Pathogen Identification in Lower Respiratory Tract Infection

Thesis submitted for the degree of Doctor of Philosophy

April 2014

John Matthew Wrightson

Green Templeton College

Nuffield Department of Medicine

University of Oxford

1 Table of Contents

2 Figure legends

3 Table legends

4 Abstract

Treatment of lower respiratory tract infection (pneumonia and pleural infection) relies on the use of empirical broad spectrum antibiotics, primarily because reliable pathogen identification occurs infrequently. Another consequence of poor rates of pathogen identification is that our understanding of the microbiology of these infections is incomplete.

This thesis addresses some of these issues by combining the acquisition of high quality lower respiratory tract samples, free from nasooropharyngeal contamination, with novel molecular microbiological techniques in an attempt to increase rates of pathogen identification.

Four main areas are examined;

1. The role of so-called 'atypical pneumonia' bacteria in causing pleural infection. These pathogens have been previously identified in the pleural space infrequently and routine culture usually fails to isolate such bacteria. High sensitivity nested polymerase chain reaction (PCR) is a culture-independent technique which is used to undertake a systematic evaluation for these pathogens in pleural infection samples.
2. The role of *Pneumocystis jirovecii* in pleural infection, either as a co-infecting pathogen or in monomicrobial infection. This fungus causes severe pneumonia, particularly in the immunosuppressed, but is increasingly recognised as a co-pathogen in community-acquired pneumonia, and is frequently isolated in the upper and lower respiratory tract in health. A high sensitivity real-time PCR assay is used to examine for this fungus.

3. Ultra-deep sequencing of the 16S rRNA gene is used to perform a comprehensive microbial survey in samples taken from the multi-centre MIST2 study of pleural infection. The techniques employed allow analysis of polymicrobial samples and give very high taxonomic resolution, whilst incorporating methods to control for potential contamination. Further, these techniques provide confirmation of the results from the ‘atypical’ bacteria nested PCR study.
4. Bedside ultrasound-guided percutaneous transthoracic needle aspiration (TNA) of consolidated lung is undertaken in patients with pneumonia, as part of the PIPAP study. An evaluation is undertaken of the efficacy and acceptability of TNA. Aspirate samples acquired are also processed using ultra-deep sequencing of the 16S rRNA gene. Other samples obtained as part of the PIPAP study, such as ‘control’ lung aspirates and ‘control’ pleural fluid samples, are similarly processed to enable calculation of sensitivity and specificity of the sequencing methodology.

5 Acknowledgements

None of this work would have been possible without the vision, inspiration and support of the late Prof Rob Davies, who initially supervised my DPhil studies and encouraged me to develop as a Researcher and Respiratory Physician. Prof Derrick Crook co-supervised me, and took the lead when Rob died. He provided much needed guidance on molecular techniques and, as a Microbiologist, was endlessly patient with me as a Respiratory Physician. Dr Steve Chapman very kindly took over from Rob as co-supervisor, providing much needed support and Respiratory overview for my studies. I thank them all.

I would also like to thank Dr Teresa Street and Dr Jess Wray, both post-doctoral Scientists in the group who were enormously patient as I developed a molecular skill set.

Dr Naj Rahman and Prof Tim Peto have very graciously given their time, inspiration and encouragement and have provided significant mentorship and encouragement, particularly looking forwards to future work.

Finally, I am very grateful to the support, encouragement and time given by my wife, Ruth, and children, Sophie and Tommy, and our parents.

5.1 Funding bodies

I would like to express my gratitude to the Oxford NIHR Biomedical Research Centre for awarding me a DPhil Clinical Research Fellow studentship and to the Oxford Health Services Research Committee for awarding me a Grant to fund initial pilot work.

6 Publications and Conference Attendances

6.1 Publications over period of DPhil studentship

6.1.1 Journal publications

- Du Rand IA, Blaikley J, Booton R, Chaudhuri N, Gupta V, Khalid S, Mandal S, Martin J, Mills J, Navani N, Rahman NM, Wrightson JM, Munavvar M. Summary of the British Thoracic Society guideline for diagnostic flexible bronchoscopy in adults. *Thorax* **68,** 786–787 (2013).
- Du Rand IA, Blaikley J, Booton R, Chaudhuri N, Gupta V, Khalid S, Mandal S, Martin J, Mills J, Navani N, Rahman NM, Wrightson JM, Munavvar M. British Thoracic Society guideline for diagnostic flexible bronchoscopy in adults: accredited by NICE. *Thorax* **68,** i1–i44 (2013).
- Wrightson JM, Bateman KM, Hooper C, Gleeson FV, Rahman NM, Maskell NA. Development and efficacy of a 1-d thoracic ultrasound training course. *Chest* **142,** 1359–1361 (2012).
- Fysh ET, Wrightson JM, Lee YC, Rahman NM. Complications of Removal of Indwelling Pleural Catheters - Response. *Chest* **142,** 1071 (2012).
- Davies HE, Mishra EK, Kahan BC, Wrightson JM, Stanton AE, Guhan A, Davies CW, Grayez J, Harrison R, Prasad A, Crosthwaite N, Lee YC, Davies RJ, Miller RF, Rahman NM. Effect of an indwelling pleural catheter vs chest tube and talc pleurodesis for relieving dyspnea in patients with malignant pleural effusion: the TIME2 randomized controlled trial. *JAMA: The Journal of the American Medical Association* **307,** 2383–2389 (2012).
- Fysh ET, Wrightson JM, Lee YC, Rahman NM. Fractured indwelling pleural catheters. *Chest* **141,** 1090–1094 (2012).

- Wrightson JM, Maskell NA. Pleural infection. *Clin Med* **12,** 82–86 (2012).
- Tobin CL, Porcel JM, Wrightson JM, Waterer GW, Light RW, Lee YCG. Diagnosis of pleural infection: state-of-the-art. *Current Respiratory Care Reports* **1,** 101–110 (2012).
- Rahman NM, Maskell NA, West A, Teoh R, Arnold A, Mackinlay C, Peckham D, Davies CW, Ali N, Kinnear W, Bentley A, Kahan BC, Wrightson JM, Davies HE, Hooper CE, Lee YC, Hedley EL, Crosthwaite N, Choo L, Helm EJ, Gleeson FV, Nunn AJ, Davies RJ. Intrapleural Use of Tissue Plasminogen Activator and DNase in Pleural Infection. *N. Engl. J. Med.* **365,** 518–526 (2011).
- Menzies SM, Rahman NM, Wrightson JM, Davies HE, Shorten R, Gillespie SH, Davies CW, Maskell NA, Jeffrey AA, Lee YC, Davies RJ. Blood culture bottle culture of pleural fluid in pleural infection. *Thorax* **66,** 658–662 (2011).
- Wrightson JM, Davies HE. Outcome of patients with nonspecific pleuritis at thoracoscopy. *Curr Opin Pulm Med* **17,** 242–246 (2011).
- Wrightson JM, Maskell NA. Thoracic ultrasound for beginners: utility and training issues for clinicians. *Br J Hosp Med (Lond)* **72,** 325–330 (2011).
- Morel A, Mishra E, Medley L, Rahman NM, Wrightson J, Talbot D, Davies RJ. Chemotherapy should not be withheld from patients with an indwelling pleural catheter for malignant pleural effusion. *Thorax* **66,** 448–448 (2011).
- Wrightson JM, Rahman NM, Novak T, Huggett JF, Maskell NA, Zumla A, Miller RF, Davies RJ. Pneumocystis jirovecii in pleural infection: a nucleic acid amplification study. *Thorax* **66,** 450–451 (2011).

- Mahmood F, Davies RJO, Wrightson JM. Examination of the Respiratory System. *Journal of Clinical Examination* **11,** 69-89 (2011).
- Wrightson JM, Davies RJO. The approach to the patient with a parapneumonic effusion. *Semin Respir Crit Care Med* **31,** 706–715 (2010).
- Wrightson JM, Stanton AE, Maskell NA, Davies RJ, Lee YC. Could Decortication Become Necessary in Cases of Pseudochylothorax? - Response. *Chest* **138,** 1023–1024 (2010).
- Wrightson JM, Fysh E, Maskell NA, Lee YC. Risk reduction in pleural procedures: sonography, simulation and supervision. *Curr Opin Pulm Med* **16,** 340–350 (2010).
- Rahman NM, Singanayagam A, Davies HE, Wrightson JM, Mishra EK, Lee YC, Benamore R, Davies RJ, Gleeson FV. Diagnostic accuracy, safety and utilisation of respiratory physician-delivered thoracic ultrasound. *Thorax* **65,** 449–453 (2010).
- Wrightson JM, Stanton AE, Maskell NA, Davies RJ, Lee YC. Pseudochylothorax, an Unknown Disease - Response. *Chest* **137,** 1005–1005 (2010).
- Wrightson JM, Stanton AE, Maskell NA, Davies RJ, Lee YC. Pseudochylothorax Without Pleural Thickening: Time to Reconsider Pathogenesis? *Chest* **136,** 1144–1147 (2009).
- Wrightson JM, Davies RJO. Preventing, Diagnosing and Treating Pleural Infection and Malignant Pleural Effusion. *European Respiratory Disease* (2009).
- Wrightson JM, Helm EJ, Rahman NM, Gleeson FV, Davies RJ. Pleural procedures and pleuroscopy. *Respirology* **14,** 796–807 (2009).

6.1.2 Book chapters

- Wrightson JM, Davies HE, Lee YCG. Pleural Effusion, Empyema and Pneumothorax. Book chapter in: *Clinical Respiratory Medicine* (ed. Spiro SG, Silvestri GA, Agusti A; ISBN 1455707929).
- Wrightson JM, Mishra EK, Davies RJO. Pneumonia. Book chapter in: Understanding Medical Research: The Studies That Shaped Medicine (ISBN 0470654481).

6.2 Conference presentations

- Wrightson JM, Wray JA, Rahman NM, Crook DWM. Low Prevalence of "Atypical" Pathogens in Pleural Infection. *Am J Respir Crit Care Med* **185,** A5342 (2012).
- Wrightson JM, Rahman NM, Novak T, Huggett JF, Miller RF, Davies RJO. Absence of Pneumocystis jirovecii in pleural infection: A genetic amplification study. *Thorax* **63,** A138 (2008).

6.3 Conference posters

- Wrightson JM, Rahman NM, Crook DWM, Wray JA. Improving Pathogen Identification in Pleural Infection – Application of Molecular Techniques. *Am J Respir Crit Care Med* **185,** A5244 (2012).
- Davies H, Mishra EK, Wrightson JM, Stanton A, Guhan A, Davies C, Grayez J, Harrison R, Prasad A, Crosthwaite N, Lee YC, Miller R, Kahan B, Rahman NM. The Second Therapeutic Intervention in Malignant Effusion Trial (TIME2):

A Randomised Controlled Trial to Assess the Efficacy and Safety of Patient Controlled Malignant Effusion Drainage by Indwelling Pleural Catheter Compared to Chest Tube and Talc Slurry Pleurodesis. *Am J Respir Crit Care Med* **185,** A6861 (2012).

- Wrightson JM, Bateman K, Hooper C, Gleeson FV, Rahman NM, Maskell NA. Training in Thoracic Ultrasound – Development of a National Ultrasound Course. *Am J Respir Crit Care Med* **185,** A1559 (2012).
- Hallifax RJ, Nagendran M, Maruthappu M, Manuel A, Wrightson JM, Maskell NA, Davies HE, Rahman NM. Effusion Size on the Chest Radiograph at Day 7 Post-Pleural Drainage is a Good Predictor of Size at 3 Months, Regardless of Initial Intrapleural Therapy. *Thorax* **67,** A116 (2012).
- Corcoran JP, Lockey JW, Zheng SL, Wrightson JM, Rahman NM. The bacteriology and epidemiology of pleural infection–a review of the literature and its relevance to clinical practise. *Thorax* **67,** A9 (2012).
- Morel A, Mishra E, Medley L, Rahman NM, Wrightson J, Talbot D, Davies RJO. Chemotherapy should not be withheld from patients with an indwelling pleural catheter for malignant pleural effusion. *Thorax* **66,** 448–448 (2011).
- Rahman NM, Maskell N, Davies CWH, West A, Teoh R, Arnold A, Peckham D, Ali N, Bentley A, Mackinlay CI, Kinnear WJ, Wrightson JM, Davies HE, Miller RF, Lee YCG, Hedley EL, Crosthwaite N, Choo L, Darbyshire J, Gleeson FV, Nunn AJ, Davies RJO. Primary Result of the 2nd Multi-centre Intrapleural Sepsis Trial (MIST2); Randomised Trial of Intrapleural tPA and DNAse in Pleural Infection. *Am J Respir Crit Care Med* **181,** A6143 (2010).
- Mishra EK, Morel A, Medley L, Rahman NM, Wrightson JM, Talbot DC, Davies RJO. Chemotherapy does not increase the risk of pleural infection in

patients treated with an indwelling pleural catheter for malignant pleural effusion. *Am J Respir Crit Care Med* **181,** A4345 (2010).

- Morel A, Mishra EK, Medley L, Rahman NM, Wrightson JM, Talbot DC, Davies RJ. Chemotherapy does not increase the risk of pleural infection in patients managed with an indwelling pleural catheter for malignant pleural effusion. *Lung Cancer* **67,** S6–S7 (2010).
- Rahman NM, Maskell N, Davies CWH, West A, Teoh R, Arnold A, Peckham D, Ali N, Bentley A, Mackinlay CI, Kinnear WJ, Wrightson JM, Davies HE, Miller RF, Lee YCG, Hedley EL, Crosthwaite N, Choo L, Darbyshire J, Gleeson FV, Nunn AJ, Davies RJO. Primary result of the second Multicentre Intrapleural Sepsis (MIST2) Trial; randomised trial of intrapleural TPA and DNase in pleural infection. *Thorax* **64,** A1 (2009).
- Rahman NM, Davies HE, Wrightson JM, Singanayagam A, Singanayagam A, Benamore R, Gleeson FV, Davies RJO. Activity, efficacy and safety of a physician-based thoracic ultrasound (PBUS) service. *Am J Respir Crit Care Med* **179,** A4456 (2009).
- Wrightson JM, Rahman NM, Novak T, Huggett JF, Miller RF, Davies RJO. Polymerase Chain Reaction (PCR) Demonstrates No Evidence of Pneumocystis jirovecii in Pleural Infection. *Am J Respir Crit Care Med* **179,** A4469 (2009).

7 Attributions

My direct involvement in the work represented by each chapter is described below.

Chapter 10 – Atypical pathogens in Pleural Infection

Direct contribution – concept development; recruitment of some participants; sample management; extraction of DNA from some samples; data analysis and manuscript drafting. Dr J Wray (Nuffield Department of Medicine, University of Oxford) extracted DNA from the majority of samples and performed nucleic acid amplification and analysis. Dr T Street (Nuffield Department of Medicine, University of Oxford) performed the capillary sequencing.

Chapter 10 – *Pneumocystis jirovecii* in Pleural Infection

Direct contribution – sample management; data analysis; manuscript drafting. Drs T Novak, JF Huggett, A Zumla, RF Miller (University College London, UK) undertook the nucleic acid assays and analysis.

Chapter 12 – MIST2 pleural infection ultra-deep sequencing

Direct contribution – concept development; recruitment of some participants; sample management; extraction of DNA from many samples; development of qPCR, capillary sequencing and ultra-deep sequencing assays; undertaking PCR assays; undertaking emPCR assays; undertaking 454 FLX sequencing; data analysis. Dr J Wray (Nuffield Department of Medicine, University of Oxford) extracted DNA from further samples.

Chapter 13 – PIPAP study

Direct contribution – concept development; protocol, consent form and case record form drafting; recruitment of all participants; sample acquisition and management; extraction of DNA from all samples; development of qPCR, capillary sequencing and ultra-deep sequencing assays; undertaking PCR assays; undertaking emPCR assays; undertaking 454 FLX sequencing; data analysis.

8 Glossary of Abbreviations

AIDS	Acquired immunodeficiency syndrome
AUC	Area under the curve
BAL	Bronchoalveolar lavage
bp	Base pair
BTS	British Thoracic Society
CA-MRSA	Community-acquired MRSA
CAP	Community-acquired pneumonia
CFT	Complement fixation testing
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
Cq	Quantification cycle
CRP	C-reactive protein
CT	Computed tomography
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
EIA	Enzyme immunoassay
emPCR	Emulsion PCR
GNAB	Gram negative anaerobic bacilli
GPAB	Gram positive anaerobic bacilli
GPAC	Gram positive anaerobic cocci
HAP	Hospital-acquired pneumonia
HCAI	Healthcare associated infection
HCAP	Healthcare associated pneumonia
HIV	Human immunodeficiency virus
hMPV	Human metapneumovirus
HSP70	Heat shock protein 70
IL	Interleukin
IQR	Interquartile range
iv	Intravenous
LDH	Lactate dehydrogenase
MID	Multiplex identifier
MRSA	Methicillin-resistant *Staphylococcus aureus*
NAAT	Nucleic acid amplification test

NMDS	Non-metric multidimensional scaling
NPA	Negative percent agreement
nt	Nucleotide
OR	Odds ratio
OTU	Operational Taxonomic Unit
PAI	Plasminogen activator inhibitor
PCoA	Principal Coordinate Analysis
PCR	Polymerase chain reaction
PF	Pleural fluid
PIPAP	Pathogen Identification in Pneumonia and Pleural Infection study
PIV	Parainfluenza virus
po	Per os (oral)
PPA	Positive percent agreement
PSI	Pneumonia severity index
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative (real time) PCR
RDP	Ribosomal database project
REC	Research ethics committee
RNA	Ribonucleic acid
ROC	Receiver operator curve
rRNA	Ribosomal RNA
RSV	Respiratory syncytial virus
RT PCR	Reverse transcriptase PCR
SD	Standard deviation
TNA	Transthoracic needle aspiration
TNF	Tumour necrosis factor
tPA	Tissue plasminogen activator
UV	Ultraviolet
VAS	Visual analogue scale
VATS	Video-assisted thoracoscopic surgery
WCC	White blood cell count
WGS	Whole genome sequencing

9 Introduction

9.1 Pneumonia

9.1.1 Overview

Pneumonia is a major global health problem – it accounts for more than 6% of the total burden of all disease and causes more disease than cancer, myocardial infarction, stroke, HIV/AIDS or malaria (figure 9-1). Despite this, it has attracted a smaller quantity of research activity and funding than these other areas. In both the developed and developing world, pneumonia causes the highest disease burden of all infectious disease(1). It is the UK's third leading cause of death in women and the fifth leading cause of death in men. Between 6 and 13% of patients admitted to hospital with community-acquired pneumonia die(2-5).

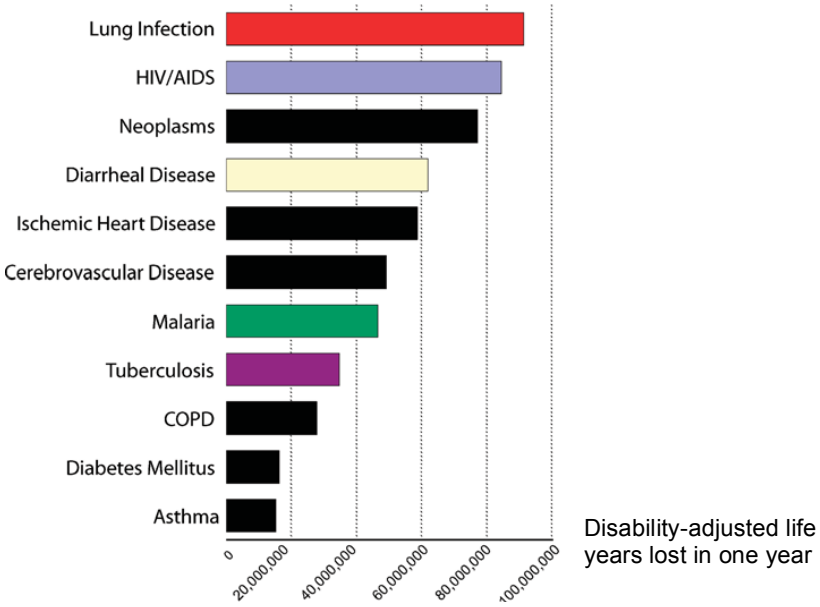


Figure 9-1 – The global burden of major diseases (from Mizgerd 2006(1))

Even patients who do not die of pneumonia often experience major morbidities. 10% of patients require admission to the intensive care unit(6). Many of these

require ventilatory support for respiratory failure. Sepsis may cause shock requiring inotropic and vasopressor support. Other patients suffer from renal failure. Patients develop other major complications, such as lung abscess or pleural infection/empyema.

One of the earliest descriptions of pneumonia is attributed to Hippocrates of Cos. Sir William Osler, who himself died of pneumonia complicated by pleural infection, stated that pneumonia was “one of the most widespread and fatal of all acute diseases” and continued “pneumonia has become the ‘Captain of the Men of Death’ to use the phrase applied by John Bunyan to consumption”(7).

Community-acquired pneumonia (CAP) incidence is estimated at ~5-10 per 1000 adult population per year, with a much higher incidence in the very young and elderly(8-11). 20-42% of adults with CAP are hospitalised(8,12). English Hospital Episode Statistic data demonstrated a 34% increase in age-standardised incidence of hospitalisation between 1997 and 2005(13). Average hospital stay was 5.8 to 7.8 days in one study for patients under and over 65 years respectively(14).

9.1.2 Pathogenesis

The human lower respiratory tract is a common site of infection by bacteria, viruses and (less commonly) fungi, predominantly because it is continuity with the environment and the upper respiratory tract. Approximately 8000 litres of air enter the lung per day during normal ventilation. Pathogens contained within this air, and saliva and other secretions aspirated from the upper airways give multiple

opportunities for the development of lower respiratory tract infection, particularly in the context of immunosuppression (e.g. older patients, general debility, diabetes, alcoholism, malignancy, HIV infection and immunosuppressive drugs), smokers and those with any underlying respiratory disease(15).

9.1.3 Diagnosis

Hospital-based diagnosis of pneumonia is based on a constellation of symptoms, laboratory investigations (including white cell count and C-reactive protein) and radiological imaging. The latest British Thoracic Society (BTS) guidelines(16) define hospital-treated community-acquired pneumonia as follows;

Symptoms and signs consistent with an acute lower respiratory tract infection associated with new radiographic shadowing for which there is no other explanation

The illness is the primary reason for hospital admission and is managed as pneumonia.

Pneumonia is confirmed using chest X-ray, or occasionally using computed tomography (CT) imaging, although chest X-ray changes (deterioration and improvement) may lag behind clinical symptoms and examination findings.

9.1.4 Treatment

Prior to the advent of antibiotic therapy (initially sulphonamides and subsequently penicillins), serum therapy reduced the mortality of pneumonia from 31%

untreated to 15%. Use of sulphonamides further reduced mortality to 8% in a series between 1938-1941(17). A study from 1964 showed that the mortality from bacteraemic pneumonia due to *Streptococcus pneumoniae* (the 'pneumococcus') fell from 83% to 47% with serum therapy, and to 15% with penicillin(17).

The treatment of most bacterial infectious disease depends on identification of causative pathogens and the selection of targeted antibiotics – such as in urinary tract infections and meningitis. This is in striking contrast to the usual treatment of pneumonia, where the organism is infrequently identified (only 5-14% of blood cultures are positive(4,18-20)). In view of the poor rates of organism identification in pneumonia, most cases are usually assumed to be caused by bacteria (rather than viruses or fungi) and antibiotic selection is necessarily empirical and broad-spectrum for the full course of treatment.

A severity assessment is normally undertaken to help determine risk of complications (including mortality), and hence optimal location for treatment and antibiotic regime. Commonly used severity scoring systems (outcome predictive models) include the CURB65 score(21) and the Pneumonia Severity Index (PSI)(22). The PSI derives a score based on 20 variables, which assigns patients to one of five classes, each with a distinct estimate of 30-day mortality. CURB65 is a point system based on presence of the following clinical or laboratory features: **C**onfusion, **U**rea >7 mmol/l, **R**espiratory rate ≥30/min, **B**lood pressure <90 mm Hg systolic or ≤60 mm Hg diastolic, and age ≥**65** years. Addition of points scored places patients in one of six classes, each with a distinct mortality.

For patients admitted to hospital, Table 9-1 summarises the British Thoracic Society's suggested empirical antibiotic selection(16).

Pneumonia severity (clinical judgement combined with CURB65 score)	**Preferred treatment** (assuming no allergy)
Low severity (e.g. CURB65 = 0-1)	Amoxicillin 500mg tds po (iv if po not possible)
Moderate severity (e.g. CURB65 = 2)	Amoxicillin 500mg-1g tds po + clarithromycin 500mg bd po (iv amoxicillin or benzylpenicillin if po not possible)
High severity (e.g. CURB65 ≥ 3)	Co-amoxiclav 1.2g tds iv + clarithromycin 500mg bd iv (+ levofloxacin if legionella strongly suspected)

Table 9-1 – Empirical antibiotic therapy for CAP, modified from BTS guidance

Treatment for 7 days is suggested for most patients with low or moderate severity uncomplicated pneumonia, but an extended course may be required for severe pneumonia.

De-escalation of broad spectrum antibiotics to specific pathogen directed therapy is encouraged when a pathogen is discovered to both maximise the efficacy of antibiotics, but also to narrow antimicrobial spectrum as much as possible. The clinical reality is that this is achievable infrequently due to failure to identify an infecting pathogen.

9.2 Pleural infection

9.2.1 Overview

'Parapneumonic' pleural effusions develop in up to 57% of patients with pneumonia(23-25). These effusions vary in size from small sub-centimetre effusions to large effusions that may cause ventilatory embarrassment.

The majority of parapneumonic effusions are non-infected *simple* parapneumonic effusions that usually resolve with antibiotic therapy. However, about 40%(25) of

cases progress to pleural infection – either infected *complicated* parapneumonic effusions or frank pus (empyema), the incidence of which appears to be increasing(26,27). Pleural infection affects more than 65,000 patients per year in the UK and USA(28). Complicated effusions and empyema require prompt treatment, being associated with significant morbidity and mortality; 15% of patients require surgery to recover and 22% die(29). Early identification and treatment of pleural infection is also essential as effective resolution becomes more difficult to achieve, with higher associated healthcare costs, as the disease progresses(27). The median length of stay in hospital is 12 to 15 days and 25% of these patients are in hospital for more than one month.

The first description of pleural infection is attributed to Hippocrates of Cos, recognising that 'pneumonia coming on pleurisy is bad'. Mortality associated with pleural infection was up to 70%, although the advent of closed tube drainage and antibiotics led to a gradual reduction to 30% by the 1950s(30).

9.2.2 Increasing prevalence

Pleural infection is more common in children and the elderly. Many cohort studies suggest that incidence is increasing in all age groups, particularly in children(27,31-34). One USA-based study found that hospitalisation rates had doubled (to 6 per 100,000/year) between 1996 and 2008(35). Another study found that pleural infection rates have been increasing by 3% per year for the past two decades(26). *Steptococcus pneumoniae* vaccination also appears to be affecting the rate of pleural infection, particularly in children – following the introduction of the 7-valent pneumococcal conjugate vaccine (PCV), a replacement phenomenon

is likely occurring, with a reduction in invasive pneumococcal disease due to PCV-7 serotypes (4, 6B, 9V, 14, 18C, 19F and 23F) but increase in rates of pleural infection caused particularly by serotype 1 (and also 3, 5, 7F and 19A)(36-40). In adults, diabetes, the immunosuppressed state, alcohol misuse and intravenous drug abuse have been shown to be risk factors for pleural infection. Aspiration and poor oral hygiene are more common in anaerobic pleural infection. A prospective study found serum albumin <30g/l, C-reactive protein >100mg/l, platelet count >400x10^9/l, serum sodium <130mmol/l, intravenous drug use and alcohol misuse were predictive of development of complicated effusions in patients with pneumonia(25).

9.2.3 Pathogenesis

The natural history of an untreated parapneumonic effusion may be considered in terms of evolving infection, associated with progressive fibrin deposition and septation development, ultimately resulting in pleural fibrosis if left untreated and the patient survives.

A parapneumonic effusion is thought to start as a ‘simple’ non-infected exudative effusion. This develops when pneumonia-associated neutrophils and monocytes migrate across the pleural mesothelial cell layer into the pleural space. These cells cause an increase in intrapleural inflammatory cytokines, e.g. interleukin-8 (IL-8) and tumour necrosis factor-α (TNF-α)(41,42), leading to increased vascular permeability, ultimately resulting in an initially sterile pleural effusion.

Subsequent bacterial migration causes a complicated effusion. Bacteria create a pro-coagulant, depressed fibrinolysis state, with increased levels of plasminogen activator inhibitor (PAI) and decreased tissue-type plasminogen activator (tPA)(43). Consequently, septations are formed as fibrin is produced – this is known as the 'fibrinopurulent' stage. Intrapleural bacterial metabolism and host cell phagocytic activity utilises glucose and produces lactic acid resulting in the characteristic biochemical findings found in pleural infection(44,45), namely low pleural fluid pH and glucose with elevated lactate dehydrogenase (LDH). Left untreated, frank pus may develop within the pleural space, known as empyema.

The final stage of progression of an untreated parapneumonic effusion is the 'organising' stage in which fibroblast proliferation creates a solid fibrous pleural peel, often requiring surgical decortication to obliterate the infected pleural space and allow lung re-expansion.

Early medical treatment of parapneumonic effusions is aimed at arresting the progression towards a heavily fibrotic septated pleural space using antibiotic treatment and appropriate tube drainage.

9.2.4 Diagnosis

Key to the diagnosis and treatment of pleural infection is the distinction between non-infected simple parapneumonic effusions and infected complicated parapneumonic effusions and empyema.

The initial investigation when pleural infection is suspected is a chest X-ray. 200-300ml of pleural fluid may be detected using a posteroanterior X-ray(46,47). There

are no radiographic features with high specificity for pleural infection, although a new loculated effusion is suspicious for infection in the context of sepsis. Thoracic ultrasound has a higher sensitivity than chest X-ray in detecting pleural fluid(48,49) and is able to accurately locate loculated fluid(50), estimate pleural fluid depth and volume(51-54). Sonographic features allow effusions to be characterised into anechoic, complex nonseptated, complex septated and homogenously echogenic effusions(55). None of these features rules out pleural infection, but a septated or heavily echogenic effusion is strongly suspicious of pleural infection in the appropriate clinical context (although the differential still includes any cause of inflammatory effusions, including pleural malignancy).

Pleural fluid septation density has been proposed to predict outcome, in particular chest tube drainage success. Studies have shown a success rate of drainage for 92-100% of anechoic effusions, 79-82% of complex non-septated effusions, 80% of homogenously echogenic effusions and 51-63% of complex septated effusions(56-58). Chen *et al*.(59) found a higher treatment failure rate for complex septated effusions compared with non-septated effusions. However, others have failed to demonstrate an association with outcome or surgical requirements(60). Overall, septated effusions should not deter physicians from using chest tubes given that the majority of such effusions still drain well.

Having imaged the pleural space, fluid aspiration allows the effusion to be characterised and infection to be confirmed. Parapneumonic effusions are usually neutrophil-predominant exudates. Culture is the gold-standard for diagnosis of pleural infection, but is an imperfect test having a sensitivity of only 30-40% and taking 24-72 hours for bacterial culture. Given these issues, biochemical proxies of infection are required to enable immediate decisions regarding treatment.

A variety of pleural fluid laboratory tests have been advocated for detecting pleural infection requiring urgent fluid drainage, including glucose, lactate dehydrogenase (LDH), adenosine deaminase, C-reactive protein (CRP) and white blood cell differential, but none are superior to pleural fluid pH(61). Bacterial metabolism and host leucocyte response causes a characteristic pleural space biochemical profile, with low pH and glucose and high LDH and CRP. A meta-analysis showed that a receiver operator curve (ROC) for pleural fluid pH had the highest AUC (area under the ROC curve) compared with glucose and LDH (0.89 vs. 0.71 vs. 0.71, respectively)(62), and pH <7.20 is conventionally taken as a cut-off for pleural infection(63). Pleural fluid glucose (<2.2mmol/L) and LDH (>1000 U/L) do not improve diagnostic reliability but are occasionally used when pH is unavailable. Caution must be exercised in taking these values as rigid; advanced malignant effusions, rheumatoid arthritis, lupus pleuritis and tuberculosis can all create a low pH, low glucose pleural effusion. Further, urea splitting organisms, such as *Proteus* spp. may cause an alkalotic pleural effusion. Patients regarded as having a 'simple' non-infected parapneumonic effusion who are failing to improve should be considered for chest tube drainage regardless of pleural fluid biochemical values. The differing locules of a multiloculated effusion have been shown to have differing biochemical profiles(64), which may fall on either side of cut-off values for drainage. Indeed, a loculated effusion is highly suggestive of a fibrinopurulent stage complicated effusion and associated with a poorer outcome(44,65,66).

Table 9-2 summarises characteristics of parapneumonic effusions.

	Simple parapneumonic effusions	Complicated parapneumonic effusions	Empyema
Proportion of pneumonia-associated effusions	~60%	~40%	
Appearance	Clear or slightly turbid	Usually cloudy	Pus
pH	≥7.20	<7.20	Not usually measured
Glucose	≥2.2mmol/L	<2.2mmol/L	Not usually measured
LDH	≤1000U/L	>1000U/L	Not usually measured
Microbiological positivity	No	~25%	~70%
Usual treatment	Antibiotic therapy alone	Broad spectrum antibiotics and chest tube drainage	

Table 9-2 – Characterisation of parapneumonic effusions

Various other biomarkers of infection, including tumour necrosis factor-α, myeloperoxidase, matrix metalloproteinase-2, neutrophil elastase, interleukin-8, lipopolysaccharide binding protein and soluble triggering receptor expressed on myeloid cells all have been proposed to have potential utility(67-73), but have not been robustly validated.

9.2.5 Treatment

Treatment for pleural infection requires urgent fluid drainage and antibiotic therapy, which is initially broad spectrum given the difficulty isolating and broad range of possible causative organisms.

Antibiotic therapy

Patients with pleural infection should be treated with empiric broad-spectrum antibiotic therapy until culture results are available. Antibiotic choice will be determined in accordance with local antibiotic policy and resistance patterns but community-acquired pathogens are normally covered by a beta-lactam antibiotic in conjunction with the beta-lactamase inhibitor, such as amoxicillin & clavulinic

acid or piperacillin-tazobactam. Metronidazole is frequently added to increase anaerobic coverage. Healthcare-associated pleural infection is often associated with resistant bacteria, and a reasonable choice of antibiotic is a carbapenem combined with vancomycin. Although it has not been robustly tested in a randomised controlled trial, antibiotic treatment duration is usually given for a total of 3 to 4 weeks. Initial therapy is with intravenous antibiotics, with oral agent conversion being guided by clinical course.

Pleural fluid drainage

Pleural infection requires prompt tube drainage. Simple non-infected parapneumonic effusions do not usually require drainage. Recent evidence suggests that small bore tubes (<15 F) have a similar efficacy to large bore tubes in draining pus, and are associated with less pain(74). A chest tube flush regime, such as 20mL 0.9% sodium chloride solution every six hours, is often used for small bore tubes, together with suction using a dedicated thoracic suction unit.

Adjunctive intrapleural medication

The role of intrapleural fibrinolytics in improving the drainage of poorly resolving heavily septated pleural infection has been investigated. Despite small studies suggesting that the fibrinolytic streptokinase may improve fluid drainage when instilled into the pleural space, a large randomised controlled trial, MIST1, showed that intrapleural streptokinase did not improve clinical outcome(28).

The MIST2 randomised controlled trial of intrapleural tissue plasminogen activator (tPA) and deoxyribonuclease (DNase; an enzyme which disrupts DNA) suggests that the combination of tPA and DNase increase the quantity of pleural fluid

drained(75). However, further studies are required to define the clinical treatment effect, and to define those patients who should undergo immediate ‘up-front’ surgical pleural debridement.

Consideration of surgery

30% of patients fail to respond to medical management and these patients should be considered for an early surgical opinion. There is no published data on the timing or clinical criteria for surgical referral, although it is common practice to refer patients with ongoing signs of sepsis and incomplete pleural drainage on the fifth day of medical management. Conversely, patients who have residual pleural fluid, but are otherwise well with improving clinical and laboratory parameters will normally have gradual resolution of the pleural fluid from the pleural space over time. Video-assisted thoracoscopic surgery (VATS) enables decortication of pleural thickening, septation division and fluid removal, allowing lung re-expansion. VATS is performed under general anaesthesia with single lung ventilation, but in expert centres this can be performed using regional anaesthesia. Several studies have investigated the role of primary VATS versus chest tube drainage on initial presentation with pleural infection, but the methodological limitations of these studies means that definitive evidence is lacking(76,77).

Nutritional support

Weight loss and low serum albumin concentration are commonplace in pleural infection and associated with poorer outcome(78). Whilst specific nutritional therapy has not been subject to a randomised controlled trial in this setting,

nutritional support is likely to be important in counteracting the catabolic state associated with infection.

9.3 Microbiology of pneumonia and pleural infection

9.3.1 Microbiology of pneumonia

Comparison of studies examining the pathogens causing pneumonia is problematic, predominantly due to differences in clinical samples collected and investigations performed in the varying studies. Many other factors are also relevant, including whether patients are hospitalised, patient demographic characteristics, and pneumonia severity assessment. Clinical features are not specific enough to reliably predict the infective aetiology. Another factor to be considered is that some infections may be polymicrobial, and identification of one infecting organism does not necessarily out-rule other pathogens.

Table 9-3 summarises the pathogens found in five studies of CAP patients admitted to hospital in the UK. Sample limitations, the relative insensitivity associated with conventional microbiological testing and a lack of a gold standard means that this is almost certainly an incomplete picture. It should also be noted that pathogens are frequently (~30%) not found, even in these dedicated pathogen identification studies.

Pathogen	% frequency of identification, mean (95% CI)
Streptococcus pneumoniae	39 (36.1-41.8)
Haemophilus influenzae	5.2 (4.0-6.6)
Legionella spp.	3.6 (2.6-4.9)
Staphylococcus aureus	1.9 (1.2-2.9)
Moraxella catarrhalis	1.9 (0.6-4.3)
Enterobacteriaceae	1.0 (0.5-1.7)
Mycoplama pneumoniae	10.8 (9.0-12.6)
Chlamydophila pneumoniae	13.1 (9.1-17.2)
Chlamydophila psittaci	2.6 (1.7-3.6)
Coxiella burnetii	1.2 (0.7-2.1)
Viruses	12.8 (10.8-14.7)
Mixed	14.2 (12.2-16.3)
Other	2.0 (1.3-3.0)
None	30.8 (28.1-33.5)

Table 9-3 – Frequency of pathogen identification in five studies (total 1137 patients) for patients with CAP admitted to hospital in the UK(79-83), modified from Lim *et al.*(16)

An ongoing large prospective multicentre pneumonia study, CAPNETZ, has the potential to add to our understanding of the aetiology of pneumonia and factors predictive of a poor outcome. The CAPNETZ investigators have developed a network of 17 (mostly German) centres and have recruited more than 10000 patients with community-acquired pneumonia since 2002. Aetiological investigation using sputum and blood cultures, pharyngeal aspiration, urinary antigen testing and also nucleic acid amplification tests and serological studies demonstrate a similar distribution of pathogens to those shown in Table 9-3(84). The power associated with such a large study has enabled an assessment of the correlation of inflammatory markers and aetiological pattern, finding that atypical and viral aetiologies are associated with a significantly lower level of procalcitonin, CRP and WCC. Further, procalcitonin concentration (but not WCC or CRP) are associated with CAP severity(84). The investigators have also found that other serum biomarkers, such as IL-6, IL-10 and lipopolysaccharide-binding protein are

associated with severe CAP(85). Several other community-acquired pneumonia networks exist, including the CAPUCI group (community-acquired pneumonia in the Intensive Care Unit, a Spanish multicentre group(86,87)) and the CAPO (community-acquired pneumonia organization) international cohort study(88).

Previous studies and guidelines(16) have suggested that microbiology partly varies by region. For example, *Klebsiella pneumoniae* is commmonly seen in South Africa(89), *Burkholderia pseudomallei* is particularly seen in the Asiopacific region(90,91) and *Coxiella burnetii* is more common in Spain(92).

Specific patterns of pathogens may be seen in certain groups, with a possible higher rate of Enterobacteriaceae and anaerobic infection associated with Nursing Home residents. Alcoholism has also been variably associated with *S. pneumoniae*, Enterobacteriaceae, anaerobic infection and atypical infections. Environmental sources are also of relevance, with *Legionella* spp. being associated with building water-containing systems, *Chlamydophila psittaci* being associated with birds and animals and *Coxiella burnetii* being associated with farm workers in contact with sheep(16).

Staphylococcus aureus may be associated with influenza, with evidence of influenza infection in 39% of patients admitted with *S. aureus* pneumonia(80-83). Rates of community-acquired methicillin resistant *Staphylococcus aureus* (CA-MRSA) infection appear to be increasing(93-95), and this is of significant concern due to its inherent resistance to standard empirical CAP treatment.

9.3.2 Microbiology of pleural infection

Gram-positive aerobic bacteria are the commonest cause of both community and healthcare-associated pleural infection. *Streptococcus* 'milleri' (*constellatus-intermedius-anginosus*) group are the commonest in community-acquired infection, being isolated in 24-40% of cases(28,29,96-100). Further species identification of these bacteria is often not performed in microbiological practice, and molecular differentiation is problematic as they share very similar 16S rRNA gene sequences. Streptococcus 'milleri' group bacteria are a heterogeneous group of facultatively Gram-positive anaerobic cocci of the viridans streptococci group, often found in the oropharynx. *Streptococcus pneumoniae* is also commonly isolated, with a prevalence of 21% in the MIST1 study (using culture and molecular techniques)(28,29,96). Gram negative organisms, such as *Escherichia coli*, *Haemophilus influenzae* and *Klebsiella pneumoniae* occur in about 10% of community-acquired infections(29,101). Anaerobic organisms, including *Bacteroides fragilis*, are detected in about 12-34%(101,102). Polymicrobial infection is fairly common, particularly in the elderly and those with co-morbidities. The isolation of 'atypical' pneumonia pathogens has previously been reported in the pleural space(103-108), but appears to be rare in larger studies.

Similar to pneumonia, there also appears to be worldwide variation in bacteriology – for example, two studies found that *Klebsiella pneumoniae* was the commonest cause of pleural infection in adults in Taiwan(109,110).

In healthcare-associated pleural infection, resistant pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and Gram negative organisms, such as

E. coli, *Enterobacter* spp. and *Pseudomonas* spp., are more common(28,29), with decreased isolation of *S. pneumoniae* and the S. 'milleri' group.

MRSA pneumonia may be particularly prone to development of pleural infection with 35-75% of patients being affected; such infections are associated with very high mortality(111-113).

Fungal pleural infection is apparently rare, but important to consider in immunosuppressed patients and has mortality in excess of 70% with *Candida* spp. being the most commonly isolated.

9.4 Current aetiological diagnostic strategies

A variety of techniques are currently used in attempting to establish an aetiological diagnosis for pneumonia and pleural infection, using a variety of clinical samples.

For pneumonia, only ~20% of patients have a microbiological diagnosis as part of routine clinical practice(114,115). Studies which have analysed the utility of conventional microbiological testing for pneumonia have highlighted their low sensitivity(116). Approximately 30-40% of patients with pleural infection have an aetiological diagnosis established using pleural or blood culture(29).

It is important to examine why establishing an aetiological diagnosis is so problematic;

1. A '**syndromic approach**', based on clinical presentation features, is occasionally advocated to establish a likely aetiology, but this lacks sensitivity and specificity. For example, *S. pneumoniae* pneumonia is

reported to be associated with increasing age, comorbidity, high fever and pleuritic chest pain; *L. pneumophila* pneumonia is stated to be particularly associated with young smokers, with diarrhoea, neurological symptoms and disorders of other systems (e.g. liver function test derangement)(16). Unfortunately, the significant overlap in clinical features between a variety of microbiologically-diverse lower respiratory tract infections prevents such an approach from being helpful in antimicrobial therapy selection(16).

2. **Sputum samples** frequently yield false positive and negative results in pneumonia. Expectorated sputum passes through the upper airways, colonised by many organisms (including those that commonly cause pneumonia). Therefore, culture of sputum is frequently unhelpful(117), yielding bacterial and fungal contaminants. Furthermore, more than 40% of patients with pneumonia are unable to expectorate sputum(6,118). Prior antibiotic therapy may prevent the identification of the true infecting pathogen.
3. **Culture of blood** is occasionally helpful in diagnosing the cause of pneumonia and pleural infection. A positive blood culture (a normally sterile site) is usually a strongly significant finding. However, only 5-14% of blood cultures are positive(4,18-20), making this non-diagnostic for the vast majority of patients.
4. **Immunochromatographic urinary antigen testing** for *Legionella pneumophila* and *Streptococcus pneumoniae* has a relatively high sensitivity >80%. Unfortunately, particularly in children, *S. pneumoniae* is a frequent coloniser of the upper airways, frequently giving false positive

results for patients with pneumonia(119). Furthermore, the antigen test for *L. pneumophila* only diagnoses serogroup 1.

5. **Culture of pleural fluid** for patients with pleural effusion is essential for attempting to establish an aetiology in pleural infection, successfully culturing a pathogen in 30-40% of patients using conventional aerobic and anaerobic culture of fluid transported in a universal container(28). Bedside inoculation of pleural fluid into aerobic and anaerobic 'blood culture bottles' increases sensitivity by 21% when combined with conventional culture(120). Addition of *S. pneumoniae* antigen testing to pleural fluid may increase the sensitivity and specificity beyond testing of urine alone(121-123).
6. **Bronchoscopy** may be carried out to suction bronchial secretions and perform a bronchoalveolar lavage for microbiological analysis. This technique is performed relatively infrequently for pneumonia, for a number of reasons;
    a. It usually requires sedation, which may be hazardous in acutely unwell patients.
    b. The procedure is normally performed in a staffed, dedicated procedure room or operating theatre once or twice per week. Additional procedures beyond these sessions are usually possible, but pose logistical issues.
    c. Samples may also be contaminated by upper airways colonisers, due to the bronchoscope passage through the nose or mouth.
7. **Molecular assays** (Nucleic Acid Amplification Tests (NAATs)) are performed sporadically for select pathogens. A variety of NAATs are

available, usually in reference laboratories, for pathogens including *M. pneumoniae*, *Chlamydophila* spp. and viruses. The performance characteristics of these assays vary, and are critically dependent on the quality of the clinical sample being tested (e.g. nasopharyngeal swab vs. sputum vs. bronchoalveolar lavage)(16). The potential role of NAAT for analysing blood has been examined, but has not been shown to be superior to current methodologies – one assay examined three distinct nucleic acid amplification assays for *S. pneumoniae* in blood, finding that sensitivity was lower than conventional blood culture(124).

8. A variety of **serological assays** have been used to detect specific host immunological response, to enable diagnosis of infection by atypical and viral pathogens. These include complement fixation testing (CFT) and enzyme immunoassays (EIAs). Full interpretation requires acute and convalescent samples (taken 4-6 weeks after the acute illness), and is therefore usually unhelpful in acute diagnosis. Performance characteristics of individual tests vary considerably, and many assays are non-commercial 'in-house' tests(16).

9.4.1 British Thoracic Society recommended investigations

The 2009/2010 BTS guidelines recommend the following investigations for suspected pneumonia and pleural infection;

1. For moderate severity pneumonia (e.g. CURB65 score 2), the BTS recommends blood cultures, sputum culture, pneumococcal urinary antigen

testing, pleural fluid culture and pneumococcal antigen testing (if available) and consideration of urinary legionella antigen testing.

2. For high severity pneumonia (e.g. CURB65 score 3-5), the BTS additionally recommends consideration of investigations for atypical and viral pathogens, including PCR/direct immunofluorescence for *M. pneumoniae*, *Chlamydophila* spp., influenza, parainfluenza, adenovirus, RSV ± *Pneumocystis jirovecii* and also serological testing.
3. For suspected pleural infection, the BTS recommend aerobic and anaerobic culture of pleural fluid with additional inoculation of fluid into aerobic and anaerobic 'blood culture bottles'. Fluid should also have standard biochemical analyses, cytological analyses and a low threshold for mycobacterial investigations.

9.5 Rationale for improving rates of aetiological diagnosis

In view of the poor rates of organism identification in pneumonia and pleural infection, most cases are usually assumed to be caused by bacteria (rather than viruses or fungi) and antibiotic selection is empirical and broad-spectrum to cover most of the standard causative bacteria, given the significant mortality and morbidity associated with these diseases. This strategy has a number of weaknesses;

- Therapy is not targeted to the specific organism causing pneumonia. A proportion of cases will be caused by unexpected organisms, not treated by the empirical antibiotic choice; this usually only becomes evident when the patient deteriorates or fails to improve. As many as 15% of patients with

community-acquired pneumonia may not respond to initial antibiotic therapy, a proportion of these due to failure of empirical antibiotics(125,126). Increased mortality(127) and increased risk of clinical failure(128,129) are more common with inappropriate antibiotic therapy.

- A strategy of multiple empirical antibiotics has a higher likelihood of side effects than narrow, targeted therapy(126). Broad-spectrum antibiotic usage leads to increased rate of antibiotic-resistant healthcare-associated infections (HCAI)(130), such as MRSA, vancomycin-resistant enterococci (VRE), and *Clostridium difficile* associated diarrhoea(131). The UK Department of Health has made reduction of HCAI such as *MRSA* and *C. difficile* a key healthcare priority.
- There are infection control and public health implications associated with certain pathogens, including *L. pneumophila*, *C. psittaci*, *C. burnetii* and multidrug resistant organisms.
- Some pathogens are associated with a particularly high mortality(19); such identification could potentially allow risk stratification and enhanced monitoring of these patients.
- On a population-scale, empirical therapy does not allow evaluation of changing patterns of bacterial infection and resistance, guiding future antibiotic decision.

9.6 Novel diagnostic strategies

Given the imperative above to reduce the reliance on empirical broad spectrum antibiotics for treating lower respiratory tract infection, a variety of clinical and

laboratory techniques may have a potential role in improving rates of pathogen identification.

9.6.1 Ultrasound-guided respiratory sampling

Rapid, complication free acquisition of uncontaminated lower respiratory tract samples is likely to be pivotal in allowing accurate, timely pathogen identification. As discussed, any pleural fluid present in association with pneumonia should be sampled, and this has been traditionally undertaken using clinical examination alone without any direct radiological guidance. However, there has been significant recent increase in the use of bedside ultrasonography by respiratory physicians to identify and sample pleural fluid(132). These increases have been stimulated by technological advances, which have reduced the cost of portable high quality ultrasound machines. Sonography allows rapid real-time radiation-free thoracic imaging (including pleural fluid and consolidated lung), and allows guided sampling, which is safer and more sensitive than clinical examination combined with chest X-ray review(133-135) (an approach which risks organ perforation in 10% of cases(136)).

Given such benefits, thoracic ultrasound is now strongly recommended prior to any procedure for pleural fluid by the recent British Thoracic Society Pleural Disease guidelines(63,137) and a NHS National Patient Safety Agency Rapid Response Report(138). UK respiratory trainees are now required to become competent at thoracic ultrasound during their training program, and trainees in a variety of specialities use ultrasound for other purposes including vascular access and lymph node aspiration.

9.6.2 Lung aspiration

Bedside ultrasound also gives excellent views of consolidated lung, and could enable samples to be taken from such lung for microbiological analysis in pneumonia, using transthoracic needle aspiration (TNA) in a similar manner to that used for sampling possible lung tumours.

Such an approach has significant theoretical advantages, being a rapid, bedside test that avoids the nasooropharyngeal contamination associated with sputum and bronchoscopic sampling. TNA was first reported in 1883 by Leyden, but is not currently a routine investigation in the UK for pneumonia. Despite this, medical literature describes such a procedure being performed in over 6000 children(139,140) and 1000 adults (mostly in the developing world)(141-146) with minimal risks to patients. Furthermore, the stated microbiological yield is often high – a diagnosis is reached in up to 66% of cases in children(140) and around 50% of cases in adults (usually without using direct radiological guidance, and only using standard microbiological techniques)(142-145). Using polymerase chain reaction (PCR) technology, sensitivity has been reported as >90% for *S. pneumoniae*(142). Prior antibiotic therapy reduces yield rate of standard TNA culture from 53% to 23%(146); however, PCR-based techniques may mitigate such antibiotic use.

TNA for pneumonia has largely been performed at the bedside, almost uniquely without direct radiological control; even in these circumstances, the yield is stated to be high with low complication rates. TNA is a rapid procedure in which a needle is passed into the lung percutaneously. Thereafter, a lung aspirate is taken and the needle is removed. Previous studies in children found that patients find it no

worse than a blood test(139). Some investigators use local anaesthetic at the skin and pleura, whereas others undertake the procedure without anaesthetic. Potential discomfort is likely mainly at the skin and the pleura; the lung parenchyma lacks pain receptors. Such a technique is claimed to be well tolerated and associated with a very low rate of side effects or complications. Use of bedside thoracic ultrasound to accurately localise consolidation may further decrease risks associated with the procedure while maximising microbiological return. In particular, the risk of pneumothorax is likely to be substantially reduced by ensuring that the TNA needle only enters consolidated, non-aerated, lung.

9.6.3 Molecular techniques to identify pathogens

There has been recent interest in rapid technologies to diagnose infection-causing organisms within hours, rather than days (usually required for culture techniques). One such technology, widely used in research and increasingly used in clinical practice, is nucleic acid amplification testing (NAAT), such as the polymerase chain reaction (PCR), which identifies pathogens by detecting pathogen nucleic acid in clinical samples. NAATs have been shown to detect respiratory pathogens with high sensitivity and specificity(147,148).

The basic components of PCR include nucleic acid (DNA/RNA) extraction and exponential amplification of a target gene using polymerase enzymes, followed by detection of the amplified gene (amplicon). Single pathogen NAATs have been used to detect one pathogen (usually a clinically highly significant pathogen), but give no further information about other possible infective aetiologies. Recent advances with ‘multiplex’ PCR and other multi-pathogen assays mean that a large

battery of bacteria and viruses may be now identified simultaneously, within a few hours(149). NAATs are often more sensitive than standard culture techniques, and have particular utility for those pathogens which require specialised culture (e.g. atypical pathogens and viruses); they are also useful for those patients who have received previous antibiotic therapy.

Commercially available assays use a variety of molecular techniques, including;

- **Real-time PCR-based assays**. These are much less time- and labour-intensive than the conventional method of detecting amplicons (using gel electrophoresis). Rather, real-time assays are able to simultaneously amplify nucleic acids and detect the product amplicons in a single reaction tube, with high sensitivity using fluorescently-labelled DNA probes. For example, the Cepheid GeneXpert MTB assay can detect *M. tuberculosis* (including a gene target associated with rifampicin resistance) in clinical respiratory samples in less than two hours. This cartridge-based low complexity real-time PCR system has high sensitivity compared with AFB smears and also incorporates nucleic acid extraction within the cartridge(150). Similar Xpert assays for influenza have been developed for influenza A and B, with high (>90%) sensitivity for detection, when compared with non-automated laboratory-developed assays(151,152).
- **Multiplex PCR assays**. These can detect multiple targets in one reaction and have been commercially available for detecting respiratory viruses for some time. Respiratory viruses (e.g. influenza, respiratory syncytial virus (RSV), human metapneumovirus (hMPV), parainfluenza virus (PIV) and adenoviruses) may be detected with sensitivity >90%(153) using the multiplex reverse transcriptase (RT) real-time PCR tests deployed in the

Prodesse commercialised Taqman-based assays from Gen-probe. Samples are processed on a Cepheid SmartCycler real-time instrument using nucleic acids extracted from nasopharyngeal swabs using automated extraction systems (e.g. MagNA Pure). Other commercially-available multiplex assays exist, including the xTAG respiratory virus panel (RVP) (Luminex Molecular Diagnostics), which combines RT-PCR with bead hybridisation on the Luminex xMAP instrument to detect 12 viral targets, with high sensitivity(153).

Recently, the FilmArray system has been developed, combining nucleic acid extraction, PCR detection and data analysis for multiple viral and bacterial targets. Nested PCR is followed by real-time PCR followed by melt-curve analysis to sensitively detect multiple pathogens (15 respiratory viruses, *B. pertussis*, *C. pneumoniae* and *M. pneumoniae*) in a fully automated system(154). This system showed an overall agreement with individual pathogen real-time PCR assays of 99%(155).

Multiplex PCR-electrospray ionization mass spectrometry is an emerging technology designed to identify pathogens directly in clinical samples. Nucleic acids are extracted, followed by either RT-PCR or PCR of viral or bacterial targets using pathogen specific primers. Generated amplicons are injected into an electrospray ionisation time-of-flight mass spectrometer, where the quantity and mass of each amplicon can be detected(156).

Sample type is likely to be critical for interpretation of the results of NAATs. For example, PCR for autolysin and pneumolysin genes in *S. pneumoniae* have a high

sensitivity (>80%) but low specificity (~30%) when analysing *sputum*(157). However, *pleural fluid* pneumolysin gene PCR has sensitivity 78% and specificity 93%(158), emphasising the importance of using contamination-free samples to ensure test specificity. Furthermore, assay design and sample extraction techniques are likely to be key in determining the performance characteristics of each assay.

The strategies described above examine for (multiple) specific pathogens, but provide no information as to other potential pathogens not included in the assay design. Conversely, DNA **sequencing-based** strategies provide a relatively 'assumption-free' strategy in pathogen discovery, which has many advantages in pathogen discovery.

One such strategy uses **targeted amplicon sequencing of the 16S rRNA gene** present in all bacteria(159). PCR is used to amplify and then sequence the 16S rRNA gene, allowing taxonomic assignment of individual organisms and documentation of evolutionary history(160-162). Several large curated databases exist for 16S analyses, including the Ribosomal Database Project (RDP)(163), SILVA(164) and the GreenGenes consortium(165), and these allow taxonomic identification of bacteria present in a clinical sample with high confidence. The clinical utility and validity of 16S rRNA gene sequencing is confirmed in that Public Health England reference laboratories offer such an assay for pathogen identification. 16S rRNA gene sequencing has been previously used for analysing pleural fluid samples (but not lung aspirates), using capillary sequencing(29).

A major disadvantage of capillary sequencing to sequence the 16S gene is that it is only able to reliably identify one pathogen per clinical sample. This inability to resolve multiple pathogens is a particular problem in pleural infection, which is frequently polymicrobial. The 'one sequence per sample' limitation is overcome by next-generation sequencing, in which thousands of individual sequencing reads per clinical sample are achievable. Various researchers have examined the lower respiratory tract's *microbiome* – the collection of microbial genes present within a defined environment – using next-generation sequencing of the 16S gene in clinical samples (sputum and bronchoalveolar lavages) for patients with COPD, cystic fibrosis and asthma(166-176).

There is no current literature examining the use of next-generation sequencing techniques for analysing lower respiratory tract samples from patients with pneumonia, and there is only one small study (18 patients) using these techniques for patients with pleural infection(177).

Other 16S rRNA gene assays also have utility. For example, quantification of this gene in clinical samples (using quantitative real-time PCR) acts as a surrogate for 'total bacterial load' within a given sample. Such uses have been advocated(178), and may help differentiate inflammatory from infectious clinical conditions (e.g. malignant pleural effusions from pleural infection).

9.7 Thesis outline

This thesis describes a novel body of work that combines high quality lower respiratory tract sampling (lung aspirates and pleural fluid samples) with the

development of molecular techniques to further our understanding of the microbiology of lower respiratory tract infection.

Chapter 10 systematically examines the role of the so-called ‘atypical pneumonia’ bacteria (*Mycoplasma pneumoniae, Legionella pneumophila, Chlamydophila pneumoniae, Chlamydophila psittaci* and *Coxiella burnetii*) and the fungus *Pneumocystis jirovecii* (a cause of severe pneumonia, particularly in immunosuppressed patients) in pleural infection using high sensitivity nested PCR and real-time assays. Methods, results and discussion are all included within this chapter.

