., 2012, McCrea et al., 2008, Sandstedt et al., 2008, Reddington et al., 2015, Price et al., 2015). Absences of the *hpd*, *lgtC*, *omp2*, *iga* and the genes of the fucose operon have been reported and the ability of the *omp6* gene disputed due variation (Chang et al., 2010, Chang et al., 2011). Similarly, recombination between NTHi and *H. haemolyticus* has been highlighted in 16S rRNA analysis (de Gier et al., 2015). Jointly these molecular assays enable an informed way to distinguish between the two species however additional methods result in a burden on time and resources.

MLST was used to attempt to accurately classify isolates into NTHi and *H. haemolyticus* (Table 3.1). MLST uses sequences from seven housekeeping genes, the resulting sequence fragment at each of the seven loci defining the sequence type. There were eight new STs novel to this study, seven of which (STs 353, 356, 357, 1314, 1441, 1442, 1664) have been curated to the MLST database. However, one set of isolates that were defined as a new ST were those negative for a *fucK* sequence but positive for known sequences at the remaining six loci of the MLST schema. Due to the absence of this loci within these isolates it cannot be defined as an official ST but was classified as *fucK* negative NTHi. These isolates were also negative for the *fucP* gene. This concurs with previous studies reporting the absence of the *fucK* gene and the complete fucose operon that has led to the use of MLST being questioned for its ability to fully categorise NTHi (Meats et al., 2003, de Gier et al., 2015, Ridderberg et al., 2010).

The *fucK* gene is not expected to be present within the *H. haemolyticus* genome. In addition to the 54 isolates from the *fucK* negative ST, ninety two isolates were *fucK* negative but also returned no identifiable sequence from the other six loci (*adk, atpG,frdB, mdh, pgi, recA*) and were therefore identified as potential *H. haemolyticus* (Table 3.2). These isolates also included the only twelve haemolytic isolates from the study, a phenotype associated with *H. haemolyticus* (Kilian, 1976a). When further investigated using molecular markers this group could be classified as *H. haemolyticus* with twelve haemolytic and 80 non-haemolytic isolates. This classification was confirmed due to the presence of *sodC* in all of the isolates and the absence of *omp2*, *lgtC*, *fucK* and *fucP* in addition to the clustering of the *H. haemolyticus* into separate clades from the NTHi in the phylogeny of the *hpd*, *smpB* and *omp6* genes (Figure 3.11, Figure 3.12Figure 3.13). Atypically, twelve isolates returned truncated *iga* sequences and two isolates returned sequences for 86% *smpB* however the remaining genotype for these isolates enabled the classification of *H. haemolyticus* (Figure 3.6). MLST was also able to identify this group from NTHi with *H. haemolyticus* sequence fragments being sufficiently variable from those known for NTHi (Figure 3.3). Therefore, despite the existence of *fucK* negative strains of NTHi, MLST could still be used as a tool to differentiate between morphologically similar *Haemophilus* spp. once growth factor requirements had been observed.

Genomic heterogeneity has been repeatedly demonstrated with missing NTHi-specific genes reported such as *hpd*, *fucK*, *iga*, *lgtC*, *omp2*, *omp6* and positive results for the *H. haemolyticus* marker *sodC* (Price et al., 2015, Reddington et al., 2015, Wang et al., 2011, Binks et al., 2012, Latham et al., 2015, McCrea et al., 2008). The molecular markers *fucP* and *smpB* are relatively new assays and literature is limited, however the *fucP* gene is part of the fucose operon which is absent in some strains of NTHi (de Gier et al., 2015). As expected the majority (n=1142, 83%) of NTHi returned the typical NTHi genotype of *iga*, *lgtC*, *fucK*, *fucP* and *omp2* positive and *sodC* negative (Table 3.2). In addition, all isolates clustered with other NTHi sequences in *hpd*, *smpB* and *omp6* with *H. haemolyticus* clustering into a separate clade as discussed previously (Figure 3.11, 3.12, Figure 3.13). The phylogenetic tree resulting from the *omp6* sequences from all isolates showed also the diversity of sequences of this gene throughout the isolates by comparison with distribution of ST (Figure 3.13). *H. haemolyticus* clustered together relatively closely compared to that of NTHi which formed different clades all containing isolates from many different STs (Figure 3.13). All clades contained >3 STs displaying the diversity of *omp6* sequences throughout STs indicating that variation in genes within NTHi is not reflected by the MLST gene sequence variation, recombination may have a role to play in the distribution of the variation *omp6* throughout the STs.

A number of atypical NTHi genotypes were also observed. There were 234 isolates of NTHi in which one or more of the expected molecular markers were absent, and in two cases where *sodC* was present, a gene expected within *H. haemolyticus* but not NTHi (Figure 3.8,Table 3.2). These atypical isolates could be considered as NTHi based on the consistent phylogeny on examination of *omp6* and *smpB* in addition to the presence of the remainder of the genotypic markers (Figure 3.12, Figure 3.13).

From the 104 isolates that were negative for *omp2*, a group of 76 were seen to repeatedly cluster separately from the other NTHi isolates when the allelic variation in *hpd* and *smpB* was taken into account (Figure 3.11, Figure 3.12). This group consisted of all isolates of ST353 (n=2), ST356 (n=51) and ST1314 (n=23) which were novel to this study (Figure 3.1). These were not the only isolates observed to have an atypical genotype but no other group of atypical isolates were seen to cluster separately within phylogenetic analysis (Figure 3.11, 3.12 and Figure 3.13). This could indicate a potential group of atypical isolates that may share other genomic characteristics.

All isolates of ST 819 (n=14) were also *omp2* negative and were also negative for *hpd*. ST 925 (n=7) were *hpd* negative only. STs 925 and 819 both clustered with other NTHi *omp6* and *smpB* sequences. Other atypical isolates were singletons from time points that contained isolates that displayed the typical NTHi genotype (Table 3.2). These may represent an artefact within the sequencing analysis rather than a true negative due to the difference being so varied from the other NTHi isolated at the same time point with the same ST. However, ST types were isolated within the same time point, furthermore isolates with a recognised ST and isolates of *H. haemolyticus* were also observed simultaneously in a time point therefore it is feasible that these single isolates displaying atypical traits may indeed be true negatives and indicate intra ST variation within a time point. (Appendix 2). Simultaneous isolation of different STs have been reported previously (Murphy et al., 1999).

NTHi isolates devoid of the genes *fucK, fucP, lgtC, hpd* and *omp2* and positive for *sod*C were observed, these are atypical traits for NTHi and in the case of absent markers may demonstrate potential molecular redundancy of these genes in the NTHi genome. The *lgtC* gene is one of many genes that encodes for and contributes to the heterogeneity of the lipooligosaccharide (LOS) of NTHi, an important virulence factor (Rahman et al., 1999, Morey et al., 2013). The galactosyltransferase expressed by *lgtC* results in an epitope of the LOS important for serum resistance by inhibiting C4b deposition within the classical complement pathway (Erwin et al., 2006a, Ho et al., 2007). However, effects of other genes such as *losA*, *lex2A*, *lex2B* and *lic3A,* on the LOS structure, in addition to mechanisms such as sialic acid production and binding to members of the extra cellular matrix, may also result in increased serum resistance (Griffin et al., 2005, Erwin et al., 2006a, Erwin et al., 2006b, Deadman et al., 2009, Barthel et al., 2012b, Hallstrom et al., 2009, Su et al., 2016, Hood et al., 2001). Likewise, *omp2* shares the ability for mucin binding with *ompP5* and *pilA* (Reddy et al., 1996, Kubiet et al., 2000). Additionally, Protein D, encoded for by *hpd* has been found non-essential for survival of NTHi but is thought to play a major role in pathogenesis with mutant isolates requiring 100x more colony forming units to cause infection in a mouse model (Forsgren et al., 2008).

An important marker for NTHi used frequently as a PCR target for identification and also as part of the MLST schema is *fucK*. This is part of the fucose operon, both *fucK* and the full fucose operon have been reported absent from some strains of NTHi (de Gier et al., 2015, Ridderberg et al., 2010, LaCross et al., 2013). The fucose operon consists of *fucU*, *fucR* which encode for the fucose operon protein, L-fucose operon activator and *fucA*, *fucK*, *fucP* and *fucI*, which encode for enzymes L-fucose phosphate aldolase, L-fucose kinase, L-fucose permease and L-fucose isomerase (Ridderberg et al., 2010). Fucose is used in the *O*-glycosylation process in mucin production within the respiratory tract and has been shown to be metabolised by gut bacteria and used for adherence (Rose and Voynow, 2006, Dwivedi et al., 2016, Muraoka and Zhang, 2011, Stahl et al., 2011). However, the deletion of both *fucP* and *fucK* in 54 of the isolates from this study and the previously reported deletion of the whole operon points to alternative sources for these mechanisms for NTHi within the respiratory tract.

For identification purposes the use of any of these molecular markers as single PCR assays would have resulted in inaccurate speciation. All *H. haemolyticus* were positive for *sod*C however two NTHi were also positive for *sodC* sequences. All NTHi samples returned a sequence for *iga* and *omp6* but both genes have been reported absent from strains in previous research and twelve *H. haemolyticus* displayed truncated *iga* sequences (Binks et al., 2012). The largest gene absence was that of *omp2* with 104 strains. Recently, duplex PCR assays have been presented as an option for differentiating between NTHi and *H. haemolyticus*. Hu *et al.* (2016) have developed an assay using *purT* and *hpd*, De Gier *et al.* (2016) suggest *hpd* and *fucP* (de Gier et al., 2016, Hu et al., 2016). With the heterogeneity of NTHi leading to previous reported occurrences of absence of *hpd* and *fucose* genes the use of even two molecular markers may be questionable (Smith-Vaughan et al., 2014, de Gier et al., 2015, Ridderberg et al., 2010). The genotypes returned by this study identified isolates absent for more than one molecular marker.

From the original 1460 *Haemophilus* spp. only 92 (6.3%) were identified as *H. haemolyticus* (Table 3.1, Table 3.2). Twelve (13%) of these would have been discarded with the observation of haemolysis and the addition of a culture assay which would reduce the number to 80/1460 (5.5%) Murphy *et al.* (2007) reported a misidentification rate of 39.5% (102/258) in COPD sputa samples and 27.3% (12/44) in paediatric nasopharyngeal samples. Although this represents a significant misidentification rate, other studies have reported a much lower amount. In a study of respiratory samples taken from cystic fibrosis patients misidentification numbers were much smaller at 0.5% (Fenger et al., 2012). An Australian study into 447 clinical NTHi isolated across the country over seven years discovered only 1.5% (7/447) of the isolates to be *H. haemolyticus* and interestingly were all from cystic fibrosis patients (Zhang et al., 2014). A study into otitis media reported an isolation rate of 11.7% of *H. haemolyticus* however it has also been reported that children suffering from recurrent otitis media are more likely to carry *H. haemolyticus* in their nasopharynx (Kirkham et al., 2010). A Danish study returned 2/480 (0.4%) clinically relevant NTHi isolates as misidentified non-haemolytic *H. haemolyticus* (Norskov-Lauritsen, 2009, Kirkham et al., 2010). Therefore, the amount of *H. haemolyticus* identified within this study is far less than the initially reported 39.5% from Murphy *et al.* (2007) and is more comparable with more recent research (Murphy et al., 2007).

3.6 Conclusions and Future Work

From this study it can be concluded that only 92/1460 isolates of the *Haemophilus* spp. initially identified by culture methods and growth requirements were re-classified as *H. haemolyticus.* Therefore, it has been shown that the potential for misidentification of *H. haemolyticus* as NTHi within sputa collected from individuals with COPD is much less than the initially reported 39.5% (Murphy et al., 2007).

The optimum incubation time for haemolysis to become evident in *H. haemolyticus* was 48 h which was visible in 13% of the 92/1460 *H. haemolyticus*. All *H. haemolyticus* were negative for *fucK* and returned unrecognisable sequences for the remaining six MLST loci therefore MLST was able to distinguish *H. haemolyticus* from NTHi, even those NTHi that were *fucK* negative. MLST was not able to identify isolates with atypical genotypes or genetic variation within the molecular markers.

No single PCR assay for presence or absence of gene was able to speciate fully. Absences were observed in *omp2*, *fucK*, *fucP* and *lgtC.* Unexpected truncated sequences for *iga* beta core were observed in NTHi and two isolates were positive for *sodC. omp6* was the only single PCR assay based on allelic variation that returned a sequence for all isolates from mapping and was able to speciate. Absences in *hpd* and variations in two isolates for *smpB* were observed.

Methods of differentiating *H. haemolyticus* from NTHi have differed between studies. Single gene targets cannot be relied upon due to the heterogeneity of NTHi and 16S rRNA analysis is also not able to accurately distinguish between the two species due to recombination therefore the ability to examine the relationships between strains using multiple gene targets and phylogeny resulting from sequencing data may provide a more accurate method of species differentiation and enable identification of atypical strains (de Gier et al., 2015). However, bioinformatic analysis requires the functional assays and proteomic investigations to substantiate the absence/presence of genes and products from findings such as truncated *iga* sequences observed in *H. haemolyticus* for example. To ascertain the true negative nature of singletons of NTHi negative for an expected gene in amongst multiple other isolates from the same time point would require phenotypic assays to confirm the negative nature for the phenotype associated with the gene.

Genotyping using molecular markers was able to delve deeper and identify atypical isolates but with such a limited number of genes used this really only scratches the surface of the potential variance within the NTHi strains. Whole genome analysis of the strains would enable a further depth to the investigation into the relationship between atypical NTHi and *H. haemolyticus* and also ascertain any interspecies recombination events that may also explain the variation shown within STs for *omp6* (Figure 3.13). Whole genome analysis will be addressed in Chapter 6.

4 Adhesin Genotypes Present in NTHi and *H. haemolyticus*

4.1 Introduction

The difference in capabilities for attachment and invasion between NTHi and *H. haemolyticus* has recently been observed (Pickering et al., 2016, Singh et al., 2016a). NTHi has been shown to display a greater attachment and capacity to invade *in vitro* nasopharyngeal and bronchoalveolar epithelial cells *in vitro* than *H. haemolyticus* (Pickering et al., 2016, Singh et al., 2016a).

Adhesins enable bacteria to colonise the human host, the first step to infection. In the case of NTHi and *H. haemolyticus*, colonisation occurs initially in the upper respiratory tract via adhesins that invoke attachment of bacteria to epithelial cells within the nasopharynx (Tenenbaum et al., 2012). NTHi and *H. haemolyticus* exhibit phenotypic and genotypic differences including a divergence in their capacity for causing disease, which NTHi achieves by using a number of virulence factors that includes adhesins (Rodriguez et al., 2003, Cardines et al., 2007, Avadhanula et al., 2006, Barthel et al., 2012b, Ecevit et al., 2004, Euba et al., 2015). In addition to colonisation, most adhesins have secondary mechanisms that enable NTHi to internalise and migrate, form biofilms and evade immune clearance by integration with the extra cellular matrix (ECM) or by phase variable expression (Atack et al., 2015, Davis et al., 2014b, Hallstrom et al., 2009, Su et al., 2016, Singh et al., 2014). Adhesins therefore are far more than just adherence proteins they form an assorted arsenal of tools for pathogenesis. It is therefore suggested that these phenotypic differences between NTHi and *H. haemolyticus* may in part be reflected in their respective adhesin genotypes.

4.2 Hypothesis

The adhesin genotype for NTHi and *H. haemolyticus* will vary with those adhesin genes associated with virulence being more associated with NTHi.

4.3 Aims

To characterise and compare the adhesin genotypes observed within NTHi and *H. haemolyticus.* The adhesin genes investigated are detailed in Table 4.1. Whole genome sequences will be analysed for presence and absence of adhesin genes and phylogenetic relationships of shared genes between the species will also be investigated. Association between the onset of acute exacerbation and the prevalence of adhesin genes within the study isolates of NTHi will also be reported.

Table 4.1 Adhesin genes investigated in study.

Details of thirteen adhesin genes with secondary functions within NTHi. GenBank reference refers to sequences downloaded from the GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) and used for mapping against to identify gene of interest.

Gene name	Protein name	Function	GenBank reference
hmwA	High weight molecule	High adherence capacity. Mediates protein expression by phase variation. Associated with Otitis media.(Cholon et al., 2008, Davis et al., 2014b, Giufre et al., 2008, Vuong et al., 2013)	AY601283.1
hmw2A	High weight molecule	High adherence capacity mediates protein expression by phase variation. Associated with Otitis media. (Cholon et al., 2008, Davis et al., 2014b, Giufre et al., 2008, Vuong et al., 2013)	U08875.1:352-4785
hia	*H. influenzae* adhesin	Associated with invasive disease. Small numbers of positive strains reported. Binding capabilities to ECM unknown. Homologue of *hsf* (Cardines et al., 2007, Rodriguez et al., 2003, Ecevit et al., 2004, Cotter et al., 2005)	AY078090.1 U38617.2 (poly T tract)
hsf	*Haemophilus* surface fibril	Encapsulated adhesin only. Binds to vitronectin for immune evasion. Homologue of *hia* (Hallstrom et al., 2006, Satola et al., 2008, Cotter et al., 2005)	U41852.2
hap	*Haemophilus* adhesin protein	Binds to Fibronectin, laminin and collagen IV for immune evasion and adherence with ECM. Forms microcolonies before auto proteolysis. Aids in invasion and migration. (Fink et al., 2003, Fink et al., 2002)	(U11024.1, CP002277.1:326101330315 and LN831035.1: 815431-819606
omp1	Outer membrane protein 1	Binds to CEACAM-1(Tchoupa et al., 2015)	CP000057.2: 491079-492455
omp2	Outer membrane protein 2	Binds to mucins, immune evasion through spontaneous point mutations (Reddy et al., 1996, Duim et al., 1994)	CP00057: 216249-217346 M93269.1
hel	Outer membrane protein 4	Binds to vitronectin, fibronectin for immune evasion and vitronectin and laminin for increased adherence(Su et al., 2016)	M68502.1
ompA	Outer membrane protein 5	CEACAM-1 and ICAM-1 ligand. Also binds to mucins. (Hill et al., 2001, Bookwalter et al., 2008, Avadhanula et al., 2006, Reddy et al., 1996)	L20309.1
omp6	Outer membrane protein 6		M19391.1
hif	*H. influenzae* fimbriae	Phase variation mediates expression. Binds to mucins. (Kubiet et al., 2000, van Ham et al., 1993)	U19730.1
pilA	Type IV pilus	Binds to adhesion molecule ICAM-1. Used in competence for recombination.(Novotny and Bakaletz, 2016, Carruthers et al., 2012)	CP000057.2:396415-396864
pE	protein E	Binds to adhesion molecule ICAM-1. Also binds to plasminogen, vitronectin and laminin in the ECM enabling immune evasion. (Barthel et al., 2012b, Hallstrom et al., 2009, Hallstrom et al., 2011)	CP002276.1:463373-463855.

4.4 Results

In Chapter 3, 1460 *Haemophilus* spp. isolated from sputa of COPD patients were speciated into NTHi and *H. haemolyticus* by genotyping sequence data (Chapter 3, Table 3.2). This resulted in 1368 NTHi and 92 *H. haemolyticus*. The NTHi were further grouped into typical NTHi, a group of fucK negative NTHi (n=52) and a group of atypical isolates that were *omp2* negative and grouped separately from the other NTHi in phylogenetic analysis of molecular markers *hpd* and *smpB* (n=76) (Chapter 3, section 3.4.6). This atypical group was made up of the full complement of STs 353, 356 and 1314 from MLST analysis and will be referred to as the atypical NTHi group in this chapter.

In order to ascertain the adhesin genotype sequence data for each isolate was mapped against GenBank sequences for each adhesin listed in Table 4.1 to identify presence/absence status and return a consensus nucleotide sequence (Materials and Methods, section 2.3.7). Sequencing data was also assembled into draft genomes, annotated and annotations searched for adhesin gene alleles (Materials and Methods, section 2.3.7). In the case of *omp5*, *hel*, *pilA*, *pE* and *omp6*, which were observed in both species, sequences were aligned and phylogenetic analysis carried out to investigate the relationship between the sequences from both species where available (Materials and Methods, section 2.3.8)

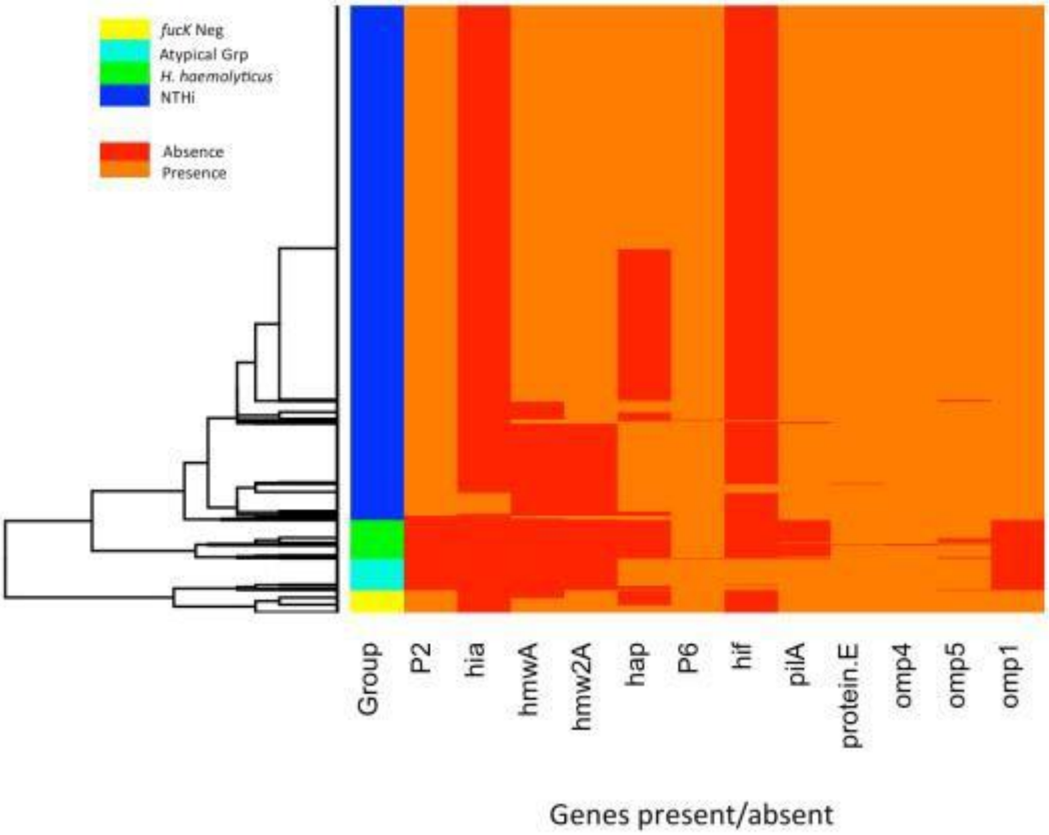


Figure 4.1 Heatmap showing presence and absence distribution of adhesin genes throughout the 1460 isolates split into four groups.

The four groups consist of atypical NTHi group (n=76), *fucK* negative NTHI (n=52), *H. haemolyticus* (n=92) and the remaining NTHi. With the exception of the absence of *hap* and *hif* from *H. haemolyticus*, the atypical NTHi group displays more homogeneity in adhesion genotype to *H. haemolyticus* than to NTHi or the group of *fucK* negative NTHi isolate

4.4.1 The Outer Membrane Proteins – *omp1*, *omp2*, *hel*, *ompA*, *omp6* and *pE*

The presence of *omp2* and *omp6* status for all isolates was established in Chapter 3. From the 1368 NTHi, 104 isolates were negative for *omp2*, including 76 categorised into the atypical NTHi group. All *H. haemolyticus* were negative for *omp2*. *Omp6* was identified in all isolates of both NTHi and *H. haemolyticus*.

Outer membrane protein 1 has been identified as the main ligand for CEACAM-1, a human cell adhesion molecule (Tchoupa et al., 2015). Mapping to a reference sequence found *omp1* in 1291/1368 (94.4%) of the NTHi isolates (Table 4.1). No *H. haemolyticus* isolate returned a sequence for *omp1*. From the 77 isolates negative for *omp1* all 77 were also negative for *omp2* and were consisted predominantly of the atypical NTHi isolates (n=76). The one exception was an isolate of ST 513 (Figure 4.1)

Outer membrane protein 4, expressed by the *hel* gene, has been reported to bind to vitronectin and fibronectin which shields the bacteria from avoidance of clearance by complement (Su et al., 2016). All 1368 NTHi returned a sequence for *hel* from mapping to the reference sequence (Table 4.1,Figure 4.1) Two *H. haemolyticus* isolates were negative for mapping. In order to confirm the status of these two isolates the genomes were assembled from the sequencing data and annotated and the reference sequence was searched for within the annotations (Materials and Methods, section 2.3.7). Both isolates were found to harbour a sequence for *hel*. These sequences were submitted to BlastN using the megablast algorithm and returned 82% identity to GenBank sequence LN831035.1:346077-346901 for *hel* (Materials and Methods, section 2.3.7).

ompA, which codes the outer membrane protein 5, has been associated with CEACAM-1 binding and is also thought to play an important role in attachment to mucins (Hill et al., 2001, Reddy et al., 1996). Mapping to the GenBank reference (Table 4.1) returned sequences for *ompA* in all 1368 NTHi (Figure 4.1). Sequences were discovered in 74 of 92 *H. haemolyticus* isolates through mapping. Sequences for the eighteen mapping negative isolates were obtained by investigating the annotated assembled genomes (Materials and Methods, section 2.3.7). These annotated sequences were compared to those of mapping positive isolates to confirm relationship, all eighteen were deemed positive for *ompA* (Figure 4.2).

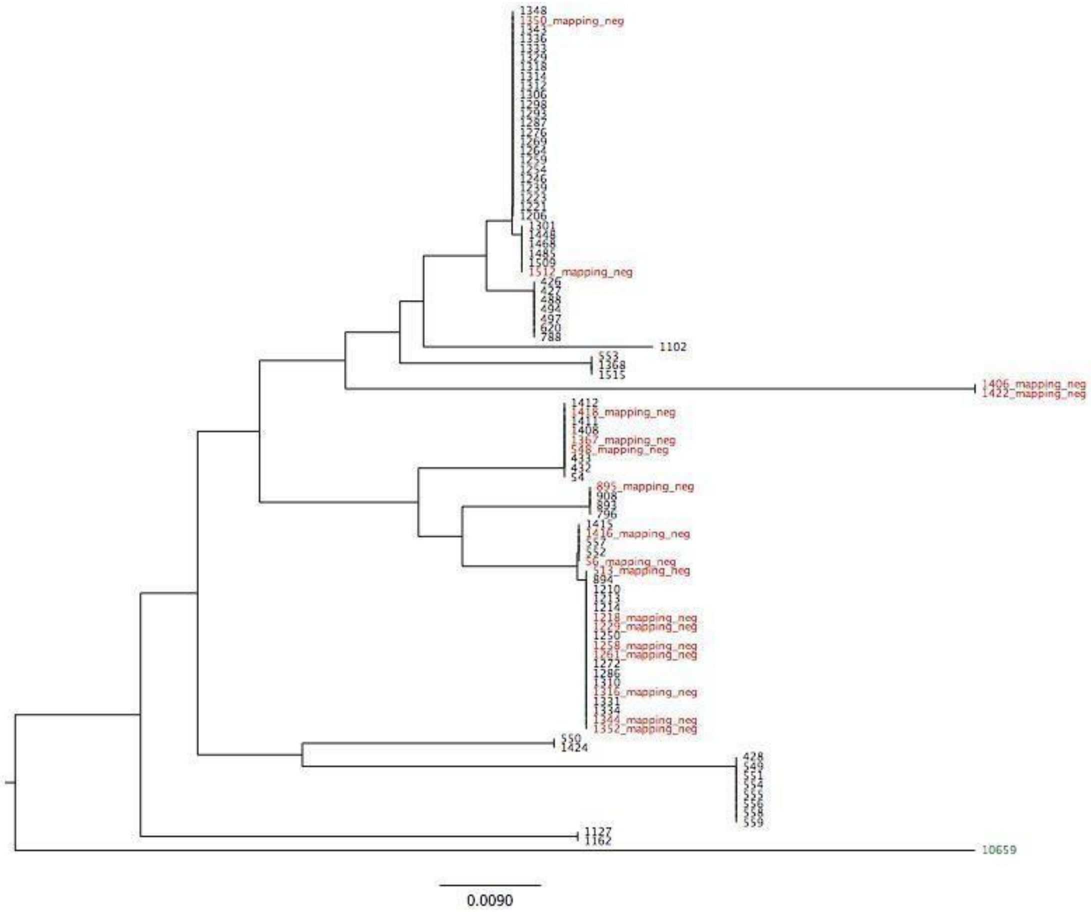


Figure 4.2 *ompA* sequences extracted from all *H. haemolyticus* annotations.

Neighbour joining tree created in MEGA 6.06 from MUSCLE alignment with bootstrapping set to 500. Sequences from annotations for the mapping negative isolates (red) were compared to sequences from annotations from mapping positive isolates (black) to confirm relationship of *ompA* sequences. All annotated sequences from mapping negative isolates mapped closely to other sequences confirming the presence of *ompA* within these isolates. Sequence 10659 (green) is taken from the annotation of NCTC reference isolate 10659. Scale bar shows 0.009 substitutions per 100 bp.

Protein E is a surface protein that has displayed the capability to bind to vitronectin and plasminogen to enable avoidance of immune clearance, in addition to binding to ICAM-1 and laminin (Barthel et al., 2012a, Barthel et al., 2012b, Hallstrom et al., 2011, Ronander et al., 2009). All NTHi isolates returned a sequence for *pE* (Table 4.1). Two *H. haemolyticus* did not return a sequence through mapping but were identified as positive by extracting a sequence from an annotated assembly (Materials and Methods, section 2.3.7). The resulting sequences were submitted to BLAST and returned 82% identity to adhesin protein E (GenBank reference CP002277.1:414179-414662) (Materials and Methods, section 2.3.7).

4.4.2 The Autotransporters – *hmwA*, *hmw2A*, *hsf*, *hia* and *hap*

Two loci, *hmwA* and *hmw2A*, are responsible for the high molecule weight proteins Hmw1 and Hmw2 (Barenkamp and St Geme, 1994, Winter and Barenkamp, 2014, Rempe et al., 2016). The autotransporter adhesin genes *hmwA*, *hmw2A*, *hia* and *hap* have been reported to be more associated with disease such as otitis media and meningitis (Ecevit et al., 2004, Cardines et al., 2007, Rodriguez et al., 2003, Davis et al., 2014b). Mapping to the reference sequence resulted in 1005/1368 (73.5%) of NTHi returning an *hmwA* sequence and 1070 (78.2%) returning an *hmw2A* sequence (Table 4.1, Figure 4.1). All isolates positive for *hmwA* were also positive for *hmw2A* leaving 65 isolates positive for *hmw2A* only. All 92 *H. haemolyticus* isolates were negative for both *hmwA* and *hmw2A* mapping (Figure 4.1).

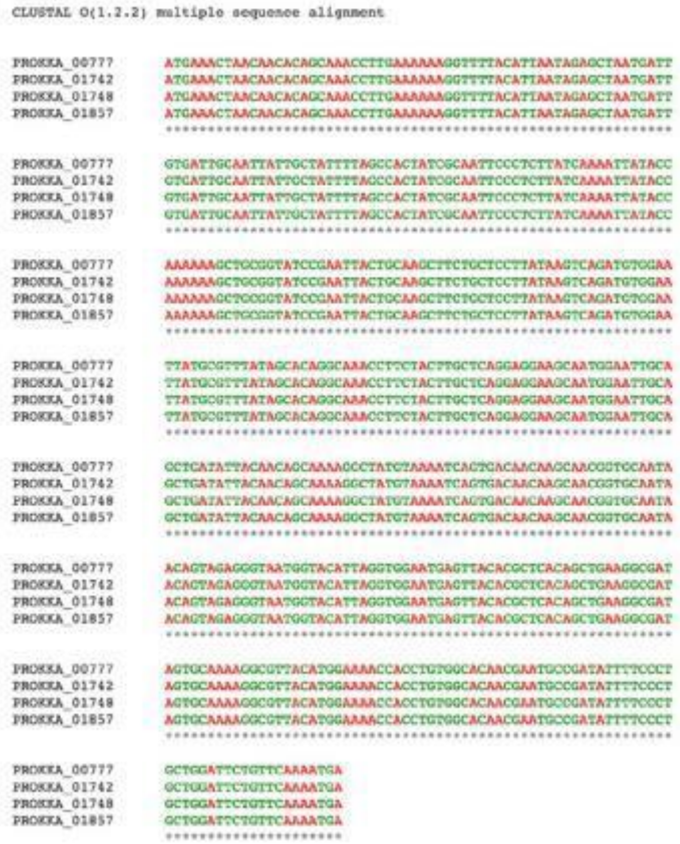
hsf is predominantly found in capsulated strains only and not generally observed in NTHi, this was reflected in the results from mapping with all 1460 isolates of both species negative for the presence of *hsf* (Cotter et al., 2005, St Geme et al., 1996).

The *H. influenzae* adhesin, Hia, has also been associated with invasive disease and is predominantly observed in strains that do not contain the *hmw* genes (St Geme et al., 1998, Cardines et al., 2007, Satola et al., 2008, Rodriguez et al., 2003). *hia* is historically reported in low amounts within strains with phase variation in the form of a long poly thymidine tract (poly T tract) mediating protein expression. Mapping for *hia* was carried out against two GenBank reference sequences where one contained an example of the Poly T tract and one did not (Table 4.1). No *H. haemolyticus* were positive for either *hia* mapping sequence and only 50/1368 (3.7%) of NTHi were positive (Figure 4.1).

The *Haemophilus* adhesin protein (*hap*) is thought to be ubiquitous throughout NTHi however the expression of the Hap protein has also been associated with invasive disease alongside the presence of Hia in encapsulated strains (Rodriguez et al., 2003). Three reference sequences from GenBank were used after initial mapping analysis to U11024.1 resulted in only 733/1368 NTHi returning a sequence (Table 4.1). From the further mapping using all three sequences all NTHi isolates returned a *hap* sequence, and 628 mapped to all three sequences. All *H. haemolyticus* were negative for the *hap* mapping (Figure 4.1).

4.4.3 Protusions - *hif* and *pilA*

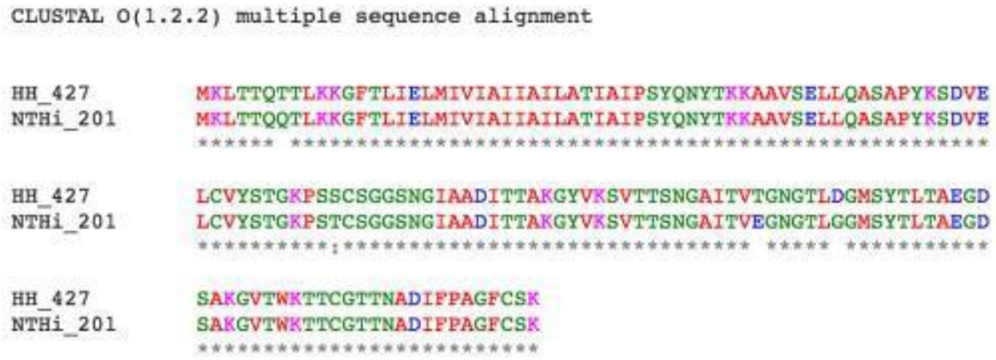
PilA is the gene responsible for the major pilus of the type IV secretion system and is thought to bind to cell adhesion molecule ICAM-1 (Novotny and Bakaletz, 2016). Mapping to the reference from GenBank returned sequences for 1365/1368 NTHi and 4/92 *H. haemolyticus* (Table 4.1,Figure 4.1). Further investigation was carried out into the three negative NTHi isolates. Genomes were assembled and annotated and *pilA* searched for in the annotations using the reference sequence (Materials and Methods, section 2.3.7). All three NTHi isolates negative for mapping were found to harbour a sequence for *pilA*. These were aligned with a representative sequence from a positive mapping isolate to confirm *pilA* presence. All four sequences shared 100% identity (Figure 4.3). All annotated genomes for *H. haemolyticus* were further investigated for *pilA* which resulted in a similar sequence to the NTHi *pilA* sequence being identified in all *H. haemolyticus*. Representative sequences from NTHi and were translated into protein sequences and when compared were seen to show protein homology of 97.3% (Figure 4.4) (Materials and Methods, section 2.3.7).



```
CLUSTAL O(1.2.2) multiple sequence alignment

PROKKA_00777      ATGAAACTAACAACACAGCAAACCTTGAAAAAAGGTTTTACATTAATAGAGCTAATGATT
PROKKA_01742      ATGAAACTAACAACACAGCAAACCTTGAAAAAAGGTTTTACATTAATAGAGCTAATGATT
PROKKA_01748      ATGAAACTAACAACACAGCAAACCTTGAAAAAAGGTTTTACATTAATAGAGCTAATGATT
PROKKA_01857      ATGAAACTAACAACACAGCAAACCTTGAAAAAAGGTTTTACATTAATAGAGCTAATGATT
                  ************************************************************

PROKKA_00777      GTGATTGCAATTATTGCTATTTTAGCCACTATCGCAATTCCCTCTTATCAAAATTATACC
PROKKA_01742      GTGATTGCAATTATTGCTATTTTAGCCACTATCGCAATTCCCTCTTATCAAAATTATACC
PROKKA_01748      GTGATTGCAATTATTGCTATTTTAGCCACTATCGCAATTCCCTCTTATCAAAATTATACC
PROKKA_01857      GTGATTGCAATTATTGCTATTTTAGCCACTATCGCAATTCCCTCTTATCAAAATTATACC
                  ************************************************************

PROKKA_00777      AAAAAAGCTGCGGTATCCGAATTACTGCAAGCTTCTGCTCCTTATAAGTCAGATGTGGAA
PROKKA_01742      AAAAAAGCTGCGGTATCCGAATTACTGCAAGCTTCTGCTCCTTATAAGTCAGATGTGGAA
PROKKA_01748      AAAAAAGCTGCGGTATCCGAATTACTGCAAGCTTCTGCTCCTTATAAGTCAGATGTGGAA
PROKKA_01857      AAAAAAGCTGCGGTATCCGAATTACTGCAAGCTTCTGCTCCTTATAAGTCAGATGTGGAA
                  ************************************************************

PROKKA_00777      TTATGCGTTTATAGCACAGGCAAACCTTCTACTTGCTCAGGAGGAAGCAATGGAATTGCA
PROKKA_01742      TTATGCGTTTATAGCACAGGCAAACCTTCTACTTGCTCAGGAGGAAGCAATGGAATTGCA
PROKKA_01748      TTATGCGTTTATAGCACAGGCAAACCTTCTACTTGCTCAGGAGGAAGCAATGGAATTGCA
PROKKA_01857      TTATGCGTTTATAGCACAGGCAAACCTTCTACTTGCTCAGGAGGAAGCAATGGAATTGCA
                  ************************************************************

PROKKA_00777      GCTGATATTACAACAGCAAAAGGCTATGTAAAATCAGTGACAACAAGCAACGGTGCAATA
PROKKA_01742      GCTGATATTACAACAGCAAAAGGCTATGTAAAATCAGTGACAACAAGCAACGGTGCAATA
PROKKA_01748      GCTGATATTACAACAGCAAAAGGCTATGTAAAATCAGTGACAACAAGCAACGGTGCAATA
PROKKA_01857      GCTGATATTACAACAGCAAAAGGCTATGTAAAATCAGTGACAACAAGCAACGGTGCAATA
                  ************************************************************

PROKKA_00777      ACAGTAGAGGGTAATGGTACATTAGGTGGAATGAGTTACACGCTCACAGCTGAAGGCGAT
PROKKA_01742      ACAGTAGAGGGTAATGGTACATTAGGTGGAATGAGTTACACGCTCACAGCTGAAGGCGAT
PROKKA_01748      ACAGTAGAGGGTAATGGTACATTAGGTGGAATGAGTTACACGCTCACAGCTGAAGGCGAT
PROKKA_01857      ACAGTAGAGGGTAATGGTACATTAGGTGGAATGAGTTACACGCTCACAGCTGAAGGCGAT
                  ************************************************************

PROKKA_00777      AGTGCAAAAGGCGTTACATGGAAAACCACCTGTGGCACAACGAATGCCGATATTTTCCCT
PROKKA_01742      AGTGCAAAAGGCGTTACATGGAAAACCACCTGTGGCACAACGAATGCCGATATTTTCCCT
PROKKA_01748      AGTGCAAAAGGCGTTACATGGAAAACCACCTGTGGCACAACGAATGCCGATATTTTCCCT
PROKKA_01857      AGTGCAAAAGGCGTTACATGGAAAACCACCTGTGGCACAACGAATGCCGATATTTTCCCT
                  ************************************************************

PROKKA_00777      GCTGGATTCTGTTCAAAATGA
PROKKA_01742      GCTGGATTCTGTTCAAAATGA
PROKKA_01748      GCTGGATTCTGTTCAAAATGA
PROKKA_01857      GCTGGATTCTGTTCAAAATGA
                  *********************
```

Figure 4.3 *pilA* alignment of mapping negative isolates alongside mapping positive representative.

Alignment carried out using Clustal Omega. PROKKA_0077 is the annotated sequence extracted from a representative isolate positive for mapping for *pilA*. PROKKA_01742,01748 and 01857 are *pilA* sequences extracted from the annotated assemblies of three isolates negative for mapping. There is 100% identity between the four sequences confirming the *pilA* positive status of the three isolates negative for mapping.



```
CLUSTAL O(1.2.2) multiple sequence alignment

HH_427          MKLTTQTTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKSDVE
NTHi_201        MKLTTQQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKSDVE
                ****** *****************************************************

HH_427          LCVYSTGKPSSCSGGSNGIAADITTAKGYVKSVTTSNGAITVTGNGTLDGMSYTLTAEGD
NTHi_201        LCVYSTGKPSTCSGGSNGIAADITTAKGYVKSVTTSNGAITVEGNGTLGGMSYTLTAEGD
                **********:******************************* ***** ***********

HH_427          SAKGVTWKTTCGTTNADIFPAGFCSK
NTHi_201        SAKGVTWKTTCGTTNADIFPAGFCSK
                **************************
```

Figure 4.4 Protein comparison of translated nucleotide *pilA* sequences from representative isolates 201 (NTHi) and 427 (*H. haemolyticus*).

Clustal omega alignment. 142/146 amino acids resulted in protein homology between NTHi and *H. haemolyticus* of 97.3% between the sequences.

Hif is a fimbriae protein known to bind to respiratory mucins and is isolated more predominantly in throats of healthy people rather than that of the middle ear of otitis media patients (Ecevit et al., 2004, Kubiet et al., 2000). From the mapping to the reference strain, sequences were returned for 102 (7.5%) isolates and no sequence was observed in any of the 92 *H. haemolyticus* isolates (Figure 4.1). The 102 *hif* positive isolates included the atypical group (n=76) and isolates from ST 154 and ST 156. All isolates of ST 154 (n=20) were *hif* positive plus an additional six isolates from ST 156. ST was previously identified in Chapter 3 from MLST analysis (Chapter 3, section 3.4.1). All *hif* positive isolates were designated STs novel to this study except the six isolates from ST 156.

4.4.4 Summary of Presence and Absence of Adhesins

All *H. haemolyticus* were negative for *omp1*, *hmwA*, *hmw2A*, *hif*, *omp2*, *hia* and *hap.* Only *hap* resulted in a sequence for all 1368 NTHi from these markers (Figure 4.1). For the NTHi isolates 1266/1368 (92.5%) were positive for *omp1*, 1005 (73.5%) *hmwA*, 1070(78.2%) *hmw2A*, 102 (7.5%) *hif*, 1264 (92.4%) *ompP* and 50 (3.7%) *hia*. The atypical NTHi group of isolates were all *omp2*, *hmwA*, *hmw2A*, *hia* and *omp1* negative but also made up 74.5% (76/102) of the group that were *hif* positive (Figure 4.1). All isolates from both species were positive for *omp6*, *pilA*, *pE*, *hel* and *ompA* (Figure 4.1). These adhesins were therefore chosen for further analysis for allelic variation of the sequences to determine the relationship between the species at each loci.

4.4.5 Allelic Variation of *ompA*, *omp6*, *pilA, hel* and *pE*

Isolates for both NTHi and *H. haemolyticus* returned results for *ompA*, *omp6*, *pilA*, *hel* and *pE* and were further investigated to determine the relationship of the alleles found within these adhesins between the two species. Sequences returned from either mapping or annotations were aligned for each gene and phylogenetic analysis carried out to visualise how the isolates clustered together (Figure 4.5-Figure 4.9). *H. haemolyticus* isolates clustered within their own clade for *ompA*, *omp6*, *pilA and hel* demonstrating sufficient sequence variation between the two species to separate them into identifiable clades. In the case of *pE* however this is not as clear with a branch of NTHi sitting in amongst the clustering of the *H. haemolyticus,* this NTHi branch is made up of all NTHi isolates from STs 513, 311, 1163 and 704. The group of atypical NTHi isolates clustered independently on branches within phylogenetic trees for *ompA*, *hel* and *pE* but not in *pilA* or *omp6* (Figure 4.5-Figure 4.9). ST 503 clustered with the atypical NTHi isolates in *ompA* but showed expected genotype for all adhesins indicating a relationship between ST 503 and the atypical NTHi isolates at the *ompA* sequence (Figure 4.8).

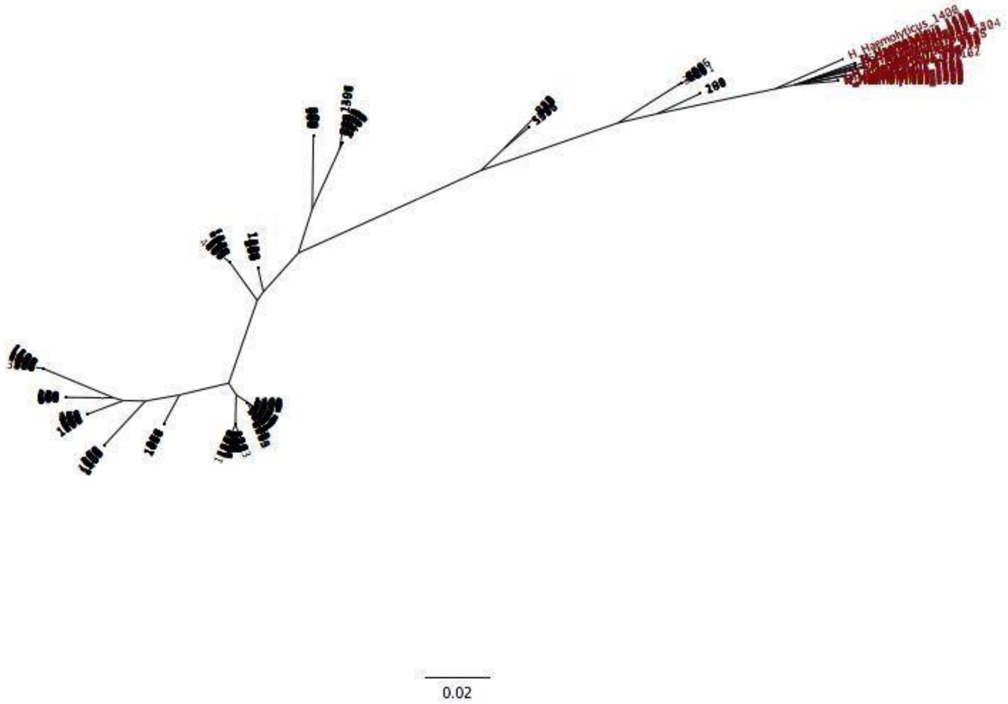


Figure 4.5 Relationship of *pilA* sequences extracted from annotations.

Neighbour joining tree, bootstrapping set at 500. *H. haemolyticus* in red and NTHi in black. *H. haemolyticus* can be seen to cluster away from NTHi displaying the variation between the two species within *pilA*. These sequences when translated however show protein homology at 97.3%. Sequences from atypical NTHi isolate did not cluster separately from other NTHi. Substitutions at 0.02 per 100 bp.

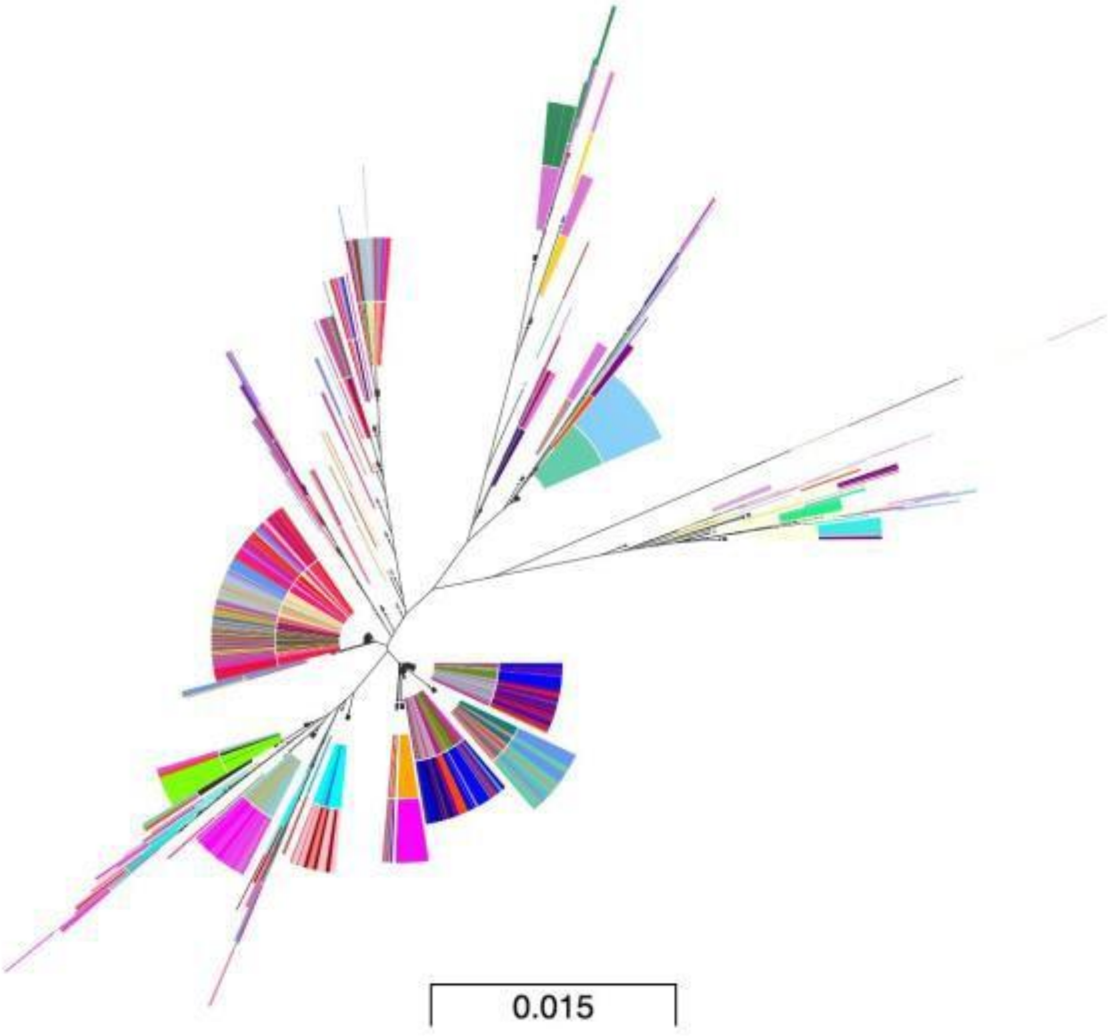


Figure 4.6 Distribution of sequences for *omp6* alongside ST's.

Neighbour joining tree with bootstrapping set to 500. ST's are represented by colours in the inner circle, patients are depicted by outer circle. *H. haemolyticus* are coloured light yellow and cluster separately to the right of the tree. This is a still of the interactive tree available for more in depth investigation created in microreact (https://microreact.org/project/ryTBYIEHx) The NTHi are distributed in different branches however the branches do not cluster into STs and isolates from the same ST are found in different branches.

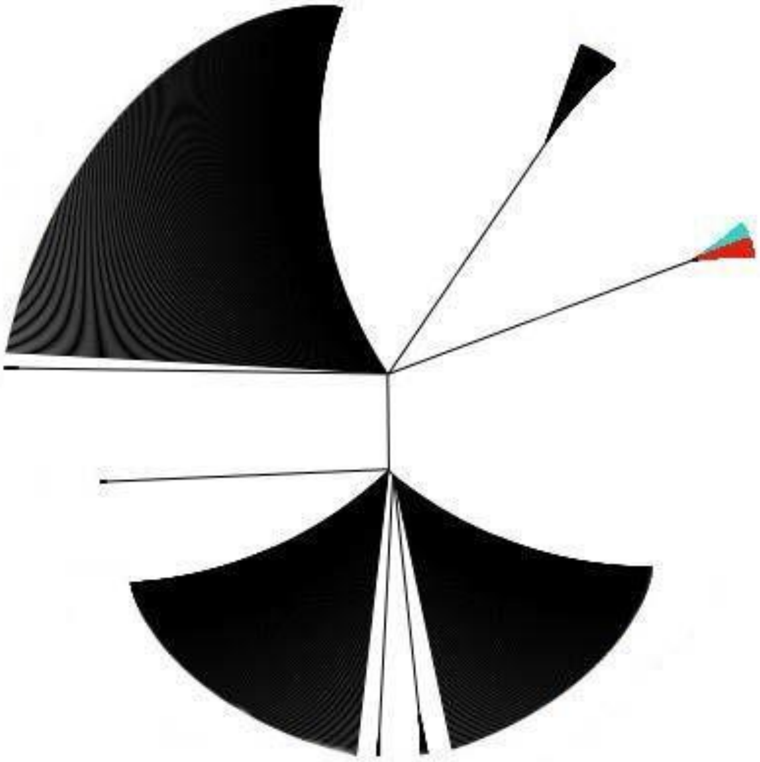


Figure 4.7 Sequences for *hel* resulting from mapping.

Neighbour joining tree bootstrapping set to 500. NTHi represented by black clades, red represents *H. haemolyticus* and green atypical isolates from ST 1314, ST 353 and ST 356. The atypical isolates clustered in the same clade as *H. haemolyticus* away from the other NTHi illustrating a relationship between the atypical NTHi isolates and *H. haemolyticus* within the *hel* sequences.

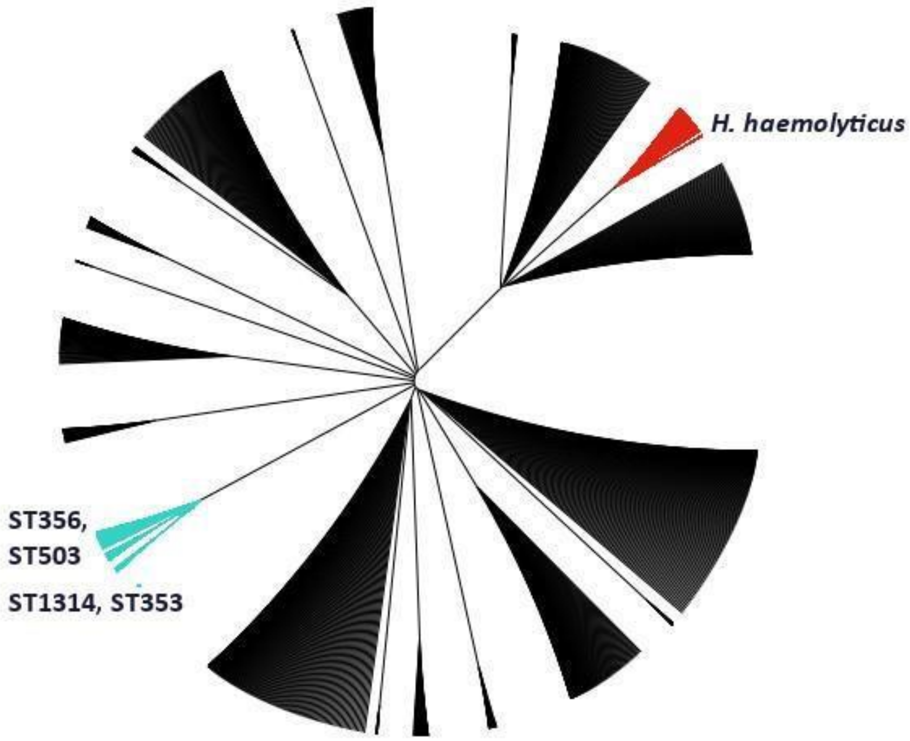


Figure 4.8 Sequences for *ompA*.

Neighbour joining tree bootstrapping at 500. The *H. haemolyticus* cluster in their own clade. The atypical NTHi isolates sit within the same clade (green) however all isolates with ST 503 are also present, indicating a relationship between the atypical NTHi isolates and ST 503 isolates at the *ompA* sequence.

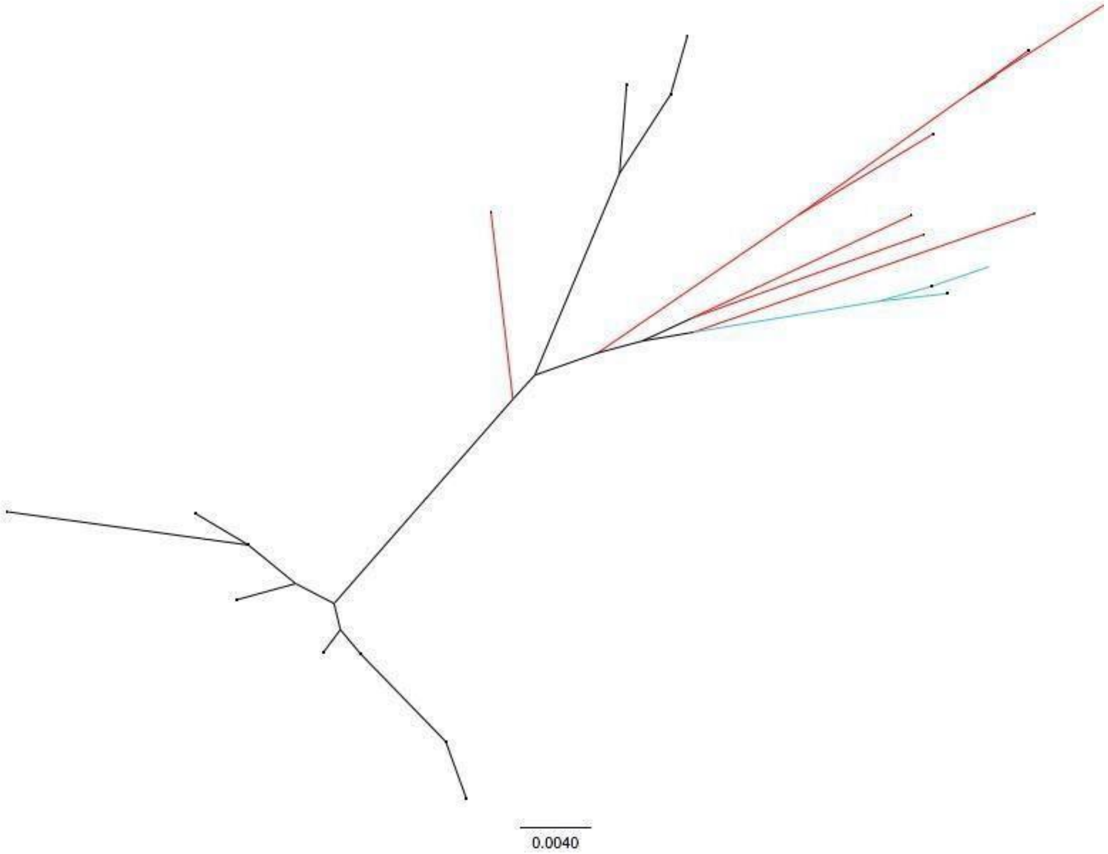


Figure 4.9 Protein E sequences.

Neighbour joining tree of protein E sequences, bootstrapping set to 500. A subset of NTHi branching from within the *H. haemolyticus* (red) clade composed of atypical isolates (green) from ST 1314, ST 356, ST 353. A branch of NTHi can be seen in amongst the *H. haemolyticus* illustrating a closer phylogenetic relationship within the protein E sequences between *H. haemolyticus* and these NTHi isolates from STs 513, 311, 1163 and 704 identified in MLST analysis Chapter 3.

4.4.6 Adhesin Genotypes Are Not Associated With the Onset of Acute Exacerbations

The isolates from this study were collected from patients either on a scheduled monthly appointment at stable periods or on the event of acute exacerbation. *Hia, hmwA, hmw2A, hap* and *hif* adhesin genes have been associated with disease. To determine whether the adhesins were more associated with acute exacerbation events in COPD the genotyping data was statistically investigated however this was not found to be the case. Chi square analysis was carried out on an adjusted sample set to remove duplicates. Presence or absence of an adhesin was not found to be significantly associated with the onset of acute exacerbation in patients (Materials and Methods, section 2.5)(Appendix 2, Appendix 3.2).

4.4.7 *hia* - Intrapatient Variation

Hia has been shown to be highly immunogenic but has a survival strategy in the form of phase variation (Atack et al., 2015). This is achieved by reducing Hia protein expression due to the extension of a polythymine (T) tract downstream from the promoter region. Despite this phase variation however a surprisingly low number of *hia* positive NTHi were isolated 50/1368 (3.7%) from this study (Figure 4.1). From the 50 positive isolates *hia* positive strains were only observed in two patients, patient 29 and patient 32, at more than one time point (Appendix 2). These were further broken down into sequences containing the poly T tract and sequences where it was absent. The poly T tract was found, on average (n=25/50), in half of all the *hia* sequences from each time point (Table 4.2). All adhesin genotypes were inspected for these two patients at each time point (Figure 4.10). In patient 32 *hia* was the only adhesin present in the first two time points and was not present after this. In the remaining time points where NTHi was isolated *hap* and *hif*, and *hmw* and *hap* positive isolates were the adhesins present, at each change of adhesin the ST also changed (Figure 4.10). NTHi isolated from patient 29 however maintained the same ST 311 throughout all time points, *hia* was observed in the first three time points simultaneously with *hap* but gave way to *hap* positive isolates only in the remaining time points (Figure 4.10).

Table 4.2 *hia* positive strains and the distribution of poly T tract indicative of phase variation.

Patient	Time point	No of *hia* positive isolates	Poly –t tract observed
29	1	14	7
29	2	8	4
29	3 (exacerbation)	10	5
32	1	9	3
32	2	9	6

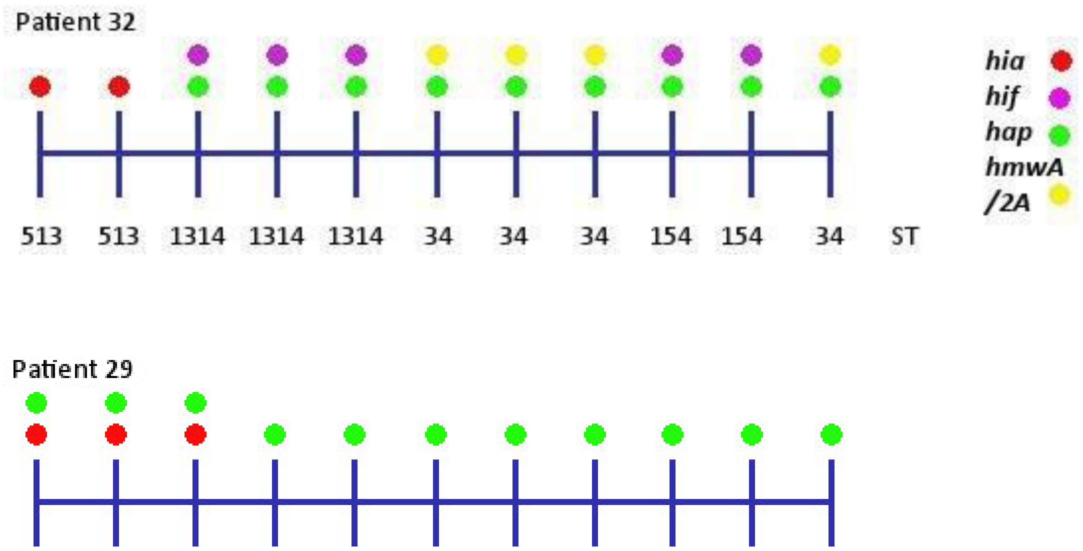


Figure 4.10 Distribution of poly T tract in hia positive isolates.

hia positive isolates were observed in more than one time point in two patients. However, *hia* was not seen again after these initial time points and replaced by other adhesin genotypes. Change in adhesin genotype also equated to a change in ST type within patient 32, however patient 29 ST 311 was the continual single ST throughout all time points despite a change in adhesin genotype.

4.5 Discussion

Here it has been demonstrated that *H. haemolyticus* and NTHi isolated from sputa from the COPD lung do indeed have distinct adhesin genotypes between the two species (Figure 4.1). Adhesins that have been previously reported to be more present in disease such as *hmwA* and *hmw2A, hia, hif* and *hap* are absent from in *H. haemolyticus* along with outer membrane protein genes *omp2* and *omp1* (Figure 4.1). Genes shared between them, *ompA, hel, pilA* and *omp6*, display allelic variation sufficient for speciation (Figure 4.5-Figure 4.9).

Adhesins are vital for bacteria to achieve colonization (Avadhanula et al., 2006, Barenkamp and St Geme, 1996, Gilsdorf et al., 1997, Hendrixson and St Geme, 1998, Davis et al., 2014a, Rempe et al., 2016). These adhesins also offer secondary mechanisms to provide methods of immune evasion, cellular internalisation and tissue destruction, facilitating pathogenicity (Fink et al., 2003, Atack et al., 2015, Davis et al., 2014a, Fink et al., 2002, Hallstrom et al., 2011, Spahich et al., 2014, Su et al., 2016, Barthel et al., 2012b, Hallstrom et al., 2009).

H. haemolyticus and NTHi are genetically very similar however despite this NTHi is a known causative organism of both non-invasive and invasive disease whereas *H. haemolyticus* is not classified as a pathobiont. There have been eight incidences of *H. haemolyticus* isolated from invasive disease but these are far from the norm (Hotomi et al., 2010, Murphy et al., 2007, Anderson et al., 2012, Morton et al., 2012, Jordan et al., 2011). NTHi uses its arsenal of adhesins to aid in causing disease therefore a difference in these adhesins associated with pathogenicity between NTHi and *H. haemolyticus* may begin to answer the contrast in disease causing ability between them.

1368 NTHi and 92 *H. haemolyticus* isolates were investigated for the presence of thirteen adhesin genes. All isolates were negative for the adhesion most associated with encapsulated *H. influenzae*, *hsf,* and all *H. haemolyticus* were negative for *hmwA*, *hmw2A* and *hia, hap, hif, omp2 and omp1* (Figure 4.1). Both *hmwA* and *hmw2A* were observed in 73.5% of the NTHi isolates from this study with another 4.8% positive for *hmw2A* only, this concedes with other studies reporting *hmw* positive rates of 37.7%-79.6% (St Geme et al., 1998, Satola et al., 2008, Ecevit et al., 2004, Erwin et al., 2005, Erwin et al., 2008, Davis et al., 2014b, Shahini Shams Abadi et al., 2016). Hmw proteins are thought to play an important role both in colonisation and pathogenesis in otitis media, they are reportedly immunogenic but have shown to avoid immune clearance by mediation of protein expression using phase variation in the form of 7bp repeats (Davis et al., 2014a, Cholon et al., 2008, Giufre et al., 2008, Winter and Barenkamp, 2014, Rempe et al., 2016, St Geme et al., 1993, Davis et al., 2014b, Ecevit et al., 2004).

Although a small number of isolates have been reported to harbor both *hmw* and *hia* this is not usually the case and this was also reflected in our results shown in Figure 4.1 (Barenkamp and St Geme, 1996, St Geme et al., 1998, Satola et al., 2008, Erwin et al., 2005, Ecevit et al., 2004, Erwin et al., 2008). From the 50/1368 *hia* positive isolates all were negative for *hmw* genes. This equated to 3.7% of all isolates being positive for *hia* which is relatively low in comparison to the 9-54.7% previously reported (Erwin et al., 2008, Ecevit et al., 2004, Erwin et al., 2005, Satola et al., 2008). Hia has been shown to be an immunogenic protein that is cleared by opsonophagocytic mechanisms however as mentioned previously, *hia* is also able to mediate protein expression by phase variation enabling an immune avoidance mechanism, but this can still result in low levels of protein being produced (Atack et al., 2015). Phase variation in Hia comes in the form of a poly T tract that diminishes protein expression with extension of the tract (Atack et al., 2015, Surana et al., 2004). From the 50 isolates which were positive for *hia* 50% also had a poly T tract present indicating that Hia expression was being reduced (Table 4.2).

Haemophilus surface fibril (*Hsf*) is a conserved homologue of *hia* found ubiquitously throughout encapsulated *H. influenzae* (Cotter et al., 2005, St Geme et al., 1996, Radin et al., 2009). *Hsf* is able to bind to vitronectin, a glycoprotein within the extra cellular matrix, which enables avoidance of the complement system (Singh et al., 2014, Hallstrom et al., 2006, Laarmann et al., 2002, Cotter et al., 2005). *Hsf* and *hia* display homology in two binding pockets therefore the binding capability of *hia* to vitronectin would be a worthwhile investigation to understand how Hia works within the ECM. However, the potential for *hia* to bind to vitronectin does not explain the disappearance of *hia* positive strains or indeed the relatively small percentages reported within studies, quite the opposite. C-reactive protein (CRP) is an acute phase reactant and inflammatory marker shown to predict mortality in COPD and coronary artery disease (Yaghoubi et al., 2015). The reduction of vitronectin in the presence of C-Reactive protein that has been reported in bacterial pneumonia could point to a case of competitive binding between CRP and *hia* for vitronectin (Kawahira, 1990). Hia may therefore also have more of a role in initial colonisation or instigating pathogenesis than in persistence in NTHi.

The atypical NTHi group contributed to 76/77 *omp1* negative isolates and 76/102 *omp2* negative. No *omp1* or *omp2* sequence was returned for any *H. haemolyticus* isolate. Currently there is limited literature with regards to distribution and prevalence of *omp1* throughout NTHi isolates although. It has however been identified as a potential vaccine candidate and therefore the absence of *omp1* in three STs (the atypical NTHi group) is an important consideration (Bolduc et al., 2000, Roier et al., 2012). Similarly, *omp2* has been identified as an important potential for vaccine efficacy, however previous studies have reported absence of *omp2* from NTHi isolates (Neary and Murphy, 2006, Yi and Murphy, 1997, Roier et al., 2012, Binks et al., 2012). *omp2* has also shown a high level of genetic variation in NTHi explaining the necessity to use more than one reference sequence when mapping (Forbes et al., 1992, Duim et al., 1994, Hiltke et al., 2002, Regelink et al., 1999, Troelstra et al., 1994).

Sequences for *hel*, *omp6*, *pilA*, *pE* and *hel* were identified in all NTHi and *H. haemolyticus*. All have been indicated in vaccine development however it has been revealed that co-colonisation of *H. haemolyticus* may have a detrimental effect on NTHi infection due to competitive binding to epithelial cells, therefore a vaccine candidate that targets a protein found within both species may have an unexpected deleterious effect on preventing NTHi infection (Hotomi et al., 2005, Esmaily et al., 2006, Badr et al., 1999, Berenson et al., 2005, Bertot et al., 2004, DeMaria et al., 1996, Hotomi et al., 1998, Hotomi et al., 2002, Hua et al., 2016, Pichichero et al., 2010, Novotny et al., 2009, Green et al., 2005, Green et al., 1991, Pickering et al., 2016). Further phylogenic analysis of the returned sequences demonstrated that allelic variance between the species were able to differentiate NTHi from *H. haemolyticus* except in the case of *pE* which contained an NTHi branch in amongst *H. haemolyticus* (Figure 4.9). Separate clustering between the two species for *pilA*, *hel*, *ompA* and *omp6* indicated a sufficiently conserved sequence structure within each species to enable speciation between them (Figure 4.5-Figure 4.8). Despite the difference in nucleotide sequence protein translation of returned sequences for *pilA* exhibited 97.3% homology between NTHi and *H. haemolyticus* (Figure 4.4).

pE and *omp6* have been reported as ubiquitous throughout NTHi although *omp6* has demonstrated a high level of sequence variation (Chang et al., 2010, Chang et al., 2011, Singh et al., 2010).

A sequence for *hap* was present in all of the NTHi isolates but not *H. haemolyticus.* This supports previous studies that report the ubiquitous nature of hap throughout the NTHi population (Fink et al., 2002, De Chiara et al., 2014). Hap is able to form microcolonies by adhering to itself. At an optimum concentration Hap overcomes secretory leucocyte protease inhibitor (SLPI) and is able to initiate autoproteolysis (Meng et al., 2011, Finket al., 2003, Fink and St Geme, 2003, Hendrixson and St Geme, 1998, DALYs et al., 2015). Sequence variation in *hap* has also been reported and this has been reflected in our isolates with three reference sequences being required to ascertain sequences within all NTHi isolates (Cutter et al., 2002).

The presence of *hif* was observed in the 76 atypical isolates (ST 1314, ST 353 and ST 356) and also the full complement of ST 154 and ST 156 of which only ST 156 was not novel to this study (Figure 4.1, Chapter 3). No *H. haemolyticus* isolates contained a *hif* sequence. The presence of the *hif* cluster of genes is reportedly more prevalent in Hib isolates (64%) than in NTHi (18%-28%) however our study reports an even lower rate of 7.5% (Figure 4.1). There is limited literature into the prevalence of the *hif* cluster in NTHi isolated from COPD and *hif* cluster positive isolates have been shown to be more prevalent in NTHi isolated from the nasopharynx of healthy carriers than otitis media, therefore our relativelylow figure may reflect selective pressures on NTHi from the COPD lung environment (Geluk et al., 1998, Ecevit et al., 2004, Nakamura et al., 2006).

4.6 Conclusions and Future Work

In this chapter, five genes have been identified (*omp6, pilA, hel, pE* and *ompA*) that are shared between NTHi and *H. haemolyticus*, four (*omp6, pilA, hel,* and *ompA*) of which could be used to develop assays for identification between NTHi and *H. haemolyticus* due to allelic variation. *H. haemolyticus* isolates were negative for *hia*, *hmw*, *hmw2A* and *hap* which are observed in disease in addition to absences of *omp1* and *omp2.* In NTHi, *hap* was observed in all isolates however *omp1*, *omp2*, *hia*, *hmwA*, *hmw2A* and *hif* were not ubiquitous in NTHi. This has therefore illustrated that *H. haemolyticus* does not possess the same adhesin genotype as NTHi which may in turn propose an explanation for the difference in pathogenic capability between NTHi and *H. haemolyticus* but also highlights the genetic diversity of adhesin genotype throughout NTHi. It may be concluded that *omp6*, *pilA*, *hel* and *ompA* have limited roles in virulence due to the ubiquitous nature of the genes in both NTHi and *H. haemolyticus*.

The atypical NTHi isolates identified in Chapter 3 continue to show atypical traits within adhesin genotypes (Figure 4.1). The adhesin genotype for the atypical NTHi group was unusual for NTHi and more homologous to that of *H. haemolyticus* (negative for *hmwA, hmw2A*, *omp2*, *omp1* and *hia*) except for the presence of *hap* and *hif* within the atypical NTHi isolates which were not observed in *H. haemolyticus*. Gene markers do not enable full investigation of the isolates therefore full genome analysis would be important to determine whether these are examples of the hypothesised strains that bridge the gap between NTHi and *H. haemolyticus* (Binks et al., 2012). This will be addressed in Chapter 6.

Protein analysis and adherence assays of mutant strains to fully understand the differences between the two species would result in phenotypic confirmation of the genotypes reported in this study to give a better understanding of how the two species differ in their adhesion capability and how this aids pathogenesis in NTHi.
Investigating the mechanisms of Hia and its interactions with proteins vitronectin and CRP may assist in the understanding of the pathogenesis of the strains more associated with invasive disease such as paediatric meningitis and discover if the COPD lung environment subjects strains of NTHi positive for Hia production to selective pressures (Cardines et al., 2007, Satola et al., 2008).

5 Associating Genetic Mechanisms of Phenotypic Azithromycin Resistance within NTHi and *H. haemolyticus* Isolated from COPD Using Genomic Data

5.1 Introduction

Azithromycin is a macrolide used in the prophylactic treatment of patients with cystic fibrosis, bronchiestasis, supporative chronic lung disease, asthma, malaria and COPD. Adoption of long-term macrolide treatment has been questioned due to the potential for the accumulation of resistance to antibiotic treatment (Serisier, 2013, Serisier et al., 2013, Albert et al., 2011, Hare et al., 2015, Samson et al., 2016, Anderson et al., 1995b, Stokholm et al., 2016, Hansen et al., 2009, Kuschner et al., 1994, Davies and Wilson, 2004, Anwaret al., 2008). Resistance to azithromycin and treatment failure has been observed in many different bacterial species including *S. pneumoniae*, *S. aureus*, *Campylobacter* spp., *Salmonella enterica*, *Treponema palladium*, *Neisseria gonorrhoea* and *N. meningitidis.*

Whilst information on macrolide resistance in NTHi is limited what is available is contradictory and the observations differ by study (Chisholm et al., 2009, Chisholm et al., 2010, Hoge et al., 1998, Katz et al., 2012, Katz and Klausner, 2008, Van Damme et al., 2009, Lukehart et al., 2004, Galarza et al., 2010, Molloy et al., 2010, Zarantonelli et al., 2001, Luna et al., 2000, Phaff et al., 2006). Increased amounts of *Haemophilus* spp. isolated from cystic fibrosis were observed to harbour resistance to macrolides however this was not the case in an alternative study into NTHi isolated from cystic fibrosis patients on long term azithromycin treatment over a four year period whereby all isolates were susceptible to azithromycin (Phaff et al., 2006, Hansen et al., 2009). In a study into resistance of NTHi isolated from COPD, all 267 isolates from 77 patients were found to be susceptible at first isolation with resistance developing after exposure to treatment in five out of 27 strains further investigated (Pettigrew et al., 2016). However, studies have shown that there is a decrease in number of positive cultures after treatment with azithromycin inclusive of *Haemophilus* spp. (Hare et al., 2015, Davies and Wilson, 2004, Anwar et al.,2008).

There is very limited data available on the antimicrobial resistance levels of *H. haemolyticus,* which is known to be a commensal of the nasopharynx, especially in children, and is also isolated from the sputa collected from COPD patients (Murphy et al., 2007, Kirkham et al., 2010). The first incidence of resistance to ciprofloxacin within *H. haemolyticus* was recently reported but no information is available on the susceptibility of *H. haemolyticus* to macrolide treatment, more specifically azithromycin, in COPD (Marti et al., 2016). The capability for NTHi and *H. haemolyticus* to co-colonise the COPD lung, combined with the reported incidences of recombination between the two species may suggest a reciprocal potential for gene transfer associated with virulence or resistance as has been reported between Shigella and enteroinvasive *E. coli* (Murphy et al., 2007, de Gier et al., 2015, van den Beld and Reubsaet, 2012).

Macrolide resistance has been associated with the presence of a group of genes referred to as the acquired macrolide resistance genes (AMRG). These genes such as the *erm* operon and *mef*(A), *mef*(E), *mel*, acrAB, *mtr*(R) and *msr*(A) are responsible for efflux pumps or ribosomal modifications (Zarantonelli et al., 2001, Tait-Kamradt et al., 2000a, Weisblum, 1995b, Weisblum, 1995a, Peric et al., 2003, Sanchez et al., 1997, Seyama et al., 2016, Schroeder and Stephens, 2016, Grad et al., 2016). The *mel* gene in *S. pneumoniae* is also known as the *mrs(*D) homologue of *msr(*A) present within *S. aureus* (Schroeder and Stephens, 2016). The genetic mechanism for resistance in NTHi to macrolides however has been disputed. Roberts *et al*. (2011) reported a 100% positivity rate for at least one AMRG in 106 NTHi isolated from paediatric cystic fibrosis patients whereas Atkinson *et al*.(2015) were unable to identify any AMRGs in a study of 172 NTHi isolated from cystic fibrosis, non cystic fibrosis bronchiestasis and other clinical samples (Roberts et al., 2011, Atkinson et al., 2015). This suggests perhaps that AMRG is not widespread amongst NTHi and therefore other mechanisms of macrolide resistance may be prevalent. For example, modifications of the L4 and L22 ribosomal proteins which form a bottle neck of the peptide exit tunnel, as well as mutations of the domain V of the 23s rRNA have been associated with macrolide resistance (Diner and Hayes, 2009, Tait-Kamradt et al., 2000b, Clark et al., 2002, Peric et al., 2003). Modifications of L4 and L22 have been shown to affect the functionality of the large ribosome and result in resistance to macrolide treatment (Lovmar et al., 2004, Garza-Ramos et al., 2001). In addition azithromycin alters the assembly of the 50s subunit resulting in ribosome degradation (Petropoulos et al., 2009, Chittum and Champney, 1995).

In order to understand the prevalence and mechanisms of azithromycin resistance within our study, 104 isolates including 21 *H. haemolyticus* isolated from COPD were tested for susceptibility to azithromycin. This subset was chosen to represent all occurrences of *H. haemolyticus* and one example of each ST from each time point isolated from each patient isolated within the first six months. In addition, whole genome sequencing data was used to examine the presence and variation of genes and mutations previously reported to be associated with azithromycin resistance. Primarily the genes demonstrating resistance in NTHi have been reported as presence of the *erm* and *mef* genes or L4, L22 or 23srRNA mutations (Roberts et al., 2011, Peric et al., 2003).

5.2 Hypothesis

H. haemolyticus and NTHi will have similar susceptibility patterns when isolated from the same environment due to previously reported genetic mechanisms present.

5.3 Aims

- To identify levels of susceptibility to azithromycin in *H. haemolyticus* and NTHi isolated from COPD.

- To determine the genetic mechanisms of resistance in *H. haemolyticus* and NTHi by mapping genome sequence data to reference sequences of macrolide resistance genes as well as investigating mutations in ribosomal proteins L4 and L22 and 23s rRNA.

5.4 Results

5.4.1 Study Isolates

In Chapter 3 1460 *Haemophilus* spp. isolated from sputa collected from patients on the AERIS COPD study were differentiated into NTHi and *H. haemolyticus* (Bourne et al., 2014). The NTHi were also sequence typed using MLST on sequencing data (Chapter 3)(Meats et al., 2003). A subset of 104 isolates representing one NTHi isolate from each ST (n=83) and all *H. haemolyticus* that were isolated from the 16 patients over a six month period (defined in chapter 3) were chosen for Etest® and genetic analysis of sequencing data (Chapter 3, Appendix 2). Minimum inhibitory concentrations (MIC) using Etests® and EUCAST breakpoints were determined for these isolates (Testing., Version 6.0, 2016.).

5.4.2 Azithromycin Susceptibility by MIC

Etest® strips were placed on spread *Haemophilus* test agar plates of separate isolates and incubated over night (Materials and Methods, section 2.2.5) The MICs were identified from the Etest® at the relevant point and noted (Materials and Methods, section 2.2.5). Breakpoints for resistance were ascertained from EUCAST and were noted as azithromycin susceptibility at ≤ 0.125mg/L resistance at > 4mg/L (Testing., Version 6.0, 2016.).

MICs above 4 mg/L (6-24mg/L) were classified as resistant isolates and this was observed in 53/104 (51%) isolates. These consisted of 44/83 (53%) of NTHi and 9/21 (43%) of *H. haemolyticus.* No isolate was susceptible to azithromycin and the remaining 39/83 (47%) NTHi and 12/21 (57%) *H. haemolyticus* displayed intermediate resistance to azithromycin with MICs from 1-4mg/L (Figure 5.1).

Twelve patients had intermediate and resistant isolates. *Haemophilus* spp. with MIC's all deemed intermediate resistance were isolated from three patients and isolates with MIC's of ≥ 12mg/L were observed in *Haemophilus* spp. isolated from seven patients.

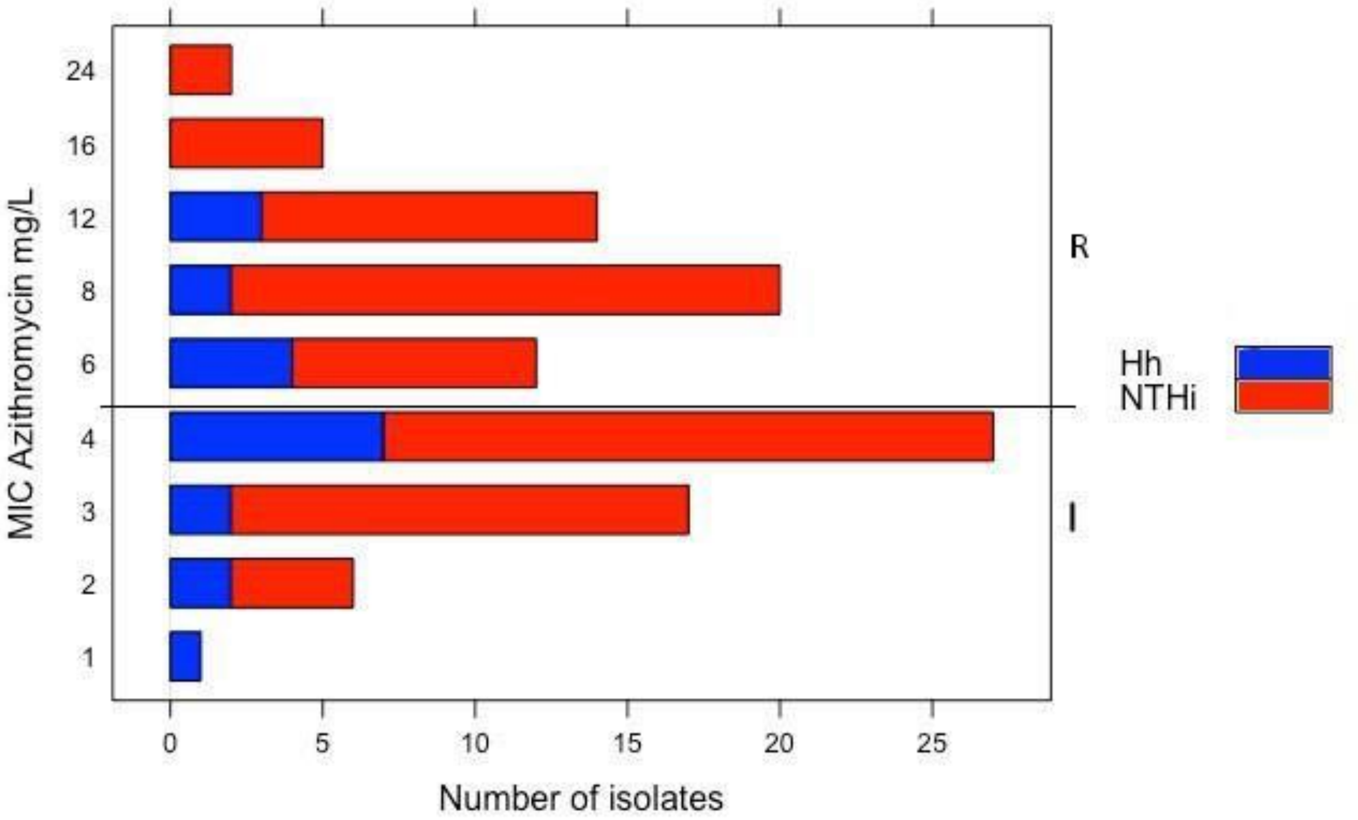


Figure 5.1 MIC of Azithromycin in NTHi and *H. haemolyticus*.

Fifty-three of 104 isolates with MIC's over 4mg/L were classed as resistant to azithromycin using EUCAST breakpoints. Of those 53 isolates, 44 (83%) were NTHi and 9 (17%) were H. *haemolyticus*. This equated to 42.9% of all *H. haemolyticus* and 53% of all NTHi isolates within the study. No isolate was susceptible for azithromycin (≤0.125mg/L). The remaining 39/83 (47%) of NTHi and 57.1% of *H. haemolyticus* isolates were classed as having intermediate resistance with MICs of over 0.125mg/L and up to and including 4mg/L.

5.4.3 Genetic Mechanisms

In silico PCR was used to extract sequences for genes L4 and L22 from the sequence data of the 104 isolates used for this analysis. This was done using primers that had previously been published from Peric *et al.* (2003). Primers from Atkinson *et al.* (2015) were used to ascertain presence of macrolide resistance genes *erm(a)*, *erm(b)*, *erm(c)*, *erm(F)* and *mef(a)* (Atkinson et al., 2015, Peric et al., 2003). No macrolide resistant gene was found within any isolate as has been previously reported in *H. influenzae*, primers were tested against positive controls (Materials and Methods, section 2.3.7). The majority of isolates returned a gene sequence for L22 (102/104, 98.1%) and L4 (102/104, 98.1%) (Peric et al., 2003).

It has been reported that a deletion of amino acids MKR (methionine, lysine and arginine) at position 82 in the L22 protein sequence is associated with the increased resistance to macrolides. Macrolides bind to a site close to the constriction of the peptide exit tunnel formed by the L4 and L22 proteins and inhibit protein synthesis on binding. The deletion of the MKR amino acids is thought to result in a widening of the peptide exit tunnel allowing the entry of nascent proteins despite the presence of bound macrolides (Clark et al., 2002, Tu et al., 2005, Gabashvili et al., 2001, Davydova et al., 2002).

L22 sequences were collated from annotated genomes (Material and Methods, section 2.3.7). The resulting sequences were translated into protein sequences using EMBOSS and then aligned in Clustal Omega (Materials and Methods, section 2.3.7). Three isolates did not return a sequence, 98 were 100% identical with one isolate truncated with 14 amino acids removed from the end of the sequence but identical to the remaining otherwise. All protein translations contained the MKR protein sequence.

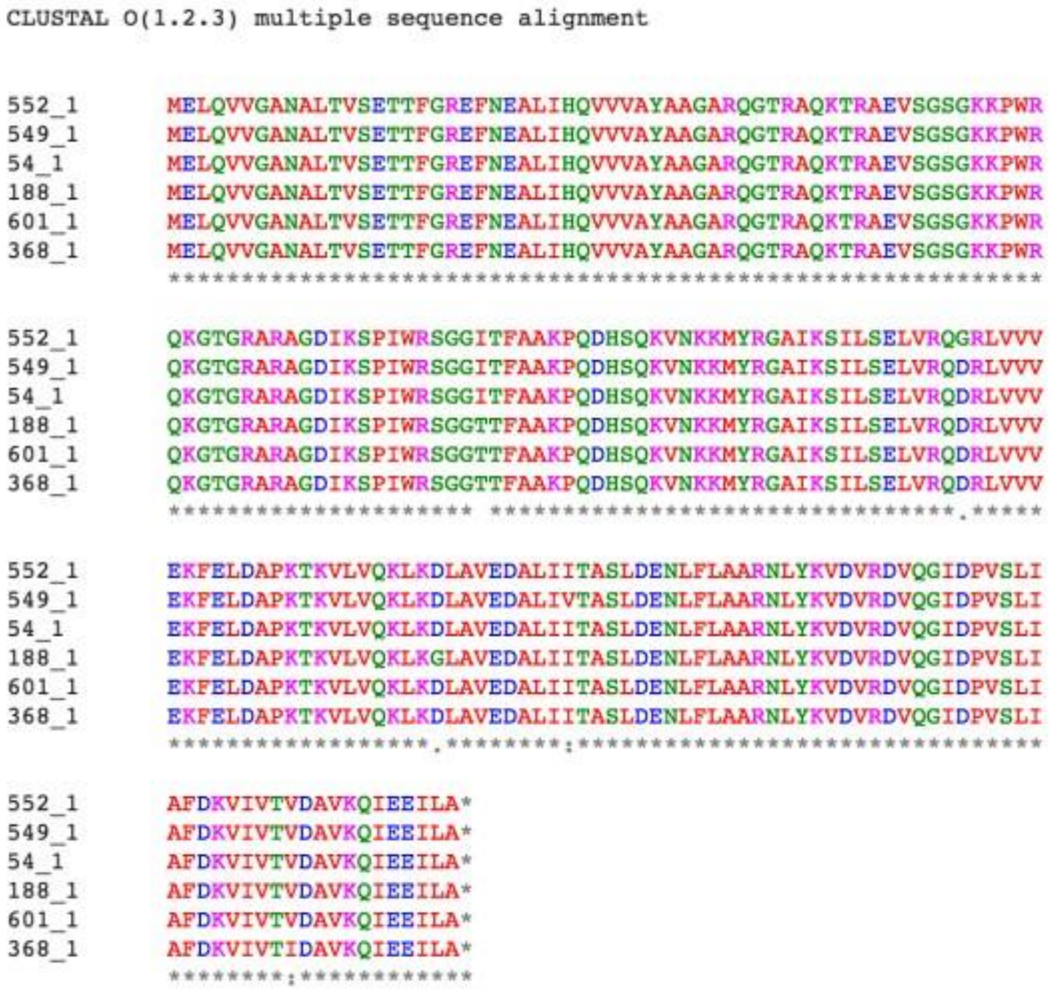
No substitution was observed at position 54 from EVSG to EVSA and no mutations were observed in the conserved L4 sequence KPWRQKGTGRARA associated with macrolide resistance as previously reported in *H. influenzae* and as previously observed in *S. pneumoniae* (Tait-Kamradt et al., 2000b, Clark et al., 2002).

Protein translations from annotated sequences for L4 were also carried out. One isolate returned a truncated sequence with the first 50 amino acids absent. Compared to NTHi sequences, all *H. haemolyticus* isolates displayed an isoleucine instead of a threonine at position 82 indicating a species based substitution rather than a resistance associated mutation (figure 5.5). At amino acid position 116, three *H. haemolyticus* displayed an aspartic acid rather than a glycine represented by isolate 552 in figure 5.5, and in seven *H. haemolyticus* an isoleucine was observed instead of a valine at position 149 (isolate 549 in figure 5.5). Two further isolates of NTHi exhibited novel protein sequences with one showing an aspartic acid to glycine at position 139 (isolate 188, Figure 5.5) and the other a valine instead of an isoleucine in position 189 (isolate 368, Figure 5.5). None of the aforementioned substitutions were associated with an increased MIC to azithromycin.

METIAKHRYARTSAQKARLVADLIRGKKVAQALEILTFTNKKAAALVKKVLESAIANAEH
NDGADIDDLKVAKIFVDEGPS**MKR**VMPRAKGRADRILKRTSHITVVVSDR*

Figure 5.4 Protein Translation of L22 Nucleotide Sequences.

All isolates positive for L22 showed 100% homology except one isolate missing the final 14 amino acids. The MKR deletion reportedly responsible for the increased resistance to macrolides was not observed in any of the isolates.



```
CLUSTAL O(1.2.3) multiple sequence alignment


552_1           MELQVVGANALTVSETTFGREFNEALIHQVVVAYAAGARQGTRAQKTRAEVSGSGKKPWR
549_1           MELQVVGANALTVSETTFGREFNEALIHQVVVAYAAGARQGTRAQKTRAEVSGSGKKPWR
54_1            MELQVVGANALTVSETTFGREFNEALIHQVVVAYAAGARQGTRAQKTRAEVSGSGKKPWR
188_1           MELQVVGANALTVSETTFGREFNEALIHQVVVAYAAGARQGTRAQKTRAEVSGSGKKPWR
601_1           MELQVVGANALTVSETTFGREFNEALIHQVVVAYAAGARQGTRAQKTRAEVSGSGKKPWR
368_1           MELQVVGANALTVSETTFGREFNEALIHQVVVAYAAGARQGTRAQKTRAEVSGSGKKPWR
                ************************************************************

552_1           QKGTGRARAGDIKSPIWRSGGITFAAKPQDHSQKVNKKMYRGAIKSILSELVRQGRLVVV
549_1           QKGTGRARAGDIKSPIWRSGGITFAAKPQDHSQKVNKKMYRGAIKSILSELVRQDRLVVV
54_1            QKGTGRARAGDIKSPIWRSGGITFAAKPQDHSQKVNKKMYRGAIKSILSELVRQDRLVVV
188_1           QKGTGRARAGDIKSPIWRSGGTTFAAKPQDHSQKVNKKMYRGAIKSILSELVRQDRLVVV
601_1           QKGTGRARAGDIKSPIWRSGGTTFAAKPQDHSQKVNKKMYRGAIKSILSELVRQDRLVVV
368_1           QKGTGRARAGDIKSPIWRSGGTTFAAKPQDHSQKVNKKMYRGAIKSILSELVRQDRLVVV
                ********************* ********************************.*****

552_1           EKFELDAPKTKVLVQKLKDLAVEDALIITASLDENLFLAARNLYKVDVRDVQGIDPVSLI
549_1           EKFELDAPKTKVLVQKLKDLAVEDALIVTASLDENLFLAARNLYKVDVRDVQGIDPVSLI
54_1            EKFELDAPKTKVLVQKLKDLAVEDALIITASLDENLFLAARNLYKVDVRDVQGIDPVSLI
188_1           EKFELDAPKTKVLVQKLKGLAVEDALIITASLDENLFLAARNLYKVDVRDVQGIDPVSLI
601_1           EKFELDAPKTKVLVQKLKDLAVEDALIITASLDENLFLAARNLYKVDVRDVQGIDPVSLI
368_1           EKFELDAPKTKVLVQKLKDLAVEDALIITASLDENLFLAARNLYKVDVRDVQGIDPVSLI
                ******************.********:********************************

552_1           AFDKVIVTVDAVKQIEEILA*
549_1           AFDKVIVTVDAVKQIEEILA*
54_1            AFDKVIVTVDAVKQIEEILA*
188_1           AFDKVIVTVDAVKQIEEILA*
601_1           AFDKVIVTVDAVKQIEEILA*
368_1           AFDKVIVTIDAVKQIEEILA*
                ********:************
```

Figure 5.5 Protein alignment showing substitutions in L4.

Representative isolate 601 (MIC=16mg/L) for NTHi and representative isolate 54 (MIC=2mg/L) for *H. haemolyticus* display the common protein sequence observed in both species which are identical apart from a substitution of isoleucine instead of a threonine at position 82. Isolate 552 (MIC=2mg/L) is a *H. haemolyticus* containing a substitution of aspartic acid rather than glycine at position 116. Isolate 549 (MIC =4 mg/L) is a *H. haemolyticus* that contains a substitution of an isoleucine instead of valine at position 149. Two NTHi isolates contained substitutions not shared anywhere else, these were isolates 188 (MIC=3mg/L) and 368 (MIC= 3mg/L) with substitutions of aspartic acid to glycerine and valine instead of isoleucine at positions 139 and 189 respectively. Again no substitution was associated with a change in MIC to azithromycin.

Observations of substitutions of arginine for glycerine at positions 2058 and 2059 and cysteine to threonine at position 2116 of the 23s rRNA domain V have been associated with macrolide resistance in *H. influenzae*, *Neisseria gonorrhoeae* and *Mycoplasma pneumoniae* (Clark et al., 2002, Chisholm et al., 2010, Morozumi et al., 2005) GenBank sequence LN831035.1 (270311-273445) from NTHi NCTC 8134 was used to map against the sequencing data for the 104 isolates. A sequence for the 23s rRNA gene was present in all isolates which were then aligned and investigated for the previously reported substitutions associated with resistance in the 23s rRNA gene. On further investigation of the alignments no substitutions were observed at position #2058, #2059 or #2611, all isolates were positive for adenine at positions #2058 and #2059 and a cytosine at #2611 as previously reported (*E. coli* numbering) (Almofti et al., 2011, Alonso et al., 2005, Bartkus et al., 2003, Begovic et al., 2009, Chan et al., 2007, Chisholm et al., 2010). Alignment of the resulting sequences were used to construct a phylogenetic maximum likelihood tree in RAxML using the GTRCAT model for heterogeneity on CIPRES (https://www.phylo.org/)(Materials and Methods, section 2.3.8) and visualised in microreact as shown in figure 5.6 (viewable at **https://microreact.org/project/HyC5cj8xe**) . Similarly to L4 and L22 there was no correlation between resistance and allele variation of the 23s rRNA gene. Instead the isolates were seen to fall into clades specific to their ST. Variation in MIC between isolates with the same ST were seen despite being in the same clade.

Further phylogenetic investigation into the L4 and L22 sequence variation was carried out to determine whether variation in sequences were associated with MIC levels for resistance levels within the isolates.

L22 variation separates NTHi from *H. haemolyticus* rather than clustering those of both species with the same level of resistance this can be seen from the maximum likelihood tree based on the alignment of L22 sequences (Figure 5.2). Isolates were labelled as intermediate or resistant depending on their MIC and nodes coloured by ST. For the NTHi isolates no ST was found in more than one cluster. Of the isolates in the largest cluster of NTHi approximately half (53.5%) of 43 isolates displayed MIC levels (>4mg/L) that are indicative of resistance to azithromycin despite the high degree of sequence identity of the L22 sequence (Figure 5.2). Five isolates that displayed high levels of resistance all clustered together on the same branch and were all designated ST 176 in Chapter 3. These were however all from the same patient and therefore could be a consistently carried strain rather than a characteristic linked to a novel mechanism of L22 mediated resistance.

The clade containing isolates of ST 353, 356 and 1314; the atypical NTHi group identified in chapter 3, was predominantly found to be of intermediate resistance to azithromycin. This was in addition to ST 513 and one isolate of ST 201. This clade was phylogenetically closer to *H. haemolyticus* sequences than the majority of the remaining NTHi. Branch lengths indicate more variation within NTHi than in *H. haemolyticus* although this could be attributed to smaller sample size for the latter species (Figure 5.2). The *H. haemolyticus* isolates cluster into group by patient which also gives the appearance of clustering by resistance levels. However, it is more likely that the isolates are clustering together because they are the same strain of *H. haemolyticus* from the same patient rather than due to an L22 resistance mediated mutation. The visualisation of this phylogeny can be accessed and further interrogated online at https://microreact.org/project/r1vwRWkee.

L4 sequences from 102 isolates and 2 references (NTHi NCTC 4842 and *H. haemolyticus* NCTC 10659) were aligned and a maximum likelihood tree constructed using RAxML with the GTRCAT model for heterogeneity on CIPRES (https://www.phylo.org/)(Materials and Methods, section 2.3.8). Similar to the observations made for the aligned L22 sequences, the isolates clustered by ST rather than MIC. Similarly, no ST was seen in more than one branch. Isolates from different patients appeared in different branches indicating that sequence variation in L4 is more related to ST type than resistance level or MIC (Figure5.3). The visualisation of this phylogeny can be accessed and further interrogated online at **https://microreact.org/project/rJatxIbex**

In conclusion no variation in L4, L22 or 23s rRNA genes were observed to be associated with level of MIC within this study.

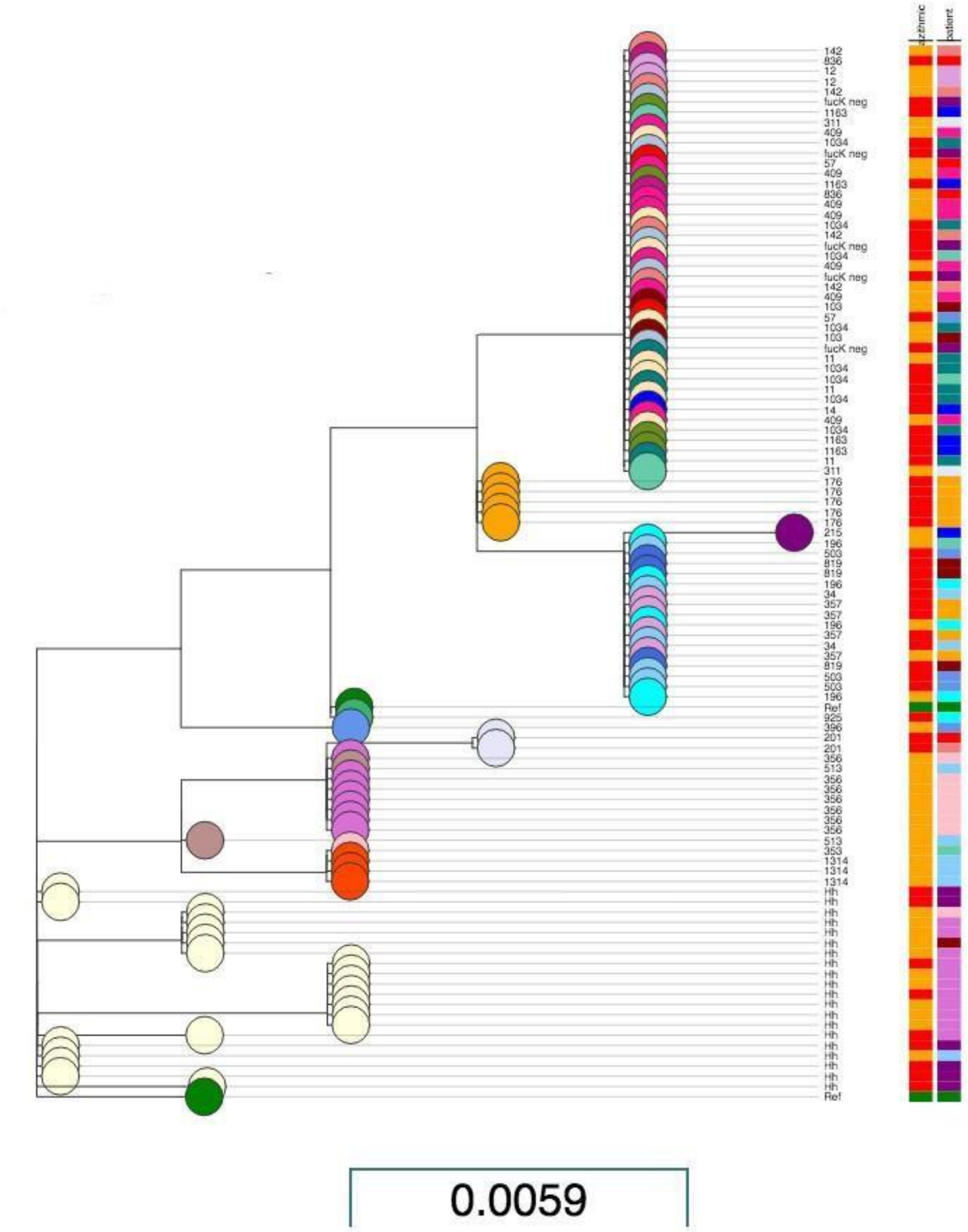


Figure 5.2 Phylogeny of L22 sequences compared to ST, MIC to azithromycin, and patient of origin.

Maximum likelihood tree created in RAxML and visualised in microreact. (https://microreact.org/project/r1vwRWkee) from alignments of 102 NTHi and Hh isolates and two NCTC references. The coloured nodes depict ST (along with label). The orange and red refer to the intermediate (MIC 0.125 mg/L – 4mg /L) and resistant (MIC >4mg /L) levels of MIC identified by E-test, respectively. The final metadata column is that of patient of origin - each colour depicts a different patient. No ST appears in more than one branch of L22 variation. Resistance levels within these isolates do not group unequivocally by L22 variation. Clustering of resistance in small incidences may be likely to same patient of origin. References included were NTHi NCTC 4842 and *H. haemolyticus* NCTC 10659.

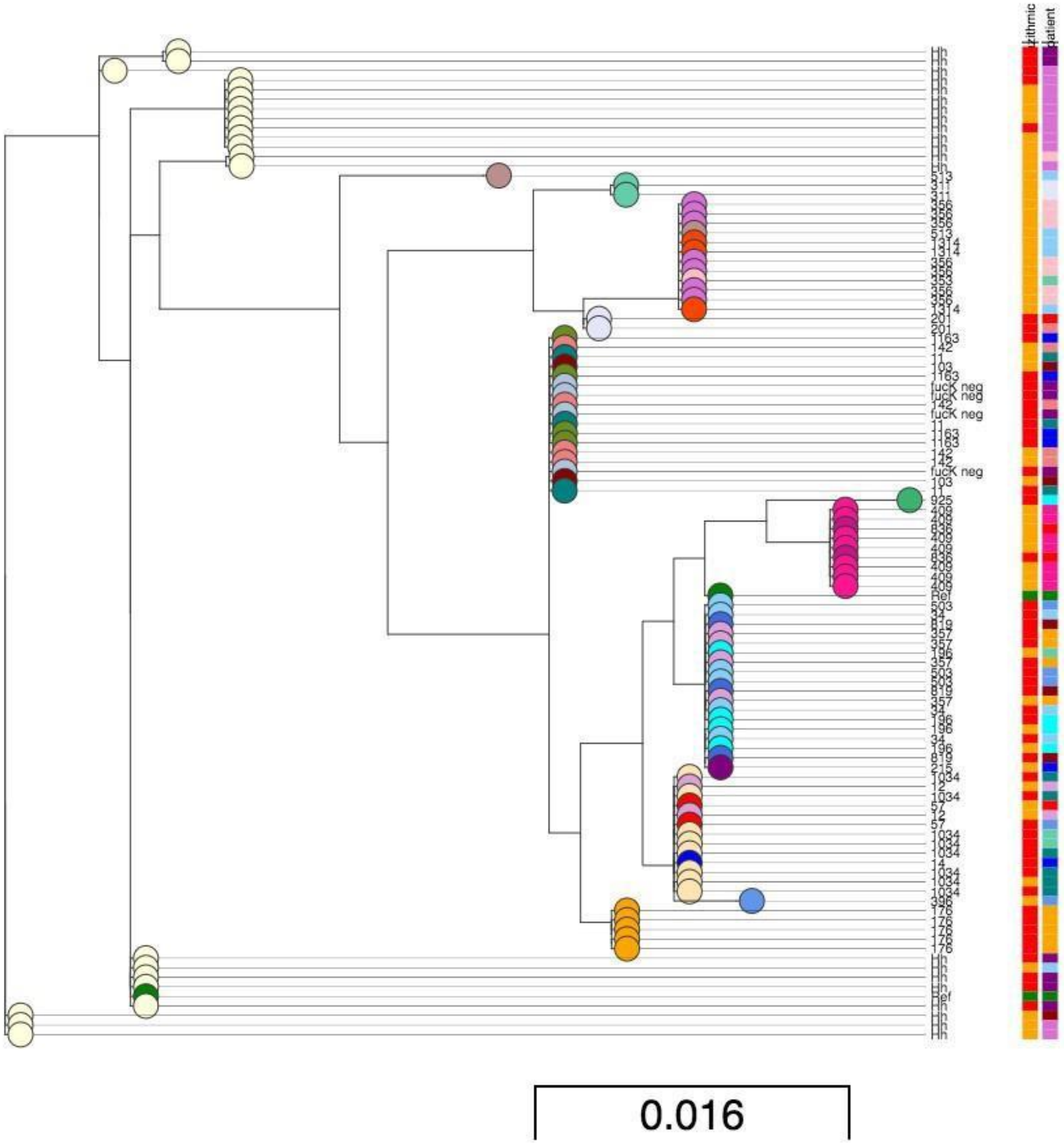


Figure 5.3 Phylogeny of L4 variations compared to ST, MIC to Azithromycin and patient.