Chapter 11 describes the methods used for the subsequent studies presented in chapters 12 and 13, in which 16S rRNA gene ultra-deep sequencing is used to define the bacteriology of lower respiratory tract samples.

Chapter 12 uses ultra-deep sequencing to examine the bacteriology of pleural infection using samples acquired as part of the MIST2 pleural infection randomised controlled trial. The results from this study confirm the bacteriological results of Chapter 10.

Finally, chapter 13 describes a further study that evaluates the applicability and acceptability of bedside ultrasound-guided lung aspiration in pneumonia within an NHS setting, and analyses obtained samples using 16S rRNA gene ultra-deep sequencing. Additional pleural fluid samples were also obtained, including further pleural infection samples and a variety of ‘control’ pleural fluid samples – these are used to assess the sensitivity and specificity of the ultra-deep sequencing technique. Results from this chapter are consistent with chapters 10 and 12.

10 The role of 'atypical pneumonia' pathogens and *Pneumocystis jirovecii* in pleural infection

10.1 Introduction

As discussed in chapter 9 (and further in chapter 12), 40-60% of patients with pleural infection have a causative pathogen identified in the pleural space, commonly Streptococcus 'milleri' group bacteria, *Streptococcus pneumoniae*, *Staphylococcus aureus* and anaerobic bacteria. Consequently, about 40% of patients with pleural infection have no pathogen identified by standard culture or gram staining techniques.

The negative microbiology may be due to antibiotic therapy prior to pleural fluid sampling, varying bacterial prevalence in different pleural fluid locules (which are known to vary biochemically and possibly, therefore, microbiologically) or the presence of organisms that are difficult to culture or detect using conventional techniques.

So-called 'atypical pneumonia' bacteria (such as *Mycoplasma pneumoniae, Legionella pneumophila, Chlamydophila pneumoniae, Chlamydophila psittaci and Coxiella burnetii*) and the fungus *Pneumocystis jirovecii* (a cause of severe pneumonia, particularly in immunosuppressed patients) are all causes of pneumonia, which are particularly difficult or impossible to culture, and may explain a proportion of the culture-negative cases of pleural infection. This chapter explores the potential role of these organisms in pleural infection, investigating firstly *P. jirovecii,* followed by the 'atypical pneumonia' bacteria.

10.2 Role of *Pneumocystis jirovecii* in pleural infection

10.2.1 Introduction

Pneumocystis jirovecii (previously known as *Pneumocystis carinii* forma specialis *hominis*)(179), is a fungal cause of pneumonia, particularly in immunosuppressed individuals. Recent studies have also found evidence of a high prevalence of *P. jirovecii* in asymptomatic *non*-immunosuppressed individuals, both in oropharyngeal washes and bronchoalveolar lavage (BAL) specimens – it has been isolated from BAL fluid using nucleic acid amplification techniques in 18% of patients with lung disease (but without HIV) undergoing bronchoscopy(180), in BAL fluid from 4.4% of general medical patients with community-acquired bacterial pneumonia(181), and in the oropharyngeal washes of 20% of a healthy population(182).

Identification of *P. jirovecii* requires specific diagnostic techniques (e.g. Grocott-Gomori methenamine silver staining or nucleic acid amplification techniques, such as polymerase chain reaction (PCR)). PCR assays for detecting *P. jirovecii* have a higher sensitivity and specificity than conventional cytochemical staining(183-185).

Pleural effusion occurs in 2 to 15% of patients with *Pneumocystis* pneumonia(186-189) and, to date, *P. jirovecii* has only been found in pleural fluid from the immunocompromised: Several case reports describe detection of *P. jirovecii* within pleural fluid of HIV-infected individuals using standard cytological and histological techniques(190-195).

There has been no previous systematic examination for *P. jirovecii* in the pleural fluid of non-immunosuppressed patients presenting with pleural infection. Given

the prevalence of *P. jirovecii* in chronic lung disease and asymptomatic healthy people, it was hypothesised that it might be a passenger or co-pathogen in infected pleural fluid. This study assessed the prevalence of *P. jirovecii* in the pleural fluid of patients with established pleural infection, using a highly sensitive quantitative polymerase chain reaction targeting the *P. jirovecii* heat shock protein 70 (HSP70) gene sequence.

10.2.2 Methods

133 samples of pleural fluid from 126 adult patients with pleural infection were analysed. Patients were recruited in 52 UK centres between 1999 and 2002, as part of a previous randomised trial of intrapleural therapy in pleural infection(28). Patients had clinical evidence of infection and pleural fluid that was macroscopically purulent, positive on culture for infection, positive for pathogens on Gram staining, or with pH <7.2.

Microbiological information included results of pleural fluid Gram stain and aerobic and anaerobic pleural fluid culture, performed in the recruiting centre. Pleural fluid was collected and transferred to the coordinating center and frozen at -70°C. Samples were frozen within 48 h of being taken. The trial was approved by the Anglia and Oxford Multicentre Research Ethics Committee (MREC) (ref: 98/5/61). All participants gave informed, written consent.

10.2.2.1 Nucleic Acid amplification and assessment of inhibition

All molecular analyses were performed in accordance with the *M*inimum *I*nformation for Publication of *Q*uantitative Real-Time PCR *E*xperiments (MIQE) guidelines(196).

10.2.2.2 DNA extraction

Samples were thawed and DNA was extracted using Qiagen QIAamp Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The extracted DNA was then eluted in 100 µl of buffer(29). All extracts were diluted 1:10 using Tris-EDTA (Sigma, UK) prior to analysis by quantitative PCR (qPCR).

10.2.2.3 Molecular assessment of *P. jirovecii* DNA in pleural fluid

A highly sensitive probe-based qPCR technique was used, which targets the *P. jirovecii* Heat Shock Protein 70 (HSP70) gene to detect and quantify the presence of *P. jirovecii* DNA, as previously described(185). Both positive controls (BAL fluid samples from patients with *Pneumocystis* pneumonia confirmed by identification of cystic forms of *P. jirovecii* by Grocott-Gomori methenamine silver staining and response to specific anti-*Pneumocystis* therapy) and negative controls (ultrapure DNase/RNase-free water, Sigma) were included.

10.2.2.4 Inhibition analysis

Inhibition was assessed separately by spiking qPCR reactions with 1000 copies per reaction of linearised HSP70 *P. jirovecii* plasmid. Inhibition was assessed by comparing the effect of PCR of the sample with the same reaction in which water was added instead of sample. Any increase (shift to the right) in quantification cycle (C_q) as a result of adding the sample was due to sample inhibition.

10.2.3 Results

10.2.3.1 Participants

Clinical and laboratory characteristics of the participants with pleural infection are shown in table 10-1.

Characteristic	**Statistic**
Age, yr, median (IQR)	56.0 (38.7-71.4)
Male, n (%)	89 (71)
Duration of symptoms prior to presentation, median (IQR)	14 (7-28)
Co-morbidity, n (%)	86 (68)
Chronic respiratory disease	15 (12)
Excess alcohol consumption	17 (14)
Diabetes mellitus	13 (10)
Patients with neutropenia on admission blood tests	0 (0)
Pleural fluid characteristics	
Visibly purulent, n (%)	104 (83)
Positive standard microbiology, n (%)	69 (55)
Streptococcus pneumoniae	12
Streptococcus 'milleri' group	13
Other streptococci	5
Staphylococcus aureus	9
Anaerobic or mixed aerobic/anaerobic infection	16
Mixed aerobic bacteria	8
Gram negative bacteria	6
pH in patients without frankly purulent fluid, median (IQR)	6.9 (6.7-7.1)
Glucose (mg/dL), median (IQR)	18 (11-61)

LDH (IU/L), median (IQR)	6000 (1629-20000)
Patient laboratory characteristics	
Total white cell count (x 10^9/L), median (IQR)	14.7 (10.3-22.0)
C-reactive protein (mg/L), median (IQR)	187 (83-271)
Albumin (g/L), median (IQR)	27 (22-31)

Table 10-1 – Characteristics of participants (n = 126)

10.2.3.2 Pleural fluid PCR

There was no evidence of *P. jirovecii* DNA in any of the 133 pleural fluid samples analyzed from 126 patients with pleural infection using the qPCR HSP70 assay. Positive and negative controls performed as expected.

No inhibition of PCR was identified in 131 samples; amplification of spiked 1000 copies of *P. jirovecii* HSP70 was within 0 - 1.5 C_q of the spiked water controls. Two pleural fluid samples showed evidence of inhibition of the qPCR reaction; a 2.71 increase in C_q in one patient and a 4.68 increase in C_q in the other.

10.2.4 Discussion

This study found no *P. jirovecii* in pleural fluid samples from 126 patients with pleural infection. This is an important and interesting negative finding, since among healthy individuals *P. jirovecii* has been identified in oropharyngeal washes of 20%(182), in 18% of BAL samples(180), is present in 4.4% of those with bacterial pneumonia(181) and in case reports has been identified in pleural fluid of severely immunocompromised patients(190-195). By contrast, this study found no evidence of *P. jirovecii* in the pleural space using highly sensitive nucleic acid amplification techniques in immunocompetent adults with pleural infection. These

data suggest there is no need to investigate for *P. jirovecii,* nor for its empiric treatment, in patients with pleural infection who are not significantly immunosuppressed.

Unlike many similar molecular analyses this study also investigated the influence of co-purified inhibitors on the PCR reaction, this is essential for accurate assessment of the specific nucleic acid within the sample(196). The same molecular target was used to conduct this inhibition analysis, in light of recent findings that inhibition of PCR is target specific(197). Our strategy was to dilute the sample in order to compensate for inhibition, which will also dilute the template DNA, however, as dilutions were not major (1:10 sample to water dilution) this did not likely impact on the sensitive qPCR detection of *P. jirovecii* DNA.

The striking absence of *P. jirovecii* in the pleural space, despite its prevalence in the lung parenchyma which is separated by only the visceral pleura, is particularly interesting. This may be due to its tropism for the lung, where it exists primarily as an alveolar pathogen, usually without host invasion(198). The trophic forms of *P. jirovecii* are strongly adherent to the type I alveolar cells, as a result of cell membrane interdigitation(199). The glycoprotein A (major surface glycoprotein) molecule has an integral role in the attachment of *Pneumocystis* to host alveolar cells(179). Such an attachment to alveolar cells is thought to be a requirement for proliferation(200), probably via signalling pathways, including the PCSTE20 kinase gene, shown to be involved in fungal mating and proliferation(201). Sustained experimental propagation outside the host lung has been very difficult to achieve(202) and perhaps the avidity of *P. jirovecii* for alveolar cells makes it unable to reproduce in the pleural space, unless in the setting of overwhelming immunosuppression.

The acidic pleural environment of pleural infection is not the likely explanation for the absence of *P. jirovecii* in the fluid, since *Pneumocystis* proliferates more rapidly at pH 4.0 than 7.0(203) and the organism possesses the *PCA1* gene, which encodes an H^+-ATPase that, in yeasts, allow growth at a low pH(204). Limited cell surface binding capacity to mesothelial cells may prevent *P. jirovecii* entering the pleural space at all. Adhesion to pleural mesothelial cells is pivotal for penetration of the mesothelial cell monolayer prior to entering the pleural space. Lack of appropriate cell surface ligands on the surface of *P. jirovecii* would make the visceral pleura a critical barrier, preventing entry of *P. jirovecii* into the pleural space.

10.3 Role of ‘atypical pneumonia’ pathogens in pleural infection

10.3.1 Introduction

Mycoplasma pneumoniae, Legionella pneumophila, Chlamydophila pneumoniae, Chlamydophila psittaci and Coxiella burnetii are common causes of pneumonia (both sporadic and epidemic), stated to account for up to 20-40% of cases(16,205). These organisms are often described as ‘atypical’ pathogens due to difficulties in their identification and lack of sensitivity to beta-lactams.

Atypical pathogens are widely distributed in the environment. In addition to causing pneumonia, these and related pathogens have been identified at other sites, including the pleural space(103-108), pericardium(206,207), heart valves(208), liver, aneurysms, joint space(209-211) and blood.

Parapneumonic effusions accompany atypical pneumonia in up to 60% of cases(212-216) although it is unclear how many of these effusions represent pleural infection (‘complicated’ parapneumonic effusions and empyema) rather than reactive uninfected effusions given that standard culture techniques fail to isolate such pathogens.

Routine pleural fluid investigations include aerobic and anaerobic culture. However, *Legionella* spp. & *Mycoplasma* spp. are nutritionally fastidious and require prolonged incubation on specialist media. *Chlamydophila* spp. & *Coxiella burnetii* are obligate intracellular organisms that require cell culture techniques to grow in the laboratory. None of these techniques are routinely used in laboratory pleural fluid testing. Other atypical diagnostic tests are available including serological antibody tests and urinary antigen tests, but these fail to localise

pathogens to the pleural space and have several limitations. Commercial urinary antigen enzyme immunoassays are only available for *L. pneumophila* serogroup 1. Retrospective diagnoses may be made using serology studies, but these require peak and convalescent samples and up to 25% of patients fail to demonstrate diagnostic titres(217).

Given the significant mortality and morbidity associated with pleural infection(28,75), an understanding of the role of atypical pathogens in pleural infection is essential to allow clinicians to make a rational choice of empiric antibiotics and to determine whether routine atypical laboratory investigations are warranted. A highly sensitive nested PCR-based study was undertaken to estimate the prevalence of atypical pathogens in pleural infection, either as a sole pathogen or in polymicrobial infection. The organism-specific PCR probes were complementary to regions of the 16S ribosomal RNA (rRNA) gene. As further discussed in chapter 11, the 16S rRNA gene is common to all bacteria, having both conserved and hypervariable regions unique to each species, which allows for primers to be designed having specificity for particular organisms, thereby allowing for pathogen identification.

10.3.2 Methods

10.3.2.1 Patients

Pleural fluids from 374 patients meeting standard criteria for pleural infection were analysed. Patients were recruited in 52 UK centres between 1999 and 2008, as

part of previous randomised trials of intrapleural therapy in pleural infection(28,75). Pleural fluid was stored at -80°C prior to analysis.

Inclusion criteria were clinical evidence of infection and pleural fluid that was macroscopically purulent, positive on culture for bacterial infection, or positive for bacteria on Gram's staining, or pleural fluid that had a pH of less than 7.2 (measured using a blood-gas analyser). Evidence of infection, which was assessed by the recruiting physician, included the presence of fever and elevated serum levels of inflammatory markers such as C-reactive protein or an elevated white-cell count.

Baseline microbiological information included the results of pleural fluid Gram stain, aerobic and anaerobic pleural fluid culture.

All participants gave informed, written consent and pleural fluid collection was approved by the Multicentre Research Ethics Committees (98/5/61 & 04/MRE05/53).

10.3.2.2 Nucleic Acid Amplification Techniques

10.3.2.2.1 Pleural fluid DNA extraction

Pleural fluids were thawed and DNA was extracted from 200 µl of fluid using the FastDNA SPIN Kit technique (MP Biomedicals, Solon, OH), following the manufacturer's instructions. This technique uses a bead-based cell lysis technique combined with a silica column-based DNA purification process. Two 40 s homogenisation steps at speed 6.0 were used and DNA was eluted in 100 µl of kit-provided buffer. DNA was quantified using a NanoDrop 1000

spectrophotometer (Thermo Fisher Scientific, Wilmington, USA). The extracted DNA was either amplified immediately or stored at -20°C for subsequent analysis.

10.3.2.2.2 Nested polymerase chain reaction (PCR)

Nested PCRs were used to detect atypical pathogens. The first-round PCRs used primers complementary to sequences of the 16S gene which are widely conserved in bacteria, surrounding hypervariable regions V3-V8 (~1070 bp). The second-round PCRs used primers targeting genus specific 16S sequences.

10.3.2.2.3 First-round PCR

First round primers, Nested16S_F and Nested16S_R, are detailed in table 10-2. The PCR reaction mixtures consisted of 2.5 µl 10x PCR buffer, final concentrations of 3.0 mM $MgCl_2$, 0.2 mM dNTPs (Applied Biosystems, California, USA), 0.4 µM each primer (Eurofins MWG Operon, Ebersberg, Germany) and 0.625 units per reaction HotStarTaq polymerase (Qiagen, Hilden, Germany). Molecular grade water and 100 ng pleural fluid DNA were added to give a final volume of 25 µl. The cycling conditions were as follows: 95°C for 5 mins followed by 30 cycles of 95°C for 30 s, 64°C for 30 s, and 72°C for 1.5 mins, with a final extension of 72°C for 5 mins (Techne TC-512 Thermal Cycler; Bibby Scientific, Staffordshire, UK).

The size and presence of amplicons were established using electrophoresis on 1% agarose gel alongside a 100 bp ladder (Promega, Madison, USA), with ethidium bromide and ultraviolet light visualisation. When no amplicon was

detected (suggesting possible inhibition), the first-round PCR was repeated using 1:10 diluted template. For the samples in which this strategy failed to produce a detectable amplicon, a bead-based human DNA depletion step with mass amplification was undertaken as previously described(218), followed by repeated nested PCR.

10.3.2.2.4 Second-round PCR

The products of the first-round PCR reactions were diluted 1:100 in molecular grade water and 1 µl was used as template in multiple second-round PCRs using primer pairs specific for *Mycoplasma* spp., *Legionella* spp., *Chlamydophila* spp. and *Coxiella burnetii* (table 10-2). The same cycling conditions as for the first-round PCR reactions were used, except for a 60°C annealing step. Presence of amplicons was established as before.

10.3.2.2.5 Sensitivity analysis

The sensitivity of the nested PCR was estimated by spiking culture- and PCR-negative pleural fluid samples with a dilution series of atypical pathogens. Tenfold dilutions of *Chlamydophila pneumoniae* (ATCC VR-1360), *Chlamydophila psitacci* (ATCC VR-125), *Mycoplasma pneumoniae* (ATCC 15531), *Coxiella burnetii* (ATCC VR-616) and *Legionella pneumophila* (NCTC 12821) were spiked into extracted negative pleural fluids. First round PCR reactions were performed with ten-fold dilutions, starting at 10^5 genome copies of each organism, and second round reactions were performed as previously described.

10.3.2.2.6 Sequencing

Pleural fluid samples which generated an amplicon during second-round PCR underwent capillary sequencing to verify species identity. Extracted DNA underwent a separate PCR, using the same PCR conditions as previously described, except for a 58°C annealing step and using 16S gene sequencing primers (Sequencing16S_F & Sequencing16S_R, table 10-2) that amplified a 625 bp product, covering hypervariable regions V5-V8.

Primer name	Primer sequence (5'-xxx-3')	Product size (bp)
Nested16S_F	CTCCTACGGGAGGCWGCAGT	~1070
Nested16S_R	TGACGGGCGGTGTGTACAA	
Chlamydophila_F	TCCAATCGATTTGAGCGTAC	~347
Chlamydophila_R	TTACGGCAAGGACTACCAGG	
Coxiella_F	TCTCAAGGGTAATATCCTTGG	~337
Coxiella_R	CTAATCCTGTTTGCTCCCCA	
Legionella_F	GGTTGATAGGTTAAGAGCTGAT	~427
Legionella_R	TCAACTTATCGCGTTTGCTG	
Mycoplasma_F	TACATAGGTCGCAAGCGTTA	~257
Mycoplasma_R	CTAATCCTATTTGCTCCCCA	
Sequencing16S_F	AGGATTAGAKACCCTGGTA	~625
Sequencing16S_R	TGACGGGCGGTGTGTACAAG	

Table 10-2 – Primer sequences and expected product sizes

Subsequently, the amplicons were purified using polyethylene glycol-based precipitation. Sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, California, USA) using the manufacturer's guidelines and products were cleaned using ethanol precipitation. Sequences were analysed on a 3730 DNA Analyser (Applied Biosystems, California, USA). Sequencing was undertaken in both the forward and reverse direction, and the sequences aligned to ensure base pair accuracy using the Staden package(219). Bacterial identification was performed by comparison of the

16S rRNA gene sequences with the Ribosomal Database Project. Reference sequences, stated to be of good quality and ≥1200 bp length, were used. A positive identification was recorded when the sequence matched a database entry and other matches showed significantly less homology(163).

10.3.3 Results

10.3.3.1 Patient and pleural fluid characteristics

The demographic, clinical and microbiologic characteristics of the patients and their pleural fluid samples are shown in table 10-3. 34/373 (9%) of patients consumed excess alcohol, 51/373 (14%) were diabetic, 18/373 (5%) were immunosuppressed and 75/373 (20%) had a chronic respiratory condition. 89% of patients acquired their pleural infection in the community. 41% of pleural fluid samples were positive on aerobic or anaerobic culture.

Characteristic	**Statistic**
Age (years) – Median (IQR)	61 (47-74)
Male sex – % (no.)	70 (262/374)
Duration of symptoms before enrolment (days) – Median (IQR)	14 (7-28)
Community-acquired infection – % (no.)	89 (326/368)
Purulent pleural fluid – % (no.)	69 (253/369)
Positive pleural fluid Gram's stain – % (no.)	14 (43/298)
Positive pleural fluid culture – % (no.)	41 (153/374)
Pleural fluid pH – Median (IQR)	6.9 (6.8-7.1)
Pleural fluid glucose (mmol/L) – Median (IQR)	1.0 (0.6-3.5)
Pleural fluid lactate dehydrogenase (IU/L) – Median (IQR)	3830 (1100-14062)
Serum C-reactive protein (mg/L) – Median (IQR)	160 (115-245)
White cell count ($x10^9$/L) – Median (IQR)	14.1 (10.5-19.9)

Table 10-3 – Patient characteristics. IQR = interquartile range

Table 10-4 summarizes the bacteriology of these samples, with streptococci (S. 'milleri' group and *S. pneumoniae*), *S. aureus*, Enterobacteriaceae and obligate anaerobes being the most frequently identified using standard culture techniques.

Organism	**Frequency of isolation in positive samples***
Streptococcus 'milleri' group	20% (30/153)
Streptococcus pneumoniae	14% (22/153)
Pyogenic group streptococci	4% (6/153)
Enterococcus spp.	4% (6/153)
Other/undifferentiated streptococci	6% (9/153)
Staphylococcus aureus	16% (25/153)
Other gram positives	5% (7/153)
Enterobacteriaceae	18% (27/153)
Pseudomonads	3% (5/153)
Other gram negatives	4% (6/153)
Obligate anaerobes	14% (22/153)
Mycobacterium tuberculosis	3% (4/153)
Fungal	1% (1/153)

Table 10-4 – Organisms identified in culture-positive pleural fluid samples

*Total greater than 100% due to the polymicrobiality of some samples

10.3.3.2 Nested PCR sensitivity analysis

Nested PCR assay sensitivity, evaluated using atypical pathogens spiked into sterile pleural fluids, is shown in table 10-5. A minimum of 10-100 genome copies of each pathogen could be reliably detected per 25 µl first round PCR reaction.

Organism	**Sensitivity (genome copies detectable by nested PCR)**
Legionella pneumophila, NCTC 12821	≥10-≤100
Coxiella burnetii, ATCC VR-616	≥10-≤100
Mycoplasma pneumoniae, ATCC 15531	≤10
Chlamydophila pneumoniae, ATCC VR-1360	≥10-≤100
Chlamydophila psittaci, ATCC VR-125	≥10-≤100

Table 10-5 – Estimated limits of detection of nested PCR assays

10.3.3.3 Pleural fluid testing

Of 374 pleural fluid samples analysed, 201 samples yielded a ~1070bp product after first-round PCR, confirming amplification of the 16S gene. 173 samples were initially negative on first-round PCR, and were subject to repeated PCR after 1:10 dilution and, if still negative, human DNA depletion with subsequent mass amplification. This process allowed the detection of a 16S product in a further 131 samples. The 16S gene was not amplified in 42 samples.

All *Legionella* and *Chlamydophila* spp. assays were negative. Nested PCR of two pleural fluid samples produced a faint band on agarose gel with *Coxiella* spp. primers; both of these samples were negative on standard culture. 16S gene sequencing revealed *Pseudomonas* spp. in one sample and *Prevotella* spp. in the other. Both samples were retested and found to be negative using published *Coxiella burnetii* real time(220) and standard PCR(221) assays, suggesting a false positive result.

Two other samples produced an amplicon with the second-round *Mycoplasma* spp. primers. Species identity was confirmed by sequencing the 16S gene and comparing the sequences with those held by the Ribosomal Database Project. The first sample (labelled as sample 'A' in figure 10-1), had 100% similarity with the published sequence for *Mycoplasma salivarium* (NCBI accession AF125583), and a one base-pair difference to *Mycoplasma arthritidis* (accession CP001047). Sample 'B' did not have a perfect match, but had a one base-pair difference to both *Mycoplasma salivarium* and *Mycoplasma arthritidis*. These pleural fluid samples had been extracted, amplified and sequenced on separate days.

Previous bacterial culture of these samples revealed *Bacteroides* spp. in sample A and MRSA with *Enterococcus* spp. in sample B.

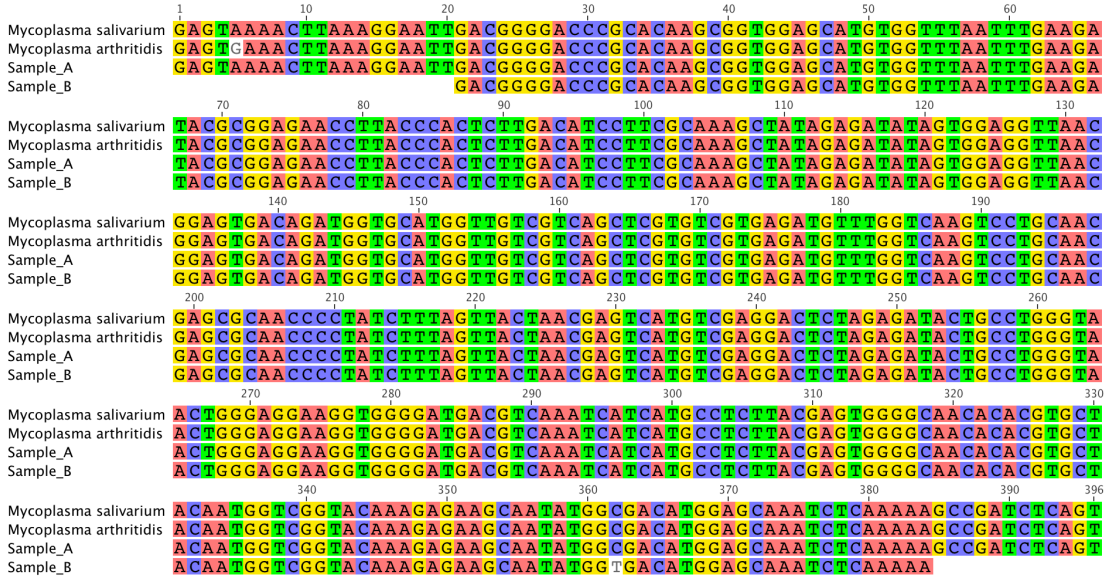


Figure 10-1 – Comparison of sequences of *M. salivarium and M. arthritidis*, with sequences from two study samples (Samples 'A' and 'B')

Using a conservative approach of including *Mycoplasma salivarium* and *arthritidis* within the definition of atypical pathogens, therefore gives an atypical prevalence estimate of 0.53% (95% confidence interval 0.06-1.92%).

10.3.4 Discussion

This study has established a very low prevalence of atypical pathogens in pleural infection and helps inform antibiotic choices for the 65,000 people who develop this infection each year in the UK and US(75). It is the first study to robustly examine for atypical pathogens in a large generalizable cohort. Previously, the

role of atypical pathogens in pleural infection was unknown, given difficulties isolating such pathogens as part of routine pleural fluid analyses.

The patients studied were recruited on the basis of standard clinical criteria used for the diagnosis of pleural infection(28,75) from 52 centres in the United Kingdom, including both teaching and community hospitals, over several years (including mycoplasma peak periods(222)), and are likely to be representative. The techniques employed included nested PCR, maximising sensitivity, and confirmatory DNA sequencing to ensure specificity.

Our findings lend further weight to the hypothesis that the bacteriology of pleural infection may differ from that of pneumonia. It is often assumed that pleural infection occurs secondary to bacterial migration across the visceral pleural from pneumonic lung. Given the paucity of atypical pathogens found in pleural infection, despite a 20-40% rate in pneumonia, this adds further evidence that the pathology of pleural infection is not so straightforward and the bacteriology (and antibiotic choice) of each disease should be viewed distinctly.

10.3.4.1 Detection of atypicals

Routine clinical laboratory isolation of fastidious atypical organisms is difficult, and not part of standard pleural fluid investigations. Mycoplasma culture requires media enriched with precursors for nucleic acids, proteins and lipids. *Legionella* spp. require L-cysteine and iron salts for growth (contained in buffered charcoal yeast extract agar supplemented with alpha-ketoglutarate). Being obligately

intracellular organisms, *Chlamydophila* and *Coxiella* spp. require complex and time-consuming cell culture techniques for growth.

Our approach to pathogen identification relies on the characteristics of the bacterial 16S rRNA gene. This gene has highly conserved regions, with sequences common to nearly all bacteria, and hypervariable regions, with sequences specific to each genus/species. 16S sequence analysis is increasingly used in clinical practice for bacterial identification(223).

A nested PCR format increases detection sensitivity. Our first round PCR primers targeted highly conserved 16S sequences and second round primers targeted genus specific 16S sequences, thereby conferring specificity. With the exception of *C. burnetii* (the only member of the Coxiella genus), our primers gave resolution to the genus-level, allowing maximal opportunity to detect related species. Species-level identification (and result validation) was possible by confirmatory 16S gene sequencing of positive pleural fluid samples.

Another advantage of nucleic acid amplification techniques is that prior antibiotic therapy, which frequently hinders culture, is unlikely to affect pathogen detection. However, pleural fluid is rich in non-target human DNA and other substances such as haemoglobin and immunoglobulins which can cause inhibition of PCR and reduce sensitivity. Our nested format allowed inhibition to be assessed by the first round PCR. Where necessary, template dilution and human DNA depletion strategies were used to mitigate such inhibition.

10.3.4.2 Implications for antibiotic selection

Pleural infection is treated with empiric broad-spectrum antibiotics to cover likely pathogens, given that 40-60% of pleural infection is culture-negative. Whilst practice varies geographically, many physicians would treat community-acquired pleural infection with a beta-lactam antibiotic in conjunction with a beta-lactamase inhibitor, such as co-amoxiclav (amoxicillin & clavulinic acid) or piperacillin-tazobactam. Metronidazole is usually additionally given to ensure adequate anaerobic coverage. Being intracellular organisms, atypicals require specific antibiotics, such as macrolides or fluoroquinolones, which are concentrated within cells to ensure adequate treatment. Our data demonstrate that such atypical coverage is not routinely required for the empirical treatment of pleural infection.

10.3.4.3 Mycoplasma salivarium

Two samples were identified containing a non-pneumoniae mycoplasma species. Sequence analysis was identical to *Mycoplasma salivarium* for one sample, and had equal similarity with both *M. salivarium* and *M. arthritidis* in the other. *M. salivarium* has been isolated in the oropharynx and nasopharynx, particularly within the gingival crevices, and may play a role in periodontal disease(224). Its DNA has been isolated in synovial fluid of patients with arthritis(225), and has also been found in polymicrobial brain abscesses (along with *Streptococus* spp., *Staphylococcus epidermidis*, *Prevotella* spp., *Bacteroides* spp., *Clostridium* spp. and *Peptoniphilus asaccharolyticus*)(226). *M. arthritidis* causes acute and chronic arthritis in rats and mice, has not previously been demonstrated to cause human disease, and seems unlikely to be the implicated pathogen in the second case.

There has only been one previous case report of *M. salivarium* in pleural fluid(227). A retrospective case review was therefore undertaken of our two patients to identify any unusual characteristics. The first patient was under 25 years old and had community acquired pleural infection. Dental caries was denied. The pleural fluid was purulent, cultured *Bacteroides* spp. and had a particularly high LDH (46000 IU/L). Recovery from pleural infection was uncomplicated, without the need for surgery, and there were no further hospital admissions during the next year. The second patient had recently undergone a thoracotomy as part of pharyngo-oesophageal surgery. The pleural fluid was purulent and grew methicillin-resistant *Staphylococcus aureus* and enterococci.

Identification of *M. salivarium* in the second patient is perhaps unsurprising given the history of recent surgery, and that the organism has been found in the oropharynx. Detection in the first case is unexpected, given the lack of complicating features. Further studies will be required to establish whether this organism represents a significant pleural pathogen or, perhaps more likely, a bystander organism in polymicrobial infection.

10.4 Conclusions

These two studies represent the first systematic assessment for ‘atypical pneumonia’ pathogens and *P. jirovecii* in the pleural space using highly sensitive nucleic acid amplification techniques. The multicentre nature of both studies, together with the use of conventionally accepted criteria for defining pleural infection, ensures that the patient groups are representative of clinical practice; in these patients an extremely low prevalence of such organisms was found. This is

in contrast to the relative frequency with which they are detected in the lung parenchyma in patients with pneumonia (and also in health, for *P. jirovecii*).

'Atypical pneumonia' pathogens are implicated in 20-40% of cases of pneumonia, and yet there was no convincing evidence of these bacteria (finding only two cases of *M. salivarium*, a mouth commensal). *P. jirovecii* is increasingly implicated as a commensal organism in the upper and lower respiratory tract in health, and studies have detected it in non-immunosuppressed patients with pneumonia. This study found no evidence of *P. jirovecii* in the pleural space.

We can conclude that such pathogens are not an explanation for the ~25-40% of cases of pleural infection where a causative pathogen cannot be identified (even after nucleic acid amplification testing), and this suggests that these organisms either do not cross the visceral pleural membrane or survive in the pleural space in immunocompetent individuals

Further, these findings demonstrate that there is no need to perform routine specialist investigations for *Legionella pneumophila*, *Mycoplasma pneumoniae*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci* and *Coxiella burnetii* or *Pneumocystis jirovecii* in pleural infection unless a patient is severely immunocompromised or there are specific features suggesting the possibility of such an infection.

From a treatment perspective, pleural infection is usually treated with empirical antibiotics. These studies provide evidence that there is no need to ensure that the empirical antibiotic choice includes coverage for these pathogens.

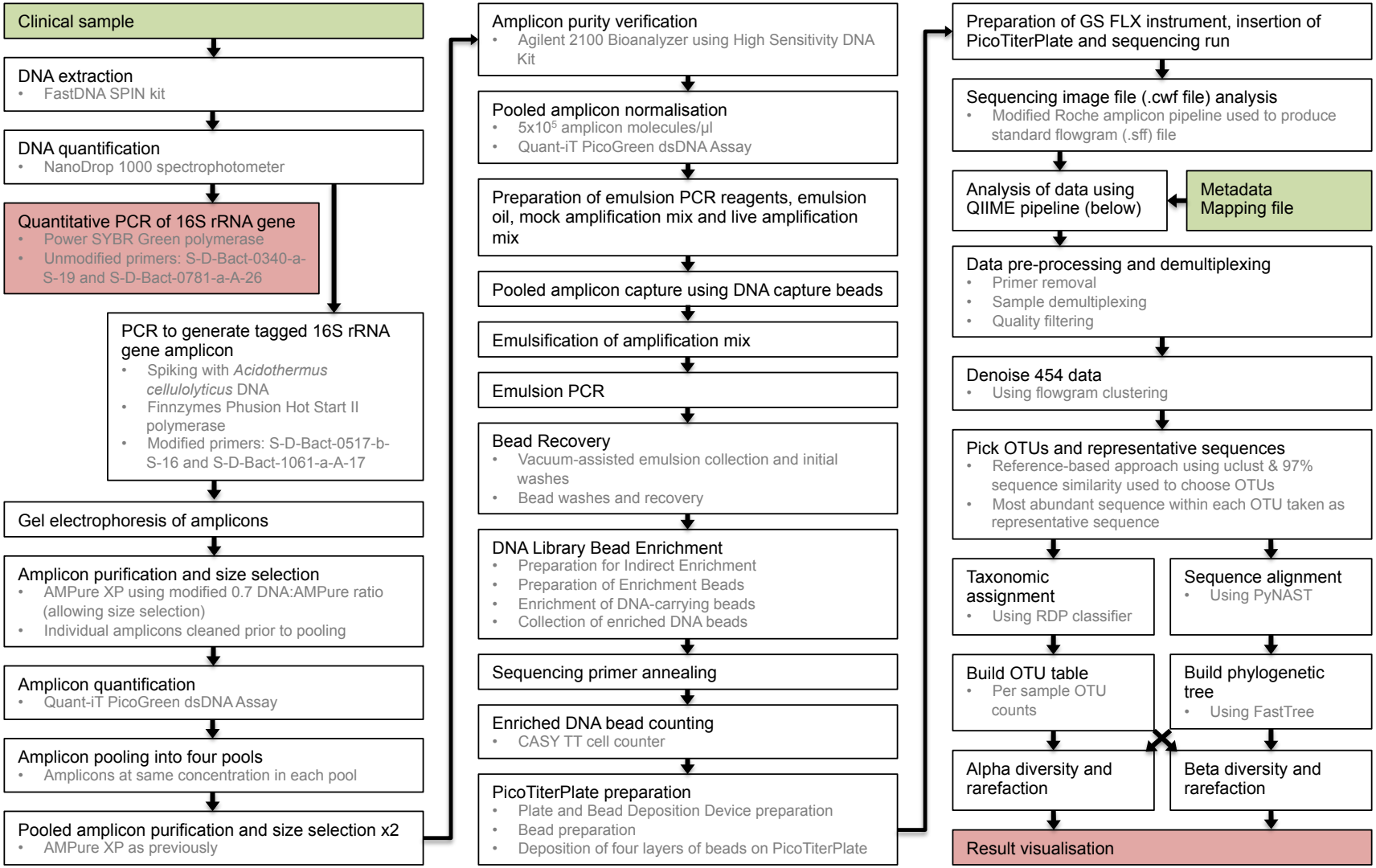


Figure 11-1 – Overview of processes involved in ultra-deep sequencing the 16S rRNA gene. Primer nomenclature throughout this section is consistent with Alm(228) and Klindworth(229)

11 Ultra-deep sequencing of 16S rRNA gene - Methods

11.1 Introduction

Chapter 10 discusses two studies in which nucleic acid amplification techniques were used to detect pathogens in clinical samples, using highly specific primers. A key advantage of this methodology is its sensitivity and specificity, but the experimental design requires pre-selection of the pathogens of interest to enable primer design. Any other organisms present in a clinical sample will not be detected.

This, and subsequent, chapters use ultra-deep sequencing techniques designed to detect all bacteria present within clinical samples (the 'microbiota') using a relatively 'assumption-free' methodology. This chapter further discusses these techniques.

11.1.1 Analysis of microbiota and microbiome in clinical samples

A *microbiota* is the collection of organisms present within a given environment. A variety of strategies can be used to investigate a clinical sample's microbiota, including microscopy, culture, MALDI-TOF (matrix-assisted laser desorption/ionisation-time of flight) mass spectrometry and nucleic acid based techniques. Nucleic acid sequencing techniques investigate a microbiota by examining the sample's *microbiome* – the collection of genes present within the microbiota.

The current diagnostic 'gold-standard' for bacterial identification – bacterial culture – is far from perfect, being affected by the type of culture techniques employed

(e.g. growth media used, anaerobic/aerobic conditions) and concomitant antimicrobial therapy used. Furthermore, some microorganisms have yet to be characterised by culture techniques or are currently unculturable, potentially biasing culture-based results. When one includes environmental species, the vast majority of nature's microbes are yet to be identified using culture(230).

Nucleic acid amplification techniques vary considerably in the range of microorganisms detectable, dependent on assay design. For example, the use of highly specific primer pairs will detect the presence or absence of a particular microorganism, but give no further information about other microorganisms present within a sample. Conversely, other techniques can be used to evaluate all microorganisms present within a sample in an 'assumption-free' manner, both utilising next-generation sequencing: (1) targeted amplicon (usually 16S rDNA amplicon) sequencing and (2) shotgun metagenomic sequencing. Metagenomic approaches, in which DNA is fragmented and then randomly sequenced, provide rich functional data based on genes present within a sample, but provide a relatively low taxonomic resolution of the microorganisms present.

16S rDNA sequencing, in which the 16S rRNA genes in a microbiome are amplified and then sequenced(159), is the most commonly used culture-independent method for taxonomic assignment of individual organisms and documentation of evolutionary history(160-162). While it provides excellent taxonomic resolution for microorganisms present within a sample it fails to provide the functional information given by metagenomic approaches.

Spurred by significant technological advances in sequencing instruments, microbiomic approaches to clinical sample analyses are now widespread, and

have been employed to analyse microbial composition of non-sterile sites, such as the skin, nasooropharynx and gastrointestinal tract, which have been shown to have significant numbers and characteristic patterns of bacteria. Projects which have employed these techniques include the Human Microbiome Project(231,232), and MetaHIT(233).

Other body sites, such as the pleural space, are characteristically sterile in health and are unlikely to have an appreciable microbiota. Such sites may become infected, but culture is not universally reliable at detecting an infecting pathogen, being positive in only 40-60% of cases. An advantage of a 'microbiomic' 16S rDNA sequencing approach to these samples would be its relatively assumption free nature, and the sensitivity of nucleic acid amplification in detecting pathogens.

11.1.2 16S rRNA gene amplification and sequencing

The rRNA gene is the most commonly used gene for targeted amplicon based sequencing for microbiomic microorganism identification, due to the ubiquitous nature of ribosomes. Original sequencing efforts were based on the short 5S rRNA gene; subsequent technological advances led to interest being diverted to the longer 16S rRNA gene, present in all bacteria(234). Bacterial 16S rRNA genes contains nine rapidly evolving hypervariable regions (V1-V9), flanked by conserved sequences (Figures 11-2 and 11-3)(235).

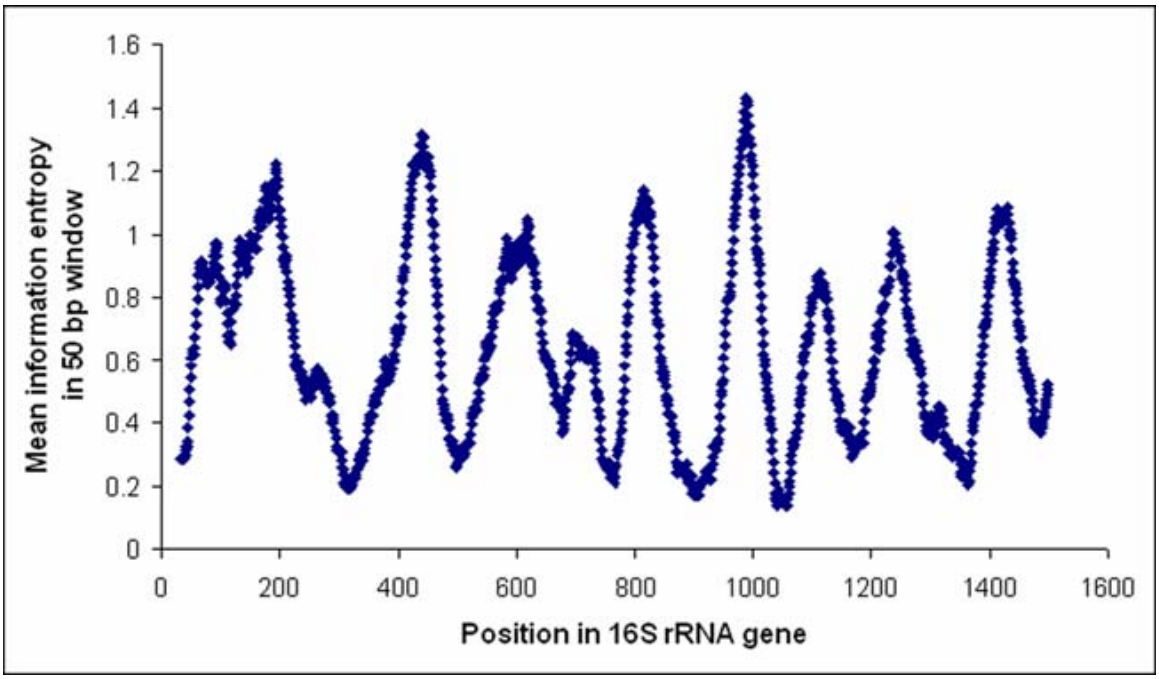


Figure 11-2 – Demonstration of hypervariable and conserved regions within the 16S rRNA gene. Variability across species at each sequence position is demonstrated on the y-axis, measured using Shannon information entropy averaged over 50 bp windows (higher entropy, more variability at a given position). Taken from Andersson *et al.*(236)

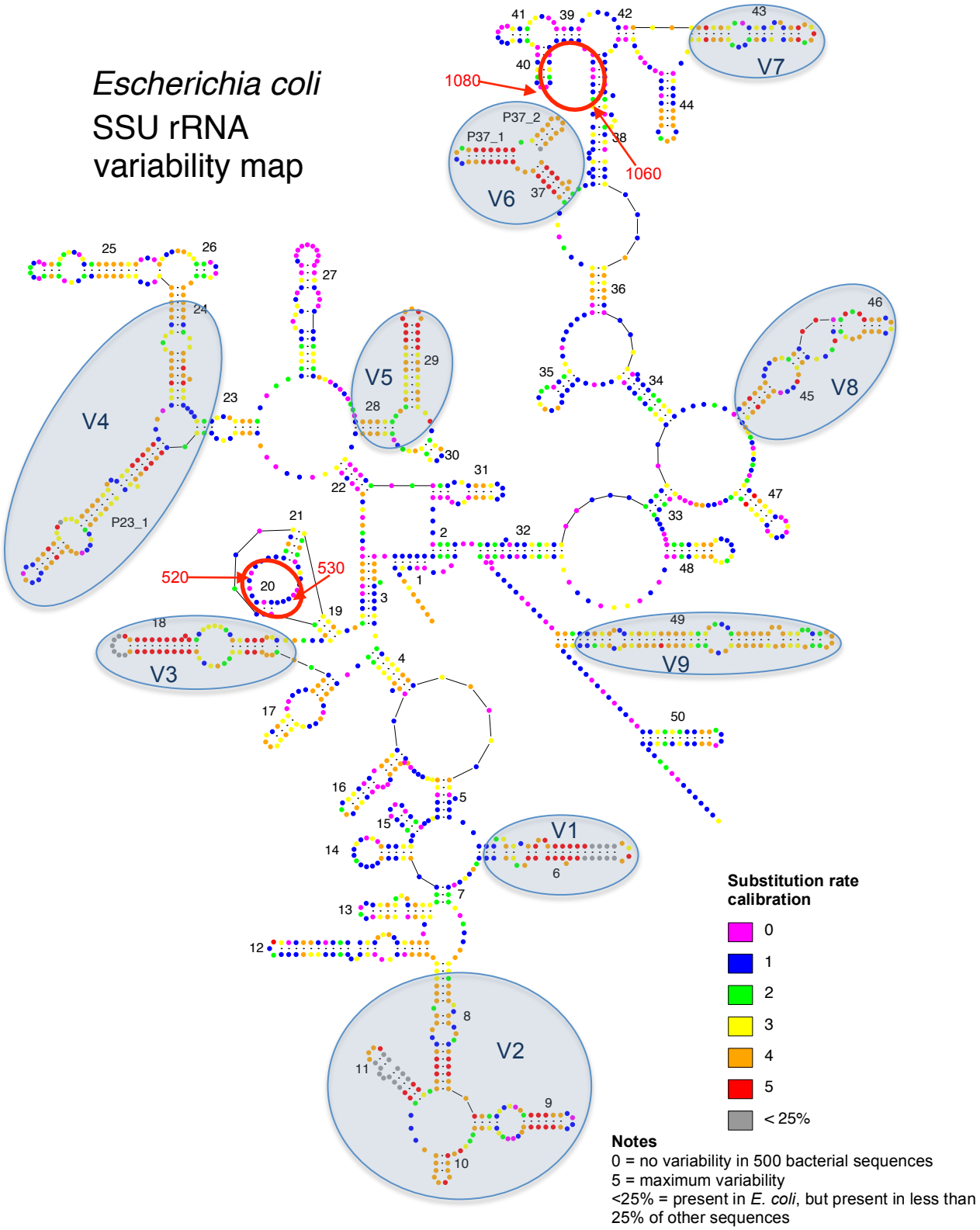


Figure 11-3 – Variability (substitution rate) of nucleotide sites superimposed on a secondary structure model of *Escherichia coli* 16S (small sub-unit (SSU)) rRNA. Red circles represent primer sites used for sequencing in this chapter. Modified from Van de Peer *et al.*(237)

The conserved regions allow ‘universal’ primer design, enabling amplification and sequencing of the hypervariable regions. Comparison of the sequences of the hypervariable regions with reference databases enables identification of individual bacterial species(237).

Several large curated databases exist for 16S analyses, including the Ribosomal Database Project (RDP)(163), SILVA(164) and the GreenGenes consortium(165).

Previous sequencing approaches to the 16S rRNA gene used capillary (Sanger) sequencing, frequently using the ABI 3730, or similar instruments(238). The major limitation of this platform is that only one read is obtained for each amplicon, a critical limitation if more that one microorganism (each with a different 16S rDNA sequence) is present in a clinical sample. Polymicrobial ‘microbiomic’ analyses have traditionally been performed by PCR-based amplification of 16S rRNA genes within a sample, followed by cloning the amplicons using a vector transformed into *Escherichia coli*. Individual transformed colonies are randomly selected, plasmids are purified and the 16S rRNA inserts are sequenced using capillary sequencing. This technique has significant disadvantages, including time required to sequence a relatively low number of transformed colonies for each sample, annealing bias(239) and cloning bias(240). Complex microbiomes may have low abundance species overlooked using such a technique. This limitation is overcome by next-generation sequencing, in which thousands of individual sequencing reads per clinical sample are achievable. A variety of sequencing platforms have been used for 16S rRNA gene microbiomic studies, each with differing performance characteristics (table 11-1).

	Read length (nt)	Run hours	Reads per run	Relative cost per Mb (using 454 as reference)	Raw error rate (%)			
					Total	Insertions	Deletions	Mismatches
ABI 3730	800	2 h	96	100	0.001	<<0.1	<<0.1	<<0.1
454 FLX Titanium	400-500	9 h	10^6	1	1	<1	<<0.1	<<1
Illumina HiSeq 2000	101-151	9-15 d	10^9	0.002	<1	<<1	<<1	<1
Illumina MiSeq	36-151	4-27 h	10^7	0.06	<1	<<1	<<1	<1
PacBio	1100	1.5 h	10^7	1.5	15	13	1	1
IonTorrent	200	2-3 h	10^6	0.4	2	1	1	<1

Table 11-1 – Comparison of sequencing platforms used and potentially suitable for microbiome studies. Adapted from Kuczynski *et al*.(241)

The 454 FLX Titanium platform was chosen for this study, primarily due to its long read lengths which enable taxonomic assignment with high resolution(242).

In this study, the 454 FLX Titanium platform uses emulsion PCR (emPCR) to clonally amplify 16S amplicons attached to microscopic beads. Subsequently, these beads are individually deposited into one of the 3.4 million wells of a PicoTiter Plate and each bead is pyrosequenced in parallel with other beads. Pyrosequencing sequentially washes all four deoxynucleoside triphosphates over the beads which, if complementary to the template strand, are incorporated by DNA polymerase. This reaction releases pyrophosphate, which is converted by two enzymatic reactions to produce light. Light production from each bead is detected in parallel using a charge-coupled device camera (CCD), which is then analysed to generate the template sequence.

11.2 Overall strategy used to ultra-deep sequence the 16S rRNA gene

Figure 11-1 depicts the overall strategy used to ultra-deep sequence the 16S rRNA gene of clinical samples and subsequently perform taxonomic assignment. Several thousand sequences are generated for each clinical sample. Due to errors in the sequencing process, denoising, quality filtering and error correction steps are undertaken – these are discussed subsequently. Sequences were collapsed into 'Operational Taxonomic Units' (OTUs), a proxy for species, on the basis of 97% sequence similarity to enable manageable further analysis, including taxonomic assignment. Whilst there is no clear-cut definition as to what threshold value corresponds to species-level similarity, 97% is a commonly accepted value for species level divergence(241).

11.3 Contamination considerations

16S rRNA gene amplification, being 'broad range', reveals sample and reagent contamination. Where possible (and for all PIPAP (chapter 13) samples), freshly collected clinical samples were immediately placed into 2 ml DNA/RNA, DNase/RNase-free microcentrifuge tubes (HydroLogix, Molecular BioProducts, San Diego, USA). Primers were divided into small volume aliquots upon receipt and, similarly, low volume reagents were used to minimise opportunities for contamination. 1.5 ml tubes of pre-aliquoted quality-assured reverse transcriptase-PCR grade water (Ambion, Carlsbad, USA) were used for making up master mixes.

All reagents and consumables (apart from polymerases, dNTPs and DNA template) underwent exposure to 254nm ultraviolet light for 15 minutes to reduce DNA contamination. Amplicons were prepared in a “master-mix only” PCR6 Vertical Laminar Flow Cabinet with built in ultraviolet sterilisation (Labcaire, North Somerset, UK) and extractions were undertaken in a separate PCR6 hood.

11.4 DNA extraction

There is ongoing debate as to the optimal technique to extract DNA in 16S-based studies(243), although all extraction techniques use the same basic principles of cell lysis, removal of proteins and other non-DNA macromolecules, and DNA collection. Various factors affect ease of bacterial DNA extraction from bacteria, in particular cell wall thickness (characteristically greater in Gram positive bacteria). Given this, a mechanical cell disruption technique was employed to extract DNA from the clinical samples in this project. Sample DNA was extracted using the FastDNA SPIN Kit technique (MP Biomedicals, Solon, OH), following the manufacturer’s instructions. This technique uses a bead-based cell lysis technique combined with a silica column-based DNA purification process. Two 40 s homogenisation steps at speed setting “6” were used and DNA was eluted in 100 µl of kit-provided buffer. DNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, USA). The extracted DNA was either amplified immediately or stored at -20°C for subsequent analysis.

11.5 16S rRNA gene quantitative PCR

Separate to the sequencing experiments, quantitative PCR of the 16S rRNA gene was undertaken on clinical samples to obtain an estimate of overall 'bacterial load'. Power SYBR Green PCR reagent was used, a master mix that contains an ultra-pure version of AmpliTaq Gold DNA Hot Start polymerase (minimising potential bacterial DNA contamination), SYBR Green I Dye, dNTPs and buffer.

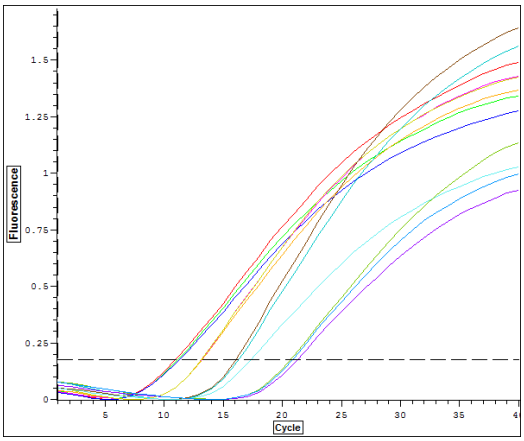
Previously published 16S rRNA gene qPCR primers were used (table 11-2), which amplify a 467 nt portion of the 16S rRNA gene, incorporating hypervariable regions V3 and V4 (figure 11-7).

Primer	Sequence	Estimated Tm (°C)[1]	Reference
S-D-Bact-0340-a-S-19 (qPCR S primer)	5'-TCCTACGGGAGGCAGCAGT-3'	62.6	(245)
S-D-Bact-0781-a-A-26 (qPCR A primer)	5'-GGACTACCAGGGTATCTAATCCTGTT-3'	61.2	

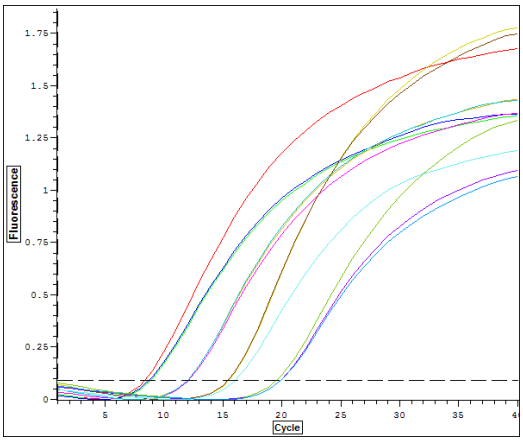
Table 11-2 – Quantitative PCR primers used

Given that this primer pair generates an amplicon longer than often desired for real-time PCR, both 2- and 3-step thermal amplification profiles were evaluated (using both combined and separate annealing and extension steps) – 3-step profiles were found to give more favourable sigmoidal fluorescence curves and were therefore subsequently used (figure 11-4).

[1] Estimated using salt correction formula of SantaLucia(244)



2-step thermal cycling profile
Combined annealing/extension step
(60 secs at 60°C)



3-step thermal cycling profile
Separate annealing and extension steps
(20 secs at 60°C, 30 secs at 72°C)

Figure 11-4 – Evaluation of 2- and 3-step thermal cycling profiles for 16S rRNA gene quantitative PCR, performed using triplicate repeat qPCRs of 10 fold dilutions of *Escherichia coli*

Acidothermus cellulolyticus DNA was used to estimate sample 16S rRNA gene concentration. This organism has only one copy of the 16S rRNA gene per genome, and ten-fold serial dilutions of the 16S rRNA gene were used (10^7-10^2 copies).

All PCRs were performed in duplicate, and the 20 µl PCR reaction mixtures consisted of 7.8 µl RT-PCR grade water (Ambion, Carlsbad, USA), 10 µl 2x Power SYBR Green PCR Master Mix, 0.3 µM of each primer (Eurofins MWG Operon, Ebersberg, Germany), and 1 µl of 1/10 diluted DNA template. The cycling conditions were as follows: 95°C for 10 mins (activation) followed by 40 cycles of 95°C for 15 s (denaturation), 60°C for 20 s (annealing), 72°C for 30 s (extension) with a final extension of 72°C for 7 mins. A melt curve was performed at 0.3°C increments between 72 and 95°C (Opticon 2 DNA Engine, Bio-Rad; California, USA).

Melt-curve analyses and agarose gel electrophoresis of qPCR amplicons were used to ensure absence of non-specific PCR products, such as primer dimers.

11.6 Amplicon preparation

11.6.1 Reference organism spike

All samples were spiked with a low concentration of DNA from *Acidothermus cellulolyticus* (ATCC 43068, obtained from LGC Standards, London, UK), to enable a contamination threshold to be defined during sequencing. *A. cellulolyticus* is a thermophilic, acidophilic bacteria obtained from acidic hot springs, with an optimal growth at pH 5 and temperature 55ºC(246). This bacteria has never been reported in clinical samples.

11.6.2 Polymerase selection

Errors during PCR, including incorporation of incorrect bases or chimera formation could lead to incorrect conclusions being drawn regarding bacterial species present within a clinical sample, including a falsely high estimate of bacterial diversity. Conventional Taq polymerases have been shown to have a relatively high error rate, and a high fidelity polymerase (with 3'→5' proofreading exonuclease activity) was chosen in an attempt to reduce errors(247,248). Two enzymes were compared – Qiagen HotStar HiFidelity DNA Polymerase (Qiagen, Hilden, Germany), an enzyme derived from *Pyrococcus* spp. stated to have an enzyme fidelity ~10x standard *Taq* polymerase(249); and Finnzymes Hot Start II DNA Polymerase, also derived from *Pyrococcus* spp., but stated to have an

additional DNA binding domain and a 52x higher fidelity than standard *Taq* polymerase(250) and found to have the lowest error rate of multiple high fidelity enzymes when used to 454 sequence a clonal TP53 plasmid(251). In practice, Qiagen HotStar HiFidelity DNA polymerase performed less well than Finnzymes Hot Start II DNA polymerase, and reactions frequently suffered with additional undesired short fragments even after varying PCR annealing temperatures and extension phase duration (figures 11-5 and -6).