Maximum likelihood tree created in RAxML and visualised in microreact. (https://microreact.org/project/rJatxIbex) from alignments of 102 study isolates and two NCTC references. The coloured nodes depict ST (along with label). The orange and red in the bar refer to the intermediate (MIC 0.125 mg/L 4 mg/L) and resistant (MIC >4mg/L) levels of MIC identified by E-test, respectively. The patient bar shows each patient of origin as a different colour. The light yellow nodes at the top of the tree are *H. haemolyticus,* the green nodes are NCTC 10659 and NC-4842 reference strains clustered with *H. haemolyticus* and NTHi respectively. The L4 sequences for the isolates do not cluster by resistance but do cluster within groups of the same STs. No ST appears on more than one branch.

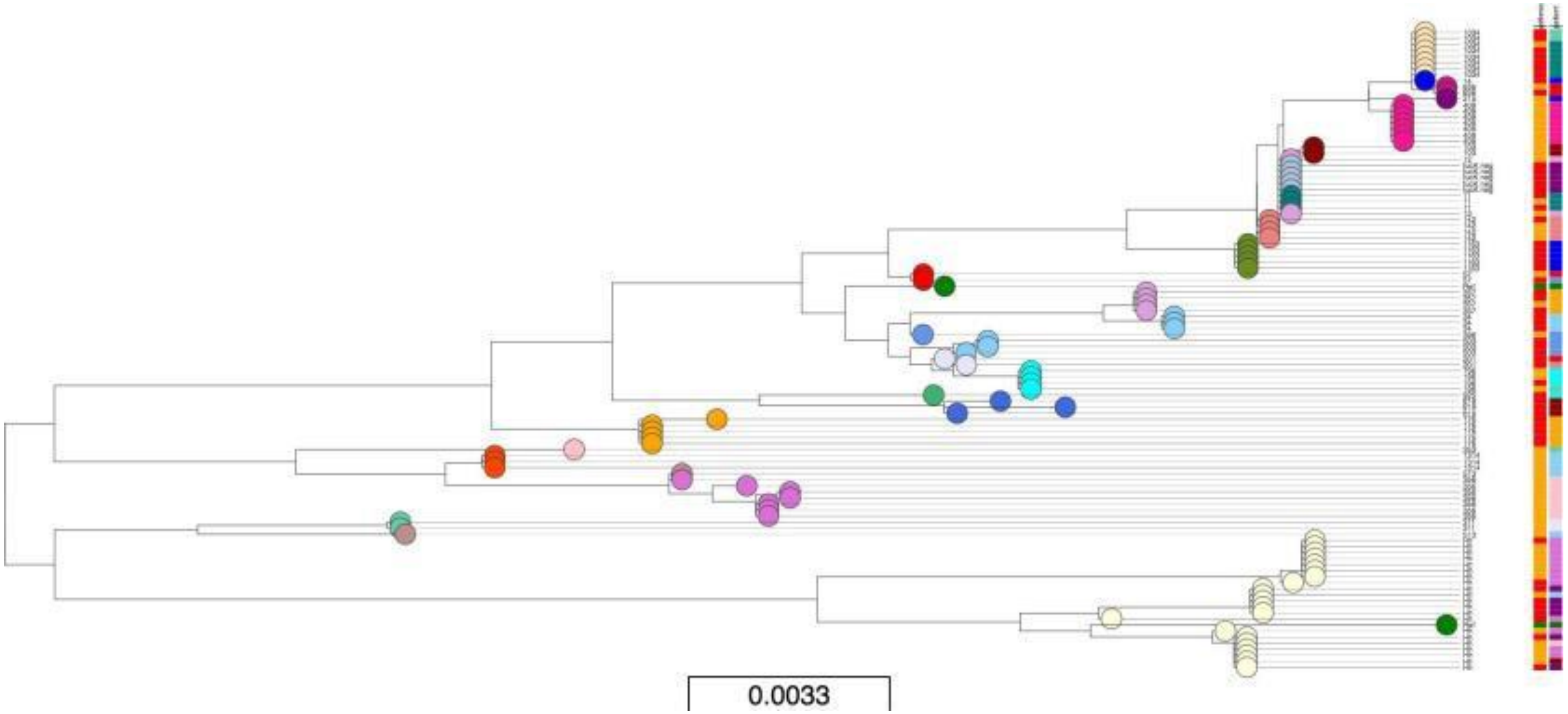


Figure 5.6 Allele variation of 23s rRNA Gene for 104 isolates compared to azithromycin MIC, patient and ST.

Maximum likelihood tree created in RAxML from MUSCLE alignment. Scale is 0.0033 substitution per bp. (Tree can be viewed here https://microreact.org/project/HyC5cj8xe along with metadata for each variable). Node colours are attributed to the different ST types, the light yellow nodes are *H. haemolyticus*. Resistance (MIC >4mg/L) is indicated by red and intermediate (MIC 0.125 mg/L - 4 mg/L) resistance indicated by orange. The different colours on the outer metdata bar depict different patients. Predominantly the isolates tend to cluster within ST/patient of origin rather than resistance status. Isolates with different STs but same resistance status are found in the same patient. Occurrences of the same ST having different resistance levels and isolated from different patients are also apparent. References for NTHi (NCTC 4842) and *H. haemolyticus* (NCTC 10659) were also include

5.4.4 Identification of Genes Associated with Resistance to Azithromycin

No gene previously associated with resistance to macrolide therapy was found to be associated with resistance in the isolates from this study. Consequently, further investigation was carried out to identify other loci that may be associated.

Annotated genomes for the isolates were assembled into two pan genomes for ten NTHi and ten *H. haemolyticus* isolates using ROARY (Materials and Methods, section 2.3.9). For each species, the ten isolates were made up from the five isolates with the highest MIC and the five isolates with the lowest MIC to enable an optimum comparison from the isolates available. The isolates were assigned into one of two trait groups, low MIC (NTHi 2mg/L, Hh 1-3mg/L) and high MIC (NTHi 16-24 mg/L, Hh 8-12mg/L) and a metadata csv file created consisting of trait type and isolate number. Where possible isolates were chosen from different patients at different time points (Appendix X). The resulting pan genomes were run through Scoary to ascertain statistical significance of genes annotated within the pan genome to the MIC trait defined within the metadata file. (Materials and Methods, section 2.3.9). For NTHi there were 1344 genes that were identified as statistically significantly associated with the MIC trait ($p< 0.005$). For *H. haemolyticus* there were 87 genes identified as statistically significantly associated with MIC trait. Twenty seven genes were found to be present in both. From this list seven genes were identified in both species with high specificity and sensitivity, a measure of the potential predictive power of the significant genes in identifying isolates with high MICs (Table 5.1).

Two of the seven genes, *purL* and *purH*, encode for phosphoribosylformyl-glycineamide synthetase and phosphoribosylaminoimidazolecarboxamide formyltransferase proteins that are involved in the de novo synthesis of purine. The gene *rluA* encodes for pseudouridine synthase A within the large ribosomal subunit. The genes *galK*, *mazG*, *focA* encode the proteins galactokinase and nucleoside triphosphate pyrophosphohydrolase and formate transporter respectively. A putative metallo-hydrolase was also identified in both species associated with high MIC level.

Genes *rlmM* and *rlmC* with 100% sensitivity and 80% specificity both encode for 23s rRNA methyltransferases, these proteins are responsible for methylation of antibiotic binding sites and associated with resistance (Vester, 2000). Other significant genes have also been previously associated with resistance, the dam gene encodes for DNA adenine methylase that has been reported to have a role in virulence but also regulating the expression of other genes, including multi efflux pumps associated with resistance (Heithoff et al., 1999, Adam et al., 2008). A resistance mechanism to mupirocin is associated with isoleucyl-tRNA synthetase encoded by gene *ileS* furthermore *emrA* encodes for an efflux pump which increases antibiotic resistance on overexpression due to mutation within its negative regulator *ermR* (Lomovskaya et al., 1995, Serafini et al., 2011). Glycerol phosphate encoded for by *glpG* has been shown to increase resistance due to mutations (Clemmer et al., 2006). Currently *dsb_D2*, the gene that encodes for the thiol disulphide interchange protein involved in the folding of secreted proteins in the periplasmic space, is being investigated for a possible link to resistance in *N. meningitidis* and *P. aeruginosa* (http://gtr.rcuk.ac.uk/projects?ref=MR%2FM009505%2F1) (Nagy, 2013).

No remaining significant gene from Table 5.1 has an association with resistance reported, *dnaJ* is a heat shock protein and *msbA* is also involved in maintaining viability of bacteria with temperature change (Caplan et al., 1993, Polissi and Georgopoulos, 1996). *bamA* encodes for a protein that assists in the assemblage of ABC transporter proteins within the outer membrane, furthermore *pot_A* gene encodes for PotA protein which enables energy dependent spermidine uptake in intact cells necessary for growth (Kashiwagi et al., 1995). Additionally, ung encodes for uracil-N- glycosylase which is involved in DNA repair (Ieva and Bernstein, 2009, Kashiwagi et al., 1995, Johnson et al., 2007) and *cpdA* encodes cAMP phosphodiesterase which catalyses the cAMP to AMP conversion. (Kim et al., 2009a). *tabA* is associated with a toxin-antitoxin biofilm protein that reduces biofilm production on deletion of five other toxin-antitoxin systems and *guaA* encodes for a GMP synthase which is involved in purine salvaging pathways (Jewett et al., 2009, Kim et al., 2009b). *glnS* Encodes for glutamine –tRNA ligase which is involved with amino acid biosynthesis and transportation (Hoben et al., 1982, Yamao et al., 1982). *metC* is also involved with biosynthesis of amino acid methionine from cysteine (Ferla and Patrick, 2014).

Table 5.1 Genes significantly ($p < 0.005$) associated with high resistance found in both NTHi and *H. haemolyticus.* Genes highlighted in blue have both high sensitivity/specificity in both species.

Gene name	Hh sensitivity	Hh specificity	NTHi sensitivity	NTHi specificity
ribonuclease I	100	80	0	0
bamA	100	80	0	20
metC	100	80	0	0
rlmC	100	80	0	0
ssb_1	100	80	0	0
dsbD_2	100	80	0	0
rlmM	100	80	0	0
mazG	100	80	80	100
guaA	100	80	0	0
glnS	100	80	0	0
potA_1	100	80	0	20
ileS	100	80	0	0
emrA	100	80	0	0
purl	100	80	100	100
rluA_1	100	80	100	100
dnaJ	100	80	0	0
galK	100	80	80	100
ung	100	80	0	0
putative metallo-hydrolase	100	80	100	100
dam	100	80	0	0
murein-L putative	100	80	0	0
focA	100	80	100	100
purH	100	80	100	100
msbA	100	80	0	0
glpG	0	20	80	100
cpdA	0	20	80	100
tabA	0	20	80	100

5.5 Discussion

To the authors best knowledge this is the first study into azithromycin resistance in both *H. haemolyticus* and NTHi isolated from the COPD lung. No isolate was found to be susceptible for azithromycin from the 104 tested (Figure 5.1). More than half of the NTHi (53%) were found to have intermediate resistance with MICs 1-4 mg/L and the remaining isolates displaying MICs of 6-24mg/L (Figure 5.1). This is high in comparison to Pettigrew *et al*. (2015) whereby all newly acquired strains of NTHi isolated from sputa from COPD patients were found to be susceptible to azithromycin. Although five of 27 susceptible strains they identified were seen to develop resistance after exposure to azithromycin (Pettigrew et al., 2016). The *H. haemolyticus* isolates from our study also displayed high levels of resistance at 43% with an MIC of above 4mg/L (Figure 5.1).

Previous studies have highlighted a number of resistance mechanisms including the presence of the *erm* and *mef* genes. However, no acquired macrolide resistance genes were identified in any of the isolates tested here. This finding is in agreement with the findings of Atkinson *et al.* (2015) and Clark *et al.* (2002) but contrasts that of Roberts *et al.* (2011) who were able to identify an AMRG in all isolates by PCR (Atkinson et al., 2015, Clark et al., 2002, Roberts et al., 2011). Variations in L4, L22 and 23s rRNA genes were all investigated for previously reported mutations, however these too were absent in the isolates from this study (Figures 5.4 and 5.5). Furthermore, when phylogenetically analysed it was observed that the variations in sequence for these genes did not correlate with MIC indicating an absence of effect of these genes on the azithromycin resistance observed and a non-clonal phenotype of resistance, demonstrated by MIC not clustering by ST (Figures 5.2, 5.3, 5.6, 5.7).

In silico analysis of the genome of ten isolates from both species returned seven genes that were statistically significantly ($p<0.005$) associated with MIC trait and which had high specificity and sensitivity in both species. Five isolates were chosen with the highest MIC's within the subset of isolates for each species and five with the lowest MIC. Where possible these were taken from different time points and patients. These genes could potentially identify those isolates with high resistance levels to azithromycin. Two of these seven genes, *purL* and *purH* are important for de novo purine synthesis which has interestingly been indicated in resistance to rifampicin in *S. aureus* (Yee et al., 2015). There is also an association with vancomycin resistance in *S. aureus* however this has since been questioned (Mongodin et al., 2003, Fox et al., 2007). Another gene of significance, *rluA* is responsible for pseudouridine synthase A important for post transcriptional modification in the 23s rRNA, the target for macrolide intervention (Hamma and Ferre-D'Amare, 2006, Tait-Kamradt et al., 2000a). The remaining genes *galK*, *mazG* and *focA* which express galactokinase, nucleoside triphosphate pyrophosphohydrolase and a formate transporter have no immediate associations with potential mechanisms or previous associations with antibiotic resistance (Wang et al., 2009, Debouck et al., 1985, Zhang and Inouye, 2002). One last gene putative for a metallo-hydrolase in both species was identified which have been previously reported to have a role as beta lactamases and resistance to penicillin (Daiyasu et al., 2001). The absence of L4 and L22 genes from this analysis also supports the phylogenetic analysis of these isolates.

Interest in antimicrobial resistance in *H. haemolyticus* has historically been low, presumably due to its limited pathogenic capability. Recently however the first strains of *H. haemolyticus* resistant to ciprofloxacin have been reported (Marti et al.,2016). Recombination between *H. haemolyticus* and NTHi involving genes responsible for resistance has been observed previously therefore there is hypothetically a potential for this mechanism to pass on azithromycin resistance and for *H. haemolyticus* to act as a pool of resistance genes to aid the capacity for antimicrobial resistance in the more pathogenic NTHi (Witherden et al., 2014). The identification of genes statistically significant to the MIC of the isolates shared between the two species may indicate a strong starting point for further investigation into the resistance levels to azithromycin, especially those who have been previously indicated in mechanisms for resistance in other bacteria or to other antibiotics, or those that effect 23s rRNA which houses the macrolide binding site. As of yet, however, the resistance mechanism for the isolates from this study cannot be identified.

5.6 Conclusions and Future Work

The levels of resistance observed within the isolates were unexpectedly high compared with prior limited research into macrolide resistance within NTHi and *H. haemolyticus*. These isolates represent 16 patients only and further investigation into treatment regimes may uncover a history of azithromycin therapy resulting in accrued levels of exposure as previous studies have displayed that exposure to azithromycin is the driving factor for acquisition of resistance. The treatment regimes for the patients were not accessible for the purposes of this PhD (Bergman et al., 2006, Phaff et al., 2006).

Although this is an *in silico* investigation based on statistical analysis it does identify genes that could potentially be associated with the MIC trait and form a good starting point for further investigation into resistance within these isolates. The next step would be to employ methods that are able to link phenotype with genotype. Transposon mutagenesis sequencing methods such as TraDIS or HITS have been historically used to identify virulence factors in bacteria and would facilitate the requirement to link genotype with the observed resistance phenotype (Mei et al., 1997, Langridge et al., 2009, Barquist et al., 2016, Autret et al., 2001, Gawronski et al., 2009).

6 Whole Genome Analysis for Speciation Between NTHi and *H. haemolyticus*

6.1 Introduction

Due to the failure of molecular markers to unambiguously differentiate between NTHi and *H. haemolyticus* it has been hypothesised that the two species may share sufficient genetic content to represent a genomic continuum rather than distinct species separation (Binks et al., 2012). Despite the reported close genetic relationship between the two species *H. haemolyticus* nevertheless lacks pathogenic behaviour and is able to cause a cytotoxic response in epithelial cells in comparison to the invasive nature of NTHi (Pickering et al., 2016). Similarly, *E.coli* and Shigella are different in pathogenic ability and are closely genetically related. These two can be clearly differentiated using biochemical and physiological tests however the differentiation can be clouded by disease causing intermediate strains of enteroinvasive *E.coli* (EIEC). The ability to differentiate between EIEC and *E.coli* comes with the PCR assay for marker gene *ipaH* present in all Shigella and EIEC strains but absent from non dysentery causing *E.coli* (van den Beld and Reubsa et, 2012). In chapters 3 and 4 heterogeneity in NTHi and atypical NTHi was demonstrated by using limited molecular marker data. The numbers of molecular markers used were few in comparison to the number of genes present in the NTHi genome and therefore in this chapter we explore the comparison of the whole genomes and turn to methods of taxonomy.

6.2 Hypothesis

We hypothesise that whole genome analysis will enable clear speciation between NTHi and *H. haemolyticus* despite the atypical genotypes of some isolates.

6.3 Aims

The aim of this chapter was to use whole genome sequencing data to investigate the genetic relatedness of *H. haemolyticus* and NTHi. This will:-

- Establish speciation boundaries between NTHi, *H. haemolyticus* and atypical isolates using average nucleotide identity (ANI) analysis
- Determine the extent of the shared core and pan genome between the two species

6.4 Results

In Chapter 3, 1460 isolates were categorised into NTHi (n=1241), atypical NTHi (n=76), *fucK* negative NTHi (n=51) and *H. haemolyticus* (n=92). These were isolated from sputa collected from COPD patients on monthly appointments or on an occurrence of acute exacerbation in a two-year longitudinal study (Bourne et al., 2014). The classification of the isolates into these four groups was based on gene marker information including MLST analysis as detailed in Chapter 3.

Whole genome sequence (WGS) data from the isolates was assembled and interrogated by MetaPhlAn (Materials and Methods, section 2.4.1). Five out of 1460 isolates were not identified to species level but were denoted as 97.28%-99.92% similar to *Pasteurellaceae*. The remaining 1455 returned varying levels of percentage identity to the MetaPhlAn reference *H. influenzae* marker genes (70.94%-100%) with the remaining percentage of the genome being identified as *Haemophilus* genus but unable to specify to species level. Only 68 of the study isolates were 100% identical to the 260 taxonomic marker genes of *H. influenzae* within the MetaPhlAn reference database. The *H. haemolyticus* isolates returned an overall lower percentage identity for these marker genes, ranging from 70.94%-82.09% (Figure 6.1). The atypical NTHi isolates were situated in between *H. haemolyticus* and NTHi with identity matches ranging from 82.94%-89.97%. NTHi isolates and *fucK* negative NTHi ranged from 90.21%-100% and did not cluster separately, with the exception of one isolate of ST 513 with an identity of 86.5% (Figure 6.1). Both the atypical and *H. haemolyticus* groups show clear separation (Figure 6.1).

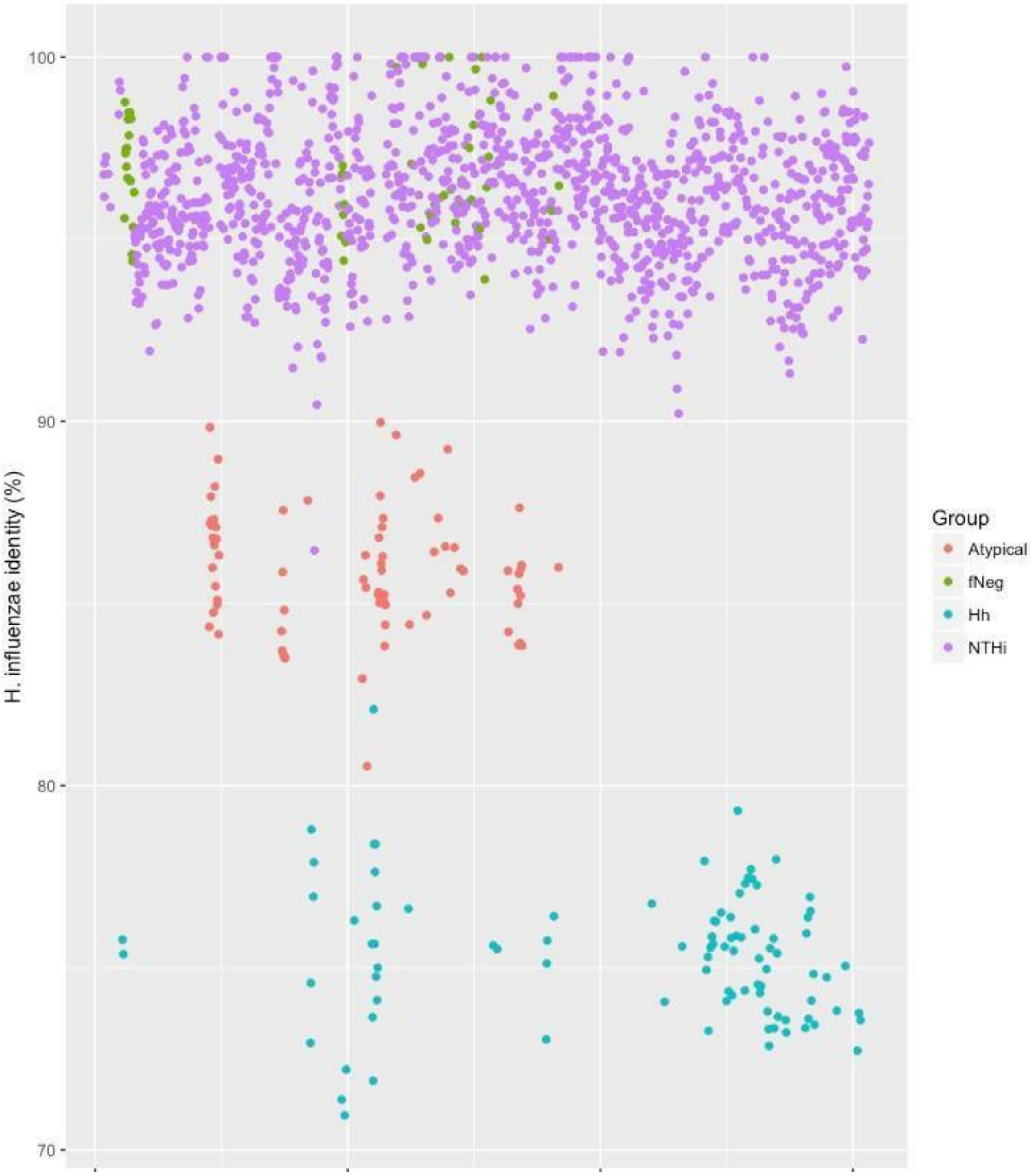


Figure 6.1 Dot plot showing percentage identity of assembled genomes of 1460 study *Haemophilus* isolates against 260 *H. influenzae* marker genes using MetaPhlAn. NTHi and *fucK* negative isolates returned 90.21%-100% identity except one isolate of ST 513 that returned 86.5%. *H. haemolyticus* isolates returned 70.94%-82.09% identity. The atypical isolates were observed to fall centrally with a range of 82.94%-89.97%.

6.4.1 Taxonomy Threshold Identification Using Average Nucleotide Identity

Initially ANI was calculated for *H. haemolyticus* and a subset of NTHi composed of a representative isolate from each ST (Materials and Methods, section 2.4.2). Reference genomes for *H. influenzae* NCTC 4842, 7279, 8467 and *H. haemolyticus* NCTC 10659 and NCTC 10839 were also included. An all-by-all comparison was done and the ANI was calculated for each pair of isolates. The *H. haemolyticus* reference strains were <92% similar to the reference NTHi and all >97% similar to each other. The NTHi references were also >97% similar to each other. ANI was calculated for all isolates (n=1460).

6.4.2 *H. haemolyticus*

Two *H. haemolyticus* reference isolates (NCTC 10839 and NCTC 10659) and 90 study isolates were >95%-100% genetically similar. The lowest identity between two *H. haemolyticus* isolates was 94.6%. This was seen in two isolates only and the remainder were >95% (Figure 6.2).

6.4.3 NTHi ST Representatives and *H. haemolyticus*

An isolate from each ST was selected for ANI comparison. The range of ANI calculations between the NTHi were 94.8%-100%. This included representatives of the *fucK* negative and atypical isolates. A representative of *H. haemolyticus* was also included andobserved to be 91.5%-91.9% similar to the NTHi isolates. The representatives for the atypical group ST356, ST353 and ST1314 were the least identical to the other ST representatives andwere just below 95% identity to other typical and *fucK* neg NTHi isolates (Figure 6.3). The *fucK* negative isolate displayed a large identity match to the larger group of NTHi (Figure 6.3). Representative isolates for ST 311, ST 704 and ST 513 clustered together and were seento be of a lower percentage identity to the majority of the isolates.

6.4.4 ANI of Complete Study Isolates

The ANI calculations for all 1460 *Haemophilus* spp. shown in Figure 6.4 supports the findings of the ANI calculations for the subset representing each ST from the NTHi group and one representative *H. haemolyticus* as shown in Figure 6.3. The STs that make up the atypical group of isolates, namely ST 353, 356 and 1314 were the group that had the lowest identity to the remaining NTHi. The *H. haemolyticus* were all classified as below threshold for inclusion as a same species to the NTHi and can be seen highlighted in blue in Figure 6.4.

6.4.5 Phylogenetic Analyses of Core Genome

The resulting core genome alignment from ROARY was used to create a maximum likelihood tree in RAxML using GTRCAT method on CIPRES (https://www.phylo.org/) (Materials and Methods, section 2.3.8). A metadata file was created linking ST with patient and isolate number. Both metadata file and Newick tree files were submitted to microreact for analysis (Materials and Methods, section 2.3.8). The *H. haemolyticus* in the resulting tree clearly cluster separately from the NTHi supporting the ANI findings. The atypical isolates can be seen to cluster separately at the edge of the NTHi clade also as marked in Figure 6.5. The *fucK* negative isolates cluster within the other NTHi supporting findings in the previous chapters that the *fucK* negative status of these isolates does not represent a large variation from typical NTHi (Figure 6.5). The visualisation of this phylogeny can be accessed and further interrogated online athttps://microreact.org/project/rJV586A5e.

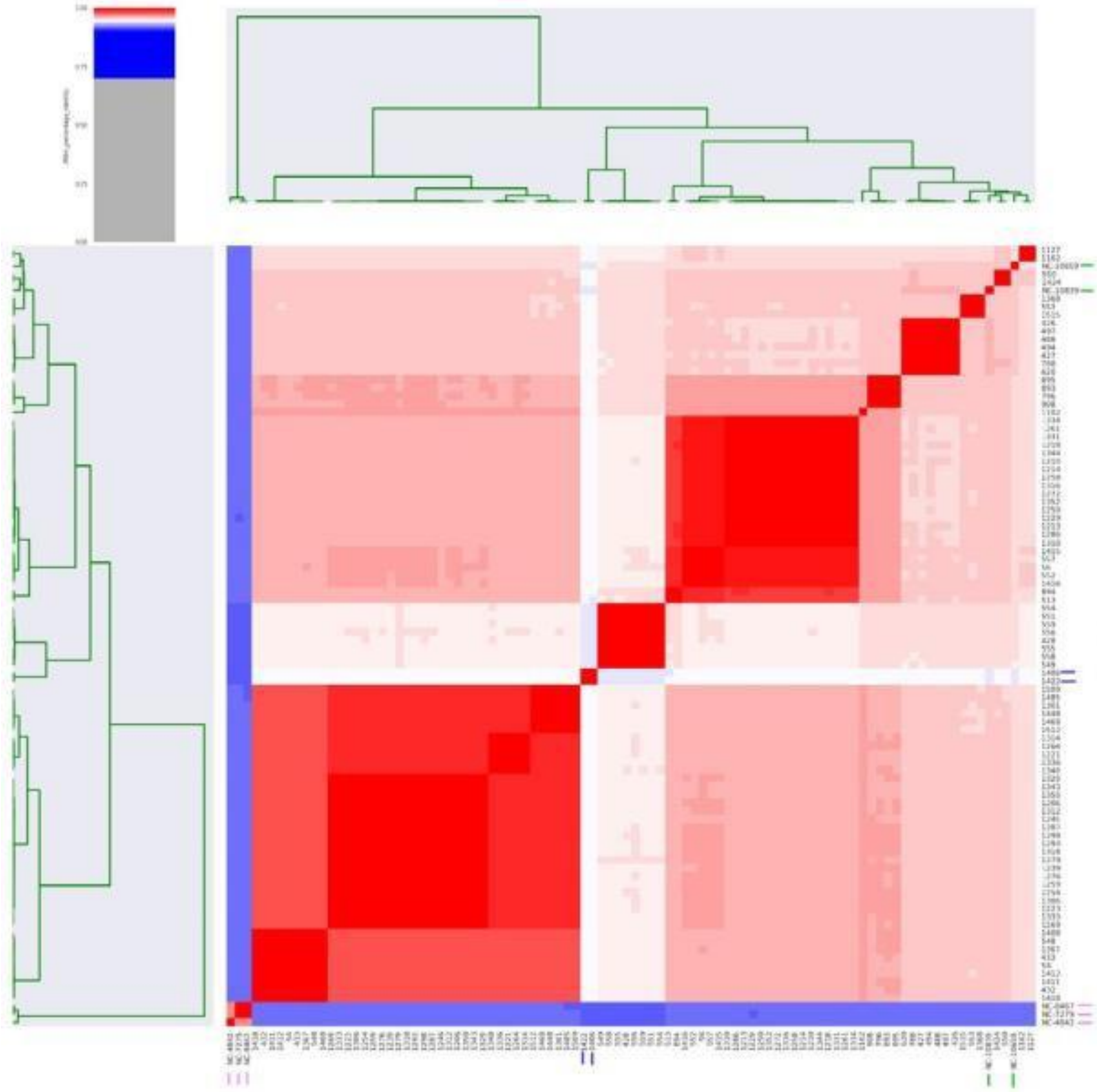


Figure 6.2 Heatmap of ANI calculations for 92 *H. haemolyticus* isolates.

H. influenzae NCTC strains marked in lilac, *H. haemolyticus* reference strains marked in green and two atypical *smpB* negative strains highlighted with blue. All *H. haemolyticus* were 94.6%-99.9% identical.

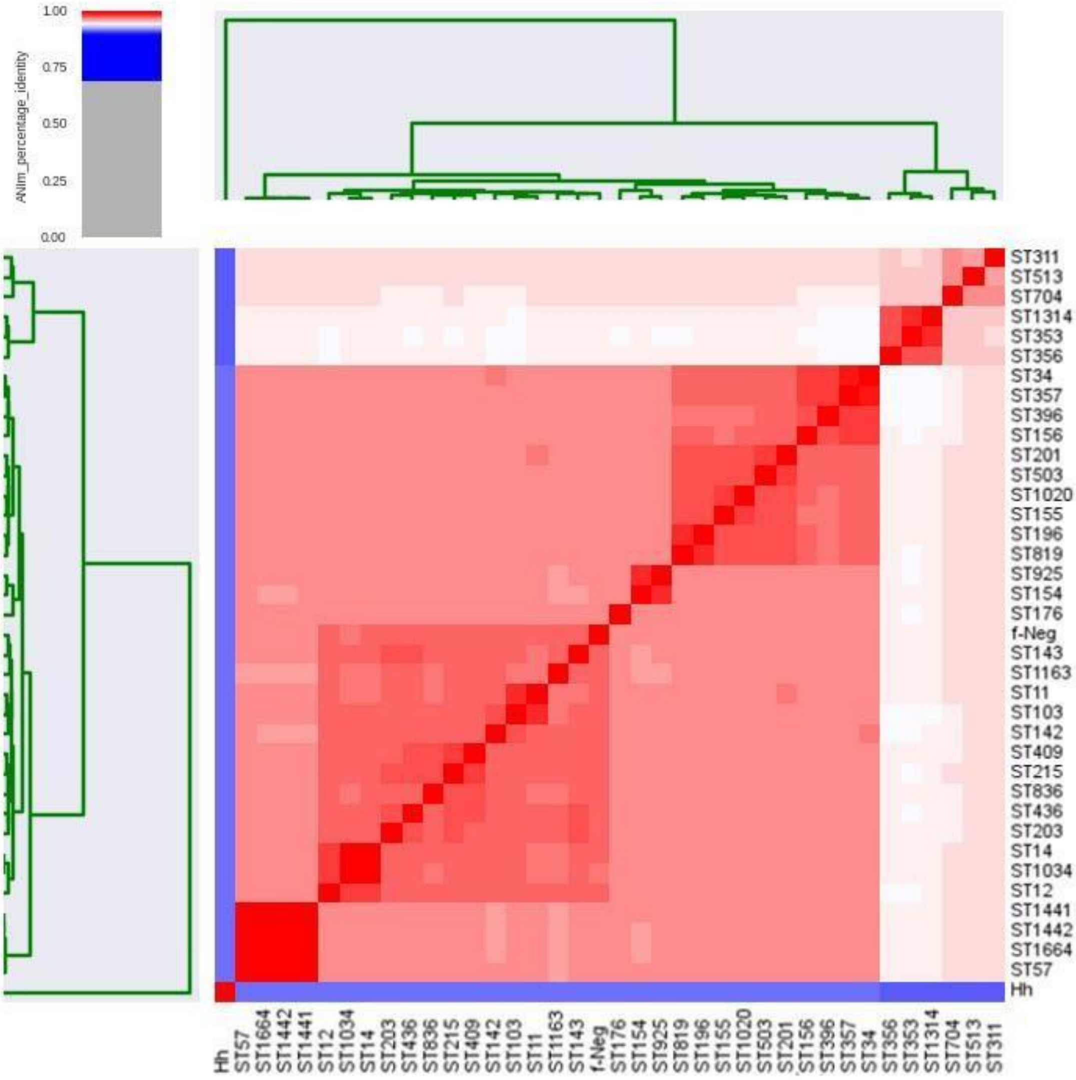


Figure 6.3 Heatmap of ANI calculations for representative isolate of *H. haemolyticus* and representative isolates of each ST from MLST analysis of NTHi.

All NTHi isolates were <94% identity. The percentage identity for *H. haemolyticus* compared to the representative NTHi isolates was <92% (shown in blue).

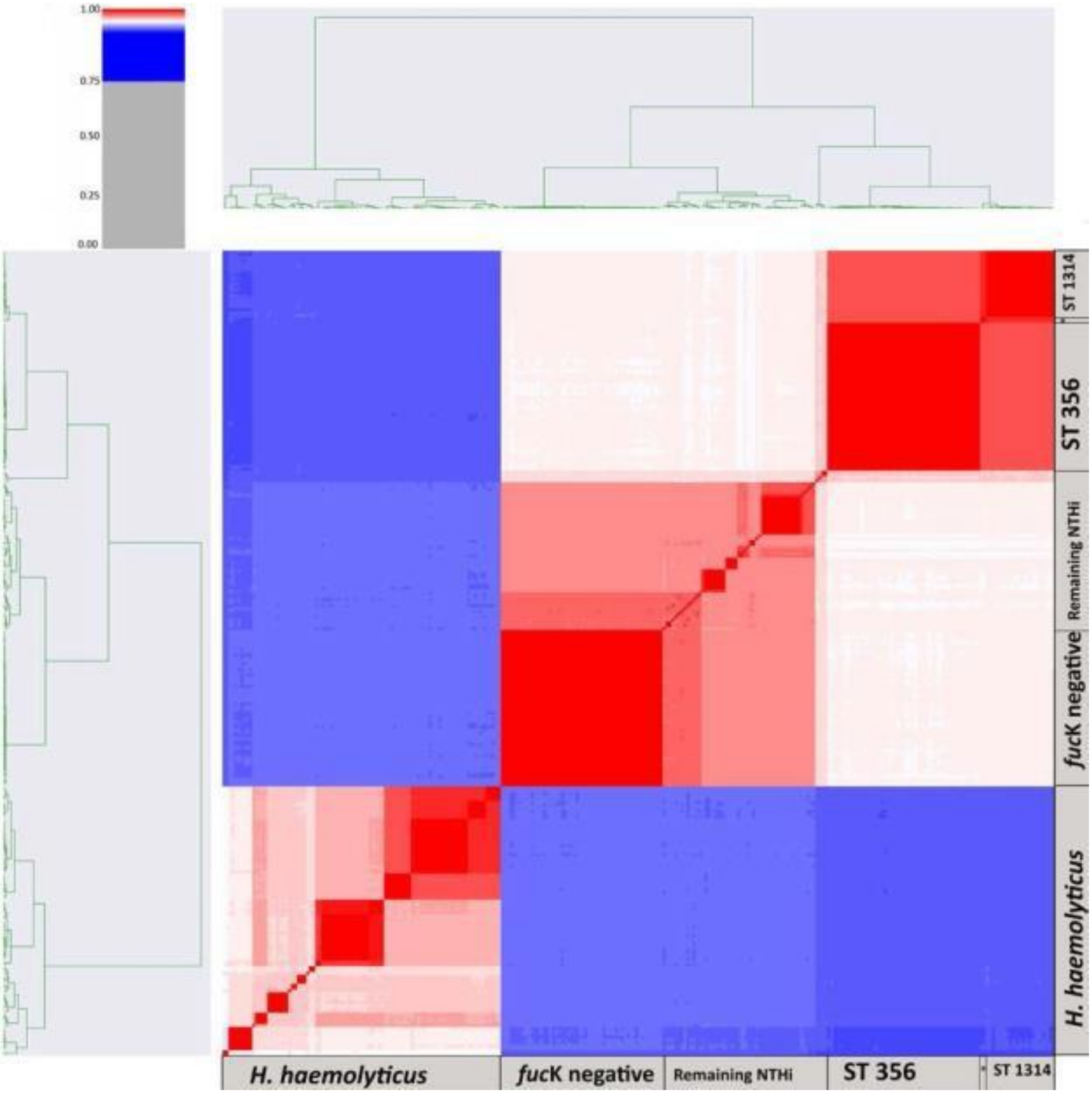


Figure 6.4 Heatmap of ANI calculations of 1460 *Haemophilus* spp.

The *H. haemolyticus* compared to the NTHi has percentage identity which is <92% and therefore shown in blue. Despite the heterogeneity displayed within the NTHi isolates, they are all observed to be classified as the same species (in red) using ANI.

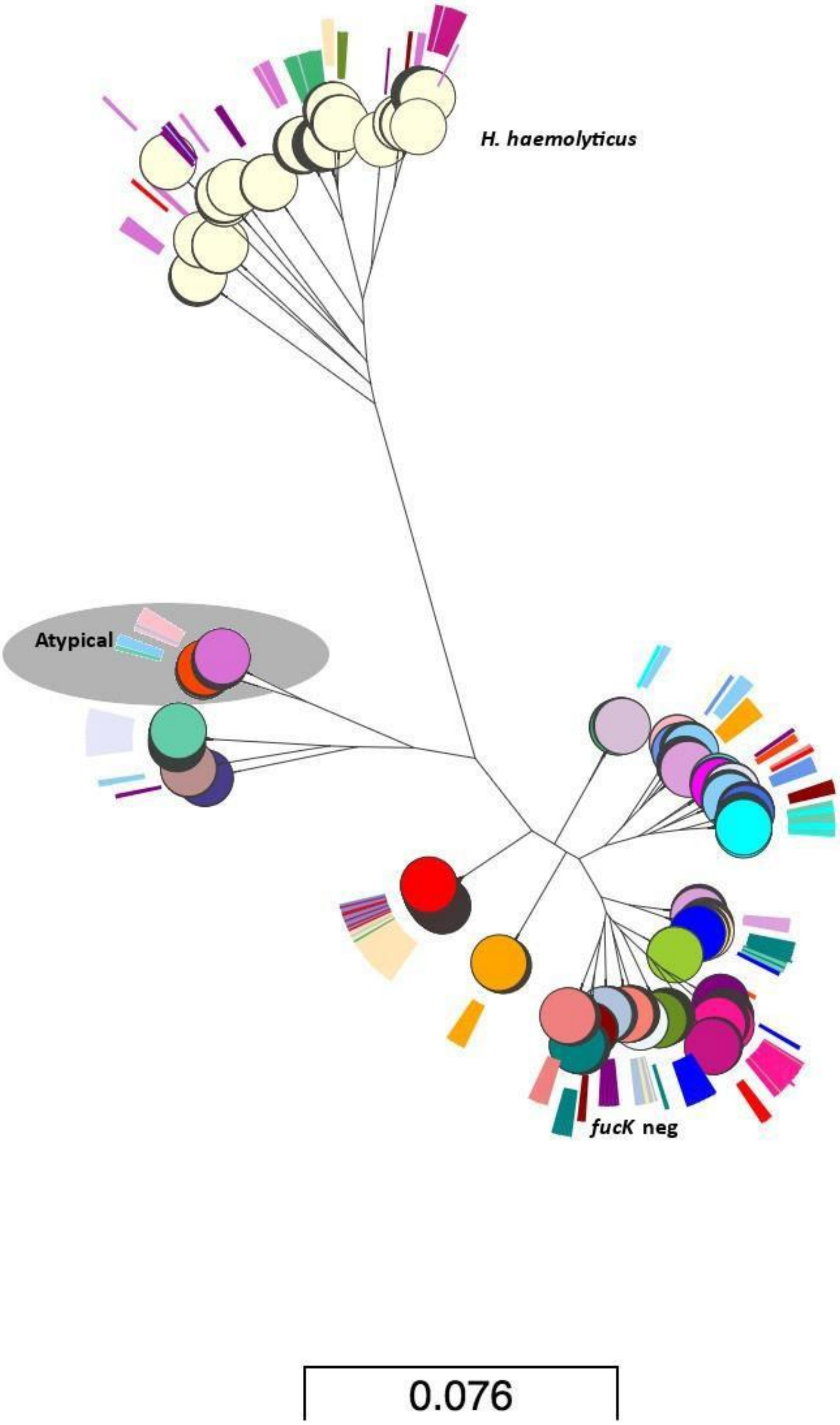


Figure 6.5 Core genome phylogeny between NTHi and *H. haemolyticus.*
Maximum likelihood phylogeny created using subset of 91 *H. haemolyticus* and 297 NTHi representing 37 STS at 134 time points from 24 patients. As per the ANI analysis the *H. haemolyticus* (light yellow nodes) can be seen to cluster separately from the NTHi (remaining coloured node, colours are representative of ST). The colours on the external bar are representative of the patient of isolation. Atypical, *H. haemolyticus* and location of *fucK* negative isolates are highlighted.

6.4.6 Shared Core Genome between NTHi and *H. haemolyticus*

A pan genome was generated including annotated genomes from *H. haemolyticus* using ROARY (Materials and Methods, section 2.3.9). In summary, a total of 7479 different genes were identified, 1064 of those were detected in 95-100% of all the isolates and defined as core genes giving a core/genome ratio of 14.2% (Figure 6.6). The distributions of the genes across isolates are shown in Figure 6.7.

Using NTHi isolates only, a total of 13869 genes were identified, of which only 573 designated as shared core summary core genes in 95-100% of the isolates resulting in a core/genome ratio of 4.13% (Figure 6.6, Figure 6.7).

The genomes from 91 *H. haemolyticus* and 297 NTHi were used to generate a joint pan genome for both species. The NTHi were selected by choosing two isolates from every time point of each patient. Where there was evidence of more than one ST in a time point, as determined by the MLST analysis in chapter 3, two isolates were taken for each ST for that time point where available. From a total of 19301 genes identified 324 were found to be shared as core genes in 95-100% isolates giving a core/genome ratio of 1.67% (Figure 6.6). Distribution of genes can be seen in Figure 6.7.

The summary core genes were further split into core, those shared in 99-100% of isolates, and soft core, those shared in 95-99% of isolates (Figure 6.6).

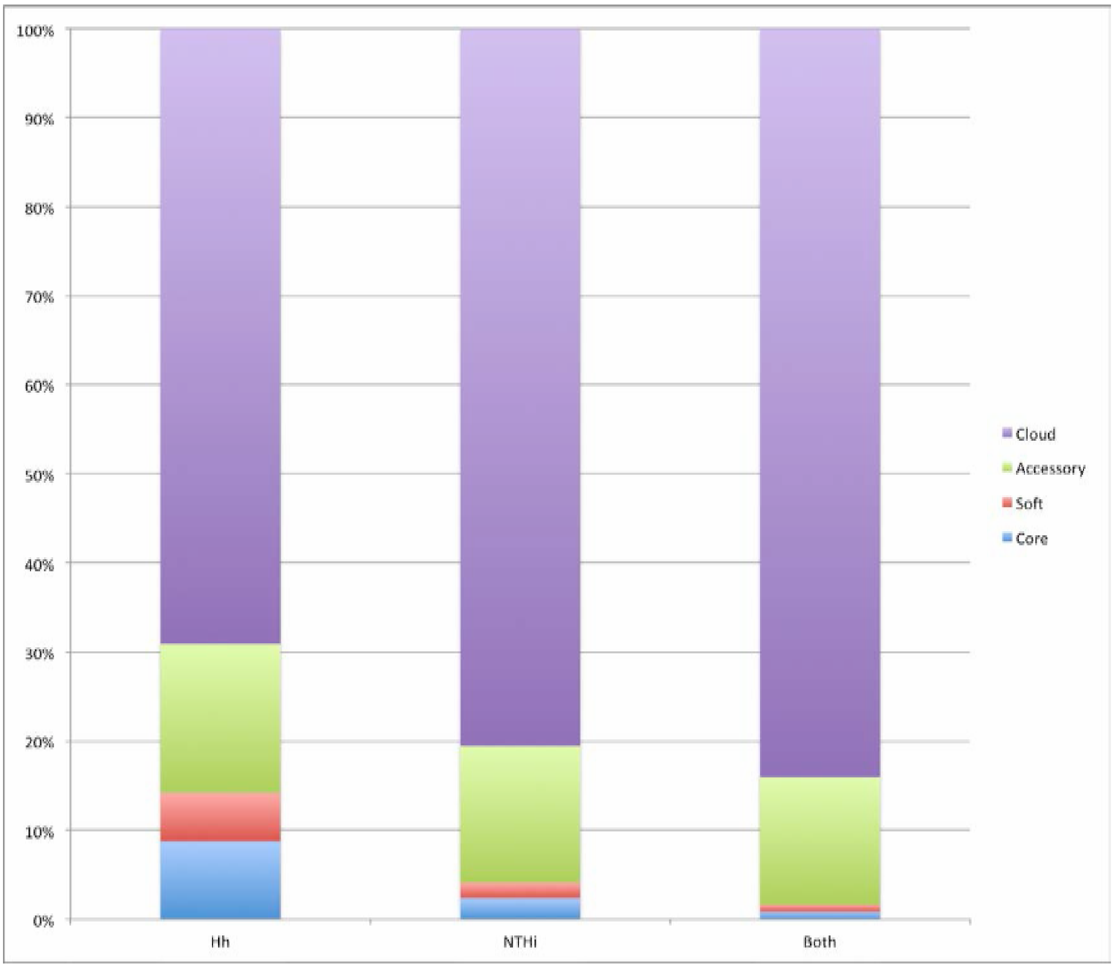


Figure 6.6 Composition of pan genomes of 91 *H. haemolyticus*, 297 NTHi and a subsetof both species.

Stacked bars represent the percentage of genes designated as core (detected in 99-100% of isolates -blue) soft core (95-99% - red) and accessory (15-95%- green)genes and accessory and cloud genes (<15% - purple).

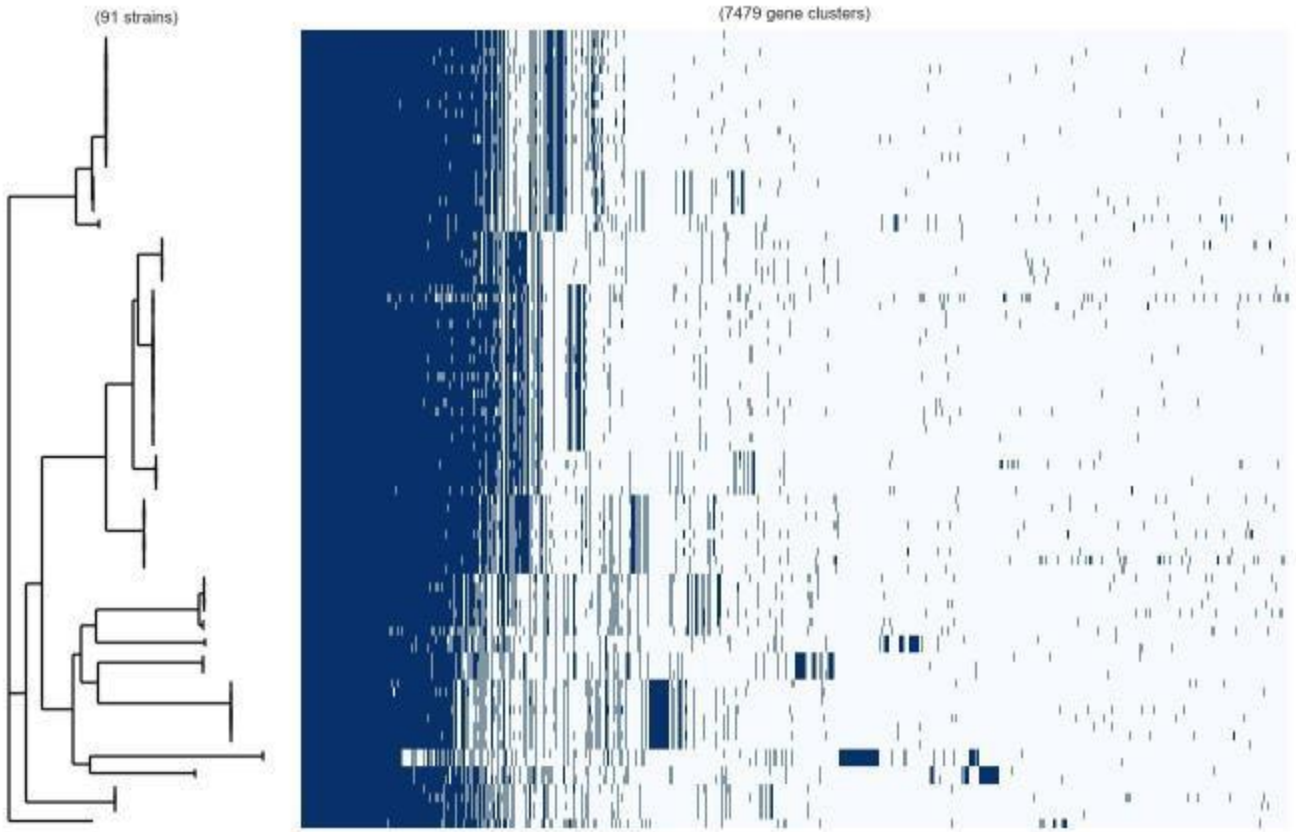


Figure 6.7 Heatmap visualising the distribution of the pan genome of 91 *H. haemolyticus* isolates visualising distribution of shared genes throughout isolates.

Solid dark blue vertical areas show the extent of the core (shared) genes.

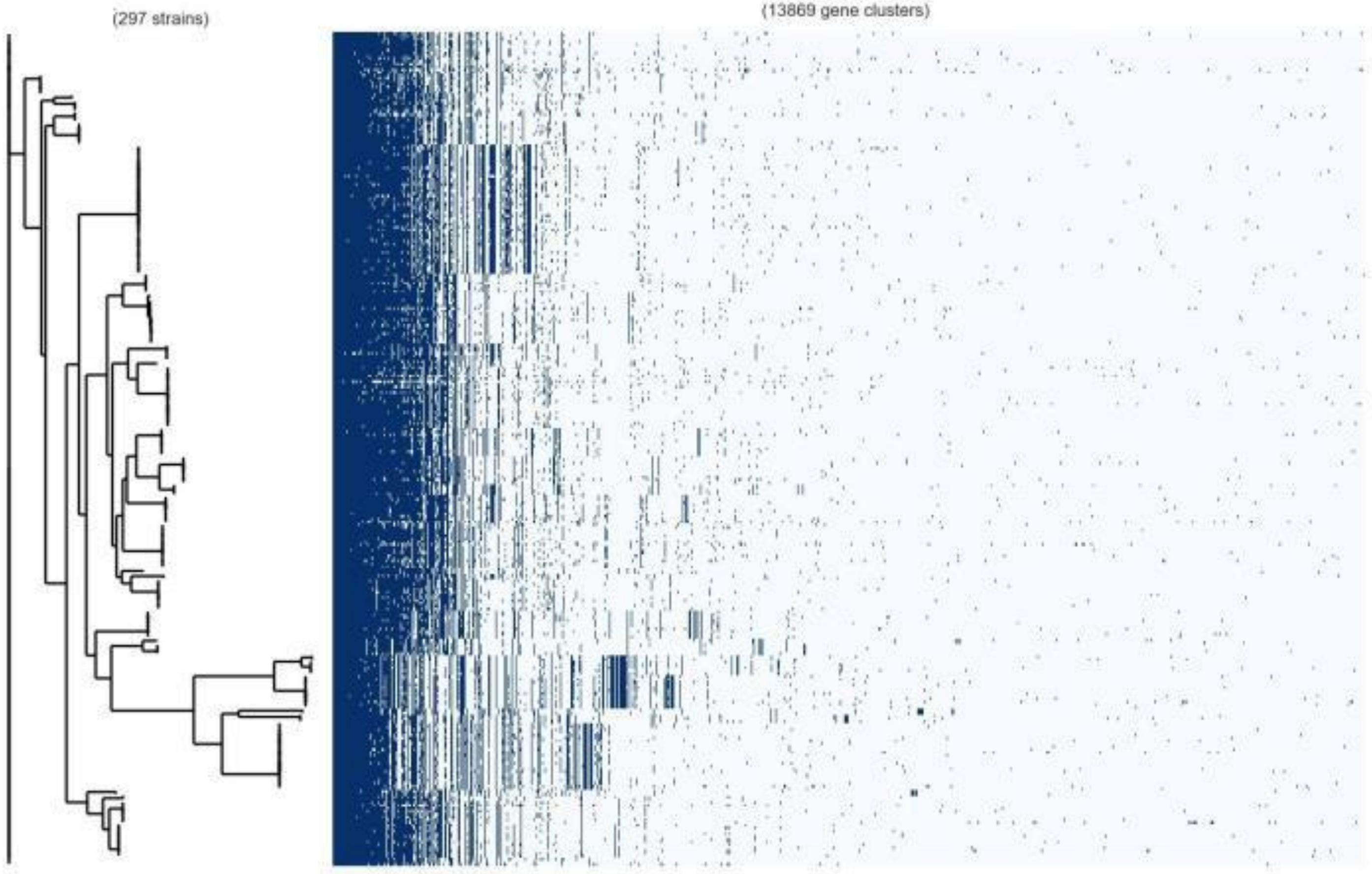


Figure 6.8 Heatmap visualising the distribution of the pan genome of 297 NTHi visualising distribution of shared genes throughout isolates. Solid dark blue vertical areas show the extent of the core (shared) genes.

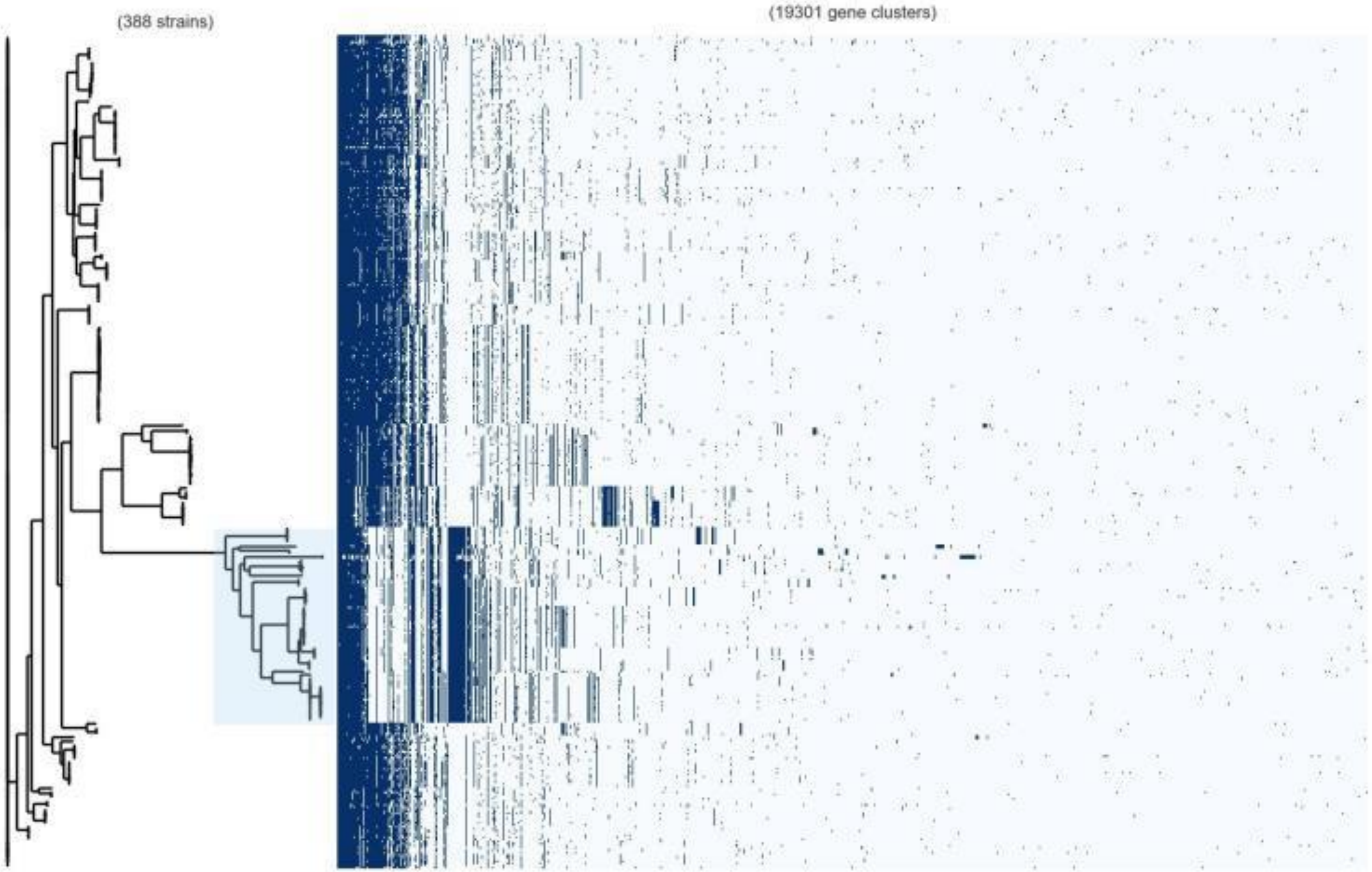


Figure 6.9 Heatmap visualising the distribution of the pan genome of 297 NTHi and 91 *H. haemolyticus* throughout sample set. Solid dark blue vertical areas show the extent of the core (shared) genes. The *H. haemolyticus* isolates are highlighted in light blue

The accumulation of total genes relative to isolates included in the analysis can be seen in Figures 6.10, 6.11 and 6.12. This shows an accumulative pan genome that does not achieve an asymptote. This indicates that this is an open pan genome with scope for further increase with the inclusion of additional strains.

Figures 6.13, 6.14 and 6.15 show the accumulation of unique genes into the pan genome. The plots show how many new genes have been found for each genome examined. The unique plot represents the cumulative number of new genes on first occurrence and is a sum of all new genes found in all genomes.

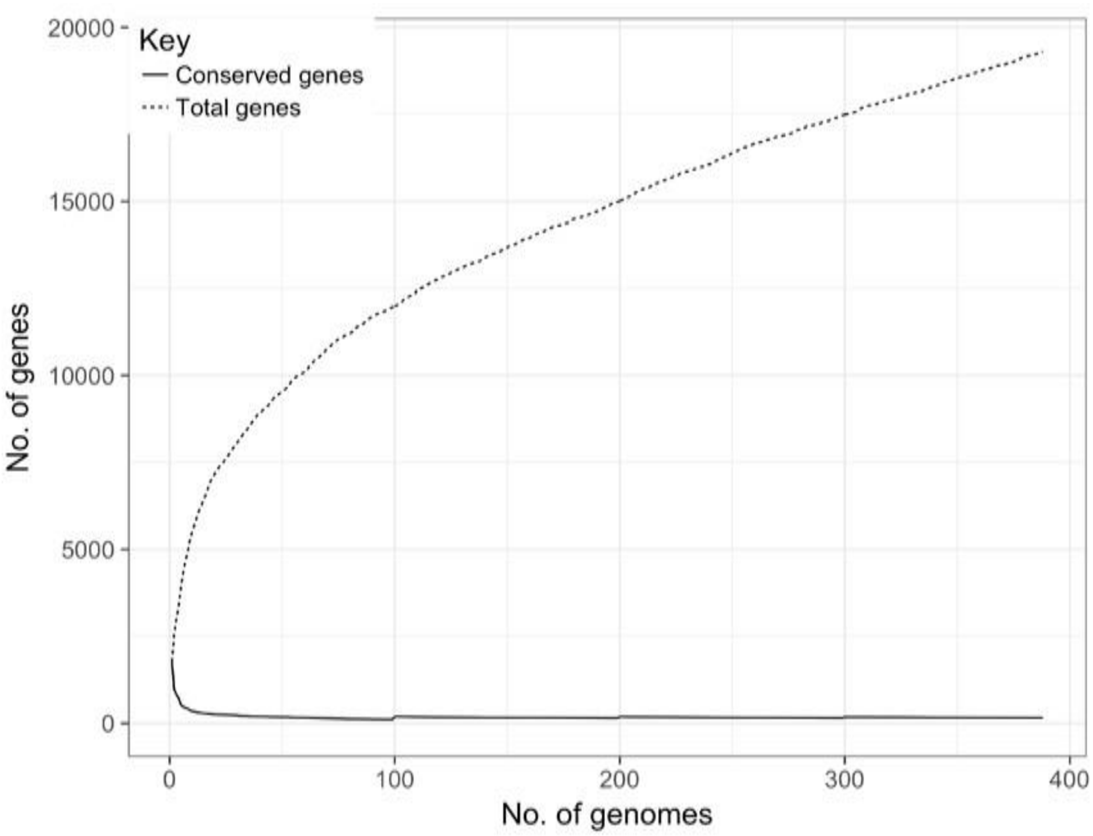


Figure 6.10 Accumulation curves showing 19301 total (dotted line) and 324 conserved genes (solid line) in the pan genome derived from 91 *H. haemolyticus* and 297 NTHi isolates.

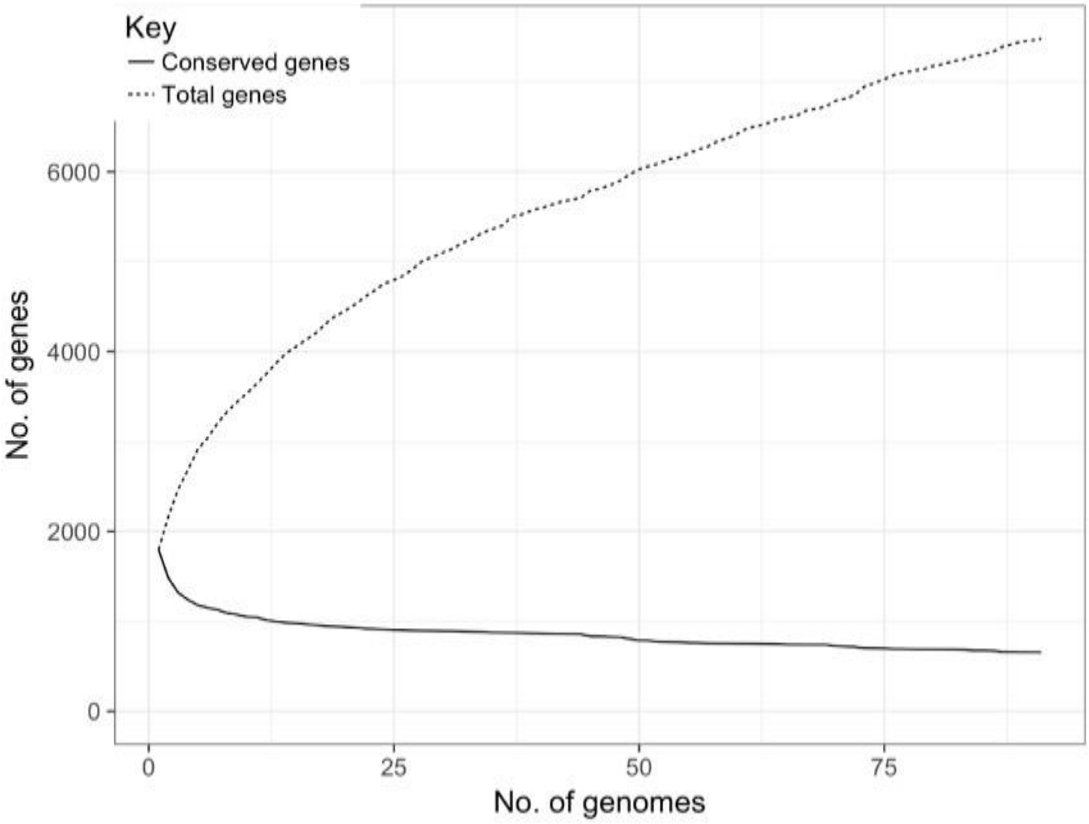


Figure 6.11 Accumulation curves showing 7479 total (dotted line) and 1064 conserved (solid line) genes in the pan genome derived from 91 *H. haemolyticus* isolates.

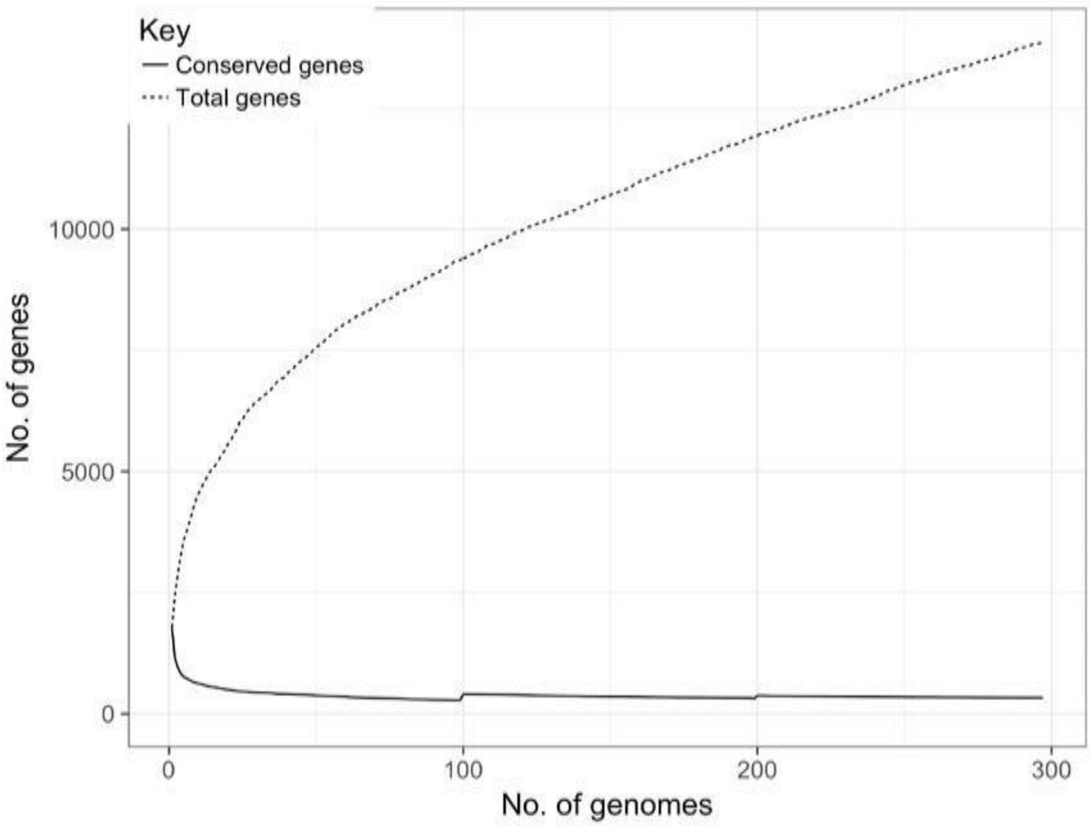


Figure 6.12 Accumulation curves showing 13869 total (dotted line) and 373 conserved genes (solid line) in the pan genome derived from 297 NTHi isolates.

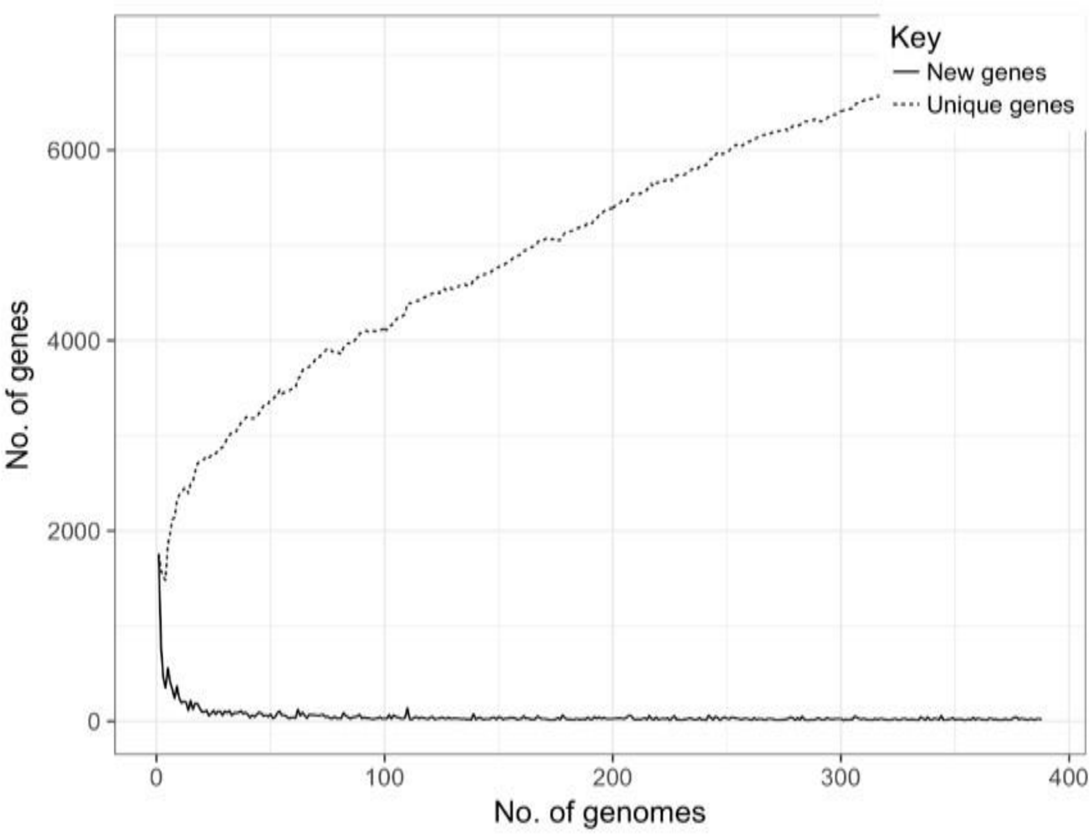


Figure 6.13 Accumulation of genes on initial identification into pan genome for 297NTHi and 91 *H. haemolyticus*.

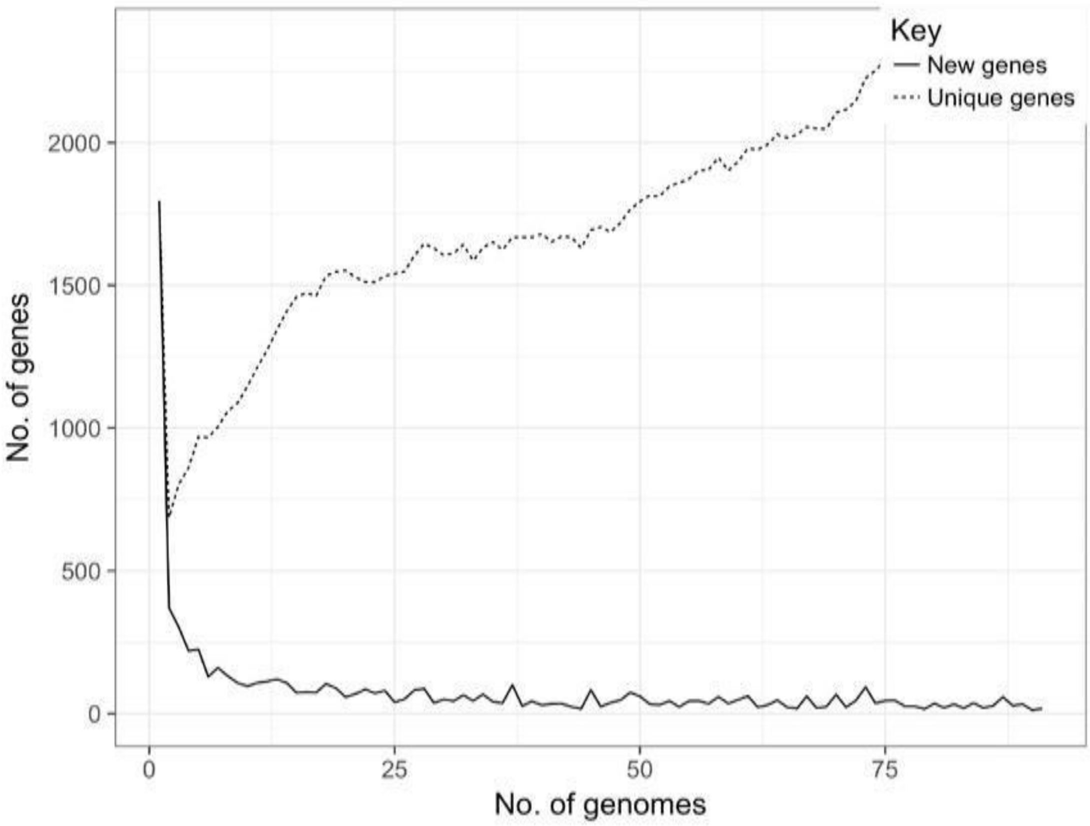


Figure 6.14 Accumulation of genes on initial identification into pan genome for 91 *H. haemolyticus*.

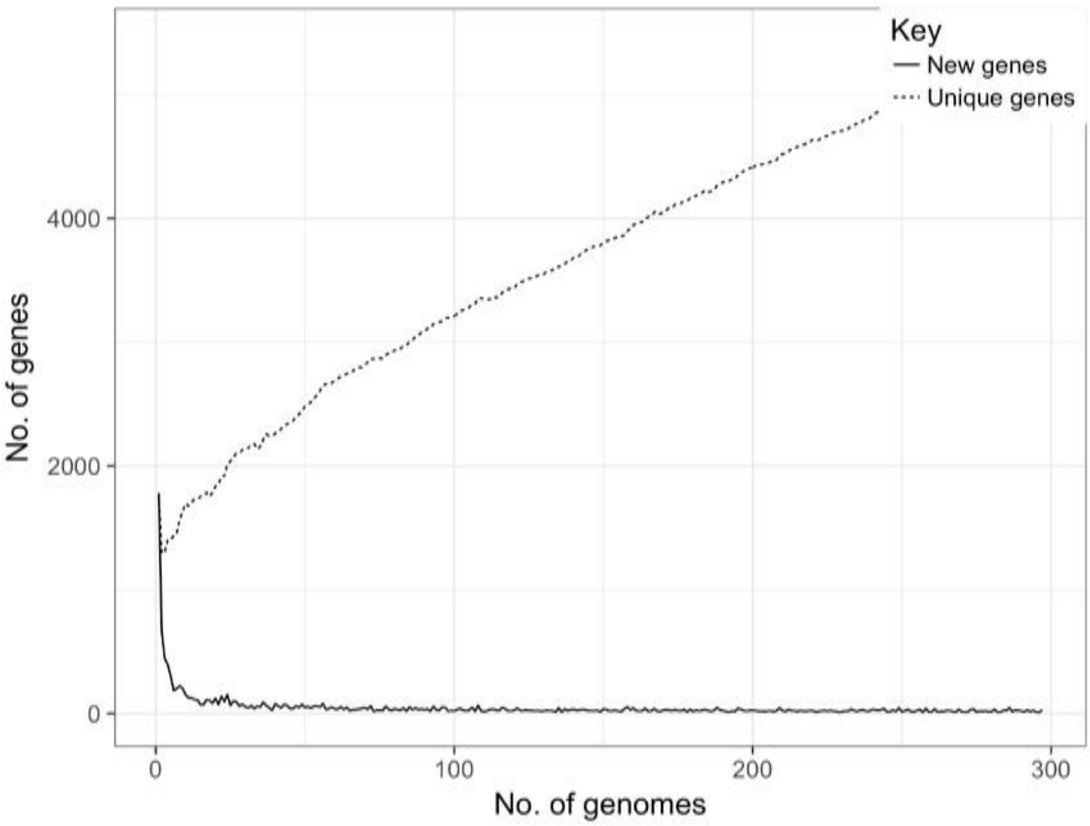


Figure 6.15 Accumulation of genes on initial identification into pan genome for 297 NTHi.

6.4.7 Gene Markers for Differentiation

Using the output from ROARY, Scoary was used to identify genes in both species that were statistically significantly associated with either NTHi or *H. haemolyticus*, isolate datashown in (Materials and Methods, 2.3.9). Results are shown in Table 6.1.

The genes with higher sensitivity were those that had a greater measure of predictive power in identifying NTHi from *H. haemolyticus*. Those with a high specificity were those that were able to more reliably identify NTHi in their absence and were therefore present in *H. haemolyticus*. From the 19301 genes identified from the isolates, 33 were found tobe 100% sensitive and 100% specificity and suitable for differentiating between the two species within this isolate set. However, none of the previous gene markers published and discussed in Chapter 3 were identified. Seven genes were unidentifiable by annotation and reported as hypothetical proteins (Table 6.1).

Table 6.1 Genes with 100% specificity and 100% sensitivity associated with identifying NTHi from *H. haemolyticus*.

Gene name	Annotation
ampD	1,6-anhydro-N-acetylmuramyl-L-alanine amidase AmpD
qseC	Sensor protein QseC
ftsL	Cell division protein FtsL
ureR	Urease operon transcriptional activator
vqgF	Putative Holliday junction resolvase
ddpD	putative D,D-dipeptide transport ATP-binding protein DdpD
glmS	Glutamine--fructose-6-phosphate aminotransferase [isomerizing]
fur	Ferric uptake regulation protein
ruvB	Holliday junction ATP-dependent DNA helicase RuvB
rsmE	Ribosomal RNA small subunit methyltransferase E
murE	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase
lptF	Lipopolysaccharide export system permease protein LptF
lptC	Lipopolysaccharide export system protein LptC
mraZ	Protein MraZ
rnfE	Electron transport complex protein RnfE
ptsN	Nitrogen regulatory protein
argR	Arginine repressor
ligA_1	DNA ligase
dsbB	Disulfide bond formation protein B
trpG_2	Anthranilate synthase component 2
murG	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
group_12981	EamA-like transporter family protein
group_11942	Isoprenylcysteine carboxyl methyltransferase (ICMT) family protein
group_10792	Sulfite exporter TauE/SafE
group_10040	transcriptional regulator BolA
group_7712	Competence Protein A
group_6435	hypothetical protein
group_7562	hypothetical protein
group_11914	hypothetical protein
group_9044	hypothetical protein
group_10875	hypothetical protein
group_11723	hypothetical protein
group_10039	hypothetical protein

6.5 Discussion

The objective of this chapter was to ascertain whether *H. haemolyticus* was genetically related to NTHi to such a degree that species boundaries were blurred or whether the two could continue to be classified as separate species. NTHI exhibits great genetic diversity yet *H. haemolyticus* displays a large genetic similarity to NTHi. However, using taxonomic methods and whole genomes it has been shown that NTHi and *H. haemolyticus* are indeed two separate species.

The MetaPhlAn analysis displayed the variation between the four different groups identified in Chapter 3; typical, atypical, *fucK* negative and *H. haemolyticus.* The atypical group was observed to have less genetic similarity to the reference marker genes with a range of 80.53%-89.97%. The typical and *fucK* negative groups displayed a percentage identity match of 86.46%-100% and 93.9% -100% and the *H. haemolyticus* ranged from 70.95%-82.09% with means of 85.95%, 96.25%, 96.82% and 75.34% respectively (Figure 6.1). The 260 genes used for this classification is a small percentage of those available in the NTHi genome but much broader than the gene markers used initially for classification (Chapter 3). Despite the increase in genes for comparison the *H. haemolyticus* andatypical isolates were seen to cluster in their groups whereas the *fucK* negative and NTHi didnot (Figure 6.1).

No previously reported gene marker was found to be statistically significant for the differentiation 100% sensitivity or 100% specificity. This could however be a reflection of the isolates within this study. As NTHi is such a heterogeneous species, attempting to establish a single gene marker to use for identification appears to be redundant. Therefore, the use of whole genome analysis is more suitable when investigating the differences between NTHi and *H. haemolyticus*.

When applying ANI to whole genome data, the atypical isolate group and the *fucK* negative group were greater in similarity to the typical NTHi than that of *H. haemolyticus.* The atypical group were shown to be more diverse from the majority of the other NTHiisolates but a greater similarity to NTHi rather than *H. haemolyticus* was observed. The majorityof NTHi were 94-96% identical to the atypical group whereas the *H. haemolyticus* isolates were <92% match to all NTHi (Figure 6.4). The similarity between the two species islarge with an average of 91% identity match between NTHi and *H. haemolyticus* (Figure 6.4)and this is reflected in the results of previous chapters and historical research into gene markers.

The thresholds for speciating isolates by ANI has been recommended as 95%, anything below is thought to be indicative of isolates from a different species although certain species with high levels of heterogeneity such as the *E.coli* - Shigella group have been shown to have an ANI <95% (Konstantinidis and Tiedje, 2005a, Richter and Rossello-Mora, 2009, Kim et al., 2014). The *H. haemolyticus* isolates can clearly be seen to sit in their own clade in Figure 6.5 and all have <92% identity to NTHi (Figure 6.4), including the atypical isolates, therefore in the case of NTHi and *H. haemolyticus* a threshold of 94% similarity would be more descriptive for the intra species relationship for both NTHi and *H. haemolyticus*. The speciation can be seen in the phylogeny of the isolates in Figure 6.5 which supports the clear definition between the isolates into NTHi and *H. haemolyticus*.

Phylogeny of the 297 NTHi and 91 *H. haemolyticus* isolates support the findings of ANIwith *H. haemolyticus* clustering separately to NTHi. The atypical isolates are also seen tocluster but closer to the NTHi than the *H. haemolyticus* (Figure 6.1).

The resulting number of shared core genes was found to be low between the species. From an identified 19301 genes, only 153 genes were detected in 99-100% of the isolates and 171 genes detected in between 95-99% of the isolates (Figure 6.6). A similar study has reported a larger number of core genes between NTHi isolates, this may be indicative of the number of isolates used in analysis or representative of the variation within the isolates it also may reflect the unclosed status of the pan genome described in this study (De Chiara et al., 2014). The pan genome for *H. influenzae* has been estimated as 4452-6052 genes in size however this was from 13 strains only (Hogg et al., 2007). Here we have used 22 times that number of isolates and this has resulted in a pan genome of 13689, 2.26x larger than the estimated pan genome from the smaller study. The resulting pan genome shows the diversity within NTHi and to a lesser extent *H. haemolyticus* but it also questions the number of strains to be used for pan genome analysis in species that exhibit such diversity. A pan genome is declared a closed pan genome at the point where there is no more accumulation of genes despite the addition of further strains. The minimal number of isolates required to characterise a closed pan genome is thought to depend on the heterogeneity of the species in question (Rouli et al., 2015). A species such as *E.coli* that exhibits a great deal of heterogeneity is thought to be infinite by definition whereas a reasonably conserved species such as *Yersinia pestis* may require few strains to close the pan genome (Rouli et al., 2015). The heterogeneity of *H. influenzae* has been reported and this is supported by the results from this chapter with the open pan genome resultingfrom 297 NTHi isolates (Figure 6.12). Similarly, the 91 isolates used for the *H. haemolyticus* pan genome were not sufficient to close the pan genome inferring that *H. haemolyticus* also displays genetic diversity throughout the species (Figure 6.11).

6.6 Conclusion and Future Work

The pan genome of both species requires further investigation. The resulting coregenomes of the isolates from this study are very small in comparison to the limited studies intothe pan genome of *H. influenzae* (Rouli et al., 2015, Hogg et al., 2007). Species diversity would explain a small core genome which when analysing both species together would be expected however the core genomes are still much lower than previously reported (Rouli et al., 2015, Hogg et al., 2007). Recombination analysis may result in a smaller core genome also so applying a tool that initiates a recombination filter such as GUBBINS may be beneficial (Croucher et al., 2015).

MetaPhlAn, ANI and phylogeny all resulted in clear separation of *H. haemolyticus* isolates from NTHi, including the atypical isolates. The *fucK* negative isolates were not found to be diverse enough to separate themselves from the remaining typical NTHi. ANI calculations define *H. haemolyticus* as <92% similar to NTHi however atypical isolates can be <95% to typical NTHi. These isolates however were only just below at 94% similarity and show more similarity to typical NTHi than *H. haemolyticus*. It can be concluded that there issufficient genetic diversity between NTHi and *H. haemolyticus* to enable the separate speciation of both species.

7 Final Discussion

NTHi is an opportunistic pathogen that colonises the respiratory tract and can cause both invasive and non invasive disease (Collins et al., 2015, Cardines et al., 2007, Erwin et al., 2005, Angoulvant et al., 2015, Barkai et al., 2009, Cody et al., 2003, Leach et al., 2016, Leach et al., 2015). It is also associated with the onset of acute exacerbations in COPD, the third largest cause of global mortality which is reported by the WHO to affect approximately 65 million people worldwide (WHO, 2017). Both NTHi and its closest phylogenetic relative, *H. haemolyticus* can be isolated from the COPD lung (Murphy et al., 2007). Atypical results from PCR assays developed for identification purposes led to a suggestion that the two may form a genetic continuum and not two separate species (Binks et al., 2012), however the relationship between the two species had not been thoroughly investigated.

The primary aim of this thesis therefore was to characterise the relationship between NTHi and *H. haemolyticus* isolated from the COPD lung and to ascertain whether or not the genetic similarity reported warranted the reclassification of the two species as a single species. This required differentiating the isolates as accurately as possible into typical NTHi and *H. haemolyticus* isolates and identifying any outside these classifications using both culture techniques and molecular markers (Chapter 3). Following this, adhesin genotypes and azithromycin resistance were investigated (Chapters 4 and 5) to identify respectively potential explanations for variation in carriage prevalence or virulence. Finally to fully appreciate the shared genetic content of the isolates whole genome analysis and *in silico* taxonomy methods (e.g ANI) were employed (Chapter 6).

This thesis reveals, for the first time, the extent of the genomic heterogeneity of NTHi longitudinally recovered from COPD patients. The delineation of NTHi and *H. haemolyticus* as separate species has been firmly established. Based on the analysis of singular molecular markers for differentiation between the species this thesis proposes that only whole genome analysis is adequate to differentiate between NTHi and *H. haemolyticus*.

In spite of this clear species delineation, similar high-level resistance to azithromycin, unexpected in the context of previous reports, was observed in both species. However, the genetic mechanism underlying this resistance could not be determined. Finally, this thesis also used results from the comparative genomic analyses as a starting point into understanding why NTHi and *H. haemolyticus* interact differently with epithelial cells through the examination of adhesin genotypes (Pickering et al., 2016, Binks et al., 2012).

7.1 Identification

How important it is to differentiate between the two species is open for debate. Initially the need to differentiate between the two species arose to ensure correct treatment prescription, accurate epidemiology data and to monitor potentially pathogenic cases of *H. haemolyticus*. However, since the initial description of a 39.5% misclassification of *H. haemolyticus* misidentified as NTHi by Murphy *et al.* 2007 other studies have retrospectively looked at collected NTHi isolates and discovered much smaller numbers of cases of misidentification. The second largest number of *H. haemolyticus* (11%) reported was discovered within *Haemophilus* spp. isolated from Aboriginal children but this particular indigenous group has previously been shown to carry unexplainably high levels of *Haemophilus* spp. (Kirkham et al., 2010, Pickering et al., 2014b). Levels of misidentification in clinical samples have been very low, 1.5% in an Australian study of varied clinical samples, 0.5% in cystic fibrosis isolates and 0.4% in Danish clinical isolates (Zhang et al., 2014, Fenger et al., 2012, Norskov-Lauritsen, 2009). Given this low level of misidentification it raises the question as to whether the extra expense and manpower to carry out assays to differentiate between the species is worthwhile, particularly when simple overnight culture assays for growth factors and haemolysis can ensure timely treatment for invasive, potentially life threatening *Haemophilus* disease. Singular molecular markers such as *hpd* and *fucK* would result in certain strains of NTHi being incorrectly classified as *H. haemolyticus* and therefore risk delaying the treatment of the patient. Ideally multiple markers would be used, or metagenomics applied to the sample to identify all potential pathogens present but currently this would be too expensive and lengthy for clinical usage. Currently the extra cost and manpower do not seem appropriate for the reported cases of clinical misidentification of *H. haemolyticus* as NTHi.

The only phenotype used to differentiate *H. haemolyticus* from NTHi in culture-based methods is haemolysis on blood agar. However, *H. haemolyticus* is known to lose its haemolytic ability when cultured on blood agar (Kilian, 1976a). Both species require NAD and Hemin growth factors to grow *in vitro* and both are morphologically similar, nonetheless in a large study it has been shown that occasionally haemolytic isolates go unnoticed when identification and isolate collection is taken from an original sample plate. This could be due to overcrowding or isolation from a haemolysed agar plate (Murphy et al., 2007). The isolates identified as *Haemophilus* spp. due to morphology and growth factors initially were put through an extra step of innoculation of a single isolate on blood agar to check for haemolysis. This simple step identified twelve isolates that could have been disregarded as *H. haemolyticus* when looking for clinically relevant NTHi (Chapter 3).

As previously stated Murphy *et al.* identified close to 40% of *Haemophilus* spp. isolated from sputa collected from the COPD lung retrospectively as *H. haemolyticus* using molecular methods. Of this resulting group, 50% of these isolates were also revealed to be haemolytic (Murphy et al., 2007). The present study also revealed haemolytic *H. haemolyticus* within the *Haemophilus* spp. indicating that a small culture test to eliminate haemolytic isolates would be beneficial (Chapter 3). An optimum time frame of 48 hours to identify all haemolytic samples was observed in this thesis, previously the standard has been overnight culture (Chapter 3)(PHE, 2015).

Previous studies comparing the sensitivity and specificity rates of single molecular markers for differentiating NTHi from *H. haemolyticus* have concluded that no single marker is capable of unambiguously identifying NTHi from *H. haemolyticus*, this was also confirmed in Chapter 3 (Binks et al., 2012, Chang et al., 2010, de Gier et al., 2015, Hu et al., 2016). The use of single molecular markers to differentiate between NTHi and *H. haemolyticus* has been widely unsuccessful due to the heterogeneity of NTHi and the similarity to *H. haemolyticus* with no single molecular marker displaying 100% sensitivity and specificity repeatedly (Binks et al., 2012, McCrea et al., 2008, Sandstedt et al., 2008).

The use of MLST is to characterise NTHi has been questioned after the discovery of fucK negative strains, one of the MLST (de Gier et al., 2015, Meats et al., 2003, Ridderberg et al., 2010). Indeed, strains of NTHi have been identified for which the entire fucose operon is absent, as is the norm in *H. haemolyticus*. Whole genome analysis has been employed to determine single molecular markers that are considered suitable for identification purposes however the results described in Chapter 3 do not reflect the findings of these studies (Price et al., 2015, Hu et al., 2016). When identifying a gene more statistically associated with identification of NTHi, the origin and limitations of the actual isolates themselves must be taken into account. The isolates used in the present study were taken from a cohort of COPD patient from Hampshire, UK, suffering from moderate to severe COPD (Bourne et al., 2014). As this thesis specifically describes isolates from the associated clinical study, with its geographical limitations and specific patient inclusion criteria, the findings cannot be generalised broadly for NTHi and *H. haemolyticus*.

Initial characterisation of the isolates by identification of genotype using molecular markers and MLST enabled the classification of the isolates into four groups; typical NTHi, atypical NTHi, *H. haemolyticus* and *fucK* negative NTHi (Chapter 3). The discovery of the *fucK* negative strains has been previously reported, however this group did not cluster separately in phylogenetic analysis of marker or adhesin genes and was not seen to separate from typical NTHi when compared using MetaPhlAn or whole genome analysis (Chapter 6) (de Gier et al., 2015, Ridderberg et al., 2010). This suggests that the absence of the *fucK* gene does not necessarily correspond to a divergent strain. A sequence for all seven housekeeping genes of the MLST schema is required for inclusion; therefore, currently there is no mechanism to type *fucK* negative strains (Meats et al., 2003). As these are frequently reported the schema falls short of suitably characterising NTHi (de Gier et al., 2015, Ridderberg et al., 2010).

7.2 Adhesin Genotypes

The difference in pathogenesis between NTHi and *H. haemolyticus* could be related, in part, to the recent findings of the different way in which they interact with epithelial cells (Pickering et al., 2016, Singh et al., 2016a).

Primarily *H. haemolyticus* has been classified as non pathogenic and has been shown to interact with epithelial cells very differently from NTHi, resulting in cytosis rather than cell invasion (Pickering et al., 2016). Adhesin genes were investigated to ascertain a possible mechanism for this difference in behaviour between the two species (Chapter 4). This thesis is the first study into the adhesin genotype of *H. haemolyticus* and showed some clear differences in adhesin genotype between the two species (Chapter 4). Certain genes (*hia*, *hmwA* and *hmw2A*) that have been more associated with strains of NTHi isolated from invasive disease were found present in a portion of the NTHi but none were identified in *H. haemolyticus* (Chapter 4) (Vuong et al., 2013, Rodriguez et al., 2003, St Geme et al., 1998). The *fucK* negative group of NTHi did not show any difference in adhesin genotype from the typical NTHi. Despite the absence of the *fucK* gene from NTHi being the focus for atypical strains in previous research the *fucK* negative NTHi from this study appears not to reflect a large genetic diversity from typical NTHi and therefore hypothetically they should display the same mechanism of adherence. Currently it is not possible to determine this relationship for the atypical group of NTHi where an alternative complement of adhesins may reflect mechanisms more associated with *H. haemolyticus*.

7.3 Azithromycin Resistance

There is a recent trend of prophylactic use of azithromycin in inflammatory respiratory disease such as asthma, cystic fibrosis and COPD (Albert et al., 2011, Altenburg et al., 2013, Anwar et al., 2008). Although previously for NTHi attention has been focussed on amoxillicin resistance and beta lactamases (Hotomi et al., 2006, Mendelman et al., 1987), this use of azithromycin has raised the issue of the potential for increased resistance to this antibiotic treatment. We also hypothesised that the exposure of *H. haemolyticus* to azithromycin due to misidentification as NTHi, or given that it was demonstrated that the two species can be co-carried (Chapter 3), might result in increased resistance in this species as well. This thesis presents the first study that looks into the resistance of both NTHi and *H. haemolyticus* in the COPD lung. NTHi and *H. haemolyticus* displayed surprisingly high levels of resistance and intermediate resistance to azithromycin when compared to previous studies (Chapter 5).

The azithromycin resistant phenotype has been associated with AMRG in *S. pneumoniae* and NTHi, and mutations in the L4 and L22 and 23s rRNA genes (Chen et al., 2013, Chisholm et al., 2010, Galarza et al., 2010, Roberts et al., 2011). Bioinformatic analysis of all isolates found no AMRG and no variation within the L4, L22 or 23s rRNA genes, associated with resistance levels (Chapter 5).

Furthermore, no previously reported mutations associated with L4, L22 or the 23s rRNA gene were identifiable in the isolates tested. These previous studies have failed to follow up investigations to ascertain the protein expression of these isolates and the interaction with azithromycin, and are further compromised by small sample sizes. Due to the lack of obvious genetic mechanism for the resistance levels statistical analysis was carried out to identify any gene within the isolates that were more associated with the resistance phenotype. This thesis provides a list of genes associated with high levels of resistance in both NTHi and *H. haemolyticus* as a starting point for further investigation into resistance mechanisms in *Haemophilus* spp.

7.4 Taxonomy and Whole Genome Analysis

This thesis for the first time reveals the level of genomic similarity between NTHi and *H. haemolyticus*. Whilst the accumulation of separate gene markers helps build a picture of the genetic characteristics of the *Haemophilus* spp., comparing the whole genome data resulted in a more comprehensive comparison. MetaPhlAn analysis using 260 marker genes demonstrated no clear variation between the typical NTHi and the *fucK* negative isolates. Furthermore, the *H.haemolyticus* isolates were shown to exhibit a lower percentage of similarity to the gene markers used and the atypical group were seen to fit between the NTHi and *H. haemolyticus* groups. Average nucleotide identity showed that NTHi were over 94% similarity to each other and <92% similar to the *H. haemolyticus* isolates. This method therefore defines the two groups as separate species. This is in spite of the substantial heterogeneity within the NTHi as shown throughout this thesis. *H. haemolyticus* and NTHi were seen to share 91% of their genomes supporting the close genetic relationship previously reported (Chapter 6) (Binks et al., 2012, Hinz et al., 2015). No gene was found to be associated with differentiation of NTHi and *H. haemolyticus* to 100% sensitivity however, 11 genes were found to have sensitivity levels of >99%. No previously published markers, as used in Chapter 3, were identified in this analysis.

7.5 Future Work

In silico analysis is putative and therefore requires biological validation. Although the isolate set presented here enabled robust comparative genomic analyses, topics such as adhesins and antibiotic resistance require corroboration from more traditional, laboratory-based analysis. The expansion of the adhesin genotype investigation into determining the associated phenotypes e.g. performing adherence assays for the isolates, would allow for further understanding as to the differences in pathogenicity between NTHi and *H. haemolyticus*. This would also aid in supporting and understanding the recent research revelation into their different mannerisms in interacting with epithelial cells (Pickering et al., 2016, Singh et al., 2016b).

The requirement for extensive understanding of antibiotic resistance is now more important than ever due to the emergence of azithromycin resistant strains of NTHi and *H. haemolyticus* in the COPD lung (Chapter 5). It would be beneficial to expand this to investigate NTHi and *H. haemolyticus* from different disease states and sites of isolation as well as carriage. This would ascertain whether the high levels of resistance observed in Chapter 5 are widespread within the species or are specific to the COPD lung or the locality of the study. The mechanism for the observed resistance could not be identified. In this study a gene-based statistical analysis was undertaken using the program Scoary to try to identify loci associated with this phenotype. This relies on the presence or absence of genes derived from core-genome annotations alone, neither accounting for variability across these gene sets nor the variability within the accessory genome. As such a more powerful approach would be to adopt a kmer searching strategy such as implemented in SEER (Lees et al., 2016). Here the association between kmers and phenotype can indicate loci within whole genomes, not just core alignments that are associated with the phenotype in question. Once identified these loci can be further interrogated using laboratory-based methods for the identification of function.

Lastly it would then be necessary to compare the mechanisms used by NTHi and *H. haemolyticus* to determine whether it is orthologous and if so is this due to recombination between the two species. In understanding these mechanisms tests can be developed for the genes in question so treatment can be targeted to improve patient outcomes. Furthermore, this information could assist the development of vaccines or alternative treatment options such as replacement colonisation in response to these antibiotic resistant strains.

7.6 Conclusions

NTHi is a heterogeneous bacterial species with a large pan genome and an ability to evolve through mutations and recombination. The study size took into account 1460 *Haemophilus* spp. of which 1368 were defined as NTHi, including the atypical and *fucK* negative groups. However, the sputa samples from which the bacteria were isolated were extracted from a limited number of patients from the same geographical area (Hampshire, UK). This is a limitation of this study since only the genetic content of a limited population of NTHi was examined. Based on a previous COPD study which reported 39.5% misidentification rate of *H. haemolyticus* as NTHi, the level of *H. haemolyticus* misidentified within this thesis was a lot lower at (Murphy et al., 2007). Therefore, a certain amount of bias is recognised due to the high numbers of NTHi compared to *H. haemolyticus* but also due to using all *H. haemolyticus* regardless of time point or patient in subsets. The limited number of *H. haemolyticus* may also explain the limited variability of *H. haemolyticus* compared to that of NTHi. Expanding this investigation to incorporate different geographical locations, disease sources, carriage isolates and site locations would be beneficial to get a wider range of NTHi and *H. haemolyticus* and therefore a more accurate genomic picture.

This thesis was heavily reliant on the *in silico* analysis of a large quantity of sequencing data and with such a large dataset comes its own set of problems. Some bioinformatic tools cannot handle the volume of isolates within the study, some could not handle the variation between the two species to enable comparison and some took a long timeframe to complete.

Bioinformatics is a rapidly growing field and tools are currently being developed to investigate the ever-increasing size of the datasets. These tools are not generally built as a part of a business model as software would be required to for commercial purposes but as a response to a need for data analysis developed for a research purpose within a study group (Seeman, 2014, Alikhan et al., 2011, Seemann, 2014, Treangen et al., 2014, Page et al., 2015, Stamatakis et al., 2012, Inouye et al., 2014). This results in software tools that are on going projects with bugs and fixes often requiring troubleshooting and further development. Some are not sufficiently user friendly to understand errors, or limitations in the capability of the tool. However, the science community is a friendly one and the exchange of emails has enabled me to use many of these tools on my data where my first attempt has failed.

This thesis would further benefit from protein expression analyses for biological confirmation of putative results. Genome wide association studies to identify loci linked with antibiotic resistance could lead to targeted gene knock-out and functional assessments of their role in this phenotype. Where phenotypes have been found to be important, such as the binding affinity of the various outer membrane adhesins to different cellular receptors, then screening the range of genotypes observed against these known associations would shed further light on the interplay between microbe and host for these two species. The isolates as previously mentioned are from a limited isolate set all from the same geographical location and disease state. It does, however provide an in depth bioinformatics investigation into NTHi and *H. haemolyticus* from a large longitudinal study resulting in a large, clearly speciated sample set.

7.7 Concluding Remarks

This thesis characterised and described 1460 *Haemophilus* spp. isolated from COPD and confirmed once and for all that NTHi and *H. haemolyticus* are indeed two separate species. Although the two species are separate by taxonomy standards they do share a large percentage of genetic identity and there is evidence of an atypical subset of NTHi that exhibits a genotype containing characteristics that are more associated with *H. haemolyticus*. This thesis has furthermore shown differences between the two species in an important virulence genotype and highlighted worrying levels of resistance in both NTHi and *H. haemolyticus* isolated from the COPD lung.

In conclusion, NTHi and *H. haemolyticus* although genetically similar, show marked differences in adhesin genotype, which may explain their different reported interactions with epithelial cells (Pickering et al., 2016). Further investigation into adherence may indicate a survival advantage and assist in developing replacement therapies for *H. haemolyticus* to replace NTHi in carriage. Both species however, display a worrying level of resistance to azithromycin in the COPD lung of which the mechanism could to be determined. This would benefit further investigation as this may affect prophylactic treatment regimes to manage exacerbations in COPD patients with azithromycin.

Appendices

Appendix 1: Bioinformatics scripts and usage

/path/= path leading to required file

Renaming files

The MiSeq assigns an ‘S number’ naming convention from S1 – S24 which identifies it within the run. To enable ease of batching files for later analysis the S number is renamed to S24 (the number is arbitrary, the filename requires an s number but it does not matter which). When performing these tasks on single files changing the S number is not required.

```
for i in????_S*_L001_R?_001.fastq.gz
do
        x=$(sed -E 's/(_S[0-9]{1,2}_)/_S24_/' <<< "$i")
        echo "$i renames to: $x"
        mv $i $x
done
```

Trimming adapters off sequence data

Illumina adds adapters to sequence data, before assembling these are removed as they are not part of the genome and not required for analysis.

To run a single set of paired end fastq files through trimmomatic

```
java -jar /local/software/trimmomatic/0.32/trimmomatic-0.32.jarPE 0076_S2_L001_R1_001.fastq.gz 0076_S2_L001_R2_001.fastq.gz 76-paired_1.fastq.gz 76-unpaired_1.fastq.gz 77-paired_2.fastq.gz 76-unpaired_2.fastq.gz SLIDINGWINDOW:10:20 MINLEN:50 ILLUMINACLIP:nextera.fa:1:40:15
```

This results in 4 files:-

isolateNumber-paired_1.fastq.gz
isolateNumber-unpaired_1.fastq.gz
isolateNumber-paired_2.fastq.gz
isolateNumber-unpaired_2.fastq.gz

To run multiple trims.

Run_trimmomatic.pbs is a batch script that runs a loop through a set of numbers passed to the script through the qsub command on Iridis high performance cluster.

```
qsub –t x-y run_trimmomatic.pbs
```

where x is first isolate number in the sequence and and y is the last number.

Run_trimmomatic.pbs

```
!/bin/bash
=============================================

PBS -l nodes=1:ppn=1
PBS -l walltime=01:00:00
Change to directory from which job was submitted
cd $PBS_O_WORKDIR
 load software module
module load jdk/1.7.0
adapter_file="nextera.fa"
fastq_file=" 0"$PBS_ARRAYID"_S24_L001_R1_001.fastq.gz"
R2fastq_file=" 0"$PBS_ARRAYID"_S24_L001_R2_001.fastq.gz"
if [ -f $fastq_file ]
then
 if file exist - run script
  java -jar /local/software/trimmomatic/0.32/trimmomatic-0.32.jar PE $fastq_file
$R2fastq_file $PBS_ARRAYID-paired_1.fastq.gz$PBS_ARRAYID-unpaired_1.fastq.gz
$PBS_ARRAYID-paired_2.fastq.gz PBS_ARRAYID-unpaired_2.fastq.gz
SLIDINGWINDOW:10:20 MINLEN:50 ILLUMINACLIP:nextera.fa:1:40:15
      else
 if file not found - print error message and exit
              echo 'File '$fastq_file' not found, quit'
exit
fi
```

Check Quality Data for Sequencing

The fastQC programme returns statistics on the raw sequencing data.

```
fastqc seqfile1.fastq.gz seqfile2.fast.gz > seqfileoutput
```

Running MLST Analysis using SRST2

Dependencies - MLST db –requires downloading

```
Module load srst2
getmlst.py - -species  "Haemophilus influenzae"
```

The genome numbers for the trimmed sequence read files needs to be passed through to the PBS_ARRAY. (x-y)

```
qsub –t x-y srst2.pbs
```

SRST2.pbs

```
!/bin/bash
=====================================================


PBS -l nodes=1:ppn=1
PBS -l walltime=00:30:00
cd $PBS_O_WORKDIR
module load srst2
fastq_file=$PBS_ARRAYID"-paired_1.fastq.gz"
R2fastq_file=$PBS_ARRAYID"-paired_2.fastq.gz"
if [ -f $fastq_file ]
then
 if file exist - run script
srst2 --input_pe $fastq_file $R2fastq_file --output $PBS_ARRAYID --log --mlst_db
Haemophilus_influenzae.fasta --mlst_definitions haemophilus.txt
else
	echo 'File '$fastq_file' not found, quit'
exit
fi
```

Once SRST2 is complete the following will enable all results to be collated into one text file and converted into a spreadsheet for further analysis.

```
for i in ? mlst Haemophilus_influenzae  results.txt
        do
        LineToAppend=$(sed -n -e '2p' $i)
        echo $LineToAppend >> collated_mlst_results.txt
done
```

Gene mapping using SRST2 for establishing gene sequences

SRST2_gene.pbs

```
!/bin/bash
==============================================


PBS -l nodes=1:ppn=1
PBS -l walltime=00:30:00
cd $PBS_O_WORKDIR
module load srst2
fastq_file=$PBS_ARRAYID"-paired_1.fastq.gz"
R2fastq_file=$PBS_ARRAYID"-paired_2.fastq.gz"
if [ -f $fastq_file ]
then
 if file exist - run script
srst2 --input_pe $fastq_file $R2fastq_file --output $PBS_ARRAYID --log --gene_db
GenBank_seq.fasta --min_coverage 60 --max_divergence 40--report_all_consensus
else
	echo 'File '$fastq_file' not found, quit'
exit
```

Assembling Genomes De Novo with MaSuRCA

The following script creates the necessary directory, moves the paired fastqs into this directory along with the required configuration file and amends the configuration file ready for assembly.

```
for i in ABC-???-paired_1.fastq.gz
        do
                x=$(sed -E 's/(ABC-)//' <<< "$i")
                y=$(sed -E 's/-paired_1.fastq.gz//' <<< "$x")
        mkdir -p "$y"
        cp ABC-"$y"-paired_1.fastq.gz $y
        cp ABC-"$y"-paired_2.fastq.gz $y
        cp sr_config_example.txt $y
        sed -i 's/filename1/'ABC-"$y"-paired_1.fastq.gz'/g' $y/sr_config_example.txt
        sed -i 's/filename2/'ABC-"$y"-paired_2.fastq.gz'/g' $y/sr_config_example.txt
done
```

MaSuRCA example configuration file

```
 DATA is specified as type {PE,JUMP,OTHER} and 5 fields:
 1)two_letter_prefix 2)mean 3)stdev 4)fastq(.gz)_fwd_reads
 5)fastq(.gz)_rev_reads. The PE reads are always assumed to be
 innies, i.e. --->.<---, and JUMP are assumed to be outties
 <---.--->. If there are any jump libraries that are innies, such as
 longjump, specify them as JUMP and specify NEGATIVE mean. Reverse reads
 are optional for PE libraries and mandatory for JUMP libraries. Any
 OTHER sequence data (454, Sanger, Ion torrent, etc) must be first
 converted into Celera Assembler compatible .frg files (see
 http://wgs-assembler.sourceforge.com)
DATA
PE= pe 251 38 filename1 filename2
END
PARAMETERS
this is k-mer size for deBruijn graph values between 25 and 101 are supported, auto will
compute the optimal size based on the read data and GC content
GRAPH_KMER_SIZE=auto
set this to 1 for Illumina-only assemblies and to 0 if you have 1x or more long (Sanger, 454)
reads, you can also set this to 0 for large data sets with high jumping clone coverage, e.g.
>50x
USE_LINKING_MATES=1
this parameter is useful if you have too many jumping library mates. Typically set it to 60
for bacteria and something large (300) for mammals
LIMIT_JUMP_COVERAGE = 60
these are the additional parameters to Celera Assembler.  do not worry about
performance, number or processors or batch sizes -- these are computed automatically. for
mammals do not set cgwErrorRate above 0.15!!!
CA_PARAMETERS = ovlMerSize=30 cgwErrorRate=0.25 ovlMemory=4GB
minimum count k-mers used in error correction 1 means all k-mers are used. onecan
increase to 2 if coverage >100
KMER_COUNT_THRESHOLD = 1
auto-detected number of cpus to use
NUM_THREADS= $NUM_THREADS
```

```
NUM_THREADS= 16
this is mandatory jellyfish hash size
JF_SIZE=100000000
this specifies if we do (1) or do not (0) want to trim long runs of homopolymers (e.g.
GGGGGGGG) from 3' read ends, use it for high GC genomes
DO_HOMOPOLYMER_TRIM=0
END
```

To run one assembly from inside the folder containing the relevant paired readfiles.

```
Module load masurca
masurca configuration.txt
./assemble.sh
```

To run multiple assemblies via a batch script

```
qsub masurca_oneloop.pbs
```

masurca_oneloop.pbs

```
!/bin/bash
============================================

PBS -l nodes=1:ppn=16
PBS -l walltime=03:00:00
module load masurca
Change to directory
cd $PBS_O_WORKDIR
for i in {genome_numbers};
do
   cd $i
   masurca sr_config_example.txt
   ./assemble.sh
   cd /path/
done
```

The genome assembly files can be found in CA/10-gapclose and are calledgenome.ctg.fasta and genome.scf.fasta. The following script renames the assemblies to include identifiers and collates all the genomes into one folder.

```
for i in {genome_numbers};
do
cd $i/CA/10-gapclose
pwd
mv genome.ctg.fasta $i-genome.ctg.fasta
mv genome.scf.fasta $i-genome.scf.fasta
        cp $i-genome.ctg.fasta /path/assemblies/.
        cp $i-genome.scf.fasta /path/assemblies/.
        cd /path/
done
```

Assembly metrics

Running assemblathon.pl on each assembly will return a text file with relevant statistics on the assembly. This study uses the N50 contig number, NG50 contig number, number of contigs and size of assembly as benchmarks for quality.