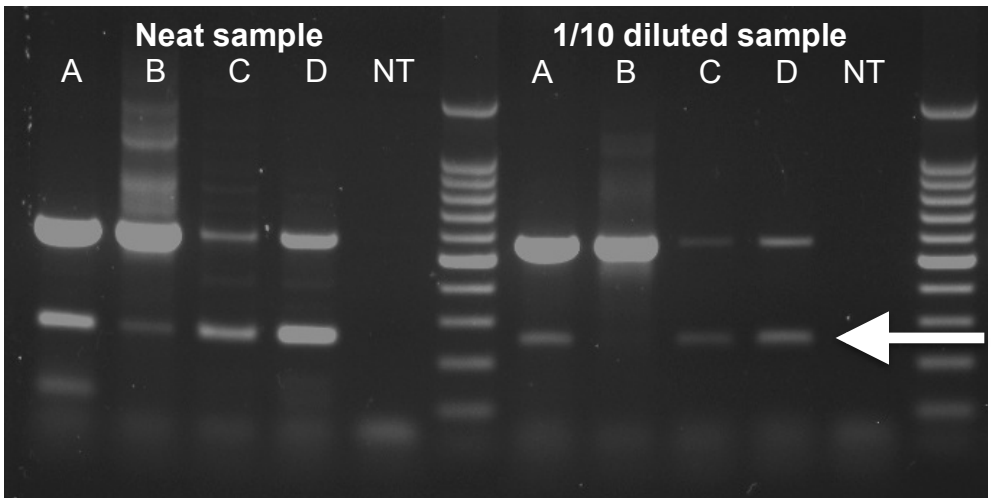


Figure 11-5 – Electrophoretic gel of PCR products (representative samples) using Qiagen HotStar HiFidelity DNA polymerase. Note a prominent unwanted additional band at ~250nt (arrow)

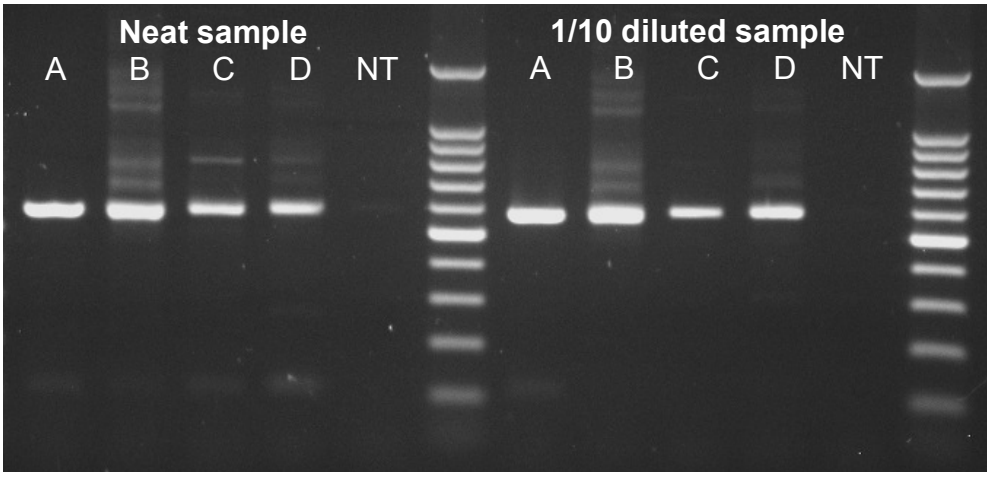


Figure 11-6 – Electrophoretic gel of PCR products (representative samples) using Finnzymes Phusion Hot Start II High-Fidelity DNA polymerase

11.6.3 Primer selection

There are a large choice of published primers for 16S rDNA, differing in PCR product characteristics (such as length, compatibility with sequencing platforms and taxonomic assignment resolution) and specificity of primers (e.g. specificity to bacteria vs. human host sequences and broad-range 'universal' bacterial primers vs. those specific to certain species).

11.6.3.1 Amplicon length and taxonomic coverage

The importance of primer selection for 16S-based diagnostic and metagenomic studies has been previously highlighted(229,252).

Whilst population level analyses, such as sequence clustering and Unifrac analyses are relatively insensitive to read length(253), detailed species-level (operational taxonomic unit-level) analyses are dependent on read length(254).

16S rRNA gene read lengths >400nt, typical of 454 FLX Titanium pyrosequencing have been shown to give accurate species-level classification(252).

Of the hypervariable regions, V4 has been consistently shown to give among the highest taxonomic classification accuracy and confidence estimates(242,255). The V2 and V6 regions also provide useful information. Another consideration is that amplification of differing hypervariable regions can give skewed estimates of number of OTUs in a sample; one cloning-based study proposed that V4 and V5+V6 provide OTU estimates most similar to amplification of a nearly full-length 16S rRNA gene fragment(256). For this study, primers were selected which amplify the V4-V6 region (table 11-3, figure 11-7)

Primer	Sequence	Estimated Tm (°C)[2]	Reference
S-D-Bact-0517-b-S-16 (Sequencing S primer)	5'-GCCAGCMGCNGCGGTA-3'	59.7-65.7	(257)
S-D-Bact-1061-a-A-17 (Sequencing A primer)	5'-CRRCACGAGCTGACGAC-3'	55.6-60.9	

Table 11-3 – Sequencing primers used in this study

[2] Estimated using salt correction formula of SantaLucia(244)

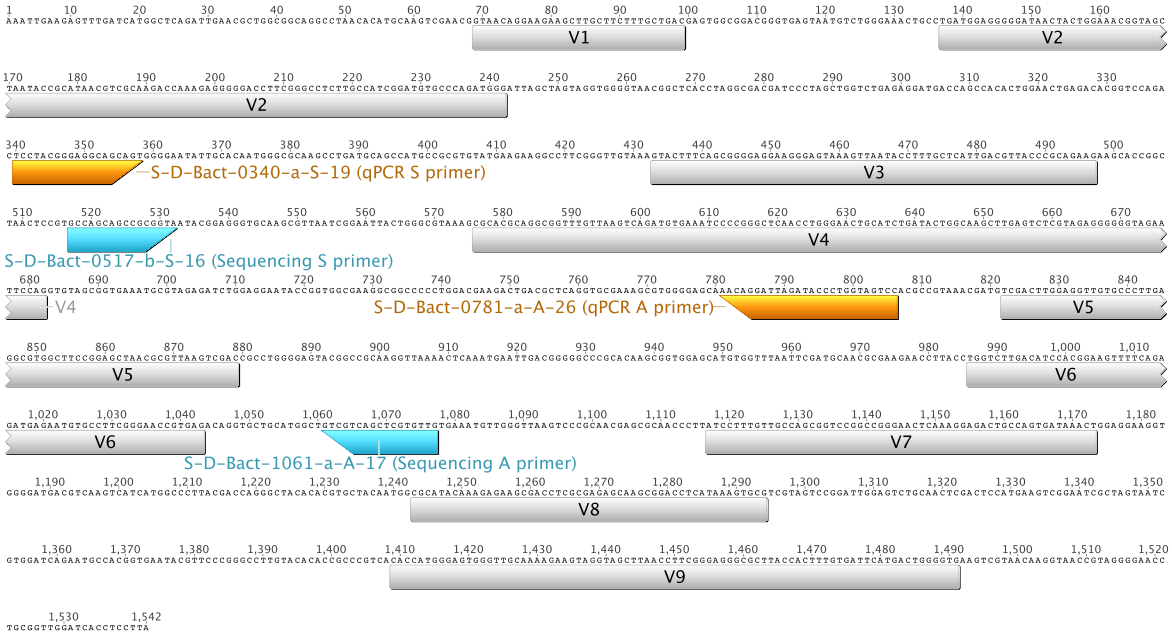


Figure 11-7 – Primers used in this study and hypervariable regions with positions corresponding to the 16S rRNA gene of Escherichia coli (accession number J01695).

Both the sequencing and qPCR primers were evaluated with BLASTn to ensure that they had no complementarity to the human genome[3].

The utility of the primer pairs used in this study in detecting all bacterial species was evaluated by testing the primer sequences against the 16S rRNA gene database at the Ribosomal Database Project (http://rdp.cme.msu.edu). Version 10.31 (December 2012) of the database has a collection of 16S rRNA genes from 9196 type-strain bacterial isolates and 1186838 high quality bacterial sequences 1200 nt long. Tables 11-4 and 11-5 show the results of both primer pairs used for this study. The sequencing primers used in this study had an identical match to 95.7% of all good quality, ≥1200 bp sequences obtained from culture isolates.

[3] Using "word size" = 7, "expect value" = 1000 and "low complexity filter" disabled and all possible permutations of primers (when ambiguous bases were present)

Assuming a PCR is likely to tolerate 2 primer-template mismatches, the primer pair would anneal to 99.6% of sequences (table 11-4). The qPCR primers have an identical match to 86.8%; assuming tolerance to two primer-template mismatches, the primer pair would anneal to 96.6% of sequences (table 11-5).

Primers: S-D-Bact-0517-b-S-16 (sequencing S primer), S-D-Bact-1061-a-A-17 (sequencing A primer)							
		Primer mismatches allowed					
		0 mismatches		**1 mismatch**		**2 mismatches**	
Phyllum	**Sequences within RDP[4]**	**Sequence matches**	**%**	**Sequence matches**	**%**	**Sequence matches**	**%**
Bacteria (domain)	**179984**	**172200**	**95.7**	**178111**	**99.0**	**179245**	**99.6**
Actinobacteria	29560	28270	95.6	29298	99.1	29443	99.6
Aquificae	177	167	94.4	174	98.3	176	99.4
Bacteroidetes	8710	8143	93.5	8470	97.2	8634	99.1
Caldiserica	3	3	100.0	3	100.0	3	100.0
Chlamydiae	324	318	98.1	322	99.4	322	99.4
Chlorobi	152	9	5.9	151	99.3	152	100.0
Chloroflexi	175	169	96.6	172	98.3	174	99.4
Chrysiogenetes	10	9	90.0	10	100.0	10	100.0
Deferribacteres	43	42	97.7	42	97.7	42	97.7
Deinococcus-Thermus	488	477	97.7	483	99.0	486	99.6
Dictyoglomi	16	16	100.0	16	100.0	16	100.0
Elusimicrobia	3	3	100.0	3	100.0	3	100.0
Fibrobacteres	72	66	91.7	71	98.6	72	100.0
Fusobacteria	546	400	73.3	535	98.0	543	99.5
Gemmatimonadetes	7	7	100.0	7	100.0	7	100.0
Lentisphaerae	11	10	90.9	11	100.0	11	100.0
Nitrospira	120	116	96.7	120	100.0	120	100.0
Planctomycetes	395	384	97.2	392	99.2	394	99.7
Proteobacteria	81443	78806	96.8	80754	99.2	81156	99.6
Spirochaetes	1791	1731	96.6	1771	98.9	1780	99.4
Synergistetes	127	118	92.9	123	96.9	124	97.6
Tenericutes	2390	1871	78.3	2377	99.5	2387	99.9
Thermodesulfobacteria	16	16	100.0	16	100.0	16	100.0
Thermotogae	162	160	98.8	162	100.0	162	100.0
SR1	7	7	100.0	7	100.0	7	100.0
TM7	11	0	0.0	11	100.0	11	100.0
Armatimonadetes	4	4	100.0	4	100.0	4	100.0
Verrucomicrobia	151	146	96.7	150	99.3	150	99.3
Acidobacteria	159	157	98.7	159	100.0	159	100.0
Firmicutes	48843	46809	95.8	48265	98.8	48630	99.6
Cyanobacteria/Chloroplast	3691	3621	98.1	3676	99.6	3686	99.9

Table 11-4 – Sequences from the Ribosomal Database Project (Release 10, update 31, 7th December 2012) matching sequencing primers used within the study

[4] Using 16S rRNA gene sequences classified by RDP as good quality, ≥1200 bp, obtained from culture isolates

Primers: S-D-Bact-0340-a-S-19 (qPCR S primer), S-D-Bact-0781-a-A-26 (qPCR A primer)							
		Primer mismatches allowed					
		0 mismatches		1 mismatch		2 mismatches	
Phyllum	Sequences within RDP[5]	Sequence matches	%	Sequence matches	%	Sequence matches	%
Bacteria (domain)	**179984**	**156158**	**86.8**	**168234**	**93.5**	**173838**	**96.6**
Actinobacteria	29560	25616	86.7	29136	98.6	29412	99.5
Aquificae	177	0	0.0	0	0.0	169	95.5
Bacteroidetes	8710	7411	85.1	8212	94.3	8574	98.4
Caldiserica	3	3	100.0	3	100.0	3	100.0
Chlamydiae	324	2	0.6	208	64.2	216	66.7
Chlorobi	152	135	88.8	148	97.4	151	99.3
Chloroflexi	175	2	1.1	19	10.9	151	86.3
Chrysiogenetes	10	9	90.0	10	100.0	10	100.0
Deferribacteres	43	39	90.7	43	100.0	43	100.0
Deinococcus-Thermus	488	0	0.0	0	0.0	478	98.0
Dictyoglomi	16	0	0.0	0	0.0	16	100.0
Elusimicrobia	3	0	0.0	0	0.0	3	100.0
Fibrobacteres	72	57	79.2	70	97.2	72	100.0
Fusobacteria	546	493	90.3	515	94.3	545	99.8
Gemmatimonadetes	7	7	100.0	7	100.0	7	100.0
Lentisphaerae	11	0	0.0	7	63.6	7	63.6
Nitrospira	120	108	90.0	120	100.0	120	100.0
Planctomycetes	395	0	0.0	1	0.3	62	15.7
Proteobacteria	81443	75824	93.1	79137	97.2	80977	99.4
Spirochaetes	1791	10	0.6	1358	75.8	1402	78.3
Synergistetes	127	0	0.0	0	0.0	122	96.1
Tenericutes	2390	1786	74.7	1925	80.5	2278	95.3
Thermodesulfobacteria	16	0	0.0	0	0.0	16	100.0
Thermotogae	162	0	0.0	0	0.0	156	96.3
SR1	7	6	85.7	6	85.7	7	100.0
TM7	11	0	0.0	0	0.0	11	100.0
Armatimonadetes	4	0	0.0	0	0.0	4	100.0
Verrucomicrobia	151	0	0.0	0	0.0	54	35.8
Acidobacteria	159	141	88.7	142	89.3	159	100.0
Firmicutes	48843	44433	91.0	47070	96.4	48456	99.2
Cyanobacteria/Chloroplast	3691	0	0.0	4	0.1	18	0.5

Table 11-5 – Sequences from the Ribosomal Database Project (Release 10, update 31, 7th December 2012) matching qPCR primers used within the study

Using another 16S reference dataset (the SILVA 16S/18S rRNA gene non-redundant reference dataset; SSURef 108 NR), Klindworth and colleagues evaluated individual 175 primers and 512 primer pairs for their utility in

[5] Using 16S rRNA gene sequences classified by RDP as good quality, ≥1200 bp, obtained from culture isolates

sequencing-based diversity studies for both Bacteria and Archaea(229). The authors pointed out that suboptimal primer pairs can lead to under-representation or selection against certain species. Of published primers, they found that the reverse primer used in this study (S-D-Bact-1061-a-A-17, using their nomenclature) had the highest overall coverage and specificity for the domain Bacteria, with a 96.4% coverage of all bacteria within the SILVA NR database. The primer referenced as S-D-Bact-0517-a-S-17 had 95.7% coverage for bacteria, and was most similar to the forward primer used in this study (S-D-Bact-0517-a-S-17 sequence GCCAGCAGCCGCGGTAA vs. S-D-Bact-0517-b-S-16 sequence GCCAGCMGCNGCGGTA, differences highlighted).

11.6.4 Design of fusion primers used for ultra-deep sequencing

To enable sequencing and sample identification, modified 'fusion' primers are used to amplify the 16S rRNA gene (Roche technical bulletin TCB No. 013-2009). These fusion primers are modified from the sequencing primers and include a 10 nt multiplex identifier (MID) tag for sample identification and one of two 25 nt sequences ('Primer A' or 'Primer B' sequences), which are used to hybridise amplicons to DNA capture beads, and to enable annealing of emPCR amplification primers and sequencing primers during ultra-deep sequencing (figure 11-8).

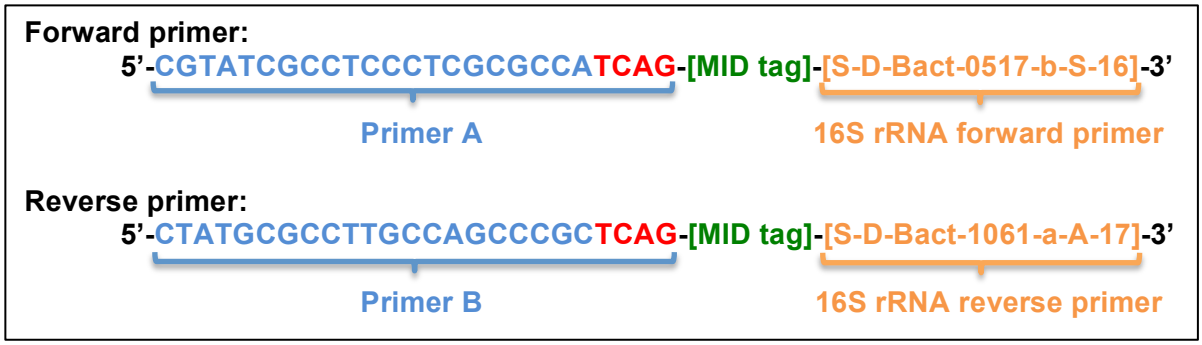


Figure 11-8 – Fusion primer design for ultra-deep sequencing

11.6.5 Optimisation of PCR using fusion primers

Using fusion primers, a variety of PCR conditions were systematically adjusted to optimise PCR yield while minimising spurious PCR product generation. Variables that were evaluated included denaturation time (10 s found to be optimal), extension time and temperature (24 s at 72°C optimal), polymerase concentration (2 U per 50 ul reaction optimal), primer concentration (0.5 uM optimal), magnesium concentration (1.5 mM optimal) and use of betaine/DMSO as PCR additives (both found to be detrimental). Both two- and three-step thermal cycling is possible with this enzyme, and a two-step profile was found to be favourable.

11.6.6 Sequencing PCR mix composition and thermal cycler settings

The PCR reaction mixtures consisted of RT-PCR grade water (Ambion, Carlsbad, USA) to take total volume to 50 µl, 10 µl 5x Phusion HF buffer, 200 µM dNTPs (New England BioLabs, Ipswich, USA), 0.5 µM of each primer (Eurofins MWG Operon, Ebersberg, Germany), 2 U of Phusion Hot Start II DNA polymerase, 100 copies of *Acidothermus cellulolyticus* DNA and 5 µl DNA template. The cycling

conditions were as follows: 98°C for 30 s (initial denaturation) followed by 35 cycles of 98°C for 10 s (denaturation) and 72°C for 24 s (combined annealing and extension), with a final extension of 72°C for 10 mins (Applied Biosystems 2720 Thermal Cycler; Carlsbad, USA).

454 ultra-deep sequencing runs were designed to sequence 96 samples in one run, configured with 24 different samples (each with a different MID tag) in 4 sequencing lanes on the PicoTiterPlate (sequencing plate). All sequencing reactions were performed in duplicate, to minimise potential amplification biases, and reaction products were pooled post-PCR prior to any further processing.

11.6.7 Amplicon purification

Amplicon purification broadly followed Roche document '*Amplicon Library Preparation Method Manual (GS FLX Titanium Series), revision October 2009*' with minor modifications, detailed below.

The size and presence of amplicons were established using electrophoresis on 1% agarose gel alongside a 100 bp ladder (Promega, Madison, USA), with ethidium bromide and ultraviolet light visualisation. When no amplicon was detected (possibly due to PCR inhibition), the PCR was repeated in duplicate using 1 µl of sample DNA template.

Individual amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter, California, USA). These are Solid Phase Reversible Immobilisation (SPRI) para-magnetic beads in PEG, which reversibly bind DNA, enabling efficient purification using ethanol wash phases prior to elution in Tris-EDTA. The

manufacturer recommends a DNA volume:Ampure XP volume of 1:1.8, which will remove all primers and DNA fragments below 100 nt, but this would fail to remove the longer primer-dimer fragments seen secondary to fusion primers, characteristically ~150 nt. Changing the DNA volume:Ampure XP volume ratio to 1:0.7 has been shown to efficiently remove all fragments below 200-300 nt(258), and was used in this study. The Roche protocol was followed, except; (a) 90 µl of amplicon products was mixed with 63 µl Ampure XP beads; (b) ethanol washes were undertaken with 200 µl (rather than 100 µl) 70% ethanol; and (c) DNA was eluted in 45 µl Tris-EDTA buffer.

Subsequently, amplicons were quantified using the fluorometric Quant-iT PicoGreen dsDNA Assay (Invitrogen, Carlsbad, USA) following Roche's instructions, and amplicons were combined in four pools at equal concentration (each pool destined to be sequenced on one lane of the PicoTiter Plate).

The four amplicon pools were twice purified further with Ampure XP using the same ratio as previously, and DNA was finally eluted in 12 µl Tris-EDTA buffer. Amplicon purity was verified using an Agilent 2100 Bioanalyzer using the High Sensitivity DNA Assay Kit (Agilent, Santa Clara, USA) (figure 11-9).

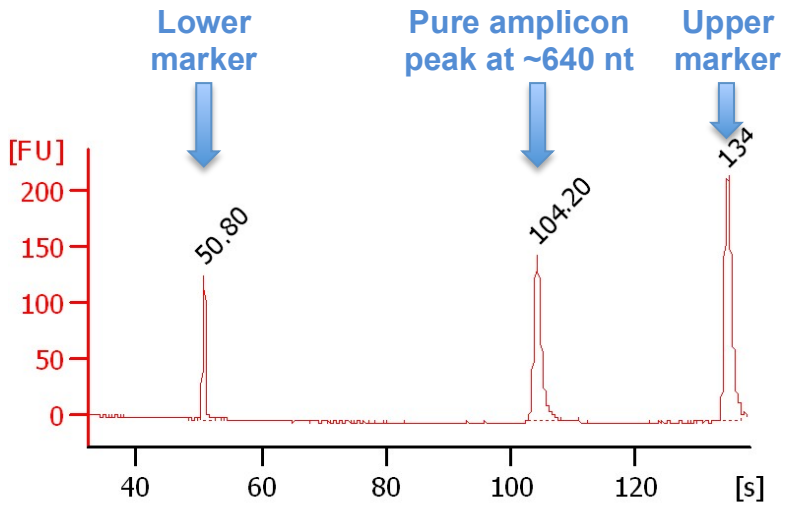


Figure 11-9 – Typical electropherogram of an amplicon pool showing 2 marker ladders and an expected single prominent peak at ~640 nt.

As previously, DNA was quantified using the Quant-iT PicoGreen dsDNA Assay and DNA was diluted to 1 x 10^9 molecules/µl in Molecular Biology grade water. This was stored at -20°C until required for emulsion PCR, at which point DNA was further diluted to 5 x 10^5 molecules/µl.

11.6.8 Emulsion PCR (emPCR)

Emulsion PCR was performed as per Roche document '*emPCR Method Manual – Lib-A MV (GS FLX Titanium Series), revision January 2010*'. Samples were prepared for loading in Medium-sized regions using a four lane gasket (referred to as 1xAmp MVE within the referenced document). Modifications to the Live Amplification Mix were made as per Roche technical bulletin 2011-001 '*Amplicon Sequencing with Various emPCR Amplification Conditions*' to optimise the sequencing of the relatively long amplicons used in this project. These changes

included increases to emPCR amplification primers, amplification mixes and alteration of thermocycler conditions.

Using the titration procedures described in the Roche literature, it was empirically determined that 0.37 amplicon molecules (DNA library) per DNA capture beads provided optimal DNA capture bead enrichment during emPCR (defined by Roche as 8% DNA capture bead enrichment) – this quantity of amplicons balances over-enrichment (which risks several differing amplicon molecules annealing to a single capture bead, leading to failed sequencing) versus under-enrichment (which reduces the amount of usable sequencing data). Bead counting was undertaken on a CASY TT cell counter (Roche, Burgess Hill, UK), in accordance with the manufacturer's instructions.

11.7 Ultra-deep sequencing and data processing

11.7.1 Sequencing on Roche 454 GS FLX instrument

Sequencing of enriched DNA capture beads was undertaken on a Roche 454 GS FLX instrument (Roche, Burgess Hill, UK), following the methods within Roche document '*Sequencing Method Manual, revision November 2010*), for medium-sized regions using a four lane gasket.

11.7.2 454 .cwf image file processing

Pyrosequencing flow signal information was extracted for each raw well and was organised into composite well format (.cwf) for further processing. Subsequently,

these files underwent signal processing using the Roche GS Run Processor tool (version 2.6). Signal processing performs a series of normalisation, correction and quality filtering steps and then outputs remaining high quality signals into flowgrams for each well. The signal processing step also generates basecalls with associated quality scores for the individual reads, and outputs this data as standard flowgram format (.sff) files that contain all signal trace data for the reads. Signal processing was invoked using the `runAnalysisPipeAmplicons` command line interface launch script, with a modified xml file (found experimentally to maximise the number and length of reads passing filter) (appendix 16).

11.7.3 Further data analysis using QIIME

Given the large amount of data produced by ultra-deep sequencing, direct manipulation of data (e.g. manually aligning DNA sequences) is impossible without using specific data analysis tools.

Data analyses were therefore partly performed using QIIME (Quantitative Insights Into Microbial Ecology – www.qiime.org), an open source software platform for analyses of microbial communities based on high-throughtput amplicon sequencing data(259).

QIIME uses .sff files and sample mapping files (including primer sequences and MID tags attributed to each sample) and performs the following;

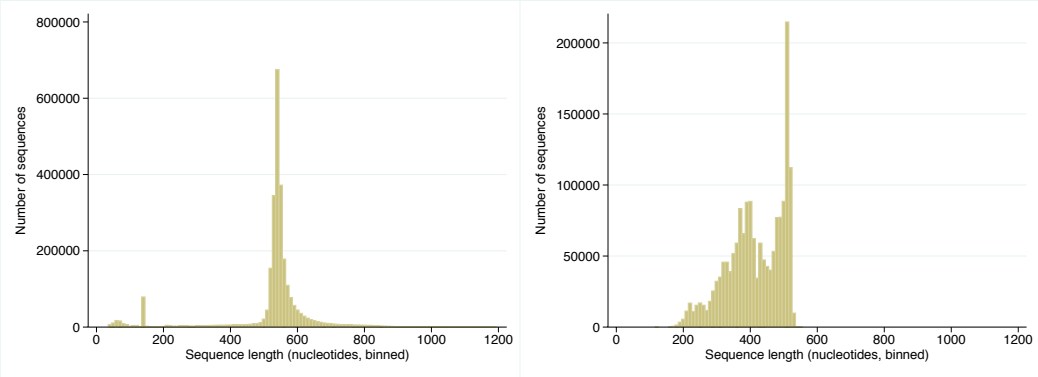
1. Data pre-processing
    a. Sample ‘demultiplexing’

    b. Primer removal
    c. Quality filtering
2. Data denoising using flowgram clustering
3. OTU picking, and identification of representative sequences for each OTU
4. Taxonomic assignment using the Ribosomal Database Project classifier
5. Sequence alignment using the PyNAST aligner
6. Phylogenetic tree construction using FastTree
7. Calculation of alpha and beta diversity with rarefaction
8. Result visualisation

The full script and options used within QIIME are in appendix 17, but a brief summary of the workflow follows. Appendix 19 contains example mapping files used for demultiplexing and analyses.

11.7.3.1 Data pre-processing

As previously discussed, each 454 ultra-deep sequencing run analysed 96 samples, 24 samples (each with a unique MID tag) in each of 4 lanes. Each lane generates a .cwf and .sff files, and these were demultiplexed using `split_libraries.py` to bin sequences according to sample and remove the MID tag. Sequences also underwent quality filtering for length parameters (minimum 200 nt, maximum 650 nt), minimum quality score and both forward and reverse primers were removed (figure 11-10).



(a) Raw sequence lengths (median 540 nt, IQR 530-560 nt)

(b) Lengths of sequences that pass quality filters post demultiplexing (median 420 nt, IQR 360-490 nt).

Figure 11-10 – Length distribution for sequences generated by 454 FLX pyrosequencing as part of this study

11.7.3.2 Data denoising using flowgram clustering

Pyrosequencing produces occasional characteristic sequencing errors, mostly at homopolymer repeats where the number of repeats may be misinterpreted and may artificially create additional erroneous OTUs. `denoise_wrapper.py` within QIIME enables flowgram clustering in an attempt to reduce this artificial noise(260). This process is computationally intensive, and typically takes several days even on a multicore system for a full 454 run.

11.7.3.3 OTU picking & selection of representative sequences for each OTU

Having denoised sequencing reads, reference-based OTU clustering was performed to assign the reads to microbial species. The `pick_otus.py` script

was used to pick OTUs, using the reference-based uclust method(261). This method uses reference sequences as 'seeds', to which sample sequences are clustered at 97% identity. The Greengenes database (October 2012 release, http://greengenes.secondgenome.com) is an international consortium curated database of ≥1200 nt 16S rRNA gene sequences for the kingdoms Bacteria and Archaea, and was used to provide reference sequences and taxonomic assignments(165,262,263). The 1075170 sequences within this database have had extensive chimera filtering, and have been clustered at varying sequence identities to create OTUs. A 97% sequence identity threshold was used in this study, corresponding to 86548 OTUs. It might be expected that a higher sequence identity threshold would identify a greater number of OTUs, but a sensitivity analysis comparing 97% with 99% sequence identity thresholds consistently showed slightly more OTUs were identified using a 97% identity threshold.

Such a reference-based OTU picking approach is particularly useful for the bidirectional sequencing approaches used in this study; further, it reduces the impact of chimeric sequencing reads.

`Pick_rep_set.py` was subsequently used to pick a representative sequence for each OTU identified (chosen to be the most abundant sequence in each OTU).

11.7.3.4 Taxonomic assignment

Taxonomic assignment to genus level was based on the Greengenes taxonomy and reference database(262,263) using the RDP classifier (version 2.2)(242). This

is a naïve Bayesian classifier, which does not require sequence alignment and has been shown to be rapid while providing high accuracy (~98%).

After contaminant removal (see below), species-level taxonomic assignment was performed for representative sequences using both Megablast [6] and RDP-Seqmatch[7]. Where a representative sequence had two or more sequences with the same pairwise identity, type-strain sequences listed on the database "List of Prokaryotic names with Standing in Nomenclature" (http://www.bacterio.cict.fr/index.html) were used to arbitrate where possible.

11.7.3.5 OTU table construction and graphical representations

`Make_otu_table.py`, `summarize_taxa.py` and `plot_taxa_summary.py` were used to construct OTU tables and display these results.

11.7.3.6 Sequence alignment and phylogenetic tree construction

Sequences were aligned to the Greengenes Core reference alignment using PyNAST(165,264), a python implementation of the NAST alignment algorithm (`align_seqs.py`).

---

[6] Using the nr database, with maximum e-value 1e-1, and the following argument: `all[filter] NOT (uncultured bacterium[orgn] OR uncultured organism[orgn])`. All sequences had % pairwise identity >98% (with most having pairwise identity ≥99%)

[7] Using the following parameters – "source" = isolates; "size" ≥1200 nt; "quality" = good

The phylogenetic relationships among DNA sequences can be inferred de novo using FastTree(265); this allows the use of phylogenetically-aware analyses such as UniFrac(266), which measure the distance between two samples in terms of the fraction of evolutionary history that separates the organisms in the two samples (or environments). For each pair of samples, UniFrac measures the fraction of total branch length in a phylogenetic tree that leads to sequences from one community or the other(253). The UniFrac value is determined for all pairs of samples to produce a distance matrix, and this can be used to perform dimensionality reduction using Principal Coordinates Analysis (PCoA), in which the matrix of distances are converted into points on a projection that maximises the amount of variation along a series of orthogonal axes.

11.7.3.7 Contaminant removal

All sequencing runs included control samples, either control pleural fluid samples or control lung aspirates. These underwent PCR amplification as normal, typically producing very faint bands on gel electrophoresis (unlike 'no template' samples which universally failed to produce an amplicon); these control sample amplicons were subsequently sequenced. Examination of sequences found a similar pattern of contamination in all control samples, in which *Ralstonia* spp. corresponded to the most common OTU.

As previously discussed, all samples were spiked with 100 copies of *Acidothermus cellulolyticus*. This enabled calculation of a ratio of *Ralstonia spp.* sequences to *Acidothermus cellulolyticus* sequences for each sample, as follows;

$$x = \frac{Number\ of\ sequences\ corresponding\ to\ Ralstonia\ spp.}{Number\ of\ sequences\ corresponding\ to\ A.cellulolyticus}$$

Taking the values of the ratio *x* for each sample, a threshold was defined (two standard deviations above the mean ratio of *Ralstonia* spp. to *Acidothermus cellulolyticus*);

$$threshold = \bar{x} + 2\sigma_x$$

This threshold was used to define OTUs considered to be contaminants and those considered to be of importance.

Such an approach is novel, but necessary, given the ubiquity of contamination. This strategy is possible due to a close correlation between number of ultra-deep sequencing reads attributable to a given species and the actual number of gene copies present in a clinical sample from the same pathogen. Several previous authors have noted such a correlation across a variety of gene targets(267-273). Use of a spike sequence, as an internal control, is common practice in diagnostic microbiology(274), however the threshold definition strategy is novel to this study. Our threshold definition should ensure that there is less than 3% chance that a contaminant remains after threshold filtering (assuming the ratio *x* is normally distributed). Sensitivity and specificity analyses of this approach are discussed in chapters 12 and 13.

11.7.3.8 Alpha and beta diversity calculations

`Alpha_rarefaction.py` was used to generate rarefied OTU tables, compute alpha diversity metrics and collate these results.

`Beta_diversity_through_plots.py` subsampled each sample's sequences and performed beta diversity and principal coordinate analyses (PCoA). Such PCoA plots had taxa superimposed (weighted by relative abundance of taxa in each sample) using `make_3d_plots.py`.

11.7.4 Phylogenetic tree visualisation

Rarefied OTU tables and phylogenetic trees were visualised using TopiaryExplorer version 1.0 (https://github.com/qiime/Topiary-Explorer), using the polar plot option. Aesthetic improvements were made using CorelDRAW Graphics Suite X6 (Corel Corp., Ottawa, Canada).

11.7.5 Heatmap analysis of data using Phyloseq

Heatmap analysis of abundances of bacterial lineages was undertaken using non-metric multidimensional scaling (NMDS) ordination techniques based on Bray-Curtis distances. The `plot_heatmap` function of the R-based Phyloseq package (http://joey711.github.io/phyloseq/) was used. Absolute sequence abundances were converted to relative abundances and both pre- and post-threshold sequence plots were generated (using both linear and logarithmic scales). See appendix 18 for example scripts used.

11.7.6 Error considerations

Compared with whole genome sequencing, in which each region of the genome is sequenced many times, 16S rRNA gene fragments may only be sequenced a few times for very rare organisms (particularly contaminants). The effect of PCR errors (including chimera generation in up to 8% of reactions(275)), conflated by sequencing errors is potentially great, as every error could be portrayed as arising from a novel organism(276,277). Various strategies were used to reduce the effect of potential errors;

1. Use of the Roche GS Run Processor tool to quality filter reads and remove errors characteristic of pyrosequencing, including homopolymer repeat errors.
2. Further quality filtering within QIIME, including filtering by average quality score and removal of reads with significant primer/MID mismatches, and those reads which are particularly short or long.
3. Sequence read denoising using flowgram clustering. Such denoising algorithms have been shown to provide the best technique to reduce error rate, improve sequence length and number of sequences when known mock community samples were sequenced(275).
4. Reference-based OTU picking is a particularly efficacious method for removing chimeric sequences.

11.8 Further statistical analyses

Unless specified otherwise, all further statistical analyses and graph plotting were performed using STATA version 12.1 (StataCorp, Texas, USA) and Prism 6 (GraphPad Software, California, USA).

12 Ultra-deep sequencing – MIST2 Pleural Infection samples

12.1 Introduction

Chapter 10 used NAAT with high sensitivity and specificity to investigate atypical causes of pleural infection. This approach, using pathogen-specific primers, is excellent for rule-in/rule-out hypothesis testing, but gives no information about all the other potential causes of pleural infection.

Current laboratory techniques for broad-range identification of bacteria in cases of pleural infection uses culture of pleural fluid, but this is positive in only approximately 30-40% of cases. Public Health England (the successor to the Health Protection Agency) recommends that fluid should be aerobically cultured on blood and chocolate agar and anaerobically cultured on fastidious anaerobic agar to maximise test yield(278). Additional bedside inoculation of pleural fluid into 'blood culture bottle' (BCB) media is increasingly used, having been shown to increase culture positivity (by 21% to 59% in one study(120)). This increased yield is beyond that expected from repeated testing, and may be due to the relatively large volume of pleural fluid cultured (5-10 ml) and antibiotic binding resins within the media. However, such culture techniques remain imperfect for a number of reasons;

- Prior antibiotic usage is likely to be a significant factor in culture negativity, particularly with conventional (non-BCB) culture
- Some organisms may become non-viable during sample transport. *Streptococcus pneumoniae* has been shown to be particularly susceptible to delays in initiating culture(279)

- Fastidious organisms with specific nutritional requirements (such as *Haemophilus* spp.) are poorly cultured by BCB. Other organisms, such as *Treponema* spp. are unculturable using standard culture techniques.
- Sodium polyanetholsulphonate (SPS) is used as an anticoagulant in BCB media; in the absence of blood (e.g. when pleural fluid is used as the inoculating specimen), SPS may have a detrimental effect on the culture of *Neisseria* spp. and *Moraxella catarrhalis*(280).

Pleural fluid 16S rRNA gene amplification using 'universal' bacterial primers and subsequent capillary sequencing has been previously used in an attempt to further increase broad-range bacterial identification(29,281), being unaffected by prior antibiotics and not requiring initial culture. However, capillary sequencing only provides a single sequence (or pair of sequences) per sample and is therefore unable to resolve polymicrobiality without use of time-consuming 16S rRNA-based cloning techniques. Such cloning techniques have been applied in limited numbers previously and have suggested significant pleural infection polymicrobiality, particularly for anaerobes(29,282).

With technological advances, such as those deployed in the Roche 454 platform, it is now possible to use ultra-deep sequencing of the 16S rRNA gene to significantly increase the rate of accurate pathogen identification, while simultaneously gaining a full understanding of the polymicrobiality of complex infections.

The aim of this study was to use these techniques to provide a definitive understanding of pleural infection and, in particular, its polymicrobial nature. The

Roche 454 FLX platform was used to pyrosequence 16S rRNA gene amplicons generated from samples from the Second Multicentre Intrapleural Sepsis Trial (MIST2)(75), a randomised controlled trial investigating the effects of intrapleural tissue plasminogen activator (tPA) and DNase. Samples were sequenced to a depth of almost 5000 reads per sample with a representative read length of ~500 nt, thereby allowing species-level taxonomic assignment. Further, I sought to gain increased understanding about the characteristics of culture- and sequencing-negative 'pleural infection' samples using 16S rRNA quantitative PCR (designed to give an estimate of 'bacterial burden').

12.2 Methods

12.2.1 Sample collection

Pleural fluid samples were obtained from patients with pleural infection in the Second Multicentre Intrapleural Sepsis Trial (MIST2)(75), a randomised controlled trial investigating the effects of intrapleural tissue plasminogen activator (tPA) and DNase. 210 patients with pleural infection were recruited from 11 UK centres between December 2005 and November 2008. Eligibility criteria were clinical evidence of infection and either purulent pleural fluid, culture positive for bacterial infection, Gram staining positivity or fluid with pH <7.2. Baseline microbiological information included Gram staining results, aerobic and anaerobic pleural fluid culture, performed in the recruiting centre. At randomisation, pleural fluid was collected and transferred to the coordinating centre and frozen at -80°C, within 48 hours of being taken. The study was approved with NHS Research Ethics Committee number 04/MRE5/53 and all participants gave informed written

consent. Transudative control non-infected pleural fluid was also taken from 8 patients with heart failure and a low C-reactive protein, as part of the PIPAP study (see chapter 13).

12.2.2 Pleural fluid sample processing

Pleural fluid samples were processed as detailed in chapter 11.

12.3 Results

12.3.1 Patient characteristics

172 pleural fluid samples were available from 210 patients enrolled in the MIST2 study. Table 12-1 describes the characteristics of these 172 patients and pleural fluid samples.

Characteristic	Value
Age – years ±SD	59±18
Male sex – no. (%)	122 (71)
Duration of symptoms before randomisation – days	
Median	14
Interquartile range	7-29
Community-acquired infection – no. (%)	150 (87)
Purulent pleural fluid – no. (%)	83 (50)
Positive Gram's stain – no. (%)	7 (4)
Positive culture of pleural fluid – no. (%)	52 (30)
Pleural fluid pH	
Median	6.9
Interquartile range	6.8-7.1
Pleural fluid lactate dehydrogenase (IU/L)	
Median	2701
Interquartile range	867-6705
Patient white cell count (x 10^9/L)	
Median	13.8
Interquartile range	10-19.5
Patient C-reactive protein (mg/L)	
Median	>160
Interquartile range	118->160

Table 12-1 – Patient characteristics of 172 patients from the MIST2 study for whom pleural fluid was available for analysis

12.3.2 Predictors of bacterial load in pleural fluid samples

qPCR was used to estimate the number of copies of 16S rRNA gene (a surrogate of 'bacterial load') for each sample (see chapter 11). Table 12-2 shows these results, tabulated against patient and pleural fluid characteristics. Pleural fluid pH, culture status, appearance, LDH and glucose were all predictive of bacterial load. Patient C-reactive protein (CRP) and white cell count (WCC) were not significantly associated with bacterial load.

	Number	Copies of 16S rRNA gene, measured by qPCR				
		% change per unit[1]	95% CI	Geometric mean	95% CI	*P* value
PF pH	97	-89.6	-97.6, -55.5			0.003
PF culture status						<0.0001
Negative	110			2.8×10^5	1.9×10^5, 4.0×10^5	
Positive	44			2.2×10^6	1.2×10^6, 4.0×10^6	
PF appearance						<0.0001
Non-purulent	74			1.9×10^5	1.2×10^5, 2.9×10^5	
Purulent	76			1.4×10^6	8.8×10^5, 2.2×10^6	
PF LDH						0.007
≤1000	26			1.4×10^5	6.9×10^4, 2.8×10^5	
1000-5000	41			4.1×10^5	2.3×10^5, 7.1×10^5	
>5000	36			6.2×10^5	3.4×10^5, 1.1×10^6	
PF glucose						0.007
≤1.0	43			5.9×10^5	3.4×10^5, 1.0×10^6	
1.0-2.2	14			2.9×10^5	1.1×10^5, 7.3×10^5	
>2.2	35			1.5×10^5	8.5×10^4, 2.8×10^5	
Patient CRP						0.372
<100	28			3.5×10^5	1.6×10^5, 7.5×10^5	
100-160	27			3.5×10^5	1.6×10^5, 7.7×10^5	
≥160	87			5.8×10^5	3.7×10^5, 8.9×10^5	
Patient WCC						0.215
≤11.0	41			3.7×10^5	1.9×10^5, 7.0×10^5	
11.0-16.5	58			4.6×10^5	2.7×10^5, 7.9×10^5	
>16.5	53			7.7×10^5	4.4×10^5, 1.4×10^6	

[1] % change in 16S rRNA gene copies number per unit increase in the specified variable. CI = confidence interval. *P* values for tests of linear trend (continuous variables) and for tests of heterogeneity (categorical variables). LDH units – IU/L; glucose units – mmol/L; CRP units – mg/L; WCC units – x 10^9/L

Table 12-2 – Relationships between copies of 16S rRNA (base 10 logarithmic values) and characteristics of patients and pleural fluid samples

12.3.3 Ultra-deep sequencing – technical characteristics

PCR successfully generated MID-tagged 16S rRNA DNA amplicons for 169 samples, while three samples did not amplify the gene despite repeated attempts, including varying input DNA concentration to overcome potential PCR inhibition.

454 pyrosequencing was successful in all samples that generated a PCR amplicon, giving a mean of 4691 sequences (SD 2735) per sample, range 1082-28619. 8 non-infected control pleural fluid samples were similarly processed giving a mean of 3973 sequences (SD 873) per sample, range 2767-5182. These all generated weak amplicons that were also successfully sequenced.

Representative amplicon sequence reads were typically in excess of 450 nt. For the 90 representative sequences which represented the most abundant sequences for each Operational Taxonomic Unit (OTU) post contaminant filtering (see below), the median read length was 517 nt (range 484-535 nt), see figure 12-1.

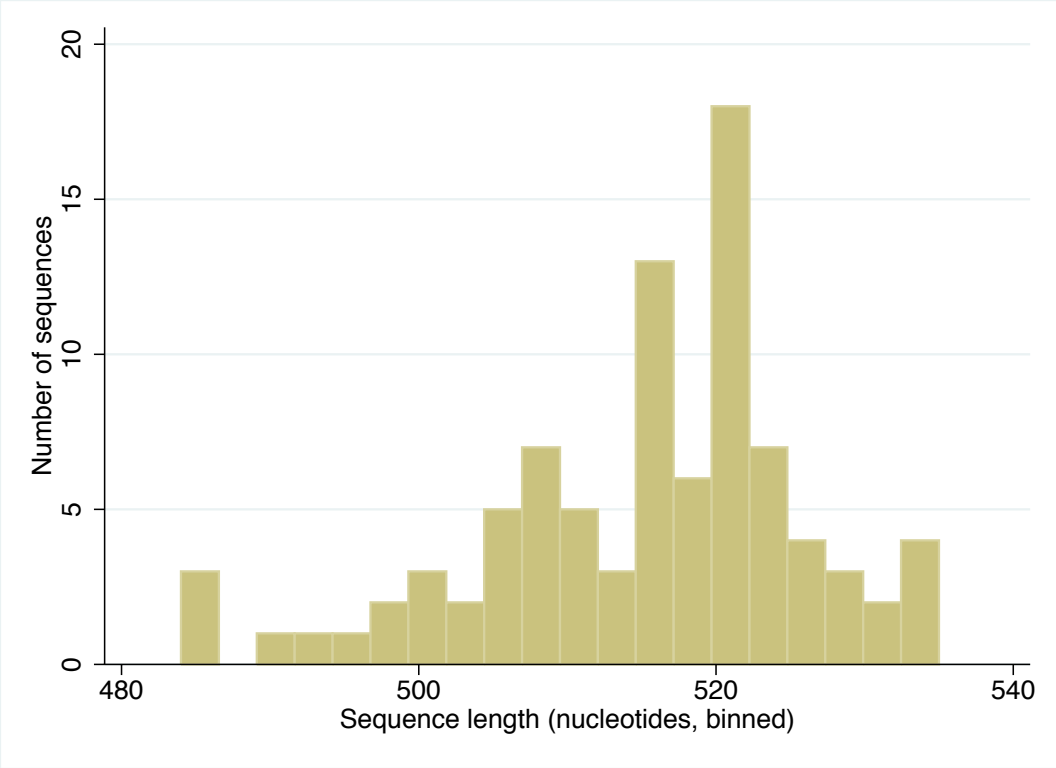


Figure 12-1 – Sequence lengths for the representative sequences of the 90 OTUs post contaminant filtering (discussed subsequently)

12.3.4 Sequencing results prior to contaminant filtering

The raw sequence data was processed as described in chapter 11, using QIIME and Phyloseq.

Figures 12-2 and 12-3 show linear and logarithmic abundance heatmap representations of the raw data, with samples and taxonomy ordered using NMDS, based on the Bray-Curtis distance. Each point on the x axis represents one sample, with its accompanying culture results displayed. The eight control non-infected pleural fluid samples are shown in red on the x axis. Sequences which match the spike DNA of *Acidothermus cellulolyticus* are shown (family Acidothermaceae, in red) and sequences of the most common contaminant in all

control pleural fluid samples, *Ralstonia* spp., are also shown (family Oxalobacteraceae, in red).

Figure 12-4 shows 3D Principal Coordinate Analysis (PCoA) plots, based on weighted UniFrac distances (a measure of phylogenetic distance between samples). A tight clustering of the eight control non-infected pleural fluid samples (red points) can be seen. Results from pleural infection samples show broad distribution of points but, significantly, some samples which show considerable overlap with the control samples. These pleural infection samples are essentially indistinguishable from control samples due to a low pathogen input DNA concentration (see subsequent section ‘Predictors of sequencing success’). Such findings highlight the importance of a data processing strategy to control for contamination.

Figure 12-2 – Heatmap of raw sequence data using a linear relative abundance scale. Oxalobacteraceae (*Ralstonia* spp.) and Acidothermaceae (*Acidothermus cellulolyticus* spike) are shown as are control non-infected pleural fluid samples.

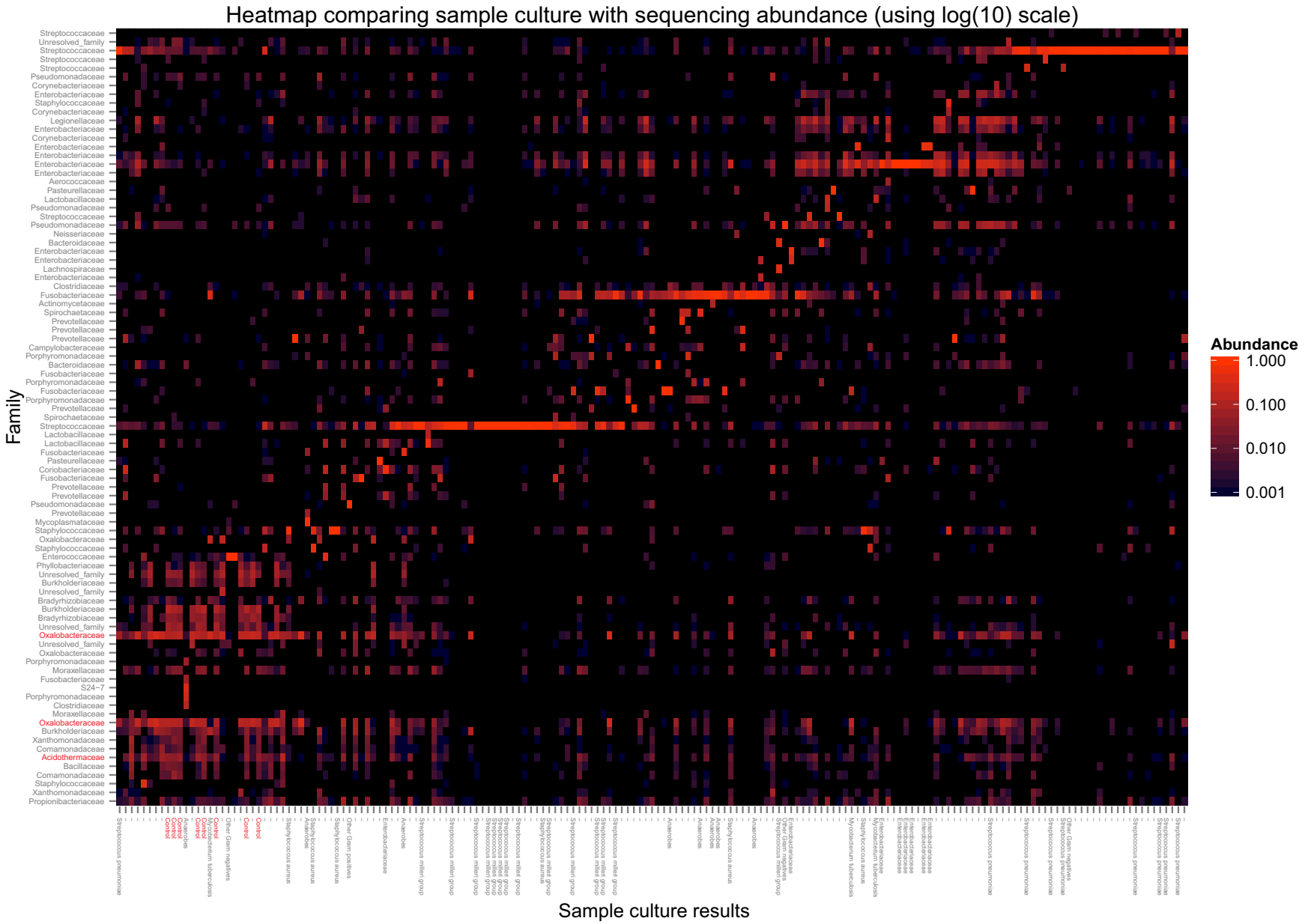


Figure 12-3 – Heatmap of raw sequence data using a logarithmic relative abundance scale. Oxalobacteraceae (*Ralstonia* spp.) and Acidothermaceae (*A. cellulolyticus* spike) are shown as are control non-infected pleural fluid samples.

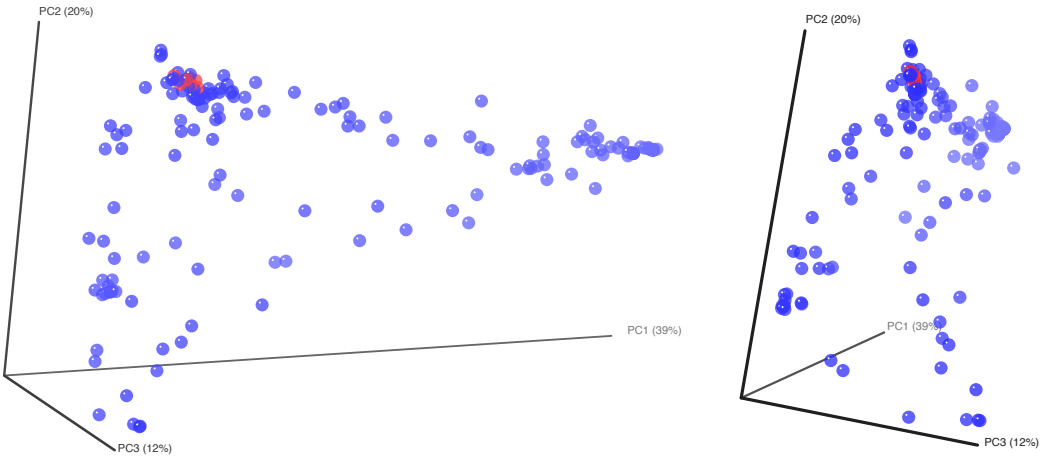


Figure 12-4 – 3D Principal Coordinate Analysis (PCoA) plots of raw sequence data using weighted UniFrac distance (a phylogenetic measure of distances between samples). Each coloured point represents one sample. Blue points represent pleural infection samples and red points represent control non-infected pleural fluid samples.

Figures 12-5 and 12-6 display species richness and diversity metrics for pleural infection and control pleural fluid samples with sequences rarefied at varying depths. Species richness in an estimate of the total number of species identified in a sample, and estimates of this include the Chao1 and Michaelis-Menten metrics. Species diversity is a function of both richness and evenness; the Shannon diversity metric (and its derivative the Brillouin index) both measure diversity with higher values representing more diverse samples(283).

Pleural infection samples have lower richness and diversity than control pleural fluid samples. This is an expected finding, representing the low grade contamination of control pleural fluid from a variety of species. Of note, however, all pleural fluid samples have *Ralstonia* spp. as a predominating contaminating bacteria.

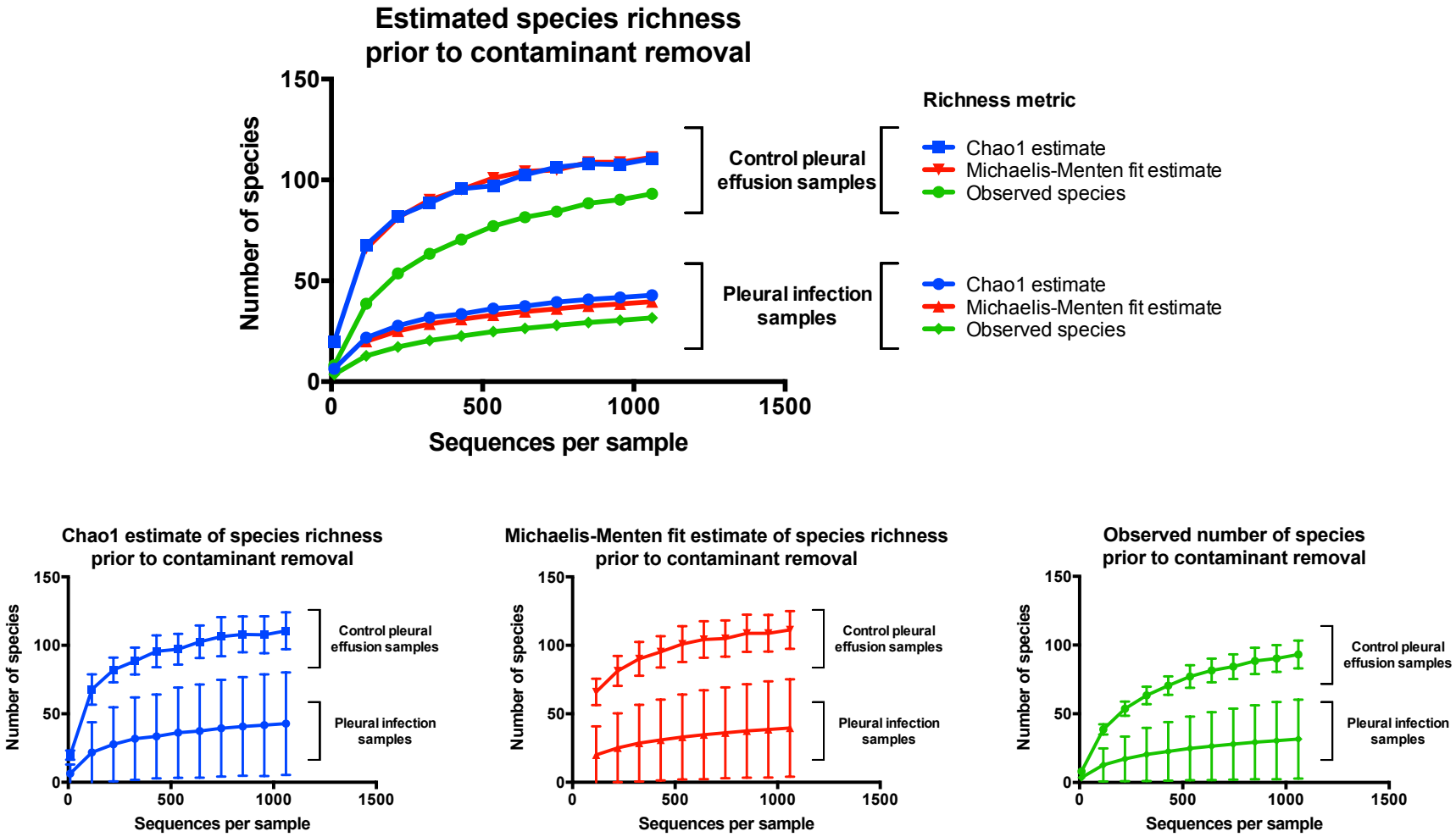


Figure 12-5 – Rarefaction curves for estimates of species richness for pleural infection samples vs. control pleural fluid samples. Standard deviation plotted in graphs 2-4. Rarefaction at 10 sequences omitted for Michaelis-Menten fit plot for clarity.