To run a single file

```
        perl assemblathon_stats.pl NNN-genome.scf.fasta > Stats_$i
```

To run multiple files passing genome size through to calculate NG50.

```
for i in {genome_numbers};
        do
        perl assemblathon_stats.pl $i-genome.scf.fasta > Stats_$i –genome_size=1800000
done
```

The following collates the results for the N50 contig, NG50, number of contigs and size of assembly to one text file for importing to excel

```
for i in {genome_numbers};
do
        fileName="Stats_"$i;
         echo $fileName;
        NoOfContigs=$(sed -n -e '15p' $fileName);
        N50ContigLength=$(sed -n -e '28p' $fileName);
        NG50=$(sed -n -e '30p' $fileName);
        TotalSizeofContigs=$(sed -n -e '18p' $fileName);
        echo $i $NoOfContigs $N50ContigLength $NG50ContigLength $TotalSizeofContigs
        >> Name_of_File.txt;
done
```

Once assembly work is complete remove all the surplus folder structure:-

```
for i in {genome_numbers};
do
  rm -r $i
```

```
cd /path/; done
```

MetaPhlAn

Metaphlan.py and Bowtie.db are required.

Module load numpy and bowtie2

To run multiple MetaPhlAn requests

```
for i in ABC-???-paired_1.fastq.gz;
do
        x=$(sed -E 's/(ABC-)//' <<< "$i") ;
        y=$(sed -E 's/-paired_1.fastq.gz//' <<< "$x") ; gunzip -c $i | metaphlan.py --
        input_type multifastq --bowtie2db /path/bowtie2db/mpa --nproc 12 --bowtie2out
        "$y".bt.out.txt > "$y"_meta.txt; echo "$y";
done
```

The returning text files can be collated into one text file to be converted into an easily manageable spreadsheet in excel.

```
for i in ???_meta.txt;
do
        LineToAppend=$(sed -n -e '7p' $i);
        secondLineToAppend=$(sed -n -e '8p' $i);
        echo $i $LineToAppend $secondLineToAppend >> MetaPhlan_results.txt;
done
```

IPCRESS - *in silico* PCR

Requires a .ipcress or .txt file containing primer sequences. Text file should be laid out in 5 columns in tab-delimited format.

Gene Name	Forward primer	Reverse Primer	Minimum Product Length	Maximum product length

IPCRESS is part of the European Bioinformatics Institute's exonerate package.

Module load exonerate

For each genome assembly the .ipcress file containing the primers is run resulting in atext file ($y) containing any results .

```
for i in {genome_numbers};
do
        y=$i_GeneName.fasta";
        ipcress geneName.ipcress -s $i-genome.ctg.fasta -p F -P T -m 3 >"$y";
done
```

Files =<30kb were identified as being empty of any results and so were negative results. The following records the size of the file for each genome assembly into a text file which can be opened in excel.

```
for i in {genome_numbers};
do
      filenameIPCRESS=$i"_ ipcress_results_GeneName.fasta";
      echo $filenameIPCRESS;
      SizeOfFile=$(stat -c %s $filenameIPCRESS);
      echo $filenameIPCRESS $SizeOfFile >> Size_Of_File_to_excel.txt;
done
```

Moving output files into one folder

```
for i in {genome_numbers};
do
        mv "$i"_ ipcressGeneName.fasta /path/output_files_Name/.
done
```

To delete the first line of the ipcress file to enable recognition for aligning:

```
for i in {genome_numbers};
do
    File=$i" _ipcress_GeneName.fasta";
    newFile=$i"_ ipcress_GeneName_new.fasta";
    echo $File;
    echo $newFile;
    echo -e "$(sed '1d' $File)\n" > $newFile
done
```

Add the isolate number to the sequence to identify (replaces first line with isolate number and gene name – needs to be hardcoded)

```
for i in {genome_numbers};
do
    FirstFile=$i"_ ipcress_GeneName_new.fasta";
    GeneName=$i"_iga";
    echo -e "$(sed '1c\ > '$GeneName'' $FirstFile)\n" >$FirstFile
done
```

Concatenate the sequences to be aligned.

```
cat *_new.fasta >> GeneName_subset_sequences.fasta
```

Manipulating fasta files to copy certain alleles to separate txt file.

```
for i in ???-genome.ctg.fasta.txt
do
    x=$(sed -E 's/-genome.ctg.fasta //' <<< "$i")
    sed –n ’17,31’p $i > $x.txt;
done
```

PROKKA – annotations of assembled genomes.

```
for i in {sequence of isolate numbers}; do
```

```
        mkdir $i ;
        cp /path/$i_contigs.fasta $i/.;
done
```

prokka.pbs

```
!/bin/bash
==========================================
PBS -l nodes=1:ppn=16
PBS -l walltime=00:30:00
Change to directory from which job was submitted
cd $PBS_O_WORKDIR
 load software module
module load prokka
echo $PBS_ARRAYID
fasta_file=$PBS_ARRAYID"-genome.ctg.fasta"
echo "processing files $fasta_file "
if [ -f /path/$fasta_file ]
then
            if file exist - run script
        prokka --outdir /path/$PBS_ARRAYID --prefix $PBS_ARRAYID $fasta_file;
        prokka --outdir $PBS_ARRAYID --prefix $PBS_ARRAYID --cpus 16/path/$fasta_file;
else


 if file no found - print error message and exit
        echo 'File '$fastq_file' not found, quit'
exit
fi
```

Identifying genes within annotations

```
module load exonerate
exonerate – bestn 1 GenBankSeq.fasta isolateNumber/isolateNumber.faa
```

Average Nucleotide Identity

Assembled genomes are located in the ani_input_files directory.

```
!/bin/bash
=============================================
PBS -l nodes=1:ppn=1
PBS -l walltime=60:00:00
cd $PBS_O_WORKDIR

load software module
module load pyani

average_nucleotide_identity.py -i ani_input_files -o ani_output_directory -m ANIm-g
```

Pan and core genome – ROARY

```
module load roary/3.4.1
module load mafft

roary -e --mafft -p 8 *.gff
```

Scoary

```
scoary -g gene_presence_absence.csv -t traits.csv -s 12
```

Appendix 2: Patient, date, MLST and subset data

Isolate number	Patient Number	Activity	Collection Date	ST	Chap 5 Subset	Chap 6 wga subset
17	5	Exac	16-Nov-11	176		17
18	5	Exac	16-Nov-11	176		18
19	5	Exac	16-Nov-11	176		
21	5	Exac	16-Nov-11	176		
26	5	Exac	16-Nov-11	176		
30	5	Exac	16-Nov-11	176		
47	9	NonExac	23-Aug-11	409		47
48	9	NonExac	23-Aug-11	409		48
50	9	NonExac	23-Aug-11	409		
54	34	Exac	12-Dec-11	Hh	54	54
56	34	Exac	12-Dec-11	Hh		56
58	11	Exac	16-Jan-12	fucKN		58
59	11	Exac	16-Jan-12	fucKN		59
60	11	NonExac	30-Jan-12	fucKN	60	60
61	11	Exac	16-Jan-12	fucKN		
62	11	Exac	16-Jan-12	fucKN		
63	11	NonExac	30-Jan-12	fucKN		63
64	11	NonExac	30-Jan-12	fucKN		
65	11	NonExac	30-Jan-12	fucKN		
66	11	NonExac	30-Jan-12	fucKN		
67	11	NonExac	30-Jan-12	fucKN		
68	11	NonExac	30-Jan-12	fucKN		
69	11	NonExac	30-Jan-12	fucKN		
70	11	NonExac	30-Jan-12	fucKN		
71	11	Exac	16-Jan-12	fucKN		
73	11	Exac	16-Jan-12	fucKN		
74	11	Exac	16-Jan-12	fucKN		
75	11	Exac	16-Jan-12	fucKN		
77	11	Exac	16-Jan-12	fucKN	77	
78	17	Exac	05-Dec-11	1163		78
79	17	Exac	23-Aug-11	1163	79	79
80	17	Exac	05-Dec-11	1163	80	80
81	17	Exac	23-Aug-11	1163		
82	17	Exac	05-Dec-11	1163		
83	17	Exac	05-Dec-11	1163		
84	17	Exac	05-Dec-11	1163		
85	17	Exac	05-Dec-11	1163		
86	17	Exac	14-Nov-11	1163		86
87	17	Exac	14-Nov-11	1163		87
88	17	Exac	23-Aug-11	1163		
89	17	Exac	14-Nov-11	1163		
90	17	Exac	23-Aug-11	1163		
91	17	Exac	14-Nov-11	2131		

92	17	Exac	14-Nov-11	1163		
93	17	Exac	14-Nov-11	1163		
94	17	NonExac	15-Aug-11	215		94
95	17	NonExac	15-Aug-11	215		95
96	17	NonExac	15-Aug-11	215		
97	17	Exac	23-Aug-11	1163		
98	17	NonExac	15-Aug-11	215		
99	17	Exac	23-Aug-11	1163		
100	17	NonExac	15-Aug-11	215		
101	17	NonExac	15-Aug-11	215		
102	17	Exac	23-Aug-11	1163		102
103	17	Exac	23-Aug-11	1163		
104	17	Exac	23-Aug-11	1163		
105	17	NonExac	15-Aug-11	215		
106	17	NonExac	15-Aug-11	215		
107	17	Exac	14-Nov-11	1163		
108	17	Exac	14-Nov-11	1163		
109	17	Exac	14-Nov-11	1163		
110	17	Exac	23-Aug-11	1163		
112	17	NonExac	15-Aug-11	215		
113	17	Exac	23-Aug-11	1163		
114	17	NonExac	15-Aug-11	215		
115	17	NonExac	15-Aug-11	215		
116	17	Exac	23-Aug-11	1163		
117	17	NonExac	15-Aug-11	215		
118	17	Exac	05-Dec-11	1163		
119	17	Exac	14-Nov-11	1163	119	
120	17	Exac	05-Dec-11	1163		
121	17	Exac	23-Aug-11	1163		
122	17	Exac	23-Aug-11	1163		
123	17	Exac	23-Aug-11	1163		
124	17	Exac	29-Dec-11	14		124
125	17	Exac	29-Dec-11	14		125
126	17	Exac	29-Dec-11	14		
127	17	Exac	29-Dec-11	14		
128	17	Exac	29-Dec-11	14		
129	17	Exac	29-Dec-11	14		
130	17	Exac	29-Dec-11	14	130	
131	17	Exac	29-Dec-11	14		
132	17	Exac	29-Dec-11	14		
133	17	Exac	29-Dec-11	14		
134	17	NonExac	15-Aug-11	215	134	
135	21	Exac	13-Jan-12	1034		135
136	21	Exac	13-Jan-12	1034		136
137	21	Exac	30-Jan-12	11		137
138	21	Exac	30-Jan-12	11		138

139	21	Exac	05-Mar-12	1034	139	139
140	21	Exac	30-Jan-12	11		
141	21	Exac	30-Jan-12	11		
142	21	Exac	13-Jan-12	1034		
143	21	Exac	13-Jan-12	1034		
144	21	Exac	13-Jan-12	1034		
145	21	Exac	30-Jan-12	11		
146	21	Exac	13-Jan-12	1034		
147	21	Exac	05-Mar-12	1034		147
148	21	Exac	05-Mar-12	1034		
149	21	Exac	30-Jan-12	11		
150	21	Exac	05-Mar-12	1034		
151	21	Exac	05-Mar-12	1034		
152	21	Exac	05-Mar-12	1034		
153	21	Exac	13-Jan-12	1034		
154	21	Exac	05-Mar-12	1034		
155	21	Exac	05-Mar-12	1034		
156	21	Exac	05-Mar-12	1034		
157	21	Exac	30-Jan-12	11		
158	21	Exac	13-Jan-12	1034		
159	21	Exac	30-Jan-12	11		
160	21	Exac	30-Jan-12	11		
161	21	Exac	13-Jan-12	1034		
162	21	Exac	30-Jan-12	11		
163	21	NonExac	23-Dec-11	1034		163
164	21	NonExac	23-Dec-11	1034		
165	21	NonExac	23-Dec-11	1034		
166	21	NonExac	23-Dec-11	1034		
167	21	NonExac	23-Dec-11	1034		
168	21	NonExac	23-Dec-11	1034	168	
169	21	NonExac	23-Dec-11	11	169	169
170	21	NonExac	23-Dec-11	1034		
171	21	NonExac	23-Dec-11	1034		
172	21	NonExac	23-Dec-11	1034		
173	21	NonExac	24-Jan-12	1034		
174	21	NonExac	24-Jan-12	1034		
175	21	NonExac	24-Jan-12	1034		
176	21	NonExac	24-Jan-12	1034		
177	21	NonExac	24-Jan-12	1034		
178	21	NonExac	24-Jan-12	1034		
179	21	NonExac	24-Jan-12	1034		
180	21	NonExac	24-Jan-12	1034	180	
181	21	NonExac	24-Jan-12	1034		
182	21	NonExac	24-Jan-12	1034		
183	21	Exac	13-Jan-12	1034	183	
184	25	Exac	28-Nov-11	396	184	184

185	25	Exac	23-Jan-12	503		
186	25	Exac	23-Jan-12	503		
187	25	Exac	23-Jan-12	503		
188	25	Exac	28-Nov-11	396	188	188
189	25	Exac	23-Jan-12	503	189	189
190	25	Exac	28-Nov-11	396	190	
191	25	Exac	23-Jan-12	503		
192	25	Exac	23-Jan-12	503		191
193	25	Exac	28-Nov-11	396		
194	25	Exac	28-Nov-11	396		
195	25	Exac	28-Nov-11	396		
196	25	Exac	23-Jan-12	503		
197	25	Exac	23-Jan-12	503		
198	25	Exac	23-Jan-12	503		
199	25	Exac	28-Nov-11	396		
200	25	Exac	28-Nov-11	396		
201	25	Exac	28-Nov-11	396		
202	25	Exac	23-Jan-12	503		
203	25	Exac	28-Nov-11	396		
204	27	NonExac	07-Dec-11	819		204
205	27	NonExac	07-Dec-11	819	205	205
206	3	Exac	20-Jan-12	57		206
207	3	Exac	20-Jan-12	57		207
208	3	Exac	20-Jan-12	57		
209	3	Exac	20-Jan-12	57		
210	3	Exac	20-Jan-12	57		
211	3	Exac	20-Jan-12	57		
212	3	Exac	20-Jan-12	57		
213	3	Exac	20-Jan-12	57		
214	3	Exac	20-Jan-12	57	214	
215	3	Exac	20-Jan-12	57		
216	30	NonExac	15-Nov-11	142	216	216
217	30	NonExac	15-Nov-11	142		217
218	30	NonExac	15-Nov-11	142		
219	30	NonExac	15-Nov-11	142		
220	30	NonExac	15-Nov-11	142		
221	30	NonExac	15-Nov-11	142		
222	30	NonExac	15-Nov-11	142		
223	30	NonExac	15-Nov-11	142		
224	30	NonExac	15-Nov-11	142		
225	30	NonExac	15-Nov-11	142		
226	32	NonExac	19-Dec-11	1314		226
227	32	NonExac	19-Dec-11	1314		227
228	32	NonExac	19-Dec-11	1314		
229	32	NonExac	19-Dec-11	1314	229	
230	32	NonExac	19-Dec-11	1314		

231	32	NonExac	19-Dec-11	1314		
232	32	NonExac	19-Dec-11	1314		
233	32	NonExac	19-Dec-11	1314		
234	32	NonExac	19-Dec-11	1314		
235	32	NonExac	19-Dec-11	1314		
236	38	Exac	19-Dec-11	356		236
237	38	Exac	19-Dec-11	356		237
238	38	Exac	19-Dec-11	356		
239	38	Exac	19-Dec-11	356		
240	38	Exac	19-Dec-11	356		
241	38	Exac	19-Dec-11	356		
242	38	Exac	19-Dec-11	356		
243	38	Exac	19-Dec-11	356		
244	38	Exac	19-Dec-11	356		
245	38	Exac	19-Dec-11	356	245	
246	5	Exac	03-Jan-12	176		246
247	5	Exac	03-Jan-12	176		247
248	9	NonExac	22-Nov-11	409		248
249	9	NonExac	22-Nov-11	409		249
250	9	NonExac	23-Aug-11	409		
251	9	NonExac	22-Nov-11	409		
252	9	NonExac	22-Nov-11	409		
253	9	NonExac	22-Nov-11	409		
254	9	NonExac	22-Nov-11	409		
255	9	NonExac	23-Aug-11	409		
256	9	NonExac	23-Aug-11	409		
257	9	NonExac	22-Nov-11	409	257	
258	9	NonExac	23-Aug-11	409		
259	9	NonExac	22-Nov-11	409		
260	9	NonExac	23-Aug-11	409		
261	9	NonExac	23-Aug-11	409		
262	9	NonExac	23-Aug-11	409		
263	9	NonExac	22-Nov-11	409		
264	9	NonExac	22-Nov-11	409		
265	17	NonExac	15-Aug-11	215		
266	21	NonExac	26-Aug-11	11		266
267	21	NonExac	26-Aug-11	11	267	267
268	21	NonExac	26-Aug-11	11		
269	21	NonExac	26-Aug-11	11		
270	21	NonExac	26-Aug-11	11		
271	21	NonExac	26-Aug-11	11		
272	21	NonExac	26-Aug-11	11		
273	21	NonExac	26-Aug-11	11		
274	21	NonExac	26-Aug-11	11		
275	21	NonExac	26-Aug-11	11		
276	21	NonExac	26-Aug-11	11		

277	21	NonExac	26-Aug-11	11		
278	21	NonExac	26-Aug-11	11		
279	21	NonExac	26-Aug-11	11		
280	21	NonExac	26-Aug-11	11		
281	24	NonExac	02-Sep-11	196	281	
282	25	NonExac	06-Sep-11	503	282	282
283	25	NonExac	06-Sep-11	503		283
284	25	NonExac	06-Sep-11	503		
285	25	NonExac	06-Sep-11	503		
286	25	NonExac	06-Sep-11	503		
287	25	NonExac	06-Sep-11	503		
288	25	NonExac	06-Sep-11	503		
289	25	NonExac	06-Sep-11	503		
290	25	NonExac	06-Sep-11	503		
291	25	NonExac	06-Sep-11	503		
292	25	NonExac	06-Sep-11	503		
293	25	NonExac	06-Sep-11	503		
294	25	NonExac	06-Sep-11	503		
295	25	NonExac	06-Sep-11	503		
296	25	NonExac	06-Sep-11	503		
297	27	NonExac	07-Sep-11	819	297	297
298	30	Exac	06-Jan-12	201	298	298
299	30	Exac	07-Feb-12	142	299	299
300	30	Exac	07-Feb-12	142		300
301	30	Exac	07-Feb-12	142		302
302	30	Exac	06-Jan-12	201		
303	30	Exac	06-Jan-12	201	303	
304	30	Exac	06-Jan-12	201		
305	30	Exac	07-Feb-12	142		
306	30	Exac	07-Feb-12	142		
307	30	Exac	07-Feb-12	142		
308	30	Exac	07-Feb-12	142		
309	30	Exac	06-Jan-12	201		
310	30	Exac	07-Feb-12	142		
311	30	Exac	07-Feb-12	142		
312	30	Exac	06-Jan-12	201		
313	30	Exac	06-Jan-12	201		
314	30	Exac	06-Jan-12	201		
315	30	Exac	07-Feb-12	142		
316	30	Exac	06-Jan-12	201		
317	30	Exac	06-Jan-12	201		
318	5	NonExac	14-Sep-11	176		318
319	5	NonExac	14-Sep-11	176		319
320	5	NonExac	14-Sep-11	176		
321	5	NonExac	14-Sep-11	176	321	
322	5	NonExac	14-Sep-11	176		

323	5	NonExac	14-Sep-11	176		
324	5	NonExac	14-Sep-11	176		
325	5	NonExac	14-Sep-11	176		
326	5	Exac	03-Oct-11	176	326	326
327	5	NonExac	14-Sep-11	176		
328	5	NonExac	14-Sep-11	176		
329	5	NonExac	14-Sep-11	176		
330	5	NonExac	14-Sep-11	176		
331	5	NonExac	14-Sep-11	176		
332	5	NonExac	14-Sep-11	176		
333	5	NonExac	14-Sep-11	176		
334	17	Exac	12-Oct-11	1163		334
335	17	Exac	12-Oct-11	1163		335
336	17	Exac	12-Oct-11	1163		
337	17	Exac	12-Oct-11	1163		
338	17	Exac	12-Oct-11	1163		
339	17	Exac	12-Oct-11	1163		
340	21	Exac	14-Sep-11	1034	340	340
341	21	Exac	14-Sep-11	1034		341
342	25	Exac	05-Oct-11	503		
343	25	Exac	05-Oct-11	503		
344	25	Exac	05-Oct-11	503		
345	25	Exac	05-Oct-11	503		
346	25	Exac	05-Oct-11	503		
347	25	Exac	05-Oct-11	503		
348	25	Exac	05-Oct-11	503		348
349	25	Exac	05-Oct-11	503		349
350	25	Exac	05-Oct-11	503		
351	25	Exac	05-Oct-11	503		
352	25	NonExac	04-Nov-11	503	352	352
353	25	NonExac	04-Nov-11	503		353
354	25	NonExac	04-Nov-11	503		
355	25	NonExac	04-Nov-11	503		
356	25	NonExac	04-Nov-11	503		
357	25	NonExac	04-Nov-11	503		
358	25	NonExac	04-Nov-11	503		
359	27	NonExac	07-Nov-11	819	359	359
360	27	NonExac	07-Nov-11	819		360
361	27	NonExac	07-Nov-11	819		
362	27	NonExac	07-Nov-11	819		
363	27	NonExac	07-Nov-11	819		
364	27	NonExac	07-Nov-11	819		
365	27	NonExac	07-Nov-11	819		
366	27	NonExac	07-Nov-11	819		
367	27	NonExac	07-Nov-11	819		
368	32	NonExac	26-Oct-11	513	368	368

369	38	NonExac	06-Oct-11	356		369
370	38	NonExac	06-Oct-11	356	370	370
371	38	NonExac	06-Oct-11	356		
372	38	NonExac	06-Oct-11	356		
373	38	NonExac	06-Oct-11	356		
374	38	NonExac	06-Oct-11	356		
375	38	NonExac	06-Oct-11	356		
376	38	NonExac	06-Oct-11	356		
377	29	NonExac	13-Oct-11	311		377
378	29	NonExac	13-Sep-11	311		378
379	29	NonExac	13-Sep-11	311		379
380	29	NonExac	13-Sep-11	311		
381	29	NonExac	13-Sep-11	311		
382	29	NonExac	13-Sep-11	311		
383	29	NonExac	13-Sep-11	311		
384	29	NonExac	13-Sep-11	311		
385	29	NonExac	13-Sep-11	311		
386	29	NonExac	13-Sep-11	311		
387	29	NonExac	13-Sep-11	311		
388	29	NonExac	13-Sep-11	311		
389	29	NonExac	13-Oct-11	311		389
390	29	NonExac	13-Sep-11	311		
391	29	NonExac	13-Sep-11	311		
392	29	NonExac	13-Oct-11	311		
393	29	NonExac	13-Oct-11	311		
394	29	NonExac	13-Oct-11	311		
396	29	NonExac	13-Sep-11	311		
397	29	Exac	04-Nov-11	311		397
398	29	Exac	04-Nov-11	311		398
399	29	Exac	04-Nov-11	311		
400	29	Exac	04-Nov-11	311		
401	29	Exac	04-Nov-11	311		
402	29	Exac	04-Nov-11	311		
403	29	Exac	04-Nov-11	311		
404	29	Exac	04-Nov-11	311		
405	29	Exac	04-Nov-11	311		
406	29	Exac	04-Nov-11	311		
407	21	Exac	14-Sep-11	1034		
408	21	Exac	14-Sep-11	1034		
409	21	Exac	14-Sep-11	1034		
410	21	Exac	14-Sep-11	1034		
411	21	Exac	14-Sep-11	1034		
412	21	Exac	14-Sep-11	1034		
413	21	Exac	14-Sep-11	1034		
414	21	Exac	14-Sep-11	1034		
415	21	Exac	14-Sep-11	1034		

416	21	Exac	14-Sep-11	1034		
417	21	Exac	14-Sep-11	1034		
418	21	Exac	14-Sep-11	1034		
420	9	NonExac	01-Mar-12	409		420
421	38	NonExac	01-Dec-11	356	421	421
422	42	Exac	13-Dec-11	12	422	422
423	3	Exac	21-Mar-12	836		423
424	5	NonExac	21-Feb-12	357		424
425	27	NonExac	06-Jan-12	103		425
426	11	Exac	26-Apr-12	Hh	426	426
427	32	NonExac	21-Feb-12	Hh	427	427
428	34	NonExac	29-Mar-12	Hh	428	428
429	3	NonExac	06-Jul-11	201		429
430	30	Exac	02-Mar-12	142		430
431	3	NonExac	06-Jul-11	201		431
432	34	NonExac	25-Nov-11	Hh		432
433	38	NonExac	08-Feb-12	Hh	433	433
434	32	NonExac	21-Nov-11	513	434	434
435	5	Exac	29-Feb-12	357		435
436	3	NonExac	07-Mar-12	836		436
437	3	NonExac	07-Mar-12	836	437	437
438	3	NonExac	07-Mar-12	836		
439	3	NonExac	06-Jul-11	201	439	
440	3	NonExac	06-Jul-11	201		
442	3	NonExac	07-Mar-12	836		
443	3	NonExac	07-Mar-12	836		
444	3	NonExac	07-Mar-12	836		
445	3	NonExac	06-Jul-11	201		
446	3	NonExac	06-Jul-11	201		
447	3	NonExac	06-Jul-11	201		
448	3	NonExac	06-Jul-11	201		
449	3	NonExac	07-Mar-12	836		
450	3	NonExac	07-Mar-12	836		
451	3	NonExac	06-Jul-11	201		
452	3	NonExac	06-Jul-11	201		
453	3	Exac	21-Mar-12	836		453
454	3	Exac	21-Mar-12	836	454	
455	3	NonExac	06-Jul-11	201		
456	3	NonExac	06-Jul-11	201		
457	3	Exac	21-Mar-12	836		
458	3	NonExac	06-Jul-11	201		
459	3	NonExac	06-Jul-11	201		
460	3	NonExac	06-Jul-11	201		
461	3	NonExac	07-Mar-12	836		
462	5	Exac	29-Feb-12	357	462	462
463	5	Exac	29-Feb-12	357		

464	5	Exac	29-Feb-12	357		
465	5	Exac	29-Feb-12	357		
466	5	Exac	29-Feb-12	357		
467	5	Exac	29-Feb-12	357		
468	5	Exac	29-Feb-12	357		
469	5	Exac	29-Feb-12	357		
470	5	Exac	29-Feb-12	357		
471	5	NonExac	21-Mar-12	357		471
472	5	NonExac	21-Mar-12	357		472
473	5	NonExac	21-Feb-12	357		473
474	5	NonExac	21-Feb-12	357		
475	5	NonExac	21-Feb-12	357		
476	5	Exac	15-Dec-11	176	476	476
477	9	NonExac	23-Aug-11	409		
478	9	NonExac	23-Aug-11	409		
479	9	NonExac	23-Aug-11	409	479	
480	9	NonExac	23-Aug-11	409		
481	9	NonExac	23-Aug-11	409		
482	9	NonExac	27-Apr-12	409	482	482
483	9	NonExac	27-Apr-12	409		483
484	9	NonExac	01-Mar-12	409		484
485	9	NonExac	01-Mar-12	409	485	
486	9	NonExac	29-Dec-11	409		486
487	11	Exac	26-Apr-12	fucKN		487
488	11	Exac	26-Apr-12	Hh	488	488
489	11	NonExac	26-Mar-12	fucKN		489
490	11	NonExac	24-Feb-12	fucKN		490
491	11	Exac	26-Apr-12	fucKN		491
492	11	Exac	26-Apr-12	fucKN		
493	11	Exac	26-Apr-12	fucKN		
494	11	Exac	26-Apr-12	Hh	494	494
495	11	Exac	26-Apr-12	fucKN		
496	11	Exac	26-Apr-12	fucKN	496	
497	11	Exac	26-Apr-12	Hh	497	497
498	17	Exac	07-Sep-11	1163	498	498
499	17	NonExac	14-Dec-11	1163		499
500	21	NonExac	27-Mar-12	1034		
501	21	NonExac	27-Mar-12	1034	501	
502	24	NonExac	30-Nov-11	196		502
503	24	NonExac	30-Nov-11	196		
504	24	NonExac	30-Nov-11	196		
505	24	NonExac	30-Nov-11	196		
506	24	NonExac	30-Nov-11	196		
507	24	NonExac	30-Nov-11	196		
508	24	NonExac	30-Nov-11	196		
509	24	NonExac	30-Nov-11	196	509	509

510	24	NonExac	30-Nov-11	196		
511	24	NonExac	30-Nov-11	196		
512	27	NonExac	06-Jan-12	103		512
513	27	NonExac	07-Dec-11	Hh		513
514	27	NonExac	06-Jan-12	103	514	
515	30	Exac	02-Mar-12	142	515	515
516	30	Exac	02-Mar-12	142		
517	30	Exac	02-Mar-12	142		
518	30	Exac	02-Mar-12	142		
519	30	Exac	02-Mar-12	142		
520	30	Exac	02-Mar-12	142		
521	30	Exac	02-Mar-12	142		
522	30	Exac	02-Mar-12	142		
523	30	Exac	02-Mar-12	142		
524	30	NonExac	04-Apr-12	142		524
525	32	NonExac	28-Mar-12	34	525	525
526	32	NonExac	28-Mar-12	34		526
527	32	NonExac	28-Mar-12	34		
528	32	NonExac	21-Nov-11	513		528
529	32	NonExac	21-Feb-12	1314	529	529
530	32	NonExac	21-Nov-11	513		
531	32	NonExac	21-Feb-12	1314	531	
532	32	NonExac	21-Nov-11	513		
533	32	NonExac	21-Nov-11	513	533	
534	32	NonExac	28-Mar-12	34		
535	32	NonExac	18-Jan-12	1314		535
536	32	NonExac	18-Jan-12	1314		
537	32	NonExac	21-Nov-11	513		
538	32	NonExac	21-Feb-12	1314		
539	32	NonExac	21-Nov-11	513		
540	32	NonExac	01-May-12	34	540	540
541	32	NonExac	21-Nov-11	513		
542	32	NonExac	21-Nov-11	513		
543	32	NonExac	21-Nov-11	513		
544	32	NonExac	28-Mar-12	34		
545	33	NonExac	21-May-12	196		545
546	33	NonExac	21-May-12	196		546
547	33	NonExac	21-May-12	196		548
548	34	NonExac	25-Nov-11	Hh		549
549	34	NonExac	29-Mar-12	Hh	549	
550	34	NonExac	25-Nov-11	Hh		550
551	34	NonExac	29-Mar-12	Hh	551	551
552	34	NonExac	25-Nov-11	Hh	552	552
553	34	NonExac	25-Nov-11	Hh	553	553
554	34	NonExac	29-Mar-12	Hh	554	554
555	34	NonExac	29-Mar-12	Hh	555	555

556	34	NonExac	29-Mar-12	Hh	556	556
557	34	NonExac	29-Mar-12	Hh	557	557
558	34	NonExac	29-Mar-12	Hh	558	558
559	34	NonExac	29-Mar-12	Hh	559	559
560	38	NonExac	01-Dec-11	356	560	560
561	38	NonExac	01-Dec-11	356		
562	38	NonExac	01-Dec-11	356		
563	38	NonExac	08-Feb-12	356		563
564	38	NonExac	08-Feb-12	356	564	
565	38	NonExac	08-Feb-12	356		
566	38	NonExac	01-Dec-11	356		
567	38	NonExac	01-Dec-11	356		
568	38	NonExac	08-Feb-12	356		
569	38	NonExac	08-Feb-12	356		
570	38	NonExac	08-Feb-12	356	570	
571	38	NonExac	01-Dec-11	356		
572	38	NonExac	01-Dec-11	356		
573	38	NonExac	01-Dec-11	356		
574	38	NonExac	04-Jan-12	356		574
575	38	NonExac	01-Dec-11	356		
576	42	Exac	13-Dec-11	12	576	576
577	42	Exac	13-Dec-11	12		
578	42	Exac	13-Dec-11	12		
579	42	Exac	13-Dec-11	12		
580	42	Exac	13-Dec-11	12		
582	42	Exac	13-Dec-11	12		
583	42	Exac	13-Dec-11	12		
584	3	NonExac	07-Mar-12	836		
585	3	NonExac	07-Mar-12	836		
586	3	NonExac	07-Mar-12	836		
587	32	NonExac	01-May-12	34		587
588	3	NonExac	07-Mar-12	836		
589	3	NonExac	07-Mar-12	836		
590	3	NonExac	07-Mar-12	836		
591	5	NonExac	17-Apr-12	357		591
592	5	Exac	16-Nov-11	176	592	
593	32	NonExac	01-May-12	34		
594	33	NonExac	21-May-12	196		
595	9	NonExac	27-Apr-12	409	595	
596	38	NonExac	04-Jan-12	356		596
597	11	NonExac	26-Mar-12	fucKN	597	597
598	5	NonExac	17-Apr-12	357	598	598
599	5	NonExac	17-Apr-12	357		
600	21	NonExac	26-Apr-12	11		600
601	5	Exac	03-Jan-12	176	601	
602	5	NonExac	21-Feb-12	357		

603	3	Exac	21-Mar-12	836		
604	3	Exac	21-Mar-12	836		
605	5	NonExac	21-Feb-12	357	605	
606	5	Exac	03-Jan-12	176		
607	21	NonExac	26-Apr-12	11	607	607
608	5	Exac	15-Dec-11	176		608
609	32	NonExac	24-May-12	34	609	609
610	17	NonExac	14-Dec-11	1163		610
611	30	NonExac	04-Apr-12	142		611
612	25	NonExac	09-May-12	57		612
613	9	NonExac	29-Dec-11	409		613
614	9	NonExac	29-Dec-11	409		
615	25	NonExac	09-May-12	57		615
616	30	NonExac	04-Apr-12	142		
617	32	NonExac	01-May-12	34		
618	17	NonExac	14-Dec-11	1163		
619	5	Exac	15-Dec-11	176		
620	11	Exac	16-May-12	Hh		620
621	29	NonExac	05-Dec-11	311		621
622	32	NonExac	18-Jan-12	1314		
623	21	NonExac	27-Mar-12	1034		
624	5	NonExac	21-Mar-12	357		
625	33	NonExac	20-Apr-12	196		625
626	11	NonExac	24-Feb-12	fucKN		626
627	5	Exac	15-Dec-11	176		
628	17	NonExac	14-Dec-11	1163		
629	29	NonExac	05-Dec-11	311	629	629
630	32	NonExac	24-May-12	34		630
631	30	NonExac	04-Apr-12	142		
632	21	NonExac	26-Apr-12	11		
633	32	NonExac	18-Jan-12	1314	633	633
634	11	NonExac	24-Feb-12	fucKN		
635	33	NonExac	20-Apr-12	196		635
636	5	NonExac	21-Mar-12	357		
637	5	NonExac	17-Apr-12	357		
638	5	Exac	03-Jan-12	176		
639	21	NonExac	26-Apr-12	11		
640	5	NonExac	17-Apr-12	357		
641	5	NonExac	21-Mar-12	357	641	
642	21	NonExac	27-Mar-12	1034		
643	32	NonExac	18-Jan-12	1314		
644	11	NonExac	24-Feb-12	fucKN		
645	33	NonExac	20-Apr-12	196		
646	3	Exac	21-Mar-12	836		
647	5	NonExac	21-Feb-12	357		
648	11	NonExac	24-Feb-12	fucKN		

649	33	NonExac	20-Apr-12	196	
650	5	NonExac	21-Feb-12	357	
651	33	NonExac	20-Apr-12	196	651
652	3	Exac	21-Mar-12	836	
653	25	NonExac	09-May-12	57	653
654	9	NonExac	29-Dec-11	409	654
655	30	NonExac	04-Apr-12	142	
656	32	NonExac	18-Jan-12	1314	
657	11	NonExac	24-Feb-12	fucKN	
658	33	NonExac	20-Apr-12	196	
659	9	NonExac	27-Apr-12	409	
660	5	NonExac	21-Mar-12	357	
661	21	NonExac	27-Mar-12	1034	
662	5	NonExac	17-Apr-12	357	
663	5	NonExac	21-Feb-12	357	663
664	11	NonExac	24-Feb-12	fucKN	
665	33	NonExac	20-Apr-12	196	
666	3	Exac	21-Mar-12	836	
667	21	NonExac	26-Apr-12	11	
668	5	NonExac	17-Apr-12	357	
669	21	NonExac	26-Apr-12	11	
670	5	NonExac	17-Apr-12	357	
671	32	NonExac	18-Jan-12	1314	
672	11	NonExac	24-Feb-12	fucKN	
673	33	NonExac	20-Apr-12	196	
674	5	NonExac	21-Mar-12	357	
675	21	NonExac	27-Mar-12	1034	
676	32	NonExac	01-May-12	34	
677	33	NonExac	21-May-12	196	
678	9	NonExac	29-Dec-11	409	
679	38	NonExac	04-Jan-12	356	
681	5	Exac	15-Dec-11	176	
682	17	NonExac	14-Dec-11	1163	
683	21	NonExac	26-Apr-12	11	
684	5	NonExac	17-Apr-12	357	
685	5	NonExac	21-Feb-12	357	
686	42	Exac	13-Dec-11	12	
687	3	Exac	21-Mar-12	836	
688	21	NonExac	26-Apr-12	11	
689	5	NonExac	17-Apr-12	357	
690	11	NonExac	26-Mar-12	fucKN	
691	32	NonExac	01-May-12	34	
692	33	NonExac	21-May-12	196	
693	38	NonExac	04-Jan-12	356	693
694	9	NonExac	27-Apr-12	409	
695	5	Exac	16-Nov-11	176	

696	32	NonExac	01-May-12	34		
697	33	NonExac	21-May-12	196		
698	38	NonExac	04-Jan-12	356	698	
699	17	NonExac	14-Dec-11	1163	699	
700	9	NonExac	27-Apr-12	409		
701	11	NonExac	26-Mar-12	fucKN		
702	5	Exac	16-Nov-11	176		
703	32	NonExac	18-Jan-12	1314		
704	11	NonExac	24-Feb-12	fucKN		
705	33	NonExac	20-Apr-12	196		
706	9	NonExac	27-Apr-12	409		
707	5	NonExac	21-Mar-12	357		
708	21	NonExac	27-Mar-12	1034		
709	9	NonExac	27-Apr-12	409		
710	33	NonExac	21-May-12	196		
711	32	NonExac	18-Jan-12	1314		
712	5	NonExac	21-Mar-12	357		
713	11	NonExac	26-Mar-12	fucKN		
714	21	NonExac	27-Mar-12	1034		
715	32	NonExac	01-May-12	34		
716	25	NonExac	09-May-12	57		
717	30	NonExac	04-Apr-12	142		
718	9	NonExac	29-Dec-11	409	718	
719	5	Exac	15-Dec-11	176		
720	17	NonExac	14-Dec-11	1163		
721	32	NonExac	01-May-12	34		
722	33	NonExac	21-May-12	196		
723	38	NonExac	04-Jan-12	356		
724	9	NonExac	27-Apr-12	409		
725	5	Exac	16-Nov-11	176		
726	11	NonExac	26-Mar-12	fucKN		
727	21	NonExac	27-Mar-12	1034		
728	9	NonExac	29-Dec-11	409		
729	32	NonExac	18-Jan-12	1314		
730	11	NonExac	24-Feb-12	fucKN	730	
731	33	NonExac	20-Apr-12	196		
732	5	NonExac	21-Mar-12	357		
733	3	NonExac	07-Mar-12	836		
734	32	NonExac	28-Mar-12	34		
735	25	NonExac	09-May-12	57		
736	32	NonExac	24-May-12	34	736	
737	9	NonExac	29-Dec-11	409		
738	5	Exac	15-Dec-11	176		
739	17	NonExac	14-Dec-11	1163		
740	3	Exac	20-Jan-12	57		740
741	11	NonExac	24-Feb-12	fucKN		

742	5	Exac	29-Feb-12	357	
743	32	NonExac	01-May-12	34	
744	3	Exac	20-Jan-12	57	744
745	11	NonExac	24-Feb-12	fucKN	
746	5	Exac	29-Feb-12	357	
747	32	NonExac	01-May-12	34	
748	3	Exac	20-Jan-12	57	
749	11	NonExac	24-Feb-12	fucKN	
750	5	Exac	29-Feb-12	357	
751	32	NonExac	01-May-12	34	
752	3	Exac	20-Jan-12	57	
753	11	NonExac	24-Feb-12	fucKN	
754	5	Exac	29-Feb-12	357	
755	32	NonExac	01-May-12	34	
756	3	NonExac	07-Mar-12	836	
757	32	NonExac	28-Mar-12	34	
758	3	NonExac	07-Mar-12	836	
759	32	NonExac	28-Mar-12	34	
760	3	Exac	20-Jan-12	57	
761	11	NonExac	24-Feb-12	fucKN	
762	5	Exac	29-Feb-12	357	
763	32	NonExac	01-May-12	34	
764	3	Exac	20-Jan-12	57	
765	11	NonExac	24-Feb-12	fucKN	
766	5	Exac	29-Feb-12	357	
767	32	NonExac	01-May-12	34	
768	3	NonExac	07-Mar-12	836	
769	32	NonExac	28-Mar-12	34	
770	3	Exac	20-Jan-12	57	
771	11	NonExac	24-Feb-12	fucKN	
772	5	Exac	29-Feb-12	357	
773	32	NonExac	01-May-12	34	
774	3	Exac	20-Jan-12	57	
775	11	NonExac	24-Feb-12	fucKN	
776	5	Exac	29-Feb-12	357	
777	32	NonExac	28-Mar-12	34	
778	3	Exac	20-Jan-12	57	
779	11	NonExac	24-Feb-12	fucKN	
780	5	Exac	29-Feb-12	357	
781	32	NonExac	01-May-12	34	
782	3	Exac	20-Jan-12	57	
783	11	NonExac	24-Feb-12	fucKN	
784	5	Exac	29-Feb-12	357	
785	32	NonExac	01-May-12	34	
786	25	NonExac	09-May-12	57	
787	30	NonExac	04-Apr-12	142	

788	11	Exac	16-May-12	Hh		788
789	29	NonExac	05-Dec-11	311		
790	32	NonExac	01-May-12	34		
791	25	NonExac	09-May-12	57		
792	9	NonExac	29-Dec-11	409		
793	5	Exac	15-Dec-11	176		
794	17	NonExac	14-Dec-11	1163		
795	30	NonExac	04-Apr-12	142		
796	11	Exac	16-May-12	Hh	796	796
797	32	NonExac	24-May-12	34		
798	5	Exac	15-Dec-11	176		
799	17	NonExac	14-Dec-11	1163		
800	27	NonExac	06-Oct-11	103	800	800
801	27	NonExac	06-Oct-11	103		801
802	30	Exac	13-Oct-11	142	802	802
803	30	Exac	13-Oct-11	142		803
804	30	Exac	13-Oct-11	142		
805	30	Exac	13-Oct-11	142		
806	30	Exac	13-Oct-11	142		
807	30	Exac	13-Oct-11	142		
808	30	Exac	13-Oct-11	142		
809	30	Exac	13-Oct-11	142		
810	30	Exac	13-Oct-11	142		
811	30	Exac	13-Oct-11	142		
812	33	Exac	29-Sep-11	1034		812
813	33	Exac	29-Sep-11	1034		813
814	33	Exac	29-Sep-11	1034		
815	33	Exac	29-Sep-11	1034		
816	33	Exac	29-Sep-11	1034	816	
817	33	Exac	29-Sep-11	353	817	817
818	33	Exac	29-Sep-11	353		818
819	33	Exac	29-Sep-11	1034		
820	33	Exac	29-Sep-11	1034		
821	33	Exac	29-Sep-11	1034		
823	29	NonExac	13-Sep-11	311	823	
824	21	Exac	14-Sep-11	1034		
825	21	Exac	14-Sep-11	1034		
826	21	Exac	14-Sep-11	1034		
827	21	Exac	14-Sep-11	1034		
828	21	Exac	14-Sep-11	1034		
829	21	Exac	14-Sep-11	1034		
830	21	Exac	14-Sep-11	1034		
831	21	Exac	14-Sep-11	1034		
832	21	Exac	14-Sep-11	1034		
833	21	Exac	14-Sep-11	1034		
834	21	Exac	14-Sep-11	1034		

835	21	Exac	14-Sep-11	1034		
836	38	NonExac	03-Nov-11	356	836	836
837	38	NonExac	03-Nov-11	356		837
838	38	NonExac	03-Nov-11	356		
839	38	NonExac	03-Nov-11	356		
840	38	NonExac	03-Nov-11	356		
841	38	NonExac	03-Nov-11	356		
842	38	NonExac	03-Nov-11	356		
843	38	NonExac	03-Nov-11	356		
844	38	NonExac	03-Nov-11	356		
845	38	NonExac	03-Nov-11	356		
846	9	NonExac	22-Sep-11	409	846	846
847	9	NonExac	22-Sep-11	409		847
848	9	NonExac	22-Sep-11	409		
849	9	NonExac	22-Sep-11	409		
850	9	NonExac	22-Sep-11	409		
851	9	NonExac	22-Sep-11	409		
852	9	NonExac	22-Sep-11	409		
853	9	NonExac	22-Sep-11	409		
854	9	NonExac	22-Sep-11	409		
855	24	NonExac	07-Sep-11	196	855	
856	24	NonExac	07-Sep-11	925		856
857	17	Exac	07-Sep-11	1163		857
858	29	NonExac	05-Dec-11	311		
859	24	NonExac	07-Sep-11	925		
860	5	Exac	03-Jan-12	176		
861	24	NonExac	07-Sep-11	196		
862	5	Exac	03-Jan-12	176		
863	24	NonExac	07-Sep-11	196		
864	24	NonExac	07-Sep-11	925		
865	5	Exac	03-Jan-12	176		
866	5	Exac	03-Jan-12	176		
867	17	Exac	07-Sep-11	1163		
868	24	NonExac	07-Sep-11	925	868	
869	24	NonExac	07-Sep-11	196		
870	17	Exac	07-Sep-11	1163		
871	24	NonExac	07-Sep-11	925		871
872	17	Exac	07-Sep-11	1163		
873	17	Exac	07-Sep-11	1163		
874	17	Exac	07-Sep-11	1163		
875	24	NonExac	07-Sep-11	196		875
876	17	Exac	07-Sep-11	1163		
877	24	NonExac	07-Sep-11	925		
878	24	NonExac	07-Sep-11	196		878
879	24	NonExac	07-Sep-11	925		
880	24	NonExac	07-Sep-11	196		

881	33	NonExac	23-Sep-11	1034		881
882	33	NonExac	23-Sep-11	1034	882	882
883	33	NonExac	23-Sep-11	1034		
884	33	NonExac	23-Sep-11	1034		
885	33	NonExac	23-Sep-11	1034		
886	33	NonExac	23-Sep-11	1034		
887	33	NonExac	23-Sep-11	1034		
888	33	NonExac	23-Sep-11	1034		
889	33	NonExac	23-Sep-11	1034		
890	33	NonExac	23-Sep-11	1034		
891	5	Exac	03-Jan-12	176		
892	21	NonExac	26-Apr-12	11		
893	11	Exac	16-May-12	Hh	893	893
894	27	NonExac	06-Jan-12	Hh	894	894
895	11	Exac	16-May-12	Hh		895
896	32	NonExac	24-May-12	34		
897	29	NonExac	05-Dec-11	311		
898	25	NonExac	09-May-12	57		
899	32	NonExac	24-May-12	34		
900	29	NonExac	05-Dec-11	311		
901	11	NonExac	26-Mar-12	fucKN		
902	32	NonExac	24-May-12	34		
903	29	NonExac	05-Dec-11	311		
904	11	NonExac	26-Mar-12	fucKN		
905	32	NonExac	24-May-12	34		
906	29	NonExac	05-Dec-11	311		
907	11	NonExac	26-Mar-12	fucKN		
908	11	Exac	16-May-12	Hh		908
909	32	NonExac	24-May-12	34		
910	29	NonExac	05-Dec-11	311		
911	25	NonExac	09-May-12	57		
912	32	NonExac	24-May-12	34		
913	29	NonExac	05-Dec-11	311		
914	30	NonExac	04-Apr-12	142		
915	32	NonExac	01-May-12	34		
916	33	NonExac	21-May-12	196		
917	38	NonExac	04-Jan-12	356		
918	11	NonExac	26-Mar-12	fucKN		
919	29	NonExac	18-Jun-12	311		919
920	21	NonExac	21-Aug-12	1034		920
921	32	NonExac	25-Sep-12	154		921
922	33	Exac	21-Sep-12	196		922
923	32	Exac	24-Jul-12	34		923
924	29	NonExac	13-Apr-12	311		924
925	25	NonExac	06-Sep-12	57		925
926	25	NonExac	30-May-12	57		926

927	9	NonExac	29-May-12	409	927
928	33	NonExac	23-Aug-12	196	928
929	3	NonExac	06-Feb-12	57	929
930	9	NonExac	30-Jul-12	409	930
931	11	Exac	20-Jul-12	1020	931
932	29	NonExac	18-May-12	311	932
933	32	NonExac	23-Aug-12	154	933
934	11	Exac	20-Jul-12	1020	934
935	32	NonExac	23-Aug-12	154	935
936	29	NonExac	18-May-12	311	936
937	9	NonExac	29-May-12	409	937
938	33	NonExac	24-Jul-12	196	938
939	25	NonExac	30-May-12	57	939
940	9	NonExac	30-Jul-12	409	940
941	11	Exac	20-Jul-12	1020	
942	32	NonExac	23-Aug-12	154	
943	25	Exac	27-Aug-12	57	943
944	29	NonExac	18-May-12	311	945
945	29	NonExac	13-Apr-12	311	946
946	32	Exac	24-Jul-12	34	947
947	25	NonExac	06-Sep-12	57	948
948	33	Exac	21-Sep-12	196	949
949	21	NonExac	21-Aug-12	1034	950
950	29	NonExac	18-Sep-12	311	951
951	32	NonExac	25-Sep-12	154	952
952	5	NonExac	18-May-12	357	953
953	33	NonExac	23-Aug-12	196	954
954	3	NonExac	06-Feb-12	57	957
955	25	NonExac	30-May-12	57	
956	9	NonExac	29-May-12	409	
957	25	Exac	27-Aug-12	57	
958	29	NonExac	18-May-12	311	
959	11	Exac	20-Jul-12	1020	
960	9	NonExac	30-Jul-12	409	
961	32	Exac	24-Jul-12	34	
962	32	NonExac	28-Mar-12	34	
963	42	NonExac	06-Dec-11	12	963
964	29	NonExac	18-Jun-12	311	964
965	21	NonExac	21-Aug-12	1034	
966	32	NonExac	25-Sep-12	154	
967	11	Exac	20-Jul-12	1020	
968	9	NonExac	30-Jul-12	409	
969	32	NonExac	23-Aug-12	154	
970	29	NonExac	18-May-12	311	
971	33	Exac	21-Sep-12	196	
972	29	NonExac	18-May-12	311	

973	9	NonExac	30-Jul-12	409	
974	11	Exac	20-Jul-12	1020	
975	32	Exac	24-Jul-12	34	
976	25	Exac	27-Aug-12	57	
977	29	NonExac	18-Jun-12	311	
978	21	NonExac	21-Aug-12	1034	
979	32	NonExac	25-Sep-12	154	
980	9	NonExac	29-May-12	409	
981	25	NonExac	30-May-12	57	
982	33	NonExac	23-Aug-12	196	
983	3	NonExac	06-Feb-12	57	
984	29	NonExac	13-Apr-12	311	
985	25	NonExac	30-May-12	57	
986	3	NonExac	06-Feb-12	57	
987	33	NonExac	23-Aug-12	196	
988	25	NonExac	30-May-12	57	
989	9	NonExac	29-May-12	409	
990	33	NonExac	23-Aug-12	196	
991	25	NonExac	06-Sep-12	57	
992	33	NonExac	23-Aug-12	196	
993	3	NonExac	06-Feb-12	57	
994	5	NonExac	18-May-12	357	994
995	29	NonExac	18-Sep-12	311	995
996	21	NonExac	21-Aug-12	1034	
997	21	NonExac	21-Aug-12	1034	
998	29	NonExac	18-Jun-12	311	
999	32	NonExac	25-Sep-12	154	
1000	25	NonExac	30-May-12	57	
1001	29	NonExac	13-Apr-12	311	
1002	33	NonExac	23-Aug-12	196	
1003	3	NonExac	06-Feb-12	57	
1004	24	Exac	18-Nov-11	196	1004
1005	42	NonExac	06-Dec-11	12	1005
1006	21	NonExac	21-Aug-12	1034	
1007	32	NonExac	25-Sep-12	154	
1008	29	NonExac	18-Jun-12	311	
1009	32	NonExac	23-Aug-12	154	
1010	11	Exac	20-Jul-12	1020	
1011	29	NonExac	18-May-12	311	
1012	9	NonExac	29-May-12	409	
1013	25	NonExac	30-May-12	57	
1014	3	NonExac	06-Feb-12	57	
1015	3	NonExac	06-Feb-12	57	
1016	9	NonExac	29-May-12	409	
1017	25	NonExac	30-May-12	57	
1018	33	NonExac	23-Aug-12	196	

1019	32	NonExac	28-Mar-12	34	
1020	42	NonExac	06-Dec-11	12	
1021	32	NonExac	28-Mar-12	34	
1022	42	NonExac	06-Dec-11	12	
1023	33	Exac	21-Sep-12	196	
1024	25	Exac	27-Aug-12	57	
1025	32	Exac	24-Jul-12	34	
1026	11	Exac	20-Jul-12	1020	
1027	9	NonExac	30-Jul-12	409	
1028	29	NonExac	18-May-12	311	
1029	29	NonExac	13-Apr-12	311	
1030	33	Exac	21-Sep-12	196	
1031	3	NonExac	06-Feb-12	57	
1032	25	NonExac	06-Sep-12	57	
1033	3	NonExac	06-Feb-12	57	
1034	29	NonExac	13-Apr-12	311	
1035	25	NonExac	30-May-12	57	
1036	33	NonExac	23-Aug-12	196	
1037	9	NonExac	29-Jun-12	409	1037
1038	5	NonExac	19-Jun-12	357	1038
1039	29	NonExac	17-Jul-12	311	1039
1040	33	Exac	21-Sep-12	196	
1041	9	NonExac	30-Jul-12	409	
1042	25	Exac	27-Aug-12	57	
1043	29	NonExac	13-Apr-12	311	
1044	11	Exac	20-Jul-12	1020	
1045	32	Exac	24-Jul-12	34	
1046	25	NonExac	06-Sep-12	57	
1047	29	NonExac	13-Apr-12	311	
1048	32	Exac	24-Jul-12	34	
1049	33	Exac	21-Sep-12	196	
1050	33	Exac	21-Sep-12	196	
1051	32	Exac	24-Jul-12	34	
1052	9	NonExac	30-Jul-12	409	
1053	25	NonExac	06-Sep-12	57	
1054	29	NonExac	13-Apr-12	311	
1055	32	Exac	24-Jul-12	34	
1056	11	Exac	20-Jul-12	1020	
1057	9	NonExac	30-Jul-12	409	
1058	29	NonExac	18-May-12	311	
1059	25	Exac	27-Aug-12	57	
1060	25	NonExac	06-Sep-12	57	
1061	33	Exac	21-Sep-12	196	
1062	29	NonExac	13-Apr-12	311	
1064	25	NonExac	06-Sep-12	57	
1065	33	Exac	21-Sep-12	196	

1066	32	Exac	24-Jul-12	34	
1067	29	NonExac	17-Jul-12	311	1067
1068	5	NonExac	19-Jun-12	357	1068
1069	42	NonExac	06-Dec-11	12	
1070	29	NonExac	17-Jul-12	311	
1071	5	NonExac	19-Jun-12	357	
1072	29	NonExac	17-Jul-12	311	
1073	5	NonExac	19-Jun-12	357	
1074	29	NonExac	17-Jul-12	311	
1075	9	NonExac	29-Jun-12	409	1075
1076	5	NonExac	19-Jun-12	357	
1077	29	NonExac	17-Jul-12	311	
1078	42	NonExac	06-Dec-11	12	
1079	42	NonExac	06-Dec-11	12	
1080	32	NonExac	23-Aug-12	154	
1081	29	NonExac	18-May-12	311	
1082	29	NonExac	17-Jul-12	311	
1083	9	NonExac	29-Jun-12	409	
1084	42	NonExac	06-Dec-11	12	
1085	29	NonExac	18-Jun-12	311	
1086	32	NonExac	23-Aug-12	154	
1087	24	Exac	18-Nov-11	196	1087
1088	5	NonExac	19-Jun-12	357	
1089	29	NonExac	17-Jul-12	311	
1090	32	NonExac	25-Sep-12	154	
1091	21	NonExac	21-Aug-12	1034	
1092	9	NonExac	29-Jun-12	409	
1093	29	NonExac	18-Sep-12	311	
1094	29	NonExac	18-Sep-12	311	
1095	9	NonExac	29-Jun-12	409	
1096	42	NonExac	13-Jan-12	12	1096
1097	29	NonExac	08-Mar-12	311	1097
1098	29	NonExac	12-Jan-12	311	1098
1099	42	NonExac	13-Jan-12	12	1099
1100	29	NonExac	12-Jan-12	311	1100
1101	42	NonExac	13-Jan-12	12	
1102	11	NonExac	27-Jul-11	Hh	1102
1103	29	NonExac	08-Mar-12	311	1103
1104	42	NonExac	13-Jan-12	12	
1105	29	NonExac	07-Feb-12	311	1105
1106	32	NonExac	23-Aug-12	154	
1107	29	NonExac	18-Jun-12	311	
1108	42	NonExac	13-Jan-12	12	
1109	29	NonExac	12-Jan-12	311	
1110	42	NonExac	13-Jan-12	12	
1111	29	NonExac	12-Jan-12	311	

1112	32	NonExac	28-Mar-12	34	
1113	42	NonExac	06-Dec-11	12	
1114	33	NonExac	24-Jul-12	196	1114
1115	3	NonExac	06-Jun-12	836	
1116	9	NonExac	29-May-12	409	
1117	33	NonExac	24-Jul-12	196	
1118	32	NonExac	25-Sep-12	154	
1119	21	NonExac	21-Aug-12	1034	
1120	32	NonExac	23-Aug-12	154	
1121	29	NonExac	18-Jun-12	311	
1122	32	NonExac	25-Sep-12	154	
1123	29	NonExac	18-Jun-12	311	
1124	29	NonExac	12-Jan-12	311	
1125	42	NonExac	13-Jan-12	12	
1126	33	NonExac	24-Jul-12	196	
1127	3	NonExac	06-Jun-12	Hh	1127
1128	32	NonExac	23-Aug-12	154	
1129	29	NonExac	18-Jun-12	311	
1130	29	NonExac	17-Jul-12	311	
1131	5	NonExac	19-Jun-12	357	
1132	9	NonExac	29-May-12	409	
1133	33	NonExac	24-Jul-12	196	
1134	29	NonExac	18-Sep-12	311	
1135	5	NonExac	18-May-12	357	
1136	29	NonExac	12-Jan-12	311	
1137	42	NonExac	13-Jan-12	12	
1138	32	NonExac	25-Sep-12	154	
1139	21	NonExac	21-Aug-12	11	1139
1140	29	NonExac	12-Jan-12	311	
1141	42	NonExac	13-Jan-12	12	
1142	32	NonExac	28-Mar-12	34	
1143	42	NonExac	06-Dec-11	12	
1144	42	NonExac	13-Jan-12	12	
1145	29	NonExac	12-Jan-12	311	
1146	5	NonExac	19-Jun-12	357	
1147	29	NonExac	17-Jul-12	311	
1148	9	NonExac	29-May-12	409	
1149	33	NonExac	24-Jul-12	196	
1151	29	NonExac	18-Sep-12	311	
1152	21	NonExac	28-May-12	143	1152
1153	21	NonExac	28-May-12	143	1153
1154	21	NonExac	28-May-12	143	
1155	21	NonExac	28-May-12	143	
1156	21	NonExac	28-May-12	143	
1157	21	NonExac	28-May-12	143	
1159	33	NonExac	24-Jul-12	196	

1160	21	NonExac	28-Jun-12	1034	1160
1161	9	NonExac	29-Jun-12	409	
1162	3	NonExac	06-Jun-12	Hh	1162
1163	29	NonExac	07-Feb-12	311	1163
1164	27	NonExac	07-Feb-12	819	1164
1165	21	NonExac	28-Jun-12	1034	1165
1166	29	NonExac	07-Feb-12	311	
1167	5	NonExac	19-Jun-12	357	
1168	42	Exac	27-Jan-12	12	1168
1169	21	NonExac	28-Jun-12	1034	
1170	9	NonExac	29-Jun-12	409	
1171	33	NonExac	24-Jul-12	196	
1172	9	NonExac	29-Jun-12	409	
1173	29	NonExac	08-Mar-12	311	
1174	21	NonExac	28-Jun-12	1034	
1175	3	NonExac	06-Jun-12	836	1175
1176	33	NonExac	24-Jul-12	196	
1177	29	NonExac	08-Mar-12	311	
1178	29	NonExac	18-Sep-12	311	
1179	9	NonExac	30-Jul-12	409	
1180	29	NonExac	07-Feb-12	311	
1181	3	NonExac	06-Jun-12	836	1181
1182	33	NonExac	24-Jul-12	196	
1183	21	NonExac	28-Jun-12	1034	
1184	29	NonExac	08-Mar-12	311	
1185	5	NonExac	19-Jun-12	357	
1186	21	NonExac	28-Jun-12	1034	
1187	29	NonExac	08-Mar-12	311	
1188	29	NonExac	07-Feb-12	311	
1189	29	NonExac	08-Mar-12	311	
1190	21	NonExac	26-Apr-12	11	
1191	29	NonExac	08-Mar-12	311	
1192	29	NonExac	07-Feb-12	311	
1193	21	NonExac	28-Jun-12	1034	
1194	29	NonExac	07-Feb-12	311	
1195	29	NonExac	08-Mar-12	311	
1196	3	NonExac	06-Jun-12	836	
1197	29	NonExac	07-Feb-12	311	
1198	9	NonExac	29-Jun-12	409	
1199	3	NonExac	06-Jun-12	836	
1200	29	NonExac	07-Feb-12	311	
1201	27	NonExac	07-Feb-12	819	1201
1202	29	NonExac	08-Mar-12	311	
1203	3	NonExac	06-Jun-12	836	
1204	4	Exac	05-Apr-12	203	1204
1206	20	NonExac	25-Aug-11	Hh	1206

1207	86	Exac	17-Oct-12	155	1207
1208	73	NonExac	16-Mar-12	57	1208
1209	73	NonExac	16-May-12	1441	1209
1210	35	NonExac	26-Sep-12	Hh	1210
1211	86	Exac	18-Dec-12	155	1211
1212	73	NonExac	16-Apr-12	203	1212
1213	4	NonExac	09-Jul-12	Hh	1213
1214	35	NonExac	30-Mar-12	Hh	1214
1215	73	NonExac	16-Apr-12	57	
1216	4	NonExac	09-Jul-12	203	1216
1217	86	Exac	18-Dec-12	155	1217
1218	35	NonExac	26-Sep-12	Hh	1218
1219	73	NonExac	16-May-12	1441	1219
1220	124	Exac	17-Dec-12	409	1220
1221	75	Exac	14-Nov-12	Hh	1221
1222	4	Exac	05-Apr-12	203	1222
1223	20	NonExac	25-Aug-11	Hh	1223
1224	73	NonExac	17-Feb-12	1442	
1225	35	NonExac	21-Nov-11	Hh	1225
1226	18	NonExac	23-Nov-11	156	1226
1227	86	NonExac	11-Apr-12	436	1227
1228	4	NonExac	09-Jul-12	203	1228
1229	35	NonExac	30-Mar-12	Hh	1229
1230	73	NonExac	16-Apr-12	57	
1231	86	Exac	18-Dec-12	155	
1232	73	NonExac	16-Apr-12	1441	1232
1233	4	NonExac	09-Jul-12	203	
1234	86	Exac	18-Dec-12	155	
1235	86	Exac	18-Dec-12	155	
1236	73	NonExac	16-Apr-12	57	1236
1237	4	NonExac	09-Jul-12	203	
1238	73	NonExac	16-Mar-12	57	1238
1239	20	NonExac	25-Aug-11	Hh	1239
1240	86	Exac	17-Oct-12	155	1240
1241	4	NonExac	17-Apr-12	203	1241
1242	73	NonExac	16-Apr-12	57	1242
1243	86	Exac	18-Dec-12	155	
1245	86	Exac	17-Oct-12	155	
1246	20	NonExac	25-Nov-11	Hh	1246
1247	73	NonExac	16-Mar-12	57	
1248	4	NonExac	17-Apr-12	203	1248
1249	73	NonExac	16-Mar-12	57	
1250	35	NonExac	30-Mar-12	Hh	1250
1251	4	NonExac	17-Apr-12	203	
1252	86	Exac	18-Dec-12	155	
1253	86	Exac	18-Dec-12	155	

1254	20	NonExac	25-Nov-11	Hh	1254
1255	73	NonExac	16-Mar-12	57	
1256	4	NonExac	17-Apr-12	203	
1257	73	NonExac	16-May-12	1441	
1258	35	NonExac	26-Sep-12	Hh	1258
1259	20	NonExac	25-Aug-11	Hh	1259
1260	86	NonExac	11-Apr-12	436	1260
1261	35	NonExac	21-Nov-11	Hh	1261
1262	18	NonExac	23-Nov-11	156	1262
1263	73	NonExac	17-Feb-12	57	
1264	75	Exac	14-Nov-12	Hh	1264
1265	4	Exac	05-Apr-12	203	
1266	73	NonExac	16-Apr-12	203	1266
1267	86	Exac	18-Dec-12	155	
1268	4	NonExac	09-Jul-12	57	1268
1269	20	NonExac	25-Aug-11	Hh	1269
1270	86	NonExac	11-Apr-12	436	
1271	18	NonExac	23-Nov-11	156	
1272	35	NonExac	21-Nov-11	Hh	1272
1273	73	NonExac	17-Feb-12	1442	
1274	4	Exac	05-Apr-12	203	
1275	86	Exac	17-Oct-12	155	
1276	20	NonExac	25-Nov-11	Hh	1276
1277	73	NonExac	16-Mar-12	57	
1278	4	NonExac	17-Apr-12	203	
1279	20	NonExac	25-Nov-11	Hh	1279
1280	4	NonExac	17-Apr-12	203	
1282	73	NonExac	16-Mar-12	57	
1283	86	Exac	18-Dec-12	155	
1284	4	Exac	05-Apr-12	203	
1285	73	NonExac	16-May-12	1441	
1286	35	NonExac	26-Sep-12	Hh	1286
1287	20	NonExac	25-Aug-11	Hh	1287
1288	73	NonExac	16-Mar-12	1442	1288
1289	86	Exac	17-Oct-12	155	
1291	20	NonExac	25-Aug-11	57	1291
1292	73	NonExac	16-Mar-12	203	1292
1293	4	NonExac	17-Apr-12	Hh	1293
1294	73	NonExac	16-Apr-12	57	1294
1296	18	NonExac	23-Nov-11	156	
1298	20	NonExac	25-Aug-11	Hh	1298
1299	86	NonExac	11-Apr-12	436	
1301	73	NonExac	17-Feb-12	Hh	1301
1302	73	NonExac	16-Apr-12	1441	1302
1303	4	NonExac	09-Jul-12	203	
1304	4	NonExac	17-Apr-12	203	

1305	86	Exac	17-Oct-12	155	
1306	20	NonExac	25-Aug-11	Hh	1306
1307	73	NonExac	16-Mar-12	1442	1307
1309	73	NonExac	17-Feb-12	57	
1310	35	NonExac	21-Nov-11	Hh	1310
1311	18	NonExac	23-Nov-11	156	
1312	20	NonExac	25-Aug-11	Hh	1312
1313	86	NonExac	11-Apr-12	436	
1314	75	Exac	14-Nov-12	Hh	1314
1315	73	NonExac	17-Feb-12	57	
1316	35	NonExac	21-Nov-11	Hh	1316
1317	18	NonExac	23-Nov-11	156	
1318	20	NonExac	25-Aug-11	Hh	1318
1319	4	Exac	05-Apr-12	203	
1320	86	NonExac	11-Apr-12	436	
1321	73	NonExac	16-Apr-12	57	
1322	4	NonExac	09-Jul-12	203	
1323	4	Exac	05-Apr-12	203	
1325	4	Exac	05-Apr-12	203	
1326	4	Exac	05-Apr-12	203	
1327	86	NonExac	11-Apr-12	436	
1329	20	NonExac	25-Aug-11	Hh	1329
1330	18	NonExac	23-Nov-11	156	
1331	35	NonExac	21-Nov-11	Hh	1331
1332	73	NonExac	17-Feb-12	1442	1332
1333	20	NonExac	25-Aug-11	Hh	1333
1334	35	NonExac	21-Nov-11	Hh	1334
1336	75	Exac	14-Nov-12	Hh	1336
1337	18	NonExac	23-Nov-11	156	
1338	124	Exac	17-Dec-12	409	1338
1339	86	Exac	17-Oct-12	155	
1340	73	NonExac	17-Feb-12	1442	1340
1341	4	Exac	05-Apr-12	203	
1342	18	NonExac	23-Nov-11	156	
1343	20	NonExac	25-Aug-11	Hh	1343
1344	35	NonExac	21-Nov-11	Hh	1344
1345	4	Exac	05-Apr-12	203	
1346	86	NonExac	11-Apr-12	436	
1347	73	NonExac	17-Feb-12	57	1347
1348	75	Exac	14-Nov-12	Hh	1348
1349	86	NonExac	11-Apr-12	436	
1350	20	NonExac	25-Aug-11	Hh	1350
1351	18	NonExac	23-Nov-11	156	
1352	35	NonExac	21-Nov-11	Hh	1352
1353	73	NonExac	17-Feb-12	57	1353
1354	4	Exac	05-Apr-12	203	

1355	4	Exac	05-Apr-12	203	
1356	4	Exac	05-Apr-12	203	
1357	4	Exac	05-Apr-12	203	
1358	17	NonExac	14-Sep-11	1163	1358
1359	21	Exac	25-Oct-11	11	1359
1360	17	NonExac	14-Sep-11	1163	1360
1361	21	Exac	25-Oct-11	11	1361
1362	17	NonExac	14-Sep-11	1163	
1363	29	NonExac	13-Oct-11	311	
1364	29	NonExac	13-Oct-11	311	
1365	5	Exac	03-Oct-11	176	1365
1366	5	Exac	03-Oct-11	176	
1367	34	NonExac	27-Sep-11	Hh	1367
1368	34	NonExac	27-Sep-11	Hh	1368
1369	24	Exac	28-Oct-11	196	1369
1370	24	Exac	28-Oct-11	196	1370
1371	24	Exac	28-Oct-11	196	
1372	24	Exac	28-Oct-11	196	
1373	24	Exac	28-Oct-11	196	
1374	24	Exac	28-Oct-11	196	
1375	24	Exac	28-Oct-11	196	
1376	24	Exac	28-Oct-11	196	
1377	32	NonExac	26-Oct-11	513	1377
1378	32	NonExac	26-Oct-11	513	
1379	9	NonExac	20-Oct-11	409	1379
1380	32	NonExac	26-Oct-11	513	
1381	9	NonExac	20-Oct-11	409	1381
1382	11	Exac	24-Oct-11	704	1382
1383	32	NonExac	26-Oct-11	513	
1384	11	Exac	24-Oct-11	704	
1385	32	NonExac	26-Oct-11	513	
1386	9	NonExac	20-Oct-11	409	
1387	9	NonExac	20-Oct-11	409	
1388	11	Exac	24-Oct-11	704	
1389	17	Exac	28-Oct-11	1163	
1390	21	Exac	25-Oct-11	11	
1391	32	NonExac	26-Oct-11	513	
1392	29	NonExac	13-Oct-11	311	
1394	11	Exac	24-Oct-11	704	
1395	9	NonExac	20-Oct-11	409	
1396	32	NonExac	26-Oct-11	513	
1397	11	Exac	24-Oct-11	704	
1398	9	NonExac	20-Oct-11	409	
1399	11	Exac	24-Oct-11	704	
1400	9	NonExac	20-Oct-11	409	
1401	9	NonExac	20-Oct-11	409	

1402	11	Exac	24-Oct-11	704	
1403	17	Exac	28-Oct-11	1163	1403
1404	11	Exac	24-Oct-11	704	1404
1405	11	Exac	24-Oct-11	704	
1406	34	NonExac	26-Oct-11	Hh	1406
1407	17	Exac	28-Oct-11	1163	1407
1408	34	NonExac	26-Oct-11	Hh	1408
1409	17	Exac	28-Oct-11	1163	
1410	17	Exac	28-Oct-11	1163	
1411	34	NonExac	26-Oct-11	Hh	1411
1412	34	NonExac	26-Oct-11	Hh	1412
1413	17	Exac	28-Oct-11	1163	
1414	17	NonExac	14-Sep-11	1163	
1415	34	NonExac	26-Oct-11	Hh	1415
1416	34	NonExac	26-Oct-11	Hh	1416
1417	17	NonExac	14-Sep-11	1163	
1418	34	NonExac	26-Oct-11	Hh	1418
1419	17	NonExac	14-Sep-11	1163	
1420	21	Exac	25-Oct-11	11	
1421	17	NonExac	14-Sep-11	1163	
1422	34	NonExac	26-Oct-11	Hh	1422
1423	21	Exac	25-Oct-11	11	
1424	34	NonExac	26-Oct-11	Hh	1424
1425	21	Exac	25-Oct-11	11	
1426	17	NonExac	14-Sep-11	1163	
1427	17	NonExac	14-Sep-11	1163	
1428	21	Exac	25-Oct-11	11	
1429	21	Exac	25-Oct-11	11	
1430	17	NonExac	14-Sep-11	1163	
1431	21	Exac	25-Oct-11	11	
1432	17	NonExac	14-Sep-11	1163	
1433	17	NonExac	14-Sep-11	1163	
1434	17	NonExac	14-Sep-11	1163	
1435	17	NonExac	14-Sep-11	1163	
1436	17	NonExac	14-Sep-11	1163	
1437	29	NonExac	13-Oct-11	311	
1438	73	NonExac	13-Jul-12	1441	1438
1439	4	NonExac	17-May-12	203	1439
1440	73	NonExac	13-Jul-12	1441	1440
1441	73	NonExac	13-Jul-12	1441	
1442	73	NonExac	15-Jun-12	1441	1442
1443	73	NonExac	16-May-12	1441	
1445	73	NonExac	15-Jun-12	1441	1445
1446	73	NonExac	16-May-12	1441	
1447	73	NonExac	15-Jun-12	1441	
1448	73	Exac	17-Sep-12	Hh	1448

1449	73	NonExac	13-Jul-12	1441	
1450	73	NonExac	15-Jun-12	1441	
1452	73	NonExac	17-Dec-12	1441	1452
1453	4	NonExac	17-May-12	203	1453
1454	73	NonExac	17-Dec-12	1441	1454
1455	73	NonExac	15-Jun-12	1441	
1456	73	NonExac	15-Jun-12	1441	
1457	73	NonExac	16-May-12	1441	
1458	73	NonExac	15-Jun-12	1441	
1459	73	NonExac	16-May-12	1441	
1460	4	NonExac	17-May-12	203	
1461	73	Exac	17-Sep-12	1441	1461
1462	4	NonExac	17-May-12	203	
1463	73	NonExac	13-Jul-12	1441	
1464	73	NonExac	17-Dec-12	1441	
1465	73	NonExac	17-Oct-12	1441	1465
1466	73	NonExac	21-Jan-13	1441	1466
1467	73	NonExac	15-Jun-12	1441	
1468	73	Exac	17-Sep-12	Hh	1468
1469	4	NonExac	17-May-12	203	
1470	73	NonExac	13-Jul-12	1441	
1471	73	NonExac	16-May-12	1441	
1472	73	NonExac	15-Jun-12	1441	
1473	4	NonExac	17-May-12	203	
1474	73	NonExac	15-Jun-12	1441	
1475	4	NonExac	17-May-12	203	
1476	4	NonExac	17-May-12	203	
1477	73	NonExac	17-Oct-12	1441	1477
1478	4	NonExac	17-May-12	203	
1479	73	NonExac	17-Oct-12	1441	
1480	73	NonExac	13-Jul-12	1441	
1481	4	NonExac	17-May-12	203	
1482	73	NonExac	17-Dec-12	1441	
1483	73	NonExac	17-Oct-12	1441	
1484	73	NonExac	21-Jan-13	1441	1484
1485	73	Exac	17-Sep-12	Hh	1485
1486	73	NonExac	21-Jan-13	1441	
1487	73	NonExac	21-Jan-13	1441	
1488	73	NonExac	21-Jan-13	1441	
1489	73	NonExac	17-Oct-12	1441	
1490	73	NonExac	13-Jul-12	1441	
1491	73	NonExac	21-Jan-13	1441	
1492	73	NonExac	21-Jan-13	1441	
1493	73	NonExac	17-Dec-12	1441	
1494	73	NonExac	17-Oct-12	1441	
1495	73	NonExac	17-Dec-12	1441	

1496	73	NonExac	21-Jan-13	1441	
1497	73	Exac	17-Sep-12	1441	1497
1498	73	NonExac	13-Jul-12	1664	1498
1499	73	NonExac	17-Dec-12	1441	
1500	73	NonExac	17-Dec-12	1441	
1501	73	NonExac	13-Jul-12	1441	
1502	73	NonExac	17-Dec-12	1441	
1503	73	NonExac	17-Dec-12	1441	
1504	73	NonExac	21-Jan-13	1441	
1505	73	Exac	17-Sep-12	1441	
1506	73	NonExac	17-Oct-12	1441	
1507	73	NonExac	17-Oct-12	1441	
1508	73	NonExac	17-Oct-12	1441	
1509	73	Exac	17-Sep-12	Hh	1509
1510	73	NonExac	21-Jan-13	1441	
1511	73	NonExac	17-Oct-12	1441	
1512	73	Exac	17-Sep-12	Hh	1512
1513	73	Exac	17-Sep-12	1441	
1514	73	Exac	18-Feb-13	1441	1514
1515	34	NonExac	27-Sep-11	Hh	1515
1516	9	NonExac	20-Oct-11	409	
1517	17	Exac	28-Oct-11	1163	
1518	21	Exac	25-Oct-11	1034	1518
1519	11	Exac	24-Oct-11	704	
1520	29	NonExac	13-Oct-12	311	
1521	24	Exac	28-Oct-11	196	
1522	32	NonExac	26-Oct-11	513	
1523	73	Exac	18-Feb-13	1441	1523
1524	73	Exac	18-Feb-13	1441	
1525	73	Exac	18-Feb-13	1441	
1526	17	Exac	28-Oct-11	1163	
1527	17	Exac	28-Oct-11	1163	
1528	17	Exac	28-Oct-11	1163	
1529	73	Exac	18-Feb-13	1441	
1530	73	Exac	18-Feb-13	1441	
1531	73	Exac	18-Feb-13	1441	
1532	73	Exac	18-Feb-13	1441	

Appendix 3: Contingency tables for Chi^2

Appendix 3.1: Chapter 3. Haemolysis, H_2S and MLST

Contingency table for Chi square analysis of established ST from MLST analysis determined on recognized variations on sequences from seven housekeeping genes versus hydrogen sulphide production and haemolysis

		ST	
		Yes	**No**
H_2S positive	**Yes**	45.8% (33) 95% CI (39.14, 52.46)	54.2% (39) 95% CI (47.54, 60.86)
	No	89.7% (35) 95% CI (85.64, 93.76)	10.3% (4) 95% CI (6.24, 14.36)
	***p*-value**	*p*=<0.001	
Haemolysis	**Yes**	0% (0)	100% (12)
	No	69.7% (69) 95% CI (63.56, 75.84)	30.3% (30) 95% CI (24.16, 36.44)
	***p*-value**	*p*=0.001	

Appendix 3.2: Chapter 4: Adhesin genes and Exacerbation

Contingency table for Pearson Chi square analysis of adhesion genes present or absent in NTHi isolated from exacerbation events using Fisher's exact 2-sided test where counts <5 (noted *).

		Gene present in NTHi isolated from exacerbation.	
		Absent	**Present**
omp2	**Yes**	4.5% (3) 95% CI (1.73, 7.27)	95.5% (63) 95% CI (92.73, 98.27)
	No	10.7% (16) 95% CI (6.57, 14.83)	89.3% (133) 95% CI (85.17, 93.43)
	***p*-value**	*p*=0.140	
hia	**Yes**	97% (64) 95% CI (94.72, 99.28)	3% (2) 95% CI (0.72, 5.28)
	No	96.6% (144) 95% CI (94.18, 99.02)	3.4% (5) 95% CI (0.98, 5.82)
	***p*-value**	*p*=0.901	
hmwA	**Yes**	28.8% (19) 95% CI (22.75, 34.85)	71.2% (47) 95% CI (65.15, 77.25)
	No	34.9% (52) 95% CI (28.53, 41.27)	65.1% (97) 95% CI (58.73, 71.47)
	***p*-value**	*p*=0.379	
hmw2A	**Yes**	15.2% (10) 95% CI (10.4, 20)	84.8% (56) 95% CI (80, 80.96)
	No	20.1% (30) 95% CI (14.74, 25.46)	79.9% (119) 95% CI (74.54, 85.26)
	***p*-value**	*p*=0.387	
hap	**Yes**	12.1% (8) 95% CI (7.74, 16.46)	87.9% (58) 95% CI (83.54, 92.26)
	No	12.8% (19) 95% CI (8.33, 17.27)	87.2% (130) 95% CI (82.73, 91.67)
	***p*-value**	*p*=0.898	
hif	**Yes**	97% (64) 95% CI (94.72, 99.28)	3% (2) 95% CI (0.72, 5.28)
	No	91.3% (136) 95% CI (87.53, 95.07)	8.7% (13) 95% CI (4.93, 12.47)
	***p*-value**	*p*=0.157 *	
omp1	**Yes**	3% (2) 95% CI (0.72, 5.28)	97% (64) 95% CI 94.72, 99.28)
	No	6.7% (10) 95% CI (3.36, 10.04)	93.3% (139) 95% CI (89.96, 96.64)
	***p*-value**	*p*=0.352 *	

Appendix 3.3: Chapter 5: Isolates chosen for ROARY analysis.

Isolates from both species were chosen for high or low MIC and split into two trait groups.

Isolate Number	Species	MIC	Trait
557	Hh	1	0
552	Hh	2	0
54	Hh	2	0
427	Hh	3	0
559	Hh	3	0
426	Hh	8	1
553	Hh	8	1
488	Hh	12	1
494	Hh	12	1
497	Hh	12	1
421	NTHi	2	0
560	NTHi	2	0
836	NTHi	2	0
529	NTHi	2	0
531	NTHi	2	0
597	NTHi	24	1
592	NTHi	24	1
352	NTHi	16	1
598	NTHi	16	1
298	NTHi	16	1

References

ADAM, M., MURALI, B., GLENN, N. O. & POTTER, S. S. 2008. Epigenetic inheritance based evolution of antibiotic resistance in bacteria. *BMC Evol Biol,* 8**,** 52.

ADAMS, W. G., DEAVER, K. A., COCHI, S. L., PLIKAYTIS, B. D., ZELL, E. R., BROOME, C. V. & WENGER, J. D. 1993. Decline of childhood *Haemophilus influenzae* type b (Hib) disease in the Hib vaccine era. *JAMA,* 269**,** 221-6.

AHMED, S., SHAPIRO, N. L. & BHATTACHARYYA, N. 2014. Incremental health care utilization and costs for acute otitis media in children. *Laryngoscope,* 124**,** 301-5.

ALBERT, R. K., CONNETT, J., BAILEY, W. C., CASABURI, R., COOPER, J. A., JR., CRINER, G. J., CURTIS, J. L., DRANSFIELD, M. T., HAN, M. K., LAZARUS, S. C., MAKE, B., MARCHETTI, N., MARTINEZ, F. J., MADINGER, N. E., MCEVOY, C., NIEWOEHNER, D. E., PORSASZ, J., PRICE, C. S., REILLY, J., SCANLON, P. D., SCIURBA, F. C., SCHARF, S. M., WASHKO, G. R., WOODRUFF, P. G., ANTHONISEN, N. R. & NETWORK, C. C. R. 2011. Azithromycin for prevention of exacerbations of COPD. *N Engl J Med,* 365**,** 689-98.

ALIKHAN, N. F., PETTY, N. K., BEN ZAKOUR, N. L. & BEATSON, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics,* 12**,** 402.

ALLIANCE, G. 2016. *Hib Initiative: a GAVI success story.* [Online]. Available: http://www.gavi.org/library/news/roi/2010/hib-initiative--a-gavi-success-story/. [Accessed 17th January 2017].

ALMOFTI, Y. A., DAI, M., SUN, Y., HAIHONG, H. & YUAN, Z. 2011. Impact oferythromycin resistance on the virulence properties and fitness of Campylobacter jejuni. *Microb Pathog,* 50**,** 336-42.

ALONSO, R., MATEO, E., CHURRUCA, E., MARTINEZ, I., GIRBAU, C. & FERNANDEZ-ASTORGA, A. 2005. MAMA-PCR assay for the detection of point mutations associated with high-level erythromycin resistance in *Campylobacter jejuni* and *Campylobactercoli* strains. *J Microbiol Methods,* 63**,** 99-103.

ALTENBURG, J., DE GRAAFF, C. S., STIENSTRA, Y., SLOOS, J. H., VAN HAREN, E. H., KOPPERS, R. J., VAN DER WERF, T. S. & BOERSMA, W. G. 2013. Effect of azithromycin maintenance treatment on infectious exacerbations among patients with non-cystic fibrosis bronchiectasis: the BAT randomized controlled trial. *JAMA,* 309**,** 1251-9.

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

ANDERSON, E. C., BEGG, N. T., CRAWSHAW, S. C., HARGREAVES, R. M., HOWARD, A. J.& SLACK, M. P. 1995a. Epidemiology of invasive *Haemophilus influenzae* infections in England and Wales in the pre-vaccination era (1990-2). *Epidemiol Infect,* 115**,** 89-100.

ANDERSON, R., WANG, X., BRIERE, E. C., KATZ, L. S., COHN, A. C., CLARK, T. A., MESSONNIER, N. E. & MAYER, L. W. 2012. *Haemophilus haemolyticus* isolates causing clinical disease. *J Clin Microbiol,* 50**,** 2462-5.

ANDERSON, S. L., BERMAN, J., KUSCHNER, R., WESCHE, D., MAGILL, A., WELLDE, B., SCHNEIDER, I., DUNNE, M. & SCHUSTER, B. G. 1995b. Prophylaxis of *Plasmodium falciparum* malaria with azithromycin administered to volunteers. *Ann Intern Med,* 123**,** 771-3.

ANDREWS, S. 2012. *Babraham Bioinformatics: FASTQC, a quality control tool for thehigh throughput sequence data.* [Online]. Available: http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/ [Accessed 01/01/2013 2014].

ANGOULVANT, F., COHEN, R., DOIT, C., ELBEZ, A., WERNER, A., BECHET, S., BONACORSI, S., VARON, E. & LEVY, C. 2015. Trends in antibiotic resistance of *Streptococcus pneumoniae* and *Haemophilus influenzae* isolated from nasopharyngeal flora in children with acute otitis media in France before and after 13 valent pneumococcal conjugate vaccine introduction. *BMC Infect Dis,* 15**,** 236.

ANWAR, G. A., BOURKE, S. C., AFOLABI, G., MIDDLETON, P., WARD, C. & RUTHERFORD, R. M. 2008. Effects of long-term low-dose azithromycin in patients with non-CF bronchiectasis. *Respir Med,* 102**,** 1494-6.

ATACK, J. M., WINTER, L. E., JURCISEK, J. A., BAKALETZ, L. O., BARENKAMP, S. J. & JENNINGS, M. P. 2015. Selection and Counterselection of Hia Expression Reveals a Key Role for Phase-Variable Expression of Hia in Infection Caused by Nontypeable *Haemophilus influenzae*. *J Infect Dis*.

ATKINSON, C. T., KUNDE, D. A. & TRISTRAM, S. G. 2015. Acquired macrolide resistance genes in *Haemophilus influenzae*? *J Antimicrob Chemother,* 70**,** 2234-6.

AUTRET, N., DUBAIL, I., TRIEU-CUOT, P., BERCHE, P. & CHARBIT, A. 2001. Identification of new genes involved in the virulence of *Listeria monocytogenes* by signature-tagged transposon mutagenesis. *Infect Immun,* 69**,** 2054-65.

AVADHANULA, V., RODRIGUEZ, C. A., ULETT, G. C., BAKALETZ, L. O. & ADDERSON, E. E. 2006. Nontypeable *Haemophilus influenzae* adheres to intercellular adhesion molecule 1 (ICAM-1) on respiratory epithelial cells and upregulates ICAM-1 expression. *Infect Immun,* 74**,** 830-8.

BADR, W. H., LOGHMANEE, D., KARALUS, R. J., MURPHY, T. F. & THANAVALA, Y. 1999. Immunization of mice with P6 of nontypeable *Haemophilus influenzae*: kinetics of the antibody response and IgG subclasses. *Vaccine,* 18**,** 29-37.

BAKALETZ, L. M., RS. 2014. *Chimeric vaccine for Haemophilus influenzae induceddisease*. US patent application. June 3rd 2014.

BANDI, V., APICELLA, M. A., MASON, E., MURPHY, T. F., SIDDIQI, A., ATMAR, R. L.& GREENBERG, S. B. 2001. Nontypeable *Haemophilus influenzae* in the lower respiratory tract of patients with chronic bronchitis. *Am J Respir Crit Care Med,* 164**,** 2114-9.

BANDI, V., JAKUBOWYCZ, M., KINYON, C., MASON, E. O., ATMAR, R. L., GREENBERG, S. B.& MURPHY, T. F. 2003. Infectious exacerbations of chronic obstructive pulmonary disease associated with respiratory viruses and non-typeable *Haemophilus influenzae*. *FEMS Immunol Med Microbiol,* 37**,** 69-75.

BANERJEE, D., KHAIR, O. A. & HONEYBOURNE, D. 2005. The effect of oral clarithromycin on health status and sputum bacteriology in stable COPD. *Respir Med,* 99**,**208-15.

BARANY, F. & KAHN, M. E. 1985. Comparison of transformation mechanisms of *Haemophilus parainfluenzae* and *Haemophilus influenzae*. *J Bacteriol,* 161**,** 72-9.

BARENKAMP, S. J. & ST GEME, J. W., 3RD 1994. Genes encoding high-molecular-weight adhesion proteins of nontypeable *Haemophilus influenzae* are part of gene clusters. *Infect Immun,* 62**,** 3320-8.

BARENKAMP, S. J. & ST GEME, J. W., 3RD 1996. Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable *Haemophilus influenzae*. *Mol Microbiol,* 19**,** 1215-23.

BARKAI, G., LEIBOVITZ, E., GIVON-LAVI, N. & DAGAN, R. 2009. Potential contribution by nontypable *Haemophilus influenzae* in protracted and recurrent acute otitismedia. *Pediatr Infect Dis J,* 28**,** 466-71.

BARQUIST, L., MAYHO, M., CUMMINS, C., CAIN, A. K., BOINETT, C. J., PAGE, A. J., LANGRIDGE, G. C., QUAIL, M. A., KEANE, J. A. & PARKHILL, J. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics,* 32**,** 1109-11.

BARTHEL, D., SCHINDLER, S. & ZIPFEL, P. F. 2012a. Plasminogen is a complement inhibitor. *J Biol Chem,* 287**,** 18831-42.

BARTHEL, D., SINGH, B., RIESBECK, K. & ZIPFEL, P. F. 2012b. *Haemophilus influenzae* uses the surface protein E to acquire human plasminogen and to evade innate immunity. *J Immunol,* 188**,** 379-85.

BARTKUS, J. M., JUNI, B. A., EHRESMANN, K., MILLER, C. A., SANDEN, G. N., CASSIDAY, P. K., SAUBOLLE, M., LEE, B., LONG, J., HARRISON, A. R., JR. & BESSER, J. M. 2003. Identification of a mutation associated with erythromycin resistance in *Bordetella pertussis*: implications for surveillance of antimicrobial resistance. *J Clin Microbiol,* 41**,** 1167-72.

BEGOVIC, J., HUYS, G., MAYO, B., D'HAENE, K., FLOREZ, A. B., LOZO, J., KOJIC, M., STRAHINIC, I. & TOPISIROVIC, L. 2009. Human vaginal *Lactobacillusrhamnosus* harbor mutation in 23S rRNA associated with erythromycin resistance. *Res Microbiol,* 160**,** 421-6.

BENSON, D. A., CAVANAUGH, M., CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL,J. & SAYERS, E. W. 2016. GenBank. *Nucleic Acids Res*.

BERENSON, C. S., GARLIPP, M. A., GROVE, L. J., MALONEY, J. & SETHI, S. 2006a. Impaired phagocytosis of nontypeable *Haemophilus influenza*e by human alveolar macrophages in chronic obstructive pulmonary disease. *J Infect Dis,* 194**,** 1375-84.

BERENSON, C. S., MURPHY, T. F., WRONA, C. T. & SETHI, S. 2005. Outer membrane protein P6 of nontypeable *Haemophilus influenzae* is a potent and selective inducer of human macrophage proinflammatory cytokines. *Infect Immun,* 73**,** 2728-35.

BERENSON, C. S., WRONA, C. T., GROVE, L. J., MALONEY, J., GARLIPP, M. A., WALLACE, P. K., STEWART, C. C. & SETHI, S. 2006b. Impaired alveolar macrophage response to *Haemophilus* antigens in chronic obstructive lung disease. *Am J Respir Crit Care Med,* 174**,** 31-40.

BERGMAN, M., HUIKKO, S., HUOVINEN, P., PAAKKARI, P., SEPPALA, H. & FINNISHSTUDY GROUP FOR ANTIMICROBIAL, R. 2006. Macrolide and azithromycin use are linked to increased macrolide resistance in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother,* 50**,** 3646-50.

BERTOT, G. M., BECKER, P. D., GUZMAN, C. A. & GRINSTEIN, S. 2004. Intranasalvaccination with recombinant P6 protein and adamantylamide dipeptide as mucosal adjuvant confers efficient protection against otitis media and lung infection by nontypeable *Haemophilus influenzae*. *J Infect Dis,* 189**,** 1304-12.

BHANDARI, R. & SHARMA, R. 2012. Epidemiology of chronic obstructive pulmonary disease: a descriptive study in the mid-western region of Nepal. *Int J Chron Obstruct Pulmon Dis,* 7**,** 253-7.

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.

BLASI, F., BONARDI, D., ALIBERTI, S., TARSIA, P., CONFALONIERI, M., AMIR, O., CARONE,M., DI MARCO, F., CENTANNI, S. & GUFFANTI, E. 2010. Long-term azithromycin use in patients with chronic obstructive pulmonary disease and tracheostomy. *Pulm Pharmacol Ther,* 23**,** 200-7.

BOLDUC, G. R., BOUCHET, V., JIANG, R. Z., GEISSELSODER, J., TRUONG-BOLDUC, Q. C., RICE, P. A., PELTON, S. I. & GOLDSTEIN, R. 2000. Variability of outer membrane protein P1 and its evaluation as a vaccine candidate against experimental otitis media due to nontypeable *Haemophilus influenzae*: an unambiguous, multifaceted approach. *Infect Immun,* 68**,** 4505-17.

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

BONJOUR, S. E. A. 2013. Solid Fuel Use for Household Cooking: Country and Regional Estimates for 1980–2010. *Environmental Health Perspectives*,784-790.

BOOKWALTER, J. E., JURCISEK, J. A., GRAY-OWEN, S. D., FERNANDEZ, S., MCGILLIVARY, G.& BAKALETZ, L. O. 2008. A carcinoembryonic antigen-related cell adhesion molecule 1 homologue plays a pivotal role in nontypeable *Haemophilus influenzae* colonization of the chinchilla nasopharynx via the outer membrane protein P5-homologous adhesin. *Infect Immun,* 76**,**48-55.

BOOY, R., HODGSON, S. A., SLACK, M. P., ANDERSON, E. C., MAYON-WHITE, R. T.& MOXON, E. R. 1993. Invasive *Haemophilus influenzae* type b disease in the Oxford region (1985-91). *Arch Dis Child,* 69**,** 225-8.

BOURNE, S., COHET, C., KIM, V., BARTON, A., TUCK, A., ARIS, E., MESIA-VELA, S.,DEVASTER, J. M., BALLOU, W. R., CLARKE, S. C. & WILKINSON, T. 2014. Acute Exacerbation and Respiratory InfectionS in COPD (AERIS): protocol for a prospective, observational cohort study. *BMJ Open,* 4**,** e004546.

BRANDILEONE, M. C., ZANELLA, R. C., ALMEIDA, S. C., BRANDAO, A. P., RIBEIRO, A. F., CARVALHANAS, T. M., SATO, H., ANDRADE, A. L., VERANI, J. R. & PNEUMOCOCCAL CARRIAGE STUDY, G. 2016. Effect of 10-valent pneumococcal conjugate vaccineon nasopharyngeal carriage of *Streptococcus pneumoniae* and *Haemophilus influenzae* among children in Sao Paulo, Brazil. *Vaccine,* 34**,**5604-5611.

BRYNILDSRUD, O., BOHLIN, J., SCHEFFER, L. & ELDHOLM, V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol,* 17**,**238.

BURNS, J. L. & SMITH, A. L. 1987. A major outer-membrane protein functions as a porin in *Haemophilus influenzae*. *J Gen Microbiol,* 133**,** 1273-7.

CAPLAN, A. J., CYR, D. M. & DOUGLAS, M. G. 1993. Eukaryotic homologues of *Escherichia coli* dnaJ: a diverse protein family that functions with hsp70 stress proteins. *Mol Biol Cell,* 4**,** 555-63.

CARDINES, R., GIUFRE, M., MASTRANTONIO, P., CIOFI DEGLI ATTI, M. L. & CERQUETTI, M. 2007. Nontypeable *Haemophilus influenzae* meningitis in children: phenotypic and genotypic characterization of isolates. *Pediatr Infect Dis J,* 26**,** 577-82.

CARDINES, R., GIUFRE, M., POMPILIO, A., FISCARELLI, E., RICCIOTTI, G., DI BONAVENTURA, G. & CERQUETTI, M. 2012. *Haemophilus influenzae* in children with cysticfibrosis: antimicrobial susceptibility, molecular epidemiology, distribution of adhesins and biofilm formation. *Int J Med Microbiol,* 302**,**45-52.

CARO-QUINTERO, A., RODRIGUEZ-CASTANO, G. P. & KONSTANTINIDIS, K. T. 2009. Genomic insights into the convergence and pathogenicity factors of *Campylobacter jejuni* and *Campylobacter coli* species. *J Bacteriol,* 191**,** 5824-31.

CARRUTHERS, M. D., TRACY, E. N., DICKSON, A. C., GANSER, K. B., MUNSON, R. S., JR. & BAKALETZ, L. O. 2012. Biological roles of nontypeable *Haemophilus influenzae*type IV pilus proteins encoded by the pil and com operons. *J Bacteriol,* 194**,**1927-33.

CASEY, R. M., DUMOLARD, L., DANOVARO-HOLLIDAY, M. C., GACIC-DOBO, M., DIALLO, M. S., HAMPTON, L. M. & WALLACE, A. S. 2016. Global Routine Vaccination Coverage, 2015. *MMWR Morb Mortal Wkly Rep,* 65**,** 1270-1273.

CATCHPOLE, I., THOMAS, C., DAVIES, A. & DYKE, K. G. 1988. The nucleotide sequenceof *Staphylococcus aureus* plasmid pT48 conferring inducible macrolide-lincosamide-streptogramin B resistance and comparison with similar plasmids expressing constitutive resistance. *J Gen Microbiol,* 134**,** 697-709.

CHAN, K., MILLER, W. G., MANDRELL, R. E. & KATHARIOU, S. 2007. The absence of intervening sequences in 23S rRNA genes of *Campylobacter coli* isolatesfrom Turkeys is a unique attribute of a cluster of related strains which also lack resistance to erythromycin. *Appl Environ Microbiol,* 73**,** 1208-14.

CHANG, A., ADLOWITZ, D. G., YELLAMATTY, E. & PICHICHERO, M. 2010. *Haemophilus influenzae* outer membrane protein P6 molecular characterization may not differentiate all strains of *H. Influenzae* from *H. haemolyticus*. *J Clin Microbiol,*48**,** 3756-7.

CHANG, A., KAUR, R., MICHEL, L. V., CASEY, J. R. & PICHICHERO, M. 2011. *Haemophilus influenzae* vaccine candidate outer membrane protein P6 is not conserved in all strains. *Hum Vaccin,* 7**,** 102-5.

CHEN, L., SONG, Y., WEI, Z., HE, H., ZHANG, A. & JIN, M. 2013. Antimicrobial susceptibility, tetracycline and erythromycin resistance genes, and multilocus sequence typing of *Streptococcus suis* isolates from diseased pigs in China. *J Vet Med Sci,* 75**,**583-7.

CHISHOLM, S. A., DAVE, J. & ISON, C. A. 2010. High-level azithromycin resistance occurs in *Neisseria gonorrhoeae* as a result of a single point mutation in the 23S rRNAgenes. *Antimicrob Agents Chemother,* 54**,** 3812-6.

CHISHOLM, S. A., NEAL, T. J., ALAWATTEGAMA, A. B., BIRLEY, H. D., HOWE, R. A. & ISON, C. *A.* 2009. Emergence of high-level azithromycin resistance in *Neisseriagonorrhoeae* in England and Wales. *J Antimicrob Chemother,* 64**,** 353-8.

CHITTUM, H. S. & CHAMPNEY, W. S. 1995. Erythromycin inhibits the assembly of the large ribosomal subunit in growing *Escherichia coli* cells. *Curr Microbiol,* 30**,** 273-9.

CHOLON, D. M., CUTTER, D., RICHARDSON, S. K., SETHI, S., MURPHY, T. F., LOOK, D. C. & ST GEME, J. W., 3RD 2008. Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing quantities of the HMW1 and HMW2 adhesins. *Infect Immun,* 76**,** 4463-8.

CHUNG, K. F. & ADCOCK, I. M. 2008. Multifaceted mechanisms in COPD: inflammation, immunity, and tissue repair and destruction. *Eur Respir J,* 31**,** 1334-56.

CLANCY, J., PETITPAS, J., DIB-HAJJ, F., YUAN, W., CRONAN, M., KAMATH, A. V., BERGERON, J. & RETSEMA, J. A. 1996. Molecular cloning and functional analysis of a novel macrolide-resistance determinant, mefA, from *Streptococcus pyogenes*. *Mol Microbiol,* 22**,** 867-79.

CLARK, C., BOZDOGAN, B., PERIC, M., DEWASSE, B., JACOBS, M. R. & APPELBAUM, P. C. 2002. In vitro selection of resistance in *Haemophilus influenzae* by amoxicillin-clavulanate, cefpodoxime, cefprozil, azithromycin, and clarithromycin. *Antimicrob Agents Chemother,* 46**,** 2956-62.

CLEMENTI, C. F., HAKANSSON, A. P. & MURPHY, T. F. 2014. Internalization and trafficking of nontypeable *Haemophilus influenzae* in human respiratory epithelial cells and roles of IgA1 proteases for optimal invasion and persistence. *Infect Immun,* 82**,** 433-44.

CLEMMER, K. M., STURGILL, G. M., VEENSTRA, A. & RATHER, P. N. 2006. Functional characterization of *Escherichia coli* GlpG and additional rhomboid proteins using an aarA mutant of *Providencia stuartii*. *J Bacteriol,* 188**,** 3415-9.

CODY, A. J., FIELD, D., FEIL, E. J., STRINGER, S., DEADMAN, M. E., TSOLAKI, A. G., GRATZ, B., BOUCHET, V., GOLDSTEIN, R., HOOD, D. W. & MOXON, E. R. 2003. High rates of recombination in otitis media isolates of non-typeable *Haemophilusinfluenzae*. *Infect Genet Evol,* 3**,** 57-66.

COHEN, R., BINGEN, E., LEVY, C., THOLLOT, F., BOUCHERAT, M., DERKX, V. & VARON, E. 2012. Nasopharyngeal flora in children with acute otitis media before and after implementation of 7 valent pneumococcal conjugate vaccine in France. *BMCInfect Dis,* 12**,** 52.

COLLABORATORS, G. B. D. M. M. 2016. Global, regional, and national levels ofmaternal mortality, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet,* 388**,** 1775-1812.

COLLINS, S., LITT, D. J., FLYNN, S., RAMSAY, M. E., SLACK, M. P. & LADHANI, S. N. 2015. Neonatal Invasive *Haemophilus influenzae* Disease in England and Wales: Epidemiology, Clinical Characteristics, and Outcome. *Clin Infect Dis*.

CONNOR, T. R., CORANDER, J. & HANAGE, W. P. 2012. Population subdivision and the detection of recombination in non-typable *Haemophilus influenzae*. *Microbiology,* 158**,** 2958-64.

COTTER, S. E., YEO, H. J., JUEHNE, T. & ST GEME, J. W., 3RD 2005. Architecture and adhesive activity of the *Haemophilus influenzae* Hsf adhesin. *J Bacteriol,* 187**,** 4656-64.

CRIPPS, A. W. 2010. Nontypeable *Haemophilus influenzae* and childhood pneumonia. *P NG Med J,* 53**,** 147-50.

CROSBIE, P. A. & WOODHEAD, M. 2013. Should we give long-term macrolide therapy for COPD? *Thorax,* 68**,** 966.

CROUCHER, N. J., PAGE, A. J., CONNOR, T. R., DELANEY, A. J., KEANE, J. A., BENTLEY, S. D., PARKHILL, J. & HARRIS, S. R. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res,* 43**,** e15.

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

CUTTER, D., MASON, K. W., HOWELL, A. P., FINK, D. L., GREEN, B. A. & ST GEME, J. W., 3RD 2002. Immunization with *Haemophilus influenzae* Hap adhesin protects against nasopharyngeal colonization in experimental mice. *J Infect Dis,* 186**,** 1115-21.

DAIYASU, H., OSAKA, K., ISHINO, Y. & TOH, H. 2001. Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS Lett,* 503**,** 1-6.

DALYS, G. B. D., COLLABORATORS, H., MURRAY, C. J., BARBER, R. M., FOREMAN, K. J., ABBASOGLU OZGOREN, A., ABD-ALLAH, F., ABERA, S. F., ABOYANS, V., ABRAHAM, J. P., ABUBAKAR, I., ABU-RADDAD, L. J., ABU-RMEILEH, N. M., ACHOKI, T., ACKERMAN, I. N., ADEMI, Z., ADOU, A. K., ADSUAR, J. C., AFSHIN, A., AGARDH, E. E., ALAM, S. S., ALASFOOR, D., ALBITTAR, M. I., ALEGRETTI, M. A., ALEMU, Z. A., ALFONSO-CRISTANCHO, R., ALHABIB, S., ALI, R., ALLA, F., ALLEBECK, P., ALMAZROA, M. A., ALSHARIF, U., ALVAREZ, E., ALVIS-GUZMAN, N., AMARE, A. T., AMEH, E. A., AMINI, H., AMMAR, W., ANDERSON, H. R., ANDERSON, B. O., ANTONIO, C. A., ANWARI, P., ARNLOV, J., ARSIC ARSENIJEVIC, V. S., ARTAMAN, A., ASGHAR, R. J., ASSADI, R., ATKINS, L. S., AVILA, M. A., AWUAH, B., BACHMAN, V. F., BADAWI, A., BAHIT, M. C., BALAKRISHNAN, K., BANERJEE, A., BARKER-COLLO, S. L., BARQUERA, S., BARREGARD, L., BARRERO, L. H., BASU, A., BASU, S., BASULAIMAN, M. O., BEARDSLEY, J., BEDI, N., BEGHI, E., BEKELE, T., BELL, M. L., BENJET, C., BENNETT, D. A., BENSENOR, I. M., BENZIAN, H., BERNABE, E., BERTOZZI-VILLA, A., BEYENE, T. J., BHALA, N., BHALLA, A., BHUTTA, Z. A., BIENHOFF, K., BIKBOV, B., BIRYUKOV, S., BLORE, J. D., BLOSSER, C. D., BLYTH, F. M., BOHENSKY, M. A., BOLLIGER, I. W., BORA BASARA, B., BORNSTEIN, N. M., BOSE, D., BOUFOUS, S., BOURNE, R. R., BOYERS, L. N., BRAININ, M., BRAYNE, C. E., BRAZINOVA, A., BREITBORDE, N. J., BRENNER, H., BRIGGS, A. D., BROOKS, P. M., BROWN, J. C., BRUGHA, T. S., et al. 2015. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. *Lancet,* 386**,** 2145-91.

DARCH, S. E., MCNALLY, A., HARRISON, F., CORANDER, J., BARR, H. L., PASZKIEWICZ, K., HOLDEN, S., FOGARTY, A., CRUSZ, S. A. & DIGGLE, S. P. 2015. Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. *Sci Rep,* 5**,** 7649.

DAVIES, G. & WILSON, R. 2004. Prophylactic antibiotic treatment of bronchiectasis with azithromycin. *Thorax,* 59**,** 540-1.