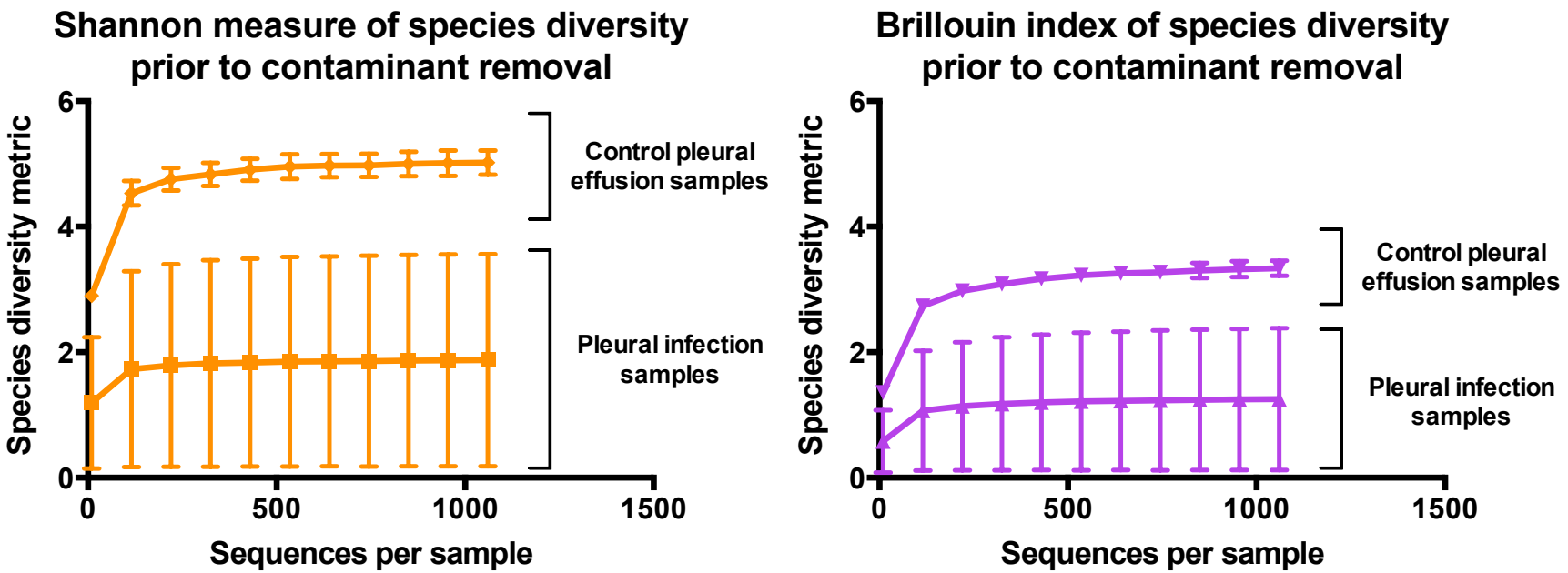


Figure 12-6 – Rarefaction curves for estimates of species diversity (with standard deviation) for pleural infection samples vs. control pleural fluid samples.

12.3.5 Contaminant removal

Contaminant sequence reads were filtered from sample reads using a threshold, below which sequences were discarded, as described in chapter 11. A threshold of 396 x (no. of *Acidothermus cellulolyticus* sequences in a sample) was used[8]. This resulted in 98 samples that had sequences passing threshold (figure 12-7).

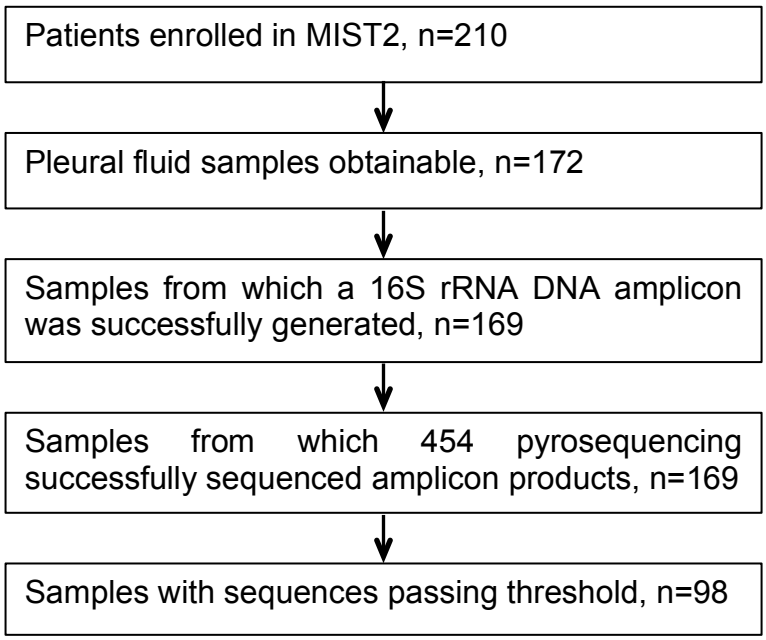


Figure 12-7 – Flow chart of pleural infection samples in MIST2

12.3.6 Predictors of sequencing success

Logistic regression was used to examine the relationship between pleural fluid/patient characteristics and subsequent sequencing success (table 12-3). Pleural fluid culture positivity, purulence, higher lactate dehydrogenase (LDH), lower glucose, lower pH and increased number of copies of 16S rRNA gene were all associated with a higher odds ratio of sequencing success. Patient

[8] Ratio of *Acidothermus cellulolyticus* to *Ralstonia* spp. mean = 52.9, standard deviation = 171.7 giving a threshold of 52.9 + (2 x 171.7) = 396.

characteristics such as WCC and CRP were not significantly associated with sequencing success.

	Sequencing successful, number	Sequencing unsuccessful, number	Odds ratio of sequencing success	95% CI	*P* value
PF pH (0.1 unit decrease)	47	62	4.14	0.48, 35.42	0.001
PF culture status					<0.0001
Negative	54	66	1.00	referent	
Positive	44	8	6.72	2.92, 15.49	
PF appearance					<0.0001
Non-purulent	33	51	1.00	referent	
Purulent	64	19	5.21	2.65, 10.21	
PF LDH					0.003
≤1000	10	20	1.00	referent	
1001-5000	23	20	2.30	0.87, 6.05	
>5000	28	10	5.60	1.96, 15.97	
PF glucose					0.036
≤1.0	29	17	3.15	1.28, 7.76	
1.1-2.2	9	7	2.37	0.72, 7.85	
>2.2	13	24	1.00	referent	
Patient CRP					0.403
<100	14	17	1.00	referent	
100-160	20	14	1.73	0.65, 4.64	
≥160	55	39	1.71	0.76, 3.88	
Patient WCC					0.103
≤11.0	21	26	1.00	referent	
11.0-16.5	37	26	1.76	0.82, 3.78	
>16.5	39	21	2.30	1.05, 5.03	
Tenfold increase in copies of 16S rRNA gene	89	65	350.47	42.36, 2899.83	<0.0001

LDH units – IU/L; glucose units – mmol/L; CRP units – mg/L; WCC units – x 10^9/L

Table 12-3 – Association between sequencing success and patient and pleural fluid characteristics

12.3.7 Sequencing results post contaminant removal

12.3.7.1 Overall bacteriology

The heatmaps in figures 12-8 and 12-9 display the bacteriology of 98 pleural infection samples post contaminant removal. The following bacteriologically-distinct patterns are evident;

1. Fusobacteriaceae (predominantly *Fusobacterium nucleatum*)
2. Prevotellaceae (predominantly *Prevotella oris*)
3. *Streptococcus pneumoniae*
4. Streptococcus 'milleri' group (*Streptococcus anginosus-constellatus-intermedius*)
5. *Staphylococcus aureus*
6. Enterobacteriaceae (predominantly *Escherichia coli*)

Figure 12-10 displays 3D Principal Coordinate Analysis (PCoA) plots using weighted UniFrac distance (a phylogenetic measure of distances between samples), with sample colouring determined by culture results. Figure 12-11 displays the overall bacteriology of the 172 samples, including identification of sequencing negative samples. Table 12-4 displays the frequency of each bacterial grouping on the basis of predominant sequence taxonomic assignment.

Predominant sequence taxonomic assignment	**Detection frequency as % of all samples (n=172)**	**Detection frequency as % of all samples successfully sequenced (n=98)**
Anaerobes	16.3	28.6
Streptococcus pneumoniae	14.5	25.5
Streptococcus 'milleri' group	13.4	23.5
Enterobacteriaceae	6.4	11.2
Staphylococcus aureus	2.3	4.1
Mycoplasma salivarium	0.6	1.0
Other Gram negatives	2.3	4.1
Other Gram positives	1.2	2.0

Table 12-4 – Frequency of bacterial groups by taxonomic assignment of predominant sequence

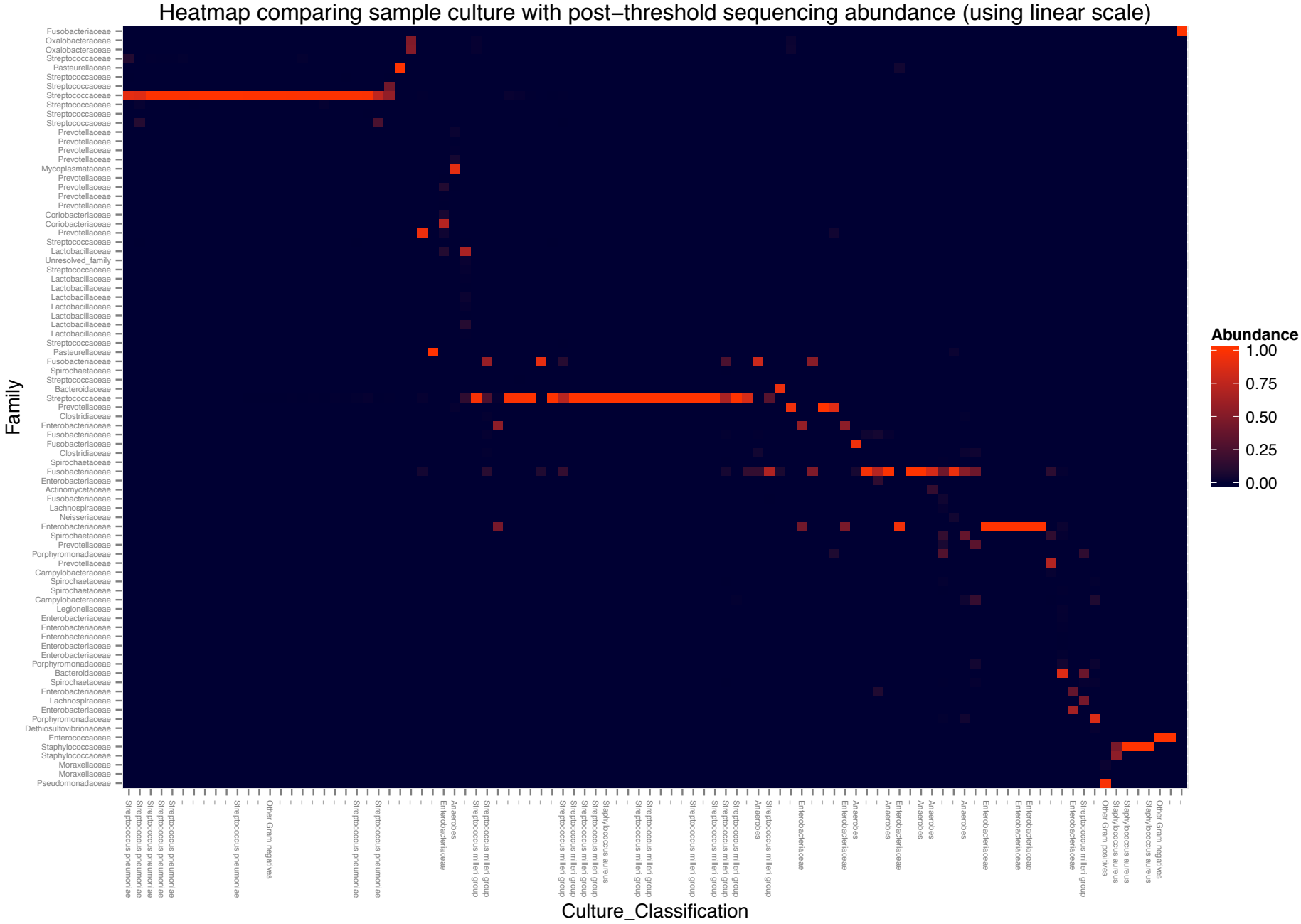


Figure 12-8 – Heatmap of sequence data post contaminant removal using a linear relative abundance scale

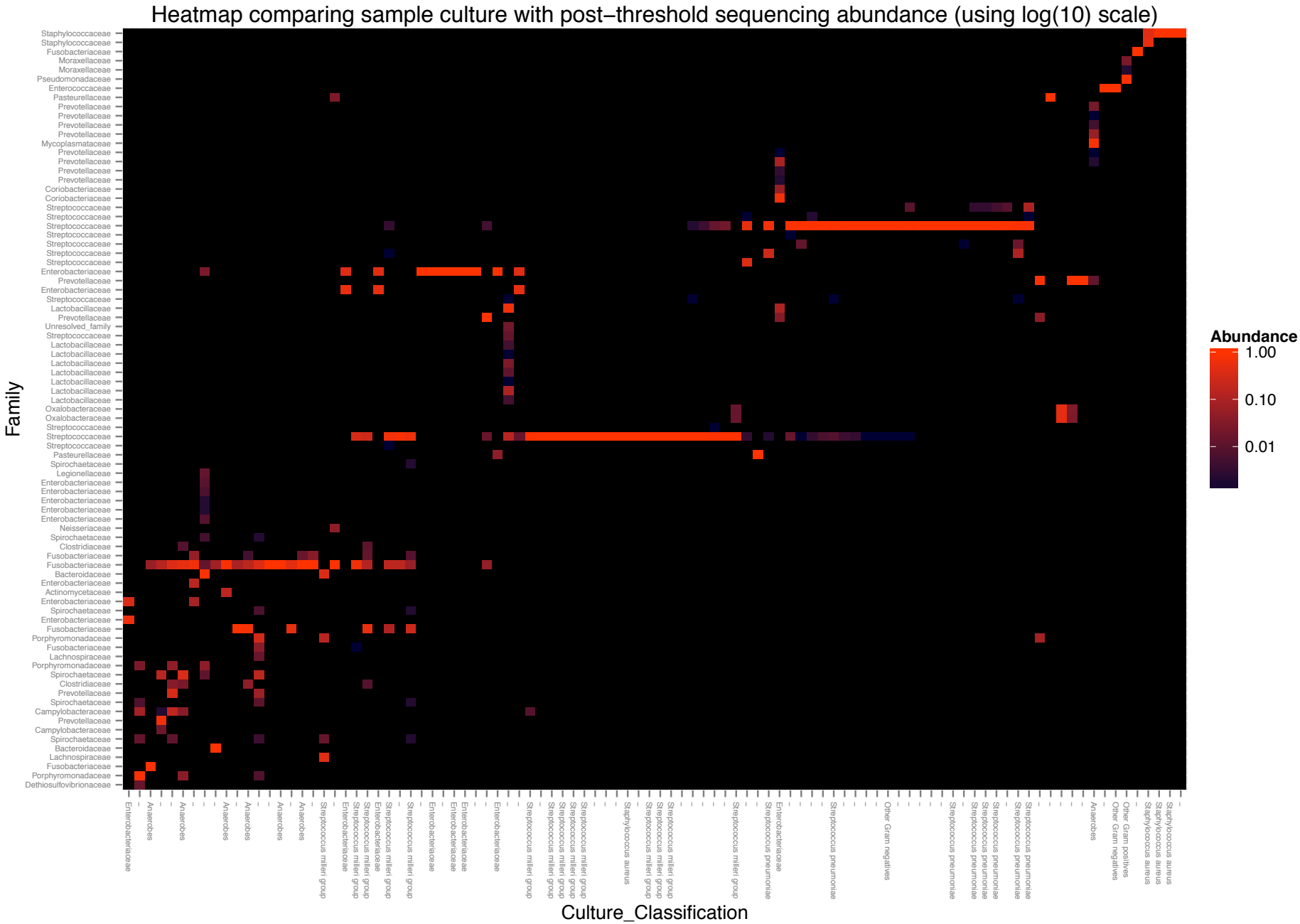


Figure 12-9 – Heatmap of sequence data post contaminant removal using a logarithmic relative abundance scale

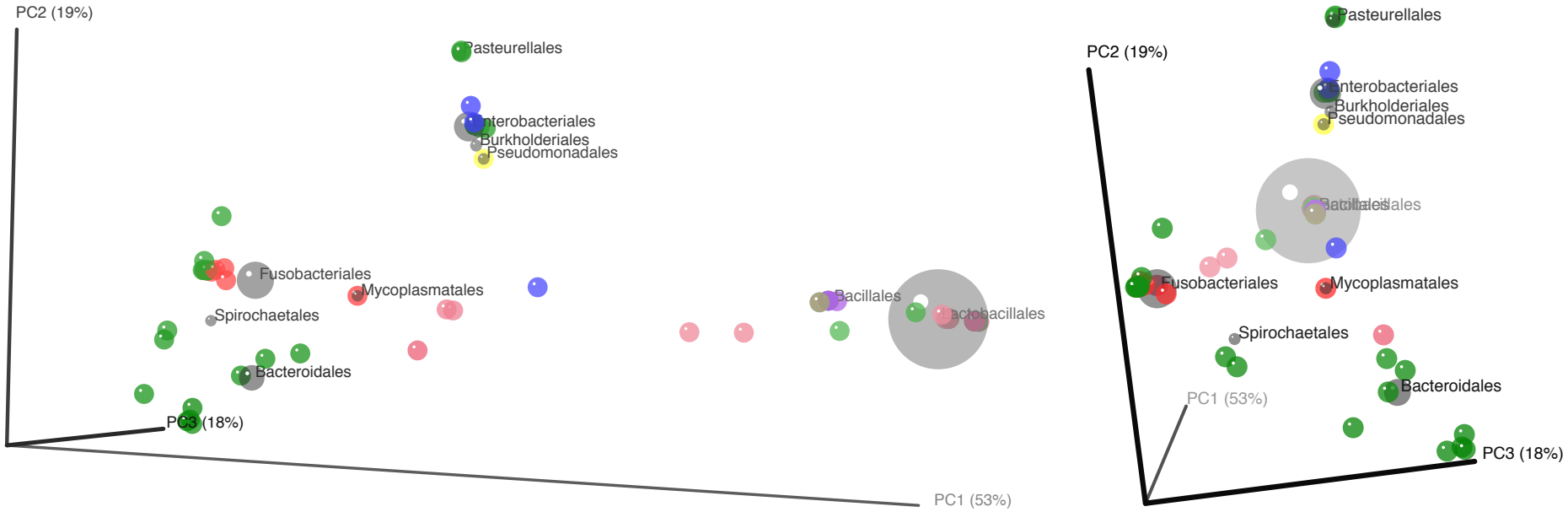


Figure 12-10 – 3D Principal Coordinate Analysis (PCoA) plots using weighted UniFrac distance (a phylogenetic measure of distances between samples). Each coloured point represents one sample, the colour of which is determined by culture results (green = no growth; red = anaerobes; pink = Streptococcus milleri group; maroon = *Streptococcus pneumoniae*; purple = *Staphylococcus aureus*; blue = Enterobacteriaceae; yellow = other Gram positives; brown = other Gram negatives). The ten most abundant Orders are plotted as a weighted average of the coordinates of all samples, where the weights are the relative abundances of the Order in the samples. Sphere size is proportional to the mean relative abundance of the Order across all samples.

Overall bacteriology of 172 pleural fluid samples: Results of pyrosequencing compared with culture

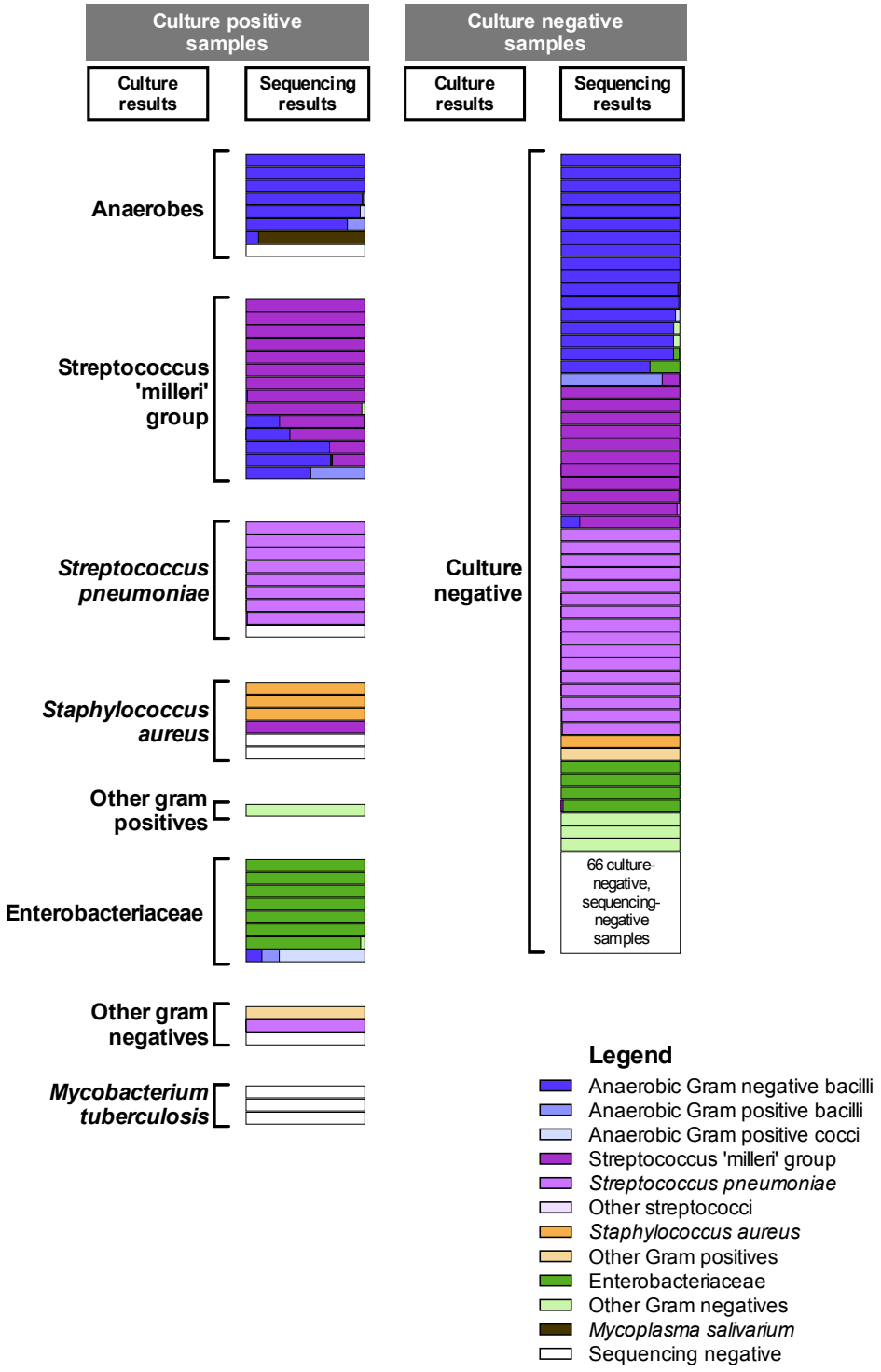


Figure 12-11 – Sequencing and culture results for 172 individual pleural fluid samples from the MIST2 study

12.3.7.2 Culture negative samples

Table 12-5 compares the performance of sequencing against culture status, demonstrating significantly more success with sequencing when culture is positive ($p<0.001$).

	Culture negative	Culture positive
Sequencing negative	66	8
Sequencing positive	54	44

$p<0.001$ by chi-square test ($\chi 2$ [1 df] = 23.2)

Table 12-5 – comparison of overall performance of sequencing versus culture status

Figure 12-12 depicts a PCoA plot of culture-negative sequencing-positive samples. Comparison of this figure with figure 12-10 suggests that sequencing is increasing detection rates across all pathogens (i.e. without marked bias). Table 12-6 compares the predominant sequence taxonomy for culture positive and culture negative samples, finding no significant differences between bacterial distribution patterns ($p=0.401$).

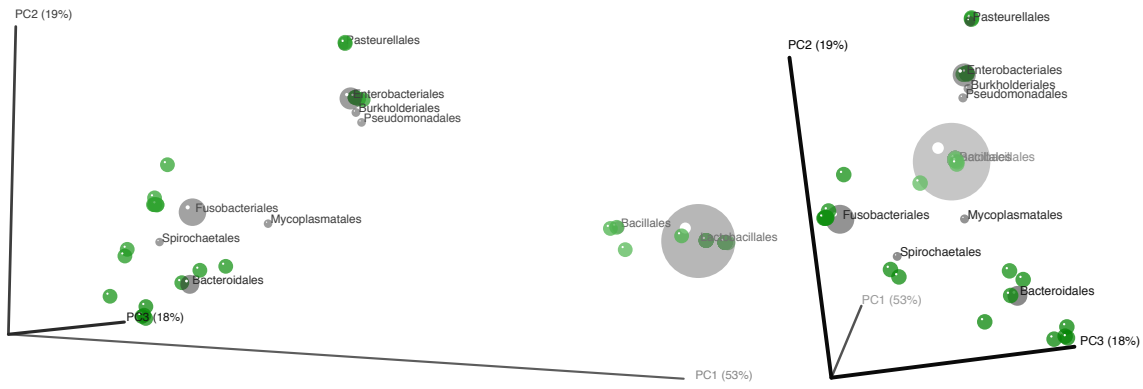


Figure 12-12 – 3D PCoA plot of culture negative, sequencing positive samples. A broad distribution of sample bacteriology is evident.

		Culture negative	Culture positive	Total
Predominant sequence	Anaerobes	18 (33)	10 (23)	28 (29)
	Enterobacteriaceae	4 (7)	7 (16)	11 (11)
	Mycoplasma salivarium	0 (0)	1 (2)	1 (1)
	Other Gram negatives	3 (6)	1 (2)	4 (4)
	Other Gram positives	1 (2)	1 (2)	2 (2)
	Staphylococcus aureus	1 (2)	3 (7)	4 (4)
	Streptococcus 'milleri' group	11 (20)	12 (27)	23 (23)
	Streptococcus pneumoniae	16 (30)	9 (20)	25 (26)
	Total	54 (100)	44 (100)	98 (100)

p=0.401 by Fisher's exact test

Table 12-6 – Comparison of predominant sequence and culture status

12.3.7.3 Agreement between culture and sequencing results

Figures 12-8 to 12-11 demonstrate broad agreement between culture results and sequencing results. This broad agreement is confirmed in the tabulated comparison of table 12-7.

		Culture results										
		Anaerobes	**Enterobacteriaceae**	**Other Gram negatives**	**Other Gram positives**	***Staphylococcus aureus***	**Streptococcus 'milleri' group**	***Streptococcus pneumoniae***	***Mycobacterium tuberculosis***	***Mycoplasma salivarium***	**No growth**	**Total**
Predominant sequence	**Anaerobes**	6	1	0	0	0	3	0	0	0	18	28
	Enterobacteriaceae	0	7	0	0	0	0	0	0	0	4	11
	Other Gram negatives	0	0	0	1	0	0	0	0	0	3	4
	Other Gram positives	0	0	1	0	0	0	0	0	0	1	2
	Staphylococcus aureus	0	0	0	0	3	0	0	0	0	1	4
	Streptococcus 'milleri' group	0	0	0	0	1	11	0	0	0	11	23
	Streptococcus pneumoniae	0	0	1	0	0	0	8	0	0	16	25
	Mycobacterium tuberculosis	0	0	0	0	0	0	0	0	0	0	0
	Mycoplasma salivarium	1	0	0	0	0	0	0	0	0	0	1
	Sequencing negative	1	0	1	0	2	0	1	3	0	66	74
	Total	8	8	3	1	6	14	9	3	0	120	172

Table 12-7 – Comparison of taxonomy of predominant sequence versus culture results

Comparison of sequencing results with culture results is hampered by a lack of a gold standard given that culture is frequently not positive in pleural infection, precluding an easy assessment of sensitivity and specificity. However, it is possible to compare sequencing with a non-referent standard (culture results) to calculate estimates of agreement using methods suggested by the FDA(284).

Multiple 2x2 comparisons, comparing sequencing and culture results for bacterial classes are calculated, e.g. for *S. pneumoniae*;

		Streptococcus pneumoniae on culture	
		Yes	No
Streptococcus pneumoniae predominant sequencing type	Yes	a	b
	No	c	d
	Total	a+c	b+d

Positive Percent Agreement (PPA) = 100% x a/(a+c)

Negative Percent Agreement (NPA) = 100% x d/(b+d)

Table 12-8 tabulates the PPA and NPA for the common bacteriological groups. Note that Anaerobes and Enterobacteriaceae have been compared as a group (rather than at species level) as culture reports may only provide resolution to the level of 'anaerobes' and 'coliforms'.

Bacteriological class	PPA	NPA
Anaerobes	75%	87%
Enterobacteriaceae	88%	98%
Staphylococcus aureus	50%	99%
Streptococcus 'milleri' group	79%	92%
Streptococcus pneumoniae	89%	90%

Table 12-8 – Positive and negative percent agreements (PPA and NPA) comparing sequencing with culture results

12.3.7.4 Pleural infection diversity and richness metrics

Examination of the linear scale heatmap (figure 12-8) and figure 12-11 reveals that pleural infection is usually dominated by one bacteria on sequencing. However, the logarithmic scale heatmap (figure 12-9) shows that many samples are polymicrobial, particularly those that are anaerobic. Richness metrics (figure 12-13) suggest 1-2 operational taxonomic units (OTUs) for *Streptococcus pneumoniae*, Streptococcus 'milleri' group, *Staphylococcus aureus* and Enterobacteriaceae. Anaerobic pleural infections have significantly higher richness, and are estimated to have 4-5 OTUs per sample. Diversity indices suggest a similar pattern.

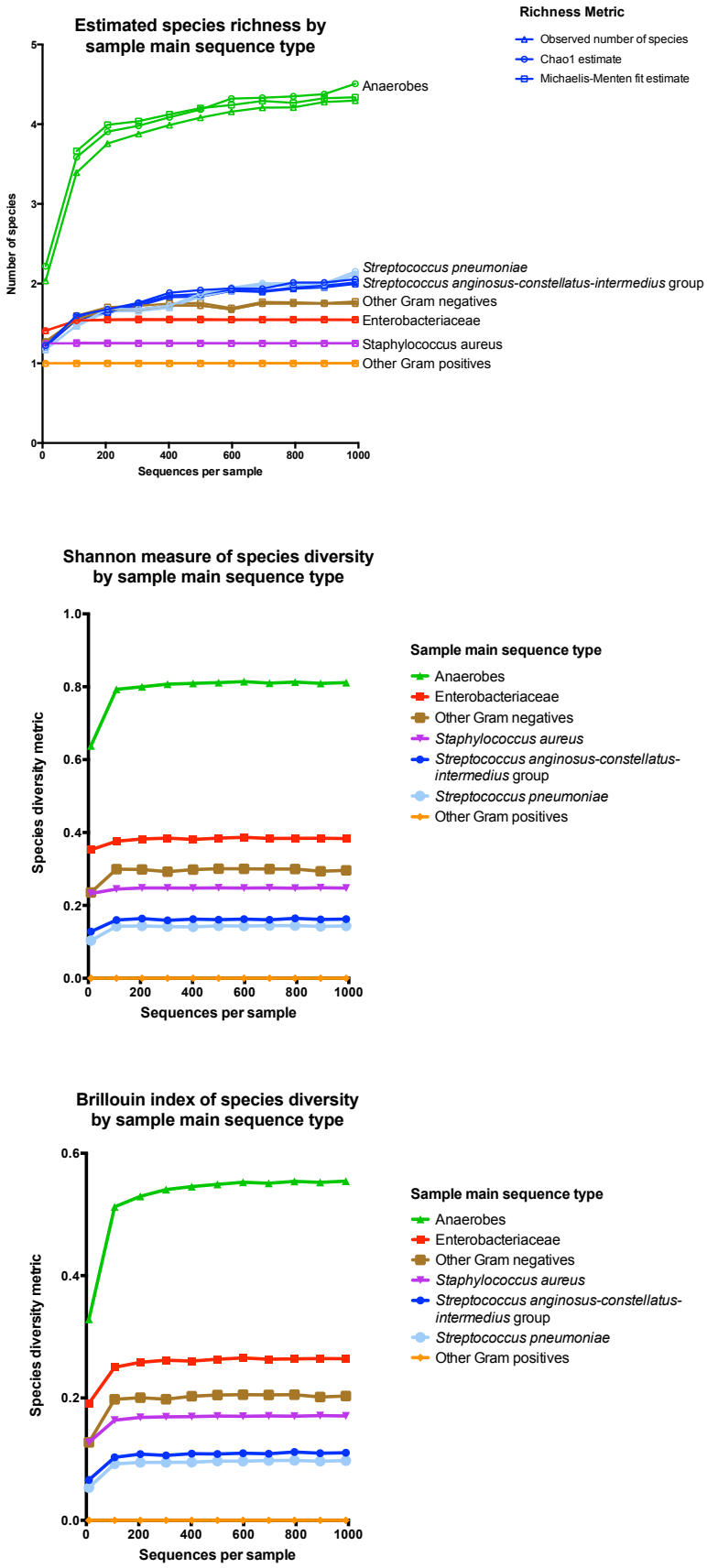


Figure 12-13 – Species diversity and richness by taxonomy of predominant sequence

12.3.8 Analysis of common bacterial groups in pleural infection

12.3.8.1 Anaerobes

Anaerobes were seen as the predominant sequence taxonomy in 28 out of 98 successfully sequenced samples and at ≥10% of individual sample sequences in 32 out of 98 samples. Figures 12-14 shows a polar phylogenetic plot of sequencing results when anaerobes were the main sequence taxonomy found. Figures 12-15 and 12-16 are 3D PCoA plots for samples that have anaerobes on culture and as the main sequence taxonomy respectively.

Fusobacteriales, particularly *Fusobacterium nucleatum*, and Bacteroidales, particularly *Prevotella* spp. were commonly found although other anaerobes were found from the following Families: Actinomycetaceae, Bacteroides, Campylobacteraceae, Clostridiaceae, Coriobacteriaceae, Lachnospiraceae, Porphyromonas, Spirochaetaceae and Synergistetes.

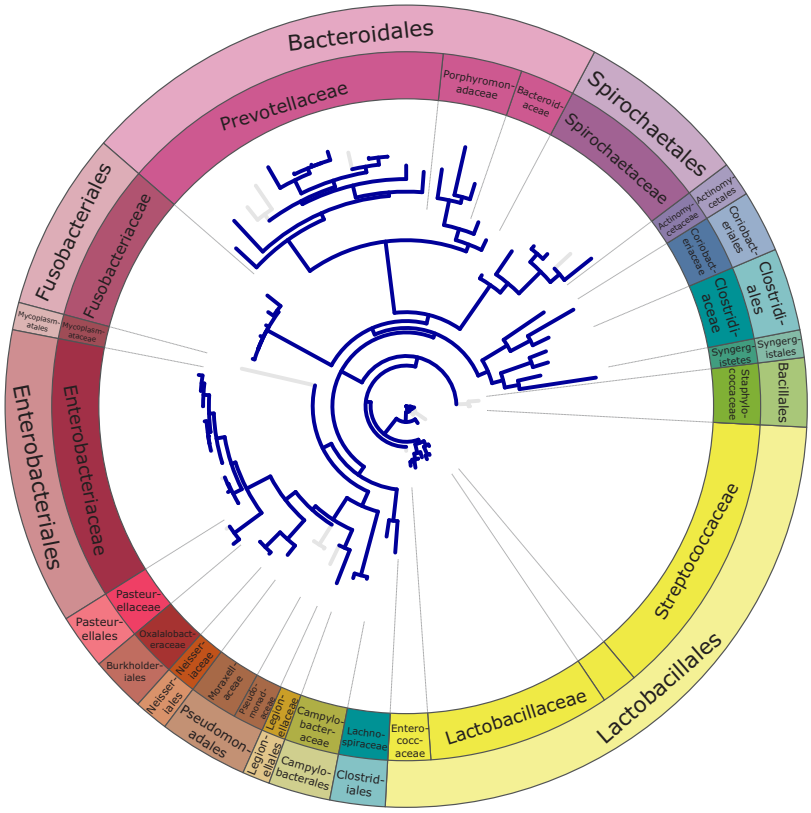


Figure 12-14 – Polar phylogenetic plot for samples that had an anaerobe as the predominant sequence taxonomy

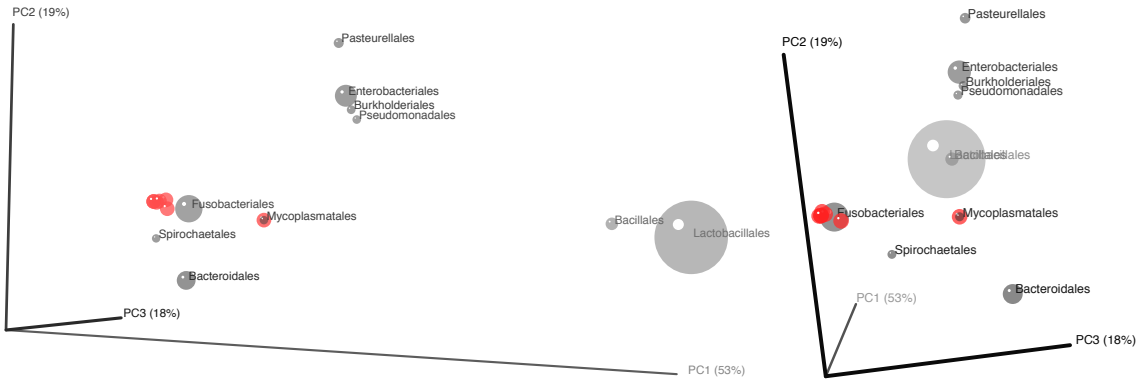


Figure 12-15 – 3D PCoA plots for samples that cultured anaerobic bacteria

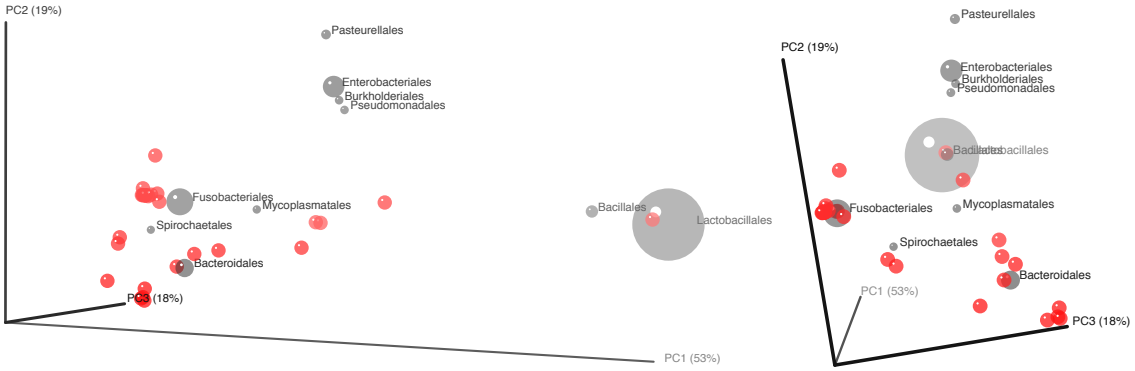


Figure 12-16 – 3D PCoA plots for samples that had an anaerobe as the predominant sequence taxonomy

The length of representative OTU sequences allowed taxonomic assignment to species level for the vast majority of sequences. Table 12-9 and figure 12-17 illustrate these results for samples that had anaerobes accounting for at least 10% of all sequence reads.

Fusobacterium nucleatum was the most common anaerobe seen, being found in 24 of the 32 anaerobic samples sequenced. *Prevotella oris* was found in 6/24 samples, as was *Treponema socranksii.*

Polymicrobiality was the norm for the majority of anaerobic samples. *Fusobacterium nucleatum* was commonly seen together with Streptococcus 'milleri' group bacteria and also with *Prevotella* spp. and *Porphyromonas* spp.

Many species were found that have not been previously documented in pleural infection (table 12-9), such as;

- *Atopobium rimae*
- *Cryptobacterium curtum*
- *Lactobacillus gasseri/taiwanensis*

- *Lactobacillus iners*
- *Stomatobaculum* spp.
- *Oribacterium* spp.
- *Prevotella baroniae*
- *Prevotella dentalis/Hallella seregens*
- *Prevotella scopos*
- *Fretibacterium* spp.
- *Tanerella forsythia*
- *Treponema denticola*
- *Treponema lecithinolyticum*
- *Treponema maltophilum*
- *Treponema medium*
- *Treponema socranskii*

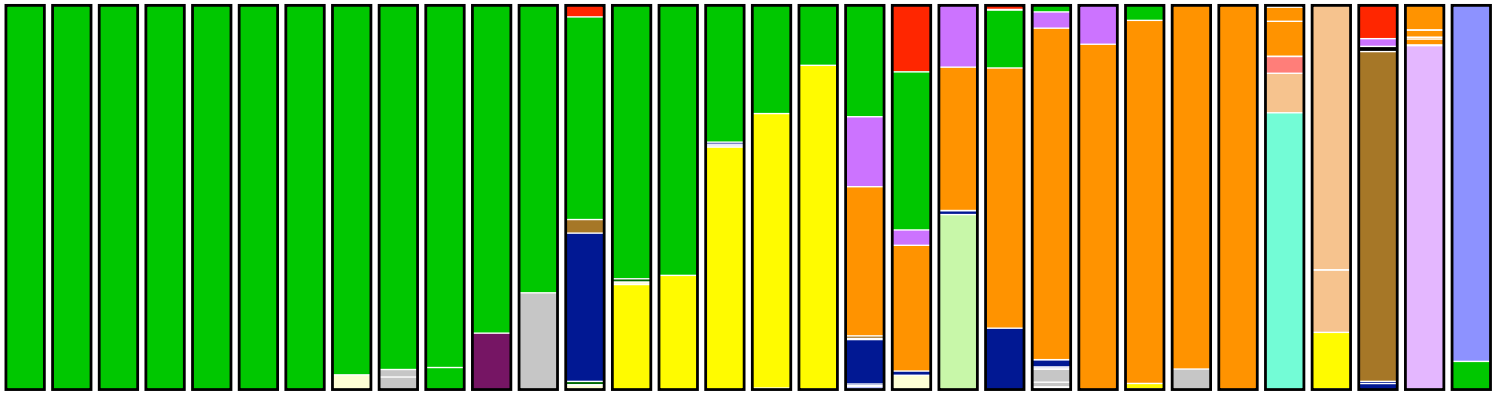
Table 12-9 further documents the source of each anaerobe's type strain – almost all of these were originally found in the oral cavity, particularly in gingival crevices. Notably, five sequences had closest homology to GenBank/Ribosomal Database Project sequences for hitherto unnamed species (e.g. *Stomatobaculum* sp. oral taxon 373, *Oribacterium* sp. oral taxon 078, *Prevotella* sp. clone 303A09, *Fretibacterium* sp. oral taxon 360 and *Treponema* sp. oral taxon 270). These sequences were all isolated from oral (particularly gingival crevice) bacteria.

Anaerobic species	Classification	Pairwise identity (%)	Maximum proportion in one sample (%)	No. of samples containing anaerobe	Type strain		Pleural infection	
					Source	References	Cited cases[9] (+ 1-10; ++ >10)	References
Actinomyces meyeri	GPAB	99.4	15	1	Pleural effusion; principal habitat is human periodontal sulci	Cato(285)	+	Fazili(286)
Atopobium rimae	GPAB	99.6	72	1	Gingival crevices	Olsen(287)	-	NA
Bacteroides fragilis	GNAB	99.8	93	1	Appendix abscess	Skerman(288)	++	Bartlett(289-291); Boyanova(292); Brook(101); Civen(102); De(293); Finegold(294)
Campylobacter gracilis	GNAB	100	1	1	Gingival crevices	Tanner(295)	++	Boyanova(292); Civen(102); Johnson(296); Lee(297)
Campylobacter rectus/showae	GNAB	99.6	17	4	Gingival crevices	Tanner(295); Etoh(298)	+	Lam(299); Spiegel(300)
Cryptobacterium curtum	GPAB	100	4	1	Gingival crevices	Nakazawa(301)	-	NA
Eubacterium brachy	GPAB	99.6	0.9	2	Gingival crevices	Holdeman(302)	+	Civen(102); Rochford(303)
Eubacterium saphenum	GPAB	100	0.02	1	Gingival crevices	Uematsu(304)	+	Civen(102)
Fusobacterium necrophorum	GNAB	99.2	94	1	Not stated	Skerman(288)	++	Bartlett(290); Boyanova(292); Finegold(294)
Fusobacterium nucleatum	GNAB	100	100	24	Cervico-facial lesion	Skerman(288)	++	Bartlett(289-291); Boyanova(292); Civen(102); De(293); Finegold(294); Kawanami(282)
Lactobacillus gasseri/taiwanensis	GPAB	99.4	69	2	*L. gasseri* – human source; *L. taiwanensis* – silage	Lauer(305); Wang(306)	-	NA
Lactobacillus iners	GPAB	100	0.03	1	Urine, vagina and medical care products	Falsen(307)	-	NA
Lactobacillus paracasei/casei/zeae	GPAB	99.8	16	1	*L. casei* and *paracasei* – dairy products; *L. zeae* – corn steep liquor	Hansen(308); Collins(309); Dicks(310)	+	Gouriet(311); Shoji(312)
Stomatobaculum sp. oral taxon 373	GPAB	100	46	1	Endodontic infections	Munson(313)	-	NA
Mycoplasma salivarium	Other	100	90	1	Saliva	Edward(314)	+	Baracaldo(227)
Oribacterium sp. oral taxon 078	GPAB	100	0.9	1	Gingival crevice in 2 patients with refractory periodontitis and 2 healthy subjects	Paster(315)	-	NA
Parvimonas micra	GPAC	98.7	4	4	Purulent pleurisy. Stated to be part of normal mouth flora.	Murdoch(316)	++	Boyanova(292); Civen(102); Kawanami(282)
Porphyromonas endodontalis	GNAB	99.8	18	3	Infected root canal	Shah(317)	+	Kawanami(282)
Porphyromonas gingivalis	GNAB	99.8	4	3	Gingival sulcus	Shah(317)	+	Civen(102); Kawanami(282)
Prevotella baroniae	GNAB	100	6	1	Dental plaque and endodontic and periodontal infections	Downes(318)	-	NA

[9] NCBI PubMed search strategy = ($GENUS OR $BASONYM OR $SYNONYM) AND (parapneumonic OR pleura* OR pleuri* OR empyema)

Prevotella buccae	GNAB	99.8	0.04	2	Gingival sulcus, moderate periodontitis	Holdeman(319)	+	Boyanova(292); Civen(102)
Prevotella dentalis/Hallella seregens	GNAB	100	2	1	Dental root canals	Willems(320)	-	NA
Prevotella denticola	GNAB	100	0.2	1	Dental plaque	Shah(321)	+	Civen(102)
Prevotella heparinolytica	GNAB	99.8	86	2	Periodontal sites	Okuda(322)	+	Civen(102)
Prevotella loescheii	GNAB	98.3	39	2	Gingival crevices	Holdeman(323)	+	Civen(102); Kawanami(282)
Prevotella melaninogenica	GNAB	98.9	0.2	2	Gingival crevices and human clinical specimens	Holdeman(323)	++	Bartlett(289-291); Boyanova(292); Brook(101); Civen(102); De(293); Finegold(294); Petty(324)
Prevotella nigrescens	GNAB	99.8	0.4	1	Vincent's gingivitis	Shah(325)	+	Civen(102); Matto(326)
Prevotella oris	GNAB	99.7	100	6	Gingival sulcus, moderate periodontitis	Holdeman(319)	++	Boyanova(292); Civen(102); Kawanami(282)
Prevotella sp. (clone 303A09)	GNAB	99.8	9	2	Subgingival plaque	Genbank: AM420097. Bolivar(327)	-	NA
Prevotella pleuritidis	GNAB	98.5	68	1	Pleural fluid	Sakamoto(328)	+	Sakamoto(328)
Prevotella scopos	GNAB	98.2	0.04	1	Exudate associated with failing dental implant	Downes(329)	-	NA
Fretibacterium sp. oral taxon 360 (Clone BH017)	GNAB	100	1	1	Oral cavity	Genbank: AF125199. Paster(315)	-	NA
Tannerella forsythia	GNAB	100	86	3	Advanced periodontal lesions	Tanner(330)	-	NA
Treponema denticola	GNAB	99.8	0.5	3	Periodontal pockets	Chan(331)	-	NA
Treponema lecithinolyticum	GNAB	99.8	39	4	Deep periodontal lesions	Wyss(332)	-	NA
Treponema maltophilum	GNAB	99.8	0.6	4	Subgingival plaque in periodontal disease	Wyss(333)	-	NA
Treponema medium	GNAB	99.6	0.3	3	Subgingival plaque in periodontitis	Umemoto(334)	-	NA
Treponema socranskii	GNAB	99.8	2	6	Gingival crevices and sulci	Smibert(335)	-	NA
Treponema sp. oral taxon 270 (Clone DD012)	GNAB	100	0.3	1	Dental plaque	Genbank: GQ422733. Dewhirst(336)	-	NA

Table 12-9 – Analysis of sequences corresponding to anaerobes in samples where anaerobes represent at least 10% of sequences



Legend

Anaerobic Gram negative bacilli

Order	Species
Fusobacteriales	*Fusobacterium* spp.
Bacteroidales	*Bacteroides fragilis*
	Porphyromonas spp.
	Prevotella spp.
	Tannerella forsythia
Campylobacterales	*Campylobacter* spp.
Synergistales	*Fretibacterium* spp.
Spirochaetales	*Treponema* spp.

Anaerobic Gram positive bacilli

Order	Species
Actinomycetales	*Actinomyces meyeri*
Coriobacteriaceae	*Atopobium rimae*
	Cryptobacterium curtum
Lactobacillales	*Lactobacillus* spp.
Clostridiales	*Eubacterium* spp.
	Stomatobaculum spp.
	Oribacterium spp.

Anaerobic Gram positive cocci

Clostridiales	*Parvimonas micra*

Other

- Streptococcus 'milleri' group
- *Mycoplasma salivarium*
- Other (non-anaerobes)

Figure 12-17 – Individual bacteriology of all samples that had anaerobes account for at least 10% of sequence reads

12.3.8.2 Streptococcus ‘milleri’ group

Streptococcus ‘milleri’ group were seen as the predominant sequence taxonomy in 23 out of 98 successfully sequenced samples and at ≥10% of individual sample sequences in 26 out of 98 samples (with mean proportion in these samples of 87.8% (SD 25.3)). This was most frequently detected as a monomicrobial infection although it was also seen in association with other bacteria, such as anaerobes (particularly *Fusobacterium nucleatum*, but also *Lactobacillus* spp., *Treponema* spp. and *Campylobacter* spp.) (see figures 12-18 to 12-21).

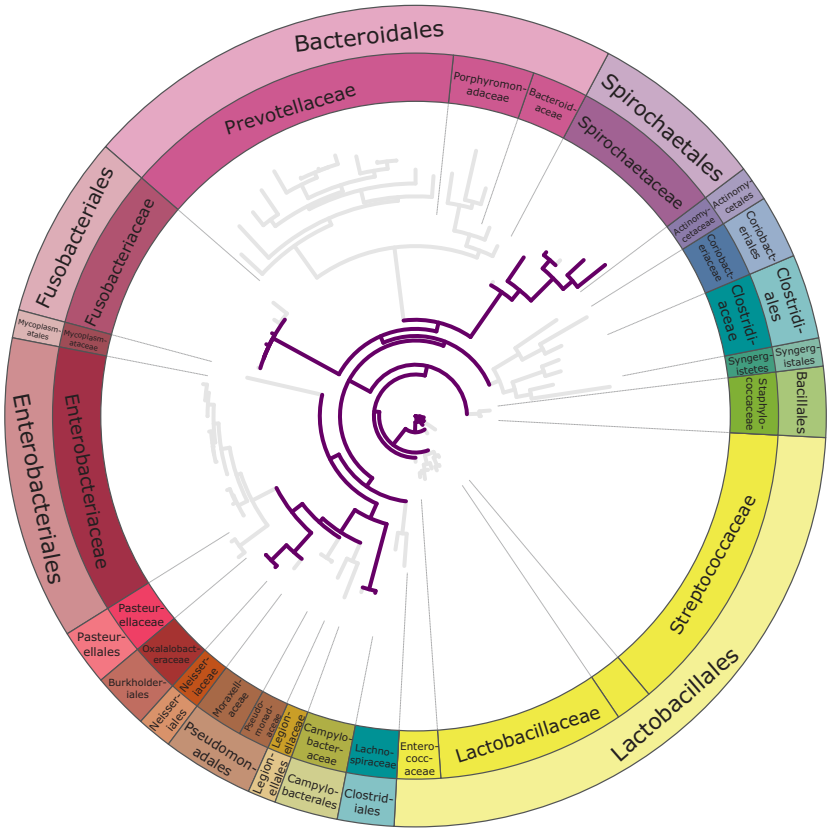


Figure 12-18 – Polar phylogenetic plot for samples that had Streptococcus ‘milleri’ group as the predominant sequence taxonomy

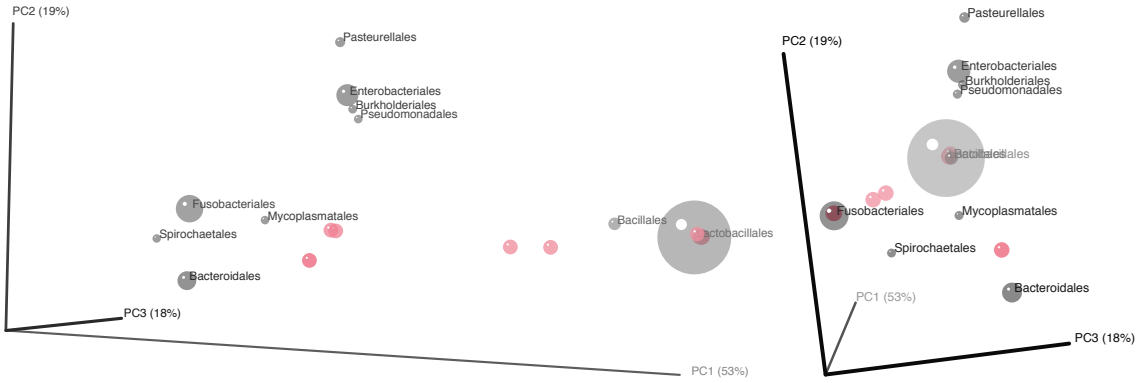


Figure 12-19 – 3D PCoA plots for samples that cultured Streptococcus 'milleri' group bacteria

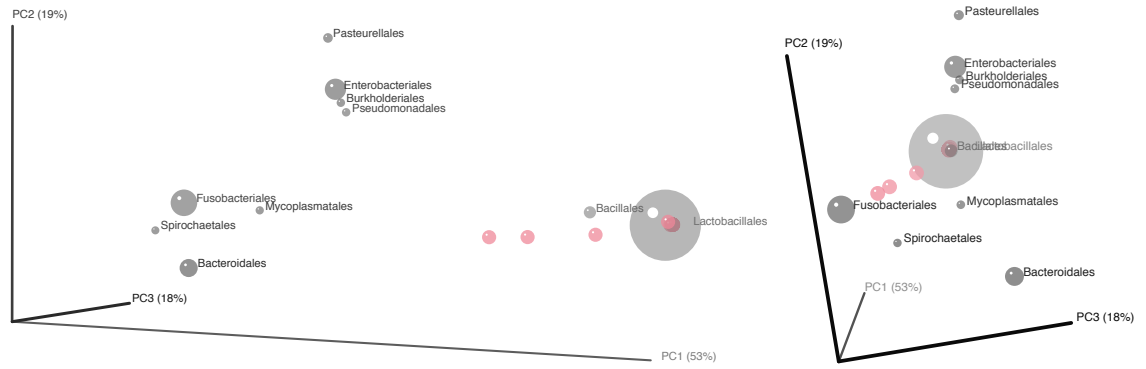
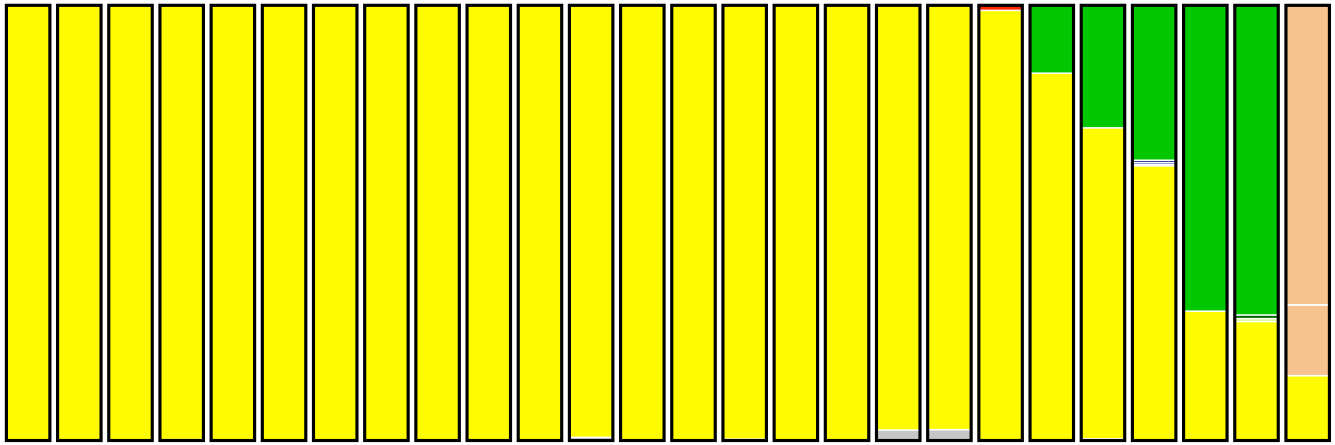


Figure 12-20 – 3D PCoA plots for samples that had Streptococcus 'milleri' group as the predominant sequence taxonomy



Legend

Streptococcus 'milleri' group

Anaerobic Gram negative bacilli

- *Campylobacter* spp.
- *Fusobacterium* spp.
- *Treponema* spp.

Anaerobic Gram positive bacilli

- *Eubacterium* spp.
- *Lactobacillus* spp.

Anaerobic Gram positive cocci

- *Parvimonas micra*

Other

- Other (non-anaerobes)

Figure 12-21 – Individual bacteriology of all samples that had Streptococcus 'milleri' group bacteria account for at least 10% of sequence reads

12.3.8.3 *Streptococcus pneumoniae*

Streptococcus pneumoniae (figures 12-22 to 12-24) was seen as the predominant sequence taxonomy in 25 out of 98 successfully sequenced samples and at ≥10% of individual sample sequences in 25 out of 98 samples (with mean proportion in these samples of 99.8% (SD 0.29)). The few remaining sequences in these samples, of questionable significance, were from other streptococci, predominantly Streptococcus 'milleri' group.

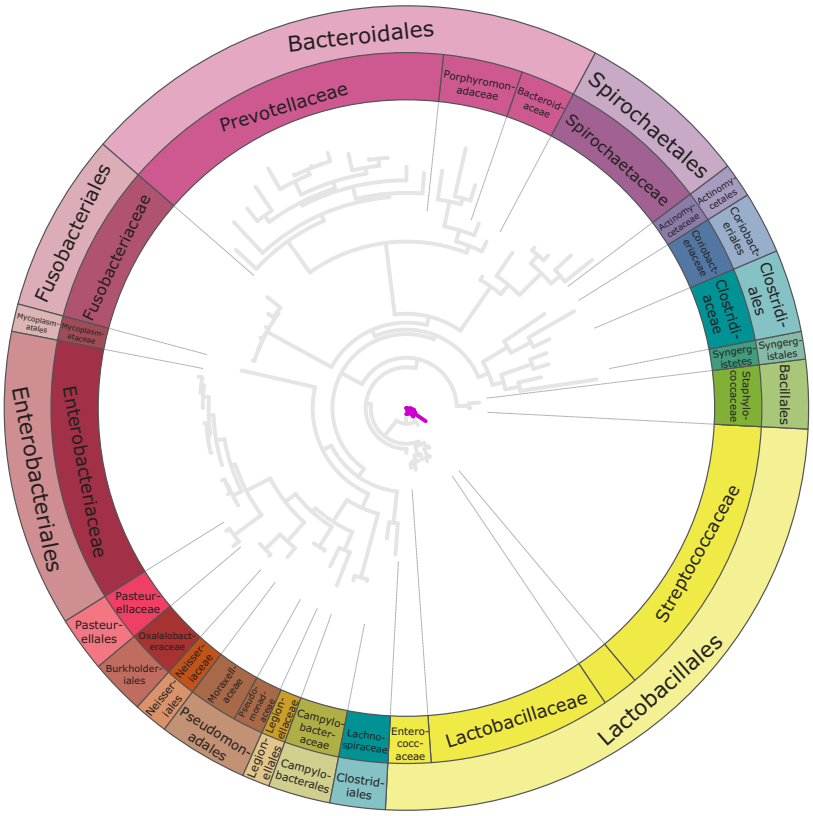


Figure 12-22 – Polar phylogenetic plot for samples that had *Streptococcus pneumoniae* as the predominant sequence taxonomy

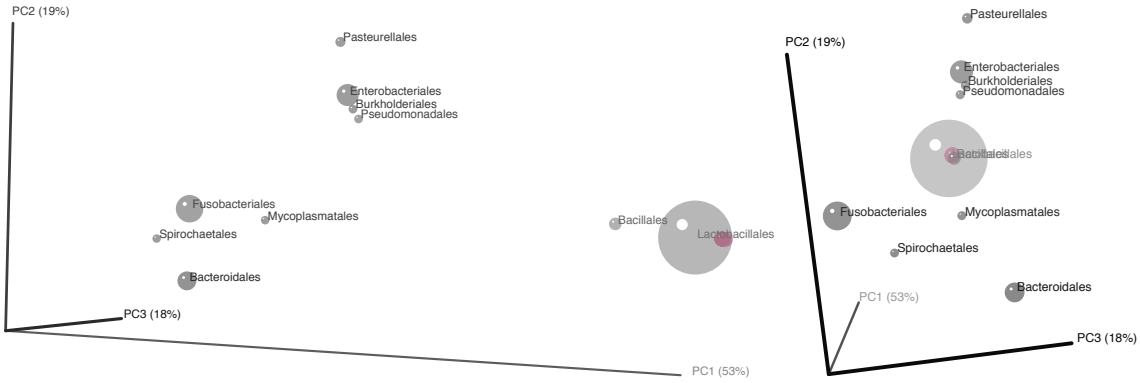


Figure 12-23 – 3D PCoA plots for samples that cultured *Streptococcus pneumoniae*

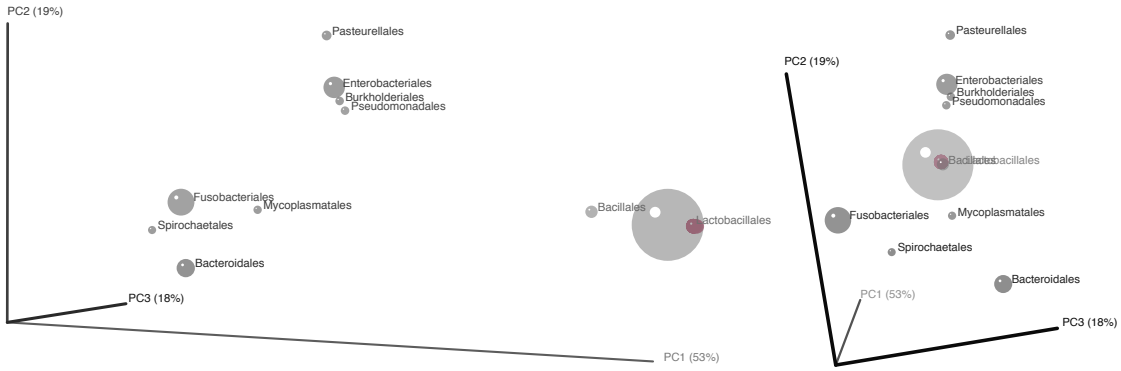


Figure 12-24 – 3D PCoA plots for samples that had *Streptococcus pneumoniae* as the predominant sequence taxonomy

12.3.8.4 *Staphylococcus aureus*

Staphylococcus aureus (Figures 12-25 to 12-27) was seen as the predominant sequence taxonomy in 4 out of 98 successfully sequenced samples and at ≥10% of individual sample sequences in 4 out of 98 samples (with no other sequences detected in these samples, i.e. 100% proportion).

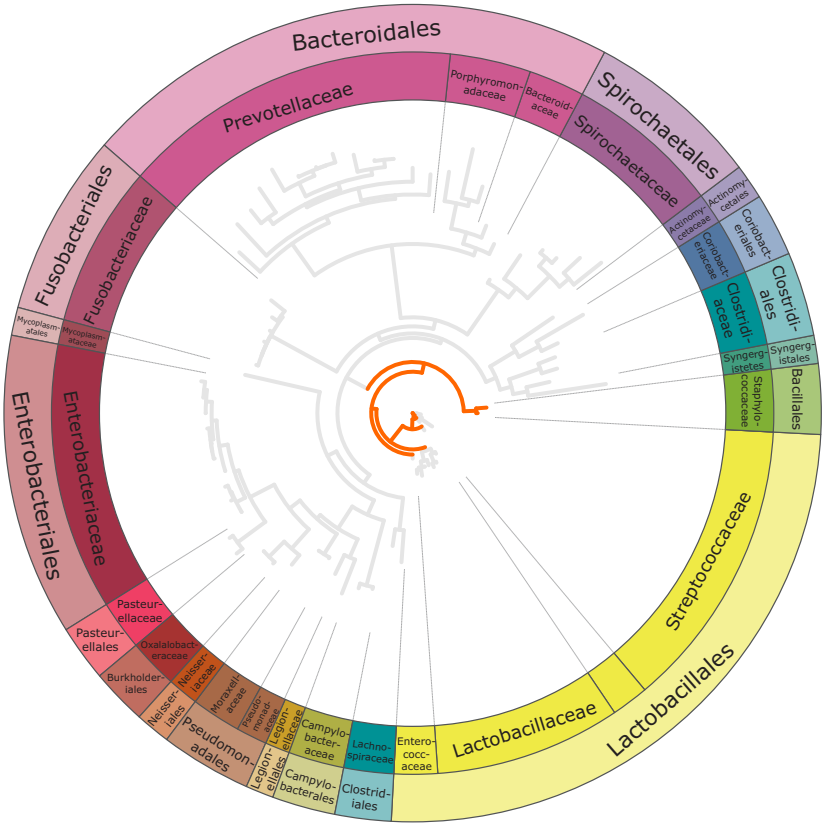


Figure 12-25 – Polar phylogenetic plot for samples that had *Staphylococcus aureus* as the predominant sequence taxonomy

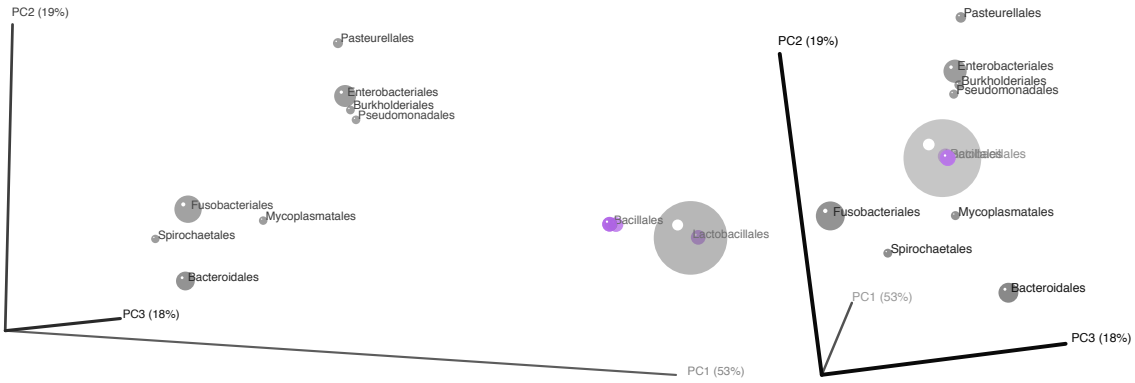


Figure 12-26 – 3D PCoA plots for samples that cultured *Staphylococcus aureus*

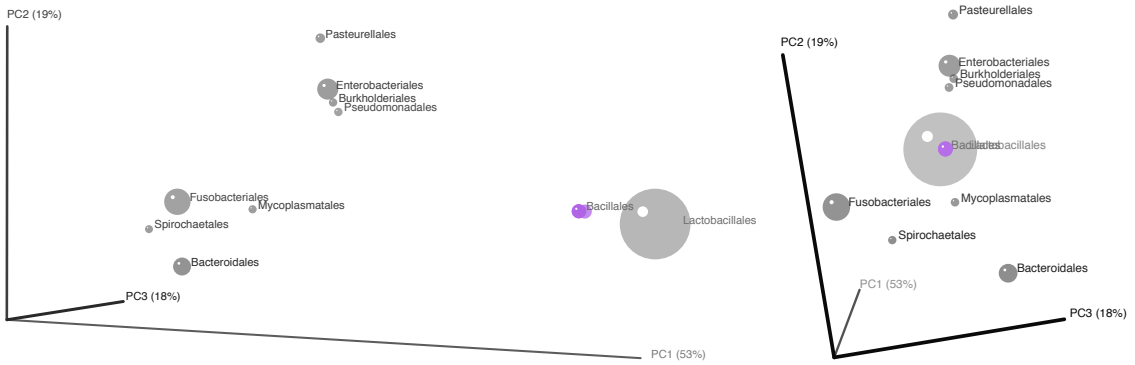


Figure 12-27 – 3D PCoA plots for samples that had *Staphylococcus aureus* as the predominant sequence taxonomy

12.3.8.5 Enterobacteriaceae

Enterobacteriaceae (Figures 12-28 to 12-30) were seen as the predominant sequence taxonomy in 11 out of 98 successfully sequenced samples and at ≥10% of individual sample sequences in 11 out of 98 samples (with mean proportion in these samples of 99.5% (SD 1.1)). Ten samples had sequences matching *Escherichia coli* and one sample had sequences matching *Proteus mirabilis*. One sample also had 3.5% of sequences which matched *Haemophilus influenzae* and another had 1.6% of sequences matching Streptococcus 'milleri' group.

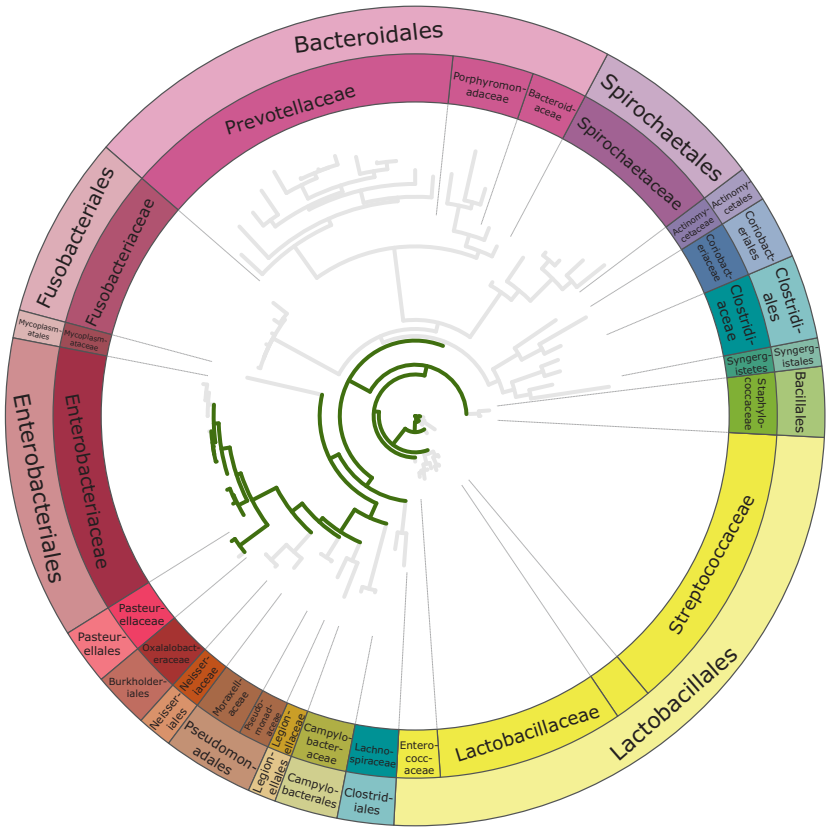


Figure 12-28 – Polar phylogenetic plot for samples that had Enterobacteriaceae as the predominant sequence taxonomy

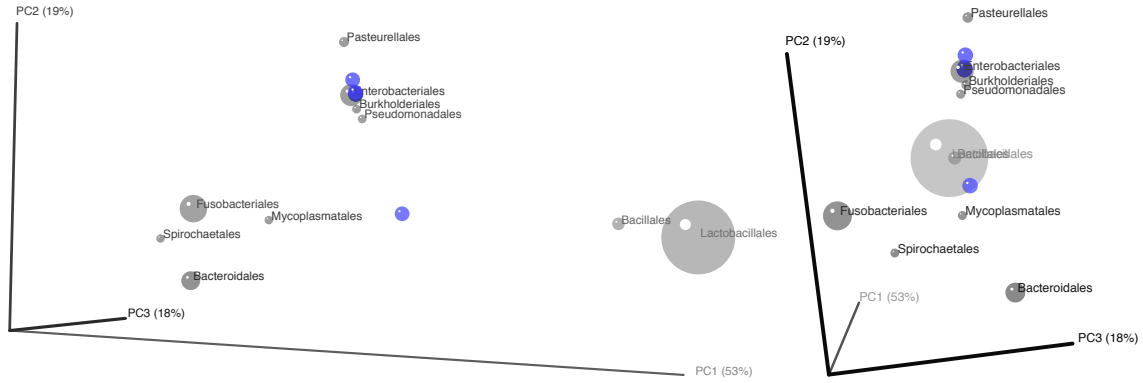


Figure 12-29 – 3D PCoA plots for samples that cultured Enterobacteriaceae

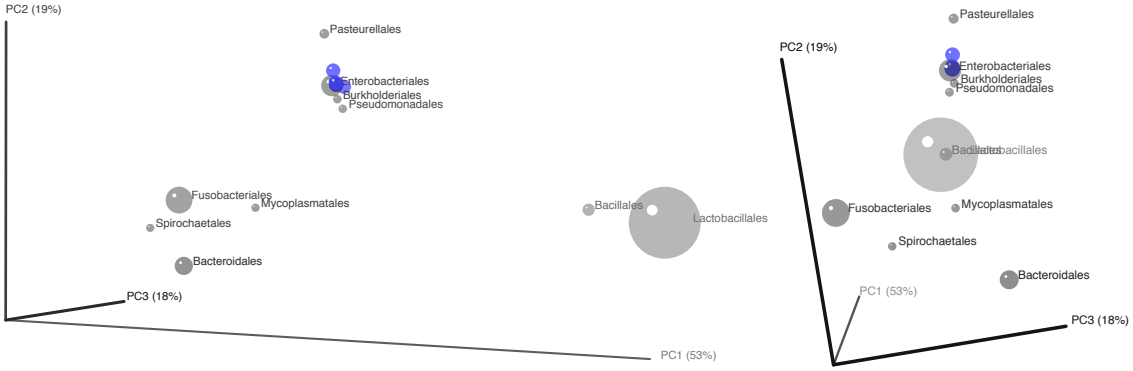


Figure 12-30 – 3D PCoA plots for samples that had Enterobacteriaceae as the predominant sequence taxonomy

12.3.9 Relationship between bacteriology and comorbidity

Table 12-10 charts patterns of comorbidities, separated by bacterial groupings. Formal statistical evaluation of association is not possible given the low number of patients in each group. However, there is no clear association between anaerobic or Streptococcus 'milleri' group infection and active dental infection/dental caries, even though these bacterial groups are common oral pathogens.

	Comorbidity						
	Active dental infection/ dental caries	**Pharyngitis**	**Respiratory conditions**	**Excessive alcohol intake**	**Immunosuppression**	**Diabetes**	**Neurological conditions**
Anaerobes (n=28)	4 (14)	3 (11)	5 (18)	3 (11)	0 (0)	8 (29)	6 (21)
Enterobacteriaceae (n=10)	3 (30)	0 (0)	3 (30)	0 (0)	1 (10)	1 (10)	0 (0)
Staphylococcus aureus (n=4)	3 (75)	0 (0)	2 (50)	1 (25)	0 (0)	2 (50)	0 (0)
Streptococcus 'milleri' group (n=23)	4 (17)	1 (4)	5 (22)	0 (0)	1 (4)	4 (17)	3 (13)
Streptococcus pneumoniae (n=25)	3 (12)	2 (8)	7 (28)	3 (12)	3 (12)	0 (0)	1 (4)
Other (n=7)	1 (14)	0 (0)	3 (43)	0 (0)	0 (0)	0 (0)	1 (14)
Sequencing negative (n=73)	11 (15)	1 (1)	16 (22)	10 (14)	3 (4)	9 (12)	6 (8)

Table 12-10 – Frequency (number (%)) of comorbidities for patients with pleural infection caused by different bacterial groups

12.4 Discussion

This is the first study to use ultra-deep sequencing to characterise the bacteriology of pleural infection using a large number of well-characterised samples. It definitively resolves the issue of sample polymicrobiality and has found involvement of many previously unrecognised anaerobes in pleural infection. It is also the first 'microbiomic' study that includes strategies to minimise the effects of contamination.

12.4.1 Bacteriological patterns

Several principle patterns of infection are evident from the data. These patterns can be summarised as follows;

- 52% – Anaerobes or Streptococcus 'milleri' group (frequently as a polymicrobial infection)
    - 29% – Anaerobes predominantly
    - 23% – Streptococcus 'milleri' group predominantly
- 26% – *Streptococcus pneumoniae* alone
- 11% – Enterobacteriaceae alone
- 4% – *Staphylococcus aureus* alone

This data is in broad agreement with data from the largest previous pleural infection study, MIST1 (figure 12-31)(96), but provides significantly increased taxonomic resolution and understanding of polymicrobiality.

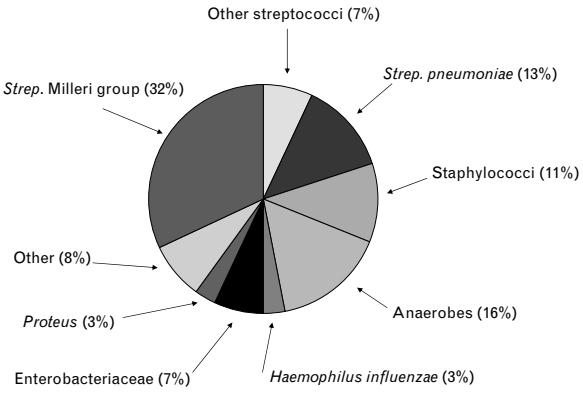


Figure 12-31 – (Reproduced from Foster et al.(96)) MIST1 community-acquired pleural infection data, using combined 16S rRNA gene capillary sequencing and culture data.

A particular advantage of ultra-deep sequencing over conventional sequencing is the ability to resolve sample polymicrobiality, given that samples were sequenced at a depth of nearly 5000 sequences per sample. Despite such resolution, infection with *Streptococcus pneumoniae*, *Staphylococcus aureus* and Enterobacteriaceae is essentially a monomicrobial process. Conversely, anaerobic and Streptococcus 'milleri' group infection are frequently polymicrobial, either with multiple anaerobic bacterial species &/or Streptococcus 'milleri' group bacteria. These data are concordant with small 16S rRNA gene cloning-based studies(29,282).

12.4.2 Anaerobic infection

Anaerobic bacteria are associated with a number of pleuropulmonary diseases, including aspiration pneumonitis, necrotising pneumonia and pleural infection,

being first reported in 1904 by Guillemot in fourteen patients with ‘putrid empyema’. A variety of techniques have been used to improve anaerobic detection rates, including bedside inoculation of ‘blood culture bottle’ media and inoculation of Stuart Transport Media (a semi-solid, non-nutrient medium supportive of anaerobic bacteria)(292); these both fail to detect some fastidious anaerobes. Nucleic acid amplification techniques (NAAT) have also been previously used to increase anaerobic detection, including capillary sequencing of the 16S rRNA gene and clone library analysis of the 16S gene (using primer pairs E341F and E907R)(282), however capillary sequencing fails to resolve polymicrobiality and clone library analysis is incredibly labour-intensive, failing to achieve a sequencing depth comparable with ultra-deep sequencing.

Using ultra-deep sequencing, 33% of samples (32/98) in this study had anaerobic bacteria present at ≥10% frequency. Previous studies have reported anaerobes in up to 76% of cases, although the authors included Streptococcus ‘milleri’ group bacteria among anaerobes. The MIST1 randomised controlled trial of pleural infection found anaerobes in only 20%(29,289,292) and previous clone library studies found anaerobes in 44% (7/16) cases(282).

Fusobacterium spp. (particularly *Fusobacterium nucleatum*) followed by *Prevotella* spp. (particularly *Prevotella oris*) were the most commonly isolated Gram negative anaerobic bacilli (GNAB) found. Other bacteria seen from the GNAB group include *Bacteroides fragilis*, *Campylobacter* spp., *Porphyromonas endodontalis* and *gingivalis*, *Tannerella forsythia*, and multiple *Treponema* spp. Multiple Gram positive anaerobic bacilli (GPAB) were also found, including *Eubacterium* spp. and *Lactobacillus* spp. *Parvimonas micra* was the only Gram positive anaerobic coccus (GPAC) found.

In comparison, previous studies examining the anaerobic microbiology of pleural infection reported three particularly common bacteria – *Fusobacterium nucleatum*, *Bacteroides melaninogenicus* and *Peptostreptococcus* spp. (this includes species that have been reclassified including *Parvimonas micra* and *Finegoldia magna*)(337). A 2004 culture-based study of the pleural infection(292) found that *Fusobacterium nucleatum* was the most commonly isolated GNAB, followed by *Bacteroides fragilis* and *Fusobacterium necrophorum*. Among GPAC, *Parvimonas micra*, *Finegoldia magna* and *Peptostreptococcus anaerobius* were most common.

Sequencing detected numerous anaerobic bacteria never previously reported in the pleural space, including *Atopobium rimae*, *Cryptobacterium curtum*, *Lactobacillus gasseri*/*taiwanensis*, *Lactobacillus iners*, *Stomatobaculum* spp., *Oribacterium* spp., *Prevotella baroniae*, *Prevotella dentalis*/*Hallella seregens*, *Prevotella scopos*, *Fretibacterium* spp., *Tannerella forsythia*, *Treponema denticola*, *Treponema lecithinolyticum*, *Treponema maltophilum*, *Treponema medium* and *Treponema socranskii*. Intriguingly, the original isolation and description of almost all the detected anaerobes were from the oropharynx and some have never been detected at other body sites. Within the oropharynx, the gingival crevices are sites of particularly high anaerobe concentration, harbouring up to 10^8 bacteria/mg plaque(338). *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola*, described as 'red complex' anaerobes by dental bacteriologists, are noted to be particularly associated with refractory periodontal disease(339). Being nutritionally fastidious, *Treponema* spp. are very difficult to culture, and would be very unlikely to be detected using routine diagnostic culture techniques. *Mycoplasma salivarium*, also a commensal of the orophyarynx, has

also been isolated (having only been reported in the pleural space once previously).

The majority of pleural fluid samples containing anaerobes were polymicrobial, with estimates of 4-5 OTUs for such samples, using multiple richness metrics. Particular patterns of co-infection are *Fusobacterium nucelatum* & Streptococcus 'milleri' group although *Prevotella* spp. ± *Fusobacterium* spp. ± *Porphyromonas* spp. ± *Treponema* spp. also appear to co-infect several samples.

The method of bacterial entry into the pleural space remains contentious, with many proposing direct spread across the visceral pleural from consolidated lung, while others propose haematogenous spread. Given the significant polymicrobiality of anaerobic pleural infection, haematogenous spread of such bacteria is unlikely given that there would need to be continual ongoing bacteraemia to enable such polymicrobiality; such conditions would likely be associated with multiple sites of abscess formation. A more likely mechanism would be recurrent microaspiration and defective mucociliary clearance in the context of pulmonary collapse or consolidation, allowing spread across the visceral pleura. Studies from the 1920s by Smith investigated this by inoculating the trachea of mice, guinea pigs, rabbits and dogs with human periodontal material, successfully causing pulmonary abscess and some cases of empyema, lending support to the aspiration theory(340).

12.4.3 Evaluation of molecular techniques

Sequencing was successful in 57% (98/172) of available pleural fluid samples, a significant improvement on 30% (52/172) culture positivity. Importantly, the bacteriological profile of culture negative and culture positive samples was similar. Comparison between sequencing and culture results was favourable with high positive and negative percent agreement values, despite the higher rates of pathogen identification with sequencing.

Bacterial load, as estimated by 16S rRNA gene qPCR, was associated with acknowledged predictors for pleural infection, i.e. pleural fluid culture status, appearance, high LDH, low pH and low glucose. Such pleural fluid characteristics also predicted sequencing success – the odds ratio (OR) of sequencing success increased by 4.1 for every 0.1 unit decrease in pH; purulent vs. non-purulent pleural fluid has OR 5.2; pleural fluid LDH >5000 vs. <1000 IU/L has OR 5.6; pleural fluid glucose ≤1.0 vs. >2.2 has OR 3.15. Such findings add further support to the current criteria for defining pleural infection and could be useful in future studies in determining appropriateness of sequencing different samples. Patient WCC and CRP were associated neither with bacterial load nor sequencing success.

The workflow used allowed nearly 5000 sequences per sample to be processed and used a variety of data visualisation strategies to examine both patterns and individual species within samples. A particular advantage of the Roche 454 FLX technology is the long sequence reads (significantly in excess of Illumina sequencing techniques). For this project, the mean length of representative

sequences of 517 nt allowed species-level resolution for the vast majority of sequences.

Previous, predominantly environmental, microbiomic studies do not address possible sample contamination, which can occur at multiple time-points, including sample collection, DNA extraction and PCR reagent contamination. The tight clustering of control pleural fluid samples show a consistent high diversity contamination profile dominated by *Ralstonia* spp. Given the significant sequence overlap between these control samples and a few pleural infection samples (with a low 16S rRNA qPCR estimation of bacterial load), our strategy of spiking each sample allowed us to define a spike threshold ratio below which sequences were considered contaminants; this is the first study to use such a technique.

66 samples remain negative by both culture and ultra-deep sequencing. Almost all of these samples successfully generated a 16S rRNA amplicon, but sequences fell below threshold and many samples were indistinguishable from control pleural fluid samples. Analysis of 16S rRNA qPCR results confirm that DNA extraction from these samples yielded lower quantities of bacterial DNA suggesting either a lower starting concentration of bacteria within the clinical sample or failure of DNA extraction for other reasons. Given the significant association between pleural fluid parameters associated with pleural infection and bacterial DNA load (such as low pH, low glucose, high LDH, purulence and culture positivity), the former explanation is more likely. It is possible that inflammatory mimics of pleural infection (such as Rheumatoid pleuritis and effusions associated with Systemic Lupus Erythematosus) may contribute to some of these negative samples.

12.4.4 Areas for further evaluation

The primary unresolved issue is determining whether a bacterial species is causing active infection or is possibly a bystander, particularly in polymicrobial pleural infection. Large scale association studies are required to answer such questions with sufficient power.

Whilst this study has used strategies to mitigate low grade contamination, further work could be undertaken to fully understand the source of contamination. Each PCR included ‘no template’ samples (molecular water), which reliably produced no amplicon, suggesting that contamination occurred prior to this, probably during sample extraction. Further studies using other extraction kits and techniques (including automated systems) would be worthwhile.

Published protocols and bead-based lysis techniques were used in an attempt to ensure uniform DNA extraction (from both Gram positive and negative samples). Further studies would usefully assess any biases associated with different extraction techniques.

12.5 Conclusions

This is the first study to comprehensively survey the bacteriology of pleural infection using deep sequencing techniques and quantitative assessment of pleural fluid bacterial load. The pleural fluid samples were obtained as part of the MIST2 multi-centre randomised controlled trial, and should be representative of routine clinical practice.

Pleural fluid sequencing has revealed characteristic bacteriological patterns, predominated by either anaerobes (± a minority of Streptococcus 'milleri' group bacteria), Streptococcus 'milleri' group bacteria (± a minority of anaerobes), *Streptococcus pneumoniae* alone, Enterobacteriaceae alone or *Staphylococcus aureus* alone. Sequencing has revealed many bacteria never previously isolated in the pleural space, predominantly anaerobic species. The vast majority of these anaerobic bacteria have a strong association with the mouth, particularly the gingival crevices. The degree of polymicrobility of these anaerobic pleural fluid samples strongly suggests direct soiling of the pulmonary parenchyma with subsequent trans-pleural spread of bacteria.

Excellent agreement between culture and sequencing was demonstrated. Of particular importance, sequencing patterns of bacteriology was similar for both culture positive and negative samples, suggesting that sequencing is not picking up irrelevant bystander organisms.

Traditionally held surrogate markers for pleural infection have been assessed in this study. Pleural fluid pH, culture status, LDH and glucose are all significantly associated with pleural fluid bacterial load and sequencing success. Such association adds validity to the accepted biochemical definition of pleural infection (based on pH, LDH and glucose), but may also provide useful information as sequencing moves towards deployment in the routine laboratory.

13 Ultra-deep sequencing – samples from PIPAP study

13.1 Introduction

Chapter 12 investigated the microbiology of pleural infection using ultra-deep sequencing of the 16S rRNA gene. This chapter further evaluates these molecular techniques, exploring their utility for analysing a wider range of samples taken from the lower respiratory tract, including both samples conventionally regarded as non-infected and infected taken from sites including the lung parenchyma and the pleural space, with two main aims;

1. Evaluation of these techniques in identifying the aetiological cause of pneumonia
2. Evaluation of the frequency with which a pathogen is discovered in 'control' pleural samples.

13.1.1 Aetiological causes of pneumonia

Readers are also referred to chapter 9 for a detailed review of pneumonia aetiology.

Pneumonia is a major global health problem, accounting for more than 6% of the total burden of all disease and causing more disease than cancer, myocardial infarction, stroke, HIV/AIDS or malaria. In both the developed and developing world, pneumonia causes the highest disease burden of infectious disease(1). It is the UK's third leading cause of death in women and the fifth leading cause of death in men. Between 6 and 13% of patients admitted to hospital with community-acquired pneumonia die(2-5).

Despite the significant disease burden that pneumonia presents, an aetiological diagnosis is infrequent established for reasons that have been fully discussed in chapter 9;

- Sputum samples frequently yield false positive results due to upper airway contamination. Furthermore, a significant minority – more than 40% – of patients with pneumonia are unable to expectorate sputum(6,118).
- Only 5 to 14% of blood cultures are positive in CAP(4,18-20).
- While immunochromatographic urinary antigen testing for *Streptococcus pneumoniae* has a high sensitivity, it lacks specificity, being affected by frequent upper airways colonisation(119). The antigen test for *Legionella pneumophila* only diagnoses serogroup 1.
- Bronchoscopic sampling requires considerable resources in terms of time and manpower, and sampling may be contaminated by upper airway colonisation. Further, sedation for bronchoscopy is potentially hazardous in those with respiratory failure.

13.1.1.1 Novel strategies in pneumonia

Bedside thoracic ultrasound gives excellent views of consolidated lung, and could enable rapid bedside ultrasound-guided sampling from consolidated lung using transthoracic needle aspiration (TNA) while avoiding the nasooropharyngeal contamination associated with sputum and bronchoscopic sampling.

A full discussion of TNA is given in section 9.6.2. Culture of the aspirate is stated to be associated with a culture positivity rate of ~50% in adults(142-145), although

prior antibiotic therapy reduces this to 23%(146). TNA for pneumonia has largely been performed at the bedside, almost uniquely without direct radiological control. Although apparently well tolerated and safe, use of bedside thoracic ultrasound to accurately localise consolidation may further decrease risks associated with the procedure while maximising microbiological return. In particular, the risk of pneumothorax is likely to be substantially reduced by ensuring that the TNA needle only enters consolidated, non-aerated, lung.

This study evaluates the safety, acceptability and applicability of TNA combined with conventional microbiological techniques and NAAT.

13.1.2 Analysis of other lower respiratory tract samples

Chapter 12 established a role for 16S rRNA gene analysis to investigate pleural infection. A further aspect of this chapter evaluates these techniques in analysing other lower respiratory tract samples (including transudative and malignant pleural fluid samples (thought biochemically and microbiologically to be 'non-infected'), simple parapneumonic effusion samples (conventionally held to be non-infected), and further samples in participants with pleural infection).

A previous study by Cremades *et al*.(341) successfully amplified the 16S rRNA gene in 'non-infected' pleural effusion samples, but it remains to be evaluated whether this represents contamination or the true presence of bacterial DNA in such clinical samples.

13.2 Methods

Samples were obtained as part of the PIPAP (Pathogen Identification in Pneumonia and Pleural Infection) study (REC No. 09/H0605/12). This chapter represents the results of this study.

Dependent on clinical presentation and diagnosis, a variety of samples were taken from different participants (see figure 13-3; further details also follow);

- Participants with pneumonia – transthoracic needle aspirates (TNA) of consolidated lung (± pleural fluid sampling, when present)
- Participants without pneumonia ('control' participants) – lung aspirates taken from non-infected surgically-excised lung. Typically, lobectomy was being performed for a small solitary pulmonary nodule (without any radiological or clinical evidence of obstructive pneumonitis)
- Participants with pleural infection (empyema or complicated parapneumonic effusions) – pleural fluid sampling
- Participants with non-infected pleural effusions (transudates and exudates) – pleural fluid sampling
- Pleural biopsy samples in participants undergoing thoracoscopy for possible malignant pleural disease.

13.2.1 Participants with pneumonia

Participants with community- or hospital-acquired pneumonia who met inclusion criteria and had no exclusion criteria (appendix 20) and who gave informed written

consent underwent radiologically-guided transthoracic needle aspirate (TNA) of the lung.

13.2.1.1 TNA technique

TNA was undertaken at the bedside using real-time ultrasound-guidance (or using computed tomography (CT), for one patient undergoing CT for another indication).

Prior to the procedure, participants completed a baseline assessment of chest pain and, separately, pain associated with phlebotomy using a 10 cm visual analogue scale. Post procedure, participants similarly assessed pain associated with TNA, and undertook a Likert-based evaluation of the procedure (figure 13-1).

All visual analogue scale (VAS) assessments were 10 cm lines, anchored at 0 cm ('least possible pain') and 10 cm ('worst possible pain')

Questions prior to procedure

'Please indicate how much pain you currently have in your chest by making a mark on the line below'

'Please indicate how much pain you experienced having your blood tests by making a mark on the line below'

Questions post procedure

'Please indicate how much pain you experienced having your lung fluid sample by making a mark on the line below'

'Please indicate how much you agree or disagree with each of these statements by ticking a box.

The lung fluid sample to diagnose your pneumonia was tolerable:

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
↓	↓	↓	↓	↓
☐1	☐2	☐3	☐4	☐5

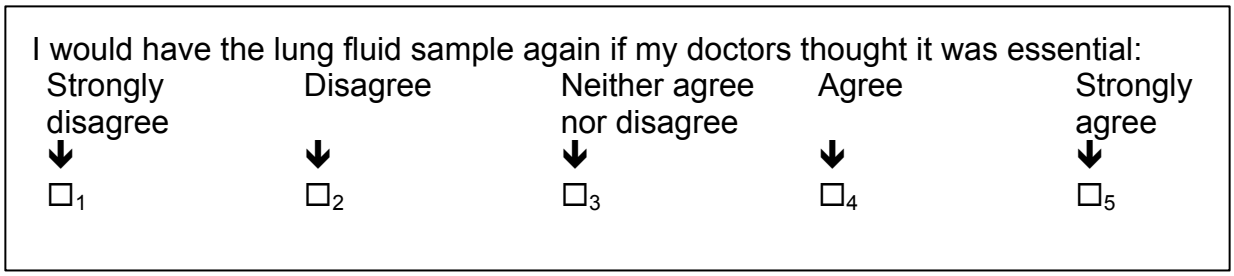


Figure 13-1 – Participant assessment of tolerability of TNA

Typically, participants were positioned in the lateral decubitus position. Ultrasound was used to select a site to perform TNA (figure 13-2). Using full asepsis and following skin preparation, up to 3 mg/kg of 1 or 2% lidocaine local anaesthetic was used to anaesthetise the skin and pleura using 25G followed by 21G needles. A 9 or 10 cm ultrafine 25G needle, attached to a 20 ml luer lock syringe containing 3.5 ml sterile non-bacteriostatic 0.9% sodium chloride solution was inserted into a consolidated (‘infected’) area of the lung under direct ultrasound guidance (using sterile ultrasound gel and probe sheath). 0.5 ml of the sodium chloride was injected followed by aspiration with gentle agitation (3 mL of sodium chloride remaining in the syringe as a carrier solution). The needle was then withdrawn.

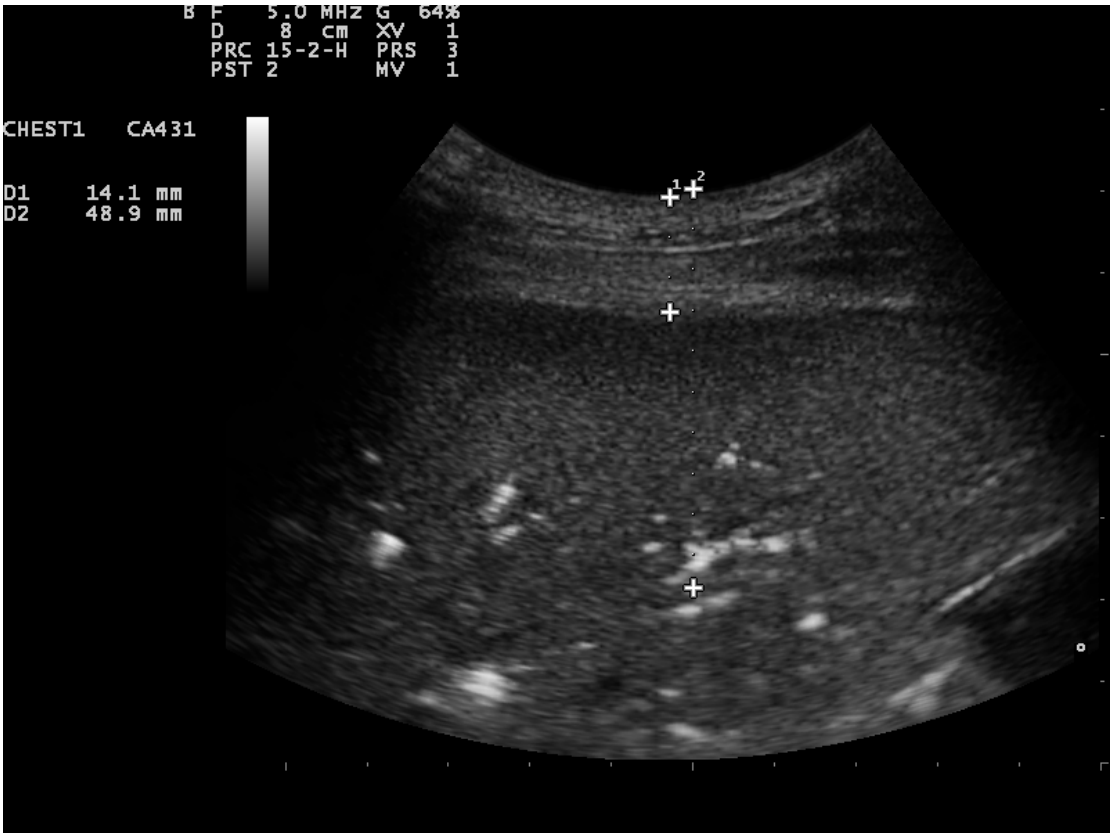


Figure 13-2 – Representative ultrasound image obtained for participant with pneumonia undergoing lung aspiration. Significant consolidated lung with air bronchograms are visible.

All participants had a follow-up chest X-ray 2-4 hours and 24 hours post procedure to evaluate for pneumothorax. Participants were systematically assessed while inpatient, and again at 30 days as an outpatient, to assess for any potential adverse events associated with TNA, including pneumothorax, haemoptysis or pain. Baseline and subsequent patient characteristics, clinical observations and laboratory investigations (including microbiology, haematology, biochemistry and radiology) were recorded on Case Record Forms.

The lung aspirate was processed as follows;

- ~1 ml underwent microscopic evaluation and culture (including aerobic, anaerobic, *Legionella* and mycobacterial and fungal culture)
- ~1 ml was inoculated into blood culture media (BD BACTEC Peds Plus) and incubated using the BD automated blood culture system
- ~1 ml underwent 16S rRNA gene analysis, as described in chapter 11
- The remaining aspirate underwent analysis using the Alere BinaxNOW *Legionella* and *Streptococcus pneumoniae* antigen card (following the manufacturer's instructions provided for analysis of urine).

Any pleural fluid present was also aspirated (separately), and underwent processing identical to lung aspirates.

Other best-practice microbiological investigations undertaken were;

- Two sets of blood cultures
- Sputum culture and sensitivity (when expectorated)
- Urinary *Legionella pneumophila* (serogroup 1) and *Streptococcus pneumoniae* antigen testing (using Alere BinaxNOW Legionella and Streptococcus pneumoniae Urinary Antigen cards)
- Baseline and four week serology for atypical pneumonia.

13.2.2 Other participants

Other participants provided a variety of samples from the lower respiratory tract as part of the PIPAP study, as summarised in figure 13-3.

- ‘Control’ lung aspirates were performed on surgically-excised lobectomy specimens for participants undergoing thoracic surgery for solitary pulmonary nodules without clinical or radiographic evidence of infection. Such aspirates were performed in theatre immediately upon lobe removal. These samples were used to define a threshold for contaminant removal purposes (details follow).
- ‘Control’ pleural fluid samples were taken from patients with transudative and non-infected exudative effusions (e.g. heart failure, hypoalbuminaemia, cirrhosis and malignant effusions). Pleural fluid samples all had pH >7.2 and glucose >3.3 mmol/L. These samples were also used for the purposes of threshold calculation for contaminant removal purposes.
- Samples were also taken from patients with simple parapneumonic effusions (an exudate with normal pH and glucose in association with pneumonia) and those with complicated parapneumonic effusions (pH <7.2 or glucose <2.2 mmol/L in association with pneumonia) and empyema (frank pus).
- Pleural biopsy samples were taken from those undergoing thoracoscopic biopsy for a possible diagnosis of malignancy, who had no systemic or pleural fluid features of infection.

All samples were processed as described in chapter 11.

13.2.3 Non-clinical control samples

Parallel to the processing of clinical samples, several non-clinical samples underwent identical DNA extraction, amplification and sequencing, including 0.9%

sodium chloride for injection, sterile water for injection and phosphate buffered saline.

13.3 Results

13.3.1 Participant characteristics

Figure 13-3 summarises the participants recruited into the PIPAP study, and table 13-1 displays their baseline characteristics.

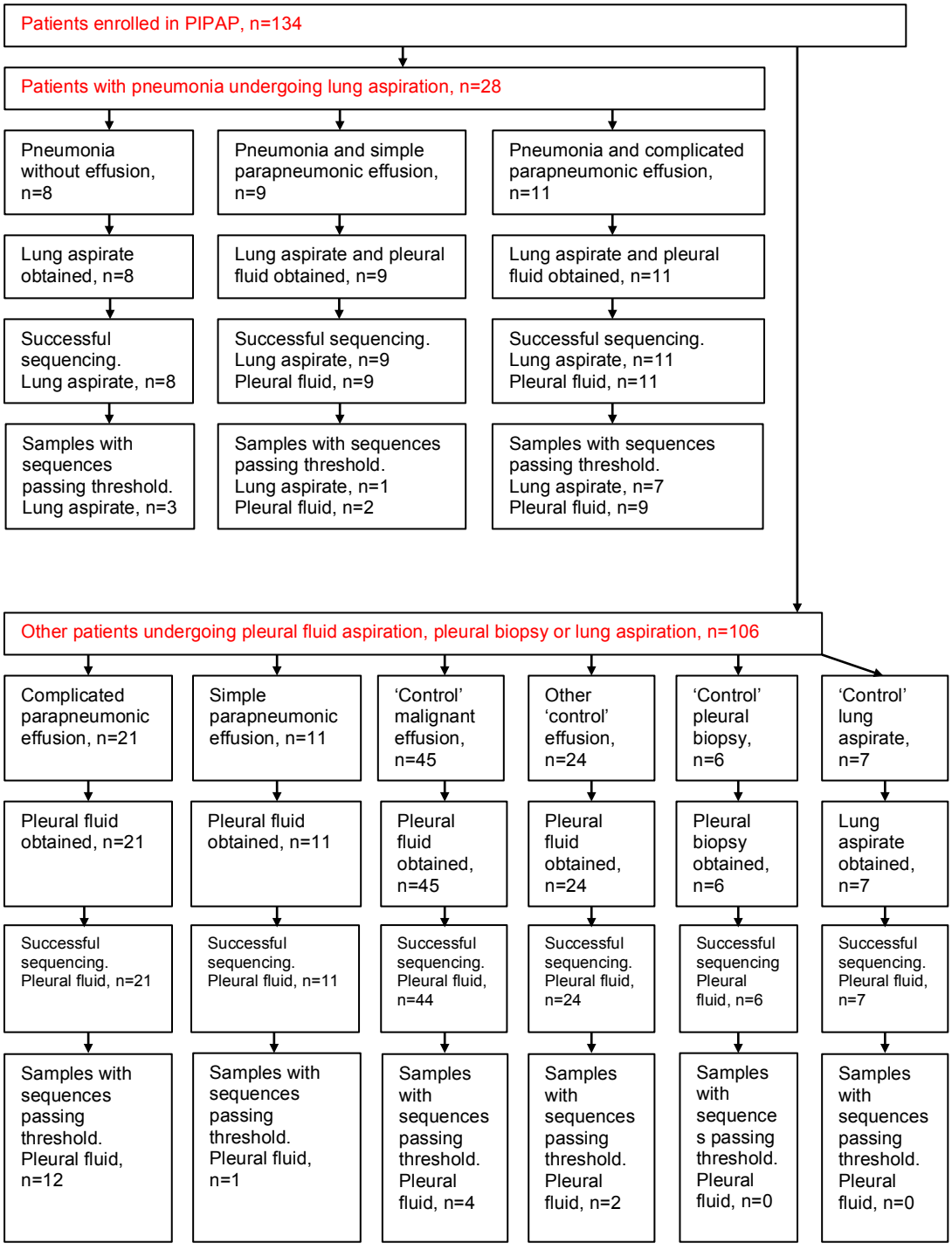


Figure 13-3 – Participants recruited into PIPAP study

Characteristic	Pneumonia (n=28)	Pleural infection (n=32)	SPPE (n=20)	Malignant effusion (n=45)	Other 'control' effusion (n=24)	'Control' lung aspirate (n=7)
Age – years ±SD	58±22	58±17	67±24	73±13	83±8	68±8
Male sex – no. (%)	19 (68)	18 (56)	11 (55)	19 (42)	16 (66)	4 (57)
Duration of symptoms before enrolment – days Median Interquartile range	9.5 5.5-17					
Community-acquired infection – no. (%)	25 (89)					
FEV1 – litres ±SD	2.2 (0.9)					
% predicted FEV1 – % ±SD	69 (19)					
Smoking history – no. (%) Never Ex Current	9 (32) 15 (54) 4 (14)					
Patient white cell count (x 10^9/L) Median Interquartile range	14.1 11.1-19.1	12.3 8.3-21.1	10.6 8.9-18.1	8.6 6.8-10.9	8.9 7.6-11.2	7.9 7.8-10.9
Patient neutrophil count (x 10^9/L) Median Interquartile range	12.5 8.6-16.9	10.8 5.7-18.5	8.5 6.6-15.0	6.2 5.0-8.6	6.3 5.5-9.6	4.7 4.6-7.2
Patient C-reactive protein (mg/L) Median Interquartile range	>160 147->160	>160 122->160	111 53-149	41 9-83	39 17-71	
Pneumonia severity index (PSI) score Median Interquartile range	2.5 1.5-4					
CURB-65 severity score Median Interquartile range	1 0-2					
Pleural fluid pH Median Interquartile range		7.03 6.75-7.17	7.37 7.33-7.39	7.36 7.32-7.41	7.42 7.38-7.47	
Pleural fluid glucose Median Interquartile range		1.6 0.4-4.2	5.7 5.1-7.0	5.7 4.9-6.3	6.3 5.4-7.0	
Pleural fluid LDH Median Interquartile range		1758 647-3807	239 211-346	220 154-464	105 71-173	
Pleural fluid total protein Median Interquartile range		40 34-50	44 32-48	44 36-48	26 21-36	

Table 13-1 – Characteristics of participants in the PIPAP study

13.3.2 Participants undergoing lung aspiration for pneumonia.

28 participants with pneumonia underwent lung aspiration, 27 using bedside thoracic ultrasound (and one using CT-guidance). No patient experienced haemoptysis or pneumothorax, and patients found the procedure tolerable (table 13-2 and figure 13-4). The VAS-assessed pain of lung aspiration was significantly lower (p=0.001, Wilcoxon matched-pairs signed-rank test) than any pre-existing chest pain, being similar to any pain associated with venesection (p=0.1).

Assessment of tolerability	Median (IQR)
'The lung fluid sample to diagnose your pneumonia was tolerable' 1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree	5 (4-5)
'I would have the lung fluid sample again if my doctors thought it was essential' 1 – Strongly disagree 2 – Disagree 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree	5 (4-5)
VAS (0-10 cm) pain in chest prior to procedure	4.0 cm (0.75-6.9 cm)
VAS (0-10 cm) pain having blood tests	1.2 cm (0.6-2.2 cm)
VAS (0-10 cm) pain having lung aspirate	1.1 cm (0.45-2.3 cm)

Table 13-2 – Patient assessment of tolerability of lung aspiration

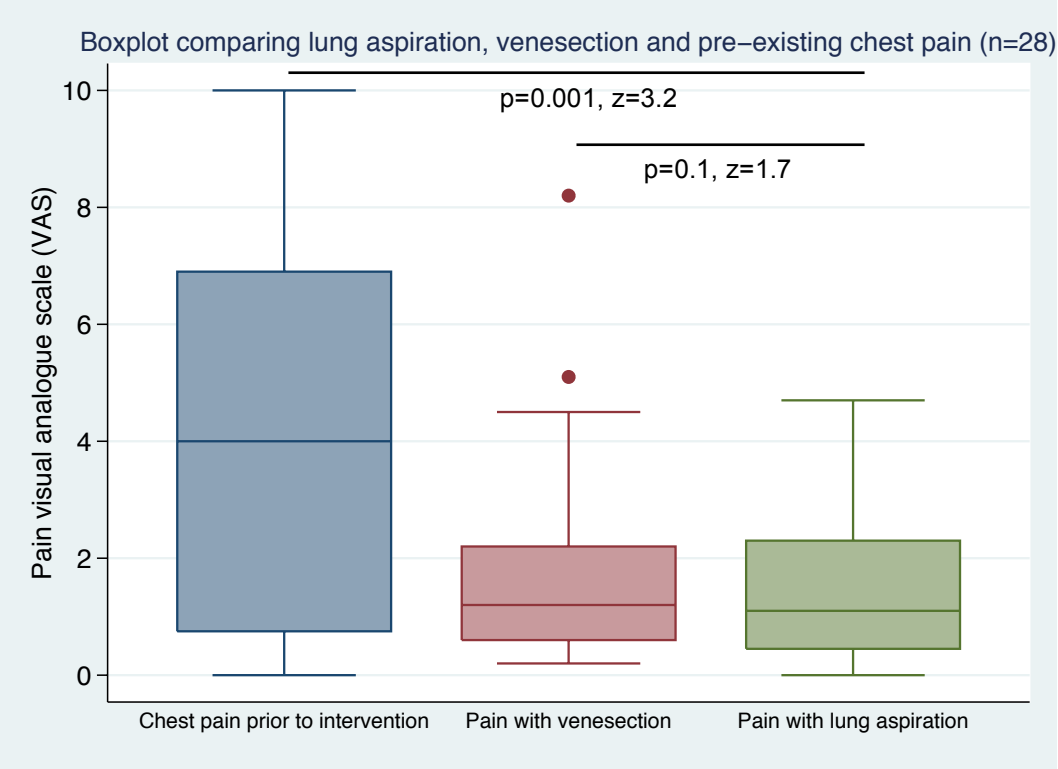


Figure 13-4 – Boxplot comparing patient VAS-assessed pain associated with lung aspiration, venesection and pre-existing chest pain. Wilcoxon matched-pairs signed-rank test was used to compare paired VAS scores as shown.

All patients were followed until day 30 for complications potentially attributable to lung aspiration. At day 30, one patient had mild ongoing pain at the site of TNA *and* chest tube insertion, although the relative contribution of each procedure to this pain was unclear.

13.3.3 Other participants recruited to the PIPAP study

106 further participants were enrolled in the PIPAP study (figure 13-3) – 21 patients with pleural infection, 11 with simple parapneumonic effusion, 45 with malignant pleural effusions, 24 with other transudative or exudative effusion, 7 'non-infected' participants undergoing control lung aspiration and 6 patients who underwent a pleural biopsy for possible malignancy.

13.3.4 Generation of 16S rRNA gene amplicon and subsequent sequencing

It was possible to generate a gene amplicon for all samples, apart from one pleural fluid sample taken from a patient with malignant pleural effusion. The amplicon product strength varied in a predictable manner, determined by the clinical presentation (i.e. pleural infection samples yielded very strong products, whereas control samples yielded very weak products and often required several attempts at amplification to generate a sequenceable product).

Figure 13-5 shows 3D PCoA plots of the raw sequencing data from the PIPAP study. Red points represent 'control' samples (including malignant pleural effusions, 'non-infected' exudative and transudative pleural effusions, control lung aspirates, control lung biopsies, and normal saline/water for injection/phosphate

buffered saline which underwent identical extraction and sequencing to clinical samples). Orange points represent simple parapneumonic effusions. Blue points represent pleural infection samples and lung aspirates from participants with pneumonia. Clear clustering is evident for red control samples.

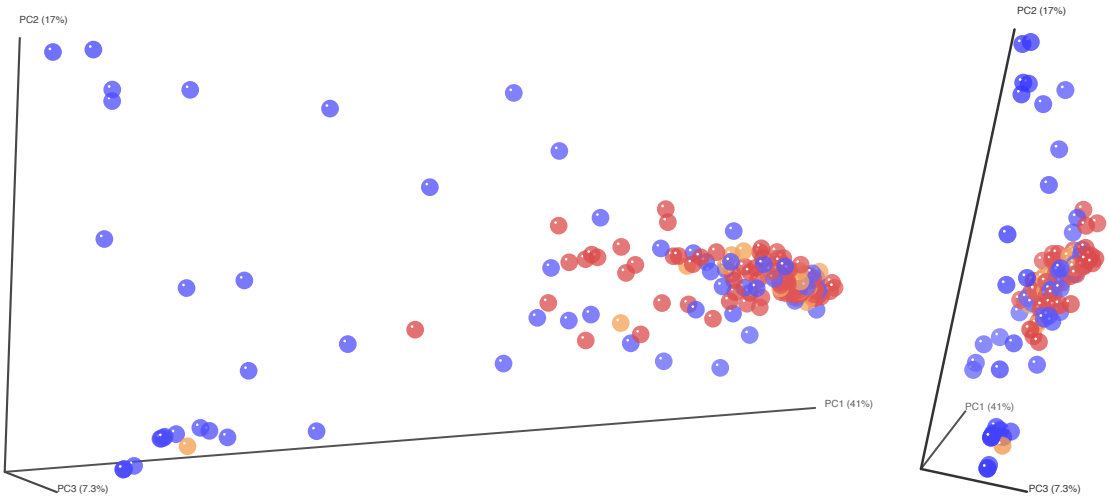


Figure 13-5 – 3D Principal Coordinate Analysis (PCoA) plots using weighted UniFrac distance. Each coloured point represents one sample, the colour of which is determined by sample type (blue = pleural infection and lung aspirates in pneumonia; orange = simple parapneumonic effusion; red = 'control' samples).

13.3.4.1 Non-clinical control samples

The bacteriology of raw sequencing results from control normal saline, water for injection and phosphate-buffered saline (which have undergone DNA extraction and sequencing identical to clinical samples) demonstrates a relatively uniform distribution of contamination (figure 13-6). Similar to chapter 12, contamination is predominated by *Ralstonia* spp. (Family – Oxalobacteraceae). Figure 13-6 also

shows neat *Acidothermus cellulolyticus* samples (which were used to spike all samples), confirming the purity of this DNA. As discussed in chapters 11 and 12, the ratio of *Ralstonia* spp. to *Acidothermus cellulolyticus* sequences was used to define a threshold above which sequences were taken to be non-contaminants[10].

[10] Threshold *Ralstonia* spp.:*Acidothermus cellulolyticus* ratio of 10.3 (TNA), 10.7 (pleural fluid), 4.3 (pleural biopsy), 2.0 (normal saline), 1.4 (water for injection).

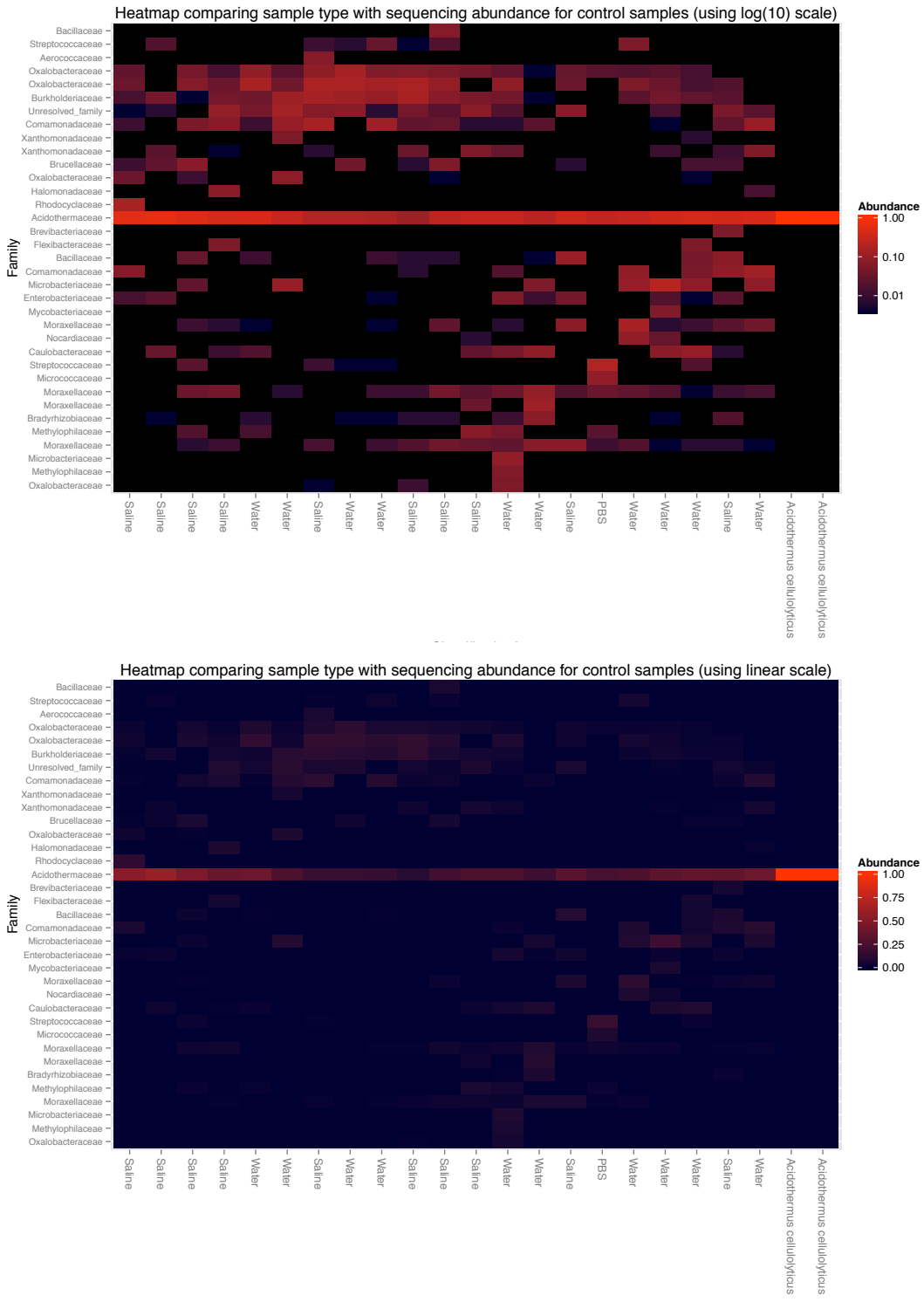


Figure 13-6 – Heatmaps (logarithmic and linear scales) showing raw sequencing results for 'extracted' saline, water and phosphate buffered saline. Neat *Acidothermus cellulolyticus* (spike used in all samples) is also shown to demonstrate purity.