DAVIS, G. S., MARINO, S., MARRS, C. F., GILSDORF, J. R., DAWID, S. & KIRSCHNER, D. E. 2014a. Phase variation and host immunity against high molecular weight (HMW) adhesins shape population dynamics of nontypeable *Haemophilus influenzae* within human hosts. *J Theor Biol,* 355**,** 208-18.

DAVIS, G. S., PATEL, M., HAMMOND, J., ZHANG, L., DAWID, S., MARRS, C. F. & GILSDORF, J. R. 2014b. Prevalence, distribution, and sequence diversity of hmwA among commensal and otitis media non-typeable *Haemophilus influenzae*. *Infect Genet Evol,* 28**,** 223-32.

DAVYDOVA, N., STRELTSOV, V., WILCE, M., LILJAS, A. & GARBER, M. 2002. L22 ribosomal protein and effect of its mutation on ribosome resistance to erythromycin. *JMol Biol,* 322**,** 635-44.

DE CHIARA, M., HOOD, D., MUZZI, A., PICKARD, D. J., PERKINS, T., PIZZA, M., DOUGAN, G., RAPPUOLI, R., MOXON, E. R., SORIANI, M. & DONATI, C. 2014. Genomesequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. *Proc Natl Acad Sci U S A,* 111**,** 5439-44.

DE GIER, C., KIRKHAM, L. A. & NORSKOV-LAURITSEN, N. 2015. Complete Deletion ofthe Fucose Operon in *Haemophilus influenzae* Is Associated with a Cluster in Multilocus Sequence Analysis-Based Phylogenetic Group II Related to *Haemophilus haemolyticus*: Implications for Identification and Typing. *J Clin Microbiol,* 53**,**3773-8.

DE GIER, C., PICKERING, J. L., RICHMOND, P. C., THORNTON, R. B. & KIRKHAM, L. A.2016. Duplex Quantitative PCR Assay for Detection of *Haemophilus influenzae* That Distinguishes Fucose- and Protein D-Negative Strains. *J Clin Microbiol,* 54**,** 2380-3.

DEADMAN, M. E., HERMANT, P., ENGSKOG, M., MAKEPEACE, K., MOXON, E. R., SCHWEDA, E. K. & HOOD, D. W. 2009. Lex2B, a phase-variable glycosyltransferase, addseither a glucose or a galactose to *Haemophilus influenzae* lipopolysaccharide. *Infect Immun,* 77**,** 2376-84.

DEBOUCK, C., RICCIO, A., SCHUMPERLI, D., MCKENNEY, K., JEFFERS, J., HUGHES, C., ROSENBERG, M., HEUSTERSPREUTE, M., BRUNEL, F. & DAVISON, J. 1985. Structure of the galactokinase gene of *Escherichia coli*, the last (?) gene of the gal operon. *Nucleic Acids Res,* 13**,** 1841-53.

DEL GROSSO, M., IANNELLI, F., MESSINA, C., SANTAGATI, M., PETROSILLO, N., STEFANI, S., POZZI, G. & PANTOSTI, A. 2002. Macrolide efflux genes mef(A) and mef(E) are carried by different genetic elements in *Streptococcus pneumoniae*. *J Clin Microbiol,* 40**,** 774-8.

DEMARIA, T. F., MURWIN, D. M. & LEAKE, E. R. 1996. Immunization with outer membrane protein P6 from nontypeable *Haemophilus influenzae* induces bactericidal antibody and affords protection in the chinchilla model of otitis media. *Infect Immun,* 64**,** 5187-92.

DINER, E. J. & HAYES, C. S. 2009. Recombineering reveals a diverse collection ofribosomal proteins L4 and L22 that confer resistance to macrolide antibiotics. *J Mol Biol,* 386**,** 300-15.

DJOKIC, S., KOBREHEL, G. & LAZAREVSKI, G. 1987. Erythromycin series. XII. Antibacterial in vitro evaluation of 10-dihydro-10-deoxo-11-azaerythromycin A: synthesis and structure-activity relationship of its acyl derivatives. *J Antibiot (Tokyo),* 40**,** 1006-15.

DOBRZANSKI, W. T. & PULAWSKA, E. 1960. [Studies on the role of mutation inerythromycin resistance in strains of staphylococci belonging to various phage groups]. *Med Dosw Mikrobiol,* 12**,** 331-7.

DUIM, B., VAN ALPHEN, L., EIJK, P., JANSEN, H. M. & DANKERT, J. 1994. Antigenic drift of non-encapsulated *Haemophilus influenzae* major outer membrane protein P2 in patients with chronic bronchitis is caused by point mutations. *Mol Microbiol,* 11**,** 1181-9.

DWIVEDI, R., NOTHAFT, H., GARBER, J., XIN KIN, L., STAHL, M., FLINT, A., VAN VLIET, A. H., STINTZI, A. & SZYMANSKI, C. M. 2016. L-fucose influences chemotaxis and biofilm formation in *Campylobacter jejuni*. *Mol Microbiol,* 101**,** 575-89.

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.

ECEVIT, I. Z., MCCREA, K. W., MARRS, C. F. & GILSDORF, J. R. 2005. Identification of new *hmwA* alleles from nontypeable *Haemophilus influenzae*. *Infect Immun,* 73**,**1221-5.

ECEVIT, I. Z., MCCREA, K. W., PETTIGREW, M. M., SEN, A., MARRS, C. F. & GILSDORF, J. R. 2004. Prevalence of the *hifBC*, *hmw1A*, *hmw2A*, *hmwC*, and *hia* Genes in *Haemophilus influenzae* Isolates. *J Clin Microbiol,* 42**,** 3065-72.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

EHRLICH, G. D., AHMED, A., EARL, J., HILLER, N. L., COSTERTON, J. W., STOODLEY, P., POST, J. C., DEMEO, P. & HU, F. Z. 2010. The distributed genome hypothesis as a rubric for understanding evolution in situ during chronic bacterial biofilm infectious processes. *FEMS Immunol Med Microbiol,* 59**,** 269-79.

EHRLICH, G. D., VEEH, R., WANG, X., COSTERTON, J. W., HAYES, J. D., HU, F. Z., DAIGLE, B. J., EHRLICH, M. D. & POST, J. C. 2002. Mucosal biofilm formation on middle-ear mucosa in the chinchilla model of otitis media. *JAMA,* 287**,** 1710-5.

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.

ERWIN, A. L., ALLEN, S., HO, D. K., BONTHUIS, P. J., JARISCH, J., NELSON, K. L., TSAO, D.L., UNRATH, W. C., WATSON, M. E., JR., GIBSON, B. W., APICELLA, M. A. & SMITH, A. L. 2006a. Role of lgtC in resistance of nontypeable *Haemophilus influenzae* strain R2866 to human serum. *Infect Immun,* 74**,** 6226-35.

ERWIN, A. L., BONTHUIS, P. J., GEELHOOD, J. L., NELSON, K. L., MCCREA, K. W., GILSDORF, J. R. & SMITH, A. L. 2006b. Heterogeneity in tandem octanucleotides within *Haemophilus influenzae* lipopolysaccharide biosynthetic gene losA affects serum resistance. *Infect Immun,* 74**,** 3408-14.

ERWIN, A. L., NELSON, K. L., MHLANGA-MUTANGADURA, T., BONTHUIS, P. J., GEELHOOD, J. L., MORLIN, G., UNRATH, W. C., CAMPOS, J., CROOK, D. W., FARLEY, M. M., HENDERSON, F. W., JACOBS, R. F., MUHLEMANN, K., SATOLA, S. W., VAN ALPHEN, L., GOLOMB, M. & SMITH, A. L. 2005. Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. *Infect Immun,* 73**,** 5853-63.

ERWIN, A. L., SANDSTEDT, S. A., BONTHUIS, P. J., GEELHOOD, J. L., NELSON, K. L., UNRATH, W. C., DIGGLE, M. A., THEODORE, M. J., PLEATMAN, C. R., MOTHERSHED, E. A., SACCHI, C. T., MAYER, L. W., GILSDORF, J. R. & SMITH, A. L. 2008. Analysis of genetic relatedness of *Haemophilus influenzae* isolates by multilocus sequence typing. *J Bacteriol,* 190**,** 1473-83.

ESMAILY, F., VAEZZADEH, F. & SHARIFI-YAZDI, M. K. 2006. Efficacy of immunization with outer membrane proteins for induction of pulmonary clearance of nontypeable

Haemophilus influenzae in a rat respiratory model. *Iran J Allergy Asthma Immunol,* 5**,** 57-61.

EUBA, B., MOLERES, J., VIADAS, C., RUIZ DE LOS MOZOS, I., VALLE, J., BENGOECHEA, J. A.& GARMENDIA, J. 2015. Relative Contribution of P5 and Hap Surface Proteins to Nontypable *Haemophilus influenzae* Interplay with the Host Upper and Lower Airways. *PLoS One,* 10**,** e0123154.

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.

FENGER, M. G., RIDDERBERG, W., OLESEN, H. V. & NORSKOV-LAURITSEN, N. 2012. Low occurrence of 'non-haemolytic *Haemophilus haemolyticus'* misidentified as *Haemophilus influenzae* in cystic fibrosis respiratory specimens, and frequent recurrence of persistent *H. influenzae* clones despite antimicrobial treatment. *Int J Med Microbiol,* 302**,** 315-9.

FERLA, M. P. & PATRICK, W. M. 2014. Bacterial methionine biosynthesis. *Microbiology,*160**,** 1571-84.

FERNAAYS, M. M., LESSE, A. J., CAI, X. & MURPHY, T. F. 2006a. Characterization of igaB, a second immunoglobulin A1 protease gene in nontypeable *Haemophilusinfluenzae*. *Infect Immun,* 74**,** 5860-70.

FERNAAYS, M. M., LESSE, A. J., SETHI, S., CAI, X. & MURPHY, T. F. 2006b. Differential genome contents of nontypeable *Haemophilus influenzae* strains from adults with chronic obstructive pulmonary disease. *Infect Immun,* 74**,** 3366-74.

FERNANDES, L. M., A. 2014. Understanding Gender Differences in the Clinical Presentation of Chronic Obstructive Pulmonary Disease: A Cross Sectional Study. *International Journal of Medical Science and Public Health* 3.

FERNANDES, P. B. & HARDY, D. J. 1988. Comparative in vitro potencies of nine new macrolides. *Drugs Exp Clin Res,* 14**,** 445-51.

FINK, D. L., BUSCHER, A. Z., GREEN, B., FERNSTEN, P. & ST GEME, J. W., 3RD 2003. The *Haemophilus influenzae* Hap autotransporter mediates microcolony formation and adherence to epithelial cells and extracellular matrix via binding regions in the C-terminal end of the passenger domain. *Cell Microbiol,* 5**,** 175-86.

FINK, D. L., COPE, L. D., HANSEN, E. J. & GEME, J. W., 3RD 2001. The *Haemophilusinfluenzae* Hap autotransporter is a chymotrypsin clan serine protease and undergoes autoproteolysis via an intermolecular mechanism. *J Biol Chem,* 276**,** 39492-500.

FINK, D. L., GREEN, B. A. & ST GEME, J. W., 3RD 2002. The *Haemophilus influenzae* Hap autotransporter binds to fibronectin, laminin, and collagen IV. *Infect Immun,* 70**,** 4902-7.

FINK, D. L. & ST GEME, J. W., 3RD 2003. Chromosomal expression of the *Haemophilus influenzae* Hap autotransporter allows fine-tuned regulation of adhesive potential via inhibition of intermolecular autoproteolysis. *J Bacteriol,* 185**,** 1608-15.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & ET AL. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* 269**,** 496-512.

FLETCHER, H. M., SCHENKEIN, H. A., MORGAN, R. M., BAILEY, K. A., BERRY, C. R.& MACRINA, F. L. 1995. Virulence of a *Porphyromonas gingivalis* W83 mutant defective in the prtH gene. *Infect Immun,* 63**,** 1521-8.

FLETCHER, M. J., UPTON, J., TAYLOR-FISHWICK, J., BUIST, S. A., JENKINS, C., HUTTON, J., BARNES, N., VAN DER MOLEN, T., WALSH, J. W., JONES, P. & WALKER, S. 2011. COPD uncovered: an international survey on the impact of chronic obstructive pulmonary disease [COPD] on a working age population. *BMC Public Health,* 11**,** 612.

FORBES, K. J., BRUCE, K. D., BALL, A. & PENNINGTON, T. H. 1992. Variation in lengthand sequence of porin (ompP2) alleles of non-capsulate *Haemophilus influenzae*. *Mol Microbiol,* 6**,** 2107-12.

FORBES KJ, S. D., WEISSFIELD AS. 2002. Diagnostic Microbiology Eleventh ed. St. Louis.

FOREY, B. A., THORNTON, A. J. & LEE, P. N. 2011. Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. *BMC Pulm Med,* 11**,** 36.

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, K. L., ATACK, J. M., SRIKHANTA, Y. N., ECKERT, A., NOVOTNY, L. A., BAKALETZ, L. O. & JENNINGS, M. P. 2014. Selection for phase variation of LOS biosynthetic genes frequently occurs in progression of non-typeable *Haemophilus influenzae* infection from the nasopharynx to the middle ear of human patients. *PLoS One,* 9**,**e90505.

FOX, P. M., CLIMO, M. W. & ARCHER, G. L. 2007. Lack of relationship between purine biosynthesis and vancomycin resistance in *Staphylococcus aureus*: a cautionary tale for microarray interpretation. *Antimicrob Agents Chemother,* 51**,**1274-80.

FRANCISCO, A. P., VAZ, C., MONTEIRO, P. T., MELO-CRISTINO, J., RAMIREZ, M. & CARRICO, J. A. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics,* 13**,** 87.

FRICK, A. G., JOSEPH, T. D., PANG, L., RABE, A. M., ST GEME, J. W., 3RD & LOOK, D. C. 2000. *Haemophilus influenzae* stimulates ICAM-1 expression on respiratory epithelial cells. *J Immunol,* 164**,** 4185-96.

FRICKMANN, H., PODBIELSKI, A., ESSIG, A., SCHWARZ, N. G. & ZAUTNER, A. E. 2014. Difficulties in species identification within the genus *Haemophilus* - A pilot study addressing a significant problem for routine diagnostics. *Eur J Microbiol Immunol (Bp),* 4**,** 99-105.

FU, L., NIU, B., ZHU, Z., WU, S. & LI, W. 2012. CD-HIT: accelerated for clustering thenext-generation sequencing data. *Bioinformatics,* 28**,** 3150-2.

FUNG, W. W., O'DWYER, C. A., SINHA, S., BRAUER, A. L., MURPHY, T. F., KROLL, J. S.& LANGFORD, P. R. 2006. Presence of copper- and zinc-containing superoxide dismutase in commensal *Haemophilus haemolyticus* isolates can be used as a marker to discriminate them from nontypeable *H. influenzae* isolates. *J Clin Microbiol,* 44**,** 4222-6.

GABASHVILI, I. S., GREGORY, S. T., VALLE, M., GRASSUCCI, R., WORBS, M., WAHL, M. C., DAHLBERG, A. E. & FRANK, J. 2001. The polypeptide tunnel system in the ribosome and its gating in erythromycin resistance mutants of L4 and L22. *Mol Cell,* 8**,** 181-8.

GALARZA, P. G., ABAD, R., CANIGIA, L. F., BUSCEMI, L., PAGANO, I., OVIEDO, C. & VAZQUEZ, J. A. 2010. New mutation in 23S rRNA gene associated with high level of azithromycin resistance in *Neisseria gonorrhoeae*. *Antimicrob Agents Chemother,* 54**,** 1652-3.

GALLAHER, T. K., WU, S., WEBSTER, P. & AGUILERA, R. 2006. Identification of biofilm proteins in non-typeable *Haemophilus Influenzae*. *BMC Microbiol,* 6**,** 65.

GARMENDIA, J., VIADAS, C., CALATAYUD, L., MELL, J. C., MARTI-LLITERAS, P., EUBA, B., LLOBET, E., GIL, C., BENGOECHEA, J. A., REDFIELD, R. J. & LINARES, J. 2014. Characterization of nontypable *Haemophilus influenzae* isolates recoveredfrom adult patients with underlying chronic lung disease reveals genotypic and phenotypic traits associated with persistent infection. *PLoS One,* 9**,**e97020.

GARZA-RAMOS, G., XIONG, L., ZHONG, P. & MANKIN, A. 2001. Binding site ofmacrolide antibiotics on the ribosome: new resistance mutation identifies a specific interaction of ketolides with rRNA. *J Bacteriol,* 183**,** 6898-907.

GAVI, A. 2016. *pentavalent vaccine support* [Online]. Available: http://www.gavi.org/support/nvs/pentavalent/ [Accessed 17th January 2017].

GAWRONSKI, J. D., WONG, S. M., GIANNOUKOS, G., WARD, D. V. & AKERLEY, B. J. 2009. Tracking insertion mutants within libraries by deep sequencing and agenome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci U S A,* 106**,** 16422-7.

GELUK, F., EIJK, P. P., VAN HAM, S. M., JANSEN, H. M. & VAN ALPHEN, L. 1998. The fimbria gene cluster of nonencapsulated *Haemophilus influenzae*. *Infect Immun,* 66**,** 406-17.

GIELEN, V., JOHNSTON, S. L. & EDWARDS, M. R. 2010. Azithromycin induces anti-viral responses in bronchial epithelial cells. *Eur Respir J,* 36**,** 646-54.

GILSDORF, J. R., MCCREA, K. W. & MARRS, C. F. 1997. Role of pili in *Haemophilusinfluenzae* adherence and colonization. *Infect Immun,* 65**,** 2997-3002.

GIUFRE, M., CARATTOLI, A., CARDINES, R., MASTRANTONIO, P. & CERQUETTI, M. 2008. Variation in expression of HMW1 and HMW2 adhesins in invasive nontypeable *Haemophilus influenzae* isolates. *BMC Microbiol,* 8**,** 83.

GIUFRE, M., MUSCILLO, M., SPIGAGLIA, P., CARDINES, R., MASTRANTONIO, P.& CERQUETTI, M. 2006. Conservation and diversity of HMW1 and HMW2 adhesin binding domains among invasive nontypeable *Haemophilus influenzae* isolates. *Infect Immun,* 74**,** 1161-70.

GODIER, A. & HUNT, B. J. 2013. Plasminogen receptors and their role in the pathogenesis of inflammatory, autoimmune and malignant disease. *J Thromb Haemost,* 11**,**26-34.

GOLDSBY, R. K., TJ. OSBOURNE, BA. 2000. *Kuby Immunology,* New York, W.H Freeman and Company.

GOMEZ, J., BANOS, V., SIMARRO, E., LORENZO CRUZ, M., RUIZ GOMEZ, J., LATOUR, J., GARCIA MARTIN, E., CANTERAS, M. & VALDES, M. 2000. [Prospective, comparative study (1994-1998) of the influence of short-term prophylactic treatment with azithromycin on patients with advanced COPD]. *Rev Esp Quimioter,* 13**,**379-83.

GOODGAL, S. H. & HERRIOTT, R. M. 1961. Studies on transformations of *Haemophilus influenzae*. I. Competence. *J Gen Physiol,* 44**,** 1201-27.

GORIS, J., KONSTANTINIDIS, K. T., KLAPPENBACH, J. A., COENYE, T., VANDAMME, P.& TIEDJE, J. M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol,* 57**,** 81-91.

GRAD, Y. H., HARRIS, S. R., KIRKCALDY, R. D., GREEN, A. G., MARKS, D. S., BENTLEY, S. D., TREES, D. & LIPSITCH, M. 2016. Genomic Epidemiology of Gonococcal Resistance to Extended-Spectrum Cephalosporins, Macrolides, and Fluoroquinolones in the United States, 2000-2013. *J Infect Dis,* 214**,** 1579-1587.

GRASS, S. & ST GEME, J. W., 3RD 2000. Maturation and secretion of the non-typable *Haemophilus influenzae* HMW1 adhesin: roles of the N-terminal and C-terminal domains. *Mol Microbiol,* 36**,** 55-67.

GREEN, B. A., BARANYI, E., REILLY, T. J., SMITH, A. L. & ZLOTNICK, G. W. 2005. Certain site-directed, nonenzymatically active mutants of the *Haemophilus influenzae* P4 lipoprotein are able to elicit bactericidal antibodies. *Infect Immun,* 73**,**4454-7.

GREEN, B. A., FARLEY, J. E., QUINN-DEY, T., DEICH, R. A. & ZLOTNICK, G. W. 1991. The e(P4) outer membrane protein of *Haemophilus influenzae*: biologic activity of anti-e serum and cloning and sequencing of the structural gene. *Infect Immun,* 59**,** 3191-8.

GRIFFIN, R., COX, A. D., MAKEPEACE, K., RICHARDS, J. C., MOXON, E. R. & HOOD, D. W. 2005. Elucidation of the monoclonal antibody 5G8-reactive, virulence-associated lipopolysaccharide epitope of *Haemophilus influenzae* and its role in bacterial resistance to complement-mediated killing. *Infect Immun,* 73**,** 2213-21.

HABIB, M., PORTER, B. D. & SATZKE, C. 2014. Capsular serotyping of *Streptococcus pneumoniae* using the Quellung reaction. *J Vis Exp*, e51208.

HAIGHT, T. H. & FINLAND, M. 1952. Laboratory and clinical studies on erythromycin. *NEngl J Med,* 247**,** 227-32.

HALL, W. H. 1960. Rate of development of resistance in *Staphylococcus aureus* to erythromycin in combination with oleandomycin, vancomycin, or dihydrostreptomycin. *J Lab Clin Med,* 56**,** 83-104.

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-ear mucosa of children with chronic otitis media. *JAMA,* 296**,** 202-11.

HALLSTROM, T., BLOM, A. M., ZIPFEL, P. F. & RIESBECK, K. 2009. Nontypeable *Haemophilus influenzae* protein E binds vitronectin and is important for serum resistance. *J Immunol,* 183**,** 2593-601.

HALLSTROM, T., SINGH, B., RESMAN, F., BLOM, A. M., MORGELIN, M. & RIESBECK, K. 2011. *Haemophilus influenzae* protein E binds to the extracellular matrix by concurrently interacting with laminin and vitronectin. *J Infect Dis,* 204**,** 1065-74.

HALLSTROM, T., TRAJKOVSKA, E., FORSGREN, A. & RIESBECK, K. 2006. *Haemophilus influenzae* surface fibrils contribute to serum resistance by interacting with vitronectin. *J Immunol,* 177**,** 430-6.

HAMMA, T. & FERRE-D'AMARE, A. R. 2006. Pseudouridine synthases. *Chem Biol,* 13**,**1125-35.

HAMMITT, L. L., AKECH, D. O., MORPETH, S. C., KARANI, A., KIHUHA, N., NYONGESA, S., BWANAALI, T., MUMBO, E., KAMAU, T., SHARIF, S. K. & SCOTT, J. A. 2014. Population effect of 10-valent pneumococcal conjugate vaccine on nasopharyngeal carriage of *Streptococcus pneumoniae* and non-typeable *Haemophilus influenzae* in Kilifi, Kenya: findings from cross-sectional carriage studies. *Lancet Glob Health,*2**,** e397-405.

HANSEN, C. R., PRESSLER, T., HOIBY, N. & JOHANSEN, H. K. 2009. Long-term, low-dose azithromycin treatment reduces the incidence but increases macrolide resistance in *Staphylococcus aureus* in Danish CF patients. *J Cyst Fibros,* 8**,** 58-62.

HARDY, D. J., HENSEY, D. M., BEYER, J. M., VOJTKO, C., MCDONALD, E. J. & FERNANDES, P. B. 1988. Comparative in vitro activities of new 14-, 15-, and16-membered macrolides. *Antimicrob Agents Chemother,* 32**,** 1710-9.

HARE, K. M., BINKS, M. J., GRIMWOOD, K., CHANG, A. B., LEACH, A. J. & SMITH-VAUGHAN, H. 2012. Culture and PCR detection of *Haemophilus influenzae* and *Haemophilus haemolyticus* in Australian Indigenous children with bronchiectasis. *J Clin Microbiol,* 50**,** 2444-5.

HARE, K. M., GRIMWOOD, K., CHANG, A. B., CHATFIELD, M. D., VALERY, P. C., LEACH, A. J., SMITH-VAUGHAN, H. C., MORRIS, P. S., BYRNES, C. A., TORZILLO, P. J. & CHENG, A. C. 2015. Nasopharyngeal carriage and macrolide resistance in Indigenous children with bronchiectasis randomized to long-term azithromycin or placebo. *Eur J Clin Microbiol Infect Dis,* 34**,** 2275-85.

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.

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.

HARRISON, A., DYER, D. W., GILLASPY, A., RAY, W. C., MUNGUR, R., CARSON, M. B., ZHONG, H., GIPSON, J., GIPSON, M., JOHNSON, L. S., LEWIS, L., BAKALETZ, L. O. & MUNSON, R. S., JR. 2005. Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. *J Bacteriol,* 187**,** 4627-36.

HARVALA, H., ROSENDAL, T., LAHTI, E., ENGVALL, E. O., BRYTTING, M., WALLENSTEN, A.& LINDBERG, A. 2016. Epidemiology of *Campylobacter jejuni* infections in Sweden, November 2011-October 2012: is the severity of infection associated with *C.jejuni* sequence type? *Infect Ecol Epidemiol,* 6**,** 31079.

HE, Z. Y., OU, L. M., ZHANG, J. Q., BAI, J., LIU, G. N., LI, M. H., DENG, J. M., MACNEE, W. & ZHONG, X. N. 2010. Effect of 6 months of erythromycin treatment oninflammatory cells in induced sputum and exacerbations in chronic obstructive pulmonary disease. *Respiration,* 80**,** 445-52.

HEITHOFF, D. M., SINSHEIMER, R. L., LOW, D. A. & MAHAN, M. J. 1999. An essential role for DNA adenine methylation in bacterial virulence. *Science,* 284**,** 967-70.

HENDRIXSON, D. R. & ST GEME, J. W., 3RD 1998. The *Haemophilus influenzae* Hapserine protease promotes adherence and microcolony formation, potentiated by a soluble host protein. *Mol Cell,* 2**,** 841-50.

HILL, D. J., TOLEMAN, M. A., EVANS, D. J., VILLULLAS, S., VAN ALPHEN, L. & VIRJI, M. 2001. The variable P5 proteins of typeable and non-typeable *Haemophilusinfluenzae* target human CEACAM1. *Mol Microbiol,* 39**,** 850-62.

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.

HINZ, R., ZAUTNER, A. E., HAGEN, R. M. & FRICKMANN, H. 2015. Difficult identification of *Haemophilus influenzae*, a typical cause of upper respiratory tract infections, inthe microbiological diagnostic routine. *Eur J Microbiol Immunol (Bp),* 5**,** 62-7.

HO, D. K., RAM, S., NELSON, K. L., BONTHUIS, P. J. & SMITH, A. L. 2007. *lgtC* expression modulates resistance to C4b deposition on an invasive nontypeable *Haemophilus influenzae*. *J Immunol,* 178**,** 1002-12.

HOBEN, P., ROYAL, N., CHEUNG, A., YAMAO, F., BIEMANN, K. & SOLL, D. 1982. Escherichia coli glutaminyl-tRNA synthetase. II. Characterization of the *glnS* gene product. *JBiol Chem,* 257**,** 11644-50.

HOGE, C. W., GAMBEL, J. M., SRIJAN, A., PITARANGSI, C. & ECHEVERRIA, P. 1998. Trends in antibiotic resistance among diarrheal pathogens isolated in Thailand over 15 years. *Clin Infect Dis,* 26**,** 341-5.

HOGG, J. S., HU, F. Z., JANTO, B., BOISSY, R., HAYES, J., KEEFE, R., POST, J. C. & EHRLICH, G. D. 2007. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol,* 8**,** R103.

HOOD, D. W., COX, A. D., GILBERT, M., MAKEPEACE, K., WALSH, S., DEADMAN, M. E.,CODY, A., MARTIN, A., MANSSON, M., SCHWEDA, E. K., BRISSON, J. R., RICHARDS, J. C., MOXON, E. R. & WAKARCHUK, W. W. 2001. Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from *Haemophilus influenzae*. *Mol Microbiol,* 39**,** 341-50.

HOTOMI, M., FUJIHARA, K., SAKAI, A., BILLAL, D. S., SHIMADA, J., SUZUMOTO, M.& YAMANAKA, N. 2006. Antimicrobial resistance of *Haemophilus influenzae* isolated from the nasopharynx of Japanese children with acute otitis media. *Acta Otolaryngol,* 126**,** 240-7.

HOTOMI, M., IKEDA, Y., SUZUMOTO, M., YAMAUCHI, K., GREEN, B. A., ZLOTNICK, G., BILLAL, D. S., SHIMADA, J., FUJIHARA, K. & YAMANAKA, N. 2005. A recombinant P4 protein of *Haemophilus influenzae* induces specific immune responses biologically active against nasopharyngeal colonization in mice after intranasalimmunization. *Vaccine,* 23**,** 1294-300.

HOTOMI, M., KONO, M., TOGAWA, A., ARAI, J., TAKEI, S., IKEDA, Y., OGAMI, M., MURPHY, T. F. & YAMANAKA, N. 2010. *Haemophilus influenzae* and *Haemophilus haemolyticus* in tonsillar cultures of adults with acute pharyngotonsillitis. *Auris Nasus Larynx,* 37**,** 594-600.

HOTOMI, M., SAITO, T. & YAMANAKA, N. 1998. Specific mucosal immunity and enhanced nasopharyngeal clearance of nontypeable *Haemophilus influenzae* afterintranasal immunization with outer membrane protein P6 and cholera toxin. *Vaccine,* 16**,** 1950-6.

HOTOMI, M., YAMANAKA, N., SAITO, T., SHIMADA, J., SUZUMOTO, M., SUETAKE, M.& FADEN, H. 1999. Antibody responses to the outer membrane protein P6 of non-typeable *Haemophilus influenzae* and pneumococcal capsular polysaccharides in otitis-prone children. *Acta Otolaryngol,* 119**,** 703-7.

HOTOMI, M., YAMANAKA, N., SHIMADA, J., SUZUMOTO, M., IKEDA, Y., SAKAI, A., ARAI, J.& GREEN, B. 2002. Intranasal immunization with recombinant outer membrane protein P6 induces specific immune responses against nontypeable *Haemophilus influenzae*. *Int J Pediatr Otorhinolaryngol,* 65**,** 109-16.

HOTOMI, M., YOKOYAMA, M., KUKI, K., TOGAWA, A. & YAMANAKA, N. 1996. Study on specific mucosal immunity by intranasal immunization of outer membrane protein P6 of *Haemophilus influenzae* with cholera toxin B subunit. *Acta Otolaryngol Suppl,* 523**,** 150-2.

HU, F., RISHISHWAR, L., SIVADAS, A., MITCHELL, G. J., JORDAN, I. K., MURPHY, T. F., GILSDORF, J. R., MAYER, L. W. & WANG, X. 2016. Comparative genomic analysis of *Haemophilus haemolyticus* and non-typeable *Haemophilus influenzae* and a new testing scheme for their discrimination. *J Clin Microbiol*.

HUA, C. Z., HU, W. L., SHANG, S. Q., LI, J. P., HONG, L. Q. & YAN, J. 2016. Serum Concentrations of Antibodies against Outer Membrane Protein P6, Protein D, and T- and B-Cell Combined Antigenic Epitopes of Nontypeable *Haemophilusinfluenzae* in Children and Adults of Different Ages. *Clin Vaccine Immunol,* 23**,**155-61.

IEVA, R. & BERNSTEIN, H. D. 2009. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci U S A,* 106**,** 19120-5.

INGHAMMAR, M., EKBOM, A., ENGSTROM, G., LJUNGBERG, B., ROMANUS, V., LOFDAHL, C. G. & EGESTEN, A. 2010. COPD and the risk of tuberculosis--a population-based cohort study. *PLoS One,* 5**,** e10138.

INOUYE, M., CONWAY, T. C., ZOBEL, J. & HOLT, K. E. 2012. Short read sequencetyping (SRST): multi-locus sequence types from short reads. *BMC Genomics,* 13**,**338.

INOUYE, M., DASHNOW, H., RAVEN, L. A., SCHULTZ, M. B., POPE, B. J., TOMITA, T., ZOBEL, J. & HOLT, K. E. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med,* 6**,** 90.

JALALVAND, F., LITTORIN, N., SU, Y. C. & RIESBECK, K. 2014. Impact of immunization with Protein F on pulmonary clearance of nontypeable *Haemophilusinfluenzae*. *Vaccine,* 32**,** 2261-4.

JALALVAND, F., SU, Y. C., MORGELIN, M., BRANT, M., HALLGREN, O., WESTERGREN-THORSSON, G., SINGH, B. & RIESBECK, K. 2013. *Haemophilus influenzae* protein F mediates binding to laminin and human pulmonary epithelial cells. *J Infect Dis,* 207**,** 803-13.

JANSON, H., RUAN, M. & FORSGREN, A. 1993. Limited diversity of the protein D gene (hpd) among encapsulated and nonencapsulated *Haemophilus influenzae* strains. *Infect Immun,* 61**,** 4546-52.

JEWETT, M. W., LAWRENCE, K. A., BESTOR, A., BYRAM, R., GHERARDINI, F. & ROSA, P. A. 2009. GuaA and GuaB are essential for Borrelia burgdorferi survival in the tick-mouse infection cycle. *J Bacteriol,* 191**,** 6231-41.

JIANG, Z., NAGATA, N., MOLINA, E., BAKALETZ, L. O., HAWKINS, H. & PATEL, J. A. 1999. Fimbria-mediated enhanced attachment of nontypeable *Haemophilus influenzae* to respiratory syncytial virus-infected respiratory epithelial cells. *Infect Immun,* 67**,** 187-92.

JOHNSON, S. S., HEBSGAARD, M. B., CHRISTENSEN, T. R., MASTEPANOV, M., NIELSEN, R., MUNCH, K., BRAND, T., GILBERT, M. T., ZUBER, M. T., BUNCE, M., RONN, R., GILICHINSKY, D., FROESE, D. & WILLERSLEV, E. 2007. Ancient bacteria show evidence of DNA repair. *Proc Natl Acad Sci U S A,* 104**,** 14401-5.

JORDAN, I. K., CONLEY, A. B., ANTONOV, I. V., ARTHUR, R. A., COOK, E. D., COOPER, G. P., JONES, B. L., KNIPE, K. M., LEE, K. J., LIU, X., MITCHELL, G. J., PANDE, P. R., PETIT, R. A., QIN, S., RAJAN, V. N., SARDA, S., SEBASTIAN, A., TANG, S., THAPLIYAL, R., VARGHESE, N. J., YE, T., KATZ, L. S., WANG, X., ROWE, L., FRACE, M. & MAYER, L. W. 2011. Genome sequences for five strains of the emerging pathogen *Haemophilus haemolyticus*. *J Bacteriol,* 193**,** 5879-80.

JURCISEK, J. A. & BAKALETZ, L. O. 2007. Biofilms formed by nontypeable *Haemophilus influenzae* in vivo contain both double-stranded DNA and type IV pilin protein. *J Bacteriol,* 189**,** 3868-75.

KAPATAI, G., SHEPPARD, C. L., AL-SHAHIB, A., LITT, D. J., UNDERWOOD, A. P., HARRISON, T. G. & FRY, N. K. 2016. Whole genome sequencing of *Streptococcus pneumoniae*: development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ,* 4**,** e2477.

KARALUS, R. J. & MURPHY, T. F. 1999. Purification and characterization of outer membrane protein P6, a vaccine antigen of non-typeable *Haemophilus influenzae*. *FEMS Immunol Med Microbiol,* 26**,** 159-66.

KASHIWAGI, K., ENDO, H., KOBAYASHI, H., TAKIO, K. & IGARASHI, K. 1995. Spermidine-preferential uptake system in Escherichia coli. ATP hydrolysis by PotA protein and its association with membrane. *J Biol Chem,* 270**,** 25377-82.

KASTENBAUER, S. & PFISTER, H. W. 2003. Pneumococcal meningitis in adults: spectrum of complications and prognostic factors in a series of 87 cases. *Brain,* 126**,**1015-25.

KATZ, A. R., KOMEYA, A. Y., SOGE, O. O., KIAHA, M. I., LEE, M. V., WASSERMAN, G. M., MANINGAS, E. V., WHELEN, A. C., KIRKCALDY, R. D., SHAPIRO, S. J., BOLAN, G. A.& HOLMES, K. K. 2012. *Neisseria gonorrhoeae* with high-level resistance to azithromycin: case report of the first isolate identified in the United States. *Clin Infect Dis,* 54**,** 841-3.

KATZ, K. A. & KLAUSNER, J. D. 2008. Azithromycin resistance in *Treponema pallidum*. *Curr Opin Infect Dis,* 21**,** 83-91.

KAUR, R., CHANG, A., XU, Q., CASEY, J. R. & PICHICHERO, M. E. 2011. Phylogenetic relatedness and diversity of non-typable *Haemophilus influenzae* in the nasopharynx and middle ear fluid of children with acute otitis media. *JMed Microbiol,* 60**,** 1841-8.

KAWAHIRA, M. 1990. [Clinical and experimental studies on vitronectin in bacterial pneumonia]. *Kansenshogaku Zasshi,* 64**,** 741-51.

KENJALE, R., MENG, G., FINK, D. L., JUEHNE, T., OHASHI, T., ERICKSON, H. P., WAKSMAN, G. & ST GEME, J. W., 3RD 2009. Structural determinants of autoproteolysis of the *Haemophilus influenzae* Hap autotransporter. *Infect Immun,* 77**,** 4704-13.

KILIAN, M. 1976a. The haemolytic activity of *Haemophilus* species. *Acta Pathol Microbiol Scand B,* 84B**,** 339-41.

KILIAN, M. 1976b. A taxonomic study of the genus *Haemophilus*, with the proposal ofa new species. *J Gen Microbiol,* 93**,** 9-62.

KIM, H. S., KIM, S. M., LEE, H. J., PARK, S. J. & LEE, K. H. 2009a. Expression of the *cpdA*gene, encoding a 3',5'-cyclic AMP (cAMP) phosphodiesterase, is positively regulated by the cAMP-cAMP receptor protein complex. *J Bacteriol,* 191**,** 922-30.

KIM, M., OH, H. S., PARK, S. C. & CHUN, J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol,* 64**,** 346-51.

KIM, Y., WANG, X., MA, Q., ZHANG, X. S. & WOOD, T. K. 2009b. Toxin-antitoxin systems in *Escherichia coli* influence biofilm formation through YjgK (TabA) and fimbriae. *J Bacteriol,* 191**,** 1258-67.

KIMURA, A. & HANSEN, E. J. 1986. Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. *Infect Immun,* 51**,** 69-79.

KIRKHAM, L. A., WIERTSEMA, S. P., MOWE, E. N., BOWMAN, J. M., RILEY, T. V. & RICHMOND, P. C. 2010. Nasopharyngeal carriage of *Haemophilus haemolyticus* in otitis-prone and healthy children. *J Clin Microbiol,* 48**,** 2557-9.

KLEMM, P., VEJBORG, R. M. & SHERLOCK, O. 2006. Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol,* 296**,**187-95.

KONG, K. & COATES, H. L. 2009. Natural history, definitions, risk factors and burden ofotitis media. *Med J Aust,* 191**,** S39-43.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2005a. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A,* 102**,** 2567-72.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2005b. Towards a genome-based taxonomy for prokaryotes. *J Bacteriol,* 187**,** 6258-64.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol,* 10**,**504-9.

KOZITSKAYA, S., CHO, S. H., DIETRICH, K., MARRE, R., NABER, K. & ZIEBUHR, W. 2004. The bacterial insertion sequence element IS256 occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: association with biofilm formation and resistance to aminoglycosides. *Infect Immun,* 72**,** 1210-5.

KROLL, J. S., HOPKINS, I. & MOXON, E. R. 1988. Capsule loss in *H. influenzae* type boccurs by recombination-mediated disruption of a gene essential for polysaccharide export. *Cell,* 53**,** 347-56.

KROLL, J. S., LANGFORD, P. R. & LOYNDS, B. M. 1991. Copper-zinc superoxide dismutase of *Haemophilus influenzae* and *H. parainfluenzae*. *J Bacteriol,* 173**,** 7449-57.

KUBIET, M. & RAMPHAL, R. 1995. Adhesion of nontypeable *Haemophilus influenzae* from blood and sputum to human tracheobronchial mucins and lactoferrin. *Infect Immun,* 63**,** 899-902.

KUBIET, M., RAMPHAL, R., WEBER, A. & SMITH, A. 2000. Pilus-mediated adherence of Haemophilus influenzae to human respiratory mucins. *Infect Immun,* 68**,** 3362-7.

KURTZ, S., PHILLIPPY, A., DELCHER, A. L., SMOOT, M., SHUMWAY, M., ANTONESCU, C.& SALZBERG, S. L. 2004. Versatile and open software for comparing largegenomes. *Genome Biol,* 5**,** R12.

KUSCHNER, R. A., HEPPNER, D. G., ANDERSEN, S. L., WELLDE, B. T., HALL, T., SCHNEIDER, I., BALLOU, W. R., FOULDS, G., SADOFF, J. C., SCHUSTER, B. & ET AL. 1994. Azithromycin prophylaxis against a chloroquine-resistant strain of *Plasmodium falciparum*. *Lancet,* 343**,** 1396-7.

LAARMANN, S., CUTTER, D., JUEHNE, T., BARENKAMP, S. J. & ST GEME, J. W. 2002. The *Haemophilus influenzae* Hia autotransporter harbours two adhesive pockets that reside in the passenger domain and recognize the same host cell receptor. *Mol Microbiol,* 46**,** 731-43.

LACROSS, N. C., MARRS, C. F. & GILSDORF, J. R. 2013. Population structure in nontypeable *Haemophilus influenzae*. *Infect Genet Evol,* 14**,** 125-36.

LACROSS, N. C., MARRS, C. F., PATEL, M., SANDSTEDT, S. A. & GILSDORF, J. R. 2008. High genetic diversity of nontypeable *Haemophilus influenzae* isolates from twochildren attending a day care center. *J Clin Microbiol,* 46**,** 3817-21.

LADHANI, S., SLACK, M. P., HEATH, P. T., VON GOTTBERG, A., CHANDRA, M., RAMSAY, M. E. & EUROPEAN UNION INVASIVE BACTERIAL INFECTION SURVEILLANCE, P. 2010. Invasive *Haemophilus influenzae* Disease, Europe, 1996-2006. *Emerg Infect Dis,* 16**,** 455-63.

LAHTI, E., LOFDAHL, M., AGREN, J., HANSSON, I. & OLSSON ENGVALL, E. 2016. Confirmation of a Campylobacteriosis Outbreak Associated with Chicken Liver Pate Using PFGE and WGS. *Zoonoses Public Health*.

LANGFORD, P. R., SHEEHAN, B. J., SHAIKH, T. & KROLL, J. S. 2002. Active copper- and zinc-containing superoxide dismutase in the cryptic genospecies of *Haemophilus* causing urogenital and neonatal infections discriminates them from *Haemophilus influenzae* sensu stricto. *J Clin Microbiol,* 40**,** 268-70.

LANGRIDGE, G. C., PHAN, M. D., TURNER, D. J., PERKINS, T. T., PARTS, L., HAASE, J., CHARLES, I., MASKELL, D. J., PETERS, S. E., DOUGAN, G., WAIN, J., PARKHILL, J.& TURNER, A. K. 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res,* 19**,** 2308-16.

LATHAM, R., ZHANG, B. & TRISTRAM, S. 2015. Identifying *Haemophilus haemolyticus* and *Haemophilus influenzae* by SYBR Green real-time PCR. *J Microbiol Methods,* 112**,** 67-9.

LAZAREVSKI, G., KOBREHEL, G., NARANDA, A., BANIC-TOMISIC, Z. & METELKO, B. 1998. Acid catalyzed ring opening reactions of 6-deoxy-9-deoxo-9a-aza-9a-homo-erythromycin A 6,9-cyclic imino ether. *J Antibiot (Tokyo),* 51**,**893-6.

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., WIGGER, C., BEISSBARTH, J., WOLTRING, D., ANDREWS, R., CHATFIELD, M. D., SMITH-VAUGHAN, H. & MORRIS, P. S. 2016. General health, otitis media, nasopharyngeal carriage and middle ear microbiology in Northern Territory Aboriginal children vaccinated during consecutive periods of 10-valent or 13-valent pneumococcal conjugate vaccines. *Int J Pediatr Otorhinolaryngol,* 86**,** 224-32.

LEACH, A. J., WIGGER, C., HARE, K., HAMPTON, V., BEISSBARTH, J., ANDREWS, R., CHATFIELD, M., SMITH-VAUGHAN, H. & MORRIS, P. S. 2015. Reduced middle ear infection with non-typeable *Haemophilus influenzae*, but not *Streptococcus pneumoniae*, after transition to 10-valent pneumococcal non-typeable *H. influenzae* protein D conjugate vaccine. *BMC Pediatr,* 15**,** 162.

LEANORD, A. & WILLIAMS, C. 2002. *Haemophilus influenzae* in acute exacerbations of chronic obstructive pulmonary disease. *Int J Antimicrob Agents,* 19**,** 371-5.

LEE, C. H., LEE, M. C., LIN, H. H., SHU, C. C., WANG, J. Y., LEE, L. N. & CHAO, K. M. 2012. Pulmonary tuberculosis and delay in anti-tuberculous treatment are important risk factors for chronic obstructive pulmonary disease. *PLoS One,* 7**,** e37978.

LEES, J. A., VEHKALA, M., VALIMAKI, N., HARRIS, S. R., CHEWAPREECHA, C., CROUCHER, N. J., MARTTINEN, P., DAVIES, M. R., STEER, A. C., TONG, S. Y., HONKELA, A., PARKHILL, J., BENTLEY, S. D. & CORANDER, J. 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat Commun,* 7**,** 12797.

LINKE, D., RIESS, T., AUTENRIETH, I. B., LUPAS, A. & KEMPF, V. A. 2006. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol,* 14**,** 264-70.

LIU, S., ZHOU, Y., WANG, X., WANG, D., LU, J., ZHENG, J., ZHONG, N. & RAN, P. 2007. Biomass fuels are the probable risk factor for chronic obstructive pulmonary disease in rural South China. *Thorax,* 62**,** 889-97.

LIU, Y. Y., CHEN, C. C. & CHIOU, C. S. 2016. Construction of a Pan-Genome AlleleDatabase of Salmonella enterica *Serovar Enteritidis* for Molecular Subtyping and Disease Cluster Identification. *Front Microbiol,* 7**,** 2010.

LOMOVSKAYA, O., LEWIS, K. & MATIN, A. 1995. EmrR is a negative regulator of the *Escherichia coli* multidrug resistance pump EmrAB. *J Bacteriol,* 177**,**2328-34.

LOVMAR, M., TENSON, T. & EHRENBERG, M. 2004. Kinetics of macrolide action: the josamycin and erythromycin cases. *J Biol Chem,* 279**,** 53506-15.

LUKEHART, S. A., GODORNES, C., MOLINI, B. J., SONNETT, P., HOPKINS, S., MULCAHY, F., ENGELMAN, J., MITCHELL, S. J., ROMPALO, A. M., MARRA, C. M. & KLAUSNER, J. D. 2004. Macrolide resistance in *Treponema pallidum* in the United States and Ireland. *N Engl J Med,* 351**,** 154-8.

LUNA, V. A., COUSIN, S., JR., WHITTINGTON, W. L. & ROBERTS, M. C. 2000. Identification of the conjugative *mef* gene in clinical *Acinetobacter junii* and *Neisseria gonorrhoeae* isolates. *Antimicrob Agents Chemother,* 44**,** 2503-6.

LUNDBACK, B., LINDBERG, A., LINDSTROM, M., RONMARK, E., JONSSON, A. C., JONSSON, E., LARSSON, L. G., ANDERSSON, S., SANDSTROM, T., LARSSON, K. & OBSTRUCTIVE LUNG DISEASE IN NORTHERN SWEDEN, S. 2003. Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med,* 97**,** 115-22.

MAIDEN, M. C. 2006. Multilocus sequence typing of bacteria. *Annu Rev Microbiol,* 60**,**561-88.

MALHOTRA-KUMAR, S., LAMMENS, C., COENEN, S., VAN HERCK, K. & GOOSSENS, H. 2007. Effect of azithromycin and clarithromycin therapy on pharyngeal carriage of macrolide-resistant streptococci in healthy volunteers: a randomised, double-blind, placebo-controlled study. *Lancet,* 369**,** 482-90.

MAMMEN, M. J. & SETHI, S. 2012. Macrolide therapy for the prevention of acute exacerbations in chronic obstructive pulmonary disease. *Pol Arch Med Wewn,* 122**,** 54-9.

MARJANOVIC, N., BOSNAR, M., MICHIELIN, F., WILLE, D. R., ANIC-MILIC, T., CULIC, O., POPOVIC-GRLE, S., BOGDAN, M., PARNHAM, M. J. & ERAKOVIC HABER, V. 2011. Macrolide antibiotics broadly and distinctively inhibit cytokine and chemokine production by COPD sputum cells in vitro. *Pharmacol Res,* 63**,** 389-97.

MARTI, S., PUIG, C., DE LA CAMPA, A. G., TIRADO-VELEZ, J. M., TUBAU, F., DOMENECH, A., CALATAYUD, L., GARCIA-SOMOZA, D., AYATS, J., LINARES, J. & ARDANUY, C. 2016. Identification of *Haemophilus haemolyticus* in clinical samples and characterization of their mechanisms of antimicrobial resistance. *J Antimicrob Chemother,* 71**,**80-4.

MAUGHAN, H. & REDFIELD, R. J. 2009a. Extensive variation in natural competence in *Haemophilus influenzae*. *Evolution,* 63**,** 1852-66.

MAUGHAN, H. & REDFIELD, R. J. 2009b. Tracing the evolution of competence in *Haemophilus influenzae*. *PLoS One,* 4**,** e5854.

MCCREA, K. W., WANG, M. L., XIE, J., SANDSTEDT, S. A., DAVIS, G. S., LEE, J. H., MARRS, C. F. & GILSDORF, J. R. 2010. Prevalence of the sodC gene in nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus* by microarray-based hybridization. *J Clin Microbiol,* 48**,** 714-9.

MCCREA, K. W., XIE, J., LACROSS, N., PATEL, M., MUKUNDAN, D., MURPHY, T. F., MARRS, C. F. & GILSDORF, J. R. 2008. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. *J Clin Microbiol,* 46**,** 406-16.

MCGEHEE, R. R., BARRE, F. F. & FINLAND, M. 1968. Resistance of *Staphylococcus aureus* to lincomycin, clinimycin, and erythromycin. *Antimicrob Agents Chemother (Bethesda),* 8**,** 392-7.

MCNALLY, A., CHENG, L., HARRIS, S. R. & CORANDER, J. 2013. The evolutionary path to extraintestinal pathogenic, drug-resistant *Escherichia coli* is marked by drastic reduction in detectable recombination within the core genome. *Genome Biol Evol,* 5**,** 699-710.

MCVERNON, J., SLACK, M. P. & RAMSAY, M. E. 2006. Changes in the epidemiology of epiglottitis following introduction of *Haemophilus influenzae* type b (Hib) conjugate vaccines in England: a comparison of two data sources. *Epidemiol Infect,* 134**,** 570-2.

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.

MEI, J. M., NOURBAKHSH, F., FORD, C. W. & HOLDEN, D. W. 1997. Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. *Mol Microbiol,* 26**,** 399-407.

MENDELMAN, P. M., CHAFFIN, D. O., MUSSER, J. M., DE GROOT, R., SERFASS, D. A.& SELANDER, R. K. 1987. Genetic and phenotypic diversity among ampicillin-resistant, non-beta-lactamase-producing, nontypeable *Haemophilus influenzae* isolates. *Infect Immun,* 55**,** 2585-9.

MENG, G., SPAHICH, N., KENJALE, R., WAKSMAN, G. & ST GEME, J. W., 3RD 2011. Crystal structure of the *Haemophilus influenzae* Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *EMBO J,* 30**,** 3864-74.

MENG, G., ST GEME, J. W., 3RD & WAKSMAN, G. 2008. Repetitive architecture ofthe *Haemophilus influenzae* Hia trimeric autotransporter. *J Mol Biol,* 384**,**824-36.

MENZEL, M., AKBARSHAHI, H., BJERMER, L. & ULLER, L. 2016. Azithromycin induces anti-viral effects in cultured bronchial epithelial cells from COPD patients. *Sci Rep,* 6**,** 28698.

MHLANGA-MUTANGADURA, T., MORLIN, G., SMITH, A. L., EISENSTARK, A. & GOLOMB, M. 1998. Evolution of the major pilus gene cluster of *Haemophilus influenzae*. *J Bacteriol,* 180**,** 4693-703.

MICHEL, L. V., SNYDER, J., SCHMIDT, R., MILILLO, J., GRIMALDI, K., KALMETA, B., KHAN, M. N., SHARMA, S., WRIGHT, L. K. & PICHICHERO, M. E. 2013. Dual orientation ofthe outer membrane lipoprotein P6 of nontypeable *haemophilus influenzae*. *J Bacteriol,* 195**,** 3252-9.

MILLER, M. A., SCHWARTZ, T., PICKETT, B. E., HE, S., KLEM, E. B., SCHEUERMANN, R. H., PASSAROTTI, M., KAUFMAN, S. & O'LEARY, M. A. 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evol Bioinform Online,* 11**,**43-8.

MOGHADDAM, S. J., OCHOA, C. E., SETHI, S. & DICKEY, B. F. 2011. Nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease and lung cancer. *Int J Chron Obstruct Pulmon Dis,* 6**,** 113-23.

MOLLOY, A., NAIR, S., COOKE, F. J., WAIN, J., FARRINGTON, M., LEHNER, P. J. & TOROK, M. E. 2010. First report of *Salmonella enterica* serotype paratyphi Aazithromycin resistance leading to treatment failure. *J Clin Microbiol,* 48**,**4655-7.

MONASTA, L., RONFANI, L., MARCHETTI, F., MONTICO, M., VECCHI BRUMATTI, L., BAVCAR, A., GRASSO, D., BARBIERO, C. & TAMBURLINI, G. 2012. Burden of disease caused by otitis media: systematic review and global estimates. *PLoS One,* 7**,** e36226.

MONGODIN, E., FINAN, J., CLIMO, M. W., ROSATO, A., GILL, S. & ARCHER, G. L. 2003. Microarray transcription analysis of clinical *Staphylococcus aureus* isolatesresistant to vancomycin. *J Bacteriol,* 185**,** 4638-43.

MOREY, P., VIADAS, C., EUBA, B., HOOD, D. W., BARBERAN, M., GIL, C., GRILLO, M. J., BENGOECHEA, J. A. & GARMENDIA, J. 2013. Relative contributions of lipooligosaccharide inner and outer core modifications to nontypeable *Haemophilus influenzae* pathogenesis. *Infect Immun,* 81**,** 4100-11.

MORLEY, L., MCNALLY, A., PASZKIEWICZ, K., CORANDER, J., MERIC, G., SHEPPARD, S. K., BLOM, J. & MANNING, G. 2015. Gene Loss and Lineage-Specific Restriction-Modification Systems Associated with Niche Differentiation in the *Campylobacter jejuni* Sequence Type 403 Clonal Complex. *Appl Environ Microbiol,* 81**,**3641-7.

MOROZUMI, M., HASEGAWA, K., KOBAYASHI, R., INOUE, N., IWATA, S., KUROKI, H., KAWAMURA, N., NAKAYAMA, E., TAJIMA, T., SHIMIZU, K. & UBUKATA, K. 2005. Emergence of macrolide-resistant *Mycoplasma pneumoniae* with a 23S rRNAgene mutation. *Antimicrob Agents Chemother,* 49**,** 2302-6.

MORTON, D. J., HEMPEL, R. J., WHITBY, P. W., SEALE, T. W. & STULL, T. L. 2012. Aninvasive *Haemophilus haemolyticus* isolate. *J Clin Microbiol,* 50**,** 1502-3.

MUIR, P., LI, S., LOU, S., WANG, D., SPAKOWICZ, D. J., SALICHOS, L., ZHANG, J., WEINSTOCK, G. M., ISAACS, F., ROZOWSKY, J. & GERSTEIN, M. 2016. The real cost ofsequencing: scaling computation to keep pace with data generation. *Genome Biol,* 17**,**53.

MUKUNDAN, D., ECEVIT, Z., PATEL, M., MARRS, C. F. & GILSDORF, J. R. 2007. Pharyngeal colonization dynamics of *Haemophilus influenzae* and *Haemophilushaemolyticus* in healthy adult carriers. *J Clin Microbiol,* 45**,** 3207-17.

MUNSON, R., JR., BRODEUR, B., CHONG, P., GRASS, S., MARTIN, D. & PROULX, C. 1992. Outer membrane proteins P1 and P2 of *Haemophilus influenzae* type b: structure and identification of surface-exposed epitopes. *J Infect Dis,* 165 Suppl 1**,** S86-9.

MUNSON, R. S., JR., GRASS, S. & WEST, R. 1993. Molecular cloning and sequence of the gene for outer membrane protein P5 of *Haemophilus influenzae*. *Infect Immun,* 61**,** 4017-20.

MURAOKA, W. T. & ZHANG, Q. 2011. Phenotypic and genotypic evidence for L-fucose utilization by *Campylobacter jejuni*. *J Bacteriol,* 193**,** 1065-75.

MURPHY, E., HUWYLER, L. & DE FREIRE BASTOS MDO, C. 1985. Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. *EMBO J,* 4**,** 3357-65.

MURPHY, T. F. & BARTOS, L. C. 1988. Purification and analysis with monoclonal antibodies of P2, the major outer membrane protein of nontypable *Haemophilusinfluenzae*. *Infect Immun,* 56**,** 1084-9.

MURPHY, T. F., BRAUER, A. L., SCHIFFMACHER, A. T. & SETHI, S. 2004. Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 170**,** 266-72.

MURPHY, T. F., BRAUER, A. L., SETHI, S., KILIAN, M., CAI, X. & LESSE, A. J. 2007. *Haemophilus haemolyticus*: a human respiratory tract commensal to be distinguished from *Haemophilus influenzae*. *J Infect Dis,* 195**,** 81-9.

MURPHY, T. F. & KIRKHAM, C. 2002. Biofilm formation by nontypeable *Haemophilus influenzae*: strain variability, outer membrane antigen expression and role ofpili. *BMC Microbiol,* 2**,** 7.

MURPHY, T. F., KIRKHAM, C. & LESSE, A. J. 2006. Construction of a mutant and characterization of the role of the vaccine antigen P6 in outer membrane integrity of nontypeable *Haemophilus influenzae. Infect Immun,* 74**,** 5169-76.

MURPHY, T. F., LESSE, A. J., KIRKHAM, C., ZHONG, H., SETHI, S. & MUNSON, R. S., JR. 2011. A clonal group of nontypeable *Haemophilus influenzae* with two IgA proteases is adapted to infection in chronic obstructive pulmonary disease. *PLoS One,* 6**,** e25923.

MURPHY, T. F., SETHI, S., KLINGMAN, K. L., BRUEGGEMANN, A. B. & DOERN, G. V. 1999. Simultaneous respiratory tract colonization by multiple strains of nontypeable *Haemophilus influenzae* in chronic obstructive pulmonary disease: implications for antibiotic therapy. *J Infect Dis,* 180**,** 404-9.

MUSSER, J. M., BARENKAMP, S. J., GRANOFF, D. M. & SELANDER, R. K. 1986a. Genetic relationships of serologically nontypable and serotype b strains of *Haemophilus influenzae*. *Infect Immun,* 52**,** 183-91.

MUSSER, J. M., HEWLETT, E. L., PEPPLER, M. S. & SELANDER, R. K. 1986b. Genetic diversity and relationships in populations of *Bordetella* spp. *J Bacteriol,* 166**,** 230-7.

MUSSER, J. M., KROLL, J. S., GRANOFF, D. M., MOXON, E. R., BRODEUR, B. R., CAMPOS, J., DABERNAT, H., FREDERIKSEN, W., HAMEL, J., HAMMOND, G. & ET AL. 1990. Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. *Rev Infect Dis,* 12**,** 75-111.

NAGY, P. 2013. Kinetics and mechanisms of thiol-disulfide exchange covering direct substitution and thiol oxidation-mediated pathways. *Antioxid Redox Signal,* 18**,** 1623-41.

NAKAMURA, M., ASAKA, T., KIRITA, A., MIYAZAKI, H., SENDA, Y., FUJITA, S., FUKUSHIMA, R., WATANABE, K., KARASAWA, T., KAWAHARA, E., SHIMURA, S. & YAMAGISHI, T. 2006. Occurrence of the Fimbria Gene *hifA* in clinical isolates ofnonencapsulated *Haemophilus influenzae*. *Microbiol Immunol,* 50**,** 327-9.

NEARY, J. M. & MURPHY, T. F. 2006. Antibodies directed at a conserved motif in loop 6 of outer membrane protein P2 of nontypeable *Haemophilus influenzae* recognize multiple strains in immunoassays. *FEMS Immunol Med Microbiol,* 46**,**251-61.

NEARY, J. M., YI, K., KARALUS, R. J. & MURPHY, T. F. 2001. Antibodies to loop 6 of the P2 porin protein of nontypeable *Haemophilus influenzae* are bactericidal against multiple strains. *Infect Immun,* 69**,** 773-8.

NELSON, M. B., APICELLA, M. A., MURPHY, T. F., VANKEULEN, H., SPOTILA, L. D. & REKOSH, D. 1988. Cloning and sequencing of *Haemophilus influenzae* outer membrane protein P6. *Infect Immun,* 56**,** 128-34.

NELSON, M. B., MUNSON, R. S., APICELLA, M. A., SIKKEMA, D. J., MOLLESTON, J. P.& MURPHY, T. F. 1991. Molecular Conservation of the P6 Outer-Membrane Protein among Strains of *Haemophilus-Influenzae* - Analysis of Antigenic Determinants, Gene-Sequences, and Restriction-Fragment-Length-Polymorphisms. *Infectionand Immunity,* 59**,** 2658-2663.

NHS. 2012. *COPD Commissioning Toolkit* [Online]. http://www.dh.gov.uk/publications-electronic formatt only.: Department of Health. Available: https://www.gov.uk/government/uploads/system/uploads/attachment data/file/212876/chronic-obstructive-pulmonary-disease-COPD-commissioning-toolkit.pdf [Accessed November 14th 2014].

NICE. 2010. *Chronic obstructive pulmonary disease in over 16s: diagnosis andmanagement* [Online]. NICE Available: https://www.nice.org.uk/guidance/CG101/chapter/1-Guidance - management-of-exacerbations-of-copd [Accessed 20th January 2017 2017].

NORSKOV-LAURITSEN, N. 2009. Detection of cryptic genospecies misidentified as *Haemophilus influenzae* in routine clinical samples by assessment of markergenes *fucK*, *hap*, and *sodC*. *J Clin Microbiol,* 47**,** 2590-2.

NORSKOV-LAURITSEN, N., OVERBALLE, M. D. & KILIAN, M. 2009. Delineation of the species *Haemophilus influenzae* by phenotype, multilocus sequence phylogeny, and detection of marker genes. *J Bacteriol,* 191**,** 822-31.

NOVOTNY, L. A., ADAMS, L. D., KANG, D. R., WIET, G. J., CAI, X., SETHI, S., MURPHY, T. F.& BAKALETZ, L. O. 2009. Epitope mapping immunodominant regions of the PilA protein of nontypeable *Haemophilus influenzae* (NTHI) to facilitate the design of two novel chimeric vaccine candidates. *Vaccine,* 28**,** 279-89.

NOVOTNY, L. A. & BAKALETZ, L. O. 2003. The fourth surface-exposed region of the outer membrane protein P5-homologous adhesin of nontypable Haemophilus influenzae is an immunodominant but nonprotective decoying epitope. *J Immunol,* 171**,** 1978-83.

NOVOTNY, L. A. & BAKALETZ, L. O. 2016. Intercellular adhesion molecule 1 serves as a primary cognate receptor for the Type IV pilus of nontypeable *Haemophilus influenzae*. *Cell Microbiol*.

NOVOTNY, L. A., CLEMENTS, J. D. & BAKALETZ, L. O. 2011. Transcutaneous immunization as preventative and therapeutic regimens to protect against experimental otitismedia due to nontypeable *Haemophilus influenzae*. *Mucosal Immunol,* 4**,** 456-67.

OROZCO-LEVI, M., GARCIA-AYMERICH, J., VILLAR, J., RAMIREZ-SARMIENTO, A., ANTO, J. M. & GEA, J. 2006. Wood smoke exposure and risk of chronic obstructive pulmonary disease. *Eur Respir J,* 27**,** 542-6.

OSTBERG, K. L., RUSSELL, M. W. & MURPHY, T. F. 2009. Mucosal immunization of micewith recombinant OMP P2 induces antibodies that bind to surface epitopes of multiple strains of nontypeable *Haemophilus influenzae*. *Mucosal Immunol,* 2**,** 63-73.

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T., FOOKES, M., FALUSH, D., KEANE, J. A. & PARKHILL, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* 31**,** 3691-3.

PANKUCH, G. A., JUENEMAN, S. A., DAVIES, T. A., JACOBS, M. R. & APPELBAUM, P. C.1998. In vitro selection of resistance to four beta-lactams and azithromycin in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother,* 42**,** 2914-8.

PAPI, A., BELLETTATO, C. M., BRACCIONI, F., ROMAGNOLI, M., CASOLARI, P., CARAMORI, G., FABBRI, L. M. & JOHNSTON, S. L. 2006. Infections and airway inflammation in chronic obstructive pulmonary disease severe exacerbations. *Am J Respir Crit Care Med,* 173**,** 1114-21.

PEREZ-PADILLA, R., REGALADO, J., VEDAL, S., PARE, P., CHAPELA, R., SANSORES, R.& SELMAN, M. 1996. Exposure to biomass smoke and chronic airway disease in Mexican women. A case-control study. *Am J Respir Crit Care Med,* 154**,**701-6.

PERIC, M., BOZDOGAN, B., JACOBS, M. R. & APPELBAUM, P. C. 2003. Effects of an efflux mechanism and ribosomal mutations on macrolide susceptibility of *Haemophilus influenzae* clinical isolates. *Antimicrob Agents Chemother,* 47**,** 1017-22.

PETROPOULOS, A. D., KOUVELA, E. C., STAROSTA, A. L., WILSON, D. N., DINOS, G. P.& KALPAXIS, D. L. 2009. Time-resolved binding of azithromycin to *Escherichia coli* ribosomes. *J Mol Biol,* 385**,** 1179-92.

PETTIGREW, M. M., TSUJI, B. T., GENT, J. F., KONG, Y., HOLDEN, P. N., SETHI, S. & MURPHY, T. F. 2016. Effect of Fluoroquinolones and Macrolides on Eradication and Resistance of *Haemophilus influenzae* in Chronic Obstructive Pulmonary Disease. *Antimicrob Agents Chemother,* 60**,** 4151-8.

PHAFF, S. J., TIDDENS, H. A., VERBRUGH, H. A. & OTT, A. 2006. Macrolide resistance of *Staphylococcus aureus* and *Haemophilus* species associated with long-term azithromycin use in cystic fibrosis. *J Antimicrob Chemother,* 57**,**741-6.

PHE. 2014. *Health Protection Report* [Online]. Available: http://webarchive.nationalarchives.gov.uk/20140722222558/http://www.hpa.org.uk/hpr/archives/2014/hpr0814.pdf [Accessed 88].

PHE 2015. Identification of *Haemophilus* species and the HACEK Group of Organisms. . *UK Standards for Microbiology Investigations. .*

PICHICHERO, M. E., KAUR, R., CASEY, J. R., SABIROV, A., KHAN, M. N. & ALMUDEVAR, A. 2010. Antibody response to *Haemophilus influenzae* outer membrane protein D, P6, and OMP26 after nasopharyngeal colonization and acute otitis media in children. *Vaccine,* 28**,** 7184-92.

PICHICHERO, M. E., LOEB, M., ANDERSON & SMITH, D. H. 1982. Do pili play a role in pathogenicity of *Haemophilus influenzae* type B? *Lancet,* 2**,** 960-2.

PICKERING, J., BINKS, M. J., BEISSBARTH, J., HARE, K. M., KIRKHAM, L. A. & SMITH-VAUGHAN, H. 2014a. A PCR-high-resolution melt assay for rapid differentiation of nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. *J Clin Microbiol,* 52**,** 663-7.

PICKERING, J., SMITH-VAUGHAN, H., BEISSBARTH, J., BOWMAN, J. M., WIERTSEMA, S., RILEY, T. V., LEACH, A. J., RICHMOND, P., LEHMANN, D. & KIRKHAM, L. A. 2014b. Diversity of Nontypeable *Haemophilus influenzae* Strains Colonizing Australian Aboriginal and Non-Aboriginal Children. *J Clin Microbiol,* 52**,** 1352-7.

PICKERING, J. L., PROSSER, A., CORSCADDEN, K. J., DE GIER, C., RICHMOND, P. C., ZHANG, G., THORNTON, R. B. & KIRKHAM, L. A. 2016. *Haemophilus haemolyticus* Interaction with Host Cells Is Different to Nontypeable *Haemophilus influenzae*and Prevents NTHi Association with Epithelial Cells. *Front Cell Infect Microbiol,* 6**,**50.

PITTMAN, M. 1931. Variation and Type Specificity in the Bacterial Species *Haemophilus Influenzae*. *The Journal of Experimental Medicine,* 4**,** 471-492.

PITTMAN, M. 1953. A classification of the hemolytic bacteria of the genus *Haemophilus: Haemophilus haemolyticus* Bergey et al. and *Haemophilus parahaemolyticus* nov spec. *J Bacteriol,* 65**,** 750-1.

POLISSI, A. & GEORGOPOULOS, C. 1996. Mutational analysis and properties of the *msbA* gene of *Escherichia coli*, coding for an essential ABC family transporter. *Mol Microbiol,* 20**,** 1221-33.

POMARES, X., MONTON, C., ESPASA, M., CASABON, J., MONSO, E. & GALLEGO, M. 2011. Long-term azithromycin therapy in patients with severe COPD and repeated exacerbations. *Int J Chron Obstruct Pulmon Dis,* 6**,** 449-56.

POST, J. C. 2001. Direct evidence of bacterial biofilms in otitis media. *Laryngoscope,* 111**,** 2083-94.

POULSEN, K., REINHOLDT, J. & KILIAN, M. 1992. A comparative genetic study of serologically distinct *Haemophilus influenzae* type 1 immunoglobulin A1 proteases. *J Bacteriol,* 174**,** 2913-21.

POWER, P. M., BENTLEY, S. D., PARKHILL, J., MOXON, E. R. & HOOD, D. W. 2012. Investigations into genome diversity of *Haemophilus influenzae* using whole genome sequencing of clinical isolates and laboratory transformants. *BMC Microbiol,* 12**,** 273.

PRESTON, A., MANDRELL, R. E., GIBSON, B. W. & APICELLA, M. A. 1996. The lipooligosaccharides of pathogenic gram-negative bacteria. *Crit Rev Microbiol,* 22**,** 139-80.

PRICE, E. P., SAROVICH, D. S., NOSWORTHY, E., BEISSBARTH, J., MARSH, R. L., PICKERING, J., KIRKHAM, L. A., KEIL, A. D., CHANG, A. B. & SMITH-VAUGHAN, H. C. 2015. *Haemophilus influenzae*: using comparative genomics to accurately identify a highly recombinogenic human pathogen. *BMC Genomics,* 16**,** 641.

PUIG, C., DOMENECH, A., GARMENDIA, J., LANGEREIS, J. D., MAYER, P., CALATAYUD, L., LINARES, J., ARDANUY, C. & MARTI, S. 2014. Increased biofilm formation by nontypeable *Haemophilus influenzae* isolates from patients with invasive disease or otitis media versus strains recovered from cases of respiratory infections. *Appl Environ Microbiol,* 80**,** 7088-95.

RADIN, J. N., GRASS, S. A., MENG, G., COTTER, S. E., WAKSMAN, G. & ST GEME, J. W., 3RD 2009. Structural basis for the differential binding affinities of the HsfBD1 and

HsfBD2 domains in the *Haemophilus influenzae* Hsf adhesin. *J Bacteriol,* 191, 5068-75.

RAHMAN, M. M., GU, X. X., TSAI, C. M., KOLLI, V. S. & CARLSON, R. W. 1999. The structural heterogeneity of the lipooligosaccharide (LOS) expressed by pathogenic non-typeable *Haemophilus influenzae* strain NTHi 9274. *Glycobiology,* 9,1371-80.

RAMBAUT, A. FigTree.

RAMIREZ-VENEGAS, A., SANSORES, R. H., PEREZ-PADILLA, R., REGALADO, J., VELAZQUEZ, A., SANCHEZ, C. & MAYAR, M. E. 2006. Survival of patients with chronic obstructive pulmonary disease due to biomass smoke and tobacco. *Am J Respir Crit Care Med,* 173, 393-7.

RAVIN, A. W. & IYER, V. N. 1961. The genetic relationship and phenotypic expression of mutations endowing Pneumococcus with resistance to erythromycin. *J Gen Microbiol,* 26, 277-301.

RAVIN, A. W., ROTHEIM, M. B. & COULTER, D. M. 1969. Genetic and biochemical studies of suppression of ribosomal resistance to streptomycin and erythromycin in Pneumococcus. *Genetics,* 61, 23-40.

REDDINGTON, K., SCHWENK, S., TUITE, N., PLATT, G., DAVAR, D., COUGHLAN, H., PERSONNE, Y., GANT, V., ENNE, V. I., ZUMLA, A. & BARRY, T. 2015. Comparison of Established Diagnostic Methodologies and a Novel Bacterial *smpB* Real-Time PCR Assay for Specific Detection of *Haemophilus influenzae* Isolates Associated with Respiratory Tract Infections. *J Clin Microbiol,* 53, 2854-60.

REDDY, M. S., BERNSTEIN, J. M., MURPHY, T. F. & FADEN, H. S. 1996. Binding between outer membrane proteins of nontypeable *Haemophilus influenzae* and human nasopharyngeal mucin. *Infect Immun,* 64, 1477-9.

REGELINK, A. G., DAHAN, D., MOLLER, L. V., COULTON, J. W., EIJK, P., VAN ULSEN, P., DANKERT, J. & VAN ALPHEN, L. 1999. Variation in the composition and pore function of major outer membrane pore protein P2 of *Haemophilusinfluenzae* from cystic fibrosis patients. *Antimicrob Agents Chemother,* 43, 226-32.

REMPE, K. A., PORSCH, E. A., WILSON, J. M. & ST GEME, J. W., 3RD 2016. The HMW1 and HMW2 adhesins enhance the ability of nontypeable *Haemophilus influenzae* to colonize the upper respiratory tract of rhesus macaques. *Infect Immun*.

RETSEMA, J., GIRARD, A., SCHELKLY, W., MANOUSOS, M., ANDERSON, M., BRIGHT, G., BOROVOY, R., BRENNAN, L. & MASON, R. 1987. Spectrum and mode of action of azithromycin (CP-62,993), a new 15-membered-ring macrolide with improved potency against gram-negative organisms. *Antimicrob Agents Chemother,* 31, 1939-47.

REVEZ, J., LLARENA, A. K., SCHOTT, T., KUUSI, M., HAKKINEN, M., KIVISTO, R., HANNINEN, M. L. & ROSSI, M. 2014a. Genome analysis of *Campylobacter jejuni* strainsisolated from a waterborne outbreak. *BMC Genomics,* 15, 768.

REVEZ, J., ZHANG, J., SCHOTT, T., KIVISTO, R., ROSSI, M. & HANNINEN, M. L. 2014b. Genomic variation between *Campylobacter jejuni* isolates associated with milk-borne-disease outbreaks. *J Clin Microbiol,* 52, 2782-6.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16, 276-7.

RICHTER, M. & ROSSELLO-MORA, R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A,* 106, 19126-31.

RIDDERBERG, W., FENGER, M. G. & NORSKOV-LAURITSEN, N. 2010. *Haemophilusinfluenzae* may be untypable by the multilocus sequence typing scheme due to a complete deletion of the fucose operon. *J Med Microbiol,* 59, 740-2.

ROBERTS, M. C., SOGE, O. O. & NO, D. B. 2011. Characterization of macrolide resistance genes in *Haemophilus influenzae* isolated from children with cystic fibrosis. *J Antimicrob Chemother,* 66, 100-4.

RODRIGUEZ, C. A., AVADHANULA, V., BUSCHER, A., SMITH, A. L., ST GEME, J. W., 3RD & ADDERSON, E. E. 2003. Prevalence and distribution of adhesins in invasive non-type b encapsulated *Haemophilus influenzae*. *Infect Immun,* 71**,** 1635-42.

ROIER, S., BLUME, T., KLUG, L., WAGNER, G. E., ELHENAWY, W., ZANGGER, K., PRASSL, R., REIDL, J., DAUM, G., FELDMAN, M. F. & SCHILD, S. 2014. A basis for vaccine development: Comparative characterization of *Haemophilus influenzae* outer membrane vesicles. *Int J Med Microbiol*.

ROIER, S., BLUME, T., KLUG, L., WAGNER, G. E., ELHENAWY, W., ZANGGER, K., PRASSL, R., REIDL, J., DAUM, G., FELDMAN, M. F. & SCHILD, S. 2015. A basis for vaccine development: Comparative characterization of *Haemophilus influenzae* outer membrane vesicles. *Int J Med Microbiol,* 305**,** 298-309.

ROIER, S., LEITNER, D. R., IWASHKIW, J., SCHILD-PRUFERT, K., FELDMAN, M. F., KROHNE, G., REIDL, J. & SCHILD, S. 2012. Intranasal immunization with nontypeable *Haemophilus influenzae* outer membrane vesicles induces cross-protective immunity in mice. *PLoS One,* 7**,** e42664.

RONANDER, E., BRANT, M., ERIKSSON, E., MORGELIN, M., HALLGREN, O., WESTERGREN-THORSSON, G., FORSGREN, A. & RIESBECK, K. 2009. Nontypeable *Haemophilus influenzae* adhesin protein E: characterization and biological activity. *J Infect Dis,* 199**,** 522-31.

RONANDER, E., BRANT, M., JANSON, H., SHELDON, J., FORSGREN, A. & RIESBECK, K. 2008. Identification of a novel *Haemophilus influenzae* protein important for adhesion to epithelial cells. *Microbes Infect,* 10**,** 87-96.

ROSE, M. C. & VOYNOW, J. A. 2006. Respiratory tract mucin genes and mucinglycoproteins in health and disease. *Physiol Rev,* 86**,** 245-78.

ROULI, L., MERHEJ, V., FOURNIER, P. E. & RAOULT, D. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect,* 7**,**72-85.

RYCHLIK, I. & CERNA, J. 1968. Cross resistance of *Escherichia coli* B. ribosomes toinhibition of the puromycin reaction by erythromycin, spiramycin and chloramphenicol. *Hoppe Seylers Z Physiol Chem,* 349**,** 958-9.

SABIROV, A., CASEY, J. R., MURPHY, T. F. & PICHICHERO, M. E. 2009. Breast-feeding is associated with a reduced frequency of acute otitis media and high serum antibody levels against NTHi and outer membrane protein vaccine antigen candidate P6. *Pediatr Res,* 66**,** 565-70.

SABIROV, A., KODAMA, S., SABIROVA, N., MOGI, G. & SUZUKI, M. 2004. Intranasal immunization with outer membrane protein P6 and cholera toxin induces specific sinus mucosal immunity and enhances sinus clearance of nontypeable *Haemophilus influenzae*. *Vaccine,* 22**,** 3112-21.

SALCEDO, S. P. & CID, V. J. 2011. Nontypable *Haemophilus influenzae*: an intracellular phase within epithelial cells might contribute to persistence. *Microbiology,* 157**,**1-2.

SAMSON, C., TAMALET, A., THIEN, H. V., TAYTARD, J., PERISSON, C., NATHAN, N., CLEMENT, A., BOELLE, P. Y. & CORVOL, H. 2016. Long-term effects of azithromycin in patients with cystic fibrosis. *Respir Med,* 117**,** 1-6.

SANCHEZ, L., PAN, W., VINAS, M. & NIKAIDO, H. 1997. The *acrAB* homolog of *Haemophilus influenzae* codes for a functional multidrug efflux pump. *J Bacteriol,* 179**,**6855-7.

SANDSTEDT, S. A., ZHANG, L., PATEL, M., MCCREA, K. W., QIN, Z., MARRS, C. F. & GILSDORF, J. R. 2008. Comparison of laboratory-based and phylogenetic methods to distinguish between *Haemophilus influenzae* and *H. haemolyticus*. *JMicrobiol Methods,* 75**,** 369-71.

SATOLA, S. W., NAPIER, B. & FARLEY, M. M. 2008. Association of IS1016 with the *hia* adhesin gene and biotypes V and I in invasive nontypeable *Haemophilusinfluenzae*. *Infect Immun,* 76**,** 5221-7.

SCHROEDER, M. R. & STEPHENS, D. S. 2016. Macrolide Resistance in *Streptococcus pneumoniae. Front Cell Infect Microbiol,* 6**,** 98.

SEEMAN, T. 2014. *Wombac* [Online]. Available: https://github.com/tseemann/wombac [Accessed].

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SEGAL, L. N., CLEMENTE, J. C., WU, B. G., WIKOFF, W. R., GAO, Z., LI, Y., KO, J. P., ROM, W. N., BLASER, M. J. & WEIDEN, M. D. 2016. Randomised, double-blind, placebo-controlled trial with azithromycin selects for anti-inflammatory microbial metabolites in the emphysematous lung. *Thorax*.

SEGATA, N., WALDRON, L., BALLARINI, A., NARASIMHAN, V., JOUSSON, O. & HUTTENHOWER, C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods,* 9**,** 811-4.

SERAFINI, F., BOTTACINI, F., VIAPPIANI, A., BARUFFINI, E., TURRONI, F., FORONI, E., LODI, T., VAN SINDEREN, D. & VENTURA, M. 2011. Insights into physiological and genetic mupirocin susceptibility in bifidobacteria. *Appl Environ Microbiol,* 77**,**3141-6.

SERISIER, D. J. 2013. Risks of population antimicrobial resistance associated withchronic macrolide use for inflammatory airway diseases. *Lancet Respir Med,* 1**,**262-74.

SERISIER, D. J., MARTIN, M. L., MCGUCKIN, M. A., LOURIE, R., CHEN, A. C., BRAIN, B., BIGA, S., SCHLEBUSCH, S., DASH, P. & BOWLER, S. D. 2013. Effect of long-term, low-dose erythromycin on pulmonary exacerbations among patients with non-cystic fibrosis bronchiectasis: the BLESS randomized controlled trial. *JAMA,* 309**,** 1260-7.

SETHI, S., MALONEY, J., GROVE, L., WRONA, C. & BERENSON, C. S. 2006. Airway inflammation and bronchial bacterial colonization in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 173**,** 991-8.

SEYAMA, S., WAJIMA, T., NAKAMINAMI, H. & NOGUCHI, N. 2016. Clarithromycin Resistance Mechanisms of Epidemic beta-Lactamase-Nonproducing Ampicillin-Resistant *Haemophilus influenzae* Strains in Japan. *Antimicrob Agents Chemother,* 60**,** 3207-10.

SHAHINI SHAMS ABADI, M., SIADAT, S. D., VAZIRI, F., DAVARI, M., FATEH, A., POURAZAR, S., ABDOLRAHIMI, F. & GHAZANFARI, M. 2016. Distribution and Diversity of hmw1A Among Invasive Nontypeable *Haemophilus influenzae* Isolates in Iran. *Avicenna J Med Biotechnol,* 8**,** 99-102.

SHAW, J. H. & CLEWELL, D. B. 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in *Streptococcus faecalis*. *J Bacteriol,* 164**,** 782-96.

SHEN, K., ANTALIS, P., GLADITZ, J., SAYEED, S., AHMED, A., YU, S., HAYES, J., JOHNSON, S., DICE, B., DOPICO, R., KEEFE, R., JANTO, B., CHONG, W., GOODWIN, J., WADOWSKY, R. M., ERDOS, G., POST, J. C., EHRLICH, G. D. & HU, F. Z. 2005. Identification, distribution, and expression of novel genes in 10 clinical isolates of nontypeable *Haemophilus influenzae*. *Infect Immun,* 73**,** 3479-91.

SHINKAI, M., FOSTER, G. H. & RUBIN, B. K. 2006. Macrolide antibiotics modulate ERK phosphorylation and IL-8 and GM-CSF production by human bronchial epithelial cells. *Am J Physiol Lung Cell Mol Physiol,* 290**,** L75-85.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SODING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiplesequence alignments using Clustal Omega. *Mol Syst Biol,* 7**,** 539.

SIKKEMA, D. J. & MURPHY, T. F. 1992. Molecular analysis of the P2 porin proteinof nontypeable *Haemophilus influenzae*. *Infect Immun,* 60**,** 5204-11.

SIKKEMA, D. J., NELSON, M. B., APICELLA, M. A. & MURPHY, T. F. 1992. Outer membrane protein P6 of *Haemophilus influenzae* binds to its own gene. *Mol Microbiol,* 6**,** 547-54.

SINGH, B., AL-JUBAIR, T., MORGELIN, M., THUNNISSEN, M. M. & RIESBECK, K. 2013. The unique structure of *Haemophilus influenzae* protein E reveals multiple binding sites for host factors. *Infect Immun,* 81**,** 801-14.

SINGH, B., BRANT, M., KILIAN, M., HALLSTROM, B. & RIESBECK, K. 2010. Protein E of *Haemophilus influenzae* is a ubiquitous highly conserved adhesin. *J Infect Dis,* 201**,** 414-9.

SINGH, B., JALALVAND, F., MORGELIN, M., ZIPFEL, P., BLOM, A. M. & RIESBECK, K. 2011. *Haemophilus influenzae* protein E recognizes the C-terminal domain ofvitronectin and modulates the membrane attack complex. *Mol Microbiol,* 81**,** 80-98.

SINGH, B., JUBAIR, T. A., MORGELIN, M., SUNDIN, A., LINSE, S., NILSSON, U. J. & RIESBECK, K. 2015. *Haemophilus influenzae* surface fibril (Hsf) is a unique twisted hairpin-like trimeric autotransporter. *Int J Med Microbiol,* 305**,** 27-37.

SINGH, B., SU, Y. C., AL-JUBAIR, T., MUKHERJEE, O., HALLSTROM, T., MORGELIN, M., BLOM, A. M. & RIESBECK, K. 2014. A fine-tuned interaction between trimeric autotransporter *Haemophilus* surface fibrils and vitronectin leads to serum resistance and adherence to respiratory epithelial cells. *Infect Immun,* 82**,**2378-89.

SINGH, N. K., KUNDE, D. & TRISTRAM, S. 2016a. Inability of *H. haemolyticus* toinvade respiratory epithelial cells in vitro. *J Med Microbiol*.

SINGH, N. K., KUNDE, D. A. & TRISTRAM, S. G. 2016b. Effect of epithelial cell type on in vitro invasion of non-typeable *Haemophilus influenzae*. *J Microbiol Methods,* 129**,** 66-9.

SKAARE, D., ALLUM, A. G., ANTHONISEN, I. L., JENKINS, A., LIA, A., STRAND, L., TVETEN, Y.& KRISTIANSEN, B. E. 2010. Mutant *ftsI* genes in the emergence of penicillin-binding protein-mediated beta-lactam resistance in *Haemophilus influenzae* in Norway. *Clin Microbiol Infect,* 16**,** 1117-24.

SLATER, G. S. & BIRNEY, E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics,* 6**,** 31.

SMITH-VAUGHAN, H. C., CHANG, A. B., SAROVICH, D. S., MARSH, R. L., GRIMWOOD, K., LEACH, A. J., MORRIS, P. S. & PRICE, E. P. 2014. Absence of an important vaccine and diagnostic target in carriage- and disease-related nontypeable *Haemophilus influenzae*. *Clin Vaccine Immunol,* 21**,** 250-2.

SOLUTIONS, H. Q. 2014. *COPD Education - Management of Exacerbations* [Online]. Southampton University Hospitals Trust Available: http://www.copdeducation.org.uk/-Category-668/COPD-Exacerbations/Treatment [Accessed 6th September 2014].

SPAHICH, N. A., HOOD, D. W., MOXON, E. R. & ST GEME, J. W., 3RD 2012. Inactivation of *Haemophilus influenzae* lipopolysaccharide biosynthesis genes interferes with outer membrane localization of the hap autotransporter. *J Bacteriol,* 194**,**1815-22.

SPAHICH, N. A., KENJALE, R., MCCANN, J., MENG, G., OHASHI, T., ERICKSON, H. P. & ST GEME, J. W., 3RD 2014. Structural determinants of the interaction between the *Haemophilus influenzae* Hap autotransporter and fibronectin. *Microbiology,* 160**,** 1182-90.

ST GEME, J. W., 3RD & CUTTER, D. 2000. The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated. *J Bacteriol,* 182**,** 6005-13.

ST GEME, J. W., 3RD, CUTTER, D. & BARENKAMP, S. J. 1996. Characterization of the genetic locus encoding *Haemophilus influenzae* type b surface fibrils. *J Bacteriol,* 178**,** 6281-7.

ST GEME, J. W., 3RD, DE LA MORENA, M. L. & FALKOW, S. 1994. A *Haemophilusinfluenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. *Mol Microbiol,* 14**,** 217-33.

ST GEME, J. W., 3RD, FALKOW, S. & BARENKAMP, S. J. 1993. High-molecular-weight proteins of nontypable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci U S A,* 90**,** 2875-9.

ST GEME, J. W., 3RD, KUMAR, V. V., CUTTER, D. & BARENKAMP, S. J. 1998. Prevalence and distribution of the *hmw* and *hia* genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*. *Infect Immun,* 66**,** 364-8.

STAHL, M., FRIIS, L. M., NOTHAFT, H., LIU, X., LI, J., SZYMANSKI, C. M. & STINTZI, A. 2011. L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc Natl Acad Sci U S A,* 108**,** 7194-9.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

STAMATAKIS, A., ABERER, A. J., GOLL, C., SMITH, S. A., BERGER, S. A. & IZQUIERDO-CARRASCO, F. 2012. RAxML-Light: a tool for computing terabyte phylogenies. *Bioinformatics,* 28**,** 2064-6.

STOKHOLM, J., CHAWES, B. L., VISSING, N. H., BJARNADOTTIR, E., PEDERSEN, T. M., VINDING, R. K., SCHOOS, A. M., WOLSK, H. M., THORSTEINSDOTTIR, S., HALLAS, H. W., ARIANTO, L., SCHJORRING, S., KROGFELT, K. A., FISCHER, T. K., PIPPER, C. B., BONNELYKKE, K. & BISGAARD, H. 2016. Azithromycin for episodes with asthma-like symptoms in young children aged 1-3 years: a randomised, double-blind, placebo-controlled trial. *Lancet Respir Med,* 4**,** 19-26.

SU, Y. C., JALALVAND, F., MORGELIN, M., BLOM, A. M., SINGH, B. & RIESBECK, K. 2013. *Haemophilus influenzae* acquires vitronectin via the ubiquitous Protein F tosubvert host innate immunity. *Mol Microbiol,* 87**,** 1245-66.

SU, Y. C., MUKHERJEE, O., SINGH, B., HALLGREN, O., WESTERGREN-THORSSON, G., HOOD, D. & RIESBECK, K. 2016. *Haemophilus influenzae* P4 Interacts With Extracellular Matrix Proteins Promoting Adhesion and Serum Resistance. *J Infect Dis,* 213**,**314-23.

SURANA, N. K., CUTTER, D., BARENKAMP, S. J. & ST GEME, J. W., 3RD 2004. The *Haemophilus influenzae* Hia autotransporter contains an unusually short trimeric translocator domain. *J Biol Chem,* 279**,** 14679-85.

SUZUKI, T., YANAI, M., YAMAYA, M., SATOH-NAKAGAWA, T., SEKIZAWA, K., ISHIDA, S.& SASAKI, H. 2001. Erythromycin and common cold in COPD. *Chest,* 120**,** 730-3.

TAIT-KAMRADT, A., CLANCY, J., CRONAN, M., DIB-HAJJ, F., WONDRACK, L., YUAN, W.& SUTCLIFFE, J. 1997. *mefE* is necessary for the erythromycin-resistant Mphenotype in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother,* 41**,** 2251-5.

TAIT-KAMRADT, A., DAVIES, T., APPELBAUM, P. C., DEPARDIEU, F., COURVALIN, P., PETITPAS, J., WONDRACK, L., WALKER, A., JACOBS, M. R. & SUTCLIFFE, J. 2000a. Two new mechanisms of macrolide resistance in clinical strains of *Streptococcus pneumoniae* from Eastern Europe and North America. *Antimicrob Agents Chemother,* 44**,** 3395-401.

TAIT-KAMRADT, A., DAVIES, T., CRONAN, M., JACOBS, M. R., APPELBAUM, P. C.& SUTCLIFFE, J. 2000b. Mutations in 23S rRNA and ribosomal protein L4 account for resistance in pneumococcal strains selected in vitro by macrolide passage. *Antimicrob Agents Chemother,* 44**,** 2118-25.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol,* 30**,**2725-9.

TAUBMAN, S. B., JONES, N. R., YOUNG, F. E. & CORCORAN, J. W. 1966. Sensitivity and resistance to erythromycin in *Bacillus subtilis* 168: the ribosomal binding of erythromycin and chloramphenicol. *Biochim Biophys Acta,* 123**,**438-40.

TAUBMAN, S. B., YOUNG, F. E. & CORCORAN, J. W. 1963. Antibiotic Glycosides. Iv. Studies on the Mechanism of Erythromycin Resistance in Bacillus Subtilis. *Proc Natl Acad Sci U S A,* 50**,** 955-62.

TAYLOR, P. S., FAETH, I., MARKS, M. K., DEL MAR, C. B., SKULL, S. A., PEZZULLO, M. L., HAVYATT, S. M. & COATES, H. L. 2009. Cost of treating otitis media in Australia. *Expert Rev Pharmacoecon Outcomes Res,* 9**,** 133-41.

TCHOUPA, A. K., LICHTENEGGER, S., REIDL, J. & HAUCK, C. R. 2015. Outer membrane protein P1 is the CEACAM-binding adhesin of *Haemophilus influenzae*. *Mol Microbiol,* 98**,** 440-55.

TEAM, R. 2015. *RStudio: Integrated development for R* [Online]. Available: http://www.rstudio.com [Accessed 29th December 2016].

TENENBAUM, T., FRANZ, A., NEUHAUSEN, N., WILLEMS, R., BRADE, J., SCHWEITZER-KRANTZ, S., ADAMS, O., SCHROTEN, H. & HENRICH, B. 2012. Clinical characteristics of children with lower respiratory tract infections are dependent on the carriage of specific pathogens in the nasopharynx. *Eur J Clin Microbiol Infect Dis,* 31**,** 3173-82.

TESTING., T. E. C. O. A. S. Version 6.0, 2016. Breakpoint tables for interpretation of MICs and zone diameters.

THEODORE, M. J., ANDERSON, R. D., WANG, X., KATZ, L. S., VUONG, J. T., BELL, M. E., JUNI, B. A., LOWTHER, S. A., LYNFIELD, R., MACNEIL, J. R. & MAYER, L. W. 2012. Evaluation of New Biomarker Genes for Differentiating *Haemophilusinfluenzae* from *Haemophilus haemolyticus*. *Journal of Clinical Microbiology,* 50**,**1422-1424.

THUN, M., PETO, R., BOREHAM, J. & LOPEZ, A. D. 2012. Stages of the cigarette epidemic on entering its second century. *Tob Control,* 21**,** 96-101.

TREANGEN, T. J., ONDOV, B. D., KOREN, S. & PHILLIPPY, A. M. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands ofintraspecific microbial genomes. *Genome Biol,* 15**,** 524.

TROELSTRA, A., VOGEL, L., VAN ALPHEN, L., EIJK, P., JANSEN, H. & DANKERT, J. 1994. Opsonic antibodies to outer membrane protein P2 of nonencapsulated *Haemophilus influenzae* are strain specific. *Infect Immun,* 62**,** 779-84.

TU, D., BLAHA, G., MOORE, P. B. & STEITZ, T. A. 2005. Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell,* 121**,** 257-70.

VALERY, P. C., MORRIS, P. S., BYRNES, C. A., GRIMWOOD, K., TORZILLO, P. J., BAUERT, P. A., MASTERS, I. B., DIAZ, A., MCCALLUM, G. B., MOBBERLEY, C., TJHUNG, I., HARE, K. M., WARE, R. S. & CHANG, A. B. 2013. Long-term azithromycin for Indigenous children with non-cystic-fibrosis bronchiectasis or chronic suppurative lung disease (Bronchiectasis Intervention Study): a multicentre, double-blind, randomised controlled trial. *Lancet Respir Med,* 1**,** 610-20.

VAN DAMME, K., BEHETS, F., RAVELOMANANA, N., GODORNES, C., KHAN, M., RANDRIANASOLO, B., RABENJA, N. L., LUKEHART, S., COHEN, M. & HOOK, E. 2009. Evaluation of azithromycin resistance in *Treponema pallidum* specimens from Madagascar. *Sex Transm Dis,* 36**,** 775-6.

VAN DEN BELD, M. J. & REUBSAET, F. A. 2012. Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive Escherichia coli. *Eur J Clin Microbiol Infect Dis,* 31**,** 899-904.

VAN DEN BERGH, M. R., SPIJKERMAN, J., SWINNEN, K. M., FRANCOIS, N. A., PASCAL, T. G., BORYS, D., SCHUERMAN, L., IJZERMAN, E. P., BRUIN, J. P., VAN DER ENDE, A., VEENHOVEN, R. H. & SANDERS, E. A. 2013. Effects of the 10-valent pneumococcal nontypeable *Haemophilus influenzae* protein D-conjugate vaccine on nasopharyngeal bacterial colonization in young children: a randomized controlled trial. *Clin Infect Dis,* 56**,** e30-9.

VAN HAM, S. M., VAN ALPHEN, L., MOOI, F. R. & VAN PUTTEN, J. P. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell,* 73**,** 1187-96.

VAN TONDER, A. J., MISTRY, S., BRAY, J. E., HILL, D. M., CODY, A. J., FARMER, C. L., KLUGMAN, K. P., VON GOTTBERG, A., BENTLEY, S. D., PARKHILL, J., JOLLEY, K. A., MAIDEN, M. C. & BRUEGGEMANN, A. B. 2014. Defining the estimated coregenome of bacterial populations using a Bayesian decision model. *PLoS Comput Biol,* 10**,** e1003788.

VAN VLIET, A. H. 2016. Use of pan-genome analysis for the identification of lineage-specific genes of *Helicobacter pylori*. *FEMS Microbiol Lett*.

VANAPORN, M., WAND, M., MICHELL, S. L., SARKAR-TYSON, M., IRELAND, P., GOLDMAN, S., KEWCHAROENWONG, C., RINCHAI, D., LERTMEMONGKOLCHAI, G. & TITBALL, R. W. 2011. Superoxide dismutase C is required for intracellular survival and virulence of *Burkholderia pseudomallei*. *Microbiology,* 157**,** 2392-400.

VANNUCCHI, V. 1952. [Clinical study of a new antibiotic; erythromycin]. *Riv Crit ClinMed,* 52**,** 128-36.

VESIKARI, T., FORSTEN, A., SEPPA, I., KAIJALAINEN, T., PUUMALAINEN, T., SOININEN, A., TRASKINE, M., LOMMEL, P., SCHOONBROODT, S., HEZAREH, M., MOREIRA, M., BORYS, D. & SCHUERMAN, L. 2016. Effectiveness of the 10-Valent Pneumococcal Nontypeable *Haemophilus influenzae* Protein D-Conjugated Vaccine (PHiD-CV) Against Carriage and Acute Otitis Media-A Double-Blind Randomized Clinical Trial in Finland. *J Pediatric Infect Dis Soc,* 5**,** 237-48.

VESTER, B. 2000. Antibiotic Resistance in Bacteria Caused by Modified Nucleosides in 23S Ribosomal RNA. *In:* GROSJEAN, H. (ed.) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution.* Austin: Landes Bioscience.

VITOVSKI, S., DUNKIN, K. T., HOWARD, A. J. & SAYERS, J. R. 2002. Nontypeable *Haemophilus influenzae* in carriage and disease: a difference in IgA1 protease activity levels. *JAMA,* 287**,** 1699-705.

VITOVSKI, S., READ, R. C. & SAYERS, J. R. 1999. Invasive isolates of *Neisseriameningitidis* possess enhanced immunoglobulin A1 protease activity compared to colonizing strains. *FASEB J,* 13**,** 331-7.

VU, H. T., YOSHIDA, L. M., SUZUKI, M., NGUYEN, H. A., NGUYEN, C. D., NGUYEN, A. T., OISHI, K., YAMAMOTO, T., WATANABE, K. & VU, T. D. 2011. Association between nasopharyngeal load of *Streptococcus pneumoniae*, viral coinfection, and radiologically confirmed pneumonia in Vietnamese children. *Pediatr Infect Dis J,* 30**,** 11-8.

VUONG, J., WANG, X., THEODORE, J. M., WHITMON, J., GOMEZ DE LEON, P., MAYER, L. W., CARLONE, G. M. & ROMERO-STEINER, S. 2013. Absence of high molecular weight proteins 1 and/or 2 is associated with decreased adherence among non-typeable *Haemophilus influenzae* clinical isolates. *J Med Microbiol,* 62**,** 1649-56.

WAHL, R. & FOUACE, J. 1961a. [Mutations of a strain of *Staphylococcus aureus* towards resistance to erythromycin with loss of prophages]. *Ann Inst Pasteur (Paris),* 100**,** 553-68.

WAHL, R. & FOUACE, J. 1961b. [Phenotypic unstable resistance to erythromycin in *Staphylococcus pyogenes*]. *Ann Inst Pasteur (Paris),* 101**,** 722-37.

WALKOWIAK, S., ROWLAND, O., RODRIGUE, N. & SUBRAMANIAM, R. 2016. Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics,* 17**,** 1014.

WANG, X., MAIR, R., HATCHER, C., THEODORE, M. J., EDMOND, K., WU, H. M., HARCOURT, B. H., CARVALHO MDA, G., PIMENTA, F., NYMADAWA, P., ALTANTSETSEG, D., KIRSCH, M., SATOLA, S. W., COHN, A., MESSONNIER, N. E. & MAYER, L. W. 2011. Detection of bacterial pathogens in Mongolia meningitis surveillance with a new real-time PCR assay to detect *Haemophilus influenzae*. *Int J Med Microbiol,* 301**,** 303-9.

WANG, Y., HUANG, Y., WANG, J., CHENG, C., HUANG, W., LU, P., XU, Y. N., WANG, P., YAN, N. & SHI, Y. 2009. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature,* 462**,** 467-72.

WANG, Z., BAFADHEL, M., HALDAR, K., SPIVAK, A., MAYHEW, D., MILLER, B. E.,TAL-SINGER, R., JOHNSTON, S. L., RAMSHEH, M. Y., BARER, M. R., BRIGHTLING, C. E. & BROWN, J. R. 2016. Lung microbiome dynamics in COPD exacerbations. *Eur Respir J,* 47**,** 1082-92.

WEAVER, J. R. & PATTEE, P. A. 1964. Inducible Resistance to Erythromycin in *Staphylococcus aureus*. *J Bacteriol,* 88**,** 574-80.

WEBER, A., HARRIS, K., LOHRKE, S., FORNEY, L. & SMITH, A. L. 1991. Inability toexpress fimbriae results in impaired ability of *Haemophilus influenzae* b to colonize the nasopharynx. *Infect Immun,* 59**,** 4724-8.

WEBSTER, P., WU, S., GOMEZ, G., APICELLA, M., PLAUT, A. G. & ST GEME, J. W., 3RD 2006. Distribution of bacterial proteins in biofilms formed by non-typeable *Haemophilus influenzae*. *J Histochem Cytochem,* 54**,** 829-42.

WEISBLUM, B. 1995a. Erythromycin resistance by ribosome modification. *Antimicrob Agents Chemother,* 39**,** 577-85.

WEISBLUM, B. 1995b. Insights into erythromycin action from studies of its activity as inducer of resistance. *Antimicrob Agents Chemother,* 39**,** 797-805.

WELCH, H., RANDALL, W. A., REEDY, R. J. & KRAMER, J. 1952. Bacterial spectrum of erythromycin, carbomycin, chloramphenicol, aureomycin, and terramycin. *Antibiot Chemother (Northfield),* 2**,** 693-6.

WHO. 2017. WHO. Available: http://www.who.int/respiratory/copd/burden/en/ [Accessed 10th March 2017 2017].

WILKINSON, T. M., 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.

WINTER, L. E. & BARENKAMP, S. J. 2014. Antibodies to the HMW1/HMW2 and Hia adhesins of nontypeable *Haemophilus influenzae* mediate broad-based opsonophagocytic killing of homologous and heterologous strains. *Clin Vaccine Immunol*.

WITHERDEN, E. A., BAJANCA-LAVADO, M. P., TRISTRAM, S. G. & NUNES, A. 2014. Role of inter-species recombination of the *ftsI* gene in the dissemination of altered penicillin-binding-protein-3-mediated resistance in *Haemophilus influenzae* and *Haemophilus haemolyticus*. *J Antimicrob Chemother,* 69**,** 1501-9.

WITHERDEN, E. A. & TRISTRAM, S. G. 2013. Prevalence and mechanisms of beta-lactam resistance in *Haemophilus haemolyticus*. *J Antimicrob Chemother,* 68**,**1049-53.

WOODRUFF, P. G., BARR, R. G., BLEECKER, E., CHRISTENSON, S. A., COUPER, D., CURTIS, J. L., GOUSKOVA, N. A., HANSEL, N. N., HOFFMAN, E. A., KANNER, R. E., KLEERUP, E., LAZARUS, S. C., MARTINEZ, F. J., PAINE, R., 3RD, RENNARD, S., TASHKIN, D. P., HAN, M. K. & GROUP, S. R. 2016. Clinical Significance of Symptoms in Smokers with Preserved Pulmonary Function. *N Engl J Med,* 374**,** 1811-21.

WU, S., BAUM, M. M., KERWIN, J., GUERRERO, D., WEBSTER, S., SCHAUDINN, C., VANDERVELDE, D. & WEBSTER, P. 2014. Biofilm-specific extracellular matrix proteins of nontypeable *Haemophilus influenzae*. *Pathog Dis,* 72**,** 143-60.

WU, T., CHEN, J., MURPHY, T. F., GREEN, B. A. & GU, X. X. 2005. Investigation of non-typeable *Haemophilus influenzae* outer membrane protein P6 as a new carrier for lipooligosaccharide conjugate vaccines. *Vaccine,* 23**,** 5177-85.

XIE, J., JULIAO, P. C., GILSDORF, J. R., GHOSH, D., PATEL, M. & MARRS, C. F. 2006. Identification of new genetic regions more prevalent in nontypeable *Haemophilus influenzae* otitis media strains than in throat strains. *J Clin Microbiol,* 44**,** 4316-25.

YAGHOUBI, A., GHOJAZADEH, M., ABOLHASANI, S., ALIKHAH, H. & KHAKI-KHATIBI, F. 2015. Correlation of Serum Levels of Vitronectin, Malondialdehyde and Hs- CRP With Disease Severity in Coronary Artery Disease. *J Cardiovasc Thorac Res,* 7**,**113-7.

YAMAO, F., INOKUCHI, H., CHEUNG, A., OZEKI, H. & SOLL, D. 1982. Escherichia coli glutaminyl-tRNA synthetase. I. Isolation and DNA sequence of the *glnS* gene. *JBiol Chem,* 257**,** 11639-43.

YAMAUCHI, K., HOTOMI, M., BILLAL, D. S., SUZUMOTO, M. & YAMANAKA, N. 2006. Maternal intranasal immunization with outer membrane protein P6 maintains specific antibody level of derived offspring. *Vaccine,* 24**,** 5294-9.

YAMAYA, M., AZUMA, A., TANAKA, H., TAKIZAWA, H., CHIDA, K., TAGUCHI, Y., MIKASA, K., KADOTA, J. & KUDOH, S. 2008. Inhibitory effects of macrolide antibiotics on exacerbations and hospitalization in chronic obstructive pulmonary disease in Japan: a retrospective multicenter analysis. *J Am Geriatr Soc,* 56**,**1358-60.

YANG, Y. P., LOOSMORE, S. M., UNDERDOWN, B. J. & KLEIN, M. H. 1998. Nasopharyngeal colonization with nontypeable *Haemophilus influenzae* in chinchillas. *Infect Immun,* 66**,** 1973-80.

YEE, R., CUI, P., SHI, W., FENG, J. & ZHANG, Y. 2015. Genetic Screen Reveals the Role of Purine Metabolism in *Staphylococcus aureus* Persistence to Rifampicin. *Antibiotics (Basel),* 4**,** 627-42.

YI, K. & MURPHY, T. F. 1997. Importance of an immunodominant surface-exposed loopon outer membrane protein P2 of nontypeable *Haemophilus influenzae*. *Infect Immun,* 65**,** 150-5.

YIN, P., ZHANG, M., LI, Y., JIANG, Y. & ZHAO, W. 2011. Prevalence of COPD and its association with socioeconomic status in China: findings from China Chronic Disease Risk Factor Surveillance 2007. *BMC Public Health,* 11**,** 586.

ZARANTONELLI, L., BORTHAGARAY, G., LEE, E. H., VEAL, W. & SHAFER, W. M. 2001. Decreased susceptibility to azithromycin and erythromycin mediated by a novel *mtr(R)* promoter mutation in *Neisseria gonorrhoeae*. *J Antimicrob Chemother,* 47**,** 651-4.

ZHANG, B., KUNDE, D. & TRISTRAM, S. 2014. *Haemophilus haemolyticus* is infrequently misidentified as *Haemophilus influenzae* in diagnostic specimens in Australia. *Diagn Microbiol Infect Dis,* 80**,** 272-3.

ZHANG, J. & INOUYE, M. 2002. MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with Era, an essential GTPase in *Escherichia coli*. *J Bacteriol,* 184**,** 5323-9.

ZHANG, L., SU, Z., ZHANG, Z., LIN, J., LI, D. Q. & PFLUGFELDER, S. C. 2015. Effects of azithromycin on gene expression profiles of proinflammatory and anti-inflammatory mediators in the eyelid margin and conjunctiva of patients with meibomian gland disease. *JAMA Ophthalmol,* 133**,** 1117-23.

ZHONG, N., WANG, C., YAO, W., CHEN, P., KANG, J., HUANG, S., CHEN, B., WANG, C., NI, D., ZHOU, Y., LIU, S., WANG, X., WANG, D., LU, J., ZHENG, J. & RAN, P. 2007. Prevalence of chronic obstructive pulmonary disease in China: a large, population-based survey. *Am J Respir Crit Care Med,* 176**,** 753-60.

ZHOU, Y. & CHEN, R. 2013. Risk factors and intervention for chronic obstructive pulmonary disease in China. *Respirology,* 18 Suppl 3**,** 4-9.

ZIMIN, A. V., MARCAIS, G., PUIU, D., ROBERTS, M., SALZBERG, S. L. & YORKE, J. A. 2013. The MaSuRCA genome assembler. *Bioinformatics,* 29**,** 2669-77.