13.3.4.2 Lung aspirates from participants with pneumonia

13.3.4.2.1 Raw results

The raw sequencing results from lung aspirates are shown in figure 13-7. This figure includes samples from participants with pneumonia, control lung aspirates and control saline and water samples. Contaminant clustering is evident.

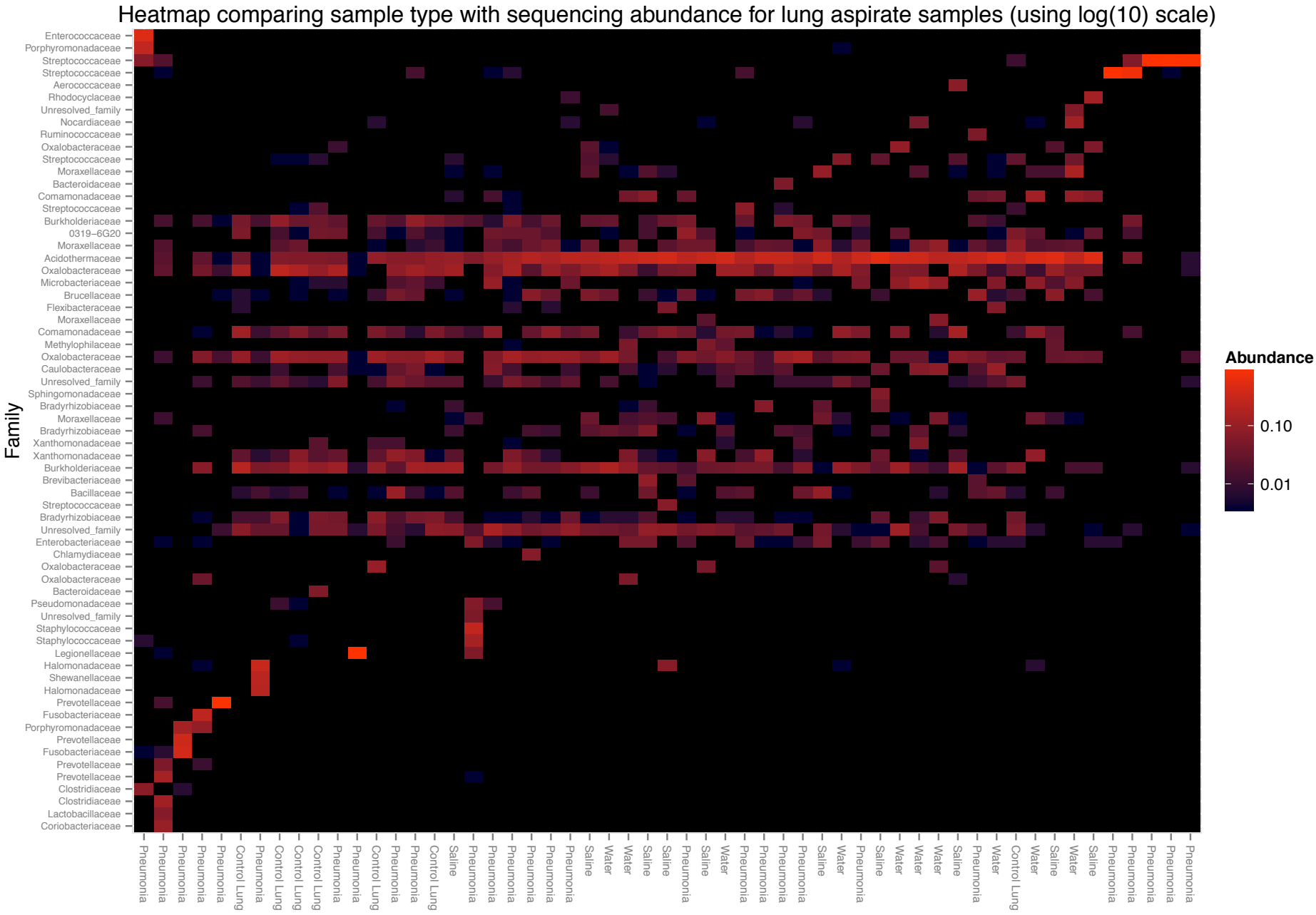


Figure 13-7 – Raw sequencing results from lung aspirate samples (including controls), saline and water controls.

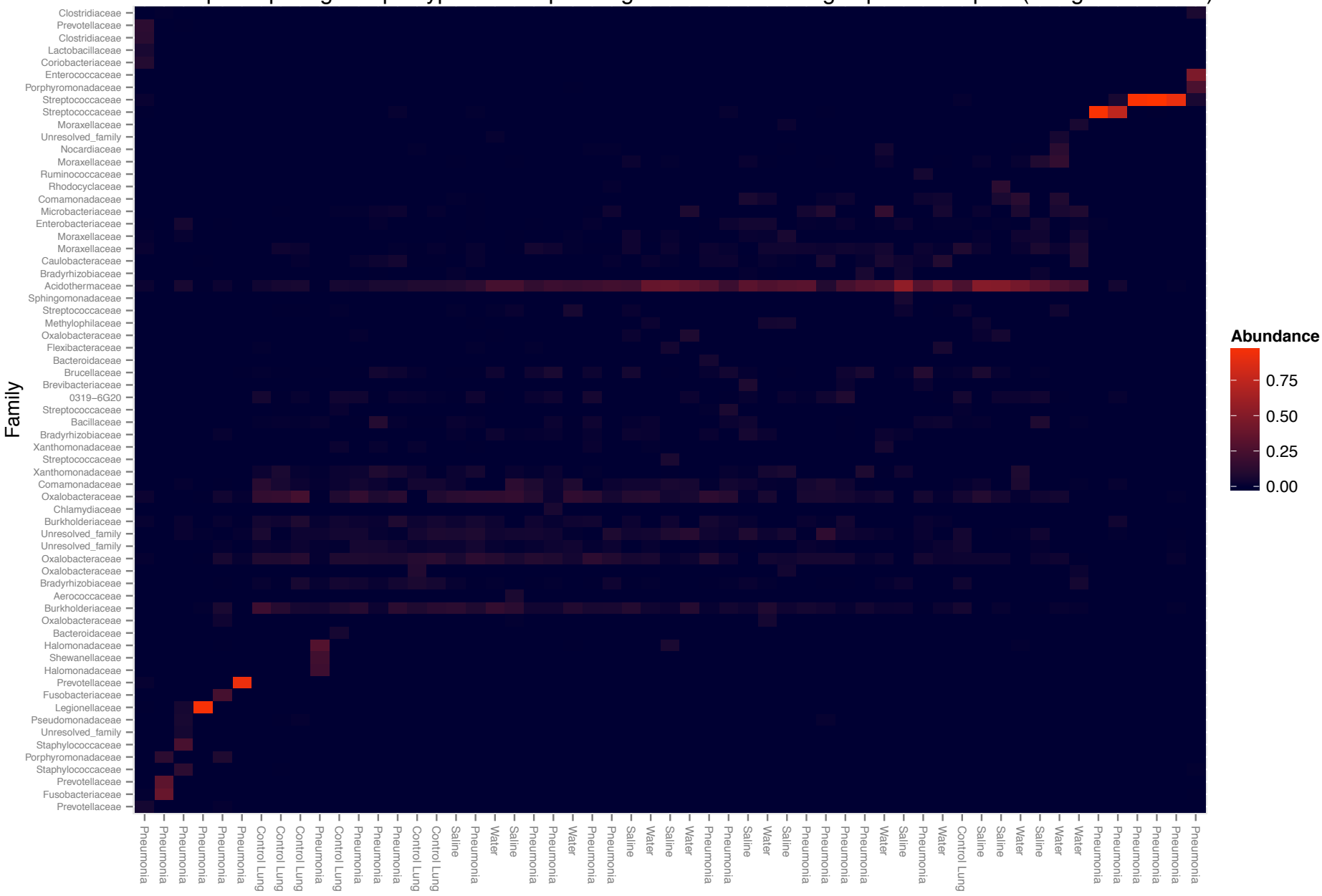


Figure 13-7 – continued

13.3.4.2.2 Post-threshold results

Figures 13-8 to 13-10 summarise the post-threshold lung aspirate results (and, in figure 13-8, compare lung aspirate results with pleural fluid results, where applicable).

Overall bacteriology of transthoracic lung needle aspirate samples and accompanying pleural fluid samples

Figure 13-8 – Sequencing and culture results from lung aspirates and any accompanying pleural fluid results for patients with pneumonia

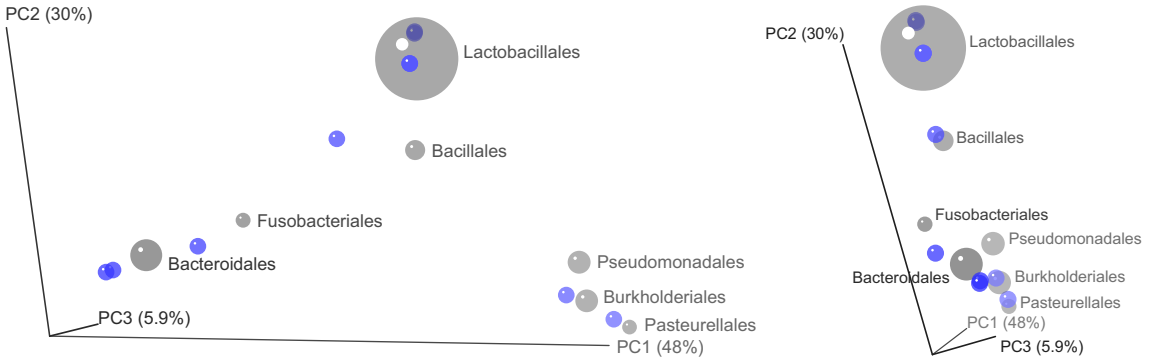


Figure 13-9 – 3D Principal Coordinate Analysis (PCoA) plots using weighted UniFrac distance. Each blue coloured point represents one lung aspirate sample for a patient with pneumonia. The seven most abundant Orders are plotted as a weighted average of the coordinates of all samples passing threshold, where the weights are the relative abundances of the Order in the samples. Sphere size is proportional to the mean relative abundance of the Order across all samples.

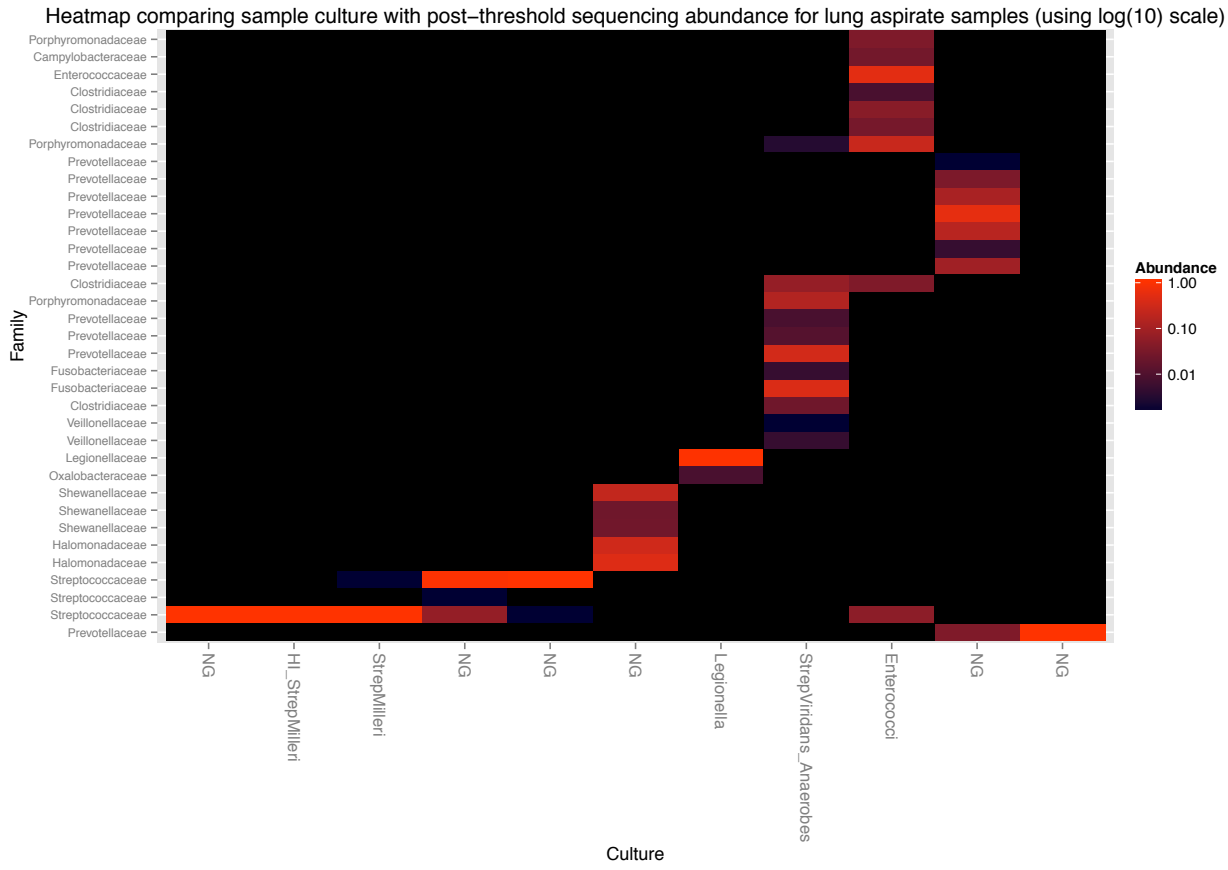


Figure 13-10 – Heatmap showing post-threshold lung aspirate sequencing results, with culture results plotted on the x axis

Several patterns are evident;

1. *Streptococcus pneumoniae* is a significant pathogen. 5/28 participants had *S. pneumoniae* in either TNA or pleural fluid samples, and an additional 6/28 participants had other results suggestive of *S. pneumoniae* involvement (e.g. blood cultures and urinary or lung antigen testing)
2. Streptococcus 'milleri' group bacteria are common. 5/28 participants had this group of bacteria in TNA/pleural fluid samples.
3. Polymicrobial anaerobic infection is common, being found in 5/28 participants.
4. *Legionella pneumophila* was only found by antigen testing in one sample, and this was confirmed by culture (figure 13-11) and sequencing.

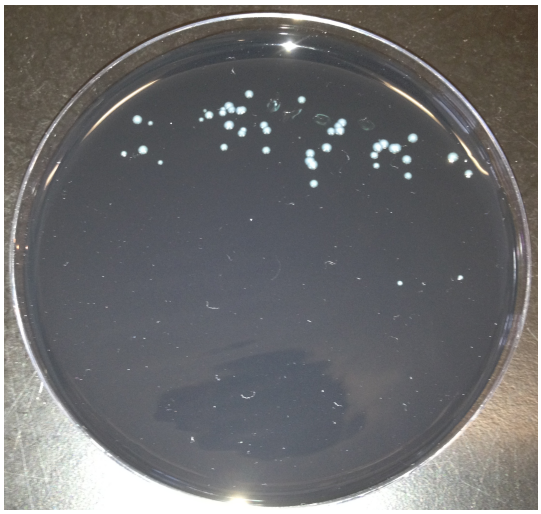


Figure 13-11 – White colonies of *Legionella pneumophila* growing from lung aspirate from one patient, using Legionella buffered charcoal yeast extract agar

5. Serological testing was frequently positive for viral (or atypical) pathogens (figure 13-8) – 10/28 participants had either Influenza A, B, RSV or *Chlamydophila pneumoniae*. Many of these samples were positive on

bacterial sequencing for other pathogens, so the true relevance of such results are unclear, in particular whether bacterial/viral co-infection is occurring.

6. Sequencing from one sample found sequences with a closest match to *Halomonas* and *Shewanella* spp., both Gram negative rods rarely found in human infections.

16 of 28 patients had a pathogen identified by sequencing, either on lung aspirate or pleural fluid aspiration. Pleural fluid samples from two patients were negative on sequencing, for whom lung aspiration yielded a result. Sequencing of 12 patients' lung aspirates and/or pleural fluid samples did not provide any bacteriological result.

Table 13-3 summarises sequencing and culture results for TNA and pleural fluid analyses combined.

	Sequencing positive	**Sequencing negative**
Culture positive	7	1
Culture negative	9	11

Table 13-3 – Summary of sequencing and culture results for patients with pneumonia (combining results from TNA and pleural fluid samples)

The incremental benefit of the different aetiological tests is as follows;

a) Only 3/28 participants had an aetiological cause of pneumonia established using blood or sputum cultures. Additional use of *S. pneumoniae* and *L. pneumophila* urinary antigen testing increased this by 9/28 participants, to a total of 12/28

b) Addition of lung aspirate sequencing and/or culture to a) increased the total by 7/28 participants (total 19/28 participants)
c) Addition of pleural fluid sequencing and/or culture to a) also increased the total by 7/28 participants (total 19/28 participants)
d) Combined addition of lung aspirate and pleural fluid analyses to a) increased the total by 9/28 participants (total 21/28 participants).

16S rRNA qPCR results were compared for sequencing positive and negative TNA samples to further understand why some samples were negative on sequencing. Table 13-4 demonstrates median 16S rRNA gene copies categorised by sequencing success. Table 13-5 demonstrates results of a logistic regression to examine the relationship between 16S rRNA gene copy and sequencing success. It can be seen that increased number of copies of 16S rRNA gene was associated with a higher odds ratio of sequencing success, suggesting that sample acquisition and DNA extraction is critical to success.

	Sequencing successful n=11	Sequencing unsuccessful n=17
Median (IQR) 16S rRNA gene copy estimate	34000 (5600-1200000)	2700 (2000-3700)

Table 13-4 – mean 16S gene copy estimation for samples, determined by sequencing success (values shown to 2 significant figures)

	Odds ratio of sequencing success	95% CI	*P* value
Tenfold increase in copies of 16S rRNA gene	4.95	1.11, 22.1	0.0015

Table 13-5 – Association between sequencing success and TNA 16S rRNA copy number

13.3.4.3 Pleural fluid sampling

13.3.4.3.1 Raw results

The raw sequencing results for pleural fluid samples from participants is shown in figure 13-12. This figure includes samples from participants with pleural infection, SPPE, malignancy, other control pleural fluid samples and control saline and water samples.

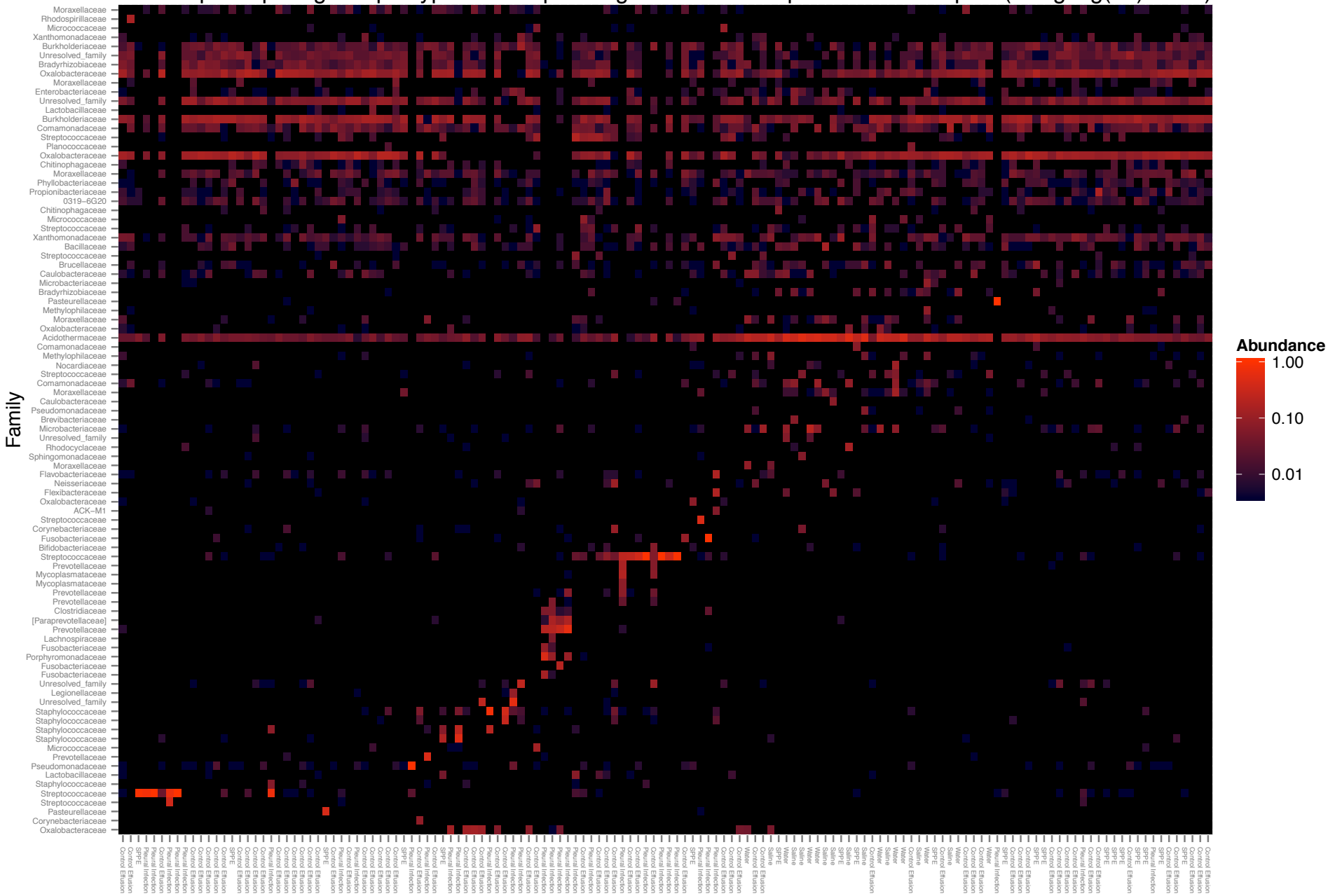


Figure 13-12 – Raw sequencing results from pleural fluid samples (including controls), saline and water controls

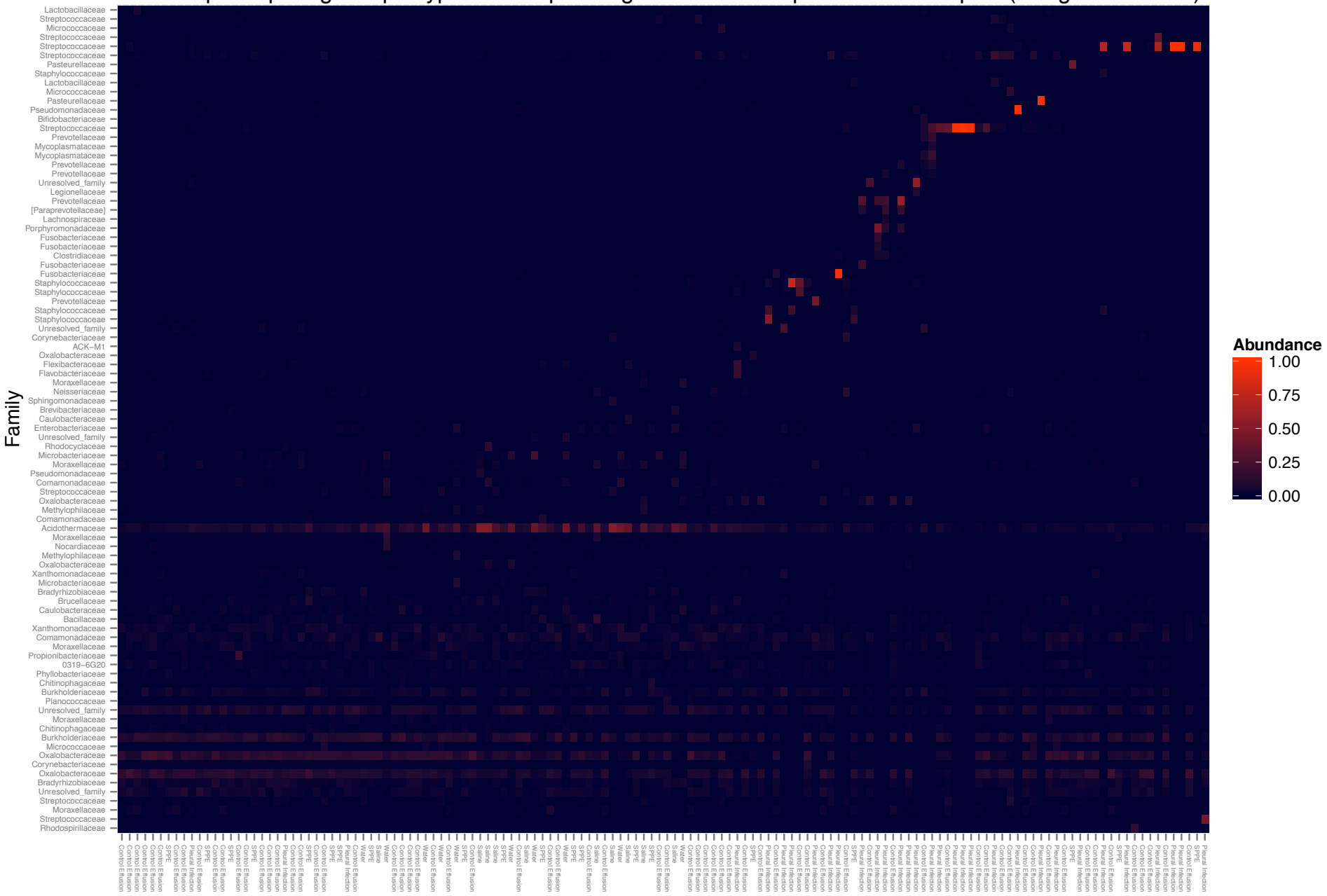


Figure 13-12 – continued

13.3.4.3.2 Post-threshold results

Figures 13-13 to 13-15 summarise the post-threshold lung aspirate results (and, in figure 13-13, include a specific comparison with culture performance).

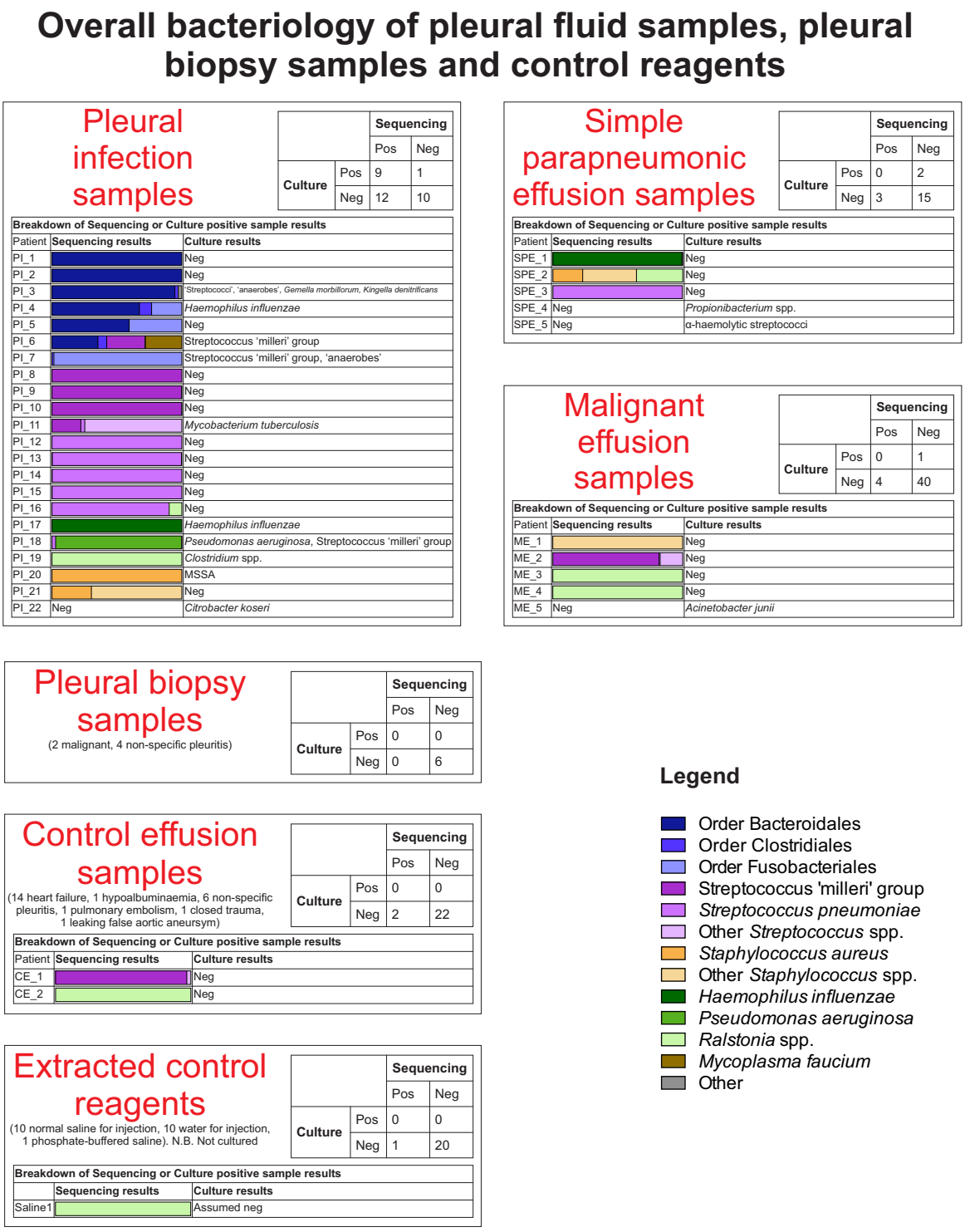


Figure 13-13 – Sequencing and culture results for pleural fluid samples, pleural biopsy samples and control reagents

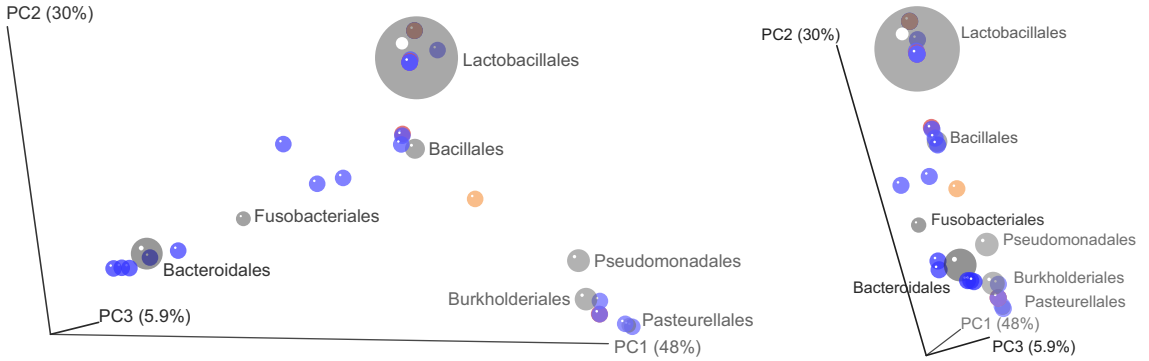


Figure 13-14 – 3D Principal Coordinate Analysis (PCoA) plots using weighted UniFrac distance. Each coloured point represents one sample, the colour of which is determined by sample type (blue = pleural infection; orange = simple parapneumonic effusion; red = 'control' pleural effusions).

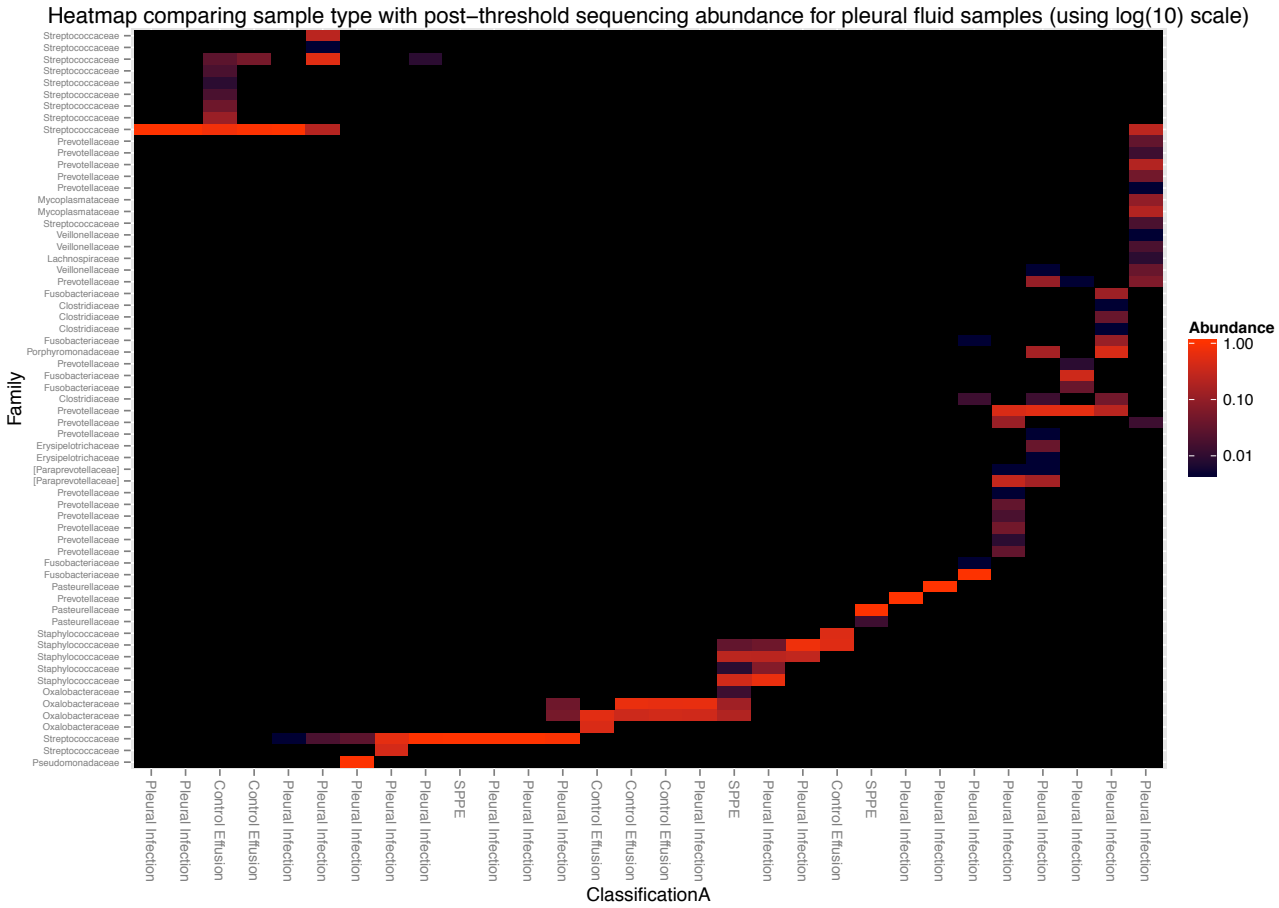


Figure 13-15 – Heatmap showing post-threshold pleural fluid sequencing results

The following may be concluded from the results;

1. A similar distribution of pathogens are seen in pleural infection as to those noted in the MIST2 study, notably multiple instances of polymicrobial anaerobic infection (with or without involvement of Streptococcus ‘milleri’ group bacteria).
2. One pleural infection sample showed polymicrobial infection including *Mycoplasma faucium*, a pathogen never previously identified within the pleural space. This sample represented an empyema that cultured Streptococcus ‘milleri’ group bacteria. Sequencing revealed polymicrobiality with the following;
    a. Streptococcus ‘milleri’ group
    b. Mycoplasma faucium
    c. *Prevotella* spp. (multiple species including *nigrescens*, *oris*, *multiformis*, *buccae*)
    d. Dialister pneumosintes
    e. Moryella indoligenes
3. 21/32 pleural infection samples were microbiologically positive by sequencing (± culture), compared with (1/32) that were positive only by culture.
4. Most simple parapneumonic effusions were negative by both culture and sequencing (15/20). Simple parapneumonic effusions are conventionally thought to be non-infected parapneumonic effusions, but probably represent an intermediate stage before the development of frank pleural infection. It is likely that bacterial ingress into a simple effusion followed by sustained growth then causes the development of pleural infection.

5. Malignant effusions and control effusions all had a predictably very low rate of sequencing positivity (4/45) and (2/24), respectively. These values are lower if *Ralstonia* spp. are taken to be contaminants (as they are assumed to be in threshold calculations).

13.4 Discussion

13.4.1 Analysis of lung aspirate samples

This study is the first to evaluate the efficacy of radiologically-guided transthoracic needle aspiration for the bacteriological diagnosis of pneumonia while also systematically assessing for adverse events and patient tolerability.

The rationale for ultrasound guidance is two-fold;

1. Ensuring accuracy of needle localisation in an attempt to maximise bacteriological return
2. Improving safety – the likelihood of adverse events (including pneumothorax) is theoretically reduced by ensuring that the TNA needle only enters non-aerated consolidated lung.

Bedside ultrasound-guided lung aspiration proved to be achievable and to be associated with very low levels of discomfort (similar to those associated with venesection). Further, the majority of patients strongly agreed that TNA was tolerable and would be willing to undergo a repeat procedure if necessary. No patients suffered from pneumothorax or haemoptysis, and only one patient (who had also had a chest drain inserted for pleural infection) had mild ongoing pain at day 30 follow-up. Previous non-guided TNA studies found that 2% of patients had a pneumothorax with 5% of patients suffering from haemoptysis (usually minor and transient)(141,144,145). In the developing world, there was a 0.2% death rate; such patients were profoundly unwell prior to the TNA and the contribution of TNA to deaths is unclear. A 0.2% mortality rate is well below the published 6-13%

pneumonia mortality rate. Meta-analyses of cases in children demonstrated similar rates of complications, but no deaths in the developed world(140).

The technique of TNA is very applicable to the UK. Bedside thoracic ultrasound is increasingly being used by Respiratory Physicians and Trainees in the UK and US for pleural procedures and assessment of the lung parenchyma. Guided TNA represents a natural extension of this skill set and all TNA procedures in this study were undertaken by a Respiratory Physician rather than a Radiologist.

Significantly, the use of ultrasound to guide the TNA procedure during this study has highlighted that needle trajectory and depth of insertion is key to accurately target consolidated lung, rather than entering non-consolidated lung, pleural fluid, liver or spleen. 'Blind' TNA studies should be interpreted with this in mind – it is possible that many 'lung aspirates' in these studies are actually pleural fluid samples.

Use of TNA increased bacteriological positivity from 43% (achieved with blood/sputum cultures and urinary antigen testing) to 68%. When combined with pleural fluid analysis (when pleural effusions were also present), bacteriological positivity increased to 75%. The overall test characteristics for TNA alone combined with 16S deep sequencing is estimated as sensitivity 39%, with specificity 100%.

Despite the ability to precisely target consolidated lung, there were still participants in whom lung aspirate culture and sequencing yielded no bacteriological aetiology, and several factors may account for this;

- The 25G needle used for lung aspiration is fine bore such that fluid return was often of low volume
- Sonographically consolidated lung represents a combination of bacterial and host immune cells, in a varying ratio. Some areas of consolidated lung will have predominantly inflammatory (rather than bacterial) cells and would have a low bacteriological yield.

13.4.2 Bacteriological patterns in pneumonia

This is the first study to combine highly accurate ultrasound-guided TNA with 16S rRNA based ultra-deep sequencing to analyse lung aspirates in an 'assumption free' manner to evaluate for previously under-recognised pathogens. Previous TNA studies variably used aspirate culture, antigen testing, and/or individual pathogen PCRs(139-142,144,342-347), although one smaller study used capillary sequencing on the 16S rRNA gene(348).

Lung aspiration and pleural fluid analysis found an expected rate of detection of *Streptococcus pneumoniae* in pneumonia. Some patients also had positive urinary and lung antigen testing for *Streptococcus pneumoniae*, but were negative on either sequencing or culture. The relevance of such antigen positivity is unclear, particularly given that oropharygneal carriage is known to cause false positives with urine antigen testing, particularly in children(119).

Contrary to large pneumonia aetiology studies (discussed in chapter 9), there was a surprising discovery of (often polymicrobial) anaerobic infection and Streptococcus 'milleri' group bacteria in a number of lung aspirates. This pattern of

infection, combined with finding the same pathogens in both lung aspirates and pleural fluid samples in several participants, gives further insight into the development of pleural infection – Anaerobes and Streptococcus 'milleri' bacteria are frequently isolated in the mouth and oropharynx; given the rarity with which these pathogens have been previously reported in pneumonia, some have suggested that direct haematogenous spread (entering the blood stream possibly via poor dentition) to the pleural space may be causing pleural infection for many patients. Our co-discovery of these pathogens in both the lung parenchyma and pleural space is more suggestive of pleural infection developing via pathogen aspiration into collapsed/consolidated lung (with accompanying defects in muco-ciliary pathogen clearance) followed by trans-pleural spread.

Antigen testing for *Legionella pneumophila* proved accurate. The only patient with antigen positivity (urine and lung aspirate) also had *L. pneumophila* identified on lung aspirate culture and sequencing.

Halomonas and *Shewanella* spp. were found in the lung aspirate from one patient, of uncertain significance. These are both Gram negative bacteria which are only infrequently reported as a cause of human infection. *Halomonas* spp. has previously been found to cause bacteraemia associated with dialysis(349), but not respiratory tract infection. *Shewanella* spp. are usually found in marine microflora, but have occasionally been reported to cause soft tissue, ear, abdominal and biliary tract infections together with bacteraemia(350). One case of traumatic pleural space infection has also been recently documented with *Shewanella* spp.(351).

Serological assays at baseline and 4 weeks post admission suggested recent acute viral infection in 9/28 patients (including influenza A, B and RSV). The relevance of such a finding is unclear, and it may be that a preceding viral upper or lower respiratory tract infection predisposed to subsequent bacterial infection. Previous studies have suggested associations between *S. aureus* and *S. pneumoniae* and viruses, and this is an area of ongoing study(352-354).

13.4.3 Analysis of pleural fluid samples

The overall sensitivity of sequencing was 66% with specificity 96%. 69% of pleural infection samples were positive by sequencing or culture. Similar to the MIST2 study, common pathogen groups included *Streptococcus pneumoniae*, Streptococcus 'milleri' group bacteria and (usually polymicrobial) anaerobic infection.

Our study found several bacteria not previously identified in the pleural space, including *Mycoplasma faucium* and *Moryella indoligenes*. Both of these bacteria have been previously described in the oropharynx.

Only 15% of biochemically-defined simple parapneumonic samples were positive by sequencing. Such effusions are conventionally regarded as non-infected. However, the development of pleural infection is thought to occur when simple parapneumonic effusions undergo bacterial invasion, followed by replication, and the discovery of bacterial DNA in such samples is therefore not unsurprising.

When *Ralstonia* spp. is classified as a contaminant (a reasonable strategy given this is the most common contaminant in non-clinical control samples) only 4% of

‘control’ transudative and exudative pleural effusions (including malignant effusion) were positive by sequencing, adding significant validity to our sequencing strategy and contaminant removal process.

13.4.4 Contamination

Examination of raw sequencing data shows contamination to be a uniquitous issue in 16S rRNA gene sequencing, despite robust strategies for contamination minimisation (including use of ultrapure plasticware and UV hoods and pre-exposure of reagents to UV light). Previous studies have found traces of bacterial DNA in polymerase enzymes used during PCR, which are potentially relevant to any assay based on 16S rRNA gene amplification(355-357). Furthermore, previous studies have suggested that there is bacterial DNA in the pleural space of many non-infective effusions(341); data from this thesis suggest that reagent contamination (including those reagents used in DNA extraction) may account for these findings.

Instead of ignoring such issues, our strategy of acknowledging potential contamination by spiking samples and defining a contamination threshold has enabled a high sequencing test specificity of 96% for pleural fluid samples and 100% for lung aspirate samples.

13.4.5 Conclusions

I have shown bedside ultrasound-guided TNA to be deliverable at the bedside, being associated with a very favourable adverse event profile and being acceptable to patients. This study also further explored the role of 16S rRNA gene

ultra-deep sequencing in analysing lower respiratory tract samples (including lung aspirates, pleural fluid samples and pleural biopsy samples). Combination of this sequencing strategy with a threshold strategy to remove contaminants has allowed a marked increase in bacterial identification whilst ensuring a high test specificity. It seems likely that the contribution of anaerobes and Streptococcus ‘milleri’ group bacteria to pneumonia is higher than previously suspected. Such pathogen identification is becoming increasingly important given a trend toward increasing bacterial resistance driven by broad spectrum antibiotic use.

14 Overall discussion and conclusions

14.1 Introduction

This thesis has explored the microbiology of the lower respiratory tract by combining the acquisition of high quality lung and pleural aspirates with high sensitivity nucleic acid amplification techniques.

Chapter 10 examined a potential role for *P. jirovecii* in pleural infection, either as a sole pathogen or as a co-infecting pathogen, finding no evidence of this fungus in the pleural space despite its increasing recognition in health and disease in both the upper and lower respiratory tract.

Chapter 10 also used nested PCR to examine for 'atypical' pneumonia pathogens (*Mycoplasma pneumoniae*, *Legionella pneumophila*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci* and *Coxiella burnetii*) in pleural infection. Again, no evidence of these pathogens was found in 374 patients with pleural infection recruited in multiple centres over many months.

Chapter 12 used ultra-deep sequencing of the 16S rRNA gene in participants with pleural infection recruited as part of the MIST2 study. This represents the only systematic evaluation of pleural fluid microbiology using such sequencing methodology, and found several distinct patterns of pleural infection, including significant polymicrobiality associated with anaerobic infection. Previous large studies have used capillary sequencing of the 16S rRNA gene (with the inherent difficulties in resolving polymicrobiality)(29,281). One previous study examined pyogenic infection at a variety of body sites using ultra-deep sequencing, and included 18 pleural fluid samples(177). This study also provided some evidence of

polymicrobial pleural infection, but did not include any methods for contamination control. Evidence was found of many pathogens not previously recognised as playing a role in pleural infection; such bacteria's usual human site of infection is almost always the oropharynx (particularly the gingival crevices).

Chapter 13 undertook a robust sensitivity and specificity assessment of 16S rRNA ultra-deep sequencing, by evaluating many lung and pleural samples conventionally regarded as non-infected (based on clinical presentation, final diagnosis and laboratory test results). I found that the sequencing techniques employed, combined with a technique of defining a threshold below which bacteria are regarded as contaminants, gave a high test specificity, while ensuring sensitivity remained adequate.

Chapter 13 also combined bedside ultrasound-guided lung aspiration for patients with pneumonia with ultra-deep sequencing. Various researchers have previously used such sequencing to examine sputum and bronchoalveolar lavage samples taken from patients without pneumonia (but suffering with COPD, cystic fibrosis and asthma)(166-176). These researchers used similar 16S rRNA gene sequencing techniques, but did include threshold techniques to control for contamination. Predictable evidence of *Streptococcus pneumoniae* was found in a significant proportion of samples. Interestingly, I also found a significant amount of polymicrobial anaerobic infection and infection with Streptococcus 'milleri' group bacteria; such pathogens have previously been rarely recognised to play a significant role in pneumonia, being more associated with cavitatory lung abscesses. Comparison of pleural fluid and lung aspirate microbiology suggests similarities amongst many patients. This adds to the body of evidence as to the mechanism of development of pleural infection.

14.2 Patterns of infection, novel pathogens and antibiotic choices

Pleural infection is usually treated with empirical broad spectrum antibiotics, but physicians vary as to whether such antibiotic selection includes atypical organism coverage. These large studies, which include participants recruited from multiple centres over many months, have found no evidence of *Mycoplasma pneumoniae*, *Legionella pneumophila*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci* or *Coxiella burnetii* and also found no evidence of *Pneumocystis jirovecii*, which is increasingly recognised to be a common airway fungus in both disease and health. There is therefore no need to either routinely include specific testing for atypical pathogens nor to routinely empirically treat for these among patients presenting with pleural infection.

'Microbiomic' study of the pleural space confirmed the above results, and found specific bacteriological patterns in the pleural space (figures 12-8 to 12-11). Significantly, extensive polymicrobiality was shown in anaerobic pleural infection. Amongst this polymicrobiality, many bacteria have been identified which have never been previously recognised to cause pleural infection, including *Atopobium rimae*, *Cryptobacterium curtum*, *Lactobacillus gasseri*/*taiwanensis*, *Lactobacillus iners*, *Stomatobaculum* spp., *Moryella indoligenes, Mycoplasma faucium, Oribacterium* spp., *Prevotella baroniae*, *Prevotella dentalis*/*Hallella seregens*, *Prevotella scopos*, *Fretibacterium* spp., *Tannerella forsythia*, *Treponema denticola*, *Treponema lecithinolyticum*, *Treponema maltophilum*, *Treponema medium* and *Treponema socranskii*. Almost universally, these pathogens have been described in the oropharynx, particularly within the gingival crevices.

Given such anaerobic polymicrobiality, it is essential that empirical antibiotic choices in pleural infection include adequate anaerobic coverage.

14.3 Aetiology of pneumonia and pleural infection

Results from lung aspiration have significant clinical relevance. Firstly, the relative frequency with which Streptococcus ‘milleri’ group bacteria and anaerobic bacteria were isolated is in contrast to the reported frequency in scientific literature (often based on serological studies, which fail to detect such pathogens)(16). This finding, combined with finding similar microbiological patterns in pleural fluid in some patients who underwent both lung aspiration and pleural aspiration lends weight to the likely trans-pleural spread of bacteria into the pleural space from consolidated lung.

Additionally, the significant polymicrobiality in many cases of anaerobic pleural infection caused by bacteria usually found in the gingival crevices suggests a likely mechanism for development of pleural infection. Recurrent microaspiration of colonised oropharyngeal secretions into lung, followed by spread across the visceral pleura into the pleural space is the most likely cause of pleural infection (rather than recurrent episodes of bacteraemia, which would probably cause multiple abscesses at several body sites).

Anaerobes and Streptococcus ‘milleri’ group bacteria are part of the human oropharyngeal flora, and are commonly isolated in associated with dental disease

and teeth abscesses(358). The milleri group are recognised to cause deep-seated infection, including abscesses at other sites(359,360). Isolates of *Streptococcus anginosus* have been shown to have thrombin-like activity, which may be of relevance in the development of pleural septations(361). Investigators suggest that the virulence of Streptococcus 'milleri' bacteria are increased in the presence of anaerobes(362), which could explain our findings of co-infection. Host factors may encourage co-infection with anaerobes and the milleri group bacteria – impaired mucociliary clearance and areas of pulmonary atelectasis or consolidation are likely to create anaerobic conditions to allow such pathogens to flourish. Given a likely oropharyngeal source, the role of teeth hygiene/clearance would be a potential future area for research, as would preventative chest physiotherapy aimed at reducing basal atelectasis and improving respiratory defence.

Finally, it should be noted that pneumonia participants recruited into PIPAP all had discrete consolidation (visible on ultrasound) rather than patchy airways-centric changes (which would be invisible on ultrasound). Such requirements may potentially skew bacteriological patterns seen.

14.4 Overall success of TNA and sequencing in pneumonia

The clinical utility and acceptability of ultrasound-guided bedside TNA was evaluated. Overall identification of pathogens (by culture or sequencing) for patients with pneumonia increased from 43% (using standard investigations,

including two sets of blood cultures and urinary antigen testing) to 68% with the addition of TNA, and to 75% when TNA was combined with pleural fluid sampling when possible. The overall sensitivity of TNA sequencing alone was 39%, with specificity 100%.

Patients generally found TNA to be tolerable and would be prepared to have a repeat procedure if required. Pain associated with TNA was comparable with phlebotomy and, on average, less than any pre-existing chest pain. No patients suffered pneumothorax or haemoptysis.

14.5 Overall success of pleural fluid sequencing

Ultra-deep sequencing also doubled rates of identification of pleural fluid pathogens in patients with pleural infection. Baseline pleural fluid culture positivity increased from ~30% (MIST2 – 30%, PIPAP – 31%) to ~66% (MIST2 – 62%; PIPAP – 69%). Sequencing characteristics were estimated for pleural fluid sequencing, with a sensitivity estimate of 66% and specificity 96%.

Exploration of factors associated with both bacterial load (estimated by 16S rRNA qPCR) and sequencing success adds further validity to the conventional biochemical proxies for defining pleural infection. Our studies found that both odds ratio of sequencing success and bacterial load increased with culture positivity, purulent pleural fluid, high LDH, low glucose and low pH. CRP and WCC were not significantly associated with sequencing success, although there appeared to be a possible weak trend.

14.6 Sequencing considerations

The optimal technique of DNA extraction from bacteria would be bias-free and remove all PCR inhibitors from a clinical sample while maximising DNA return. Many techniques have previously been advocated, some using pure chemical means to ensure cell lysis and DNA recovery, while others use mechanical cell disruption(243,363). None are clearly superior, and I adopted a mechanical disruption technique in chapters 11-13 in the hope of maximising cell lysis while reducing the time required to extract DNA.

Considerable time was spent in choosing and optimising PCR conditions and enzymes used. A high fidelity polymerase enzyme (incorporating proofreading) was chosen. All sequencing PCRs were carried out in duplicate (followed by pooling) to minimise any potential biases that could occur in a given PCR. Despite PCR optimisation, some reactions (particularly those performed using low concentrations of input DNA) generated faint primer-dimer products when using sequencing fusion primers. Modification of the AmpureXP clean-up protocol allowed us to use this simple clean-up protocol without having to resort to a (contamination prone) gel excision technique to remove such primer dimers.

The importance of primer selection for 16S-based diagnostic and metagenomic studies has been previously highlighted(229,252). My primer choice and use of 454 FLX Titanium pyrosequencing enabled read lengths in excess of 400nt, which is essential to allow detailed species-level (operational taxonomic unit-level) taxonomic assignment(252,254). Primers were chosen to enable amplification of the V4-6 region of the 16S rRNA gene; V4 has been consistently shown to give

among the highest taxonomic classification accuracy and confidence estimates(242,255). The V2 and V6 regions also provide useful information. Further, V4 and V5+V6 have been shown to minimise potential bias associated with amplification of smaller segments of the 16S rRNA gene (rather than the gene in its entirety)(256).

The validated QIIME pipeline was used to analyse the data from sequencing experiments, ultimately clustering sequences at a 97% similarity level against the curated GreenGenes 16S rRNA gene database to identify Operational Taxonomic Units present in clinical samples. Sequencing-based species definition is a subject of much debate, but many use a 97% similarity to define a species-level OTU(241).

Overall, comparison culture and sequencing results from pleural infection samples shows that sequencing does not appear to be skewing patterns of pathogen discovery, but rather increases the rate of pathogen identification and taxonomic resolution.

14.7 Contamination considerations

Acknowledgement of potential contamination is essential to all 16S rRNA-based studies. Previous studies have identified multiple sites of potential contamination, including within reagents, plasticware and polymerase enzymes (which are produced using genetically-modified bacteria)(355-357,364,365). Reagents were used which were stated to be ultra-pure and all primers were aliquoted into small volumes at an early stage to minimise any potential for contamination. All pipette

tips were pre-sterilised and included filters. Separate UV hoods were used for all Mastermix and DNA extaction processes. All reagents (apart from dNTPs and enzymes) underwent UV irradiation prior to use.

In spite of these considerations, control samples showed a characteristic pattern of contamination with multiple OTUs, predominated by *Ralstonia* spp. The source of this contamination is likely related to extraction processes given that all no-template control PCRs did not generate an amplicon.

Our strategy of spiking and threshold definition was used to minimise any problems created by contamination. Indeed, these strategies ensured a specificity of 100% for TNA samples and 96% for pleural fluid samples, while achieving a sensitivity of 39% and 66% respectively. Ensuring a reasonable balance between sensitivity and specificity is problematic, but specificity should win in a clinical context.

14.8 Why is ultra-deep sequencing sensitivity less than 100%?

It is worth considering why ultra-deep sequencing of the 16S rRNA gene has sensitivity less than 100%. Several factors are likely to be of relevance;

- Results from 16S qPCR suggest that a low input DNA concentration is the main predictor for sequencing failure. Biochemical pleural fluid parameters of infection are also associated with sequencing success. Both of these strongly suggest that low levels of bacteria (and hence their DNA) in clinical samples are the main factor determining sequencing failure. Particularly for lung aspirates, a low volume return is likely to be a significant factor

(particularly given that the aspirate is diluted in carrier saline); pleural fluid return is usually of high volume and a carrier solution is not required.

- The contaminant:pathogen bacteria DNA ratio will be critical in determining a threshold value, and hence the proportion of samples passing threshold. Further investigations using other extraction techniques, in particular automated solutions, may improve this ratio.
- Samples from lower respiratory tract sites are likely to be rich in potential inhibitors of PCR, such as a high concentration of host genomic DNA released from lysed neutrophils, and other factors including IgG and haemoglobin. Chapter 10 found that 1.5% of pleural nucleic acid extracts showed minor inhibition of PCR. Despite this, only one PIPAP sample failed to generate an amplicon suggesting that PCR inhibition was not a major factor.

14.9 Future perspectives

Our studies used primer-based and 16S rRNA sequencing based strategies to detect and identify pathogens in lower respiratory tract samples. Given the decreasing cost and increasing speed associated with whole genome sequencing (WGS), it becomes apparent that future diagnostic strategies are likely to include an element of WGS-based analyses. Significant advances in DNA library preparation are likely to simplify DNA preparation, thereby reducing ‘hands-on’ time and requirements for experiment optimisation, bringing such technology closer to the bedside.

Such a WGS strategy would, however, present significant bioinformatic issues. Firstly, the quantity of sequence data produced requires considerable resources to both process and store. Analysis of polymicrobial clinical samples (rather than the more usually analysed monomicrobial samples) would require an extremely high depth of sequencing to ensure accurate genome assembly, particularly if host (human) DNA concentration remains high after DNA extraction. There are many theoretical advantages to WGS, however, including the ability to sequences genes associated with adverse clinical outcomes, such as those associated with antibiotic resistance. Metagenomic-based approaches (based on shotgun-type sequencing, but without attempting to assemble the genome) are a potential intermediate step, and may potentially allow analysis of genes associated with resistance. However, such an approach lacks considerable taxonomic resolution and is in a state of relative infancy being highly reliant on existing annotated data sets, which are less developed than current 16S rRNA gene databases.

Specific further areas of investigation suggested as a result of this work could include;

- Assessment whether it is possible to predict pleural infection microbiology from baseline characteristics. For example, one may propose that a prolonged history of illness in a patient with multiple co-morbidities and prior antibiotic usage may be particularly associated with anaerobic infection, whereas a previously-well patient presenting ‘toxic’ with a short history of illness may be more likely to be suffering with S. pneumoniae pleural infection. Further examination of the MIST1 and 2 datasets may allow such analyses, particularly when combined with the ultra-deep sequencing used in this thesis.

- WGS may be used to directly compare an individual's oropharyngeal bacteria with those in the same patient's pleural space. The discovery of genetically-identical bacteria would add significant weight to the assertion that many pleural space bacteria originate in the oropharynx. This could be investigated in a relatively small study.
- It remains unclear why only some patients with pneumonia develop pleural infection. In particular, the relative contribution of host (genetic) and environmental factors would be worthy of further study. From a physician's perspective, the ability to predict on admission whether a patient is likely to succumb to pleural infection would be useful in tailoring treatment and follow-up. A case-control study would provide a suitable study design to investigate this.

14.10 Conclusions

The studies in this thesis have combined high quality lower respiratory tract sampling with nucleic acid amplification techniques to further our understanding of lower respiratory tract infection. I have found distinctive patterns of disease in both the lung and pleural space, and such data offer significant insights into the mechanisms underlying the development of such infection and provide further data that can be used for empirical antibiotic selection.

15 References

1. Mizgerd JP. Lung infection--a public health priority. PLoS Med. 2006 Feb;3(2):e76.

2. Menéndez R, Torres A, Zalacaín R, Aspa J, Martín-Villasclaras JJ, Borderías L, et al. Guidelines for the treatment of community-acquired pneumonia: predictors of adherence and outcome. American Journal of Respiratory and Critical Care Medicine. 2005 Sep 15;172(6):757–62.

3. Kothe H, Bauer T, Marre R, Suttorp N, Welte T, Dalhoff K, et al. Outcome of community-acquired pneumonia: influence of age, residence status and antimicrobial treatment. European Respiratory Journal. 2008 Mar 5;32(1):139–46.

4. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Timing of antibiotic administration and outcomes for Medicare patients hospitalized with community-acquired pneumonia. Arch Intern Med. 2004 Mar 22;164(6):637–44.

5. Kaplan V, Angus DC, Griffin MF, Clermont G, Scott Watson R, Linde-Zwirble WT. Hospitalized community-acquired pneumonia in the elderly: age- and sex-related patterns of care and outcome in the United States. American Journal of Respiratory and Critical Care Medicine. 2002 Mar 15;165(6):766–72.

6. van der Eerden MM, Vlaspolder F, de Graaff CS, Groot T, Bronsveld W, Jansen HM, et al. Comparison between pathogen directed antibiotic treatment and empirical broad spectrum antibiotic treatment in patients with community acquired pneumonia: a prospective randomised study. Thorax. 2005 Aug;60(8):672–8.

7. Osler W, McCrae T. The Principles and Practice of Medicine: Designed for the Use of Practitioners and Students of Medicine. D. Appleton and Company; 1918.

8. Woodhead MA, Macfarlane JT, McCracken JS, Rose DH, Finch RG. Prospective study of the aetiology and outcome of pneumonia in the community. Lancet. 1987 Mar 21;1(8534):671–4.

9. Jokinen C, Heiskanen L, Juvonen H, Kallinen S, Karkola K, Korppi M, et al. Incidence of community-acquired pneumonia in the population of four municipalities in eastern Finland. Am J Epidemiol. 1993 May 1;137(9):977–88.

10. Foy HM, Cooney MK, Allan I, Kenny GE. Rates of pneumonia during influenza epidemics in Seattle, 1964 to 1975. JAMA. 1979 Jan 19;241(3):253–8.

11. Macfarlane JT, Colville A, Guion A, Macfarlane RM, Rose DH. Prospective

study of aetiology and outcome of adult lower-respiratory-tract infections in the community. Lancet. 1993 Feb 27;341(8844):511–4.

12. Guest JF, Morris A. Community-acquired pneumonia: the annual cost to the National Health Service in the UK. Eur Respir J. 1997 Jul;10(7):1530–4.

13. Trotter CL, Stuart JM, George R, Miller E. Increasing hospital admissions for pneumonia, England. Emerg Infect Dis. 2008 May;14(5):727–33.

14. Niederman MS, McCombs JS, Unger AN, Kumar A. The cost of treating community-acquired pneumonia. Clin Ther. 1998 Jul-Aug;20(4):820-37.

15. Mizgerd JP. Acute lower respiratory tract infection. N Engl J Med. 2008 Feb 14;358(7):716–27.

16. Lim WS, Baudouin SV, George RC, Hill AT, Jamieson C, Le Jeune I, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. Thorax. BMJ Publishing Group Ltd and British Thoracic Society; 2009. pp. iii1–55.

17. Fass RJ. Aetiology and treatment of community-acquired pneumonia in adults: an historical perspective. J Antimicrob Chemother. 1993 Jul;32 Suppl A:17–27.

18. Campbell SG, Marrie TJ, Anstey R, Dickinson G, Ackroyd-Stolarz S. The contribution of blood cultures to the clinical management of adult patients admitted to the hospital with community-acquired pneumonia: a prospective observational study. Chest. American College of Chest Physicians; 2003 Apr;123(4):1142–50.

19. Fine MJ, Smith MA, Carson CA, Mutha SS, Sankey SS, Weissfeld LA, et al. Prognosis and outcomes of patients with community-acquired pneumonia. A meta-analysis. JAMA: The Journal of the American Medical Association. 1996 Jan 10;275(2):134–41.

20. Waterer GW, Wunderink RG. The influence of the severity of community-acquired pneumonia on the usefulness of blood cultures. Respiratory Medicine. 2001 Jan;95(1):78–82.

21. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003 May;58(5):377–82.

22. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997 Jan 23;336(4):243–50.

23. Light RW, Girard WM, Jenkinson SG, George RB. Parapneumonic effusions. The American Journal of Medicine. 1980 Oct;69(4):507–12.

24. Taryle DA, Potts DE, Sahn SA. The incidence and clinical correlates of parapneumonic effusions in pneumococcal pneumonia. Chest. 1978 Aug;74(2):170–3.

25. Chalmers JD, Singanayagam A, Murray MP, Scally C, Fawzi A, Hill AT. Risk factors for complicated parapneumonic effusion and empyema on presentation to hospital with community-acquired pneumonia. Thorax. 2009 Jun 26;64(7):592–7.

26. Farjah F, Symons RG, Krishnadasan B, Wood DE, Flum DR. Management of pleural space infections: A population-based analysis. The Journal of Thoracic and Cardiovascular Surgery. 2007 Feb;133(2):346–351.e1.

27. Finley C, Clifton J, Fitzgerald JM, Yee J. Empyema: an increasing concern in Canada. Can Respir J. 2008 Mar;15(2):85–9.

28. Maskell NA, Davies CWH, Nunn AJ, Hedley EL, Gleeson FV, Miller R, et al. U.K. Controlled trial of intrapleural streptokinase for pleural infection. N Engl J Med. 2005 Mar 3;352(9):865–74.

29. Maskell NA, Batt S, Hedley EL, Davies CWH, Gillespie SH, Davies RJO. The bacteriology of pleural infection by genetic and standard methods and its mortality significance. American Journal of Respiratory and Critical Care Medicine. 2006 Oct 1;174(7):817–23.

30. Tassi GF, Marchetti GP. Pleural disease: historical perspective. Textbook of Pleural Diseases Second Edition. London : Hodder Arnold; 2008. pp. 1–12.

31. Bender JM, Ampofo K, Sheng X, Pavia AT, Cannon-Albright L, Byington CL. Parapneumonic empyema deaths during past century, Utah. Emerg Infect Dis. 2009 Jan;15(1):44–8.

32. Li S-TT, Tancredi DJ. Empyema hospitalizations increased in US children despite pneumococcal conjugate vaccine. Pediatrics. American Academy of Pediatrics; 2010 Jan;125(1):26–33.

33. Goldbart AD, Leibovitz E, Porat N, Givon-Lavi N, Drukmann I, Tal A, et al. Complicated community acquired pneumonia in children prior to the introduction of the pneumococcal conjugated vaccine. Scand J Infect Dis. 2009;41(3):182–7.

34. Wu P-S, Huang L-M, Chang I-S, Lu C-Y, Shao P-L, Tsai F-Y, et al. The epidemiology of hospitalized children with pneumococcal/lobar pneumonia and empyema from 1997 to 2004 in Taiwan. Eur J Pediatr. Springer-Verlag; 2010 Jul;169(7):861–6.

35. Grijalva CG, Zhu Y, Nuorti JP, Griffin MR. Emergence of parapneumonic empyema in the USA. Thorax. 2011 Aug;66(8):663–8.

36. Kyaw MH, Lynfield R, Schaffner W, Craig AS, Hadler J, Reingold A, et al. Effect of introduction of the pneumococcal conjugate vaccine on drug-

resistant Streptococcus pneumoniae. N Engl J Med. 2006 Apr 6;354(14):1455–63.

37. Hendrickson DJ, Blumberg DA, Joad JP, Jhawar S, McDonald RJ. Five-fold increase in pediatric parapneumonic empyema since introduction of pneumococcal conjugate vaccine. The Pediatric Infectious Disease Journal. 2008 Nov;27(11):1030–2.

38. Byington CL, Hulten KG, Ampofo K, Sheng X, Pavia AT, Blaschke AJ, et al. Molecular epidemiology of pediatric pneumococcal empyema from 2001 to 2007 in Utah. Journal of Clinical Microbiology. American Society for Microbiology; 2010 Feb;48(2):520–5.

39. Rello J, Lujan M, Gallego M, Belmonte Y, Fontanals D, Vallès J, et al. Influence of pneumococcal serotype group on outcome in adults with bacteraemic pneumonia. European Respiratory Journal. European Respiratory Society; 2010 Nov;36(5):1073–9.

40. Burgos J, Lujan M, Falcó V, Sánchez A, Puig M, Borrego A, et al. The spectrum of pneumococcal empyema in adults in the early 21st century. Clinical Infectious Diseases. Oxford University Press; 2011 Aug 1;53(3):254–61.

41. Alemán C, Alegre J, Monasterio J, Segura RM, Armadans L, Anglés A, et al. Association between inflammatory mediators and the fibrinolysis system in infectious pleural effusions. Clin Sci. 2003 Nov;105(5):601–7.

42. Kroegel C, Antony VB. Immunobiology of pleural inflammation: potential implications for pathogenesis, diagnosis and therapy. European Respiratory Journal. European Respiratory Society; 1997 Oct 1;10(10):2411–8.

43. Idell S, Girard W, Koenig KB, McLarty J, Fair DS. Abnormalities of pathways of fibrin turnover in the human pleural space. Am Rev Respir Dis. 1991 Jul;144(1):187–94.

44. Potts DE, Taryle DA, Sahn SA. The glucose-pH relationship in parapneumonic effusions. Arch Intern Med. 1978 Sep;138(9):1378–80.

45. Sahn SA, Reller LB, Taryle DA, Antony VB, Good JT. The contribution of leukocytes and bacteria to the low pH of empyema fluid. Am Rev Respir Dis. 1983 Nov;128(5):811–5.

46. Emamian SA, Kaasbol MA, Olsen JF, Pedersen JF. Accuracy of the diagnosis of pleural effusion on supine chest X-ray. European Radiology. 1997 Jan 30;7(1):57–60.

47. Blackmore CC, Black WC, Dallas RV, Crow HC. Pleural fluid volume estimation: a chest radiograph prediction rule. Acad Radiol. 1996 Feb;3(2):103–9.

48. Lipscomb DJ, Flower CD, Hadfield JW. Ultrasound of the pleura: an

assessment of its clinical value. Clinical Radiology. 1981 May;32(3):289–90.

49. Eibenberger KL, Dock WI, Ammann ME, Dorffner R, Hörmann MF, Grabenwöger F. Quantification of pleural effusions: sonography versus radiography. Radiology. 1994 Jun;191(3):681–4.

50. Wu RG, Yuan A, Liaw YS, Chang DB, Yu CJ, Wu HD, et al. Image comparison of real-time gray-scale ultrasound and color Doppler ultrasound for use in diagnosis of minimal pleural effusion. American Journal of Respiratory and Critical Care Medicine. 1994 Aug;150(2):510–4.

51. Balik M, Plasil P, Waldauf P, Pazout J, Fric M, Otahal M, et al. Ultrasound estimation of volume of pleural fluid in mechanically ventilated patients. Intensive Care Med. 2006 Jan 24;32(2):318–21.

52. Remérand F, Dellamonica J, Mao Z, Ferrari F, Bouhemad B, Jianxin Y, et al. Multiplane ultrasound approach to quantify pleural effusion at the bedside. Intensive Care Med. 2010 Feb 6;36(4):656–64.

53. Vignon P, Chastagner C, Berkane V, Chardac E, Francois B, Normand S, et al. Quantitative assessment of pleural effusion in critically ill patients by means of ultrasonography*. Critical Care Medicine. 2005 Aug;33(8):1757–63.

54. Roch A, Bojan M, Michelet P, Romain F, Bregeon F, Papazian L, et al. Usefulness of ultrasonography in predicting pleural effusions > 500 mL in patients receiving mechanical ventilation. Chest. 2005 Jan;127(1):224–32.

55. Hirsch JH, Rogers JV, Mack LA. Real-time sonography of pleural opacities. AJR Am J Roentgenol. 1981 Feb;136(2):297–301.

56. Akhan O, Ozkan O, Akinci D, Hassan A, Ozmen M. Image-guided catheter drainage of infected pleural effusions. Diagn Interv Radiol. 2007 Dec;13(4):204–9.

57. Shankar S, Gulati M, Kang M, Gupta S, Suri S. Image-guided percutaneous drainage of thoracic empyema: can sonography predict the outcome? European Radiology. 2000;10(3):495–9.

58. Chen C-H, Chen W, Chen H-J, Yu Y-H, Lin Y-C, Tu C-Y, et al. Transthoracic ultrasonography in predicting the outcome of small-bore catheter drainage in empyemas or complicated parapneumonic effusions. Ultrasound in Medicine & Biology. Elsevier; 2009 Sep;35(9):1468–74.

59. Chen KY, Liaw YS, Wang HC, Luh KT, Yang PC. Sonographic septation: a useful prognostic indicator of acute thoracic empyema. J Ultrasound Med. 2000 Dec;19(12):837–43.

60. Kearney SE, Davies CW, Davies RJ, Gleeson FV. Computed tomography and ultrasound in parapneumonic effusions and empyema. Clinical Radiology. 2000 Jul;55(7):542–7.

61. Porcel JM, Ruiz-González A, Bielsa S, Esquerda A, Falguera M. Pleural fluid C-reactive protein contributes to the diagnosis and assessment of severity of parapneumonic effusions. Eur J Intern Med. 2012 Jul;23(5):447–50.

62. Heffner JE, Brown LK, Barbieri C, DeLeo JM. Pleural fluid chemical analysis in parapneumonic effusions. A meta-analysis. American Journal of Respiratory and Critical Care Medicine. 1995 Jun;151(6):1700–8.

63. Davies HE, Davies RJO, Davies CWH, on behalf of the BTS Pleural Disease Guideline Group. Management of pleural infection in adults: British Thoracic Society pleural disease guideline 2010. Thorax. 2010 Aug 9;65(Suppl 2):ii41–ii53.

64. Maskell NA, Gleeson FV, Darby M, Davies RJO. Diagnostically significant variations in pleural fluid pH in loculated parapneumonic effusions. Chest. American College of Chest Physicians; 2004 Dec;126(6):2022–4.

65. Himelman RB, Callen PW. The prognostic value of loculations in parapneumonic pleural effusions. Chest. 1986 Dec;90(6):852–6.

66. Huang H-C, Chang H-Y, Chen C-W, Lee C-H, Hsiue T-R. Predicting Factors for Outcome of Tube Thoracostomy in Complicated Parapneumonic Effusion or Empyema. Chest. American College of Chest Physicians; 1999 Mar 1;115(3):751–6.

67. Porcel JM. Pleural fluid tests to identify complicated parapneumonic effusions. Curr Opin Pulm Med. 2010 Jul;16(4):357–61.

68. Porcel JM, Vives M, Esquerda A. Tumor necrosis factor-alpha in pleural fluid: a marker of complicated parapneumonic effusions. Chest. 2004 Jan;125(1):160–4.

69. Alegre J, Jufresa J, Alemán C, Segura R, Armadans L, Marti R, et al. Pleural fluid myeloperoxidase as a marker of infectious pleural effusions. Eur J Intern Med. 2001 Jul;12(4):357–62.

70. Oikonomidi S, Kostikas K, Kalomenidis I, Tsilioni I, Daenas C, Gourgoulianis KI, et al. Matrix metalloproteinase levels in the differentiation of parapneumonic pleural effusions. Respiration. Karger Publishers; 2010;80(4):285–91.

71. Alemán C, Alegre J, Segura RM, Armadans L, Suriñach JM, Varela E, et al. Polymorphonuclear elastase in the early diagnosis of complicated pyogenic pleural effusions. Respiration. 2003 Sep;70(5):462–7.

72. Porcel JM, Ruiz-González A, Galindo C, Esquerda A, Trujillano J, Falguera M, et al. Pleural fluid interleukin-8 and C-reactive protein for discriminating complicated non-purulent from uncomplicated parapneumonic effusions. Respirology. Blackwell Publishing Asia; 2008 Jan;13(1):58–62.

73. Porcel JM, Vives M, Cao G, Bielsa S, Ruiz-González A, Martinez-Iribarren A, et al. Biomarkers of infection for the differential diagnosis of pleural effusions. European Respiratory Journal. 2009 Nov 30;34(6):1383–9.

74. Rahman NM, Maskell NA, Davies CWH, Hedley EL, Nunn AJ, Gleeson FV, et al. The relationship between chest tube size and clinical outcome in pleural infection. Chest. 2010 Mar;137(3):536–43.

75. Rahman NM, Maskell NA, West A, Teoh R, Arnold A, Mackinlay C, et al. Intrapleural Use of Tissue Plasminogen Activator and DNase in Pleural Infection. N Engl J Med. 2011 Aug 11;365(6):518–26.

76. Wait MA, Sharma S, Hohn J, Dal Nogare A. A randomized trial of empyema therapy. Chest. 1997 Jun;111(6):1548–51.

77. Bilgin M, Akcali Y, Oguzkaya F. Benefits of early aggressive management of empyema thoracis. ANZ J Surg. Blackwell Publishing Asia; 2006 Mar;76(3):120–2.

78. Ferguson AD, Prescott RJ, Selkon JB, Watson D, Swinburn CR. The clinical course and management of thoracic empyema. QJM. 1996 Apr;89(4):285–9.

79. Lim WS, Macfarlane JT, Boswell TC, Harrison TG, Rose D, Leinonen M, et al. Study of community acquired pneumonia aetiology (SCAPA) in adults admitted to hospital: implications for management guidelines. Thorax. 2001 Apr;56(4):296–301.

80. McNabb WR, Shanson DC, Williams TD, Lant AF. Adult community-acquired pneumonia in central London. J R Soc Med. 1984 Jul;77(7):550–5.

81. Community-acquired pneumonia in adults in British hospitals in 1982-1983: a survey of aetiology, mortality, prognostic factors and outcome. The British Thoracic Society and the Public Health Laboratory Service. Q J Med. 1987 Mar;62(239):195–220.

82. White RJ, Blainey AD, Harrison KJ, Clarke SK. Causes of pneumonia presenting to a district general hospital. Thorax. 1981 Aug;36(8):566–70.

83. Macfarlane JT, Finch RG, Ward MJ, Macrae AD. Hospital study of adult community-acquired pneumonia. Lancet. 1982 Jul 31;2(8292):255–8.

84. Kruger S, Ewig S, Papassotiriou J, Kunde J, Marre R, Baum von H, et al. Inflammatory parameters predict etiologic patterns but do not allow for individual prediction of etiology in patients with CAP: results from the German competence network CAPNETZ. Respir Res. 2009;10(1):65.

85. Zobel K, Martus P, Pletz MW, Ewig S, Prediger M, Welte T, et al. Interleukin 6, lipopolysaccharide-binding protein and interleukin 10 in the prediction of risk and etiologic patterns in patients with community-acquired pneumonia: results from the German competence network

CAPNETZ. BMC Pulm Med. 2012;12(1):6.

86. Bodí M, Rodríguez A, Solé-Violán J, Gilavert MC, Garnacho J, Blanquer J, et al. Antibiotic prescription for community-acquired pneumonia in the intensive care unit: impact of adherence to Infectious Diseases Society of America guidelines on survival. Clinical Infectious Diseases. 2005 Dec 15;41(12):1709–16.

87. Rello J, Gattarello S, Souto J, Solé-Violán J, Vallès J, Peredo R, et al. Community-acquired Legionella Pneumonia in the intensive care unit: Impact on survival of combined antibiotic therapy. Med Intensiva. 2013 Jun;37(5):320–6.

88. Arnold FW, Brock GN, Peyrani P, Rodríguez EL, Díaz AA, Rossi P, et al. Predictive accuracy of the pneumonia severity index vs CRB-65 for time to clinical stability: results from the Community-Acquired Pneumonia Organization (CAPO) International Cohort Study. Respiratory Medicine. Elsevier; 2010 Nov;104(11):1736–43.

89. Feldman C, Ross S, Mahomed AG, Omar J, Smith C. The aetiology of severe community-acquired pneumonia and its impact on initial, empiric, antimicrobial chemotherapy. Respiratory Medicine. 1995 Mar;89(3):187–92.

90. Thompson JE. Community acquired pneumonia in north eastern Australia--a hospital based study of aboriginal and non-aboriginal patients. Aust N Z J Med. 1997 Feb;27(1):59–61.

91. Tan YK, Khoo KL, Chin SP, Ong YY. Aetiology and outcome of severe community-acquired pneumonia in Singapore. Eur Respir J. 1998 Jul;12(1):113–5.

92. Montejo Baranda M, Corral Carranceja J, Aguirre Errasti C. Q fever in the Basque Country: 1981-1984. Rev Infect Dis. 1985 Sep;7(5):700–1.

93. Mongkolrattanothai K, Aldag JC, Mankin P, Gray BM. Epidemiology of community-onset Staphylococcus aureus infections in pediatric patients: an experience at a Children's Hospital in central Illinois. BMC Infect Dis. BioMed Central Ltd; 2009 Jul 16;9(1):112.

94. Gonzalez BE, Rueda AM, Shelburne SA, Musher DM, Hamill RJ, Hulten KG. Community-associated strains of methicillin-resistant Staphylococccus aureus as the cause of healthcare-associated infection. Infect Control Hosp Epidemiol. 2006 Oct;27(10):1051–6.

95. Micek ST, Dunne M, Kollef MH. Pleuropulmonary Complications of Panton-Valentine Leukocidin-Positive Community-Acquired Methicillin-Resistant Staphylococcus aureusImportance of Treatment With Antimicrobials Inhibiting Exotoxin Production. Chest. American College of Chest Physicians; 2005 Oct 1;128(4):2732–8.

96. Foster S, Maskell N. Bacteriology of complicated parapneumonic

effusions. Curr Opin Pulm Med. 2007 Jul;13(4):319–23.

97. Wong CA, Donald F, Macfarlane JT. Streptococcus milleri pulmonary disease: a review and clinical description of 25 patients. Thorax. 1995 Oct;50(10):1093–6.

98. Ahmed RA, Marrie TJ, Huang JQ. Thoracic empyema in patients with community-acquired pneumonia. The American Journal of Medicine. 2006 Oct;119(10):877–83.

99. Meyer CN, Rosenlund S, Nielsen J, Friis-Møller A. Bacteriological aetiology and antimicrobial treatment of pleural empyema. Scand J Infect Dis. 2011 Mar;43(3):165–9.

100. Lindstrom ST, Kolbe J. Community acquired parapneumonic thoracic empyema: predictors of outcome. Respirology. 1999 Jun;4(2):173–9.

101. Brook I, Frazier E. Aerobic and anaerobic microbiology of empyema. A retrospective review in two military hospitals. Chest. 1993 May 1;103(5):1502–7.

102. Civen R, Jousimies-Somer H, Marina M, Borenstein L, Shah HN, Finegold SM. A retrospective review of cases of anaerobic empyema and update of bacteriology. Clin Infect Dis. 1995 Jun;20 Suppl 2:S224–9.

103. Buesching WJ, Brust RA, Ayers LW. Enhanced primary isolation of Legionella pneumophila from clinical specimens by low-pH treatment. Journal of Clinical Microbiology. 1983 Jun;17(6):1153–5.

104. McKinney RM, Wilkinson HW, Sommers HM, Fikes BJ, Sasseville KR, Yungbluth MM, et al. Legionella pneumophila serogroup six: isolation from cases of legionellosis, identification by immunofluorescence staining, and immunological response to infection. Journal of Clinical Microbiology. 1980 Sep;12(3):395–401.

105. Saravolatz LD, Russell G, Cvitkovich D. Direct immunofluorescence in the diagnosis of Legionnaires' disease. Chest. 1981 May;79(5):566–70.

106. Narita M, Matsuzono Y, Itakura O, Yamada S, Togashi T. Analysis of mycoplasmal pleural effusion by the polymerase chain reaction. Archives of Disease in Childhood. 1998 Jan 1;78(1):67–9.

107. Augenbraun MH, Roblin PM, Mandel LJ, Hammerschlag MR, Schachter J. Chlamydia pneumoniae pneumonia with pleural effusion: diagnosis by culture. The American Journal of Medicine. 1991 Oct;91(4):437–8.

108. Toorians AW, Pneumatikos JA, Zaaijer HL, Strack van Schijndel RJ. Bilateral pleural effusion and a subsegmental infiltrate due to Chlamydia pneumoniae in a mechanically ventilated patient. Neth J Med. 2001 Aug;59(2):62–5.

109. Lin YT, Chen TL, Siu LK, Hsu SF, Fung CP. Clinical and microbiological

characteristics of community-acquired thoracic empyema or complicated parapneumonic effusion caused by Klebsiella pneumoniae in Taiwan. Eur J Clin Microbiol Infect Dis. 2010 May 27;29(8):1003–10.

110. Tu C-Y, Hsu W-H, Hsia T-C, Chen H-J, Chiu K-L, Hang L-W, et al. The changing pathogens of complicated parapneumonic effusions or empyemas in a medical intensive care unit. Intensive Care Med. 2006 Apr;32(4):570–6.

111. Gonzalez BE, Hulten KG, Dishop MK, Lamberth LB, Hammerman WA, Mason EO, et al. Pulmonary manifestations in children with invasive community-acquired Staphylococcus aureus infection. Clinical Infectious Diseases. Oxford University Press; 2005 Sep 1;41(5):583–90.

112. Thomas R, Ferguson J, Coombs G, Gibson PG. Community-acquired methicillin-resistant Staphylococcus aureus pneumonia: a clinical audit. Respirology. Blackwell Publishing Asia; 2011 Aug;16(6):926–31.

113. Carrillo-Marquez MA, Hulten KG, Hammerman W, Lamberth L, Mason EO, Kaplan SL. Staphylococcus aureus pneumonia in children in the era of community-acquired methicillin-resistance at Texas Children's Hospital. The Pediatric Infectious Disease Journal. 2011 Jul;30(7):545–50.

114. Ewig S, Bauer T, Hasper E, Marklein G, Kubini R, Luderitz B. Value of routine microbial investigation in community-acquired pneumonia treated in a tertiary care center. Respiration. 1996;63(3):164–9.

115. Woodhead MA, Arrowsmith J, Chamberlain-Webber R, Wooding S, Williams I. The value of routine microbial investigation in community-acquired pneumonia. Respiratory Medicine. 1991 Jul;85(4):313–7.

116. Sanyal S, Smith PR, Saha AC, Gupta S, Berkowitz L, Homel P. Initial microbiologic studies did not affect outcome in adults hospitalized with community-acquired pneumonia. American Journal of Respiratory and Critical Care Medicine. 1999 Jul;160(1):346–8.

117. Taylor E, Marrie T, Fine M, Obroskyl D, Kapoor W, Coley C, et al. Observations from a multicentre study on the use of the sputum specimen in patients hospitalized with community-acquired pneumonia. Can J Infect Dis. 1999 Jan;10(1):39–46.

118. García-Vázquez E, Marcos MA, Mensa J, de Roux A, Puig J, Font C, et al. Assessment of the usefulness of sputum culture for diagnosis of community-acquired pneumonia using the PORT predictive scoring system. Arch Intern Med. 2004 Sep 13;164(16):1807–11.

119. Navarro D, García-Maset L, Gimeno C, Escribano A, García-de-Lomas J, Spanish Pneumococcal Infection Study Network. Performance of the Binax NOW Streptococcus pneumoniae urinary antigen assay for diagnosis of pneumonia in children with underlying pulmonary diseases in the absence of acute pneumococcal infection. Journal of Clinical Microbiology. 2004

Oct;42(10):4853–5.

120. Menzies SM, Rahman NM, Wrightson JM, Davies HE, Shorten R, Gillespie SH, et al. Blood culture bottle culture of pleural fluid in pleural infection. Thorax. 2011 Aug;66(8):658–62.

121. Maataoui N, Bidet P, Doit C, De Lauzanne A, Lorrot M, Mariani-Kurkdjian P, et al. A multiplex polymerase chain reaction method for rapid pneumococcal serotype determination in childhood empyema. Diagnostic Microbiology and Infectious Disease. 2011 Mar;69(3):245–9.

122. Martinón-Torres F, Dosil-Gallardo S, Perez del Molino-Bernal ML, Sánchez FP, Tarrago D, Alvez F, et al. Pleural antigen assay in the diagnosis of pediatric pneumococcal empyema. J Crit Care. 2012 Jun;27(3):321.e1–4.

123. Strachan RE, Cornelius A, Gilbert GL, Gulliver T, Martin A, McDonald T, et al. A bedside assay to detect streptococcus pneumoniae in children with empyema. Pediatr Pulmonol. Wiley Subscription Services, Inc., A Wiley Company; 2011 Feb;46(2):179–83.

124. Sheppard CL, Harrison TG, Kearns AM, Guiver M, Creek M, Evans R, et al. Diagnosis of invasive pneumococcal infection by PCR amplification of Streptococcus pneumoniae genomic fragments in blood: a multi-centre comparative study. Commun Dis Public Health. 2003 Sep;6(3):221–7.

125. Menendez R, Torres A, Zalacain R, Aspa J, Martín Villasclaras JJ, Borderías L, et al. Risk factors of treatment failure in community acquired pneumonia: implications for disease outcome. Thorax. 2004 Nov;59(11):960–5.

126. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines on the Management of Community-Acquired Pneumonia in Adults. Clinical Infectious Diseases. 2007 Mar 1;44(Supplement 2):S27–S72.

127. Kollef MH, Sherman G, Ward S, Fraser VJ. Inadequate antimicrobial treatment of infections: a risk factor for hospital mortality among critically ill patients. Chest. 1999 Feb;115(2):462–74.

128. Rosón B, Carratalà J, Fernández-Sabé N, Tubau F, Manresa F, Gudiol F. Causes and factors associated with early failure in hospitalized patients with community-acquired pneumonia. Arch Intern Med. 2004 Mar 8;164(5):502–8.

129. Arancibia F, Ewig S, Martinez JA, Ruiz M, Bauer T, Marcos MA, et al. Antimicrobial treatment failures in patients with community-acquired pneumonia: causes and prognostic implications. American Journal of Respiratory and Critical Care Medicine. 2000 Jul;162(1):154–60.

130. Wise R, Hart T, Cars O, Streulens M, Helmuth R, Huovinen P, et al.

Antimicrobial resistance. Is a major threat to public health. BMJ. 1998 Sep 5;317(7159):609–10.

131. Gould IM. A review of the role of antibiotic policies in the control of antibiotic resistance. J Antimicrob Chemother. 1999 Apr;43(4):459–65.

132. Rahman NM, Singanayagam A, Davies HE, Wrightson JM, Mishra EK, Lee YCG, et al. Diagnostic accuracy, safety and utilisation of respiratory physician-delivered thoracic ultrasound. Thorax. 2010 Apr 30;65(5):449–53.

133. Duncan DR, Morgenthaler TI, Ryu JH, Daniels CE. Reducing Iatrogenic Risk in Thoracentesis: Establishing Best Practice Via Experiential Training in a Zero-Risk Environment. Chest. 2009 May 5;135(5):1315–20.

134. Raptopoulos V, Davis LM, Lee G, Umali C, Lew R, Irwin RS. Factors affecting the development of pneumothorax associated with thoracentesis. AJR Am J Roentgenol. 1991 May;156(5):917–20.

135. Barnes TW, Morgenthaler TI, Olson EJ, Hesley GK, Decker PA, Ryu JH. Sonographically guided thoracentesis and rate of pneumothorax. J Clin Ultrasound. 2005 Dec;33(9):442–6.

136. Diacon AH, Brutsche MH, Solèr M. Accuracy of pleural puncture sites: a prospective comparison of clinical examination with ultrasound. Chest. 2003 Feb;123(2):436–41.

137. Havelock T, Teoh R, Laws D, Gleeson F, on behalf of the BTS Pleural Disease Guideline Group. Pleural procedures and thoracic ultrasound: British Thoracic Society pleural disease guideline 2010. Thorax. 2010 Aug 9;65(Suppl 2):i61–i76.

138. Rapid Response Report - Risks of chest drain insertion [Internet]. NHS National Patient Safety Agency. London, UK: NHS National Patient Safety Agency; 2008 May. Report No.: 15th May 2008. Available from: http://www.nrls.npsa.nhs.uk/resources/type/alerts/?entryid45=59887

139. Vuori-Holopainen E, Salo E, Saxén H, Hedman K, Hyypiä T, Lahdenperä R, et al. Etiological diagnosis of childhood pneumonia by use of transthoracic needle aspiration and modern microbiological methods. Clinical Infectious Diseases. 2002 Mar 1;34(5):583–90.

140. Vuori-Holopainen E, Peltola H. Reappraisal of lung tap: review of an old method for better etiologic diagnosis of childhood pneumonia. Clin Infect Dis. 2001 Mar 1;32(5):715–26.

141. Dorca J, Manresa F, Esteban L, Barreiro B, Prats E, Ariza J, et al. Efficacy, safety, and therapeutic relevance of transthoracic aspiration with ultrathin needle in nonventilated nosocomial pneumonia. American Journal of Respiratory and Critical Care Medicine. 1995 May;151(5):1491–6.

142. García A, Rosón B, Pérez JL, Verdaguer R, Dorca J, Carratala J, et al.

Usefulness of PCR and antigen latex agglutination test with samples obtained by transthoracic needle aspiration for diagnosis of pneumococcal pneumonia. Journal of Clinical Microbiology. 1999 Jan 30;37(3):709–14.

143. Ishida T, Hashimoto T, Arita M, Osawa M, Tachibana H, Nishioka M, et al. Efficacy of transthoracic needle aspiration in community-acquired pneumonia. Intern Med. 2001 Sep;40(9):873–7.

144. Ruiz-González A, Falguera M, Nogues A, Rubio-Caballero M. Is Streptococcus pneumoniae the leading cause of pneumonia of unknown etiology? A microbiologic study of lung aspirates in consecutive patients with community-acquired pneumonia. The American Journal of Medicine. 1999 Apr;106(4):385–90.

145. Scott JAG, Hall AJ. The value and complications of percutaneous transthoracic lung aspiration for the etiologic diagnosis of community-acquired pneumonia. Chest. American College of Chest Physicians; 1999 Dec;116(6):1716–32.

146. Zalacain R, Llorente JL, Gaztelurrutia L, Pijoan JI, Sobradillo V. Influence of Three Factors on the Diagnostic Effectiveness of Transthoracic Needle Aspiration in Pneumonia. Chest. 1995 Jan 1;107(1):96–100.

147. Cloud JL, Carroll KC, Pixton P, Erali M, Hillyard DR. Detection of Legionella species in respiratory specimens using PCR with sequencing confirmation. Journal of Clinical Microbiology. 2000 May;38(5):1709–12.

148. Hardegger D, Nadal D, Bossart W, Altwegg M, Dutly F. Rapid detection of Mycoplasma pneumoniae in clinical samples by real-time PCR. Journal of Microbiological Methods. 2000 Jun;41(1):45–51.

149. Gaydos CA. What is the role of newer molecular tests in the management of CAP? Infect Dis Clin North Am. 2013 Mar;27(1):49–69.

150. Dorman SE, Chihota VN, Lewis JJ, Shah M, Clark D, Grant AD, et al. Performance characteristics of the Cepheid Xpert MTB/RIF test in a tuberculosis prevalence survey. Wilkinson RJ, editor. PLoS ONE. Public Library of Science; 2012;7(8):e43307.

151. Popowitch EB, Rogers E, Miller MB. Retrospective and prospective verification of the Cepheid Xpert influenza virus assay. Journal of Clinical Microbiology. American Society for Microbiology; 2011 Sep;49(9):3368–9.

152. Novak-Weekley SM, Marlowe EM, Poulter M, Dwyer D, Speers D, Rawlinson W, et al. Evaluation of the Cepheid Xpert Flu Assay for rapid identification and differentiation of influenza A, influenza A 2009 H1N1, and influenza B viruses. Journal of Clinical Microbiology. American Society for Microbiology; 2012 May;50(5):1704–10.

153. Ginocchio CC. Strengths and weaknesses of FDA-approved/cleared diagnostic devices for the molecular detection of respiratory pathogens. Clinical Infectious Diseases. Oxford University Press; 2011 May;52 Suppl

4(Supplement 4):S312–25.

154. Poritz MA, Blaschke AJ, Byington CL, Meyers L, Nilsson K, Jones DE, et al. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection. Costa C, editor. PLoS ONE. Public Library of Science; 2011;6(10):e26047.

155. Pierce VM, Elkan M, Leet M, McGowan KL, Hodinka RL. Comparison of the Idaho Technology FilmArray system to real-time PCR for detection of respiratory pathogens in children. Journal of Clinical Microbiology. American Society for Microbiology; 2012 Feb;50(2):364–71.

156. Wolk DM, Kaleta EJ, Wysocki VH. PCR-electrospray ionization mass spectrometry: the potential to change infectious disease diagnostics in clinical and public health laboratories. J Mol Diagn. Elsevier; 2012 Jul;14(4):295–304.

157. Butler JC, Bosshardt SC, Phelan M, Moroney SM, Tondella ML, Farley MM, et al. Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. J Infect Dis. 2003 May 1;187(9):1416–23.

158. Falguera M, López A, Nogués A, Porcel JM, Rubio-Caballero M. Evaluation of the polymerase chain reaction method for detection of Streptococcus pneumoniae DNA in pleural fluid samples. Chest. 2002 Dec;122(6):2212–6.

159. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991 Jan;173(2):697–703.

160. Woese CR. Bacterial evolution. Microbiol Rev. 1987 Jun;51(2):221–71.

161. Woese C. The universal ancestor. Proc Natl Acad Sci USA. 1998 Jun 9;95(12):6854–9.

162. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA. 1990 Jun;87(12):4576–9.

163. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research. 2009 Jan;37(Database issue):D141–5.

164. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research. 2007 Nov 14;35(21):7188–96.

165. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and

workbench compatible with ARB. Applied and Environmental Microbiology. 2006 Jul;72(7):5069–72.

166. Tunney MM, Klem ER, Fodor AA, Gilpin DF, Moriarty TF, McGrath SJ, et al. Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. Thorax. 2011 Jun 15;66(7):579–84.

167. Tunney MM, Einarsson GG, Wei L, Drain M, Klem ER, Cardwell C, et al. Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation. American Journal of Respiratory and Critical Care Medicine. 2013 May 15;187(10):1118–26.

168. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. American Journal of Respiratory and Critical Care Medicine. 2013 May 15;187(10):1067–75.

169. Pattison SH, Rogers GB, Crockard M, Elborn JS, Tunney MM. Molecular detection of CF lung pathogens: Current status and future potential. J Cyst Fibros. 2013 Feb 9.

170. Fodor AA, Klem ER, Gilpin DF, Elborn JS, Boucher RC, Tunney MM, et al. The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. PLoS ONE. 2012;7(9):e45001.

171. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine. 2012 May 15;185(10):1073–80.

172. Huang YJ, Nelson CE, Brodie EL, DeSantis TZ, Baek MS, Liu J, et al. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. J Allergy Clin Immunol. 2011 Feb;127(2):372–381.e1–3.

173. Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, et al. Disordered microbial communities in the upper respiratory tract of cigarette smokers. PLoS ONE. 2010;5(12):e15216.

174. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. American Journal of Respiratory and Critical Care Medicine. 2011 Oct 15;184(8):957–63.

175. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. Neyrolles O, editor. PLoS ONE. Public Library of Science; 2010;5(1):e8578.

176. Rogers GB, Skelton S, Serisier DJ, van der Gast CJ, Bruce KD.

Determining cystic fibrosis-affected lung microbiology: comparison of spontaneous and serially induced sputum samples by use of terminal restriction fragment length polymorphism profiling. Journal of Clinical Microbiology. American Society for Microbiology; 2010 Jan;48(1):78–86.

177. Sibley CD, Church DL, Surette MG, Dowd SE, Parkins MD. Pyrosequencing reveals the complex polymicrobial nature of invasive pyogenic infections: microbial constituents of empyema, liver abscess, and intracerebral abscess. Eur J Clin Microbiol Infect Dis. Springer-Verlag; 2012 Oct;31(10):2679–91.

178. Waterer GW, Rello J. Why should we measure bacterial load when treating community-acquired pneumonia? Current Opinion in Infectious Diseases. 2011 Apr;24(2):137–41.

179. Thomas CF, Limper AH. Current insights into the biology and pathogenesis of Pneumocystis pneumonia. Nat Rev Micro. 2007 Apr;5(4):298–308.

180. Maskell NA, Waine DJ, Lindley A, Pepperell JCT, Wakefield AE, Miller RF, et al. Asymptomatic carriage of Pneumocystis jiroveci in subjects undergoing bronchoscopy: a prospective study. Thorax. 2003 Jul;58(7):594–7.

181. Helweg-Larsen J, Jensen JS, Dohn B, Benfield TL, Lundgren B. Detection of Pneumocystis DNA in samples from patients suspected of bacterial pneumonia--a case-control study. BMC Infect Dis. 2002 Nov 25;2:28.

182. Medrano FJ, Montes-Cano M, Conde M, la Horra de C, Respaldiza N, Gasch A, et al. Pneumocystis jirovecii in general population. Emerg Infect Dis. 2005 Feb;11(2):245–50.

183. Ortona E, Margutti P, De Luca A, Peters SE, Wakefield AE, Tamburrini E, et al. Non specific PCR products using rat-derived Pneumocystis carinii dihydrofolate reductase gene-specific primers in DNA amplification of human respiratory samples. Molecular and Cellular Probes. 1996 Jun;10(3):187–90.

184. Wakefield AE, Guiver L, Miller RF, Hopkin JM. DNA amplification on induced sputum samples for diagnosis of Pneumocystis carinii pneumonia. Lancet. 1991 Jun 8;337(8754):1378–9.

185. Huggett JF, Taylor MS, Kocjan G, Evans HE, Morris-Jones S, Gant V, et al. Development and evaluation of a real-time PCR assay for detection of Pneumocystis jirovecii DNA in bronchoalveolar lavage fluid of HIV-infected patients. Thorax. 2007 Sep 17;63(2):154–9.

186. Joseph J, Strange C, Sahn SA. Pleural effusions in hospitalized patients with AIDS. Ann Intern Med. 1993 Jun 1;118(11):856–9.

187. Miller RF, Howling SJ, Reid AJ, Shaw PJ. Pleural effusions in patients with AIDS. Sex Transm Infect. 2000 Apr;76(2):122–5.

188. Stover DE, White DA, Romano PA, Gellene RA, Robeson WA. Spectrum of pulmonary diseases associated with the acquired immune deficiency syndrome. The American Journal of Medicine. 1985 Mar;78(3):429–37.

189. Suster B, Akerman M, Orenstein M, Wax MR. Pulmonary manifestations of AIDS: review of 106 episodes. Radiology. 1986 Oct;161(1):87–93.

190. Balachandran I, Jones DB, Humphrey DM. A case of Pneumocystis carinii in pleural fluid with cytologic, histologic and ultrastructural documentation. Acta Cytol. 1990 Jun;34(4):486–90.

191. Elwood LJ, Dobrzanski D, Feuerstein IM, Solomon D. Pneumocystis carinii in pleural fluid. The cytologic appearance. Acta Cytol. 1991 Oct;35(6):761–4.

192. Jayes RL, Kamerow HN, Hasselquist SM, Delaney MD, Parenti DM. Disseminated pneumocystosis presenting as a pleural effusion. Chest. 1993 Jan;103(1):306–8.

193. Ewig S, Rockstroh J. Diagnosis of Pneumocystis carinii infection in HIV-seropositive patients by identification of P carinii in pleural fluid. Chest. 1994 Aug;106(2):644.

194. Horowitz ML, Schiff M, Samuels J, Russo R, Schnader J. Pneumocystis carinii pleural effusion. Pathogenesis and pleural fluid analysis. Am Rev Respir Dis. 1993 Jul;148(1):232–4.

195. Schaumberg TH, Schnapp LM, Taylor KG, Golden JA. Diagnosis of Pneumocystis carinii infection in HIV-seropositive patients by identification of P carinii in pleural fluid. Chest. 1993 Jun;103(6):1890–1.

196. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry. 2009. pp. 611–22.

197. Huggett JF, Zumla A, Novak T, Garson JA, Green C, Morris-Jones SD, et al. Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. BMC Res Notes. 2008;1(1):70.

198. Thomas CF, Limper AH. Pneumocystis pneumonia. N Engl J Med. 2004 Jun 10;350(24):2487–98.

199. Walzer PD. Attachment of microbes to host cells: relevance of Pneumocystis carinii. Lab Invest. 1986 Jun;54(6):589–92.

200. Limper AH, Martin WJ. Pneumocystis carinii: inhibition of lung cell growth mediated by parasite attachment. J Clin Invest. 1990 Feb;85(2):391–6.

201. Kottom TJ, Köhler JR, Thomas CF, Fink GR, Limper AH. Lung epithelial cells and extracellular matrix components induce expression of Pneumocystis carinii STE20, a gene complementing the mating and pseudohyphal growth defects of STE20 mutant yeast. Infection and

Immunity. 2003 Nov;71(11):6463–71.

202. Cushion MT, Beck JM. Summary of Pneumocystis research presented at the 7th International Workshop on Opportunistic Protists. J Eukaryotic Microbiology. 2001;Suppl:101S–105S.

203. Cushion MT, Ebbets D. Growth and metabolism of Pneumocystis carinii in axenic culture. Journal of Clinical Microbiology. 1990 Jun;28(6):1385–94.

204. Grigore D, Meade JC. Functional complementation of the yeast P-type H-ATPase, PMA1, by the Pneumocystis carinii P-type H-ATPase, PCA1. J Eukaryotic Microbiology. 2006 May;53(3):157–64.

205. File TM Jr. Community-acquired pneumonia. The Lancet. 2003 Dec;362(9400):1991–2001.

206. Kenney RT, Li JS, Clyde WA, Wall TC, O'Connor CM, Campbell PT, et al. Mycoplasmal pericarditis: evidence of invasive disease. Clin Infect Dis. 1993 Aug;17 Suppl 1:S58–62.

207. Torrús Tendero D, Gutiérrez Fernández J, Díez Ruiz A, Bermúdez García JM, Rico Irles J. [Pleuropericarditis as the only manifestation of Legionella pneumophila infection]. Arch Bronconeumol. 1995 May;31(5):249–51.

208. Mühlemann K, Matter L, Meyer B, Schopfer K. Isolation of Coxiella burnetii from heart valves of patients treated for Q fever endocarditis. Journal of Clinical Microbiology. 1995 Feb;33(2):428–31.

209. Davis CP, Cochran S, Lisse J, Buck G, DiNuzzo AR, Weber T, et al. Isolation of Mycoplasma pneumoniae from synovial fluid samples in a patient with pneumonia and polyarthritis. Arch Intern Med. 1988 Apr;148(4):969–70.

210. Tully JG, Rose DL, Baseman JB, Dallo SF, Lazzell AL, Davis CP. Mycoplasma pneumoniae and Mycoplasma genitalium mixture in synovial fluid isolate. Journal of Clinical Microbiology. 1995 Jul;33(7):1851–5.

211. Bemer P, Leautez S, Ninin E, Jarraud S, Raffi F, Drugeon H. Legionella pneumophila Arthritis: use of medium specific for Mycobacteria for isolation of L. pneumophila in culture of articular fluid specimens. Clinical Infectious Diseases. 2002 Jul 1;35(1):E6–7.

212. Tan MJ, Tan JS, Hamor RH, File TM, Breiman RF. The radiologic manifestations of Legionnaire's disease. The Ohio Community-Based Pneumonia Incidence Study Group. Chest. 2000 Feb;117(2):398–403.

213. Macfarlane JT, Miller AC, Roderick Smith WH, Morris AH, Rose DH. Comparative radiographic features of community acquired Legionnaires' disease, pneumococcal pneumonia, mycoplasma pneumonia, and psittacosis. Thorax. 1984 Jan;39(1):28–33.

214. Fine NL, Smith LR, Sheedy PF. Frequency of pleural effusions in

mycoplasma and viral pneumonias. N Engl J Med. 1970 Oct 8;283(15):790–3.

215. McConnell CT, Plouffe JF, File TM, Mueller CF, Wong KH, Skelton SK, et al. Radiographic appearance of Chlamydia pneumoniae (TWAR strain) respiratory infections. CBPIS Study Group. Community-based Pneumonia Incidence Study. Radiology. 1994 Sep;192(3):819–24.

216. Voloudaki AE, Kofteridis DP, Tritou IN, Gourtsoyiannis NC, Tselentis YJ, Gikas AI. Q fever pneumonia: CT findings. Radiology. 2000 Jun;215(3):880–3.

217. Harrison TG, Taylor AG. Timing of seroconversion in Legionnaires' disease. Lancet. 1988 Oct 1;2(8614):795.

218. Lin B, Wang Z, Vora GJ, Thornton JA, Schnur JM, Thach DC, et al. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Research. 2006 Apr;16(4):527–35.

219. Bonfield JK, Smith KF, Staden R. A new DNA sequence assembly program. Nucleic Acids Research. 1995 Dec 25;23(24):4992–9.

220. Mediannikov O, Fenollar F, Socolovschi C, Diatta G, Bassene H, Molez J-F, et al. Coxiella burnetii in Humans and Ticks in Rural Senegal. Small PL, editor. PLoS Negl Trop Dis. 2010 Apr 6;4(4):e654.

221. Stein A, Raoult D. Detection of Coxiella burnetti by DNA amplification using polymerase chain reaction. Journal of Clinical Microbiology. 1992 Sep;30(9):2462–6.

222. Health Protection Agency, UK. Laboratory reports to HPA Colindale of infections due to Mycoplasma pneumoniae, England and Wales by date of report 1990-2011 (4 weekly). [Internet]. hpa.org.uk. [cited 2012 Jan 30]. Available from: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1194947359371

223. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. Journal of Clinical Microbiology. 2007 Sep;45(9):2761–4.

224. Engel LD, Kenny GE. Mycoplasma salivarium in human gingival sulci. J Periodont Res. 1970;5(3):163–71.

225. Johnson SM, Bruckner F, Collins D. Distribution of Mycoplasma pneumoniae and Mycoplasma salivarium in the synovial fluid of arthritis patients. Journal of Clinical Microbiology. 2007 Mar;45(3):953–7.

226. Ørsted I, Gertsen JB, Schønheyder HC, Jensen JS, Nielsen H. Mycoplasma salivarium isolated from brain abscesses. Clin Microbiol Infect. 2011 Jul;17(7):1047–9.

227. Baracaldo R, Foltzer M, Patel R, Bourbeau P. Empyema caused by

Mycoplasma salivarium. Journal of Clinical Microbiology. 2012 May;50(5):1805–6.

228. Alm EW, Oerther DB, Larsen N, Stahl DA, Raskin L. The oligonucleotide probe database. Applied and Environmental Microbiology. 1996 Oct;62(10):3557–9.

229. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research. 2012 Dec 26;41(1):e1–e1.

230. Pace NR. A molecular view of microbial diversity and the biosphere. Science. 1997 May 2;276(5313):734–40.

231. NIH HMP Working Group, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, et al. The NIH Human Microbiome Project. Genome Research. 2009 Dec;19(12):2317–23.

232. Blaser MJ. Harnessing the power of the human microbiome. Proceedings of the National Academy of Sciences. 2010 Apr 6;107(14):6125–6.

233. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010 Mar 4;464(7285):59–65.

234. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995 Mar;59(1):143–69.

235. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of Microbiological Methods. 2007 May;69(2):330–9.

236. Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L. Comparative Analysis of Human Gut Microbiota by Barcoded Pyrosequencing. Ahmed N, editor. PLoS ONE. 2008 Jul 30;3(7):e2836.

237. Van de Peer Y, Chapelle S, De Wachter R. A quantitative map of nucleotide substitution rates in bacterial rRNA. Nucleic Acids Research. 1996 Sep 1;24(17):3381–91.

238. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975 May 25;94(3):441–8.

239. Suzuki MT, Giovannoni SJ. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and Environmental Microbiology. 1996 Feb;62(2):625–30.

240. Zoetendal EG, Akkermans AD, de Vos WM. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals

stable and host-specific communities of active bacteria. Applied and Environmental Microbiology. 1998 Oct;64(10):3854–9.

241. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, et al. Experimental and analytical tools for studying the human microbiome. Nature Reviews Genetics. 2012 Jan;13(1):47–58.

242. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology. 2007 Aug;73(16):5261–7.

243. McOrist AL, Jackson M, Bird AR. A comparison of five methods for extraction of bacterial DNA from human faecal samples. Journal of Microbiological Methods. 2002 Jul;50(2):131–9.

244. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA. 1998 Feb 17;95(4):1460–5.

245. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology (Reading, Engl). 2002 Jan;148(Pt 1):257–66.

246. Mohagheghi A, Grohmann K, Himmel M, Leighton L, Updegraff DM. Isolation and characterization of Acidothermus cellulolyticus gen. nov., sp. nov., a new genus of thermophilic, acidophilic, cellulolytic bacteria. Int J Syst Bacteriol. Soc General Microbiol; 1986;36(3):435–43.

247. Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, et al. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology. 2001 Feb;67(2):880–7.

248. Wu J-Y, Jiang X-T, Jiang Y-X, Lu S-Y, Zou F, Zhou H-W. Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiol. BioMed Central Ltd; 2010 Oct 12;10(1):255.

249. QIAGEN. Qiagen HotStar HiFidelity Polymerase Kit [Internet]. qiagen.com. Hilden, Germany; [cited 2013 Jan 29]. Available from: http://www.qiagen.com/products/pcr/hotstartaqsystem/hotstarhifidelitypolymerasekit.aspx#Tabs=t1

250. Thermo Scientific. Phusion High-Fidelity DNA Polymerases [Internet]. thermoscientificbio.com. [cited 2013 Jan 29]. Available from: http://www.thermoscientificbio.com/phusion/

251. Vandenbroucke I, Van Marck H, Verhasselt P, Thys K, Mostmans W, Dumont S, et al. Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques. 2011 Sep;51(3):167–77.

252. Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, DeSantis TZ, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World J Gastroenterol. 2010 Sep 7;16(33):4135–44.

253. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Research. 2007;35(18):e120.

254. Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. Applied and Environmental Microbiology. 2008 Mar;74(5):1453–63.

255. Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Research. Oxford University Press; 2008 Oct;36(18):e120–0.

256. Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, Elshahed MS. Comparison of Species Richness Estimates Obtained Using Nearly Complete Fragments and Simulated Pyrosequencing-Generated Fragments in 16S rRNA Gene-Based Environmental Surveys. Applied and Environmental Microbiology. 2009 Aug 4;75(16):5227–36.

257. Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. Bereswill S, editor. PLoS ONE. 2010 Feb 5;5(2):e9085.

258. Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, et al. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. Genome Biol. BioMed Central Ltd; 2010;11(2):R15.

259. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010 Apr 11;7(5):335–6.

260. Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nature Methods. Nature Publishing Group; 2010 Sep;7(9):668–9.

261. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010 Oct 1;26(19):2460–1.

262. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal. 2012 Mar;6(3):610–8.

263. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, et al. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME Journal. 2012 Jan;6(1):94–103.

264. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010 Jan 15;26(2):266–7.

265. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490.

266. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology. 2005 Dec;71(12):8228–35.

267. Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z, et al. Oral microbiome profiles: 16S rRNA pyrosequencing and microarray assay comparison. PLoS ONE. 2011;6(7):e22788.

268. Amend AS, Seifert KA, Bruns TD. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol. 2010 Dec;19(24):5555–65.

269. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. Sci Transl Med. 2009 Dec 23;1(12):12ra23.

270. Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, Smidt H, et al. Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. 2009;4(8):e6669. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19693277&retmode=ref&cmd=prlinks

271. Pilloni G, Granitsiotis MS, Engel M, Lueders T. Testing the limits of 454 pyrotag sequencing: reproducibility, quantitative assessment and comparison to T-RFLP fingerprinting of aquifer microbes. PLoS ONE. 2012;7(7):e40467.

272. Wolcott RD, Gontcharova V, Sun Y, Dowd SE. Evaluation of the bacterial diversity among and within individual venous leg ulcers using bacterial tag-encoded FLX and Titanium amplicon pyrosequencing and metagenomic approaches. BMC Microbiol. 2009;9(1):226.

273. Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. Human gut microbiota in obesity and after gastric bypass. Proceedings of the National Academy of Sciences. 2009 Feb 17;106(7):2365–70.

274. Health Protection Agency, UK. Guidance on the Development and Validation of Diagnostic Tests that depend on Nucleic Acid Amplification and Detection. 2012 Sep pp. 1–52.

275. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE. 2011;6(12):e27310.

276. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. BioMed Central Ltd; 2007;8(7):R143.

277. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology. 2010 Jan;12(1):118–23.

278. Public Health England. UK Standards for Microbiological Investigations - Investigation of Fluids from Normally Sterile Sites. 5 ed. hpa.org.uk. London, UK: Standards Unit, Microbiology Services Division, Public Health England; 2012 Jul. Report No.: B 26.

279. Akan OA, Yildiz E. Comparison of the effect of delayed entry into 2 different blood culture systems (BACTEC 9240 and BacT/ALERT 3D) on culture positivity. Diagnostic Microbiology and Infectious Disease. 2006 Mar;54(3):193–6.

280. Reimer LG, Wilson ML, Weinstein MP. Update on detection of bacteremia and fungemia. Clinical Microbiology Reviews. 1997 Jul;10(3):444–65.

281. Saglani S, Harris KA, Wallis C, Hartley JC. Empyema: the use of broad range 16S rDNA PCR for pathogen detection. Archives of Disease in Childhood. 2005 Jan;90(1):70–3.

282. Kawanami T, Fukuda K, Yatera K, Kido M, Mukae H, Taniguchi H. A higher significance of anaerobes: the clone library analysis of bacterial pleurisy. Chest. 2011 Mar;139(3):600–8.

283. Magurran AE, McGill BJ. Biological Diversity. OUP Oxford; 2010. 1 p.

284. Center for Devices and Radiological Health. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. 13 ed. fda.gov. Center for Devices and Radiological Health, U.S. Food and Drug Administration; 2007 Mar. Report No.: FDA docket No. 2003D-0044.

285. Cato EP, Moore WEC, Nygaard G, Holdeman LV. Actinomyces meyeri sp. nov., Specific Epithet rev. Int J Syst Bacteriol. 1984;34(4):487–9.

286. Fazili T, Blair D, Riddell S, Kiska D, Nagra S. Actinomyces meyeri infection: Case report and review of the literature. Journal of Infection. 2012 Oct;65(4):357–61.

287. Olsen I, Johnson JL, Moore LVH, Moore WEC. Lactobacillus uli sp. nov. and Lactobacillus rimae sp. nov. From the Human Gingival Crevice and Emended Descriptions of Lactobacillus minutus and Streptococcus parvulus. Int J Syst Bacteriol. 1991 Apr;41(2):261–6.

288. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. Int J Syst Bacteriol. 1980;30(1):225–420.

289. Bartlett JG, Gorbach SL, Thadepalli H, Finegold SM. Bacteriology of

empyema. Lancet. 1974 Mar 2;1(7853):338–40.

290. Bartlett JG. Anaerobic bacterial infections of the lung. Chest. 1987 Jun;91(6):901–9.

291. Bartlett JG. Anaerobic bacterial infections of the lung and pleural space. Clin Infect Dis. 1993 Jun;16 Suppl 4:S248–55.

292. Boyanova L, Vladimir Djambazov, Gergova G, Dragomir Iotov, Petrov D, Osmanliev D, et al. Anaerobic microbiology in 198 cases of pleural empyema: a Bulgarian study. Anaerobe. 2004 Oct;10(5):261–7.

293. De A, Varaiya A, Mathur M. Anaerobes in pleuropulmonary infections. Indian J Med Microbiol. 2002 Jul;20(3):150–2.

294. Finegold SM, George WL, Mulligan ME. Anaerobic infections. Part I. Dis Mon. 1985 Oct;31(10):1–77.

295. Tanner ACR, Badger S, Lai CH, Listgarten MA, Visconti RA, Socransky SS. Wolinella gen. nov., Wolinella succinogenes (Vibrio succinogenes Wolin et al.) comb. nov., and Description of Bacteroides gracilis sp. nov., Wolinella recta sp. nov., Campylobacter concisus sp. nov., and Eikenella corrodens from Humans with Periodontal Disease. Int J Syst Bacteriol. Soc General Microbiol; 1981;31(4):432–45.

296. Johnson CC, Reinhardt JF, Edelstein MA, Mulligan ME, George WL, Finegold SM. Bacteroides gracilis, an important anaerobic bacterial pathogen. Journal of Clinical Microbiology. 1985 Nov;22(5):799–802.

297. Lee D, Goldstein EJ, Citron DM, Ross S. Empyema due to Bacteroides gracilis: case report and in vitro susceptibilities to eight antimicrobial agents. Clin Infect Dis. 1993 Jun;16 Suppl 4:S263–5.

298. Etoh Y, Dewhirst FE, Paster BJ, Yamamoto A, Goto N. Campylobacter showae sp. nov., isolated from the human oral cavity. Int J Syst Bacteriol. 1993 Oct;43(4):631–9.

299. Lam JYW, Wu AKL, Ngai DC, Teng JLL, Wong ESY, Lau SKP, et al. Three cases of severe invasive infections caused by Campylobacter rectus and first report of fatal C. rectus infection. Journal of Clinical Microbiology. 2011 Apr;49(4):1687–91.

300. Spiegel CA, Telford G. Isolation of Wolinella recta and Actinomyces viscosus from an actinomycotic chest wall mass. Journal of Clinical Microbiology. Am Soc Microbiol; 1984;20(6):1187–9.

301. Nakazawa F, Poco SE, Ikeda T, Sato M, Kalfas S, Sundqvist G, et al. Cryptobacterium curtum gen. nov., sp. nov., a new genus of Gram-positive anaerobic rod isolated from human oral cavities. Int J Syst Bacteriol. 1999 Jul 1;49(3):1193–200.

302. Holdeman LV, Cato EP, Burmeister JA, Moore WEC. Descriptions of

Eubacterium timidum sp. nov., Eubacterium brachy sp. nov., and Eubacterium nodatum sp. nov. isolated from human periodontitis. Int J Syst Bacteriol. Soc General Microbiol; 1980;30(1):163–9.

303. Rochford JC. Pleuropulmonary infection associated with Eubacterium brachy, a new species of Eubacterium. Journal of Clinical Microbiology. 1980 Nov;12(5):722–3.

304. Uematsu H, Nakazawa F, Ikeda T, Hoshino E. Eubacterium saphenus sp. nov., isolated from human periodontal pockets. Int J Syst Bacteriol. Soc General Microbiol; 1993;43(2):302–4.

305. Lauer E, Kandler O. Lactobacillus gasseri sp. nov., a new species of the subgenus Thermobacterium. Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. Reihe C. 1980;1:75-78.

306. Wang L-T, Kuo H-P, Wu Y-C, Tai C-J, Lee F-L. Lactobacillus taiwanensis sp. nov., isolated from silage. Int J Syst Bacteriol. 2009 Aug;59(Pt 8):2064–8.

307. Falsen E, Pascual C, Sjoden B, Ohlen M, Collins MD. Phenotypic and phylogenetic characterization of a novel Lactobacillus species from human sources: description of Lactobacillus iners sp. nov. Int J Syst Bacteriol. 1999 Jan 1;49(1):217–21.

308. Hansen PA, Lessel EF. Lactobacillus casei (Orla-Jensen) comb. nov. Int J Syst Bacteriol. Soc General Microbiol; 1971;21(1):69–71.

309. Zanoni P, Collins MD, Phillips BA. Deoxyribonucleic Acid Homology Studies of Lactobacillus casei, Lactobacillus paracasei sp. nov., subsp. paracasei and subsp. tolerans, and Lactobacillus rhamnosus sp. nov., comb. nov. Int J Syst Bacteriol. Soc General Microbiol; 1989;39(2):105–8.

310. Dicks LM, Plessis Du EM, Dellaglio F, Lauer E. Reclassification of Lactobacillus casei subsp. casei ATCC 393 and Lactobacillus rhamnosus ATCC 15820 as Lactobacillus zeae nom. rev., designation of ATCC 334 as the neotype of L. casei subsp. casei, and rejection of the name Lactobacillus paracasei. Int J Syst Bacteriol. 1996 Jan;46(1):337–40.

311. Gouriet F, Million M, Henri M, Fournier PE, Raoult D. Lactobacillus rhamnosus bacteremia: an emerging clinical entity. Eur J Clin Microbiol Infect Dis. 2012 Sep;31(9):2469–80.

312. Shoji H, Yoshida K, Niki Y. Lung abscess and pleuritis caused by Lactobacillus rhamnosus in an immunocompetent patient. J Infect Chemother. Springer Japan; 2010 Jan 14;16(1):45–8.

313. Munson MA, Pitt-Ford T, Chong B, Weightman A, Wade WG. Molecular and Cultural Analysis of the Microflora Associated with Endodontic Infections. Journal of Dental Research. 2002 Nov 1;81(11):761–6.

314. Edward DGFF. A suggested classification and nomenclature for organisms

of the pleuropneumonia group. Int J Syst Bacteriol. Soc General Microbiol; 1955;5(2):85–93.

315. Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, et al. Bacterial diversity in human subgingival plaque. J Bacteriol. 2001 Jun;183(12):3770–83.

316. Murdoch DA, Shah HN. Reclassification of Peptostreptococcus magnus (Prevot 1933) Holdeman and Moore 1972 as Finegoldia magna comb. nov. and Peptostreptococcus micros (Prevot 1933) Smith 1957 as Micromonas micros comb. nov. Anaerobe. Elsevier; 1999;5(5):555–9.

317. Shah HN, Collins MD. Proposal for reclassification of Bacteroides asaccharolyticus, Bacteroides gingivalis, and Bacteroides endodontalis in a new genus, Porphyromonas. Int J Syst Bacteriol. Soc General Microbiol; 1988;38(1):128–31.

318. Downes J, Sutcliffe I, Tanner ACR, Wade WG. Prevotella marshii sp. nov. and Prevotella baroniae sp. nov., isolated from the human oral cavity. Int J Syst Bacteriol. 2005 Jul;55(Pt 4):1551–5.

319. Holdeman LV, Moore WEC, Churn PJ, Johnson JL. Bacteroides oris and Bacteroides buccae New Species from Human Periodontitis and Other Human Infections. Int J Syst Bacteriol. Soc General Microbiol; 1982;32(1):125–31.

320. Willems A, Collins MD. Notes: 16S rRNA Gene Similarities Indicate that Hallella seregens (Moore and Moore) and Mitsuokella dentalis (Haapasalo et al.) Are Genealogically Highly Related and Are Members of the Genus Prevotella: Emended Description of the Genus Prevotella (Shah and Collins) and Description of Prevotella dentalis comb. nov. Int J Syst Bacteriol. Soc General Microbiol; 1995;45(4):832–6.

321. Shah HN, Collins DM. Notes: Prevotella, a new genus to include Bacteroides melaninogenicus and related species formerly classified in the genus Bacteroides. Int J Syst Bacteriol. Soc General Microbiol; 1990;40(2):205–8.

322. Okuda K, Kato T, Shiozu J, Takazoe I, Nakamura T. Bacteroides heparinolyticus sp. nov. isolated from humans with periodontitis. Int J Syst Bacteriol. Soc General Microbiol; 1985;35(4):438–42.

323. Holdeman LV, Johnson JL. Description of Bacteroides loescheii sp. nov. and emendation of the descriptions of Bacteroides melaninogenicus (Oliver and Wherry) Roy and Kelly 1939 and Bacteroides denticola Shah and Collins 1981. Int J Syst Bacteriol. Soc General Microbiol; 1982;32(4):399–409.

324. Petty BG, Smith CR. The syndrome of inappropriate secretion of antidiuretic hormone associated with anaerobic thoracic empyema. Am Rev Respir Dis. 1977 Apr;115(4):685–8.

325. Shah HN, Gharbia SE. Biochemical and chemical studies on strains designated Prevotella intermedia and proposal of a new pigmented species, Prevotella nigrescens sp. nov. Int J Syst Bacteriol. 1992 Oct;42(4):542–6.

326. Mättö J, Asikainen S, Väisänen ML, Rautio M, Saarela M, Summanen P, et al. Role of Porphyromonas gingivalis, Prevotella intermedia, and Prevotella nigrescens in extraoral and some odontogenic infections. Clin Infect Dis. 1997 Sep;25 Suppl 2:S194–8.

327. Bolivar I, Whiteson K, Stadelmann B, Baratti-Mayer D, Gizard Y, Mombelli A, et al. Bacterial diversity in oral samples of children in niger with acute noma, acute necrotizing gingivitis, and healthy controls. PLoS Negl Trop Dis. 2012;6(3):e1556.

328. Sakamoto M, Ohkusu K, Masaki T, Kako H, Ezaki T, Benno Y. Prevotella pleuritidis sp. nov., isolated from pleural fluid. Int J Syst Bacteriol. 2007 Aug;57(Pt 8):1725–8.

329. Downes J, Wade WG. Prevotella fusca sp. nov. and Prevotella scopos sp. nov., isolated from the human oral cavity. Int J Syst Bacteriol. 2011 Apr;61(Pt 4):854–8.

330. Tanner ACR, Listgarten MA, Ebersole JL, Strzempko MN. Bacteroides forsythus sp. nov., a slow-growing, fusiform Bacteroides sp. from the human oral cavity. Int J Syst Bacteriol. Soc General Microbiol; 1986;36(2):213–21.

331. Chan ECS, Siboo R, Keng T, Psarra N, Hurley R, Cheng SL, et al. Treponema denticola (ex Brumpt 1925) sp. nov., nom. rev., and identification of new spirochete isolates from periodontal pockets. Int J Syst Bacteriol. 1993 Apr;43(2):196–203.

332. Wyss C, Choi BK, Schupbach P, Moter A, Guggenheim B, Göbel UB. Treponema lecithinolyticum sp. nov., a small saccharolytic spirochaete with phospholipase A and C activities associated with periodontal diseases. Int J Syst Bacteriol. 1999 Oct;49 Pt 4:1329–39.

333. Wyss C, Choi BK, Schupbach P, Guggenheim B, Göbel UB. Treponema maltophilum sp. nov., a small oral spirochete isolated from human periodontal lesions. Int J Syst Bacteriol. 1996 Jul;46(3):745–52.

334. Umemoto T, Nakazawa F, Hoshino E, Okada K, Fukunaga M, Namikawa I. Treponema medium sp. nov., isolated from human subgingival dental plaque. Int J Syst Bacteriol. 1997 Jan;47(1):67–72.

335. Smibert RM, Johnson JL, Ranney RR. Treponema socranskii sp. nov., Treponema socranskii subsp. socranskii subsp. nov., Treponema socranskii subsp. buccale subsp. nov., and Treponema socranskii subsp. paredis subsp. nov. isolated from the human periodontia. Int J Syst Bacteriol. Soc General Microbiol; 1984;34(4):457–62.

336. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, et al. The human oral microbiome. J Bacteriol. 2010 Oct;192(19):5002–17.

337. Bartlett JG. Anaerobic bacterial infection of the lung. Anaerobe. 2012 Apr;18(2):235–9.

338. Hirsch RS, Clarke NG. Infection and periodontal diseases. Rev Infect Dis. 1989 Sep;11(5):707–15.

339. Socransky SS, Smith C, Haffajee AD. Subgingival microbial profiles in refractory periodontal disease. J Clin Periodontol. 2002 Mar;29(3):260–8.

340. Smith DT. Experimental aspiratory abscess. Archives of Surgery. Am Med Assoc; 1927;14(1):231.

341. Cremades R, Galiana A, Rodriguez JC, Santos A, Lopez P, Ruiz M, et al. Identification of bacterial DNA in noninfectious pleural fluid with a highly sensitive PCR method. Respiration. 2011;82(2):130–5.

342. Scott JAG, Marston EL, Hall AJ, Marsh K. Diagnosis of pneumococcal pneumonia by psaA PCR analysis of lung aspirates from adult patients in Kenya. Journal of Clinical Microbiology. American Society for Microbiology; 2003 Jun;41(6):2554–9.

343. Manresa F, Dorca J. Needle aspiration techniques in the diagnosis of pneumonia. Thorax. 1991 Aug;46(8):601–3.

344. Ideh RC, Howie SRC, Ebruke B, Secka O, Greenwood BM, Adegbola RA, et al. Transthoracic lung aspiration for the aetiological diagnosis of pneumonia: 25 years of experience from The Gambia. Int J Tuberc Lung Dis. 2011 Jun;15(6):729–35.

345. Falguera M, Ruiz-González A, Puig T, Nogues A, Garcia M. Detection of Mycoplasma pneumoniae by Polymerase Chain Reaction in Lung Aspirates From Patients With Community-Acquired Pneumonia. Chest. 1996 Oct 1;110(4):972–6.

346. Carrol ED, Mankhambo LA, Guiver M, Banda DL, IPD Study Group, Denis B, et al. PCR improves diagnostic yield from lung aspiration in Malawian children with radiologically confirmed pneumonia. Nizami Q, editor. PLoS ONE. 2011;6(6):e21042.

347. Bella F, Tort J, Morera MA, Espaulella J, Armengol J. Value of bacterial antigen detection in the diagnostic yield of transthoracic needle aspiration in severe community acquired pneumonia. Thorax. 1993 Dec;48(12):1227–9.

348. Hernes SS, Hagen E, Tofteland S, Finsen NT, Christensen A, Giske CG, et al. Transthoracic fine-needle aspiration in the aetiological diagnosis of community-acquired pneumonia. Clin Microbiol Infect. 2010 Jul;16(7):909–11.

349. Stevens DA, Hamilton JR, Johnson N, Kim KK, Lee J-S. Halomonas, a newly recognized human pathogen causing infections and contamination in a dialysis center: three new species. Medicine (Baltimore). 2009 Jul;88(4):244–9.

350. Vignier N, Barreau M, Olive C, Baubion E, Théodose R, Hochedez P, et al. Human infection with Shewanella putrefaciens and S. algae: report of 16 cases in Martinique and review of the literature. Am J Trop Med Hyg. 2013 Jul;89(1):151–6.

351. Jacob-Kokura S, Chan CY, Kaplan L. Bacteremia and Empyema Caused by Shewanella algae in a Trauma Patient. Ann Pharmacother. 2014 Jan;48(1):128–36.

352. Madhi SA, Klugman KP, The Vaccine Trialist Group. A role for Streptococcus pneumoniae in virus-associated pneumonia. Nat Med. 2004 Jul 11;10(8):811–3.

353. Zhou H, Haber M, Ray S, Farley MM, Panozzo CA, Klugman KP. Invasive Pneumococcal Pneumonia and Respiratory Virus Co-infections. Emerg Infect Dis. 2012 Feb;18(2):294–7.

354. Madhi SA, Ludewick H, Kuwanda L, Niekerk NV, Cutland C, Little T, et al. Pneumococcal coinfection with human metapneumovirus. J Infect Dis. 2006 May 1;193(9):1236–43.

355. Mühl H, Kochem A-J, Disqué C, Sakka SG. Activity and DNA contamination of commercial polymerase chain reaction reagents for the universal 16S rDNA real-time polymerase chain reaction detection of bacterial pathogens in blood. Diagnostic Microbiology and Infectious Disease. 2010 Jan;66(1):41–9.

356. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB, Fox AJ. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. Journal of Clinical Microbiology. 2000 May;38(5):1747–52.

357. Newsome T, Li B-J, Zou N, Lo S-C. Presence of bacterial phage-like DNA sequences in commercial Taq DNA polymerase reagents. Journal of Clinical Microbiology. 2004 May;42(5):2264–7.

358. Băncescu G, Dumitriu S, Băncescu A, Skaug N. Streptococci species of anginosus group isolated from oral and maxillofacial infections. Roum Arch Microbiol Immunol. 1999 Jan;58(1):49–55.

359. Shinzato T, Saito A. The Streptococcus milleri group as a cause of pulmonary infections. Clin Infect Dis. 1995 Dec;21 Suppl 3:S238–43.

360. Porta G, Rodríguez-Carballeira M, Gómez L, Salavert M, Freixas N, Xercavins M, et al. Thoracic infection caused by Streptococcus milleri. Eur Respir J. 1998 Aug;12(2):357–62.

361. Willcox MD. Potential pathogenic properties of members of the

"Streptococcus milleri" group in relation to the production of endocarditis and abscesses. Journal of Medical Microbiology. 1995 Dec;43(6):405–10.

362. Shinzato T, Saito A. A mechanism of pathogenicity of "Streptococcus milleri group" in pulmonary infection: synergy with an anaerobe. Journal of Medical Microbiology. 1994 Feb;40(2):118–23.

363. Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. PLoS ONE. 2012;7(4):e34605.

364. Witt N, Rodger G, Vandesompele J, Benes V, Zumla A, Rook GA, et al. An assessment of air as a source of DNA contamination encountered when performing PCR. J Biomol Tech. 2009 Dec;20(5):236–40.

365. Goldenberger D, Altwegg M. Eubacterial PCR: contaminating DNA in primer preparations and its elimination by UV light. Journal of Microbiological Methods. Elsevier; 1995;21(1):27–32.

16 Appendix – Altered amplicon pipeline settings

The following settings were used for the Roche `runAnalysisPipeAmplicons` command line interface launch script

Default amplicon processing settings	Modified amplicon processing settings used in this study
`default_template.xml`	`modified_template.xml`
`<recursiveCafieCorrector>` `<enable>false</enable>`	`<recursiveCafieCorrector>` `<enable>true</enable>`
`<vfBadFlowThreshold>4</vfBadFlowThreshold>`	`<vfBadFlowThreshold>6</vfBadFlowThreshold>`
`<vfScanAllFlows>tiOnly</vfScanAllFlows>`	`<vfScanAllFlows>false</vfScanAllFlows>`

Command to launch script

```
runAnalysisPipeAmplicons --pipe=$PATH/modified_template.xml

$PATH/File_imageProcessingOnly &
```

17 Appendix – Representative scripts used in QIIME

17.1 Main script

```
#!/bin/bash
set -e -x                                          # Enables verbose mode
exec 3>&1 4>&2 >Scriptlog_PLATExx 2>&1             # Enables logging

REF="/home/ubuntu"
DIR="/home/ubuntu/data/PLATExx"
NINETYSEVENDIR="/home/ubuntu/data/PLATExx/97"

# Check mapping files for errors and copying to correct directory
check_id_map.py -m Lane01ForwardMapping.txt -o Lane01ForwardMappingOutput
check_id_map.py -m Lane01ReverseMapping.txt -o Lane01ReverseMappingOutput
check_id_map.py -m Lane02ForwardMapping.txt -o Lane02ForwardMappingOutput
check_id_map.py -m Lane02ReverseMapping.txt -o Lane02ReverseMappingOutput
check_id_map.py -m Lane03ForwardMapping.txt -o Lane03ForwardMappingOutput
check_id_map.py -m Lane03ReverseMapping.txt -o Lane03ReverseMappingOutput
check_id_map.py -m Lane04ForwardMapping.txt -o Lane04ForwardMappingOutput
check_id_map.py -m Lane04ReverseMapping.txt -o Lane04ReverseMappingOutput
mkdir Lane01
mkdir Lane02
mkdir Lane03
mkdir Lane04
cp Lane01ForwardMapping.txt Lane01/
cp Lane01ReverseMapping.txt Lane01/
cp Lane02ForwardMapping.txt Lane02/
cp Lane02ReverseMapping.txt Lane02/
cp Lane03ForwardMapping.txt Lane03/
cp Lane03ReverseMapping.txt Lane03/
cp Lane04ForwardMapping.txt Lane04/
cp Lane04ReverseMapping.txt Lane04/

# Convert .sff files to .fasta and .qual files using Roche sffinfo tool
./sffinfo    HV01QUF01.sff > Lane01/Lane01.sff.txt
./sffinfo -s HV01QUF01.sff > Lane01/Lane01.fasta
./sffinfo -q HV01QUF01.sff > Lane01/Lane01.qual

./sffinfo    HV01QUF02.sff > Lane02/Lane02.sff.txt
./sffinfo -s HV01QUF02.sff > Lane02/Lane02.fasta
./sffinfo -q HV01QUF02.sff > Lane02/Lane02.qual

./sffinfo    HV01QUF03.sff > Lane03/Lane03.sff.txt
./sffinfo -s HV01QUF03.sff > Lane03/Lane03.fasta
./sffinfo -q HV01QUF03.sff > Lane03/Lane03.qual

./sffinfo    HV01QUF04.sff > Lane04/Lane04.sff.txt
./sffinfo -s HV01QUF04.sff > Lane04/Lane04.fasta
./sffinfo -q HV01QUF04.sff > Lane04/Lane04.qual


# Processing Lane01 (similar commands for further lanes)
cd Lane01/

# split_libraries.py enables demultiplexing for both forward and reverse sequences
# Mapping file specifies Multiplex identifier (MID) tags and forward and reverse primers
# -l 200 (Minimum sequence length accepted = 200nt), -L 650 (Maximum sequence length accepted = 650nt)
# -M 2 (2 primer mismatches accepted); -b 10 (MID length = 10nt);
# -w 50 (sliding window quality score using value 50); -z (remove reverse primers and subsequent sequences)
split_libraries.py -m Lane01ForwardMapping.txt -f Lane01.fasta -q Lane01.qual -l 200 -L 650
-M 2 -b 10 -o Forward -w 50 -z truncate_only --reverse_primer_mismatches 2 -d
split_libraries.py -m Lane01ReverseMapping.txt -f Lane01.fasta -q Lane01.qual -l 200 -L 650
-M 2 -b 10 -o Reverse -w 50 -z truncate_only --reverse_primer_mismatches 2 -d

# 454 denoiser based on flowgram clustering
denoise_wrapper.py -i Lane01.sff.txt -f Forward/seqs.fna -o Forward/denoised/ -m
Lane01ForwardMapping.txt -n 8 --titanium
cd Forward/
inflate_denoiser_output.py -c denoised/centroids.fasta -s denoised/singletons.fasta -f
seqs.fna -d denoised/denoiser_mapping.txt -o inflated_denoised_seqs.fna
cd ..
```

```
denoise_wrapper.py -i Lane01.sff.txt -f Reverse/seqs.fna -o Reverse/denoised/ -m
Lane01ReverseMapping.txt -n 8 --titanium
cd Reverse/
inflate_denoiser_output.py -c denoised/centroids.fasta -s denoised/singletons.fasta -f
seqs.fna -d denoised/denoiser_mapping.txt -o inflated_denoised_seqs.fna
sed -i 's/Sample_/Sample_100000/g' inflated_denoised_seqs.fna
adjust_seq_orientation.py -i inflated_denoised_seqs.fna
cd ..
```

 # Combine forward and reverse sequences (reverse complemented) into one file

```
cat Forward/inflated_denoised_seqs.fna Reverse/inflated_denoised_seqs_rc.fna >
../Lane01combined_seqs.fna
cd ..
```

 # …other lanes processed identically (not shown)

 # Combine sequences from all lanes

```
cat Lane01combined_seqs.fna Lane02combined_seqs.fna Lane03combined_seqs.fna
Lane04combined_seqs.fna > combined_seqs.fna
```

 # OTU picking by clustering sequences based on 97% similarity to Greengenes core database using uclust. This is a closed-reference process, essential given bidirectional sequencing and also acts to remove chimeras.

```
parallel_pick_otus_uclust_ref.py -i $DIR/combined_seqs.fna -o
$NINETYSEVENDIR/uclust_ref_picked_otus -r $REF/97_otus.fasta -s 0.97
```

 # Picks the most abundant sequences for each OTU

```
pick_rep_set.py -i $NINETYSEVENDIR/uclust_ref_picked_otus/combined_seqs_otus.txt -f
$DIR/combined_seqs.fna -o $NINETYSEVENDIR/uclust_ref_picked_otus/rep_set.fna -m
most_abundant
```

 # Aligns representative sequences using PyNAST (a python implementation of the NAST alignment algorithm)

```
parallel_align_seqs_pynast.py -i $NINETYSEVENDIR/uclust_ref_picked_otus/rep_set.fna -o
$NINETYSEVENDIR/uclust_ref_picked_otus/pynast_aligned/
```

 # Assign taxonomy to representative sequences using the RDP classifier and genus-level taxonomic mapping file. This assigns taxonomies by matching sequence segments of length 8 to a database of assigned sequences with a minimum confidence score of 0.8(242)

```
parallel_assign_taxonomy_rdp.py -t $REF/97_otu_taxonomy_rdp.txt -r $REF/97_otus.fasta --
rdp_max_memory 5000 -i $NINETYSEVENDIR/uclust_ref_picked_otus/rep_set.fna -o
$NINETYSEVENDIR/uclust_ref_picked_otus/rdp22_assigned_taxonomy/
```

 # Script to remove sequence positions that are gaps in every sequence, and to remove positions which are typically non-conserved (and therefore uninformative for tree building) using lanemask file

```
filter_alignment.py -i
$NINETYSEVENDIR/uclust_ref_picked_otus/pynast_aligned/rep_set_aligned.fasta -m
/home/ubuntu/qiime_software/lanemask_in_1s_and_0s -o
$NINETYSEVENDIR/uclust_ref_picked_otus/pynast_aligned/filtered_alignment/
```

 # Phylogenetic tree construction for representative sequences using FastTree(265)

```
make_phylogeny.py -i
$NINETYSEVENDIR/uclust_ref_picked_otus/pynast_aligned/filtered_alignment/rep_set_aligned_pf
iltered.fasta -o
$NINETYSEVENDIR/uclust_ref_picked_otus/pynast_aligned/filtered_alignment/rep_phylo.tre
```

 # OTU table construction

```
make_otu_table.py -i $NINETYSEVENDIR/uclust_ref_picked_otus/combined_seqs_otus.txt -o
$NINETYSEVENDIR/uclust_ref_picked_otus/otu_table.biom -t
$NINETYSEVENDIR/uclust_ref_picked_otus/rdp22_assigned_taxonomy/rep_set_tax_assignments.txt

per_library_stats.py -i $NINETYSEVENDIR/uclust_ref_picked_otus/otu_table.biom -m
$DIR/CombinedMapping.txt -o
$NINETYSEVENDIR/uclust_ref_picked_otus/OTU_Stats_CombinedMapping.txt --otu_counts

per_library_stats.py -i $NINETYSEVENDIR/uclust_ref_picked_otus/otu_table.biom -m
$DIR/CombinedMapping.txt -o
$NINETYSEVENDIR/uclust_ref_picked_otus/Stats_CombinedMapping.txt
```

 # Summarises taxa at Order (L4), Family (L5) and Genus (L6) taxonomic levels

```
summarize_taxa.py -i $NINETYSEVENDIR/uclust_ref_picked_otus/otu_table.biom -o
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary -L 4,5,6 -a
```

 # Plots graphical representations of taxa summaries

```
plot_taxa_summary.py -i
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary/otu_table_L4.txt -o
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary/taxa_summary_plots_L4 -c bar,pie
```

```
plot_taxa_summary.py -i
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary/otu_table_L5.txt -o
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary/taxa_summary_plots_L5 -c bar,pie
plot_taxa_summary.py -i
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary/otu_table_L6.txt -o
$NINETYSEVENDIR/uclust_ref_picked_otus/wf_taxa_summary/taxa_summary_plots_L6 -c bar,pie
```

17.2 Script to generate alpha and beta diversity measures

```
#!/bin/bash
set -e -x
exec 3>&1 4>&2 >Diversitylog_PLATExx 2>&1

echo "alpha_diversity:metrics
brillouin_d,michaelis_menten_fit,simpson,shannon,PD_whole_tree,chao1,chao1_confidence,obser
ved_species" > alpha_params.txt
```

 # Generates rarefied OTU tables, compute alpha diversity metric, collate results and generate plots (using maximum rarefaction depth specified by –e)

```
alpha_rarefaction.py -i otu_table.biom -m CombinedMapping.txt -o wf_arare/ -p
alpha_params.txt -t rep_phylo.tre -a -O 7 -e 1062
```

 # Performs beta diversity and principal coordinate analysis, subsampled at –e number of sequences per sample

```
beta_diversity_through_plots.py -i otu_table.biom -m CombinedMapping.txt -o
wf_bdiv_even1062/ -t rep_phylo.tre -a -O 7 -e 1062
```

 # Perform weighted and unweighted 3D principal coordinate analysis plots and add taxa from taxon summary files, in which tax are plotted as a weighted average of coordinates of all samples, weighted by relative abundance of taxa in samples. Size of a taxon sphere is proportional to mean relative abundance of taxon across all samples

```
make_3d_plots.py -i wf_bdiv_even1062/unweighted_unifrac_pc.txt -m CombinedMapping.txt -t
wf_taxa_summary/otu_table_L4.txt -o 3d_biplot_unweighted -k white -s scaled
make_3d_plots.py -i wf_bdiv_even1062/weighted_unifrac_pc.txt -m CombinedMapping.txt -t
wf_taxa_summary/otu_table_L4.txt -o 3d_biplot_weighted -k white -s scaled
```

17.3 Further QIIME commands used

 # Convert OTU file from *.txt to *.biom format

```
convert_biom.py -i otu_table.txt -o otu_table.biom --biom_table_type="otu table" --
process_obs_metadata taxonomy
```

 # Convert OTU table from *.biom to *.txt format

```
convert_biom.py -i otu_table.biom -o otu_table.txt --biom_table_type "otu table" --
header_key "taxonomy" -b
```

 # Filter samples without any OTUs from OTU table

```
filter_samples_from_otu_table.py -i otu_table_unfiltered.biom -o otu_table.biom -n 1
```

 # Filter OTUs not represented by any samples from OTU table

```
filter_otus_from_otu_table.py -i otu_table_unfiltered.biom -o otu_table.biom -n 1
```

 # Generate Topiary Explorer project files

```
make_tep.py -i otu_table.biom -m CombinedMapping.txt -t rep_phylo.tre
```

18 Appendix – Representative scripts used in phyloseq in R

18.1 Heatmap analysis of raw sequencing data

```
library("phyloseq")
library("ape")
library("ggplot2")
library("scales")
library("grid")
uzdir <- "MIST2raw/"
biom_file <- paste(uzdir, "otu_table.biom", sep = "")
tree_file <- paste(uzdir, "rep_phylo.tre", sep = "")
map_file <- paste(uzdir, "CombinedMapping.txt", sep = "")
my_biom_withoutmapping <- import_biom(biom_file, tree_file, parseFunction =
parse_taxonomy_greengenes)
my_biom_withoutmapping
my_mappingfile <- import_qiime_sample_data(map_file)
class(my_mappingfile)
dim(my_mappingfile)
my_biom <- merge_phyloseq(my_biom_withoutmapping, my_mappingfile)
my_biom

set.seed(2)
my_biom_rarefy <- rarefy_even_depth(my_biom)
my_biom_relativeabundance = transform_sample_counts(my_biom_rarefy, function(x) x/sum(x))
set.seed(2)
phy_tree(my_biom_relativeabundance) <- root(phy_tree(my_biom_relativeabundance),
sample(taxa_names(my_biom_relativeabundance),1), resolve.root = TRUE)
phy_tree(my_biom_relativeabundance)
sample_variables(my_biom_relativeabundance)

wh001 <- genefilter_sample(my_biom_relativeabundance, filterfun_sample(function(x) x >
0.05), A = 1)
my_biom_relativeabundance_filter001 <- prune_taxa(wh001, my_biom_relativeabundance)

r <- plot_heatmap(my_biom_relativeabundance_filter001, "NMDS", "bray",
"Culture_Classification", "Family", low = "#000033", high = "#FF3300",
trans=identity_trans(), title="Heatmap comparing sample culture with sequencing abundance
(using linear scale)")
r + geom_tile(linetype = 0)
Sys.sleep(5)
ggsave(file="MIST2RawIdentity.pdf", width=11.69, height=8.27)

s <- plot_heatmap(my_biom_relativeabundance_filter001, "NMDS", "bray",
"Culture_Classification", "Family", low = "#000033", high = "#FF3300", trans=log_trans(10),
title="Heatmap comparing sample culture with sequencing abundance (using log(10) scale)")
s + geom_tile(linetype = 0)
Sys.sleep(5)
ggsave(file="MIST2RawLog.pdf", width=11.69, height=8.27)
```

18.2 Heatmap analysis of post-threshold sequence data

```
library("phyloseq")
library("ape")
library("ggplot2")
library("scales")
library("grid")
uzdir <- "MIST2_postfiltering/"
biom_file <- paste(uzdir, "otu_table.biom", sep = "")
tree_file <- paste(uzdir, "rep_phylo.tre", sep = "")
map_file <- paste(uzdir, "CombinedMapping.txt", sep = "")
my_biom_withoutmapping <- import_biom(biom_file, tree_file, parseFunction =
parse_taxonomy_greengenes)
my_biom_withoutmapping
my_mappingfile <- import_qiime_sample_data(map_file)
class(my_mappingfile)
dim(my_mappingfile)
my_biom <- merge_phyloseq(my_biom_withoutmapping, my_mappingfile)
```

```
my_biom

set.seed(2)
my_biom_rarefy <- rarefy_even_depth(my_biom)
my_biom_relativeabundance = transform_sample_counts(my_biom_rarefy, function(x) x/sum(x))
set.seed(2)
phy_tree(my_biom_relativeabundance) <- root(phy_tree(my_biom_relativeabundance),
sample(taxa_names(my_biom_relativeabundance),1), resolve.root = TRUE)
phy_tree(my_biom_relativeabundance)
sample_variables(my_biom_relativeabundance)

p <- plot_heatmap(my_biom_relativeabundance, "NMDS", "bray", "Culture_Classification",
"Family", low = "#000033", high = "#FF3300", trans=identity_trans(), title="Heatmap
comparing sample culture with post-threshold sequencing abundance (using linear scale)")
p + geom_tile(linetype = 0)
Sys.sleep(5)
ggsave(file="MIST2ThresholdIdentity.pdf", width=11.69, height=8.27)

p <- plot_heatmap(my_biom_relativeabundance, "NMDS", "bray", "Culture_Classification",
"Family", low = "#000033", high = "#FF3300", trans=log_trans(10), title="Heatmap comparing
sample culture with post-threshold sequencing abundance (using log(10) scale)")
p + geom_tile(linetype = 0)
Sys.sleep(5)
ggsave(file="MIST2ThresholdLog.pdf", width=11.69, height=8.27)
```

19 Appendix – Sample mapping files

19.1 Example combined mapping file

(CombinedMapping.txt)

#SampleID	ClassificationA	ClassificationB	CulturePos	Culture	Description
Plate04.Lane01.MID01.Sample	PleuralInfection	MIST2	Positive	StreptococcusMilleri	Plate04.Lane01.MID01.Sample
Plate04.Lane01.MID02.Sample	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID02.Sample
Plate04.Lane01.MID23.Sample	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID23.Sample
Plate04.Lane01.MID24.Sample	PleuralInfection	MIST2	Positive	StreptococcusMilleri	Plate04.Lane01.MID24.Sample
Plate04.Lane02.MID01.Sample	PleuralInfection	MIST2	NG	NG	Plate04.Lane02.MID01.Sample
Plate04.Lane02.MID02.Sample	PleuralInfection	MIST2	NG	NG	Plate04.Lane02.MID02.Sample

etc…

19.2 Example mapping file for forward sequencing

(Lane01ForwardMapping.txt)

#SampleID	BarcodeSequence	LinkerPrimerSequence	ReversePrimer	ClassificationA	ClassificationB	CulturePos	Culture	Description
Plate04.Lane01.MID01.Sample	ACGAGTGCGT	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	Positive	StreptococcusMilleri	Plate04.Lane01.MID01.Sample
Plate04.Lane01.MID02.Sample	ACGCTCGACA	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID02.Sample
Plate04.Lane01.MID03.Sample	AGACGCACTC	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	Positive	MRSA	Plate04.Lane01.MID03.Sample
Plate04.Lane01.MID04.Sample	AGCACTGTAG	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	Positive	AnaerobicStreptococci	Plate04.Lane01.MID04.Sample
Plate04.Lane01.MID05.Sample	ATCAGACACG	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	Positive	Ecoli	Plate04.Lane01.MID05.Sample
Plate04.Lane01.MID06.Sample	ATATCGCGAG	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID06.Sample
Plate04.Lane01.MID07.Sample	CGTGTCTCTA	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID07.Sample
Plate04.Lane01.MID09.Sample	TAGTATCAGC	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	Positive	KlebsiellaSp	Plate04.Lane01.MID09.Sample
Plate04.Lane01.MID10.Sample	TCTCTATGCG	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID10.Sample
Plate04.Lane01.MID11.Sample	TGATACGTCT	GCCAGCMGCNGCGGTA	CRRCACGAGCTGACGAC	PleuralInfection	MIST2	Positive	StreptococcusMilleri	Plate04.Lane01.MID11.Sample

etc…

19.3 Example mapping file for reverse sequencing

(Lane01ReverseMapping.txt)

#SampleID	BarcodeSequence	LinkerPrimerSequence	ReversePrimer	ClassificationA	ClassificationB	CulturePos	Culture	Description
Plate04.Lane01.MID01.Sample	ACGAGTGCGT	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	Positive	StreptococcusMilleri	Plate04.Lane01.MID01.Sample
Plate04.Lane01.MID02.Sample	ACGCTCGACA	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID02.Sample
Plate04.Lane01.MID03.Sample	AGACGCACTC	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	Positive	MRSA	Plate04.Lane01.MID03.Sample
Plate04.Lane01.MID04.Sample	AGCACTGTAG	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	Positive	AnaerobicStreptococci	Plate04.Lane01.MID04.Sample
Plate04.Lane01.MID05.Sample	ATCAGACACG	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	Positive	Ecoli	Plate04.Lane01.MID05.Sample
Plate04.Lane01.MID06.Sample	ATATCGCGAG	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID06.Sample
Plate04.Lane01.MID07.Sample	CGTGTCTCTA	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID07.Sample
Plate04.Lane01.MID09.Sample	TAGTATCAGC	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	Positive	KlebsiellaSp	Plate04.Lane01.MID09.Sample
Plate04.Lane01.MID10.Sample	TCTCTATGCG	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	NG	NG	Plate04.Lane01.MID10.Sample
Plate04.Lane01.MID11.Sample	TGATACGTCT	CRRCACGAGCTGACGAC	GCCAGCMGCNGCGGTA	PleuralInfection	MIST2	Positive	StreptococcusMilleri	Plate04.Lane01.MID11.Sample

etc…

20 Appendix – Inclusion and exclusion criteria for PIPAP study

20.1 Inclusion criteria for participants with pneumonia

Clinically confident diagnosis of pneumonia ± overlying pleural effusion/infection based on all of a, b and c.

a) Acute illness

AND

b) One of;
    1. Raised inflammatory markers (white cell count (>11.0 x 10^9/l) or C-reactive protein (>50 mg/l)
    2. Fever ≥37.5°C (or hypothermia <36.0°C)
    3. New febrile symptoms – rigors or sweats
    4. New/worsening dyspnoea
    5. New/worsening cough (with or without sputum expectoration)
    6. New chest pain

AND

c) Radiographic consolidation/infiltrate consistent with pneumonia (not previously noted) ± pleural effusion

20.2 Exclusion criteria for participants with pneumonia

a) Evidence of hypercapnia ($PaCO_2$ >6 kPa)
b) Mechanical ventilation
c) Life expectancy < 2 months
d) Known bronchiectasis

e) Known tuberculosis

f) Age < 18 years

g) Severe immunosuppression – HIV positive, recent/current (< 30 days) chemotherapy, prednisolone >35mg/day for >30 days

h) Participants who are pregnant or lactating

i) Irreversible bleeding diathesis or platelet count <50 x 10^9/l

j) Irreversible visual impairment

k) Bullous lung disease that interferes with safe path for TNA

l) Contralateral severe lung disease such that pneumothorax may be poorly tolerated

m) Contralateral pneumonectomy such that pneumothorax may be poorly tolerated

n) Inability to give informed consent or comply with the protocol

21 Appendix – PIPAP study representative consent form

Form C

MRC Clinical Trials Unit

REC 09/H0605/12
Patient's sticker:
Patient's Trial No.: P

Oxford Radcliffe Hospitals NHS
NHS Trust
RESPIRATORY TRIALS UNIT
Oxford Centre for Respiratory Medicine
Churchill Hospital
Old Road
Headington
Oxford
OX3 7LJ
Tel: 01865 225205 Fax: 01865 857109

Pathogen identification in pneumonia and pleural infection study
(*A study evaluating efficacy and safety of transthoracic needle aspiration, used in conjunction with nucleic acid amplification techniques, to improve rates of timely pathogen identification and antibiotic selection in pneumonia and pleural infection*)

PATIENT INFORMATION SHEET – Participants with Pneumonia

Key Points
- This is a study trying to improve how we diagnose the bacteria ('bugs') causing pneumonia.
- Identifying the bacteria may help us to choose appropriate antibiotics. These antibiotics may have a lower risk of side effects than other antibiotics.
- Participants will have an exceptionally fine needle passed into the infected lung using local anaesthetic to numb the skin. This is standard practice in some countries, but not currently in the U.K.
- A small amount of infected lung fluid will be drawn into a syringe and analysed in the laboratory. The laboratory tests will include standard tests and some additional tests which look at the bacteria's genetic material ('DNA'), to help us identify the bacteria.
- These laboratory tests may allow us to identify bacteria causing the pneumonia, and may help us with antibiotic choices.
- The above points are only a brief outline – please do read the remainder of the information sheet.

PART 1

1. Invitation

You are being invited to take part in a research study. Before you decide whether to take part or not, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please feel free to ask us if any questions or if you would like more information.

PART 1 tells you the purpose of this study and what will happen to you if you take part.

PART 2 gives you more detailed information about the conduct of the study.

2. What is the purpose of the study?

Pneumonia is a potentially serious infection of the lung. It is commonly treated by antibiotics, which kill many bacteria (broad-spectrum antibiotics). These antibiotics are normally chosen *without knowing which bacteria are causing the pneumonia*.

Doctors perform many tests to find a cause for pneumonia, including tests on sputum, blood and urine samples. These tests do not often give results that allow us to choose or change antibiotics.

Without knowing which bacteria are causing pneumonia, doctors may choose antibiotics that are either wrong or too broad-ranging. The use of broad-ranging antibiotics has potential problems – they increase resistance rates (the ability of bacteria to survive despite treatment with an antibiotic) and side effects. This is the main factor increasing rates of infection from MRSA (methicillin-resistant *Staphylococcus aureus*) and C.diff (*Clostridium difficile*) diarrhoea. As well as encouraging these serious infections, broad-ranging antibiotics can cause other unpleasant types of diarrhoea.

Form C

This study aims to see if a new technique can help us accurately identify the bacteria causing pneumonia. Potentially, pneumonia could then be targeted with specific antibiotics, which are narrow-spectrum (i.e. active against only a few types of bacteria). This may then increase the efficiency of therapy and reduce rates of resistance development and possibly reduce side effects.

We will use a technique called 'transthoracic needle aspiration' (TNA) to sample lung fluid in pneumonia, to try to make a fast and accurate diagnosis of the cause of pneumonia. TNA is usually a straightforward procedure. Essentially, an exceptionally fine needle is passed into the infected lung – this needle is smaller than a blood-taking needle. This technique has been performed many times world-wide with a very low complication rate; it has the great advantage of taking a direct sample of the infected lung fluid. The lung fluid will be analysed using standard microbiology techniques and new techniques (called 'nucleic acid amplification techniques' or 'polymerase chain reaction') to try to find a cause for your pneumonia, hopefully enabling careful antibiotic selection.

TNA (sampling lung fluid) is not routinely performed in the UK at present. We would like to make sure that the procedure is effective and safe to patients in the UK. We think it will be, but assessing this is the purpose of this study.

This study will be led from the Oxford Centre for Respiratory Medicine, Oxford Radcliffe Hospitals NHS Trust. In total, approximately 100 patients will participate in the study – approximately 50 patients with pneumonia (like you), and approximately 50 'control' patients (who provide samples to ensure that the 'polymerase chain reaction' is accurate).

3. Why have I been chosen?

You have been chosen to consider taking part because you have pneumonia that requires treatment with antibiotics. Some patients in this study will also have infected fluid around the lung in addition to pneumonia.

The doctors treating you think that the TNA may help us to diagnose the bacteria causing your pneumonia and may enable us to tailor your antibiotic treatment. Your doctors also think that a TNA is unlikely to cause any significant problems to you.

4. Do I have to take part?

No. It is up to you to decide whether or not to take part. This information sheet is to help you make this decision. If you do decide to take part you will be asked to sign a consent form to confirm that you understand what is involved. If you decide to take part, you are still free to withdraw at any time without giving a reason. If you withdraw, we will still keep records relating to the treatment given to you, as this is valuable to the study. A decision to withdraw at any time, or a decision not to take part, will not affect your future medical care outside the trial.

5. What will happen to me if I take part?

You would have the usual investigations undertaken for your pneumonia, including blood, sputum and urine tests. We will also take an additional 10 ml (two teaspoons) of blood taken for testing. In addition, you would have the additional test – a TNA (transthoracic needle aspirate).

TNA is a common procedure for diagnosing lung conditions. TNA has been performed for pneumonia in other countries for many years, successfully. However, TNA is not currently performed in the UK routinely for pneumonia – this is the reason for this study.

TNA is performed using an ultrasound machine (or occasionally a CT machine, if you are having a CT scan anyway) to make sure that we target the infected part of the lung. The doctor performing the procedure will inject a small amount of local anaesthetic under the skin using a very fine needle – this is initially uncomfortable, but the area then goes numb within a few seconds. You should then feel minimal pain during the procedure. An exceptionally fine needle (much smaller than a blood-taking needle) is passed into the lung; a small amount of fluid is taken from the lung and the needle is quickly removed. The whole procedure is over within a few minutes. Participants will have a chest X-ray after the procedure. Participants will also have a swab of their nasal passages, looking for bacteria and viruses.

The rest of your hospital stay will be largely unaffected by this study – you should not stay in hospital any longer because of the TNA. Participants will be asked some routine questions about their health and about how they found the TNA. Participants will also have a chest X-ray the day after the procedure.

Hopefully the TNA will give us useful information about your pneumonia. We will apply any results so that you get the benefit. It is hoped that we may be able to change your antibiotics to a more suitable choice than you would have otherwise received.

As with our standard practice, you will be followed-up after you are discharged from hospital, to ensure that you are recovering well. We will ask you some more information about how you are feeling at this point and perform a follow-up blood test and chest X-ray – this is standard practice.

Our study period ends when we see you in clinic, about one month after discharge. If we feel that continued follow-up is required (but not as part of the study), this will be arranged.

TNA samples will be analysed in the microbiology laboratory to try to detect the bacteria causing your pneumonia using standard techniques. Parts of the samples will also be analysed by 'polymerase chain reaction' (PCR, a type of 'nucleic acid amplification technology'); a rapid technique being developed to allow us to diagnose bacteria very quickly. In some patients, this may yield results quickly enough to allow us to alter antibiotic therapy. If we get any fluid from the space around the lung – 'pleural fluid' – this will be dealt with in the same way. Participants will also be asked if we can store samples of blood, lung fluid and pleural fluid for subsequent testing, to help us develop methods to diagnose or treat diseases. We will also store blood samples for entirely anonymous genetic tests to help us understand why some people develop certain diseases such as pneumonia.

Patients who have pneumonia **and also infected fluid around the lung (pleural infection)** will also have a small piece taken of the lining of the infected chest fluid using a needle (called a biopsy), at the time of the TNA. You should not feel any more discomfort by having a biopsy, as the area is already numb (anaesthetised). This biopsy is necessary to diagnose some types of infection; furthermore, it can sometimes find causes for your fluid that are unexpected. We will look at the samples to see if we can identify any bacteria in the biopsy; we will also use a very high power microscope (electron microscope) to examine the biopsy for layers of bacteria, to help us understand how pleural infection (infection around the lung) develops. The risks of having such a biopsy are very small and little greater than only having the TNA performed. Some people may feel discomfort – if this is the case, we can give you more anaesthetic. There are rarer side-effects, including bleeding, or potentially damage to the lung, nerves or other structures – these are minimised by having the procedure done with the ultrasound machine. This risk of a significant problem is less than 1 in 100. The doctor undertaking the procedure will go through this, and answer any of your questions.

6. What do I have to do?

Apart from undergoing the TNA, there is little that you need to do. We will ask you to complete a simple questionnaire enquiring about your health status, three times during the study. We will also ask you a few simple questions about your opinion of the TNA.
After discharge, we would normally see you just once – about a month after your pneumonia – to make sure you are recovering well.

7. What is the procedure being tested?

Patients will undergo a TNA – 'transthoracic needle aspiration'. This is a procedure that is commonly performed for diagnosing lung conditions. TNA has been performed for pneumonia in other countries for many years, successfully. However, TNA is not currently performed in the UK routinely for pneumonia – this is the reason for this study.

The procedure is described in detail in section 5.

8. What are the alternatives for diagnosis?

The current routinely-performed procedures for diagnosing pneumonia include blood, sputum and urine tests. Blood tests (cultures) are only helpful 10% of the time. Sputum tests (cultures) are often unhelpful; either because the patient has no sputum or because the samples are contaminated by the throat or mouth. Urine tests are helpful only for two particular types of pneumonia (out of many).

Another test can be performed to diagnose pneumonia – this is called a bronchoscopy. This test is a camera examination of the airways and lungs, and requires a camera to be inserted through the nose or mouth into the lungs. This is a fairly invasive procedure, which normally requires heavy sedation. Sedation is potentially hazardous in those with low oxygen levels and pneumonia.

9. What are the side effects of the procedure?

TNA is a common procedure performed world-wide for a variety of respiratory conditions. Studies looking at TNA performed for pneumonia show that it infrequently causes any problems requiring treatment.

A small proportion of patients will have a small air leak – air leaking from the lung into the space around the lung. The medical name for this is pneumothorax. This usually causes no problems and resolves itself without any specific treatment. We shall routinely look for this side effect by carrying out a chest X-ray one hour after the procedure, and the day following the procedure. Even if a moderate amount of air leaks out, this can be treated by removing such air with a temporary tube.

Some patients cough up some blood; this almost never requires any specific treatment. Pneumonia itself can cause patients to cough up blood.

Some patients may have a bit of pain following the procedure – this should be easily managed with pain killers. We shall be routinely evaluating whether you have had any pain.

Most doctors do not use X-ray imaging techniques to help guide the TNA procedure. We are planning to use these to potentially decrease the (already small) rate of side effects.

10. What are the possible disadvantages and risks of taking part?

Patients will have a TNA – there are both potential disadvantages and advantages of this technique. We anticipate that the vast majority of patients will have no problems at all with the TNA. A small proportion of patients may suffer a side effect (listed above) – most of these resolve without any specific treatment.

11. What are the possible benefits of taking part?

It is hoped that the TNA will improve how we diagnose your pneumonia. It may then help us tailor your antibiotic treatment. It may be that the lung aspirate allows us to diagnose causes of pneumonia that we would not have previously expected. Furthermore, it is possible that the tailored treatment may have fewer side effects than the standard treatment. We cannot guarantee that this additional procedure will help, since answering these questions is an aim of the trial.

When the trial is completed, the information that it gives us will hopefully allow us to treat patients with pneumonia more effectively in the future.

12. What happens when the research study stops and what happens to the results of the study?

At the end of the study the results will be made available to all healthcare professionals through publication in scientific journals, and presentations at conferences. In all instances, the data will be anonymous and none of the patients involved in the study will be identified in any report or publication.

If you wish, we would be pleased to write to you personally to explain the results of the study. If you would like us to write to you in this way, please ensure this is recorded on your consent form and let the trial team doctors know.

13. What if there is a problem?

If there are any problems, or you have any complaints during the course of this trial, your hospital doctors will do whatever needs to be done to help you. If you remain unhappy, and wish to complain formally, you can do this through the Respiratory Trials Unit team directly or contact the University of Oxford Clinical Trials and Research Governance office on 01865 743005.

Compensation for harm arising from an accidental injury occurring as a consequence of your participation in the study will be covered by the University of Oxford. If you are harmed and this is due to someone's negligence then you may have grounds for legal action for compensation against the University of Oxford (in respect of any harm arising out of the participation in the clinical study).

14. Will my taking part in this study be kept confidential?

Yes. All the information about your participation in this study will be kept confidential. The details are included in Part 2.

15. Contact details

If you require any further information or have any concerns while taking part in the study please contact us;
Oxford Respiratory Trials Unit office (01865 225205). Your trial doctor is Dr Stephen Chapman.
Oxford Centre for Respiratory Medicine, Churchill Hospital, Oxford, OX3 7LJ.

This completes Part 1 of the Information Sheet.

If the information in Part 1 has interested you and you are considering participation, please continue to read the additional information in Part 2 before making any decision.

Form C

PART 2

16. What if new information becomes available?

The team running this study will continue to review all new research data. If any new information that influences the study becomes available, alterations will be made accordingly to the study.

17. What will happen if I don't want to carry on with the study?

If you decide you do not want to continue in the study your future medical care will not be affected. Routine follow up for your condition will continue with the medical team as would be standard practice. We would seek your permission to keep and use information and samples collected up to your withdrawal, but you can ask us to destroy all your identifiable samples.

18. Will my taking part in this study be kept confidential?

If you consent to take part in this study, your medical records may be inspected by the research trial team for purposes of collecting information and analysing the results. Authorised personnel from the sponsor (University of Oxford) and Oxford Radcliffe Hospitals NHS Trust may also look at the study data and your medical records; this is to ensure that the study is carried out to the highest possible scientific standards. All will have a duty of confidentiality to you as a research participant.

All information, which is collected about you during the course of the research, will be kept strictly confidential; it will be held securely on paper and electronically under the provisions of the 1998 Data Protection Act. The information will be stored with a code number (a 'trial number') and not with your name and address. This is to protect your confidentiality.

In line with Good Clinical Practice guidelines, at the end of the study, your data will be securely archived for a minimum of 5 years.

If you withdraw consent from further study treatment, your data and samples will remain on file and will be included in the final study analysis. Identifiable samples will be destroyed, at your request.

19. Informing your General Practitioner and other clinicians

We will ask your permission to inform your GP and other medical practitioners involved in your care that you are participating in this study.

20. What will happen to any samples I give?

Lung fluid samples (and pleural fluid, pleural biopsies and blood samples, where appropriate) will be processed in hospital laboratories, as previously described. In addition, samples will also be stored anonymously in locked freezers within the Oxford Respiratory Trials Unit; such samples would contain your genetic material ('DNA'). All samples will be coded so they cannot be identified as yours, other than by the specified investigators. Scientists working with the samples will only know their code number, and would not be able to trace the sample to you. Your samples may be stored indefinitely and are regarded as a 'gift' to the research team.

Samples will be used for the purposes indicated in this information sheet, and may also be used to help us develop methods to diagnose or treat diseases. All future research on samples would be 'ethically-approved' by an independent Research Ethics Committee, including any research on stored DNA (your genetic material). Future research may include working with commercial organisations, or tests in laboratories outside Oxford and the European Union. All tests will still be under the control of the Oxford Clinical Trials Unit, and results of any tests remain strictly confidential at all times.

21. Who is organising and funding the research?

The trial is run by staff within the Oxford Centre for Respiratory Medicine. The infrastructure is being funded by the Oxford Biomedical Research Centre, using a grant from the Department of Health's National Institute for Health Research (NIHR) under the programme 'Best Research for Best Health'.

22. Who has reviewed the study?

This study was given a favourable ethical opinion for conduct in the NHS by the Oxford Research Ethics Committee.

23. Contact for further information

You are encouraged to ask any questions you wish, before, during or after your participation in the study. If you have any questions about the study, please speak to your study doctor, who will be able to provide you with up to date information about the

procedure involved. If you wish to see a more detailed background to the study, we are happy to provide a copy of the full study protocol.

If you require any further information or have any concerns while taking part in the study please contact us;
Oxford Respiratory Trials Unit office (01865 225205).
Oxford Centre for Respiratory Medicine, Churchill Hospital, Oxford, OX3 7LJ.

24. Consent

If you decide you would like to take part then please read and sign the consent form. You will be given a copy of this information sheet and the consent form to keep. A copy of the consent form will be filed in your patient notes and one will be filed with the study records.

Thank you very much for considering taking part in this study.

Form C

MRC Clinical Trials Unit

Oxford Radcliffe Hospitals NHS
NHS Trust

REC 09/H0605/12
Patient's sticker:
Patient's Trial No.: P

Pathogen identification in pneumonia and pleural infection study

(*A study evaluating efficacy and safety of transthoracic needle aspiration, used in conjunction with nucleic acid amplification techniques, to improve rates of timely pathogen identification and antibiotic selection in pneumonia and pleural infection*)

PATIENT CONSENT FORM – Participants with Pneumonia

(Please initial box to agree)

1. I confirm that I have read and understand the information sheet dated 03/02/11 (version 1.2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my medical care or legal rights being affected. ☐
3. If I am found to have infection in the fluid around the lung (pleural infection) at the time of my procedure (TNA), I confirm that I am willing to have a biopsy of the lining of this fluid during the procedure. ☐
4. I agree that samples of my blood, nasal swabs, sputum, lung aspirates, pleural fluid and pleural biopsies taken for the study may be stored indefinitely and used in other research in the future, which may include analysis outside Oxford and the European Union and may involve commercial companies. This would always be done in a completely anonymous manner. I agree to these being considered a gift to the trial team. ☐
5. I agree that samples of my genetic material ('DNA') will be stored as above, and may be used anonymously in future ethically-approved genetic studies. ☐
6. I am happy for my GP and other medical practitioners involved in my care to be informed I am participating in this study. ☐
7. If you would like to know the results of this study, please initial the box and we will write to you when the study is completed. ☐
8. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from Oxford University, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐
9. I agree to take part in the above study. ☐

All handling, storage and destruction of data is in accordance with the Data Protection Act 1998

_____	_____	_____
Name of Patient	Date	Signature
_____	_____	_____
Name of Person taking consent	Date	Signature

1 copy for participant, 1 copy for researcher's site file, original to be kept in medical notes